
Dynamic aspects of DNA

DNA-slippage and nucleosome dynamics

Richard Neher



München 2007

Dynamic aspects of DNA

DNA-slippage and nucleosome dynamics

Richard Neher

Dissertation
an der Fakultät für Physik
der Ludwig-Maximilians-Universität
München

vorgelegt von
Richard Neher
aus Göttingen

München, den 17. April 2007

Erstgutachter: Prof. Dr. Erwin Frey
Zweitgutachter: Prof. Dr. Ulrich Gerland
Tag der mündlichen Prüfung: 16.05.2007

Contents

Zusammenfassung	ix
Abstract	xi
1. Properties of DNA	1
1.1. The structure of DNA	2
1.2. Thermodynamics of DNA	3
1.2.1. Binding free energies of double stranded DNA	4
1.2.2. Denaturation of long DNA molecules	5
1.2.3. DNA melting of homogenous sequences	8
1.3. Mechanical properties of DNA	10
1.3.1. Single molecule force spectroscopy	11
1.3.2. Stretching double stranded DNA	13
1.3.3. Stretching single stranded DNA	15
1.3.4. DNA unzipping	15
2. Dynamics of repetitive DNA	17
2.1. The biological role of repetitive sequences	18
2.1.1. The number of repeats changes rapidly in evolution	18
2.1.2. SSRs are versatile genetic markers	22
2.1.3. SSR expansion is related to hereditary diseases	22
2.1.4. SSRs in prokaryotes	23
2.1.5. <i>In vitro</i> evidence for local strand slippage	24
2.2. Force induced DNA-slippage	25
2.2.1. Sliding dynamics of perfectly repetitive sequences	26
2.3. Single molecule experiments on DNA-slippage	29
2.4. DNA sliding in presence of sequence disorder	30
2.5. Repetitive DNA as a visco-elastic nanoelement	31
2.6. Intermediate phase in DNA-melting	32
2.7. Conclusion & Outlook	32
2.8. R.A. Neher and U. Gerland, <i>Phys. Rev. Lett.</i> , 93 , 198102	35
2.9. F. Kuehner <i>et al.</i> , <i>Biophys. J.</i> , 92 , 2491, 2007	39
2.10. R.A. Neher and U. Gerland, <i>Biophys. J.</i> , 89 , p. 3846	47
2.10.1. Supplementary Material to <i>Biophys. J.</i> , 89 , p. 3846.	57
2.11. R.A. Neher and U. Gerland, <i>Phys. Rev. E</i> , 73 , 030902(R)	63

3. Dynamics of nucleosomal DNA	67
3.1. DNA compactification	67
3.1.1. The nucleosome core particle	69
3.2. Gene regulation in eukaryotes	71
3.3. Experiments on single nucleosome dynamics	72
3.4. Kinetic accessibility of protein binding sites in nucleosomal DNA	74
3.4.1. Kinetics of site exposure	75
3.5. Flexibility assisted conformational transitions	77
3.6. Conclusion & Outlook	78
3.7. W. Möbius, R.A. Neher and U. Gerland, <i>Phys. Rev. Lett.</i> , 97 , 20102	79
3.8. R.A. Neher <i>et al.</i> , submitted, 2007	83
A. Partition sums of repetitive DNA	87
A.1. Thermal denaturation of repetitive DNA	89
A.2. Pulling on repetitive DNA	90
A.2.1. Determining loop densities.	91
B. Generalized Kramers-Langer rates	93
B.1. The flux over the barrier	95
B.2. Stochastic dynamics of constrained systems	97
B.2.1. Langevin equations for a multisegment chain	100
Bibliography	101
Glossary	110
Danksagung	113

List of Figures

1.1. Nucleotides, the elementary subunits of DNA.	3
1.2. Structure of dsDNA	4
1.3. DNA melting curve.	6
1.4. Contour integration in the fugacity plane.	10
1.5. Sketch of an optical trap and an AFM.	12
2.1. Repetitive DNA has many binding configurations.	18
2.2. Repeat number can increase or decrease during replication.	20
2.3. Evidence for fast chain sliding reaction.	26
2.4. Shearing repetitive DNA.	27
2.5. Microscopic dynamics of DNA sliding.	28
2.6. Supp.: The lifetime distribution of two random walkers.	58
2.7. Supp.: Measuring rates of mutation opening and closing.	60
2.8. Supp.: Mutation opening is a trade-off between entropy and energy.	61
3.1. Hierarchical organization of chromatin.	68
3.2. Crystal structure of Nucleosomes	70
3.3. Site exposure mechanism	72
A.1. Calculating the partition sum.	88
A.2. Calculating loop densities.	92
B.1. The effect of noise-induced drift on the stochastic separatrix.	95

Zusammenfassung

DNA ist keine steife und festgefügte Einheit, sondern ändert fortlaufend ihre Konformation. Die Dynamik von DNA auf unterschiedlichen Längenskalen ist Thema dieser Dissertation. Der erste Teil der Arbeit befasst sich mit der Dynamik der Basenpaarung zwischen zwei DNA Strängen, deren Sequenz eine mehrfache Wiederholung eines kurzen Motivs ist. Im zweiten Teil der Arbeit wird die Dynamik von Chromatin, d.h. DNA, die mit Hilfe von Proteinen in Chromosome gepackt ist, diskutiert.

Vielfache Wiederholungen eines kurzen Motivs von ein bis sechs Basen sind sehr häufig in eukaryotischen Genomen und haben die erstaunliche Eigenschaft, dass sich die Zahl der wiederholten Einheiten ausserordentlich schnell von Generation zu Generation ändert. Die Rate solcher Mutationen übersteigt die von Punktmutationen um mehrere Größenordnungen. Diese Hypervariabilität repetitiver Sequenzen hat eine Reihe von biologischen Konsequenzen und ist unter anderem für einige menschliche Erbkrankheiten verantwortlich. Repetitive DNA mutiert um so vieles schneller als gewöhnliche DNA, da die beiden Stränge gegeneinander versetzt binden können und dadurch Fehler bei der DNA Replikation auftreten. Dieses versetzte Binden heisst *DNA-slippage*. Wir haben die Dynamik von DNA-slippage theoretisch untersucht und Experimente vorgeschlagen, mit denen DNA-slippage in einzelnen Molekülen detektiert werden kann. Zwei kurze repetitive DNA Stränge können sich durch Propagation von Defekten gegeneinander bewegen und daher durch eine Scherkraft aneinander entlang gezogen werden. Die Defekte werden durch DNA-slippage an den Enden des Doppelstrangs erzeugt. Die Rate, mit der Defekte produziert werden und damit die Geschwindigkeit, mit der die Stränge sich gegeneinander bewegen, hängt sehr empfindlich von der angelegten Kraft ab. Unsere theoretische Analyse hat gezeigt, dass es vier dynamische Regime gibt, in denen die typischen Abrisszeiten unterschiedlich mit Länge des Moleküls anwachsen. Ferdinand Kühner und Julia Morfill aus dem Labor von Prof. H.E. Gaub haben kürzlich mit Hilfe eines Kraftmikroskops (AFM) experimentell gezeigt, dass DNA-slippage tatsächlich durch Scherkräfte ausgelöst werden kann [1]. Über die biologische Relevanz hinaus könnte repetitive DNA auch Anwendungen in der Nanotechnologie finden, denn sie verhält sich wie ein kontraktiles visko-elastisches Element. Die Kenngrößen eines solchen Elements können durch Wahl der Sequenz und der Länge der Stränge programmiert werden. Durch einzelne Punktmutationen, die die Periodizität der Sequenz unterbrechen, kann die mechanische Antwort des Systems gezielt verzögert werden.

Der zweite Teil der Dissertation behandelt die Dynamik des elementaren Baustein von Chromatin, dem Nukleosom. Damit das Genom eukaryotischer Zellen in den Zellkern passt, ist die DNA dicht gepackt. Trotzdem muss die in der DNA gespeicherte Information für die Zelle zugänglich sein. Daher ist die Frage, wie oft und wie lang ein bestimmter Teil

der DNA sich von dem Nukleosom löst von großer biologischer Relevanz. Ein Nukleosom besteht aus einem zylindrischen Proteinkomplex mit einem Durchmesser von ca. 6 nm, um den die DNA in etwa 1.7 mal gewickelt ist. Es wurde kürzlich experimentell gezeigt [2, 3], dass sich Teile der DNA eines Nukleosoms auf eine Skala von Millisekunden bis Sekunden vom Proteinzylinder lösen. Diese Dynamik könnte Teil des Mechanismus sein, mit Hilfe dessen die Zelle Zugang zu kompaktifizierter DNA erlangt. Komplementär zu diesen Experimenten haben wir Nukleosom-Dynamik theoretisch untersucht. Unsere Studien haben gezeigt, dass die wenn auch kleine Flexibilität der DNA einen außerordentlich großen Einfluss auf die Dynamik solcher DNA-Protein Komplexe hat. Der wesentliche Prozess des Auf- und Abwickelns ist thermisch aktiviertes Überqueren einer Potentialbarriere, in dessen Verlauf sich die DNA reorientiert. Die reichhaltige Phänomenologie und die Allgegenwärtigkeit solcher Prozesse hat uns motiviert thermisch aktiviertes Überqueren einer Potentialbarriere gekoppelt an die Rotation eines flexiblen Arms genauer zu studieren. Die Rate für das Überqueren der Barriere wird maximal bei einer intermediären Steifigkeit. Solche optimalen Parameter könnten in biologischen Makromolekülen wie z.B. molekularen Motoren realisiert sein.

Im ersten Kapitel dieser Arbeit werden die chemische Zusammensetzung von DNA, ihre Struktur, sowie ihre thermodynamischen und mechanischen Eigenschaften diskutiert. Das zweite Kapitel befasst sich mit der Dynamik repetitiver DNA Sequenzen. Zu Beginn wird die biologische Rolle repetitiver DNA und ihre Verbindung zu menschlichen Erbkrankheiten vorgestellt. Dann diskutiere ich die Grundzüge unserer theoretischen Arbeit sowie die erste experimentelle Bestätigung von DNA-slippage, gefolgt von unseren Publikationen zu diesem Themenkomplex. Das dritte Kapitel befasst sich mit der Dynamik von Chromatin. Nach der Struktur von Chromatin werden die Experimente zur Nukleosom-Dynamik diskutiert und im Anschluss unsere theoretische Arbeit und unsere Publikationen vorgestellt.

Abstract

DNA is not a rigid entity, but a highly dynamic molecule. The dynamics of DNA on different length scales is the objective of this thesis. The first part of this thesis addresses the dynamics of the base pairing patterns of DNA, the sequence of which is a multi-fold repetition of a short motif. In the second part of this thesis, we discuss the dynamics of chromatin.

Repetitions of short motifs of one to six bases are very common in eukaryotic genomes and the number of repeated units changes extraordinarily fast from generation to generation. The rate of such contractions or deletions is orders of magnitudes larger than the rate of ordinary point mutations. This hyper-variability of repetitive DNA has a number of implications in biology and is the cause of certain human hereditary diseases. The reason why repetitive DNA mutates so rapidly is related to the fact, that two complementary strands with repetitive sequence can bind to each other, even when shifted relative to each other. Locally shifted binding is called *DNA-slippage* and leads to errors during DNA replication. We studied the dynamics of DNA-slippage theoretically and suggest experiments that probe DNA-slippage in single DNA molecules. The propagation of small bulge loops in the double helix of repetitive DNA allows the two strands to move relative to each other. Application of a shear force to repetitive DNA should therefore induce a strand motion. The bulge loops are produced by DNA slippage at the ends of the double strand. We show, that the bulge loop production rate and hence the relative velocity of the two strands depends sensitively on the applied shear force. We uncover four dynamical regimes, where the rupture times scale differently with the system size. Ferdinand Kühner and Julia Morfill from the lab of Prof. H.E. Gaub succeeded in measuring force induced DNA-slippage in single molecules using an atomic force microscope [1]. In addition to its biological relevance, repetitive DNA has intriguing mechanical properties that might find applications in nanotechnology. Repetitive DNA acts as a contractile visco-elastic element, the characteristics of which can be programmed by its length and sequence composition. Rare point mutations that interrupt the repetitive sequence allow to delay the response in a controlled manner.

The second part of this thesis addresses the dynamics of nucleosomes, which are the elementary packing units of chromatin. Eukaryotic cells compactify their genome to make it fit into the cell's nucleus. Nevertheless, the cell has to access the information in the DNA. Since most proteins cannot bind to DNA buried in nucleosomes, the question how often and how rapidly a particular stretch of DNA detaches from the nucleosome is of great biological relevance. A nucleosome consists of a cylindrical protein core with 6 nm in diameter. The DNA is wrapped around this protein cylinder approximately 1.7 times. Recent experiments measured the rates, at which the DNA detaches and attaches partly

from the protein core [2, 3]. We studied the dynamics of DNA wrapping and unwrapping in single nucleosomes theoretically. We show, that the small but finite flexibility of the DNA drastically enhances the rates of the wrapping and unwrapping kinetics. The rich phenomenology and the ubiquity of similar processes in biology motivated us to study transition that involve the rotation of flexible lever-like object in more detail. The transition rate displays an optimum at an intermediate stiffness. The optimal stiffness parameters could be realized by evolution in biological macromolecules such as molecular motors.

In the first chapter of this thesis, I present general features of DNA such as its chemical composition, its structure, its thermodynamics and its mechanical properties. The second chapter is on the dynamics of repetitive DNA. First, we discuss the biological significance of repetitive DNA and existing experimental evidence for DNA-slippage. Then we present our theoretical analysis and our publications on repetitive DNA. The third chapter is on chromatin dynamics. To begin with, chromatin structure and its implication for gene regulation in eukaryotes are discussed. This is followed by a discussion of recent experiments on single nucleosome dynamics, our theoretical work, and our publications.

1. Properties of DNA

Every known biological cell uses *deoxyribonucleic acid*, in short DNA, as carrier of its hereditary information. DNA is a double stranded heteropolymer into which information can be coded by the sequence of its four elementary subunits known as bases or nucleotides. The information encoded in DNA has three primary functions. First of all, DNA codes for proteins. The structure of a protein is determined by its amino acid sequence, which is coded as a nucleotide sequence in DNA. Most organisms known use 20 different amino acids to build their proteins. Since there are only four different bases, a multi-letter code has to be used to describe a sequence of amino acids by a DNA sequence. Biology uses three successive bases, called *codons*, to code for one amino acid. To produce a protein, the double stranded DNA is locally opened and the nucleotide sequence is transcribed by a protein complex called *polymerase* into messenger RNA (mRNA). The mRNA is then translated into a sequence of amino acids, which folds into the functional protein. In addition to being the storage medium for protein sequences, DNA has a pivotal role in cellular information processing. At every instant in time, a cell has to determine how much of each gene is to be expressed. To accomplish this feat, DNA contains regulatory regions to which specialized proteins, so called transcription factors (TF), can bind. These proteins either preclude or enhance the assembly of the transcription machinery and thereby regulate the expression of genes. Different signals associated with different TFs can be logically combined by arranging their binding sites on the DNA such that the TFs bind cooperatively or exclude each other [4, 5]. Yet a different class of DNA regions codes for RNA sequences that are not translated into proteins but have important functions themselves. RNA can form complicated secondary and tertiary structures, which make certain RNA molecules powerful catalysts. For example the ribosome, the cellular machine that translates the RNA into proteins, is a complex of folded RNA molecules and proteins. Its catalytic activity is performed by RNA parts and the proteins merely stabilize the RNA complex. Another important example are transfer RNAs (tRNA) that decipher the genetic code into amino acids. In addition to these three roles, many other and to date unknown functions of DNA might exist. Indeed, only a fraction of the genome of higher organism can be linked to any function, whereas the role of the largest part, often referred to as junk DNA, is largely unknown [6]. For a comprehensive and fairly up-to-date source of information on molecular biology, I refer the reader to the classic textbook by Alberts et al.

The objectives of this thesis are dynamical aspects of DNA. On one hand, we are going to discuss the dynamics of the base pairing pattern of two complementary DNA strands, whose sequence is the multi-fold repetition of units of one to six bases in length. Such repetitive sequences play important roles in various processes in biology and are related to a certain class of human hereditary diseases. What distinguishes them from sequences without this

specific order is the dynamics of the base pairing pattern which renders repetitive sequences hyper-variable in evolution. On the other hand, we will study the dynamics of DNA on a much larger scale. In eukaryotes DNA is highly compactified. However, in order to be of any use for the cell, the information encoded in the DNA sequence needs to be accessible for read out. In chapter 3, we will discuss physical aspects of the dynamics of compactified DNA, which has profound implication to the accessibility of genetic information in eukaryotes.

Since the dominant theme of this thesis is DNA and since its chemical and physical properties are needed throughout, we will compile the basics of DNA in this introductory chapter. We will only discuss aspects that are related or prerequisite to our work and we do not attempt to provide a comprehensive survey. The biological background and the relevance of the specific questions we address is discussed in the introductory part of the two main chapters of the thesis.

1.1. The structure of DNA

The elementary subunits of DNA are single nucleotides, which can be connected to each other to form a polymer. Each nucleotide consists of the sugar *deoxyribose*, a *phosphate group* and one of the four bases *adenine (A)*, *cytosine (C)*, *guanine (G)* and *thymine (T)*, cf. Fig. 1.1. The phosphate group is attached to the 5' carbon atom of the sugar. Nucleotides with different numbers of phosphates and slightly modified sugars play pivotal roles as energy storage and as signaling molecules in all known biological cells. Here, we focus on the role of nucleotides as building blocks of DNA. Two nucleotides can be linked to each other by formation of a phosphate bond between the phosphate group of one and the hydroxyl group at the 3' carbon atom of the other nucleotide (cf. Fig. 1.2). The formation of this bond is independent of the base that is attached to the sugar, and long chains with an arbitrary sequence of bases can be formed. The polynucleotide chain ends with a free phosphate only at one end. The phosphorylated end is commonly called the 5' end after the carbon atom, to which the phosphate is attached. Correspondingly, the other end is called the 3' end. This polarity has important implications for DNA replication, since the DNA polymerase attaches nucleotides only to 5' ends, cf. Fig. 2.2.

Single stranded DNA (ssDNA) alone is not suitable as a storage medium for hereditary information because single phosphate bonds are not stable enough and a damaged single strand is hard to repair without a backup copy. Both problems are solved elegantly by the ability of ssDNA to pair with a *complementary* strand. The geometry of the bases adenine and thymine is such that they can form two hydrogen bonds, whereas cytosine and guanine interact via three hydrogen bonds, cf. Fig. 1.2. The interaction between these bases occurs only, if they are aligned in opposite polarity. Therefore a DNA strand binds selectively to a strand, the sequence of which is the complementary sequence in opposite order. After base pairing, the double stranded DNA (dsDNA) winds into a double helix, which brings consecutive base pairs closer together and shortens the duplex from 0.7 nm per base in single strand to 0.34 nm in double strand. The double helix has a diameter of approximately 2 nm and a helical pitch of about 10.5 bp or 3.5 nm. By base pair stacking,

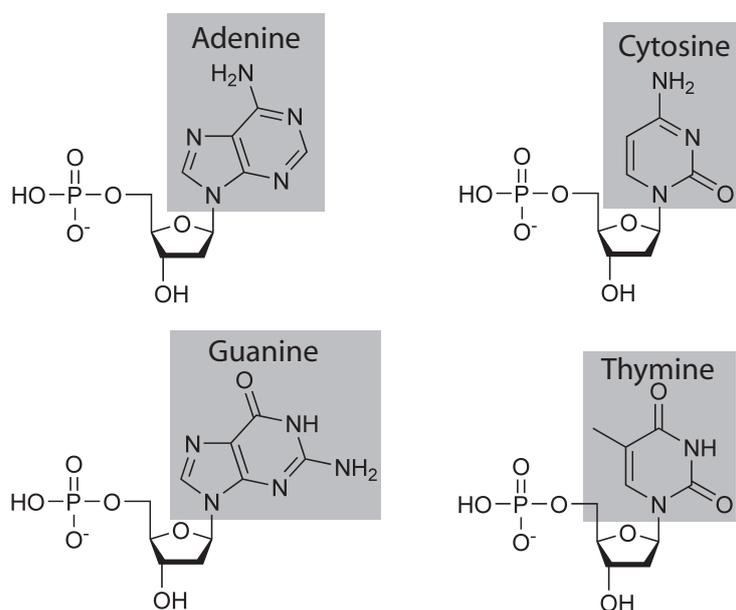


Figure 1.1: The four different nucleotides of DNA: Each nucleotide consists of a phosphate group, the sugar desoxyribose and one of the bases adenine, cytosine, guanine or thymine. Source of images: Wikipedia.

water is driven out of the space between base pairs and the carbon rings of bases align, which is the major contribution to the DNA binding free energy. The DNA double helix is not completely symmetric, meaning the two bases of a base pair do not form an angle of 180° . Thereby, dsDNA has a major and a minor groove, as illustrated in Fig. 1.2b.

In contrast to ssDNA, damage to dsDNA is easy to repair. The complementary strands can serve as a template for reconstruction of a damaged strand and for replication of the molecules. The stability of dsDNA and its potential to be repaired enable cells to maintain genomes as long as 10^{10} nucleotides, resulting in molecules of macroscopic length.

1.2. Thermodynamics of DNA

In the previous section we discussed how the chemical architecture of DNA makes it an ideal carrier of genetic information. We will now turn to thermodynamic properties of DNA which have implications for the ability of cells to read the sequence information from the DNA. Efficient information read out is possible only when the double helix is opened and the unpaired bases are exposed. Hence, the cell has to separate dsDNA locally into single strands at ambient temperature. How this can be achieved is mainly determined by binding affinity of the two strands and their fluctuation properties, which therefore have been studied extensively, both experimentally and theoretically.

The basic ingredients to DNA thermodynamics are the free energy contributions of base pairing and stacking interactions. Using these parameters, the properties of short molecules

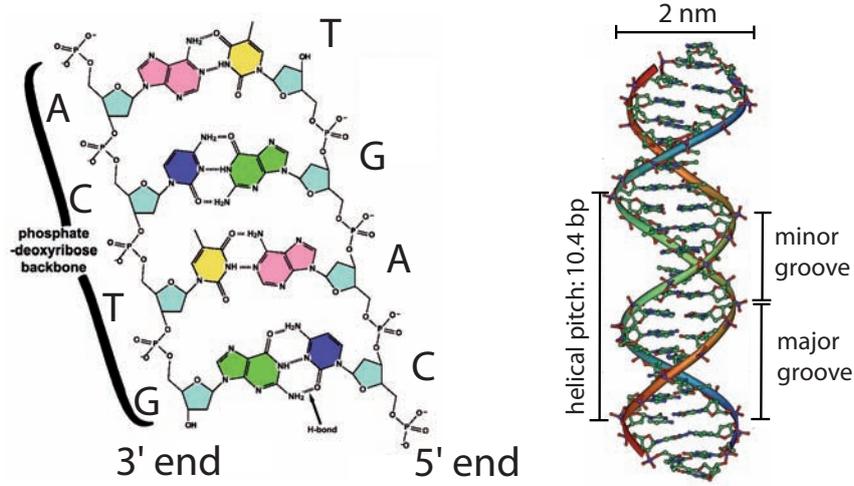


Figure 1.2: Left: Two oppositely aligned DNA strands bind specifically to each other if their sequences are complementary, i.e. if each base A faces a T and each base G faces a C. Hydrogen bonds between bases are indicated by dashed lines. Right: The ladder-like structure compactifies further by winding into a double helix. The pitch of the double helix 3.5 nm, which corresponds to ≈ 10.5 bps. Since base pairs are not perfectly straight the helix displays a major and a minor groove. Source of images: Wikipedia.

can be well understood with simple two state models. Subsequently we discuss the melting behavior of long molecules and calculate the partition sum of a dsDNA molecule within the framework of the Poland-Scheraga model.

1.2.1. Binding free energies of double stranded DNA

The dominant contributions to the binding free energy of double stranded DNA are H-bonds between bases in Watson-Crick base pairs and the stacking interaction between subsequent base pairs. The binding of the two strands goes along with a significant reduction in entropy, since two floppy single strands are forced into a much more rigid double stranded conformation. Assuming constant specific heat c_p , the free energy of a particular structure is given by

$$\Delta G = \Delta H - T\Delta S, \quad (1.1)$$

where enthalpies, entropies and free energies are measured with respect to the dissociated case. To a good approximation, ΔH and ΔS can be calculated as sums from contributions of consecutive base pairs, such as AG/CT. Additional free energy contributions stem from penalties for mismatches and loops or the lack of stacking interactions at the first and last base pair, often called initiation and termination costs.

$$\Delta G = \Delta G_{init} + \sum_{\text{basepairs}} \Delta G_{bp} + \sum_{\text{loops}} \Delta G_{loop} + \sum_{\text{mismatches}} \Delta G_{mm} + \Delta G_{term}. \quad (1.2)$$

Many of the parameters have been carefully measured and are reviewed by SantaLucia Jr. in [8] and [9]. While the precise binding energies depend on at least two consecutive base pairs, a good rule of thumb is that a **CG** base pair contributes approximately $3k_B T$ and an **AT** pair about $2k_B T$ at physiological salt concentrations. The penalty for initiating a loop, that is interrupting base pair stacking, is typically between 3 and $10k_B T$. Extrapolation formulas of the parameters to different salt concentrations are also available [9]. The complete set of parameters has been fed into software packages that predict melting temperatures and plausible secondary structures of short DNA oligonucleotides, see for example [10].

Two-state models. Short DNA oligonucleotides occur in essentially two different states. The two strands are either dissociated and float freely in solution, or the two strands are bound in their most stable binding configuration since any suboptimal base pairing is unstable. For such molecules, it is particularly easy to predict their melting temperature. The melting temperature T_m is commonly defined as the temperature where half of the single strands are part of duplexes. Setting ΔG in Eq. (1.2) to zero and accounting for the total concentration of DNA strands c_T , one finds [8]

$$T_m = \Delta H / (\Delta S + R \ln c_T), \quad (1.3)$$

where the gas constant $R = 1.9872 \frac{\text{cal}}{\text{K}\cdot\text{mol}}$. Due to significant contributions from terminal ends, the melting temperature of short molecules strongly depends on the length. At larger length (above 30 base pairs), the melting temperature mainly depends on the bulk binding energy and hence on the **CG** content of the sequence. Melting temperatures range from 20°C for very short (5 base pairs) sequences to 90°C for long **CG**-rich molecules.

1.2.2. Denaturation of long DNA molecules

The assumption that two DNA strands are either firmly bound in the most stable state, or completely dissociated is not justified for long sequences. Long sequences might have regions with different **CG**-content that melt at different temperatures. Even homogenous molecules will once in a while open their double stranded structure locally and form denatured bubbles as illustrated in Fig. 1.3. Two-state models are therefore not suitable for long molecules, but many different configurations including partly melted patches contribute significantly. The most important experimentally accessible quantity is the degree of base pairing of the DNA strands, which can be monitored by the absorption of UV light. Unpaired bases absorb UV light more efficiently than bases stacked in the double helical conformation and any change in the absorption coefficient can be directly related to the degree of base pairing $\Theta(T)$ between the two strands [11]. If the local **CG**-content is constant along the molecule, the derivative $-\frac{d\Theta(T)}{dT}$ of the fraction of bound base pairs Θ has a single peak. However, the **CG**-content of DNA varies considerably along the genome¹.

¹ There are differences in **CG**-content between coding and non-regions, as well as between incorporated viral DNA and proper DNA.

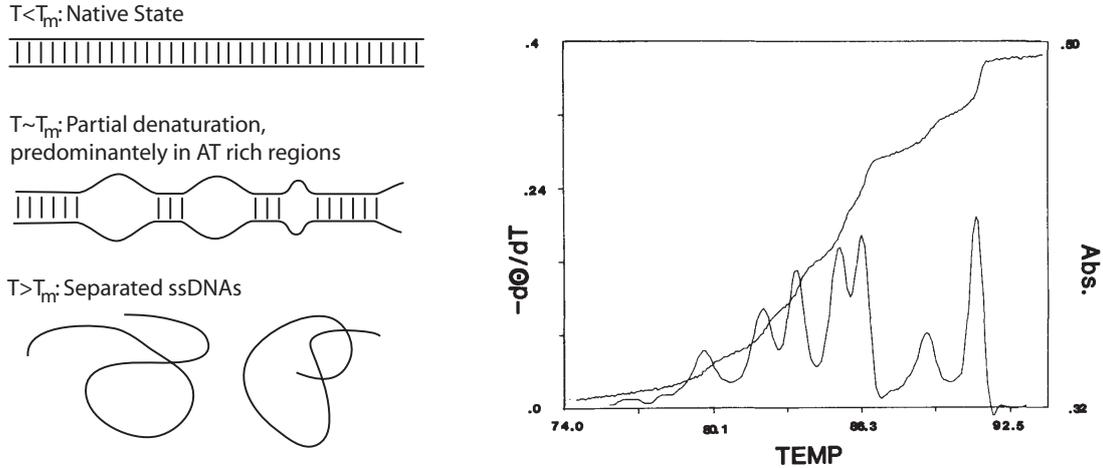


Figure 1.3: Left: With increasing temperature the double stranded structure of DNA is interrupted by denatured loops and the two strands eventually separate. AT rich regions tend to denature at lower temperatures due to their smaller binding free energy. Left: The UV absorbance and the negative differential of the fraction of base pairs vs. temperature, see main text. Reproduced from [11].

In this case, the differential melting curve has many peaks corresponding to different regions of the DNA that melt at different temperatures. A typical melting curve is shown in Fig. 1.3.

Attempts to describe the melting transition of DNA theoretically date back to the late 1950s and resulted in a class of models that are now commonly referred to as Poland-Scheraga models [12, 13, 14, 15] or Ising type models. These models describe a particular configuration of the DNA by the set of base pairs formed. In general, base pairs can be formed between any two complementary bases on different strands as well as within one strand, that folds back onto itself. The latter is particularly important for RNA, but is rarely relevant for two complementary DNA strands since a high degree of self complementarity within a single strand is unlikely. Poland-Scheraga models are usually restricted to native base pairs, *i.e.* only base pairs that are present in the ground state are allowed. The restriction to native base pairs is a good approximation, since stable base pairing requires several consecutive base pairs and the chance of finding two non-native complementary stretches that are several base pairs long is slim. A convenient way to denote a base pairing configuration of dsDNA of length N is by an ordered subset of the integers $\mathcal{S} = i_1, \dots, i_m \subset [1, \dots, N]$, which corresponds to the base pairs present in the DNA duplex. Since we are not interested in reproducing experimental data as faithfully as possible, but rather seek generic explanations to general features of DNA denaturation, it is reasonable to simplify the free energy model. In the following, we assume that stacking interactions do not depend on the base pair type and include them through a cost ε_ℓ^0 for initiating a loop. Furthermore, we assume that the loop penalty is independent of the bases in the loop and only depends on the loop size. The free energy model for a DNA

configuration then simplifies to

$$G[\mathcal{S}] = - \sum_{i \in \mathcal{S}} \varepsilon_b(i) + \sum_{\text{loops}} \varepsilon_\ell(n_l), \quad (1.4)$$

where $\varepsilon_b(i)$ is the binding energy of base i and $\varepsilon_\ell(n)$ is the free energy cost of a loop of size n . However, even with this simple free energy model the explicit summation of all configurations is infeasible, since their number increases exponentially with the length. Fortunately, there is a much more clever way to calculate the partition sum. Any allowed base pairing configuration is a sequence of double stranded stems and denatured loops. Furthermore, the free energy of a particular configuration is a sum of local contributions from base pairs and loops. These properties allow the calculation of the partition function using recursion relations. Let Z_n be the partition function of a dsDNA molecule of length n , where the first and the last base pair are formed. Z_n obeys the recursion relation

$$Z_n = e^{\frac{\varepsilon_b(n)}{k_B T}} Z_{n-1} + \sum_{m=1}^{n-2} e^{\frac{\varepsilon_b(n) - \varepsilon_\ell(m)}{k_B T}} Z_{n-m-1}, \quad (1.5)$$

where the first term includes any structure that can be obtained by adding base pair n to any configuration in Z_{n-1} and the sum includes all structures that are obtained when adding the base pair n followed by a loop of size m to any structure in Z_{n-m-1} . This recursion relation allows the calculation of the partition sum of arbitrary sequences of length N in $\mathcal{O}(N^2)$ steps. Similar recursion relations have been used to study the statistical physics of RNA strands that fold back onto themselves [16, 17].

The length dependence of the loop cost. When describing a dsDNA molecule by its set of base pairs it is implicitly assumed, that all other degrees of freedom, *e.g.* the conformation of single stranded ends or denatured loops equilibrate rapidly compared to major rearrangements in the base pairing patterns. This results in a subtle dependence of the free energy of a loop on its length. Single stranded DNA is rather flexible and changes its orientations typically every 2 to 3 base pairs, see Sec. 1.3.3. The possible conformations of ssDNA can therefore be mapped to the configurations of a random walk². The number of random walks increases as $\sim s^n$ with its length n . For ssDNA, s has to be chosen such that $\ln s$ is the decrease in entropy when a single stranded monomer is forced into dsDNA. A denatured bubble in a dsDNA can also be described by a random walk, however, subject to the constraint that the random walk forms a closed loop. This reduces the number of allowed configurations by a factor n^{-c} giving rise to a loop penalty of entropic origin of the form $c \ln n$ [18]. For ordinary random walks in d dimensions c has the value $d/2$. When self-avoidance is included, c is given by d times the Flory exponent ν . At first sight this logarithmic correction to entropy appears to be of minor importance, but it is responsible for a genuine denaturation transition in Poland-Scheraga models (see below). In particular, a discontinuous transition requires $c > 2$. Kafri et al. claim that c is indeed larger than

²Excluded volume effects of free single strand are not essential for the denaturation transition.

2 when mutual excluded volume effects of different loops are taken into account [19]. In a nutshell the argumentation is as follows: The denatured loops in a DNA molecule are not independent self-avoiding polymer loops, but are linked by the double stranded stem to form a polymer network. Excluded volume effects in a connected polymer network are stronger than for non-interacting loops, resulting in a higher value of the exponent c . Kafri et al. calculated a value of $c = 2.15$ for DNA denaturation, resulting in a discontinuous transition [20]. However, the applicability of the scaling theory of polymer networks to DNA has been questioned [21, 22]. The objection is, that denatured loops are rare and far apart such that their interaction should be negligible. The rigid double stranded stems are essentially one dimensional objects, which are irrelevant for scaling. In any case, corrections to the loop exponent will only become important when studying DNA melting using extremely long molecules with very homogenous sequences. For now, we treat c as a variable and use a loop initiation cost of the form

$$\varepsilon_\ell(n) = \varepsilon_\ell^0 - n \ln s + c \ln n, \quad (1.6)$$

where ε_ℓ^0 is a constant loop initiation cost due to the loss of base pair stacking when a loop is formed.

1.2.3. DNA melting of homogenous sequences

While the recursion relations are indispensable when studying the thermodynamics of a particular sequence, they do not provide insight into the universal properties of DNA melting. To this end, we now demonstrate how the partition sum can be calculated in closed form if the binding energy per base is the same for every base. This might appear to be a very restrictive and unrealistic assumption. However, we can coarse grain our description even further and lump a small number of bases together and treat them as a single entity. Given the sequence is random, the relative fluctuations of the binding energy of such “super bases” become small. At the same time, each super base is likely to have a unique binding partner, as assumed when choosing the set of allowed configurations. The assumption made is thus not that restrictive and the homogenous Poland-Scheraga model is adequate to study the melting transition³.

Poland-Scheraga models are essentially one-dimensional, similar to an Ising model. It is well known, that one dimensional models do not exhibit genuine phase transitions. This fact is in conflict with the observed melting behavior of DNA and the apparent contradiction troubled (theoretical) physicists a while. A genuine melting transition is only obtained if the proper dependence of the loop cost on the loop length is included in the model. The logarithmic term in Eq. (1.6) introduces an effective long range interaction, that gives rise to an order-disorder phase transition in such one dimensional models [23]. The detailed thermodynamics of DNA was worked out by Poland and Scheraga in the publications [14, 15] and later summarized in their book [24]. We will now briefly summarize the statistical physics of homogenous DNA following Poland and Scheraga. For a homogenous

³Obviously, this assumption breaks down when macroscopic regions differ in CG content.

DNA molecule the recursion relation Eq. (1.5) simplifies to

$$Z_n = qZ_{n-1} + \sum_{m=1}^{n-2} \frac{qg^2s^{2m}}{m^c} Z_{n-m-1}, \quad (1.7)$$

where $q = e^{\frac{\varepsilon_b}{k_B T}}$, $g^2 = e^{-\frac{\varepsilon_\ell^0}{k_B T}}$ and the starting value of the recursion is set to $Z_1 = q$. This recursion relation can be solved by z -transformation. The z -transform of Z_n is defined as $\hat{Z}(z) = \sum_{n=0}^{\infty} Z_n z^n$ and is also known as generating function or discrete Laplace transform. Multiplying both sides of Eq. (1.7) by z^n and summing over n yields after some algebra

$$\frac{\hat{Z}(z) - qz}{z} = q\hat{Z}(z) + qg^2\Phi_c(zs^2)\hat{Z}(z), \quad (1.8)$$

where $\Phi_c(z) = \sum_{n=1}^{\infty} \frac{z^n}{n^c}$ is the polylogarithm. Eq. (1.8) is readily solved for $\hat{Z}(z)$

$$\hat{Z}(z) = \sum_{n=0}^{\infty} Z_n z^n = \frac{qz}{1 - qz - qg^2z\Phi_c(zs^2)}. \quad (1.9)$$

This z -transformed partition sum is nothing but the grand-canonical partition sum of a DNA molecule coupled to a fictive nucleotide reservoir with fugacity z . The original partition sum of a molecule of length N can now be recovered from $\hat{Z}(z)$ by contour integration around the origin of the complex plane.

$$Z_N = \frac{1}{2\pi i} \oint dz \frac{\hat{Z}(z)}{z^{N+1}} = \frac{1}{2\pi i} \oint dz \sum_{n=1}^{\infty} \frac{Z_n}{z^{N-n+1}} \quad (1.10)$$

The function $\hat{Z}(z)$ is analytic everywhere, except on $[s^{-2}, \infty[$ and possibly at isolated singularities, *i.e.* zeroes of the denominator of Eq. (1.9). Having identified the singularities and branch-cuts, the contour integral can be evaluated by calculating the residuals and the integral encircling the branch-cut, as illustrated in Fig. 1.4. A graphical solutions for zeroes of the denominator for different values of c are given in Fig. 1.4. At low temperatures, that is large q , the denominator has a real root z^* . If $c > 1$, this root merges with the branch-cut at some critical temperature and does not exist in the high temperature regime. The partition function of a DNA molecule of length N is therefore of the form

$$Z_N = \frac{\text{Res}(\hat{Z}(z), z^*)}{z^{*N+1}} + As^{2N} \quad \text{or} \quad Z_N = As^{2N}, \quad (1.11)$$

depending on whether the root z^* exists or not. If z^* exists, the fraction Θ of base pairs present in the structure is given by the logarithmic derivative of $\ln Z_N$ with respect to q

$$\Theta = \frac{1}{N} \frac{\partial \ln Z_N}{\partial \ln q} = -\frac{\partial z^*}{\partial \ln q}. \quad (1.12)$$

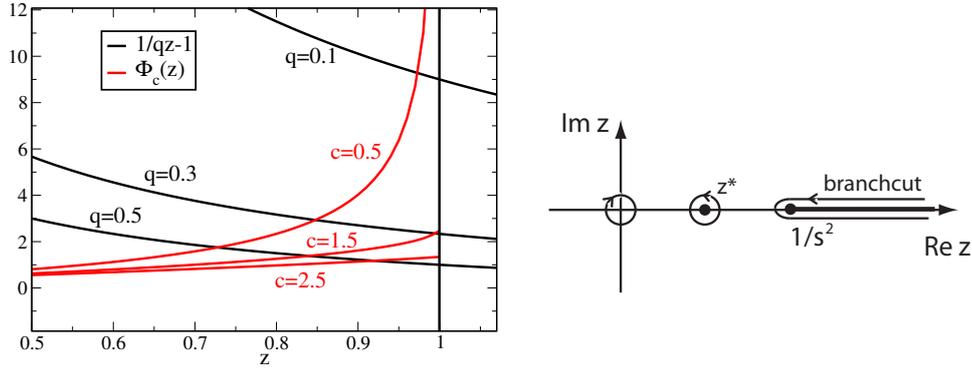


Figure 1.4: Left: Zeros z^* of the denominator of Eq. (1.9) are given by the intersections of $g\Phi_c(zs^2)$ and $q^{-1}z^{-1} - 1$ (g and s are set to 1 for simplicity). Right: Contour integration in the fugacity plane. The contour integral around the origin is the sum of the residue at $z = z^*$ and the integral encircling the branch cut. For large N , the integral is dominated by the residue.

If the isolated singularity does not exist Z_N does not depend on q and Θ vanishes. The existence of z^* is therefore connected to the phase where the two strands are bound and the temperature at which z^* ceases to exist corresponds to the melting temperature T_m . The order of the melting transition is determined by the value of the loop closure exponent c [15, 20]: If $c \leq 1$, $\Phi_c(z)$ diverges as $z \rightarrow 1$, hence there is always a solution z^* and no melting transition exists. If $1 < c \leq 2$, $\Phi_c(z)$ remains finite as $z \rightarrow 1$ but approaches its limiting value with infinite slope, resulting in a melting transition where Θ approaches zero as $T \rightarrow T_m$ and the melting transition is continuous. If $c > 2$, $\Phi_c(z)$ tends to its limiting value with finite slope and Θ drops from a finite value to zero at $T = T_m$, giving rise to a first order melting transition. Experimental melting curves of DNA are very steep, *i.e.* the fraction of bound bases vanishes very rapidly, and denaturation appears to be a first order transition. The additional contributions to c from loop interactions might therefore be relevant to reconcile the Poland-Scheraga models with experimental data. Available experimental data has been reexamined using $c = 2.15$ instead of $c = 3\nu \approx 1.8$, resulting in a smaller estimate of the effective loop initiation cost [25].

1.3. Mechanical properties of DNA

So far, we discussed the chemical and thermodynamic properties of DNA and neglected the organization of DNA in space. A typical human chromosome is 10^8 bp long, corresponding to a string of 3 cm in length. Forty-six of these strings have to fit into the cell's nucleus, which is only several micrometers in diameter. Packaging and compactifying DNA is thus a nontrivial issue to cells, especially since they have to keep their genome, or at least the relevant parts, accessible. This problem is addressed in the second part of this thesis. Obviously, the mechanical properties of DNA play an important role in DNA compactification and the dynamics of compactified DNA. During the last 15 years, it has become possible to study the mechanical properties of DNA by manipulating single

molecules and measuring their response to pico Newton forces. These single molecule force spectroscopy techniques provided unprecedented insight into the static and dynamic properties of biological macromolecules and even allow to study cellular machinery such as polymerases or topoisomerases life on stage. The dynamics of repetitive DNA sequences has also been studied using such techniques, which will be discussed in chapter 2. I will therefore give a brief overview over such techniques and then discuss the mechanical properties of DNA.

1.3.1. Single molecule force spectroscopy

At the molecular level, biological processes involve energy differences of the order of the thermal energy $1k_B T \approx 4 pN nm$ and length scales on the order of nano meters. To probe biological macromolecules mechanically, instrumentation is needed that is capable of applying forces in the pico Newton regime and measure distance with nano meter resolution. By now, a variety of different techniques are available to achieve this feat and I will briefly discuss their basic mechanism as well as their advantages and drawbacks. Atomic force microscopy is discussed in a little more detail, since the slippage of repetitive DNA was detected using this technique (see also Sec. 2.3). For comprehensive reviews of the different techniques see for example [26, 27, 28].

Optical tweezers. When an object with an optical density higher than the surrounding media is placed in a non-uniform electric field, it feels a force towards the stronger field. This effect is exploited in optical traps, where a small spherical bead is held in a laser focus. As soon as the bead is no longer centered in the focus, it experiences a restoring force. Although the explanation illustrated in Fig. 1.5 is not exactly applicable to beads of sizes of the order of a micrometer, it conveys the essence of the method. The laser light is refracted by the bead and thereby transmits momentum to the bead. If the bead is not centered, the laser intensity on the two sides of the bead are not equal and hence the transmitted momenta do not balance, resulting in a net force towards the focus. To study the response of a system to mechanical force, it is attached to the bead and the exerted force can be determined by measuring the deviation of the bead from the trap center. Interferometric methods allow to determine the bead position to nm resolution, which for typical trap stiffnesses results in force resolution of pN and below. The maximal forces optical traps can apply depend on the bead size and are in the range of 20 to 150pN. One important application of optical tweezers has been the unzipping of single DNA molecules [29], which is discussed in more detail below in Sec. 1.3.4. The motion of single processive molecular motors has also been studied using optical tweezers [30].

Magnetic tweezers. Similar to optical tweezers, magnetic tweezers exploit the fact that magnetic dipoles are attracted to high field regions and therefore experience a force in a gradient field. In addition to force, magnetic fields exert torque on permanent magnetic dipoles. This opens up the possibility to twist biomolecules by rotating the magnet that

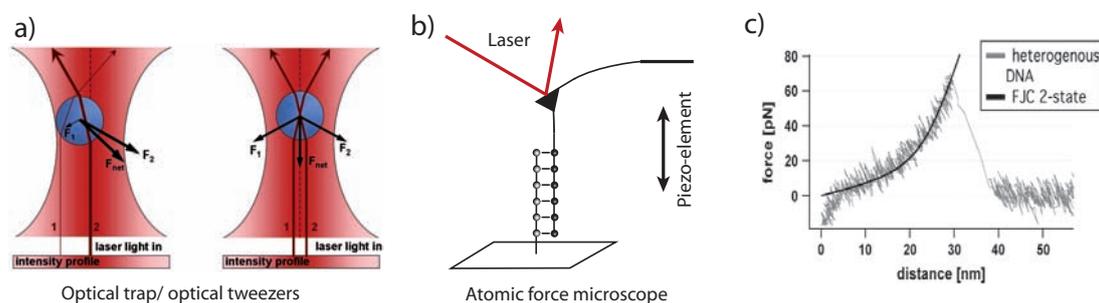


Figure 1.5: Part a) Ray optics explanation of an optical trap. Image source: Wikipedia. b) Sketch of an AFM in a single molecule experiment. The different parts are extremely out of scale. c) A typical force-extension curve recorded with an AFM (taken from ref. [1]).

generates the field. The position of the beads can be detected using similar interferometric techniques as for optical tweezers, reaching nm resolution. The force can be sensitively controlled through the field gradient and forces as small as 10 fN can be applied. Magnetic tweezers have been used to study the stretching response of supercoiled DNA [31] and to observe topoisomerases, the enzymes that disentangle DNA, in action [32].

Biomembrane force probes. Yet another technique of measuring small forces are biomembrane force probes. Small lipid vesicles or red blood cells are partially sucked into a micropipette to establish a well controlled tension of the vesicle membrane. The system to be studied is attached to the other end of the vesicle and pulled away. The deformation of the vesicle can be related to the force applied to the sample. This technique has been used for measuring the binding strength and kinetics of receptor-ligand systems [33].

Atomic force microscope. While optical and magnetic techniques excel at small forces with exquisite resolution, the realm of the atomic force microscope (AFM) are forces above 5pN. Among all force spectroscopy techniques, the AFM has by far the greatest spatial resolution, which can be as good as the diameter of an atom. Historically, AFMs were invented to map surfaces at atomic resolution, and only later became important tools to study mechanical properties of biological macromolecules or molecular complexes. An AFM used for force spectroscopy consists of a tiny solid state cantilever with an even tinier tip. The substrate surface is mounted onto a piezo element which allows to move the sample with respect to the cantilever with extremely high precision. A force exerted on the cantilever will cause a slight bend in the cantilever. This minute deflection can be measured by shining a laser beam onto the reflective upper side of the cantilever, as sketched in Fig. 1.5. A deflection of the cantilever changes the reflection angle of the beam, which in turn can be sensitively detected by a split photodiode. Since control and detection is done by fast electronic devices, the bandwidth of AFM measurements can be as high as 100kHz and is limited by the viscous damping of the cantilever.

The substrate and the cantilever have to be prepared such that upon bringing the can-

cantilever in contact with the substrate, the sample attaches to both, the substrate and the cantilever. Often, the connection to the cantilever and substrate is established via well characterized and chemically functionalized linker molecules. Using linker molecules increases the distance of the sample from the surface and thereby reduces surface effects. The sample density has to be chosen such that a single contact between cantilever and substrate is more probable than multiple linkages. The substrate surface is then retracted and the force is measured as a function of extension. Depending on the question to address, the force extension relation or the dependence of rupture forces on retract speed are informative quantities.

1.3.2. Stretching double stranded DNA

On a microscopic scale, dsDNA is a very stiff polymer and thermal forces bend DNA only on length scales long compared to the helical pitch or even individual base pairs. To a good approximation, the DNA bendability is continuously distributed along the DNA and the typical curvature radius is large compared to molecular dimensions. On the other hand, up to forces of about 50pN, dsDNA is virtually inextensible. Ignoring excluded volume effects the equilibrium conformations of DNA are therefore well described by the ensemble of inextensible contour lines with a linear bending stiffness, a model commonly referred to as *worm-like-chain model* (WLC) [34, 35]. The energy of a particular contour $\mathbf{r}(s)$ with a force \mathbf{f} pulling the endpoints apart is given by

$$E = \frac{\kappa}{2} \int_0^L ds \mathbf{t}'(s)^2 - \mathbf{f} \cdot (\mathbf{r}(L) - \mathbf{r}(0)), \quad (1.13)$$

where κ is the bending stiffness and $\mathbf{t}'(s)$ denotes the derivative of the tangent vector with respect to the arclength s . To calculate the equilibrium properties of such a chain immersed in a heat bath, one would have to calculate the integral over all possible paths $\mathbf{r}(s)$, which in general is infeasible. Some quantities, however, can be calculated exactly. In the absence of force, the most important exactly known quantity is the tangent correlation function at different points of the contour.

$$\langle \mathbf{t}(s) \cdot \mathbf{t}(s') \rangle = e^{-\frac{|s-s'|}{\ell_p}}, \quad (1.14)$$

where $\ell_p = \frac{\kappa}{k_B T}$ is called the persistence length. The persistence length is the length scale at which the correlations of different parts of the chain decay and a molecule is considered flexible, if its total length is large compared to ℓ_p . Conversely, a chain several times smaller than the persistence length is typically straight. The persistence length of double stranded DNA under physiological conditions is $\ell_p = 50\text{nm}$.

