
Physical aspects of chromatin constituents

DNA dynamics and nucleosome positioning

Wolfram Möbius



München 2010

Physical aspects of chromatin constituents

DNA dynamics and nucleosome positioning

Wolfram Möbius

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

vorgelegt von
Wolfram Möbius
aus Eisenach

München, den 5. März 2010

Erstgutachter: Prof. Dr. Ulrich Gerland
Zweitgutachter: Prof. Dr. Jens Michaelis
Tag der mündlichen Prüfung: 6. Mai 2010

Contents

Zusammenfassung	ix
Abstract	xi
1. DNA and nucleosomes	1
1.1. The structure of DNA	2
1.2. DNA and the worm-like chain model	4
1.3. DNA as the substrate for nucleosome formation	6
2. Dynamics of a knotted polymer under strong confinement	13
2.1. Knots	13
2.2. Polymers under confinement	15
2.3. Knots in strongly confined polymers	16
2.4. Outlook: two overlapping polymers under strong confinement	19
2.5. Conclusion	20
2.6. W. Möbius, E. Frey, and U. Gerland, <i>Nano Letters</i> , 8 , 4518 (2008)	21
3. Dynamics of nucleosomal DNA	27
3.1. Introduction to the site exposure mechanism	27
3.1.1. Thermodynamics	27
3.1.2. Interactions mediated by the nucleosome via site exposure	29
3.1.3. Relevance beyond enabling protein binding	30
3.2. Dynamics of site exposure	31
3.2.1. Experiments	31
3.2.2. Theoretical description	33
3.3. Outlook: site exposure in nucleosome arrays	37
3.4. Optimal flexibility for rotational barrier crossing	37
3.5. Conclusion	39
3.6. W. Möbius, R. A. Neher, and U. Gerland, <i>Phys. Rev. Lett.</i> , 97 , 208102 (2006)	41
3.7. R. A. Neher, W. M., E. Frey, U. Gerland, <i>Phys. Rev. Lett.</i> , 99 , 178101 (2007)	45
4. Nucleosome organization in yeast	49
4.1. Methods to determine nucleosome positions	50
4.2. Nucleosome organization close to start sites of genes	51
4.3. Nucleosome positioning determinants	53
4.3.1. Models predicting nucleosome positions based on the DNA sequence	54
4.3.2. In vitro – in vivo comparisons	56

4.4. Statistical positioning	57
4.4.1. Statistical positioning from a biological perspective	57
4.4.2. Physics of statistical positioning: the Tonks gas model	59
4.4.3. Quantitative test of statistical positioning	63
4.5. Outlook: the inverse approach	67
4.6. Conclusion	69
4.7. W. Möbius and U. Gerland, <i>PLoS Comput. Biol.</i> , 6 , e1000891 (2010)	71
A. Simulations and numerical methods	87
A.1. Simulations of polymer dynamics	87
A.2. More general Langevin dynamics	89
A.3. Recursion relations for partition functions	90
Bibliography	93
Danksagung	105
Curriculum Vitae	107

List of Figures

1.1. Structure of double-stranded DNA	3
1.2. Steps of DNA compaction in a eukaryotic cell	6
1.3. Microscopic structure of the nucleosome core particle	7
1.4. Geometries of models proposed to describe the 30-nm fiber	10
2.1. Table of prime knots with at most seven crossings	14
2.2. Regimes of confinement for long worm-like chains with self-exclusion	15
3.1. Illustration of the site exposure mechanism	28
3.2. Equilibrium constants for site exposure, Cooperativity and negative cooperativity mediated by the nucleosome	29
3.3. Sketch of dwell times in open and closed nucleosome configurations	34
4.1. Typical nucleosome organization close to transcription start sites	52
4.2. Particle density for a semi-infinite system of hard rods	57
4.3. One realization of a gas of hard rods placed between two fixed particles	59
4.4. Mapping between binding energy landscape and position dependent density	69
A.1. Bead-spring model for a polymer	88
A.2. Gas of hard rods on a lattice	90

Zusammenfassung

In Eukaryoten findet sich fast die gesamte DNA im Zellkern, jedoch nicht als Polymerknäuel, sondern in dicht und kontrolliert gepackter Form. Die DNA bildet einen Komplex mit verschiedenen Proteinen, der Chromatin genannt wird. Wichtiger Bestandteil von Chromatin sind Histonproteine, welche eine herausragende Rolle bei der Kompaktifizierung spielen. In regelmäßigen Abständen sind ungefähr 150 Basenpaare fast zweimal um einen Komplex aus acht Histonproteinen gewunden. Die Einheit aus DNA und Histonoktamer bezeichnet man als Nukleosom. Die DNA, welche die einzelnen Nukleosomen verbindet, ist viel kürzer als 150 Basenpaare, so dass sich der größte Teil der DNA innerhalb von Nukleosomen befindet und damit für die Bindung von Proteinen und molekularen Maschinen zunächst nicht zugänglich ist. Alle drei Teile dieser Doktorarbeit beschäftigen sich mit physikalischen Aspekten von Chromatin beziehungsweise dessen elementaren Bestandteilen: dem Biopolymer DNA, einzelnen Nukleosomen und schliesslich der Positionierung von vielen Nukleosomen auf der DNA.

Als vielleicht bekanntestes Beispiel eines semiflexiblen Polymers ist DNA seit langem Gegenstand theoretischer und experimenteller Untersuchungen. Seit einigen Jahren ist es möglich, einzelne, möglicherweise verknotete, DNA-Moleküle in sehr engen Kanälen experimentell zu beobachten. Motiviert durch solche Experimente untersuchen wir die Dynamik eines langen verknoteten semiflexiblen Polymeres in einem engen Kanal. Wir finden, dass aufgrund thermischer Fluktuationen sowohl die Größe des Knotens als auch seine Position entlang der DNA variieren. Die gekoppelte Dynamik von Größe und Position führt schlussendlich zum Auflösen des Knotens. Unsere Arbeit zeigt, dass die Größe des Knotens im Mittel mit der Polymerlänge vergleichbar ist, kurz bevor der Knoten verschwindet. Eine zusätzliche Kraft, angelegt an die Enden des Polymeres, verhindert hingegen ein Wachsen des Knotens, was sich in einem schnelleren Auflösen des Knotens niederschlägt.

Polymerdynamik steht auch im zweiten Teil dieser Arbeit im Vordergrund, jedoch motiviert von einer gänzlich anderen Fragestellung. Wie bereits angedeutet, ist ein großer Teil eukaryotischer DNA in Nukleosomen ‘verborgen’. Auf den ersten Blick steht somit der größte Teil der DNA-Sequenz nicht für das Binden regulatorischer Proteine zur Verfügung. Jedoch sind Nukleosomen keine statischen Gebilde. Thermische Fluktuationen sorgen dafür, dass sich Teile der DNA von dem Histonkomplex lösen können. Dieser Prozess wird *site exposure* oder *nucleosome breathing* genannt und ermöglicht das Binden von Proteinen auch an nukleosomale DNA. Mit Hilfe eines vergrößerten Modells untersuchen wir die Dynamik der DNA im Nukleosom, insbesondere das Ablösen vom und das Binden an den Histonkomplex. Beim Ablösen bzw. Binden vollführt die DNA eine Rotationsbewegung, wobei durch thermische Fluktuationen induziert eine Energiebarriere überquert wird. Mit Hilfe eines weiter vereinfachten Modells zeigen wir, dass das Überqueren der Barriere beschleunigt wird durch die Flexibilität der DNA auf Längenskalen deutlich kleiner als die Persistenzlänge. Damit können wir die Ergebnisse der Simulation unseres Nukleosommodells verstehen. Zukünftige Experimente werden unsere Vorhersagen bestätigen oder falsifizieren können. In einer Folgearbeit ersetzen wir innerhalb unseres vereinfachten Modells das semiflexible Polymer durch eine Scharnierstruktur. Auch hier finden wir, dass bereits kleine Flexibilitäten großen Ein-

fluss auf die Rate des Überquerens der Energiebarriere haben. Zudem wird die Rate bei einer endlichen Flexibilität maximal.

Der dritte und letzte Teil der Arbeit beschäftigt sich mit der Positionierung von Nukleosomen auf der DNA. Die Anordnung von Nukleosomen spielt eine wichtige Rolle bei der Regulation von Genen und wurde in den letzten Jahren intensiv experimentell untersucht, insbesondere für die Hefe *Saccharomyces cerevisiae*. Besonderes Augenmerk galt dabei den Regionen um Transkriptionsstartpunkte. Dort findet man (im Mittel) eine nukleosomfreie Region und davon ausgehend zu beiden Seiten Oszillationen der Nukleosomdichte. Wir untersuchen die Frage, ob sich die Dichteoszillationen allein mit Hilfe des Tonksgases (d.h., mit einem eindimensionalen System harter Stäbe) beschreiben lassen. Dichteoszillationen treten im Tonksgas in der Nähe einer Wand auf, allein aufgrund des Zusammenspiels von Entropie mit der endlichen Breite der Stäbe, ein Effekt, der in der Biologie als “statistische Positionierung” bekannt ist. Unsere systematische quantitative Untersuchung vorhandener experimenteller Daten für Hefe [77] zeigt, dass die Dichteoszillationen in der Tat mit dem Tonksgasmodell beschrieben werden können. Man muss jedoch unterschiedliche Randbedingungen auf beiden Seiten der nukleosomfreien Region in Betracht ziehen. Das Modell des Tonksgases beschreibt die Daten quantitativ, wenn das erste Nukleosom auf der einen Seite der nukleosomfreien Region (*downstream*) direkt positioniert ist, während dies für das erste Nukleosom auf der anderen Seite (*upstream*) nicht gilt.

Diese kumulative Doktorarbeit ist wie folgt strukturiert: In Kapitel 1 werden die Struktur der DNA, ihre Beschreibung als wurmartige Kette sowie Nukleosomen vorgestellt. Es soll als Grundlage für die folgenden Kapitel dienen, welche die einzelnen Veröffentlichungen sowie Einleitungen in die verschiedenen Fragestellungen beinhalten. Kapitel 2 beschäftigt sich mit der Dynamik eines Polymerknotens in eingeschränkter Geometrie. In Kapitel 3 wird die Dynamik nukleosomaler DNA und dessen theoretische Beschreibung diskutiert. In Kapitel 4 steht schlussendlich nicht mehr die Dynamik von DNA im Vordergrund, sondern die Anordnung der Nukleosomen auf der DNA sowie das Modell des Tonksgases.

Abstract

In eukaryotes, most of the DNA is found inside the cell nucleus. There, the DNA does not form a random coil but is densely packed in a highly controlled way. DNA, together with different proteins, forms a complex called chromatin. Histone proteins are an important component of chromatin and play an important role in DNA compaction. About 150 base pairs of DNA are wrapped almost twice around eight histone proteins. DNA and histone proteins together comprise the smallest structural subunit of chromatin, a nucleosome. The linker DNA between individual nucleosomes is much shorter than 150 base pairs. Consequently, most of the DNA is buried inside nucleosomes and is, at least at first sight, not accessible for binding proteins and molecular machines. All three parts of this thesis consider physical aspects of chromatin or its elementary constituents: the biopolymer DNA, single nucleosomes, and positioning of nucleosomes on the DNA.

DNA is often considered the archetype of a semi-flexible polymer and has been intensively studied theoretically and experimentally for a long time. Technological progress has recently made it possible to study single, possibly knotted, DNA molecules confined to very narrow channels. Such experiments motivated us to study the dynamics of a long, knotted semi-flexible polymer confined to a narrow channel. We find that thermal fluctuations lead to variations in both the size of the knot and its position along the polymer contour. The coupled dynamics between size and position eventually leads to disappearance of the knot. Our study shows that the size of the knot when disappearing is on average comparable to the length of the polymer. However, additionally applying a tension at the ends of the polymer prevents knot growth and speeds up spontaneous unknotting.

Polymer dynamics also takes center stage in the second part of this thesis, but motivated by a totally different scientific question. As stated above, most of the eukaryotic DNA is buried inside nucleosomes, which makes it mostly inaccessible for binding of regulatory proteins. However, nucleosomes are not static entities, and thermal fluctuations lead to transient unwrapping of DNA from the histone complex, a process called *site exposure* or *nucleosome breathing*. The exposure of DNA allows for binding of regulatory proteins to DNA in nucleosomes. Here, we theoretically study the dynamics of site exposure using a coarse-grained nucleosome model. When wrapping and unwrapping from the histone core, the DNA rotationally crosses a free energy barrier. Using a more simplified toy model we show that flexibility on length scales much shorter than the persistence length accelerates barrier crossing, an effect which explains the results of our simulations. In a subsequent study, we replace the semi-flexible polymer by a hinged lever. This modified toy model also displays flexibility-assisted barrier crossing, but in addition the crossing rate becomes maximal at intermediate flexibility.

The third and last part of this thesis addresses the positioning of nucleosomes on the DNA. The organization of nucleosomes is known to play an important role in gene regulation and has extensively been addressed experimentally within the last few years, especially for the yeast *Saccharomyces cerevisiae*. Special attention has been paid to regions around transcription start sites. On average, one finds a nucleosome-free region there, flanked by oscillations in

nucleosome density. We address the question of whether the oscillations can be described solely within the Tonks gas model (a one-dimensional system of hard rods). Close to a boundary, density oscillations arise in the Tonks gas purely from the interplay between entropy and the finite width of the rods, an effect known as “statistical positioning” in biology. Our systematic and quantitative reanalysis of experimental data for yeast [77] shows that the density oscillations are indeed captured by the Tonks gas model. However, one must consider different boundary conditions on both sides of the nucleosome-free region. The Tonks gas model describes the data quantitatively if the first nucleosome on one side (downstream) is directly positioned while this does not apply to the first nucleosome on the other side (upstream).

This cumulative thesis is structured as follows: In Chapter 1, the biochemical structure of DNA, its description in terms of a worm-like chain polymer, and nucleosomes are introduced. It is meant to provide background information for subsequent chapters. Those include the publications and introductions into the individual scientific questions. Chapter 2 addresses the knot dynamics in strongly confined polymers. In Chapter 3, the site exposure mechanism and its theoretical description are discussed. In Chapter 4, nucleosome organization and the Tonks gas model take center stage.

1. DNA and nucleosomes

DNA serves as the storage medium for hereditary information in all organisms on earth, from bacteria to fungi to plants and animals. The building blocks of the DNA polymer are four different nucleotides, and it is the sequence of the monomers along the DNA molecule which carries the hereditary information. DNA encodes for proteins, RNA molecules, and concomitantly for regulatory information specifying under which conditions the information is read out. How the long DNA polymer is stored inside the cell differs between different organisms and in fact classifies them. In eukaryotes, the vast majority of DNA is stored inside the cell nucleus, a compartment inside the cell that is bounded by a double layer membrane. The name 'eukaryote' is of greek origin and refers to this fact. Eukaryotes may be categorized into fungi, plants, and animals. Prokaryotes (bacteria and archaea), by contrast, do not have a cell nucleus.

In eukaryotes, the nuclear DNA is divided between individual chromosomes and is tightly packed. As an illustrative example: Human DNA is about 2 m long in total (though only 2 nm wide) and packed into a nucleus with a diameter of about 6 μm [3]. However, the DNA is not forced into the cell nucleus as a set of long randomly coiled polymers, but it is associated with different proteins, classified into histones and non-histone chromosomal proteins. The complex of DNA and proteins is called chromatin whose name originates from the greek word for color because it can be stained. Histone proteins are present in a large quantity. Their total molecular mass is about equal to the mass of DNA the histone proteins are associated with. DNA is wrapped around a complex of eight histone proteins this way forming nucleosomes, the smallest and most basic subunit of eukaryotic chromatin.

Although the DNA is densely packed, it remains accessible for enzymatic machines carrying out transcription (the read out of the genetic information), DNA repair, and DNA replication in a highly controlled way. The interplay between compaction and accessibility as well as the regulatory role of chromatin are intriguing questions of molecular biology about which much is already known; however, much also remains to be explored and understood [3].

Though the composition and function of eukaryotic chromatin is a question of originally and primarily biological interest, chromatin and its constituents are also fascinating from a physics perspective. This, on the one hand, means that chromatin constituents, most importantly the DNA molecule as the archetype of a semi-flexible polymer, are subject to intense physical research in its own regard and not necessarily within the biological context. On the other hand, physical approaches help to better understand chromatin, its constituents and their functions. Within this cumulative thesis, I address physical aspects of chromatin and its constituents adopting both perspectives.

In the first part of this thesis, the dynamics of a long semi-flexible polymer (e.g., DNA), being both knotted and confined to a very narrow channel, is investigated. Simulations of the polymer and a coarse-grained stochastic model are employed to characterize the dynamics of the knot and its disassembly by thermal fluctuations.

The dynamic behavior of DNA also is subject of the second part of the thesis. It is known that DNA may transiently unwrap from the histone complex this way allowing protein

binding to DNA stretches that are otherwise buried inside nucleosomes. Here, the mechanism of DNA unwrapping is theoretically studied using a coarse-grained nucleosome model. The process which sets the rates for wrapping and unwrapping, stochastic multidimensional barrier crossing, is investigated using a simplified toy model. At the end of this part, another related toy model is presented, which could be understood within a similar framework.

The third and last part of this thesis concerns the distribution of nucleosomes on the DNA, especially around the start sites of transcription. This part of the thesis differs from the remaining work in different respects. First, recent experimental data [77] are reanalyzed and quantitatively compared to an already well-characterized physical model. Second, an equilibrium observable – the density of nucleosomes – is investigated instead of dynamic properties.

The work is presented in Chapters 2–4, which also include the publications [84–86, 90]. Each chapter gives a general introduction to the scientific questions addressed, describes related experimental and theoretical work, and outlines the different publications. The purpose of this first chapter is to introduce the structure of DNA, its description in terms of the worm-like chain polymer model, and its packaging into nucleosomes. It is meant to provide background material for the main part of the thesis. Since the thesis touches very different topics and work from different scientific communities including statistical and soft condensed matter physics, molecular biology, and bioinformatics, it would be beyond the scope of the thesis to treat all topics on an equal footing while aiming to give extended background information for all topics. I therefore decided to be brief on ‘classical’ statistical and single polymer physics, here and in the remainder of this thesis, and in return allow more space for the discussion of nucleosomes and especially the fast growing field of nucleosome positioning.

1.1. The structure of DNA

Deoxyribonucleic acid (short DNA) is composed of two polynucleotide chains consisting of individual monomers. Each monomer, called nucleotide, consists of a sugar (the deoxyribose), a phosphate group, and one of the four bases adenine (A), cytosine (C), guanine (G), or thymine (T). The phosphate group of each nucleotide is attached to the 5'C atom of the sugar and covalently binds to the hydroxyl group at the 3'C atom of the neighbored nucleotide's sugar. This way, a phosphate-sugar-phosphate-sugar-... backbone forms. The bases are attached to this backbone at the individual sugar molecules, and any order of the bases is possible. Importantly, since every nucleotide is different on both ends within the backbone, a polarity is introduced into the chain. Hence, single stranded DNA has a so called 5' end and a 3' end.

Two of these single stranded DNAs form a double stranded DNA molecule (dsDNA) as depicted in Fig. 1.1. Thereby, the two backbones are located at the outside. The bases face each other in the inside interacting via hydrogen bonds and forming base pairs. However, this requires that a thymine is always bound to an adenine whereby two hydrogen bonds are formed, and a guanine is always bound to a cytosine by formation of three bonds. In short, only the base pairs TA and GC are possible, or equivalently stated, the two strands are complementary to each other. Moreover, the two strands run in opposite directions, they are antiparallel. One end of the dsDNA thus consists of the 5' end of one strand and the 3' end of the other. This has an interesting topological consequence: one cannot create a Möbius band by half-twisting a long dsDNA molecule and joining the two ends since this would

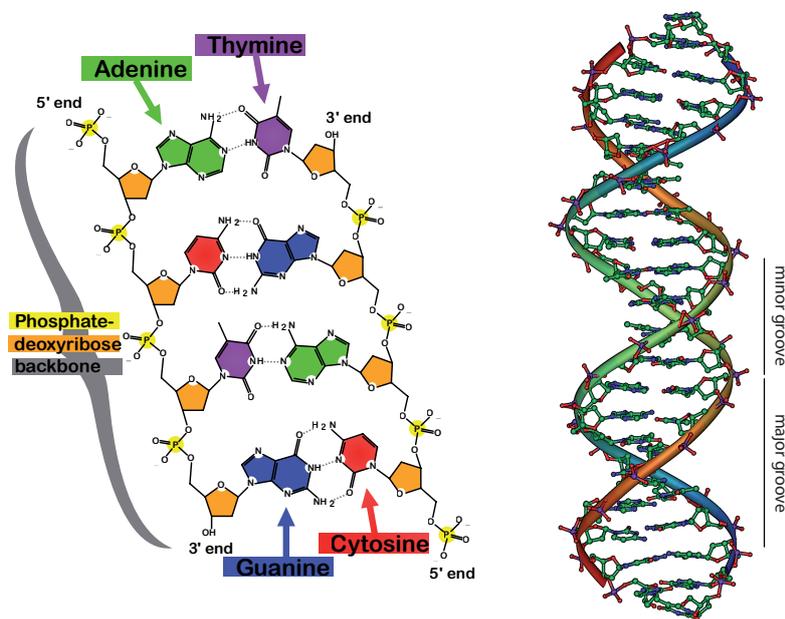


Figure 1.1.: Structure of double-stranded DNA. **(Left)** Two antiparallel strands of single-stranded DNA (a polymer with a phosphate-deoxyribose backbone and bases attached to it) are bound to each other if their sequence of bases is complementary, i.e., adenine binds exclusively to thymine, as does guanine to cytosine. Two and three hydrogen bonds are formed within the base pairs, respectively. **(Right)** In three dimensions, dsDNA forms a double-helix with the two backbone strands located on the outside and the base pairs in between, stacked on top of each other. Since the chemical bonds between bases and the backbone sugar are not located diametrically opposite to each other (i.e., the base pair is ‘wider’ on one side than on the other), the double helix displays a major and a minor groove. Images adapted from Wikipedia (January 2010).

require connecting a 5’ end to a 5’ end and a 3’ end to a 3’ end, respectively [1]. Focusing on the biological role of dsDNA, the sequence of base pairs carries the genetic information. It specifies where proteins (preferentially) bind to DNA and it encodes for amino acids and eventually for proteins. When reading out the genetic information, one of the strands is read from the 5’ to the 3’ end. For the replication of DNA, both strands serve as a template for a new dsDNA molecule, respectively.

In three dimension, dsDNA takes on the form of the celebrated double helix originally proposed by James D. Watson and Francis H. C. Crick in 1953 [148], where the two backbones run on the outside and the base pairs are found in the center of the helix. Today, different variants of the double helix are known. In the cell, the B-form is generally found, depicted in Fig. 1.1: a left-handed double-helix with a diameter of 2.4 nm, a rise of 0.34 nm per base pair, and 10.4 base pairs per helical turn [12].¹ The formation of the double helix is achieved by staggering the base pairs, which are similar in shape, on top of each other. The hydrogen bonds within the base pairs contribute little to the formation of the double helical structure,

¹The reported parameters for the B-form of DNA vary, the original Watson-and-Crick structure of DNA, for example, has 10 base pairs per helical turn.

more important are van der Waals interactions between neighboring base pairs as well as hydrophobic effects. Since the chemical bonds between the bases of one base pair and the sugar molecules are not located diametrically opposite to each other, the double helix displays a minor and a major groove. dsDNA does not form a perfect double helix, however, its precise shape and elastic properties (for us most importantly, its bending properties) depend on the nucleotides incorporated. Since the staggering of base pairs plays an important role in the formation of the double helix, the smallest sequence pattern considered for DNA bending properties are dinucleotides as we will see below. Last but not least, as the phosphate groups in the backbone carry negative charge, DNA is a negatively charged polyelectrolyte.

1.2. DNA and the worm-like chain model

On very small scales, i.e., on the scale of individual monomers, double-stranded DNA is a stiff molecule. On larger scales, however, DNA is bendable and thermal fluctuations lead to deviations from a straight contour, even in the absence of external forces. For these reasons DNA is considered a *semi-flexible polymer*. The bending properties of DNA are usually described by the *worm-like chain* model [57, 116], which briefly is outlined below.

Mathematically, the worm-like chain (WLC) is represented by a continuous and differentiable space curve $\mathbf{r}(s)$, where $0 \leq s \leq L$, and L is the contour length, i.e., the length of the backbone. The local tangent vector $\mathbf{t}(s) = \partial\mathbf{r}/\partial s$ fulfills an inextensibility constraint, $|\mathbf{t}(s)| = 1$. The energy cost of bending is that of a thin elastic rod with bending stiffness κ ,

$$H = \frac{\kappa}{2} \int_0^L \left(\frac{\partial^2 \mathbf{r}(s)}{\partial s^2} \right)^2 ds. \quad (1.1)$$

The hall-mark of the worm-like chain is the exponential decay of the tangent-tangent correlation function,

$$\langle \mathbf{t}(s) \cdot \mathbf{t}(s') \rangle = \exp\left(-\frac{|s - s'|}{l_p}\right), \quad (1.2)$$

where $l_p = \kappa/k_B T$ is called the persistence length.² The persistence length is the length-scale on which correlations between the tangent vectors decay.

If $L \ll l_p$, the WLC's configuration is essentially straight, and the contour may be described by small deviations \mathbf{r}_\perp perpendicular to the global orientation. In this limit, the deterministic part of the equation of motion may be written as

$$\zeta_\perp \frac{\partial \mathbf{r}_\perp(s, t)}{\partial t} = -k_B T l_p \frac{\partial^4 \mathbf{r}_\perp(s, t)}{\partial s^4}, \quad (1.3)$$

where ζ_\perp is the friction coefficient per length for motion perpendicular to the overall orientation (see, e.g., [58, 152]). From this weakly-bending rod approximation one may directly read off the longest relaxation time:

$$\tau_r \sim \frac{\zeta_\perp L^4}{k_B T l_p}. \quad (1.4)$$

²This holds for three-dimensional embedding space. In two dimensions: $\kappa = k_B T l_p / 2$. Alternatively, one can use $l_p = \kappa / k_B T$ and redefine the persistence length in the tangent-tangent correlation.

This means that a small perturbation does not immediately propagate to the end of the polymer. Conversely, the polymer does not ‘feel’ its end on very short time scales. This fact will be exploited in a scaling argument in Sec. 3.2.2. Though at small length-scales the WLC may be considered essentially stiff, external constraints such as confinement or a potential barrier can give rise to the emergence of length scales which are smaller than the persistence length and which characterize the WLC’s shape (Odijk length, Sec. 2.2) or the dynamics (characteristic length when rotationally crossing a barrier, Sec. 3.2.2).

If, conversely, the length of the WLC is much larger than the persistence length ($L \gg l_p$), different parts of the chain become statistically independent. Then, the WLC can be described as a flexible chain (a succession of flexibly joined stiff segments) on large length scales. In this regime, the chain forms a coil with a typical size of $R \sim L^\nu$ with $\nu = 1/2$. Self-avoidance (the fact that different parts of the chain cannot occupy the same region in space) leads to an effective swelling of the coil and $\nu \approx 0.588$ (e.g., [114]). The limit of a flexible chain, however, has to be used with caution. For example, when confined to very narrow channels, the WLC may not be described as a flexible chain any longer and self-exclusion ceases to be important, even if $L \gg l_p$.³

As already mentioned, DNA is commonly modeled as a WLC. A direct and single-molecule based method to test the bending properties of DNA, i.e., the applicability of Eq. (1.1), are stretching experiments which became possible almost 20 years ago. Mechanically pulling on long individual DNA molecules resulted in a relation between force and extension which is well-described by the WLC model [73, 130] (for the highest salt concentration used in the measurements). The fit of the data to the WLC model revealed a persistence length of $l_p \approx 50$ nm (≈ 150 base pairs), a value often referred to for DNA under physiological conditions. More recent experiments tested whether the description of DNA as a WLC with a persistence length of 50 nm is valid on small length scales and found DNA to be more flexible than expected (e.g., [153, 154]).

Though DNA is often modeled just as a WLC, it is important to keep in mind that DNA is a charged polymer and thus a polyelectrolyte. When salt ions are abundant in the solution, as is the case in biological systems, the interactions between different charges are not long-ranged, but effectively screened, with the screening length depending on the salt concentration. Still, the charge on the polymer backbone cannot be neglected and affects the shape of the polyelectrolyte [91]. Importantly, the charges lead to an effective increase of the persistence length and of the polymer diameter. Both effects, of course, depend on the bare persistence length, the charge density on the backbone, and ambient salt conditions. As an illustrative example: The dependence of the effective DNA diameter on salt conditions has been studied in an innovative experiment [115]. Rybenkov et al. compared the experimentally determined knotting probability of DNA molecules to the knotting probability obtained in computer simulations for polymers with varying width. For a solution with the physiological concentration of NaCl, the authors estimated the effective diameter to be 5 nm, which is considerably larger than the bare DNA diameter of about 2 nm. Concluding this brief discussion of charge effects, it is often sufficient to include the effects of the charged DNA backbone by choosing an appropriate persistence length and effective diameter. It can, however, be important to explicitly include the charges at least on a coarse-grained level by screened electrostatic interactions, depending on the system under consideration and the salt conditions used.

³This is the reason why I here use the expression ‘semi-flexible polymer’ even for long polymers and not only if $L < l_p$ as is sometimes done.

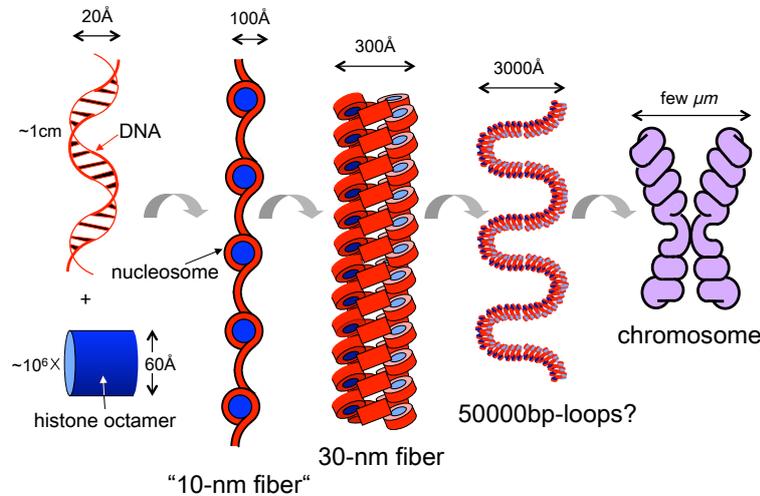


Figure 1.2.: Levels of DNA compaction in a eukaryotic cell. (From left to right) Double stranded DNA forms, together with histone octamers, nucleosomes which are arranged like beads on a string. On larger scales, DNA is thought to be packaged into the 30-nm fiber and other higher-order structures. The most compact form is the chromosome as it appears during cell division. Image: Courtesy of Helmut Schiessel.

1.3. DNA as the substrate for nucleosome formation

As mentioned in the introduction, nuclear DNA in eukaryotic cells is not found as long extended biopolymers, but is tightly packaged with the help of various proteins and in this way forms chromosomes. The packaging occurs on different length scales as illustrated in Fig. 1.2. On the largest scale, which is present only during cell division, the packaging is manifested by the existence of very compact structures that are colloquially known as chromosomes. Staining of these compact structures yields reproducible patterns of bands, indicating a controlled mechanism of packaging rather than uncontrolled DNA compaction. But also during interphase, DNA is compacted in a controlled way. A recent study revealed which of the three billion base pairs in the human genome are found close to each other within the nucleus and determined a fascinating non-random structure [66], deepening previous knowledge about DNA organization within the nucleus.

The nucleosome As noted earlier, DNA compaction on the smallest scale is achieved by histone proteins, which – beside nonhistone chromosomal proteins and the DNA itself – constitute the chromatin [3]. As early as 1974, in a visionary paper, Roger D. Kornberg reviewed the knowledge of that time and proposed the packaging to be achieved by a “repeating unit of histones and DNA” with a repeat length of about 200 base pairs (bp) [53]. It took almost a quarter of a century until a high resolution structure of this packaging complex, referred to as a *nucleosome* or more precisely as a *nucleosome core particle*, could be resolved, first at 2.8 Å [72] and some years later at 1.9 Å resolution [23, 108]. The structure of a nucleosome core particle is displayed in Fig. 1.3. 146 bp of DNA are tightly wrapped around a complex of histones in about 1 3/4 turns on a left-handed superhelix.⁴ The histone complex consists

⁴The two most well-known structures of the nucleosome core particle include 146 bp [72] and 147 bp of DNA [23], respectively, which is why both numbers are commonly reported.

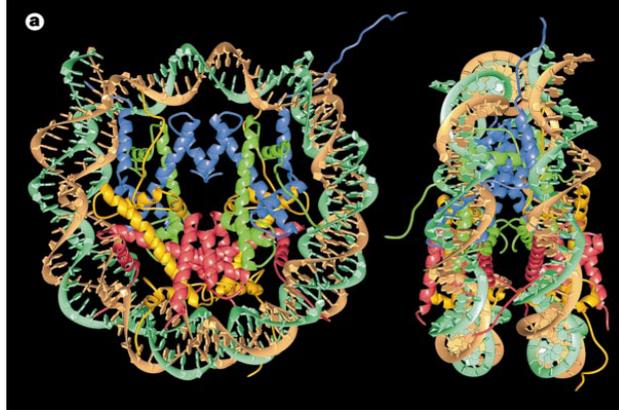


Figure 1.3.: Microscopic structure of the nucleosome core particle viewed along and vertical to the superhelical axis. The backbones of the two DNA strands and the main chains of the eight core histone proteins are displayed. The axis of pseudo-two-fold symmetry, the dyad axis, runs from top to bottom in both views. Reprinted with permission from Ref. [72].

of two of each histone proteins H2A, H2B, H3, and H4. The histones H2A and H2B form heterodimers as do the H3 and H4 histones. The H3-H4 dimers aggregate to a tetramer that forms the template to which the two H2A-H2B dimers bind. Each histone consists of a core and a long amino-acid tail that sticks out from the histone core or between the two DNA turns. The tails are subject to covalent modifications and contribute to the interaction between nucleosomes. The overall structure of the nucleosome is symmetric under a rotation of 180° around the so called *dyad axis* (pseudo-two-fold symmetry of the nucleosome). The dyad axis runs through the nucleosome core particle and crosses the center of the DNA wrapped up. This location therefore often is referred to as the *dyad*.

Different nucleosome core particles are joined by linker DNA, thereby forming a structure similar to beads on a string as displayed in Fig. 1.2. An additional histone, called H1, can complete a nucleosome at the entry and exit site of the DNA. This linker histone is much less conserved than the core histones and is involved in forming higher order structures.

Fig. 1.3 implies that the DNA must be very strongly bent within a nucleosome. Indeed, the bending energy required to bring DNA into the observed shape is huge [122]: Treating the DNA as a worm-like chain as described above, the bending energy is estimated to be $k_B T l_p L / 2R^2$, where l_p stands for the DNA's persistence length ($\approx 500 \text{ \AA}$), L for the length of DNA bent within the nucleosome ($\approx 431 \text{ \AA}$), and R for the radius of curvature given by the histone octamer ($\approx 43 \text{ \AA}$). Inserting the numbers yields an estimated elastic bending energy of about $58 k_B T$ which corresponds to about $4 k_B T$ for each 10 bp. Note that this number has to be treated as a rough estimate since, for example, it is not clear whether DNA may be regarded as a worm-like chain at those strong curvatures. Despite this huge bending energy, nucleosomes readily form in nature and the interaction between histones and DNA accounts for an additional net adsorption energy which is estimated to be $1.5 - 2 k_B T$ per 10 bp [122] (see Sec. 3.1.1).

This large interaction energy is due in part to the electrostatic interaction between the positively charged histone complex and the negatively charged DNA. This is demonstrated by the fact that the stability of the nucleosome is salt-dependent ([122] and therein). This salt dependence of nucleosome stability is exploited to assemble nucleosomes from histones

and DNA *in vitro*. The electrostatic interaction between DNA and the histone complex and additionally between the different parts of the DNA has motivated several theoretical studies. The wrapping of a negatively charged polyelectrolyte around a positively charged sphere has been applied as a toy model for the histone-DNA complex [14, 122].

However, the picture of DNA smoothly wrapped around a sphere or cylinder stabilized by electrostatic attraction is simplified too much as the reported microscopic nucleosome structure reveals [23, 72, 108]. First of all, the interactions between DNA and the histone octamer are mainly localized at the 14 points where the minor groove of the DNA faces towards the histone complex and consists of different types of more specific interactions such as hydrogen bonds and salt linkages. Second, the DNA structure deviates from a smooth helical path around the histone complex, the DNA is neither uniformly bent nor uniformly twisted.⁵ In our coarse-grained nucleosome model, which we present in Sec. 3.2.2, we take the localized interactions into account. However, we disregard deviations from the superhelical path.

Sequence dependence of nucleosome binding affinity and nucleosome positioning Despite the strong bending of DNA around the histone octamer, nucleosomes form on essentially all DNA sequences. The binding affinity, however, depends on the sequence which is incorporated. Some sequences bind about 1000-fold better than others [134]. Although this effect is small compared to the total net binding affinity, it is of great importance and subject to intensive research. Sequence influences the unwrapping of DNA from nucleosomes (see Sec. 3.1.1) and the organization of nucleosomes on the genome (see Sec. 4.3).

DNA sequence can lead to an increased binding affinity by different mechanisms. First, the sequence can just make more or stronger bonds with the histone complex. Second, the DNA sequence can (by its intrinsic structure) support the shape enforced by the histone complex, hence decreasing the bending and twisting energy. This can be achieved by DNA being locally more bendable or pre-bent as well as by being more twistable or pre-twisted. It is believed that the sequence dependence of nucleosome binding affinity is primarily determined by DNA mechanics and that the first mechanism plays a minor role [151].

Dinucleotides are the shortest sequence elements which can capture local DNA properties like bendability and bendedness. Therefore, special emphasis has been put in determining how dinucleotides are distributed in DNA sequences incorporated in nucleosomes. As early as 1986, Satchwell *et al.* determined dinucleotide frequencies in isolated natural nucleosomes [120]. They found, amongst others, a modest enrichment of AA=TT dinucleotides where the minor groove of the DNA faces the histone complex. Simultaneously, the authors observed an enrichment of GC, but with a phase-shift of 180°. This is consistent with the rule of thumb that stretches comprised of A and T tend to bend into the minor groove, whereas stretches consisting of G and C tend to bend into the major groove.⁶ In this way, an increased binding affinity may be explained by local bending properties which add up to a global curvature and facilitate incorporation of DNA into the nucleosome. Due to the helical structure of the DNA, this is achieved by a pattern with about 10 bp periodicity rather than by a homogeneous sequence. In consequence, a high affinity sequence can incorporate a nucleosome relatively well if shifted by 10 bp but significantly poorer if shifted by only 5 bp.

⁵On average, DNA in the nucleosome is slightly underwound. The overall twist is about 10.2 bp per turn compared to 10.5 for free DNA.

⁶For a more thorough discussion see the comprehensive though not very recent review by Widom [151].

Though dinucleotides represent the shortest sequence elements influencing binding affinity, longer sequence elements are also important. Arguably the most prominent representatives of longer sequence elements are poly(dA:dT) strands (A on one strand, T on the other). In their early study, Satchwell *et al.* found these elements to preferentially occupy the ends and to be excluded from the center of the DNA wrapped around the histone octamers. Today, poly(dA:dT) stretches are known to be depleted of nucleosomes *in vivo* and are thought to serve the biological function of keeping certain genomic regions free of nucleosomes. For a recent review see Ref. [126].

The early work by Satchwell *et al.* has inspired many studies trying to determine the sequence rules underlying nucleosome binding affinity both *in vivo* and *in vitro* [151]. Today, such sequence rules are made quantitative in models trying to predict nucleosome binding affinities *de novo* for arbitrary sequences. For an introduction see Sec. 4.3.1.

Evidently, nucleosome binding affinity is closely linked to positioning of nucleosomes on stretches of DNA longer than 147 bp [69]. One distinguishes between *translational positioning* and *rotational positioning*. Translational positioning, by convention, refers to preference of certain locations on the DNA over neighboring positions. Rotational positioning is seen as a special case of translational positioning where multiple positions, about 10 bp apart from each other, are preferred over other possible locations. In this case, disregarding the precise location of the nucleosome, the same parts of the nucleosomal DNA are exposed to solution and are oriented towards the histone complex, respectively. Rotational positioning may be induced by a 10 bp periodic arrangement of dinucleotides as explained above.

Nucleosome positioning sequences (i.e., sequences with high binding affinity) have become work-horses for *in vitro* experiments on nucleosomes, some of which are discussed in Chapter 3. Most prominent positioning sequences are the natural sequence coding for the 5S rRNA and the artificial 601 positioning sequence derived in *in vitro* selection experiments [70]. Importantly, positioning of nucleosomes plays a vital role in gene regulation, see Chapter 4.

Histone variants, histone modifications, and chromatin remodelers Nucleosomes are not static entities, placed on the genome after replication and remaining unmodified until the next cell division occurs. Rather, they are subject to regulated changes concerning composition, modification and position. Though these ‘dynamic properties’ are not subject of this thesis, a very brief outline helps to understand the biological context.

First, not all nucleosomes are composed of the four canonical core histones H2A, H2B, H3, and H4. There are different histone proteins, among them H3.3 (a variant of H3) and H2A.Z (in yeast often termed Htz1, a variant of H2) [118]. The variant Htz1 is enriched at promoter regions [105] (see Sec. 4.2).

The histone tails sticking out from the histone complex allow for a second type of variability. They are subject to a number of covalent modifications, such as acetylation, methylation, and phosphorylation. These covalent modifications are added and removed by enzymes, nucleosome-modifying complexes which are regulated by DNA-binding proteins. Nucleosome modifications are known to influence the structure of chromatin and play an important role in gene regulation. Since nucleosomes are randomly distributed on the daughter strands upon replication, such modifications may be inherited by daughter cells. The wealth of possible histone modifications with in principle every nucleosome carrying different modifications provoked the question of a *histone code*. Today it is known, however, that modifications often occur in domains and that many modifications occur together, challenging the idea of such a code. For a short review on patterns of histone modifications see Ref. [103].

