# Probing the Mechanics of Linker DNA in Folded Chromatin Fibers with Monte Carlo Simulations

# Probing the Mechanics of Linker DNA in Folded Chromatin Fibers with Monte Carlo Simulations

**Willem Jan de Voogd**

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

January 28, 2019

## Abstract

Chromatin is a dense structure of DNA and histone proteins. DNA wraps in units of 147 bps around 8 histones forming the nucleosomes. Strings of nucleosomes stack into dense 30 nm fibers, but the structure of these fibers remains disputed. Force spectroscopy experiments show that the linker length affects the characteristic unfolding of nucleosome-arrays and suggests two alternative conformations. However, the precise influence of linker base pairs remains unclear. Here we use Monte Carlo simulations and show that DNA twist-energy largely determines whether nucleosomes stack in a 1-start or 2-start conformation. In addition, we show with Mutation Monte Carlo simulations that the sequence of the linker DNA may affect nucleosome stacking. Thus both the sequence and length of the linker DNA are important for the higher-order structure of 30 nm chromatin fibers. Furthermore, the simulations form a bridge between theory and experiment as predicted changes in chromatin folding can be verified experimentally.

# Contents

# Chapter 1

# Introduction

## 1.1 DNA to Chromatin

All of the genetic information of cells is encoded on strings of deoxyribonucleic acid (DNA). The macromolecule comes in two strands of molecules, and each strand consists of 4 different types basic nucleotides. In double stranded DNA these bases form hydrogenbonds with their counter-nucleotides. These base pairs (bps) consist of adenine (A) bonded with Thymine (T) and guanine (G) bonded with cytosine (C). As such, both strands carry the same information encoded on their sequence. Two strands form a left-handed helical structure with a pitch of 10.4 bps. Two subsequent nucleotides in one strand are called dinucleotides [1, 2].

The DNA needs to be safely stored to avert errors in the sequence. Each human cell has approximately two meters of DNA condensed into the nucleus, which has a diameter of around 6 $\mu$m. This extreme condensation is achieved by storing the DNA in chromatin [3].

This is a higher order structure where the DNA in vitro is found wrapped around nucleosomes. These nucleosomes form dense arrays [5], with a diameter of approximately 30 nm [1], see figure 1.1. Chromatin not only condenses DNA it also regulates gene expression and replication of DNA. For a better understanding of these processes, it is important to know more about the structural and mechanical properties of chromatin. Over the last decades' researchers have speculated about many of the features of the 30 nm fiber, but there is still no consensus on its structure and even its existence in vivo has been questioned [6, 7]. Here, we will review the structure of nucleosomes, which has been resolved in great detail and describe a coarse-grained model that captures the mechanical properties of DNA. By combining these, we aim to get a better, quantitative understanding of
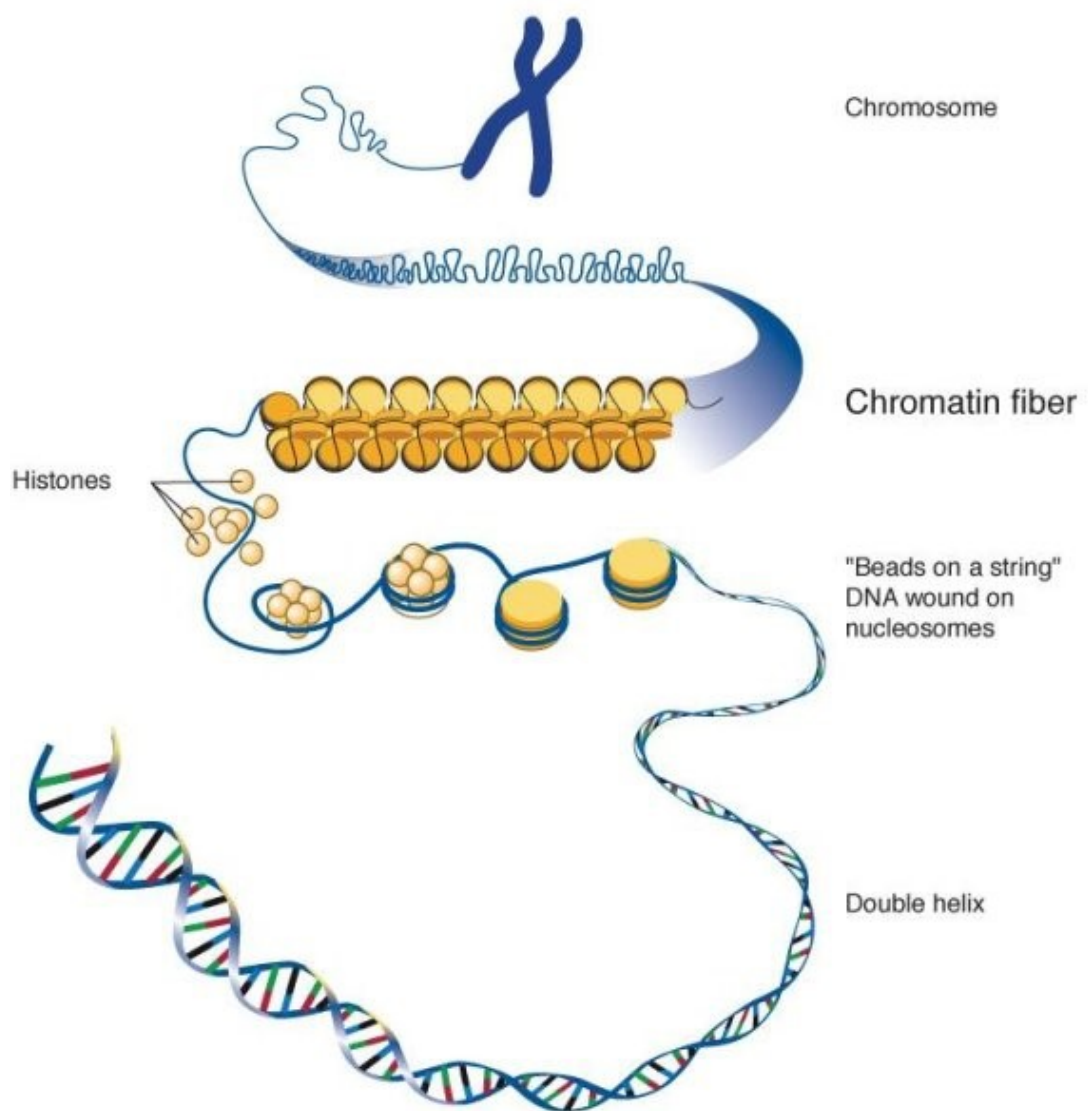
the folding of chromatin fibers.



**Figure 1.1:** *The general accepted view of DNA folding into the 30 nm chromatin fibers. Figure from Creative Diagnostics Group [4].*

## 1.2 The Nucleosome Core Particle

X-ray images have uncovered the structure of the nucleosome at 2.8 Å resolution [8]. 146 bps are wrapped in a superhelix 1.65 turns around proteins forming a nucleosome core particle (NCP). The core of the nucleosome is built of two pairs of the four histones H2A, H2B, H3 and H4. Linker DNA forms the connection between two subsequent nucleosomes. In vivo, nucleosomes can be positioned anywhere on the DNA, though it is now clear that sequence preferences drive nucleosomes to preferred positions. In synthetic chromatin fibers, the Widom 601 sequence is often used to reconstitute chromatin fibers with perfectly controlled spacing between the nucleosomes. This spacing is called the nucleosome repeating length (NRL). Using such regular nucleosomal arrays, cryo-EM resolved the structure of chromatin fibers in more detail [9]. However, such studies only investigated a very limited number of NRL's, providing a biased view on chromatin structure. Moreover, single-molecule force spectroscopy studies on such fibers suggest a different structure than that resolved by cryo-EM, which could be due to different preparation conditions. This shows our limited understanding of chromatin folding.