Short molecules. Polymers that are short compared to the persistence length are often referred to as *semi-flexible*. The typical contours of these polymers are deviations from a straight line. If the straight contour is the z -axis, the contour can be parameterized by two

single valued functions $x(z)$ and $y(z)$. Furthermore, longitudinal contraction is only of second order, such that we can identify the arclength with z . Within this weakly bending approximation, the equation of motion of the polymer is given by [36, 37, 38]

$$\frac{\partial x(z, t)}{\partial t} = -\frac{k_B T \ell_p}{\zeta} \frac{\partial^4 x(z, t)}{\partial z^4}, \quad (1.15)$$

where ζ is the friction coefficient per length (analogously for $y(z, t)$). The eigenfunctions of this equation are of the form $W_n(z) = a_1 \sin k_n z + a_2 \cos k_n z + a_3 \sinh k_n z + a_4 \cosh k_n z$ with a discrete set of wave numbers k_n fixed by the boundary conditions. The corresponding relaxation times are $\tau_n = \zeta / (k_n^4 k_B T \ell_p)$

Long molecules. According to Eq. (1.14) the correlation length of the tangent vectors $\mathbf{t}(s)$ along the backbone is the persistence length ℓ_p . Hence, a polymer that is far longer than its persistence length will form a random coil where the number of independent segments is given by $\sim L/\ell_p$. The diameter of the coil increases with length as $\sim \ell_p (L/\ell_p)^\nu$, where $\nu \approx 0.588$ is the Flory exponent. The end-to-end vector is a sum of independent increments and hence Gaussian distributed. The number of possible chain configurations for a given end-to-end distance is maximal at zero separation and decreases rapidly as the ends are pulled apart. Entropy therefore favors small end-to-end distances and gives rise to a restoring force opposing stretching. The force extension relation of a dsDNA molecule several micrometers in length has been measured by Smith et al. using magnetic tweezers [39]. At distances Δr small to the backbone length the polymer responds like a linear spring with entropic spring constant $k = \frac{3k_B T}{2\ell_p L}$. The force-extension relation becomes non-linear as soon as the force exceeds $k_B T/\ell_p$. At very strong stretching, Δr approaches the contour length and the undulations of the of shorter and shorter wavelength are straightened out. The stretching force diverges quadratically as Δr approaches L [40].

Overstretching DNA. DNA ceases to be well described by an inextensible WLC model at stretching forces of about 65pN, where the molecule suddenly extends by a factor of 1.7 [41, 42]. The transition is reversible and very little hysteresis is seen when the molecule is first overstretched and subsequently relaxed. Upon overstretching DNA changes from its ordinary structure called B-form to S-form. For this reason, the transition is called B-S-transition. Since the mechanical properties of S-DNA are different from one single DNA strand, two separated single DNA strands, and ordinary B-DNA [43], it is generally believed that S-DNA is double stranded but has a structure distinct from B-DNA. The true structure of S-DNA is not completely resolved. Rief et al. report another conformational transition at forces of about 150pN, which is irreversible on experimental time scales [44]. The force-extension trace of relaxation suggests that the two strands separate during the transition and only one single DNA strand remains attached between the substrate and the cantilever. This force induced unpeeling is strongly sequence dependent.

1.3.3. Stretching single stranded DNA

Single stranded DNA responds differently to stretching than dsDNA. Inspection of the chemical structure of ssDNA sketched in Fig. 1.2 already hints at the great flexibility of ssDNA. The monomers are attached to each other via a single chemical bond, about which the bases can rotate and bend. In fact, ssDNA in solution reorientates about every two to three bases. As opposed to dsDNA, the bendability of ssDNA is no longer continuously distributed along the chain but concentrated at the joints between the bases. A suitable model for such a system is the *freely jointed chain* (FJC) model which describes a polymer by a chain of rigid rods which are connected at hinges. The length of single stranded DNA corresponding to one segment of the FJC is about 1.5 to 2 nm.

Without a stretching force, the FJC model is equivalent to a random walk in space or, if mutual exclusion of the monomers is accounted for, a self-avoiding random walk, as already discussed for the long WLC polymer. At large stretching force, the response of the FJC differs from that of the WLC due to the fact that WLC polymer displays undulations at all wavelengths whereas the FJC has a lower cut-off length given by the monomer length. The statistical mechanics of a FJC polymer under tension is very simple and is equivalent to that of a paramagnet in an external magnetic field. The partition function of a single monomer of length b is given by

$$Z = \frac{1}{4\pi} \int d\phi d\cos\theta e^{-\frac{fb\cos\theta}{k_B T}} = \frac{k_B T}{fb} \sinh \frac{fb}{k_B T}, \quad (1.16)$$

where the force is parallel to the z -axis. The partition function of a N -monomer chain is simply Z^N . From this, the force extension relation is readily calculated

$$\Delta r = -Nk_B T \frac{\partial \ln Z}{\partial f} = Nb \left(\coth \frac{fb}{k_B T} - \frac{k_B T}{fb} \right). \quad (1.17)$$

As Δr approaches Nb , the force diverges as $f \sim (Nb - \Delta r)^{-1}$. The functional dependence of Δr on $\frac{fb}{k_B T}$ is known as Langevin function. For a thorough discussion of this and similar models, see [45].

1.3.4. DNA unzipping

Using single molecule manipulation techniques, one can unzip a single dsDNA. While separating the two strands, the force needed for unzipping is recorded. Earlier experiments achieved a spatial resolution of hundreds of base pairs [46], which was later improved to tens of base pairs [29]. To interpret these experiments, it is helpful to consider the time scales involved. The unzipping speeds used in these experiments are on the order of 100nm/s, which corresponds to 300 bp per second. On the other hand, the intrinsic dynamics of base pair formation is faster than 10^6 bp per second [47, 48]. Hence, unzipping is slow compared to the base pair formation and the unzipping fork is essentially in equilibrium. The opening of one base pair adds two bases to the single stranded part. The free energy per base of the single stranded DNA under tension can be calculated using Eq. (1.16).

The force adjusts itself such that this free energy equals half the binding free energy of a base pair. Hence, the binding free energy can be calculated from the measured force, yielding results in agreement with bulk thermodynamics. The coupling of the dsDNA to the measurement device is soft, such that the fork averages over many base pairs. As expected, the estimated local binding free energies correlate with the GC-content of the sequence. Unzipping forces range between 10pN for AT rich sequences to about 15pN for GC-rich sequences.

These unzipping experiments attracted the attention of many theoretical physicists which studied the nature of the unzipping transition [49] and in particular focussed on the effect of sequence heterogeneity [50, 51, 52]. The unzipping transition is a first order phase transition. The double helical state is stable at low force and the completely unzipped state is favorable at high force. If the experiments are performed in the constant extension ensemble, the opening fork of the unzipped DNA is the analog of a meniscus separating two phases. While the phase diagram is extremely simple, the nature of the transition and the unzipping dynamics is sensitive to sequence disorder. When unzipping homopolymers, every part of the molecule becomes unstable at the critical force and the number of unzipped bases diverges as $m \sim (f - f_c)^{-1}$ as the transition is approached from below. If the sequence consists of a random mixture of weakly and strongly binding base pairs, the local binding energy fluctuates. Even though the energy landscape for unzipping is flat on average at the critical force, it fluctuates up and down like an unbiased random walk. Since the standard deviation of an unbiased random walk grows with square root of the number of steps, the energy barriers the unzipping force has to overcome to proceed m bases are typically of height $\Delta E \sqrt{m}$, where ΔE is the difference in binding energy between the strongly and weakly binding base pairs. It can be shown, that the number unzipped bases m diverges quadratically as the transition is approached [49]. Due to energy barriers on all scales, unzipping at constant force is often interrupted by long pauses and the unzipping fork exhibits anomalous dynamics.

2. Dynamics of repetitive DNA

At first sight DNA with repetitive sequences seems to be a rather artificial concept and one would not expect such DNA to be relevant in biology. Consider a DNA sequence such as 5'-CACACACACACACACACA-3' and its complementary counterpart 3'-GTGTGTGTGTGTGTGTGTGTGT-5'. In a randomly assembled sequence of typical genome length the probability to find this particular sequence is small (there are $4^{20} \approx 10^{12}$ ways to assemble a 20 bp sequence, a mammalian genome is about 10^9 bps long and hence the chance of occurrence is on the order of 10^{-3}). Nevertheless, perfectly periodic sequences, *i.e.* repetitions of short motifs of one to six nucleotides, are extremely common in eukaryotic genomes [53] and account for up to 3% of the human genome [6]. This drastic overrepresentation of repetitive sequences cannot be linked to any particular function, since most of these repetitive sequences have been found in non-coding regions of the genome. Instead, what makes repetitive DNA special compared to ordinary DNA is its much richer dynamics. While two complementary single stranded DNA molecules with a sequence that is not particularly ordered form base pairs only when correctly aligned, repetitive sequences can bind out of register and form asymmetric loops as illustrated in Fig. 2.1. In particular, two complementary repetitive single strands can slip, meaning they can bind to each other when locally shifted. This phenomenon of DNA-slippage is the key to understand the peculiarities of repetitive DNA. In the following, we will outline the role of repetitive DNA in biology and discuss how it is linked to human hereditary diseases.

Using a simple model of repetitive DNA we explore the potential of single molecule experiments to study the dynamics of DNA slippage. Our theoretical analysis suggests, that DNA-slippage can be probed by applying a shear force to a repetitive dsDNA. We find that the two repetitive DNA strands start moving relative to each other when a sufficiently high force is applied. The observed sliding speed can be related to the microscopic dynamics of DNA-slippage. Hence, a thorough understanding of this sliding motion might give insight into the molecular basis of DNA-slippage and shed light on the evolutionary dynamics of repetitive sequences. The peculiar properties of repetitive DNA could also be exploited in nanotechnology as visco-elastic elements and force generators. Our theoretical study is complemented by a collaboration with the lab of Prof. H.E. Gaub, where Ferdinand Kühner and Julia Morfill succeeded in measuring DNA sliding using an atomic force microscope. These experiments are also discussed briefly.

Not only the dynamical properties of repetitive DNA are different from ordinary DNA, but also its equilibrium thermodynamics is richer. If two repetitive and complementary DNA strands of different length bind to each other, they undergo an additional temperature driven phase transition before they separate into two single strands at high temperatures. This additional transition will be discussed at the end of this chapter.

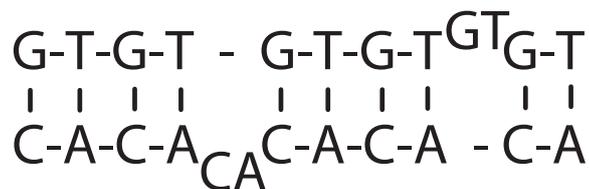


Figure 2.1: Two complementary DNA strands with repetitive sequence can bind in many different configuration, since strands are complementary even when locally shifted by an integral multiple of a repeat unit. In particular, repetitive DNA can form asymmetric loops and bulge loops, resulting in local strand slippage.

2.1. The biological role of repetitive sequences

Repetitive sequences have first been observed in the early 80's [54] and since then have been found in every eukaryotic organism that was investigated. Repetitive sequences with short repeat motifs (one to six base pairs) are commonly called *microsatellites*¹, *simple sequence repeats* (SSR) or *short tandem repeats*. To me, simple sequence repeat (SSR) appears to be the most natural name and I will try to stick to it. Most of the simple sequence repeats are found in non-coding DNA and are believed to evolve more or less neutrally, that is the reproductive fitness of the organism is independent of length of the SSR. Only very little is known about possible functional roles of repetitive DNA, see below in Sec. 2.1.4. Within non-coding DNA, mono- and di-nucleotide repeats are the most abundant, while within coding DNA, predominantly tri-nucleotide repeats are found. Tri-nucleotide repeats constitute a special class of repeats, since the genetic code assigns amino acids to combinations of three bases, so called codons. A tri-nucleotide repeat in coding DNA therefore corresponds to a repeated amino acid in the protein. An extension or contraction of a tri-nucleotide repeat results in the deletion or insertion of an amino acid but leaves other parts of the poly-peptide sequence unaltered. This is very different for most other repeat lengths, where expansions or deletions result in frameshift mutations, *i.e.* the interpretation of the DNA sequence as three base codons is changed for the entire part downstream of the repeat expansion. This most certainly results in a useless protein or premature termination of transcription. The special role of tri-nucleotide repeats and their relation to human hereditary diseases will be discussed in greater detail below.

2.1.1. The number of repeats changes rapidly in evolution

The key to understand the importance and ubiquity of SSRs is the extraordinarily large rate at which the number of repeats changes from generation to generation. Although numbers have to be taken with care, rates of contractions and expansions of SSRs in mammals

¹DNA containing long repeated motifs forms 'satellite' peaks in centrifugation experiments and such DNA was called satellite DNA. Later shorter and very short repeat motifs were observed and became known as mini- and microsatellites.

can be as high as 10^{-2} per locus and generation [55]. This is orders of magnitude higher than the typical rate for base substitutions which in mammals is about 10^{-9} per base and generation. This hyper-variability can be linked to a peculiarity of the mechanism by which DNA is replicated prior to cell division. To replicate DNA, the double stranded molecule is separated into two single strands by a helicase and the two single strands serve as templates to which the complementary strands are added by the DNA polymerase. However, the DNA polymerase operates only from the 5' to the 3' end. Therefore, only one strand, the so called *leading* strand, is copied continuously while the other strand, the *lagging* strand, is copied piecewise as illustrated in the upper panel of Fig. 2.2. The pieces of DNA that are added at a time are known as *Okazaki fragments*. The 5' end of an Okazaki fragment is fairly unprotected and a couple of bases will frequently detach from the template strand by thermal fluctuations. Whenever the 5' end of an Okazaki fragment happens to have a repetitive sequence, it is possible that it rebinds in a misaligned manner, forming a bulge loop containing one or more repeat units. If the DNA polymerase fills in the next Okazaki fragment while such a bulge loop is present the number of repeat unit on the copied strand has changed with respect to the template strand. A bulge loop on the template strand results in the deletion of one repeat, whereas a loop on the newly synthesized strand adds a repeat, as illustrated in the lower panel of Fig. 2.2.

Energetically, SSR contraction is greatly favored over repeat expansion, since only one repeat unit has to open before a loop on the template strand can be formed, compared to two repeat units that have to dissociate to form a loop on the nascent strand. Hence, one would expect SSRs to contract and disappear quickly if the mutation mechanism was adequately described by the sketches in Fig. 2.2. While this conclusion obviously contradicts the abundance of SSRs in eukaryotic genomes, it explains the fact that SSRs are rare in prokaryotes and tend to contract during PCR, as discussed below in Sec. 2.1.5. The high ratio of expansion to contractions observed in eukaryotes is probably connected to the DNA mismatch repair machinery, which checks the double helical structure of the newly synthesized DNA. Only mutations that escape this repair machinery or are falsely corrected persist to the next generation. Experiments in yeast have shown that a malfunctioning DNA mismatch repair system causes an increase of SSR mutations of 100-700 fold [56]. These findings provide further indirect evidence that misaligned rebinding during replication, *i.e.* replication slippage, is an important source of SSR expansion and contraction. *In vivo* rates of SSR evolution depend on a variety of different factors, most of which are still heavily debated in the literature. For a concise summary I recommend ref. [57].

SSR genesis and the length dependence of the mutation rate. The longer a SSR, the higher the probability that an open end during replication lies within the SSR. One would therefore expect a linear increase of the expansion and contraction rate with the number of repeats. Observations confirm a positive correlation between the repeat number and the mutation rate, the precise length dependence of the mutation rate, however, is less clear and a large body of contradicting evidence exists (for a review, see [53]). Before a SSR can start growing by replication slippage, it has to contain at least two repeat units. It has

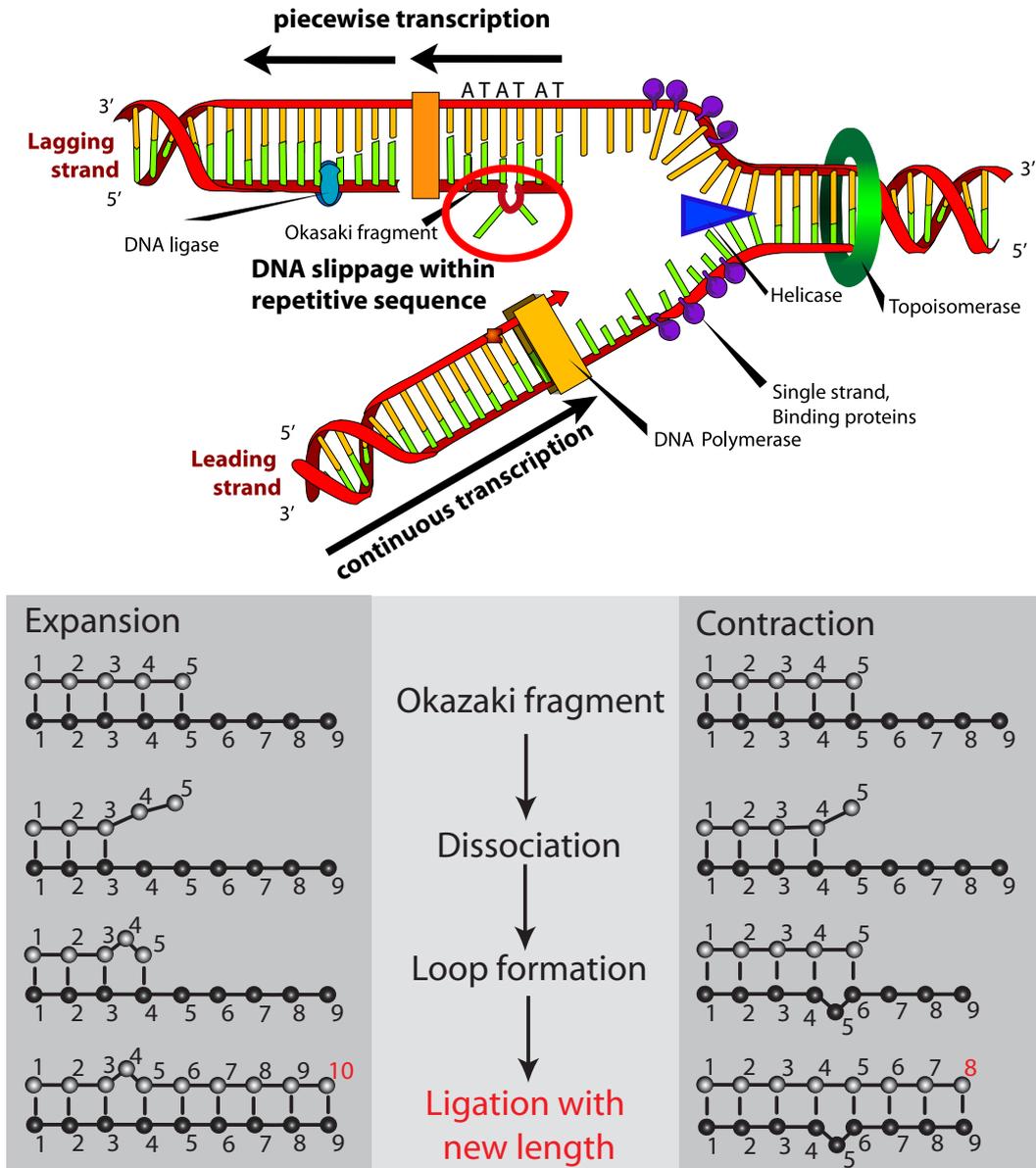


Figure 2.2: Upper panel: Replication of the lagging strand is done piecewise and the stretches replicated in one round are called *Okazaki fragments*. When an Okazaki fragment has a repetitive sequence, the two strands can dissociate and subsequently rebind in a misaligned manner, as illustrated by the bulge loop inside the red circle. Image adapted from Wikipedia. Lower panel: Depending on whether the loop occurs on the template strand or the new strand, the number of repeat units is increased or decreased in the copy.

been shown, that expansion of SSR does actually occur for SSRs as short as two units. It is generally believed that these initial seeds for SSR expansion are assembled by chance.

Dependence of mutation rate on the repeat unit length. Another feature one would expect to have drastic effects on the SSR mutation rate, is the length of the elementary repeat unit. The longer a repeat unit, the more bases have to dissociate before DNA-slippage can occur. The activation free energy for DNA-slippage therefore increases with repeat length and rates are expected to be small for SSRs with long repeat units. This reasoning is very well supported by *in vitro* experiments (cf. Sec. 2.1.5), but *in vivo* evidence is less conclusive. Some experiments seem to confirm that shorter repeat units mutate faster than long repeat units [58]. A bioinformatics study also reports strongly decreasing mutation rates with repeat length [59]. Repeated motifs that are longer than five or six bases are not known to mutate significantly by replication slippage.

Length distributions of SSRs. A comparative study of the length distributions of SSRs of different repeat length in humans, mice, fruit-flies, and yeast revealed significant differences in abundance and distribution between different organisms [59]. In all cases, short SSRs are most abundant. The longest SSRs are found in mice, but even for mice the frequency drops rapidly to zero beyond 40 repeats. The absence of very long SSRs is somewhat puzzling, since there appears to be a bias towards expansions of repeats². Furthermore, the mutation dynamics becomes faster as the length increases. Very long SSRs are therefore expected. One possible resolution to this puzzle are point mutations, which split a SSR into two smaller ones. The frequency of such a point mutation within one locus increases with increasing length and thus provide a plausible explanation for stationary length distributions [59]. However, one should keep in mind that many other influences, most importantly selection or unanticipated properties of the mismatch repair system, might be just as important to understand SSR length distributions.

The mechanism of SSR expansion discussed so far, *replication slippage*, changes the length of an SSR usually by one repeat unit, sometimes by a few, but never by many. However, the length of some special classes of SSRs tends to expand by a large number of repeats in one generation. A common feature of such SSRs is their ability to form hairpins, *i.e.* the ability of the single strand to fold back onto itself and form a stable structure. Possible mechanism of SSR expansion due to hairpin formation are reviewed in [60, 61]. Another source of length variations of SSRs is unequal crossing-over during meiosis. When a diploid organism produces haploid gametes, the genes on the chromosomes are reshuffled by building new chromosomes out of pieces of the old ones. In repetitive parts of the genome, these recombination sites are ambiguous, which can give rise to two new chromosomes with SSRs of different length.

²As noted above contractions are less costly energetically, but more expansions seem to escape the mismatch repair machinery, resulting in an expansion bias.

2.1.2. SSRs are versatile genetic markers

The human genome contains hundreds of thousands SSRs, each of which changes its length with a probability of $10^{-6} - 10^{-2}$ in each round of replication. The chance, that two humans have the same set of SSRs is therefore negligibly small, even between closely related individuals. For this reason SSRs are ideally suited as genetic markers and have acquired great popularity in phylogeny studies, paternity testing and forensic sciences. To measure the length of a set of SSRs, one exploits the fact that each SSR can be uniquely identified by its flanking sequences. A short fragment of DNA containing the SSR to be analyzed is cut from the sample using restriction enzymes that cut DNA specifically at the flanking sequences of the SSR. The short fragments are amplified by PCR and their length is measured using gel electrophoresis. The resolving power of these techniques is high enough to detect insertion and deletions of single repeat units [62]. The great advantage of SSRs based genotyping techniques is the small length of the sequence fragments that need to be analyzed. Since short sequences can be efficiently amplified by PCR, minute DNA samples suffice for reliable data. Depending on the questions one wants to address, different genetic markers with different mutation rates, chromosome type (autosomes, X-, or Y-chromosome) or flanking sequence are more suitable than others.

2.1.3. SSR expansion is related to hereditary diseases

Most SSRs reside in non-coding regions of the genome and mutations of these SSRs have little or no effect on the fitness of the individual. A special class of tri-nucleotide SSRs, however, occurs in coding regions or is part of introns, *i.e.* regions of a gene that are transcribed but spliced from the mRNA before translation, and mutations of these can lead to severely impaired phenotypes. Only tri-nucleotide repeats are found in coding regions due to strong selection against frameshift mutations, which result by insertion or deletion of a number of bases that is incommensurate with three. Mutations of these tri-nucleotide SSRs are linked to a number of severe human hereditary diseases, such as *Huntington's*, *fragile X* or *Friedenreich's ataxia*. These diseases fall into two different categories [63]. The first class of tri-nucleotide related diseases is caused by expansions of **CAG** repeats that code for the amino acid glutamine. The most prominent member of this class is Huntington's, which we will discuss in little more detail. Huntington's is a neurodegenerative disease that develops gradually. At early stages, patients suffer from rapid uncontrollable movements. As the disease proceeds, patients lose virtually all motor control, including the ability to speak, eat, or facial expression. The gene containing the **CAG** repeat codes for the protein *huntingtin*, whose function is largely unknown. The number of repeats in healthy individuals ranges between 6 and 34, people with 35 to 39 repeats have an increasing risk of developing the disease during their lifetime and people with 40 or more repeats almost certainly suffer from Huntington's by the age of 40. Due to late onset of Huntington's disease, patients usually have children before the disease is detected. The disease tends to become worse from generation to generation as the **CAG** repeat tends to grow longer and longer. The rate of elongation is correlated with the

number of divisions in the paternal germ line [64]. The molecular basis of the pathology of mutated huntingtin is still not completely resolved. The most popular hypothesis is, that proteins with a long poly-glutamine stretch tend to aggregate and that those aggregates are toxic. Such aggregates have been found in the brains of deceased patients. It is unclear, however, whether these aggregates are key to the pathology or an unimportant byproduct [65]. In any case, it is extremely astonishing that the protein works fine with any number of glutamines between 6 and 34 and is almost certainly lethal with just 6 copies more.

The second class of pathological tri-nucleotide repeats are located in introns. In some way or the other, the expansion of the SSR prevents transcription of the gene or the processing of the mRNA for translation into protein. Pathological expansions often reach repeat numbers as high as 2000, which is to be compared to the normal range of 5 to 50 [63]. The most prominent member of this class is the fragile X syndrome causing mental retardation. Fragile X results from an expansion of a CCG unit beyond 200 repeats. The pathology of fragile X is believed to be related to methylation of CG di-nucleotides, which might silence the transcription of the gene.

2.1.4. SSRs in prokaryotes

In sharp contrast to eukaryotes, repetitive DNA is extremely rare in prokaryotes and seems to occur only at loci, where there is a strong selective advantage to keep it. The prime purpose of SSRs in prokaryotes are *contingency genes*, *i.e.* bacteria exploit the high mutation rate of SSRs to maintain genetic and phenotypic diversity within a population. Such diversity is essential to any organism subject to environmental changes, which may lead to extinction if not a small fraction of the population happens to be prepared for the new conditions [66]. SSRs are exceptionally well suited for contingency genes, since SSR mutations are frequent and lead to repeat number changes only. A change in repeat number differs from ordinary base substitution mutations, since they are easily reversed. Assuming unbiased expansion or contraction, there is a 50% chance that a mutation is undone by the subsequent mutation, or put into more academic terms, random walks in one dimension are recurrent and almost surely to return to the origin in finite time. This is very different for ordinary mutations, which correspond to a random walk in a very high dimensional space (every base can be of four different types), where the chance of returning to a prior state is negligibly small.

If there is a way to couple SSR contraction and expansion to switching genes on and off or to change protein function, rapid SSR mutations would result in phenotypic variation within populations and at the same time ensure easy recovery of temporarily switched off traits. Not surprisingly biology has found several ways to do so. Among the best studied examples are contingency loci of human pathogens such as *Haemophilus influenzae*³ or *Neisseria meningitidis* [67]. To evade the immune system, these bacteria frequently exchange proteins in their outer membrane. This variability is often achieved by placing SSRs either

³Although its name suggests that *H. influenzae* is the cause of the flu and hence a virus, *H. influenzae* is a gram-negative bacterium. It was mistakenly associated with the flu until 1933.

in the promoter region or the in coding sequence itself. A change in length within a promoter region can preclude necessary interactions of transcription factors or destroy the RNA polymerase binding site. SSRs within the coding regions often cause frameshift mutations, which results in transcription termination or an “gibberish” mRNA.

2.1.5. In vitro evidence for local strand slippage

By now, it is fairly well established that replication slippage plays a pivotal role in the mutational dynamics of SSRs. However, the change in repeat length from one generation to the next results always from an interplay of replication slippage and the DNA mismatch repair machinery. Only those mutations that escape mismatch repair or are falsely corrected can be detected. One way to study replication slippage alone is to knock out the DNA repair machinery [56]. Another way is to study DNA replication *in vitro* [68]. Three examples of such experiments are described below.

PCR slippage. The invention of the *polymerase chain reaction* (PCR) was one of the most important steps towards modern biotechnology. PCR allows to faithfully amplify minute amounts of DNA fast and cheaply. During one cycle of PCR, the DNA template strands are separated by thermal denaturation, subsequently the temperature is lowered such that short primer strands hybridize at the 3' ends of both strands. A heat resisting DNA polymerase then extends the short primers and copies the template. By multiple repetitions of these steps, the initial template is amplified exponentially. By now, PCR is a highly automated and very reliable technique. Only when amplifying repetitive DNA the amplification is error prone [69, 70]. The product DNA is a mixture of the faithfully copied template DNA and DNA sequences where the repetitive part has shortened by some repeat units. Why does the PCR lose repeat units while amplification? Whenever the template strand has a repetitive sequence and the polymerase happens to fall off the strand while transcribing the DNA, one strand can slip with respect to the other and form a bulge loop. Similarly to replication slippage *in vivo*, the number of repeats changes if the polymerase resumes the replication while such a bulge loop is present (cf. Fig. 2.2). Only contractions are observed due to the fact that the formation of a bulge loop has a lower activation energy on the template strand. Given a binding free energy per repeat unit ε_b , contraction is more likely than expansion by a factor of $e^{-\frac{\varepsilon_b}{k_B T}}$.

SSR synthesis via DNA-slippage. Schlötterer and Tautz succeeded in synthesizing repetitive sequences by exploiting DNA-slippage *in vitro*. They mixed short repetitive DNA with DNA polymerase and the required nucleotides in a suitable medium. After some incubation time, they measured the length distributions of the DNA strands and found that the DNA strands tend to grow [68]. The elongation rate primarily depends on the length and the binding strength of the elementary repeat unit. Di-nucleotide repeats grow at a rate of 4 to 6 base pairs per minute while tri-nucleotide repeats grow at a rate ranging from 0.5 to

3 base pairs per minute. The elongation rate of tri-nucleotide repeats correlates strongly with the number of AT base pairs in the repeat unit.

The observations can be explained by the following mechanism: The ends of the repetitive dsDNA undergo DNA-slippage and form a bulge loop which diffuses inside the double strand. The single stranded overhang produced by this slippage event is then filled in by the DNA polymerase. When the bulge leaves the double strand again, it produces another single stranded overhang, which is then filled in by the DNA polymerase. Since the rate of bulge loop formation depends exponentially on the binding energy of one repeat unit, dinucleotide repeats are expected to grow faster than tri-nucleotide repeats. High AT content should also enhance slippage, as observed.

Evidence for chain sliding. In the 1970s, Pörschke measured the hybridization kinetics of short repetitive RNA oligomers [71]. When complementary strands are mixed, the rate limiting step for hybridization is the formation of a critical nucleus of a few base pairs. Once such a nucleus is formed, the remaining bases rapidly close in a zipper-like manner. When sequences have no particular order, a stable nucleus and subsequent zipping is only possible if the two strands are correctly aligned. This is very different for repetitive sequences, since the two strands can bind with arbitrary shift relative to each other, see Fig. 2.3a. Once such a misaligned duplex is formed, it is stable since both strands are bound by many base pairs. Hence, one would expect to find a large number of misaligned double stranded intermediates with a different number of base pairs. The relaxation dynamics of these intermediates to the fully aligned state would further be expected to occur at markedly different rates, since the time required to dissociate the strands by thermal activation depends exponentially on the number of base pairs. However, no such slow multi-exponential relaxation is observed [71]. Pörschke himself provided a very plausible explanation for his results. As already discussed several times, repetitive sequences can form mobile bulge loops. The propagation of a bulge loop from one end to another shifts both strands by the length of the loop, very much like a rug can be moved by propagating slack from one side to another, see Fig. 2.3b. The energy cost for the nucleation of such a bulge is small and in particular does not depend on the length of the molecule.

2.2. Force induced DNA-slippage

The observations and experiments reported above provide good evidence that DNA-slippage is indeed happening and that it plays a crucial role during SSR evolution. However, the evidence for DNA-slippage is more or less indirect and based on bulk observations. One goal of this thesis was to suggest experiments that allow to observe DNA-slippage in single molecules using modern force spectroscopy techniques (cf. Sec. 1.3.1).

We suggest, that DNA-slippage can be induced by applying a shear force to repetitive DNA. In a nutshell, application of a sufficiently high shear force fosters the formation of bulge loops on both unstretched ends of the DNA duplex, which then travel across the duplex and exit on the other side, as illustrated in Fig. 2.3b. The duplex lengthens stepwise,

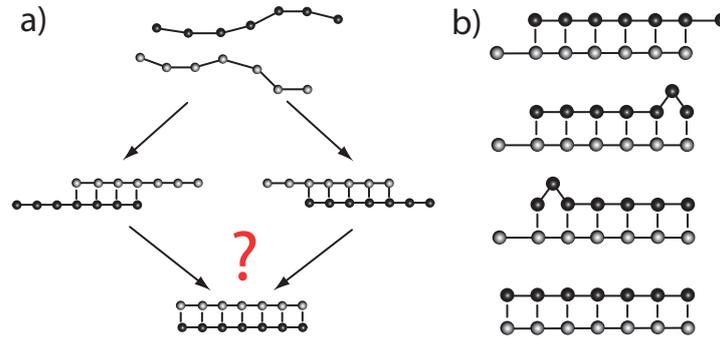


Figure 2.3: Left: During the hybridization reaction of repetitive RNA oligonucleotides, many misaligned intermediates will be formed. Right: To explain the fast relaxation to the completely aligned state, Pörschke suggested that the two strands can slide by the propagation of bulge loops from one end to the other.

where each step corresponds to an individual bulge loop, the length of which can be one or several repeat units. We devise a theoretical model, that allows us to predict experimental signatures and relate measurements to microscopic parameters of DNA-slippage. Using kinetic Monte Carlo simulations, methods from statistical mechanics, drift-diffusion, and reaction-diffusion theory, we uncover four different force regimes with distinct characteristic behavior. The model we use is simple enough to be amenable to analytic treatment, which allows us to calculate the threshold forces and the average sliding speed exactly. We further investigate how this sliding behavior is affected by rare mutations that destroy the perfect repetitivity of the sequence. Such mutations do not necessarily impede sliding, but, depending on the frequency of such alien bases, delay the mechanical response and require larger forces.

2.2.1. Sliding dynamics of perfectly repetitive sequences

In a typical force spectroscopy experiment, a force extension curve is recorded until rupture. In such experiments, either the distance of the cantilever from the surface, *i.e.* the extension of the sample, or the applied force is controlled, while the other is recorded. Though in principle the same information can be gained from either of the two approaches, there are significant practical differences between the two. The former is easier to implement experimentally, since distance can be precisely controlled using piezo-elements. However, applying a constant force and monitoring extension is easier to interpret and handle analytically or numerically. In the following, we will study the response of repetitive dsDNA to a constant shear force, as illustrated in Fig. 2.4a.

DNA sliding exhibits four different force regimes

No matter how small forces are applied to the molecule, it will rupture eventually since the state of large separation has the lowest free energy. However, to separate the two strands, an activation barrier has to be overcome. The height of this barrier depends on the applied

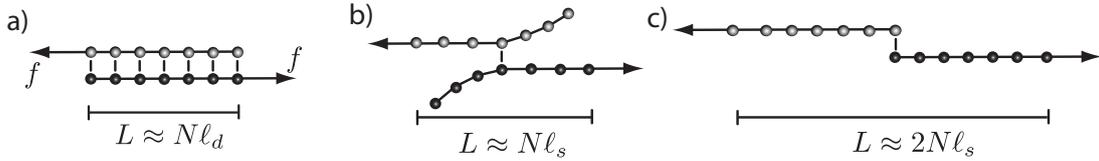


Figure 2.4: Part a): A dsDNA molecule sheared by a force f . Part b): When the DNA duplex ruptures without sliding, the last base pair before rupture will be a native base pair. Part c): If the two sequences slide along each other, the transition state has a larger extension L , see text for details. Shorter duplexes rupture in a cooperative manner [72].

force, as well as on the internal dynamics of the system. In the case of repetitive DNA, the main question is whether the two strands stayed in register or have shifted relative to each other before rupture. Two possible transition states, *i.e.* the state prior to rupture, with and without sliding are sketched in Fig. 2.4 b&c. The free energies difference of these states to the ground states are given by

$$\Delta E_{non-sliding} = N(\varepsilon_b - f(\ell_s - \ell_d)) \quad \text{and} \quad \Delta E_{sliding} = N(\varepsilon_b - f(2\ell_s - \ell_d)). \quad (2.1)$$

The parameters ℓ_s and ℓ_d are effective lengths of ssDNA and dsDNA chosen such that the stretching free energy per base or base pair is given by $f \cdot \ell_s$ and $f \cdot \ell_d$, respectively. Obviously, the transition state after sliding is always lower in free energy than the transition state, when both strands stick. At low force, however, even $\Delta E_{sliding}$ is positive and the dissociation of the two strands is a thermally activated barrier crossing process, no matter which dissociation path is taken. The rupture times are exponentially distributed with a mean time $\langle \tau \rangle$ that increases exponentially with ΔE and hence exponentially with the length of the molecule. The situation changes, when the force is increased to

$$f_c = \frac{\varepsilon_b}{2\ell_s - \ell_d}, \quad (2.2)$$

where the $\Delta E_{sliding}$ ceases to be positive while $\Delta E_{non-sliding}$ is still positive. If the duplex ruptures via sliding the dissociation should no longer be a thermally activated barrier crossing process but some sort of creeping motion from the ground state to the transitionstate. While sliding still involves local energy barriers such as bulge loop formation, there is no longer an extensive barrier. Hence, the mean rupture time no longer increases exponentially with the length but is determined by how rapidly the two strands can move relative to each other. Simulations suggest, that $\langle \tau \rangle$ scales as N^3 at the critical force⁴ f_c . At forces above f_c , we observe a quadratic increase of $\langle \tau \rangle$ with N , see Sec. 2.8. How can the cubic and quadratic scaling be rationalized? As suggested by Pörschke the two strands can be shifted relative to each other by propagation of bulge loops from one end to the other end. But a loop that is nucleated at one end most likely leaves the duplex again at the same end. It travels all the way to the other end only with probability M^{-1} , where M is the number of base pairs (see Fig. 2.5a for illustration). Since the nucleation rate of loops

⁴ This force is actually slightly different from the expression given in Eq. (2.2) due to entropic effects.

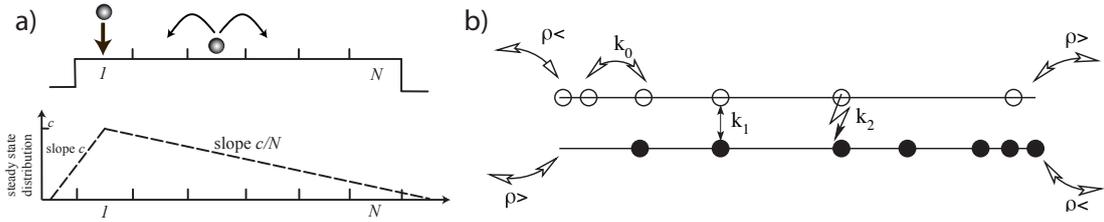


Figure 2.5: Left: A particle placed at site 1 will be exit at site $N + 1$ rather than at site 0 with probability N^{-1} . This can be seen from the steady state distribution resulting when particles are injected at a constant rate. The particle fluxes to the left or right are proportional to the slopes of the distribution, the ratio of which is N^{-1} . Right: A particle-antiparticle model for bulge loop dynamics, see main text.

at the end is independent of the total length, the mobility of the two strands relative to each other is inversely proportional to the overlap length M . This mechanistic explanation of strand mobility is consistent with the intuitive expectation, that the friction coefficient of a one dimensional object should linearly depend on its length. At the critical force, the nucleation rates of loops at stretched or unstretched ends are equal and the duplex shortens and lengthens at equal rates, resulting in an undirected diffusive motion. Since the diffusion constant itself is inversely proportional to N , the time needed to overcome the distance N increases as N^3 . At forces above or below the critical force, bulge loops are nucleated more frequently on unstretched or stretched strands, respectively, than they are on the opposite strand. This induces a directed motion either extending or contracting the duplex. The effective drift velocity is inversely proportional to the overlap length M . The time required to overcome a distance N with a velocity proportional to M^{-1} scales as N^2 .

This quadratic scaling does not persist to arbitrarily high forces, but crosses over to a linear scaling. The threshold force f^* for this crossover is given by the force, at which also the $\Delta E_{non-sliding}$ becomes negative.

$$f^* = \frac{\varepsilon_b}{\ell_s - \ell_d}. \quad (2.3)$$

In this case, it is no longer energetically expensive to open base pairs from both ends. Since consecutive opening of base pairs until rupture is faster than sliding, this mode of unravelling wins over sliding dynamically and repetitive sequences behave similarly to random sequences.

On a macroscopic level, the sliding dynamics of the two strands is well described by a drift-diffusion equation, where the drift and the diffusion coefficient are inversely proportional to the instantaneous length x of the overlap of the two strands at any instant.

$$\frac{\partial}{\partial t} \mathcal{P}(x, t) = \frac{\partial}{\partial x} \left(\frac{D_0(f)}{x} \frac{\partial}{\partial x} - \frac{v_0(f)}{x} \right) \mathcal{P}(x). \quad (2.4)$$

By fitting the solution of this drift-diffusion equation to the simulated rupture time distributions, we obtain the drift and diffusion coefficients $v_0(f)$ and $D_0(f)$ that are independent

of x and depend on the force only. The dependence of the drift coefficient on the force can be understood by microscopic modeling of the bulge loop dynamics. In essence, bulge loops on opposite strands behave as particles and anti-particles, cf. Fig. 2.5b: They annihilate on encounter forming one double stranded repeat unit. Bulge loops can also be produced in pairs, when a spontaneously nucleated bubble separates into two bulges. The nucleation of bulge loops at the ends is mimicked by a coupling to particle reservoirs, the density of which depends on the force applied to that particular end. The difference of particles and antiparticles fluxes is conserved and directly related to the sliding velocity of the two DNA strands: The sliding velocity is given by the difference of the reservoir densities on stretched and unstretched ends, divided by the length of the double stranded region. The reservoir densities are determined by the pseudo-equilibrium loop densities on stretched and unstretched ends, which be calculated from the partition sum of our model. These results and the corresponding plots are included in our publication entitled “Dynamics of Force-Induced DNA-slippage” in *Physical Review Letters* [73], which is reprinted in Sec. 2.8. The details of the calculation of the partition sum, defect densities, and critical forces are presented in the Appendix A.

2.3. Single molecule experiments on DNA-slippage

Our theoretical study was intended to trigger experiments that study DNA-slippage in single molecules. We are very happy, that Ferdinand Kühner and Julia Morfill from the group of Hermann Gaub were willing to perform such experiments and collaborate with us. The experiments show very clearly, that two strands with repetitive DNA can slide along each other once the applied shear force exceeds a certain threshold value. Sliding proceeds in stepwise manner and the observed steps are compatible with a shift by one repeat unit. The observations can be convincingly explained by the force induced formation of a bulge loop which is propagated to the opposite end and thereby lengthens the duplex. The experiments were performed with two different sequences, a 10 fold repeat of GTT and a 15 fold repeat of GT.

In the vicinity of the threshold force f_c for sliding, the sliding velocity is linearly related to $f - f_c$, and a sliding mobility μ can be defined via

$$v(f) = \mu \cdot (f - f_c) \quad (2.5)$$

In the experiments the situation is reversed, as the velocity is imposed by the speed at which the cantilever is retracted and the force adjusts itself. Higher forces at a given speed correspond to higher “friction”, *i.e.* lower mobility. The forces measured at different speed confirm an approximately linear relationship. As expected, the tri-nucleotide repeat slipped slower than the di-nucleotide repeat.

The threshold forces observed in the experiments were considerably higher than expected from theory. We expect this discrepancy to be the result of deformations of the duplex, which is not accounted for by the theory. When a force is applied to one strand of dsDNA it will take a few bases, probably of the order of one helical turn of the DNA, to distribute

the force evenly to both backbones. In this boundary region the double stranded structure is distorted. The DNA sequences used are only 30 bps long and the two boundary regions take up the whole molecule. It is therefore not surprising, that the observed forces deviate from the theoretical estimates which assume an undistorted structure. The short sequences also limited the number of possible sliding steps to four or five⁵. Therefore, the predicted scaling behavior for the mean rupture time could not be tested. Given the biological importance of repetitive sequences and putative applications as active nano-scale building blocks, mechanical properties and the kinetics of repetitive sequences continue to be an interesting and challenging field for single molecule studies. The publication containing these results is reprinted in Sec. 2.9 and the interested reader is referred to this publication for details [1].

2.4. DNA sliding in presence of sequence disorder

After having discussed the sliding dynamics of perfectly repetitive DNA, the question whether sliding is robust to mutations that destroy the repetitive pattern, arises naturally. We show that DNA sliding persists even in presence of such disorder. However, the onset of sliding is delayed by a waiting time, during which all mutated base pairs are opened.

To begin with, we simulated the response to a shear force of a molecule with repetitive sequence where once in a while a repeat unit has been exchanged by bases, that bind only to their native binding partner and not to any other bases in the sequence. We find, that the extension of the molecule remains constant for some time until suddenly a fairly normal sliding behavior sets in. The existence of some delay is obvious, since all mutated bases have to be opened before the molecule yields. But it is less clear how the state with all mutations open is established and how the delay times are distributed. By monitoring the state of individual mutations during the waiting stage, we reveal that mutations open consecutively from both ends and that sliding starts, as soon as the last mutation has opened. The mechanism by which a mutation opens is illustrated in Fig. 2.7.

Since the mutations open from both ends of the molecule, the state of all mutations can be described by the outermost mutations on both sides. If the total number of mutations in the molecule is M , the outermost mutations perform a random walk on $[1, 2, \dots, M]$. Sliding starts, when both of these random walkers meet, that is no more mutations are bound. The rate, at which these random walkers hop, *i.e.* the outermost mutations open and close, depends on the force and the spacing between mutations. At low force, the random walkers are biased away from each other and mutations are preferentially closed. In this case, the waiting time before sliding increases exponentially with the size of the system. At high forces, the mutations are preferentially open and the waiting time increases as a power law [74]. We can therefore identify different dynamical regimes in the plane of mutation density and applied force, where rupture is fast or exponentially slow.

⁵The terminal bases on both ends are opening and closing very frequently and a duplex with fewer than 10 bp overlap rapidly dissociates under force before sliding can be observed [72].

DNA sliding in presence of sequence disorder is treated in detail in our publication entitled “DNA as a Programmable Viscoelastic Nanoelement” in the *Biophysical Journal* [75]. Details of the two random walker model are presented in the supplementary material to this article. The publication and the supplementary are reprinted in Sec. 2.10.

2.5. Repetitive DNA as a visco-elastic nanoelement

In recent years DNA has become a popular material to build elaborate structures on a molecular scale [76, 77, 78]. These applications exploit the specificity of complementary base pairing to guide an ensemble of DNA strands with carefully designed sequence into the desired conformation. DNA has also been used to construct devices that undergo conformational transitions in response to a change in the chemical composition of the environment. A very versatile approach is to construct strands that bind competitively to a scaffold strand with different binding affinities. If the device is in a particular conformation including a weakly binding strand, the addition of a more strongly binding strand will replace the weakly binding strand from the structure and, if designed properly, will result in the desired conformational change [79]. Such competitive binding has also been used to reversibly cross-link acrylamid gels [80]. Other structures are sensitive to variation of pH and such reversible pH-driven transitions have recently been coupled to an oscillatory chemical reaction [81, 82].

Here, we want to discuss briefly the potential of repetitive DNA in nano-mechanical applications. Most applications mentioned above are rigid and conformational transitions occur only between well defined states or require the replacement of one strand by another. Repetitive DNA might be useful when building more flexible devices that respond dynamically to mechanical forces. We have seen above that repetitive DNA lengthens by one repeat unit at a time, if subject to a sufficiently high shear force. Conversely, it contracts against a sub-threshold force until maximal overlap of the two strands is reached. In this way, mechanical energy is reversibly transformed into base pairing energy. Effectively, repetitive DNA acts as a contractile visco-elastic element with a viscosity that can be programmed by choice of the length of the individual repeat unit and the overall length of the molecule. The contractile force is determined by the sequence composition. Such a visco-elastic element might find applications as a molecular tie-rope that keeps a connection between two parts straight and at the same time adjusts its length. Another conceivable application is a molecular force sensor that responds to forces exceeding the threshold force. The read out signal could either be rupture or simply a relative shift of the two strands measured by FRET. The sliding response can be delayed by mutations in the repetitive sequence that transiently lock the two strands in a particular relative position. Cross-linking gels with repetitive DNA might also result in material with novel mechanical properties. A detailed characterization of the visco-elastic properties of repetitive DNA can be found in the article reprinted in Sec. 2.10 [75].

2.6. Intermediate phase in DNA-melting

So far, we have been predominantly interested in dynamic features of repetitive DNA, which proved to be much richer than DNA with sequences without a particular order. Here, we show that not only the dynamics but also equilibrium properties of repetitive DNA are different from random DNA.

The most prominent difference is, at least within Poland-Scheraga models, that the order of the melting transition is different for repetitive sequences than for ordinary sequences. We have seen in Sec. 1.2.2 that the order of the melting transition of DNA depends on the entropy of denatured loops, which in turn is governed by the loop closure exponent c . For random sequences, a loop of a given size at a given position corresponds to one base pairing configuration. When sequences are repetitive, however, loops can be asymmetric and the bases of a loop of size n can be distributed between the two strands in $n + 1$ ways. This effectively reduces c by one [83]. Hence, no transition is observed if c is smaller than 2, denaturation is continuous if $2 \leq c < 3$ and a first order transition is observed only, if $c \geq 3$. A brief discussion of physical values of c was given in Sec. 1.2.2.

This argument can be formalized by calculating the partition sum of all possible pairings between the two strands⁶, as we already did to study DNA sliding. When the two strands have repetitive sequences, there is no reason to consider only strands of equal lengths. Our study revealed an additional phase transition, which occurs when the two strands are of different length. At low temperatures, the two strands form a rigid double helix and the excess bases of the longer strand reside in a single stranded overhang. As the temperature rises, more and more of these unbound bases are absorbed into bulge loops within the double helix and the overhang becomes shorter. At a certain temperature all the bases are absorbed and the length of the overhang is no longer extensive. This transition is a continuous phase transition which formally and conceptually resembles Bose-Einstein condensation. The overhang corresponds to those particles that condense into the ground state, while bulge loops within the double helix correspond to populated excited states. The intermediate phase persists in presence of weak sequence disorder. Our work on phase transitions in repetitive dsDNA is published in *Physical Review E* [84], which is reprinted in Sec. 2.11.

2.7. Conclusion & Outlook

The dynamics of repetitive DNA is a fascinating research area with many open questions remaining to be addressed. Due to their structural simplicity, such sequences are amenable to methods from statistical mechanics and their generic properties can be elucidated without reference to a particular realization. We studied equilibrium and dynamical features of repetitive DNA using both analytical and computational tools. Many quantities such as threshold forces, defect densities and phase diagrams can be calculated exactly within our

⁶Only pairings corresponding to alternating stems and loops (no crossing base pairs) are allowed. This is usually a good assumption due to steric constraints.

model. When repetitive DNA is sheared with sufficiently high forces the two strands start moving relative to each other. This motion is mediated by propagation of bulge loops from one end to the other, very similar to defect diffusion in crystals. The dynamics of the two strands can be described by a drift-diffusion equation with drift and diffusion coefficients that are inversely proportional to the length of the double stranded overlap of the two single strands. The drift and diffusion coefficients can be related to the microscopic bulge loop dynamics using a reaction-diffusion model. These two different levels of description provide a link between the microscopic dynamics of a bulge loop inside double stranded DNA to the DNA sliding dynamics, that can be measured in single molecule experiments. First experiments confirm that two repetitive DNA strands slide relative to each other when sheared [1], but more experiments with longer strands are necessary to test the prediction of different scaling regimes and to infer quantities such as bulge loop mobilities.

