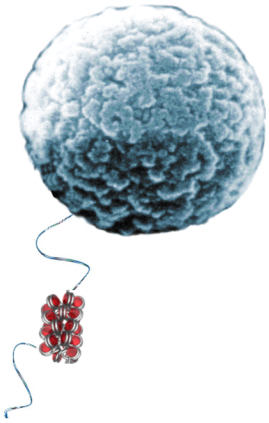
Quantification of Discrete States in Chromatin Fiber Folding



THESIS

submitted in partial fulfillment of the
requirements for the degree of

BACHELOR OF SCIENCE
in
PHYSICS

Author :	Rick Rodrigues de Mercado
Student ID :	1687115
Supervisor :	Prof.dr.ir. S.J.T. van Noort
2^{nd} corrector :	Prof.dr. H. Schiessel

Leiden, The Netherlands, June 30, 2018

Quantification of Discrete States in Chromatin Fiber Folding

Rick Rodrigues de Mercado

Huygens-Kamerlingh Onnes Laboratory, Leiden University
P.O. Box 9500, 2300 RA Leiden, The Netherlands

June 30, 2018

Abstract

The accessibility of DNA is regulated by means of dynamic folding and unfolding of chromatin fibers. Quantification of chromatin unfolding is a key in understanding the higher order structure of chromatin. Single molecule force spectroscopy is an ideal tool to study chromatin dynamics. The extension of reconstituted chromatin fibers under force was measured using Magnetic Tweezers. Here I show a model to fit discrete states in chromatin unfolding, and use this model to quantify the influence of linker histone H1 on chromatin fibers. Stepsize analysis revealed that linker histone H1 induces cooperative unstacking of nucleosomes at forces less than 10 pN. Chromatin fibers in presence of linker histone H1 must be stretched 4.5 ± 0.8 nm for two nucleosomes to unstack. This model can be used to quantify natively folded chromatin fibers.

Contents

Version of June 30, 2018– Created June 30, 2018 - 20:00

Chapter 1

Introduction

DNA is the fundamental building block of most living organisms. It contains the genetic information that is used during DNA transcription to make RNA, which is translated to functional proteins. In order to fit several meters of DNA in the nucleus of each cell, a DNA molecule is folded into a higher order structure to decrease the volume it occupies. In the first level of DNA condensation nucleosomes are formed. A nucleosome consists of 147 base pairs of DNA wrapped in approximately 1.7 turns around a histone octamer core [1–3]. Nucleosomes unwrap in two steps: the outer turn wrap contains 68 bp and the inner turn wrap consists of one full turn of 79 bp [4]. The histone core consists of two H2A-H2B dimers and two H3-H4 dimers. Nucleosomes are connected by linker DNA, forming a "beads on a string" array (compare with a pearl necklace). Linker DNA is flexible and can bend in order to stack nucleosomes on top of each other. Due to nucleosome-nucleosome interactions, tightly packed 30-nm chromatin fibers are formed [5, 6]. The structure of chromatin is complex, but of great importance. The structure of chromatin regulates the availability of the chromosomal DNA, which influences various processes such as replication, transcription, recombination and DNA repair. The accessibility of DNA is regulated by dynamic folding and unfolding of chromatin fibers [7, 8].

The higher order structure of chromatin has been studied intensively. A cryo-EM study revealed that chromatin fibers consist of tetranucleosomal units in a two-start superhelix [9]. A shortcoming of cryo-EM is that chromatin fibers are frozen at cryogenic temperatures, making it impossible to research chromatin dynamics. Intermediate states of the (de)formation of chromatin fibers is studied with single molecule force spectroscopy. By stretching a chromatin fiber, using Magnetic Tweezers,

and modelling the force versus extension, it has been shown that the structure of the 30-nm fiber depends on the nucleosome repeat length (NRL) [10]. Also using force spectroscopy, it was shown that the inner turn wrap of nucleosomes unwraps stepwise at forces above 15 pN with a stepsize of 79 bp [4].

1.1 The Linker Histone

The linker histone is a protein that binds to linker DNA and the histone octamer core [9]. Several variants, including H1 and H5 exist. Figure 1.1 shows a schematic picture of a nucleosome with a linker histone bound. The exact structure of linker histones binding to nucleosomes is still unclear. It has been shown that different linker histones interact differently; linker histone H5 binds symmetrically to the nucleosome, while H1 binds asymmetrically [11]. Linker histone H1 increases stacking interactions and moderates folding and unfolding kinetics of the outer turn wrap of nucleosomes [12]. Previous studies by Kim Vendel [13] and Artur Kaczmarczyk have shown that linker histone H1 induces discrete unstacking of nucleosomes in chromatin fibers.

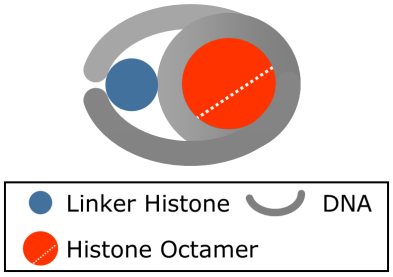


Figure 1.1: *A schematic picture of a nucleosome with a linker histone attached.*

The stepsizes and the free energy involved in stepwise unstacking of nucleosomes in the presence of linker histone have not been quantified yet. I used force spectroscopy to trigger mechanically unfold and fold chromatin fibers, and modeled the structure of chromatin fibers in presence of linker histone H1. The aim of this thesis is to introduce a model to quantify chromatin folding to get a better understanding of the higher order structure of chromatin. The model can be used to study natively folded chromatin of which the NRL is not known a priori. Using these data and the new data analysis method, I show that unstacking events feature an unanticipatedly broad distribution of stepsizes and the thermodynamical model suggest a maximum extension before ruptering of 4.5 nm. In this thesis I introduce a novel method to quantify discrete steps in chromatin unfolding and extract stepsizes from force-extension measurements.

Chapter 2

Theory

To model the mechanical disassembly of chromatin fibers under force, we therefore describe the mechanics of both bare DNA and compacted chromatin fibers. In our experiments, chromatin fibers are flanked by long DNA handles of bare DNA. The behaviour of bare DNA under force is described by the wormlike chain model [14]. For small forces is was shown that the chromatin fiber behaves like a Hookean spring [15]. The extension of the entire tether is the sum of the extension of the wormlike chain and the extension of the Hookean spring. With increasing force, the nucleosomes unstack from the chromatin fiber and the DNA unwraps from the nucleosomes. This leads to a more dominant contribution of the wormlike chain model. In the theory section, I will describe the freely jointed chain, which yields in a Hookean fashion for small forces, the wormlike chain and a model for chromatin fibers under force.

2.1 Freely Jointed Chain

Like chromatin fibers, a freely jointed chain (FJC) extends linear at low force [16]. Therefore, we used a FJC to model the extension of chromatin fibers. The advantage to use a FJC over a Hookean spring is that the total extension is limited, which prevents unrealistic extensions of the chromatin fiber at large forces. A FJC consists of N freely jointed rigid segments with each a Kuhn length b. At zero force, each segment is orientated randomly, independent of the surrounding segments. At small forces ($fb \ll k_BT$), the stiffness of the fiber is related to the Kuhn length [17]:

$$q = 3\frac{k_BT}{bL_{nuc}} \tag{2.1}$$

We define q as the stiffness per nucleosome, k_B is the Boltzmann constant, T is the absolute temperature and L_{nuc} is the length in nm per nucleosome. This relation, however, does not imply that chromatin fibers can be seen as randomly orientated rigid segments, but rather that the low force behaviour of chromatin fibers matches that of a FJC. The extension z_{FJC} of a FJC as a function of force f and length L is given by the Langevin function:

$$z_{FJC}(f, L) = L\left[\coth\left(\frac{fb}{k_BT}\right) - \frac{k_BT}{fb}\right] \tag{2.2}$$

Where $\coth(x)$ is the hyperbolic cotangent. At low force, $z_{FJC} \to \frac{f}{q}$; a FJC behaves like a Hookean spring which has been used to model chromatin fibers previously [15]. At high force, $z_{FJC} \to L = Nb$; the extension is limited to the contour length of the polymer. By integrating eq. 2.2 (Apendix A), an expression for the free energy G_{FJC} as a function of force f and length L is obtained:

$$G_{FJC}(f, L) = fz_{FJC}(f, L) - L\frac{k_BT}{d}\left[\ln\left(\sinh\left(\frac{fb}{k_BT}\right)\right) - \ln\left(\frac{fb}{k_BT}\right)\right] \tag{2.3}$$