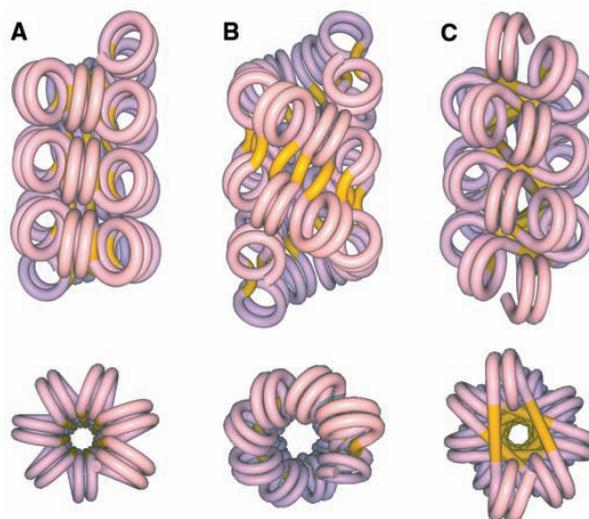


Figure 1.4.: Geometries of models proposed to describe the 30-nm fiber. DNA wrapped up in nucleosomes is displayed in red to blue, linker DNA in yellow. **(A)** One-start solenoid model: nucleosomes neighbored on the DNA follow each other on a helical path creating a one-start solenoid. The linker DNA is bent. **(B)** Helical ribbon model: nucleosomes form a two-start helix where the linker is essentially straight and connects the two starts parallel to the fiber axis. **(C)** Crossed-linker model: nucleosomes form a two-start helix where the straight linker DNA connects nucleosomes on opposite sides of the fiber. Reprinted with permission from Ref. [28].

Last but not least, chromatin remodeling should be mentioned as a type of variability. Remodelers assist in the assembly of chromatin. They also reposition and eject nucleosomes and lead to localized unwrapping of DNA. This way they regulate access to DNA for other DNA-binding proteins. Moreover, remodelers also change nucleosome composition [21]. Remodelers are divided into different families and groups based on their composition and function. One example are SWR1 remodelers; they substitute histones within a nucleosome, replacing H2A-H2B dimers with variant H2A.Z-H2B dimers. The action of chromatin remodelers is regulated, not all remodeling processes occur in the cell at the same time. In order to perform their function, all remodelers require ATP and seem to use a common mechanism associated with creation of a DNA loop in the nucleosome. It is an interesting future task to better understand how different remodelers can use a common basic mechanism but perform so different functions [21].

Higher-order structures While much is known about single nucleosomes, no common picture has evolved for the structure on the next larger compaction level of chromatin. This structure is called the 30-nm fiber due to the diameter of chromatin structure that can be observed after extraction from the nuclei. For decades, there has been an ongoing debate about the nature and geometry of this structure, which is reflected in a myriad of models and many experimental studies. Traditionally, there are two competing classes of models: the one-start and two-start structures ([28] and references therein). In the one-start class, the solenoid as displayed in Fig. 1.4A is the prominent representative. Nucleosomes which neighbor each other on the DNA sequence face each other within the *solenoid structure*. As one can easily

imagine, this structure is not very sensitive to changes in linker length, the excess DNA is just bent within the fiber. This is different in two-start geometries as displayed in Fig. 1.4B and Fig. 1.4C. The DNA goes back and forth, it ‘zigzags’, either vertically (*helical ribbon model*) or horizontally (*crossed-linker model*) with respect to the fiber axis; variations in linker length should affect the dimensions of the structure. There is well-known experimental evidence for both, one start structures [110] and two-start geometries [28, 121]. But maybe, such a distinction between two geometries is simplified too much as illustrated by the fact that a recent study combining experiments and modeling provides evidence that both straight and bent linkers may be present in a heteromorphic fiber [37].

Since the 30-nm fiber is of only minor importance for the work presented here, I refer at this point to reviews reporting on models and the physics of the 30-nm fiber [30] and on force spectroscopy experiments on chromatin [17]. Last but not least, biochemical aspects of fiber formation and the question whether the fiber exists *in vivo* are discussed in Ref. [139].

While already the nature of the 30-nm fiber is under debate, even less is known about higher-order levels of chromatin organization. It is believed that in interphase the 30-nm fiber forms looped domains in some regions and is subject to additional levels of packaging in other regions [3]. On the largest scale, as indicated in the beginning of this section, chromatin forms a non-random structure in the nucleus during interphase and a very compact structure during cell division. See Ref. [30] for a physical perspective of the structures beyond the scale of the 30-nm fiber.

2. Dynamics of a knotted polymer under strong confinement

Recent advances in nanotechnology offer enormous prospects for analyzing, sorting, and manipulating single molecules [9, 22]. One specific example is the confinement of individual biopolymers into channels with cross-sections on the micrometer or nanometer scale. In bulk, long semi-flexible polymers such as DNA coil up for entropic reasons. When confined to channels much smaller than their overall globular shape, the polymers elongate along the channel axis. This change of shape offers a big technological advantage. The distances between different parts of the molecule observed under the microscope may be mapped to chemical distances along the polymer contour. When applied to DNA, this method allows one to study sequence-dependent properties such as the binding of proteins to single DNA molecules.

The action of restriction enzymes is an illustrative example. These kind of enzymes cut DNA at specific sequences. Traditionally, the lengths of the resulting DNA fragments are measured using gel electrophoresis, a classical bulk experiment. In a proof-of-principle experiment, Riehn *et al.* measured the lengths of the resulting DNA fragments using nanochannels and only a few molecules [109]. The restriction enzymes cut the DNA within the nanochannels. In consequence, the individual DNA fragments diffused along the channel and their lengths could be determined under the microscope.

The mapping between spatial position along the confining channel and chemical position, however, may be impeded by anomalous polymer shapes within the channel. This becomes particularly important when the polymer is so strongly confined that its elongation along the channel becomes close to its contour length, i.e., when the polymer is almost completely stretched out. While undulations and overlays between different parts may relax rapidly, topological defects like knots pose a more severe obstacle since they are expected to have much longer relaxation times. Motivated by this consideration, we theoretically studied the dynamics of a knot in a polymer confined to a nanochannel [84].

I will start out this chapter with a brief introduction into knots and the physics of confined polymers. Subsequently, the main ideas and results of our work will be outlined and set into the context of a number of related studies concerning knots in elongated polymers. At the end of the chapter, I will give a short outlook on an ongoing project concerning the physics of two overlapping strongly confined polymers.

2.1. Knots

Knots are familiar in everyday life, as a means of fastening objects or of connecting two ropes, as a metaphor for hard problems, and often as an annoyance if spontaneously formed in one's shoelace. In mathematics, a knot is a closed space curve in three dimensions which does not intersect itself. Two different space curves represent the same knot if they may be transformed into one another by continuous deformations without passing the curve through itself. In this formal sense, the simplest knot is the *unknot* or *trivial knot*, simply a circle. The most simple

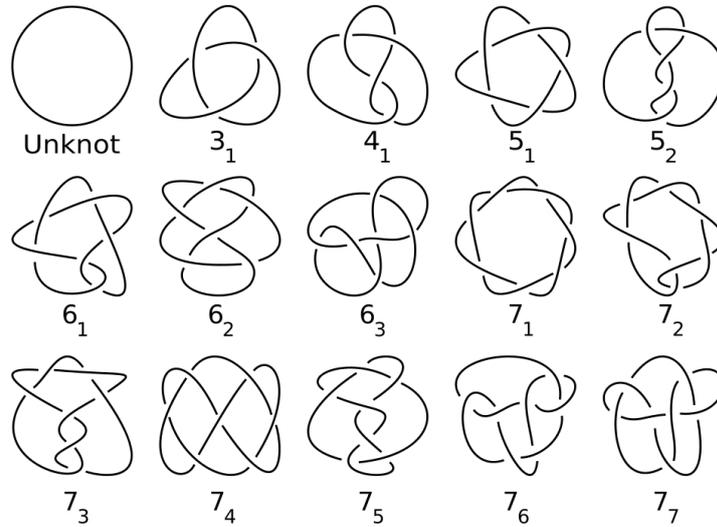


Figure 2.1.: Table of prime knots with at most seven crossings. The knots are labeled by their crossing number (minimum number of crossings in a projection), the index refers to the type of knot. The 3_1 knot is called the trefoil knot, the 4_1 knot is referred to as the figure-eight knot. As commonly done, mirror images are not included. Source: Wikipedia (2010).

‘real’ knots are the well-known *trefoil knot* and the *figure-eight knot*. Both, the trefoil knot and the figure-eight knot are displayed in Fig. 2.1, which also includes all prime knots with up to seven crossings.¹ The crossing number states the minimum number of crossings present in a projection of the knot. The crossing number is one knot invariant used to distinguish knots. It may be degenerate as is apparent from Fig. 2.1. Other knot invariants used to distinguish knots are the knot polynomials such as the Alexander and Jones polynomials [1, 80].

Curiously, an hypothesis about the nature of atoms, which soon turned out to be wrong, led to early interest in knot theory. Lord Kelvin (William Thomson) suggested that atoms are knots in the ether and that different knots correspond to different elements, which motivated an early tabulation of knots by Tait. Today, knot theory is part of the mathematical field of topology with a number of open questions which are simple to formulate but hard to solve [1]. This does not mean, however, that the study of knots is limited to mathematics. Knots and knot-like structures are of interest in biology, chemistry, and physics [1]. A number of interesting questions arise when extending the mathematical definition of a knot to a more physical one, distinguishing different representations of the same knot by their shape and spatial extension and considering knots in open structures [11]. In a polymer physics context, it is interesting that long cyclised polymers are almost always knotted.² Many studies addressed the question of how large these knots are, specifically the scaling of knot size with polymer length. One motivation for this kind of studies is the aim to understand how topoisomerases (enzymes changing the topology of DNA [3]) can affect the knottedness of DNA by acting locally; see two recent reviews [80, 95].

¹The term *prime knot* refers to the fact that these knots cannot not be ‘decomposed’ into non-trivial knots in analogy to prime numbers, which may not be decomposed into two factors both different from unity [1].

²This fact is known as the ‘Frisch-Wasserman-Delbrück’ conjecture originating from the early 1960s, which could later be proved for self-avoiding chains, see Refs. [80, 95] and therein.

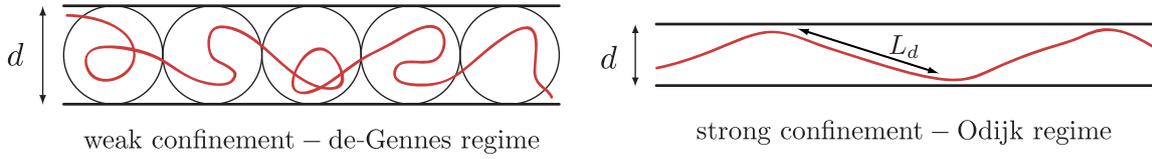


Figure 2.2.: Regimes of confinement for long worm-like chains with self-exclusion. **(Left)** Under weak confinement, commonly referred to as the de-Gennes regime, the chain may be represented as a succession of blobs. **(Right)** Under strong confinement, in the Odijk regime, the worm-like chain is mainly oriented along the channel axis and may be described as being deflected back and forth from the channel boundaries. The equivalent of the blobs is a succession of short elements of length L_d (Odijk length).

I would be remiss to conclude this section without mentioning that knots have even motivated macroscopic experiments [11], a rather rare occasion in 21st century physics. For example, Raymer and Smith addressed the “Spontaneous knotting of an agitated spring” [106]. The authors placed a string in a box, rotated the box for a while, and studied the probability of knot formation and the types of knots created. This work received the 2008 “Physics Ig Nobel Prize”.³

2.2. Polymers under confinement

Consider a long semi-flexible polymer, modeled as a worm-like chain, with contour length L much larger than the persistence length l_p . As already described in Sec. 1.2, the polymer may be treated as flexible on scales larger than l_p , and it adopts a globular shape, the size of which increases with contour length as $R \sim L^\nu$ ($\nu \approx 0.588$). Placing this globule into a channel with diameter d much larger than R has no effect on the shape of the polymer, it is just restricted in its translational motion.

The situation changes when the diameter is decreased such that $d \ll R$, but still $d \gg l_p$. In this regime, the polymer is often described by *de-Gennes’ blob picture* [24] illustrated in Fig. 2.2 (left). The conformation of the polymer is represented by a succession of blobs, treated as hard spheres. Inside each blob, the polymer does not ‘feel’ the effects of the channel boundaries and can be described by a real flexible chain, i.e., a flexible chain with self-exclusion. In this regime, the extension of the polymer along the channel axis scales linearly with the total length since each blob incorporates the same amount of contour length. Note that excluded volume is responsible for this elongation; the blob picture does not apply to ideal chains where excluded volume interactions are absent.

A further decrease of the channel diameter to $d \ll l_p$ brings us into the regime of strong confinement, called the *Odijk regime*. Naively, one would think that under this strong confinement, the chain could be treated as a hard rod on length scales smaller than l_p and that the polymer could thus be described as a succession of L/l_p segments of length l_p oriented roughly along the channel axis and deflected from the boundaries. This is not the correct, however. As Odijk described, one must instead introduce a new length scale L_d , today called the Odijk length, the (contour) length on which deflections from the wall occur on average [92].

³<http://improbable.com/ig/>

Originally, Odijk determined this length scale by comparing mean-square deviations of the worm-like chain from the channel axis with the radius of the channel and obtained

$$L_d^3 \sim d^2 l_p. \quad (2.1)$$

Fig. 2.2 (right) sketches the situation in the strong confinement regime. The polymer is described by a succession of L/L_d segments oriented along the channel axis and deflected from the boundaries. The extension along the channel axis in the Odijk regime scales linearly with the contour length L as in the de-Gennes regime, but for a different reason. In the latter, it is excluded volume interactions which lead to a stacking of the blobs, in the former it is the finite persistence length which leads to an extension along the channel, self-exclusion plays no role for the extended chain. The polymer is very stiff and may be considered free on length scales below the Odijk length, but different sections of the polymer being further apart from each other than L_d are statistically independent. The typical relaxation time of undulations perpendicular to the channel axis may thus be estimated by the relaxation time of one Odijk segment, i.e., the relaxation time scales as L_d^4/l_p (see Eq. (1.4)).

In 1983, Odijk's motivation was to study the motion of one worm-like chain in a network. To that end, he considered the chain trapped in a cylindrical pore (tube) as described above. Today, rectangular channels with cross-sections ranging from micrometer to nanometer dimensions may be fabricated using lithography techniques. This technological progress allows for the study of semi-flexible filaments and polymers (actin filaments with $l_p \approx 20 \mu\text{m}$, DNA with $l_p \gtrsim 50 \text{ nm}$) down to or close to the Odijk regime of confinement [55, 107]. The opportunity to experimentally study polymers in different regimes of confinement including the Odijk regime has led to a simultaneous interest from the theoretical side, e.g., a study combining Monte Carlo simulations with analytical approximations (aiming to go beyond the scaling picture) [147] and a more refined scaling analysis between the de-Gennes and Odijk regimes [94].⁴

2.3. Knots in strongly confined polymers

In our theoretical study of a confined semi-flexible polymer, we did not aim to explore the thermodynamic or dynamic properties of a totally elongated polymer in a nanochannel. Instead, as stated above, we were interested in the dynamics of a topological perturbation, more precisely in the dynamics of a simple (trefoil) knot and the modes of spontaneous unknotting. Here, I will outline the main ideas; details are found in Ref. [84] reprinted in Sec. 2.6.

Let me begin by characterizing the system. Since the channel leads to an overall orientation of the polymer along the channel axis, an unambiguous definition of the knot becomes possible for the linear polymer: simply connect the ends virtually in infinity. We considered a trefoil knot which is loose but not tightened. This structure implies the existence of at least two U-turns, imagine an elongated knotted cord (or see Fig. 1 in Sec. 2.6). Such U-turns are energetically costly, a rough estimate yields a bending energy of $\pi l_p k_B T/d$ (apply Eq. (1.1) to a worm-like chain with the shape of a semi-circle with diameter d given a persistence length l_p), which is (much) larger than $k_B T$ for strong confinement.⁵ This suggests that

⁴See references in Ref. [94] for more experimental and theoretical studies on nanoconfined DNA until 2008.

⁵Two remarks: First, note that in our study, $l_p/d = 3$, i.e., the persistence length is rather comparable to the channel diameter, but U-turns are still costly. Second, Odijk argued that the free energy cost for U-turns is even larger than estimated by considering bending energy alone since entropic forces cause the U-turns to tighten further [93].

U-turns and also knots are rare. It is therefore reasonable to concentrate on the dynamics of one knot (i.e., two U-turns) in the otherwise elongated polymer. Moreover, the U-turns allow us to define the size and position of the knot. A suitable measure for knot size is the contour length between the two U-turns, and knot position is conveniently defined as the midpoint between the U-turns on the polymer contour. How can the U-turns and thus the knot disappear by thermal fluctuations? The U-turns cannot ‘annihilate’ each other due to the topology of the knot. They can, however, both diffuse to the same end of the polymer contour and disappear, which corresponds to a change of knot position leading to unknotting. Alternatively, both U-turns can diffuse to opposite ends of the polymer, which corresponds to swelling of the knot, i.e., a change in knot size. This ultimately also leads to a loss of the knotted configuration. In our study, we concentrated on the dynamics of knot size and knot position and asked which of the two scenarios described leads to unknotting of the polymer.

To address these questions, we took a multi-scale approach both in coarse-graining of the model and in the time scales considered. On the small scale, we treated the polymer as a worm-like chain with excluded volume interactions, needed to preserve the topology (full model). We used standard Brownian dynamics (see Appendix A.1) to characterize the dynamics of knot size and knot position on small time scales. On the large scale, we constructed a coarse-grained model *de novo*. The coarse-grained model is based on the expectation that on long time scales the slow modes (i.e., movements of large polymer segments relative to the solution) are relevant. For the trefoil knot under consideration, there are three large segments: the polymer region between the two U-turns and from the two U-turns to the polymer ends, respectively. Our coarse-grained model considers these three segments and their (overdamped) motion relative to solution with friction (hydrodynamic drag) being proportional to its length. While the three segments move independently, they exchange length. Since the sum of the lengths of the three segments must equal the contour length L of the polymer, there are only two degrees of freedom to consider. These may be mapped to knot size k and knot position p . Making these considerations quantitative and taking thermal fluctuations into account, we obtained a coupled system of two stochastic differential equations for knot size and knot position with multiplicative noise.⁶ Taking into account boundary conditions (one reflecting boundary corresponding to minimal knot size and absorbing boundaries corresponding to unknotting), this describes our (parameter-free) coarse-grained model.

To validate the coarse-grained model, we compared the short time dynamics of the full model measured using Brownian dynamics simulations with the analytically determined short time dynamics of the coarse-grained model. In particular, we found surprisingly good agreement for the diffusion coefficient D_k for knot size, the diffusion coefficient D_p for knot position, and the coefficient $D_{kp} = D_{pk}$ describing the cross-correlation between each. All these coefficients are gathered in the diffusion matrix which is at the core of the coarse-grained model:

$$\begin{pmatrix} D_k & D_{kp} \\ D_{pk} & D_p \end{pmatrix} = \begin{pmatrix} \frac{1}{4} \left[\frac{1}{l} + \frac{1}{r} \right] & \frac{1}{8} \left[\frac{1}{r} - \frac{1}{l} \right] \\ \frac{1}{8} \left[\frac{1}{r} - \frac{1}{l} \right] & \frac{1}{16} \left[\frac{4}{k} + \frac{1}{l} + \frac{1}{r} \right] \end{pmatrix}. \quad (2.2)$$

For clarity, the individual components have been expressed using the lengths l and r of the polymer segment to the left and right of the knot, respectively, which are just functions of k and p as mentioned above. For example, a small knot ($k \ll L$) located in the middle of the polymer ($p = L/2$) implies $l, r \approx L/2$. (Note that units of D are chosen as in Sec. 2.6.)

⁶Beside the publication reprinted in Sec. 2.6, see also the more general description for the set of equations in Appendix A.2.

The dependence of the individual coefficients on knot size and position can easily be understood qualitatively. For example, the diffusion coefficient associated with knot position is dominated by a term $\sim 1/k$ for small knots, arising from diffusion of the segment between the two U-turns. Moreover, the diffusion coefficient associated with knot size depends on the length of the polymer, $D_k \sim 1/L$, since a change in knot size requires the polymer stretches to the right and left of the knot to change position and $r, l \approx L$. The immediate question arises: Are changes in knot size significant during the unknotting process, given a small knot initially located in the middle of the polymer? Or in other words, does knot swelling contribute to unknotting? In a time τ , the typical time span until the knot dissolves, the knot grows to a typical size $k \sim \sqrt{D_k \tau} \sim \sqrt{\tau/L}$. However, the typical time until unknotting by diffusion to the polymer end is given by $\tau \sim L^2/D_p \sim L^2 k$. Taken together, these estimates suggest that the knot swells to macroscopic sizes, $k \sim L$, and the unknotting time scales as $\tau \sim L^3$.

To test this scaling prediction, we simulated the coarse-grained model capturing the full dynamic interplay between k and p , starting out with a small knot in the middle of a long polymer. The simulations revealed that both mechanisms, diffusion of the knot along the polymer and swelling of the knot, contribute to unknotting. However, the simulations confirmed the scaling relations derived above, i.e., the mean final knot size scales linearly with the length of the polymer, the mean unknotting time scales cubically.

As a corollary, we considered the effect of tension applied in addition to the confinement. If the external force is sufficiently large to constrain the knot to small sizes, swelling cannot occur. The knot can only disappear by diffusion to the end of the polymer, which results in a quadratic scaling of unknotting time with polymer length, $\tau \sim L^2$. This means that, counterintuitively, an external force which suppresses one possible mode of unknotting effectively speeds up the disappearance of the knot.

Caveats and simplifications So far, one caveat our study faces has not been mentioned. As stated above, the full model and the coarse-grained model agree very well concerning the diffusion coefficients of knot size and knot position. There is, however, a small systematic discrepancy one observes comparing drift velocities. The results of the Brownian dynamics simulations suggest a drift towards smaller knot sizes, an effect not included in the coarse-grained model. Since, on long time scales, drift dominates diffusion, even a small drift to smaller knot sizes can potentially prevent diffusive knot swelling to macroscopic sizes for long polymers and thus affect the scaling behavior. We showed that the measured drift arises from the excluded volume interactions in the knot region (and not from the topology of the knot). Assuming that in experimental situations excluded volume interactions can be made much smaller than in our simulations, the observed drift effects do not necessarily qualitatively affect the knot dynamics in experiments. To quantitatively understand the origin of the drift we are currently employing a toy model described below in Sec. 2.4.

In addition, our model is very simplistic since it does not take hydrodynamic and electrostatic interactions into account. While we expect that hydrodynamic interactions have only a quantitative influence on the absolute times scale of unknotting, the case may be different concerning electrostatic interactions. Electrostatic effects between the walls and the polymer as well as between different parts of the polymer should be weak for very small screening lengths. Yet, even weak effects could prevent knot swelling to macroscopic sizes as explained above. A careful analysis of the effect of electrostatic interactions in this geometry may help to clarify this issue, similar to an existing study on knotted unconfined polymers [27].

Relation to other studies While our research was in progress, another study addressing knot dynamics in nanochannels was published [81]. The work had been motivated by an observation of bright ‘features’ in fluorescent images of confined DNA which the authors attributed to knots. The authors suggested that two different modes contribute to knot motion which involve movements of different parts of the polymer but the knot region stays constant in size, i.e., the knot does not swell. Without going into further detail here, let me conclude that the authors predicted a quadratic scaling of unknotting time with length for long confined chains, contrary to our results. Since we used simulations to verify our coarse-grained model, especially diffusive knot growth, we are confident that our study captures the right physics. Experiments, however, will be necessary to shed further light on the problem of knotted polymers under confinement.

Closely related to the dynamics of knots in nanochannels is the dynamic of knots in polymers elongated by tension.⁷ A couple of years ago, Bao *et al.* succeeded in experimentally tying knots of defined topology into individual DNA molecules and followed the dynamics of these knots under tension [10]. This experiment motivated a number of theoretical studies employing simulations [43, 75, 146] and scaling arguments [38]. This experimental and theoretical work differs in many respects, including the questions addressed and the models used. However, there seems to be agreement on diffusion of compact knots along the polymer contour with the diffusion coefficient depending on details of the knot but not on the length of the polymer under tension. No clear picture, however, emerges about what enables knot motion. There are arguments for a snake-like motion (‘self-reptation’) [10] and a sliding motion model [43].

How is this connected to our study? First, diffusion of a compact knot along the polymer contour is fundamentally different from knot swelling combined with diffusion along the contour. Thus, if our analysis is correct and effects driving knots to compact structures are negligible, knot dynamics in polymers under tension is fundamentally different from knot dynamics in confined polymers. Second, the vague picture of what mechanism enables knot mobility in polymers under tension provides an impetus to address this question for confined polymers as well.

2.4. Outlook: two overlapping polymers under strong confinement

In order to address the drift towards smaller knot sizes, which we observed in the Brownian dynamics simulations, we started to study a related but simplified problem. We considered two polymers which cannot intersect themselves (excluded volume interactions, finite polymer width), but are overlapping along the channel axis. Aiming to characterize the thermodynamics and the dynamics of this system, we applied Monte Carlo and Brownian dynamics methods. Monte Carlo simulations confirmed a free energy penalty for overlap which is extensive. Brownian dynamics simulations showed that the system can be described as effectively one-dimensional, the measured drift towards smaller overlap can be approximated by the gradient of the free energy taking the mobility of the two polymers into account. Both overlap free energy and drift become smaller when decreasing the ratio between polymer width and channel diameter (unpublished results by Clemens Veiglhuber [145]). Taken together, these observations suggest that the drift towards smaller knot sizes can become very small and

⁷For the dynamics of knots in closed ring polymers see Ref. [95].

thus excluded volume effects do not necessarily lead to significant knot shrinking. This fact favors our coarse-grained model for knot dynamics and the contribution of knot swelling to unknotting.

Though initially proposed as a toy model to understand the drift towards smaller knot sizes, the results of our study have broader relevance. The study sheds light on the segregation process of two confined polymers, the dynamics of two polymers interchanging places along the channel axis, and entropic unfolding of molecules within channels. Such questions have recently been investigated theoretically and experimentally for large channel widths, beyond the strong confinement regime [7, 62].

2.5. Conclusion

Confined DNA and knotted DNA each demonstrate interesting physics. While confinement to nanochannels offers interesting technological prospects by allowing a mapping between channel axis and the DNA backbone, knots destroy this advantageous mapping. This motivated us to address knot dynamics in strongly confined polymers, focusing on the mechanisms by which knots dissolve and the scaling of unknotting time with polymer length. In this chapter, I gave a brief introduction into both knots and the physics of confined polymers and subsequently outlined the main results and limitations of our study.

We used a multi-scale approach to study the dynamics of a confined and knotted polymer employing both Brownian dynamics simulations and a coarse-grained model. Our analysis revealed that the polymer unknots by both swelling of the knot and diffusion of the knot to the end of the polymer. During the unknotting process, the knot typically grows to macroscopic sizes, and the unknotting time scales cubically with polymer length. This is fundamentally different from the situation of a knotted polymer under tension where the knot (at sufficiently large tension) remains small and disappears by diffusion to the end of the polymer; the unknotting time in this case scales quadratically with polymer length. The detailed mechanisms enabling knot mobility, however, are not well understood, which leaves room for follow-up studies on knotted polymers under confinement and under tension. At the end of the chapter, I gave a brief outlook on an ongoing study on the thermodynamics and dynamics of two overlapping polymers under strong confinement.

Due to copyright reasons, a reprint of

Wolfram Möbius, Erwin Frey, and Ulrich Gerland
Spontaneous unknotting of a polymer confined in a nanochannel
Nano Letters, **8**, 4518 (2008)

DOI: 10.1021/nl802559q

cannot be included in the electronic version of this thesis.

Please download the publication from the journal's website.

Due to copyright reasons, a reprint of

Wolfram Möbius, Erwin Frey, and Ulrich Gerland
Spontaneous unknotting of a polymer confined in a nanochannel
Nano Letters, **8**, 4518 (2008)

DOI: 10.1021/nl802559q

cannot be included in the electronic version of this thesis.

Please download the publication from the journal's website.

Due to copyright reasons, a reprint of

Wolfram Möbius, Erwin Frey, and Ulrich Gerland
Spontaneous unknotting of a polymer confined in a nanochannel
Nano Letters, **8**, 4518 (2008)

DOI: 10.1021/nl802559q

cannot be included in the electronic version of this thesis.

Please download the publication from the journal's website.

Due to copyright reasons, a reprint of

Wolfram Möbius, Erwin Frey, and Ulrich Gerland
Spontaneous unknotting of a polymer confined in a nanochannel
Nano Letters, **8**, 4518 (2008)

DOI: 10.1021/nl802559q

cannot be included in the electronic version of this thesis.

Please download the publication from the journal's website.

Due to copyright reasons, a reprint of

Wolfram Möbius, Erwin Frey, and Ulrich Gerland
Spontaneous unknotting of a polymer confined in a nanochannel
Nano Letters, **8**, 4518 (2008)

DOI: 10.1021/nl802559q

cannot be included in the electronic version of this thesis.

Please download the publication from the journal's website.

3. Dynamics of nucleosomal DNA

In the first chapter of this thesis, DNA was described as the substrate for the formation of nucleosomes, the smallest unit of DNA compaction in eukaryotic cells. We saw that nucleosomes give rise to higher order chromatin structures, that their binding specificity depends on the DNA sequence, and that nucleosomes are subject to variability in their composition, subject to covalent modifications and active remodeling. An important point not yet explicitly emphasized is that most of the genomic DNA is truly buried inside nucleosomes. The length of linker DNA between nucleosomes is much shorter than the length of DNA wrapped up inside a nucleosome. This poses an obstacle for enzymatic machines that need to access DNA for transcription, replication, and DNA repair. The obstacle is even larger for proteins regulating these processes since they passively bind to specific sites on the DNA and cannot directly make use of ATP hydrolysis to perform their function. Thus, nucleosomes control access to binding sites on the DNA by virtue of their mere presence.

There are different effects which hinder protein binding: the presence of the bound histone complex, the strong deformation of the wrapped DNA, and the proximity of the second turn of DNA wrapped up in a nucleosome. This is not to say that binding of molecules to nucleosomal DNA is impossible [29]. However, it is limited to small molecules which match the structure of the DNA within a nucleosome. This picture changes when one takes into account the fact that nucleosomes are not static entities as suggested by the X-ray crystal structure displayed in Fig. 1.3. Thermal fluctuations can lead to transient unwrapping of DNA from the histone core, allowing for protein binding to the unwrapped DNA. This mechanism is termed *nucleosome breathing* or *site exposure* [99] and has been studied extensively *in vitro* for more than a decade. Theoretical work, by contrast, is sparse, a fact which motivated us to study the site exposure mechanism within a coarse-grained model [86].

The site exposure mechanism in single nucleosomes and its modeling take center stage in this chapter. I will first outline the key experiments characterizing site exposure and, in particular, its dynamics. Then, I will describe our model and the main results of our study. Our results will be related to experiments, and possible extensions of our model will be discussed. Within this chapter, I will also discuss a phenomenon we call *flexibility-assisted barrier crossing* which helped us to understand the dynamics within our model, but also displays interesting physics itself.

3.1. Introduction to the site exposure mechanism

3.1.1. Thermodynamics

As described above, the site exposure mechanism allows proteins to passively bind to DNA which is usually buried inside the nucleosome as illustrated in Fig. 3.1. In the first step, a stretch of DNA is transiently exposed to solution at a rate k_{unwrap} , solely driven by thermal fluctuations. In the second step, a protein binds to this stretch of DNA at a rate k_{bind} , just as it would bind to naked DNA. The opposite processes also occur, i.e., rewinding of DNA at rate

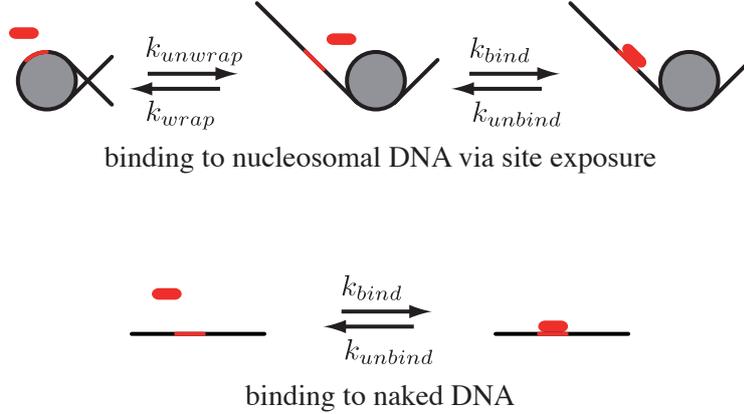


Figure 3.1.: (Top) Illustration of the site exposure mechanism. Passive binding of a protein (red blob) to a site buried inside a nucleosome (marked in red) is made possible by transient unwrapping of a DNA stretch on which the protein can bind. (Bottom) Compare with the situation of protein binding to naked DNA.

k_{wrap} and unbinding of the protein at rate k_{unbind} . Note that the mechanism, and especially the rates k_{wrap} and k_{unwrap} , in principle do not depend on the specific protein involved, but on the properties of the nucleosome and the location of the binding site. At a sufficiently low protein concentration, unwrapping and rewrapping occur on much faster time scales than protein binding ($k_{unwrap} \gg k_{bind}$, $k_{wrap} \gg k_{bind}$). The binding constant of the protein to the DNA and the effective rate of binding are then suppressed by the factor $K_{nuc} = k_{unwrap}/k_{wrap}$ relative to the case of binding to the same stretch of ‘naked’ DNA, analogous to the well-known Michaelis-Menten kinetics. K_{nuc} , naturally, represents the fraction of time the DNA is unwrapped.

Polach and Widom tested the proposed mechanism using a digestion experiment [99].¹ They constructed mononucleosomes using the 5S positioning sequence and introduced restriction sites, i.e., sites on the DNA where cleavage enzymes cut naked DNA. Indeed, DNA buried inside nucleosomes was cut, an observation which was attributed to site exposure. By comparing the cutting rate of nucleosomal DNA to that of naked DNA, the authors could estimate the equilibrium constant K_{nuc} for unwrapping, depending on the location of the restriction site within the nucleosome. In a follow-up study, Anderson and Widom placed several restriction sites onto a high-affinity 601.2 sequence (derived from the 601 positioning sequence) and again determined the equilibrium constants for unwrapping [6]. Fig. 3.2A shows the combined results. The equilibrium constants vary between about 10^{-2} and 10^{-6} depending on the position of the restriction site and depending on the DNA sequence: First, the further inside the nucleosome the restriction site is located, the larger the suppression is. This is consistent with a progressive unwrapping mechanism, i.e., ends are more often unwrapped than parts further inside the nucleosomal DNA. Second, suppression of protein binding is larger for the 601.2 sequence which has higher binding affinity than the 5S sequence variants. Thus, increased binding affinity results in reduced unwrapping of DNA from the histone core complex.

¹For a review of the early studies on site exposure see Ref. [151].

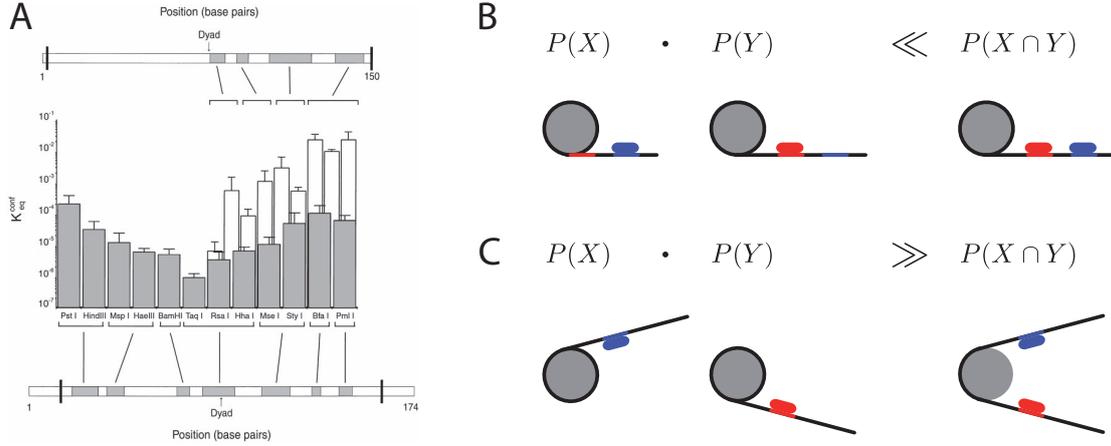


Figure 3.2.: (A) Equilibrium constants $K_{nuc} \equiv K_{eq}^{conf}$ for site exposure of nucleosomes with 5S sequence variants (light) and the 601.2 sequence (dark). K_{nuc} depends on the sequence and on the position of the binding site within the nucleosome (indicated at top and bottom). Reproduced with permission from Ref. [6]. (B) Cooperative interaction between two proteins X and Y mediated by a nucleosome. Binding of protein X (blue blob) is facilitated by protein Y being bound (red blob) and vice versa if both binding sites are placed on the same end of nucleosomal DNA. (C) In contrast, negative cooperativity between X and Y can emerge if both binding sites are located at opposite ends.

Sequence and position of the binding site are not the only determinants of site exposure equilibrium constants. Site exposure, for instance, has also been shown to depend on the state of the tails; hyperacetylation of nucleosomes leads to 1.4-fold increase in accessibility on average [5]. Regarding the physics of site exposure, however, the site dependence of K_{nuc} is most important since it implies a progressive unwrapping mechanism as indicated above. Reiterating that the interaction between DNA and histone complex is centered at the 14 points where the minor groove faces the histone complex, Fig. 3.2A may be used to estimate the effective adsorption/binding energy per contact point (or equivalently the free energy cost to unwrap a stretch of about 10 base pairs (bp)) to $\approx 1.5 - 2 k_B T$ [122].

3.1.2. Interactions mediated by the nucleosome via site exposure

Beyond allowing proteins to bind to DNA usually buried inside the nucleosome, the site exposure mechanism is able to mediate *cooperative binding* between two proteins [100]. Imagine two binding sites located on the same end of nucleosomal DNA as displayed in Fig. 3.2B, one for protein X and one for protein Y . Protein X alone binds to the DNA with probability $P(X)$, protein Y alone binds with probability $P(Y)$. The probability of both proteins binding simultaneously, $P(X \cap Y)$, is much larger than the product $P(X) \cdot P(Y)$. This is due to the fact that binding of X alone already implies that a stretch of DNA is unwrapped which in turn facilitates binding of Y and vice versa. In more general terms: Proteins X and Y ‘share’ the free energy cost of DNA unwrapping. In the same fashion as site exposure does not depend on the specific protein binding, the cooperativity mediated by the nucleosome is independent of the two proteins and does not require direct interaction between them. It has been shown that such cooperativity functions *in vivo* [83] which today is referred to as one indication that site exposure itself functions *in vivo*, see Ref. [97].

Negative cooperativity can emerge if the binding sites of X and Y are located at different ends of the nucleosomal DNA, as shown in Fig. 3.2C. Reiterating that DNA is (negatively) charged reveals the mechanism behind this negative cooperativity. Both turns of the DNA wrapped around the histone octamer effectively repel each other, and this repulsion effectively eases the transient unwrapping of DNA. Unwrapping a stretch of given length is thus more probable in the presence of the neighboring DNA turn than in its absence. Imagine now that protein X is bound, a scenario that would imply a stretch of DNA being already unwrapped. Binding of protein Y can thus be suppressed if its concomitant binding would leave less than one turn of DNA wrapped up. The same situation holds vice versa, formally: $P(X) \cdot P(Y) \gg P(X \cap Y)$. This negative cooperativity is also independent of the specific proteins involved. However, it depends on the location of the binding sites. They must be located sufficiently far inside the nucleosome for less than one turn of DNA to remain wrapped in the case of concomitant binding.

3.1.3. Relevance beyond enabling protein binding

Though site exposure has been proposed as a mechanism for proteins to passively access buried DNA, its relevance and applicability reaches beyond in several ways. First of all, site exposure is related to the unpeeling of DNA from the histone octamer when pulling at both ends; experiments together with theoretical studies allow for further investigation of interactions between DNA and histones [17, 59].

Second, and biologically more relevant, site exposure may be related to nucleosome ‘sliding’, i.e., to nucleosomes passively diffusing (or translocating) along the underlying DNA substrate. The most obvious mechanism of mobilizing nucleosomes, translocation of nucleosomes by dissociation and reformation would be too costly given the large energy required to break all contacts between DNA and histones. Rolling of the nucleosome along the DNA is not possible for geometrical reasons. Therefore, it has been proposed that site exposure leads to unwrapped DNA which – from time to time – forms a ‘bulge loop’ when rewrapping (instead of returning to the fully wrapped configuration). Such a loop can then diffuse over the histone surface and annihilate at either end. If the loop disappears at the opposite end from where it was created, the nucleosome is displaced. Although this is a fascinating mechanism, recent experiments suggest that passive nucleosome translocation is enabled by a similar mechanism which involves diffusion of twist defects instead of bulge loops; for a detailed description of mechanisms and experiments see the reviews [34, 59, 122] and references therein.

Third, site exposure becomes important when active machinery progressing along the DNA encounters a nucleosome. A recent *in vitro* experiment has shown that transcription through a single positioned nucleosome is related to site exposure. The RNA polymerase does not actively separate DNA from the histone complex but waits for spontaneous unwrapping before advancing (and at the end, the nucleosome is translocated upstream by loop formation) [42]. A theoretical study by Chou focused on the interplay between site exposure, a progressing motor which can advance only if the DNA in front is not attached to the histone, and also loop formation within the nucleosome [19]. In this study, the progressing motor was meant to represent a remodeling factor. Though the working mechanisms of chromatin remodelers are more complicated, their action is associated with the breaking of histone-DNA contacts and the formation of small DNA loops in nucleosomes. Hence, their mechanism is related to nucleosome breathing at least to some extent [21].

3.2. Dynamics of site exposure

3.2.1. Experiments

In 1997, shortly after the site exposure mechanism had been proposed, Protacio et al. were able to indirectly estimate lower bounds for the rate of DNA unwrapping [101]. They found that site exposure occurs on a time scale faster than seconds. However, almost a decade elapsed before direct and precise measurements of the dwell times of site exposure were reported [49, 51, 63, 136, 137]. Direct observation of site exposure dynamics can be achieved by *Fluorescence Resonance Energy Transfer* (FRET). This method is based on radiationless energy transfer between a donor and an acceptor fluorophore with overlapping donor emission and acceptor absorption spectra.² The donor is excited and, in the absence of an acceptor, emits light at its own emission frequency. If an acceptor molecule is in close proximity, energy is transferred without radiation from the donor to the acceptor through dipolar coupling. Consequently, the donor emits less light and emission from the acceptor can be detected. The efficiency of the energy transfer depends as $1/(1 + (r/R_0)^6)$ on the distance r between acceptor and donor, where R_0 is called the Förster radius which is characteristic for a specific donor-acceptor pair. In biomolecular applications of FRET, R_0 typically is about 6 nm, which makes FRET a useful tool to measure inter- and intramolecular distances. For this reason, FRET is often used as a ‘molecular ruler’. In the following, I will discuss the recent experiments probing site exposure using FRET. In all those studies, the fluorophore pairs are biochemically attached to the nucleosome allowing for the detection of conformational changes as they occur during site exposure.