From an engineering perspective, repetitive DNA has intriguing mechanical properties which could be exploited in nano-scale devices. In essence, repetitive DNA acts as a visco-elastic element with a force offset. The characteristics of such an element can be chosen through sequence composition and length. By introducing mutations in the perfectly repetitive sequence, the response of the element can be delayed in a controlled way.

The principle incentive to study repetitive DNA is to understand the evolutionary dynamics of simple sequence repeats. Our research is focussed on microscopic properties of DNA under well controlled conditions. On the other hand, the length distribution of SSRs in various genomes and the mutation dynamics *in vivo* are being actively investigated. The gap between these two approaches is huge, but I think that bridging this gap is not completely infeasible. After measuring the microscopic rates in single molecule experiments, one can faithfully parameterize models for *in vitro* slippage as described in Sec. 2.1.5 [68]. By successively adding a mismatch repair system, single stranded binding proteins, etc., it might be possible to understand how these components work together *in vivo*. Another possible role of DNA-slippage could be in prokaryotic transcription termination. Intrinsic terminator sequences in prokaryotes include a “slippery” and weakly binding poly-A stretch, usually 7 to 9 bases in length. Upstream of this of this slippery sequence is a palindromic sequence which forms a hairpin in the RNA transcript [85]. A popular hypothesis is, that this hairpin exerts a force on the RNA still inside the polymerase and thereby terminates transcription [86]. The precise mechanism how this is happening is unclear, but it is conceivable that the force exerted by the hairpin induces DNA-slippage, which then enables the RNA to slide along the DNA out of the polymerase. It should be possible to address this question by pulling on the nascent RNA strand as presented in ref. [86] while having the polymerase transcribe repetitive and non-repetitive sequences with different binding energies.

Dynamics of Force-Induced DNA Slippage

Richard A. Neher* and Ulrich Gerland†

Department of Physics and CENS, LMU München, Theresienstrasse 37, 80333 München, Germany
(Received 7 July 2004; published 1 November 2004)

We study the base pairing dynamics of DNA with repetitive sequences where local strand slippage can create, annihilate, and move bulge loops. Using an explicit theoretical model, we find a rich dynamical behavior as a function of an applied shear force f : reptationlike dynamics at $f = f_c$ with a rupture time τ scaling as N^3 with its length N , drift-diffusion dynamics for $f_c < f < f^*$, and a *dynamical* transition to an unraveling mode of strand separation at $f = f^*$. We predict a viscoelastic behavior for periodic DNA with time and force scales that can be *programmed* into its sequence.

DOI: 10.1103/PhysRevLett.93.198102

PACS numbers: 87.15.-v, 05.10.Gg, 87.80.Fe

The dynamics of base pairing in DNA and RNA molecules plays an important role in biological processes such as DNA replication, transcription, and RNA folding [1]. These dynamics can be probed in detail with modern single molecule techniques to exert and measure piconewton forces with nanometer spatial resolution [2]. For instance, double-stranded DNA (dsDNA) can be forced to open either by pulling on the two strands from the same end of the dsDNA (“unzipping”) [3–5] or from opposite ends (“shearing”) [6]. In the case of unzipping, the dynamics involves the consecutive opening of *native* base pairs, i.e., those present in the ground state of the molecule, and is well understood theoretically [7]. Here, we consider instead the shearing of dsDNA and focus specifically on *periodic* DNA sequences. This case is particularly interesting both from a physical and a biological point of view, since (i) periodic sequences have many non-native base pairing conformations where one strand is shifted with respect to the other, (ii) shearing probes the transitions between such states, i.e., the dynamics of DNA *slippage*; see Fig. 1, and (iii) DNA slippage during genome replication allows the expansion of nucleotide repeats, and, for certain repeats inside genes, triggers a variety of diseases including Huntington’s disease [8].

The mechanism for DNA slippage has already been suggested by Pörschke [9]; see Fig. 1(a): small bulge loops can form at one end of the molecule when a few bases spontaneously unbind and rebind shifted by one or several repeat units. Once formed, a bulge loop may diffuse along the molecule and anneal at the other end, effectively sliding the two strands against each other by a length equal to the size of the bulge loop. This mechanism involves only small energetic barriers compared to the large barrier for complete unbinding and reassociation. Here, we present a detailed theoretical study of *force-induced* DNA slippage, which has so far not been studied experimentally. We show that this system displays a rich dynamical behavior that can be controlled experimentally by adjusting the force, sequence length, and sequence composition.

Model.— We consider a dsDNA of two perfectly complementary periodic sequences with N repeat units, each consisting of m nucleotides (for simplicity, we refer to repeat units also as “bases”). Assuming that base pairing within a strand is negligible, a base pairing configuration is specified by the set of the $n \leq N$ interstrand base pairs $S = \{(u_i, l_i)\}$ with $1 \leq u_1 < u_2 < \dots < u_n \leq N$ for the “upper” strand and analogously for the l_i in the “lower” strand. We assign a binding energy $-\varepsilon_b < 0$ to each base pair and a loop cost $E_\ell(j) > 0$ when there are $j > 0$ unpaired bases (total on both strands) between two consecutive base pairs. With a constant shear force f , see Fig. 1(b), the energy of a configuration S is

$$E[S] = -\varepsilon_b n[S] + \sum_{i=2}^{n[S]} E_\ell(\Delta u_i + \Delta l_i - 2) - fL[S], \quad (1)$$

where $\Delta u_i = u_i - u_{i-1}$ and $\Delta l_i = l_i - l_{i-1}$. The loop cost $E_\ell(j)$ increases with the loop length, starting from $E_\ell(0) = 0$. Free DNA ($f = 0$) is described by $E_\ell(j) = \varepsilon_\ell + 3\nu k_B T \ln(j)$, with a loop initiation cost $\varepsilon_\ell > 0$ and a logarithmic asymptotic behavior derived from polymer theory ($\nu \approx 0.6$ is the Flory exponent) [10]. An applied force can affect $E_\ell(j)$; however, our qualitative results are insensitive to its precise form [11]. Unless stated otherwise, we keep only the constant term, $E_\ell(j > 0) = \varepsilon_\ell$, for simplicity. The total extension L is

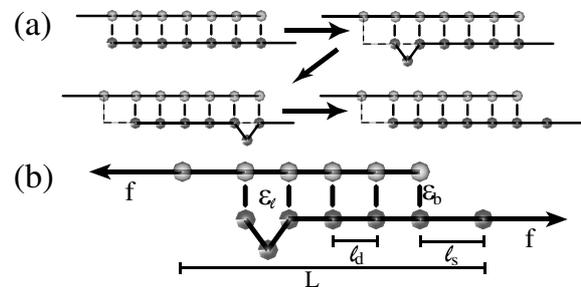


FIG. 1. Sketch of periodic dsDNA, where each bead represents one repeat unit consisting of one or several bases. (a) Many microscopic slippage events can lead to macroscopic sliding. (b) An applied shear force.

$$L[S] = \ell_s(u_1 - 1 + N - l_n) + \ell_d \sum_{i=2}^n \min(\Delta u_i, \Delta l_i), \quad (2)$$

where ℓ_d and $\ell_s > \ell_d$ are the effective lengths (in the direction of the force) per single and double-stranded unit, respectively. The entropic elasticity of DNA [12] causes both ℓ_d and ℓ_s to depend on the applied force; however, since the DNA is almost fully stretched at the forces of interest here, we use the constant values $\ell_d/m = 3.4 \text{ \AA}$ and $\ell_s/m = 7 \text{ \AA}$ for simplicity [13].

We study the dynamics of our model with analytical methods (see below) and a Monte Carlo approach using three elementary moves [14]: opening, closing, and slippage of a base pair; i.e., a pair (u_i, l_i) is removed from the set S or added to it, or, if the base pair is adjacent to a loop, either u_i or l_i can be changed to another base inside the loop. The absolute time scale of these dynamics is hard to predict, but comparison with bulk reannealing experiments [9] suggests that our simulation time step is on the order of μs in real time.

Scaling of mean rupture times.— With a constant applied force $f > 0$, eventually every finite dsDNA will rupture, since complete separation of the strands ($L \rightarrow \infty$) is the state of minimal free energy. However, both the time scale and the nature of the rupture dynamics depend drastically on the force. Figure 2 displays the scaling of the mean rupture time $\langle \tau \rangle$ with the number of bases N for a number of different forces (see caption for parameters). We observe four distinct asymptotic behaviors: an exponential increase with N for small forces, a cubic scaling

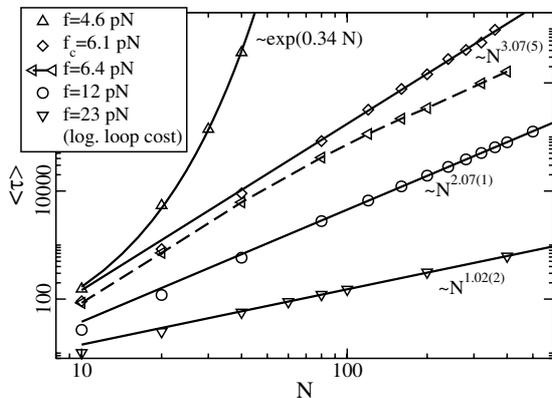


FIG. 2. Scaling of the mean rupture time $\langle \tau \rangle$ with the number of bases N for different shear forces (with $\varepsilon_b = 1.11$, $\varepsilon_\ell = 2.8$, which roughly corresponds to adenine-thymine (AT) sequences at 50°C ; see Fig. 5). The symbols represent Monte Carlo data (error less than symbol size). The solid lines for $f \geq f_c$ are power law fits (exponent with error in least significant digit is given; data with $N \leq 40$ show significant finite size deviations and are excluded). For $f < f_c$ the rupture time increases exponentially. The data for $f = 6.4 \text{ pN} \geq f_c$ (connected by the dashed line) demonstrate the crossover from diffusive to drift behavior; see main text. The data for $f = 23 \text{ pN}$ are calculated including the logarithmic loop cost, which becomes relevant at large forces [11].

with N at a certain force f_c , a nearly quadratic scaling above f_c but below a second threshold f^* , and linear scaling above f^* . The behavior in the two extremes is easily interpreted: for small f , rupture is driven by thermal fluctuations across a large free energy barrier with an associated Kramers time that scales exponentially with N , and linear scaling at large f is expected when individual bonds break sequentially at a constant rate. We now characterize the rich behavior in the intermediate force regime, including the nature of the two transitions.

The thermodynamic energy barrier disappears at a force f_c , which can be estimated by balancing the binding energy per base pair with the mechanical work exerted when sliding both strands against each other by one step,

$$f_c \approx \varepsilon_b / (2\ell_s - \ell_d). \quad (3)$$

f_c is a critical force in the thermodynamic sense, if the state of complete rupture is excluded (see below for the exact calculation including all base pairing configurations). At $f = f_c$, the rupture dynamics is best understood by analogy with the reptation problem [15], since bulge loops in the DNA structure behave similarly to the “stored length” excitations of a single chain in a polymer network: these excitations are generated at the ends of the polymer with constant rate independent of N , diffuse along the polymer, and reach the other end with a probability $\sim N^{-1}$. Therefore, the macroscopic diffusion constant for the relative motion of the two DNA strands should scale as $D \sim N^{-1}$ and the time for diffusion over distance N is $\sim N^3$.

For $f > f_c$, strand separation is energetically a downhill process, which induces a drift velocity v between the two strands. In linear response, we expect $v = \mu \Delta f$ for small $\Delta f = f - f_c$ with a mobility mediated by bulge loop diffusion, $\mu = D/k_B T \sim N^{-1}$ (from the Einstein relation), leading to $\langle \tau \rangle \sim N^2$. Why does this behavior not persist for large forces? The second transition in the scaling behavior is due to a change in the rupture *mode*: at forces larger than $f^* \approx \varepsilon_b / (\ell_s - \ell_d)$, the double strand can open by *unraveling* from both ends; i.e., the energy cost ε_b of opening a base pair at the end is outweighed by the gain $f(\ell_s - \ell_d)$ from a longer base-to-base distance in the single strand. In this unraveling mode, the rupture time scales linearly with N . The dynamical transition from sliding to unraveling is clearly reflected in the length at rupture, $L[S(\tau)]$; see Fig. 3(a), which is roughly a factor of 2 larger for sliding.

Rupture time distributions.— Single molecule setups are ideally suited to record the full distribution of rupture times, $P(\tau)$, which is a sensitive characteristic of the dynamics and permits a close examination of the physical picture introduced above. The histograms in Fig. 4 show $P(\tau)$ from simulations at $f = f_c$ and a larger force $f_c < f < f^*$; see caption for parameters. We observe that fluctuations play a dominant role at $f = f_c$, i.e., the width of $P(\tau)$ is comparable to the mean, while the rupture

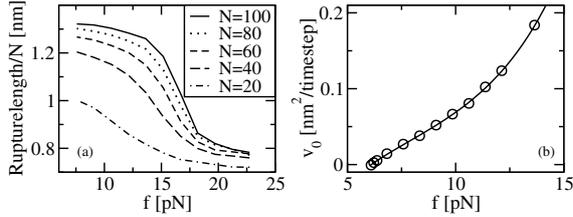


FIG. 3. (a) Rupture length as a function of applied force f (parameters as in Fig. 2). (b) Drift coefficient $v_0(f)$ extracted from simulations with $N = 150$ (circles) and analytical curve (solid line, $k_0 = 1.87$); see main text.

dynamics is drift dominated at the larger force, with a localized peak in $P(\tau)$.

To formulate the drift-diffusion dynamics quantitatively, we treat the number of bases in the double-stranded region as a continuum variable x with $0 < x < N$, and consider the probability distribution $\mathcal{P}(x, t)$, which satisfies the continuity equation $\partial_t \mathcal{P}(x, t) = -\partial_x j(x, t)$ with a force-dependent current

$$j(x, t) = -D(f, x) \partial_x \mathcal{P}(x, t) - v(f, x) \mathcal{P}(x, t). \quad (4)$$

The above discussion suggests a diffusion coefficient of the form $D(f, x) = D_0(f)/x$ and similarly a drift $v(f, x) = v_0(f)/x$. We have an absorbing boundary at $x = 0$ and it is natural to choose a reflecting boundary at $x = N$ and a delta peak at $x = N$ as initial condition. The solution $\mathcal{P}(x, t)$, which must in general be obtained numerically, determines the rupture time distribution through $P(\tau) = j(0, \tau)$.

We can determine the force dependence of the diffusion coefficient and drift empirically by fitting the calculated $P(\tau)$ to the simulation data using D_0 and v_0 as adjustable parameters. The solid lines in Fig. 4 show that the drift-diffusion theory describes the simulation data well. Figure 3(b) shows the fitted v_0 as a function of f (circles). The drift vanishes at the critical force, $v_0(f_c) = 0$, confirming the physical picture. The drift-diffusion theory also explains the crossover behavior in the vicinity of $f = f_c$; see Fig. 2: the drift is significant only when the system size N is larger than the diffusive length D_0/v_0 [16]. Hence, with $v_0 \sim \Delta f$, reptationlike dynamics is expected in a force interval $\delta f \sim N^{-1}$ around f_c .

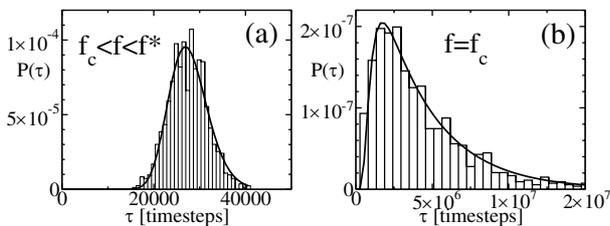


FIG. 4. Histogram of rupture times for two different forces, (a) 34.4 pN and (b) $f_c = 16.6$ pN, but the same set of DNA parameters, $N = 80$, $\varepsilon_b = 3.75$, $\varepsilon_\ell = 2.6$, which roughly correspond to a CG sequence at room temperature; see Fig. 5.

Microscopic dynamics.— Next, we study how the macroscopic drift in Eq. (4) emerges from the microscopic bulge loop dynamics and determine $v_0(f)$ in terms of our system parameters. Since bulge loops on opposite strands annihilate each other when they meet, the bulge loop dynamics is equivalent to a reaction-diffusion system of particles and antiparticles in one dimension. Both particles and antiparticles are created at each end, however, with different rates determined by the applied force. From the underlying master equation for these processes one obtains the mean-field equations [11]

$$\begin{aligned} \partial_t u(y, t) &= k_0 \partial_y^2 u(y, t) - k_1 u(y, t) l(y, t) + k_2, \\ \partial_t l(y, t) &= k_0 \partial_y^2 l(y, t) - k_1 u(y, t) l(y, t) + k_2, \end{aligned} \quad (5)$$

where $u(y, t)$ and $l(y, t)$ denote the bulge loop density on the upper/lower strand, $y \in [0, x]$ is the position within the double-stranded region, and k_0, k_1, k_2 are the rates for hopping, annihilation, and pair creation, respectively. At the boundaries, the densities take on constant values, $u(0, t) = l(x, t) = \rho_<$ and $u(x, t) = l(0, t) = \rho_>$, where $\rho_<(f)$ and $\rho_>(f)$ are calculated below by assuming a local equilibrium of the DNA at the edges. The macroscopic drift is determined by the stationary solution and depends only on the difference between the loop densities on the upper/lower strand, $v(f, x) = k_0 \partial_y [u(y) - l(y)]$. Using Eq. (5), this yields $v_0(f) = 2k_0 [\rho_>(f) - \rho_<(f)]$. Figure 3(b) shows that this result is in excellent agreement with the empirical $v_0(f)$ obtained above.

Since the loop cost $E_\ell(j)$ is larger for two separate loops than for a single one of the combined length, bulge loops on the same strand feel a short-range attraction. However, the interaction is not strong enough to cause a significant aggregation of the loops in our Monte Carlo simulations. This is consistent with the observation that with our DNA parameters, the interaction energy ε_ℓ is never significantly larger than the entropic cost $\sim \log \rho$ of colocalization at loop density ρ . While $v_0(f)$ is apparently robust to interaction effects, the diffusion coefficient $D_0(f)$ is sensitive to interactions as well as correlations. Both are neglected in Eq. (5), leaving the calculation of $D_0(f)$ as a challenge for the future.

Critical force.— To obtain the exact critical force, we need the partition function $Z = \sum e^{-E[S]}/k_B T$ summed over all configurations S with at least one base pair. It is useful to allow for different numbers of bases in the two strands, e.g., $1 \leq u_i \leq N$ and $1 \leq l_i \leq M$, with a corresponding partition function

$$Z(N, M) = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} b_s^{i+j} \sum_{n=1}^{N-i} \sum_{m=1}^{M-j} Z_p(n, m), \quad (6)$$

where $b_s = e^{f \ell_s / k_B T}$ is the Boltzmann factor for a stretched base, and $Z_p(n, m)$ is the partition function for the central, double-stranded section starting with the first and ending with the last base pair cf. Figure 1(b). We calculate $Z_p(n, m)$ recursively by introducing a comple-

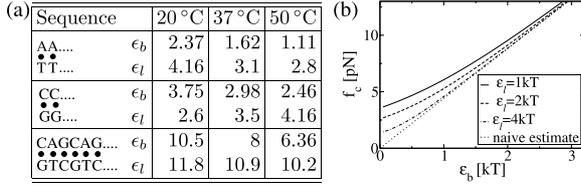


FIG. 5. (a) Model parameters for different DNA sequences and temperatures as obtained by fitting to a detailed thermodynamic model [11,18] (all energies in units of kT). (b) The exact critical force compared to the estimate of Eq. (3).

mentary partition function $Z_u(n, m)$ containing only structures where the last of the n upper bases is *not* bound to the last of the m lower bases:

$$Z_p(n+1, m+1) = qb_d Z_p(n, m) + qb_d g Z_u(n, m),$$

$$Z_u(n+1, m+1) = g \sum_{k=1}^n Z_p(k, m+1) + g \sum_{k=1}^m Z_p(n+1, k) + gb_d Z_p(n, m) + b_d Z_u(n, m). \quad (7)$$

Here, the Boltzmann factors $q = e^{\epsilon_b/k_B T}$, $g = e^{-\epsilon_t/2k_B T}$, and $b_d = e^{f \ell_d/k_B T}$ account for base pairing, loop costs, and stretching of double strand, respectively. To obtain the critical behavior for $N \rightarrow \infty$, we take the z transform $\hat{Z}(z, y) = \sum_{N, M} Z(N, M) z^N y^M$. The inverse z transform is then determined by the simultaneous poles of $\hat{Z}(z, y)$ in z and y . For large N , the pair of poles with the smallest $|zy|$ dominates. A detailed analysis of the critical behavior will be presented elsewhere [11]; here we are interested in f_c , the force where the dominant pole switches. We find that f_c is the nontrivial root of

$$\left(\frac{b_s^2}{b_d} - q\right)\left(\frac{b_s^2}{b_d} - 1\right) - g^2 q \left(\frac{2}{b_s - 1} \frac{b_s^2}{b_d} + 1\right) = 0. \quad (8)$$

When ϵ_b or $\epsilon_t \gg k_B T$, the second term is negligible and the nontrivial root of (8) is $b_s^2/b_d = q$, recovering the naive estimate (3). However, for smaller ϵ_b , ϵ_t one finds significant deviations from (3); see Fig. 5.

Loop densities.— Using the same approach as above, we can calculate the loop densities $\rho_<$, $\rho_>$ introduced above. Assuming equilibration between all possible conformations of the two strands with a fixed central base pair, we find $\rho_< = \sum_{a,b} P(a, b) a / \lambda$ and $\rho_> = \sum_{a,b} P(a, b) b / \lambda$, where $\lambda = \min(a, b) + 1$ and $P(a, b) = b_s^{b-a} b_d^a q g Z_p(N - b - 1, N - b - 1) / Z_p(N, N)$. The sums can be evaluated exactly for large N [11].

Conclusions.— We find a response of periodic dsDNA to shear forces that is very distinct from that for nonperiodic sequences. Above a thermodynamic critical force f_c , but below a *dynamic* critical force f^* , bulge loop diffusion allows periodic DNA to open by *sliding*. This mechanism leads to a much lower thermodynamic critical force than the *unraveling* mechanism by which nonperiodic DNA opens. Within our model, we have calculated f_c exactly and characterized the associated

dynamics, which is effectively *viscoelastic* with a creep compliance $\sim N^{-1}$ for $f_c < f < f^*$. Above f^* , periodic dsDNA also opens predominantly by unraveling (this dynamical transition may be regarded as a remnant of the thermodynamic transition for nonperiodic sequences). Interestingly, periodic DNA could be used as a viscoelastic nanomechanical element with properties that are *programmable* by choosing sequence length and composition. This may lead to applications in microstructured devices, similar to the programmable DNA-based force sensors reported in Ref. [17].

We thank T. Hwa, F. Kühner, and M. Rief for fruitful discussions and the DFG for financial support.

*Electronic address: Richard.Neher@physik.lmu.de

†Electronic address: Ulrich.Gerland@physik.lmu.de

- [1] B. Alberts *et al.*, *Molecular Biology of the Cell* (Garland, New York, 2002).
- [2] C. Bustamante, J. C. Macosko, and G. J. Wuite, *Nat. Rev. Mol. Cell Biol.* **1**, 130 (2000); H. Clausen-Schaumann, M. Seitz, R. Krautbauer, and H. E. Gaub, *Curr. Opin. Chem. Biol.* **4**, 524 (2000); R. Merkel, *Phys. Rep.* **346**, 343 (2001); R. Lavery, A. Lebrun, J.-F. Allemand, D. Bensimon, and V. Croquette, *J. Phys. Condens. Matter* **14**, R383 (2002).
- [3] C. Danilowicz *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 1694 (2003).
- [4] R. Krautbauer, M. Rief, and H. E. Gaub, *Nano Lett.* **3**, 493 (2003).
- [5] U. Bockelmann, B. Essevaz-Roulet, and F. Heslot, *Phys. Rev. E* **58**, 2386 (1998).
- [6] T. Strunz, K. Oroszlan, R. Schäfer, and H.-J. Güntherodt, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 11277 (1999).
- [7] D. K. Lubensky and D. R. Nelson, *Phys. Rev. Lett.* **85**, 1572 (2000); *Phys. Rev. E* **65**, 031917 (2002).
- [8] S. T. Lovett, *Mol. Microbiol.* **52**, 1243 (2004); L. S. Kappen *et al.*, *Biochemistry* **42**, 2166 (2003).
- [9] D. Pörschke, *Biophysical Chemistry* **2**, 83 (1974).
- [10] T. Hwa, E. Marinari, K. Sneppen, and L. Tang, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 4411 (2003).
- [11] R. Neher and U. Gerland (to be published).
- [12] S. B. Smith, Y. Cui, and C. Bustamante, *Science* **271**, 795 (1996).
- [13] Here we assume the B-DNA form. Before melting, DNA with a high cytosine-guanine (CG) content undergoes a transition to a stretched S-DNA form at large forces. This transition was found to be rapid compared to melting [H. Clausen-Schaumann *et al.*, *Biophys. J.* **78**, 1997 (2000)], so that S-DNA could be treated by different effective parameters.
- [14] C. Flamm, W. Fontana, I. Hofacker, and P. Schuster, *RNA* **6**, 325 (2000).
- [15] P.-G. de Gennes, *J. Chem. Phys.* **55**, 572 (1971).
- [16] D. K. Lubensky and D. R. Nelson, *Biophys. J.* **77**, 1824 (1999).
- [17] C. Albrecht *et al.*, *Science* **301**, 367 (2003).
- [18] M. Zuker, *Nucleic Acids Res.* **31**, 3406 (2003).

Force-Induced DNA Slippage

Ferdinand Kühner,* Julia Morfill,* Richard A. Neher,[†] Kerstin Blank,* and Hermann E. Gaub*

*Chair for Applied Physics and Center for NanoScience, and [†]Arnold Sommerfeld Center for Theoretical Physics and Center for Nanoscience, Ludwig-Maximilians Universität München, Munich, Germany

ABSTRACT DNA containing repetitive sequences displays richer dynamics than heterogeneous sequences. In the genome the number of repeat units of repetitive sequences, known as microsatellites, may vary during replication by DNA slippage and their expansion gives rise to serious disorders. We studied the mechanical properties of repetitive DNA using dynamic force spectroscopy and found striking differences compared with ordinary heterogeneous sequences. Repetitive sequences dissociate at lower forces and elongate above a certain threshold force. This yield force was found to be rate dependent. Following the rapid stretching of the DNA duplex, the applied force relaxes by stepwise elongation of this duplex. Conversely, contraction of the DNA duplex can be observed at low forces. The stepwise elongation and shortening is initiated by single slippage events, and single-molecule experiments might help to explain the molecular mechanisms of microsatellites formation. In addition to the biological importance, the remarkable properties of repetitive DNA can be useful for different nanomechanical applications.

INTRODUCTION

Not only is DNA the key molecule for life, it has also become an extremely versatile tool kit for man made nanoscale structures and devices (1). Despite the fact that structure and dynamics of DNA were studied extensively, many of the discovered intramolecular processes, which exhibit complex dynamics and a distinct biological function, still lack satisfactory explanation. Microsatellites formation and bulge loop propagation in repetitive sequences are prominent examples (2). Two complementary DNA strands with heterogeneous sequences can only bind in a well-defined, unique conformation. Thermodynamic fluctuations lead to excitations in the double-stranded DNA, which results in a fast opening and closing of short stretches of basepairs (3,4). These fluctuations are localized and do not propagate through the DNA duplex.

In contrast, double-stranded DNA, containing short repetitive sequences, so-called microsatellites, displays a more complex dynamic behavior (5–9) with potential applications in nanotechnology. Two complementary strands can hybridize in various different conformations in which sufficiently long stretches are aligned to build up thermodynamically stable structures. Rapid transitions between these different conformations may occur. This so-called bulge loop formation and propagation (see Fig. 1) is called DNA slippage. It is considered to play a central role in the evolution of microsatellites, which can be found throughout the genome (10).

The repeat units of these microsatellites usually consist of one to six bases, e.g., (A)_N, (GT)_N, or (GTT)_N. The corresponding length of the microsatellites, i.e., the number of consecutive identical repeat units, *N*, changes rapidly in evolution, presumably due to DNA slippage events during

replication. Because of this length variability, the microsatellites frequently are used as genetic markers, e.g., for forensic purposes, or to determine the genetic similarity between different populations. On the other hand, certain human neurodegenerative diseases, such as fragile X or Chorea Huntington, are related to expansions of trinucleotide repeats of microsatellites beyond certain thresholds (11).

Investigations of DNA slippage in vitro (2,12) showed that DNA bulge loop formation at the end of the duplex occurs on a timescale of microseconds. As a result, the two strands can be shifted relative to each other by propagation of a bulge loop toward the other end of the duplex (compare Fig. 1). In recent years, single molecule techniques have been used to study the mechanical properties of single DNA molecules. For example, the elasticity (13–16) and unzipping of λ -phage DNA (17,18), the interactions between proteins and double-stranded DNA (19,20) and the dissociation forces of short DNA duplexes (21–23) have been measured using different experimental setups. Recently, a theoretical work suggested studying DNA slippage with an atomic force microscope (AFM) (24): The two complementary strands of a DNA duplex with a repetitive sequence are predicted to move relative to each other if the externally applied force exceeds a critical force, the slipping threshold f_c . This slipping threshold can be estimated by balancing the work performed by the external force with the binding free energy, which is lost if both strands are shifted relative to each other by one repeat unit. When shifting the two strands, the contour length of the single-stranded parts of the duplex elongates by twice the length of one repeat, whereas the double-stranded part shortens by one repeat unit. This simple argument leads to the slipping threshold f_c :

$$f_c = \frac{\epsilon_b}{2l_s - l_d}, \quad (1)$$

Submitted August 24, 2006, and accepted for publication December 1, 2006.

Address reprint requests to Ferdinand Kühner, E-mail: ferdinand.kuehner@physik.uni-muenchen.de.

© 2007 by the Biophysical Society

0006-3495/07/04/2491/07 \$2.00

doi: 10.1529/biophysj.106.095836

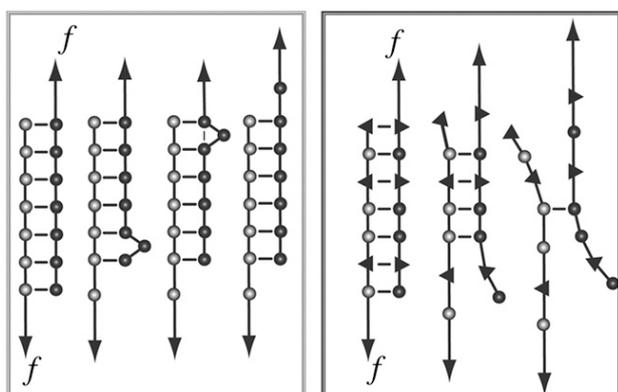


FIGURE 1 Comparison of the behavior of repetitive (*left*) and heterogeneous (*right*) DNA sequences under an externally applied force. Repetitive DNA sequences can form bulge loops. These bulge loops can propagate to the other end of the DNA duplex and therefore cause a lengthening of the molecule. In contrast, this dissociation path is not available for heterogeneous DNA sequences. Heterogeneous DNA sequences simply dissociate in an all or none mode.

where ε_b is the binding free energy of one repeat unit, l_s is the effective length of one unit when unbound and stretched by the force f , and l_d the length of the hybridized repeat unit. For a trinucleotide (GTT) the basepairing energy is $\varepsilon_b \approx 7\text{--}8 k_B T$, the length of three basepairs in the duplex is $l_d \approx 1$ nm, and the effective length of three single-stranded bases is $l_s \approx 1.5$ nm. Inserting these values in Eq. 1 a slipping threshold f_c of roughly 15 pN can be predicted. For a dinucleotide (GT) f_c is roughly 13 pN, with $l_s \approx 1$ nm, $l_d \approx 0.7$ nm, and $\varepsilon_b \approx 4\text{--}5.5 k_B T$. However, this estimate has to be taken with care, since the deformation of the duplex and finite size effects will affect the true value of f_c . The rate of this motion of the two DNA strands relative to each other is determined by the diffusion of bulge loops from one end of the strand to the other (see Fig. 1). The slipping process can be characterized with the following parameters: the slipping rate, which describes the speed of the movement of the bulge loops along the DNA duplex; the slipping length, which describes the length increase or decrease that is determined by the number of bases in one repeat unit; and the slipping threshold, which describes the critical force for the appearance of slipping.

Here we report on an investigation on the response of short DNA duplexes to an externally applied shear force and

compare these repetitive sequences with heterogeneous sequences with respect to their slipping rate, slipping length, and slipping threshold with the intent to test the concept of bulge loop mediated elongation. The dependence of the dynamics on the number of repeat units and the number of bases in one repeat unit is investigated.

MATERIALS AND METHODS

Oligonucleotides modified with a thiol group at the 5'-terminus (for details see Table 1; IBA GmbH, Göttingen, Germany; Metabion GmbH, Martinsried, Germany) were immobilized on amino-functionalized surfaces using a heterobifunctional poly(ethylene glycol) (PEG) spacer. One oligonucleotide was immobilized on the cantilever and the complementary sequence was coupled to the surface. Note that such a chemical functionalization leaves the molecule the freedom to rotate because of the single covalent bonds in the PEG chain. The cantilevers (Bio-lever, Olympus, Tokyo, Japan) were cleaned and functionalized as described previously (25). Instead of epoxy-functionalized cantilevers, amino-modified surfaces on the cantilevers were prepared using 3-aminopropyl-dimethylethoxysilane (ABCR GmbH, Karlsruhe, Germany). Commercially available amino-functionalized slides (Slide A, Nexterion, Mainz, Germany) were used.

From this step on, the surfaces of cantilever and slide were treated in parallel as described in Blank et al. (26). They were incubated in borate buffer pH 8.5 for 1 h. This step was necessary to deprotonate the amino groups for coupling to the *N*-hydroxysuccinimide groups (NHS) of the heterobifunctional NHS-PEG-maleimide (MW 5000 g/mol; Nektar, Huntsville, AL). The PEG was dissolved in a concentration of 50 mM in borate buffer at pH 8.5 and incubated on the surfaces for 1 h. In parallel, the oligonucleotides were reduced using TCEP beads (Perbio Science, Bonn, Germany) to generate free thiols. After washing with ultrapure water, a solution of the oligonucleotides (1.75 μM) was incubated on the surfaces for 1 h. Finally, the surfaces were rinsed with phosphate buffered saline (PBS) to remove noncovalently bound oligonucleotides and stored in PBS until use.

All force measurements were performed with a MFP-3D atomic force microscope (AFM) (Asylum Research, Santa Barbara, CA) at room temperature in PBS. Force-clamp, distance jump experiments and analysis of the data were carried out in Igor 5.3 with self-written procedures. Cantilever spring constants were determined by thermal calibration (6–8 pN/nm).

RESULTS

For the AFM experiments, the complementary DNA strands were covalently anchored via poly(ethylene glycol) (PEG) spacers. One strand was bound to the surface of a glass slide (26) and the complementary strand was coupled to the cantilever tip, respectively.

TABLE 1 DNA sequences

DNA duplex	Sequence (cantilever)	Sequence (slide)
(X) _{ID}	5'SH-TTTTTTTTTTTTTTTTTTTTCGTTGGTGCGGA TATCTCGGTAGTGGGATACGACGATACOGAAG ACAGCTCATGTTATATTATG-3'	5'SH-TTTTTTTTTTTATCCCACTACCGAGATATCCGCAC CAACG-3'
(GT) _{ID}	5'SH-TTTTTTTTTTGTGTGTGTGTGTGTGTGTGTGT-3'	5'SH-TTTTTTTTTTACACACACACACACACACACA-3'
(GT) _{ID}	5'SH-TTTTTTTTTTGTGTGTGTGTGTGTGTGTGTGT GTGTGT-3'	5'SH-TTTTTTTTTTACACACACACACACACACACA CACAC-3'
(GGT) _{ID}	5'SH-TTTTTTTTTTGGTGGTGGTGGTGGTGGTGGTGGT GGTGGT-3'	5'SH-TTTTTTTTTTACCACCACCACCACCACCACCACC ACC-3'

In all experiments the slide was approached with the tip of the cantilever, allowing the two single strands to hybridize and form a duplex. To avoid double rupture events the binding probability was adjusted to $<30\%$ by the applied force and duration time on the surface. Subsequently, the cantilever was retracted and the DNA duplex was loaded with a gradually increasing force until it finally ruptured and the cantilever relaxed back into its equilibrium position. The force applied to the DNA duplex via the PEG spacers was recorded as a function of the distance between the cantilever tip and the surface (Fig. 2). This curve was fitted with a two-state freely jointed chain (FJC) model, which describes the enthalpic and entropic behavior of polymers under an applied force (27).

Because most biologically relevant interactions are comparable in strength to thermal energies, force-induced processes such as the separation of receptor-ligand systems or in our case DNA duplexes are fluctuation-assisted processes (28). Therefore the distribution of the unbinding forces is broadened significantly (29). At a given force rate and at a fixed bond energy, a shift of the histograms directly reflects the difference in the effective width of the binding potentials (30,31) and indicates different unbinding pathways in the energy landscape.

To investigate, whether DNA duplexes with repetitive sequences have different unbinding pathways and therefore show different unbinding forces than heterogeneous sequences, both systems were analyzed. Fig. 3 shows the resulting distributions of the rupture forces of a heterogeneous (X)₃₀ and a repetitive DNA sequence (GT)₁₅ recorded at approximately the same pulling speed. Although both sequences have similar thermodynamic properties, which mainly correlate with the GC content of the sequence, their rupture force distributions differ drastically. The histogram for the repetitive DNA sequence (blue) is shifted toward lower dissociation forces. The DNA complex typically dissociates at forces below 40 pN. We conclude that an

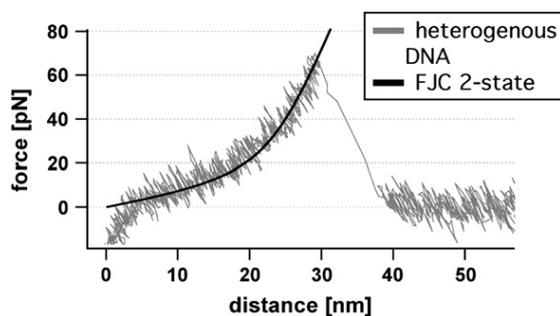


FIGURE 2 Example of a force-extension curve of a heterogeneous DNA duplex. While retracting the cantilever from the surface the polymer spacer and the DNA duplex are set under stress. The elastic behavior of the polymer-DNA duplex can be described with the FJC fit (black dashed line). At a force of 62 pN the double-stranded DNA dissociates and the cantilever drops back into its relaxed state.

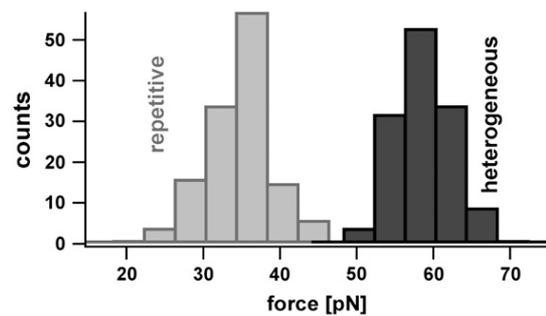


FIGURE 3 Histograms of the unbinding forces of DNA duplexes with a heterogeneous (X)₃₀ and a repetitive (GT)₁₅ sequence measured at similar pulling speeds. The duplex with the repetitive sequence dissociates at markedly lower forces, although its binding energy equals the binding energy of the heterogeneous sequence. The force distribution of the repetitive DNA is detruncated at a force of 40 pN. This gives evidence for an additional unbinding path, which is favored if an external force is applied.

additional dissociation path is available for the repetitive sequence. Note that the repetitive sequence might also bind fractionally and therefore might lead to lower dissociation forces. But without the assumption of an additional unbinding path this effect would lead to a broadening of the force distribution containing also higher forces similar to those of the heterogeneous sequence. The specificity of the measured interactions was proven, by replacing one single DNA strand with a noncomplementary sequence. This leads to $<0.5\%$ interactions.

Having established that repetitive DNA has characteristics, which are absent in heterogeneous sequences, two repetitive sequences with a different number of bases per repeat unit were compared with a heterogeneous sequence. The study of the unbinding mechanism of (GTT)₁₀ and (GT)₁₅ should reveal more detailed insights in the unbinding mechanism.

The theoretically predicted unbinding path represents a stepwise elongation of the repetitive DNA duplex by moving both strands against each other (see Fig. 1) as soon as the externally applied force exceeds a certain threshold (slipping threshold f_c). Such an elongation can indeed be observed in the recorded force-extension curves. Fig. 4 shows several typical force-extension curves obtained for two different repetitive and one heterogeneous DNA sequence. The force-extension curves for repetitive DNA deviate from the FJC behavior at forces above 40 pN, whereas the curves for heterogeneous DNA follow the FJC fit up to much higher forces. Apparently, repetitive DNA gets elongated at a slipping threshold between 35 and 40 pN. In the following we will use the expression “slipping threshold” for the value of the applied force beyond which the DNA duplex starts to slip or creep.

Whereas the results described above show further proof that repetitive sequences slip under load, the following experiment was carried out to examine the dependence of the slipping process on the length of the elementary repeat unit

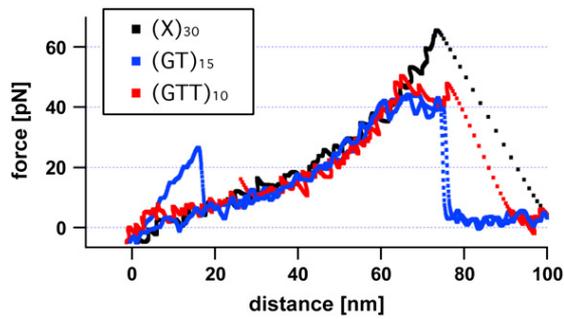


FIGURE 4 Typical force-extension curves of a heterogeneous $(X)_{30}$, a repetitive $(GT)_{15}$, and a $(GTT)_{10}$ sequence. In contrast to the force-extension curve of the heterogeneous DNA sequence, which follows the two-state FJC behavior, the repetitive DNA duplexes elongate at a force (slipping threshold) of ~ 35 – 40 pN until they finally dissociate (rupture force).

and the number of repeats. These experiments were carried out with the following sequences: $(GTT)_{10}$, $(GT)_{10}$, and $(GT)_{15}$. These DNA duplexes were probed at different pulling speeds because the slipping threshold is expected to be speed dependent. Because the slipping thresholds only differ by a few piconewtons, each data set was recorded with one cantilever to avoid calibration errors. In Fig. 5 the maxima of the slipping force distributions are plotted against the pulling speed of the cantilever. As can be seen in this figure, the slipping threshold shows a weaker dependence on the pulling speed for dinucleotide than for trinucleotide repeat units. Furthermore, the slipping thresholds are lower for the shorter repetitive sequence $(GT)_{10}$ than for the sequence containing 15 repeat units. In a linear response, the velocity of a particle in a viscous environment is given by the product, containing the mobility of the particle and the applied force. Close to the slipping threshold f_c , the relationship between the measured force and the velocity of DNA slippage can be treated similarly. Here an effective friction for relative strand motion

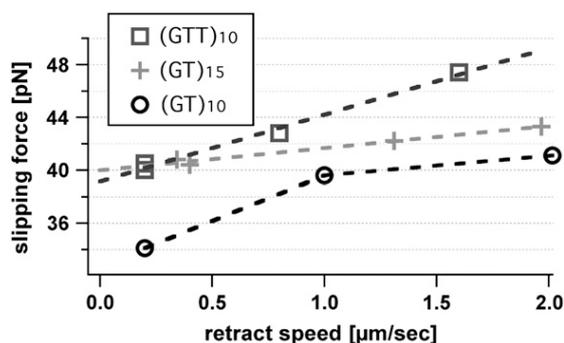


FIGURE 5 Pulling speed dependence of the slipping threshold for different repetitive DNA sequences. To avoid spring calibration errors every data set for one sequence is performed in a single experiment with the same cantilever. The maxima of the slipping threshold histograms, containing 80–150 force curves, are plotted against the pulling speed of the cantilever. The slipping threshold depends on the pulling speed and shows a linear time dependency as a first approximation.

arises from the need to nucleate bulge loops to shift both strands.

Close to slipping threshold f_c the slipping velocity $v(f)$ increases linearly with the measured force due to $v(f) = dv/df \times (f - f_c)$. The variable $u_0 = dv/df$ represents the effective slipping mobility, which depends on the bulge loop nucleation rate, microscopic slipping rate, and the number of repeat units. In our experiments, we correlated the slipping velocity $v(f)$ with the retract speed of the cantilever and measured the resulting slipping threshold.

From a linear fit, we achieve a slipping mobility u_0 of 580 nm/s-pN for a dinucleotide and 250 nm/s pN for a trinucleotide sequence. This is in agreement with the theoretical predictions and with bulk experiments that observed faster expansions for shorter repeat units. To form a bulge loop in a dinucleotide sequence, fewer basepairs have to open up than in a trinucleotide sequence and hence the rate to create these defects is smaller for longer repeat units. However, the additional length increase per step for longer repeat units does not compensate the lower rate.

So far the experimental results confirm that repetitive DNA strands can slide against each other and that the slipping threshold can be determined for different retract speeds. The values obtained for the slipping mobility are in good agreement with theoretical predictions. However, the time resolution in an usual force-extension measurement is not sufficient enough to discriminate individual steps, which would give direct evidence of the stepwise microscopic sliding mechanism.

Initial force clamp (32) measurements (data not shown) performed with the AFM only showed a lengthening of the different DNA duplexes at forces of 35–40 pN, but failed to resolve the expected individual steps. Therefore, a new measurement protocol was implemented, whose time resolution is limited only by the relaxation of the cantilever. These measurements were carried out as follows: i), The cantilever was lowered, to allow the DNA to hybridize and form a duplex. ii), The cantilever was gradually retracted from the surface allowing a certain force, well below the slipping threshold, to build up. iii), Then, in one step, the cantilever was retracted an additional 3–7 nm away from the surface. As a result of this distance jump, the force acting on the DNA duplex rises almost instantaneously to a new higher value. Initially, the contour length, which gives the total length under force, does not change. If, in response, the DNA duplex elongates due to slipping, an increase of the contour length is observed. In addition, the applied force drops, which can be detected by the cantilever.

Fig. 6 A shows two typical curves, force versus time and distance versus time, for a 15-times repetitive dinucleotide DNA duplex $(GT)_{15}$ recorded with the measurement protocol described above. First, the force acting on the DNA molecule is fluctuating around 38 pN, a value close to the previously observed slipping threshold. As indicated with the blue arrow in the force versus time graph a 4-nm distance

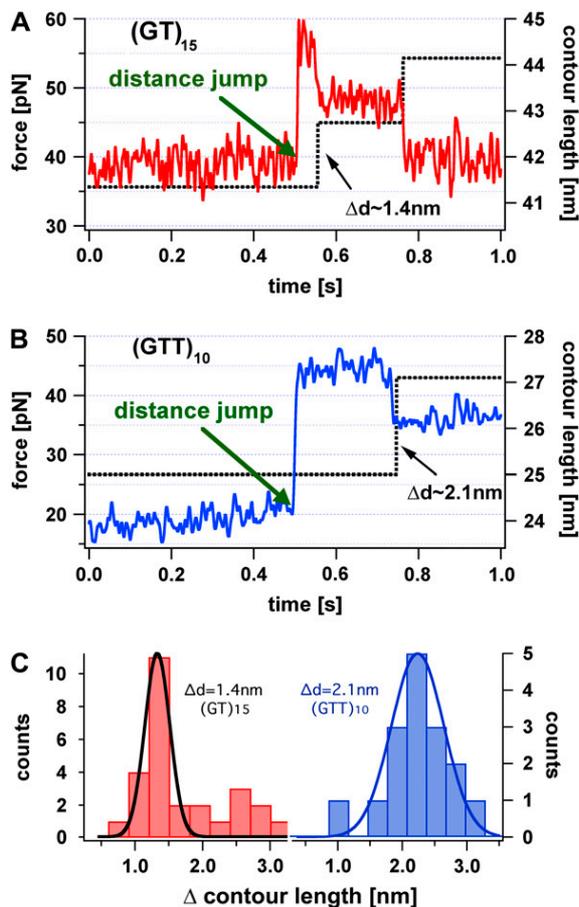


FIGURE 6 Force versus time (*red*) and contour length versus time curve (*black*) of a repetitive (GT)₁₅ and (GTT)₁₀ DNA duplex, initially held at a force below the slipping threshold. After 0.5 s a distance jump of the cantilever was performed resulting in a force step above the slipping threshold, but leaving the contour length of the molecule constant. In panel A the contour length of the (GT)₁₅ DNA complex relaxes in two discrete elongation steps and the force acting on the duplex drops below the slipping threshold. In panel B the contour length (*black curve*) of the (GTT)₁₀ DNA complex elongates in one discrete step and the force drops below the slipping threshold. Panel C shows the distribution of the contour length change for the di- and trinucleotide sequence. The additional peak at (~ 2.8 nm) in the dinucleotide sequence may occur from double slipping events.

jump was performed. As a result of the distance jump the force (*red*) increases to nearly 57 pN, but the contour length stays constant. Within a fraction of a second the measured force decreases to a value below 40 pN in discrete steps. The final force is again close to the observed slipping threshold. This rapid and discrete decrease of the force can only be explained with a stepwise lengthening of the DNA molecule, which compensates the performed distance jump. These observations show that a single relaxation process increases the contour length of the dinucleotide DNA system by $\sim 1.4 \pm 0.3$ nm. This value was obtained by a Gaussian fit of the contour length increase histogram (see Fig. 6 C) of several experiments, with a confidence interval of 90% certainty. This effect can be well explained by a relative sliding of one

repeat unit ($dl = 2l_s - l_d = 4 \times 0.5 \text{ nm} - 2 \times 0.34 \text{ nm} = 1.4 \text{ nm}$). Unfortunately, only a limited number of steps can be observed, because the probability of holding these DNA duplexes under such a high force for a long time is very low and decreases further with every step.

Fig. 6 B shows the equivalent experiment for a trinucleotide (GTT)₁₀ sequence. As expected, the contour length increase of $\sim 2.1 \pm 0.3$ nm, determined analogous to the (GT) sequence, is higher than for the dinucleotide sequence. Analogous experiments performed with the heterogeneous sequence did not show any discrete steps (data not shown).

To exclude the possibility, that the observed steps are artifacts of multiple binding the following arguments are pointed out. First of all the overall elasticity of the measured PEG polymer spacer would be much stiffer. Secondly, the presumption of three bound molecules in parallel mimicking the three steps of the single molecule shown in Fig. 5 would require the respective PEG polymer spacers to differ in length by < 2 nm. This would mean that the total force acting on the cantilever would be distributed on three duplexes and as a consequence the lifetime for the duplexes would be much longer than our experimental findings. Dissociation of one duplex increases the split force applied to the remaining duplexes and reduces their lifetime drastically. For this reason multiple binding as potential artifact can be excluded with a very high certainty.

Having shown that all characteristic parameters describing the slipping process can be determined experimentally, we wanted to obtain more detailed information about the behavior near the slipping threshold. The system for the native and elongated conformation of the repetitive DNA duplex can be described with a two-state potential illustrated in Fig. 7. Application of an external force allows the tuning of the potential, so that the Gibbs free energy of these two states is the same as shown with the dashed line in Fig. 7. If this force equals the slipping threshold f_c the system can fluctuate in equilibrium. This was achieved in the measurement shown in Fig. 8. Using the above-mentioned measurement protocol a distance jump is performed and the force on the DNA duplex first increases over the slipping threshold limit f_c . As a result the DNA duplex elongates and the force drops to the slipping threshold. At this force the system starts to perform multiple back and forth slipping events. An additional distance jump forces the DNA duplex in its elongated conformation until it finally ruptures completely.