2.2 Wormlike Chain

The extension of bare DNA under an external force is described by the wormlike chain (WLC) model [15, 18]. A WLC describes a continues flexible rod. To accommodate stretching deformations of the DNA, this WLC was modified to the extendable WLC: eWLC [17]. The extension z_{WLC} of an eWLC as a function of force f and length L is given by:

$$z_{WLC}(f, L) = L\left[1 - \frac{1}{2}\sqrt{\frac{k_BT}{Af}} + \frac{f}{S}\right] \tag{2.4}$$

Where A is the persistence length of the polymer, which is a measure of its rigidity. At length scales shorter than the persistence length, chains are almost straight. S is the stretching modulus. The stretching modulus is the force at which the WLC would stretch to twice its original length (if it would not break). The free energy is calculated by integrating the force for every extension (Apendix A). The free energy G_{WLC} as a function of force f and length L is given by:

$$G_{WLC}(f, L) = fz_{WLC}(f, L) - L\left[f - \sqrt{\frac{fk_BT}{A}} + \frac{f^2}{2S}\right] \tag{2.5}$$

2.3 Nucleosome unstacking

During an unstacking transition, the outer turn of the nucleosomes can unwrap [19]. Unstacking of a pair of nucleosomes effectively increases the contour length of the free DNA. Figure 2.1 displays three possible ways for two nucleosomes to unwrap and unstack. The number of base pairs in the outer turn is $147 - 79 = 68$. However, unwrapping can occur symmetrically from both side, or asymmetrically from one side. This has large implications for the measured stepsize. For single 601 sequence nucleosomes, it was shown that unwrapping occurs asymmetrically [19]. For an array of stacked nucleosomes, asymmetric unwrapping can also occur. The first possibility is that of both nucleosomes half an outer turn unwraps. For nucleosomes of known nucleosome repeat length (NRL), the total increase in linker DNA is then given by $\Delta L = 1 \times outer\ turn + 1 \times linker = NRL - 79$ bp. Second, the outer turn of one nucleosome stays wrapped, while of the other nucleosome half an outer turn unwraps. The total increase of free DNA is then given by $\Delta L = \frac{1}{2} \times outer\ turn + 1 \times linker = NRL - 113$ bp. A third possibility is that both the outer turns stay wrapped, and only the linker extends, which gives an increase of $\Delta L = 1 \times linker = NRL - 147$ bp.

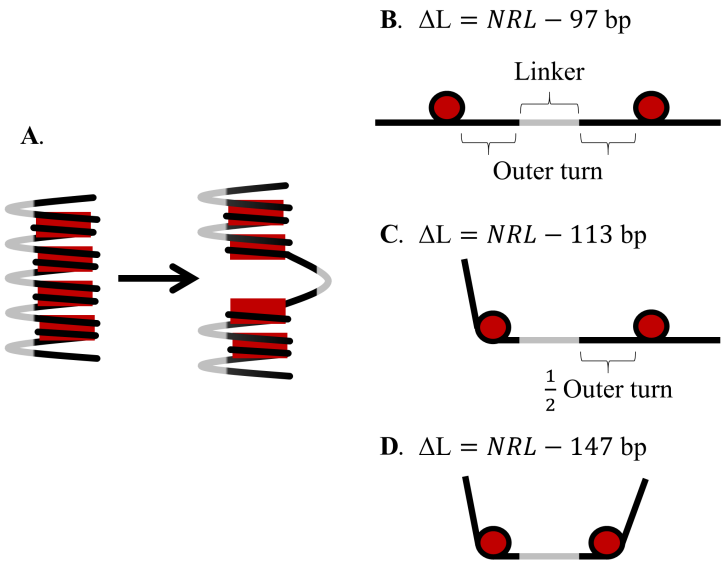


Figure 2.1: *Schematic picture of a pair of nucleosome unstacking.* ***A.*** *Side view of nucleosomes unstacking in a chromatin fiber.* ***B.*** *Half an outer turn unwraps from both nucleosomes.* ***C.*** *Half an outer turn unwraps of just one nucleosome.* ***D.*** *No outer turn unwrapping. The increase of the free DNA ΔL is given in bp and depends on the NRL.*

Version of June 30, 2018– Created June 30, 2018 - 20:00

2.4 Force-Extension Model of Chromatin

In previous work, it has been shown that folded chromatin fibers unfold in three steps as force increases [15]. Before the first transition, the total extension equals that of the DNA plus the extension per stacked nucleosome times the number of nucleosomes. We model the extension per nucleosome in subsequent unfolding steps as a function of force by distinguishing three different states for each nucleosome as displayed in figure 2.2. This is a simplification of the model used by [15], where four states were used. In the first state, nucleosomes interact with each other and stack in a tightly packed chromatin fiber. When the nucleosomes unstack and the outer turn unwraps, the new state is ΔG_1 higher in free energy. This leaves a single wrap of DNA around the histone core. When also the inner turn unwraps, a bare DNA molecule with histones attached remains at an energy cost of ΔG_2. The extensions of these conformations is written as:

$$z_{nuc}(f) = z_{FJC}(f, L_{nuc}) \tag{2.6a}$$
$$z_{DNA}(f) = z_{WLC}(f, L_{DNA}) \tag{2.6b}$$

Where L_{nuc} is the effective contour length of a nucleosome stacked in a chromatin fiber and L_{DNA} is the contour length of the free DNA given by:

$$L_{DNA} = N_{singlywrapped} L_{singlywrapped} + N_{unwrapped} L_{unwrapped} + L_{handles} \tag{2.7}$$

Where N_i is the number of nucleosomes in the i^{th} state, $L_{singlywrapped} = NRL - 79$ is the contour length of a nucleosome with a single wrap of DNA and $L_{unwrapped} = NRL$ is the contour length of an unwrapped nucleosome. The contour lengths are indicated by arrows in figure 2.2. $L_{handles}$ is the contour length of the DNA handles. In this thesis, each state is identified by the number of free base pairs L_{DNA}. This has the advantage that each transition between states is quantified by the number of base pairs released in the transition, which ultimately is the most interesting metric giving structural insight. The total extension $z_{tot}(f)$ of the tether is sum of the extension of the stacked fiber plus the extension of the free DNA:

$$z_{tot}(f) = N_{stacked} z_{nuc}(f) + z_{DNA}(f) \tag{2.8}$$

With $N_{stacked}$ the number of nucleosomes that is stacked in the chromatin fiber. The free energy of each nucleosome in either of the three states is expressed as [15]:

Version of June 30, 2018– Created June 30, 2018 - 20:00

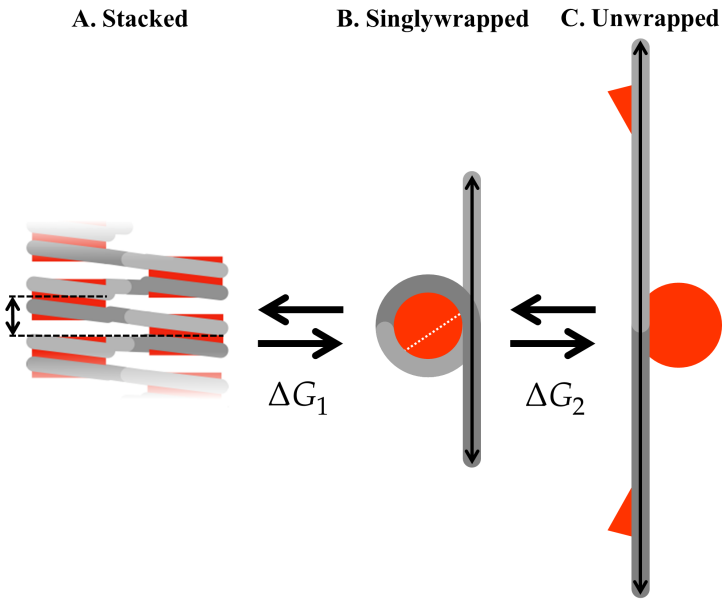


Figure 2.2: *A Schematic picture of the three stage unwrapping of nucleosomes in equilibrium.* ***A.*** *Stacked nucleosomes forming a folded chromatin fiber.* ***B.*** *An unstacked nucleosome with a single wrap of DNA around the histone core.* ***C.*** *A fully unwrapped nucleosome.*