The effect of nucleosome wrapping has also been investigated in simulations. In molecular dynamics (MD) simulations a numerical approach is used to solve for the laws of motion for each atom in the system of molecules. As a result, MD provides a way of assessing the dynamical evolution of a system. MD simulations showed, in agreement with experiment, that the DNA-histone interactions are localized at specific positions in the nucleosome. Furthermore, the wrapping of DNA is stabilized by the interaction of the histonetails of H3 within the NCP units [10, 11]. However, MD is a computational very expensive method and time-scales are much smaller than experimental time-scales.

To address larger structures and time scales, it is therefore necessary to coarse-grain the system. An obvious level of coarse-graining is the base-pair, which we will use in this thesis. In addition to coarse-graining, one can abandon the laws of motion and probe statistically independent structures, using Monte Carlo (MC) simulations. Using this method, De Bruin et al. showed that the asymmetric unwrapping of nucleosomes can be explained by the sequence of nucleosomal DNA [12]. Furthermore, MC-simulations were used to compute nucleosome position preferences in the DNA [13, 14].

All these results provide a better understanding of the NCP. However, the higher order structures, that are important for the regulation of the gene expression, cannot be understood from evaluation of single nucleosomes.

## 1.3   Configuration of Nucleosome-Arrays

The NCP-units tend to stack in pairs. Experimental and theoretical data indicates that this interaction is mainly mediated by the tail of the H4-histone [15, 16]. Coarse grained (CG) MD demonstrated that electrostatic potentials are decisive for the interactions of the H2A and H4 histoneterminals with the NCP units [17, 18].

However, the configuration of stacked nucleosome-arrays remains disputed. In structural studies of fixed chromatin fibers two types of configurations were seen. Measurements with electron microscopy indicated a 1-start fiber, where the nucleosomes form one stack [19]. On the other hand, other EM-studies [20], as well as X-ray experiments indicated the existence of 2-start fibers [21]. In the 2-start fiber nucleosomes interact with their next neighbor, while 1-start fibers interact with their direct neighbor. Both configurations wind up to a left-handed superhelix [22].

Single-molecule force spectroscopy experiments included the dynamics of chromatin. In these experiments, forces of several pico-Newtons are applied to unfold the chromatin. Physical models for the experimental force-extension data couple the dynamics to structural features. For instance in the WLC-model the DNA-helix is approximated as a continuous flexible rod, where the potential energy depends only upon curvature. The WLC-model can quantify DNA flexibility by fitting the persistence length, which was found to be 50 *nm* [23]. The unfolding of chromatin fibers depends on this DNA flexibility but also on the unwrapping and unstacking of nucleosomes and therefore requires a much more intricate model. Meng et al. developed a statistical mechanics model which takes these fiber states into account [24].

With these spectroscopy experiments it became possible to assess the higher order structure of chromatin, by statistically analyzing the unfolding. Magnetic tweezer experiments implied the existence of two different fiber structures. Unfolding of 197 NRL nucleosome-arrays pointed to a solenoid fiber. For arrays with a 167 NRL, the results indicated a zig-zag fiber, where two stacks are formed [24, 25].

More recently, experiments with linker-lengths that increased in steps of 5 bps indicated that the 10 bps periodicity of DNA affects the folding of 30 nm fibers [26]. These mechanics suggested that nucleosome stacking

may also depend on the sequence of the linker DNA. For a better understanding of the higher order structure of 30 nm fibers, we thus need to include the mechanics of the linker DNA on the base pair level. And these mechanics are in turn dependent upon the sequence of the DNA. De Jong et al. developed a rigid basepair chromatin model to perform Monte Carlo simulations that mimicked the force spectroscopy experiments. Results were to a large degree consistent with experimental data [27]. Here we use this model to assess the impact of the linker DNA on fiber configurations at thermodynamic equilibrium. In addition we extended this model to allow for sequence-dependent simulations. With Mutation Monte Carlo [13] we assessed the dependence of sequence in the linker DNA upon stacking.
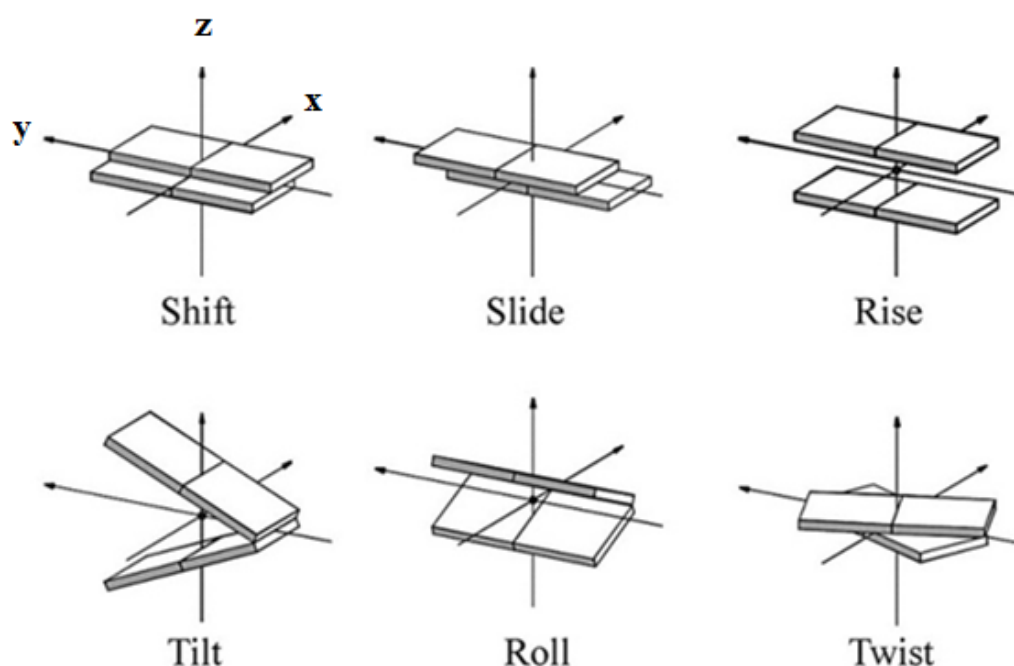
**Figure 1.2:** *The 6 degrees of freedom determine the step between two subsequent bps in the Rigid Base Pair Model. Figure from the Molecular Modelling and Bioinformatics Group [28].*

## 1.4   Rigid Basepair Monte Carlo Simulations

In our problem we want to know the structure of a very complex system in thermodynamic equilibrium. The system consists of many different atoms, each with their own potentials that are possibly correlated. To relate theoretical calculations to experimentally accessible size and time scales, we need to simplify the problem with a model that still incorporates the important physics. Since we are mainly interested in the mechanics on the base pair level we use a model that precisely fits to this scale, the Rigid Base Pair model (RBP) [28]. In the RBP model, base pairs are simplified as rigid bodies. Their dynamics are broken down to movements for all 6 degrees of freedom as showed in figure 1.2.

Mean step parameters and the distributions of these 6 steps have been determined experimentally. In our approach we made use of previous work that curated these distributions from crystallography experiments [27, 29]. The mean step parameters describe the ground state, the shape of the DNA-helix. The potentials of each degree of freedom are assumed to

be harmonic.