Li et al. constructed a mononucleosome using a 147 bp long positioning sequence which additionally contained a binding site for LexA (a protein from *Escherichia coli*) extending from base pair 8 to base pair 27 at one end of the DNA [63]. In addition, they biochemically attached fluorophores to the nucleosome. A donor was attached to the end of the DNA close to the LexA binding site, and an acceptor to the histone octamer. The geometry was designed such that donor and acceptor were close and thus FRET efficiency was high in the fully wrapped state, but lower in the case where DNA was partially unwrapped and the LexA protein was bound. As expected, the authors found a significant decrease in FRET efficiency upon titration with LexA, consistent with a conformational change of the DNA end required for (and stabilized by) LexA binding. To characterize the kinetics of this conformational change, a stopped-flow experiment was performed: the authors measured the time course of FRET decrease after rapid mixing of nucleosomes with LexA. From that experiment, the effective rate of LexA binding could be estimated to $k_{unwrap} \sim 4 \text{ s}^{-1}$. Repeating the experiment with another LexA concentration resulted in the same rate indicating that, indeed, nucleosome unwrapping and not protein binding was the rate limiting step (in contrast to the digestion experiments described above). Thus, the measured rate could be attributed to the unwrapping of at least 30 bp of DNA. Using an equilibrium constant of $4.5 \cdot 10^{-2}$ for site exposure (based on estimates from preceding work [64]) permitted determination of the rewinding rate to $k_{wrap} \sim 90 \text{ s}^{-1}$. In order to confirm the results, the authors performed a fluorescence correlation spectroscopy (FCS) experiment. They measured the auto-correlation function of donor fluorescence within a small confocal volume. By repeating these measurements using a nucleosome construct with the acceptor missing, the authors were

²In general, the method is termed *Förster Resonance Energy Transfer* and does not necessarily require fluorophores.

able to separate diffusion of nucleosomes from site exposure dynamics. The values for the rates obtained are similar to those stated above, $k_{unwrap} = 3.6 \text{ s}^{-1}$ and $k_{wrap} = 20 \text{ s}^{-1}$. In conclusion and expressed in terms of dwell times: the nucleosome stays in the fully wrapped state for $\sim 250 \text{ ms}$ before exposing a stretch of DNA for $\sim 10 - 50 \text{ ms}$.

The experiment by Li et al. [63] for the first time allowed for a direct measurement of site exposure dynamics, but nevertheless was a bulk measurement. Tracking the state of individual particles using single-pair FRET offers a deeper understanding of nucleosome dynamics. For example, it could permit one to directly distinguish between multi-step and single-step opening transitions. Ideally, the experiment should be performed on nucleosomes freely floating in solution. However, the time scales of interest are too large relative to diffusion times. On the time scales of interest (milliseconds or larger) the nucleosome will have diffused out of the confocal volume before a change in FRET efficiency is observed. Thus, two different groups attempted studying site exposure using nucleosomes immobilized on a surface.

Koopmans et al. constructed nucleosomes containing a positioning sequence and about 30 bp of linker DNA [49]. The nucleosomes were tethered to the surface using a biotin linker. Acceptor and donor fluorophores were attached to the DNA, at the site where DNA exits the nucleosome to the linker and close to the dyad. On the extended DNA strand, both fluorophores are about 80 bp apart. Within the nucleosome, however, their distance is only about 4 nm. The authors claim that the positioning at the end of the wrapped DNA makes their construct comparable to the one used in the study by Li et al. [63]. Two obstacles made the experiment difficult. First, photoblinking (the acceptor stochastically goes into a dark state) interfered with the observation of conformational changes within the nucleosome and had to be suppressed biochemically. Second, only a very small fraction of all nucleosomes immobilized on the surface were intact and showed FRET dynamics which could be associated with site exposure. From those, however, Koopmans et al. were able to deduce a dwell time of 120 ms in the open state. The onset of photobleaching impeded quantification of the time the nucleosome stays in the closed state.

In a follow-up study, Koopmans et al. systematically addressed the challenges of immobilizing nucleosomes and compared different surface passivation strategies using the same nucleosome construct as before [51]. In this study, dwell time of 25 ms in the open state and 280 ms in the closed state are reported, contrary to the preceding study, but comparable to the work by Li et al. [63]. For a more careful discussion of the results see Ref. [51].

An early study by Tomschik et al. [137] also used immobilized nucleosomes, but with acceptor and donor located further inside the nucleosome, close to the dyad axis. Fluctuations bringing those fluorophores apart should involve DNA unwrapping to a much greater extent than in the experiments described so far and thus would permit testing site dependence of nucleosome dynamics. Originally, Tomschik et al. did in fact report such large conformational changes including the corresponding dwell times [137]. Recently, however, most events earlier interpreted as changes in nucleosome conformation were reinterpreted as photoblinking [137, correction].³ In a follow-up study, Tomschik et al. reevaluated their results [136]. They used cross-linked nucleosomes which cannot undergo unwrapping and (normal) control nucleosomes to estimate the effect of photoblinking. Concluding, they could argue for the existence of unwrapping events where a long stretch of DNA had to be involved. However, they were not able to measure dwell times.

³The correction appeared after publication of our study [86] where we report the dwell times originally published in Ref. [137].

Taken together, after fruitful years of studying the equilibrium constants of site exposure and its biological significance, recent studies now focus on the dynamics of the mechanism. Though direct measurements imply that site exposure of short DNA pieces happens on the scale of milliseconds, the overall picture remains incomplete. In particular, a systematic study of dwell times depending on the location of the putative protein binding site (or equivalently the location of the FRET pair) is missing but is likely to be published in near future.⁴

3.2.2. Theoretical description

In 2005, the publication of the first studies reporting dwell times in site exposure [63, 137] called for a theoretical description.⁵ In order to fill the gap, we chose to build a coarse-grained model and used it to determine dwell times depending on the length of DNA involved in unwrapping. Thereon, we were able to capture the underlying physics within a more simplified model. The main ideas of our study are outlined in the following, details are found in our publication [86] reprinted in Sec. 3.6.

Coarse-grained model As described in Sec. 1.3, the DNA wrapped up in a nucleosome is forced into a helical ramp by the histone octamer, and interactions between the two are localized to the 14 points where the minor groove faces the histone complex [72]. Since we were interested in the DNA dynamics, we had to explicitly model the DNA. As is commonly done, we represented the DNA as a system of beads with springs in between and standard energy terms accounting for bending stiffness and the electrostatic self-interaction in solution (see Appendix A.1). The histone complex, in contrast, was reduced to its interaction with the DNA. Specifically, we introduced 14 attractive regions located on a helical ramp and summarized quantitatively in the equation

$$U = \gamma k_B T \sum_{n=1}^{14} \left(1 - e^{-|\mathbf{r}_{i(n)} - \mathbf{c}_n|/\rho}\right)^2. \quad (3.1)$$

This sum comprises the 14 contact points located at \mathbf{c}_n and interacting with the DNA beads at positions \mathbf{r} . The single terms in Eq. (3.1) represent parts of Morse potentials which are characterized by a potential well at \mathbf{c}_n and, importantly, a finite dissociation energy.⁶ The potential's characteristic width, $\rho = 0.5$ nm, was chosen such that it is in between the range of chemical and electrostatics interactions, both of which are known to keep the nucleosome together. The beads representing the DNA and the location of the 14 contact points, i.e., the geometry of the coarse-grained model, is sketched in Fig. 1a of Sec. 3.6.

Since the model is far from being analytically solvable, we employed Brownian dynamics (summarized in Appendix A.1) to study the thermodynamic properties and the dynamics of

⁴Very recently, a study used a variety of techniques to determine equilibrium constants of unwrapping with FRET pairs at different positions in the nucleosome [50, Table 2]. Yet, measurements of dwell times have not been possible so far.

⁵At that time, the only existing study focusing on transient DNA dissociation suggested an all-or-none process for DNA dissociation but did not include the localized interaction between DNA and the histone complex [74]. Later, another model employing continuous adhesion energy between DNA and histone complex was used to compute equilibrium constants depending on the degree of unwrapping [35], but dwell times do not seem to be reported.

⁶This potential had been proposed by Morse to describe vibrational spectra in diatomic molecules [88].

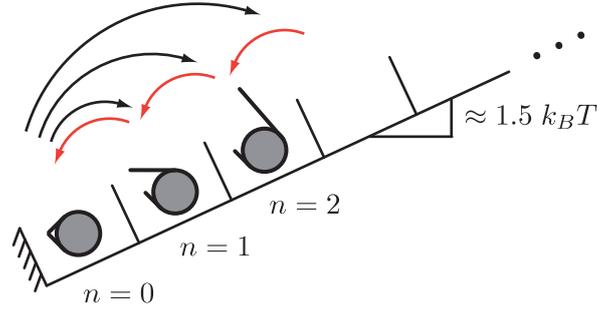


Figure 3.3.: Reduced description of the dynamics in the coarse-grained nucleosome model. The state of the nucleosome may be described by n , the number of open contact points (or, equivalently, the number of unwrapped DNA segments). Since the free energy increases with n , the system spends most of the time in the fully wrapped state ($n = 0$). Thus, the average time until contact n opens corresponds to the mean waiting time starting from the fully wrapped state (unwrapping times, black arrows). In contrast, the average time until contact n closes, i.e., the dwell time in the open state, is the average time until one single segment rewraps (rewrapping times, red arrows).

the coarse-grained nucleosome model. First of all, we observed that DNA unwraps progressively and in discrete steps. This is not surprising due to the localized attractive interactions. Bending and electrostatic repulsion between the two DNA turns counteract the attractive interaction between the histone complex and the DNA. Since the latter is comparably short-ranged, unwrapping a stretch of DNA (or equivalently ‘opening’ one contact point) involves crossing a free-energy barrier which leads to the observed discrete (and progressive) unwrapping. The two-state approximation, describing a contact point being either ‘open’ or ‘closed’, made it convenient to adjust the prefactor γ in Eq. (3.1) to resemble a net adsorption energy of $\approx 1.5 k_B T$ as estimated from experiments (see above).

Simulation results Before reporting on the dynamics, i.e., the dwell times, let us recapitulate the situation using Fig. 3.3. The state of the nucleosome may be described by n , the number of contact points with DNA not bound. Thinking of protein binding or a FRET experiment as described above, the dwell times in the open and closed states thus depend on n , the number of contact points which must be unbound to make a binding site accessible or lead to a change in FRET efficiency. Since the nucleosome stays fully wrapped most of the time, the dwell time in the closed state is represented by the mean waiting time starting from the fully wrapped state with $n = 0$ (unwrapping times, black arrows). The situation is different regarding the dwell times in the open state. To make a certain binding site inaccessible it suffices to close one contact point, i.e., to rewrap one DNA segment (rewrapping times, red arrows).

Since progressive unwrapping of DNA in this picture corresponds to running up a free energy ramp, one expects the unwrapping times to depend exponentially on n , which was confirmed by extensive simulations.⁷ The situation is more subtle regarding the rewrapping times. Naively, and based on energetic considerations alone, one would expect those dwell times not to depend on n at all. Considering, however, that rewrapping involves overdamped

⁷Because spontaneous unwrapping of multiple DNA segments is a very rare event, we had to employ the *Forward Flux Sampling* method recently developed by Allen et al. [4].

rotational motion of a stretch of DNA (see Fig. 3.3), a second thought reveals that dwell times in the open states should increase with n . More precisely, the length of DNA involved in the rewinding process should be the important factor setting the dwell time's dependence on n . This was confirmed by simulations. The observed about linear dependence of the rewinding time on DNA length, however, was hard to understand at first sight. On the length scales under consideration (note: the whole DNA wrapped up corresponds to about one persistence length), the DNA should resemble the same rotational behavior as a rod. A rod's rotational friction coefficient depends cubically on the length [26], and since the friction coefficient is the prefactor in barrier crossing, dwell times should depend on DNA length cubically, not about linearly.

Flexibility assisted barrier crossing We found this discrepancy to be explained by an effect we termed *flexibility-assisted barrier crossing* which can nicely be understood using a toy model named *semi-flexible Brownian rotor*. Imagine a semi-flexible polymer in two dimensions which is hinged at the origin and allowed to rotate freely. Additionally, a periodic potential $V(\phi)$ is acting on the attachment angle ϕ making some orientations more probable than others (see Fig. 1b in Sec. 3.6). Barrier crossing, i.e., transition from one minimum in $V(\phi)$ to another, then mimics a rewinding event described above.

Let us for a moment treat this barrier crossing as effectively one-dimensional. Then, Kramers theory [56] may be applied: The rate for barrier crossing is proportional to the probability of being in the transition region (close to the top of the barrier) times the rate with which the particle relaxes away from the barrier.⁸ The former does not depend on the length L of the polymer. Thus, the barrier crossing rate is determined by the time τ the polymer spends at the barrier. Since τ is finite, the polymer not necessarily relaxes completely. Only a polymer part of length l_c gets deformed during relaxation away from the barrier and contributes to friction. The scaling behavior of the new length scale l_c may be determined using a simple argument: First, reiterating that the relaxation time of the polymer scales as L^4/l_p (Eq. (1.4)) one may write $l_c \sim (l_p\tau)^{1/4}$. Second, the deterministic overdamped relaxation away from the barrier at $\phi = 0$ is described by $\dot{\phi} = -\mu(l_c)\partial V/\partial\phi$, where $\mu(l_c)$ is the effective mobility of the polymer with a stretch of length l_c contributing to friction. The mobility scales as $1/l_c^3$ due to the rotational motion. Thus, $\phi/\tau \sim -\phi/l_c^3 \cdot \partial^2 V/\partial\phi^2|_{\text{barrier}}$. After elimination of τ one obtains:

$$l_c \sim \frac{k_B T l_p}{|\partial^2 V/\partial\phi^2|_{\text{barrier}}}. \quad (3.2)$$

The new characteristic length l_c scales with the persistence length l_p and inversely with the potential's curvature at the top of the barrier. For polymers with length L much shorter than l_c , the whole polymer relaxes at the top of the barrier and behaves like a stiff rod. The dwell time in this case scales as L^3 , just as for a rod. For polymers with length $L \gtrsim l_c$, only a segment of length l_c is involved in the actual barrier crossing making the dwell time independent of L

⁸Recall that the Kramers rate for the (overdamped) escape from a well over a potential barrier can be written as $\sim (\omega_b/\omega_t) \exp(-\Delta U/k_B T) \cdot \mu \omega_t^2$, where μ is the particle's mobility, ΔU is the height of the potential barrier, and ω_b^2 and ω_t^2 are the curvatures of the potential at the bottom of the well and the top of the barrier, respectively [36, 39, 56]. The first term states the probability to find the particle at the top of the barrier relative to the bottom of the well (harmonic potential approximations). The second term, the mobility multiplied by the curvature at the top, determines the rate of relaxation away from the barrier. See also Ref. [39, p. 277].

and effectively accelerating barrier crossing. Since (in our case) $l_c \ll l_p$, flexibility accelerates barrier crossing even for polymers which can be treated as rods otherwise.

This reasoning was put on more quantitative grounds by applying Langer theory for multidimensional barrier crossing (reviewed in Ref. [39]) to our model at hand. The result agrees with our considerations above and with Brownian dynamics simulations of the *semi-flexible Brownian rotor*. Simulations also revealed a third regime not discussed so far; for large polymer lengths, $l_c \ll L < l_p$, barrier crossing becomes diffusion limited. In this regime, the dwell time again scales as L^3 . Close inspection reveals that within the coarse-grained nucleosome model we are in the intermediate regime. Barrier crossing involves only a length l_c of the polymer, which explains the only weak dependence of the dwell times on DNA length and thus on n .

Comparison with experiments As mentioned above, measurements of dwell times for site exposure and their dependence on the position of the FRET pair will probably soon be reported. How should the results be interpreted in the light of our theoretical study? Based on our model, we expect only a weak dependence of the open state's dwell time on the length of the DNA involved in rewinding. If such a weak dependence is indeed observed in experiments, should it be attributed to flexibility-assisted barrier crossing or could it simply be explained by the fact that histone-DNA interactions are not the same at all contact points? Complementary experiments can resolve this question. If polymer physics rather than the histone-DNA interaction determines the site dependence of rewinding times, then the same length dependence should be observed when measuring rewinding at the very end of the DNA with varying linker lengths. If, contrary to our expectations, not a weak but a cubic dependence of rewinding times on DNA length will be reported, this very likely will be due to the rotational motion of the DNA involved in site exposure, either due to rod-like barrier crossing or diffusion limited motion of the DNA discussed above.

What about absolute time scales? Rescaling dwell times from simulation time units to real times reveals that the dwell times predicted by our model are orders of magnitude smaller than the dwell times observed in the experiments. Though our model is very coarse-grained (involving DNA modeled as a WLC on small length scales and strong curvatures, negative charges placed along the WLC backbone, etc.) it seems unlikely that more detailed modeling alone could explain the long dwell times observed in experiments. It is much more likely that another effect, not incorporated in our model at all, is responsible for this discrepancy. First, DNA is over- and undertwisted in the nucleosome [72].⁹ Twist of the DNA therefore has to change when attaching to or detaching from the histone octamer. Imagine now that the DNA bends towards the histone octamer, but not with the correct twist needed for a successful binding event. In such a scenario, only a fraction of rewinding attempts would be successful – potentially explaining the large dwell times observed. Second, we did not include hydrodynamic interactions between the two turns of the DNA and between the octamer and the DNA, which may be too simple. It is known that hydrodynamic interactions lead to a dramatic slow-down of a sphere moving towards a wall [15]. Similarly, in a rewinding event, the DNA comes very close to the histone complex. Hydrodynamic interactions therefore are expected to slow down rewinding and could potentially explain the large dwell times observed in the experiments.

⁹On average, DNA is slightly unwound and we estimated the torsional energy of bound DNA to be about $1 k_B T$ for 10 bp and included this effect in the contact point potential.

3.3. Outlook: site exposure in nucleosome arrays

A number of extensions of our model spontaneously come into one's mind. For example, it is not clear whether the dynamics of nucleosomes tethered to a surface as in the experiments by Koopmans et al. [49] can be described by our model and extending the model to such geometries would be a worthwhile endeavor. Also, our study concentrated on DNA unwrapping at one end. It is expected that the electrostatic interaction between the two DNA turns leads to a dynamic (anti)correlation of unwrapping at the two different ends, similar to the effect of negative cooperativity mediated by the nucleosome (see above). Such a correlation could be measured in our simulation and in principle in experiments, too.

Arguably of largest biological interest is the extension from mononucleosomes to multiple nucleosomes. On the experimental side, Poirier et al. recently investigated binding of proteins to DNA in nucleosome arrays [97]. In an assay similar to the early studies more than a decade ago, the authors used restriction enzymes to estimate the extent to which nucleosomes suppress protein binding. They introduced target sites in the middle of a compact array of 17 nucleosomes. Protein binding to DNA buried inside the center nucleosome occurs with probabilities similar to that observed in a mononucleosome. In contrast, binding to the linker DNA is up to about 50-fold suppressed compared to naked DNA. Concerning the biological relevance, the authors nicely conclude: "Thus, nucleosome positioning dramatically influences the accessibility of target sites located inside nucleosomes, while chromatin folding dramatically regulates access to target sites in linker DNA." As with mononucleosomes, a FRET study followed the restriction enzyme study [98]. It was shown that a compact system of three nucleosomes is dynamic and that protein binding can occur to the center nucleosome via site exposure.

These experimental results, and the outlook for more to come, call for a theoretical description. The most obvious approach, modeling the nucleosome array as a succession of our coarse-grained mononucleosome, is not applicable. Our coarse-grained model does not include any nucleosome-nucleosome interactions, which are known to be important. Models for nucleosome arrays including the latter are numerous and range from geometrical considerations (e.g., [25]) to strongly coarse-grained models (e.g., [48, 79]) to detailed mesoscopic models [8, 37]. To my knowledge, none of them allows for unwrapping of DNA, nucleosomes are always modeled as fixed entities with linker DNA in between.¹⁰ A study of site exposure in nucleosome arrays would therefore probably entail extending an existing model which then can in principle be studied using extensive Brownian dynamics simulations or (more efficiently and valid if one is interested in equilibrium constants only) by Monte Carlo methods. Last but not least, one might want to explicitly include proteins in the model to serve as probes for site exposure, but also as entities which decompact the structure by non-specific binding [98].

3.4. Optimal flexibility for rotational barrier crossing

In the last section, we saw how even little flexibility in a polymer can strongly influence its barrier crossing rate. More precisely, due to flexibility, only a part of the polymer is involved in crossing the barrier, effectively reducing the friction and leading to an increased crossing rate. Originally, we had used a simpler model termed the *two-segment lever* to try to understand this phenomenon. Indeed, the two-segment lever captures some of the features

¹⁰There are published studies which include unwrapping of DNA from the histone core (e.g., [65]). However, to my knowledge, none is applicable to model a nucleosome array.

discussed above, but in addition shows richer phenomenology, especially optimal flexibility for barrier crossing. The most important aspects of our study are outlined in the following. For details, I refer to the reprint of our publication [90] in Sec. 3.7.

As the name suggests, the lever consists of two segments, hinged at the origin in two dimensions. The attachment angle ϕ is subject to an external potential $V(\phi)$. The angle θ between both segments is subject to an ‘internal’ potential $\epsilon(1 - \cos\theta)$ introducing stiffness to the system. For large ϵ , both segments are aligned most of the time; for vanishing ϵ , both segments are merely hinged. Friction is generated by two beads at the ends of the two segments (see Fig. 1 in Sec. 3.7). To some extent, the two-segment lever is the discretized version of the semi-flexible Brownian rotor. Both flexibility and mobility are distributed over the whole length of the polymer in the latter while being localized in the former.

Again, we were interested in the rate with which the lever crosses a barrier (in the overdamped limit). For a given potential $V(\phi)$, the crossing rate can easily be determined by numerically solving the Langevin equations describing the system.¹¹ We found that for very large ϵ barrier crossing may be described as a one-dimensional process with the lever behaving as though it were completely stiff. This is in analogy to the semi-flexible Brownian rotor where we also observed rod-like barrier crossing in the stiff limit (i.e., for very short lengths). When ϵ becomes smaller, the numerically determined crossing rates become larger, an effect resulting from flexibility-assisted barrier crossing. In this regime, barrier crossing cannot be treated as one-dimensional. At the transition point, the full potential $V(\phi, \theta)$ exhibits a maximum in the direction of ϕ and a minimum in the direction of θ . The multidimensional generalization of Kramers theory, developed originally by Langer and reviewed in Ref. [39], can again quantitatively account for this effect.

With further decrease of stiffness ϵ , the numerically determined crossing rates display a maximum, contrary to Langer theory which predicts further monotonic increase. The discrepancy can be resolved by recognizing that the mobility matrix of the system strongly depends on the state of the system: when both segments are perpendicular, friction opposing a small change in ϕ is much larger than when both segments are aligned (Fig. 4 in Sec. 3.7). This friction dependence becomes important only at intermediate and small ϵ , where the segments are not aligned almost all the time. In a nutshell, the maximum in the crossing rate is a trade-off between the accelerating effect of bending fluctuations and decreasing average mobility in ϕ . Quantitatively, the maximum can be described when extending Langer theory, which assumes a constant mobility matrix, to the more general case of a configuration-dependent mobility matrix, see Ref. [89, Appendix B]. Note that when studying the semi-flexible Brownian rotor, we assumed an almost straight polymer ($L \ll l_p$) and were not able to account for such an effect.

In summary, the simple two-segment lever shows flexibility-assisted barrier crossing as does the semi-flexible Brownian rotor. In addition, there exists an intermediate or optimal stiffness at which barrier crossing is fastest. For intermediate stiffness, the system is also optimal from another point of view, not discussed so far: the system is also most robust at intermediate stiffness; the crossing rate changes only little when increasing the friction of the outer bead, i.e., when attaching some kind of ‘cargo’ to the lever. These two observations raise the question, whether these effects are exploited, for example, in molecular motors.

¹¹The fixed length of the segments leads to a direct coupling of the two degrees of freedom, the mobility matrix is non-diagonal. Furthermore, the mobility matrix explicitly depends on the state of the system. The general form of the two coupled Langevin equations is outlined in Appendix A.2.

3.5. Conclusion

In the introductory chapter, I presented the nucleosome as the smallest packaging unit of chromatin. Most of the DNA is found inside nucleosomes and the question of whether and how proteins get access to nucleosomal DNA immediately arises. A mechanism which enables protein binding without the use of ATP is site exposure or nucleosome breathing: thermal fluctuations lead to transient unwrapping of the DNA from the histone core which makes DNA stretches available for protein binding.

In this chapter, I reviewed the site exposure mechanism, concentrating on *in vitro* studies of mononucleosomes. While the thermodynamics of site exposure have been well-understood, much less is known about the dynamics of nucleosomal DNA. Our theoretical study, outlined and reprinted here, addresses this issue. Using a coarse-grained model we numerically determined the dependence of dwell times on the length of DNA involved in the unwrapping and rewinding process. Though drastically coarse-grained, we believe our model captures the essential physics of DNA dynamics within nucleosomes. Future experiments will be able to test our predictions while our study may help to interpret the experimental results. Finally, I discussed limitations and possible extensions of our model.

Within our nucleosome model, DNA rotationally crosses a free energy barrier when wrapping or unwrapping from the histone core. To understand the physics underlying the results of the simulation, we employed a toy model named the semi-flexible Brownian rotor. Within this toy model we were able to understand the dependence of dwell times on the length of DNA: flexibility on length scales much smaller than the persistence length facilitates rotational barrier crossing. A similar effect is observed in another toy model we addressed in a separate study, the two-segment lever, where additionally the crossing rate displays as maximum as a function of bending stiffness.

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

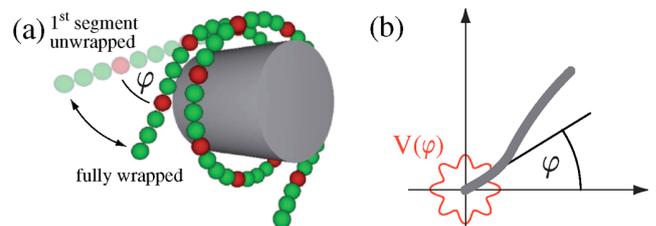


FIG. 1 (color online). (a) Illustration of our nucleosome model. The DNA-histone interaction is localized at contact points attracting the red (dark) beads. The DNA is shown in the ground state as well as a conformation where the first contact is open. (b) Illustration of the semiflexible Brownian rotor (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

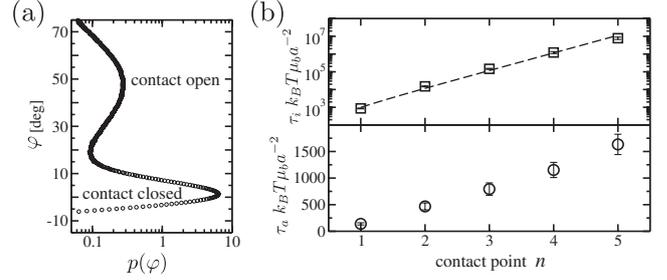


FIG. 2. (a) Equilibrium distribution of the DNA angle φ defined in Fig. 1(a). The two peaks at $\varphi = 0$ and $\varphi \approx 45$ deg correspond to the fully wrapped state and the state with contact 1 open, respectively. (b) Kinetics of DNA site exposure within our nucleosome model. The dwell time 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-

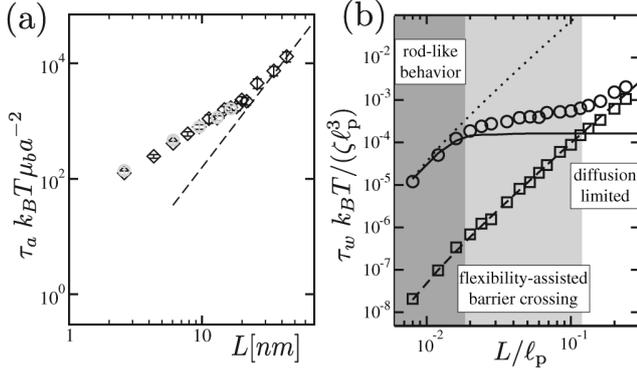


FIG. 3. (a) The dependence of the dwell time $\tau_a(n=1)$ on the overhanging DNA length (diamonds) is compatible with $\tau_a(n)$ when n is converted to contour length (gray open circles). The dashed line indicates the diffusion limit (see main text for details). (b) The average barrier crossing time 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

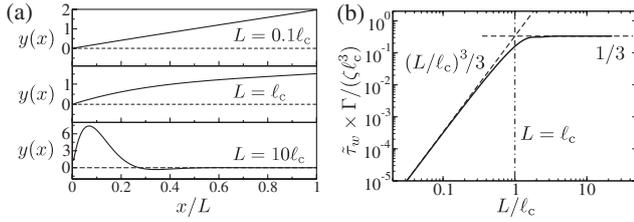


FIG. 4. Dynamics at the barrier. (a) The unstable eigenmode for three different lengths. Polymers shorter than ℓ_c rotate without significant deformation, while long polymers form a bulge of size $\sim \ell_c$ at the origin. (b) The prefactor of the Kramers time $\tilde{\tau}_w = 1/\lambda_-$ as a function 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. Neher,^{1,*} Wolfram Möbius,¹ Erwin Frey,¹ and Ulrich Gerland^{1,2}

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

²Institute for Theoretical Physics, University of Cologne, 50397 Cologne, Germany

(Received 13 April 2007; published 22 October 2007)

Conformational transitions in macromolecular complexes often involve the reorientation of leverlike 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 multidimensional barrier crossing to configuration-dependent mobility matrices.

DOI: 10.1103/PhysRevLett.99.178101

PACS numbers: 87.15.He, 82.20.Db

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 time scale 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 multidimensional transition pro-

cesses 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 first approximation to a more continuously distributed flexibility. 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 [11] from this minimum passes through the

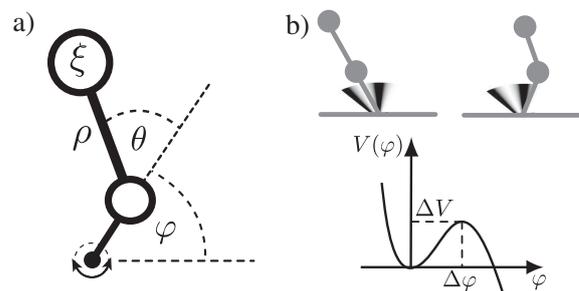


FIG. 1. Schematic illustration of the “two-segment lever” (TSL). (a) The two segments of lengths 1 and ρ are connected by a hinge and attached to the origin. The viscous drag acts on the ends of the segments as indicated by the beads. (b) Schematic illustration of the barrier crossing processes. The external metastable potential $V(\varphi)$ is indicated by shading (top; dark corresponds to high energy) and is also sketched below.

transition state at $(\varphi, \theta) = (b/a, 0)$ with a barrier height $\Delta V = b^3 k_B T / 6$.

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 length of the second segment. We choose our time unit such that the friction coefficient of the first bead is unity, and we denote the coefficient of the second bead by ξ . In general, the derivation of the correct dynamic equations can be non-trivial for stochastic systems with constraints [12,13]. 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 [13]. 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 [14]. 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 [15].

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 [16]. The rate for the conformational transition is related to the dwell time by $k(\epsilon) = 1/\tau(\epsilon)$. Figure 2 shows $k(\epsilon)$ (circles) for a barrier $\Delta V = 12k_B T$, a distance $\Delta\varphi = 0.4$ to the transition state, and $\xi = \rho = 1$ (data for different parameter values behave 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 the dynamics in the stiff limit ($\epsilon \rightarrow \infty$); see the inset. Note that the enhancement persists even at relatively large ϵ , where

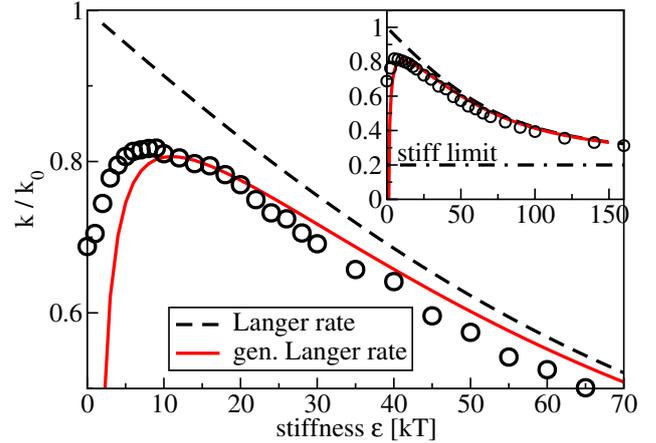


FIG. 2 (color online). Simulation data of the barrier crossing rate normalized by k_0 display a prominent peak at finite stiffness (circles, each obtained from 20 000 simulation runs initialized at the metastable minimum). The conventional Langer theory fails to describe the nonmonotonicity of the rate and overestimates the rate at small ϵ . The generalized Langer theory captures the nonmonotonicity of the rate and describes the simulation data accurately; for parameters, see the main text.

typical thermal bending fluctuations $\delta\varphi \sim \epsilon^{-(1/2)}$ are significantly smaller than $\Delta\varphi$. Surprisingly, the rate is largest 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 transition decreases; see Fig. 3(a). This decrease is most dramatic in the stiff limit (dashed-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. 3(b) shows that the ϵ -dependence of this sensitivity (measured as the slope of the curves in Fig. 3(a) at $\xi = 1$) has a pronounced minimum at $\epsilon \approx 20$. Hence, intermediate hinge stiffnesses in the TSL lead to

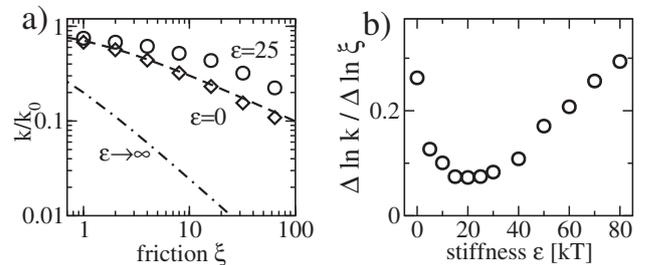


FIG. 3. The sensitivity of the rate to the friction coefficient ξ is minimal at intermediate stiffness. (a) Simulation results at $\epsilon = 0$ and $\epsilon = 25$ as well as the theoretical estimates of the rate at $\epsilon = 0$ and in the stiff limit. (b) The derivative of $\ln k$ with respect to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

maximal robustness, which is an important design constraint for many biomolecular mechanisms.

It is instructive to consider simple theoretical bounds on the transition rate. An upper bound is obtained by completely eliminating the outer bead. The Kramers rate [17] 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 the rotational friction by a factor $\zeta = 1 + (1 + \rho)^2 \xi$, so that the 1D Kramers rate becomes $k_\infty = k_0 / \zeta$; see the dashed-dotted line in Figs. 2 and 3(a). However, to fully understand the above phenomenology, we must consider the coupled dynamics of barrier crossing and bending. The multidimensional generalization of Kramers theory is Langer's formula for the escape rate over a saddle in a potential landscape [10],

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

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 $\mathbf{e}^{(s)}$. Equation (4) can be made plausible in simple terms: Given a quasiequilibrium in the metastable state, the determinants and the Boltzmann factor represent the probability of being in the transition region. 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 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 $\sim \epsilon^{-1/2}$ are small compared to the width of the barrier $\sim \gamma^{-1/2}$. 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, which 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 , 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. 3(a), indicating that the configuration-dependent mobility (2) plays an important role for the transition rate. In contrast, conventional Langer theory assumes a constant mobility matrix near the transition state. Figure 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, 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 [18]; 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 [17] for the conventional Langer theory. We use the ansatz $p(\{q_i\}) = \frac{1}{2} p_{\text{eq}}(\{q_i\}) \text{erfc}(u)$, where $p_{\text{eq}}(\{q_i\}) = Z^{-1} e^{-V(\{q_i\})/k_B T}$ 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} ,

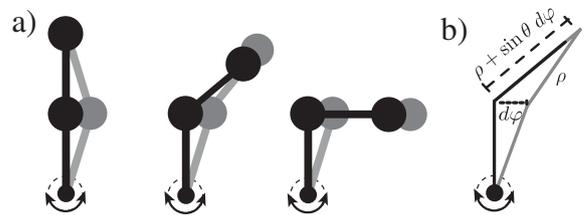


FIG. 4. (a) The friction opposing rotation of the attachment angle φ depends on the bending angle θ , since the outer bead is moved by different amounts in different configurations. (b) For an infinitesimal displacement $d\varphi$, the displacement of the outer bead is $\sin\theta d\varphi$. The projection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

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

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 [19].

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

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

where $c = U_i e_{ij}^{-1} U_j + 1 = B_{1i} e_{i1}^{-1} / M_{11}^{(s)}$ and \mathbf{e}^{-1} denotes the inverse matrix of $\mathbf{e}^{(s)}$. Equation (7) contains three corrections to (4), all of which vanish when $\mathbf{M}(\{q_{ij}\})$ is constant: The most important one is given by $\sum_{l>1} A_{1l}^{ll} / \mu_l$, 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 Eq. (7) to the TSL. It captures the peak in the transition rate and thus the essential phenomenology of the TSL [20].

Discussion.—We introduced the “two-segment lever” as a simple model for a class of conformational transitions in biomolecules and derived a generalized Langer theory to understand its behavior. The model clearly demonstrates how flexibility can enhance the rate of a conformational transition. This remains true when the hinge in the TSL is replaced by a more continuous bendability [4]. Interestingly, a discrete hinge has a stiffness regime where the transition rate is both large and robust. It is conceivable that this effect is exploited by evolution, for example, in the design of hinged molecular motors [6,7]. A promising candidate to test these ideas is myosin II, where mutations affecting the stiffness of the converter region and the activity of the motor are known [22,23]. Whether flexibility assisted barrier crossing is important in processive motors is difficult to test, since the rate of the conformational transition has to be considerably faster than the

unbinding rate from the filament. However, the mechanism could very well be exploited to establish this hierarchy of rates. We hope that our work will stimulate further experiments to clarify these intriguing questions.

We thank Andrej Vilfan for fruitful discussions and the German Excellence Initiative for financial support via the program NIM. R. N. and U. G. are grateful for the hospitality of the CTBP at UCSD, where part of this work was done, and for financial support by the DFG and the CeNS.

*neher@kitp.ucsb.edu

- [1] J. Howard, *Mechanics of Motor Proteins and the Cytoskeleton* (Sinauer Press, Sunderland, MA, 2001).
- [2] S. Sugimura and D. M. Crothers, Proc. Natl. Acad. Sci. U.S.A. **103**, 18510 (2006).
- [3] G. Li, M. Levitus, C. Bustamante, and J. Widom, Nat. Struct. Biol. **12**, 46 (2005).
- [4] W. Möbius, R. A. Neher, and U. Gerland, Phys. Rev. Lett. **97**, 208102 (2006).
- [5] X. Zhuang *et al.*, Science **296**, 1473 (2002).
- [6] M. Terrak *et al.*, Proc. Natl. Acad. Sci. U.S.A. **102**, 12718 (2005).
- [7] S. Jeney *et al.*, Chem. Phys. Chem. **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. (N.Y.) **54**, 258 (1969).
- [11] With the potential (1) this transition is irreversible; however, all of our conclusions equally apply to reversible transitions in a double-well potential.
- [12] E. Helfand, J. Chem. Phys. **71**, 5000 (1979).
- [13] N. van Kampen and J. Lodder, Am. J. Phys. **52**, 419 (1984).
- [14] H. Goldstein, C. P. Poole, and J. L. Safko, *Classical Mechanics* (Addison Wesley, Reading, MA, 2002).
- [15] C. W. Gardiner, *Handbook of Stochastic Methods* (Springer-Verlag, Berlin, 2004).
- [16] In our simulations, the particle has escaped when $\varphi > 0.7$. Results are independent of the threshold value, as long it is large enough to exclude recrossing events.
- [17] P. Hänggi, P. Talkner, and M. Borkovec, Rev. Mod. Phys. **62**, 251 (1990).
- [18] B. Gavish, Phys. Rev. Lett. **44**, 1160 (1980).
- [19] A. Berezhkovskii and A. Szabo, J. Chem. Phys. **122**, 014503 (2005).
- [20] Obviously, the evaluation of the Gaussian integral leading to Eq. (7) is meaningful only if the harmonic approximation of the mobility matrix is adequate within the saddle 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 [21].
- [21] A. N. Drozdov and P. Talkner, J. Chem. Phys. **105**, 4117 (1996); A. M. Berezhkovskii and V. Y. Zitserman, Chem. Phys. **157**, 141 (1991).
- [22] H. Yamashita *et al.*, J. Biol. Chem. **275**, 28045 (2000).
- [23] J. Kohler *et al.*, Proc. Natl. Acad. Sci. U.S.A. **99**, 3557 (2002).

4. Nucleosome organization in yeast

The presence of nucleosomes and higher order structures leads to strong compaction of eukaryotic DNA. This immediately poses the questions: How can DNA transcription, replication, and repair nevertheless occur, and, importantly, how can regulatory proteins passively bind to DNA? A number of mechanisms by which access to DNA may be granted have been mentioned or discussed in the preceding chapters: active remodeling (e.g., the shifting or eviction of nucleosomes), passive sliding of nucleosomes along the DNA, and the transient unwrapping of DNA from the histone core. Another possibility of ensuring access to certain locations on the genome is through the use of nucleosome positioning. A precise arrangement of nucleosomes can control access to binding sites of regulatory proteins. This is illustrated by the fact that most functional binding sites of transcription factors are free of nucleosomes [156]. However, nucleosome positioning can influence gene regulation also by other means than by preventing or allowing access to binding sites [102]. One example of this is the cooperative interaction between two transcription factors mediated by a nucleosome, which requires the nucleosome to be positioned with respect to the two binding sites as explained in Sec. 3.1.2. Or, as another example, a positioned nucleosome may bring two protein binding sites close together by wrapping up the DNA in between [102].