These fluctuations of the DNA duplex between the two states can be analyzed with random telegraph noise analysis similar to ion channel recordings (Fig. 8, *black curve*). The hidden Markovian process can be characterized with the transition rates from one state to the other (33). The data trace shown in Fig. 8 exhibits mean lifetimes of 0.031 s for the elongated and 0.022 s for the shortened state. Note that these lifetimes are very dependent on the applied force. The energy difference between these two states varies between the binding energy of the DNA duplex and the bending energy

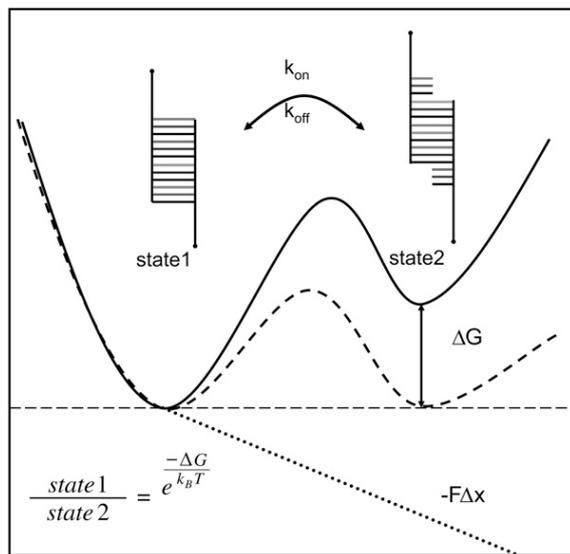


FIGURE 7 Gibbs free energy of a two-state system under an external force. The model describes the completely bound and the first lengthened state of a repetitive DNA duplex. The potential is tilted due to an externally applied force. This results in a leveling of the energy of the two states, allowing the DNA duplex to fluctuate between the two states in equilibrium.

of the cantilever. The energy was found to be $\sim 7 k_B T$, which is close to well-established values of about $\sim 8 k_B T$ for a trinucleotide GTT repeat unit. Other experiments underline this value. Due to the sensitivity of the system regarding the applied force and the low detection probability a closer examination will require substantial instrumental improvements. The observed multiple forward and backward jumps in Fig. 8 could be detected with short polymer spacers with lengths between 15 and 20 nm only. A possible reason for this finding could be that the fluctuations of the cantilever allow the duplex to form a bulge loop at lower forces, which eventually diffuses to the other end. For longer spacers these

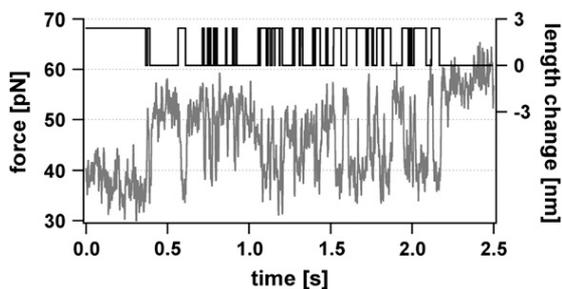


FIGURE 8 Example for the slipping of the DNA duplex between the elongated and the short state. A repetitive (GTT)₁₀ DNA duplex is held at a constant force analogous to Fig. 6. A distance jump drives the force above the slipping threshold. This results in a lengthening of the DNA duplex and the force drops down to the slipping threshold. Consequently, the duplex lengthens and shortens due to forward and backward slipping. The measured time trace (red) of the fluctuation process was analyzed with a telegraph noise algorithm to extract the dynamics of the length changes (black). The mean lifetimes were found to be 0.031 s for the fully hybridized state and 0.022 s for the lengthened state.

fluctuations are averaged by the elasticity of the polymer (34). The alternative scenario, that the observed shortening is a simple transient bulge loop formation at the stretched end, can be ruled out because the lifetime of these bulge loops, even if they travel some steps into the molecule, is orders of magnitude too small to explain the observed frequencies.

DISCUSSION

The data presented here show that repetitive DNA duplexes elongate under an applied shear force and dissociate at significantly lower forces of ~ 38 pN than for heterogeneous DNA sequences. Because of the possibility of fractional binding for repetitive sequences, lower dissociation forces are possible in regular force distance curves (see Fig. 2). This is due to an additional unbinding path that allows the repetitive DNA duplex to increase its contour length without having to overcome a large free energy barrier. It should be pointed out that this unbinding path energetically is not favored over other paths but gets populated by force.

The theoretically predicted length increase occurs in discrete slipping steps. We could show that the resulting length increase of the whole DNA duplex is consistent with the length increase obtained by shifting both strands of di- and trinucleotide sequences by one repeat unit. Slippage is faster for shorter repeat units and smaller repeat numbers. This is consistent with the theory of bulge loop diffusion because the expected slipping velocity decreases with the energy needed to produce a bulge loop. In addition, the diffusion of a bulge loop through the molecule is faster for shorter duplexes. The mechanism of relative strand motion caused by the creation, diffusion, and absorption of bulge loop defects is similar to defect propagation in crystal lattices.

The slipping threshold determined in the measurements was found to be larger than the theoretically predicted slipping threshold f_c . This may be due to the small number of repeat units used in the experiments and to the simplistic model used for the theory. For instance, deformations and conformational changes in the backbone of the DNA duplex resulting from an externally applied force are not included in the model.

The slipping velocity is expected to scale inversely with the number of repeat units. This prediction could not be unambiguously confirmed because only rather short sequences were available. Further experiments are necessary to quantify the dependence of the slipping dynamics on the repeat number and flanking sequences. More detailed experiments will shed some light on the kinetics of the processes involved in expansion of microsatellites during replication. Because of its bidirectional property DNA slippage, itself, is not the cause for the asymmetric increase effect of repeat units in the human genome.

Besides the biological importance of repetitive sequences, the remarkable properties of repetitive DNA might also be useful for different nanomechanical applications (35–37). Because the rupture force distribution for repetitive

sequences is truncated sharply at forces close to 40 pN, repetitive DNA could serve as a programmable force sensor, with a threshold force that can be fine tuned by sequence composition. Adjustable viscoelastic building blocks in DNA self-assembly structures can be realized with repetitive sequences.

Furthermore, the relaxation of a large force to a slipping threshold force f_c with a time constant that can be chosen by length and sequence composition could be used as a length independent force normal. Conversely, if extended, repetitive double-stranded DNA contracts until the slipping threshold force f_c is built up if the initial force is below f_c . Therefore, complementary repetitive single-stranded DNA could be applied for self-tightening connections in nanostructures. After initial hybridization, two single strands tend to maximize their overlap, i.e., the number of basepairs, until a tension of the order of the slipping threshold f_c is built up. These adjustable force-induced tensions at confident locations establish completely new features in nanoscale structures.

We thank Ulrich Gerland and Stefan Thalhammer for helpful discussions and Steffen Mihatsch for help with the data analysis.

This study was supported by the German Science Foundation DFG and by the Fond der Chemischen Industrie.

REFERENCES

- Seeman, N. C. 2003. DNA in a material world. *Nature*. 421:427–431.
- Schlotterer, C., and D. Tautz. 1992. Slippage synthesis of simple sequence DNA. *Nucleic Acids Res.* 20:211–215.
- Altan-Bonnet, G., A. Libchaber, and O. Krichevsky. 2003. Bubble dynamics in double-stranded DNA. *Phys. Rev. Lett.* 90:138101.
- Neher, R. A., and U. Gerland. 2006. Intermediate phase in DNA melting. *Phys. Rev. E.* 73:030902.
- Lyer, R. R., A. Pluciennik, V. Burdett, and P. L. Modrich. 2006. DNA mismatch repair: functions and mechanisms. *Chem. Rev.* 106:302–323.
- Heidenfelder, B. L., A. M. Makhov, and M. D. Topal. 2003. Hairpin formation in Friedreich's ataxia triplet repeat expansion. *J. Biol. Chem.* 278:2425–2431.
- Karthikeyan, G., K. V. R. Chary, and B. J. Rao. 1999. Fold-back structures at the distal end influence DNA slippage at the proximal end during mononucleotide repeat expansions. *Nucleic Acids Res.* 27:3851–3858.
- Levinson, G., and G. A. Gutman. 1987. Slipped-strand mispairing: a major mechanism for DNA-sequence evolution. *Mol. Biol. Evol.* 4:203–221.
- Trinh, T. Q., and R. R. Sinden. 1991. Preferential DNA secondary structure mutagenesis in the lagging strand of replication in *Escherichia coli*. *Nature*. 352:544–547.
- Li, Y. C., A. B. Korol, T. Fahima, and E. Nevo. 2004. Microsatellites within genes: structure, function, and evolution. *Mol. Biol. Evol.* 21:991–1007.
- Schlotterer, C. 2000. Evolutionary dynamics of microsatellite DNA. *Chromosoma*. 109:365–371.
- Porschke, D. 1974. Model calculations on the kinetics of oligonucleotide double helix coil transitions. Evidence for a fast chain sliding reaction. *Biophys. Chem.* 2:83–96.
- Smith, S. B., Y. J. Cui, and C. Bustamante. 1996. Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science*. 271:795–799.
- Perkins, T. T., S. R. Quake, D. E. Smith, and S. Chu. 1994. Relaxation of a single DNA molecule observed by optical microscopy. *Science*. 264:822–826.
- Rief, M., H. Clausen-Schaumann, and H. E. Gaub. 1999. Sequence-dependent mechanics of single DNA molecules. *Nat. Struct. Biol.* 6:346–349.
- Cui, S., C. Albrecht, F. Kühner, and H. E. Gaub. 2006. Weakly bound water molecules shorten single-stranded DNA. *J. Am. Chem. Soc.* 128:6636–6639.
- EssevazRoulet, B., U. Bockelmann, and F. Heslot. 1997. Mechanical separation of the complementary strands of DNA. *Proc. Natl. Acad. Sci. USA*. 94:11935–11940.
- Lubensky, D. K., and D. R. Nelson. 2002. Single molecule statistics and the polynucleotide unzipping transition. *Phys. Rev. E.* 65:031917.
- Friedsam, C., A. K. Wehle, F. Kühner, and H. E. Gaub. 2003. Dynamic single-molecule force spectroscopy: bond rupture analysis with variable spacer length. *J. Phys. Condens. Matter*. 15:S1709–S1723.
- Kühner, F., L. T. Costa, P. M. Bisch, S. Thalhammer, W. M. Heckl, and H. E. Gaub. 2004. LexA-DNA bond strength by single molecule force spectroscopy. *Biophys. J.* 87:2683–2690.
- Krautbauer, R., M. Rief, and H. E. Gaub. 2003. Unzipping DNA oligomers. *Nano Lett.* 3:493–496.
- Strunz, T., K. Oroszlan, R. Schafer, and H. J. Guntherodt. 1999. Dynamic force spectroscopy of single DNA molecules. *Proc. Natl. Acad. Sci. USA*. 96:11277–11282.
- Kühner, F., M. Erdmann, L. Sonnenberg, A. Serr, J. Morfill, and H. E. Gaub. 2006. Friction of single polymers at surfaces. *Langmuir*. 22:11180–11186.
- Neher, R. A., and U. Gerland. 2004. Dynamics of force-induced DNA slippage. *Phys. Rev. Lett.* 93:198102.
- Neuert, G., C. Albrecht, E. Pamir, and H. E. Gaub. 2006. Dynamic force spectroscopy of the digoxigenin-antibody complex. *FEBS Lett.* 580:505–509.
- Blank, K., J. Morfill, and H. E. Gaub. 2006. Site-specific immobilization of genetically engineered variants of Candida Antarctica Lipase B. *American Chemical Society*. 7:1349–1351.
- Oesterhelt, F., M. Rief, and H. E. Gaub. 1999. Single molecule force spectroscopy by AFM indicates helical structure of poly(ethylene-glycol) in water. *New J. Phys.* 1:6.
- Evans, E., and K. Ritchie. 1997. Dynamic strength of molecular adhesion bonds. *Biophys. J.* 72:1541–1555.
- Friedsam, C., A. K. Wehle, uuml, F. hner, and H. E. Gaub. 2003. Dynamic single-molecule force spectroscopy: bond rupture analysis with variable spacer length. *J. Phys. Condens. Matter*. S1709–S1723.
- Evans, E. 2001. Probing the relation between force-lifetime and chemistry in single molecular bonds. *Annu. Rev. Biophys. Biomol. Struct.* 30:105–128.
- Heymann, B., and H. Grubmuller. 2000. Dynamic force spectroscopy of molecular adhesion bonds. *Phys. Rev. Lett.* 84:6126–6129.
- Oberhauser, A. F., P. K. Hansma, M. Carrion-Vazquez, and J. M. Fernandez. 2001. Stepwise unfolding of titin under force-clamp atomic force microscopy. *Proc. Natl. Acad. Sci. USA*. 98:468–472.
- Venkataramanan, L., and F. J. Sigworth. 2002. Applying hidden Markov models to the analysis of single ion channel activity. *Biophys. J.* 82:1930–1942.
- Kühner, F., and H. Gaub. 2006. Modelling cantilever based force spectroscopy with polymers. *Polym.* 47:2555–2563.
- Neher, R. A., and U. Gerland. 2005. DNA as a programmable viscoelastic nanoelement. *Biophys. J.* 89:3846–3855.
- Albrecht, C., K. Blank, M. Lalic-Multhaler, S. Hirler, T. Mai, I. Gilbert, S. Schiffmann, T. Bayer, H. Clausen-Schaumann, and H. E. Gaub. 2003. DNA: a programmable force sensor. *Science*. 301:367–370.
- Simmel, F. C., and W. U. Dittmer. 2005. DNA nanodevices. *Small*. 1:284–299.

DNA as a Programmable Viscoelastic Nanoelement

Richard A. Neher and Ulrich Gerland

Arnold Sommerfeld Center for Theoretical Physics and Center for Nanoscience, Ludwig-Maximilians Universität München, Munich, Germany

ABSTRACT The two strands of a DNA molecule with a repetitive sequence can pair into many different basepairing patterns. For perfectly periodic sequences, early bulk experiments of Pörschke indicate the existence of a sliding process, permitting the rapid transition between different relative strand positions. Here, we use a detailed theoretical model to study the basepairing dynamics of periodic and nearly periodic DNA. As suggested by Pörschke, DNA sliding is mediated by basepairing defects (bulge loops), which can diffuse along the DNA. Moreover, a shear force f on opposite ends of the two strands yields a characteristic dynamic response: An outward average sliding velocity $v \sim 1/N$ is induced in a double strand of length N , provided f is larger than a threshold f_c . Conversely, if the strands are initially misaligned, they realign even against an external force $f < f_c$. These dynamics effectively result in a viscoelastic behavior of DNA under shear forces, with properties that are programmable through the choice of the DNA sequence. We find that a small number of mutations in periodic sequences does not prevent DNA sliding, but introduces a time delay in the dynamic response. We clarify the mechanism for the time delay and describe it quantitatively within a phenomenological model. Based on our findings, we suggest new dynamical roles for DNA in artificial nanoscale devices. The basepairing dynamics described here is also relevant for the extension of repetitive sequences inside genomic DNA.

INTRODUCTION

The basic double-helical structure of DNA is insensitive to the nucleotide sequence, but many of its biophysical properties are not. For instance, the local thermodynamic stability of double-stranded DNA (dsDNA) depends strongly on the sequence (1), and certain sequence motifs can cause permanent bends or make DNA more bendable (2). Such local modulations of the DNA properties play an important role in molecular biology, e.g., for nucleosome positioning (3) and transcription regulation through DNA looping (4). The sequence-dependent stability of DNA basepairing is also crucial for applications in nanotechnology (5–7). Clearly, since the thermodynamics of DNA basepairing is sequence-dependent, the kinetics is sequence-dependent as well. Our aim here is to show that the kinetics can display a much richer phenomenology than might be expected on the basis of the thermodynamics alone.

The dynamics of DNA basepairing can be probed experimentally on the single-molecule level with mechanical and optical techniques (8–14). One approach is to unzip dsDNA from one end of the double helix (12,13,15). However, unzipping probes only one aspect of the basepairing dynamics—the sequential opening of consecutive basepairs. In a different approach, a shear force is applied by grabbing the two strands on opposite ends of the dsDNA (16); see Fig. 1. For a heterogeneous dsDNA with a random sequence, the effect of a shear force is to unravel the basepairs from both ends (16); see Fig. 1 *a*, which is qualitatively similar to unzipping. In contrast, with a perfectly periodic sequence,

e.g., $(C)_N$ on the upper and $(G)_N$ on the lower strand or a higher-order repeat such as $(CA)_N$ and $(GT)_N$, the two strands can bind in many configurations (17). An applied shear force then facilitates local strand slippage and can induce macroscopic DNA sliding (18); see Fig. 1 *b*. (Throughout this article, we use the term “DNA slippage” for microscopic events where a few bases at the end of the double-strand unbind and rebind shifted by one or several repeat units. In contrast, “DNA sliding” refers to an average large-scale movement of the two strands against each other.)

DNA slippage is an aspect of DNA basepairing dynamics, which plays an important role in the generation of a class of genetic diseases (19,20). If local DNA slippage occurs in an Okazaki fragment during DNA replication, trinucleotide repeats inside genes can get extended beyond a threshold length for the onset of Huntington’s and other diseases. Such slippage events are possible only within the time window that DNA polymerase needs to fill in the Okazaki fragment. Thus, the kinetics of strand slippage is an important determinant for the frequency of repeat extensions.

We propose that the dynamics of periodic and nearly periodic DNA is interesting also for the design of DNA-based nanodevices. Indeed, DNA is becoming increasingly popular as a building block for the assembly of nanoscale structures and devices (5–7). These applications already exploit the specificity of the basepairing interaction, e.g., to direct the assembly of DNA strands into predefined architectures, and the dynamics of DNA branch migration, e.g., to replace a bound DNA strand by a different strand. Below, in Discussion, we consider several new possible applications of DNA in nanotechnology, based on the dynamic properties identified in the main part of this article.

Finally, DNA sliding is interesting also from a purely theoretical perspective. Since simultaneous slippage of all basepairs is kinetically inhibited by an extensive activation barrier,

Submitted June 16, 2005, and accepted for publication September 8, 2005.

Address reprint requests to Ulrich Gerland, Arnold Sommerfeld Center for Theoretical Physics, Theresienstrasse 37, 80333 München, Germany. Tel.: 49-89-2180-4514; Fax: 49-89-2180-4517; E-mail: ulrich.gerland@physik.uni-muenchen.de.

© 2005 by the Biophysical Society

0006-3495/05/12/3846/10 \$2.00

doi: 10.1529/biophysj.105.068866

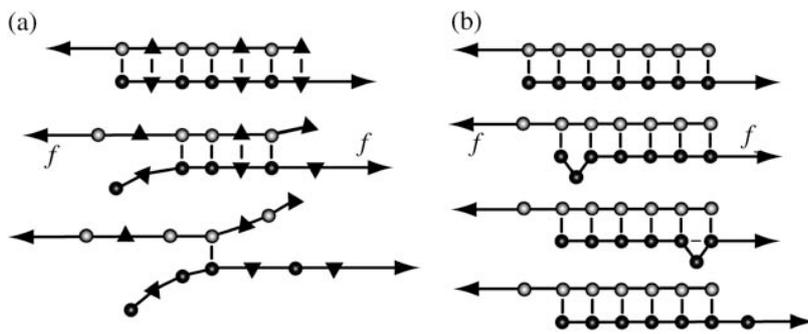


FIGURE 1 DNA under a shear force. (a) A non-periodic sequence unravels from both ends, driven by the length gain from converting stacked bases into longer single strands. (b) A periodic DNA sequence can open by sliding, mediated by bulge loops that are created at the ends and diffuse freely along the DNA. When a bulge loop reaches the opposite end, the two strands have effectively slipped against each other by a distance equal to the loop size.

the macroscopic sliding of DNA strands is a complex process involving the dynamics of many local basepairing defects (17,18). The most likely defects are bulge loops (see Fig. 1 *b*), which are created at the ends of the dsDNA, or, in pairs, anywhere along the molecule. Once formed, bulge loops diffuse freely along a periodic dsDNA until they annihilate with a loop on the opposite strand or are absorbed at an end. Mutations in the periodicity of the DNA sequence create obstacles for the diffusion of bulge loops. Effectively, the bulge-loop dynamics can be regarded as a reaction-diffusion process of particles and antiparticles in one dimension. DNA shearing experiments render certain aspects of these dynamics observable and permit a quantitative characterization.

The outline of this article is as follows. First, we describe our theoretical model for the energetics and dynamics of DNA basepairing under a shear force. We then show that our model leads to the following predictions:

1. For periodic dsDNA, the combination of polymer mechanics and basepairing dynamics gives rise to a viscoelastic response to shear forces above a threshold f_c , where both f_c and the viscosity η are programmable over a wide range through the DNA sequence. The viscoelastic behavior can be described with the help of a mechanical analog model.
2. DNA sliding is possible even when the exact sequence periodicity is destroyed by a few mutations.
3. The mutations affect the viscoelastic behavior by introducing a programmable time delay before sliding commences after a sudden force jump.
4. The mechanism for the time delay can be understood within a phenomenological model, which also permits a quantitative description of the full distribution of time delays. Taken together, we find that the sequence dependence of the basepairing dynamics allows us to adjust the mechanical response of DNA under a shear force over a broad range of behaviors. In the last section, we discuss the experimental ramifications of these findings.

DNA MODEL

To study the dynamics of DNA sliding, we consider a DNA molecule under a shear force f , which can be applied experi-

mentally by pulling the opposite 5' ends (16) or, alternatively, the opposite 3' ends. In a coarse-grained description, the configuration of the DNA is specified by its basepairing pattern S and the spatial contours of both strands. A generic configuration consists of two stretched and two unstretched single strands, and the central region from the first to the last basepair (see Fig. 1).

We will not explicitly describe the dynamics of the spatial polymer degrees of freedom, but assume rapid equilibration compared to the timescale of DNA sliding. This assumption is justified for short DNA molecules: The timescale to equilibrate a semiflexible polymer of length L and persistence length l_p in a solvent of viscosity η is $\eta L^4/72 l_p^2 f$, where f is an external force applied to its ends (21). For a DNA of 150 bp (one persistence length) in water at a 10-pN load, the equilibration time is on the order of 0.01 μ s, which is fast compared to the millisecond timescale of DNA sliding observed in the reannealing experiments of Pörschke (17). Hence, we integrate out the contour conformations to obtain a free-energy function $E(S)$ that depends only on the basepairing pattern S . The total free-energy $E(S)$ can be split up into three terms,

$$E(S) = E_{\text{stretch}}(S) + E_{\text{bp}}(S) + E_{\text{loop}}(S), \quad (1)$$

corresponding to the stretching energy, the free-energy gain due to basepairing, and the free-energy cost of (internal or bulge) loops in the pattern S , respectively.

Polymer model

The mechanical polymer properties of DNA enter only into the stretching energy, which we write in the form $E_{\text{stretch}}(S) = -fL(S)$, with an effective force-dependent total length $L(S)$ for the stretched DNA, i.e., the two single-stranded ends where the force is applied and the central DNA segment from the first to the last basepair (see Fig. 2). The unstretched single strands do not contribute to the free energy, since we take all energies relative to the unstretched and unpaired state, which is the usual convention (1). For the stretched single strands, we use a freely jointed chain polymer model with a Kuhn length twice the bare segment length $b_s \approx 0.7$ nm for a single base (22). With this model, each unbound

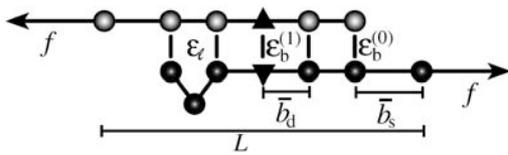


FIGURE 2 DNA free-energy model. The free energy $E(S)$ of a basepairing pattern S contains three separate contributions: first, a negative binding energy for basepairing. For simplicity, we assign the same binding energy $\varepsilon_b^{(k)}$ for every basepair of type k , regardless of the neighboring bases. Second, a positive free energy cost for internal and bulge loops. We assign the same cost ε_ℓ for every loop, regardless of its length and base sequence, since the detailed choice of the loop cost function does not affect our main findings. Third, a stretching energy. For a given pattern S , the stretching energy can be written in the form $-fL(S)$, with an effective length $L(S)$, which is obtained from force-dependent base-to-base distances $\bar{b}_d(f)$ and $\bar{b}_s(f)$ for double and single strands, respectively. Note that $L(S)$ does not correspond to the physical length of the DNA molecule (see main text).

base at the ends where the force is applied contributes an effective length $\bar{b}_s(f)$ to the total length $L(S)$, where

$$\bar{b}_s(f) = -\frac{k_B T}{2f} \ln \left(\frac{\sinh(2fb_s/k_B T)}{2fb_s/k_B T} \right) \quad (2)$$

and $k_B T$ is the thermal energy. Note that $\bar{b}_s(f)$ differs from the average extension of one segment in the direction of the force. Instead, the average total extension $\langle x \rangle$ of a DNA with basepairing pattern S is calculated as the force derivative of the stretching free energy,

$$\langle x \rangle = \frac{\partial E_{\text{stretch}}(S)}{\partial f}, \quad (3)$$

which yields the correct (Langevin) form for the extension of a freely jointed chain. For the central DNA region from the first to the last basepair, we assume a B-DNA conformation and use a wormlike-chain model with persistence length $l_p = 50$ nm and a contour length of $b_d = 0.34$ nm per base. (The length of an asymmetric loop in the central region is approximated by counting only the bases in the shorter arm of the loop.) For the forces of interest here, the effective length of a basepair, $\bar{b}_d(f)$, is given by the asymptotic formula

$$\bar{b}_d(f) = b_d \left(1 - \sqrt{\frac{k_B T}{4fl_p}} \right). \quad (4)$$

The force-dependence of the lengths $\bar{b}_s(f)$, $\bar{b}_d(f)$ is, in fact, essential only for our calculation of the viscoelastic response. For all other properties considered below, the force-dependence has no qualitative effect, and will hence be neglected (i.e., $\bar{b}_s(f) = b_s$, $\bar{b}_d(f) = b_d$ everywhere except in Viscoelastic Behavior).

Basepairing energy model

To obtain a compact theoretical description, we use a basepairing energy model which is simplified from the nearest-neighbor model of SantaLucia (1), but nevertheless permits

semiquantitative predictions. We exclude basepairs within a strand, and assign a binding (free) energy $\varepsilon_b^{(k)} > 0$ for each basepair of type k (Watson-Crick or other) between strands (see Fig. 2). Hence, $E_{\text{bp}}(S) = -\sum_k n_k(S) \varepsilon_b^{(k)}$, where $n_k(S)$ is the total number of type k basepairs in the basepairing pattern S . Similarly, we assign a loop initiation cost $\varepsilon_\ell > 0$ for each internal or bulge loop in a given pattern (we neglect an additional length-dependent loop cost, which has no qualitative effect on the results discussed below). Therefore, $E_{\text{loop}}(S) = q(S) \varepsilon_\ell$, where $q(S)$ is the total number of loops in the pattern. The numerical values of the free-energy parameters $\varepsilon_b^{(k)}$ and ε_ℓ are temperature-dependent (1). The actual values used in our simulations are given below.

Elementary kinetic steps

We support our phenomenological theory presented below by simulating the DNA basepairing dynamics in detail. To this end, we use a kinetic Monte Carlo scheme with three single base moves as elementary steps (23): basepair opening, basepair closing, and basepair slippage. Here, basepair slippage refers to a local shift of the binding partner of a base, which is possible only for basepairs next to unbound bases, i.e., inside loops or at the ends of the molecule. Clearly, basepair slippage can also be generated by a basepair opening move followed by a basepair closing move. However, the work of Pörschke (17) indicates that basepair slippage is faster than would be expected from the individual rates for basepair opening and closing (see below). Hence, we include the basepair slippage move, as has been done previously (23).

Kinetic rates

To fully specify our model for the DNA basepairing dynamics, we need to assign a rate to each elementary kinetic step. Careful relaxation experiments (24) determined the rate for basepair closing at the end of helical segments to be $1\text{--}20 \times 10^6 \text{ s}^{-1}$, where the range indicates the experimental uncertainty. In our model, we assume that this rate is independent of the identity of the basepair. To reproduce the correct equilibrium behavior from our basepairing dynamics, the rate for opening a basepair of type k at the end of a helix must be reduced by a factor $\exp(-\varepsilon_b^{(k)}/k_B T)$ with respect to the closing rate. From reannealing experiments with periodic sequences, Pörschke (17) estimated the rate for the displacement of a bulge loop by one base, i.e., the rate for basepair slippage, to be $\sim 5 \times 10^6 \text{ s}^{-1}$. Hence, the rates for basepair closing and basepair slippage are approximately equal, within experimental accuracy. In our model, we set them exactly equal, for simplicity (our main results are, in fact, insensitive to the precise value of the closing rate; see below). In general terms, our model assumes that all kinetic rates of passing from a higher energy configuration to one with lower energy are the same, whereas the reverse rates are chosen to obey detailed balance. It may be noteworthy that

recent theoretical work on the kinetics of force-induced RNA unfolding, which used similar assumptions, produced surprisingly good agreement with experiment (25).

Below, we report all of our kinetic simulation data in units of Monte Carlo steps. From Pörschke (17), our best estimate for the real-time equivalent of one Monte Carlo step is $0.2 \mu\text{s}$. However, it should be kept in mind that this estimate bears a large uncertainty.

SLIDING DYNAMICS OF PERIODIC SEQUENCES

The simplest periodic sequence is a repetition of one base on one strand, e.g., AAA. . . , and the complementary base on the other. In this case, we have only basepairs of one type (i.e., $\varepsilon_b^{(k)} \equiv \varepsilon_b$ in our model) and each base on one strand can bind to any base on the other strand. For longer repeat units, e.g., triplet repeats such as CAGCAG. . . , which play an important role in genetic diseases (19), one can treat a repeat unit as an effective base with larger associated energies ε_b , ε_ℓ and lengths b_d , b_s . We are interested in the basepairing dynamics induced by a constant shear force f that is suddenly turned on at $t = 0$. In the following, we first review the physical description of DNA sliding dynamics, which we have established already in Neher and Gerland (18). We then construct a mechanical analog model to characterize the viscoelastic response of periodic DNA and its sequence-dependence.

Quantitative phenomenological description

As illustrated in Fig. 1 *b*, sliding of periodic dsDNA is mediated by the creation, diffusion, and annihilation of bulge loops. When a force is applied, the diffusion of bulge loops within the dsDNA remains unbiased, assuming the force does not deform the dsDNA structure significantly (this assumption clearly breaks down for forces above the B-S transition around 65 pN). In contrast, the force strongly affects the rates at which bulge loops are created at the ends. When the two DNA strands are misaligned, these creation rates are imbalanced, since a bulge at an overhanging end does not reduce the number of basepairs in the structure (although it does on the opposite end). This imbalance produces a restoring force f_c , which can be obtained approximately (18) by balancing the energy cost of breaking a basepair with the gain in stretching energy, $\varepsilon_b = f \times (2\bar{b}_s(f) - \bar{b}_d(f))$. The restoring force creates an average inward drift that realigns the two strands. To obtain an outward drift velocity v , i.e., macroscopic sliding, one needs to overcome the restoring force f_c , so that $v \sim (f - f_c)$ in the vicinity of f_c . Indeed, f_c becomes a critical force in the thermodynamic sense when the limit of a large strand length N is taken and the state where the strands are completely separated is excluded.

At the critical force, the rates at which bulge loops are produced are equally large on the overhanging stretched and the unstretched ends. The average sliding velocity v van-

ishes; however, the bulge-loop dynamics still leads to a macroscopic diffusion of the two strands relative to each other, with a diffusion coefficient D . Interestingly, this diffusion coefficient scales with the total number of bases as $D \sim 1/N$, so that the rupture-time τ required to separate the two strands completely scales as $\tau \sim N^3$ instead of the usual $\tau \sim N^2$ for diffusion of a particle over a distance N . This scaling of D is due to the fact that loops are generated at the ends with a constant rate, but only result in a global shift between the strands, if they diffuse over a distance $\sim N$, either to annihilate at the other end or with a loop on the opposite strand. In both cases, the probability for an event scales as $1/N$. The $D \sim 1/N$ scaling occurs also in the reptation problem of polymer physics, and indeed the microscopic origin is closely related, as motion is mediated by defect diffusion (26).

Since the production of a loop on the stretched ends shortens the molecule, the corresponding production rate decreases with f . Hence, the rates of events extending or shortening the double-stranded region, that are equal for $f = f_c$, differ at other forces resulting in a drift. Each of these rates, and consequently the sliding velocity v as well, is proportional to $1/N$. From the Einstein relation, one expects $v \sim (f - f_c)D \sim 1/N$, in agreement with this result. With the negative (inward) drift velocity below f_c , rupture events are driven by rare fluctuations, and the rupture-time τ grows exponentially with N , as is characteristic for thermally activated transport over an extensive energy barrier. On the other hand, for forces larger than f_c , the N^{-1} scaling of v leads to rupture times increasing as $\tau \sim N^2$.

This scaling holds up to a force f^* , above which the rupture times grow only linearly with N , due to a dynamical transition in the opening mode from sliding to unraveling (i.e., the opening mode of heterogeneous dsDNA). For $f > f^*$, it is energetically favorable to break basepairs consecutively from both ends and both strands are separated after breaking N basepairs. Within our model, f^* is well approximated by the solution of $f = \varepsilon_b / [\bar{b}_s(f) - \bar{b}_d(f)]$ —i.e., the balance between the basepairing energy and the stretching energy gained by extending the molecule by the difference between the length of an unbound base and a basepair. (For DNA sequences where this force is large enough to deform the DNA structure, in particular for f^* above the ~ 65 pN of the B-S transition, the unraveling mode may not exist.)

Viscoelastic behavior

DNA sliding can be regarded as a viscous flow of the two strands relative to each other. According to the physical picture reviewed above, this flow has interesting nonlinear and sequence-dependent properties. Since the shear force also elicits an elastic response (due to the entropic elasticity of DNA), the behavior of periodic dsDNA is reminiscent of a viscoelastic material. Such materials combine solidlike and fluidlike mechanical properties when probed by external stress. In the following, we examine this analogy more closely.

The mechanical behavior of a typical viscoelastic material can be described by a Zener model (27), which is constructed, e.g., by connecting a Kelvin element (a dashpot in parallel with a spring) in series with a spring (see Fig. 3). The Zener model reproduces the two prominent characteristics of viscoelastic materials:

1. In a creep experiment, where a constant stress is suddenly applied, an instantaneous elastic strain is followed by a gradual creep toward a new equilibrium.
2. When the strain is suddenly increased, the stress rises sharply and then relaxes gradually to an equilibrium value.

On a qualitative level, periodic dsDNA displays these same characteristics in its average behavior:

1. Upon sudden application of a constant force f in the range $f_c < f < f^*$, the DNA rapidly stretches against its entropic elasticity and slowly creeps with a viscosity η that is proportional to the number of bases in the double strand. However, it will not approach a new equilibrium, but eventually rupture.
2. When the extension of the DNA is suddenly increased, the tension rapidly rises and then slowly relaxes to the critical value f_c (provided the initial rise was above f_c).

The viscoelastic behavior of periodic DNA can be described by a nonlinear generalization of the Zener model (see Fig. 3 *b*), where the Kelvin element effectively describes the basepairing dynamics, while the outer elastic element accounts for the entropic elasticity of the polymer backbone (consisting of dsDNA, single-stranded DNA, i.e., ssDNA, and, if present, linkers to the points of force application). Since the basepairing dynamics of two misaligned complementary periodic DNA strands produces a restoring force f_c , the sliding velocity v is proportional to $f - f_c$. The sliding dynamics is thus described by a dashpot in parallel with

a potential generating the restoring force and preventing contraction beyond maximal overlap. In contrast to the standard Zener model with harmonic springs, the stress in response to a strain will relax to the value f_c , independent of the displacement (within a certain range). Fig. 3 *a* shows extension-time-traces obtained from our model, both for a periodic (*bottom panel*) and a heterogeneous DNA (*center panel*); see Fig. 3 legend for parameters. Here, we have considered a creep test situation where the force is switched periodically between $f_{\min} < f_c$ and $f_{\max} > f_c$ (*top panel*). Fig. 3 *c* shows the corresponding behavior of the generalized Zener model for comparison. We observe that the average behavior of the periodic DNA resembles that of the generalized Zener model, whereas the heterogeneous DNA shows only elastic behavior. Of course the extension also displays strong thermal fluctuations, which play an important role in single-molecule dynamics, and ultimately lead to rupture even below the critical force (18).

Programmability

The viscoelastic behavior described above relies on the basepairing dynamics within the DNA molecule, and is manifestly sequence-dependent. This fact makes the mechanical behavior of dsDNA under shear-force programmable, i.e., both the force offset f_c and the viscosity η can be adjusted through sequence composition and length of the dsDNA. Even for perfectly periodic sequences, there is still a considerable freedom in the choice of the sequence composition, since a repeat unit can be several bases long and involve different combinations of Watson-Crick and other basepairs. Exploiting this freedom, the range over which the average basepairing energy ε_b can be programmed is $\sim 0.5\text{--}4 k_B T$ (1), which translates into an equally wide range of force offsets $f_c = \varepsilon_b / (2\bar{b}_s - \bar{b}_d)$. (The precise experimental range of the force offset is difficult to predict,

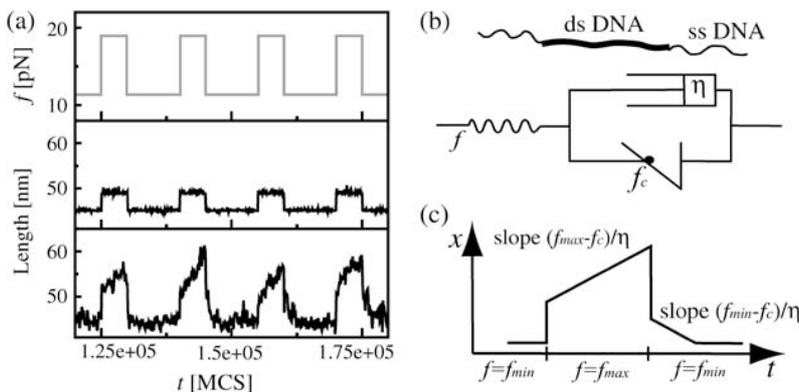


FIGURE 3 Viscoelastic response of periodic DNA. (*a*) The shear force on an 80-bp dsDNA (with two 20-bp ssDNA linkers) is switched periodically between $f_{\min} = 11.4$ pN and $f_{\max} = 19$ pN (*upper panel*). The center and bottom panels show the extension-time-trace for heterogeneous and periodic DNA, respectively (energy parameters: $\varepsilon_b = 1.11 k_B T$ and $\varepsilon_\ell = 2.8 k_B T$, roughly corresponding to AT basepairs at 50°C (18)). The time units are Monte Carlo steps, the real-time equivalent of which is discussed in Kinetic Rates (see article). The heterogeneous DNA responds only elastically to the force jumps, mostly due to stretching of the linkers. The length of the periodic DNA shows a similar elastic strain, but in addition, the molecule elongates at a finite speed due to sliding, since $f_{\max} > f_c = 16.3$ pN. When the molecule is relaxed, we find an elastic response and inward sliding,

since $f_{\min} < f_c$. The length of the periodic DNA fluctuates strongly due to loop formation and annihilation. (*b*) The viscoelastic behavior displayed by a periodic DNA molecule can be described by a generalized Zener model, where harmonic springs are substituted by anharmonic elastic elements describing polymer elasticity and the restoring force f_c . The ideal dashpot (with viscosity η) creates the viscous behavior of periodic dsDNA. (*c*) The response of the above idealized model to the same periodic force resembles the average response of periodic DNA.

since it depends sensitively on the effective ssDNA and dsDNA length. Roughly, we expect values of up to 30 pN.)

The velocity of macroscopic strand sliding is determined by four factors: 1), the mobility of defects, i.e., the rate for bulge-loop displacement; 2), the bulge-loop density; 3), the inverse strand length; and 4), the deviation of the force f from the critical force f_c . The defect density depends sensitively on the basepairing free energy and may vary roughly between 0.001 and 0.2 for different repeat lengths and temperatures, leaving great freedom to adjust the timescale of DNA sliding. Note that since only the bulge-loop density and not the individual rates for basepair closing and opening influence the sliding velocity, the rate for bulge-loop displacement is the only crucial rate parameter in our model. By increasing the strand length, the sliding velocity can be made arbitrarily small, or, equivalently, the viscosity η can be made arbitrarily large ($\eta \sim N$). Alternatively, η is increased by using longer repeat units, since η grows exponentially with the free energy cost of creating a bulge loop. An order-of-magnitude estimate for the lower bound on η yields $\sim 10^{-3}$ pN \times s/nm, based on reannealing experiments with homogeneous oligonucleotides of 10 bps (the reannealing experiments of Pörschke (17) suggest that a misaligned 10-bp molecule can slide by one basepair within 0.1 ms; assuming that the sliding velocity extends linearly from $f_c \sim 10$ pN to force zero, one obtains the estimate $\eta \sim 3 \times 10^{-3}$ pN \times s/nm). With these force- and timescales, DNA sliding should be well observable in single-molecule experiments.

PERIODIC DNA WITH WEAK SEQUENCE DISORDER

How is the basepairing dynamics affected when a few mutations destroy the perfect periodicity? Fig. 4 (top) shows two simulated extension-time-traces, one for a perfectly periodic sequence and one with $M = 7$ equidistant mutations (DNA parameters; see Fig. 4's legend). Here, we assigned the same binding energy to mutated and original basepairs, to focus on the effects that mutations exert on the basepairing dynamics rather than the energetics. Furthermore, we assumed that mutated bases can only bind to their ground-state binding partners, i.e., mutated bases cannot form basepairs with the original bases and all mutations are of a different type. The less generic effects that can result without these assumptions are discussed below.

We observe that the mutations have a drastic effect: whereas the original sequence begins to slide almost immediately after application of the force, the mutated sequence exhibits a pronounced delay before sliding sets in. Indeed, the figure suggests that the mutated sequence has two characteristic timescales: a waiting time τ_w , during which the extension fluctuates around a constant value; and a sliding time τ_s , during which the extension increases until the two strands are completely separated. Another, less drastic effect of the mutations is to reduce τ_s ; i.e., once sliding starts, it is

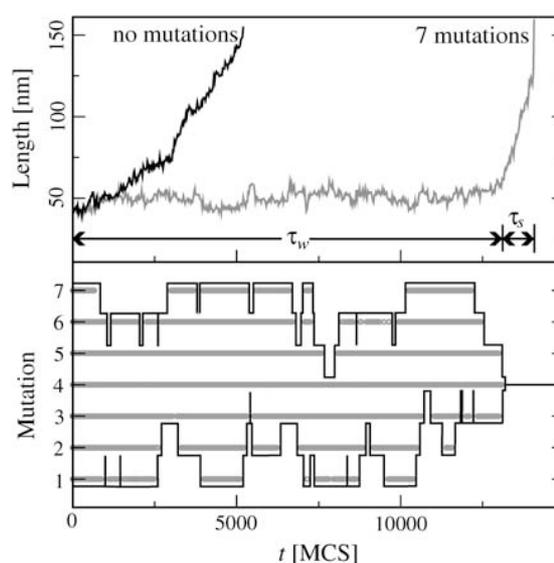


FIGURE 4 (Top) Extension-time-trace for a perfectly periodic DNA of $N = 120$ basepairs and the same DNA with seven mutations, both under a shear force of $f = 12.7$ pN (energy parameters as in Fig. 3). Whereas the molecule without mutations starts sliding almost immediately, the molecule with mutations fluctuates about its initial length for some time τ_w before sliding starts. (Bottom) The time-trace of the binding state (open/closed) for the seven mutated basepairs in the sequence. Each mutated basepair (1–7) is unbound wherever the corresponding thick horizontal line is broken, and bound where the line is shown. Note that the mutated basepairs do not open/close independently from each other. Instead, a mutated basepair opens only once all mutated basepairs to the left or right are already open. The black envelope curves emphasize the positions of the outermost bound mutation on each side. Their dynamics resembles a (biased) random walk. Sliding begins when all mutations are open.

faster than without mutations. Note that the convex shape of both sliding curves is due to the fact that the sliding velocity increases as the length of the double-stranded region decreases, $v \sim \eta^{-1} \sim N^{-1}$ (see above).

What is the physical mechanism that sets the waiting timescale (τ_w)? Clearly, sliding can begin only after all mutated basepairs have been broken, since otherwise the two strands are locked into one relative position. Arguably the simplest scenario would be that all mutations independently fluctuate between the open and closed states, and sliding commences when all mutations happen to be open simultaneously. Alternatively, the dynamics of the mutated basepairs could be correlated. To clarify the dynamical mechanism, we plotted the binding state (bound/unbound) of all mutated basepairs alongside the trajectories in Fig. 4 (bottom, shaded curves). It is evident that the mutations do not open independently. Instead, interior mutations open only once the neighboring mutation toward the exterior has already opened.

The two-random-walkers model

Inspection of Fig. 4 (bottom) suggests that the positions of the two outermost-bound mutations might, in fact, perform

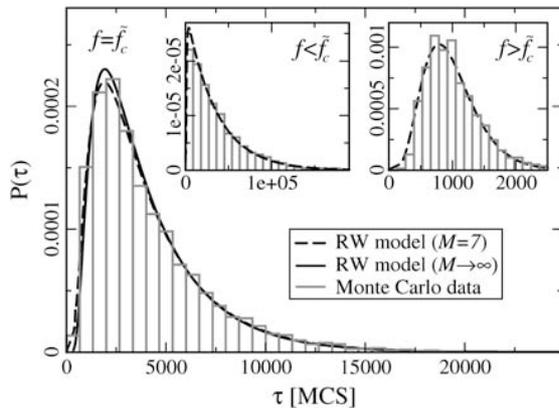
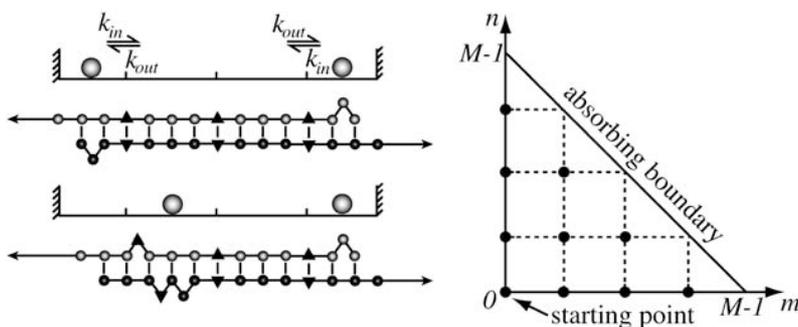


FIGURE 5 Waiting time distributions. (*Main panel*) The histogram of waiting times τ_w of a 120-bp-long DNA sequence with $M = 7$ equidistant mutations subject to a force $f = \tilde{f}_c = 12.9$ pN, is well described by the distribution of collision times (*dashed line*) of the two-random-walker model (see main text and Fig. 6). The solid line shows the parameter-free asymptotic distribution of Eq. 5 for comparison. (*Insets*) Distribution of τ_w for forces above and below \tilde{f}_c ($f = 15.2$ pN and $f = 11.4$ pN, respectively). The dashed lines are fits using the RW model with directional bias (see main text).

a (biased) random walk (see the *black curves* in Fig. 4). If true, the waiting time τ_w could be interpreted as the first collision time τ_c of two random walkers (2RW) on a row of $M + 1$ discrete sites, with force-dependent in- and outward hopping rates k_{in} and k_{out} . To test this hypothesis, we compare the histogram of τ_w values (from many simulations) with the distribution $P(\tau_c)$ of first collision times for 2RW. Fig. 5 shows three such histograms (*main panel* and *two insets*) obtained with the same DNA parameters as in Fig. 3, but with three different forces. Superimposed are the distributions $P(\tau_c)$, calculated as described below and in Supplementary Material. The case shown in the main panel of Fig. 5 corresponds to a force value for which the 2RW are unbiased—i.e., $k_{in} = k_{out}$ —whereas the left inset corresponds to a smaller force producing a bias to the outside ($k_{out} > k_{in}$) and the right inset shows the opposite case of a larger force and $k_{out} < k_{in}$. In all three cases, the observed histogram is well described by the 2RW model. Indeed, despite some caveats (see below), this model can serve as a useful coarse-grained description for the basepairing dynamics preceding the sliding stage.



The calculation of the first collision time distribution $P(\tau_c)$ belongs to the class of first-passage problems, which has been studied extensively in statistical physics (28). In the context of the helix-coil transition, Schwarz and Poland (29) (see also (30)) already solved the associated diffusion problem. Here, we use their work as a basis to treat the first-passage problem. One can replace the problem of 2RW in one dimension by the equivalent problem of one RW on a two-dimensional lattice with a triangular shape (see Fig. 6). In the following, the unbiased case ($k_{in} = k_{out} \equiv k$) is of particular interest. In this case, there is only the single rate constant k , which can be absorbed in the unit of time, so that the distribution $P(\tau_c)$ depends only on the number of lattice points (i.e., the number of mutations). However, in the limit of large M this dependence also disappears, if we use the rescaled collision time $\tilde{\tau}_c = \tau_c k / M^2$. The resulting parameter-free distribution can be expressed in the form (see Supplementary Material) of

$$P(\tilde{\tau}_c) = -\frac{16}{\pi^2} \frac{\partial}{\partial t} [Q(t)]^2 \Big|_{t=\tilde{\tau}_c}, \quad (5)$$

where $Q(t)$ is the rapidly converging series

$$Q(t) = \sum_{n=1}^{\infty} \frac{(-1)^n}{2n-1} \exp\left(-\frac{\pi^2}{2}(2n-1)^2 t\right). \quad (6)$$

This distribution is plotted as the solid line in the main panel of Fig. 5. Even when M is small, it is a good approximation to the actual distribution, as illustrated by the dashed line in Fig. 5 showing the exact distribution for the case of $M = 7$.

In the case of biased RW ($k_{in} \neq k_{out}$), we compute $P(\tau_c)$ numerically. The dashed curves in the insets of Fig. 5 show these distributions for $M = 7$ mutations, where we have used the rates k_{in} , k_{out} as fit parameters.

Scaling of mean waiting time

In the 2RW model, the mean first collision time follows the diffusive behavior $\tau_c \sim M^2$ when the RW are unbiased (see above). When the walkers have an inward bias, this changes to linear scaling $\tau_c \sim M$, whereas τ_c increases exponentially with M for an outward bias (see Supplementary Material). To test these predictions of the 2RW model, we determined the

FIGURE 6 On a coarse-grained level, the dynamics of mutation opening/closing can be described by a model of two particles hopping on a one-dimensional lattice, with inward/outward hopping rates k_{in} , k_{out} . Their positions represent the two outermost closed mutations. When the particles collide, all mutations have opened. Equivalently, one can consider a single particle hopping on a triangular two-dimensional lattice. The first collision time then corresponds to the time to reach the diagonal absorbing boundary.

mean waiting time $\langle\tau_w\rangle$ for different M and different forces f from our DNA simulations. Fig. 7 *a* shows $\langle\tau_w\rangle$ as a function of M (on a double logarithmic scale) for the same three force values as in Fig. 5. Here, we increased the total DNA length N proportional to M , to keep the mutation density constant and equal to that of Fig. 5. At the smallest force, the waiting time increases exponentially with M , as expected. At the intermediate force, corresponding to the case of unbiased walkers, we find a scaling $\langle\tau_w\rangle \sim M^\zeta$ with $\zeta \approx 2.4$, while $\zeta \approx 1.5$ for the largest force. We expect that the values of these exponents are strongly influenced by finite size effects, since we can vary M only over roughly one decade. However, our results indicate that the waiting times increase more rapidly with the system size than expected on the grounds of our phenomenological 2RW model. A possible explanation is given in Microscopic Mechanism, below.