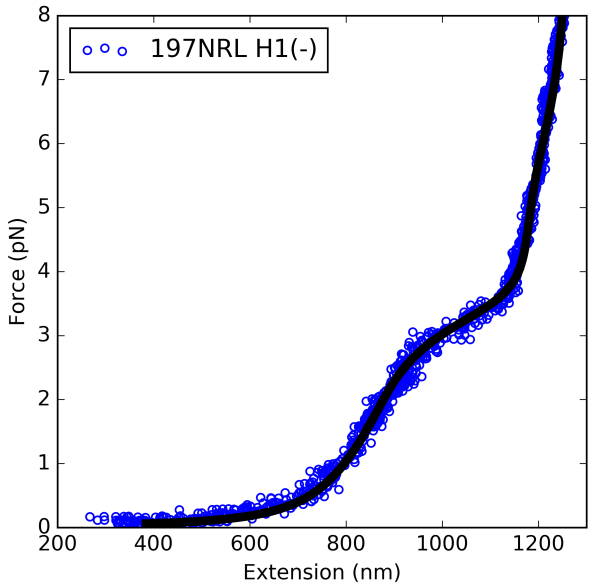


Figure 2.3: *Low force regime of a typical 197 NRL chromatin fiber. Unstacking of nucleosomes is in equilibrium and is fit (black) by eq. 2.11.*

$$G_{stacked}(f) = G_{FJC}(f, L_{nuc}) \tag{2.9a}$$
$$G_{singlywrapped}(f) = G_{WLC}(f, L_{singlywrapped}) + \Delta G_1 \tag{2.9b}$$
$$G_{unwrapped}(f) = G_{WLC}(f, L_{unwrapped}) + \Delta G_1 + \Delta G_2 \tag{2.9c}$$

The total free energy $G_{tot}(f)$ of the tether is the sum over the three states s:

$$G_{tot}(f) = \sum_s N_s G_s(f) + G_{WLC}(f, L_{handles}) \tag{2.10}$$

Where the sum is taken over $s = stacked, singlywrapped, unwrapped$. N_s is the number of nucleosomes in the s^{th} state and $G_s(f)$ is the free energy of the s^{th} state according to eq. 2.9.

Using standard statistical mechanics [15], the average extension of a chromatin fiber in equilibrium is given by:

$$< z_{tot}(f) >= \frac{\sum_s z_{tot}(f) D(s) \exp\left(-\frac{G_{tot}(f)}{k_B T}\right)}{\sum_s D(s) \exp\left(-\frac{G_{tot}(f)}{k_B T}\right)} \tag{2.11}$$

Where $D(s)$ is the degeneracy of state s. Figure 2.3 shows a force-extension curve of a 197 NRL fiber in equilibrium with a fit described by eq. 2.11.

2.5 Non-Equilibrium Model

Clearly, the 25-nm steps are not in equilibrium, since only unwrapping is observed, and no re-wrapping of the nucleosomes' inner turn. The same can be said about the low force unstacking of chromatin fibers when H1 is present; unstacking occurs in discrete steps and re-stacking of nucleosomes is not observed. The rate constant $k_D(0)$ of the transition from the wrapped state to the unwrapped state at zero force is given by the Arrhenius equation:

$$k_D(0) = \frac{1}{\tau} \exp\left(-\frac{\Delta G^{\dagger}}{k_B T}\right) \tag{2.12}$$

Where τ is the lifetime of the open state in equilibrium. The lifetime for the unstacked conformation τ is 2.1 ± 0.1 μs [20]. For the unwrapped conformation τ is typically 10^{-9}-10^{-10}s [4]. $\Delta G^{\dagger}$ is the energy barrier between two states (figure 2.4). Under an external force, the lifetime τ is increased;

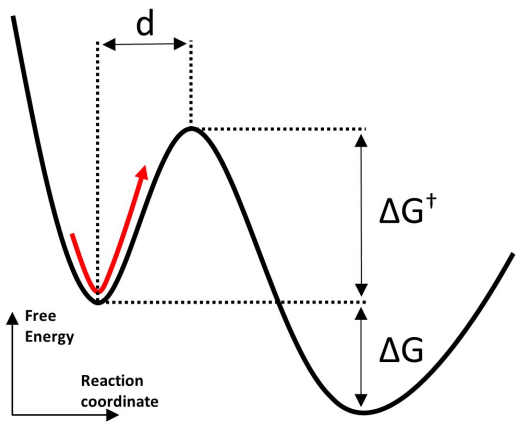


Figure 2.4: *Schematic energy landscape. The energy barrier $\Delta G^{\dagger}$ is at a distance d from the initial state.*

it is more likely to cross the energy barrier $\Delta G^{\dagger}$. The lifetime of the open state is given by [4, 21]:

$$\tau = N \left(\frac{dF}{dt}\right)^{-1} \frac{k_B T}{d} \exp\left(-\frac{\Delta G^{\dagger} - F^* d}{k_B T}\right) \tag{2.13}$$

Where N is the number of nucleosomes that can still be unstacked or unwrapped, $\frac{dF}{dt}$ is the change of force (i.e. the pulling rate), d is the distance between the energy minimum and maximum of the transition in the direction of the extension (figure 2.4) and F^* is the rupture force. The amount by which the energy barrier is decreased is described by the factor $F^* d$; that is the work done by an external force on the chromatin fiber.

The energy involved in nucleosome unstacking and inner turn unwrapping is determined by analysis of the forces at which ruptures occur and the pulling rate. By rewriting eq. 2.13, the most probable force for disruption F^* is written as [4]:

$$F^* = \frac{k_B T}{d}\left[\ln\left(\frac{1}{N}\frac{dF}{dt}\right) - \ln\left(k_D(0)\frac{k_B T}{d}\right)\right] \tag{2.14}$$

With increasing N, the first nucleosome will rupture at a lower force.

Chapter 3

Materials and Methods

3.1 Chromatin Reconstitution

Chromatin fibers were reconstituted on tandem arrays of 601 sequence DNA and prepared for force spectroscopy measurements with Magnetic Tweezers. DNA containing 15 repeats of 197 base pairs 601 sequence was digested using BsaI and BseYI enzymes (NE BioLabs). The digested DNA was labelled on the BsaI site with digoxigenin-11-ddUTP (dig) in presence of dCTP (Roche Diagnostics), klenow fragments and a klenow reaction buffer (ThermoFisher). Subsequently the BseYI site was labelled with biotin-16-ddUTP in presence of dGTP (Roche Diagnostics), klenow fragments and a klenow reaction buffer. Figure 3.1 shows a schematic picture of the DNA after labeling.

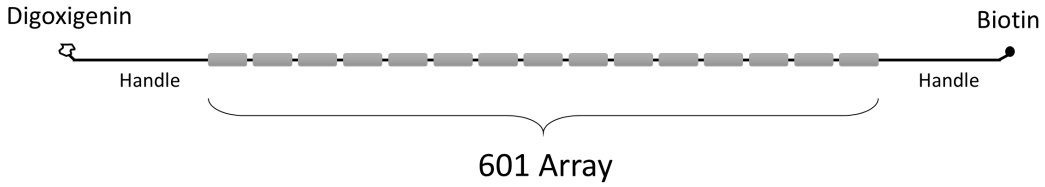


Figure 3.1: *A schematic picture of the labelled DNA. The 601 array (2955 bp) were enclosed by DNA handles (2030 bp), yielding a total length of 4985 bp. The ends were labelled with digoxigenin and biotin.*

Chromatin fibers were reconstituted with a wild-type human recombinand histone octamer (HO) (EpiChypher) consisting of two copies each of H2A, H2B, H3 and H4. The reconstitution was done by salt dialysis [22]. Membranes were placed on a floater in beaker with a high salt buffer

(2M NaCl, 1xTE). The DNA and HO were dissolved in a salt buffer (2.5M NaCl, 1xTE) and pipetted into the tubes with membrane bottoms. The ratio [DNA]:[HO] varied from 1 to 2 over different tubes to ensure that the right amount of nucleosomes were formed in at least one of the tubes. At 4°C, a low salt buffer (1xTE) was pumped into the beaker at a rate of 0.9 ml/min. To maintain a constant volume, the mixed buffer was pumped out of the beaker at the same rate. After 15 hours, the chromatin was collected into fresh tubes and stored in the fridge. The number of nucleosomes in each reconstitution was then confirmed using Magnetic Tweezers.

3.2 Flowcell Preparation

During measurements, the chromatin fibers were immobilized on the bottom of a flowcell. A flowcell consists of two cover slips in a aluminum frame, separated by a thin layer of parafilm. The flowcells can hold a volume of ~ 150 μL and were incubated for two hours with anti-digoxigenin and subsequently passivated for at least 15 hours with 4% BSA and 0.5% Tween-20. The chromatin fibers were dissolved in a measurement buffer containing:

- 100 mM KCl;
- 10mM NaN_3;
- 10 mM HEPES pH 7.5;
- 2 mM $MgCl_2$;
- 0.4% BSA;
- 0.05% Tween-20.