Until a few years ago, nucleosome positions have been known only at a limited number of genomic locations. In the last few years, however, knowledge about genome-wide nucleosome organization has dramatically increased. It has been found that a significant fraction of nucleosomes are positioned and their locations have been determined at high resolution. Large-scale or genome-wide nucleosome positions have been determined for the yeast *Saccharomyces cerevisiae* (e.g., [2, 61, 77, 129, 156]), the yeast *Candida albicans* [32], the yeast *Schizosaccharomyces pombe* [60], the fly *Drosophila melanogaster* [78], the worm *Caenorhabditis elegans* [142], the fish *Oryzias latipes* [119], and humans [123].¹ These maps of nucleosome positions either include or have been accompanied by maps of histone variants and histone modifications [2, 67]. The field is developing so fast that even last year's reviews [46, 105, 135] on nucleosome organization will probably soon become out-dated in parts.

Due to the outstanding importance in gene regulation, special emphasis has been put on nucleosome organization close to start sites of genes. How nucleosomes are organized there and what determines their positions are questions of special interest. In this chapter, I will focus on the question of the extent to which nucleosome organization around start sites of genes can quantitatively be described by a one-dimensional gas of hard rods in *S. cerevisiae* (hereafter referred to as 'yeast').

In order to address this question and discuss it in the context of the literature, a more extended introduction is necessary. Without attempting to give a comprehensive overview, I will first outline how nucleosome positions are determined experimentally and continue with a description of the canonical organization of nucleosomes around promoters. Afterwards, I will discuss nucleosome positioning determinants with focus on the relationship between

¹For *Saccharomyces cerevisiae* a reference map including and comparing data from different preceding studies has been published recently [45].

nucleosome positions and the underlying DNA sequence. Finally, I will turn to statistical positioning and explain how preferred nucleosome positions can emerge solely from the interplay between entropy and the finite width of nucleosomes. This concept will be made quantitative within the Tonks gas model (the aforementioned one-dimensional gas of hard rods), which we used to quantitatively test the statistical positioning scenario [85]. Finally, I will discuss limitations and possible extensions of our work and conclude with a proposal for a new approach towards better understanding nucleosome organization.

4.1. Methods to determine nucleosome positions

Within the last few years, genome-wide nucleosome positions have been determined at continuously increasing resolution. This has been achieved by two different experimental methods, denoted *ChIP-Chip* and *ChIP-Seq*, which share similarities, but also differ in important aspects.² In order to discuss our data-driven study below, it is useful to shortly introduce and discuss the experiments determining nucleosome positions.

In a first step, which is common to both methods, DNA buried inside nucleosomes is extracted. To that end, histones and the surrounding DNA are cross-linked and thus fixated in living cells. The resulting string of nucleosomes on the DNA should reflect the *in vivo* situation. The actual extraction of nucleosomal DNA is achieved by a DNA digesting enzyme (micrococcal nuclease, short: MNase). Since DNA inside nucleosomes is protected, MNase preferentially digests the linker DNA between nucleosomes, and the string of nucleosomes is reduced to individual and unconnected nucleosomes. At this point in the experiment, *chromatin immunoprecipitation* (*ChIP*) may be applied. Nucleosomes including histone variants or nucleosomes with certain modifications are selected using antibodies. This selection makes it possible to create maps of histone variants or histone modifications (essentially maps of subsets of nucleosomes) just like maps of nucleosome positions we are mainly interested in here. Furthermore, ChIP helps to ensure that it is indeed nucleosomes, and not other DNA bound proteins, that have been extracted by fixation and MNase digestion. Finally, the DNA is extracted and one is left with a huge number of small pieces of DNA. A method called gel electrophoresis is used to select those DNA fragments whose size corresponds to the DNA in mononucleosomes.

In a second step, the extracted DNA has to be mapped to its original genomic location in order to identify nucleosome positions. This is achieved by either hybridization on a chip (therefore the name ChIP-Chip) or sequencing (ChIP-Seq). The first studies on nucleosome organization in yeast used the hybridization approach. Sequencing has now become more popular due to higher achievable resolution and continuously decreasing costs [105].

In the hybridization technique, genomic DNA is split up in overlapping parts and distributed on a tiled chip called microarray. Every tile on the chip thus contains a probe which represents a certain location on the genome. The nucleosomal DNA fragments are hybridized against these probes. By measuring the amount of hybridization at the different parts of the chip one obtains a quantitative measure for which parts of the genome are occupied by nucleosomes.

²See Ref. [135] for an overview of the experimental methods employed in the different studies determining nucleosome positions at high resolution.

Alternatively to the hybridization approach, the nucleosomal DNA is sequenced and the resulting sequences are mapped to the reference genome using alignment algorithms. Since sequencing occurs from the 5' ends of the DNA fragments, the 5' ends of the mapped sequences represent the borders of the original nucleosomes. By merging this information about nucleosome borders from both the Watson and the Crick strand, nucleosome midpoints may be inferred. As a result, one obtains a histogram of (putative) nucleosome positions as a function of the genomic location.³ It is this kind of data we used in our study discussed further below.

Both experimental methods reveal structure in their raw data (amount of hybridization, histogram of putative nucleosome positions), but positions of individual nucleosomes cannot be directly inferred from the data. This most likely is due to the fact that the experiments are performed on ensembles of cells, not on individual cells. Hence, the raw data are expected to include cell-to-cell variation in nucleosome organization. Often, bioinformatic approaches such as Hidden Markov Models [156] or peak detection [2] are applied to identify individual nucleosomes and their locations based on the raw data. These locations should be seen as something like consensus positions of nucleosomes in ensembles of cells. Consensus positions are useful when discussing the biological relevance of individual nucleosomes, such as the closest nucleosome to a gene start. However, as previously mentioned, consensus positions neglect cell-to-cell variation.

4.2. Nucleosome organization close to start sites of genes

Proteins are omnipresent components of the cell and are encoded by the cell's DNA. The conversion of genetic information into proteins involves a number of different processes which may be summarized by transcription and translation. First, the gene coding for the protein is transcribed into RNA (ribonucleic acid), which in a second step is then translated into the protein. Since DNA is packaged into nucleosomes, transcription is naturally related to the organization of nucleosomes on the genome, the subject of this chapter.

In eukaryotes, the readout of the DNA sequence is performed by an enzyme called RNAP II (RNA polymerase II).⁴ Before transcription can occur, RNAP II must bind to the DNA at the promoter of the gene, a region 'in front' of the DNA sequence to be transcribed. By convention, one says the promoter is located *upstream* of the transcribed region and transcription itself occurs in the *downstream* direction (5' to 3' direction of the coding strand). The binding of RNAP II to the promoter is guided by the DNA sequence in the promoter region and requires a number of proteins called general transcription factors. Together with other proteins, RNAP II and the general transcription factors form the pre-initiation complex. After this complex has been formed, transcription is initiated which, as the name suggests, starts at the *transcription start site* (TSS).

Transcription is controlled by gene regulatory proteins (transcription factors) which may enhance or diminish the transcription rate. They bind directly to the DNA or bind to other protein complexes. Regulatory proteins can function by directly influencing transcription initiation as well as by recruiting histone modifiers or chromatin remodelers (which in turn control accessibility of DNA). Binding sites for gene regulatory proteins are located in regulatory regions of the DNA, which, for higher eukaryotes may be far removed from the gene controlled. In yeast, control regions are mostly located at the promoter region.

³For more details see, e.g., Box 1 in Ref. [46].

⁴In eukaryotes, there are different versions of RNA polymerase (I-III) transcribing different kind of genes. For simplicity, we limit our discussion to protein encoding genes, which are transcribed by RNAP II [3].

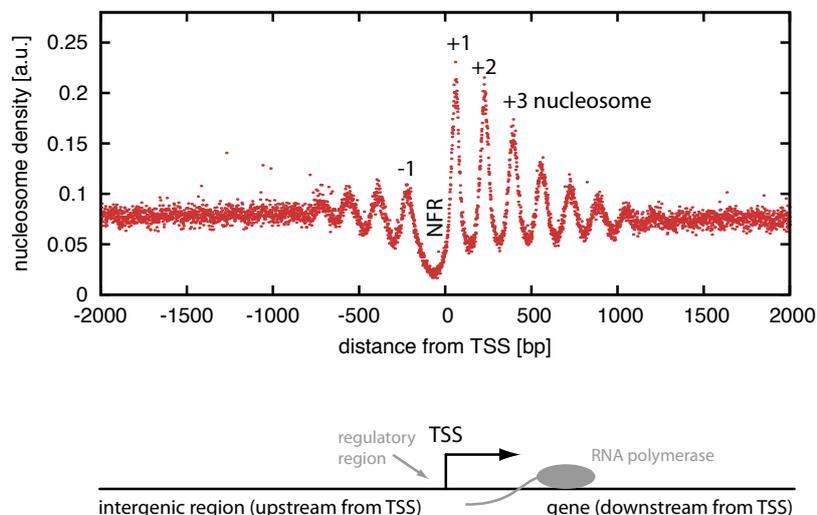


Figure 4.1.: Consensus nucleosome organization close to transcription start sites (TSS) of *S. cerevisiae*. Two features are remarkable and often discussed: First, nucleosome density displays a pronounced trough just upstream of the TSS, commonly referred to as the nucleosome-free region (NFR) and flanked by the -1 and $+1$ nucleosomes. Close to the NFR, nucleosome density displays pronounced peaks ranging about 1000 base pairs (bp) downstream of the TSS. (The plot is an average over many different TSSs and has been created like Fig. 1 in Sec. 4.7 as explained in the ‘Materials and Methods’ part there. The alignment, however, has been performed to TSSs instead of to $+1$ nucleosome positions, and the read density directly is interpreted as a nucleosome density here.)

In conclusion, transcription initiation and regulation involve the concerted action of many different proteins at the promoter. Consequently, the studies reporting on genome-wide nucleosome positions have paid special attention to promoter regions. In the last few years, an overall picture on nucleosome organization close to start sites of *S. cerevisiae* genes has emerged. Two reviews [46, 105] are the basis for the following discussion.

Fig. 4.1 displays the consensus situation of nucleosome organization around TSSs. First, just upstream of the TSS, there is a clear trough in nucleosome density, as wide as about one nucleosome. This feature commonly is referred to as the nucleosome-free region (NFR, sometimes also called nucleosome-depleted region).⁵ Second, nucleosome density displays distinct peaks which range up to about 1000 bp downstream of the TSS, i.e., nucleosomes take on characteristic positions downstream of the TSS. The origin of this positioning is the main topic of our study discussed below. For now, however, let us return to the NFR. It appears that a nucleosome is missing there, leading to an increased accessibility of DNA and facilitating binding of RNAP II. It has been found, however, that the existence of the NFR does not imply that the gene necessarily is transcribed. The two nucleosomes flanking the NFR commonly are referred to as the -1 nucleosome (upstream) and $+1$ nucleosome (downstream). Both are thought to be evicted during transcription or transcription initiation. Close to the TSS, and confined to the range of a few nucleosomes (specifically the $+1$ and -1 nucleosomes), the occurrence of the histone variants (specifically H2A.Z / Htz1) is increased as is the level of histone modification. The nucleosome organization around the NFR is related to

⁵There also typically is a nucleosome-free region at the gene end. The NFR close to the TSS is referred to as the 5' NFR while the NFR at the end of the gene is called the 3' NFR.

sequence. The +1 nucleosome is often associated with a strong nucleosome favoring sequence. The NFR typically incorporates poly(dA:dT) tracts which are hard to be accommodated for in nucleosomes, effectively leading to a sequence-dictated nucleosome depletion. It is known that binding of regulatory proteins also contributes to the creation of the NFRs.

Having described the consensus (or typical) situation of nucleosome organization close a TSS, it is now appropriate to put this into perspective with a few examples. First, nucleosome organization varies from gene to gene, the scenario depicted in Fig. 4.1 does not apply equally to all genes. Lee et al. clustered regions around promoters according to nucleosome occupancy and reported four different groups which were enriched by genes with similar function [61].⁶ Second, remodeling complexes are involved in creating the observed pattern, but do not act the same way at all genes. Whitehouse et al. found that the remodeler Isw2 leads to a shift of nucleosomes towards the intergenic region in a subset of genes [149], and Hartley and Madhani observed a shrinking (but not elimination) of the NFR in about half of all genes when depleting a subunit of the RSC remodeler [40]. Third, nucleosome organization also depends on the physiological conditions the cell is exposed to. Heat shock, a perturbation causing genome-wide transcriptional changes, does not drastically alter nucleosome organization at promoters. At some promoters, however, eviction, appearance or repositioning of one or two nucleosomes was observed [129]. Similarly, an upshift of glucose (a nutrient for yeast) affects nucleosome arrangement at some promoters (though not nearly at all which are affected by the transcriptional change) [157].

Important complementary information about nucleosomes close to TSSs and their influence on gene regulation is obtained from studies at individual promoters. Systematic experiments, e.g., modifications of promoter sequence, make it possible to disclose the underlying mechanisms for the observed nucleosome organization. For example: The *HIS3-PET56* promoter region shows a distinct nucleosome-free region. This nucleosome-free region could be recapitulated *in vitro* indicating that the DNA sequence is responsible for the observed nucleosome organization [128]. The case is different at the extensively studied *PHO5* promoter. In the uninduced state of the gene, nucleosomes are closely packed at the promoter and factors other than DNA sequence are responsible for this organization. In the induced state, however, nucleosomes at the promoter are lost [104, 105, and references therein]. As one last example, the chromatin remodeler Isw2 has been shown to continuously reposition nucleosomes onto an unfavorable sequence, thereby repressing the gene *POT1*. In a mutant lacking Isw2, nucleosome organization is very similar to the case where *POT1* is activated [150].

4.3. Nucleosome positioning determinants

The general rules and specific examples discussed above indicate that a wide range of different determinants influence nucleosome positioning. Among them are DNA sequence, the presence of transcription factors, chromatin remodelers, RNA polymerase, and ‘statistical positioning’ (to be addressed below) [102, 127]. Note however, that the action of the different positioning factors may be indirect. Transcription factors, for example, can influence nucleosome positions by competing for binding sites. But they could also recruit a chromatin remodeler that then influences nucleosome positions.

⁶More generally, one thinks that ‘housekeeping’ genes follow the canonical nucleosome organization at the promoter while ‘stress-responsive’ genes tend to deviate from this structure with their promoters being more often covered by nucleosomes, see Ref. [102] and therein.

To what extent the different determinants contribute to nucleosome organization is not clear, subject to intensive research, and strongly under debate. In this section, I would like to concentrate on the influence of sequence on nucleosome positioning for two reasons. First, the role of sequence is under debate, and hence I would like to outline some aspects of it. Second, computational models exist which predict nucleosome organization based on the underlying sequence. These models are complementary to our analysis presented later in this chapter and for this reason are discussed here in more detail.

As outlined in Sec. 1.3, nucleosome binding affinity and thus nucleosome positions are known to depend on the underlying DNA sequence. Specifically, there exist natural sequences which bind significantly better than random DNA. However, these natural sequences are outperformed by artificial sequences with an even larger binding affinity [134]. This indicates that there is no evolutionary pressure towards maximal binding affinity. Moreover, at least 95 % of bulk DNA sequences have an affinity which is similar to that of randomly synthesized DNA indicating that nucleosomes *in vivo* are not positioned one by one [69].

These results suggest that not the whole possible range of sequence-dictated nucleosome positioning is exploited in the genome. However, these observations do not satisfactorily answer the question of the extent to which sequence prescribes nucleosome organization, or, in other words, whether a *genomic code* [125] exists. In the last few years, this question has been addressed by two different means: by modeling of binding affinity and nucleosome positions as well as by comparisons of nucleosome organization *in vitro* and *in vivo*. Both approaches and their main results will be outlined in the following, for more and complementary information see Ref. [135].

4.3.1. Models predicting nucleosome positions based on the DNA sequence

If it were possible to successfully predict the majority of nucleosome positions from sequence alone, this would imply a major role of sequence on positioning nucleosomes genome-wide. With this motivation a number of studies attempted to predict binding affinities and nucleosome positions. These studies are successful in the sense that they explain a statistically significant fraction of nucleosome positions, i.e., they perform better than a null model assuming random distribution. Although a complete overview and comparison of the different models is beyond the scope of this thesis, I nevertheless would like to outline some important points in the next few paragraphs.

Arguably the best well-known models rely on some kind of pattern matching and are of probabilistic nature (e.g., [20, 33, 44, 47, 125, 155]). The models all use sequences of well-bound nucleosomes (and some of them also stretches of linker DNA) as training data and score a new, unknown sequence based on this input. The most well-known model of this type has been developed by Segal and coworkers [125]. It is based on a *position-specific scoring matrix* (or *position weight matrix*) which includes the frequency of the various dinucleotides at the different positions within a set of well-bound nucleosomes. This procedure is analogous to the approaches developed to predict transcription factor binding sites [131]. The choice to build the model on the occurrence of dinucleotides (instead of on mononucleotides as in the case of transcription factors) is manifest: dinucleotides are the shortest sequence elements governing DNA bending properties which in turn have strong influence on binding affinity (see Sec. 1.3). An extension of this model includes the preference for or aversion to 5-mers (such as AAAAA) in nucleosomes compared to linker regions [33, 47]. This way, the effect of nucleosome-repellent sequences (e.g., poly(dA:dT) tracts) can explicitly be taken into account.

Shortly after the seminal work by Segal and coworkers, Ioshikhes *et al.* employed the correlation of a given sequence to an empirically determined pattern of AA and TT dinucleotides in nucleosomes to predict *nucleosome positioning sequences* [44]. Additionally, comparative genomics was used to reduce noise, i.e., averaging over orthologous locations in related species was performed. Yuan and Liu aimed to include the periodicity of dinucleotide preference in nucleosomal DNA rather than mere over- or underrepresentation at different locations. To that end, they employed a wavelet analysis for scoring unknown sequences for their binding affinity [155]. Chung and Vingron constructed and compared different models based on dinucleotide distribution and/or GC content and found a combined model to perform best [20].

The various models differ in methodology and, by construction, can only capture some aspects of the sequence rules or nucleosome positioning signals. The same holds true concerning the training data. The early studies were only able to use a few hundred sequences of nucleosomal DNA [44, 125] or nucleosomal DNA combined with linker DNA [155] while more recent work relies on genome-wide maps of nucleosome positions *in vivo* [20, 33] and *in vitro* [47]. Equivalence of these training data sets is not a given, *a priori*. If only a few hundred stable and well-bound nucleosomes are used, important positioning signals may be missed, especially nucleosome-repellent sequences. The use of genome-wide *in vivo* maps of nucleosome positions as training data may effectively lead to training of sequence dependence of other factors that determine nucleosome positions. Possibly the best training sets are *in vitro* nucleosome maps with low numbers of nucleosomes excluding possible influences of excluded volume effects between neighboring nucleosomes (see below). As a rule of thumb, including nucleosome-repellent sequences in the training set (and, of course, in the model) significantly enhances predictability [33, 47, 155].

An alternative class of models is more biophysical and based on DNA mechanics (e.g., [87, 140]). In a nutshell, these models rely on the fact that the DNA is greatly distorted while it is wrapped up in a nucleosome and that this distortion to a large extent determines the binding affinity. The energetic cost of this distortion is computed for a given sequence. To that end, one represents the DNA as a rigid body with local angles (tilt, roll, twist) (and possibly local displacements (rise, shift, slide)). An elastic potential is used to describe the free energy cost when deviating from the intrinsic conformation using empirically determined parameters. One problem with those approaches is that the path of the DNA in the nucleosome is itself sequence-dependent and *a priori* not known for a given sequence. Allowing deviations from a defined path is one approach to circumvent this obstacle [87]. Moreover, the models rely on local DNA properties estimated for short sequences of base pairs, making it hard to implement the properties of longer sequences. On the other hand, however, these biophysical approaches do not rely on training data of nucleosome positioning and represent an attractive complementary alternative to the bioinformatic models described above.⁷

So far, I have mainly outlined how the different models score sequences with respect to their binding affinity but have not discussed how nucleosome positions predicted. Since nucleosomes occur at high densities, their finite width plays an important role. For example, two strong binding sites close together cannot be bound concomitantly. The different studies use different approaches, most importantly: filling up the genome with nucleosomes starting from highest scores [20, 44] and thermodynamic modeling equivalent to the grand-canonical ensemble [33, 47, 87, 125, 140]. From a physics perspective, the latter is the most realistic

⁷The biophysical models in general use empirical parameters which do not originate from nucleosomes. This seems not to be true for the model by Vaillant *et al.* [140], see Ref. [135].

approach since it takes thermal fluctuations (and thus cell-to-cell variations) into account. On the downside, the latter approach requires adjustment of additional parameters, namely the energy scale (temperature) and average occupancy (chemical potential).⁸

How do these models succeed in predicting large-scale nucleosome organization? Unfortunately, the various studies also differ in their measures of how the model's outcomes are compared to the experimentally observed nucleosome positions – a discrepancy which makes the field even more complex. The measures range from determining correlations between measured and predicted nucleosome occupancies to reporting the fraction of nucleosomes correctly predicted within a given error. Details aside, let me conclude with more general statements. It is commonly agreed upon that models are able to explain a statistically significant fraction of nucleosome positions genome-wide, indicating relevance of sequence on nucleosome organization. Yet, the notion of a *genomic code* dictating the majority of nucleosome positions is not commonly shared. Let me finally note that the majority of the different models predict two features close to the promoter: depletion of nucleosomes upstream of the TSS (i.e., creation of a NFR) and a strong pronucleosomal sequence at the +1 nucleosome position.

4.3.2. In vitro – in vivo comparisons

The question of the extent to which sequence dictates nucleosome positioning may also be addressed using experimental techniques only, namely by comparing nucleosome organization *in vitro* and *in vivo*. *In vitro*, nucleosomes are assembled using core histones and genomic DNA under appropriate salt conditions. The observed nucleosome organization thus is governed by sequence, excluded volume and nucleosome-nucleosome interactions only. *In vivo*, by contrast, nucleosome organization is additionally influenced by other factors like chromatin remodelers. An example for an *in vitro*–*in vivo* comparison is the work by Sekinger et al. [128], where the authors showed that the *in vivo* depletion of the *HIS3-PET56* promoter can be recapitulated *in vitro* suggesting an important role of sequence at this genomic locus.

Corresponding genome-wide studies recently were reported by Kaplan et al. [47] and Zhang et al. [159]. Both studies agree on the important point that there is nucleosome depletion just upstream of TSSs ([47, Figure 4a],[159, Figure 2]), i.e., the NFR is at least partially hard-coded into sequence. Yet, depletion is less pronounced *in vitro* than *in vivo* indicating that other factors than sequence are also important in establishing the NFR. Zhang et al. also focused on the +1 nucleosomes and found them to be strongly positioned with respect to the TSSs *in vivo*, but not *in vitro*, and concluded that sequence does not suffice to explain the strong positioning.

The two studies came to different conclusions concerning the general importance of sequence on nucleosome organization. Kaplan et al. argue for “The DNA-encoded nucleosome organization of a eukaryotic genome” while Zhang et al. state that “Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions *in vivo*”. This discrepancy is probably due to different foci of experiments, different kind of analyses, and probably also due to different experimental conditions in the *in vitro* experiment. A careful and robust parameter-free reanalysis of the data from both experiments seems necessary before a fair conclusion can be made.

⁸To my understanding the model by Segal and coworkers [125] (and subsequent studies) do not carefully consider the parameter corresponding to the temperature. A recent physics study rephrases the model in physics terms and discusses the effect of both temperature and chemical potential [124].

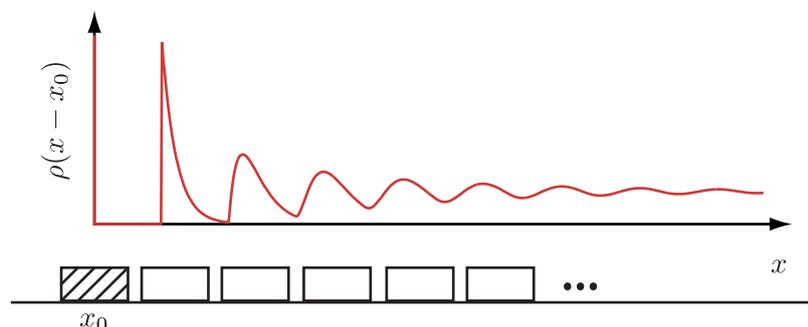


Figure 4.2.: Particle density ρ for a semi-infinite system of hard rods, close to a boundary particle located at x_0 . $\rho(x - x_0)$ is vanishing close to the wall since two particles may not overlap. The oscillations in vicinity of the boundary arise solely from the interplay between excluded volume and entropy. The distances between amplitudes and the decay depend on the mean density of particles. For dense systems (particles tightly packed), the oscillations are very pronounced and decay slowly. Note first that the peaks in the density indicate that particles are preferentially found at specific locations, in other words, they are on average ‘positioned’, just by statistical packing principles. Note second that the expected positions of the particles (indicated by the boxes underneath the graph) do not coincide with the peaks of $\rho(x - x_0)$.

Let me conclude this section by alluding to two recent reviews which address the role of sequence in nucleosome positioning [102, 127]. Both reviews comment on the success of computational modeling of nucleosome positions and the outcomes of *in vitro* – *in vivo* comparisons. The different points of views taken in these two reviews are a nice illustration of how much sequence-dictated nucleosome positioning is under debate.

4.4. Statistical positioning

4.4.1. Statistical positioning from a biological perspective

An additional mechanism by which nucleosomes may be positioned has not explicitly been discussed so far. It is known as *statistical positioning* and is commonly attributed to Roger Kornberg who suggested it in 1981 [52]. Kornberg proposed that the majority of nucleosomes are found at random positions under the condition that they neither overlap with each other nor with other DNA bound molecules. This leads to a phasing of nucleosomes relative to each other, i.e., to the generation of arrays of nucleosomes. A boundary on the DNA, such as a sequence-specific binding protein, ‘anchors’ such an array. Consequently, phasing relative to the boundary is observed, which, on average, appears as a non-random organization of nucleosomes. This ordering of nucleosomes extends over the range of several nucleosomes; for an illustration and qualitative discussion see Fig. 4.2. The concept of statistical positioning was able to reconcile a number of experimental findings of the time, including the puzzling observation of varying nucleosome spacing in different cell types (see Refs. [52, 54] and therein).

A couple of years later, the idea behind statistical positioning was made quantitative by Kornberg and Stryer [54]. Using combinatorial arguments, the authors explicitly computed

the probability that a site on the DNA is bare, i.e., not occupied by a nucleosome, close to one or between two boundaries. Such a quantitative discussion of statistical positioning, rephrased in terms of a much older physical model, will follow in the next section. Here, I would like to concentrate on the qualitative aspects of statistical positioning and discuss related experiments.

To test the statistical positioning hypothesis qualitatively and locally, Fedor *et al.* characterized nucleosome arrays close to a regulatory region in an early study [31]. The authors altered the DNA sequence adjacent to the regulatory region. This had little effect on nucleosome organization indicating that the regulatory region rather than the underlying sequence was responsible for nucleosome positioning. Furthermore, the nucleosome arrays depended on the existence of a specific binding site within the regulatory region. Taken together, both observations are consistent with statistical positioning in proximity to the binding site, but not with positioning of nucleosomes by the underlying sequence.

Very recently, an innovative experiment tested statistical positioning on a small scale *in vitro* [82]. Milani *et al.* considered an approximately 600 bp long DNA fragment containing a yeast gene. A model predicting nucleosome binding affinity based on sequence [140] revealed two nucleosome excluding regions at the end and a rather uniform region without strong pronucleosomal sequences in between. Milani *et al.* directly tested the statistical positioning scenario by loading either one or two nucleosomes onto the DNA fragment and determining positions of the nucleosomes using atomic force microscopy imaging. In the case of the mononucleosome construct, the nucleosome was preferentially found in the central region. In the case of the dinucleosome construct, two distinct peaks in nucleosome density were observable as expected for the predicted binding landscape with two excluding barriers at the gene's ends. Since the *in vivo* nucleosome occupancy [61] has the same shape, the authors argued that statistical positioning is responsible for the nucleosome organization *in vivo*.

What about the situation genome-wide? Are the majority of nucleosomes placed at random positions? Do the nucleosomes appear positioned only on average, due to a few boundaries distributed over the genome? A systematic variation of DNA sequence, as done by Fedor *et al.* [31] for one individual location, clearly is not applicable for testing this hypothesis genome-wide. Recognizing that the period and the decay of the oscillations depend on the average nucleosome density offers an alternative experimental test, at least for the situation *in vitro*. One should vary the average nucleosome density on the genome and test whether the spacing in the nucleosome arrays changes as expected. Kaplan *et al.* [47] and Zhang *et al.* [159] chose different nucleosome densities in their *in vitro* studies of nucleosome organization. A comparison of both data sets could be the first step towards a systematic analysis of this kind.

Due to the lack of systematic experimental tests and the limited practicability of changing nucleosome concentrations *in vivo*, the question of whether statistical positioning plays a significant role in nucleosome organization has to be addressed indirectly with modeling. Obvious candidates for the boundaries inducing nucleosome arrays are the NFRs close to the 5' end of genes discussed above. A number of studies focused on nucleosome organization around these NFRs and reported consistency with statistical positioning downstream of the 5' NFRs [16, 77, 141]. These studies will be discussed in the context of our work in Sec. 4.4.3.

Having described statistical positioning and (above) the notion of a genomic code, the question arises of how both are related. At first sight, both concepts seem not to be compatible. A second thought, however, reveals that statistical positioning actually is one possible way for sequence to direct nucleosome positions in parts of the genome. In fact, statistical positioning is often considered to at least contributing to nucleosome organization. This also



Figure 4.3.: One realization of a gas of hard rods with particles at positions x_1, \dots, x_N placed between two boundary particles fixed at positions x_0 and x_{N+1} , respectively.

holds for studies arguing that models or *in vitro* experiments describe the situation *in vivo* well [47, 140]. In conclusion, the question of whether a genomic code exists should therefore be refined: (1) Are nucleosomes mainly positioned one by one or, alternatively, for the most part organized in large arrays resulting from the presence of a few boundaries? (2) Are the individual positions and/or the boundaries mainly coded into the genome or are other factors than sequence more important?

4.4.2. Physics of statistical positioning: the Tonks gas model

Statistical positioning corresponds to a well-known physical model: a one-dimensional gas of hard rods – the arguably most simple many-body problem. The model is usually attributed to Lewi Tonks referring to his 1936 paper on “The complete equation of state of one, two, and three-dimensional gases of hard elastic spheres” [138]. Though he was certainly not the first to consider this system, today it is known under the name *Tonks gas*. In the following, I would like to introduce the main aspects of it, starting out with a somewhat more general model. For a review on the Tonks gas and related models see Ref. [76].

Partition sums Consider a system of N indistinguishable particles at positions x_i in between two fixed particles at positions x_0 and x_{N+1} , see Fig. 4.3. The particles are subject to a two-particle interaction energy $v(x_i, x_j) = v(x_i - x_j)$ which only depends on the distance between particles. For this system, the configurational part⁹ of the canonical partition function reads

$$\begin{aligned} Z_N &= \frac{1}{N!} \int_{x_0}^{x_{N+1}} \dots \int_{x_0}^{x_{N+1}} dx_N \dots dx_1 e^{-\beta \sum_{i>j} v(x_i - x_j)} \\ &= \int_{x_0}^{x_{N+1}} dx_N \int_{x_0}^{x_N} dx_{N-1} \dots \int_{x_0}^{x_2} dx_1 e^{-\beta \sum_{i>j} v(x_i - x_j)}. \end{aligned} \quad (4.1)$$

In the second line, the integral has been rewritten taking ordering of particles into account. Such an ordering is specific to systems in one dimension and simplifies calculations as we will see below. Further simplification arises from restricting the pair-wise potential $v(x)$ to nearest neighbors only. This condition is automatically fulfilled for hard rods of width b with an additional interaction $\phi(x)$ which does not range further than twice the rod length.

⁹We restrict ourselves to the configurational part in all what follows, the concept of kinetic energy is not applicable to our system of interest, a system of nucleosomes performing a one-dimensional diffusion along the stretch of DNA they are bound to.

Formally:

$$v(x) = \begin{cases} \infty & \text{for } |x| < b \\ \phi(x) & \text{for } b \leq |x| \leq 2b \\ 0 & \text{for } |x| > 2b \end{cases} . \quad (4.2)$$

Systems of this class are called Takahashi nearest-neighbor gases, named after Hidetoshi Takahashi, who recognized that the partition function for nearest-neighbor interactions takes on the form of a convolution [132]:

$$Z = \int_{x_0}^{x_{N+1}} e^{-\beta v(x_{N+1}-x_N)} dx_N \int_{x_0}^{x_N} e^{-\beta v(x_N-x_{N-1})} dx_{N-1} \dots \\ \dots \int_{x_0}^{x_3} e^{-\beta v(x_3-x_2)} dx_2 \int_{x_0}^{x_2} e^{-\beta v(x_2-x_1)-\beta v(x_1-x_0)} dx_1 . \quad (4.3)$$

Regarding the canonical partition function Z as a function of $L = x_{N+1} - x_0$, one can take the Laplace transform $\bar{Z}(s)$ which turns the convolution into a product:

$$\bar{Z}(s) = \int_0^\infty Z(L) e^{-sL} dL = \left(\int_0^\infty e^{-\beta v(r)-sr} dr \right)^{N+1} \equiv K(s)^{N+1} . \quad (4.4)$$

This way the partition function may be computed by taking the inverse Laplace transform. However, employing the physical meaning of the Laplace transform can simplify the matter. The Laplace variable s can be identified as βp , pressure divided by temperature, and $\bar{Z}(s)$ is the partition function corresponding to the isobaric ensemble, in the following written as:

$$Y(p) = \int_0^\infty e^{-\beta p L} Z(L) dL . \quad (4.5)$$

Hence, the Gibbs free energy,

$$G(T, p, N) = -k_B T \ln Y(p) = -(N+1) \ln \int_0^\infty e^{-\beta v(r)-\beta p r} dr , \quad (4.6)$$

and thus the equation of state,

$$L = \left. \frac{\partial G(T, p, N)}{\partial p} \right|_{T, N} = \frac{(N+1) \int_0^\infty r e^{-\beta v(r)-\beta p r} dr}{\int_0^\infty e^{-\beta v(r)-\beta p r} dr} , \quad (4.7)$$

may be computed directly, which simplifies calculations for different types of interactions as shown by Bishop and Boonstra [13].

In addition, Eq. (4.6) and Eq. (4.7) show that G and L both are single-valued functions of pressure p (for any reasonable interaction), a condensation or phase transition thus cannot occur [132, 143]. Additionally, [van Hove](#) showed that no phase transition can occur for the

more general case where the interaction is restricted to a finite number of neighbors instead of only to nearest neighbors [143]. Though we are essentially interested in the pure Tonks gas at present (with hard core repulsion only), these results imply that no phase transitions occur in more general systems where the particles (i.e., the nucleosomes) are allowed to interact on a finite range.

Considering now the Tonks gas, $\phi \equiv 0$ and $K(s) = \exp(-bs)/s$. The canonical partition function may then be calculated directly by taking the inverse Laplace transform (applying Cauchy's theorem of complex analysis). Alternatively, the interaction potential may be treated as a geometrical constraint, restricting the particle's positions relative to each other:

$$Z = \int_{x_0+Nb}^{x_{N+1}-b} dx_N \int_{x_0+(N-1)b}^{x_N-b} \dots \int_{x_0+2b}^{x_3-b} dx_2 \int_{x_0+b}^{x_2-b} dx_1. \quad (4.8)$$

A change of variables readily makes it possible to solve this integral which leads to

$$Z = \frac{1}{N!} (L - (N+1)b)^N. \quad (4.9)$$

The free energy and the equation of state are easily computed in the thermodynamic limit ($N \rightarrow \infty$, $L \rightarrow \infty$ while $N/L = \bar{\rho}$):

$$F(T, L, N) = -k_B T \ln Z \quad \Rightarrow \quad F = N k_B T \ln \frac{\bar{\rho}/e}{1 - \bar{\rho}b}, \quad (4.10)$$

$$p = - \left. \frac{\partial F(T, L, N)}{\partial L} \right|_{T, N} \quad \Rightarrow \quad \beta p = \frac{\bar{\rho}}{1 - \bar{\rho}b}. \quad (4.11)$$

The free energy is extensive and the equation of state reduces to that of an ideal gas in the low density limit $\bar{\rho} \rightarrow 0$, as expected.

Particle density and two-particle distribution function Recalling the initial motivation, we considered the Tonks gas as the physical model corresponding to statistical positioning. To test the statistical positioning scenario quantitatively, the observable of interest is the nucleosome or particle density $\rho(x)$. The density is formally equivalent to the probability of finding any particle at position x and may thus be written as

$$\rho(x) = \left\langle \sum_i \delta(x - x_i) \right\rangle. \quad (4.12)$$

As usual, brackets $\langle \dots \rangle$ denote the average over all realizations. Similarly, the two-particle distribution function states the probability of finding any pair of particles at positions x and x' simultaneously,

$$\rho^{(2)}(x, x') = \left\langle \sum_i \delta(x - x_i) \cdot \sum_{j \neq i} \delta(x' - x_j) \right\rangle. \quad (4.13)$$

In condensed matter physics, the two-particle distribution function of a system may be determined using X-ray or neutron scattering while the particle density is a quantity which, in general, is not accessible. For our system, the situation is somewhat different. The positions of nucleosomes (corresponding to the particle density) are today determined using the

molecular biology methods described in Sec. 4.1. However, to my knowledge, the two-particle distribution function is not addressed in current experiments at high resolution.¹⁰

The two-particle distribution function is a useful quantity to consider: In the bulk of a Tonks gas, far away from the boundaries, $\rho^{(2)}(x, x')$ does not independently depend on x and x' , but becomes a function of $r = |x - x'|$ alone. In this limit, $\rho^{(2)}(r)$ is closely linked to the particle density $\rho(r)$ in a semi-infinite system with a boundary particle placed at $r = 0$. The wall may be seen as an arbitrary particle, ‘anchoring’ the two-particle distribution function, just as in Fig. 4.2. At a given location, such an ‘anchoring’ particle is found with a probability which is identical to the mean density $\bar{\rho}$. Taken together:

$$\bar{\rho} \cdot \rho(r) = \rho^{(2)}(r). \quad (4.14)$$

How to obtain $\rho^{(2)}(r)$ for the Tonks gas? As Eq. (4.13) shows, $\rho^{(2)}(r)$ formally is the expectation value of δ -functions. When computing the average, the δ -functions split up the convolution integral of the partition function (Eq. (4.3)). The remaining parts, however, are still convolution integrals and may be evaluated using Laplace transforms as done by Salsburg et al. in a lengthy calculation [117]. The matter becomes much clearer when one considers the Tonks gas in the thermodynamic limit right from the beginning. Since $\rho^{(2)}(r)$ is proportional to the probability of finding any particle at a distance r , the problem may be reduced to finding the k -th neighboring particle at a distance r and subsequently summing over k . In Sec. 4.7, these considerations are made quantitative; the final expression is

$$\rho^{(2)}(r) = \frac{\bar{\rho}}{b} \sum_{k=1}^{\infty} \frac{\Theta(r/b - k) (r/b - k)^{k-1}}{(l-1)^k (k-1)!} \exp\left(-\frac{r/b - k}{l-1}\right) \quad \text{with} \quad l = \frac{1}{\bar{\rho}b}. \quad (4.15)$$

Interestingly, Zernike and Prins [158] obtained the same result by similar considerations almost a decade before Tonk’s paper [138] and well before Salsburg et al. [117].

Fig. 4.2 displays the particle density close to a boundary particle and thus the two-particle distribution function for $\bar{\rho} = 1/178$ and $b = 147$, typical values for the nucleosome system (with length units of bp). A number of oscillations are clearly visible, which decay and approach a constant. The first peak is found at $r = b$, subsequent peaks are (for very dense systems, $\bar{\rho} \rightarrow 1/b$) regularly spaced at intervals of $1/\bar{\rho}$. The limiting value of $\rho^{(2)}(r)$ for large r is $\bar{\rho}^2$ since the probabilities of finding particles at very large distances are uncorrelated. See the ‘Materials and Methods’ part of Sec. 4.7 for a more detailed characterization. Let me finally emphasize that the structure in the two-particle distribution function arises solely from the interplay between excluded volume interactions and entropy and that the shape is independent of the temperature since the Tonks gas exhibits no energy scale.

Tonks gas on lattice & in the grand-canonical ensemble So far, the introduction into the physics of the Tonks gas has been limited to ensembles with fixed particle number. Strictly speaking, this condition is not fulfilled in a system of nucleosomes bound to DNA. Nucleosomes may be evicted from and also form on the DNA; hence, the grand-canonical ensemble should be considered. However, in a semi-infinite system, the canonical and the grand-canonical ensemble are equivalent and we may use the expression derived above for the particle density

¹⁰See, however, the description of old experimental techniques which determine quantities at least related to the two-particle distribution function [54].

close to a wall to describe nucleosome organization adjacent to a barrier.¹¹ A similar consideration holds for the continuum limit we assumed so far. Nucleosomes only bind to discrete places on the DNA, separated by one base pair; a lattice is more appropriate as a binding template than the continuum. However, the average distance between nucleosomes is much larger than the lattice spacing of 1 bp, which implies that the expressions derived for the continuum limit are applicable to a system of particles bound to the lattice.

Taken together, we expect the results derived for a continuum system in the canonical ensemble to describe a lattice system in the grand-canonical ensemble. While in the former system, the particle density close to a wall may be easily derived analytically (see above), numerical computations become simple in the latter system (see Appendix A.3). We employed the numerical solution to check that equivalence between both systems is given for typical parameters of interest.

4.4.3. Quantitative test of statistical positioning

In 2008, Frank Pugh and coworkers reported a nucleosome map for *S. cerevisiae* and gave qualitative evidence that nucleosome organization in genes is governed by statistical positioning against the +1 nucleosome [77]. However, a quantitative test had been missing at that time which motivated our study on nucleosome organization [85]. Specifically, we examined whether nucleosome density downstream and upstream of TSSs is quantitatively consistent with the Tonks gas model. We made as little assumptions as possible about what constitutes the barrier which (putatively) causes statistical positioning, and merely assumed that the position of the barrier (or wall) is related to the NFR. This clearly distinguishes our approach from the computational modeling of nucleosome positions described in Sec. 4.3.1. In these ‘forward approaches’, nucleosome positions are first predicted from sequence and in a second step compared to the observed nucleosome organization. Our study, in contrast, is an example for the ‘reverse approach’. It is based on the observed distribution of nucleosomes which is compared to a minimal model in order to determine the underlying positioning determinants. Since we make almost no assumptions about what constitutes the barrier, our approach can easily incorporate other determinants contributing to the formation of the barrier besides sequence. In the following, I would like to outline the main aspects of our study. Details are found in the publication [85] reprinted in Sec. 4.7.