How does the mean waiting time depend on the applied force? Fig. 7 *b* shows three curves of $\langle\tau_w\rangle$ versus f for

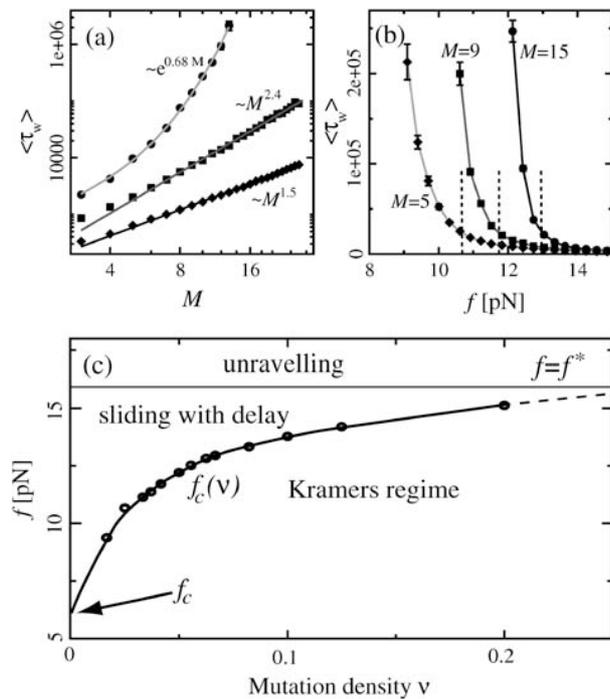


FIGURE 7 (a) The mean waiting time $\langle\tau_w\rangle$ as a function of the system size (the mutation density of $\nu = 1/15$ is kept fixed as the number of evenly spaced mutations M is increased). At low forces the scaling is exponential (circles, data for $f = 11.4$ pN; solid line, exponential fit), while we find power-law behavior at the force threshold ($f_c = 12.9$ pN, squares) and above ($f = 15.2$ pN, diamonds). (b) The mean waiting time as a function of the applied force for a sequence of $N = 240$ bp with 5, 9, and 15 mutations. The dashed lines indicate the threshold force f_c for each case. Below the threshold, $\langle\tau_w\rangle$ rises sharply. (c) Different regimes of the DNA dynamics in the parameter space (f, ν). The Kramers regime (DNA rupture becomes exponentially slow with increasing system size) is separated from the (delayed) sliding regime by the line $f_c(\nu)$ where the inward and outward hopping rates are equal, $k_{in} = k_{out}$ (circles, data; solid line, interpolation). At forces larger than f^* , the molecule dissociates by unraveling from both ends.

different mutation densities. The vertical dashed lines indicate the force value where $k_{in} = k_{out}$ for each curve. Below these values, $\langle\tau_w\rangle$ increases sharply with decreasing force. Indeed, it is reasonable to consider the force f_c where $k_{in} = k_{out}$ as a generalization of the critical force f_c to the case of weakly disordered sequences. As explained in Supplementary Material, the rates k_{in}, k_{out} can be extracted in several different ways from the simulation data, leading to f_c values which are mutually compatible.

Fig. 7 *c* summarizes the different dynamical regimes as a function of the applied force f and the mutation density ν . Without mutations ($\nu = 0$) the force axis is divided into three regimes, with rupture driven by rare fluctuations, continuous sliding, and unraveling at low, intermediate, and large forces, respectively. As mutations are introduced ($\nu > 0$), the boundary $f_c(\nu)$ between the fluctuation-driven Kramer's regime and the sliding regime rises to larger forces, and the sliding regime acquires the time delay of Fig. 4 as a new feature. It is clear from Fig. 7 *c* that the force interval displaying sliding behavior becomes narrower as the mutation density is increased. This trend can be understood within a more microscopic picture (see below). We could not determine unambiguously whether the sliding regime vanishes completely already at a finite mutation density. However, it is clear that sliding will, in practice, be unobservable for sequences with many mutations. The qualitative features depicted in Fig. 7 *c* are robust against variations in our microscopic parameters $\varepsilon_b, \varepsilon_\ell, \bar{b}_s(f)$, and $\bar{b}_d(f)$. However, the positions of the boundaries between the different regimes depend on these parameters (see below).

Microscopic mechanism

Why does the mutation dynamics of Fig. 4 (bottom) resemble the behavior of two random walkers? First, the opening of a mutation (and subsequent local shift of the two strands against each other) is always associated with the formation of two permanent loops (see Fig. 6, left). Hence, the opening of mutations is energetically expensive and mutations remain mostly closed, as long as this cost is not compensated by any gain in entropy or stretching energy. Since there is no such gain when an interior mutation opens, mutations can only open beginning from the ends toward the inside: loops are constantly created at the ends of both strands and propagate inwards until they hit a mutation, which forms a barrier to bulge-loop diffusion. On the unstretched strand, loops are generated at a higher rate than on the stretched strand, resulting in a larger quasi-equilibrium loop density (18). When the outermost bound mutation opens spontaneously, the accumulated loops on the exterior unstretched strand can suddenly penetrate to the inside. This penetration results in an entropy gain and a relative shift of the mutated bases, which prevents immediate recombination (see Fig. S10 in Supplementary Material). The size of the entropy gain and the shift

increases with the distance to the next mutation. Therefore, the mutation density, not the absolute number, is the relevant parameter that determines the relative magnitude of the hopping rates k_{in} and k_{out} in the random-walker model, and hence fixes the value of the force threshold \tilde{f}_c .

We now discuss how parameter changes affect the location of the boundaries between the different dynamical regimes in Fig. 7 *c*. First, it is clear that increasing the basepairing energy ε_b , will shift both force thresholds, f_c and f^* , toward higher forces. Furthermore, from the above microscopic picture, it follows that the inward hopping rate k_{in} is proportional to the average loop density, while the outward hopping rate k_{out} decreases with the loop density. The average loop density, in turn, is affected by our energy parameters: with increasing ε_b , ε_l , the average loop density decreases, and consequently the boundary $\tilde{f}_c(\nu)$ is shifted toward lower mutation densities, i.e., the sliding regime becomes more sensitive to mutations (this tendency is enhanced by the rising energetic cost for opening a mutation).

So far, we considered only mutations with binding energy equal to the original bases. Dropping this restriction leads to a sloped boundary $f^*(\nu)$ between the sliding and unraveling regimes, and also affects the slope of $\tilde{f}_c(\nu)$. Furthermore, we assumed above that all mutations are of a different type and bind only to their native binding partner. Without this assumption, bases belonging to different mutated basepairs can bind on encounter during the sliding phase. These basepairs have to be opened in the same way as during the waiting phase preceding sliding. When mutations are equidistant, this effect becomes particularly strong, leading to additional intervals of constant length, i.e., plateaus in the extension versus time-trace. Another important effect, caused both by variable spacing and energies of mutations, is that the hopping rates k_{in} and k_{out} become site-dependent, so that the random walks are effectively on a rugged-energy landscape (31,32).

Finally, we stress that the 2RW model is phenomenological and fails to describe certain features of the DNA dynamics. (For instance, our simple description has neglected correlations between subsequent hopping steps of an RW; see Supplementary Material.) Short-range correlations do not affect the long-time behavior, which may explain why our model describes the shape of the waiting-time distribution accurately (see Fig. 5). A more drastic approximation is that the 2RW model does not account for the time required to bring in new loops from the ends to a mutation deep inside the dsDNA. The fact that this time increases with the length of the DNA may be the cause for the waiting time to rise more rapidly with the system size than expected from the 2RW model (see Fig. 7).

CONCLUSIONS AND OUTLOOK

The basepairing dynamics in DNA and RNA molecules is only beginning to be explored. Here, we have shown that

even the seemingly simple case of periodic DNA sequences displays rich behavior, which can be revealed by applying a shear force. Our main finding is that the microscopic dynamics of bulge-loop defects endows DNA with viscoelastic properties, which can be programmed into the sequence. Weak sequence disorder does not abolish these properties, but 1), introduces a delay, since all mutations have to be broken before DNA sliding begins, and 2), effectively narrows the viscoelastic force regime. The dynamics of mutation breaking is an interesting stochastic process, with main features that can be understood by considering a first-passage problem of two random walkers. Our theoretical study has led to several experimental ramifications. For instance, we predict that periodic or nearly periodic DNA responds to sudden stress by slowly relaxing its tension to a threshold value independent of the initial stress (provided the DNA is not too short). This stress relaxation process cannot occur for heterogeneous DNA. Furthermore, we predict that the relaxation velocity is inversely proportional to the DNA length, so that the timescale of the dynamics can be easily adjusted into the range of interest for a given experimental setup. We expect the existence of the different dynamical regimes shown in Fig. 7 *c* to be independent of our detailed model assumptions. As DNA slippage is directly linked to the production rate and mobility of bulge loops, single-molecule experiments on DNA sliding would test our basic understanding of basepairing dynamics in DNA.

The same properties, which make DNA uniquely suited for reliably storing genetic information while keeping it accessible, permit many applications in nanotechnology (5). For instance, dsDNA has been used as a reversible cross-linker in polymer networks to switch between different mechanical properties (7), and even DNA-only networks with specified topologies can be constructed, exploiting the specificity of the basepairing interaction (5). In other applications, short dsDNA molecules served as programmable force sensors (6) using the sequence-dependence of the mechanical rupture force, or DNA-based nanomachines were constructed on the basis of the DNA branch migration mechanism (33). Our results render several new applications for DNA in nanotechnology conceivable. For instance, complementary periodic ssDNAs could be used as self-tightening connections in nanostructures: once two such strands found each other, they will slide to maximize their overlap until the tension reaches a value f_c . Periodic or nearly periodic DNA could also serve as a viscoelastic crosslinker in polymer networks, which should lead to different material properties from those observed in Lin et al. (7). Similarly, DNA networks could also be endowed with viscoelastic properties, and (nearly) periodic DNA might even be useful as a programmable reference molecule for kinetic measurements. Of course, which of these and other possible applications will turn out to be useful in the end is unclear at the present stage. However, we feel that there is a clear potential that should be explored.

SUPPLEMENTARY MATERIALS

An online supplement to this article can be found by visiting BJ Online at <http://www.biophysj.org>.

We thank T. Hwa, K. Kroy, F. Kühner, and F. Simmel for useful discussions.

We thank the Deutsche Forschungsgemeinschaft for financial support.

REFERENCES

- SantaLucia, J., Jr. 1998. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. USA.* 95:1460–1465.
- Hagerman, P. J. 1990. Sequence-directed curvature of DNA. *Annu. Rev. Biochem.* 59:755–781.
- Widom, J. 2001. Role of DNA sequence in nucleosome stability and dynamics. *Q. Rev. Biophys.* 34:269–324.
- Rippe, K., P. H. von Hippel, and J. Langowski. 1995. Action at a distance: DNA-looping and initiation of transcription. *Trends Biochem. Sci.* 20:500–506.
- Seeman, N. C. 2003. DNA in a material world. *Nature.* 421:427–431.
- Albrecht, C., K. Blank, M. Lalic-Multhaler, S. Hirler, T. Mai, I. Gilbert, S. Schiffmann, T. Bayer, H. Clausen-Schaumann, and H. E. Gaub. 2003. DNA: a programmable force sensor. *Science.* 301:367–370.
- Lin, D. C., B. Yurke, and N. A. Langrana. 2004. Mechanical properties of a reversible, DNA-crosslinked polyacrylamide hydrogel. *J. Biomech. Eng.* 126:104–110.
- Bustamante, C., J. C. Macosko, and G. J. Wuite. 2000. Grabbing the cat by the tail: manipulating molecules one by one. *Nat. Rev. Mol. Cell Biol.* 1:130–136.
- Clausen-Schaumann, H., M. Seitz, R. Krautbauer, and H. E. Gaub. 2000. Force spectroscopy with single bio-molecules. *Curr. Opin. Chem. Biol.* 4:524–530.
- Merkel, R. 2001. Force spectroscopy on single passive biomolecules and single biomolecular bonds. *Phys. Rep.* 346:343–385.
- Lavery, R., A. Lebrun, J.-F. Allemand, D. Bensimon, and V. Croquette. 2002. Structure and mechanics of single biomolecules: experiment and simulation. *J. Phys. Cond. Mater.* 14:R383–R414.
- Bockelmann, U., P. Thomen, B. Essevaz-Roulet, V. Viasnoff, and F. Heslot. 2002. Unzipping DNA with optical tweezers: high sequence sensitivity and force flips. *Biophys. J.* 82:1537–1553.
- Danilowicz, C., V. W. Coljee, C. Bouzigues, D. K. Lubensky, D. R. Nelson, and M. Prentiss. 2003. DNA unzipped under a constant force exhibits multiple metastable intermediates. *Proc. Natl. Acad. Sci. USA.* 100:1694–1699.
- Altan-Bonnet, G., A. Libchaber, and O. Krichevsky. 2003. Bubble dynamics in double-stranded DNA. *Phys. Rev. Lett.* 90:138101.
- Lubensky, D. K., and D. R. Nelson. 2002. Single molecule statistics and the polynucleotide unzipping transition. *Phys. Rev. E.* 65:031917.
- Strunz, T., K. Oroszlan, R. Schäfer, and H.-J. Güntherodt. 1999. Dynamic force spectroscopy of single DNA molecules. *Proc. Natl. Acad. Sci. USA.* 96:11277–11282.
- Pörschke, D. 1974. Model calculations on the kinetics of oligonucleotide double helix coil transitions. Evidence for a fast chain sliding reaction. *Biophys. Chem.* 2:83–96.
- Neher, R. A., and U. Gerland. 2004. Dynamics of force-induced DNA-sliding. *Phys. Rev. Lett.* 93:198102.
- Reddy, P. S., and D. E. Housman. 1997. The complex pathology of trinucleotide repeats. *Curr. Opin. Cell Biol.* 9:364–372.
- Lovett, S. T. 2004. Encoded errors: mutations and rearrangements mediated by misalignment at repetitive DNA sequences. *Mol. Microbiol.* 52:1243–1253.
- Hallatschek, O., E. Frey, and K. Kroy. 2005. Propagation and relaxation of tension in stiff polymers. *Phys. Rev. Lett.* 94:077804.
- Smith, S. B., Y. Cui, and C. Bustamante. 1996. Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science.* 271:795–799.
- Flamm, C., W. Fontana, I. L. Hofacker, and P. Schuster. 2000. RNA folding at elementary step resolution. *RNA.* 6:325–338.
- Craig, M. E., D. M. Crothers, and P. Doty. 1971. Relaxation kinetics of dimer formation by self-complementary oligonucleotides. *J. Mol. Biol.* 62:383–392.
- Harlepp, S., H. Isambert, D. Chatenay, S. Harlepp, T. Marchal, J. Robert, J. F. Leger, A. Xayaphoummine, D. Isambert, and H. Chatenay. 2003. Probing complex RNA structures by mechanical force. *Eur. Phys. J. E.* 12:605–615.
- de Gennes, P. G. 1971. Reptation of a polymer chain in the presence of fixed obstacles. *J. Chem. Phys.* 55:572–579.
- Zener, C. 1956. Elasticity and Anelasticity of Metals. The University of Chicago Press, Chicago, IL.
- Gardiner, C. W. 2004. Handbook of Stochastic Methods. Springer-Verlag, Berlin, Germany.
- Schwarz, M., and D. Poland. 1975. Random walk with two interacting walkers. *J. Chem. Phys.* 63:557–568.
- Anshelevich, V. V., A. V. Vologodskii, A. V. Lukashin, and M. D. Frank-Kamenetskii. 1984. Slow relaxational processes in the melting of linear biopolymers: a theory and its application to nucleic acids. *Biopolymers.* 23:39–58.
- Derrida, B. 1983. Velocity and diffusion constant of a periodic one-dimensional hopping model. *J. Stat. Phys.* 31:433–450.
- Kafri, Y., D. K. Lubensky, and D. R. Nelson. 2004. Dynamics of molecular motors and polymer translocation with sequence heterogeneity. *Biophys. J.* 86:3373–3391.
- Simmel, F. C., and B. Yurke. 2002. A DNA-based molecular device switchable between three distinct mechanical states. *Appl. Phys. Lett.* 80:883–885.

2.10.1. Supplementary Material to *Biophys. J.*, 89, p. 3846.

Waiting Time Distributions

Two Random Walker Model We consider two random walkers in one dimension confined by two reflecting boundaries $M+1$ sites apart. Since we want to model the process of mutation opening preceding the sliding stage, we seek the distribution of times until encounter of both walkers, given they started at opposite boundaries. Their motion is equivalent to the motion of one walker on a triangular piece of the two dimensional square lattice. The 2D walker on site (m, n) corresponds the state, where the left 1D walker is m steps from the left boundary and the right 1D walker n steps from the right boundary (see Fig. 3, main text). The 2D walker is reflected at the lines $m = 0$ and $n = 0$. The line, where both coordinates add up to $M - 1$ corresponds to the cases, when both walkers in 1D meet and is therefore an absorbing boundary for the 2D walker.

The case, where the rates, at which the walker moves away and towards a boundary (k_{in} and k_{out}) are independent of the site, has been solved by Schwarz and Poland [74] using the methods of image charges.

The quantity we are interested in is the distribution of the time of the first encounter of the two random walkers in 1D, or equivalently the lifetime distribution $P(\tau)$ of the random walker on the triangle. A walker sitting on any site (m, n) with $m = M - 2 - n$ can hop on two absorbing sites with rate k_{in} . The distribution of τ is therefore given by

$$P(\tau) = 2k_{in} \sum_{n=0}^{M-2} \mathcal{P}(n, M-2-n; \tau), \quad (2.6)$$

where $\mathcal{P}(n, m; \tau)$ is the probability of finding the walker on site (n, m) at time τ , given it started at site $(0, 0)$. In the following we derive approximations of the solution by Schwarz and Poland.

Unbiased Hopping When the walker has no bias, e.g. $k_{in} = k_{out} = k$, $\mathcal{P}(n, m; \tau)$ is given by a sum of $4M^2$ terms. The solution by Schwarz and Poland can be rearranged to

$$\mathcal{P}(n, m; \tilde{\tau}) = \frac{1}{M^2} \sum_{r,s=1}^{2M} e^{-2\tilde{\tau}M^2(2 - \cos \frac{\pi r}{M} - \cos \frac{\pi s}{M})} (1 - (-1)^{r+s}) \cos \frac{\pi(2n+1)r}{2M} \cos \frac{\pi r}{2M} \cos \frac{\pi(2m+1)s}{2M} \cos \frac{\pi s}{2M}, \quad (2.7)$$

where the time variable has been rescaled as $\tau = \tilde{\tau}M^2/k$. Only terms, where the argument of the cosines in the exponent are close to 0 or 2π , contribute significantly when $\tilde{\tau} > 1/M^2$. After shifting the summation interval to $r, s = -M \dots M-1$, significant terms are those the r, s close to 0. We can expand cosines with arguments $\frac{\pi r}{M}$ or $\frac{\pi s}{M}$ and keep only the first non-vanishing contribution.

$$\mathcal{P}(n, m; \tilde{\tau}) \approx \frac{1}{M^2} \sum_{r,s=-\infty}^{\infty} e^{-\tilde{\tau}\pi^2(r^2+s^2)} (1 - (-1)^{r+s}) \cos \frac{\pi(2n+1)r}{2M} \cos \frac{\pi(2m+1)s}{2M}. \quad (2.8)$$

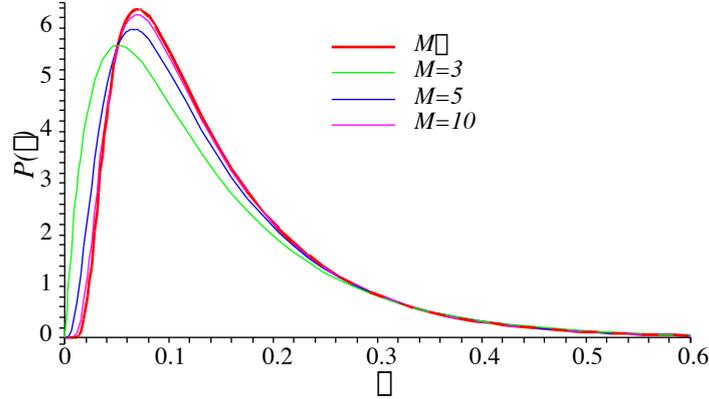


Figure 2.6: The lifetime distributions for $M = 3, 5, 10$ and the approximation for large M . The time axis is rescaled by M^2 .

The range of summation can be safely extended to $\pm\infty$, as terms with big r, s are exponentially small. Plugging this approximation into Eq. (2.6) yields, after some algebra, using similar approximations as above,

$$P(\tilde{\tau}) \approx \frac{2}{M^2} \sum_{r,s=-\infty}^{\infty} e^{-\frac{\tilde{\tau}\pi^2}{M^2}(r^2+s^2)} \frac{(1 - (-1)^{r+s})(r^2 + s^2)}{r^2 - s^2} \quad (2.9)$$

Since only those terms with odd $r + s$ contribute, we change the summation variables to $2v = r + s - 1$ and $2w = r - s - 1$.

$$P(\tilde{\tau}) \approx \frac{4}{M^2} \sum_{v=-\infty}^{\infty} e^{-\frac{\tilde{\tau}\pi^2}{2}(2v-1)^2} \frac{(-1)^v}{2v-1} \sum_{w=-\infty}^{\infty} e^{-\frac{\tilde{\tau}\pi^2}{2}(2w-1)^2} (-1)^w (2w-1) \quad (2.10)$$

From this expression, we find a parameter-free lifetime distribution

$$\tilde{P}(\tilde{\tau}) = M^2 P(\tilde{\tau}) = -\frac{16}{\pi^2} \frac{\partial}{\partial \tilde{\tau}} Q(\tilde{\tau})^2, \quad (2.11)$$

where $Q(\tilde{\tau})$ is given by

$$Q(\tilde{\tau}) = \sum_{n=1}^{\infty} \frac{(-1)^n e^{-\frac{\pi^2(2n-1)^2\tilde{\tau}}{2}}}{2n-1} \quad (2.12)$$

The approximations involved are justified for large M . However, even for small systems the agreement is excellent, as illustrated in Fig. 2.6.

Biased Hopping When the rates k_{in} and k_{out} are different, there is no compact analytical expression for $\mathcal{P}(n, m; t)$. However, the longterm behaviour of a such a biased random walker is easily understood. If k_{in} is bigger than k_{out} , the walker approaches the absorbing boundary steadily. In the opposite case, the walker will stay close to the origin and only

rare excursions will lead to absorption. Quantitatively, the hopping of the random walker on the triangle is well approximated by suitably chosen one-dimensional representation. To that end, we consider the probability to find the walker on the line ν steps away from the origin.

$$P(\nu; \tau) = \sum_{m=0}^{\nu} \mathcal{P}(m, \nu-m; \tau) \quad (2.13)$$

This amounts to projecting the motion of the random walker onto the symmetry axis of the triangle. The time derivative of this quantity is very similar to a one dimensional hopping process.

$$\begin{aligned} \partial_{\tau} P(\nu; \tau) = & -2(k_{in} + k^{-})P(\nu; \tau) + 2k_{in}P(\nu-1; \tau) + 2k_{out}P(\nu+1; \tau) \\ & + k_{out} [\mathcal{P}(0, \nu; \tau) + \mathcal{P}(\nu, 0; \tau) - \mathcal{P}(0, \nu+1; \tau) - \mathcal{P}(\nu+1, 0; \tau)] \end{aligned} \quad (2.14)$$

The contributions from the boundary terms in the second line depend on the ratio of k_{in} and k_{out} . When $k_{in} \gg k_{out}$ the walker rapidly approaches the absorbing boundary. The probability $\mathcal{P}(0, \nu; \tau)$ of finding the walker on the reflecting boundary is small, as it is unlikely to make equally many steps with high rate and a low rate. In this case the boundary terms can be neglected entirely, so that the process reduces entirely to a 1D first passage problem. Using standard methods described in ref. [87], one finds, that the mean first passage time

$$\langle \tau \rangle = \frac{M-1}{2k_{in} - 2k_{out}} - k_{out} \frac{1 - \left(\frac{k_{out}}{k_{in}}\right)^{M-1}}{2(k_{in} - k_{out})^2}, \quad (2.15)$$

increases linearly with the number of mutations M .

In the opposite limit, when $k_{in} \ll k_{out}$, $\langle \tau \rangle$ increases as $\left(\frac{k_{out}}{k_{in}}\right)^{M-1}$ with M . In this case equilibration along the line $n = \nu - m$ is fast compared to the lifetime of the walker and $\mathcal{P}(m, \nu - m; \tau)$ is almost independent of m . Setting all terms $\mathcal{P}(m, \nu - m; \tau)$ equal results in a 1D hopping process with site dependent rates. The mean first passage time of this process can be calculated in much the same way, yielding $\langle \tau \rangle \sim \left(\frac{k_{out}}{k_{in}}\right)^{M-1}$ with polynomial corrections. In summary, we find that, depending on whether the walkers have an inward bias, an outward bias or no bias, the mean lifetime scales linearly, exponentially or quadratically with time. Since the force, at which k_{in} and k_{out} are equal, separates regimes, where the waitingtime increases exponentially with M from linear scaling, we call it critical force \tilde{f}_c in the presence of mutations. The force \tilde{f}_c converges towards the critical force f_c in the limit of no mutations.

Measuring Hopping Rates

So far, we have been concerned with the waitingtime distribution given a certain set of rates, at which mutations open or close. These rates depend on the applied force and on the distance between consecutive mutations and have to be determined in simulations.

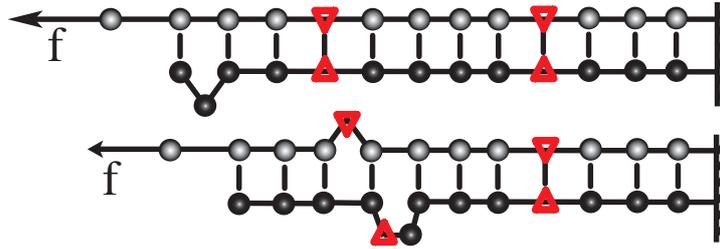


Figure 2.7: Simplified system to measure the opening and closing rates of mutations. The simulation starts from the ground state with all bases bound. We determine the first-passage time distributions of the opening of the rightmost mutation and fit these to a the first passage time distribution of a random walker.

As long as there are at least two mutations bound, the dynamics of the opening and closing of mutations at one end is independent of the other end. To measure the rates for a given pair of force and mutation density, we used a simplified system, where a dsDNA with equidistant mutations is fixed on the right hand side and a force is applied to the first base of the upper strand (see Fig. 2.7). This simplified system is useful, as finite size effects are smaller when one walker crosses M mutations as when two walkers cross $M/2$ mutations each. Furthermore, subtleties of the mutual annihilation process do not enter the measurement. We measure the distribution of the time it takes to open the rightmost mutation for the first time and fit this distribution to the lifetime distribution of a random walker in one dimension between reflecting and absorbing boundary conditions. The rates k_{in} and k_{out} are fit parameters. This is done for a range of forces and mutation densities and the critical force \tilde{f}_c for a certain mutation density can be extract from the crossing of k_{in} and k_{out} . To further pin down \tilde{f}_c , we generated data for many force values slightly above and below \tilde{f}_c and fitted a linear relation for each rate to all data sets simultaneously. The crossing of the two resulting lines yield a robust estimate of \tilde{f}_c . Using a system of $N = 240$ basepairs, energy parameters $\varepsilon_b = 1.11k_B T$, $\varepsilon_\ell = 2.8k_B T$ and different number of equidistant mutations, we determined \tilde{f}_c over broad range of mutation densities. The results are shown in Fig. 7(c) in the main text.

To check the reliability of the estimation of \tilde{f}_c , we simulated waitingtime distributions by applying the force to both ends of the DNA and fitted the two random walker model to the waitingtime distribution. The force, where k_{in} and k_{out} coincide, reproduces the previously determined force \tilde{f}_c . Furthermore, fitting the critical distribution (one fit parameter) to the waitingtime distribution with yields best fits for $f \approx \tilde{f}_c$. The absolute value of the rates shows slight dependencies on the length of the system (see below) and varies for fits to different setups.

Caveats of the Model Equilibration of the loopdensity is only possible by propagation of loops from the end beyond a newly broken mutation, or in other words by sliding the unstretched strand some distance Δd inward. The sliding velocity, however, is inversely

proportional to the length of the strand. Therefore, equilibration will slow down breaking of mutations for supercritical forces deep inside the double strand and the linear dependence of the waitingtime on the number of mutations will not persist for very large systems.

It is clear from the microscopic mechanism leading to breaking and opening of mutations (see main text and Fig. 2.8) that the rates k_{in} and k_{out} depend on the force f . The rate k_{out} also depends on the mutation density, since a great distance between mutations corresponds to a large entropy barrier for mutation closing, and hence a smaller closing rate k_{out} . The microscopic opening rate k_{in} is expected to be more or less independent of the mutation density. When looking at the opening and closing dynamics of an individual mutation, this is what we observe. However, the equilibration of loop densities after an opening or closing event takes some time. Therefore, successive microscopic opening and closing events are not entirely uncorrelated, which makes an unambiguous definition of the microscopic rates difficult. These correlations die out very quickly and it is still possible to describe the observed lifetime distribution with an uncorrelated random walker. The effective rates describing this motion both depend on mutation density and the applied force.

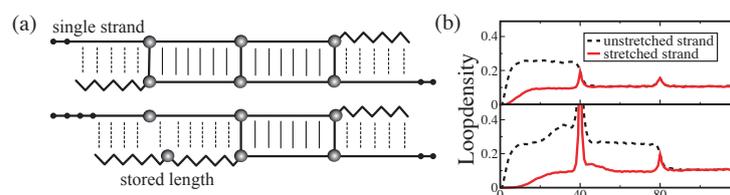


Figure 2.8: Left: Illustration of how the density of loops on the strands depends on the state of the mutated bases in the sequence. In between bound mutations loops are rare, as the formation of a loop costs initiation energy and shortens the system. The same applies to the stretched strands outside bound mutations. The only part, where a significant number of loops can be found, is the unstretched strand outside the bound mutations. When a mutation is broken, loops move across the mutation on the unstretched strand and locally both strands are shifted against each other. Thereby, the bases that previously formed the mutated basepair become permanently separated and the single strand part on the stretched strand grows. Right: To support the cartoon-like picture of part (a), we measured the time averaged loopdensity, conditioned on a certain mutation state. Mutations are located at base 40 and 80, the parameters are $\varepsilon_b = 1.11k_B T$, $\varepsilon_\ell = 2.8k_B T$ and $f = 10.7\text{pN}$. We consider only opening of mutation from the left, i.e. the rightmost base is kept fixed, as in Fig. 2.7. When all mutations are bound (upper panel), the loopdensity is high only on the unstretched strand to the left of the mutation at position 40. When this mutation is broken (lower panel), loops can spread from the left end to the mutation at position 80, yielding a fairly constant density interrupted only by the permanent loop at the position of the mutated base. The hump to the left of the broken mutation on the unstretched strand and to the right of the broken mutation on the stretched strand indicate the position of the mutated base on the opposite strand. A loop already present on one strand renders unbound bases on the other strand more likely, as no additional loop initiation has to be paid. The vanishing loopdensity at the end of the stretched strand indicates unbound ssDNA. Observe, that this is the longer, the more mutations are broken.

Intermediate phase in DNA melting

Richard A. Neher* and Ulrich Gerland

Arnold-Sommerfeld-Center for Theoretical Physics (ASC) and Center for NanoScience (CeNS), LMU München, Theresienstrasse 37, 80333 München, Germany

(Received 14 September 2005; published 23 March 2006)

We predict a temperature-driven phase transition of DNA below the melting transition. The additional, intermediate phase exists for repetitive sequences, when the two strands have different lengths. In this phase, the excess bases of the longer strand are completely absorbed as bulge loops inside the helical region. When the temperature is lowered, the excess bases desorb into overhanging ends, resulting in a contour length change. This continuous transition is in many aspects analogous to Bose–Einstein condensation. When the sequence is weakly disordered, the contour length changes discontinuously with temperature.

DOI: [10.1103/PhysRevE.73.030902](https://doi.org/10.1103/PhysRevE.73.030902)

PACS number(s): 87.14.Gg, 05.70.Fh, 87.15.Aa, 87.15.He

The base-pairing interaction between the two strands of DNA is not only pivotal to its biological function [1], but also leads to intriguing applications in nanotechnology [2]. One approach to probe this interaction is to monitor the DNA conformation as a function of temperature. Experimentally, one can observe the number of base pairs formed (using UV absorption [3,4]), as well as changes of intramolecular distances on the nanometer scale (using modern single-molecule techniques [5]). On the theoretical side, the temperature dependence of DNA conformations has been studied for almost 50 years, using models of various degrees of complexity [6–14]. Particular attention has been paid to the characteristics of the melting transition, where the two strands separate completely. Whereas early models yielded only a crossover [6], the Poland-Scheraga (PS) model [8] was the first to display a phase transition, albeit a continuous one, which appeared to be at variance with the experimentally observed sharp jump in the fraction of bound base pairs [3]. Only recently have mechanisms been proposed [11,12] which yield an abrupt, first-order transition. So far, however, most analyses of DNA melting have incorporated only native interactions, i.e., base pairs that occur in the ground state of the molecule (see [7,13,14] for notable exceptions). It is our aim here to show that such non-native interactions can introduce an intermediate phase in the melting behavior of DNA, associated with an additional conformational transition before strand separation.

Non-native interactions are particularly relevant for repetitive DNA sequences, which are common in genomes [15]. Periodic DNA, with, e.g., a single base repeat such as TTT..., or a higher-order repeat such as CAGCAG..., can take on base-pairing patterns with asymmetric loops and the two complementary strands can be shifted relative to each other. Here, we consider the general situation where the two strands can have arbitrary lengths N , M . We find that for $N \neq M$, the bound phase splits into two separate phases. The low-temperature phase is characterized by an extensive length of the unbound end on the longer strand, whereas in the new intermediate phase these overhanging bases are ab-

sorbed into the helical region. Mathematically, and also conceptually, many aspects of this transition are analogous to Bose-Einstein condensation (BEC), as “particles” (bases) condense into a single “state” (the overhanging end), which thereby acquires macroscopic “occupation” (length). Obviously, the analogy extends only to the behavior of the partition function, as there is no quantum coherence in the DNA problem. Effectively, the transition amounts to a temperature-sensitive change in the contour length of the DNA molecule, which should be observable with optical or single-molecule methods. While the transition is continuous for perfectly periodic sequences, we find that the contour length shortens discontinuously once (weak) sequence disorder is introduced.

DNA model. We consider two DNA strands with lengths N and $M \geq N$, respectively, and describe their interaction with the “imperfect matching” generalization of the PS model [8,13,14,16]. Specifically, a base $i \leq N$ of the lower strand can form a base pair (i, j) with every complementary base $j \leq M$ of the upper strand, whereas the formation of base pairs within a strand can be neglected (since we are interested only in sequences with a high degree of complementarity and a low degree of self-complementarity). Due to geometrical constraints, we may neglect the “crossing” of base pairs, e.g., two base pairs (i_1, j_1) and (i_2, j_2) with $i_1 < i_2$ but $j_1 > j_2$. The base-pairing pattern \mathcal{S} , i.e., the set of all formed base pairs, then creates a DNA conformation consisting of bound segments alternating with (possibly asymmetric) loops, see Fig. 1. To simplify the discussion, we enforce the base pair (N, M) at the right end, so that we need to consider

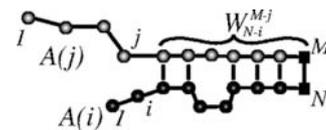


FIG. 1. A possible configuration of two complementary DNA strands with a repetitive sequence (a bead represents one repeat unit). Note that repetitive sequences can form base-pairing patterns with asymmetric loops. In general, we allow for different strand lengths N , M . The last repeat units (squares) are permanently bound.

*Electronic address: richard.neher@physik.lmu.de

only one overhanging end. However, our main findings remain equally valid without this constraint. Experimentally, this constraint could be realized, e.g., by a few particularly strong base pairs at one end.

Each base-pairing pattern \mathcal{S} , receives a statistical weight $\mathcal{Q}(\mathcal{S})$, which takes the form of a product with factors of four different types: (i) a Boltzmann factor $q=e^{\varepsilon_b/k_B T}$ for every base pair with binding energy $-\varepsilon_b < 0$, (ii) a Boltzmann factor $g^2=e^{-\varepsilon_\ell/k_B T}$ for every loop with loop initiation cost $\varepsilon_\ell > 0$, (iii) an entropic factor $B_\ell(m)=s^m m^{-c}$ for each loop, which is the increase in the number of polymer configurations when m bases form a (floppy) loop instead of being in a (rigid) double helical conformation, and (iv) a similar entropic factor $A(n)=s^n n^{-\bar{c}}$ for a single-stranded end of n bases. Here, the exponents c, \bar{c} in the entropic factors are universal in that they are independent of the detailed polymer properties, but are sensitive to excluded volume interactions. For interacting self-avoiding loops one has $c \approx 2.15$, while $\bar{c} \approx 0.1$ [17]. Whereas the value of c determines the critical behavior at the melting transition, the nonuniversal constant s has no qualitative effect on the melting behavior (we use $s=10$ in all numerical examples).

In the following, we first apply the DNA model to perfectly periodic sequences, where each repeat unit can be treated as an effective base with renormalized parameters (we use $\varepsilon_b=6$ and $\varepsilon_\ell=3$ in temperature units, $k_B=1$). We emphasize that our simplistic model for the involved energies and entropies is meant to illustrate the physical phenomena in a transparent way, but leads to an unrealistic temperature scale. With a more detailed description [4], we find that all of the interesting behavior happens at accessible temperatures [18].

Free energy of periodic DNA. To obtain the equilibrium properties of the DNA model, we calculate the partition sum over all base-pairing patterns, $Z_N^M = \sum_{\mathcal{S}} \mathcal{Q}(\mathcal{S})$. By separating the single-stranded ends from the double-stranded part, see Fig. 1, we write Z_N^M as

$$Z_N^M = \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} A(i)A(j)W_{N-i}^{M-j}. \quad (1)$$

Here, W_r^t is the partition function of two complementary and periodic strands of length r and t with the first and last base pair formed. W_r^t obeys this recursion relation

$$W_{r+1}^{t+1} = qW_r^t + g^2q \sum_{\substack{k < r, m < t \\ k+m > 0}} B_\ell(k+m)W_{r-k}^{t-m}, \quad (2)$$

with the initial conditions $W_1^1=q$ and $W_i^i=W_i^1=0$ for $i > 1$. For sequences with weak disorder considered further below, q has to be replaced by q_{r+1}^{t+1} , where $q_r^t=e^{\varepsilon_b/k_B T}$ for original bases at r and s , $q_r^s=e^{\varepsilon_b/k_B T}$ for mutated bases at $r=s$ and $q_r^t=0$ otherwise. We use the recursion relation to calculate Z_N^M and expectation values for finite lengths N, M [13,19]. To extract the thermodynamic behavior in the limit of long strands, we take the z transform $\hat{Z}(x,y)=\sum_{N,M=0}^{\infty} Z_N^M x^N y^M$ and solve for the transformed partition sum. This procedure is equivalent to the method of sequence generating function employed in Ref. [14]. One obtains

$$\hat{Z}(x,y) = \frac{\hat{A}(x)\hat{A}(y)qxy}{1 - qxy + \frac{qg^2xy}{x-y}[y\hat{B}(y) - x\hat{B}(x)]}, \quad (3)$$

where the transforms of the entropic factors are given by $\hat{A}(z)=\phi_c(sz)+1$ and $\hat{B}(z)=\phi_c(sz)$, with the polylogarithm $\phi_c(z)=\sum_{n=1}^{\infty} z^n n^{-c}$.

The Z-transform carried out above amounts to a change from the canonical to the grand canonical ensemble. The transformation variables x,y play the role of fugacities for bases in the lower and upper strands, respectively. However, for the ensuing discussion, it is advantageous to keep the length N of the shorter strand fixed as a reference. Hence, we perform the inverse transformation for the lower strand by contour integration in x , to obtain the partition sum $Z_N(y_0)$ for N bases on the lower strand and the upper strand coupled to a ‘‘nucleotide reservoir’’ with fixed fugacity y_0 . For large N and temperatures below the melting temperature, $Z_N(y_0)$ is given by $\hat{A}(y_0)x^*(y_0)^{-N}$, where $x^*(y_0)$ is the smallest real zero of the denominator of Eq. (3) for given $y=y_0$ [14]. Hence, the free energy of the bound phase is given by $Nf_b(y_0) - T \ln \hat{A}(y_0)$, where the first term is the contribution of the helical region with a free energy per length $f_b(y_0) = T \ln x^*(y_0)$, and the second term is the contribution from the unbound end of the longer strand. The free energy for given N and M is then obtained by saddle point integration,

$$\frac{F(T,N,M)}{T} = -\ln \hat{A}(y_0) + N \frac{f_b(y_0)}{T} + M \ln(y_0), \quad (4)$$

where the fugacity y_0 is determined by

$$M = \langle M \rangle_{y_0} = y_0 \frac{\partial \ln \hat{A}(y_0)}{\partial y_0} - N \frac{y_0}{T} \frac{\partial f_b(y_0)}{\partial y_0}. \quad (5)$$

Phase diagram. To extract the physical behavior of the DNA model from Eqs. (4) and (5), we focus on two observables, the total number of base pairs, $N\theta$, and the length of the single-stranded overhang. The fraction θ of bound base pairs is calculated from the free energy per length of the helical region as

$$\theta = - \frac{q}{T} \frac{\partial f_b(y_0)}{\partial q}. \quad (6)$$

To obtain the overhang length, we note that the right-hand side of (5) decomposes the total length M of the upper strand into two contributions, where the first term is the expected overhang length and the second term corresponds to the number of bases in the helical region. The dashed line in Fig. 2 (top) shows the overhang length as a function of temperature, for $N=1000$ and $M=1150$. At low temperatures, the two DNA strands are completely aligned, so that all $M-N$ excess bases of the longer strand form an overhanging end. However, we observe that the overhang length decreases with increasing temperature, dropping almost to zero before it rises again sharply at even higher temperature. We see in Fig. 2 (bottom) that this drop occurs in a temperature range where almost all possible base pairs are formed, and the rise

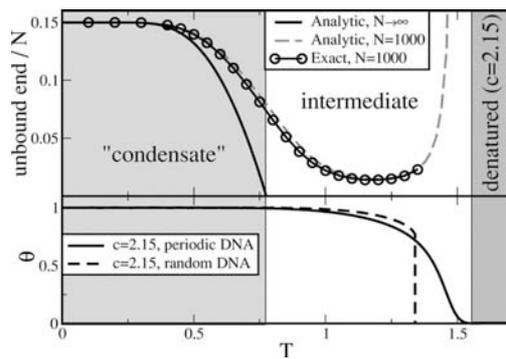


FIG. 2. Top: The length of the unbound end, normalized by the number of excess bases $N-M$ on the longer strand. For finite systems ($N=1000$, dashed line), the unbound end shrinks to a minimal value and increases again, as the melting temperature is approached. In the $N \rightarrow \infty$ limit, the overhang length diverges below $T_c=0.7752$ and is of order 1 for $T > T_c$. Expectation values calculated numerically using Eq. (2) agree well with the analytic result. Bottom: The fraction of bound base pairs θ as a function of temperature. For periodic sequences with $c=2.15$, θ vanishes with zero slope at $T_m=1.424$, whereas a random sequence shows a first-order phase transition.

occurs when the two strands separate. These observations suggest that a temperature-driven conformational transition occurs before the melting transition.

This transition is in fact completely analogous to BEC, as Eq. (5) parallels the behavior of the equation of state for an ideal Bose gas: If we divide Eq. (5) by our system size N and introduce the “particle density” $\alpha = M/N$, we obtain

$$\alpha = \frac{1}{N} \frac{\phi_{c-1}(s y_0)}{\phi_c(s y_0) + 1} + \bar{\alpha}(y_0), \quad (7)$$

where $\bar{\alpha}(y_0) = -(y_0/T)[\partial f_b(y_0)/\partial y_0] \geq 1$ is the density inside the helical region. In Eq. (7), the first term on the right-hand side corresponds to the occupation of the ground state of an ideal Bose gas, whereas $\bar{\alpha}(y_0)$ is analogous to the occupation of the excited states. The fugacities of a Bose gas and our DNA are bounded: for the former, by the energy of the ground state, and for the DNA by the weight of an unbound monomer, i.e., $y_0 \leq s^{-1}$. The population of the excited states increases monotonically with the fugacity, and attains a finite maximal value, in our case $\bar{\alpha}_{max} = \bar{\alpha}(s^{-1})$, provided the loop exponent $c > 2$ [22]. When the temperature is lowered, $\bar{\alpha}_{max}$ decreases (see bottom panel of Fig. 3), and when it falls below α , the length of the unbound end must become extensive to accommodate the remaining bases. In an analogous way, the ground state of a Bose gas is macroscopically populated at low temperatures. In this “condensate” phase, the fugacity is locked to the value s^{-1} in the thermodynamic limit ($N, M \rightarrow \infty$, $\alpha = \text{const.}$). The deviation for finite systems scales as $s^{-1} - y_0 \sim 1/N$, see Fig. 3 (top). In the opposite case, where $\alpha < \bar{\alpha}_{max}$, there is a solution to Eq. (7) with $y_0 < s^{-1}$ and the unbound end remains finite for all system sizes.

It is easily shown that $\bar{\alpha}_{max}$ approaches 1 at low temperatures, and consequently all excess bases of the longer strand are condensed in the overhang, as illustrated in Fig. 3 (bot-

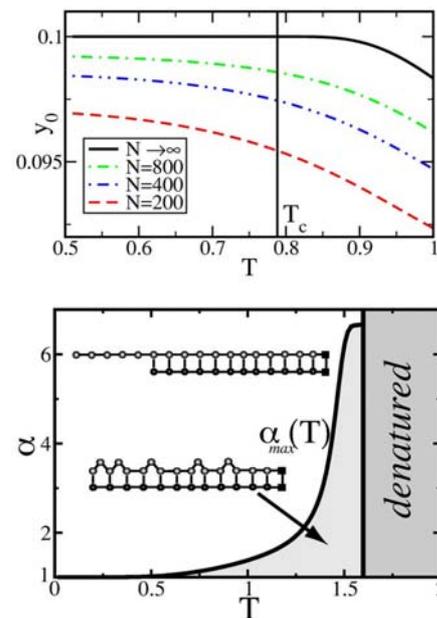


FIG. 3. (Color online) Top: The fugacity y_0 vs T for different system sizes N . In the thermodynamic limit, $y_0 = s^{-1}$ for $T < T_c$. As for BEC, y_0 approaches its limiting value as $s^{-1} - y_0 \sim 1/N$. Bottom: Phase diagram of periodic DNA. At low temperatures, both strands are completely aligned and excess bases of the longer strand form an unbound end. In the intermediate phase, all excess bases are absorbed into the helical region.

tom). As T increases, more and more bases are absorbed in the helical region ($\bar{\alpha}_{max}$ increases), and the system enters the intermediate phase at $T = T_c$, where $\bar{\alpha}_{max} = \alpha$. At T_c the condensate fraction vanishes, as the solid line shows in Fig. 2 (top). Note that the intermediate phase exists only when α is not too large.

The melting temperature T_m , where the strands separate and θ vanishes (denatured phase), is independent of α . For periodic sequences, the loop size distribution at $T = T_m$ decays as $\sim n^{-(c-1)}$ instead of $\sim n^{-c}$, since n bases of a loop can be distributed in $n+1$ ways among both strands [8]. Hence, periodic DNA displays a continuous melting transition for $2 < c \leq 3$ and a first-order transition only if $c > 3$. For $2 < c \leq 3$, we obtain $\theta \sim |T - T_m|^{(3-c)/(c-2)}$, using the same method as [20] for the standard PS model. To illustrate this, we plot θ for periodic sequences and for the standard PS model in Fig. 2 (bottom). Whereas for the latter θ drops discontinuously to zero, θ of periodic DNA vanishes with zero slope.

Weak sequence disorder. Is the intermediate phase identified above robust against sequence disorder? To address this question, we replace a small fraction of base pairs by bases that can pair with each other ($\bar{\epsilon}_b = 2$), but not with other bases in the sequence. Figure 4 shows the average overhang length calculated using the generalized Eq. (2) for sequences with evenly spaced mutations every $d = 25, 50$, and 100 bases. The unbound end keeps its ground-state length up to a certain temperature, and then shortens rapidly. The temperature at which the drop occurs increases with the density of mutations. The width of the transition region scales inversely with the system size N , see Fig. 4(b). These numerical observa-

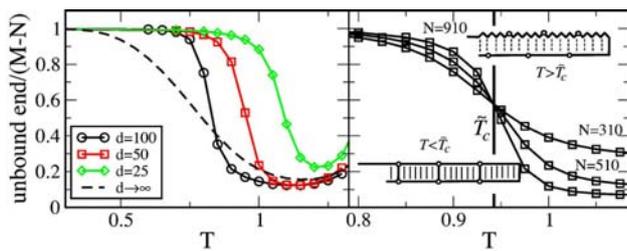


FIG. 4. (Color online) Sequence disorder renders the transition to the intermediate phase discontinuous. Both panels show the average length of the unbound end, normalized by $M-N$. The left panel shows data for different mutation densities (no mutations, mutations every $d=100$, 50, and 25 bases) for $N=510$ and $\alpha=1.15$. The right panel shows data with $d=50$ for different system sizes. The transition sharpens with increasing N and the slope of the curves in the transition region is compatible with $\sim N$.

tions suggest that in the limit of infinite system size, the transition becomes an abrupt first order transition.

Indeed, the first-order behavior can also be understood theoretically by comparing the energy barriers for forming bulge loops with and without mutations: The formation of a bulge loop on the longer strand of a perfectly periodic molecule requires only the initiation energy ε_ℓ . In the presence of mutations, however, shifting both strands breaks mutated base pairs. Hence, to form a bulge loop, all mutations to the left of the loop have to be broken and the energy barrier becomes extensive. For a sufficiently low density of muta-

tions, there is a temperature \tilde{T}_c at which the entropy gained by distributing excess bases in loops along the molecule outweighs the energetic costs to break all mutations [21]. Below \tilde{T}_c all mutations are bound, if $T > \tilde{T}_c$ as many mutations open, as are necessary to absorb all excess bases.

Discussion and outlook. We have identified a BEC-like conformational transition in periodic DNA, which occurs below the melting transition. The hallmark of this transition is the shortening of the unbound end, which could be directly observed experimentally by resonant energy transfer between fluorescent dyes located at the ends of the two strands. The transition is also associated with a change in the contour length of the DNA molecule, roughly proportional to $M-N$. The increased density of bulge loops in the helical region may additionally yield an effect on the persistence length. We found that the existence of the intermediate phase is robust against weak sequence disorder and expect that it is also independent of the details of our model. The temperature range of the intermediate phase rapidly narrows as the mutation density is increased. We conjecture that the intermediate phase disappears completely at a finite mutation density. The exploration of the critical behavior in the complete temperature—mutation density phase diagram is left as an interesting theoretical challenge for the future.