In 200 μL of measurement buffer, 1 ng of chromatin was mixed with 10 μg of paramagnetic Dynabeads^{TM} M-270 coated with streptavidin (ThermoFisher). The streptavidin bound to the biotin at the end of the chroamtin fiber. The sample was then flushed into the flowcell. After 10 minutes of incubation, the digoxigenin end of the fiber was bound to the anti-digoxigenin and 400 μL linker histone H1 (10 ng/μL) was flushed through the flowcell. The flowcell was mounted in the Magnetic Tweezers to measure the extension as a function of force.

3.3 Magnetic Tweezers

The Magnetic Tweezers consist of two permanent magnets, connected to a motorized translation stage. By varying the position of the magnets, a

force ranging from 0 to 85 pN can be achieved. Figure 3.2A displays the setup schematically. The figure is not drawn to scale; the diameter of the chromatin fiber is approximately 30 nm, while the diameter of the bead is 2.8 μm. Using a LabVIEW program, a trajectory for the magnet position was made. The magnet moved with a constant speed, resulting in an exponential increase of force. The sample was illuminated from the top. A Complementary Metal-Oxide-Semiconductor camera (CMOS Vision Condor) on the bottom of the sample detected the diffraction pattern created by the beads at a frame rate of 30 Hz [10]. Figure 3.2B shows the diffraction pattern for three different extensions. The xy-position was determined by locating the centre of the diffraction pattern. The extension of the chromatin fibers was determined by finding the maximum in the cross-correlation of the image and a reference image that consists of concentric rings [Brouwer et. al in prep].

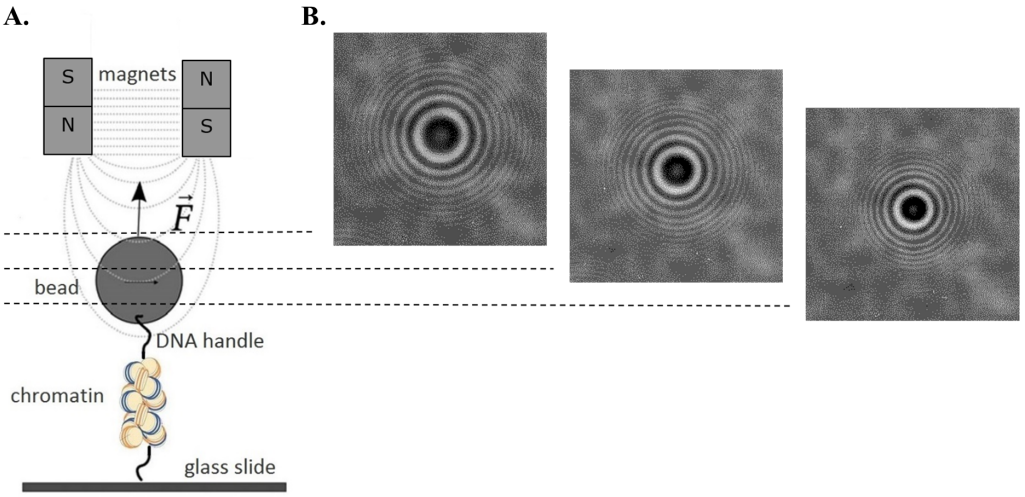


Figure 3.2: *A schematic picture of magnetic tweezers [23].* ***A.*** *A chromatin fiber was attached to the glass surface via a dig-antidig interaction and on the other side bound to a paramagnetic bead via a biotin-streptavidin interaction. A magnetic force was applied by two permanent magnets. By varying the distance between the bead and the magnets, the applied force was changed. The chromatin fiber and the bead are not drawn to scale.* ***B.*** *The extension of the fiber was determined by analysis of the diffraction pattern created by the bead. Three diffraction patterns are shown for different extensions.*

3.4 Data Analysis

The primary analysis was done in LabVIEW. Extension-time measurements were converted into force-extension curves using the simultaneously recorded magnet position and a previously calibrated force-position relation. Beads that were stuck to the surface of the glass showed no extension and were discarded. For the remaining traces, drift was corrected by manually subtracting an offset that was proportional to time until pull and release traces overlapped. An offset and stretching modulus were fitted to a WLC of known contour length to the high force regime of the traces. This was done carefully, to increase the accuracy of the rest of the analysis. The number of nucleosomes was calculated by the difference in extension between the contour length of the WLC fit and the array of singlywrapped nucleosomes state at $\sim$ 10 pN. This number also includes tetrasomes formed in the DNA handles. The number of nucleosomes in the fiber was fitted to the plateau in the low force regime. Traces with too few nucleosomes or too many tetrasomes were also discarded. The stiffness of the fiber was fitted to the initial slope in the force-extension curve. Figure 4.1 shows a typical force-extension curve. The LabVIEW script returned the fit parameters as well as force-extension information of the selected traces. For the rest of the analysis Python was used to develop an algorithm that could extract discrete states from the force-extension curves and to calculate energy barriers from rupture forces.

Version of June 30, 2018– Created June 30, 2018 - 20:00

Chapter 4

Results

The dynamic folding and unfolding of single chromatin fibers with and without H1 was measured using Magnetic Tweezers. Linker histone H1 alters the interaction between nucleosomes and therefore the stability of chromatin fibers. Moreover, it could influence the folding and unfolding the outer turn of nucleosomes [12].

Figure 4.1 shows a typical force-extension curve of a chromatin fiber with and without linker histone H1. When no force is applied to a chromatin fiber, it forms a 30-nm fiber [5, 6]. By gradually increasing the force, the fiber extends until an intermediate state is reached. At this stage, the fiber is an array of singlywrapped nucleosomes. The nucleosomes have unstacked and the outer turn of the nucleosomes have unwrapped [24]. When the force is increased to 12 pN and higher, also the inner turn of the nucleosomes unwraps. This is observed as 25-nm steps in figure 4.1B [4]. After all nucleosomes have unwrapped, a bare DNA molecule with most of the histone octamers attached remains, which can be modelled by the wormlike-chain model. By directly comparing fibers with and without linker histone H1, the influence of the linker histone has been described previously by Kim Vendel [13] and Artur Kaczmarczyk. The main difference with fibers without H1 is the stepwise unstacking of fibers in presence of linker histone. Furthermore, fibers stiffen which is seen in figure 4.1C as an increase of the initial slope. Stacking energies increase resulting in a raise of the plateau in the same figure. In the high force regime, 25 nm steps were seen in presence as well as in absence of linker histone. The overall shape of the force-extension curves for H1(+) and H1(-) is similar at forces above 10 pN. The abrupt jumps in the low force regime when linker histone is present is what distinguishes the types of fibers.