Results As a prerequisite for a quantitative test of statistical positioning, the nucleosome density at a given location on the genome is needed. The best available proxy for nucleosome density is the density of reads, i.e., the number of putative nucleosome midpoints at a given location, extracted from an ensemble of cells using ChIP-Seq experiments (see Sec. 4.1). This presumes, of course, that the average over different cells is equivalent to the thermal average in a single cell and furthermore that ChIP-Seq reliably and quantitatively determines the midpoints of nucleosomes in different cells. Neither assumption can be verified *a priori*, hence, we performed consistency checks. One limitation could not be overcome, however. The average number of reads corresponding to a nucleosome in a single cell is not known, which leaves the read density as a proxy for nucleosome density with an unknown overall normalization factor.

¹¹For particle densities and higher-order distribution functions in finite sized intervals, where the distinction between canonical and grand-canonical ensemble matters, see Ref. [112] and references therein.

In order to test the statistical positioning scenario, ideally one would aim to compare the read density to the Tonks gas model at individual genomic loci, i.e., at promoters of individual genes. The data set we used [77] (and also newer data sets reported since) do not provide enough statistics for this endeavor, the read density is not sampled well enough. To circumvent this limitation, one commonly aligns nucleosome density to a set of different genomic loci and subsequently averages. Instead of aligning to TSSs, as was done in Fig. 4.1 and also by Mavrich et al. [77], we aligned to the positions of the +1 and -1 nucleosomes surrounding the TSSs. This choice was based on the observation that not all promoters are equal and that the sizes of the NFRs differ significantly from gene to gene. Assuming that the NFR is related to the boundary element which induces statistical positioning up- and downstream, aligning to the +1 and -1 nucleosomes represents the best possible ansatz for aligning to the walls of the putative boundary element.

Both nucleosome density upstream of the -1 nucleosome and downstream of the +1 nucleosome show characteristic oscillations in qualitative agreement with the density oscillations in the Tonks gas. But quantitatively, both alignments differ significantly. Amplitudes are smaller and do not range as far upstream of the -1 nucleosome as compared to downstream of the +1 nucleosome. To make the comparison with the Tonks gas quantitative and resolve this discrepancy, we performed systematic fits of the Tonks gas particle density (Eqs. (4.14) and (4.15)). The only two essential fit parameters were the average particle density $\bar{\rho}$ and the overall normalization factor mentioned above; nucleosome width had been set to 147 bp. A systematic analysis revealed that nucleosome density both downstream and upstream of the NFR are both quantitatively consistent with statistical positioning if considering different boundary conditions on both sides of the NFR. Our results suggest that the +1 nucleosome is directly positioned and thus part of the boundary inducing statistical positioning (of the +2, +3, ... nucleosomes) downstream. The -1 nucleosome, by contrast, should be seen as the first nucleosome upstream of the boundary (formed by the NFR) and is itself statistically positioned as are the -2, -3, ... nucleosomes.¹²

It is worth rephrasing our result in physical terms (see Fig. 4D in Sec. 4.7): the NFR represents a broad effective free energy barrier, which leads to the exclusion of nucleosomes from this region. At the downstream border of the NFR, there is a localized attractive free energy well which leads to direct positioning of the +1 nucleosome. All the other nucleosomes (including the -1 nucleosome) form a Tonks gas on both sides of the NFR, and boundary effects lead to the observed oscillations in nucleosome density.

Our study made three major assumptions which should briefly be mentioned here. First, we assumed the read density to be a reliable proxy for nucleosome density (see above). Second, we performed an average over many different genomic loci to enhance statistics (see above). An additional concern here is that we mostly neglected the presence of other genetic features (e.g., other TSSs) when we performed the average and moreover when we compared the data with a semi-infinite Tonks gas. These strong assumptions are hard or even impossible to justify systematically. However, we provided consistency checks addressing these two assumptions. Third, we assumed equilibrium to hold, which is inherent not only in our work but in all thermodynamic models of nucleosome organization. Too little is known about the kinetics of the individual processes involved in nucleosome reorganization to justify the equilibrium assumption from scratch. For a more detailed discussion see Ref. [127] and Sec. 4.7.

¹²We also compared our results to the nucleosome organization around the NFRs at the end of genes. We observed the characteristic oscillations in nucleosome density but not such an asymmetry as described here.

Taken together our study revealed a surprisingly good quantitative agreement between the Tonks gas model and nucleosome organization on both sides of the NFR at the promoter. However, our study relies on a number of strong assumptions which are at least in part owed to our ‘reverse approach’ of starting our analysis with the measured data. Nevertheless, it would also be difficult to explain why the observed pattern of read density follows so closely the density pattern of a Tonks gas, which gives our results additional *a posteriori* validation.

Related studies Our results are complementary to the work by Mavrich et al. [77] and a number of newer studies which were published after our work had been finished [16, 82, 141]. All of these studies argued for statistical positioning in yeast (especially downstream of the NFRs) and should be discussed here.

Mavrich et al. presented evidence for statistical positioning downstream of the +1 nucleosome based on a collection of qualitative arguments. The authors reported the consensus position of nucleosomes (peaks in the read density) and an associated fuzziness (essentially the width of the peaks due to cell-to-cell variability). Alignment to TSSs followed by averaging revealed a pattern qualitatively very similar to that in Fig. 4.1. The presence of decaying oscillations let the authors to argue for statistical positioning. Furthermore, they found that nucleosome fuzziness is larger downstream of the TSS than close to the TSS, which they also reported to be in agreement with statistical positioning. Using similar reasoning, Mavrich et al. also argued for statistical positioning upstream of the -1 nucleosome. All of these arguments involve the identification of consensus nucleosome positions, a concept difficult to combine with statistical positioning. The authors furthermore only considered alignment to TSSs, which does not take into account the varying width of the NFRs. Taken together, these problems motivated us to put the above reasoning on a more solid basis by reanalyzing the data as described above.

Mavrich et al. provided additional support for statistical positioning using an analysis of the DNA sequence around promoters. They showed that sequence elements known to be involved in nucleosome positioning (dinucleotide patterns and poly(dA:dT) stretches) are concentrated to the NFR and the -1 and +1 nucleosomes – and are less abundant up- and downstream of this region. These findings are consistent with statistical positioning and in agreement with the picture of promoter organization emerging from our study, apart from the pro-nucleosomal sequence at the -1 nucleosome position. However, the relevance of the positioning sequence could be quite modest, which is not unlikely given that other positioning determinants than sequence are needed to explain the positions of +1 nucleosomes *in vivo* [159].

Two related studies on statistical positioning in genic regions were published after our work had been completed [16, 141]. In contrast to previous studies, the authors did not consider the common alignments to TSSs or +1 nucleosomes. Instead, they considered every transcribed region independently. They found that two soft boundaries placed at the 5’ and 3’ ends of the transcribed regions and a Tonks gas in between suffices to generate a regular pattern of nucleosomes. The positions of the minima in the predicted nucleosome occupancy agree quite well with the positions observed in an *in vivo* measurement [61]. To illustrate this, the genes were ordered by the distance between their first and last nucleosome and plotted one below each other. Using a model for nucleosome positioning [140] and comparing it to *in vitro* measurements at low nucleosome density [47], the authors provided additional evidence that sequence between the 5’ and 3’ ends of genes plays a minor role in establishing the *in vivo* nucleosome organization, consistent with statistical positioning.

Despite the compelling evidence for statistical positioning in individual genes, the work by Chevereau et al. and Vaillant et al. has to be seen as complementary to ours. The authors overcame the obstacle of averaging over genes, but focused on the positions of minima in the nucleosome occupancy instead of on the shape of the oscillations as we did. Moreover, the authors did not take into account the +1 nucleosome to be directly positioned but rather considered it to be positioned against a barrier at the 5' end of the gene.¹³

Last but not least, very recent work by Milani et al. gave experimental evidence for statistical positioning in one single yeast gene (see page 58). Though their method of imaging nucleosomes is unlikely to be scaled up to a genome-wide analysis in the future, it is interesting from another point of view. Imaging revealed the relative location of two nucleosomes on a single DNA fragment, i.e., it provided access to the two-particle distribution function.

In conclusion, the results of our study are mainly in line with recent studies addressing statistical positioning downstream of yeast promoters. However, our study is the only one concentrating on the detailed shape of the oscillations. This way, our work provides new findings about nucleosome organization upstream of the promoter and about boundary conditions close to the NFR.

The now somewhat settled picture of statistically positioned nucleosomes in genic regions allows us to return to the question of a *genomic code*: First of all, the oscillations close to NFRs have a range of about 1000 bp (see Fig. 4.1 for this rough estimate). Second, there are several thousand genes distributed on the yeast genome which has a size of about 12 million bp. Taken together, this indicates that boundaries at TSSs alone suffice to position the majority of all nucleosomes. Taking into account that the 5' NFRs are at least partially encoded into the sequence by poly(dA:dT) sequences (see above) this suggests a scenario where a significant extent of nucleosome organization is determined by a few barriers encoded in the sequence. Though in this scenario the majority of nucleosomes are not individually positioned, they are nevertheless “glued” to certain positions. This may be the explanation for genetic variation being correlated to nucleosome positions downstream of the TSS in medaka fish [119].¹⁴ One may speculate that some of the sequence patterns observed in nucleosomal DNA are a consequence of, rather than the cause for, nucleosome positioning.

Extensions So far, we and others modeled nucleosomes as hard rods of fixed length. However, this picture may be simplified too much. First, as we have seen in Chap. 3, nucleosomes are not static entities but they may undergo transient events of DNA unwrapping. The cost in free energy for this unwrapping was estimated to be rather modest: about $1.5 - 2 k_B T$ for unwrapping of each 10 bp. Second, the three-dimensional structure of a nucleosome (Fig. 1.3) suggests that, for geometrical reasons, very small linker lengths are unlikely, i.e., that they are associated with a large free energy cost. For these two reasons, it is natural to extend the Tonks gas model we considered so far. In a prescient study, Tom Chou considered one possible extension, that of a gas of hard rods with length varying between a minimum and a maximum value. A linear energy cost is associated with decreasing the rod length [18]. Since this work provides the two-particle distribution function, one has everything at hand for an analysis of statistical positioning in the context of this model.

¹³For the sake of completeness, it should be mentioned that sorting intergenic regions by length and plotting them one below each other revealed a fuzzy instead of an organized pattern of nucleosome organization in intergenic regions [141]. This is in contrast to our result of statistical positioning upstream of the -1 nucleosome, but may simply be due to the way the data were displayed.

¹⁴See Ref. [102] for more information on “sequence ‘shadows’ of nucleosomes”.

Why should one aim for a more complex analysis given that the pure Tonks gas already describes nucleosomes organization around NFRs quite well? First, the model by Chou might help to explain one peculiarity in our data analysis above. The Tonks gas model (in the parameter regime of interest) predicts the distance between peaks in the density oscillations to be very close to the mean distance between particles. Our fits, by contrast, revealed a mean inter-nucleosome distance which is slightly larger than the distance between the maxima in the read density oscillations, which means the density oscillations in data and model are slightly ‘out of phase’ (see Fig. 2 in Sec. 4.7).¹⁵ Using the Tonks gas model with varying rod length we hope to be able to resolve this discrepancy.

A second motivation comes from a comparison of nucleosome organization in related species. Oliver Rando (University of Massachusetts) and coworkers determined nucleosome positions for different yeast species and kindly provided us with their unpublished data. A preliminary analysis (results are not shown) suggests that nucleosome organization downstream of NFRs at 5’ ends of genes may (again on average) be described by a Tonks gas in the different yeast species, but with different parameters for nucleosome density $\bar{\rho}$ and nucleosome width b . We speculate that it might be possible to explain the nucleosome organization in different species within the extended model where a higher density leads to a shorter effective length of nucleosomes.

The generalized Tonks gas model, allowing for varying particle width, is one possible way to overcome the omnipresent assumption that nucleosomes have a fixed width and interact via hard core repulsion only. A similar attempt was performed on the bioinformatic side as well. The well-known bioinformatic nucleosome model by Segal and coworkers was extended to model interactions between neighboring nucleosomes by a ‘nucleosome cooperativity function’ whose parameters were learned from measured nucleosome organization [71]. The extended model improved nucleosome predictions, at the expense of higher complexity. However, it is a formidable task to combine what’s known about *in vivo* nucleosome organization with the available knowledge concerning higher order chromatin structures, especially the 30-nm fiber.

4.5. Outlook: the inverse approach

Within the preceding sections, I presented two complementary approaches of understanding nucleosome organization by modeling. First of all, there are the sequence-based models discussed in Sec. 4.3.1 (‘forward approach’). A prominent example is the model by Segal and coworkers [33], where, rephrased in physics terms, sequence around a location x is used to predict a nucleosome binding energy $u(x)$. Thereafter, the binding energy is used to compute nucleosome density $\rho(x)$ which finally is compared to experiments. Despite its reported success in describing nucleosome organization *in vitro* and *in vivo* [47], it is difficult to extend such a model to include effects other than sequence, e.g., transcription factor binding or action of chromatin remodelers. Second, there is the ‘reverse approach’ of taking the measured nucleosome density $\rho(x)$ as is and trying to quantitatively describe it in terms of the minimal Tonks gas model. An example is our study where we made only minimal assumptions on what constitutes the barrier for statistical positioning. Though we were able to describe nucleosome density $\rho(x)$ around promoters quantitatively, it is difficult to extend the approach to include weak local variations in binding affinity, for example.

¹⁵The discrepancy is due to the fact that the fitting procedure identifies the best ‘compromise’ between peak-to-peak distance, decay range of oscillations, and amplitude compared to the asymptotic value.

In this section, I would like to outline what we call the *inverse approach*, one possible way of combining the best of the two complementary strategies and overcoming some of the limitations they face. We propose taking the measured nucleosome density $\rho(x)$ and computing a corresponding (free) energy landscape $u(x)$ which is interpreted *a posteriori*.

For a system of hard rods on a lattice, a direct relation between density $\rho(x)$ and energy landscape $u(x)$ was found by Robledo [111] (see also Robledo and Varea [113]) in the grand-canonical ensemble¹⁶

$$\beta(u(x) - \mu) = -\ln \rho(x) + \ln \frac{\prod_{k=0}^{b-1} [1 - t_{b-1}(x+k)]}{\prod_{k=0}^{b-2} [1 - t_{b-2}(x+k)]} \quad \text{with} \quad t_m(x) = \sum_{l=0}^m \rho(x-l), \quad (4.16)$$

where μ is the chemical potential and b is the number of lattice sites covered by one rod. In the continuum limit, this reduces to a relation which had been found by Percus before [96],

$$\beta(u(x) - \mu) = -\ln \rho(x) + \ln \left[1 - \int_{x-b}^x \rho(w) dw \right] - \int_x^{x+b} \frac{\rho(z)}{1 - \int_{z-b}^z \rho(w) dw} dz. \quad (4.17)$$

Equations (4.16) and (4.17) provide a direct mapping from density to binding energy as needed for the inverse approach we propose. The opposite and more common mapping from binding energy to density is achieved using a recursion relation outlined in Appendix A.3 (or a dynamic programming algorithm explained in Ref. [135]) for the lattice case. If the continuum is considered, the solution to Percus' equation [144] can be applied.

Fig. 4.4 illustrates the potential of the inverse approach. Consider an energy barrier with an adjacent well in an otherwise flat energy landscape as illustrated at the left. The nucleosome density corresponding to that energy landscape for a given chemical potential is displayed in the middle of Fig. 4.4. Applying Eq. (4.16) completely reproduces the original energy landscape as is seen on the right. While the density pattern reveals a confusing diversity of oscillations which is hard to interpret, the energy landscape leading to this pattern is much less complex and easily recovered by the inverse approach.

Applying Eq. (4.16) to both *in vitro* and *in vivo* measurements of nucleosome density offers the possibility to determine the mechanisms which position nucleosomes *in vivo* only. As can be clearly seen from Eq. (4.17), all (reasonable) density landscapes are mapped to (free) energy landscapes. This means that even non-equilibrium processes such as directed action of chromatin remodelers are represented in the energy landscape. Comparing the energy landscapes for both the *in vitro* and the *in vivo* situation thus allows one to deduce where such mechanisms influence nucleosome organization *in vivo*.

In spite of all advantages the inverse approach offers, there are two major pitfalls. First, the nucleosome maps currently available do not provide enough statistics for Eq. (4.16) to be directly applied. Though statistics will very likely become much better in near future, a careful error analysis will in any case be necessary. We are currently working on error analysis using standard Monte Carlo simulations to simulate real data, but results are too preliminary

¹⁶The derivations seem to involve the assumption of large systems. Up to now, I did not invest enough time to understand the consequences of this on the applicability of Eq. (4.16) for arbitrary systems in the grand-canonical ensemble.

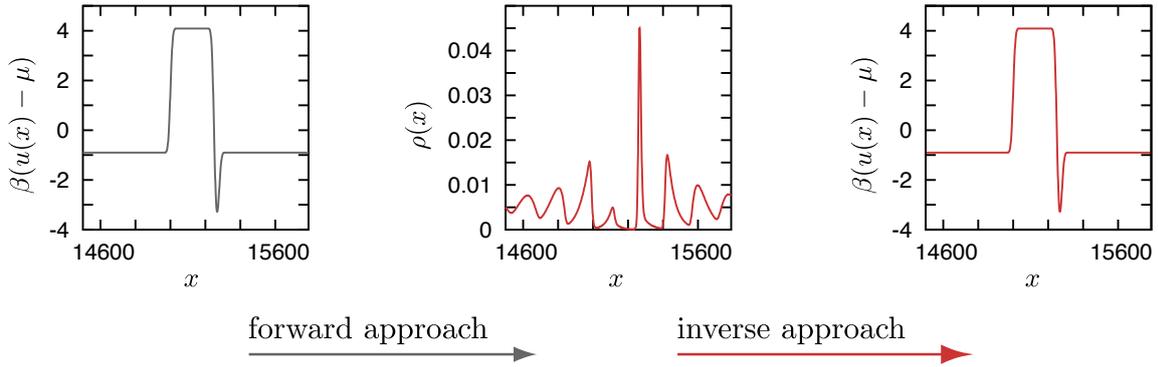


Figure 4.4.: Illustration of the mapping between binding energy $u(x)$ and density $\rho(x)$ for particles of finite width on a lattice. **(Left)** Energy barrier with an adjacent well in an otherwise flat binding energy landscape (boundaries are far apart). **(Middle)** Recursion relations (Appendix A.3) allow to compute the corresponding particle density which, due to the finite width of the particles, is complex and displays oscillations ranging far beyond the ‘feature’ in the binding energy landscape. **(Right)** Applying Eq. (4.16) to the particle density reveals the underlying energy landscape (inverse approach). Parameters: $\mu = 0.9$, $b = 147$, $\beta = 1/k_B T = 1$.

to be reported. Second, though the inverse approach makes no *a priori* assumptions about the binding energy, it assumes the interaction to be of hard core type. This is state of the art for almost all approaches addressing nucleosome organization on the genome to date. Yet, whether this assumption leads to an implicit and unphysical mapping from nucleosome-nucleosome interactions into the binding energy landscape is an important question to be explored.

4.6. Conclusion

Within the last five years, knowledge about nucleosome organization on the genome has grown dramatically. This to a large extent has been achieved by an ever increasing number of experimental studies determining nucleosome positions at high resolution. Particular attention has been paid to promoter regions where a consensus pattern of nucleosome organization has been established: a nucleosome-free region with oscillations in nucleosome density both downstream and upstream. In our theoretical study, outlined and reprinted in this chapter, we tested whether these oscillations may quantitatively be captured by a minimal physical model, a one-dimensional gas of hard rods, commonly called the Tonks gas. In this model, density oscillations occur close to a boundary at dense packing. Our systematic quantitative analysis revealed that, on average over many promoters, a Tonks gas can indeed describe nucleosome organization on both sides of the nucleosome-free region if one considers two different boundary conditions at both sides of the nucleosome-free region: while downstream, a single nucleosome is directly positioned and forms a boundary element for the gas of nucleosomes, such a barrier nucleosome is missing on the upstream side.

Our study must not be seen outside of the context of a long series of biological and bioinformatic studies, the essence of which I outlined in this chapter as well. First, describing nucleosomes as a gas of hard rods was proposed a long time ago. Since density oscillations

represent themselves in preferred nucleosome positions, the concept is known under the name of statistical positioning. Our work took up a previous study based on qualitative arguments and provided a quantitative analysis. More recent studies by another group provided a complementary quantitative analysis, mostly in line with our results. Second, a number of mainly bioinformatic studies addressed the question to what extent nucleosome organization can be predicted using sequence information alone. In a nutshell, these studies computed a binding energy for nucleosomes based on the underlying sequence, which was the basis for computing a predicted nucleosome density. Our study, by contrast, tried to infer the effective binding energy based on the measured data. At the end of the chapter, I proposed a method which combines the advantages of both approaches: deducing the binding energy landscape from measured nucleosome densities and interpreting it *a posteriori*. Though a proof of principle, i.e., application to real data, remains to be reported, I here discussed the main idea, prospects and caveats of this approach.

Quantitative Test of the Barrier Nucleosome Model for Statistical Positioning of Nucleosomes Up- and Downstream of Transcription Start Sites

Wolfram Möbius, Ulrich Gerland*

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

Abstract

The positions of nucleosomes in eukaryotic genomes determine which parts of the DNA sequence are readily accessible for regulatory proteins and which are not. Genome-wide maps of nucleosome positions have revealed a salient pattern around transcription start sites, involving a nucleosome-free region (NFR) flanked by a pronounced periodic pattern in the average nucleosome density. While the periodic pattern clearly reflects well-positioned nucleosomes, the positioning mechanism is less clear. A recent experimental study by Mavrich *et al.* argued that the pattern observed in *Saccharomyces cerevisiae* is qualitatively consistent with a “barrier nucleosome model,” in which the oscillatory pattern is created by the statistical positioning mechanism of Kornberg and Stryer. On the other hand, there is clear evidence for intrinsic sequence preferences of nucleosomes, and it is unclear to what extent these sequence preferences affect the observed pattern. To test the barrier nucleosome model, we quantitatively analyze yeast nucleosome positioning data both up- and downstream from NFRs. Our analysis is based on the Tonks model of statistical physics which quantifies the interplay between the excluded-volume interaction of nucleosomes and their positional entropy. We find that although the typical patterns on the two sides of the NFR are different, they are both quantitatively described by the same physical model with the same parameters, but different boundary conditions. The inferred boundary conditions suggest that the first nucleosome downstream from the NFR (the +1 nucleosome) is typically directly positioned while the first nucleosome upstream is statistically positioned via a nucleosome-repelling DNA region. These boundary conditions, which can be locally encoded into the genome sequence, significantly shape the statistical distribution of nucleosomes over a range of up to ~1,000 bp to each side.

Citation: Möbius W, Gerland U (2010) Quantitative Test of the Barrier Nucleosome Model for Statistical Positioning of Nucleosomes Up- and Downstream of Transcription Start Sites. *PLoS Comput Biol* 6(8): e1000891. doi:10.1371/journal.pcbi.1000891

Editor: Eran Segal, Weizmann Institute of Science, Israel

Received: February 14, 2010; **Accepted:** July 15, 2010; **Published:** August 19, 2010

Copyright: © 2010 Möbius, Gerland. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by the German Excellence Initiative via the program Nanosystems Initiative Munich (NIM). WM acknowledges funding by the Studienstiftung des deutschen Volkes and the Elite Network of Bavaria via the International Doctorate Program NanoBioTechnology (IDK-NBT). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: gerland@lmu.de

Introduction

The long DNA molecules of eukaryotic genomes are packaged into a compact structure with the help of histone proteins [1]. The fundamental unit of this structure, a nucleosome, comprises almost 150 base pairs (bp) of DNA wrapped around a histone octamer [2,3]. Individual nucleosomes are typically linked by 15–70 bp of free DNA into a “beads on a string” conformation, the primary and most stable structural level of chromatin. While packaging renders the genome compact, it also makes up to 80% of the DNA inaccessible for protein-binding at any given time [4], potentially hindering the molecular processing of genetic information. In principle, accessibility might be attained dynamically, since mechanisms are known for spontaneous unwrapping [5,6] and diffusive sliding of nucleosomes [7], as well as active remodeling [8]. However, numerous recent studies indicate that nature’s solution to the accessibility issue is based, at least in part, on the widespread use of nucleosome positioning [4,9–14]. Nucleosome positioning essentially amounts to the opposite strategy of constraining the mobility of nucleosomes, rendering a selected set of DNA sites constantly accessible.

Recent experiments measuring the distribution of nucleosomes across the genomes of several model organisms have robustly identified three salient features [11]: (i) A significant fraction of nucleosomes appears rather well positioned. In other words, the nucleosome positions determined from a large ensemble of cells do not average out to a constant density, but display many pronounced peaks. (ii) Typically, genes have a nucleosome-free region (NFR) upstream of their transcription start site (TSS). That is, when genes are aligned at the TSS and with the direction of transcription to the right, the average nucleosome density exhibits a clear dip, about one nucleosome wide, to the left of the TSS. (iii) Downstream of the TSS, the gene-averaged nucleosome density displays strong oscillations, with an amplitude that decays with the distance from the TSS. Furthermore, biochemical experiments have firmly established that the DNA-binding affinity of histones depends on the DNA sequence, largely due to the intrinsic sequence-dependence in the biophysical properties of DNA, such as its bendedness and bendability [15]. Hence, a genomic free energy landscape for nucleosome positioning can be programmed into the genome sequence by appropriate placement of nucleosome attracting and repelling sequence motifs.

Author Summary

Within the last five years, knowledge about nucleosome organization on the genome has grown dramatically. To a large extent, this has been achieved by an increasing number of experimental studies determining nucleosome positions at high resolution over entire genomes. Particular attention has been paid to promoter regions, where a canonical pattern has been established: a nucleosome-free region with pronounced adjacent oscillations in the nucleosome density. Here we tested to what extent this pattern may be quantitatively described by a minimal physical model, a one-dimensional gas of impenetrable particles, commonly referred to as the “Tonks gas.” In this model, density oscillations occur close to a boundary at dense packing. Our systematic quantitative analysis reveals that, in an average over many promoters, a Tonks gas model can indeed account for the nucleosome organization to both sides of the nucleosome-free region, if one allows for different boundary conditions at the two edges. On the downstream side, a single nucleosome is typically directly positioned such that it forms an obstacle for the neighboring nucleosomes, while such a barrier nucleosome is typically missing on the upstream side.

Indeed, bioinformatic and biophysical approaches that parameterize sequence-encoded effects on nucleosome positioning have been remarkably successful in modeling and predicting the large-scale genomic nucleosome occupancy [16–20], which has led to the notion of a genomic code for nucleosome positions [16]. Yet, the causes of the three above salient features are not yet disentangled. In particular, a recent study on nucleosome positioning in *Saccharomyces cerevisiae* [10] argued that the oscillatory pattern in the average nucleosome organization downstream of the TSS is qualitatively consistent with the statistical positioning mechanism proposed by Kornberg and Stryer [21]. With this mechanism, most nucleosomes are not individually positioned, but a non-random relative arrangement arises collectively, from statistical correlations induced by the interaction between neighboring nucleosomes. The phase of such a statistical arrangement relative to the DNA is determined by “barriers” on the genome, i.e., local disturbances of the “nucleosome gas”. A disturbance is created regardless of whether the local effect on nucleosomes is attracting or repelling, e.g., by sequences that attract or repel nucleosomes or by other bound proteins [22,23]. According to this scenario, termed the ‘barrier nucleosome model’ [10,24], sequence-encoded positioning is required only for barrier creation, whereas nucleosomes adjacent to the barriers are positioned “for free”, i.e., primarily via statistical correlations and with DNA sequence playing only a minor role.

However, while the observed oscillatory pattern downstream of the TSS is reminiscent of the pattern calculated by Kornberg and Stryer [21], there should be a similar pattern upstream of the TSS if statistical positioning is indeed the dominant force, since barriers act to both sides. Also, can the observed pattern be quantitatively explained by statistical positioning? Finally, does the precise shape of the pattern permit conclusions on the nature of the barrier, e.g., whether it is caused by an attractive or repelling effect on nucleosomes? Here, we address these quantitative questions using the yeast data of Mavrich *et al.* [10] and a quantitative description of statistical positioning, which is essentially the same as in the work of Kornberg and Stryer [21] and equivalent to the (much older) “Tonks gas” model from statistical physics [17,25–28].

Results

Quantitative barrier nucleosome model

Kornberg noted early on [24] that a nonrandom quasi-periodic nucleosome pattern arises already from the interplay of two basic biophysical constraints, (i) the constraint that the same DNA segment cannot simultaneously be incorporated into two nucleosomes, and (ii) the constraint that nucleosomes cannot form at ‘barrier’ genome locations, e.g., those already occupied by other proteins such as sequence-specific transcription factors. The significance of the first constraint is that the exclusion between nucleosomes creates correlations in their statistical distribution along the DNA. Theoretically [26], these correlations are revealed by a decaying oscillatory pattern in the two-particle distribution function, $\rho_2(r)$, which measures the probability to find, in the ensemble of all admissible nucleosome configurations, a nucleosome at a given location and another one at a distance r from it. In other words, the knowledge of the position of one nucleosome leads to a partial knowledge of the nucleosome positions in the vicinity (however, this two-particle distribution function is difficult to measure directly in experiments). The significance of the second constraint is that barriers in the “nucleosome gas” pin down the phase of the correlations, such that even the average nucleosome density $\rho(r)$ displays a decaying oscillation as a function of the distance r from the barrier [21]. Such barriers can be created by a variety of mechanisms; in particular, barriers can also be directly encoded in the DNA sequence, e.g., via poly(dA:dT)-tracts that are energetically unfavorable to incorporate into the nucleosome structure [29]. Similarly, “road block” nucleosomes that are particularly well-positioned will form a barrier for the surrounding nucleosomes.

Here, we treat the average nucleosome density $\rho(r)$ as a quantitative experimental feature that can be assayed for clues about the nature of these barriers and, more generally, about the extent to which statistical positioning is reflected in the nucleosome organization *in vivo*. This analysis must be based on a quantitative description of statistical positioning. In statistical physics, the interplay between interaction and entropy of particles in a one-dimensional configuration space has long been quantified in simple models for gas/liquid systems [25,26,30]. The classic quantitative study of statistical positioning, by Kornberg and Stryer [21], is also consistent with this general framework. The simplest model is the ‘Tonks gas’ [25] where particles with a fixed size b and a mean density $\bar{\rho}$ interact only via hard-core repulsion that makes them impenetrable. For this model, the explicit analytical expression for the average particle density at a distance r (in bp) from a perfect barrier is [26]

$$\rho(r) = \sum_{k=1}^{\infty} \frac{\left(\frac{r}{b} - k\right)^{k-1} \Theta\left(\frac{r}{b} - k\right)}{b \cdot (k-1)!} \left(\frac{\bar{\rho} b}{1 - \bar{\rho} b}\right)^k e^{-\frac{r/b-k}{1/\bar{\rho}b-1}}, \quad (1)$$

where $\Theta(r)$ denotes the Heaviside step function. This average density is related to the above-mentioned two-particle distribution function in an infinite system via $\rho(r) \cdot \bar{\rho} = \rho_2(r)$. Eq. (1) produces the decaying oscillatory pattern that is characteristic for statistical positioning, see Fig. S1 for an illustration and ‘Materials and Methods’ for a self-contained derivation and a brief discussion of the physical mechanism underlying the density oscillation. The wavelength of the oscillatory pattern and the characteristic length over which its amplitude decays are both determined by the two physical parameters of the model, i.e., the particle size b and the average particle density $\bar{\rho}$. Note that the expression (1) holds only for a perfect barrier; more general situations will be considered below.

As Eq. (1) describes a nontrivial effect that arises only from properties which the “nucleosome gas” shares with any other one-dimensional gas of impenetrable particles, it can be regarded as a ‘null model’, i.e., a quantitative reference that helps to identify relevant effects beyond the universal features for systems of this class. With this goal in mind, we wanted to compare Eq. (1) to patterns extracted from experiments.

Extraction of experimental nucleosome patterns

To extract the consensus distribution of nucleosomes around the NFR at the 5′ end of genes, previous studies aligned the genes at their TSS and averaged the nucleosome distributions over all genes [11]. This procedure is not suitable for our quantitative analysis, since the TSS cannot be mapped to a feature in the nucleosome gas. Instead, we used the positions of the NFR-flanking nucleosomes as reference points for our alignments, which permits a quantitative comparison of the averaged pattern with the nucleosome gas model (see below).

In addition to the appropriate choice of reference point for the alignment, it was important to process the experimental data in a way such that it became directly comparable to the physical density $\rho(r)$. Many studies determine nucleosome positions using a procedure of the following type [11]: First, the nucleosomal DNA is extracted from an ensemble of cells using micrococcal nuclease (MNase). The genomic positions of these DNA fragments are then located using hybridization or sequencing approaches. Usually this raw data is further processed with hidden Markov models (e.g., [9]) or peak detection algorithms (e.g., [31]), in order to infer the typical or putative nucleosome positions. These typical nucleosome positions are then used for subsequent analysis of nucleosome organization, including the consensus distribution around NFRs. However, such averages over typical nucleosome positions do not correspond to a physical observable. For qualitative analysis, the data processing algorithms are useful filters to enhance and highlight positioning effects. However, the use of a single, typical position for a nucleosome eliminates any cell-to-cell variation in the position. For our quantitative analysis, we had to use the undistorted raw data instead (i.e., the density of DNA reads along the genome for the sequencing approach), which is the best available experimental proxy to the physical density $\rho(r)$, see ‘Materials and Methods’ for details. Note that our observable, the nucleosome density, is distinct from the other frequently used observable, the nucleosome occupancy, which measures the probability to find a specified base pair covered by a nucleosome.

Fig. 1 summarizes the nature of the data from a physics perspective. As illustrated in Fig. 1A, the extracted nucleosomal DNA originates from many cells with nucleosome positions that generally differ from cell to cell. The experimentally observed read density corresponds to the histogram shown in the bottom of Fig. 1A. This histogram would be directly comparable to the theoretical density $\rho(r)$ for a nucleosome gas, if (i) the average over the different cells is equivalent to the thermal average, (ii) a DNA read identifies a nucleosome position uniquely and precisely, and (iii) the average number of reads per nucleosome is known and its fluctuations due to the random sampling are negligible. None of these conditions is entirely satisfied. Clearly, the relevant question (discussed in ‘Materials and Methods’) is how much this affects the physical interpretation of the data. Since the average number of reads per nucleosome is in fact unknown, it is already clear that one cannot readily convert the read density to an absolute nucleosome density, i.e., the experimental proxy to $\rho(r)$ is not normalized. Fig. 1B illustrates the second averaging procedure, which is akin to a “disorder average” in statistical physics, in that it

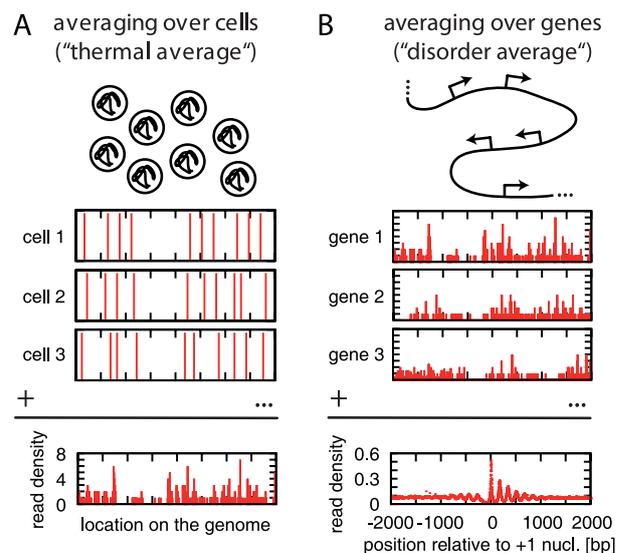


Figure 1. Illustration of the nature of the available data and its analysis. (A) Nucleosomal DNA from different cells is extracted and sequenced. The genomic positions of the sequence reads are determined, resulting in a genome-wide density of reads. This map reflects the nucleosome density averaged over an ensemble of cells. Physically, this average is akin to a thermal average. (B) To extract typical features (and to improve the statistics) genes are aligned according to a specific feature (here: the most likely position of the +1 nucleosome), and the read density is averaged over all genes. Physically, this average is akin to a disorder average. doi:10.1371/journal.pcbi.1000891.g001

involves averaging over an ensemble of different systems rather than an ensemble of different states of the same system. Clearly, each gene is intrinsically different and could display a distinct pattern of nucleosome organization. However, as illustrated in the bottom of Fig. 1B, the common pattern that emerges by aligning the genes by the position of their +1 nucleosome (the first downstream from the NFR) exposes the generic features in a large set of genes. For individual genes, this pattern is obscured by the noise due to the limited statistics of the data.

We performed our analysis on the data of Mavrich *et al.* [10]. The red dots in Fig. 2 display the average read density when the genes are aligned to the +1 nucleosome, with the direction of transcription from left to right. Our definition of the +1 nucleosome position is the most likely position of the first nucleosome downstream from the TSS based on the list of TSSs and nucleosomes by Mavrich *et al.* [10]; see ‘Materials and Methods’ for details. On a qualitative level, the pattern of Fig. 2 (red dots) closely resembles the consensus pattern from previous studies (see, e.g., Fig. 2 in Ref. [11]). In particular, both display the same salient features, i.e., the pronounced downstream oscillations, the slow decay to a constant density, the nucleosome-free region, and the weak upstream oscillations. However, on a quantitative level, the patterns are significantly different, and only the pattern of Fig. 2 is suitable for quantitative comparison with a physical model.

Our analysis leading to Fig. 2 did not include a correction for the known sequence bias of the MNase enzyme [32,33]. However, Fig. S2 compares the pattern of Fig. 2 with the result of an alternative analysis that also incorporates a correction for the MNase bias, and suggests that the MNase bias does not significantly affect the pattern; see ‘Materials and Methods’ for details. Another concern is that the entire set of genes contains a

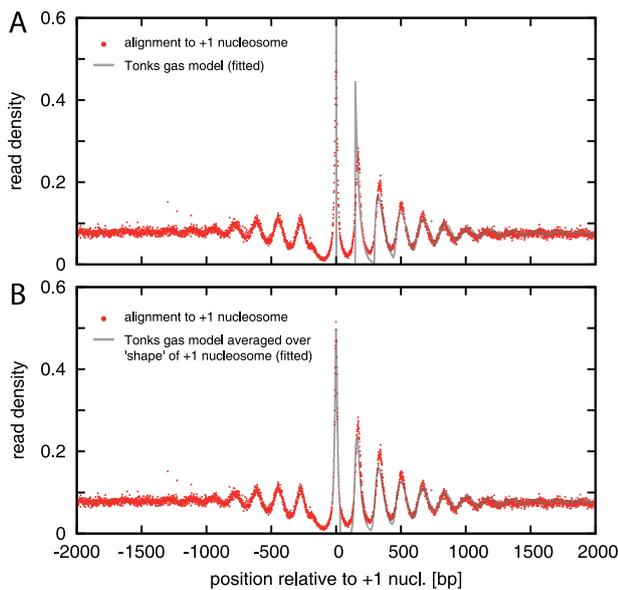


Figure 2. Comparison of the downstream nucleosome density pattern with the physical model. (A) The read density (red dots; aligned to the +1 nucleosome locations and averaged over all genes) is displayed together with the best fit by the Tonks model (gray line; least-squares fit between base pairs 200 and 2000, parameters: $b=147$ bp and $1/\bar{\rho}=177$ bp). (B) Same as in (A), but with the fit based on a convoluted Tonks gas model which takes into account the finite width of the experimental +1 peak, by convoluting the Tonks gas distribution function with the experimental probability distribution for the +1 peak in the range from -30 to 30 bp (parameters: $b=147$ bp and $1/\bar{\rho}=175$ bp). doi:10.1371/journal.pcbi.1000891.g002

significant fraction where the gene ends within the 2000 bp downstream range plotted in Fig. 2, see Fig. S3A. Therefore, we repeated our analysis on the subset of long genes with a size of more than 2000 bp in length. Fig. S3C shows that the resulting pattern is quantitatively very similar to that of Fig. 2. Taken together, these results indicate that the pattern of Fig. 2 (red dots) represents a robust quantitative signature of the nucleosome organization near transcription start sites in yeast.

Quantitative analysis

To interpret the extracted pattern within the physical model described above, we performed a nonlinear least-squares fit to Eq. (1), as described in ‘Materials and Methods’. We kept the width of the nucleosomes fixed at the value $b=147$ bp suggested by the crystal structure [3], and hence the only fit parameters were the mean nucleosome density $\bar{\rho}$ and the global normalization factor for the data (see above). The best fit is displayed as a gray line in Fig. 2A. To judge the quality of the agreement, it is useful to recall that the experimental pattern is basically described by five quantitative characteristics: the period of the oscillation, the length scale over which the oscillation decays, the asymptotic value of the density, and the amplitudes of the peaks and valleys in the density, above and below the asymptotic line. Given only two fitting parameters, the overall quantitative agreement between the physical model and the biological data is therefore remarkably good.

Fig. S4 shows the corresponding fit to only the set of long genes, with a similar result. In both cases, the most apparent deviation between the model and the data is in the shapes and the amplitudes of the first two peaks, associated with the +1 and +2

nucleosome. We wanted to test whether this is solely a consequence of the fact that Eq. (1) assumes a perfectly positioned +1 nucleosome, while the data displays a small residual positional variability for the +1 nucleosome. We therefore convoluted the model density, Eq. (1), with the shape of the +1 peak in the data (see ‘Materials and Methods’ for details). The corresponding fit of this convoluted density to the data is shown in Fig. 2B (gray line). By construction, the shape of the +1 peak now matches, but we note that the deviation in the +2 peak disappeared as well, suggesting that the finite positional variability of the +1 nucleosome is indeed sufficient to explain most of the deviation between the physical model and the biological data.