We are grateful for important comments by E. Frey and H. Wagner. We acknowledge financial support by the Deutsche Forschungsgemeinschaft through the Emmy Noether Program.

- [1] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell* (Garland Science, New York, 2002).
- [2] N. C. Seeman, *Nature (London)* **421**, 427 (2003).
- [3] R. M. Wartell and A. S. Benight, *Phys. Rep.* **126**, 67 (1985).
- [4] J. SantaLucia, Jr., *Annu. Rev. Biophys. Biomol. Struct.* **33**, 415 (2004).
- [5] X. Zhuang and M. Rief, *Curr. Opin. Struct. Biol.* **13**, 88 (2003).
- [6] B. H. Zimm and J. K. Bragg, *J. Chem. Phys.* **31**, 526 (1959).
- [7] T. L. Hill, *J. Chem. Phys.* **30**, 383 (1959).
- [8] D. Poland and H. A. Scheraga, *Theory of Helix-Coil Transitions in Biopolymers* (Academic, New York, 1970).
- [9] M. Peyrard and A. R. Bishop, *Phys. Rev. Lett.* **62**, 2755 (1989).
- [10] D. Cule and T. Hwa, *Phys. Rev. Lett.* **79**, 2375 (1997).
- [11] Y. Kafri, D. Mukamel, and L. Peliti, *Phys. Rev. Lett.* **85**, 4988 (2000).
- [12] N. Theodorakopoulos, T. Dauxois, and M. Peyrard, *Phys. Rev. Lett.* **85**, 6 (2000).
- [13] T. Garel and H. Orland, *Biopolymers* **75**, 453 (2004).
- [14] A. Litan and S. Lifson, *J. Chem. Phys.* **42**, 2528 (1965).
- [15] S. T. Lovett, *Mol. Microbiol.* **52**, 1243 (2004).
- [16] R. A. Neher and U. Gerland, *Phys. Rev. Lett.* **93**, 198102 (2004).
- [17] Y. Kafri, D. Mukamel, and L. Peliti, *Eur. Phys. J. B* **27**, 135 (2002).
- [18] R. A. Neher and U. Gerland (unpublished).
- [19] R. Bundschuh and T. Hwa, *Phys. Rev. E* **65**, 031903 (2002); V. Guttal and R. Bundschuh, *Phys. Rev. Lett.* **96**, 018105 (2006).
- [20] M. E. Fisher, *J. Chem. Phys.* **45**, 1469 (1966).
- [21] R. A. Neher and U. Gerland, *Biophys. J.* **89**, 3846 (2005).
- [22] For $y_0=s^{-1}$ and $x^*(y_0)<s^{-1}$, the loop size distribution decays as $\sim n^{-c}$. Since the mean loop size diverges for $c<2$, there is never a condensed phase in this case.

3. Dynamics of nucleosomal DNA

Every cell of a multicellular organism carries the complete genetic information in its genome, regardless of its specific role as a part of the whole. Since different cells, *e.g.* a liver cell and a neuron, need different proteins to function, cells are in need of a mechanism to control which part of the genome is transcribed and which genes are silent. This not only applies to specialized cells of multicellular organisms, but to every cell we know. Even the simplest bacteria need to adjust their metabolism to the available nutrients and require different proteins at different stages of the cell cycle. Regulation of protein production can occur either before the gene is transcribed into mRNA, or target the process of the translation of mRNA into protein. To regulate transcription, the DNA contains short sequences that bind specifically to proteins called *transcription factors* (TF). These transcription factors repress or enhance the binding of the polymerase to the promoter, which is a prerequisite for transcription. The expression levels of the cell's genes is thereby controlled by the concentrations of transcription factors in the cell.

Eukaryotic cells suffer from an additional difficulty in achieving this feat. To fit their DNA into the cell's nucleus, the DNA needs to be strongly compactified. Due to this compactification the DNA is no longer freely accessible and transcription factor binding to DNA is to some extent precluded. The precise mechanisms of transcription regulation in eukaryotes are unknown, but there is evidence that the compactified DNA is dynamic enough and exposes each part of its genome sufficiently often to allow for TF binding. On the other hand, cells exploit DNA compactification to silence subsets of their genes and to determine cell fates during development. The dynamics of the elementary compactification unit of eukaryotic DNA, the nucleosome, has recently been tested experimentally. In order to understand how the observed dynamics depends on various parameters of the system and what physical mechanisms might be responsible for the observed behavior, we investigate the dynamics of nucleosomes theoretically using a simple model. Within our model, the dynamics of nucleosomes depends drastically on the polymer properties of DNA, which could also hold true for their dynamics *in vivo*.

In this chapter, I want to discuss the basics of chromatin structure and its implications for gene regulation in eukaryotes. Then we will discuss two recent experiments, that studied the dynamics of nucleosomes and close with a discussion of our theoretical study.

3.1. DNA compactification

While bacteria have small genomes and avoid superfluous DNA, eukaryotes and in particular higher multicellular organisms need more room to store their genetic information. In

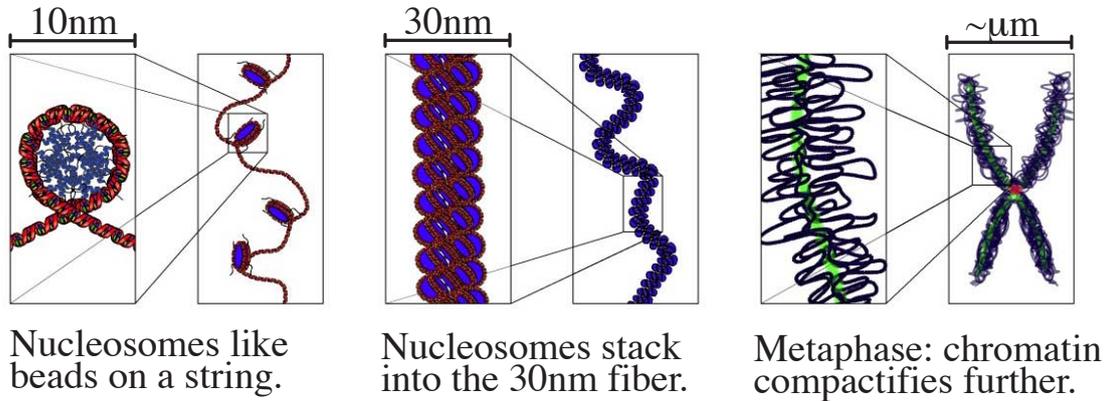


Figure 3.1: DNA is compactified into chromosomes in a hierarchic manner. See main text for details. Image source: Wikipedia.

addition to the greater number of proteins that need to be coded, eukaryotes also tend to accumulate DNA that does not code for proteins and whose function is still unclear. In any case, the genomes of eukaryotes can be as big as a few gigabases for higher mammals. Even if broken up into several chromosomes, a DNA coil of that length is several tens of micrometers in diameter, which is larger than the cell nucleus. Hence, there is a need for compactification, which is achieved by an elaborate hierarchical organization of the DNA into chromosomes, as sketched in Fig. 3.1.

At the lowest level of organization, the DNA double helix is wrapped around a protein complex of cylindrical shape with a diameter of about 6 nm. This elementary packing unit is commonly referred to as a nucleosome. Its structure is known in exquisite detail and will be discussed in Sec. 3.1.1. Nucleosomes are more or less evenly spaced on the genome with an average distance of about 30 nm. When stretched or in low salt conditions, this structure looks like a string of DNA with beads, the nucleosomes, of about 10 nm in diameter. Under physiological conditions, this array of nucleosomes is further compactified to form a fiber with 30 nm in diameter, the structure of which is still subject to debate. The two competing models differ primarily in the geometry of the linker DNA between consecutive nucleosomes. In *solenoid* models, it is assumed that nucleosomes are arranged along a helix [88], which requires the linker DNA between two nucleosomes to be strongly bent. For the second class of models, it is assumed that the linker DNA is straight and crosses the center of the chromatin fiber. In these *zig-zag* models two consecutive nucleosomes are assumed to lie on more or less opposite sides of the fiber [89]. Recently, the crystal structure of tetra-nucleosomes was resolved, providing evidence for a zig-zag structure [90, 91]. A computational study also suggests that the structure of oligo-nucleosomes is best described by an irregular zig-zag model [92]. A more comprehensive overview and a survey of the current state of the debate is given in ref. [93]. Due to its stacked structure without strong interactions along the direction of the fiber, the chromatin fiber is rather flexible and easily ripped apart by longitudinal tension. Stretching experiments on a single chromatin fiber

and comparison to an extensible worm-like-chain model suggest a persistence length of about 30 nm and a stretching modulus of 5pN [94]. This experiment further indicates, that the chromatin fiber disintegrates if tensions beyond 20pN are applied. Little is known about the intermediate levels of chromatin organization. It is believed, that the 30 nm fiber forms large loops that are arranged on some scaffold, but evidence is sparse [7]. Only during cell division in the so called metaphase, the DNA is packed into the dense structure known as chromosome¹ that is large enough to be seen in the light microscope. Our focus here is on the elementary packing unit, the nucleosome, and we will therefore describe the structure of the nucleosome in greater detail.

3.1.1. The nucleosome core particle

While the structure of chromatin at larger length scales is still under debate, the nucleosome has been studied at atomic resolution. Luger et al. succeeded in crystalizing the complex of histone proteins together with a short piece of DNA wrapped around the protein complex and resolved the structure using X-ray scattering techniques. The first study achieved a resolution of 2.8Å [95] and a subsequent experiment improved the resolution to 1.9Å [96]. The structure of the nucleosome is illustrated in Fig. 3.2. A piece of DNA, precisely 147 bps long, is wrapped around a cylindrical protein complex 1.7 times along a left-handed super-helical path. The pitch of this path is only 2.8 nm, such that the DNA comes very close to itself along the super-helix. The protein complex has a diameter of 6.5 nm and a height of about 6 nm. The cylinder is assembled out of four different histone proteins H2A, H2B, H3 and H4, each of which is present in two copies. These proteins form crescent shaped heterodimers (H2A-H2B) and (H3-H4), which are arranged such that they define a binding path for the DNA. The histone complex is positively charged and therefore attracts the negatively charged DNA. The DNA-protein interaction is concentrated in 14 well defined contact points located at positions where the minor groove of the DNA faces the protein core. Each contact points forms a variable number of hydrogen bonds with the DNA. Due to the electrostatic nature of the protein-DNA interaction, the stability of nucleosomes depends on salt concentration. With increasing salt concentration, nucleosomes disassemble into DNA and the histone core complex, before the histone complex dissociates further into the dimers [97].

The net binding free energy between DNA and the histones can be estimated using cleavage enzymes that cut DNA at specific sites. In these experiments, cleavage sites are placed at different locations on the wrapped DNA and the reduction of the cleavage rate compared to free DNA is measured. By measuring this rate reduction, one can estimate the fraction of time the DNA site is accessible to protein binding [98, 99, 100], from which the free energy difference of the wrapped and the unwrapped state is calculated. These cleavage studies also revealed a significant sequence dependence of the net binding energies, but as a rule of thumb, each contact point contributes about 1.5 to $2k_B T$ to the net binding

¹The name chromosome is derived from the greek word *chromos* for color, since chromosomes are easily stained with dyes that bind to DNA.



Figure 3.2: The nucleosome consists of 147 bp of DNA wrapped 1.7 times around a complex of eight proteins. The two strands of DNA are shown in turquoise and brown. Only the main chains of the histone proteins are shown (H3: blue, H4: green, H2A: yellow, H2B: red). Figure reprinted with kind permission by Nature [95].

free energy under physiological conditions. The net binding free energy is the amount by which the total interaction energy exceeds the free energy needed to force the DNA into the strongly bent and clamped conformation when wrapped around a histone complex. The latter can be estimated as follows. When only moderately bent, dsDNA is well described by a worm-like-chain model (WLC, comp. Sec. 1.3.2) for semi-flexible polymers with a persistence length of $\ell_p = 50\text{nm}$. Within the WLC model, the bending energy of the DNA in a nucleosome can be estimated to

$$E_{bend} = k_B T \frac{\ell_p l}{2R^2} \approx 58k_B T, \quad (3.1)$$

where $R = 4.3\text{nm}$ is the radius of curvature of the DNA contour and $l = 43\text{nm}$ is the length of the bend part. This number for E_{bend} should only be considered as an order of magnitude estimate, since it is not at all clear whether the WLC model is applicable to strongly bent DNA. The estimate for the bending free energy leads to an estimate of the total interaction free energy of about $6k_B T$ per contact point. The binding strength and the bendability of the DNA are strongly sequence dependent and special sequences, called positioning sequences, are known to bind preferentially to histones in a precise alignment.

3.2. Gene regulation in eukaryotes

In prokaryotes, the set of genes which is transcribed by the RNA polymerase into mRNA is determined by the concentration of transcription factors (TF) in the cytosol. The regulatory sequences to which TFs bind specifically are usually located from 20 to a couple of hundred base pairs upstream of the gene and either enhance the binding of the polymerase to the promoter site by attractive interaction or prevent the binding of the polymerase by steric hinderance. These regulatory mechanisms are well established for prokaryotes, where the DNA is freely accessible to passive TFs.

Transcription regulation in eukaryotes is more complicated and many additional stages of regulation exist. The regulatory sites for a specific gene can be far away from the site where transcription starts and many more signals are integrated to determine whether a gene is to be silent or not. The general picture of eukaryotic gene regulation is far from complete. Nevertheless, TFs have to find their binding sites, even if they are hidden by nucleosomes. The comparatively small net binding energy of nucleosomes led to the hypothesis, that transient unwrapping of DNA from the histone complex driven by thermal fluctuations could suffice for reliable gene regulation [99]. Polach and Widom coined the term *site exposure mechanism* for this tentative mode of gene regulation. The mechanism is illustrated in Fig. 3.3a. The site exposure mechanism allows to tune the binding affinity of a TF by the location of the binding site inside the nucleosome, the further away a site is from the entry or exit point, the harder it is to access. Indeed, it has been shown that the positioning of nucleosomes along the genome is carefully controlled [101], which might be related to the tuning of binding affinities of TFs to their sites. The nucleosome can also be exploited to mediate indirect interactions between TFs, see Fig. 3.3b&c. If

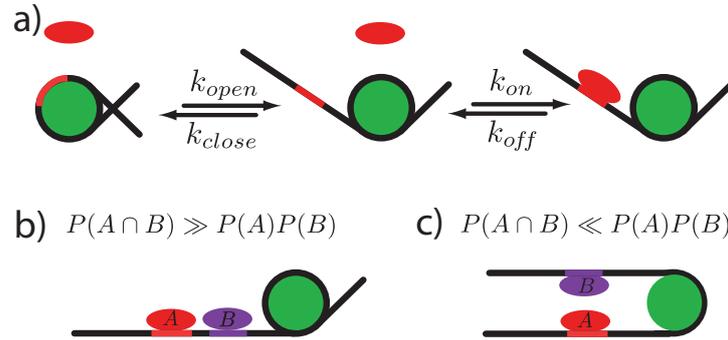


Figure 3.3: Site exposure mechanism for protein binding to nucleosomal DNA. Part a): Before the protein (red blob) can bind to its binding site (red), the DNA (black) has to detach from the histone complex (green). Part b&c): Nucleosomes can mediate transcription factor interactions, see main text.

two binding sites are located on the same side of the symmetry point of the nucleosome, exposure of the binding site further inside the nucleosome implies the exposure of the other. Hence, the joint binding probability is higher than the product of the individual binding probabilities, which is equivalent to cooperative binding of the TFs. This type of nucleosome mediated TF interaction has been shown to be a functional mode of gene regulation *in vivo* [102]. In the opposite case, where the two sites are at different ends of the piece of DNA, simultaneous binding is disfavored. To rationalize this, recall that DNA is highly charged. In a nucleosome, DNA is wrapped 1.7 times along a helical path such that the DNA comes very close (a few Å) to itself for 0.7 turns. Due to self-repulsion of DNA, the first 0.7 turns are rather easy to unwrap, while the final turn is much more stable since self repulsion is lacking. Two binding sites on opposite ends of the DNA are usually individually accessible by unwrapping less than 0.7 turns of DNA, however, when exposing both of them simultaneously only little DNA remains wrapped and one has to compensate the lacking self-repulsion. This gives rise to a joint binding probability that is less than the product of the individual binding probabilities, equivalent to repulsive interactions. However, in order to be feasible, the site exposure needs to be fast. The remainder of this chapter will address kinetic aspects of site exposure. Before presenting our theoretical study, I will discuss recent experiments, that study the dynamics of single nucleosomes *in vitro*.

3.3. Experiments on single nucleosome dynamics

A set of experiments addressing the dynamics of nucleosomes was performed by Gu Li in the group of Jonathan Widom. To study the fluctuation properties of DNA wrapped around a histone complex, they labeled a 147 bp long positioning sequence at one end with a green fluorescent dye. In addition, they labeled the appropriate histone protein with a red dye, such that the two dyes are in very close proximity when the DNA is fully

wrapped around the protein complex. The two dyes used are an efficient FRET pair, *i.e.* the excitation energy can be transferred without radiation from the green to the red dye via a dipol-dipol interaction. The efficiency of this energy transfer decreases with the distance r between the dyes as

$$E_{FRET} \sim \frac{1}{1 + (r/R_0)^6}, \quad (3.2)$$

where R_0 is the separation at which the E_{FRET} is half its maximal value. This distance is known as the Förster radius and is usually on the order of a few nanometers. The FRET efficiency drops from near to one to negligibly small values in a very narrow range surrounding R_0 , which makes FRET an extremely sensitive distance measure. The arrangement of FRET pairs on the nucleosome as realized by Li et al. allows the detection of the state of the nucleosome with optical means. In one publication [103], Li et al. convincingly showed, that the outermost part of the DNA is transiently unwrapped. It is well known, that the equilibrium constant between the wrapped and the unwrapped state can be tuned by varying the salt concentration, since mobile ions in solution screen the DNA-histone interaction. The same effect can be achieved by placing a binding site for the DNA binding protein LexA inside the nucleosome. Once the DNA unwraps from the histone and exposes the binding site, the open state is stabilized by binding of LexA to its site. The occupation of the LexA binding site can be controlled by the LexA protein concentration in solution. While this work established, that an equilibrium between the wrapped and unwrapped state exists and that proteins can access binding sites buried inside nucleosomes, it is still a bulk experiment and does not yield any information about the rates of individual wrapping and unwrapping events. This question was addressed in a subsequent publication [2] using fluorescence correlation spectroscopy (FCS) and stopped flow measurements. In the stopped flow experiments, nucleosomes are rapidly mixed with a LexA. LexA binds strongly and rapidly to a binding site located between base pair 8 and 27 of the DNA strand if and only if the site is exposed by transient DNA unwrapping from the histone complex. Since the DNA exposure is the rate limiting step, its rate can be measured by monitoring the decrease in FRET after mixing. The estimate for the exposure rate is $k_{open} = 3.9 \pm 0.9s^{-1}$. Using the equilibrium constant between the open and closed state determined in previous experiments, the rewinding rate is estimated to be $k_{close} \approx 90s^{-1}$. To corroborate these findings, a second experiment using FCS was performed. The authors compared the fluorescence autocorrelation curves of nucleosomes labeled only with the green dye to those labeled with pair of green and red dyes. In the former case, the decay of the autocorrelation function is solely due to the diffusion of nucleosomes into and out of the focal volume, while in the later case transient unwrapping events add an additional source of decorrelation. By fitting a reaction-diffusion model to the data, an independent estimation of the rates is achieved, yielding $k_{open} = 3.6s^{-1}$ and $k_{close} = 20s^{-1}$. These are within the same order of magnitude and are consistent with the previous estimates within experimental uncertainty.

Similar experiments were performed by M. Tomschik in the group of S. H. Leuba [3]. In these experiments, the red and the green dye were both attached to the DNA and their positions were chosen such that both dyes are next to each other when the DNA is

fully wrapped around the histone. The DNA used was 164 bp long with a sequence that is known to wrap symmetrically around the histone octamer. Furthermore, the DNA is functionalized at one end, such that it can be chemically ligated to a streptavidin coated glass cover slip. The green fluorescent dye can now be excited by evanescent light and individual nucleosomes show up as bright spots in the wide-field image. Depending on their conformation, the excitation energy is either transferred to the red dye or emitted as green light. Using this setup, Tomschik et al. succeeded in measuring time traces showing the opening and closing of single nucleosomes and thus were able to determine the associated rates directly. Depending on the salt concentration, the rate of unwrapping is $k_{open} = 0.2 - 0.5s^{-1}$, while the rate for rewinding is $k_{close} = 5 - 6s^{-1}$. The fact that the opening rate is much slower than the estimates by Li et al. is not surprising, because the length of the DNA segment that has to be unwrapped to change the FRET signal is much longer, at least 60 bps. However, the opening and closing rate should be related via the equilibrium constant, which is known to be larger than $k_{close}/k_{open} < 30$. What gives rise to this discrepancy is unclear. Taken together, these experiments suggest that the nucleosome undergoes rapid conformational fluctuations which involve unwrapping of the DNA and exposure of buried DNA binding sites.

3.4. Kinetic accessibility of protein binding sites in nucleosomal DNA

To help understanding the dependence of wrapping and unwrapping time scales on the DNA length involved, the DNA stiffness and the characteristics of the DNA protein interaction, we modeled the DNA-histone complex and studied the dynamics of our model using simulations. The DNA is modeled as a discretized WLC polymer with four beads per helical turn. The histone complex itself is not explicitly modeled, and only the 14 contact points, at which the DNA-histone interaction is concentrated, are included. These contact points are arranged in space along the path of the DNA deduced from the crystal structure of the nucleosome (cf. Fig. 3.2). Each contact point attracts the bead of the discretized WLC that corresponds to the appropriate location along the DNA with a short range Morse potential.

$$U_c = \gamma k_B T \sum_n \left(1 - e^{-|\mathbf{r}_i(n) - \mathbf{c}_n|/\rho}\right)^2, \quad (3.3)$$

where \mathbf{c}_n is the location of the n -th contact point, γ is the depth, and ρ the width of the contact potential. As a contact radius, we use $\rho = 0.5\text{nm}$, which is a compromise between the slightly longer ranged electrostatic interactions and the short ranged hydrogen bonding. Details of the model and the values used for the parameters are discussed in the publication reprinted in Sec. 3.7 [104].

3.4.1. Kinetics of site exposure

The wrapping and unwrapping of DNA from the protein complex is a stepwise process², where DNA detaches from one contact point at a time. In the course of unwrapping, the DNA has to overcome a transition state of high free energy, at which it no longer feels the short range attraction to the contact point but is still strongly bent. Overdamped thermally activated barrier crossing processes are well described by Kramers' rate theory, which states that the transition rate is given by the product of the pseudo-equilibrium population of the transition state and the relaxation rate out of this state [106, 107]. The former is the exponential of the free energy difference from the meta-stable state to the transition state, whereas the latter depends on the mobility of the reaction coordinate. In our case, a natural reaction coordinate is the distance of the DNA from the contact point. If the process by which the DNA detaches from the outermost contact point was purely local, *i.e.* only the part of the DNA that binds to the specific contact point is involved, one would assume that the mobility of the reaction coordinate and hence the rate was independent of the DNA length attached. However, Brownian dynamics simulations rapidly show, that this is not the case (cf. Figure 2 in the published article reprinted in Sec. 3.7). Instead, one observes a steady decrease of the rates as the attached DNA gets longer, *i.e.* for contact points that are further inside the nucleosome. A minute of thought reveals that this is what should be expected. The length of the free DNA is always far smaller than the persistence length and one expects it to move as if it was stiff. When opening or closing one contact point, this free DNA end has to rotate by about 45°. The friction coefficient associated with rotation of a rigid lever about one end increases as L^3 [108], and hence the opening and closing rate should decay with the length of the attached DNA. However, the simulation data is not compatible with such a drastic decrease of the rate, and neither of the two extreme cases, purely local vs. entirely rigid rotation, seems to be realized.

In order to describe the wrapping and unwrapping transitions faithfully, we study the rotational barrier crossing process of semi-flexible polymers taking into account the full spectrum of the polymer dynamics. The essence of the dynamics is captured by another model system, which consists of a semi-flexible polymer attached to a point about which it can rotate. The polymer experiences a potential acting on the attachment angle. This angular potential induces preferred attachment angles, separated by energy barriers. Within this model, transitions from one preferred orientation to another can be studied without interference from other aspects of the nucleosome model. We find that the dynamics of the barrier crossing process is governed by a new length scale l_c , which is given by the ratio of the polymer stiffness $\ell_p k_B T$ and the curvature γ of the angular potential at the transition state

$$l_c \sim \frac{\ell_p k_B T}{\gamma} . \quad (3.4)$$

If the overall length L of the polymer is small compared to l_c , the polymer crosses the barrier as a stiff rod with a rate that decreases as L^{-3} with the length. In the opposite case $L \gg l_c$

²Before the discrete nature of DNA-histone interactions was known, a theoretical study suggested that DNA unwrapping is an all-or-none process [105].

only the first part of length l_c is involved in the relaxation from the barrier and the rate is independent of L . If $l_c < L$, the transition rate is therefore greatly enhanced compared to the rate of a rigid lever. The dynamics of long polymers is limited by diffusion, which again results in a L^3 -dependence of the typical time of between reorientations of the polymer. In addition to this simple scaling argument, the interplay of the polymer dynamics and the relaxation from the barrier can be treated analytically taking into account the complete mode spectrum of the polymer. Comparison to the nucleosome data reveals, that the rewinding transitions in our nucleosome model fall into the crossover region between the flexibility assisted regime and the diffusion limited regime.

Caveats and pitfalls. Our model of the nucleosome is very simplistic in several aspects. First of all, it is far from obvious whether a discretized WLC model is appropriate for DNA bent as strongly as it is in nucleosomes. Furthermore, modeling DNA as a line with constant charge density is certainly not a faithful description at the nanometer scale, since the diameter of the DNA itself is 2 nm and the spatial arrangement of the charges on the double helix certainly matters. However, we are only interested in the physical mechanisms that underlie the DNA-histone dynamics and to that end, the model has to be as simple as possible to exhibit the generic features as clearly as possible. We think, that our model captures the essential physics in a satisfactory way, as it integrates polymer properties of DNA and the short range attraction of the DNA to the surface of the protein complex.

Coarse grained models like ours usually depend on reasonable choices of many unknown parameters and effective potentials. These choices can have significant impact on the time scale of the observed dynamics, which makes such modeling a very delicate task. The rates of the wrapping and unwrapping transitions surely depends on the precise form of the DNA-histone interaction potential and in particular on the nature of the transition state to unwrapping. Indeed, the dynamics of our model appears to be a factor of 100-1000 faster than real nucleosomes. Having this in mind, we can only compare different situations within the framework of our model and cannot make any statements regarding absolute timescales.

DNA is slightly unwound when wrapped around the nucleosome. We estimated the torsional energy for wrapping of one 10 bp segment to about $1k_B T$. This is far less than the energetic cost due to bending or the adsorption energy per contact point. Therefore, we implicitly included its thermodynamic effect into the effective interaction potential. Nevertheless, the fact that DNA has to be slightly unwound to match the contact potential might be responsible for the large wrapping/unwrapping times observed in experiments. Including twist deformation into our model did not seem to be justified to us, since little is known about the dependence of DNA-histone interaction on twist. While it likely affects the absolute timescales, we do not expect it to alter the qualitative picture of DNA wrapping.

3.5. Flexibility assisted conformational transitions

DNA wrapping and unwrapping in nucleosomes is a thermally activated barrier crossing process which is coupled to the lever-like rotation of the attached DNA end. While our primary motivation to study such a process was a better understanding of the dynamics of our nucleosome model, similar transitions are ubiquitous in proteins and protein-DNA complexes. One class of important examples are molecular motors such as myosins and kinesins [109], where a conformational transition in the motor head is coupled to the rotation of a lever to which the cargo is attached. Other examples are conformational changes of DNA induced by proteins such as the integration host factor (IHF), which is required for the integration of viral DNA into the genome of the host cell [110, 111]. These transitions share two generic features, which turn out to be important for the kinetics of the transition: They involve the rotation of a lever-like extended object, and this lever has some residual flexibility. This flexibility is either continuously distributed as in DNA, or localized at hinges as found in the structure of molecular motors [112, 113].

We studied such transitions using a simple but general model and revealed an unexpected non-monotonic dependence of the rate on the stiffness of the lever. Furthermore, the barrier crossing rate is fairly insensitive to the hydrodynamic drag on the tip of the lever, which might imply robustness of the speed of molecular motors to cargo size variations. Our model consists of two beads which are connected to each other and the origin. The first bead acts as a joint with a finite bending stiffness ϵ . Its friction coefficient mimics the friction associated with bending modes of the lever. The friction coefficient of the outer bead plays the role of the cargo and accounts for the hydrodynamic drag associated with rotation about the origin. In analogy to the semi-flexible Brownian rotor used to study the DNA wrapping in the nucleosome, we include an external potential acting on the attachment angle. This external potential induces preferred attachment angles separated by potential barriers.

We find, that the Kramers-Langer theory for multi-dimensional barrier crossing processes does not describe the phenomenology of our model [114]. The discrepancy results from the configuration dependent mobility matrix of our model, which is not accounted for in standard Kramers-Langer theory. We generalize the Kramers-Langer theory to a rate theory that perturbatively includes the effects of configuration dependent mobility matrices. This generalized theory captures the essential features of the observed phenomenology and in particular explains the peak. The maximal rate at finite stiffness is due to a tradeoff between an increasing average mobility of the reaction coordinate and a decreasing rate due to stronger coupling of the inner and outer bead due to higher stiffness.

Our work on this system is contained in a recently submitted publication entitled “Optimal rate in conformational transitions”, which is reprinted in Sec. 3.8. A more detailed derivation of the generalized Kramers-Langer rate is presented in the Appendix B.

3.6. Conclusion & Outlook

Understanding the way higher organisms orchestrate the expression of their genes is a formidable task and we are just beginning to get a faint idea of the elaborate mechanisms evolution came up with. Nevertheless, some molecular details such as the structure of nucleosome are known in exquisite detail. Single nucleosomes have recently been studied experimentally and were found to be very dynamics entities that undergo rapid conformational changes. We addressed this questions theoretically and extracted generic features of the dynamics of DNA unwrapping and wrapping from the protein complex. Due to the localized DNA-histone interaction, the dynamics is essentially discrete and each step involves a thermally activated barrier crossing event. In the course of this transition, the DNA is rotated like a lever. We find that the bending fluctuations of the DNA greatly enhance the barrier crossing rate and that the dynamics is governed by a new length scale l_c which emerges from the coupling of polymer modes and the relaxation dynamics from the barrier. Since similar situations are ubiquitous in conformational transitions in macromolecules, we studied such transitions in a more general context, both for continuously distributed flexibilities and hinged levers. Simulation results revealed, that the transition rates for hinged levers depend non-monotonically on the stiffness of the hinge. To describe and understand this phenomenon, we generalized the Kramers-Langer theory for multi-dimensional escape processes to account for configuration dependent mobility matrices. We hope that this generalized rate theory will find applications in other fields.

In vivo, nucleosomes are not in isolation but arranged in large arrays. They interact with each other electrostatically and via flexible protein tails. Hence, it is not at all clear, to what extend our findings carry over to *in vivo* chromatin dynamics. The next step along bottom up approach, would be to incorporate additional nucleosomes into our simulations and explore how the dynamics changes. It should be possible to test the key prediction of our study, the length dependence of the wrapping and unwrapping rate, experimentally. Another interesting question to address experimentally is the strength of the effective repulsion of transcription factors mediated by the nucleosome and whether this interaction has significant effects on gene expression.

Kinetic Accessibility of Buried DNA Sites in Nucleosomes

Wolfram Möbius, Richard A. Neher, and Ulrich Gerland

*Arnold Sommerfeld Center for Theoretical Physics (ASC) and Center for Nanoscience (CeNS),
LMU München, Theresienstrasse 37, 80333 München, Germany*

(Received 12 May 2006; published 14 November 2006)

Using a theoretical model for spontaneous partial DNA unwrapping from histones, we study the transient exposure of protein-binding DNA sites within nucleosomes. We focus on the functional dependence of the rates for site exposure and reburial on the site position, which is measurable experimentally and pertinent to gene regulation. We find the dependence to be roughly described by a random walker model. Close inspection reveals a surprising physical effect of flexibility-assisted barrier crossing, which we characterize within a toy model, the “semiflexible Brownian rotor.”

DOI: [10.1103/PhysRevLett.97.208102](https://doi.org/10.1103/PhysRevLett.97.208102)

PACS numbers: 87.15.He, 36.20.Ey, 87.16.Sr

Although the DNA in eukaryotic cells is packaged into chromatin, its genetic information must be accessible to proteins for read out and processing [1]. The structural organization of chromatin is fairly well known: the fundamental unit is a nucleosome core particle (NCP) consisting of about 150 base pairs (bp) of DNA wrapped in 1.7 turns around a cylindrical histone octamer [2], and NCPs are regularly spaced along the DNA, which is further compactified into higher order structures. In contrast, the *conformational dynamics* of chromatin is poorly understood. Recent experiments studied these dynamics on the level of individual NCPs using single-molecule force [3] and fluorescence [4,5] techniques. The latter directly observed spontaneous conformational transitions where part of the DNA unwraps reversibly, allowing proteins to access DNA sites that are normally buried. This mode of access, driven by thermal fluctuations, is particularly important for passive DNA-binding proteins, e.g., transcription factors. Here, we study spontaneous DNA unwrapping within a theoretical model; see Fig. 1(a).

Consider a buried DNA site that is accessible only when a DNA segment of length L is unwrapped. How long is the typical dwell time τ_a in the accessible state, i.e., the window of opportunity for protein binding? And what is the typical time τ_i for which it remains inaccessible? Li *et al.* [4] measured $\tau_a = 10\text{--}50$ ms and $\tau_i \approx 250$ ms for $L \sim 30$ bp, while Tomschik *et al.* [5] found $\tau_a = 100\text{--}200$ ms and $\tau_i = 2\text{--}5$ s for $L \sim 60$ bp. Taken together, these results indicate a significant dependence on L in both time scales, which cannot be reconciled with an early theoretical study [6] suggesting an all-or-none unwrapping mechanism where the nucleosome fluctuates between two conformations only. Instead, these results, as well as previous biochemical experiments [7], imply a multistep opening mechanism.

In this Letter, we propose and characterize a theoretical model for this multistep mechanism, similar in spirit to previous work on histone-DNA interactions which focused mainly on static properties or the calculation of free energy barriers [6,8,9]. Within our model, we clarify the physics

that determines the L dependence of the time scales τ_a and τ_i . We find that the dependence of τ_i can be interpreted with a simple random walker model, which may serve as a fitting model for future experiments that probe the time scales at different L values. In contrast, the L dependence of τ_a reflects the intricate coupling between the DNA polymer dynamics and the dynamics of breaking and reforming DNA-histone contacts. To analyze the effect of this coupling, we introduce a toy model, the “semiflexible Brownian rotor” (SBR); see Fig. 1(b). We identify a generic physical effect of flexibility-assisted barrier crossing, which may arise also in other contexts. It is marked by a characteristic plateau of the time scale at intermediate L . Biologically, the L dependence is relevant, because it creates a positioning effect for transcription factor binding sites relative to nucleosomes [10]. We expect that the integration of single NCPs into nucleosome arrays will alter the *absolute* time scales but not the basic physics of the DNA (un)wrapping process.

Nucleosome model.—The NCP crystal structure [2] shows that both the electrostatic and hydrogen bond interactions between the DNA and the histone complex are

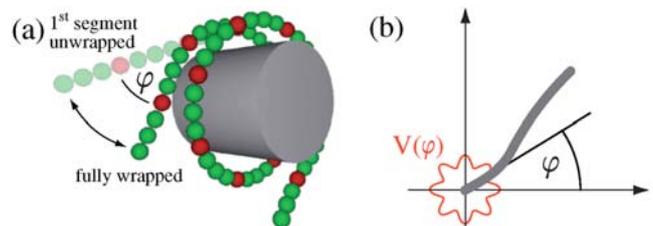


FIG. 1 (color online). (a) Illustration of our nucleosome model. The DNA-histone interaction is localized at contact points attracting the red (dark) beads. The DNA is shown in the ground state as well as a conformation where the first contact is open. (b) Illustration of the semiflexible Brownian rotor (SBR) model. In this toy model, the tradeoff between bending energy and DNA-histone interaction in the nucleosome is mimicked by an angular potential $V(\varphi)$, exerting a torque on the attachment angle φ of a semiflexible polymer at the origin.

mainly localized to 14 contact points, about evenly spaced by 10.2 bp along a superhelical contour with radius 4.2 nm and helical pitch 2.4 nm. Because we are interested only in the dynamics at a fixed (physiological) salt concentration, we combine the interactions at each of these points into a simple Morse potential [11]. The DNA-histone interaction energy is then

$$U_c = \gamma k_B T \sum_n (1 - e^{-|\mathbf{r}_{i(n)} - \mathbf{c}_n|/\rho})^2, \quad (1)$$

where \mathbf{c}_n is the n th contact point on the superhelical contour, γ is the depth, and ρ the width of the contact potential. A discrete bead-spring model with beads at positions \mathbf{r}_i models the DNA, and $i(n)$ is the bead bound to contact n in the fully wrapped state. The beads are connected by a harmonic potential $U_s = \varepsilon_s \sum_i (|\mathbf{r}_{i+1} - \mathbf{r}_i| - a)^2/2$ with a typical bead separation a and a stiffness ε_s set to $800k_B T/\text{nm}^2$. Below, we use three beads between contacts and at each end (about 2.5 bp/bead), unless stated otherwise. Increasing the discretization or ε_s raises the computational effort without affecting our results qualitatively. We account for the bending rigidity of DNA by an energy $U_b = \varepsilon_b \sum_i (1 - \cos\theta_i)$ with bending angle θ_i at bead i and a bending stiffness ε_b adjusted such that the apparent persistence length matches the known $\ell_p \approx 50$ nm for DNA at physiological salt conditions. Furthermore, we incorporate the screened electrostatic self-repulsion of DNA through a Debye-Hückel potential $U_{\text{DH}} = k_B T l_B (\tau a)^2 \sum_{i < j} e^{-\kappa|\mathbf{r}_i - \mathbf{r}_j|}/|\mathbf{r}_i - \mathbf{r}_j|$ with the Bjerrum length $l_B \approx 0.7$ nm, a charge density $\tau = 2$ charges/bp, and a screening length $\kappa^{-1} \approx 1$ nm. We use a contact radius $\rho = 0.5$ nm in between the range of hydrogen bonds and electrostatic interactions and adjust the depth γ of the Morse potential to match the binding free energy [12] of $\approx 1.5k_B T$ per contact estimated from biochemical experiments [7,9]. Taken together, the total energy is $U = U_s + U_b + U_{\text{DH}} + U_c$. To study the dynamics of our model, we perform Brownian dynamics simulations with the overdamped Langevin Eqs.

$$\dot{\mathbf{r}}_i(t) = -\mu_b \nabla_{\mathbf{r}_i} U(\{\mathbf{r}_j\}) + \boldsymbol{\eta}_i(t), \quad (2)$$

where μ_b is the bead mobility, and the absolute time scale is set by $a^2/\mu_b k_B T$. The random forces $\boldsymbol{\eta}_i$ satisfy $\langle \boldsymbol{\eta}_i(t) \cdot \boldsymbol{\eta}_j(t') \rangle = 6\mu_b k_B T \delta_{i,j} \delta(t - t')$.

Unwrapping dynamics.—A suitable reaction coordinate for the opening of a single contact is the attachment angle φ , see Fig. 1(a), which changes by $\Delta\varphi \approx 45^\circ$ in this process. The equilibrium distribution $p(\varphi)$ for the first contact is shown in Fig. 2(a). Its bimodal form suggests to approximate a contact by a 2-state system, with rates k_b , k_u for binding and unbinding, respectively. To test whether such a reduced description is sufficient, we initiate simulations in the fully wrapped state and determine the functionally relevant time scales, i.e., the average time $\tau_i(n)$ until contact n opens to expose the n th DNA segment and the average time $\tau_a(n)$ until contact n recloses [13,14]. The

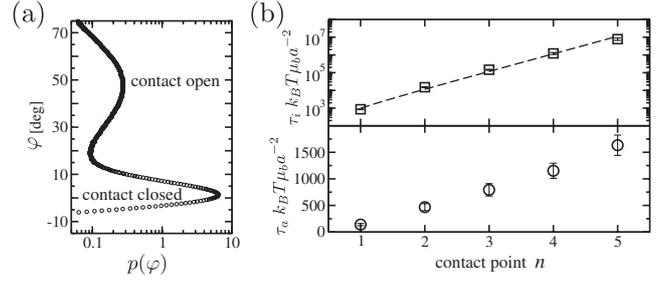


FIG. 2. (a) Equilibrium distribution of the DNA angle φ defined in Fig. 1(a). The two peaks at $\varphi = 0$ and $\varphi \approx 45$ deg correspond to the fully wrapped state and the state with contact 1 open, respectively. (b) Kinetics of DNA site exposure within our nucleosome model. The dwell time in the inaccessible state (squares) increases roughly exponentially with the number of contacts that must open to render a DNA site accessible. The dashed line is a fit to Eq. (3). The circles show the average time the n th contact point remains open.

results are shown in Fig. 2(b) for $n \leq 5$ [15]. Within the reduced description of consecutive 2-state contacts, $\tau_i(n)$ can be calculated as a mean first passage time [16] for a 1D biased random walker with hopping rates k_u , k_b . The walker starts at site zero (reflecting boundary) and reaches site n after an average time

$$\tau_i(n) = \frac{k_u^{-1}}{1 - K} \left[\frac{1 - K^n}{1 - K^{-1}} + n \right] \stackrel{K \gg 1}{\approx} \frac{K^{n-1}}{k_u}. \quad (3)$$

Here, $K = k_b/k_u$ can be interpreted as the effective equilibrium binding constant per contact. The exponential increase of $\tau_i(n)$ is clear also from the equivalence of the biased random walk with a random walk against a free energy ramp. The excellent fit of (3) to the simulation data (dashed line) indicates that the reduced description is sufficient for the dwell times in the inaccessible state. In contrast, it proves insufficient for the dwell times in the accessible state, because $\tau_a(n)$ in Fig. 2(b) is clearly not constant as one would expect with a fixed binding rate k_b . Thus, we find $\tau_a(n)$ to be a more sensitive probe for the physics of spontaneous site exposure than $\tau_i(n)$.

To probe the effect of the DNA length on the rewinding kinetics, we vary the number of overhanging beads before contact 1 and plot $\tau_a(1)$ as a function of the overhang length L in Fig. 3(a). Superimposed is the data of Fig. 2(b) (bottom) with n converted to contour length. The good agreement of these dependencies indicates that τ_a is determined by polymer dynamics. Indeed, we will now see that contact breaking and reformation of a rotating semiflexible polymer displays much richer physics than a simple 1D barrier crossing process.

Semiflexible Brownian rotor.—The essential physics of contact formation in the nucleosome is captured by the toy model depicted in Fig. 1(b): A semiflexible polymer with contour length L and persistence length ℓ_p is attached to a point about which it can rotate in a plane. The attachment angle φ experiences a periodic potential $V(\varphi) = V_0 \cos(2\pi\varphi/\Delta\varphi)$, which creates preferred angles sepa-

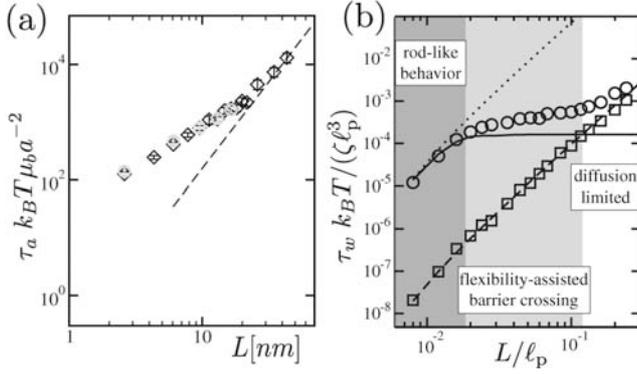


FIG. 3. (a) The dependence of the dwell time $\tau_a(n=1)$ on the overhanging DNA length (diamonds) is compatible with $\tau_a(n)$ when n is converted to contour length (gray open circles). The dashed line indicates the diffusion limit (see main text for details). (b) The average barrier crossing time τ_w (open circles) for the SBR model of Fig. 1(b). At small lengths, the barrier crossing time follows that of a stiff rod (indicated by the dotted line). Beyond a crossover length $\ell_c \ll \ell_p$, barrier crossing is much faster than for a stiff rod. For large lengths, τ_w approaches the diffusion limit, i.e., τ_w of the free SBR (open squares). With $L < \ell_p$, free diffusion of the SBR is virtually indistinguishable from free diffusion of a rigid rod (dashed line). The crossover from the rodlike regime to the intermediate regime is well described by the theoretical analysis (solid line), see main text.

rated by potential barriers as in our nucleosome model (there, the barrier for contact reformation results from the DNA bending energy and the electrostatic repulsion). The main difference is that the length of the rotating polymer is constant for this SBR, while it changes slightly when a contact breaks or reforms in the nucleosome. Also, we do not consider a directional bias in the SBR, because it is not essential for what follows. So far, barrier crossing of semiflexible polymers was studied only for situations where the entire polymer experiences an external potential [17]. In the NCP, the potential acts only on the angle at the attachment point.

To characterize the phenomenology of the SBR, we determine its barrier crossing rate $1/\tau_w$ with Brownian dynamics simulations of a discrete bead-spring model [18]. The circles in Fig. 3(b) show τ_w as a function of L/ℓ_p for $V_0 = 5k_B T$. We observe that at very short lengths, τ_w follows the stiff rod behavior $\tau_w \sim L^3$ [19] indicated by the dotted line. However, above a certain length ℓ_c , there is a regime where τ_w is nearly insensitive to L , before it rises again. Hence, for lengths $L > \ell_c$ the semiflexible polymer crosses the barrier much faster than the stiff rod. What is the physical mechanism for this acceleration? One effect of a finite flexibility is a reduced mean end-to-end distance (due to the undulations in the contour), which in turn leads to a larger rotational mobility. However, with $V(\varphi) = 0$, the rotational diffusion time of a semiflexible polymer over an angle $\Delta\varphi$ (squares) is almost identical to that of a stiff rod (dashed line) when $L < \ell_p$. Hence the acceleration is

not a mobility effect. Note that the dashed line is also the diffusion limit for τ_w , which induces a second crossover from a reaction to a diffusion controlled process. The equivalent diffusion limit is shown also in Fig. 3(a) (dashed line). It indicates that the $\tau_a(n)$ data for the nucleosome is indeed in the accelerated barrier crossing regime.

Flexibility-assisted barrier crossing.—To understand the interplay between the polymer dynamics and the barrier crossing dynamics qualitatively, we recall the basic aspects of each: (i) A semiflexible polymer of length L relaxes its conformational degrees of freedom in a time $\sim L^4/\ell_p$ [20]. Conversely, within a given time τ , a local bending deformation is “felt” only over a length $\ell \sim (\ell_p \tau)^{1/4}$. (ii) The probability current over a barrier is proportional to the quasiequilibrium occupancy of the transition state and to the relaxation rate τ^{-1} out of this state. Together, (i) and (ii) imply that ℓ_c is the length of the polymer segment that gets deformed during the relaxation process away from the potential peak. We estimate ℓ_c by noting that the attachment angle relaxes according to $\dot{\varphi} = -\mu(\ell_c)\partial V/\partial\varphi$, where $\mu(\ell_c) \sim \ell_c^{-3}$ is the rotational mobility of the deformed segment. Hence, $\tau \sim \ell_c^3(\Delta\varphi/2\pi)^2/V_0$ and with $\ell_c \sim (\ell_p \tau)^{1/4}$, we find

$$\ell_c = C\ell_p \frac{k_B T}{V_0} \left(\frac{\Delta\varphi}{2\pi}\right)^2, \quad (4)$$

where C is a constant to be determined below. For lengths below ℓ_c , the entire polymer is involved in the relaxation process, i.e., it behaves like a stiff rod.

Quantitative theory for the crossover.—To render the above picture quantitative, we employ the Langer theory for multidimensional barrier crossing processes [21]. For the case at hand, one can show [22] that the barrier crossing time simplifies to $\tau_w = \frac{\pi}{\lambda_-} e^{2V_0/k_B T}$, where λ_- is the eigenvalue associated with the unstable mode at the saddle point. We calculate λ_- using the continuous wormlike chain model in the weakly bending approximation [23]. At the transition state the chain is straight, e.g., along the x axis. We denote deviations from this configuration by $y(x, t)$. The chain dynamics follows $\partial_t y = -(k_B T \ell_p / \zeta) \partial_x^4 y$ with a friction coefficient ζ . With $\Gamma = V_0(2\pi/\Delta\varphi)^2$ denoting the curvature of the potential at the transition state, the torque on the attached polymer end is $-\Gamma \partial_x y|_{x=0}$. This torque must be balanced by a local bend resulting in the boundary condition $k_B T \ell_p \partial_x^2 y|_{x=0} = -\Gamma \partial_x y|_{x=0}$. The other boundary conditions are $y|_{x=0} = \partial_x^2 y|_{x=L} = \partial_x^3 y|_{x=L} = 0$. We find a unique unstable mode with eigenvalue $\lambda_- = k_B T \ell_p \alpha^4 / 4\zeta L^4$ and α determined by

$$\frac{\alpha[\sinh(\alpha) - \sin(\alpha)]}{\cosh(\alpha) + \cos(\alpha) + 2} = \sqrt[3]{12} \frac{L}{\ell_c}, \quad (5)$$

where ℓ_c is as in (4) with $C = \sqrt[3]{12}$. In the limit $L \ll \ell_c$, we find $\lambda_- = 3\Gamma/\zeta L^3$ independent of the stiffness, whereas in the opposite limit $\lambda_- = 3\Gamma/\zeta \ell_c^3$ independent

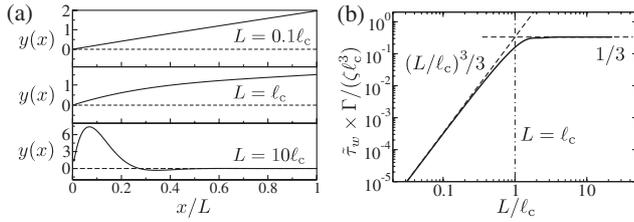


FIG. 4. Dynamics at the barrier. (a) The unstable eigenmode for three different lengths. Polymers shorter than ℓ_c rotate without significant deformation, while long polymers form a bulge of size $\sim \ell_c$ at the origin. (b) The prefactor of the Kramers time $\tilde{\tau}_w = 1/\lambda_-$ as a function of the length. The prefactor increases as L^3 if $L \ll \ell_c$ and is constant if $L \gg \ell_c$.

of L . Figure 4 shows (a) the unstable eigenmode for $L/\ell_c = \{0.1, 1, 10\}$ and (b) the crossover in the barrier crossing time. The eigenmode shape confirms our qualitative picture: stiff and short polymers respond to the torque by rotating as a whole, whereas the torque shapes a bulge of size $\sim \ell_c$ in longer polymers. For a discrete polymer model, the same analysis can be performed, but the eigenvalue λ_- must be computed numerically. The solid line in Fig. 3 shows the resulting barrier crossing time for the same discretization as used in the Brownian dynamics simulations of the SBR model. Indeed, the crossover from the rodlike to the flexibility-assisted barrier crossing is well described by this analysis. The deviations at larger L can be attributed to finite barrier corrections [24].