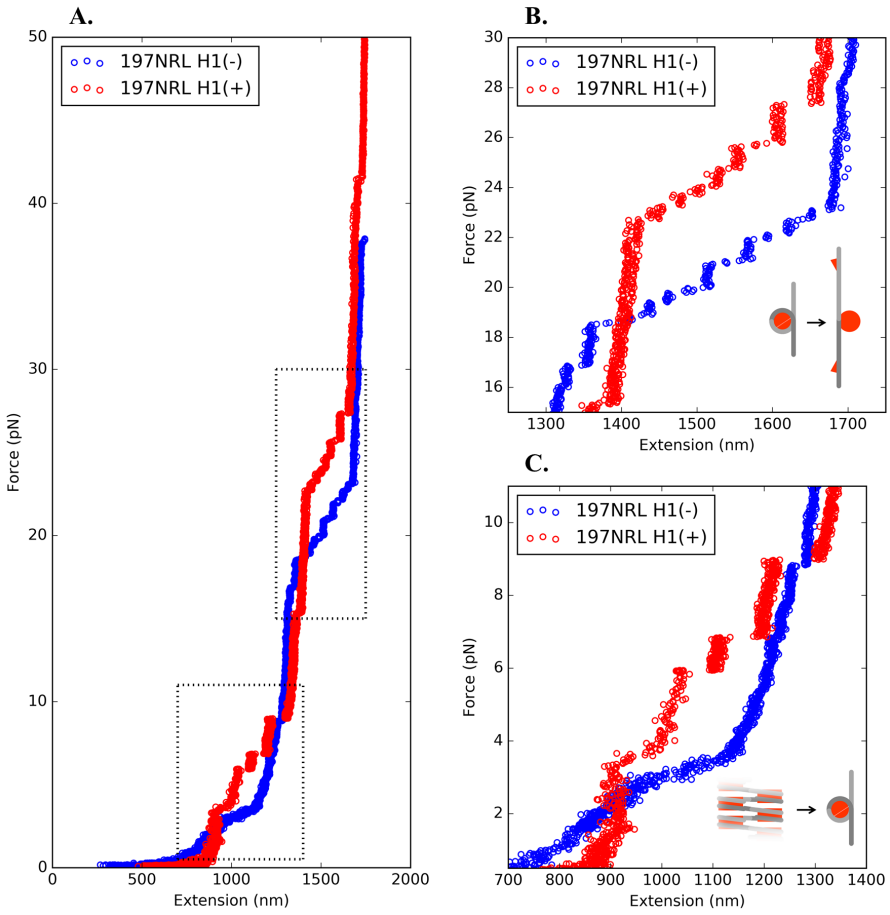


Figure 4.1: *Linker histone H1 induces discrete unstacking of 15x197 NRL 601 array fibers at forces less than* 10 pN. ***A.*** *Force-extension curve of a pulling trace of single 15x197 NRL chromatin fibers with (red) and without (blue) linker histone H1 present.* ***B.*** *Zoom at the high force regime of (A). 25-nm steps are indicative of unwrapping the inner turn of the nucleosome [4] and are similar for both molecules.* ***C.*** *Zoom at the low force regime of (A). In this force regime nucleosomes unstack. Discrete steps occur when H1 is present. (Data: A. Kaczmarczyk)*

4.1 Finding Discrete States

To model the discrete states in the force-extension curve, states were drawn according to eq. 2.8 for increasing the amount of free DNA in 1 bp steps (figure 4.2). However, because we want to develop a method for which the NRL is not know a priori, nor is the symmetry of unwrapping, the number of nucleosomes was not decremented in discrete steps, but was decreased continuously between $N = N_{tot}$ and $N = 0$. For each state, the probability that a data point belonged to it was calculated. This was done using the Z-score. The Z-score is a measure for the distance between a point and its expected value in terms of standard deviations and can be written as:

$$Z = \frac{|X - \mu|}{\sigma} \tag{4.1}$$

Where X is the measured value, μ is the expected value and σ is the standard deviation. In the case of force-extension measurements, X will be the measured extension at a certain force and μ will be the extension calculated by the combination of a WLC and a FJC (eq. 2.8). The expected standard deviation σ^2_{exp} of the extension can be expressed in terms of a measurement error and thermal fluctuations:

$$\sigma^2_{exp} = \sigma^2_{M_{err}} + \sigma^2_{TF} \tag{4.2}$$

Where $\sigma_{M_{err}} \approx 5$ nm and the thermal fluctuations σ_{TF} was calculated using equipartition for a one dimensional deviation from equilibrium:

$$\frac{1}{2}k_B T = \frac{1}{2}q\sigma^2_{TF} \tag{4.3}$$

Where q is the stiffness. The stiffness was calculated from the force-extension model using $\frac{dF}{dz}$.

The probability of a data point i belonging to a certain state S_L can now be written as:

$$P_i(S_L) = 1 - \mathrm{erf}\,(Z) \tag{4.4}$$

This yields a probability between 0 and 1; the bigger the probability, the the more likely a point is in the given state.

By summing the probabilities for all points for each state and normalizing it, a probability density distribution is obtained that tells which states are occupied during the force ramp (figure 4.2B):

$$P(S_L) = \frac{\sum_i P_i(S_L)}{\sum_S \sum_i P_i(S_L)} \tag{4.5}$$

The peaks in the probability density distribution indicate the states that are most likely to exist. States were assigned only when they were composed of two data points or more. In order to have at least two points in a single state, the probability density must be at least $2 \times (\sum_S \sum_i P_i(S_L))^{-1}$. In the example of figure 4.2, 19 states were found.

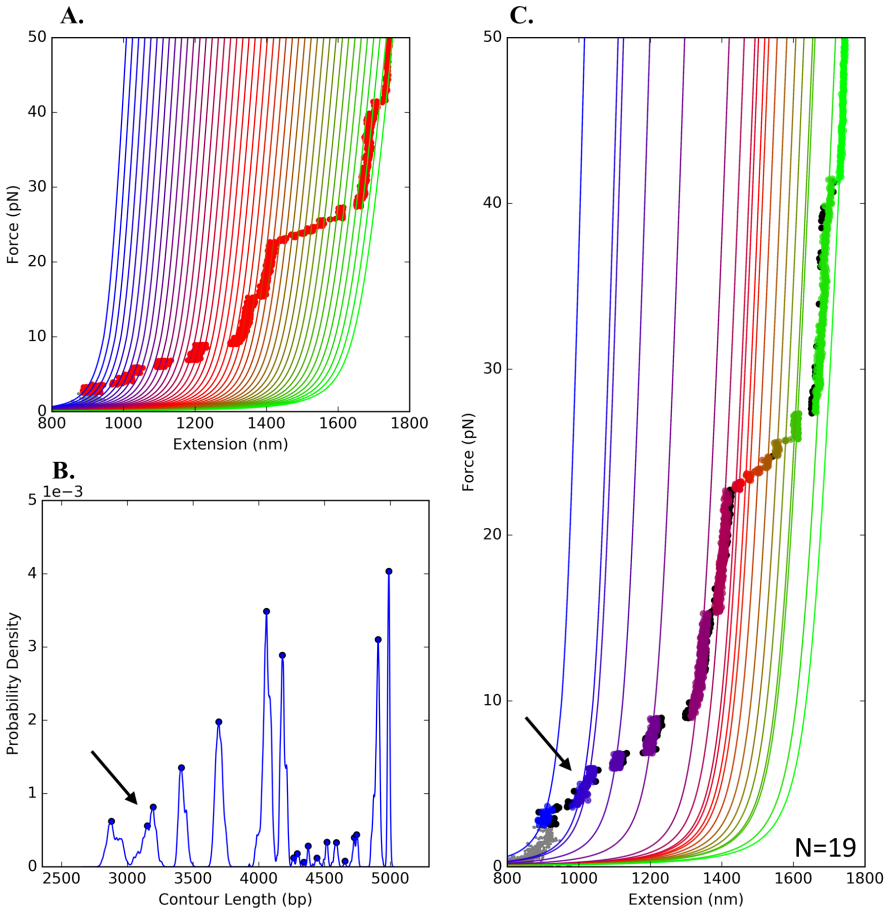


Figure 4.2: *Extraction of discrete states in force-extension curves.* ***A.*** *The same force-extension curve as in figure 4.1 A. The lines show the possible states according to eq. 2.8. To keep the figure clear, only states separated by 50 bp are shown, but we tested states with 1 bp intervals* ***B.*** *Probability distribution for all states in* 1 *bp steps. The peaks in the distribution indicate the states that are most likely to exist.* ***C.*** *Colored lines show the states corresponding to the peaks in the probability distribution. Each state is plotted in a different color and data points within* 2σ *of a state are shown in the same color. Data points that do not belong to any state are shown in black. The grey data points (*$f < 10$ *pN) were not used during the analysis.*

4.2 Merging States

Sometimes clusters of points appeared to be wrongly divided into multiple states. In figure 4.2, two seperate states were found for the group of points indicated by the arrow. In order to correct this, the states were merged if the data of the clusters fulfilled specific criteria. Since the data is expected to be normally distributed, the Z-score could again be used to attribute data points to each seperate state. The threshold was set at 2σ, which should capture 95% of the points. Alternatively, a merged state with a contour length that is calculated by the weighted average of the points belonging to both initial states was evaluated. The following criteria were set to determine whether two states were merged into this new state:

1. Each of the two initial states must have at least 50% of the data within 2σ of the other state;

2. The merged state must have at least 80% of the points within 2σ.

Two states were merged if both of these criteria were fulfilled. An example is shown in figure 4.3. Initially, two states were found for the second group of data points, as indicated by the peaks in the probability density distribution (figure 4.2B). These two states have 75% of the points within 2σ of the other state. The merged state with a contour length of free DNA that is the weighted average of the two old states, has 96% of the points within 2σ. Both criteria hold so the new state was enforced. Of the 19 states in figure 4.2C, 15 remained in figure 4.3C. Removing falsely identified states from the data results in more accurate transitions.