With the RBP model the problem is now simplified for *n* bps to a *6n*-dimensional integral to compute the lowest energy state. In MC-simulations, however, we use a reverse approach. Instead of calculating the energies, we start by sampling from given distributions. A random number generator is used for randomly sampling from these distributions. Eventually, iteration of the MC algorithm multiple times over all bps will lead to the equilibrium state of the DNA. Next to the internal energy, which is accounted for by evaluating predefined distributions of step sizes, the DNA may be exposed to external constraints. In that case, each constraint adds an extra energy term, which will be accounted for. Then each MC-step is accepted or rejected based on a certain criterion. The most commonly used criterion is the Metropolis-Hastings scheme:

$$P(accept) = \begin{cases} & \text{if } \Delta E \leq 0 \\ e^{-\frac{\Delta E}{k_B T}} & \text{if } \Delta E > 0 \end{cases} \qquad (1.1)$$

The energy-differences are defined by comparing the energy of the new MC-step with the previous one. Which constraints exist is determined by protein-DNA interactions, the environment and the force. We defined nucleosomal arrays by including stacking and wrapping as physical constraints. Upon stacking and upon wrapping into chromatin fibers DNA can reduce its energy. However, in these states the linker DNA needs to be curved, and step parameters will deviate away from the mean step parameters. The MC-simulations balance these opposing forces and yield the most likely configuration for the defined physical constraints. Such a chain of MC-steps is called a Markov Chain [30].

In our set-up each MC-step in the Markov Chain consists of *n* iterations, where *n* is the number of base pairs. The MC-simulations start with a specific configuration that may not be close to the equilibrium state. The impact of the initial configuration will however smooth out over time [30]. In the simulations we use such a relaxation period and discard the first 1000 iterations. After this period we saved the current step parameters to disk. With enough independent data-samples, we assumed the structure of the fiber to be in thermodynamic equilibrium.

# Chapter 2

# Methods

## 2.1   Monte Carlo Simulations of Chromatin-Fibers

We used the ChromatinMC code [27] for MC-simulations of chromatin-fibers. The package is an extension of the HelixMC-package [29], which was developed for MC-simulations of DNA and RNA. In this package they used the RBP model. The distributions of the 6 step parameters were then drawn from crystallographic data of B-DNA, which is the most prevalent DNA in eukaryotes. We used the curated file **\data\DNA_gau.npy**. This file was constructed by averaging over all the distributions of the step parameters and fitting these with a multivariate Gaussian. The resulting mean step parameters are shown in figure 2.1. The 6x6 covariance matrix of these step parameters was later on used to compute the rigidity of base pairs.

We sampled step parameters from the multivariate Gaussian distribution using **helixmc.random_step.RandomStepSimple()**. In ChromatinMC two extra step parameters were added to include the physics of nucleosomes, unwrapping and unstacking. They appear as additional con-

| | shift (Å) | slide (Å) | rise (Å) | tilt (deg) | roll (deg) | twist (deg) |
|---|---|---|---|---|---|---|
| mean | 0.00 | -0.32 | 3.30 | -0.05 | 1.60 | 35.21 |

**Figure 2.1:** *Mean step parameters for the bps in the Monte Carlo Simulations.*

straits for the Metropolis criterion (1.1, 2.3). Based on crystallographic data of nucleosomes, which denoted the strongest attachment points of nucleosomal DNA, 14 fixed bp-coordinates were derived. The dyad base pair was defined as the mean of these bps. For each fixed bp-position step parameters were defined relative to the dyad position, derived from the crystallographic data. The standard deviation for the translational step parameters was then defined as 1 Å, while a standard deviation of 5 degrees was used for the rotational step parameters. This flexibility was not based on direct experimental evidence, but resulted in unwrapping dynamics which resembled experimental results.

The defined wrapping step parameters forced the DNA into conformations that deviated strongly from those in B-DNA. As such, unwrapping was possible but re-wrapping was very unlikely. To alleviate this, for the first bp after a fixed bp, the steps were drawn from crystallographic data of the nucleosome, instead of from the random pool. As such, nucleosomal DNA could also re-wrap. This however, yielded asymmetric unwrapping which was resolved by alternating the iteration-direction in the fiber for each MC-step. For each of the fixed bp in the nucleosome the energy-penalty for unwrapping was set to $E_{unwrap,max} = 2.5\ k_B T$. In ChromatinMC these energies are calculated using

$$E_{unwrap} = \sum_j min(\sum_i 0.5 \frac{k_B T}{\sigma_{x_i}^2}(x_i - \bar{x}_m)^2,\ E_{unwrap,max}) \qquad (2.1)$$

The first summation is over the fixed bps, while the second one is over the 6 step parameters. $\sigma_{x_i}$, $x_i$ and $\bar{x}_m$ are respectively the standard deviation, the step and the mean step. The wrapping energy was clipped to the $E_{unwrap,max}$.

Interactions between nucleosomes were included in a similar way as the DNA base pair stacking. A reference frame of the nucleosome was defined by setting the x-axis to point through the dyad, and the z-axis to point along the symmetry axis of the nucleosome. Perfect stacking was accomplished when the nucleosome frames arranged into a perfect superhelix [27]. Average step parameters were based on experimental values of the dimensions of chromatin-fibers and for the standard deviation the same values were used as for the wrapping. For the 1-start fiber the parameters were defined relative to the neighbour and for the 2-start fiber the parameters were defined relative to the next neighbour. In that way we made sure that fibers could not alter between start configurations. In

the simulations we used fibers that consisted of an array of 8 nucleosomes. The total length was 1800 bps. Apart from the characteristic NRL, each fiber was characterised by a start-configuration. The initial conformation was defined according these two numbers. All nucleosomes started in a fully wrapped state.

The energy-penalty for unstacking was set to $E_{unstack,max} = 25\,k_BT$ per nucleosomepair. The energies for the stacking were

$$E_{unstack} = min\left(\sum_i 0.5\frac{k_BT}{\sigma_{x_i}^2}(x_i - \bar{x}_m)^2,\, E_{unstack,max}\right) \qquad (2.2)$$

The summation is over the 6 step parameters. $\sigma_{x_i}$, $x_i$ and $\bar{x}_m$ represent the standard deviation, the actual step parameter and the mean step parameter. The stacking energy was clipped to the $E_{unstack,max}$. In that way the simulations are less sensible for large fluctuations. The penalties are the default-values in ChromatinMC, and they are based on experimental data [27].

Two additional energy-penalties were used to constrain the fiber configurations. A penalty of $E_{res} = 10^6\,k_BT$ was if the z-coordinate was smaller than zero or if the z-coordinate was larger than the last base pair. As such, we averted the DNA from penetrating a surface to which the DNA was tethered, be it the flow cell or a bead. This resembles our experimental set-up. For overlapping nucleosome pairs the same energy-penalty was used. Nucleosomes overlapped if the distance between their center of masses was smaller than 55 Å. In the simulations we did not apply a force on the fibers. Each MC-step iterated over all the base pairs of the fiber. And for each iteration the step parameter samples were evaluated with a Metropolis-scheme (1.1), in which the difference in energy was computed between two subsequent iterations (2.3).

$$\Delta E = \Delta E_{unwrap} + \Delta E_{unstack} + \Delta E_{surface/top} + \Delta E_{nuc\ overlap} \qquad (2.3)$$

All fibers were set-up with a left handed chirality and the thermal energy was set to $k_BT = 41.0pN \cdot \mathring{A}$ corresponding to room temperature. The step parameters were sampled using the default random number generator of HelixMC. We used a relaxation period of 1000 MC-steps. Then, fiber poses were saved to disk every 200 MC-steps to ensure samples were independent [27]. A total of 10.000 MC-steps were used for sampling, yielding 49 independent samples.