Before discussing the obtained parameter values and the robustness of the fitting procedure, we address the immediate question that emerges from the above results: On the one hand, the agreement between model and data is consistent with the hypothesis that most of the nucleosomes downstream of the +1 nucleosome are statistically positioned. On the other hand, the statistical positioning mechanism has no intrinsic bias to a particular direction, i.e., the pattern upstream of the NFR should be described as well by a viable physical model. However, the upstream consensus pattern reported in previous studies displays much less pronounced oscillations than on the downstream side [4,10]. To test whether this is simply a consequence of the gene-to-gene variation in the distance between the -1 nucleosome and the TSS, which should smear out the averaged pattern, we analyzed the statistical distribution of these distances and realigned all genes by the position of their -1 nucleosome. The -1 position is defined here by the first nucleosome upstream from the TSS, see ‘Materials and Methods’.

Fig. 3A displays the statistics of the ± 1 nucleosome positions relative to the TSS, as derived from the nucleosome map of Ref. [10]. While +1 nucleosomes are restricted to a region about 50 bp downstream from the TSS [31,34], the -1 nucleosome position is considerably more disperse. Accordingly, the distance between the +1 and -1 nucleosomes, i.e., the gap size, also has a wide distribution, see Fig. 3B. This distribution indeed smears out an oscillatory upstream pattern, which is uncovered by an alignment to the -1 nucleosome position that eliminates the gap size variation, see Fig. 4A (blue dots). However, while this upstream pattern does display regular oscillations, the comparison to the superimposed downstream pattern from Fig. 2 demonstrates that these two patterns are significantly different. A possible concern with this upstream pattern is the frequent occurrence of another

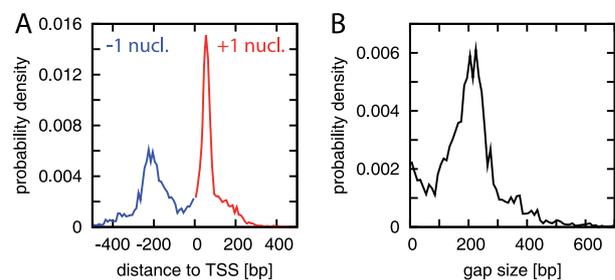


Figure 3. Distribution of +1 and -1 nucleosome positions. (A) Probability distribution of the distance of the +1 and -1 nucleosomes to the TSS, obtained as described in ‘Materials and Methods’. While the +1 nucleosome is typically found about 50 bp downstream from the TSS, the position of the -1 nucleosome is significantly more disperse. (B) Probability distribution of the gap size, i.e., the distance between the borders of the +1 and -1 nucleosomes given a nucleosome width of 147 bp. doi:10.1371/journal.pcbi.1000891.g003

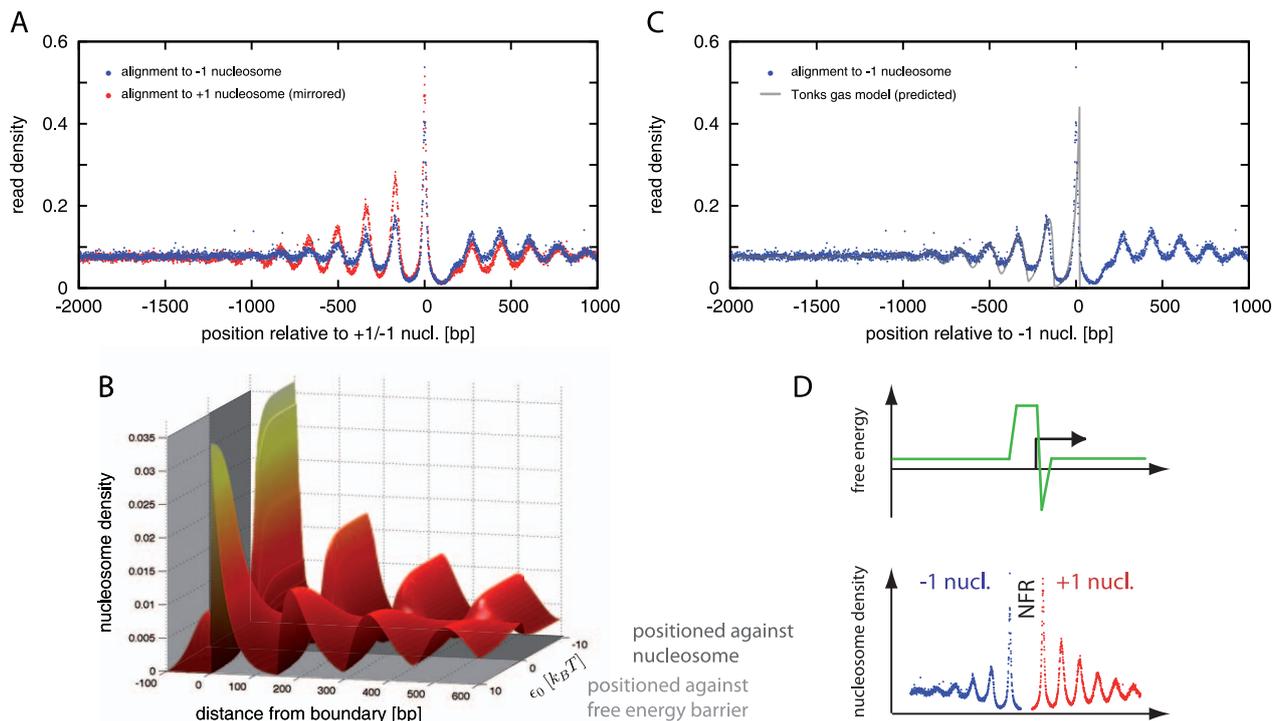


Figure 4. The upstream pattern and the effect of boundary conditions on statistical positioning. (A) Comparison of the upstream pattern in the read density (blue dots; all genes aligned to the position of their -1 nucleosome) with the (mirrored) downstream pattern of Fig. 2 (red dots). The patterns are qualitatively similar, but quantitatively significantly different. (B) 3D plot displaying the dependence of the theoretically calculated pattern on the boundary condition. The boundary condition is parameterized by the energy scale ε_0 (measured in units of $k_B T$), with $\varepsilon_0 > 0$ (light gray shaded region) representing a nucleosome repellent region, while $\varepsilon_0 < 0$ (dark gray) describes an attractive potential for a nucleosome (the width of which is chosen here to roughly correspond to the finite peak width of the $+1$ nucleosome in the data). Parameters are $b = 147$ bp and $1/\bar{\rho} \approx 175$ bp, see 'Materials and Methods' for details. (C) Comparison of the upstream pattern (blue dots) to the Tonks gas model with boundary condition for a perfectly repellent region with $\varepsilon_0 \gg 1$ (gray line; same nucleosome density and normalization as in Fig. 2A). (D) Illustration of the typical nucleosome organization around TSSs and its origin based on the conclusions of the present study. A broad repelling region combined with a localized attractive feature in the free energy landscape close to the TSS (top) leads to a NFR and a directly positioned $+1$ nucleosome (bottom). The NFR together with the $+1$ nucleosome form local boundaries which statistically position the nucleosomes in the vicinity, over ranges up to ~ 1000 bp.
doi:10.1371/journal.pcbi.1000891.g004

NFR closely upstream of the -1 nucleosome, either at the start of a divergently transcribed neighboring gene or at the $3'$ end of a gene transcribed in the same direction ($3'$ NFRs are analyzed further below). To address this concern, we selected only the subset of genes with no gene start or end within 1000 bp upstream of the TSS and compared their averaged pattern to that for all genes. Fig. S3D shows that these two patterns are quantitatively very similar (and clearly different from the downstream pattern), suggesting that the adjacent NFRs located at various distances have no significant effect on the average upstream pattern.

The difference in the up- and downstream pattern might be an indication of positioning mechanisms beyond statistical positioning. Alternatively, this difference might be due to an intrinsic asymmetry of the NFRs, caused by different molecular determinants for the up- and downstream NFR boundary. Such an asymmetry would lead to a different boundary condition for the nucleosome gas on the two sides of the NFR. To illustrate the possible effect of the boundary condition on the pattern in the nucleosome gas, Fig. 4B shows the patterns for a range of boundary conditions together in a 3D plot. Here, the different boundary conditions are parameterized by an energy scale, ε_0 , which measures the strength and the sign of the local effective free energy for nucleosome binding: Positive ε_0 (towards the front) correspond to a nucleosome repellent region, i.e., nucleosomes at positions to the left of the origin receive an energetic penalty ε_0 . In

contrast, negative ε_0 (towards the back) correspond to an attractive positioning potential that is localized to a narrow region, the width of which is chosen here to roughly correspond to the finite peak width of the $+1$ nucleosome in the data. Note that all of the patterns contained in the 3D plot of Fig. 4B are qualitatively similar, irrespective of the value of ε_0 . However, they are different on a quantitative level, and we next exploit this difference, using the experimental pattern as a quantitative signature, to infer the type of the boundary condition that is effectively implemented *in vivo*.

In particular, it is instructive to contrast the case of a perfectly repulsive barrier ($\varepsilon_0 \rightarrow \infty$) with a perfect attractive positioning potential ($\varepsilon_0 \rightarrow -\infty$). Our above analysis of the downstream pattern in Fig. 2 was based on the latter case, i.e., we assumed that most $+1$ nucleosomes are directly kept at particular positions on the genome through the action of specific molecular forces. We found that this assumption is compatible with the data. Given that the upstream pattern does not comply with this direct positioning scenario, we hypothesized that most -1 nucleosomes are instead indirectly (statistically) positioned by a repulsive barrier located at the upstream edge of the NFR. Fig. 4C displays the upstream pattern (blue dots) together with the model prediction assuming a perfectly repulsive barrier (gray line). Note that this prediction is obtained with the same values for $\bar{\rho}$ and normalization factor as inferred from the fit to the downstream pattern, i.e., it does not

involve parameter fitting, see ‘Materials and Methods’ for details. The agreement is surprisingly good, consistent with the interpretation that the positioning of most nucleosomes in the vicinity of the TSS is induced by a NFR that is intrinsically asymmetric: Our quantitative comparison suggests that the upstream boundary of the NFR is typically determined by repulsion rather than direct positioning of the -1 nucleosome.

To put these observations on a systematic basis, we performed simultaneous fits on both sides of the TSS, for all combinations of boundary conditions and compared the results quantitatively on the basis of the mean square deviation per data point, see Fig. S5, Table S3, and ‘Materials and Methods’. The results corroborate that the experimental pattern is best explained by the scenario where the $+1$ nucleosome is directly positioned whereas the -1 nucleosome is statistically positioned by a repellent region, as illustrated in Fig. 4D. The second best fit is obtained by the scenario where both the -1 and the $+1$ nucleosome are statistically positioned.

As Fig. S5 shows, both patterns are quantitatively well explained with a single average nucleosome density $\bar{\rho}=1/180$ bp for both up- and downstream of the TSS. Indeed, we find no clear evidence in the data that the average density of nucleosomes is different in intergenic and genic regions (see ‘Materials and Methods’), contrary to some observations made in other studies. We robustly obtained density values $\bar{\rho}$ in the range of one nucleosome per 172 to 180 bp, described above and independent of the detailed choice of the fitting method. These values are slightly (but consistently) larger than the “nucleosome mode” of 165 bp inferred by Mavrich *et al.* [10] by determining the typical peak to peak distance in the experimental pattern.

Finally, it is interesting to note that NFRs have also been reported at the 3′ end of genes, although their biological significance is obscure [10,35]. In order to see to what extent our findings can be generalized to this class of NFRs, we also extracted the average up- and downstream pattern for 3′ NFRs by aligning to the respective flanking nucleosomes. Fig. S6 shows these patterns; see caption for details. We observe that on neither side the pattern displays the strong features associated with the direct positioning scenario. Instead, both 3′ patterns resemble the 5′ upstream pattern, which is superimposed for comparison in Fig. S6. This suggests that the 3′ NFR is typically only a repulsive region, and hence less structured than the typical 5′ NFR.

Discussion

The recent genome-scale identification of nucleosome positions revealed that a large fraction of nucleosomes are non-randomly positioned, that a large fraction of genes have a nucleosome-free region (NFR) at their promoters, and that the NFRs are flanked by salient oscillatory patterns in the nucleosome density [9,10]. Here, we performed a quantitative analysis of the average up- and downstream patterns, to reveal hidden information about factors that affect nucleosome positioning in promoter regions. To this end, we reanalyzed previously published yeast data [10] in a physical way. We found that the up- and downstream patterns differ significantly, but both are quantitatively consistent with a minimal model where nucleosome positioning is effected only from the location of the NFRs, but radiates over a range of up to ~ 1000 bp to each side via the statistical positioning mechanism. Within this model, the difference in the average up- and downstream patterns is explained as an intrinsic asymmetry of the NFRs, which leads to different boundary conditions for the “nucleosome gas” on the two sides, see Fig. 4. In contrast, we found no evidence of such an asymmetry for 3′ NFRs at the end of genes.

That statistical positioning in the vicinity of barriers is a mechanism capable of producing a non-random nucleosome arrangement has long been established theoretically [21] and experimentally [22]. Statistical positioning of nucleosomes around promoter regions has been proposed several years ago [9], while testing of this hypothesis has started only very recently [10,36–38]. The first study [10] presented qualitative evidence for statistical positioning, but was limited by its approach relying on consensus nucleosome positions and TSS alignments. However, that study also performed a thorough statistical analysis of the DNA sequence around promoters, and found that sequence elements known to be involved in nucleosome positioning (dinucleotide patterns and poly(dA:dT) stretches) are concentrated to the NFR and the positions of the -1 and $+1$ nucleosomes, and are significantly less frequent up- and downstream from this region. This finding is consistent with our conclusions drawn from the quantitative analysis of the nucleosome patterns. Additionally, our analysis suggests that the sequence elements around the position of the -1 nucleosome are either not sufficiently widespread or not sufficiently effective to directly position the -1 nucleosome in the average pattern. This is not unlikely given that other mechanisms than direct sequence specificity are needed to obtain the precise positioning of the $+1$ nucleosome *in vivo* [39].

Two additional studies on statistical positioning in genic regions appeared after our work was completed [36,37]. These studies did not consider alignments to TSSs or $+1$ nucleosomes, but instead ranked genes by the distance between their first and last nucleosome, revealing a striking organization of the local minima in the nucleosome occupancy. This organization was found to be consistent with a Tonks gas that is constrained by repelling barriers from both sides. This analysis, with its focus on the genic regions and the positions of the minima, is complementary to ours, which focused on the quantitative shape of the average density, in particular also in the upstream intergenic region, and analyzed the difference between the up- and downstream pattern.

Taken together, our and the existing studies of statistical positioning support the view that long-range correlations in nucleosome positions produced by localized features in the effective free energy landscape for nucleosome binding are an important determinant of the genome-wide nucleosome organization. Indeed, for yeast, where TSSs are typically spaced < 2000 bp apart (Fig. S3B), statistical positioning from features encoded only at the TSSs is sufficient to obtain non-random positioning for most nucleosomes. The physical origin of statistical positioning is an interplay between the mutual exclusion and the positional entropy of nucleosomes. While this mechanism does not “glue” nucleosomes to specific locations on the DNA, it does effect that, on average, nucleosomes favor certain positions over others. It can therefore make specific (binding) sites on the DNA more (or less) accessible for proteins. Moreover, it may also cause a bias for mutation processes, thereby creating a position-dependent mutation rate [40] and possibly long-range DNA sequence correlations.

The approach taken in the present study may be classified as a “reverse approach”, which starts from the observed distribution of nucleosomes along the genome and ultimately seeks to determine from it the underlying free energy landscape for nucleosome binding (see ‘Materials and Methods’ for a discussion of the assumptions leading to the concept of an effective free energy landscape). Here, this approach has led to the typical form of local features in the landscape that is depicted in Fig. 4D. Note that by construction, our approach has two important limitations:

First, it cannot pinpoint the molecular mechanisms responsible for creating the features in the effective free energy landscape. For instance, our findings are equally compatible with sequence-

determined depletion like in the *HIS3-PET56* promoter [23], chromatin remodeler induced nucleosome organization like in the *POT1* promoter in its repressed state [41], or with varying promoter architecture in response to transcriptional perturbation [35,42]. Disentangling the molecular mechanisms on a genomic scale, requires the use of the complementary “forward approaches” based on bioinformatic methods (see, e.g., [16,18–20]) or biophysical modeling (e.g., [17,43]) to predict nucleosome positions from sequence.

Second, since reverse approaches rely on good statistics, our study is presently limited to the study of average patterns, obtained from a large number of different genes. Of course, many genes could have additional features in their free energy landscape at various positions. Again, these features could be directly encoded, by the intrinsic specificity of the DNA-histone interaction [15,16,44], or *in trans*, via competition with other specific DNA-binding proteins, biochemical histone modifications [12,45], or chromatin remodeling [8]. Such additional features do not necessarily affect the average pattern. However, our study firmly establishes the simple physical model of a Tonks gas with “programmable” boundary conditions as an excellent quantitative ‘null model’ for nucleosome positioning, which can be used as a reference point to identify specific positioning effects as deviations from it. Such a reverse approach on a gene-by-gene basis will likely be very fruitful once data with sufficient statistics and precision becomes available.

Materials and Methods

Read density as proxy for nucleosome density

The data of Mavrich *et al.* [10] is the basis for our analysis. Mavrich *et al.* extracted nucleosomal DNA from yeast cells and sequenced the DNA stretches obtaining a list of reads which they aligned to the *Saccharomyces cerevisiae* genome. Nearly perfect alignments resulted in a list of reads with start and end coordinates on the Watson or Crick strand, which we obtained from the authors. Assuming a nucleosome width of 147 bp we merged the information from both strands and assigned to each read the putative location of the midpoint of the original nucleosomal DNA sequence (see “Supplementary Information” of Ref. [31]). Originally, some reads were aligned to multiple positions on the genome giving them an artificially high weight. Therefore, we counted the number of alignments for each read (number of reads with same read identifier) and weighted the reads by the reciprocal number of their occurrence. For example, if alignment to the yeast genome resulted in 5 hits, each alignment was weighted by a factor 1/5. The frequency of reads vs. location on the genome defines the read density map serving as our proxy for nucleosome density and is denoted by Σ_{reads} below. A small region of the read density map is sketched in Fig. 1A and Fig. S2A.

Genes, ± 1 nucleosome positions and alignments

Our list of start and end sites of genes is based on the list of transcribed regions and open reading frames as reported in “Supplementary Research Data” of Ref. [10] (file Supplementary_Table_S2.xls). We combined the start sites of transcribed regions (class: *pol II*, subclass: *mRNA*) and the end sites of open reading frames (class: *pol II*, subclass: *ORF*) with same ‘feature ID’ to one ‘gene’ with a total of 4792 genes. See Figs. S3A and B for statistics of length of the genes and distances between TSSs.

We used alignments of the read density map to the positions of nucleosomes surrounding the nucleosome free region (NFR) at the TSS for a quantitative test of statistical positioning. Since the read density map we used for our analysis does not allow direct

annotation of individual nucleosomes, we had to employ the list of identified nucleosomes from the “Supplementary Research Data” of Ref. [10] (file Supplementary_Table_S1.xls). We used the definition of the +1 nucleosome as the first nucleosome at or downstream from the transcription start site (TSS) while the -1 nucleosome is defined as the first nucleosome upstream from the TSS. The probability distributions of the ± 1 nucleosome’s distance to the TSS are peaked at some distance from the TSS (Fig. 3A) such that a slightly different definition of the ± 1 nucleosome has no significant effect on the results. Next, we aligned the read density map to the position of these nucleosomes and averaged (Fig. 2 and Fig. 4A).

To test the influence of gene starts and ends close to the ± 1 nucleosomes of interest, we additionally created alignments using only genes larger than 2000 bp (Fig. S3C) and using those genes without gene start or end sites within 1000 bp upstream from the TSS (Fig. S3D).

An alternative proxy for nucleosome density

The read density map we derived does not include any correction for sequence bias of micrococcal nuclease (MNase). To test for the importance of such a correction, we performed an alternative analysis towards a nucleosome density around the ± 1 nucleosomes. To that end, we exploited the list of nucleosomes as identified by Mavrich *et al.*: Based on the reads aligned to the yeast genome, these authors identified individual nucleosomes using a peak detection algorithm after correcting for MNase bias (see “Supplementary Information” of Ref. [31]). The emerging list of nucleosomes also includes the standard deviation (measure of fuzziness) for each nucleosome (“Supplementary Research Data” of Ref. [10], file Supplementary_Table_S1.xls). Interpreting the nucleosome’s standard deviation as a cell to cell variation (instead of an experimental error), we represented each nucleosome with assigned standard deviation larger than 3 by a Gaussian distribution with standard deviation given by the nucleosome’s standard deviation. This results in an alternative proxy Σ_{peaks} for the nucleosome density as sketched in Fig. S2B. Both proxies for nucleosome density, the one based on the raw data (reads) and the one based on processed data (individual nucleosomes), significantly differ locally (compare Figs. S2A and S2B). The corresponding alignments to the +1 and -1 nucleosomes, however, are practically identical (Fig. S2C) having accounted for genome-wide normalization (997655 reads correspond to 52918 nucleosomes). This indicates that MNase bias correction as performed by Mavrich *et al.* is not essential for our analysis.

As a side-remark note that the proxy Σ_{peaks} at first sight should represent nucleosome density without any further normalization. However, repeating parts of our fitting analysis (see below) with Σ_{peaks} instead of Σ_{reads} revealed that a fit to the Tonks gas model is only possible if we allow for a normalization factor significantly different from unity (≈ 0.8), suggesting that the proxy Σ_{peaks} underestimates the number of nucleosomes. A possible explanation is that up to 20 percent of the nucleosomes were missed by the peak detection filter applied by Mavrich *et al.*. This explanation appears likely, since not all of the yeast nucleosomes are well positioned, i.e., a significant portion of nucleosomes will not lead to a clear peak in the distribution, given the average over many cells that is taken in the experiment.

Model details and assumptions

Tonks gas. Our one-dimensional gas description of statistical nucleosome positioning uses a continuous genome coordinate x , whereas in reality nucleosome positions only take on discrete values, in steps of single base pairs (bp). The continuum limit is

convenient and justified as long as the average distance between particles, i.e., the linker length, is relatively large compared to the discretization step (the average linker length is typically in the range of 15–70 bp, depending on the organism). The statistical physics of a gas of finite-sized particles in a one-dimensional continuous state space has long been worked out in detail [25–27], but for a self-contained presentation we derive the explicit form of the oscillatory pattern using a simple physical argument. To this end we consider the two-particle distribution function $\rho_2(x, x')$ which measures the probability that a particle is found at x and another particle is found at x' . Mathematically,

$$\rho_2(x, x') = \left\langle \sum_{i, j \neq i} \delta(x_i - x) \delta(x_j - x') \right\rangle,$$

where $\langle \dots \rangle$ denotes the average over all possible configurations and $\delta(x)$ denotes the Dirac delta function. In the thermodynamic limit, where the number of particles, N , and the length of the interval, L , are both large (given an average density $\bar{\rho} = N/L$) and for x and x' far away from the boundaries, $\rho_2(x, x')$ does not depend on x and x' independently, but is only a function of the distance, $\rho_2(x, x') = \rho_2(r)$ with $r = |x - x'|$. To obtain $\rho_2(r)$ explicitly, first note that, with regard to the spaces between particles, a Tonks gas of N particles with width b in an interval of length L is equivalent to a one-dimensional gas of point-like particles in an interval of length $L - Nb$ (for clarity, consider periodic boundary conditions). In the bulk, these point particles are randomly distributed, such that the gap size between neighboring particles has an exponential distribution $p_1(r) = \exp(-r/d)/d$ with mean $d = (L - Nb)/N$. For this gas, the probability $p_k(r)$ of finding the k -th neighbor particle at distance r is equivalent to the probability that the sizes of k neighboring gaps sum up to r . Since the gap sizes are independent, $p_k(r)$ can be expressed as a convolution of $p_1(r)$ distributions and the Laplace transform $\hat{p}_k(s)$ factorizes, $\hat{p}_k(s) = (\hat{p}_1(s))^k$. Inverse Laplace transformation yields

$$p_k(r) = \frac{r^{k-1} \exp(-r/d)}{d^k \cdot (k-1)!},$$

which is also referred to as the Erlang distribution. The corresponding function for the original Tonks gas is then obtained by reintroducing the particle width b . This only amounts to shifting the distance r in $p_k(r)$ by kb and assuring that the resulting function is identical to zero for $r < kb$ by use of the Heaviside step function $\Theta(r)$. The probability of finding the k -th particle at a distance r then becomes

$$p_k(r - bk)\Theta(r - bk) = : q_k(r)/\bar{\rho},$$

where we introduced the function $q_k(r)$ and used $d \equiv 1/\bar{\rho} - b$. The two-particle distribution function is obtained by multiplying the density of the first particle, $\bar{\rho}$, with the probability to find any particle at distance r , regardless of k , which amounts to the sum

$$\rho_2(r) = \sum_{k=1}^{\infty} q_k(r).$$

The first few terms $q_k(r)$ are displayed together with the total sum $\rho_2(r)$ in Fig. S1A using typical parameters for nucleosomes. Note that the density $\rho(r)$ of particles close to a boundary particle with perfectly fixed position is simply $\rho_2(r)/\bar{\rho}$. Thus, $q_k(r)/\bar{\rho}$ is the

probability of finding the k -th particle at a distance r from the boundary. It is interesting to observe that the distance between the maxima of the oscillatory pattern shown in Fig. S1A differs from $1/\bar{\rho}$. This difference is significant only at smaller densities as plotted in Fig. S1B. Note, however, that the average position of the k -th particle does not coincide with the maxima of $q_k(r)$ or the maxima of $\rho(r)$, but is simply $k/\bar{\rho}$.

Physically, the oscillations of $\rho_2(r)$ and equally of $\rho(r)$ are a signature of a collective effect, which results from an interplay between the excluded volume interaction and entropy. Very close to a given particle ($r < b$), there is a “depletion layer” which no particle midpoint can access, hence $\rho_2(r) = 0$ for $r < b$. Then, only the leftmost particle can access the first layer $b < r < 2b$. The further this first particle moves to the right, the further it compresses the remainder of the gas. In reaction, the gas exerts a pressure onto the first particle to stay close to the boundary particle, and hence $\rho_2(r) = q_1(r)$ decays within the first layer. However, in the second layer ($2b < r < 3b$), both $q_1(r)$ and $q_2(r)$ contribute and $\rho_2(r)$ increases again. Finally, an oscillatory pattern of $\rho_2(r)$ emerges from summing the individual peaked functions $q_k(r)$ (which are non-zero for $r > kb$ only and decaying for large r). The peaks in $\rho_2(r)$ wash out with increasing r since the individual $q_k(r)$ become broader and more k values contribute. The limiting value of $\rho_2(r)$ is $\bar{\rho}^2$, i.e., the square of the mean density. With increasing mean density ($\bar{\rho} \rightarrow 1/b$), the individual $q_k(r)$ become sharper and overlap less; oscillations in $\rho_2(r)$ become more pronounced and range further.

Different boundary conditions. Eq. (1) provides an analytic expression for the particle density close to a perfect boundary. This expression can in fact be interpreted and utilized in two different ways: (i) The origin, $r = 0$, can be interpreted as the location of a perfectly positioned nucleosome, which then acts as a perfect boundary for the neighboring nucleosomes. (ii) The origin can be the location of a barrier of another type, e.g., a nucleosome-repelling DNA sequence or bound transcription factors and only the series of peaks for $r > 0$ correspond to nucleosomes. The difference amounts to a horizontal shift: in the former case the $r = 0$ point of the theoretical pattern must be aligned with the first nucleosome, whereas in the latter case the $k = 1$ peak must be aligned with the first nucleosome. This simple shift switches between the two opposite extremes in the range of possible boundary conditions, i.e., perfect direct positioning vs. pure indirect positioning against a perfect barrier. For our quantitative data analysis we limited ourselves to these two extreme cases (see ‘Procedure for quantitative analysis’ below), however in Fig. 4B we also explored the effect of more realistic conditions where neither perfect attraction of a nucleosome to a single point on the genome occurs (e.g., note the finite width of the peak associated with the +1 nucleosome in Fig. 2) nor perfect repulsion. To generate Fig. 4B, we numerically determined the particle density close to a broad repellent region of varying strength, and also close to a narrow attractive region of varying depth and finite width (binding energy here is defined to act on the particle midpoint). We computed the density for a grand-canonical ensemble, using a recursion relation of the same type as described in Ref. [43], and with the chemical potential adjusted such that an average inter-particle spacing of $1/\bar{\rho} \approx 175$ bp was obtained.

Model assumptions. As stated in the main text, our application of the Tonks gas model to the nucleosome data is based on a number of simplifying assumptions. For instance, we assumed that the variation in nucleosome position indicated by the distribution of reads is a true reflection of the cell-to-cell variability. In practice, the nucleosome positions inferred from the reads have some (unknown) experimental error. However, a

posteriori our assumption appears reasonable, due to the quantitative agreement of model and data, which suggests that the decaying oscillations genuinely reflect the many-body physics of the Tonks gas – such an agreement is not expected if the variation were merely experimental error. Another assumption, shared with basically all models for nucleosome organization, is the equilibrium assumption made by associating the nucleosome distribution with a static free energy landscape. *In vivo*, transcription, DNA replication, and active remodeling processes regularly translocate and evict nucleosomes, and it is questionable to what extent these processes can be captured by a static free energy landscape. Though little is known about the kinetics of chromatin reorganization, we can consider some simple scenarios to illustrate that this assumption may not be as bad as it seems: For instance, remodeling enzymes that merely increase the mobility of nucleosomes, without preference for a certain direction or position, would only speed up the equilibration in a free energy landscape, but not affect its shape. If the remodelers do have any sort of bias, but work rapidly, their effect can be included into a modified free energy landscape. Other passive (competitive binding) and active (repositioning) processes can similarly be included in an effective free energy landscape, as long as their kinetics is rapid on the timescale of interest. Remodelers may also modify the interaction potential between the nucleosomes, beyond the simple hard-core repulsion of the Tonks model. Other effects, including transient unwrapping of the nucleosomal DNA [5,6,27] and geometric constraints in higher order structures may modify the interaction between nucleosomes as well. In statistical physics, more complicated interactions between particles in one dimensional gases have been considered [30], however due to the good agreement between the data and the simple Tonks model, we did not consider generalizations in this direction.

Procedure for quantitative analysis

To systematically compare the quantitative model to the ± 1 nucleosome alignments of the read density (i.e., our proxy \sum_{reads} for nucleosome density), we performed least squares fits using the function

$$f(r) = \lambda \cdot \rho(r - \Delta r), \quad (2)$$

where λ is a normalization factor, Δr tests for a possible horizontal offset in the data, and the function $\rho(r)$ from Eq. (1) contains the parameters $\bar{\rho}$ and b . In all our fits, the nucleosome width was kept fixed at $b = 147$ bp. We used the offset parameter Δr also to distinguish between the two opposite boundary conditions considered for our fits: As explained above, $\Delta r = 0$, corresponds to the direct positioning scenario where the first nucleosome is a fixed barrier for the neighboring nucleosomes, while a shift by one nucleosome width corresponds to the statistical positioning scenario where the boundary is not a nucleosome, but another repellent feature on the genome. (In the latter case, the different genes should in principle be aligned to the location of the boundary, but since this is not possible, our alignment to the first nucleosome is the best alternative.) For each of our fits, one of these two scenarios is imposed by choice of the starting value for Δr , since each scenario corresponds to a deep “basin” in the least-squares score function. As can be seen from the Tables in the Supporting Material, each best-fit value for Δr either clearly corresponds to the direct positioning scenario, $\Delta r \approx 0$, or to the indirect positioning scenario, $-1/\bar{\rho} < \Delta r < -b$. We performed fits to -1 nucleosome alignment data in the same way as for $+1$ nucleosome alignment data, except that we mirrored the data at the origin. For the fits, we used the data in a range from 200 to

2000 bp downstream from the $+1$ nucleosome and upstream from the -1 nucleosome, respectively. Altering the fitting range to 200–1200 bp had no significant effect on the results. To ensure best possible parameter estimates, we performed each fit 300 times from a wide range of starting parameters. Best fits are shown in Figs. 2A, S4, and 3C (where a peak at Δr has been added where applicable to indicate the directly positioned nucleosome, i.e., for the case $\Delta r \approx 0$). The corresponding parameter estimates are displayed in Table S1 where δ denotes the squared deviation per data point between data and model.

The parameter estimates from fits to the $+1$ nucleosome alignment are robust against variations in details of the fitting procedure: (i) Fitting to the average over all genes yields almost the same parameter estimates as a fit to an average where only genes larger than 2000 bp are considered. For the latter, nucleosome density is estimated slightly higher due to the slightly further ranging oscillations (Fig. S3C), but it does not significantly differ from the estimate obtained from the alignment including all genes (see Figs. 2A and S4, Table S1). (ii) Randomly partitioning the set of 4792 genes over which the average is performed into four subsets and repeating the fitting analysis yielded almost identical results, see Table S2. (iii) To account for the effect of the residual cell-to-cell variation in the position of the $+1$ nucleosome, we also performed a fit using a ‘convoluted Tonks model’, where Eq. (2) (with $\Delta r = 0$) was convoluted with a probability distribution function corresponding to the experimental nucleosome density in the range of ± 30 bp around zero. The first peak downstream from the $+1$ nucleosome, corresponding to the $+2$ nucleosome, is much better characterized by this fit (compare Figs. 2A and B) suggesting that cell-to-cell variation of the $+1$ nucleosome’s position is reflected in cell-to-cell variations of the downstream nucleosomes. Yet, parameter estimates are very similar to those obtained from the fit without convolution (Table S1) indicating that including this effect is not essential when fitting the Tonks gas model to the data in the range of 200 to 2000 bp as we do everywhere else in this study.

For the fit to just the -1 nucleosome alignment, we used the parameter estimates for nucleosome density $\bar{\rho}$ and normalization λ obtained from the fit to the $+1$ nucleosome alignment. Thus, the only remaining fit parameter here was the offset Δr (Table S1), which was started at values $\Delta r < -b$. In order to systematically test alternative scenarios (e.g., direct positioning of the -1 nucleosome and indirect positioning of the $+1$ nucleosome), we performed simultaneous fits to both the $+1$ and -1 alignment data for each of the four possible boundary conditions. Fits were carried out analogously to the procedure described above, but with the λ and $\bar{\rho}$ parameters constrained to take the same values on both sides. Fig. S5 displays the results, Table S3 shows the parameter estimates. Regarding the mean squared deviation per data point δ , scenarios C and D are similar, while scenarios A and B are less probable. In both eligible scenarios, the -1 nucleosome is indirectly positioned. In the best fit scenario C the $+1$ nucleosome is directly positioned.

Comparison of average nucleosome densities

In our systematic fitting procedure described above, we assumed the same average nucleosome density up- and downstream from the NFR. This must be justified by comparing the average density in intergenic regions to that in genic regions. To estimate their ratio, we used the proxies for nucleosome density described above, i.e., the read density (\sum_{reads}) and the representation of nucleosomes by Gaussians with appropriate width (\sum_{peaks}). To exclude the influence of the 5' NFR, which is mostly located within intergenic regions, we excluded the NFR regions. Using proxy \sum_{reads} we obtained a ratio of 1.00 for the density in intergenic to the density in

genic regions, whereas a ratio of 0.85 resulted from using \sum_{peaks} . We conclude that there is no clear indication of a density bias between intergenic and genic regions (apart from the existence of NFRs). We therefore assumed equal average density up- and downstream from the TSS for the fitting procedure.

Supporting Information

Figure S1 Characteristics of the Tonks gas two-particle distribution function. (A) Two-particle distribution function $\rho_2(r)$ for a particle size of $b=147$ and an average particle spacing $1/\bar{\rho}=178$. The first few individual terms $q_k(r)$ contributing to $\rho_2(r)$ are superimposed. (B) Distance between the individual peaks in $\rho_2(r)$ as a function of $\bar{\rho}$ for $b=147$. For dense packing, the first few maxima are equidistantly spaced by $1/\bar{\rho}$. Note that the first peak is always located at $r=b$, regardless of the particle density. Found at: doi:10.1371/journal.pcbi.1000891.s001 (0.32 MB PDF)

Figure S2 Comparison between two proxies for the nucleosome density. (A) Section of the read density map (\sum_{reads}) based on sequence reads reported by Mavrich *et al.* [10]. (B) Section of nucleosome density estimate based on the list of nucleosomes identified by Mavrich *et al.* (\sum_{peaks}): Each nucleosome is represented by a Gaussian with mean and standard deviation corresponding to the values reported. (C) Alignment of both nucleosome density proxies (red dots for \sum_{reads} , green dots for \sum_{peaks}) to ± 1 nucleosome positions and averaging over all genes leads to nearly identical results. To account for the unknown normalization, we scaled the read density map such that the genome-wide number of reads equals the genome-wide number of identified nucleosomes. Found at: doi:10.1371/journal.pcbi.1000891.s002 (0.46 MB PDF)

Figure S3 Distribution of gene start and end sites and effects on alignments. (A) Probability distribution (black) and cumulative distribution (red) for the length of genes. Typical sizes of genes are about 1000 bp, but about nearly a third is larger than 2000 bp. (B) Same distributions for the distance between neighboring TSSs. Distances are in general comparable to the size of genes, but a number of TSSs are very close to each other. (C) Alignment of read density to +1 nucleosome and average over all genes (red dots) and those 1269 genes being larger than 2000 bp only (gray dots). The averages are very similar, but close inspection shows that amplitudes are slightly larger and oscillations range further when considering large genes only. (D) Alignment of read density to -1 nucleosome and average over all genes (blue dots) and those 952 genes where no gene starts or ends were found within 1000 bp upstream of the TSS (gray dots). The averages are very similar, but amplitudes are slightly smaller when considering those genes without other gene starts or ends upstream only. Found at: doi:10.1371/journal.pcbi.1000891.s003 (0.52 MB PDF)

Figure S4 Best fit of Tonks gas model (gray line) to +1 nucleosome alignment of read density including genes larger than 2,000 bp only (red dots). Visual inspection yields good agreement between model and data, comparable to the analogous fit to the data including all genes (Fig. 2A, see also Fig. S3C). For estimated parameters see Table S1. Found at: doi:10.1371/journal.pcbi.1000891.s004 (0.41 MB PDF)

Figure S5 Best simultaneous fits (gray lines) to -1 and +1 nucleosome alignments of read density (blue and red dots, respectively) given the four possible boundary conditions (both the +1 and -1 nucleosome may be directly or indirectly positioned) with nucleosome density and normalization being equal for both alignments (see ‘Materials and Methods’ for details). Regarding the mean squared deviation per data point, scenario C describes the data

best, i.e., the scenario where the +1 nucleosome is directly positioned while the -1 nucleosome is indirectly positioned (Table S3).

Found at: doi:10.1371/journal.pcbi.1000891.s005 (1.00 MB PDF)

Figure S6 Nucleosome organization around the 3' end of genes. (A) Sketch of a typical nucleosome organization around both the 5' and 3' ends of genes. Throughout this study, the focus is primarily on the 5' NFR with its flanking -1 and +1 nucleosomes. The nucleosomes flanking the 3' NFR are here referred to as the 3' end -1 nucleosome and the 3' end +1 nucleosome. We determined the positions of 3' end ± 1 nucleosomes in analogy to the 5' end ± 1 nucleosomes: The 3' end -1 nucleosome is defined as the nucleosome at or first nucleosome upstream of the ORF end while the 3' end +1 nucleosome is the first nucleosome downstream. (B) Alignment of read density to the 3' end -1 nucleosome (left) and 3' end +1 nucleosome (right), respectively (green data points). For comparison, the alignment to the 5' end -1 nucleosome is also shown (blue data points, from Fig. 4, mirrored on the right). Overall, a good agreement is visible between the alignments to the nucleosomes flanking the 3' NFR on both sides and the alignment to the 5' end -1 nucleosome. This indicates that at the 3' end the nucleosomes are only statistically positioned against a repulsive barrier, which we found to be the most likely scenario for the pattern upstream of the 5' NFR. (Note the small bump in the read density within the nucleosome depleted region, just downstream of the 3' end -1 nucleosome and upstream of the 3' end +1 nucleosome; it indicates that the identification of 3' NFRs is not perfect or a certain fraction of genes does not display a 3' NFR.)

Found at: doi:10.1371/journal.pcbi.1000891.s006 (0.42 MB PDF)

Table S1 Parameter estimates from independent fits of Tonks gas model to ± 1 nucleosome alignments of read density based on Equation (2) (density $\bar{\rho}$, normalization λ , offset Δr , squared deviation per data point δ). Numbers in parentheses indicate values that were set fixed rather than estimated from the fit. See ‘Materials and Methods’ for details.

Found at: doi:10.1371/journal.pcbi.1000891.s007 (0.03 MB PDF)

Table S2 Parameter estimates (density $\bar{\rho}$, normalization λ , offset Δr , squared deviation per data point δ) from fits of Tonks gas model to +1 nucleosome alignments of read density using subsets of genes only. Four times (partitioning A–D), the set of 4792 genes was divided into four equal-sized subsets (subset 1–4) before fitting. Estimated parameters are very similar; see ‘Materials and Methods’ for details. Found at: doi:10.1371/journal.pcbi.1000891.s008 (0.03 MB PDF)

Table S3 Parameter estimates for simultaneous fits of Tonks gas model to ± 1 nucleosome alignments of read density. Both normalization λ and nucleosome density $\bar{\rho}$ are constrained to be equal for both alignments. Δr_{+1} and Δr_{-1} are independent parameters accounting for different boundary conditions. Regarding the mean squared deviation per data point δ , scenario C describes the data best, i.e., the scenario where the +1 nucleosome is directly positioned while the -1 nucleosome is indirectly positioned (Fig. S5). See ‘Materials and Methods’ for details. Found at: doi:10.1371/journal.pcbi.1000891.s009 (0.04 MB PDF)

Acknowledgments

We are grateful to Frank Pugh and Cizhong Jiang for providing the data and many useful comments. We also thank Ho-Ryun Chung and Jonathan Widom for valuable discussions.

Author Contributions

Conceived and designed the experiments: WM UG. Analyzed the data: WM. Wrote the paper: WM UG.