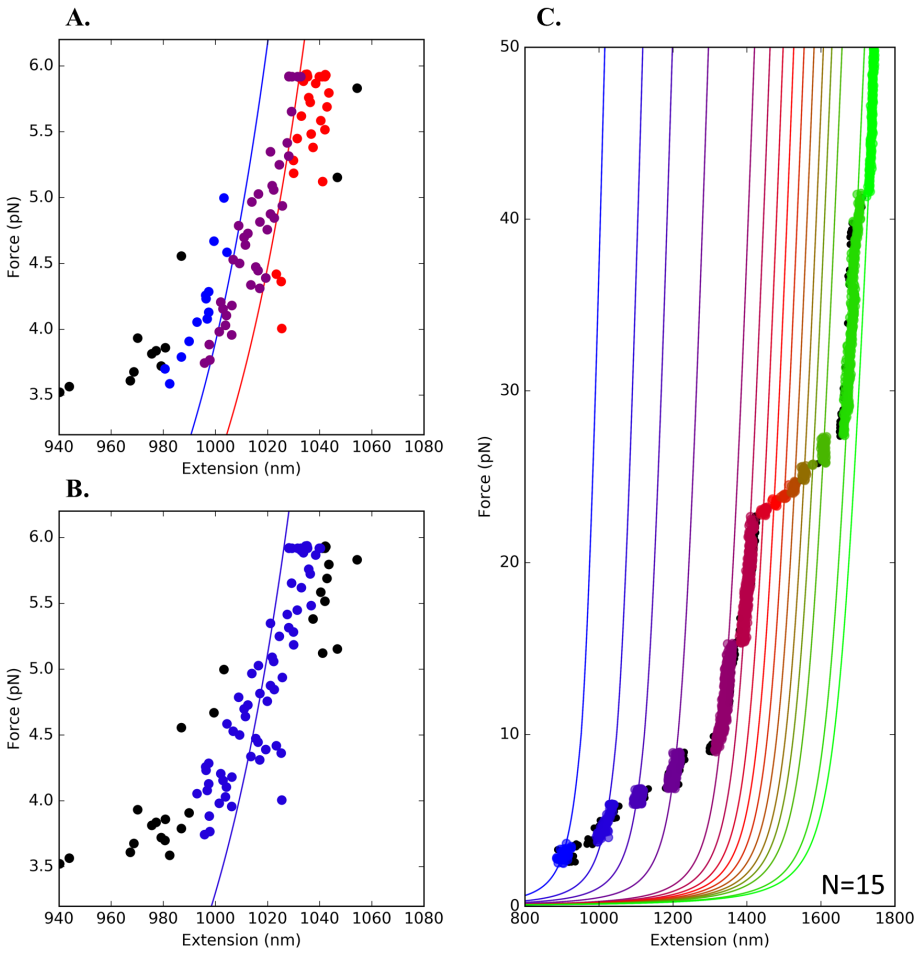


Figure 4.3: *Merging states based on overlapping probability density distributions.* ***A.*** *Close-up of the second cluster of data points of figure 4.2C. The red and blue state are in accordance with the peaks indicated by the arrow in figure 4.2B. Data points within 2σ of a state are shown in in the same color as the line representing the state. Overlapping points (i.e. within 2σ of both states) are shown in purple.* ***B.*** *Result after merging states states of (A). The cluster is now described by one state. The merged state is the weighted average of the two initial states. Note though that more points were not assigned (black).* ***C.*** *Full force-extension curve after merging. Each cluster is described by a single state.*

4.3 Unstacking Stepsizes

Two transitions were distinguished during the stepsize analysis: nucleosome unstacking and inner turn unwrapping of nucleosomes. For the analysis, 15 independent chromatin fibers were used. Stepsizes were calculated by taking the difference in contour length between two neighbouring states. The state of singlywrapped nucleosomes was used as a reference, with a contour length L_{ref}. In this state, at forces of $\sim$ 10 pN, all nucleosomes have only the inner turn wrapped.

The distinction between unwrapping steps and unstacking steps was made by comparing the contour length of the state to with L_{ref}. Steps at a contour length smaller than L_{ref} were considered unstacking steps, steps at a contour length larger than L_{ref} were considered inner turn unwrapping steps.

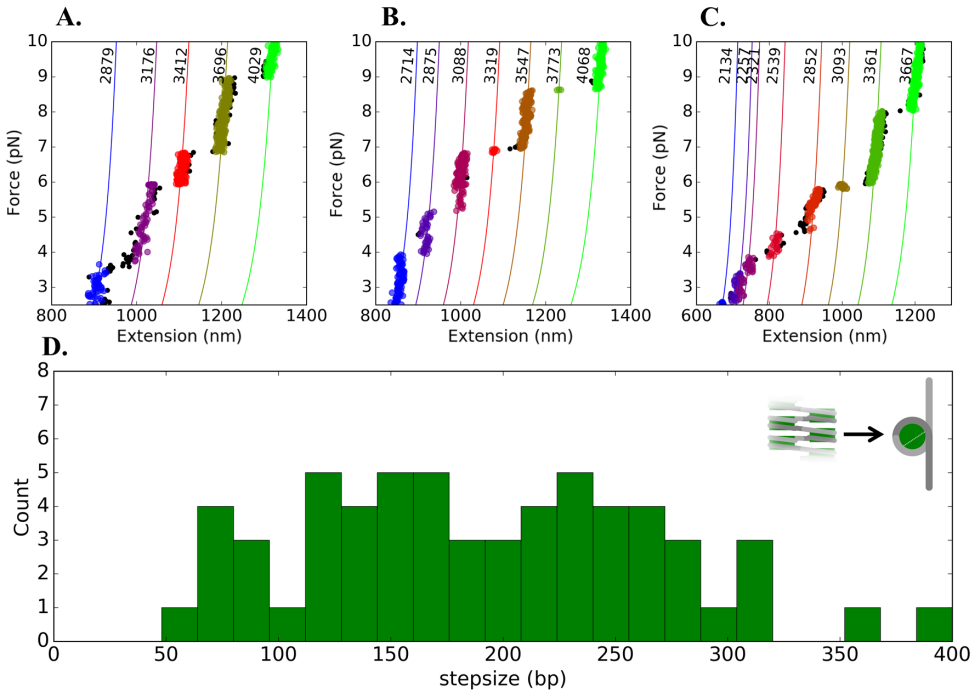


Figure 4.4: *Stepsize distribution for unstacking of nucleosomes of 197 NRL fibers with linker histones.* ***A-C.*** *Low force regime for 3 different molecules. The contour length of each state is labeled in base pairs.* ***D.*** *Histogram of steps caused by unstacking of the chromatin fiber for 15 independent molecules. The stepsize is defined by the difference in contour length between two states. (Data: A. Kaczmarczyk)*

Figure 4.4 shows the low force regime of some typical traces and a histogram of the stepsizes for unstacking transitions. The unstacking and outer turn unwrapping of a pair of nucleosomes, such that the inner turn wrap remains, gives a maximal extension of $197 - 79 = 118$ bp. This assumes that the outer turn unwraps symmetrically. Asymmetric unwrapping gives smaller stepsizes. figure 4.4D shows that in general, stepsizes are larger than these 118 bp. This indicates that nucleosomes do not unstack one at a time, but rather unstack in a cooperative way.