### 2.1.1   Clustercomputations

A typical simulation of 11.000 MC-steps for an 1800 bps chromatin fiber
takes approximately 7 hours on a Intel(R) Core(TM) i5-7500 CPU @ 3.40
GHz. We wanted to carry out these simulations for 3 different start-configurations
for 51 different NRL-numbers. We made this feasible by using the clus-
ter Maris of the Lorentz Institute [31]. The Maris cluster is divided into
several nodes. The largest nodes provide access to 96 cores and 512 GB
RAM. Access needs to be requested for each individual. After we had
been granted access, we connected via the Lorentz network or via an ssh
tunnel. All nodes are configured in a Linux environment, and the sys-
tem runs on Slurm as a workload manager for the headnode. 6 specific
nodes are reserved for Jupyterhub users. We used one of these nodes,
to be able to work in an interactive environment although the number of
cores was limited to 48. For distributing the workload over the different
cores, we used Dask.distributed. A manual for running simulations on the
Jupyterhub-cluster is included in the appendix.

## 2.2   Mutation Monte Carlo Simulations

### 2.2.1   Sequence Dependent Stepparameters

We extended the ChromatinMC-code such that it allowed for sequence-
dependent sampling of the step parameters. Instead of the average dis-
tributions, we now used different distributions for each dinucleotide. We
used the \**data**\**DNA_default.npz** file which contains 16 distribution files.
In HelixMC it is also possible to sample immediately from a given parame-
ter distribution setting **gaussian_sampling=False**. However, we assumed
normal distributions and fitted the data to a multivariate Gaussian. We
sampled step parameters from the multivariate Gaussian distribution us-
ing **helixmc.random_step.RandomStepAgg()**. Now we were able to do
simulations for a predefined sequence.

### 2.2.2   Sequence Mutation

The pool of possible sequences is enormous, since the number increases
exponentially with the number of used bps ($4^{n_{bp}}$). Eslami-Mossallam et
al., developed a method studying sequence in MC-simulations [13]. In
this method the sequence is treated as an additional degree of freedom.
As such, we can sample the sequence along with the sampling of the step

parameters described above. The initial sequence can be specified. In each iteration we used the random number generator from the python module 'random', to sample a new nucleotide. Together with the previous nucleotide, this sample defined the step parameter distribution. Then we picked a new set of step parameters and applied the Metropolis criterion showed in equation 1.1. Based on this outcome, the mutation was accepted or rejected. Each MMC-step iterated over all bps.

### 2.2.3 Set-Up

The sequence of the nucleosomes was fixed to the 601-sequence, which is the sequence also used in force spectroscopy experiments [25] and we started for the linker DNA nucleotides with 'A'. For the relaxation period we used 20.000 MMC-steps, after which we performed another 10.000 MMC-steps for sequence analysis. We saved the fiber poses for every 200 MMC-steps. We followed the acceptance of the mutations as a function of position over the whole Markov Chain.

## 2.3 Analysis

### 2.3.1 Energy Calculations

The product of our simulations is a set of structures that represent thermodynamic equilibrium. In the used RBP model the energy between pairs of nearest neighbours is defined as,

$$E_i = \frac{1}{2}(\vec{p} - \vec{p}^0) \cdot K \cdot (\vec{p} - \vec{p}^0),$$
(2.4)

where $\vec{p}$ represents the actual 6 step parameters and $\vec{p}_0$ corresponds to the mean step value. The K-matrix is the mathematical inverse of the 6x6-covariance matrix. The wrapping and stacking energies were also calculated afterwards with formulas 2.1, 2.2. All energies were expressed in units of $k_B T$.

Standard deviations are particularly important for our simulations. They show the thermal fluctuations in the system. We use independent samples, and for alike values of each sample the variance was accumulated:

$$\sigma^2 = \frac{1}{n} \sum_{s=1}^{n} \sigma_s^2$$
(2.5)

where the sum is over the $n$ samples and $\sigma_s^2$ is the variance of the calculated value in the sample $s$.

## 2.3.2  Structural Data

The computed data consisted of the step parameters of all the bps in the conformation. In another file, we saved the dyad-positions and other specifics of the fiber. As such, we had all the structural information. In ChromatinMC there is a python-script for translating the stepparameters to bp-coordinates, and subsequently to convert bp-coordinates to 3D rendered images with the help of POV-Ray. We extended this code for average conformations, where we averaged the step parameters of all samples and translated these to a new average pose. In addition, we represented the energies of the individual bps in a color scale. For standard deviations smaller than $\frac{1}{2}k_B T$, the corresponding bps got the grey default color, because we were mainly interested in the energies of the linker-DNA. As such the high energies in the nucleosome were only displayed for unwrapping parts at the beginning and end of the nucleosome.

# Results

## 3.1 The Effect of NRL on the Fiberstructure

The higher order structure of chromatin remains disputed. Experiments have both shown the existence of the solenoid and the zig-zag structure in vivo [19, 21]. Experiments with single-molecule force spectroscopy suggest that both exist for different NRL [22, 25]. Here we performed zero force MC-simulations of stacking conformations and non-stacking conformations for NRL 155-205. For each, we saved 49 independent samples after a relaxation period of 1000 MC-steps.

### 3.1.1 Structural Differences for Changes in NRL

The used fibers in the simulations consisted of 1800 bps DNA substrate, visualized as grey spheres, wrapped around 8 nucleosomes, visualized as red spheres, as seen in figure 3.1. In this figure we also see snapshots of the three configurations with increasing NRL. All the snapshots were taken at the end of the MC-simulations. Linker DNA of non-stacking configurations followed a straight path, that is a helical path, in order to minimize the energy. As such, the orientation of the nucleosomes changed upon addition of 5 bps in the linker DNA. In the 0-start conformation, i.e. when stacking interactions are absent, the orientation of the nucleosomes turned with 180 degrees. Upon stacking, strains emerged in the linker DNA, indicated by the red coloring. These colors were scaled from the total energy, which is the sum over the energies for the 6 different stepparameters. The curving of linker DNA was most striking in the 1-start conformation. In the 2-start conformation the orientation of two opposite nucleosome-pairs changed.
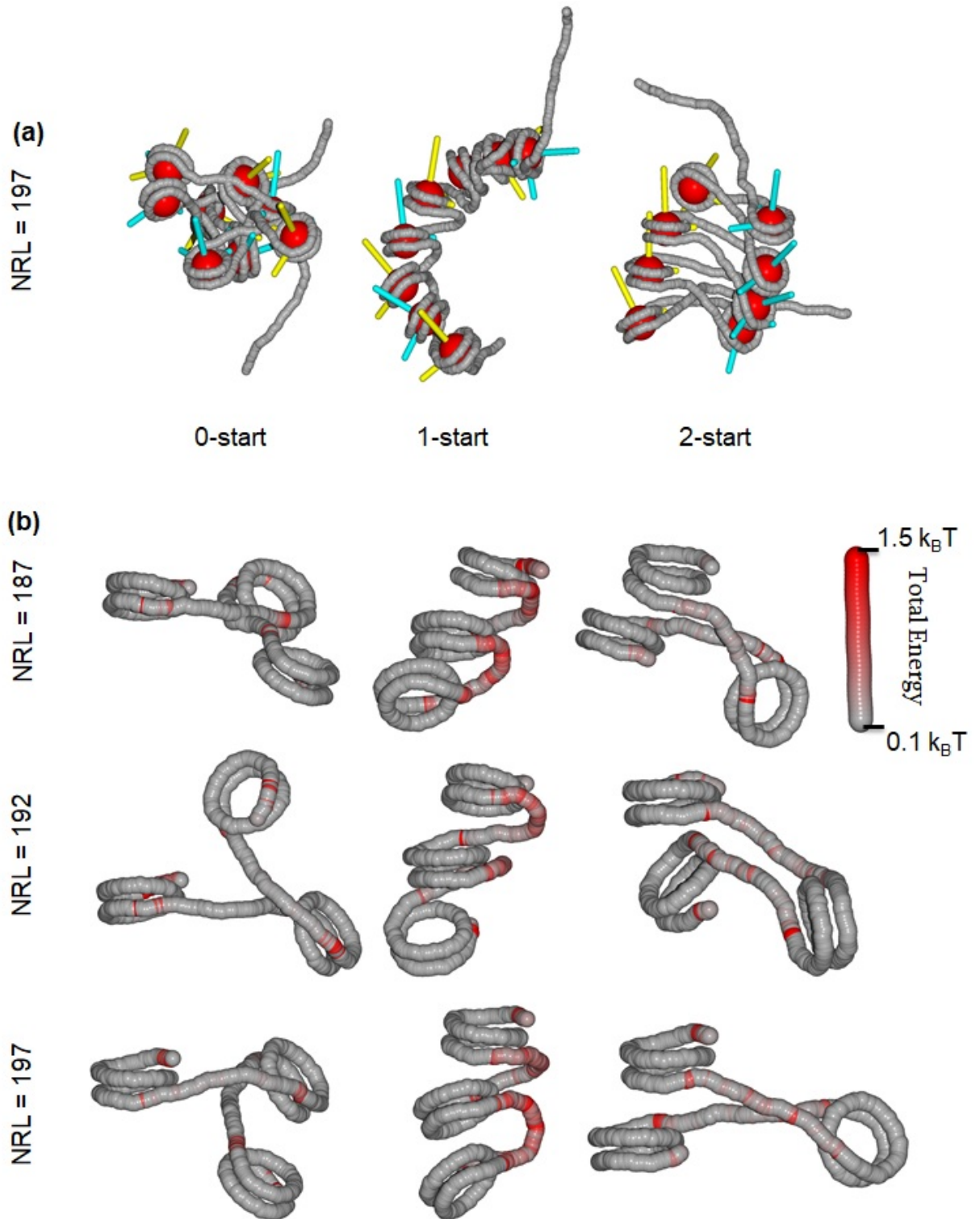
**(a)**