Discussion and outlook.—The experiments [4,5] have shown that the functionally relevant time scales τ_i and τ_a depend on the position on the nucleosomal DNA. Our theoretical study suggests that these time scales additionally depend on the total DNA length. The position dependence of τ_i should follow the random walker model (3), which is the minimal model for a gradual, multistep opening mechanism. However, we expect that the position-dependence of τ_a and the length-dependence of both time scales will reflect the polymer dynamics of the DNA. Within our toy model, the semiflexible Brownian rotor, we find three physically distinct regimes for this length dependence; see Fig. 3(b). The intermediate regime displays a striking flexibility-assisted barrier crossing effect, the onset of which is marked by the new length scale ℓ_c of Eq. (4). It can be interpreted as the length over which the polymer contour is deformed as it passes over the potential barrier. Because ℓ_c is considerably smaller than the persistence length ℓ_p , we expect that the onset of the intermediate regime will not be detectable in nucleosomes. However, nucleosomes should display the crossover from flexibility-assisted barrier crossing to diffusion-limited dynamics as shown in Fig. 3(a). All three regimes of Fig. 3(b) could be probed in an experimental realization of the SBR model, e.g., with an actin filament as the rotating polymer.

We thank H. Boroudjerdi, T. Franosch, E. Frey, O. Hallatschek, S. Leuba, R. Netz, R. Phillips, P. Reimann, P. R. ten Wolde, and J. Widom for useful discussions, and

the DFG for financial support.

- [1] K. Luger and J. C. Hansen, *Curr. Opin. Struct. Biol.* **15**, 188 (2005).
- [2] K. Luger *et al.*, *Nature (London)* **389**, 251 (1997).
- [3] B. D. Brower-Toland *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 1960 (2002).
- [4] G. Li, M. Levitus, C. Bustamante, and J. Widom, *Nat. Struct. Biol.* **12**, 46 (2005).
- [5] M. Tomschik, H. Zheng, K. van Holde, J. Zlatanova, and S. Leuba, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 3278 (2005).
- [6] N. L. Marky and G. S. Manning, *J. Mol. Biol.* **254**, 50 (1995).
- [7] K. Polach and J. Widom, *J. Mol. Biol.* **254**, 130 (1995).
- [8] K. Kunze and R. Netz, *Phys. Rev. Lett.* **85**, 4389 (2000); I. Kulic and H. Schiessel, *ibid.* **91**, 148103 (2003); **92**, 228101 (2004); W. Li, S.-X. Dou, and P.-Y. Wang, *J. Theor. Biol.* **230**, 375 (2004); D. Beard and T. Schlick, *Structure* **9**, 105 (2001).
- [9] H. Schiessel, *J. Phys. Condens. Matter* **15**, R699 (2003).
- [10] E. Segal *et al.*, *Nature (London)* **442**, 772 (2006).
- [11] Nucleosomal DNA is slightly unwound [2], with a small torsional energy $\sim 1k_B T$ per 10 bp. We include this twist contribution into the effective contact potential.
- [12] Note that our free energy cost per contact is slightly larger for the first contact and considerably larger when less than one turn of the superhelix remains wrapped, due to the electrostatic self-repulsion of DNA.
- [13] Because spontaneous unwrapping for $n > 2$ is rare, we use the forward flux sampling method to reduce the computational effort; see R. J. Allen, D. Frenkel, and P. R. ten Wolde, *J. Chem. Phys.* **124**, 024102 (2006).
- [14] We exclude transient wrapping or unwrapping events, i.e., we only count transitions that reach the metastable state in the neighboring potential valley.
- [15] For $n > 5$ the electrostatic self-repulsion of the DNA is less significant. This stabilizes the remaining contacts and modifies the exposure kinetics of the central sites (close to the dyad) [22].
- [16] C. Gardiner, *Handbook of Stochastic Methods* (Springer, Berlin, 1983).
- [17] P. Kraikivski, R. Lipowsky, and J. Kierfeld, *Europhys. Lett.* **66**, 763 (2004).
- [18] The time τ_w is operationally defined through the long-time rotational diffusion coefficient $D = \Delta\varphi^2/2\tau_w$ over many potential wells. Thus, τ_w measures the time for barrier crossing events that involve not only the passage of φ to a neighboring valley in $V(\varphi)$, but also turning of the polymer tip into the direction of the new valley, analogous to τ_a in the nucleosome model [14].
- [19] M. Doi and S. Edwards, *The Theory of Polymer Dynamics* (Clarendon Press, Oxford, 1986).
- [20] L. Le Goff, O. Hallatschek, E. Frey, and F. Amblard, *Phys. Rev. Lett.* **89**, 258101 (2002).
- [21] J. Langer, *Phys. Rev. Lett.* **21**, 973 (1968).
- [22] W. Möbius, R. A. Neher, and U. Gerland (to be published).
- [23] See, e.g., C. H. Wiggins, D. Rivelino, A. Ott, and R. E. Goldstein, *Biophys. J.* **74**, 1043 (1998).
- [24] P. Talkner, *Chem. Phys.* **180**, 199 (1994).

Optimal flexibility for conformational transitions in macromolecules

Richard A. Neher¹, Wolfram Möbius¹, Erwin Frey¹, and Ulrich Gerland^{1,2}

¹*Arnold Sommerfeld Center for Theoretical Physics (ASC) and Center for Nanoscience (CeNS), LMU München, Theresienstraße 37, 80333 München, Germany*

²*Institute for Theoretical Physics, University of Cologne, Germany*

(Dated: April 12, 2007)

Conformational transitions in macromolecular complexes often involve the reorientation of lever-like structures. Using a simple theoretical model, we show that the rate of such transitions is drastically enhanced if the lever is bendable, e.g. at a localized “hinge”. Surprisingly, the transition is fastest with an intermediate flexibility of the hinge. In this intermediate regime, the transition rate is also least sensitive to the amount of “cargo” attached to the lever arm, which could be exploited by molecular motors. To explain this effect, we generalize the Kramers-Langer theory for multi-dimensional barrier crossing to configuration dependent mobility matrices.

Many biological functions depend on transitions in the global conformation of macromolecules, and the associated kinetic rates can be under strong evolutionary pressure. For instance, the directed motion of molecular motors is based on power strokes [1], protein binding to DNA can require DNA bending [2] or spontaneous partial unwrapping of DNA from histones [3, 4], and the functioning of some ribozymes depends on global transitions in the tertiary structure [5]. These and other examples display two generic features: (i) A long segment within the molecule or complex is turned during the transition, e.g. an RNA stem in a ribozyme, the DNA as it unwraps from histones or bends upon protein binding, or the lever arm of a molecular motor relative to the attached head. (ii) The segment has a certain bending flexibility. Here, we use a minimal physical model to study the coupled dynamics of the transition and the bending fluctuations.

Our model, illustrated in Fig. 1, demonstrates explicitly how even a small bending flexibility can drastically accelerate the transition. Furthermore, if the flexibility arises through a localized “hinge”, e.g. in the protein structure of some molecular motors [6, 7] or an interior loop in an RNA stem, we find that the transition rate is maximal at an intermediate hinge stiffness. Thus, in situations where rapid transition rates are crucial, molecular evolution could tune a hinge stiffness to the optimal value. We find that an intermediate stiffness is optimal also from the perspective of robustness, since it renders the transition rate least sensitive to changes in the drag on the lever arm, incurred e.g. by different cargos transported by a molecular motor.

Our finding of an optimal rate is reminiscent of a phenomenon known as resonant activation [8, 9], where a transition rate displays a peak as a function of the characteristic timescale of fluctuations in the potential barrier. However, we will see that the peak in our system has a different origin: a trade-off between the accelerating effect of the bending fluctuations and a decreasing average mobility of the reaction coordinate. The standard Kramers-Langer theory [10] for multi-dimensional transition processes is not sufficient to capture this trade-off. A

generalization of the theory to the case of configuration-dependent mobility matrices turns out to be essential to understand the peak at intermediate stiffness.

Model.— We model the conformational transition as a thermally activated change in the attachment angle φ of a macromolecular lever, see Fig. 1. The lever has two segments connected by a hinge with stiffness ϵ , which renders the lever preferentially straight, but allows thermal fluctuations in the bending angle θ . The energy function $V(\varphi, \theta)$ of this ‘Two-Segment Lever’ (TSL) is

$$\frac{V(\varphi, \theta)}{k_B T} = \epsilon(1 - \cos \theta) - \left[\frac{(a\varphi)^3}{3} - \frac{b(a\varphi)^2}{2} \right], \quad (1)$$

where $k_B T$ is the thermal energy unit. The hinge, described by the first term, serves not only as a simple model for a protein or RNA hinge, but also as a zeroth-order approximation to a more continuously distributed flexibility; see below. The second term is the potential on the attachment angle φ , which produces a metastable minimum at $(\varphi, \theta) = (0, 0)$. The thermally-assisted escape from this minimum passes through the transition state at $(\varphi, \theta) = (b/a, 0)$ with a barrier height $\Delta V = b^3 k_B T / 6$ [20].

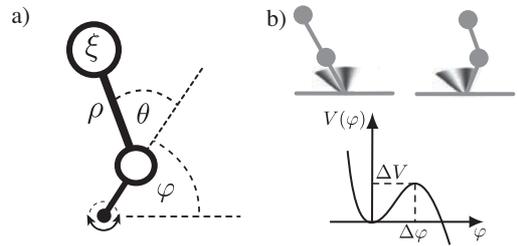


FIG. 1: Schematic illustration of the ‘Two-Segment Lever’ (TSL) model for conformational transitions. (a) The two segments of lengths 1 and ρ are connected by a hinge and attached to the origin. The viscous drag acts on the ends of the segments as indicated by the beads. (b) Schematic illustration of the barrier crossing processes. The external meta-stable potential $V(\varphi)$ is indicated by shading (top; dark corresponds to high energy) and is also sketched below.

In the present context, inertial forces are negligible, i.e. it is sufficient to consider the stochastic dynamics of the TSL in the overdamped limit. We localize the friction forces to the ends of the two segments, as indicated by the beads in Fig. 1(a). The length of the first segment defines our length unit and ρ denotes the relative length of the second segment. Similarly, we choose our time unit such that the friction coefficient of the first bead is unity, and denote the coefficient of the second bead by ξ . To describe the Brownian dynamics of the TSL, we derive the Fokker-Planck equation for the time-evolution of the configurational probability density $p(\varphi, \theta, t)$. In general, the derivation of the correct dynamic equations can be a nontrivial task for stochastic systems with constraints [11, 12]. For instance, implementing fixed segment lengths through the limit of stiff springs, leads to Fokker-Planck equations with equilibrium distributions that depend on the way in which the limit is taken [12]. However, for our overdamped system, we can avoid this problem by imposing the desired equilibrium distribution, i.e. the Boltzmann distribution $p = \exp(-V/k_B T)$, which together with the well-defined deterministic equations of motion uniquely determines the Fokker-Planck equation for the TSL.

The deterministic equations of motion take the form $\dot{q}_k = -\mathbf{M}_{kl} \partial V / \partial q_l$ with the coordinates $(q_1, q_2) = (\varphi, \theta)$ and a mobility matrix \mathbf{M} . We obtain \mathbf{M} with a standard Lagrange procedure: Given linear friction, \mathbf{M} is the inverse of the friction matrix, which in turn is the Hessian matrix of the dissipation function [13]. This yields

$$\mathbf{M} = \frac{1}{1 + \xi \sin^2 \theta} \begin{pmatrix} 1 & \frac{\rho + \cos \theta}{\rho} \\ \frac{\rho + \cos \theta}{\rho} & \frac{\rho + 2 \cos \theta}{\rho} + \frac{1 + \xi}{\xi \rho^2} \end{pmatrix}. \quad (2)$$

The Fokker-Planck equation then follows from the continuity equation $\partial_t p(\{q_i\}, t) = -\partial_k j_k(\{q_i\}, t)$ together with

$$j_k(\{q_i\}, t) = -M_{kl} \left[\frac{\partial V}{\partial q_l} + k_B T \frac{\partial}{\partial q_l} \right] p(\{q_i\}, t) \quad (3)$$

as the probability flux density. Our analytical analysis below is based directly on Eqs. (2) and (3), while we perform all Brownian dynamics simulations with a set of equivalent stochastic differential equations [14].

Transition rate.— To explore the phenomenology of the TSL, we performed simulations to determine its average dwell time τ in the metastable state, for a range of hinge stiffnesses ϵ . The rate for the conformational transition is related to the dwell time by $k(\epsilon) = 1/\tau(\epsilon)$. Fig. 2 shows $k(\epsilon)$ (circles) for a barrier $\Delta V = 12 k_B T$, a distance $\Delta\varphi = 0.4$ to the transition state, and $\xi = \rho = 1$ (data for different parameter values behaves qualitatively similar, as long as the process is reaction-limited, i.e. ΔV is sufficiently large that τ is much longer than the time for the TSL to freely diffuse over an angle $\Delta\varphi$). We observe a significant flexibility-induced enhancement of the transition rate over a broad range of stiffnesses, compared to

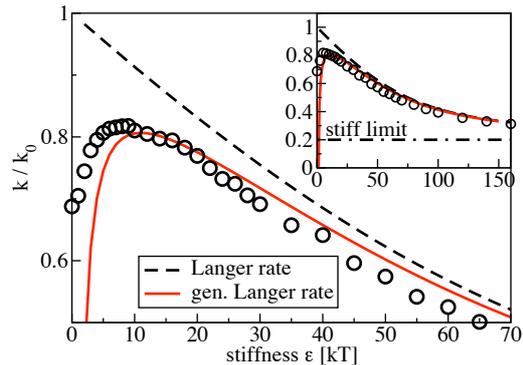


FIG. 2: Simulation data of the barrier crossing rate normalized by k_0 display a prominent peak at finite stiffness (circles, each obtained from 20000 simulation runs initialized at the metastable minimum). The conventional Langer theory fails to describe the non-monotonicity of the rate and overestimates the rate at small ϵ . The generalized Langer theory captures the non-monotonicity of the rate and describes the simulations data accurately; parameters see main text.

the dynamics in the stiff limit ($\epsilon \rightarrow \infty$), see inset. Note that the enhancement persists even at relatively large ϵ , where typical thermal bending fluctuations $\delta\varphi \sim \sqrt{\epsilon}$ are significantly smaller than $\Delta\varphi$. Surprisingly, the acceleration is strongest at an intermediate stiffness ($\epsilon \approx 10$). This observation suggests that the stiffness of molecular hinges could be used, by evolution or in synthetic constructs, to tune and optimize reaction rates.

When the friction coefficient ξ of the outer bead is increased, the rate of the conformational transition decreases; see Fig. 3a. This decrease is most dramatic in the stiff limit (dash-dotted line). In the flexible limit (diamonds) the decrease is less pronounced. Notably, the rate appears least sensitive to the viscous drag on the outer bead at intermediate ϵ (circles). Indeed, Fig. 3b shows that the ϵ -dependence of this sensitivity (measured as the slope of the curves in Fig. 3a at $\xi = 1$) has a pronounced minimum at $\epsilon \approx 20$. Hence, intermediate hinge stiffnesses in the TSL lead to maximal robustness, which is an important design constraint for many biomolecular mechanisms in the cellular context. For instance, as molecular motors transport various cargos along one-dimensional filaments, it may be advantageous to render their speed insensitive to the cargo size, e.g. to avoid “traffic jams”.

In the remainder of this letter, we seek a theoretical understanding of the above phenomenology. First, it is instructive to consider simple bounds on the transition rate. An upper bound is obtained by completely eliminating the outer bead. The Kramers rate [15] for the remaining 1D escape process, $k_0 = (a^2 b / 2\pi) e^{-\Delta V / k_B T}$, is used in Figs. 2 and 3 to normalize the transition rates. At the optimal stiffness, the transition rate in Fig. 2 comes within 20% of this upper bound. An obvious lower bound is the stiff limit: For $\epsilon \rightarrow \infty$, the second segment increases

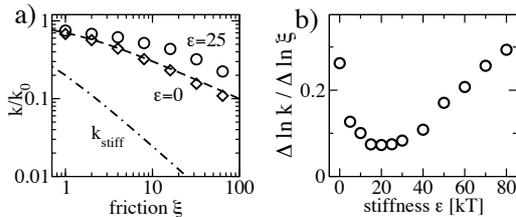


FIG. 3: The sensitivity of the rate to the friction coefficient ξ is minimal at intermediate stiffness. (a) Simulation results at $\epsilon = 0$ and $\epsilon = 25$ as well as the theoretical estimates of the rate at $\epsilon = 0$ and in the stiff limit. (b) The derivative of $\ln k$ with respect to $\ln \xi$ evaluated at $\xi = 1$, i.e. the slope of the curves in a), is minimal in an intermediate stiffness range.

the rotational friction by a factor $\zeta = 1 + (1 + \rho)^2 \xi$, so that the 1D Kramers rate becomes $k_\infty = k_0/\zeta$, as shown by the dash-dotted line in Fig. 2 and Fig. 3a. However, to understand how the dynamics of the bending fluctuations affects the transition rate, we must consider the full 2D dynamics of the TSL. The multi-dimensional generalization of Kramers theory is Langer’s formula for the escape rate over a saddle in a potential landscape [10],

$$k_{\text{Langer}} = \frac{\lambda}{2\pi} \times \sqrt{\frac{\det \mathbf{e}^{(w)}}{|\det \mathbf{e}^{(s)}|}} \exp\left(-\frac{\Delta V}{k_B T}\right). \quad (4)$$

Here, $\mathbf{e}^{(w)}$ and $\mathbf{e}^{(s)}$ denote the Hessian matrix of the potential energy, $\partial^2 V / \partial q_k \partial q_l$, evaluated at the well bottom and the saddle point, respectively, whereas λ is the unique negative eigenvalue of the product of the mobility matrix \mathbf{M} and $\mathbf{e}^{(s)}$. Eq. (4) can be made plausible in simple terms: Given a quasi-equilibrium in the metastable state, the second factor represents the probability of being in the transition region, i.e. the region within $\sim k_B T$ of the saddle. The escape rate is then given by this probability multiplied by the rate λ at which the system relaxes out of the transition state, analogous to Michaelis-Menten reaction kinetics.

For our potential (1), the determinants in (4) cancel. The eigenvalue can be determined analytically (the dashed line in Fig. 2 shows the resulting k_{Langer}), but for the present purpose it is more instructive to consider the expansions for large and small stiffness. In the stiff limit, the natural small parameter is the stiffness ratio γ/ϵ , where $\gamma = a^2 b$ is the absolute curvature or “stiffness” of the external potential at the transition state. The expansion yields $k_{\text{Langer}}/k_\infty = 1 + (\rho^2 \xi / \zeta) \gamma / \epsilon + \mathcal{O}(\gamma^2 / \epsilon^2)$. As expected, the rate approaches k_∞ , but the stiff limit is attained only when the bending fluctuations $\sim \sqrt{\epsilon}$ are small compared to the width of the barrier $\sim \sqrt{\gamma}$. In the opposite limit, $\epsilon \ll \gamma$, the rate is given by $k_{\text{Langer}}/k_0 = 1 - (1 + \rho^{-1})^2 \epsilon / \gamma + \mathcal{O}(\epsilon^2 / \gamma)$. Since the linear term is negative, Langer theory predicts that the transition rate peaks at zero stiffness, with a peak value equal to the Kramers rate k_0 for the lever without the second segment.

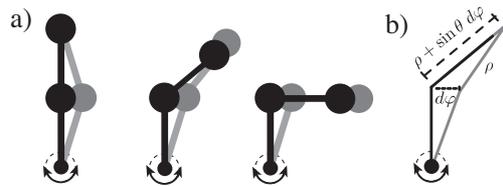


FIG. 4: The friction opposing rotation of the attachment angle φ depends on the bending angle θ , since the outer bead is moved by different amounts in different configurations. For an infinitesimal displacement $d\varphi$, the displacement of the outer bead is $\sin \theta d\varphi$. The projection of the resulting friction force onto the direction of motion adds another factor $\sin \theta$, yielding a friction coefficient for φ of $1 + \xi \sin^2 \theta$.

This prediction is clearly at variance with the simulation results. It is interesting to note, however, that the slope of the linear decay is independent of ξ . This is consistent with our observation that the transition rate is insensitive to ξ in the intermediate stiffness regime. Indeed, Fig. 2 shows that Langer theory (dashed line) describes the simulation data (circles) reasonably well for intermediate and large hinge stiffness.

To understand the origin of the peak at intermediate stiffness, it is useful to consider the flexible limit ($\epsilon = 0$). In this limit, the transition state is degenerate in θ , and it seems plausible to estimate the transition rate by using a θ -averaged mobility for the reaction coordinate φ ,

$$k(\epsilon = 0) \approx k_0 \int_{-\pi}^{\pi} \frac{d\theta}{2\pi} M_{11}(\theta) = \frac{k_0}{\sqrt{1 + \xi}}. \quad (5)$$

This estimate agrees well with the simulation data, see the dashed line in Fig. 3a, indicating that the configuration-dependent mobility (2) plays an important role for the transition rate. In contrast, the conventional Langer theory assumes the mobility matrix to be constant in the relevant region near the transition state. Fig. 4 illustrates why the mobility M_{11} of the coordinate φ is affected by the bending angle θ and gives a graphical construction for M_{11} .

Generalized Langer theory.— To account for the mobility effect identified above, we must generalize the Langer theory to configuration-dependent mobility matrices. The special case where the mobility varies only along the reaction coordinate has already been studied in [16], however the main effect in our case is due to the variation in the transverse direction. In the following, we outline the derivation of the central result, while all details will be presented elsewhere. Near the saddle, the mobility matrix takes the form $M_{ij}(\{q_i\}) = M_{ij}^{(s)} + \frac{1}{2} A_{ij}^{kl} \hat{q}_l \hat{q}_k$, where \hat{q}_i are deviations from the saddle and A_{ij}^{kl} denotes the tensor of second derivatives of the mobility matrix (we assume that the first derivatives of \mathbf{M} vanish at the saddle, which is the case for the TSL). The escape rate is given by the probability flux out of the metastable

well, divided by the population inside the well. To calculate the flux, we construct a steady state solution to the Fokker-Planck equation in the vicinity of the saddle, as described in [15] for the conventional Langer theory. We use the Ansatz $p(\{q_i\}) = \frac{1}{2}p_{eq}(\{q_i\})\text{erfc}(u)$, where $p_{eq}(\{q_i\}) = Z^{-1}e^{-V(\{q_i\})/k_B T}$ and $\text{erfc}(u)$ is the complementary error function with argument $u = U_k \hat{q}_k$. Inserting the Ansatz into the Fokker-Planck equation yields an equation for the vector \mathbf{U} ,

$$U_i(-M_{ij}e_{jk}^{(s)} + B_{ik}) - U_i M_{ij} U_j U_k = 0, \quad (6)$$

where $B_{ik} = k_B T \sum_n A_{ni}^{nk}$. $B_{ik} \hat{q}_k$ is the noise induced drift, which is absent in the conventional Langer theory. Ignoring higher order terms, this equation determines \mathbf{U} to be the left eigenvector of $-\mathbf{M}^{(s)}\mathbf{e}^{(s)} + \mathbf{B}$ to the unique positive eigenvalue λ , and requires \mathbf{U} to be normalized such that $U_i M_{ij}^{(s)} U_j = \lambda$. The directions of the left and right eigenvectors of $-\mathbf{M}^{(s)}\mathbf{e}^{(s)} + \mathbf{B}$ have a physical interpretation: \mathbf{U} is perpendicular to the stochastic separatrix, while the corresponding right eigenvector points in the direction of the diffusive flux at the saddle [17].

From $p(\{q_i\})$, the flux density is determined by (3) and the total flux is obtained by integrating the flux density over a plane containing the saddle; a convenient choice is the plane $u = 0$. Evaluation of the integral is particularly simple in a coordinate system, where the first coordinate is parallel to \mathbf{U} , and the remaining coordinates are chosen such that $\mathbf{e}^{(s)}$ is diagonal in this subspace, $e_{ij}^{(s)} = \mu_i \delta_{ij}$ for $i, j > 1$. In this coordinate system, the generalized Langer rate takes the simple form

$$k = \frac{\lambda}{2\pi} \frac{1 + \frac{1}{2M_{11}} \sum_{l>1} \frac{A_{11}^{ll}}{\mu_l}}{\sqrt{1-c}} \times \sqrt{\frac{\det \mathbf{e}^{(w)}}{|\det \mathbf{e}^{(s)}|}} e^{-\frac{\Delta V}{k_B T}}, \quad (7)$$

where $c = U_i \mathbf{e}_{ij}^{-1} U_j + 1 = B_{1i} e_{i1}^{-1} / M_{11}^{(s)}$ and \mathbf{e}^{-1} denotes the inverse matrix of $\mathbf{e}^{(s)}$. Eq. (7) contains three corrections to (4), all of which vanish when $\mathbf{M}(\{q_i\})$ is constant: The most important one is given by $\sum_{l>1} A_{11}^{ll} / \mu_l$, which changes the mobility M_{11} in the direction of \mathbf{U} to an effective mobility that is averaged over the separatrix with respect to the Boltzmann distribution. In addition, there are two corrections incurred by the noise-induced drift: the factor $\sqrt{1-c}$ and a change due to the fact that λ is now the eigenvalue to $\mathbf{M}^{(s)}\mathbf{e}^{(s)} - \mathbf{B}$ instead of $\mathbf{M}^{(s)}\mathbf{e}^{(s)}$.

The solid line in Fig. 2 shows the application of the generalized Langer formula to the TSL. We observe that it captures the peak in the transition rate and thus the essential phenomenology of the TSL. Obviously, the evaluation of the Gaussian integral that leads to Eq. (7) is only meaningful, if the harmonic approximation of the mobility matrix is reasonable within the relevant saddle point region. This integral diverges as the saddle point degenerates, which explains the behavior for $\epsilon \rightarrow 0$. At high ξ , the very anisotropic friction can also render Langer theory invalid [18, 19].

Conclusion.— We have introduced the “Two-Segment Lever” as a simple model for a class of conformational transitions in biomolecules. The model clearly demonstrates how flexibility can enhance the rate of a conformational transition. This remains true, if the hinge in the TSL is replaced by a more continuous bendability. Interestingly, a discrete hinge has a stiffness regime, where the rate is large and robust against cargo variation, which raises the question, whether these effects are exploited by evolution, for example in the design of molecular motors. To understand these effects theoretically, we derived a generalized Langer theory that takes into account configuration dependent mobility matrices. We hope that this theory will find applications also in other fields.

We thank the German Excellence Initiative for financial support via the program NIM. RN and UG are grateful for the hospitality of the CTBP at UCSD, where part of this work was done, and for financial support by the CeNS in Munich and the DFG.

-
- [1] J. Howard, *Mechanics of Motor Proteins and the Cytoskeleton* (Sinauer Press: Sunderland, MA, 2001).
 - [2] S. Sugimura and D. M. Crothers, PNAS **103**, 18510 (2006).
 - [3] G. Li, M. Levitus, C. Bustamante, and J. Widom, Nat. Struct. Mol. Biol. **12**, 46 (2005).
 - [4] W. Möbius, R. A. Neher, and U. Gerland, Phys. Rev. Lett. **97**, 208102 (2006).
 - [5] X. Zhuang, H. Kim, M. J. B. Pereira, H. P. Babcock, N. G. Walter, and S. Chu, Science **296**, 1473 (2002).
 - [6] M. Terrak, G. Rebowski, R. C. Lu, Z. Grabarek, and R. Dominguez, PNAS **102**, 12718 (2005).
 - [7] S. Jeney, E. H. K. Stelzer, H. Grubmüller, and E.-L. Florin, ChemPhysChem **5**, 1150 (2004).
 - [8] C. R. Doering and J. C. Gadoua, Phys. Rev. Lett. **69**, 2318 (1992).
 - [9] P. Reimann, Phys. Rev. Lett. **74**, 4576 (1995).
 - [10] J. S. Langer, Ann. Phys. **54**, 258 (1969).
 - [11] E. Helfand, J. Chem. Phys. **71**, 5000 (1979).
 - [12] N. van Kampen and J. Lodder, Am. J. Phys. **52**, 419 (1984).
 - [13] H. Goldstein, C. P. Poole, and J. L. Safko, *Classical Mechanics* (Addison Wesley, 2002).
 - [14] C. W. Gardiner, *Handbook of Stochastic Methods* (Springer-Verlag, 2004).
 - [15] P. Hänggi, P. Talkner, and M. Borkovec, Rev. Mod. Phys. **62**, 251 (1990).
 - [16] B. Gavish, Phys. Rev. Lett. **44**, 1160 (1980).
 - [17] A. Berezhkovskii and A. Szabo, J. Chem. Phys. **122**, 014503 (2005).
 - [18] A. N. Drozdov and P. Talkner, J. Chem. Phys. **105**, 4117 (1996).
 - [19] A. M. Berezhkovskii and V. Y. Zitserman, Chemical Physics **157**, 141 (1991).
 - [20] With the potential (1) this transition is irreversible, however all of our conclusions equally apply to reversible transitions in a double-well potential.

A. Partition sums of repetitive DNA

The analytic results obtained in the publications reprinted in Sec. 2 rely to a great extent on the calculation of the partition sum of repetitive double stranded DNA molecules. How this partition sum is calculated was discussed only briefly and I want to present the important steps in greater detail here. When the two strands have a repetitive and complementary sequence, the approximation that only native base pairs form is no longer justified. To the contrary, each repeat unit from one strand can bind to every repeat unit of the other strand. The configurations that contribute to the partition sum are therefore much more numerous and the calculation is considerably harder. The only assumption that can be reasonably made, is that base pairs do not cross, *i.e.* that the molecule can be depicted as a sequence of denatured loops and double stranded helices. However, the loops can now have a different number of bases on the two strands.

Every allowed configuration can be broken into essentially two different parts: the four open ends and the central part that is bounded by base pairs as illustrated in Fig. A.1. The partition sum of the molecule with N repeat units on one and M repeat units on the other strand is thus given by

$$Z(N, M) = \sum_{i,j,k,l=0}^{i+k < N, j+l < M} F_1(i)F_2(j)F_3(k)F_4(l)W_{M-j-l}^{N-i-k}, \quad (\text{A.1})$$

where $F_i(n)$ are the statistical weights of the single stranded ends of length n and W_s^r is the sum of all possible configuration of a double stranded part with r bases on one and s bases on the other strand. Similarly to the case where only native base pairs are allowed (see Sec. 1.2.2), W_s^r can be calculated from the recursion relation

$$W_{s+1}^{r+1} = q_{s+1}^{r+1}W_s^r + q_{s+1}^{r+1} \sum_{\substack{k < r, m < s \\ k+m > 1}} E_\ell(k, m)W_{s-m}^{r-k}, \quad (\text{A.2})$$

where q_s^r is the Boltzmann factor of the binding energy of base r and base s and $E_\ell(k, m)$ is the cost of having a loop with k bases on one and m bases on the other strand. The first term of the recursion relation accounts for all configuration where the base pair $(r+1, s+1)$ is added to any configuration in W_s^r , whereas the second term includes all configuration where the base pair $(r+1, s+1)$ followed by a loop of size (k, m) is added to any configuration in W_{s-m}^{r-k} . One easily convinces oneself, that this recursion indeed generates all allowed configurations.

This recursion relation can be solved numerically for arbitrary sequences in $\mathcal{O}(N^2M^2)$ operations [115, 116]. An analytical solution can be obtained, if the sequence is homogeneous, that is $q_s^r = q$, and if $E_\ell(k, m)$ is sufficiently well behaved. Note that every repetitive

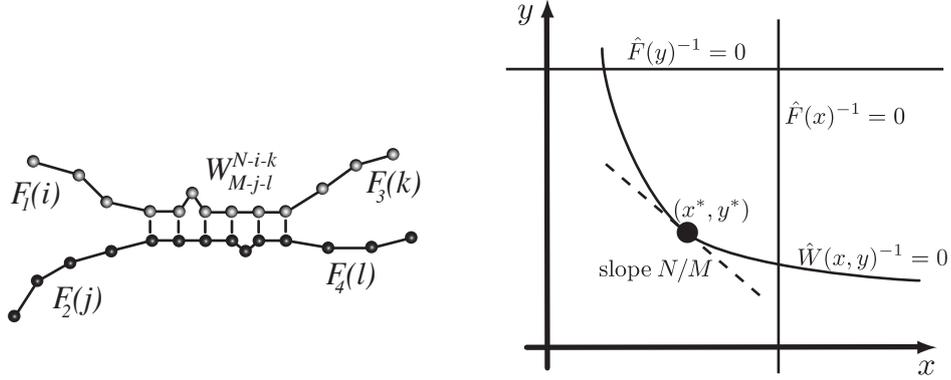


Figure A.1: Left: Every base-pairing configuration can be described by the length of the single stranded ends 1 to 4 and the part bounded by base pairs. Right: A sketch of the singularities of $\hat{F}(x)$, $\hat{F}(y)$ and $\hat{W}(x, y)$ in the positive quarter plane of real x and y .

sequence is essentially homogeneous since each repeat unit can be treated a single base which only binds to its native binding partner. Again, the recursion relation can be solved by z -transformation, but this time different fugacities x and y for the two strands are needed since they can be of different length and are not necessarily in register. To this end, we multiply both sides by $x^r y^s$ and sum over r and s to obtain

$$\frac{\hat{W}(x, y) - xyW_1^1}{xy} = q\hat{W}(x, y) + q\hat{E}_\ell(x, y)\hat{W}(x, y), \quad (\text{A.3})$$

where $\hat{W}(x, y) = \sum_{r,s=1}^{\infty} W_s^r x^r y^s$, $\hat{E}_\ell(x, y) = \sum_{r,s=1}^{\infty} E_\ell(r, s) x^r y^s$ and $W_1^i = W_i^1 = 0$ for all $i > 1$. This is readily solved for $\hat{W}(x, y)$, yielding

$$\hat{W}(x, y) = \frac{W_1^1}{\frac{1}{xy} - q - q\hat{E}_\ell(x, y)}, \quad (\text{A.4})$$

The z -transform of the partition sum in Eq. (A.5) is computed similarly and given by

$$\hat{Z}(x, y) = \hat{F}_1(x)\hat{F}_2(y)\hat{F}_3(x)\hat{F}_4(y)\hat{W}(x, y), \quad (\text{A.5})$$

where $\hat{F}_i(z)$ is the one variable z -transform of the single stranded ends.

The z -transformations in the base indices on both strands are equivalent to changing from a statistical ensemble with constant particle numbers to a grand ensemble, where the particle number is determined by the fugacities of a particle reservoir. The grand partition sum is a power series in $x^N y^M$ with coefficients $Z(N, M)$. Therefore $Z(N, M)$ can in principle be obtained from $\hat{Z}(x, y)$ by double contour integration

$$Z(N, M) = -\frac{1}{4\pi^2} \oint \oint dx dy \frac{\hat{Z}(x, y)}{x^{N+1} y^{M+1}}. \quad (\text{A.6})$$

In many cases, one of the integrals can be calculated by suitable deformation of the integration contour, but the remaining integral is usually infeasible. Nevertheless, a great deal of information can be obtained from the function $\hat{Z}(x, y)$, at least in the thermodynamic limit, which for linear molecules corresponds to the limit of long strands [117]. In this limit relative strand length fluctuations vanish and the grand ensemble is equivalent to the fixed length ensemble. The expected numbers of particles on both strands are given by

$$\langle N \rangle = x \frac{\partial \ln \hat{Z}(x, y)}{\partial x} \quad \text{and} \quad \langle M \rangle = y \frac{\partial \ln \hat{Z}(x, y)}{\partial y}, \quad (\text{A.7})$$

and for $\langle N \rangle$ or $\langle M \rangle$ to become arbitrarily large, $\hat{Z}(x, y)$ has to diverge. The thermodynamic limit thus confines the set of possible fugacities to the singularities of $\hat{Z}(x, y)$, which in general form a one dimensional, possibly multiply branched, set (cf. right part of Fig. A.1). To fix the fugacities completely, we need an additional condition. This is given by the requirement, that the ratio of the two strand length remains constant as the thermodynamic limit is approached. Otherwise, the intensive properties of the system are not preserved. We thus have to choose the fugacities (x^*, y^*) such that

$$\lim_{x, y \rightarrow x^*, y^*} \langle N \rangle = \infty \quad \text{and} \quad \lim_{x, y \rightarrow x^*, y^*} \langle M \rangle = \infty$$

$$\lim_{x, y \rightarrow x^*, y^*} \frac{\langle N \rangle}{\langle M \rangle} = c. \quad (\text{A.8})$$

Since the free energy of the DNA molecule should be extensive, the partition sum for long strands of length N and cN is of the form $Z(N, cN) = \gamma^N$. From the definition $\hat{Z}(x, y) = \sum_{N, M=1}^{\infty} Z(N, M) x^N y^M$ and the fact that only configurations with $M \approx Nc$ contribute, we have $\hat{Z}(x, y) \sim (1 - \gamma^N x^N y^{cN})^{-1}$. Hence, the singularities of $\hat{Z}(x, y)$ lie on the curve $x^* y^{*c} = \gamma^{-1}$. The limit of infinite strand length is approached from below $|xy^c| < \gamma^{-1}$. Within this approximation, which essentially is a saddle point approximation, the free energy of a molecule out of strands of length N and M is given by

$$F(N, M) = -k_B T \ln Z(N, M) = N \ln x^* + M \ln y^*, \quad (\text{A.9})$$

where the free energies per base $\ln x^*$ and $\ln y^*$ are determined by Eq. (A.8) and Eq. (A.7). Alternatively, one can perform the inverse transformation in one variable and thereby fix one strand length and then determine the remaining fugacity such that the expected length of the other strands matches the desired value. We will now apply this formalism to concrete problems of repetitive DNA under shear force and thermal denaturation of DNA. A more detailed description of this derivation is given in ref. [117].

A.1. Thermal denaturation of repetitive DNA

To describe the melting transition of repetitive DNA, we have to specify the Boltzmann factor associated with the binding of one repeat unit and the statistical weights of loops and single stranded ends. The statistical weights of loops and free ends are dominated

by the entropy of the different configurations available to the flexible polymer, as already discussed in Sec. 1.2.2. The statistical weights of free single stranded ends of length n and of loops with n and m on the two strands are given by

$$F(n) = \frac{s^n}{n^{\bar{c}}} \quad \text{and} \quad E_\ell(n, m) = g^2 \frac{s^{n+m}}{(n+m)^c}, \quad (\text{A.10})$$

where the exponents \bar{c} and c describe the excluded volume effects of an open end and a closed loop. While \bar{c} is irrelevant for the melting transition, c is pivotal and its value is assumed to lie in the range 1.8...2.15 [19] (cf. Sec. 1.2.2). The z transformations of the weights are given by

$$\hat{F}(x) = \Phi_{\bar{c}}(xs) \quad \text{and} \quad \hat{E}_\ell(x, y) = g^2 \frac{x\Phi_c(xs) - y\Phi_c(ys)}{x - y}. \quad (\text{A.11})$$

The factor g^2 accounts for a finite energy penalty to initiate a loop. The function $\Phi_c(z)$ is the polylogarithm, which is analytic everywhere in the complex plane, except on the interval $[1, \infty[$ of the real axis.

Inserting $\hat{F}(x)$ and $\hat{E}_\ell(x, y)$ into equation Eq. (A.5) yields the grand canonical partition sum two homogeneous DNA strands. As described above, the asymptotic behavior of the two strands is governed by the singularity of $\hat{Z}(x, y)$ which is closest to the origin, subject to the constraint that the ratio of the two strand length is fixed. Depending on whether this singularity is the branch-cut induced by the polylogarithm or an isolated singularity of $\hat{W}(x, y)$, the two strands are denatured or bound to each other. The phase behavior of such systems is discussed in Sec. 2.6.

A.2. Pulling on repetitive DNA

When exerting a shear force to the DNA, we have to distinguish between ends where the force is applied to and the unstretched ends. From now on, we model the ssDNA by the freely jointed chain (FJC) model, neglecting self-avoidance. Furthermore, we give all energies relative to unconstrained single strand. Consequently, for unstretched ends we have $F_2(n) = 1$ irrespective of the length n . The transformed $\hat{F}_2(x)$ is given by $(1 - x)^{-1}$. The free energy of the DNA molecule in the external force field is given by the product of an effective length L and the magnitude of the force. This effective length can be calculated separately for the single stranded ends and the double stranded region in between. The free energy of a segment of a FJC polymer with Kuhn length l_k under a tension f is given by the integral over all orientations of a segment

$$\frac{1}{2} \int e^{-\frac{fl_k \cos \theta}{k_B T}} \sin \theta d\theta = \frac{k_B T}{fl_k} \sinh \left(\frac{fl_k}{k_B T} \right). \quad (\text{A.12})$$

The free energy per monomer of length ℓ_s is thus given by $f \cdot \bar{\ell}_s = \frac{\ell_s}{l_k} \ln \left(\frac{k_B T}{fl_k} \sinh \left(\frac{fl_k}{k_B T} \right) \right)$, where $\bar{\ell}_s$ is an effective length. The contribution of stretched ssDNA of length n is therefore $F_1(n, f) = e^{n\bar{\ell}_s f} = \delta_s^n$ and its z -transform reads $\hat{F}_1(x, f) = (1 - x\delta_s)^{-1}$. dsDNA is

sufficiently stiff and long, such that we can assume that its completely aligned, yielding a free energy per base pair $f \cdot \ell_d$. Together with the binding free energy the statistical weight of a base pair is given by $q = e^{f\ell_d + \varepsilon_b} = \delta_d q_0$. The statistical weights of loops within the sequence is a bit more subtle to calculate. We assume, that the projected length of a loop is given by the length the shorter arm would have when in a double helix. As loop cost function, we use $E_\ell(n, m, f) = g^2 e^{\ell_d \min(n, m) f} = g^2 \delta_d^{\min(n, m)}$, where g^2 accounts for loop initiation and the exponential for the stretching energy. The two variable z -transformation of this quantity is a bit laborious and yields

$$\hat{E}_\ell(x, y, f) = \frac{g^2}{1 - \delta_d xy} \left(\frac{x}{1 - x} + \frac{y}{1 - y} + \delta_d \right) \quad (\text{A.13})$$

Inserting the different functions for loop cost and single strand contributions in Eq. (A.5) yields the grand canonical partition sum for sheared DNA. The single strand factors have obvious singularities at $x = 1$, $y = 1$, $x = \delta_s^{-1}$ and $y = \varepsilon_b^{-1}$, corresponding to no stable binding and completely stretched single strand, respectively. At low forces, however, $\hat{W}(x, y)$ has an additional singularity at (x^*, y^*) corresponding to a state where the two strands bind with maximal overlap. The transition from this bound state the stretched state occurs when $(x^*, y^*) = (\delta_s^{-1}, \delta_s^{-1})$. In the limit of high loop cost the contribution $\hat{E}_\ell(x, y, f)$ in the denominator can be neglected and it is easily seen that the singularity is found at $x^* y^* = \delta_d q_0$. The transition to the open state occurs at $x^* y^* = \delta_s^{-2} = \delta_d q_0$, which yields $f_c = \frac{\varepsilon_b}{2\ell_s - \ell_d}$. At finite loop cost the double stranded region is stabilized by the combinatorial entropy of the different double stranded conformations. Typically, this contribution is small.

A.2.1. Determining loop densities.

In Sec. 2, it was argued that the sliding velocities can be calculated from the pseudo-equilibrium loop densities, which in turn can be calculated from the proper equilibrium properties of a suitably chosen system. This system cannot be a double strand with a shear force applied, since at equilibrium in the supercritical regime, there is no double stranded region and no loop density can be defined. The argument leading to the pseudo-equilibrium density was, that the two strands move slowly relative to each other, such that the densities at the ends equilibrate. We now idealize this assumption, by attaching both strands to a wall at one end, and apply a force to the longer strand at the other end, as illustrated in Fig. A.2. The partition sum of this system is given by

$$\hat{Z}(x, y) = \hat{F}_s(x) \hat{F}_u(y) \hat{W}(x, y), \quad (\text{A.14})$$

where the stretched single strand contributes $\hat{F}_s(x) = (1 - x\delta_s^{-1})$, the unstretched single strand $\hat{F}_u(y) = (1 - y)^{-1}$ and $\hat{W}(x, y)$ is the same as above. Given the complete partition sum, we can now calculate the average number of base pairs via

$$\langle N_{bp} \rangle = \frac{\partial \ln \hat{Z}(x, y)}{\partial \ln q}. \quad (\text{A.15})$$

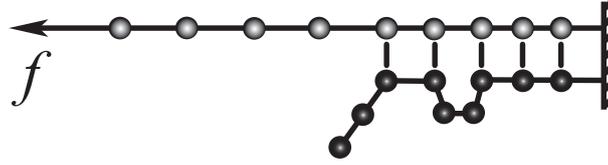


Figure A.2: Using a double strand attached to a wall, we can calculate the loop densities on the stretched and unstretched strands, even for $f > f_c$.

Furthermore, we can calculate the number of nucleotides on the upper and lower strand inside the double helical region by

$$\langle N_u \rangle = \frac{\partial \ln \hat{W}(x, y)}{\partial \ln x} \quad \text{and} \quad \langle N_l \rangle = \frac{\partial \ln \hat{W}(x, y)}{\partial \ln y}. \quad (\text{A.16})$$

Similarly, we can calculate the average number of loops inside the double helical region by differentiating $\ln \hat{Z}(x, y)$ with respect to $2 \ln g$. From these quantities the loop densities, the stored length and the mean loop size are readily calculated.

B. Generalized Kramers-Langer rates

This appendix provides a detailed derivation of the generalized Kramers-Langer-Theory presented in Sec. 3.5. The derivation is kept general and we refer the reader to Sec. 3.5 or the publication reprinted in Sec. 3.8 for a concrete application. In 1969, J. Langer published a generalization of the celebrated Kramers' formula for thermally activated escape from a one-dimensional potential well to barrier crossing processes in many dimensions [114]. A concise and clear derivation of this formula can be found in ref. [107]. The Langer formula, as presented by Hänggi et al.[107], assumes a constant mobility matrix in the relevant saddle point region. However, in many systems and especially in those described by generalized coordinates, this assumption cannot be made. Here, we seek to incorporate the configuration dependence of the mobility matrix into the Langer formula.

The escape rate out of a meta-stable potential well separated from a stable region by a unique saddle point is given by ratio of the probability current over the saddle point and the population inside the well. The evolution of the probability density $\mathcal{P}(\{\eta\})$ is determined by the Fokker-Plank equation (FPE) $\partial_t \mathcal{P}(\{\eta\}, t) = -\nabla \cdot \mathbf{j}(\{\eta\}, t)$, where the flux density $\mathbf{j}(\{\eta\}, t)$ is given by

$$j_i(\{\eta\}, t) = - \left[M_{ij} \frac{\partial V(\{\eta\})}{\partial \eta_j} + k_B T M_{ij} \partial_j \right] \mathcal{P}(\{\eta\}, t), \quad (\text{B.1})$$

The matrix \mathbf{M} is the mobility matrix which in general depends on the configuration $\{\eta\}$ and $V(\{\eta\})$ is the potential energy. We consider only the purely diffusive case and neglect any symplectic contributions to \mathbf{M} , which is justified since the applications in mind are completely overdamped. Let us now assume that particles are injected into the meta-stable well at a constant rate and removed from the stable region. In this case, $\mathcal{P}(\{\eta\})$ tends to a steady state, with a distribution that is very close the equilibrium distribution inside the meta-stable well, and which vanishes beyond the saddle where particles are removed. This steady state distribution carries a probability flux density out of the meta-stable well into the region where particles are removed. Obviously, the total flux integrated over any surface separating the insertion site from the absorbing boundary is equal to the rate of particle insertion. The non-trivial task is to relate this flux to the population inside the meta-stable well, *i.e.* the number of particles that accumulate before the steady state is attained. To this end, we solve the FPE in the vicinity of the saddle point where the probability flux is concentrated to a narrow channel, and match this solution to the approximate pseudo-equilibrium solution inside the meta-stable region. The size of the relevant saddle point region is determined by the curvature of the potential energy at the saddle point. Within this region, we can expand the potential energy, resulting in the simple FPE

$$\partial_i [M_{ij} E_{jk} \eta_k + M_{ij} k_B T \partial_j] \mathcal{P}(\{\eta\}) = 0, \quad (\text{B.2})$$

where \mathbf{E} is the Hessian of the energy near the saddle point. Since we need to match the solution near the saddle point to the equilibrium distribution inside the well, we rewrite $\mathcal{P}(\{\eta\})$ in the form $\mathcal{P}(\{\eta\}) = \mathcal{P}_{eq}(\{\eta\})\zeta(\{\eta\})$, where $\mathcal{P}_{eq}(\{\eta\})$ is the equilibrium distribution. Using this ansatz we can decompose the above equation into the equations

$$\begin{aligned} \partial_i \zeta(\{\eta\}) [M_{ij} E_{jk} \eta_k + k_B T M_{ij} \partial_j] \mathcal{P}_{eq}(\{\eta\}) &= 0 \\ \partial_i \mathcal{P}_{eq}(\{\eta\}) k_B T M_{ij} \partial_j \zeta(\{\eta\}) &= 0, \end{aligned} \quad (\text{B.3})$$

the first of which is trivially fulfilled since it includes the current of $\mathcal{P}_{eq}(\{\eta\})$.

If the mobility matrix changes significantly inside the saddle point region, its variation has to be taken into account. Here, we seek to incorporate this variations by expanding the mobility matrix about the saddle point and calculate the correction to the rate. Each entry of the mobility matrix \mathbf{M} can be expanded separately as

$$M_{ij} = M_{ij}^0 + \frac{\partial M_{ij}}{\partial \eta_k} \eta_k + \frac{1}{2} \frac{\partial^2 M_{ij}}{\partial \eta_k \partial \eta_l} \eta_k \eta_l + \mathcal{O}(\eta^2), \quad (\text{B.4})$$

where η_k are the deviations from the saddle. For symmetry reasons the linear dependence on η_k will often vanish and for the moment we will drop the linear term. In short, we have $M_{ij} = M_{ij}^0 + \frac{1}{2} A_{kl}^{ij} \eta_k \eta_l$ with the symmetric matrix \mathbf{A}^{ij} for each entry of \mathbf{M} . Inserting this expansion into Eq. (B.3) yields

$$[-E_{ik} M_{ij} + k_B T B_{jk}] \eta_k \partial_j \zeta(\{\eta\}) + k_B T M_{ij} \partial_i \partial_j \zeta(\{\eta\}) = 0, \quad (\text{B.5})$$

where the matrix \mathbf{B} is defined by

$$\partial_i M_{ij} = \frac{1}{2} \sum_k (\delta_{il} A_{lk}^{ij} \eta_k + \delta_{ik} A_{lk}^{ij} \eta_l) = \sum_k A_{ik}^{ij} \eta_k = B_{jk} \eta_k. \quad (\text{B.6})$$

$B_{kl} \eta_l$ is the noise induced drift due to the configuration dependence of the mobility matrix and which is absent in the conventional Langer theory. To solve this equation, we employ the ansatz [107]

$$\zeta(\{\eta\}) = \frac{1}{\sqrt{2\pi k_B T}} \int_u^\infty \exp\left(-\frac{z^2}{2k_B T}\right) dz, \quad (\text{B.7})$$

where the lower integration boundary depends on η_k via $u = \sum_i U_i \eta_i$, where the vector \mathbf{U} is to be determined by Eq. (B.5). This function interpolates smoothly between one inside the meta-stable region and tends to zero beyond the saddle point and therefore automatically satisfies the matching condition. Inserting this ansatz into Eq. (B.5) yields, after a bit of algebra

$$[U_j (-M_{ji} E_{ik} + k_B T B_{jk}) - U_i M_{ij} U_j U_k] \eta_k = 0 \quad (\text{B.8})$$

Since this equation should hold for any set η_k , the term in brackets itself has to vanish. Eq. (B.8) is equivalent to equation 4.71 in ref. [107], but includes the additional drift

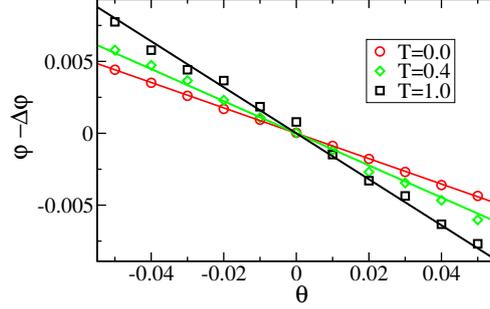


Figure B.1: The slope of stochastic separatrices at the saddle point changes with temperature. The slope calculated from the left eigenvector of $-M_{ji}^0 E_{ik} + k_B T B_{jk}$ (solid lines) agrees with simulation results (symbols) within their uncertainty.