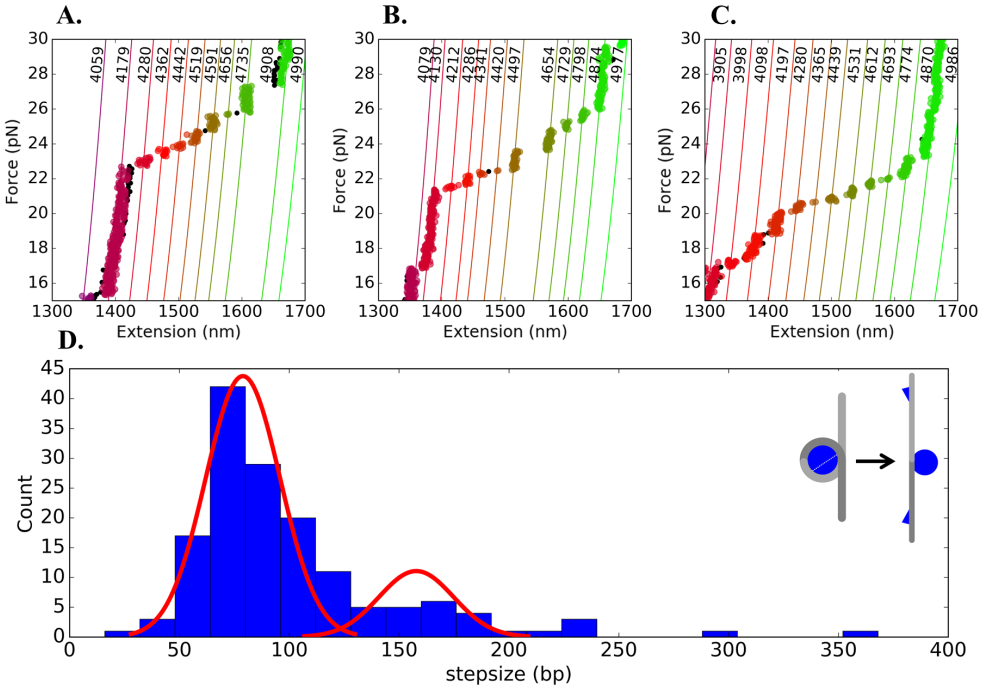


Figure 4.5: *Stepsize distribution for unwrapping the inner turn of nucleosomes of 197 NRL fibers with linker histone added.* ***A-C.*** *High force regime for 3 independent molecules. The contour length of each state is labeled in base pairs.* ***D.*** *Histogram of steps caused by inner turn unwrapping of nucleosomes for 15 independent chromatin fibers. The stepsize is defined by the difference in contour length between two states. A double Gaussian fit with mean stepsizes of $\mu = 78.9 \pm 0.3$ bp (mean $\pm$ SE) and $2\mu = 157.8 \pm 0.5$ bp (mean $\pm$ SE) is shown in red. (Data: A. Kaczmarczyk)*

4.4 Unwrapping Stepsizes

Figure 4.5 shows the high force regime of some typical traces and a histogram of the stepsizes for inner turn unwrapping transitions. A clear peak can be seen at ± 80 bp, caused by unwrapping of a single nucleosome (figure 4.5D). Nucleosomes could also unstack simultaneously. This would result in a peak at double the stepsize of a single unwrap. In the histogram, a less well-defined peak can be seen at ± 160 bp. A double error function was fitted to the cumulative stepsize distribution to retrieve the number of base pair in the inner turn wrap:

$$\mathcal{N}(x) = a_1 \left[1 + \mathrm{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right] + a_2 \left[1 + \mathrm{erf}\left(\frac{x-2\mu}{\sigma\sqrt{2}}\right)\right] \tag{4.6}$$

Where a_1, a_2 are the amplitudes of the first and second Gaussian, μ is the mean stepsize of the first Gaussian and σ is the standard deviation. The fit yields a mean stepsize $\mu = 78.9 \pm 0.3$ bp (mean $\pm$ SE).

4.5 Transition Barriers

Figure 4.6A displays the most probable force F^* versus $\ln\left(\frac{1}{N}\frac{dF}{dt}\right)$. The distance d and the rate constant $k_D(0)$ can be extracted from the linear fit. d is the distance a nucleosome can stretch before it transitions. This can be compared to stretching a rubber band; d is the distance a rubber band can stretch, before it breaks. The best fit yields $d_1 = 4.5 \pm 0.8$ nm for nucleosome unstacking. This is in the same order of magnitude as the range of the lysine-16 (H4-K16Ac) in the H4 tail ($\approx$ 3.5 nm from the H4-core), which is mostly responsible for nucleosome-nucleosome interactions [3, 25]. For inner turn unwrapping $d_2 = 0.76 \pm 0.04$ nm was obtained. The energy barrier $\Delta G^\dagger$ was calculated using equation 2.12. This yields $\Delta G_1^\dagger = 21.9 \pm 0.2\ k_B T$ for nucleosome unstacking and $\Delta G_2^\dagger = 29.6 \pm 1.2\ k_B T$ for inner turn unwrapping of nucleosomes.

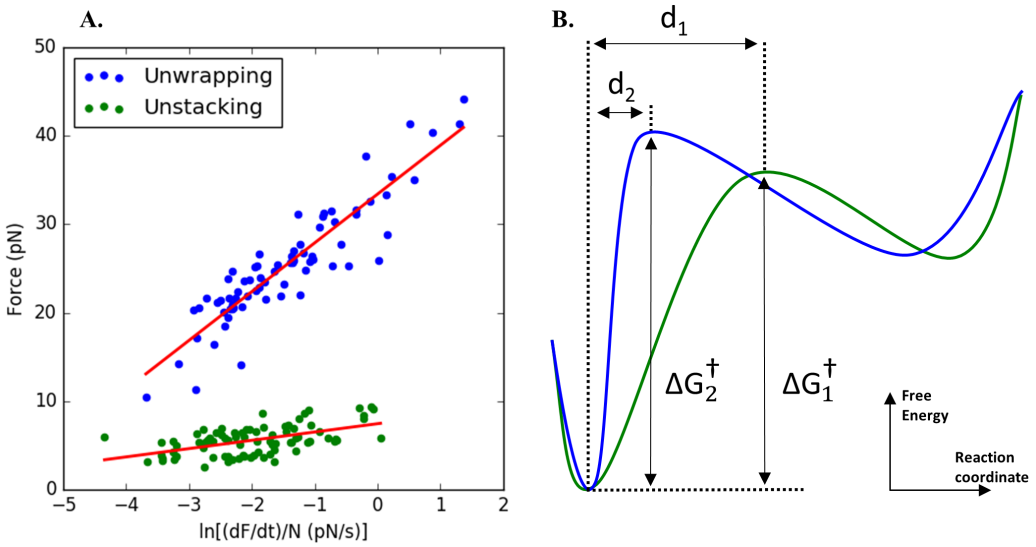


Figure 4.6: *Barrier for unstacking and unwrapping transitions.* ***A.*** *Rupture forces vs.* $\ln([dF/dt]/N)$ *for both unstacking steps (green) and unwrapping (blue). Only unwrapping of single nucleosomes, with stepsizes* 75 ± 15 *bp, were taken into account. The lines are the best linear fits for both transitions. The fits yield* $d_1 = 4.5 \pm 0.8$ *nm and* $\Delta G_1^\dagger = 21.9 \pm 0.2\ k_BT$ *for unstacking and* $d_2 = 0.76 \pm 0.04$ *nm and* $\Delta G_2^\dagger = 29.6 \pm 1.4\ k_BT$ *for unwrapping.* ***B.*** *Schematic energy landscape of the unstacking of the fiber (green) and outer turn unwrapping of the nucleosome (blue). The energy barrier* $\Delta G^\dagger$ *is at a distance d from the initial state.*

Chapter 5

Discussion & Conclusion

I have introduced a novel method to quantify discrete steps in chromatin (un)folding and extract stepsizes from force-extension measurements. The novelty of this method is that the NRL does not have to be known a priori. We tested this method on 197 NRL chromatin fibers in presence of linker histone H1 and quantified stepsizes and energies.

Unstacking events of chromatin fibers in presence of linker histone H1 feature a broad distribution of stepsizes, indicating that nucleosomes often unstack simultaneously. Less stacking steps are observed with respect to unwrapping steps. Since a fiber with a fixed NRL of 197 bp was used, the total number of nucleosomes that unstack can be calculated by assuming the steps up to 120 bp as 1 nucleosome unstacking, up to 240 bp as 2 nucleosomes unstacking, up to 360 bp as 3 nucleosomes unstacking and up to 480 bp as 4 nucleosomes unstacking, which gives a total number of 134. This is slightly less than the 160 unwrapped nucleosomes that could be counted similarly. This difference could be explained by the fact that some nucleosomes were formed in the DNA handles, which were not stacked in the fiber, and that in an array of N nucleosomes, at most $N-1$ unstacking events can take place.

The stepsize distribution of unwrapping events showed a clear peak at 79 bp. In the histogram in figure 4.5D, an asymmetry is seen under the first Gaussian. This is because not all states were merged correctly. This can result in a state that is in between two groups. During the stepsize analysis, this results in two steps that are larger than they should be. For the inner turn unwrapping, these steps are approximately $1.5 \times 79 \approx 119$ bp. The statistical error for each bin is not shown, since the fit was done to the cumulative stepsize distribution. The mean stepsize found for inner turn unwrapping of nucleosomes is in accordance with [4, 12, 15], which

is a validation of the model.

By analysis of rupture forces and pulling rate, a maximum extension before nucleosomes unstack of 4.5 nm was found. The distance from the H4 core to the lysine-16 in the H4 tail is in the same order of magnitude ($\approx$ 3.5 nm) [25]. Singlywrapped nucleosomes can be extended up to 0.8 nm before they unwrap, which is reasonable given the size of the histone core [3]. The maximum extension of a singlywrapped nucleosome significantly lower than measured by [4]. This could be caused by the difference in buffer conditions, the type of nucleosomes that were used (Avian vs. Human recombinand) or the different DNA sequence (Sea urchin 5S vs. 601 sequence). These factors play a role in the interactions between the DNA and the histones. The loading force was controlled using a velocity clamp by [4]. In that sequence, the force decreased each time a nucleosome ruptured. This effectively starts over the experiment, justifying the $1/N$ relation in eq. 2.14. In Magnetic Tweezers experiments, after each rupture the force increased, so $1/N$ might not be valid in our case. A second difference is that during velocity clamp the loading rate is constant, while in a Magnetic Tweezers experiment it is not constant, so eq. 2.14 may not be valid. Non the less, the average rupture force is significantly higher, indicating a clear difference in the transition.