NRL = 197

0-start    1-start    2-start

**(b)**

NRL = 187

NRL = 192

NRL = 197

1.5 k$_B$T

Total Energy

0.1 k$_B$T

---

***Figure 3.1* (previous page)***:** *Increasing linker length affects the higher order structure of nucleosome-arrays. (a) The three different configurations, displayed for NRL=197. The even nucleosomes are displayed with yellow nucleosome reference frames and the uneven nucleosomes are displayed with blue nucleosome reference frames. 0-start fibers formed disordered structures. The averaged stacking conformations show the solenoid and zig-zag structures. (b) Zooming in on nucleosomes 5-7, the linker DNA turned or curved upon addition of 5 linker bps. The total energies for bps in the linker DNA were displayed in color scale. The scale bar is a DNA-rod of 34 bps, corresponding to approximately 10 nm. For all snapshots the position of the basepair 72 bps further from the dyad of the 5th nucleosome was fixed. The orientation of the 6th and 7th nucleosomes in the 0-start conformation turned with 180 degrees around the axis of the linker DNA upon addition of 5 linker bps. In the 1-start conformation the linker DNA is strongly curved. In the 2-start conformation the strains in the linker DNA were lower and increasing linker length caused turns in the orientation of stacking nucleosome-pairs.*

## 3.1.2   Energy Contributions for Increasing NRL

To quantify the effect of the linker DNA on the structure of the fibers, we calculated the energetic contributions of the linker bps. These were defined as the sum of the energies from dyad-to-dyad. In figure 3.2 we show the average energies of all step parameters, and their sum, for the 1-start and 2-start conformations. To see the change in energy upon stacking, the corresponding energies of the 0-start conformations were subtracted. SD's were calculated as described in the methods. Apart from the roll and the stacking parameters, the step parameter energies showed a decrease for increasing NRL number. The decrease went along with an increase in the SD for the 6 step parameters of the bps. The changes in rotational energies were larger than changes in the translational energies. Furthermore, the rotational energies were higher for the 1-start than for the 2-start conformations. This corresponded to the strong curvature of the linker DNA in 1-start conformations, seen in figure 3.1. The twist energy of the 1-start conformations was periodic with a period of approximately 10 bps. Almost the same pattern was observed for the stacking energy of the 2-start conformation. The total energy for the 1-start conformation showed the same pattern as in the twist energy, although some maxima were now shifted with 1 bp to the right, indicating the importance of this step parameter for the 1-start conformation.
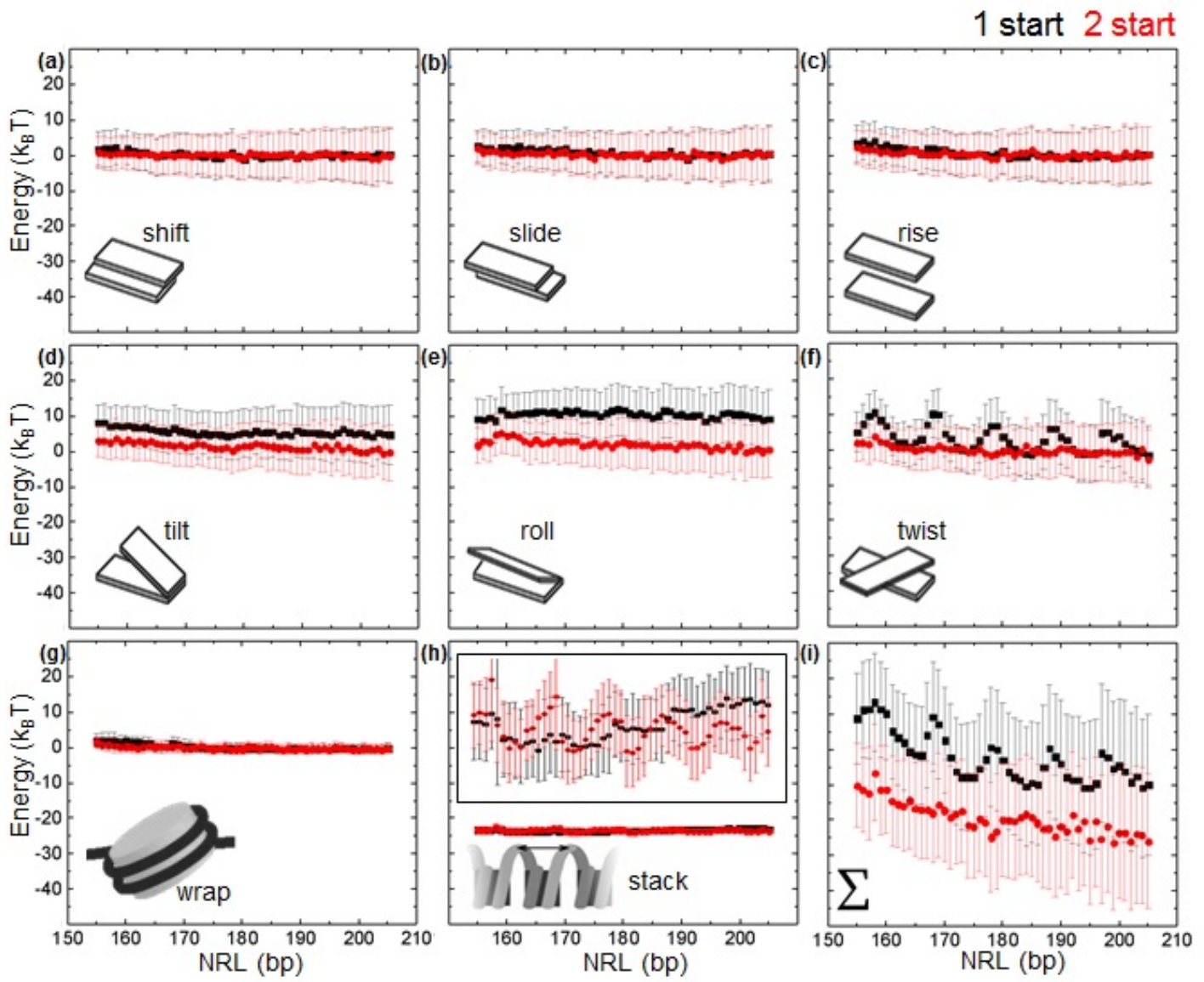
*Figure 3.2* **(previous page)***: The average energies of all 8 step parameters are shown together with the total energy in units of $k_B T$. The 6 bp-energies correspond to the average energy in the linker DNA. Errors for the energies indicate the SD's. A changing linker length predominantly yielded a periodic pattern in the average twist-energy of the linker DNA in the 1-start conformation (f), displayed in red. A similar pattern was seen for the stacking energy in the 2-start conformation (h), displayed in black, although these fluctuations were smaller. Energies for shift (a) showed a slight decrease over increasing NRL. The same applied to results for slide (b), for the rise (c), for the tilt (d) and the total energy (i). Energies of the 1-start conformations were higher than corresponding energies of the 2-start conformations for the tilt (d) and roll (e). The energies corresponding to the rotational step parameters, tilt (d), roll (e) and twist (f) were also higher than the energies for the translational step parameters shift (a), slide (b) and rise (c). For the 1-start conformation there was a periodicity in the twist energy (f). The period was approximately 10 bps. This periodicity returned in the total energy (i). The wrapping energy (g) was somewhat higher for the 1-start conformations at low NRL. The total energy (i) was higher for the 1-start conformations. However, SD's were large and for higher NRL the total energy of the 2-start could as well be higher than the 1-start.*