$k_B T B_{jk} \eta_k$ induced by the configuration dependence of \mathbf{M} . In this equation, the explicit configuration dependence of $M_{ij} = M_{ij}^0 + \frac{1}{2} A_{kl}^{ij} \eta_k \eta_l$ is of second order in η_k and can therefore be neglected. After substituting M_{ij}^0 for M_{ij} , Eq. (B.8) is an eigenvector equation for \mathbf{U} , which determines \mathbf{U} to be a left eigenvector of $-M_{ji}^0 E_{ik} + k_B T B_{jk}$. The norm of \mathbf{U} is fixed by the condition $\lambda = U_i M_{ij}^0 U_j$, where λ is the eigenvalue corresponding to the eigenvector \mathbf{U} . In particular, this condition requires λ to be positive. The necessity of the noise-induced drift term is illustrated in Fig. B.1, where the stochastic separatrix is plotted for different temperatures. The vector \mathbf{U} has an appealing interpretation: \mathbf{U} is perpendicular to the stochastic separatrix, *i.e.* the hyperplane where the probabilities to relax either into the meta-stable or stable region are both equal to 0.5. The right eigenvector to the same eigenvalue points into the direction of the diffusive flux at the transition state [118]. The right and left eigenvectors coincide if \mathbf{M} is diagonal.

B.1. The flux over the barrier

Given the approximate steady state solution of the FPE, the probability flux reads

$$j_i = - \left[M_{ij} \frac{\partial V(\{\eta\})}{\partial \eta_j} + k_B T M_{ij} \partial_j \right] \mathcal{P}(\{\eta\}) = \frac{k_B T \mathcal{P}_{eq}(\{\eta\})}{\sqrt{2\pi k_B T}} \exp\left(-\frac{u^2}{2k_B T}\right) M_{ij} U_j, \quad (\text{B.9})$$

where M_{ij} is the full configuration dependent mobility matrix. The total flux out of the meta-stable region can now be calculated by integrating the flux density over a surface surrounding that region. To calculate the total flux over the saddle, we integrate the flux density over the plane given by the separatrix $u = 0^1$ (cf. Appendix of ref. [107]). Since the flux is strongly concentrated at the saddle point, we can again expand the potential energy and the mobility matrix. The resulting integral reads

$$J = Z^{-1} \sqrt{\frac{k_B T}{2\pi}} \int_{u=0} \frac{ds}{|\mathbf{U}|} e^{-\frac{1}{2} \frac{E_{ij} \eta_i \eta_j}{k_B T}} U_i \left(M_{ij}^0 + \frac{1}{2} A_{kl}^{ij} \eta_k \eta_l \right) U_j, \quad (\text{B.10})$$

¹Any plane that separates the meta-stable well from the absorbing boundary can be used, the separatrix is a particularly convenient choice.

where $\mathcal{P}_{eq}(\{\eta\}) = Z^{-1} e^{-\frac{1}{2} \frac{E_{ij}\eta_i\eta_j}{k_B T}}$ was substituted for the equilibrium distribution. To facilitate the evaluation of this surface integral it is helpful to choose suitable coordinate system. Let the direction of the first coordinate coincide with the vector \mathbf{U} , which is then obviously of the form $U_i = \delta_{i1} U_1$. In this particular system of coordinates, the surface integral reduces to the integral over the coordinates $2, \dots, N$, with $\eta_1 = 0$.

$$J = Z^{-1} \sqrt{\frac{k_B T}{2\pi}} \int_{\eta_1=0} \prod_{l>1} d\eta_l e^{-\frac{1}{2} \frac{E_{ij}\eta_i\eta_j}{k_B T}} U_1 \left(M_{11}^0 + \frac{1}{2} A_{ij}^{11} \eta_i \eta_j \right) \quad (\text{B.11})$$

The integral over the first term in parenthesis is readily evaluated, yielding

$$J^0 = Z^{-1} \frac{\lambda}{U_1} \sqrt{\frac{k_B T}{2\pi}} \frac{1}{|\hat{\mathbf{E}}^{11}/(2\pi k_B T)|^{\frac{1}{2}}}, \quad (\text{B.12})$$

where $\hat{\mathbf{E}}^{11}$ is the matrix \mathbf{E} with the first column and row removed². The normalization of \mathbf{U} has been used to substitute $\frac{\lambda}{U_1}$ for $U_1 M_{11}^0$. The integral of the second term can be evaluated by choosing the remaining coordinates such that $\hat{\mathbf{E}}^{11}$ is diagonal.

$$\begin{aligned} J^{corr} &= Z^{-1} \sqrt{\frac{k_B T}{2\pi}} \int_{\eta_1=0} \prod_{l>1} d\eta_l U_1 \frac{1}{2} A_{ij}^{11} \eta_i \eta_j e^{-\frac{1}{2} E_{ij} \eta_i \eta_j} \\ &= Z^{-1} \sqrt{\frac{k_B T}{2\pi}} \frac{U_1}{2|\hat{\mathbf{E}}^{11}/(2\pi k_B T)|^{\frac{1}{2}}} \sum_{l>1} \frac{A_l^{11}}{\hat{\mu}_l} \end{aligned} \quad (\text{B.13})$$

where the diagonal elements of the reduced matrix $\hat{\mathbf{E}}^{11}$ are denoted by $\hat{\mu}_2, \dots, \hat{\mu}_N$. To express the determinant of $\hat{\mathbf{E}}^{11}$ in the denominator through the determinant of the full Hessian of the potential energy, we need the relation

$$U_i E_{ij}^{-1} U_j = \frac{1}{\lambda} U_l (-M_{lk}^0 E_{ki} + k_B T B_{li}) E_{ij}^{-1} U_j = -1 + \frac{k_B T}{\lambda} U_l B_{li} e_{ij}^{-1} U_j = -1 + \gamma, \quad (\text{B.14})$$

where γ is a solely due to the noise induced drift. Using the well known formula for inverse matrices $E_{kl}^{-1} = \frac{1}{|\mathbf{E}|} (-1)^{k+l} |\hat{\mathbf{E}}^{kl}|$, we have $|\hat{\mathbf{E}}^{11}| = |\mathbf{E}| E_{11}^{-1} = |\mathbf{E}| = -|\mathbf{E}| \frac{1-\gamma}{U_1^2}$. Putting everything together, we find for the total flux out of the meta-stable well

$$J = Z^{-1} \frac{\lambda}{2\pi} \frac{1}{|\mathbf{E}|/(2\pi k_B T)|^{\frac{1}{2}} (1-\gamma)^{\frac{1}{2}}} \left(1 + \frac{1}{2M_{11}} \sum_{l>1} \frac{A_l^{11}}{\hat{\mu}_l} \right), \quad (\text{B.15})$$

The population inside the meta-stable region can be calculated within a Gaussian approximation

$$N = Z^{-1} \int d\eta e^{-\frac{1}{2} \frac{E_{ij}^w \eta_i \eta_j - \Delta U}{k_B T}} = \frac{e^{\frac{\Delta U}{k_B T}}}{Z |\mathbf{E}^w / (2\pi k_B T)|^{\frac{1}{2}}}, \quad (\text{B.16})$$

²The reduced matrix $\hat{\mathbf{E}}^{11}$ is positive definite and symmetric.

where \mathbf{E}^w is the Hessian of the potential energy at the bottom of the meta-stable potential well. Dividing the flux by the population inside the meta-stable well yields the generalized Langer rate

$$k = \frac{J}{N} = \frac{\lambda}{2\pi} \sqrt{\frac{|\mathbf{E}^w|}{|\mathbf{E}^t|(1-\gamma)}} e^{-\frac{\Delta U}{k_B T}} \left(1 + \frac{1}{2M_{11}} \sum_{l>1} \frac{A_{ll}^{11}}{\hat{\mu}_l} \right), \quad (\text{B.17})$$

where we labeled the the Hessian at the transition state with a superscript t . The different terms of this rate are easily interpreted. The ratio of the determinants, the exponential factor and the correction due to noise induced drift are the probability of finding the system near the transition state. The eigenvalue λ is the relaxation rate from the saddle. The term in parenthesis quantifies by what amount the mobility of the reaction coordinate averaged over the relevant window surrounding the saddle differs from the mobility at the saddle point itself. Note, that the latter correction term is given in the coordinate system where η_1 coincides with the direction of the reaction coordinate.

B.2. Stochastic dynamics of constrained systems

Many microscopic systems such as polymer chains or proteins have some degrees of freedom that vary in a large range and others that are confined to a very narrow range. Typically, the former are bending angles while the latter are linear dimensions. It is therefore tempting to replace the strongly confined degrees of freedom by rigid constraints and describe the system with generalized coordinates corresponding to the large amplitude degrees of freedom. Such a natural choice of coordinates is often helpful to elucidate the essential features of the system. Eliminating strongly confined degrees of freedom has also practical advantages, since the steep confining potentials require very small simulation time steps. Unfortunately, the limiting procedure to eliminate the constrained coordinates is ambiguous and the subtle difficulties arise where none would be expected. Quite generally, intuition is not a very good guideline when it comes to stochastic dynamics in curvilinear coordinate systems and things go awfully wrong if insufficient care is taken. These difficulties not only affect the dynamics of the system, but also the equilibrium distribution of the spatial coordinates. To illustrate this, let us consider a system with constrained and unconstrained degrees of freedom and compare their equilibrium distribution using rigid or flexible constraints. The Hamiltonian of the rigid version is given by

$$H(\{p_i\}, \{q_i\}) = \frac{1}{2} p_k t_{kl}(\{q_i\}) p_l + V(\{q_i\}), \quad (\text{B.18})$$

where $\{q_i\}$ are the generalized coordinates, $\{p_i\}$ are the conjugate momenta, $t_{kl}(\{q_i\})$ is the quadratic form of the kinetic energy³, and $V(\{q_i\})$ is the potential energy of the unconstrained coordinates. The corresponding energy function for the system with flexible constraints in cartesian coordinates is

$$H(\{\dot{x}_i\}, \{x_i\}) = \sum_i \frac{m_i}{2} \dot{x}_i^2 + V(\{q_i(\{x_i\})\}) + U(\{x_i\}), \quad (\text{B.19})$$

³ $t_{kl}(\{q_i\})$ is the inverse of the mass matrix in generalized coordinates.

where m_i are the masses and $U(\{x_i\})$ is the confining potential for the stiff directions. Submerged in a heat bath, the equilibrium distribution of the $\{q_i\}$ and their conjugate momenta $\{p_i\}$ is the Boltzmann distribution. The same holds true for the cartesian coordinates. However, after integrating over $\{p_i\}$, the distribution of the $\{q_i\}$ alone is no longer of Boltzmann form, but reads

$$\mathcal{P}_{rigid}(\{q_i\}) \sim \frac{1}{|t_{kl}(\{q_i\})|^{\frac{1}{2}}} e^{-\frac{V(\{q_i\})}{k_B T}} \quad (\text{B.20})$$

Integrating over momenta in cartesian coordinates is trivial. To compare the distribution of the spatial coordinates in both systems, let's express the cartesian coordinates by the unconstrained coordinates $\{q_i\}$ and the coordinates along the constraints $\{r_i\}$. Integrating over the $\{r_i\}$ yields

$$\mathcal{P}_{flex}(\{q_i\}) \sim g(\{q_i\}) e^{-\frac{V(\{q_i\})}{k_B T}}, \quad (\text{B.21})$$

where $g(\{q_i\})$ is the left-over of the Jacobian determinant of the coordinate transformation. Even for very simple systems, $\mathcal{P}_{flex}(\{q_i\})$ and $\mathcal{P}_{rigid}(\{q_i\})$ are different [119, 120, 121], although the potential energies are the same. When introducing the confining potential $U(\{x_i\})$ we assigned a small range to each constrained coordinate, which is independent of the $\{q_i\}$. We then integrated over constrained momenta and coordinates, resulting in a distribution that does not favor any region of the space of $\{q_i\}$'s apart from the volume element of the $\{q_i\}$ -coordinate system. When using rigid constraints, we ignore momenta of the constrained direction from the start. The integration over the conjugate momenta reduces the complete phase space to a subspace where some regions are favored over others. Loosely speaking, in these regions the conjugates momenta have a greater number of states available than in other regions.

We are interested in the stochastic dynamics of constrained systems, but we want to interpret them as stiff limits of physical springs since there are no truly rigid constraints in classical physics. If even the equilibrium properties of constrained systems are ambiguous or incompatible with the physical picture, dynamical features certainly are too. In the case of overdamped dynamics, however, there is a remedy to this problem. The difference between stiff and rigid constraints can be compensated by a pseudo potential [119, 122, 123]. Here, we apply a somewhat reverse, but equivalent strategy. We impose an equilibrium distribution on the generalized spatial coordinates $\{q_i\}$ of the form

$$\mathcal{P}_{eq}(\{q_i\}) \sim e^{-\Omega(\{q_i\})} e^{-\frac{V(\{q_i\})}{k_B T}}, \quad (\text{B.22})$$

where $V(\{q_i\})$ is the potential energy of the system and $\Omega(\{q_i\})$ accounts for obvious dependencies of the volume element on the $\{q_i\}$, *e.g.* $\Omega(\theta) = -\ln(\sin \theta)$ for three dimensional spherical coordinates. The key observation now is, that the choice of the equilibrium distribution and the deterministic overdamped relaxation fixes the Fokker-Planck equation governing the evolution of the probability distribution. The equations describing the deterministic relaxation of constrained systems can be most conveniently obtained from the

Euler-Lagrange equations including friction [124], which in the overdamped limit simplify to

$$\frac{\partial P(\{q_i\}, \{\dot{q}_i\})}{\partial \dot{q}_j} = -\frac{\partial V(\{q_i\})}{\partial q_j}. \quad (\text{B.23})$$

Assuming linear friction, the dissipation function $P(\{q_i\}, \{\dot{q}_i\})$ is given by the kinetic energy with all masses m_i substituted by the particle mobilities μ_i . Hence, $P(\{q_i\}, \{\dot{q}_i\})$ is a quadratic form in $\{\dot{q}_i\}$ and equation Eq. (B.23) can be solved for $\{\dot{q}_i\}$

$$\dot{q}_i = -M_{ij}(\{q_i\}) \frac{\partial V(\{q_i\})}{\partial q_j}, \quad (\text{B.24})$$

where $\mathbf{M}(\{q_i\})$ is the configuration dependent mobility matrix given by the inverse of the friction matrix $\mathbf{P}(\{q_i\})$. The system obeys detailed balance, requiring that the net probability flux vanishes for $\mathcal{P}_{eq}(\{q_i\})$

$$j_k = \left[F_k + D_{kl} \frac{\partial}{\partial q_l} \right] \mathcal{P}_{eq}(\{q_i\}) = 0. \quad (\text{B.25})$$

Here, the F_k are general drift terms for each coordinate and D_{kl} is the diffusion matrix. The drift terms have to reduce to the deterministic relaxation in the low temperature limit and can be written in the form $F_k = -M_{kl} \partial_l V(\{q_i\}) + F_k^{noise}$. F_k^{noise} are stochastic drift terms absent in the low temperature limit. Since the potential $V(\{q_i\})$ is arbitrary for a particular system with mobility matrix \mathbf{M} , the fluctuation dissipation theorem $D_{kl} = k_B T M_{kl}$ follows immediately. This identification fixes the noise induced drift forces to $F_k^{noise} = -k_B T M_{kl} \partial_l \Omega(\{q_i\})$, which vanishes at zero temperature as required. The full Fokker-Plank equation of our system is therefore given by

$$\frac{\partial}{\partial t} \mathcal{P}(\{q_i\}, t) = \frac{\partial}{\partial q_k} M_{kl}(\{q_i\}) \left[\partial_l V(\{q_i\}) + k_B T \partial_l \Omega(\{q_i\}) + k_B T \frac{\partial}{\partial q_l} \right] \mathcal{P}(\{q_i\}, t) \quad (\text{B.26})$$

From this Fokker-Plank equation, a simple procedure leads to the Langevin equations, which are to be interpreted in the Ito sense [87]. Since \mathbf{M} is positive definite and symmetric, a matrix \mathbf{B} with $k_B T M_{ij} = B_{il} B_{jl}$ can be found, and the stochastic dynamics of the q_i is governed by the equation

$$\dot{q}_k = -M_{kl}(\{q_i\}) [\partial_l V(\{q_i\}) + k_B T \partial_l \Omega(\{q_i\})] + k_B T \frac{\partial}{\partial q_l} M_{kl}(\{q_i\}) + \sqrt{2} B_{kl}(\{q_i\}) \eta_l, \quad (\text{B.27})$$

where η_l is a vector of Gaussian white noise terms with unit variance. Admittedly, the simulation of these equations is computationally expensive, since the mobility matrix has to be calculated in each time step by inverting the friction matrix. Hence the computational cost increases with the third power of the system size. A different approach to the simulation of constrained systems is to over-parameterize the system using cartesian coordinates and project the solution to the appropriate subspace [125, 126]. These algorithms run in linear time, since the matrix that needs to be inverted is a band matrix [127]. Unfortunately,

these algorithms are terribly complicated and disguise the physical nature of the problem by a mind-boggling formalism. Admittedly, I was not able to or did not invest enough time to understand them. Therefore I want by no means claim that the content of this chapter is an optimal solution, nor do I claim it to be original. On the other hand, if both approaches are correct, there should be a way to reconcile them.

B.2.1. Langevin equations for a multisegment chain

A chain in two dimensions can be described by a set of $N - 1$ angles with respect to some suitably defined reference axis, and the position of one bead, with respect to which all other coordinates are measured. Here, we assume that the first bead (bead zero) is fixed, which plays the role of the reference frame taken to be the origin. Now, the position of bead i is given by

$$x_i = \sum_{j=1}^i r_j \cos \phi_j \quad y_i = \sum_{j=1}^i r_j \sin \phi_j \quad (\text{B.28})$$

Its squared velocity is given by

$$\begin{aligned} \dot{x}_i^2 + \dot{y}_i^2 &= \sum_{k,l=1}^i r_k r_l \dot{\phi}_k \dot{\phi}_l (\sin \phi_k \sin \phi_l + \cos \phi_k \cos \phi_l) = \sum_{k,l=1}^i r_k r_l \dot{\phi}_k \dot{\phi}_l \cos(\phi_k - \phi_l) \\ &= \sum_{k=1}^i r_k^2 \dot{\phi}_k^2 + \sum_{l < k}^i r_k r_l \dot{\phi}_k \dot{\phi}_l \cos(\phi_k - \phi_l) + \sum_{k < l}^i r_k r_l \dot{\phi}_k \dot{\phi}_l \cos(\phi_k - \phi_l) \end{aligned} \quad (\text{B.29})$$

Hence, the dissipation function for unit friction coefficient of each bead is given by

$$\begin{aligned} P(\{\dot{\phi}_i\}) &= \frac{1}{2} \sum_{i=1}^N \dot{x}_i^2 + \dot{y}_i^2 \\ &= \frac{1}{2} \sum_{k=1}^N (N - k + 1) r_k^2 \dot{\phi}_k^2 + \sum_{k=1}^N \sum_{l=1}^{k-1} (N - k + 1) r_k r_l \dot{\phi}_k \dot{\phi}_l \cos(\phi_k - \phi_l) \end{aligned} \quad (\text{B.30})$$

and the elements of the friction matrix \mathbf{P} are

$$P_{ii} = (N - i + 1) r_i^2 \quad \text{and} \quad P_{ij} = (N - i + 1) r_i r_j \cos(\phi_i - \phi_j) \quad i > j. \quad (\text{B.31})$$

In general, the mobility matrix $\mathbf{M} = \mathbf{P}^{-1}$ has to be computed numerically, which makes the algorithm computationally expensive. The Langevin equations for the angles ϕ_j read

$$\dot{\phi}_j = -M_{jl} \frac{\partial V(\{\phi_i\})}{\partial \phi_l} + k_B T \frac{\partial M_{jl}}{\partial \phi_l} + \sqrt{2} B_{jl} \eta_l, \quad (\text{B.32})$$

where B_{ij} is chosen such that $B_{ij} B_{kj} = k_B T M_{ik}$.

Bibliography

- [1] Ferdinand Kühner, Julia Morfill, Richard A. Neher, Kerstin Blank, and Hermann Gaub. Force-induced DNA slippage. *Biophys. J.*, 92(2):2491–9497, 2007.
- [2] G. Li, M. Levitus, C. Bustamante, and J. Widom. Rapid spontaneous accessibility of nucleosomal DNA. *Nature Structural & Molecular Biology*, 12:46, 2005.
- [3] M. Tomschik, H. Zheng, K. van Holde, J. Zlatanova, and S.H. Leuba. Fast, long-range, reversible conformational fluctuations in nucleosomes revealed by single-pair fluorescence resonance energy transfer. *Proc. Natl. Acad. Sci. U.S.A.*, 102(9):3278, 2005.
- [4] U. Gerland, J. D. Moroz, and T. Hwa. Physical constraints and functional characteristics of transcription factor-dna interaction. *Proc. Natl. Acad. Sci. U.S.A.*, 99(19):12015–12020, September 2002.
- [5] N.E. Buchler, U. Gerland, and T. Hwa. On schemes of combinatorial transcription logic. *Proc. Natl. Acad. Sci. U.S.A.*, 100(9):5136, 2003.
- [6] International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature*, 409:860, 2001.
- [7] Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter. *Molecular Biology of the Cell*. Garland Science, New York, 2002.
- [8] John SantaLucia Jr. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. U.S.A.*, 95(4):1460–1465, 1998.
- [9] John SantaLucia and Donald Hicks. The thermodynamics of DNA structural motifs. *Annual Review of Biophysics and Biomolecular Structure*, 33(1):415–440, 2004.
- [10] Michael Zuker. Mfold web server for nucleic acid folding and hybridization prediction. *Nucl. Acids Res.*, 31(13):3406–3415, 2003.
- [11] Roger M. Wartell and Albert S. Benight. Thermal denaturation of DNA molecules: A comparison of theory with experiment. *Physics Reports*, 126(2):67–107, September 1985.
- [12] B H Zimm and J K Bragg. Theory of phase transition between helix and random coil in polypeptide chains. *J. Chem. Phys.*, 31(2):526–535, 1959.

- [13] Bruno H. Zimm. Theory of “melting” of the helical form in double chains of the DNA type. *J. Chem. Phys.*, 33(5):1349–1356, 1960.
- [14] Douglas Poland and Harold A. Scheraga. Phase transitions in one dimension and the helix—coil transition in polyamino acids. *J. Chem. Phys.*, 45(5):1456–1463, 1966.
- [15] Douglas Poland and Harold A. Scheraga. Occurrence of a phase transition in nucleic acid models. *J. Chem. Phys.*, 45(5):1464–1469, 1966.
- [16] Ralph Bundschuh and Terrence Hwa. RNA secondary structure formation: A solvable model of heteropolymer folding. *Phys. Rev. Lett.*, 83(7):1479, 1999.
- [17] C. Flamm, W. Fontana, I. L. Hofacker, and P. Schuster. RNA folding at elementary step resolution. *RNA*, 6(3):325–338, 2000.
- [18] Pierre-Gilles de Gennes. *Scaling Concepts in Polymer Physics*. Cornell University Press, 1979.
- [19] Yariv Kafri, David Mukamel, and Luca Peliti. Why is the DNA denaturation transition first order? *Phys. Rev. Lett.*, 85(23):4988–4991, 2000.
- [20] Y. Kafri, D. Mukamel, and L. Peliti. Melting and unzipping of DNA. *Eur. Phys. J. B*, 27:135 – 146, May 2002.
- [21] Andreas Hanke and Ralf Metzler. Comment on “Why is the DNA denaturation transition first order?”. *Physical Review Letters*, 90(15):159801, 2003.
- [22] Y. Kafri, D. Mukamel, and L. Peliti. Kafri, Mukamel, and Peliti Reply:. *Physical Review Letters*, 90(15):159802, 2003.
- [23] Michael E. Fisher. Effect of excluded volume on phase transitions in biopolymers. *J. Chem. Phys.*, 45(5):1469–1473, 1966.
- [24] Douglas Poland and Harold A. Scheraga. *Theory of Helix-Coil Transitions in Biopolymers*. Academic Press, 1970.
- [25] Ralf Blossey and Enrico Carlon. Reparametrizing the loop entropy weights: Effect on DNA melting curves. *Phys. Rev. E*, 68(6):061911, 2003.
- [26] Rudolf Merkel. Force spectroscopy on single passive biomolecules and single biomolecular bonds. *Phys. Rep.*, 346:343–385, June 2001.
- [27] C Bustamante, J C Macosko, and G J Wuite. Grabbing the cat by the tail: Manipulating molecules one by one. *Nature Rev. Mol. Cell. Biol.*, 1:130–136, 2000.
- [28] Hauke Clausen-Schaumann, Markus Seitz, Rupert Krautbauer, and Hermann E Gaub. Force spectroscopy with single bio-molecules. *Curr. Opin. Chem. Bio.*, 4: 524–530, October 2000.

- [29] U. Bockelmann, Ph. Thomen, B. Essevaz-Roulet, V. Viasnoff, and F. Heslot. Unzipping DNA with Optical Tweezers: High Sequence Sensitivity and Force Flips. *Biophys. J.*, 82(3):1537–1553, 2002.
- [30] Anabel E.-M. Clemen, Mojca Vilfan, Johann Jaud, Junshan Zhang, Michael Barmann, and Matthias Rief. Force-Dependent Stepping Kinetics of Myosin-V. *Biophys. J.*, 88(6):4402–4410, 2005.
- [31] T. R. Strick, J. F. Allemand, D. Bensimon, A. Bensimon, and V. Croquette. The elasticity of a single supercoiled DNA molecule. *Science*, 271(5257):1835–1837, March 1996.
- [32] T. R. Strick, V. Croquette, and D. Bensimon. Single-molecule analysis of DNA uncoiling by a type ii topoisomerase. *Nature*, 404(6780):901–904, April 2000.
- [33] R. Merkel, P. Nassoy, A. Leung, K. Ritchie, and E. Evans. Energy landscapes of receptor-ligand bonds explored with dynamic force spectroscopy. *Nature*, 397(6714):50–53, January 1999.
- [34] Nobuhiko Saitô, Kunihiko Takahashi, and Yasuo Yunoki. The statistical mechanical theory of stiff chains. *Journal of the Physical Society of Japan*, 22:219–226, 1967.
- [35] O Kratky and G Porod. X-ray investigation of dissolved chain molecules. *Recueil des Travaux Chimiques des Pays-Bas et de la Belgique*, 68:1106–1122, 1949.
- [36] Mary D. Barkley and Bruno H. Zimm. Theory of twisting and bending of chain macromolecules; analysis of the fluorescence depolarization of DNA. *The Journal of Chemical Physics*, 70(6):2991–3007, 1979.
- [37] Chris H. Wiggins, D. Riveline, A. Ott, and Raymond E. Goldstein. Trapping and Wiggling: Elastohydrodynamics of Driven Microfilaments. *Biophys. J.*, 74(2):1043–1060, 1998.
- [38] Jan Wilhelm and Erwin Frey. Radial distribution function of semiflexible polymers. *Phys. Rev. Lett.*, 77(12):2581–2584, Sep 1996.
- [39] SB Smith, L Finzi, and C Bustamante. Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads. *Science*, 258(5085):1122–1126, 1992.
- [40] John F. Marko and Eric D. Siggia. Stretching DNA. *Macromolecules*, 28(26):8759–8770, 1995. ISSN 0024-9297.
- [41] Philippe Cluzel, Anne Lebrun, Christoph Heller, Richard Lavery, Jean-Louis Viovy, Didier Chatenay, and Francois Caron. DNA: An Extensible Molecule. *Science*, 271(5250):792–794, 1996.

- [42] Steven B. Smith, Yujia Cui, and Carlos Bustamante. Overstretching B-DNA: The Elastic Response of Individual Double-Stranded and Single-Stranded DNA Molecules. *Science*, 271(5250):795–799, 1996.
- [43] Simona Cocco, Jie Yan, Jean-Francois Leger, Didier Chatenay, and John F. Marko. Overstretching and force-driven strand separation of double-helix dna. *Phys. Rev. E*, 70(1):011910, 2004.
- [44] M. Rief, H. Clausen-Schaumann, and H. E. Gaub. Sequence-dependent mechanics of single DNA molecules. *Nature Structural Biology*, 6(4):346–349, April 1999.
- [45] L. Livadaru, RR. Netz, and H.J. Kreutzer. Stretching response of discrete semiflexible polymers. *Macromolecules*, 36:3732–3744, 2003.
- [46] B. Essevaz-Roulet, U. Bockelmann, and F. Heslot. Mechanical separation of the complementary strands of DNA. *Proc. Natl. Acad. Sci. U.S.A.*, 94(22):11935–11940, 1997.
- [47] Mary E Craig, Donald M. Crothers, and Paul Doty. Relaxation kinetics of dimer formation by self complementary oligonucleotides. *J. Mol. Bio.*, 62(2):383–392, 1971.
- [48] V. V. Anshelevich, A. V. Vologodskii, A. V. Lukashin, and M. D. Frank-Kamenetskii. Slow relaxational processes in the melting of linear biopolymers: A theory and its application to nucleic acids. *Biopolymers*, 23:39–58, 1984.
- [49] David K. Lubensky and David R. Nelson. Pulling pinned polymers and unzipping DNA. *Phys. Rev. Lett.*, 85(3):31917, 2000.
- [50] David K. Lubensky and David R. Nelson. Single molecule statistics and the polynucleotide unzipping transition. *Phys. Rev. E*, 65(3):31917, 2002.
- [51] Simona Cocco, Remi Monasson, and John F. Marko. Force and kinetic barriers to initiation of DNA unzipping. *Phys. Rev. E*, 65(4):41907, 2002.
- [52] Claudia Danilowicz, Vincent W. Coljee, Cedric Bouzigues, David K. Lubensky, David R. Nelson, and Mara Prentiss. DNA unzipped under a constant force exhibits multiple metastable intermediates. *Proc. Natl. Acad. Sci. U.S.A.*, 100(4):1694–1699, 2003.
- [53] H. Ellegren. Microsatellites: Simple sequences with complex evolution. *Nature Reviews Genetics*, 5(6):435–445, June 2004.
- [54] R A Spritz. Duplication/deletion polymorphism 5' - to the human beta globin gene. *Nucleic Acids Res.*, 9:5037–5047, 1981.
- [55] John F. Dallas. Estimation of microsatellite mutation rates in recombinant inbred strains of mouse. *Mammalian Genome*, V3(8):452–456, August 1992.

- [56] Micheline Strand, Tomas A. Prolla, R. Michael Liskay, and Thomas D. Petes. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature*, 365(6443):274–276, September 1993.
- [57] Christian Schlötterer. Evolutionary dynamics of microsatellite DNA. *Chromosoma*, V109(6):365–371, September 2000.
- [58] JS Lee, MG Hanford, JL Genova, and RA Farber. Relative stabilities of dinucleotide and tetranucleotide repeats in cultured mammalian cells. *Hum. Mol. Genet.*, 8(13):2567–2572, 1999.
- [59] Semyon Kruglyak, Richard T. Durrett, Malcolm D. Schug, and Charles F. Aquadro. Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. *Proc. Natl. Acad. Sci. U.S.A.*, 95(18):10774–10778, 1998.
- [60] B.A. Lenzmeier and C.H. Freudenreich. Trinucleotide repeat instability: a hairpin curve at the crossroads of replication, recombination, and repair. *Cytogenet Genome Res*, 100(1-4):7–24, 2003. ISSN 1424-8581.
- [61] Malgorzata Bzymek and Susan T. Lovett. Instability of repetitive DNA sequences: The role of replication in multiple mechanisms. *Proc. Natl. Acad. Sci. U.S.A.*, 98(15):8319–8325, 2001.
- [62] P Bennett. Demystified ...: Microsatellites. *Mol Pathol*, 53(4):177–183, 2000.
- [63] P Sanjeeva Reddy and David E Housman. The complex pathology of trinucleotide repeats. *Curr. Opin. Cell Bio.*, 9:364–372, June 1997.
- [64] Eric R. Kandel, James H. Schwartz, and Thomas M. Jessell. *Principles of Neural Science*. McGraw-Hill Medical, 2000.
- [65] Gillian Bates. Huntingtin aggregation and toxicity in huntington’s disease. *The Lancet*, 361(9369):1642–1644, May 2003.
- [66] E. Kussell and S. Leibler. Phenotypic diversity, population growth, and information in fluctuating environments. *Science*, 309(5743):2075–2078, September 2005.
- [67] Christopher D. Bayliss, Dawn Field, and E. Richard Moxon. The simple sequence contingency loci of *Haemophilus influenzae* and *Neisseria meningitidis*. *J. Clin. Invest.*, 107(6):657–666, 2001.
- [68] C. Schlötterer and D. Tautz. Slippage synthesis of simple sequence DNA. *Nucleic Acids Research*, 20(2):211–215, January 1992.
- [69] X. Y. Hauge and M. Litt. A study of the origin of shadow bands seen when typing dinucleotide repeat polymorphisms by the PCR. *Human Molecular Genetics*, 2(4):411–415, April 1993.

- [70] V. Murray, C. Monchawin, and P. R. England. The determination of the sequences present in the shadow bands of a dinucleotide repeat PCR. *Nucleic Acids Research*, 21(10):2395–2398, May 1993.
- [71] Dietmar Pörschke. Model calculations on the kinetics of oligonucleotide double helix coil transitions. EVIDENCE for a fast chain sliding reaction. *Biophys. Chem.*, 2: 83–96, 1974.
- [72] T. Strunz, K. Oroszlan, R. Schafer, and H. J. Guntherodt. Dynamic force spectroscopy of single DNA molecules. *Proc. Natl. Acad. Sci. U.S.A.*, 96(20):11277–11282, September 1999.
- [73] Richard A. Neher and Ulrich Gerland. Dynamics of force-induced DNA-sliding. *Phys. Rev. Lett.*, 93:198102, 2004.
- [74] Martin Schwarz and Douglas Poland. Random walk with two interacting walkers. *J. Chem. Phys.*, 63(1):557–568, 1975.
- [75] Richard A. Neher and Ulrich Gerland. DNA as a Programmable Viscoelastic Nanoelement. *Biophys. J.*, 89(6):3846–3855, 2005.
- [76] Nadrian C. Seeman. DNA in a material world. *Nature*, 421:427–431, 2003.
- [77] William M. Shih, Joel D. Quispe, and Gerald F. Joyce. A 1.7-kilobase single-stranded dna that folds into a nanoscale octahedron. *Nature*, 427(6975):618–621, February 2004. ISSN 0028-0836.
- [78] Paul W. K. Rothmund. Folding dna to create nanoscale shapes and patterns. *Nature*, 440(7082):297–302, March 2006. ISSN 0028-0836.
- [79] Friedrich C. Simmel and Bernard Yurke. A DNA-based molecular device switchable between three distinct mechanical states. *Appl. Phys. Lett.*, 80(5):883–885, 2002.
- [80] David C. Lin, Bernard Yurke, and Noshir A. Langrana. Mechanical properties of a reversible, DNA-crosslinked polyacrylamide hydrogel. *J. Biomech. Eng.*, 126(1): 104–110, 2004.
- [81] D. S. Liu and S. Balasubramanian. A proton-fuelled DNA nanomachine. *Angewandte Chemie-International Edition*, 42(46):5734–5736, 2003.
- [82] T. Liedl and F.C. Simmel. Switching the conformation of a dna molecule with a chemical oscillator. *Nano Letters*, 5(10):1894–1898, 2005.
- [83] D. Poland and H. A. Scheraga. *Theory of Helix-Coil Transitions in Biopolymers*. Academic, New York, 1970.
- [84] Richard A. Neher and Ulrich Gerland. Intermediate phase in DNA melting. *Phys. Rev. E*, 73(3):030902, 2006.

- [85] P. H. von Hippel. An integrated model of the transcription complex in elongation, termination and editing. *Science*, 281:660, 1998.
- [86] R. V. Dalal, M. H. Larson, K. C. Neuman, J. Gelles, R. Landick, and S. M. Block. Pulling on the nascent RNA during transcription does not alter kinetics of elongation or ubiquitous pausing. *Molecular Cell*, 23:231, 2006.
- [87] C.W. Gardiner. *Handbook of Stochastic Methods*. Springer, 2004.
- [88] J. T. Finch and A. Klug. Solenoidal Model for Superstructure in Chromatin. *Proc. Natl. Acad. Sci. U.S.A.*, 73(6):1897–1901, 1976.
- [89] CL Woodcock, SA Grigoryev, RA Horowitz, and N Whitaker. A Chromatin Folding Model that Incorporates Linker Variability Generates Fibers Resembling the Native Structures. *Proc. Natl. Acad. Sci. U.S.A.*, 90(19):9021–9025, 1993.
- [90] Benedetta Dorigo, Thomas Schalch, Alexandra Kulangara, Sylwia Duda, Rasmus R. Schroeder, and Timothy J. Richmond. Nucleosome Arrays Reveal the Two-Start Organization of the Chromatin Fiber. *Science*, 306(5701):1571–1573, 2004. doi: 10.1126/science.1103124.
- [91] Thomas Schalch, Sylwia Duda, David F. Sargent, and Timothy J. Richmond. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature*, 436(7047):138–141, July 2005. ISSN 0028-0836.
- [92] Gaurav Arya and Tamar Schlick. Role of histone tails in chromatin folding revealed by a mesoscopic oligonucleosome model. *Proc. Natl. Acad. Sci. U.S.A.*, 103(44): 16236–16241, 2006.
- [93] C. L. Woodcock. Chromatin architecture. *Current Opinion in Structural Biology*, 16 (2):213–220, April 2006.
- [94] Yujia Cui and Carlos Bustamante. Pulling a single chromatin fiber reveals the forces that maintain its higher-order structure. *Proc. Natl. Acad. Sci. U.S.A.*, 97(1):127–132, 2000.
- [95] K. Luger, A.W. Mäder, R.K. Richmond, D.F. Sargent, and T.J. Richmond. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, 389:251, 1997.
- [96] C. A. Davey, D. F. Sargent, K. Luger, A. W. Maeder, and T. J. Richmond. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 angstrom resolution. *Journal of Molecular Biology*, 319(5):1097–1113, June 2002.
- [97] H. Schiessel. The physics of chromatin. *Journal of Physics: Condensed Matter*, 15: R699, 2003.

- [98] K.J. Polach and J. Widom. Mechanism of protein access to specific DNA sequences in chromatin: A dynamic equilibrium model for gene regulation. *Journal of Molecular Biology*, 254:130, 1995.
- [99] K.J. Polach and J. Widom. A model for the cooperative binding of eukaryotic regulatory proteins to nucleosomal target sites. *Journal of Molecular Biology*, 258:800, 1996.
- [100] J.D. Anderson and J. Widom. Sequence and position-dependence of the equilibrium accessibility of nucleosomal DNA target sites. *Journal of Molecular Biology*, 296:979, 2000.
- [101] Eran Segal, Yvonne Fondufe-Mittendorf, Lingyi Chen, AnnChristine Thastrom, Yair Field, Irene K. Moore, Ji-Ping Z. Wang, and Jonathan Widom. A genomic code for nucleosome positioning. *Nature*, 442(7104):772–778, March 2006. ISSN 0028-0836.
- [102] J.A. Miller and J. Widom. Collaborative competition mechanism for gene activation in vivo. *Molecular and Cellular Biology*, 23(5):1623, 2003.
- [103] G. Li, , and J. Widom. Nucleosomes facilitate their own invasion. *Nature Structural & Molecular Biology*, 11(8):763, 2004.
- [104] Wolfram Mobius, Richard A. Neher, and Ulrich Gerland. Kinetic accessibility of buried DNA sites in nucleosomes. *Physical Review Letters*, 97(20):208102, 2006.
- [105] N.L. Marky and G.S. Manning. A theory of DNA dissociation from the nucleosome. *Journal of Molecular Biology*, 254:50, 1995.
- [106] H. A. Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7(4):284–304, April 1940.
- [107] Peter Hänggi, Peter Talkner, and Michal Borkovec. Reaction-rate theory: fifty years after kramers. *Rev. Mod. Phys.*, 62(2):251–341, Apr 1990.
- [108] M. Dio and S. Edwards. *The Theory of polymer dynamics*. Clarendon Press, Oxford, 1986.
- [109] J. Howard. *Mechanics of Motor Proteins and the Cytoskeleton*. Sinauer Press: Sunderland, MA, 2001.
- [110] Sawako Sugimura and Donald M. Crothers. Stepwise binding and bending of DNA by Escherichia coli integration host factor. *Proc. Natl. Acad. Sci. U.S.A.*, 103(49):18510–18514, 2006.
- [111] Serguei V. Kuznetsov, Sawako Sugimura, Paula Vivas, Donald M. Crothers, and Anjum Ansari. Direct observation of DNA bending/unbending kinetics in complex with DNA-bending protein IHF. *Proc. Natl. Acad. Sci. U.S.A.*, 103(49):18515–18520, 2006.

- [112] Sylvia Jeney, Ernst H. K. Stelzer, and Helmut Grubmller Ernst-Ludwig Florin. Mechanical Properties of Single Motor Molecules Studied by Three-Dimensional Thermal Force Probing in Optical Tweezers. *ChemPhysChem*, 5(8):1150–1158, 2004.
- [113] Mohammed Terrak, Grzegorz Rebowski, Renne C. Lu, Zenon Grabarek, and Roberto Dominguez. Structure of the light chain-binding domain of myosin V. *Proc. Natl. Acad. Sci. U.S.A.*, 102(36):12718–12723, 2005.
- [114] J. S. Langer. Statistical theory of decay of metastable states. *Annals of Physics*, 54(2):258–275, 1969.
- [115] Thomas Garel and Henri Orland. Generalized Poland-Scheraga model for DNA hybridization. *Biopolymers*, 75:453–467, 2004.
- [116] R. Bundschuh and T. Hwa. Statistical mechanics of secondary structures formed by random RNA sequences. *Phys. Rev. E*, 65(3):31903, 2002.
- [117] M. V. Tamm and S. K. Nechaev. Necklace-cloverleaf transition in associating RNA-like diblock copolymers. *Phys. Rev. E*, 75(3):031904, 2007.
- [118] Alexander Berezhkovskii and Attila Szabo. One-dimensional reaction coordinates for diffusive activated rate processes in many dimensions. *The Journal of Chemical Physics*, 122(1):014503, 2005.
- [119] Eugene Helfand. Flexible vs rigid constraints in statistical mechanics. *J. Chem. Phys.*, 71(12):5000–5007, 1979.
- [120] N. G. van Kampen and J. J. Lodder. Constraints. *American Journal of Physics*, 52(5):419–424, 1984.
- [121] Marshall Fixman. Classical Statistical Mechanics of Constraints: A Theorem and Application to Polymers. *Proc. Natl. Acad. Sci. U.S.A.*, 71(8):3050–3053, 1974.
- [122] Marshall Fixman. Simulation of polymer dynamics. i. general theory. *J. Chem. Phys.*, 69(4):1527–1537, 1978.
- [123] Daniel B. Roitman and Bruno H. Zimm. An elastic hinge model for dynamics of stiff chains. i. viscoelastic properties. *J. Chem. Phys.*, 81(12):6333–6347, 1984.
- [124] Herbert Goldstein, Charles P. Poole, and John L. Safko. *Classical Mechanics*. Addison Wesley, 2002.
- [125] E.J. Hinch. Constraints. *Journal of Fluid Mechanics*, 271:219, 1994.
- [126] D.C. Morse. Theory of constrained brownian motion. *Advances in Chemical Physics*, 128:65–189, 2004.

- [127] Matteo Pasquali and David C. Morse. An efficient algorithm for metric correction forces in simulations of linear polymers with constrained bond lengths. *The Journal of Chemical Physics*, 116(5):1834–1838, 2002.

Glossary

A	Adenine, page 2.
C	Cytosine, page 2.
G	Guanine, page 2.
T	Thymine, page 2.
AFM	Atomic force microscope, page 13.
Codon	The genetic code associates with three consecutive bases in DNA one specific amino acid. One element of the genetic code is known as codon.
DNA	Deoxyribonucleic acid, page 1.
dsDNA	Double stranded DNA.
FCS	Fluorescence Correlation Spectroscopy.
FJC	Freely jointed chain.
FRET	Förster Resonance Energy Transfer, page 71.
Genetic code	Since there are more amino acids than bases, a multi-letter code is used to store an amino-acid sequence. Each amino acid is encoded by three consecutive bases, known as codons. The genetic code is redundant.
Genome	The complete hereditary information of an organism encoded in DNA.
LexA	A protein that binds strongly to specific DNA sequences.
Microsatellite	Simple sequence repeat.
mRNA	Messenger RNA. Messenger RNAs are transcripts of genes that are translated into protein by the ribosome.
Nucleotide	Nucleotides are the elementary building blocks of RNA or DNA. They consist of a sugar, one or several phosphates and a base, page 2.

Okazaki fragment	Piece of DNA polymerized continuously during piecewise replication of the lagging strand, page 19.
PCR	Polymerase chain reaction. By PCR small amounts of DNA can be amplified rapidly and cheaply.
Polymerase	Protein complex that transcribes DNA into a complementary RNA or DNA strand.
PS-models	Poland-Scheraga models. A class of models for base pairing configurations of dsDNA, page 7.
Ribosome	The ribosome is a RNA-protein complex that translates the mRNA into proteins.
RNA	Ribonucleic acid. The structure of RNA and DNA are very similar. RNA nucleotides are made from a the sugar ribose instead of deoxyribose. The base complementary to adenine is uracil instead of thymine.
Short tandem repeat	Simple sequence repeat.
ssDNA	Single stranded DNA.
SSR	Simple sequence repeat. A multi-fold repetition of a short (one to six base pairs) motif, page 18.
Stacking interactions	Consecutive base pairs in dsDNA stack on top of each other and thereby drive water out of the inter-base region. These stacking interaction are a major contribution to the DNA binding free energy, page 3.
TF	Transcription factor. Transcription factors are specialized proteins that bind to DNA and to regulate gene expression.
WLC	Worm-like chain, page 13.

Danksagung

Zu vorderst gebührt mein Dank Prof. Ulrich Gerland für die vorbildliche Betreuung und die interessante Thematik der Arbeit. Insbesondere möchte ich mich für seine unermüdlichen Bemühungen, mich für die faszinierende und verwirrende Welt der Biologie zu begeistern, bedanken. Es hat am Ende doch noch funktioniert. Ich hatte während meiner Doktorarbeit alle Freiheit die ich mir wünschen konnte und gleichzeitig immer die Möglichkeit Probleme zu erörtern oder um Rat zu fragen.

Desweiteren möchte ich mich bei Wolfram Möbius für die fruchtbare Zusammenarbeit zur Dynamik von Nukleosomen bedanken. Dynamik von Nukleosomen war das Thema Wolframs Diplomarbeit, die ich anfangs als Diskussionspartner begleiten durfte und die sich mehr und mehr in ein gemeinsames Projekt entwickelte.

Julia Morfill und Ferdinand Kühner haben sich getraut während der Experimente einen Theoretiker ins Labor zu lassen und geduldig meinen bisweilen abwegigen Ideen zugehört. Dafür möchte ich mich herzlich bedanken, denn mir hat es viel Spaß gemacht. In diesem Zusammenhang gebührt auch Prof. Hermann Gaub mein Dank, an dessen Lehrstuhl diese Experimente durchgeführt wurden.

Diese Doktorarbeit wurde in erster Linie am Lehrstuhl von Prof. Erwin Frey erstellt, dem ich an dieser Stelle einerseits für die vorhandene Infrastruktur, vor allem aber auch für fortwährende Unterstützung und hilfreiche Diskussionen danken möchte. Auch bei allen übrigen Mitglieder des Lehrstuhls, insbesondere bei Georg Fritz, mit dem ich über Jahre ein Büro teilte, möchte ich mich herzlich für die angenehme und anregende Arbeitsatmosphäre bedanken. Prof. Herbert Wagner bin ich für unzählige gute Ratschläge und interessante Diskussionen zu Dank verpflichtet. Für das Korrekturlesen der Arbeit danke ich meinem Mitbewohner Tim Liedl und Wolfram Möbius.

Finanzielle Unterstützung erhielt ich über das Emmy-Noether-Stipendium von Ulrich Gerland von der DFG sowie vom Internationalen Doktoranden Kolleg Nanobiotechnologie (IDK-NBT). Das IDK-NBT hat unter anderem in großzügiger Weise viele meiner Forschungs- und Fortbildungsreisen bezahlt. Während des vergangen Jahres habe ich mehrere Wochen am Center for Theoretical Biological Physics an der UCSD und am Kavli Institute for Theoretical Physics in Santa Barbara, Kalifornien verbracht. Ich hab während beider Aufenthalte viel gelernt und möchte beiden Zentren für die Gastfreundschaft danken. Ohne die finanzielle Unterstützung und die anregenden Reisen wäre meine Doktorarbeit sicherlich nicht so fruchtbar gewesen.

Curriculum vitae

Address

Arnold-Sommerfeld-Center for Theoretical Physics,
Ludwig-Maximilians-Universität München,
Theresienstr. 37
80333 München, Germany
Tel.: +49-89-21804596
Email: r.n@lmu.de

Personal data

Date of birth: 30th of August, 1979
Place of birth: Göttingen, Germany
Nationality: German

Education

- 02/2004–06/2007: PhD. student at the University of Munich.
Thesis advisor: Prof. Ulrich Gerland.
Thesis title: Dynamic aspects of DNA.
- 07/2006: Les Houches Summer School of Theoretical Physics on Complex Systems
- 11/2003: Diploma in physics. Grade: With Distinction.
Thesis advisor: Prof. Herbert Wagner.
Thesis title: Stochastic Geometry and Percolation.
- 10/2000–11/2003: Graduate studies in physics at the University of Munich.
- 07/2000: Prediploma in physics. Grade: Very Good.
- 10/1998–09/2000: Undergraduate studies in physics at the University of Göttingen.
- 7/1998: Abitur at the Freie Waldorfschule Göttingen. Grade 1.0.