Previous studies have described chromatin unfolding by four stages [15], while I distinguish only three stage. To improve this model, an extra stage could be added after the singlywrapped state, like described by [15], or before the singlywrapped state. For the latter, the outer turn will not have fully unwrapped. Expanding the model with this extra stage will decrease the energy barrier $\Delta G_1^\dagger$, and decrease the distance d. In addition, an extra energy barrier would be added between these stages. The sum of the two new barrier should give the value of $\Delta G_1^\dagger$ as stated in the result section. Another improvement would be to fit Gaussians to the peaks in the probability density distributions. The goodness of the fit could be a better criterion for defining states. The fit error would then be a measure for the uncertainty of the contour length of a state.

It should now be possible to quantify nucleosome unstacking without knowing NRL or symmetry of unwrapping a priori. This is an important step to quantify natively assembled chromatin [26], where nucleosomes are distributed with various distances and post-translational modifications influence stacking interactions. Natively assembled chromatin is therefore a much more heterogeneous fiber compared to the 601 arrays. This model could be key to gain a better understanding in the higher order structure of natively folded chromatin.

Acknowledgements

I would like to thank Nicolaas Hermans for his patience in guiding me through my project and for answering all my questions, John van Noort for making the project possible and the whole “Van Noort group” for the valuable discussions.

Bibliography

[1] J. T. Finch, L. C. Lutter, D. Rhodes, R. S. Brown, B. Rushton, M. Levitt, and A. Klug, *Structure of nucleosome core particles of chromatin*, Nature **269**, 29 (1977).

[2] T. J. Richmond, J. T. Finch, B. Rushton, D. Rhodes, and A. Klug, *Structure of the nucleosome core particle at 7 Å resolution*, Nature **311**, 532 (1984).

[3] 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).

[4] B. D. Brower-Toland, C. L. Smith, R. C. Yeh, J. T. Lis, C. L. Peterson, and M. D. Wang, *Mechanical disruption of individual nucleosomes reveals a reversible multistage release of DNA*, Proceedings of the National Academy of Sciences **99**, 1960 (2002).

[5] H. Ris and D. F. Kubai, *Chromosome Structure*, Annual Review of Genetics **4**, 263 (1970).

[6] H. G. Davies and J. V. Small, *Structural Units in Chromatin and their Orientation on Membranes*, Nature **217**, 1122 (1968).

[7] G. Li, R. Margueron, G. Hu, D. Stokes, Y.-H. Wang, and D. Reinberg, *Highly Compacted Chromatin Formed In Vitro Reflects the Dynamics of Transcription Activation In Vivo*, Molecular Cell **38**, 41 (2010).

[8] G. Li and D. Reinberg, *Chromatin higher-order structures and gene regulation*, Current Opinion in Genetics & Development **21**, 175 (2011).

[9] F. Song, P. Chen, D. Sun, M. Wang, L. Dong, D. Liang, R.-M. Xu, P. Zhu, and G. Li, *Cryo-EM Study of the Chromatin Fiber Reveals a Double Helix Twisted by Tetranucleosomal Units*, Science **344**, 376 (2014).

[10] A. Kaczmarczyk, A. Allahverdi, T. B. Brouwer, L. Nordenskiöld, N. H. Dekker, and J. van Noort, *Single-molecule force spectroscopy on histone H4 tail-cross-linked chromatin reveals fiber folding*, Journal of Biological Chemistry **292**, 17506 (2017).

[11] B.-R. Zhou, J. Jiang, H. Feng, R. Ghirlando, T. S. Xiao, and Y. Bai, *Structural Mechanisms of Nucleosome Recognition by Linker Histones*, Molecular Cell **59**, 628 (2015).

[12] W. Li, P. Chen, J. Yu, L. Dong, D. Liang, J. Feng, J. Yan, P.-Y. Wang, Q. Li, Z. Zhang, M. Li, and G. Li, *FACT Remodels the Tetranucleosomal Unit of Chromatin Fibers for Gene Transcription*, Molecular Cell **64**, 120 (2016).

[13] K. Vendel, *The role of the linker histone in chromatin structure and dynamics*, MSc. Thesis Leiden University (2015).

[14] J. F. Marko and E. D. Siggia, *Stretching DNA*, Macromolecules **28**, 8759 (1995).

[15] H. Meng, K. Andresen, and J. van Noort, *Quantitative analysis of single-molecule force spectroscopy on folded chromatin fibers*, Nucleic Acids Research **43**, 3578 (2015).

[16] T. R. Strick, J.-F. Allemand, V. Croquette, and D. Bensimon, *Physical Approaches to the Study of DNA*, Journal of Statistical Physics **93**, 647 (1998).

[17] C. Bustamante, S. B. Smith, J. Liphardt, and D. Smith, *Single-molecule studies of DNA mechanics*, Current Opinion in Structural Biology **10**, 279 (2000).

[18] C. Bouchiat, M. Wang, J.-F. Allemand, T. Strick, S. Block, and V. Croquette, *Estimating the Persistence Length of a Worm-Like Chain Molecule from Force-Extension Measurements*, Biophysical Journal **76**, 409 (1999).

[19] T. T. Ngo, Q. Zhang, R. Zhou, J. G. Yodh, and T. Ha, *Asymmetric unwrapping of nucleosomes under tension directed by DNA local flexibility*, Cell **160**, 1135 (2015).

[20] F.-t. Chien, *Chromatin Dynamics resolved with Force Spectroscopy*, PhD Thesis Leiden University (2011).

[21] H. Evans and P. Williams, *Physics of bio-molecules and cells*, Springer, Berlin, 2002.

[22] 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*, Proceedings of the National Academy of Sciences **103**, 6506 (2006).

[23] F.-T. Chien and J. van Noort, *10 Years of Tension on Chromatin: Results from Single Molecule Force Spectroscopy*, Current Pharmaceutical Biotechnology **10**, 474 (2009).

[24] M. Kruithof, F.-T. Chien, A. Routh, C. Logie, D. Rhodes, and J. van Noort, *Single-molecule force spectroscopy reveals a highly compliant helical folding for the 30-nm chromatin fiber*, Nature Structural & Molecular Biology **16**, 534 (2009).

[25] M. Shogren-Knaak, *Histone H4-K16 Acetylation Controls Chromatin Structure and Protein Interactions*, Science **311**, 844 (2006).

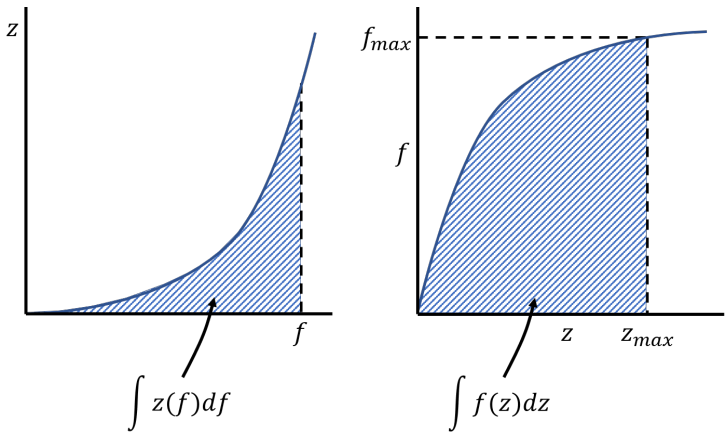
[26] N. Hermans, J. J. Huisman, T. B. Brouwer, C. Schächner, G. P. H. van Heusden, J. Griesenbeck, and J. van Noort, *Toehold-enhanced LNA probes for selective pull down and single-molecule analysis of native chromatin*, Scientific Reports **7**, 16721 (2017).

Apendix

Apendix A

Extension is measured as a function of force $z(f, L)$. A simple mathematical tool can be used to determine the free energy G, without having to invert $z(f)$ first:

$$G(f, L) = \int_0^{z_{max}} f(z)dz$$



$$G(f, L) = f_{max}z_{max} - \int_0^{f_{max}} z(f, L)df$$