### 3.1.3 Energy Contributions of Specific Linker Base Pairs

We linked the high rotational energies in the linker DNA to the structures of the 197 NRL fiber to show the positions of these base pairs. In figure 3.3 we zoomed in to just a part of the average 1-start and 2-start conformations. We only displayed the rotational energies, since these were most significant in figure 3.2. The energies were scaled to a color map. For each parameter a different color was used. The scale bars are rods of 34 bps. Wrapping energies were excluded, and thus the colors indicate the specific strains in the linker DNA. In the linker DNA of the 2-start conformation there emerged only a few of such strains. These positions appeared in the middle of the linker DNA and at the beginning and end of the nucleosome. Unlike the 1-start, where high energy bps were grouped, bps with a sufficient high energy in the 2-start conformation appeared at singular positions. The bands of high energy bps in the 1-start were periodically distributed. For the tilt parameter and roll parameter the maximum strains were separated with 5 bps, while for the twist parameter the maxima were separated with 10 bps. Most of these positions showed up in the curved part of the linker DNA, which was the upper part for the used conformation.
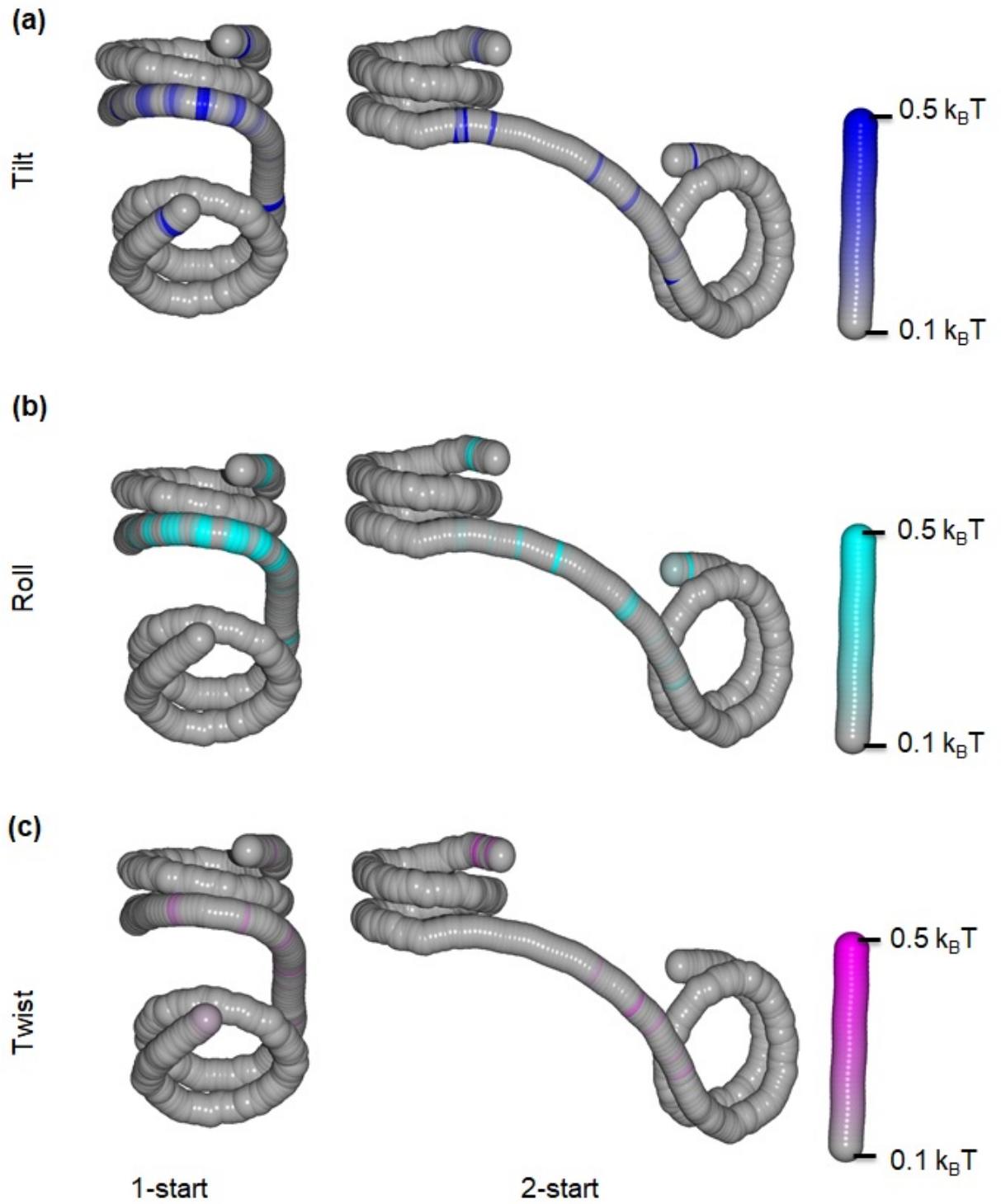
**(a)** Tilt

**(b)** Roll

**(c)** Twist

1-start          2-start

*Figure 3.3* **(previous page)***: High rotational energies of the 1-start conformation showed up in regular intervals in the linker DNA. The 2-start conformation had fewer high-energy bps in the linker DNA. Structures show the average conformations of nucleosomes 5-6 of a 197 NRL fiber. In the 1-start conformation the dyad position of the 5th nucleosome was fixed, while for the 2-start conformation the bp 72 positions further was fixed. High energies appeared at the beginning and end of the nucleosome. This is best illustrated for the tilt parameter (a), indicated in blue. High strains in the linker DNA were clearly shown for tilt and roll (b), in cyan. These strains were equally spaced within the linker DNA for the 1-start conformation. In these snapshots patterns of energetic bps repeated after approximately 5 bps with a lower energy. The region of strains did not cover the whole linker DNA but was located in the proximity of the 5th nucleosome. The twist energy (c), in violet, in the 1-start followed a different pattern over the bps of the linker DNA. There were fewer strains and the position of strains were divided by approximately 10 bps with a lower twist energy. For the 2-start conformation there were less energetic bps than for the 1-start conformation. Apart from the strains at the start of the nucleosome, some strains did appear in the middle of the linker DNA (a, c) and in the proximity of the 6th nucleosome (b, c).*