References

- Kornberg RD, Lorch Y (1999) Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell* 98: 285–94.
- Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389: 251–60.
- Davey CA, Sargent DF, Luger K, Maeder AW, Richmond TJ (2002) Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *J Mol Biol* 319: 1097–113.
- Lee W, Tillo D, Bray N, Morse RH, Davis RW, et al. (2007) A high-resolution atlas of nucleosome occupancy in yeast. *Nat Genet* 39: 1235–44.
- Li G, Levitus M, Bustamante C, Widom J (2005) Rapid spontaneous accessibility of nucleosomal DNA. *Nat Struct Mol Biol* 12: 46–53.
- Möbius W, Neher RA, Gerland U (2006) Kinetic accessibility of buried DNA sites in nucleosomes. *Phys Rev Lett* 97: 208102.
- Flaus A, Owen-Hughes T (2003) Mechanisms for nucleosome mobilization. *Biopolymers* 68: 563–578.
- Clapier CR, Cairns BR (2009) The biology of chromatin remodeling complexes. *Annu Rev Biochem* 78: 273–304.
- Yuan GC, Liu YJ, Dion MF, Slack MD, Wu LF, et al. (2005) Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science* 309: 626–30.
- Mavrich TN, Ioshikhes IP, Venters BJ, Jiang C, Tomsho LP, et al. (2008) A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Res* 18: 1073–83.
- Jiang C, Pugh BF (2009) Nucleosome positioning and gene regulation: advances through genomics. *Nat Rev Genet* 10: 161–72.
- Rando OJ, Chang HY (2009) Genome-wide views of chromatin structure. *Annu Rev Biochem* 78: 245–71.
- Segal E, Widom J (2009) What controls nucleosome positions? *Trends Genet* 25: 335–43.
- Radman-Livaja M, Rando OJ (2010) Nucleosome positioning: How is it established, and why does it matter? *Dev Biol* 339: 258–266.
- Widom J (2001) Role of DNA sequence in nucleosome stability and dynamics. *Q Rev Biophys* 34: 269–324.
- Segal E, Fondufe-Mittendorf Y, Chen L, Thåström A, Field Y, et al. (2006) A genomic code for nucleosome positioning. *Nature* 442: 772–8.
- Vaillant C, Audit B, Arneodo A (2007) Experiments confirm the influence of genome long-range correlations on nucleosome positioning. *Phys Rev Lett* 99: 218103.
- Yuan GC, Liu JS (2008) Genomic sequence is highly predictive of local nucleosome depletion. *PLoS Comput Biol* 4: e13.
- Field Y, Kaplan N, Fondufe-Mittendorf Y, Moore IK, Sharon E, et al. (2008) Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. *PLoS Comput Biol* 4: e1000216.
- Chung HR, Vingron M (2009) Sequence-dependent nucleosome positioning. *J Mol Biol* 386: 1411–22.
- Kornberg R, Stryer L (1988) Statistical distributions of nucleosomes: nonrandom locations by a stochastic mechanism. *Nucleic Acids Res* 16: 6677–90.
- Fedor MJ, Lue NF, Kornberg RD (1988) Statistical positioning of nucleosomes by specific protein-binding to an upstream activating sequence in yeast. *J Mol Biol* 204: 109–27.
- Sekinger EA, Moqtaderi Z, Struhl K (2005) Intrinsic histone-DNA interactions and low nucleosome density are important for preferential accessibility of promoter regions in yeast. *Mol Cell* 18: 735–48.
- Kornberg R (1981) The location of nucleosomes in chromatin: specific or statistical. *Nature* 292: 579–80.
- Tonks L (1936) The complete equation of state of one, two and three-dimensional gases of hard elastic spheres. *Phys Rev* 50: 955–963.
- Salsburg ZW, Zwanzig RW, Kirkwood JG (1953) Molecular distribution functions in a one-dimensional fluid. *J Chem Phys* 21: 1098–1107.
- Chou T (2003) An exact theory of histone-DNA adsorption and wrapping. *Europhys Lett* 62: 753–759.
- Schwab DJ, Bruinsma RF, Rudnick J, Widom J (2008) Nucleosome switches. *Phys Rev Lett* 100: 228105.
- Segal E, Widom J (2009) Poly(dA:dT) tracts: major determinants of nucleosome organization. *Curr Opin Struc Biol* 19: 65–71.
- Mattis DC, ed. (1993) *The Many-Body Problem: An Encyclopedia of Exactly Solved Models in One Dimension*. Singapore: World Scientific Publishing.
- Albert I, Mavrich TN, Tomsho LP, Qi J, Zanton SJ, et al. (2007) Translational and rotational settings of H2A.Z nucleosomes across the *Saccharomyces cerevisiae* genome. *Nature* 446: 572–6.
- Dingwall C, Lomonosoff GP, Laskey RA (1981) High sequence specificity of micrococcal nuclease. *Nucleic Acids Res* 9: 2659–73.
- Hörz W, Altenburger W (1981) Sequence specific cleavage of DNA by micrococcal nuclease. *Nucleic Acids Res* 9: 2643–58.
- Whitehouse I, Rando OJ, Delrow J, Tsukiyama T (2007) Chromatin remodelling at promoters suppresses antisense transcription. *Nature* 450: 1031–5.
- Shivaswamy S, Bhinge A, Zhao Y, Jones S, Hirst M, et al. (2008) Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation. *PLoS Biol* 6: e65.
- Vaillant C, Palmeira L, Chevereau G, Audit B, d'Aubenton-Carafa Y, et al. (2010) A novel strategy of transcription regulation by intra-genic nucleosome ordering. *Genome Res* 20: 59–67.
- Chevereau G, Palmeira L, Thermes C, Arneodo A, Vaillant C (2009) Thermodynamics of intragenic nucleosome ordering. *Phys Rev Lett* 103: 188103.
- Milani P, Chevereau G, Vaillant C, Audit B, Haftek-Terreau Z, et al. (2009) Nucleosome positioning by genomic excluding-energy barriers. *P Natl Acad Sci USA* 106: 22257–62.
- Zhang Y, Moqtaderi Z, Rattner BP, Euskirchen G, Snyder M, et al. (2009) Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo. *Nat Struct Mol Biol* 16: 847–852.
- Sasaki S, Mello CC, Shimada A, Nakatani Y, Hashimoto SI, et al. (2009) Chromatin-associated periodicity in genetic variation downstream of transcriptional start sites. *Science* 323: 401–4.
- Whitehouse I, Tsukiyama T (2006) Antagonistic forces that position nucleosomes in vivo. *Nat Struct Mol Biol* 13: 633–40.
- Zawadzki KA, Morozov AV, Broach JR (2009) Chromatin-dependent transcription factor accessibility rather than nucleosome remodeling predominates during global transcriptional restructuring in *Saccharomyces cerevisiae*. *Mol Biol Cell* 20: 3503–13.
- Morozov AV, Fortney K, Gaykalova DA, Studitsky VM, Widom J, et al. (2009) Using DNA mechanics to predict in vitro nucleosome positions and formation energies. *Nucleic Acids Res* 37: 4707–4722.
- Kaplan N, Moore IK, Fondufe-Mittendorf Y, Gossett AJ, Tillo D, et al. (2009) The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* 458: 362–6.
- Liu CL, Kaplan T, Kim M, Buratowski S, Schreiber SL, et al. (2005) Single-nucleosome mapping of histone modifications in *S. cerevisiae*. *PLoS Biol* 3: e328.

Supporting Material for “Quantitative test of the barrier nucleosome model for statistical positioning of nucleosomes up- and downstream of transcription start sites”

Wolfram Möbius and Ulrich Gerland

Figure S1

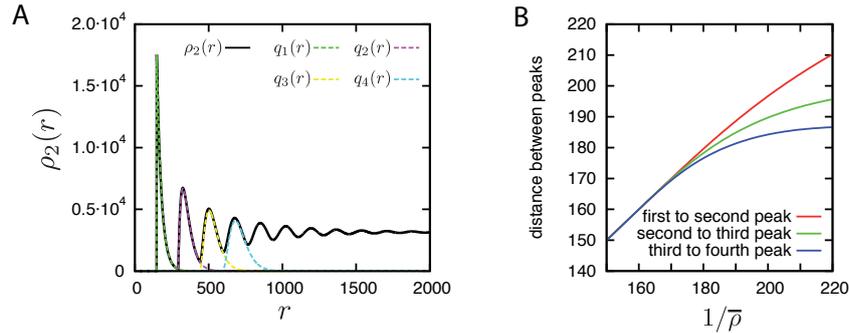


FIG. S1: Characteristics of the Tonks gas two-particle distribution function. (A) Two-particle distribution function $\rho_2(r)$ for a particle size of $b = 147$ and an average particle spacing $1/\bar{\rho} = 178$. The first few individual terms $q_k(r)$ contributing to $\rho_2(r)$ are superimposed. (B) Distance between the individual peaks in $\rho_2(r)$ as a function of $\bar{\rho}$ for $b = 147$. For dense packing, the first few maxima are equidistantly spaced by $1/\bar{\rho}$. Note that the first peak is always located at $r = b$, regardless of the particle density.

Figure S2

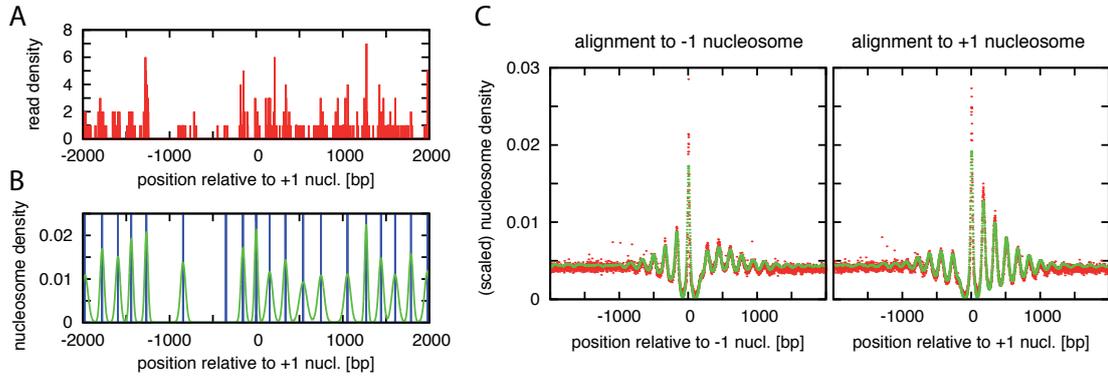


FIG. S2: Comparison between two proxies for the nucleosome density. (A) Section of the read density map (Σ_{reads}) based on sequence reads reported by Mavrich *et al.* [10] (Genome Res 18:1073–83 (2008)). (B) Section of nucleosome density estimate based on the list of nucleosomes identified by Mavrich *et al.* (Σ_{peaks}): Each nucleosome is represented by a Gaussian with mean and standard deviation corresponding to the values reported. (C) Alignment of both nucleosome density proxies (red dots for Σ_{reads} , green dots for Σ_{peaks}) to ± 1 nucleosome positions and averaging over all genes leads to nearly identical results. To account for the unknown normalization, we scaled the read density map such that the genome-wide number of reads equals the genome-wide number of identified nucleosomes.

Figure S3

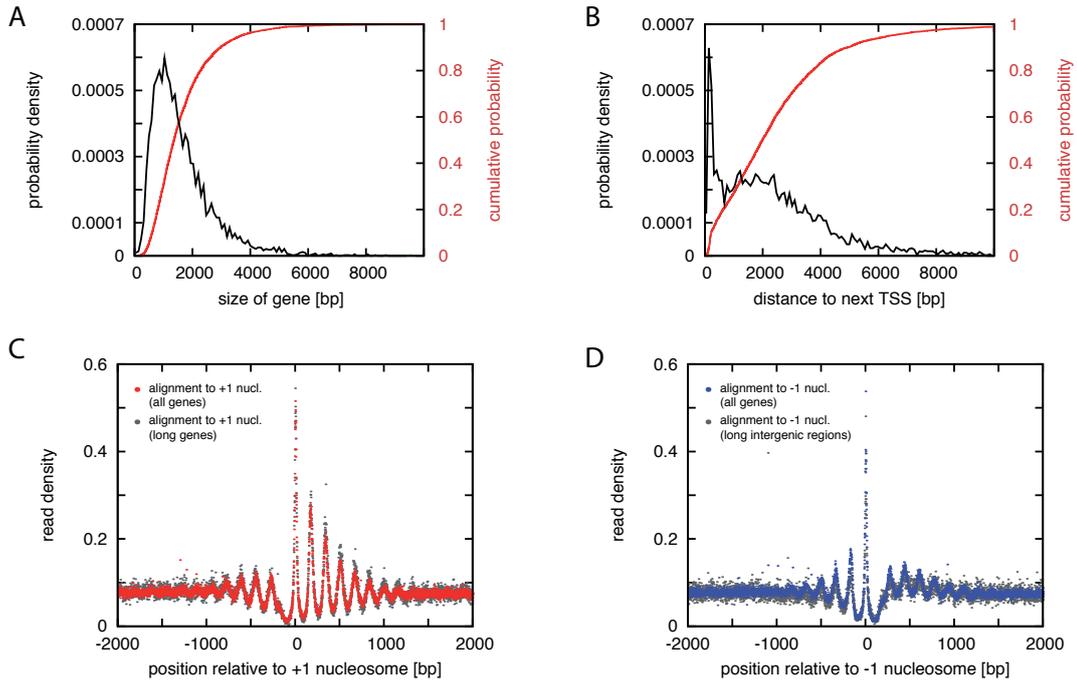


FIG. S3: Distribution of gene start and end sites and effects on alignments. (A) Probability distribution (black) and cumulative distribution (red) for the length of genes. Typical sizes of genes are about 1000 bp, but about nearly a third is larger than 2000 bp. (B) Same distributions for the distance between neighboring TSSs. Distances are in general comparable to the size of genes, but a number of TSSs are very close to each other. (C) Alignment of read density to +1 nucleosome and average over all genes (red dots) and those 1269 genes being larger than 2000 bp only (gray dots). The averages are very similar, but close inspection shows that amplitudes are slightly larger and oscillations range further when considering large genes only. (D) Alignment of read density to -1 nucleosome and average over all genes (blue dots) and those 952 genes where no gene starts or ends were found within 1000 bp upstream of the TSS (gray dots). The averages are very similar, but amplitudes are slightly smaller when considering those genes without other gene starts or ends upstream only.

Figure S4

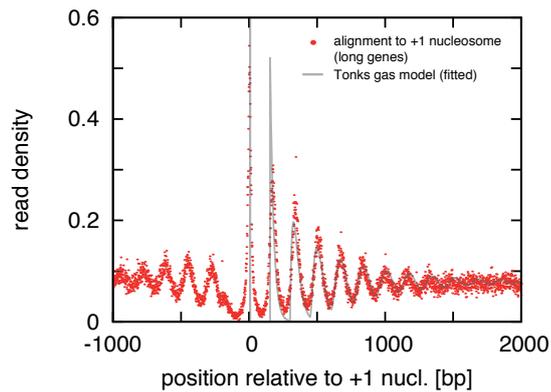


FIG. S4: Best fit of Tonks gas model (gray line) to +1 nucleosome alignment of read density including genes larger than 2000 bp only (red dots). Visual inspection yields good agreement between model and data, comparable to the analogous fit to the data including all genes (Fig. 2A, see also Fig. S3C). For estimated parameters see Table S1.

Figure S5

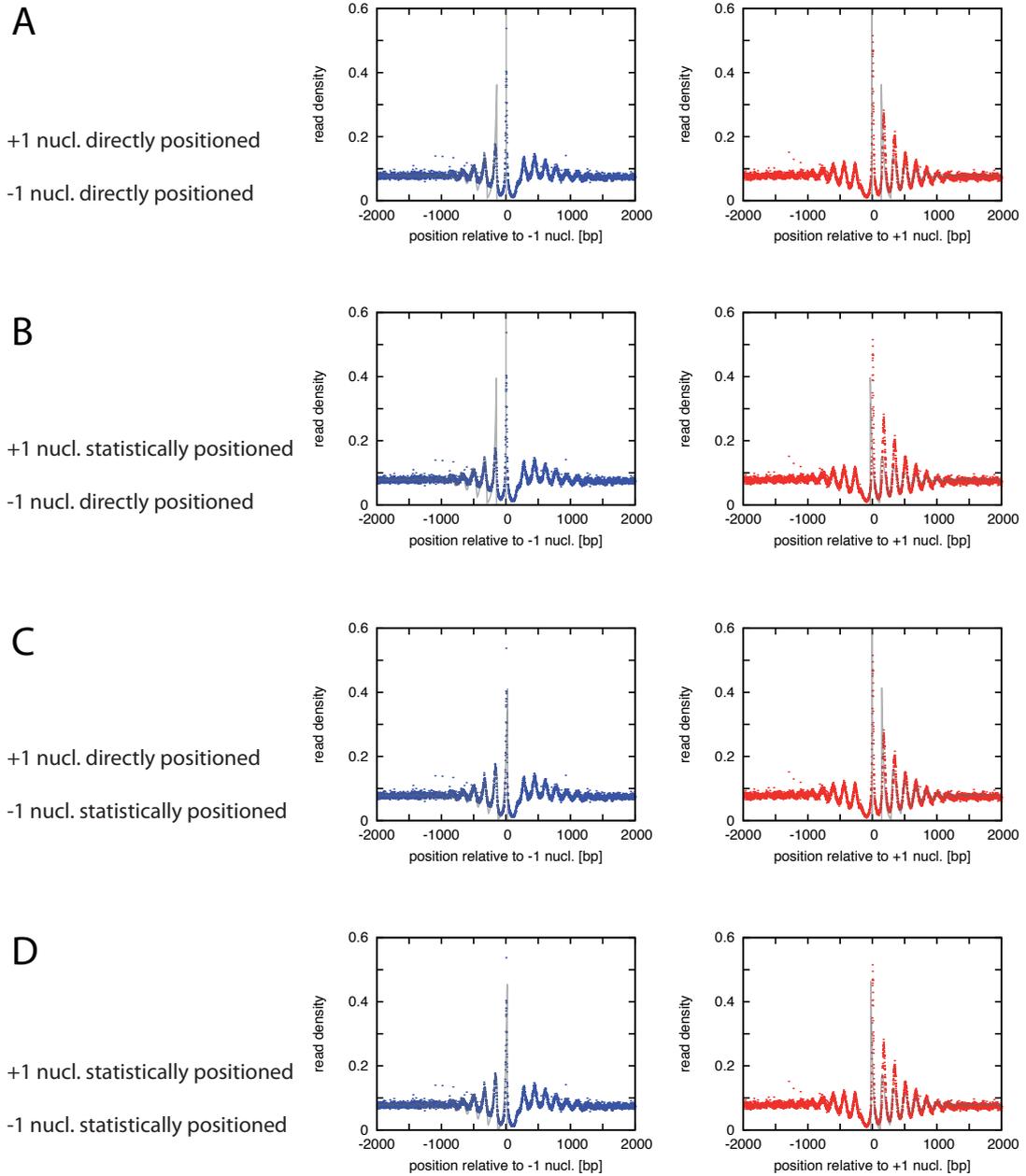


FIG. S5: Best simultaneous fits (gray lines) to -1 and +1 nucleosome alignments of read density (blue and red dots, respectively) given the four possible boundary conditions (both the +1 and -1 nucleosome may be directly or indirectly positioned) with nucleosome density and normalization being equal for both alignments (see 'Materials and Methods' for details). Regarding the mean squared deviation per data point, scenario C describes the data best, i.e., the scenario where the +1 nucleosome is directly positioned while the -1 nucleosome is indirectly positioned (Table S3).

Figure S6

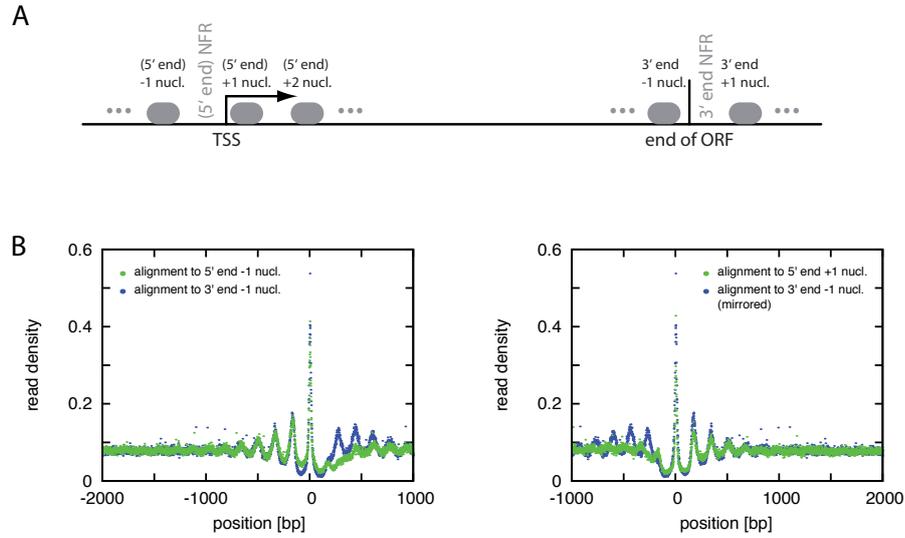


FIG. S6: Nucleosome organization around the 3' end of genes. (A) Sketch of a typical nucleosome organization around both the 5' and 3' ends of genes. Throughout this study, the focus is primarily on the 5' NFR with its flanking -1 and +1 nucleosomes. The nucleosomes flanking the 3' NFR are here referred to as the 3' end -1 nucleosome and the 3' end +1 nucleosome. We determined the positions of 3' end ± 1 nucleosomes in analogy to the 5' end ± 1 nucleosomes: The 3' end -1 nucleosome is defined as the nucleosome at or first nucleosome upstream of the ORF end while the 3' end +1 nucleosome is the first nucleosome downstream. (B) Alignment of read density to the 3' end -1 nucleosome (left) and 3' end +1 nucleosome (right), respectively (green data points). For comparison, the alignment to the 5' end -1 nucleosome is also shown (blue data points, from Fig. 4, mirrored on the right). Overall, a good agreement is visible between the alignments to the nucleosomes flanking the 3' NFR on both sides and the alignment to the 5' end -1 nucleosome. This indicates that at the 3' end the nucleosomes are only statistically positioned against a repulsive barrier, which we found to be the most likely scenario for the pattern upstream of the 5' NFR. (Note the small bump in the read density within the nucleosome depleted region, just downstream of the 3' end -1 nucleosome and upstream of the 3' end +1 nucleosome; it indicates that the identification of 3' NFRs is not perfect or a certain fraction of genes does not display a 3' NFR.)

Table S1

Fit scenario	$1/\bar{\rho}$ [bp]	λ	Δr [bp]	δ
+1 nucleosome, all genes, Fig. 2A	177	13.7	1	1.5e-4
+1 nucleosome, genes larger 2000 bp, Fig. S4	172	13.5	7	2.8e-4
+1 nucleosome, all genes, fit of convoluted funct., Fig. 2B	175	13.6	(0)	1.1e-4
-1 nucleosome, all genes, Fig. 4C	(177)	(13.7)	-169	9.8e-5

TABLE S1: Parameter estimates from independent fits of Tonks gas model to ± 1 nucleosome alignments of read density based on Equation (2) (density $\bar{\rho}$, normalization λ , offset Δr , squared deviation per data point δ). Numbers in parentheses indicate values that were set fixed rather than estimated from the fit. See ‘Materials and Methods’ for details.

Table S2

	$1/\bar{\rho}$ [bp]	λ	Δr [bp]	δ
Partitioning A, Subset 1	176.8	13.60	0.6	2.29e-4
Partitioning A, Subset 2	177.5	13.66	-0.4	2.15e-4
Partitioning A, Subset 3	176.9	13.65	1.3	2.46e-4
Partitioning A, Subset 4	177.5	13.74	0.6	2.45e-4
Partitioning B, Subset 1	176.6	13.79	1.6	2.30e-4
Partitioning B, Subset 2	176.6	13.65	0.6	2.54e-4
Partitioning B, Subset 3	177.9	13.53	-0.4	2.14e-4
Partitioning B, Subset 4	177.6	13.67	0.5	2.42e-4
Partitioning C, Subset 1	177.2	13.74	-0.3	2.36e-4
Partitioning C, Subset 2	176.6	13.60	0.7	2.31e-4
Partitioning C, Subset 3	177.9	13.74	0.6	2.48e-4
Partitioning C, Subset 4	177.3	13.59	0.6	2.24e-4
Partitioning D, Subset 1	176.9	13.65	1.4	2.26e-4
Partitioning D, Subset 2	177.3	13.76	0.4	2.39e-4
Partitioning D, Subset 3	177.2	13.61	-0.4	2.28e-4
Partitioning D, Subset 4	177.0	13.60	1.5	2.36e-4

TABLE S2: Parameter estimates (density $\bar{\rho}$, normalization λ , offset Δr , squared deviation per data point δ) from fits of Tonks gas model to +1 nucleosome alignments of read density using subsets of genes only. Four times (partitioning A-D), the set of 4792 genes was divided into four equal-sized subsets (subset 1-4) before fitting. Estimated parameters are very similar; see ‘Materials and Methods’ for details.

Table S3

Fit scenario (see Fig. S5)	$1/\bar{\rho}$ [bp]	λ	Δr_{-1} [bp]	Δr_{+1} [bp]	δ
scenario A	187	14.6	-6	-8	1.6e-4
scenario B	182	14.2	-2	-179	2.1e-4
scenario C	180	14.1	-174	-3	1.1e-4
scenario D	177	13.8	-168	-170	1.3e-4

TABLE S3: Parameter estimates for simultaneous fits of Tonks gas model to ± 1 nucleosome alignments of read density. Both normalization λ and nucleosome density $\bar{\rho}$ are constrained to be equal for both alignments. Δr_{+1} and Δr_{-1} are independent parameters accounting for different boundary conditions. Regarding the mean squared deviation per data point δ , scenario C describes the data best, i.e., the scenario where the +1 nucleosome is directly positioned while the -1 nucleosome is indirectly positioned (Fig. S5). See ‘Materials and Methods’ for details.

A. Simulations and numerical methods

A.1. Simulations of polymer dynamics

Two parts of the work presented here (Sec. 2.6 and Sec. 3.6) are based on extensive computer simulations of polymer dynamics, the basics of which are outlined in the following.

Brownian dynamics is an efficient way to simulate the dynamics of polymers, for example DNA, on length and time scales which are not microscopic and where inertia does not play a role. The polymer is represented by a coarse-grained structure. The solvent is reduced to its effects on the polymer dynamics. On the one hand, these are friction and hydrodynamic interactions mediated by the solvent, which lead to overdamped motion. On the other hand, these are stochastic forces acting on the coarse-grained polymer.

We represented the semi-flexible polymer (DNA) by a series of beads as illustrated in Fig. A.1. The beads, located at positions \mathbf{r}_i , are inter-connected by stiff springs keeping the contour length of the system approximately constant.¹ The equilibrium properties and dynamics of the polymer are determined by the total energy U , which covers the ‘internal’ energy of the polymer (such as the springs, the bending energy, the excluded volume interactions, etc.) as well as external contributions (interactions with another polymer or molecule, confinement, etc.). Considering, for example, the case of the strongly confined, knotted polymer (Sec. 2.6), the potential energy U reads $U = U_s + U_b + U_{ex} + U_{ch}$, where the different terms stand for the springs, the bending energy, excluded volume interactions, and confinement by the channel, respectively. The terms describing the springs and the bending energy should be discussed in some more detail.

The beads are kept together by springs with length b in the relaxed state. The springs contribute to the potential energy with a harmonic term,

$$U_s = \frac{K_s}{2} \sum_i (|\mathbf{r}_{i+1} - \mathbf{r}_i| - b)^2, \quad (\text{A.1})$$

where K_s is the spring stiffness. Since the springs are only introduced to keep the beads together and the polymer length approximately constant, alternative expressions for U_s are also used often.

The energy term for bending, by contrast, is directly derived by discretizing the worm-like chain Hamiltonian (Eq. (1.1)),

$$U_b = k_B T \frac{l_p}{b} \sum_i (1 - \cos \theta_i), \quad (\text{A.2})$$

where l_p is the polymer’s persistence length. The angles θ_i between consecutive segments at bead i are directly computed from the positions \mathbf{r}_i as shown in Fig. A.1.

¹In principle, the springs may be replaced by rods, which makes the simulation procedure more complex since constraints have to be taken into account, see Ref. [89, Appendix B2] and references therein.

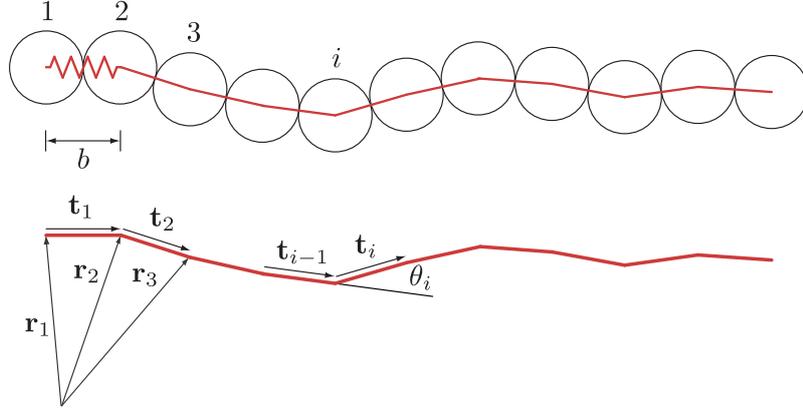


Figure A.1.: Modeling a polymer by a system of beads and springs. The beads at positions \mathbf{r}_i are interconnected by springs. The tangent vectors $\mathbf{t}_i = (\mathbf{r}_{i+1} - \mathbf{r}_i)/|\mathbf{r}_{i+1} - \mathbf{r}_i|$ determine the angles θ_i between adjacent segments by $\cos \theta_i = \mathbf{t}_{i-1} \cdot \mathbf{t}_i$.

As outlined above, in Brownian dynamics, the polymer's dynamics is determined by overdamped motion through the solvent and stochastic forces. Quantitatively, the equations of motion for the individual beads are given by a set of overdamped Langevin equations [26, Chapter 3]. If no hydrodynamic interactions between the beads are included, the equations of motion read:

$$\dot{\mathbf{r}}_i = -\mu_b \nabla_{\mathbf{r}_i} U + \sqrt{2D_b} \zeta_i(t). \quad (\text{A.3})$$

The first term describes the deterministic motion, i.e., drag through a high-viscous solvent where the velocity is given by the force multiplied by the bead's mobility μ_b . The second term describes random forces acting on the bead with $\langle \zeta_i(t) \rangle = 0$ and $\langle \zeta_i(t) \cdot \zeta_j(t') \rangle = 3\delta_{ij}\delta(t-t')$. The amplitude is determined by the bead's diffusion coefficient, related to the mobility via $D_b = k_B T \mu_b$ (Einstein relation).

In order to simulate the polymer dynamics, Eq. (A.3) has to be discretized in time. In the most simple form, one obtains (Euler algorithm [36, Chapter 10]):

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) - \mu_b \Delta t \nabla_{\mathbf{r}_i} U + \sqrt{2\mu_b k_B T \Delta t} \boldsymbol{\eta}_i, \quad (\text{A.4})$$

where the components of $\boldsymbol{\eta}_i(t)$ are independently drawn from a normal distribution with unit variance at every time step. Note that forces which only depend on the distance between two beads need only be computed once, which may save a great deal of computation time.

When implementing Eq. (A.4), the derivatives $\nabla_{\mathbf{r}_i} U$ have to be determined analytically. Note that hard core potentials as commonly found in Monte Carlo simulations cannot be used in Brownian dynamics simulations. Even reflecting boundary conditions like a hard wall or a confining channel have – in general – to be implemented using steep but soft potentials.

The time step Δt has to be chosen carefully. Excessively large values may lead to inaccurate results or even instability in the simulation. Unnecessarily small values waste CPU time. In general, accuracy of the simulation is determined by the steepest potential, typically U_s , the springs keeping the polymer together. The choice of the time step (and the simulation in general) is typically tested on the free polymer first, i.e., equilibrium observables like distribution of spring lengths or tangent-tangent correlation and dynamic observables like

center of mass diffusion are compared with the corresponding analytical values which may easily be computed for the discretized polymer. (An expression for the tangent-tangent correlation is given in Ref. [68].)

To visualize the polymer's motion, one may use a rendering software like POVray or, more conveniently, the software VMD which originally was developed for Molecular Dynamics simulations.²

A.2. More general Langevin dynamics

In Sec. 2.6 and Sec. 3.7, a more general set of coupled (overdamped) Langevin equations was used to describe the dynamics of the coarse-grained model for a knotted polymer and the two-segment lever, respectively. In its most general form, the stochastic differential equations used to describe the time evolution of the configuration $\{x_i\}$ of both systems read (Ito form, [36]):

$$\dot{x}_i = -\hat{M}_{ij}(\{x_k\}) \partial_j U(\{x_k\}) + \partial_j \hat{D}_{ij}(\{x_k\}) + \sqrt{2} \hat{B}_{ij}(\{x_k\}) \zeta_j(t). \quad (\text{A.5})$$

The first term describes the deterministic motion caused by the potential U . The mobility matrix \hat{M}_{ij} incorporates the friction and the direct coupling of the different degrees of freedom. The coupling arises from ‘geometrical constraints’ in Sec. 3.7 (see also [89, Appendix B.2]) and captures an ‘exchange of length’ in Sec. 2.6. The mobility matrix \hat{M}_{ij} is linked to the diffusion matrix \hat{D}_{ij} via the Einstein relation $k_B T \hat{M}_{ij} = \hat{D}_{ij}$. The last term in Eq. (A.5) depicts the stochastic motion. The components $\zeta_j(t)$ of the (white) noise vector are uncorrelated with $\langle \zeta_j(t) \rangle = 0$ and $\langle \zeta_j(t) \zeta_k(t') \rangle = \delta_{jk} \delta(t - t')$. The matrix \hat{B}_{ij} in front specifies the effect of the noise on the different degrees of freedom and is connected to the diffusion matrix via $\hat{D}_{ij} = \hat{B}_{il} \hat{B}_{jl}$. Note that \hat{B}_{ij} is not uniquely determined, different choices resulting in the same diffusion matrix \hat{D}_{ij} are physically equivalent. From what has said so far it is clear that knowledge of either \hat{B}_{ij} , \hat{D}_{ij} , or \hat{M}_{ij} suffices to specify the entire dynamics of the system.

The term not commented on so far, $\partial_j \hat{D}_{ij}(\{x_k\})$, becomes important if the diffusion matrix depends on the state of the system as is the case for the systems we investigated. In Sec. 2.6 we refer to this term as ‘noise induced drift’; though not caused by the potential U , it causes a real drift which in principle can be measured in the system. (Note: This ‘noise induced drift’ should not be confused with the term $-\hat{B}_{kj} \partial_k \hat{B}_{ij}$ arising when rewriting Eq. (A.5) in Stratonovich form [36, page 100] or other drift terms which vanish in the low temperature limit [89, Appendix B.2].)

Eq. (A.5) may be simulated using [36, Chapter 10]:

$$x_i(t + \Delta t) = \left[-\hat{M}_{ij}(\{x_k\}) \partial_j U(\{x_k\}) + \partial_j \hat{D}_{ij}(\{x_k\}) \right] \Delta t + \sqrt{2 \Delta t} \hat{B}_{ij}(\{x_k\}) \eta_j, \quad (\text{A.6})$$

where the η_j are, again, independently drawn from a normal distribution with unit variance at every time step (technically, they are (scaled) increments of a Wiener process).

²<http://www.ks.uiuc.edu/Research/vmd/>

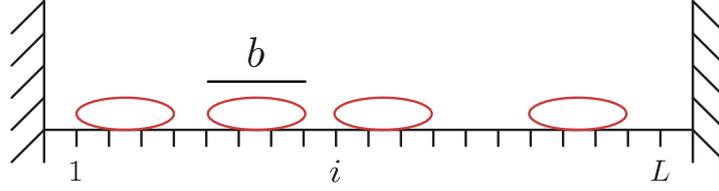


Figure A.2.: Illustration of a gas of hard rods of length b binding to a lattice of length L . (The walls are not formed by boundary particles as in Fig. 4.3 to simplify the boundary conditions for the recursion relations given in this section.)

A.3. Recursion relations for partition functions

Consider a gas of hard rods of length b binding to a lattice of length L as illustrated in Fig. A.2. Each particle is associated with a binding energy which depends on the lattice sites it covers. For clarity, let a particle covering sites i to $i + b - 1$ be considered ‘bound at site i ’ with binding energy ϵ_i .

The grand-canonical partition function (depending on the chemical potential μ) reads

$$\Xi = \sum_{\{\sigma\}} e^{-\sum_i \beta \sigma_i (\epsilon_i - \mu)} \quad \text{with} \quad \beta = \frac{1}{k_B T}, \quad (\text{A.7})$$

where $\sigma_i = 1$ if a particle is bound at site i and $\sigma_i = 0$ otherwise. For an example, see Fig. A.2, where $\sigma_1 = 1$ and $\sigma_2 = 0$. The sum is performed over all possible configurations of particles ($\{\sigma\}$) which obey excluded volume, i.e., in which two particles do not overlap. An attempt to compute Ξ by explicitly considering all possible configurations fails for $L \gg b$ since the number of terms becomes huge. Recursion relations, however, allow one to determine Ξ in a time efficient manner. One example is a linear recursion relation adapted from Ref. [41, Supplementary Information] which is derived in the following.

Consider Ξ_i , the grand-canonical partition function under the constraint that no particles are bound at sites $j > i$. One easily sees that

$$\Xi_i = \Xi_{i-1} + \Xi_{i-b} \cdot e^{-\beta(\epsilon_i - \mu)}. \quad (\text{A.8})$$

The first term covers all configurations where there is no particle bound at site i , the second covers the cases where a particle is bound at site i . The finite width is taken into account by the fact that the second term involves the factor Ξ_{i-b} which guarantees that no other particle is overlapping with site i . The recursion starts at site $i = 1$ since this is the leftmost binding site. The boundary condition may be implemented by introducing virtual sites for $i < 1$ with $\Xi_{i < 1} \equiv 1$. Finally, Ξ_{L-b+1} corresponds to the grand partition function Ξ of the full system, the recursion stops at this point.

The density of particles at site j , or equivalently the probability of finding a particle bound at site j , is given by

$$p(\sigma_j = 1) = 1 - \frac{\Xi|_{\sigma_j=0}}{\Xi}, \quad (\text{A.9})$$

where $\Xi|_{\sigma_j=0} = \tilde{\Xi}$ denotes the constrained grand partition function where no particle is bound at site j . $\tilde{\Xi}$ may be computed following the same scheme as for Ξ with the small modification

of explicitly excluding the case where a particle is bound at site j ,

$$\tilde{\Xi}_i = \begin{cases} \tilde{\Xi}_{i-1} + \tilde{\Xi}_{i-b} \cdot e^{-\beta(\epsilon_i - \mu)} & \text{for } i \neq j \\ \tilde{\Xi}_{i-1} & \text{for } i = j \end{cases} . \quad (\text{A.10})$$

Numerically, this recursion relation is fast; the computational cost to determine Ξ grows only linearly with system size L . However, to the best of my knowledge, the scheme cannot be applied in the case of periodic boundary conditions or more complex than hard core interactions. In those cases, the more general *transfer matrix* approach is appropriate. Within this scheme, the computational effort is higher than with the recursion formula described above. Instead of $\sim L$ summations one has to perform $\sim L$ multiplications of matrices of size $\sim b^2$, where b is the range of interaction. (If the binding energy landscape is simple, one may get along with much less multiplications or diagonalize the matrices first.) A nice introduction to the transfer matrix approach in protein binding problems is given by Teif [133].