### 3.1.4   Twist Energy Contributions of Linker Base Pairs

The average twist energy in the 1-start conformation was periodic over the NRL values as shown in figure 3.2. In figure 3.4 we zoomed in to extremes of the average twist-energy in the linker DNA. For these configurations we showed the positions of the bps with the highest energy in the linker DNA, by scaling the energies to a color map. The maximum energy conformations are contrasted with the minimum energy conformations, where almost no coloring appeared in the linker DNA. For the maximum energy conformations, the high-energy bps appeared in bands regularly spaced over the linker DNA. The bands were larger for NRL=178 and NRL=188, than for NRL=198. In the last, the bands were almost periodically spaced with a period of 10 bps.
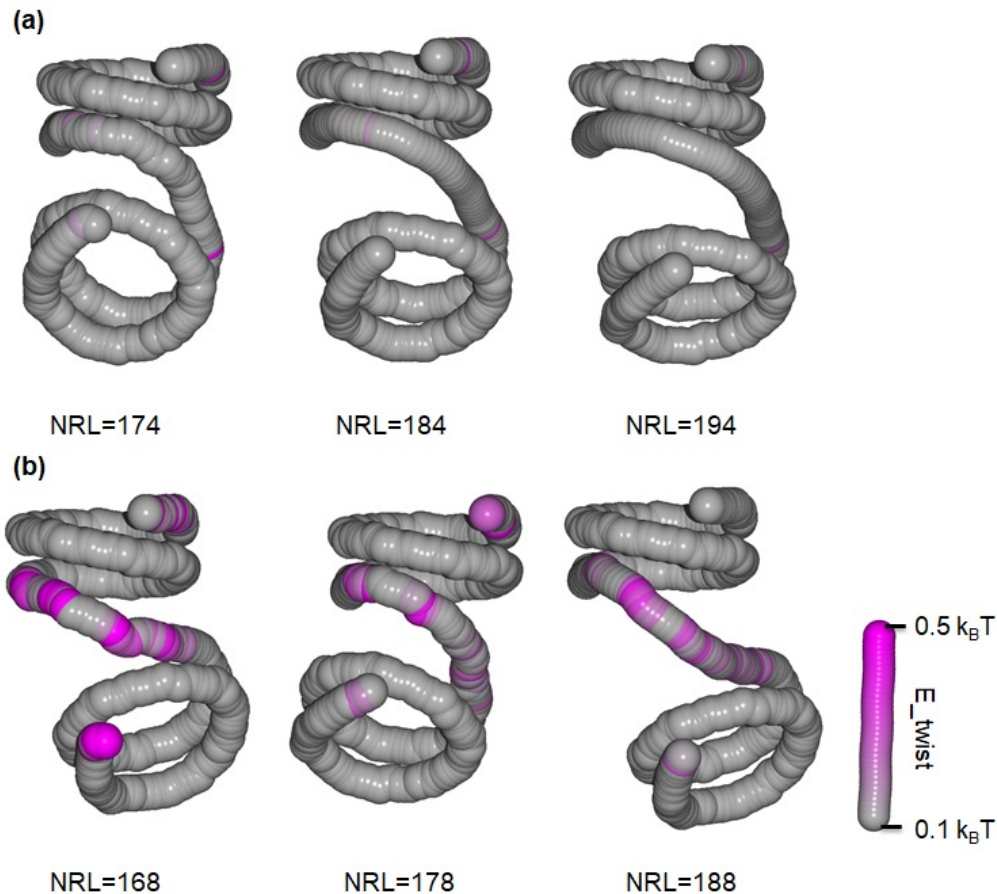
**Figure 3.4:** *Twist-energies in the linker DNA of the 1-start conformation almost vanished for NRL=174, NRL=184 and NRL=194 (a). The conformations correspond to respectively minima (a) and maxima (b) in the average twist-energy of the linker DNA. The snapshots zoomed in to the linker DNA between the 5th and 6th nucleosomes. Twist energies were scaled to a violet color map. High energies appeared at the beginning and end of the nucleosome for both the minimum energy (a) and maximum energy (b) configurations. In the maximum energy configurations, the high energy bps appeared in bands in the linker DNA. The bands were condensed and intensified for NRL=168, while for NRL=188 the bands were broader and more evenly distributed.*

# 3.2 Sequence Dependence of the Linker DNA

Previous experiments and theoretical work have shown the sequence dependent dynamics of nucleosomes [32]. Most of this work was focused on the nucleosomal DNA. We probed the sequence dependence of the linker DNA for the higher order structure of nucleosome-arrays.

## 3.2.1 Sequence-dependent Step parameters

We performed Mutation Monte Carlo to assess potential sequence selectivity of the linker DNA. For that purpose we used sequence-dependent step parameters. In figure 3.5 we compared the rotational step parameter distributions for the dinucleotides 'CC' and 'TT' with the average distributions used in the previous MC-simulations. The distributions were taken from the HelixMC-dataset. Without mutation a mean Gaussian distribution was used. The distributions of 'CC' (a) and 'TT' (b) were compared with the average distribution. Both were plotted as probability density and fitted to a Gaussian function. The step parameters appeared to follow a normal distribution.

Fluctuations for the rotational step parameters 3.5 were much smaller than for the translational step parameters, shown for reference in figure 7.1 in the appendix. Apart from the dissimilarities between different dinucleotides, the 'TT'-stepparameters were more narrowly distributed than the corresponding average parameter distributions. For the tilt parameter the mean was shifted to the right with respect to the average distribution. Sequence, thus affected the distributions of the step parameters.
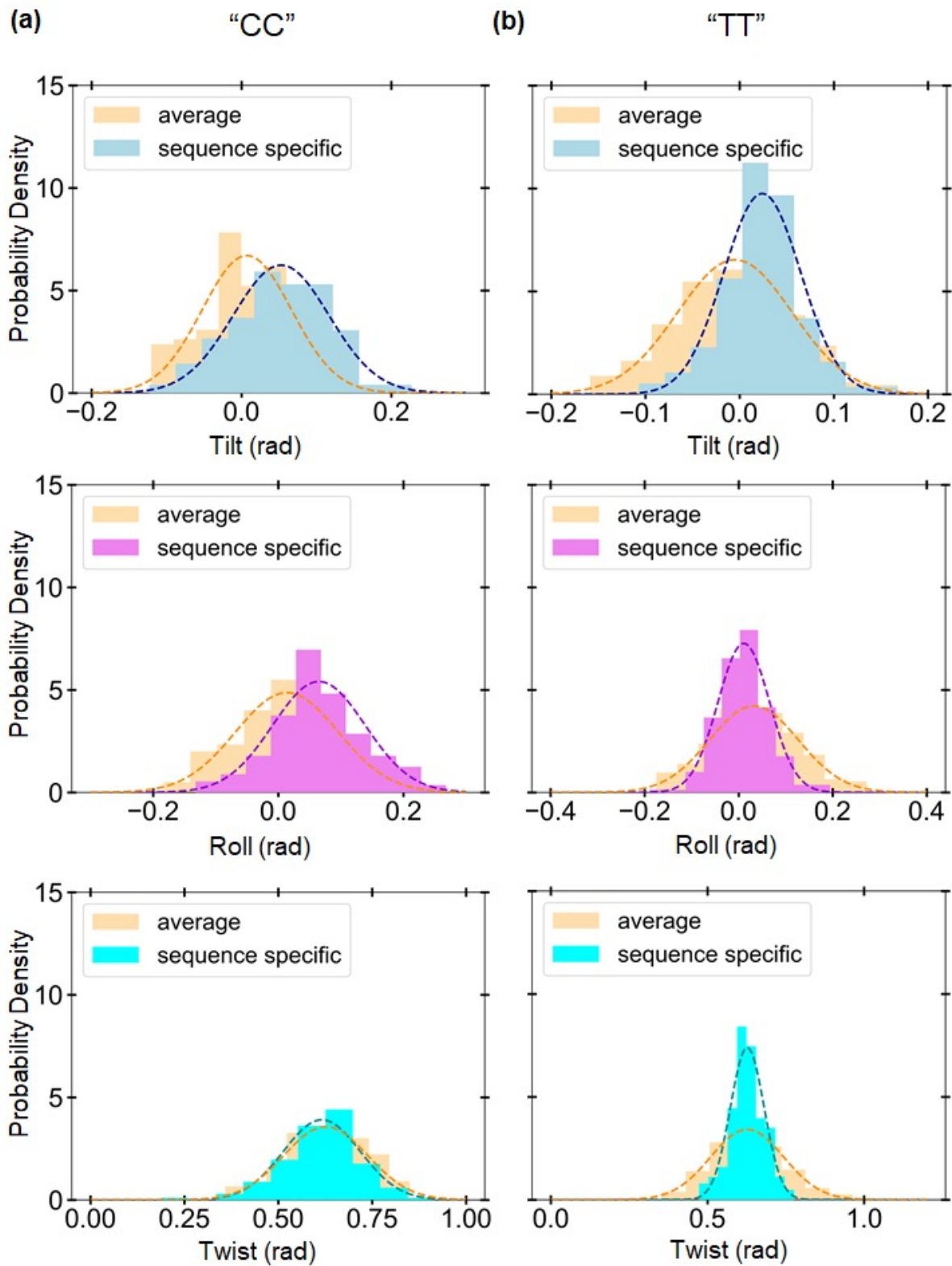
**(a)**                    "CC"                    **(b)**                    "TT"