Bibliography

- [1] C. C. Adams, “Das Knotenbuch”, (Spektrum Akademischer Verlag, Heidelberg, Berlin, Oxford, 1995), Page(s) [3](#), [14](#)
- [2] I. Albert, T. N. Mavrich, L. P. Tomsho, J. Qi, S. J. Zanton, et al., “Translational and rotational settings of H2A.Z nucleosomes across the *Saccharomyces cerevisiae* genome”, *Nature*, **446**, 572 (2007), [doi:10.1038/nature05632](https://doi.org/10.1038/nature05632), Page(s) [49](#), [51](#)
- [3] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, et al., “Molecular Biology of the Cell”, (Garland Science, New York, 2002), 4th edition, Page(s) [1](#), [6](#), [11](#), [14](#), [51](#)
- [4] R. J. Allen, D. Frenkel, and P. R. ten Wolde, “Simulating rare events in equilibrium or nonequilibrium stochastic systems”, *J. Chem. Phys.*, **124**, 024102 (2006), [doi:10.1063/1.2140273](https://doi.org/10.1063/1.2140273), Page(s) [34](#)
- [5] J. D. Anderson, P. T. Lowary, and J. Widom, “Effects of histone acetylation on the equilibrium accessibility of nucleosomal DNA target sites”, *J. Mol. Biol.*, **307**, 977 (2001), [doi:10.1006/jmbi.2001.4528](https://doi.org/10.1006/jmbi.2001.4528), Page(s) [29](#)
- [6] J. D. Anderson and J. Widom, “Sequence and position-dependence of the equilibrium accessibility of nucleosomal DNA target sites”, *J. Mol. Biol.*, **296**, 979 (2000), [doi:10.1006/jmbi.2000.3531](https://doi.org/10.1006/jmbi.2000.3531), Page(s) [28](#), [29](#)
- [7] A. Arnold and S. Jun, “Time scale of entropic segregation of flexible polymers in confinement: Implications for chromosome segregation in filamentous bacteria”, *Phys. Rev. E*, **76**, 031901 (2007), [doi:10.1103/PhysRevE.76.031901](https://doi.org/10.1103/PhysRevE.76.031901), Page(s) [20](#)
- [8] G. Arya and T. Schlick, “A tale of tails: How histone tails mediate chromatin compaction in different salt and linker histone environments”, *J. Phys. Chem. A*, **113**, 4045 (2009), [doi:10.1021/jp810375d](https://doi.org/10.1021/jp810375d), Page(s) [37](#)
- [9] R. Austin, “Nanofluidics: A fork in the nano-road”, *Nature Nanotechnol.*, **2**, 79 (2007), [doi:10.1038/nnano.2007.18](https://doi.org/10.1038/nnano.2007.18), Page(s) [13](#)
- [10] X. R. Bao, H. J. Lee, and S. R. Quake, “Behavior of complex knots in single DNA molecules”, *Phys. Rev. Lett.*, **91**, 265506 (2003), [doi:10.1103/PhysRevLett.91.265506](https://doi.org/10.1103/PhysRevLett.91.265506), Page(s) [19](#)
- [11] A. Belmonte, “The tangled web of self-tying knots”, *Proc. Natl. Acad. Sci. USA*, **104**, 17243 (2007), [doi:10.1073/pnas.0708150104](https://doi.org/10.1073/pnas.0708150104), Page(s) [14](#), [15](#)
- [12] J. M. Berg, J. L. Tymoczko, and L. Stryer, “Biochemistry”, (W. H. Freeman and Company, New York, 2007), 6th edition, Page(s) [3](#)

- [13] M. Bishop and M. A. Boonstra, “Exact partition functions for some one-dimensional models via the isobaric ensemble”, *Am. J. Phys.*, **51**, 564 (1983), doi:10.1119/1.13204, Page(s) 60
- [14] H. Boroudjerdi and R. R. Netz, “Strongly coupled polyelectrolyte-macroion complexes”, *J. Phys.: Cond. Mat.*, **17**, S1137 (2005), doi:10.1088/0953-8984/17/14/002, Page(s) 8
- [15] H. Brenner, “The slow motion of a sphere through a viscous fluid towards a plane surface”, *Chem. Eng. Sci.*, **16**, 242 (1961), doi:10.1016/0009-2509(61)80035-3, Page(s) 36
- [16] G. Chevereau, L. Palmeira, C. Thermes, A. Arneodo, and C. Vaillant, “Thermodynamics of intragenic nucleosome ordering”, *Phys. Rev. Lett.*, **103**, 188103 (2009), doi:10.1103/PhysRevLett.103.188103, Page(s) 58, 65, 66
- [17] F.-T. Chien and J. van Noort, “10 years of tension on chromatin: Results from single molecule force spectroscopy”, *Curr. Pharm. Biotechnol.*, **10**, 474 (2009), URL <http://www.bentham-direct.org/pages/content.php?CPB/2009/00000010/00000005/0003G.SGM>, Page(s) 11, 30
- [18] T. Chou, “An exact theory of histone-DNA adsorption and wrapping”, *Europhys. Lett.*, **62**, 753 (2003), doi:10.1209/epl/i2003-00437-y, Page(s) 66, 67
- [19] T. Chou, “Peeling and sliding in nucleosome repositioning”, *Phys. Rev. Lett.*, **99**, 058105 (2007), doi:10.1103/PhysRevLett.99.058105, Page(s) 30
- [20] H.-R. Chung and M. Vingron, “Sequence-dependent nucleosome positioning”, *J. Mol. Biol.*, **386**, 1411 (2009), doi:10.1016/j.jmb.2008.11.049, Page(s) 54, 55
- [21] C. R. Clapier and B. R. Cairns, “The biology of chromatin remodeling complexes”, *Annu. Rev. Biochem.*, **78**, 273 (2009), doi:10.1146/annurev.biochem.77.062706.153223, Page(s) 10, 30
- [22] H. Craighead, “Future lab-on-a-chip technologies for interrogating individual molecules”, *Nature*, **442**, 387 (2006), doi:10.1038/nature05061, Page(s) 13
- [23] 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 Å resolution”, *J. Mol. Biol.*, **319**, 1097 (2002), doi:10.1016/S0022-2836(02)00386-8, Page(s) 6, 8
- [24] P.-G. de Gennes, “Scaling Concepts in Polymer Physics”, (Cornell University Press, Ithaca and London, 1979), Page(s) 15
- [25] M. Depken and H. Schiessel, “Nucleosome shape dictates chromatin fiber structure”, *Biophys. J.*, **96**, 777 (2009), doi:10.1016/j.bpj.2008.09.055, Page(s) 37
- [26] M. Doi and S. F. Edwards, “The Theory of Polymer Dynamics”, (Oxford University Press, Oxford, 1986), Page(s) 35, 88
- [27] P. G. Dommersnes, Y. Kantor, and M. Kardar, “Knots in charged polymers”, *Phys. Rev. E*, **66**, 031802 (2002), doi:10.1103/PhysRevE.66.031802, Page(s) 18

- [28] B. Dorigo, T. Schalch, A. Kulangara, S. Duda, R. R. Schroeder, et al., “Nucleosome arrays reveal the two-start organization of the chromatin fiber”, *Science*, **306**, 1571 (2004), doi:10.1126/science.1103124, Page(s) 10, 11
- [29] R. S. Edayathumangalam, P. Weyermann, J. M. Gottesfeld, P. B. Dervan, and K. Luger, “Molecular recognition of the nucleosomal “supergroove””, *Proc. Natl. Acad. Sci. USA*, **101**, 6864 (2004), doi:10.1073/pnas.0401743101, Page(s) 27
- [30] M. Emanuel, N. H. Radja, A. Henriksson, and H. Schiessel, “The physics behind the larger scale organization of DNA in eukaryotes”, *Phys. Biol.*, **6**, 025008 (2009), doi:10.1088/1478-3975/6/2/025008, Page(s) 11
- [31] M. J. Fedor, N. F. Lue, and R. D. Kornberg, “Statistical positioning of nucleosomes by specific protein-binding to an upstream activating sequence in yeast”, *J. Mol. Biol.*, **204**, 109 (1988), doi:10.1016/0022-2836(88)90603-1, Page(s) 58
- [32] Y. Field, Y. Fondufe-Mittendorf, I. K. Moore, P. Mieczkowski, N. Kaplan, et al., “Gene expression divergence in yeast is coupled to evolution of DNA-encoded nucleosome organization”, *Nature Genet.*, **41**, 438 (2009), doi:10.1038/ng.324, Page(s) 49
- [33] Y. Field, N. Kaplan, Y. Fondufe-Mittendorf, I. K. Moore, E. Sharon, et al., “Distinct modes of regulation by chromatin encoded through nucleosome positioning signals”, *PLoS Comput. Biol.*, **4**, e1000216 (2008), doi:10.1371/journal.pcbi.1000216, Page(s) 54, 55, 67
- [34] A. Flaus and T. Owen-Hughes, “Mechanisms for nucleosome mobilization”, *Biopolymers*, **68**, 563 (2003), doi:10.1002/bip.10323, Page(s) 30
- [35] H. G. Garcia, P. Grayson, L. Han, M. Inamdar, J. Kondev, et al., “Biological consequences of tightly bent DNA: The other life of a macromolecular celebrity”, *Biopolymers*, **85**, 115 (2007), doi:10.1002/bip.20627, Page(s) 33
- [36] C. W. Gardiner, “Handbook of Stochastic Methods”, (Springer, Berlin, Heidelberg, New York, 2004), 3rd edition, Page(s) 35, 88, 89
- [37] S. A. Grigoryev, G. Arya, S. Correll, C. L. Woodcock, and T. Schlick, “Evidence for heteromorphic chromatin fibers from analysis of nucleosome interactions”, *Proc. Natl. Acad. Sci. USA*, **106**, 13317 (2009), doi:10.1073/pnas.0903280106, Page(s) 11, 37
- [38] A. Y. Grosberg and Y. Rabin, “Metastable tight knots in a wormlike polymer”, *Phys. Rev. Lett.*, **99**, 217801 (2007), doi:10.1103/PhysRevLett.99.217801, Page(s) 19
- [39] P. Hänggi, P. Talkner, and M. Borkovec, “Reaction-rate theory: fifty years after Kramers”, *Rev. Mod. Phys.*, **62**, 251 (1990), doi:10.1103/RevModPhys.62.251, Page(s) 35, 36, 38
- [40] P. D. Hartley and H. D. Madhani, “Mechanisms that specify promoter nucleosome location and identity”, *Cell*, **137**, 445 (2009), doi:10.1016/j.cell.2009.02.043, Page(s) 53

- [41] R. Hermesen, S. Tans, and P. R. ten Wolde, “Transcriptional regulation by competing transcription factor modules”, *PLoS Comput. Biol.*, **2**, e164 (2006), doi:10.1371/journal.pcbi.0020164, Page(s) 90
- [42] C. Hodges, L. Bintu, L. Lubkowska, M. Kashlev, and C. Bustamante, “Nucleosomal fluctuations govern the transcription dynamics of RNA polymerase II”, *Science*, **325**, 626 (2009), doi:10.1126/science.1172926, Page(s) 30
- [43] L. Huang and D. E. Makarov, “Langevin dynamics simulations of the diffusion of molecular knots in tensioned polymer chains”, *J. Phys. Chem. A*, **111**, 10338 (2007), doi:10.1021/jp071940+, Page(s) 19
- [44] I. P. Ioshikhes, I. Albert, S. J. Zanton, and B. F. Pugh, “Nucleosome positions predicted through comparative genomics”, *Nature Genet.*, **38**, 1210 (2006), doi:10.1038/ng1878, Page(s) 54, 55
- [45] C. Jiang and B. F. Pugh, “A compiled and systematic reference map of nucleosome positions across the *Saccharomyces cerevisiae* genome”, *Genome Biol.*, **10**, R109 (2009), doi:10.1186/gb-2009-10-10-r109, Page(s) 49
- [46] C. Jiang and B. F. Pugh, “Nucleosome positioning and gene regulation: advances through genomics”, *Nature Rev. Genet.*, **10**, 161 (2009), doi:10.1038/nrg2522, Page(s) 49, 51, 52
- [47] N. Kaplan, I. K. Moore, Y. Fondufe-Mittendorf, A. J. Gossett, D. Tillo, et al., “The DNA-encoded nucleosome organization of a eukaryotic genome”, *Nature*, **458**, 362 (2009), doi:10.1038/nature07667, Page(s) 54, 55, 56, 58, 59, 65, 67
- [48] N. Kepper, D. Foethke, R. Stehr, G. Wedemann, and K. Rippe, “Nucleosome geometry and internucleosomal interactions control the chromatin fiber conformation”, *Biophys. J.*, **95**, 3692 (2008), doi:10.1529/biophysj.107.121079, Page(s) 37
- [49] W. J. A. Koopmans, A. Brehm, C. Logie, T. Schmidt, and J. van Noort, “Single-pair FRET microscopy reveals mononucleosome dynamics”, *J. Fluoresc.*, **17**, 785 (2007), doi:10.1007/s10895-007-0218-9, Page(s) 31, 32, 37
- [50] W. J. A. Koopmans, R. Buning, T. Schmidt, and J. van Noort, “spFRET using alternating excitation and FCS reveals progressive DNA unwrapping in nucleosomes”, *Biophys. J.*, **97**, 195 (2009), doi:10.1016/j.bpj.2009.04.030, Page(s) 33
- [51] W. J. A. Koopmans, T. Schmidt, and J. van Noort, “Nucleosome immobilization strategies for single-pair FRET microscopy”, *ChemPhysChem*, **9**, 2002 (2008), doi:10.1002/cphc.200800370, Page(s) 31, 32
- [52] R. Kornberg, “The location of nucleosomes in chromatin: specific or statistical?”, *Nature*, **292**, 579 (1981), doi:10.1038/292579a0, Page(s) 57
- [53] R. D. Kornberg, “Chromatin structure: a repeating unit of histones and DNA”, *Science*, **184**, 868 (1974), doi:10.1126/science.184.4139.868, Page(s) 6

- [54] R. D. Kornberg and L. Stryer, “Statistical distributions of nucleosomes: nonrandom locations by a stochastic mechanism”, *Nucleic Acids Res.*, **16**, 6677 (1988), doi:10.1093/nar/16.14.6677, Page(s) 57, 62
- [55] S. Köster, D. Steinhauser, and T. Pfohl, “Brownian motion of actin filaments in confining microchannels”, *J. Phys.: Cond. Mat.*, **17**, S4091 (2005), doi:10.1088/0953-8984/17/49/006, Page(s) 16
- [56] H. A. Kramers, “Brownian motion in a field of force and the diffusion model of chemical reactions”, *Physica*, **7**, 284 (1940), doi:10.1016/S0031-8914(40)90098-2, Page(s) 35
- [57] O. Kratky and G. Porod, “Röntgenuntersuchung gelöster Fadenmoleküle”, *Recl. Trav. Chim. Pays-Bas*, **68**, 1106 (1949), Page(s) 4
- [58] K. Kroy and E. Frey, “Dynamic scattering from semiflexible polymers”, in W. Brown and K. Mortensen (editors), “Scattering in Polymeric and Colloidal Systems”, (Gordon and Breach, 2001), Page(s) 4
- [59] I. M. Kulić and H. Schiessel, “Opening and closing DNA: theories on the nucleosome”, in R. Dias and B. Lindman (editors), “DNA interactions with polymers and surfactants”, (John Wiley & Sons, 2008), Page(s) 30
- [60] A. B. Lantermann, T. Straub, A. Strålfors, G.-C. Yuan, K. Ekwall, et al., “*Schizosaccharomyces pombe* genome-wide nucleosome mapping reveals positioning mechanisms distinct from those of *Saccharomyces cerevisiae*”, *Nature Struct. Mol. Biol.*, **17**, 251 (2010), doi:10.1038/nsmb.1741, Page(s) 49
- [61] W. Lee, D. Tillo, N. Bray, R. H. Morse, R. W. Davis, et al., “A high-resolution atlas of nucleosome occupancy in yeast”, *Nature Genet.*, **39**, 1235 (2007), doi:10.1038/ng2117, Page(s) 49, 53, 58, 65
- [62] S. L. Levy, J. T. Mannion, J. Cheng, C. H. Reccius, and H. G. Craighead, “Entropic unfolding of DNA molecules in nanofluidic channels”, *Nano Lett.*, **8**, 3839 (2008), doi:10.1021/nl802256s, Page(s) 20
- [63] G. Li, M. Levitus, C. Bustamante, and J. Widom, “Rapid spontaneous accessibility of nucleosomal DNA”, *Nature Struct. Mol. Biol.*, **12**, 46 (2004), doi:10.1038/nsmb869, Page(s) 31, 32, 33
- [64] G. Li and J. Widom, “Nucleosomes facilitate their own invasion”, *Nature Struct. Mol. Biol.*, **11**, 763 (2004), doi:10.1038/nsmb801, Page(s) 31
- [65] W. Li, S.-X. Dou, P. Xie, and P.-Y. Wang, “Brownian dynamics simulation of the effect of histone modification on nucleosome structure”, *Phys. Rev. E*, **75**, 051915 (2007), doi:10.1103/PhysRevE.75.051915, Page(s) 37
- [66] E. Lieberman-Aiden, N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, et al., “Comprehensive mapping of long-range interactions reveals folding principles of the human genome”, *Science*, **326**, 289 (2009), doi:10.1126/science.1181369, Page(s) 6

- [67] C. L. Liu, T. Kaplan, M. Kim, S. Buratowski, S. L. Schreiber, et al., “Single-nucleosome mapping of histone modifications in *S. cerevisiae*”, PLoS Biol., **3**, e328 (2005), doi:10.1371/journal.pbio.0030328, Page(s) 49
- [68] L. Livadaru, R. R. Netz, and H. J. Kreuzer, “Stretching response of discrete semiflexible polymers”, Macromolecules, **36**, 3732 (2003), doi:10.1021/ma020751g, Page(s) 89
- [69] P. T. Lowary and J. Widom, “Nucleosome packaging and nucleosome positioning of genomic DNA”, Proc. Natl. Acad. Sci. USA, **94**, 1183 (1997), URL <http://www.pnas.org/content/94/4/1183.abstract>, Page(s) 9, 54
- [70] P. T. Lowary and J. Widom, “New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning”, J. Mol. Biol., **276**, 19 (1998), doi:10.1006/jmbi.1997.1494, Page(s) 9
- [71] S. Lubliner and E. Segal, “Modeling interactions between adjacent nucleosomes improves genome-wide predictions of nucleosome occupancy”, Bioinformatics, **25**, i348 (2009), doi:10.1093/bioinformatics/btp216, Page(s) 67
- [72] 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), doi:10.1038/38444, Page(s) 6, 7, 8, 33, 36
- [73] J. F. Marko and E. D. Siggia, “Stretching DNA”, Macromolecules, **28**, 8759 (1995), doi:10.1021/ma00130a008, Page(s) 5
- [74] N. L. Marky and G. S. Manning, “A theory of DNA dissociation from the nucleosome”, J. Mol. Biol., **254**, 50 (1995), doi:10.1006/jmbi.1995.0598, Page(s) 33
- [75] R. Matthews, A. A. Louis, and J. M. Yeomans, “Effect of topology on dynamics of knots in polymers under tension”, Europhys. Lett., **89**, 20001 (2010), doi:10.1209/0295-5075/89/20001, Page(s) 19
- [76] D. C. Mattis, “The Many-Body Problem”, (World Scientific, Singapore, New Jersey, London, Hong Kong, 1993), Page(s) 59, 102, 103
- [77] T. N. Mavrich, I. P. Ioshikhes, B. J. Venters, C. Jiang, L. P. Tomsho, et al., “A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome”, Genome Res., **18**, 1073 (2008), doi:10.1101/gr.078261.108, Page(s) x, xii, 2, 49, 58, 63, 64, 65
- [78] T. N. Mavrich, C. Jiang, I. P. Ioshikhes, X. Li, B. J. Venters, et al., “Nucleosome organization in the *Drosophila* genome”, Nature, **453**, 358 (2008), doi:10.1038/nature06929, Page(s) 49
- [79] B. Mergell, R. Everaers, and H. Schiessel, “Nucleosome interactions in chromatin: fiber stiffening and hairpin formation”, Phys. Rev. E, **70**, 011915 (2004), doi:10.1103/PhysRevE.70.011915, Page(s) 37

- [80] R. Metzler, T. Ambjörnsson, A. Hanke, Y. Zhang, and S. Levene, “Single DNA conformations and biological function”, *J. Comput. Theor. Nanos.*, **4**, 1 (2007), available under [arXiv:physics/0609139](https://arxiv.org/abs/physics/0609139), Page(s) 14
- [81] R. Metzler, W. Reisner, R. Riehn, R. Austin, J. O. Tegenfeldt, et al., “Diffusion mechanisms of localised knots along a polymer”, *Europhys. Lett.*, **76**, 696 (2006), doi:10.1209/epl/i2006-10312-5, Page(s) 19
- [82] P. Milani, G. Chevereau, C. Vaillant, B. Audit, Z. Haftek-Terreau, et al., “Nucleosome positioning by genomic excluding-energy barriers”, *Proc. Natl. Acad. Sci. USA*, **106**, 22257 (2009), doi:10.1073/pnas.0909511106, Page(s) 58, 65, 66
- [83] J. A. Miller and J. Widom, “Collaborative competition mechanism for gene activation in vivo”, *Mol. Cell. Biol.*, **23**, 1623 (2003), doi:10.1128/MCB.23.5.1623-1632.2003, Page(s) 29
- [84] W. Möbius, E. Frey, and U. Gerland, “Spontaneous unknotting of a polymer confined in a nanochannel”, *Nano Lett.*, **8**, 4518 (2008), doi:10.1021/nl802559q, Page(s) 2, 13, 16
- [85] W. Möbius and U. Gerland, “Quantitative test of the barrier nucleosome model for statistical positioning of nucleosomes up- and downstream of transcription start sites”, *PLoS Comput. Biol.*, **6**, e1000891 (2010), doi:10.1371/journal.pcbi.1000891, Page(s) 50, 63
- [86] W. Möbius, R. A. Neher, and U. Gerland, “Kinetic accessibility of buried DNA sites in nucleosomes”, *Phys. Rev. Lett.*, **97**, 208102 (2006), doi:10.1103/PhysRevLett.97.208102, Page(s) 2, 27, 32, 33
- [87] A. V. Morozov, K. Fortney, D. A. Gaykalova, V. M. Studitsky, J. Widom, et al., “Using DNA mechanics to predict in vitro nucleosome positions and formation energies”, *Nucleic Acids Res.*, **37**, 4707 (2009), doi:10.1093/nar/gkp475, Page(s) 55
- [88] P. M. Morse, “Diatomic molecules according to the wave mechanics. II. Vibrational levels”, *Phys. Rev.*, **34**, 57 (1929), doi:10.1103/PhysRev.34.57, Page(s) 33
- [89] R. Neher, “Dynamic aspects of DNA”, Ph.D. thesis, LMU München (2007), URL http://edoc.ub.uni-muenchen.de/6940/1/Neher_Richard.pdf, Page(s) 38, 87, 89
- [90] R. A. Neher, W. Möbius, E. Frey, and U. Gerland, “Optimal flexibility for conformational transitions in macromolecules”, *Phys. Rev. Lett.*, **99**, 178101 (2007), doi:10.1103/PhysRevLett.99.178101, Page(s) 2, 38
- [91] R. R. Netz and D. Andelman, “Neutral and charged polymers at interfaces”, *Phys. Reports*, **380**, 1 (2003), doi:10.1016/S0370-1573(03)00118-2, Page(s) 5
- [92] T. Odijk, “The statistics and dynamics of confined or entangled stiff polymers”, *Macromolecules*, **16**, 1340 (1983), doi:10.1021/ma00242a015, Page(s) 15
- [93] T. Odijk, “DNA confined in nanochannels: Hairpin tightening by entropic depletion”, *J. Chem. Phys.*, **125**, 204904 (2006), doi:10.1063/1.2400227, Page(s) 16

- [94] T. Odijk, “Scaling theory of DNA confined in nanochannels and nanoslits”, *Phys. Rev. E*, **77**, 060901 (2008), doi:10.1103/PhysRevE.77.060901, Page(s) 16
- [95] E. Orlandini, A. L. Stella, and C. Vanderzande, “The size of knots in polymers”, *Phys. Biol.*, **6**, 025012 (2009), doi:10.1088/1478-3975/6/2/025012, Page(s) 14, 19
- [96] J. K. Percus, “Equilibrium state of a classical fluid of hard rods in an external-field”, *J. Stat. Phys.*, **15**, 505 (1976), doi:10.1007/BF01020803, Page(s) 68
- [97] M. G. Poirier, M. Bussiek, J. Langowski, and J. Widom, “Spontaneous access to DNA target sites in folded chromatin fibers”, *J. Mol. Biol.*, **379**, 772 (2008), doi:10.1016/j.jmb.2008.04.025, Page(s) 29, 37
- [98] M. G. Poirier, E. Oh, H. S. Tims, and J. Widom, “Dynamics and function of compact nucleosome arrays”, *Nature Struct. Mol. Biol.*, **16**, 938 (2009), doi:10.1038/nsmb.1650, Page(s) 37
- [99] K. J. Polach and J. Widom, “Mechanism of protein access to specific DNA sequences in chromatin: a dynamic equilibrium model for gene regulation”, *J. Mol. Biol.*, **254**, 130 (1995), doi:10.1006/jmbi.1995.0606, Page(s) 27, 28
- [100] K. J. Polach and J. Widom, “A model for the cooperative binding of eukaryotic regulatory proteins to nucleosomal target sites”, *J. Mol. Biol.*, **258**, 800 (1996), doi:10.1006/jmbi.1996.0288, Page(s) 29
- [101] R. U. Protacio, K. J. Polach, and J. Widom, “Coupled-enzymatic assays for the rate and mechanism of DNA site exposure in a nucleosome”, *J. Mol. Biol.*, **274**, 708 (1997), doi:10.1006/jmbi.1997.1440, Page(s) 31
- [102] M. Radman-Livaja and O. J. Rando, “Nucleosome positioning: How is it established, and why does it matter?”, *Dev. Biol.*, **339**, 258 (2010), doi:10.1016/j.ydbio.2009.06.012, Page(s) 49, 53, 57, 66
- [103] O. J. Rando, “Global patterns of histone modifications”, *Curr. Opin. Genet. Dev.*, **17**, 94 (2007), doi:10.1016/j.gde.2007.02.006, Page(s) 9
- [104] O. J. Rando and K. Ahmad, “Rules and regulation in the primary structure of chromatin”, *Curr. Opin. Cell. Biol.*, **19**, 250 (2007), doi:10.1016/j.ceb.2007.04.006, Page(s) 53
- [105] O. J. Rando and H. Y. Chang, “Genome-wide views of chromatin structure”, *Annu. Rev. Biochem.*, **78**, 245 (2009), doi:10.1146/annurev.biochem.78.071107.134639, Page(s) 9, 49, 50, 52, 53
- [106] D. M. Raymer and D. E. Smith, “Spontaneous knotting of an agitated string”, *Proc. Natl. Acad. Sci. USA*, **104**, 16432 (2007), doi:10.1073/pnas.0611320104, Page(s) 15
- [107] W. Reisner, K. J. Morton, R. Riehn, Y. M. Wang, Z. Yu, et al., “Statics and dynamics of single DNA molecules confined in nanochannels”, *Phys. Rev. Lett.*, **94**, 196101 (2005), doi:10.1103/PhysRevLett.94.196101, Page(s) 16

- [108] T. J. Richmond and C. A. Davey, “The structure of DNA in the nucleosome core”, *Nature*, **423**, 145 (2003), doi:10.1038/nature01595, Page(s) 6, 8
- [109] R. Riehn, M. Lu, Y.-M. Wang, S. F. Lim, E. C. Cox, et al., “Restriction mapping in nanofluidic devices”, *Proc. Natl. Acad. Sci. USA*, **102**, 10012 (2005), doi:10.1073/pnas.0503809102, Page(s) 13
- [110] P. J. J. Robinson, L. Fairall, V. A. T. Huynh, and D. Rhodes, “EM measurements define the dimensions of the ”30-nm” chromatin fiber: Evidence for a compact, interdigitated structure”, *Proc. Natl. Acad. Sci. USA*, **103**, 6506 (2006), doi:10.1073/pnas.0601212103, Page(s) 11
- [111] A. Robledo, “Liquid-solid transition of the hard-sphere system from uniformity of the chemical potential”, *J. Chem. Phys.*, **72**, 1701 (1980), doi:10.1063/1.439281, Page(s) 68
- [112] A. Robledo and J. S. Rowlinson, “The distribution of hard rods on a line of finite length”, *Mol. Phys.*, **58**, 711 (1986), doi:10.1080/00268978600101521, Page(s) 63
- [113] A. Robledo and C. Varea, “On the relationship between the density functional formalism and the potential distribution theory for nonuniform fluids”, *J. Stat. Phys.*, **26**, 513 (1981), doi:10.1007/BF01011432, Page(s) 68
- [114] M. Rubinstein and R. Colby, “Polymer Physics”, (Oxford University Press, Oxford, New York, 2003), Page(s) 5
- [115] V. V. Rybenkov, N. R. Cozzarelli, and A. V. Vologodskii, “Probability of DNA knotting and the effective diameter of the DNA double helix”, *Proc. Natl. Acad. Sci. USA*, **90**, 5307 (1993), URL <http://www.pnas.org/content/90/11/5307.abstract>, Page(s) 5
- [116] N. Saitô, K. Takahashi, and Y. Yunoki, “The statistical mechanical theory of stiff chains”, *J. Phys. Soc. Jpn.*, **22**, 219 (1967), doi:10.1143/JPSJ.22.219, Page(s) 4
- [117] Z. W. Salsburg, R. W. Zwanzig, and J. G. Kirkwood, “Molecular distribution functions in a one-dimensional fluid”, *J. Chem. Phys.*, **21**, 1098 (1953), doi:10.1063/1.1699116, Page(s) 62
- [118] K. Sarma and D. Reinberg, “Histone variants meet their match”, *Nature Rev. Mol. Cell Biol.*, **6**, 139 (2005), doi:10.1038/nrm1567, Page(s) 9
- [119] S. Sasaki, C. C. Mello, A. Shimada, Y. Nakatani, S.-I. Hashimoto, et al., “Chromatin-associated periodicity in genetic variation downstream of transcriptional start sites”, *Science*, **323**, 401 (2009), doi:10.1126/science.1163183, Page(s) 49, 66
- [120] S. C. Satchwell, H. R. Drew, and A. A. Travers, “Sequence periodicities in chicken nucleosome core DNA”, *J. Mol. Biol.*, **191**, 659 (1986), doi:10.1016/0022-2836(86)90452-3, Page(s) 8, 9
- [121] T. Schalch, S. Duda, D. F. Sargent, and T. J. Richmond, “X-ray structure of a tetranucleosome and its implications for the chromatin fibre”, *Nature*, **436**, 138 (2005), doi:10.1038/nature03686, Page(s) 11

- [122] H. Schiessel, “The physics of chromatin”, *J. Phys.: Cond. Mat.*, **15**, R699 (2003), doi:10.1088/0953-8984/15/19/203, Page(s) 7, 8, 29, 30
- [123] D. E. Schones, K. Cui, S. Cuddapah, T.-Y. Roh, A. Barski, et al., “Dynamic regulation of nucleosome positioning in the human genome”, *Cell*, **132**, 887 (2008), doi:10.1016/j.cell.2008.02.022, Page(s) 49
- [124] D. J. Schwab, R. F. Bruinsma, J. Rudnick, and J. Widom, “Nucleosome switches”, *Phys. Rev. Lett.*, **100**, 228105 (2008), doi:10.1103/PhysRevLett.100.228105, Page(s) 56
- [125] E. Segal, Y. Fondufe-Mittendorf, L. Chen, A. Thåström, Y. Field, et al., “A genomic code for nucleosome positioning”, *Nature*, **442**, 772 (2006), doi:10.1038/nature04979, Page(s) 54, 55, 56
- [126] E. Segal and J. Widom, “Poly(dA:dT) tracts: major determinants of nucleosome organization”, *Curr. Opin. Struc. Biol.*, **19**, 65 (2009), doi:10.1016/j.sbi.2009.01.004, Page(s) 9
- [127] E. Segal and J. Widom, “What controls nucleosome positions?”, *Trends Genet.*, **25**, 335 (2009), doi:10.1016/j.tig.2009.06.002, Page(s) 53, 57, 64
- [128] E. A. Sekinger, Z. Moqtaderi, and K. Struhl, “Intrinsic histone-DNA interactions and low nucleosome density are important for preferential accessibility of promoter regions in yeast”, *Mol. Cell*, **18**, 735 (2005), doi:10.1016/j.molcel.2005.05.003, Page(s) 53, 56
- [129] S. Shivaswamy, A. Bhinge, Y. Zhao, S. Jones, M. Hirst, et al., “Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation”, *PLoS Biol.*, **6**, e65 (2008), doi:10.1371/journal.pbio.0060065, Page(s) 49, 53
- [130] S. B. Smith, L. Finzi, and C. Bustamante, “Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads”, *Science*, **258**, 1122 (1992), doi:10.1126/science.1439819, Page(s) 5
- [131] G. D. Stormo and D. S. Fields, “Specificity, free energy and information content in protein-DNA interactions”, *Trends Biochem. Sci.*, **23**, 109 (1998), doi:10.1016/S0968-0004(98)01187-6, Page(s) 54
- [132] H. Takahashi, “A simple method for treating the statistical mechanics of one-dimensional substances”, *Proc. Phys.-Math. Soc. Jap.*, **24**, 60 (1942), reprinted in [76], Page(s) 60
- [133] V. B. Teif, “General transfer matrix formalism to calculate DNA-protein-drug binding in gene regulation: application to O_R operator of phage λ ”, *Nucleic Acids Res.*, **35**, e80 (2007), doi:10.1093/nar/gkm268, Page(s) 91
- [134] A. Thåström, P. T. Lowary, H. R. Widlund, H. Cao, M. Kubista, et al., “Sequence motifs and free energies of selected natural and non-natural nucleosome positioning DNA sequences”, *J. Mol. Biol.*, **288**, 213 (1999), doi:10.1006/jmbi.1999.2686, Page(s) 8, 54

- [135] D. Tolkunov and A. V. Morozov, “Nucleosome positioning and energetics: Recent advances in genomic and computational studies”, arXiv.org (2009), [arxiv:0912.3954](https://arxiv.org/abs/0912.3954), Page(s) [49](#), [50](#), [54](#), [55](#), [68](#)
- [136] M. Tomschik, K. van Holde, and J. Zlatanova, “Nucleosome dynamics as studied by single-pair fluorescence resonance energy transfer: a reevaluation”, *J. Fluoresc.*, **19**, 53 (2009), [doi:10.1007/s10895-008-0379-1](https://doi.org/10.1007/s10895-008-0379-1), Page(s) [31](#), [32](#)
- [137] 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. USA*, **102**, 3278 (2005), [doi:10.1073/pnas.0500189102](https://doi.org/10.1073/pnas.0500189102), Page(s) [31](#), [32](#), [33](#)
- [138] L. Tonks, “The complete equation of state of one, two and three-dimensional gases of hard elastic spheres”, *Phys. Rev.*, **50**, 955 (1936), [doi:10.1103/PhysRev.50.955](https://doi.org/10.1103/PhysRev.50.955), Page(s) [59](#), [62](#)
- [139] D. J. Tremethick, “Higher-order structures of chromatin: the elusive 30 nm fiber”, *Cell*, **128**, 651 (2007), [doi:10.1016/j.cell.2007.02.008](https://doi.org/10.1016/j.cell.2007.02.008), Page(s) [11](#)
- [140] C. Vaillant, B. Audit, and A. Arneodo, “Experiments confirm the influence of genome long-range correlations on nucleosome positioning”, *Phys. Rev. Lett.*, **99**, 218103 (2007), [doi:10.1103/PhysRevLett.99.218103](https://doi.org/10.1103/PhysRevLett.99.218103), Page(s) [55](#), [58](#), [59](#), [65](#)
- [141] C. Vaillant, L. Palmeira, G. Chevereau, B. Audit, Y. d’Aubenton-Carafa, et al., “A novel strategy of transcription regulation by intra-genic nucleosome ordering”, *Genome Res.*, **20**, 59 (2010), [doi:10.1101/gr.096644.109](https://doi.org/10.1101/gr.096644.109), Page(s) [58](#), [65](#), [66](#)
- [142] A. Valouev, J. Ichikawa, T. Tonthat, J. Stuart, S. Ranade, et al., “A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning”, *Genome Res.*, **18**, 1051 (2008), [doi:10.1101/gr.076463.108](https://doi.org/10.1101/gr.076463.108), Page(s) [49](#)
- [143] L. van Hove, “Sur l’intégrale de configuration pour les systèmes de particules à une dimension”, *Physica*, **16**, 137 (1950), [doi:10.1016/0031-8914\(50\)90072-3](https://doi.org/10.1016/0031-8914(50)90072-3), reprinted in [76], Page(s) [60](#), [61](#)
- [144] T. K. Vanderlick, L. E. Scriven, and H. T. Davis, “Solution of Percus’s equation for the density of hard rods in an external field”, *Phys. Rev. A*, **34**, 5130 (1986), [doi:10.1103/PhysRevA.34.5130](https://doi.org/10.1103/PhysRevA.34.5130), Page(s) [68](#)
- [145] C. Veiglhuber, “Overlapping Polymers in Strong Confinement”, Diploma thesis, LMU München (2009), Page(s) [19](#)
- [146] A. Vologodskii, “Brownian dynamics simulation of knot diffusion along a stretched DNA molecule”, *Biophys. J.*, **90**, 1594 (2006), [doi:10.1529/biophysj.105.074682](https://doi.org/10.1529/biophysj.105.074682), Page(s) [19](#)
- [147] F. Wagner, G. Lattanzi, and E. Frey, “Conformations of confined biopolymers”, *Phys. Rev. E*, **75**, 050902 (2007), [doi:10.1103/PhysRevE.75.050902](https://doi.org/10.1103/PhysRevE.75.050902), Page(s) [16](#)
- [148] J. D. Watson and F. H. C. Crick, “Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid”, *Nature*, **171**, 737 (1953), [doi:10.1038/171737a0](https://doi.org/10.1038/171737a0), Page(s) [3](#)

- [149] I. Whitehouse, O. J. Rando, J. Delrow, and T. Tsukiyama, “Chromatin remodeling at promoters suppresses antisense transcription”, *Nature*, **450**, 1031 (2007), doi:10.1038/nature06391, Page(s) 53
- [150] I. Whitehouse and T. Tsukiyama, “Antagonistic forces that position nucleosomes in vivo”, *Nature Struct. Mol. Biol.*, **13**, 633 (2006), doi:10.1038/nsmb1111, Page(s) 53
- [151] J. Widom, “Role of DNA sequence in nucleosome stability and dynamics”, *Q. Rev. Biophys.*, **34**, 269 (2001), doi:10.1017/S0033583501003699, Page(s) 8, 9, 28
- [152] C. H. Wiggins, D. Rivelino, A. Ott, and R. E. Goldstein, “Trapping and wiggling: Elastohydrodynamics of driven microfilaments”, *Biophys. J.*, **74**, 1043 (1998), doi:10.1016/S0006-3495(98)74029-9, Page(s) 4
- [153] P. A. Wiggins, T. van der Heijden, F. Moreno-Herrero, A. Spakowitz, R. Phillips, et al., “High flexibility of DNA on short length scales probed by atomic force microscopy”, *Nature Nanotechnol.*, **1**, 137 (2006), doi:10.1038/nnano.2006.63, Page(s) 5
- [154] C. Yuan, H. Chen, X. W. Lou, and L. A. Archer, “DNA bending stiffness on small length scales”, *Phys. Rev. Lett.*, **100**, 018102 (2008), doi:10.1103/PhysRevLett.100.018102, Page(s) 5
- [155] G.-C. Yuan and J. S. Liu, “Genomic sequence is highly predictive of local nucleosome depletion”, *PLoS Comput. Biol.*, **4**, e13 (2008), doi:10.1371/journal.pcbi.0040013, Page(s) 54, 55
- [156] G.-C. Yuan, Y.-J. Liu, M. F. Dion, M. D. Slack, L. F. Wu, et al., “Genome-scale identification of nucleosome positions in *S. cerevisiae*”, *Science*, **309**, 626 (2005), doi:10.1126/science.1112178, Page(s) 49, 51
- [157] K. A. Zawadzki, A. V. Morozov, and J. R. Broach, “Chromatin-dependent transcription factor accessibility rather than nucleosome remodeling predominates during global transcriptional restructuring in *Saccharomyces cerevisiae*”, *Mol. Biol. Cell*, **20**, 3503 (2009), doi:10.1091/mbc.E09-02-0111, Page(s) 53
- [158] F. Zernike and J. A. Prins, “Die Beugung von Röntgenstrahlen in Flüssigkeiten als Effekt der Molekülanordnung”, *Z. Physik*, **41**, 184 (1927), doi:10.1007/BF01391926, Page(s) 62
- [159] Y. Zhang, Z. Moqtaderi, B. P. Rattner, G. Euskirchen, M. Snyder, et al., “Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo”, *Nature Struct. Mol. Biol.*, **16**, 847 (2009), doi:10.1038/nsmb.1636, Page(s) 56, 58, 65

Danksagung

Die vorliegende Doktorarbeit wäre ohne vielfältige Mitwirkung und Unterstützung in dieser Form nicht entstanden. Zuallererst gilt mein Dank Prof. Ulrich Gerland für die vorzügliche Betreuung und kontinuierliche Unterstützung bei der Arbeit. Seine Ideen und Vorschläge haben ganz wesentlich zum Erfolg beigetragen. Er hatte immer Zeit für die Diskussion neuer Ideen, vorläufiger Ergebnisse, aber auch technischer Details. Die ergebnisoffenen Diskussionen und sein kontinuierliches Interesse am Fortgang der Arbeit habe ich sehr geschätzt. Angestoßen von ihm – und trotz anfänglicher Widerstände meinerseits – habe ich während meiner Doktorarbeit zwei für mich wichtige Sachen kennengelernt: die zum Staunen und manchmal zum Verzweifeln komplexe Welt der Biologie – und die Stadt Köln!

Weiterhin gebührt mein Dank all denen, mit denen ich während der vergangenen Jahre zusammenarbeiten durfte, insbesondere Dr. Richard Neher und Clemens Veiglhuber. Zusammen mit Richard ist die Arbeit zur Dynamik von nukleosomaler DNA entstanden. Von Richard habe ich viel lernen dürfen und gerne denke ich an unsere gemeinsamen Projekte zurück. Aus dem Projekt zu verknoteten Polymeren ist die Idee zur Diplomarbeit von Clemens entstanden. Seine Arbeit durfte ich inhaltlich begleiten und möchte ihm für die fruchtbare Zusammenarbeit danken. Ein guter Teil der vorliegenden Arbeit beruht auf Simulationen. Für einzelne Datenpunkte musste zum Teil sehr lange gerechnet werden. Dank an alle, die dies ermöglicht haben.

Trotz des Ausfluges nach Köln habe ich den größten Teil meiner Zeit als Doktorand in München am Lehrstuhl von Prof. Erwin Frey verbracht. Für seine Unterstützung möchte ich mich ebenso bedanken wie für die Schaffung einer angenehmen Arbeitsatmosphäre sowie die Gruppenseminare in Antholz. Gleichmaßen gilt mein Dank meinen Kollegen in München und Köln, stellvertretend meinen Bürokollegen Frederik Wagner und Georg Fritz. Ich habe die angenehme Büroatmosphäre und unsere fachlichen und nichtfachlichen Diskussionen sehr genossen.

Auch beim Schreiben der eigentlichen Doktorarbeit habe ich reichlich Unterstützung bekommen. Karen Alim und Benedikt Obermayer habe ich oft um Rat fragen dürfen. Ich wünsche Euch viel Erfolg in der letzten Phase Eurer eigenen Doktorarbeiten. Korrekturgelesen und sehr geduldig auf viele meiner Fragen geantwortet haben Barbara Treutlein, Dr. Ho-Ryun Chung, PD Dr. Thomas Franosch und Brendan Osberg. Für zahlreiche Tips zum Layout der Doktorarbeit danke ich Christoph Riedl. Auch Dir viel Erfolg beim Endspurt!

Bedanken möchte ich mich auch für die Unterstützung und die vielen guten Ratschläge meiner Familie, Freundin und guten Freunden. An dieser Stelle ist es mir auch ein großes Anliegen, meinem Großvater dafür zu danken, dass er vor langer Zeit mein Interesse an Naturwissenschaften und deren Verbindungen untereinander geweckt hat. Dies trägt ganz wesentlich dazu bei, dass mir die Beschäftigung mit biologischer Physik viel Freude bereitet.

Durch die finanzielle und ideelle Unterstützung der Studienstiftung des deutschen Volkes sowie des Internationalen Doktorandenkolleg NanoBioTechnologie habe ich während meiner Doktorarbeit ausschließlich an für mich spannenden Projekten arbeiten dürfen und konnte auf Konferenzen und Seminare reisen. Das ist nicht selbstverständlich, vielen Dank!

Curriculum Vitae

Personal Data

Wolfram Möbius
born January 10, 1981 in Eisenach

Address

Arnold Sommerfeld Center for Theoretical Physics
LMU München, Germany

phone +49 89 2180 4560
email wolfram.moebius@physik.lmu.de

Education

2007 – 2008	visiting PhD student at Universität zu Köln, Germany
2006 – 2010	PhD student at LMU München, advisor: Prof. Dr. Ulrich Gerland
September 2005	physics diploma: LMU München
2002 – 2003	visiting graduate student at Cornell University, Ithaca, NY
2001 – 2005	study of physics at LMU München, Germany
1999 – 2001	study of physics at TU Dresden, Germany
July 1999	Abitur (high school diploma): Kreuzschule Dresden