---

*Figure 3.5* **(previous page)***: The distributions of the sequence specific rotational step parameters. Distributions for the dinucleotide 'TT' (b) were more concentrated around the mean than for the dinucleotide 'CC' (a). For 'CC' the roll- and twist-distributions were similar to the mean distribution, displayed in orange. 'TT' showed less resemblance with the mean distribution.*

## 3.2.2 Sequential Bias

We performed MMC simulations for the 167 NRL and 197 NRL nucleosome arrays, in which we assumed a multi-variate Gaussian distribution of the sequence-dependent step parameters. We followed the acceptance of the nucleotides over time (figure 7.1). After 1000 of such steps the number of acceptations was saved as a function of position in the fiber. This data was represented in a heatmap, where the acceptation rate was now scaled by dividing the number of acceptations by the binsize (1000 steps). For the nucleotides confined by the nucleosomes the 601-sequence was used and no mutations were allowed. As expected, almost all the MMC-steps were accepted in the handles and in the linker DNA of the 0-start conformation. Mutations for the linker nucleotides in the 1-start and 2-start conformations showed a much lower acceptance rate, due to their stacking constraints. In the 1-start conformation, these rates were not symmetric in the linker-DNA, while rates in the 2-start conformation were higher at the two ends of the fiber than in the middle of the fiber.
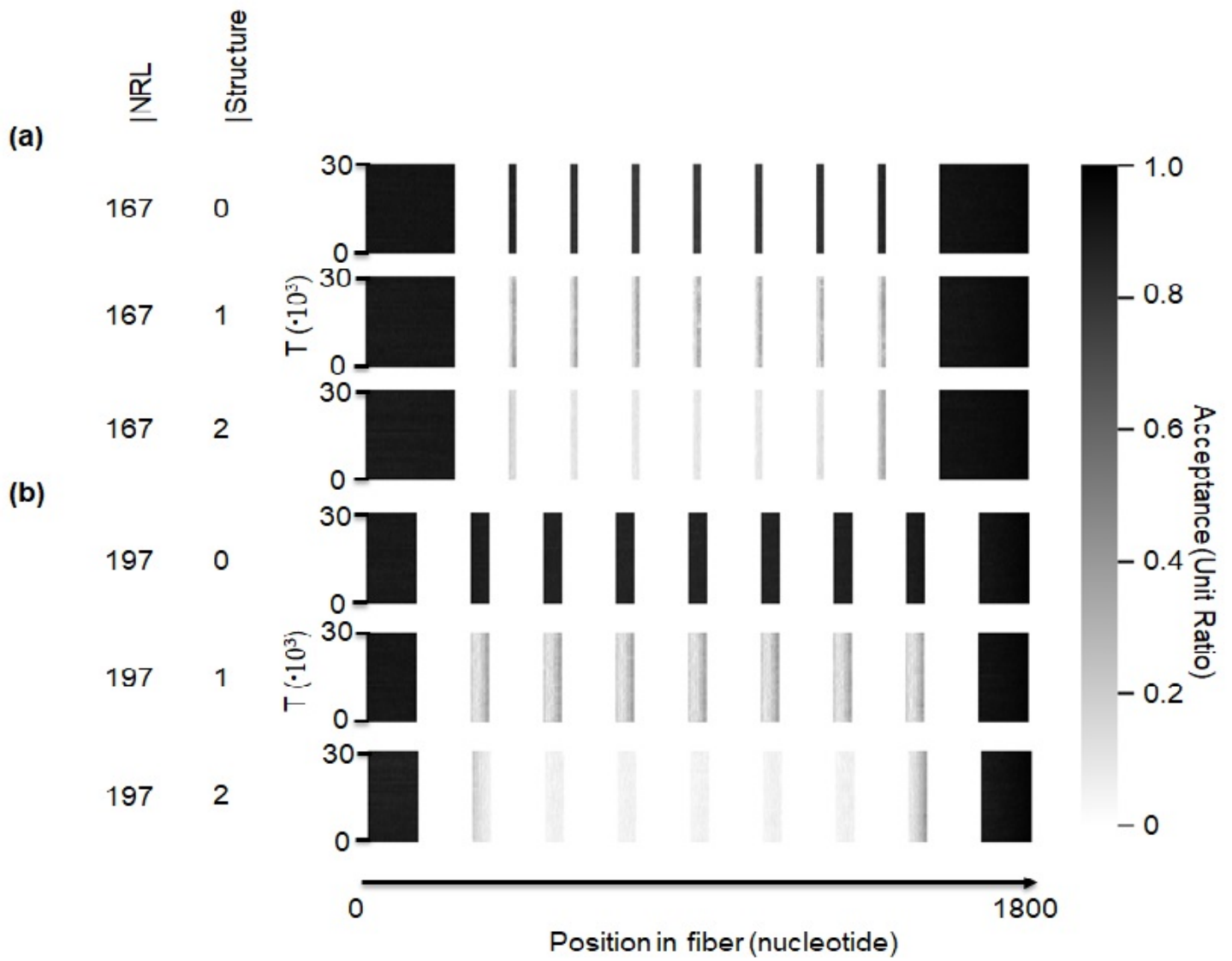
*Figure 3.6* **(previous page):** *Nucleosome stacking predominantly determined the rate of acceptance in the linker DNA. The acceptation rate did not change in MMC-steps. The DNA-handles of the 167 NRL fibers (a) were approximately 100 bps longer than those for the 197 NRL fibers (b). Almost all mutation-steps were accepted in the DNA-handles ($p \approx 0.989$). No mutations in the nucleosomes were accepted. At the positions of the linker DNA the acceptation rate was different for the different conformations. For the 0-start conformations of the 167 NRL fiber (a) and the 197 NRL fiber (b), most of the mutation-steps were accepted at these positions. For the 1-start conformations (c, d) the acceptance was around 0.3 (Unit Ratio). The acceptance rate appeared not symmetric. For both the 1-start (c) and 2-start (d) conformations, a darker band appeared at the right linker DNA positions. A different pattern existed in the 2-start conformations (e, f). In both the 1-start (e) and 2-start (f) conformations, the acceptation rate was higher for the two outer linker lengths, than the middle linker segments. In addition, the rate was higher for the last linker segments than for the first linker segments.*
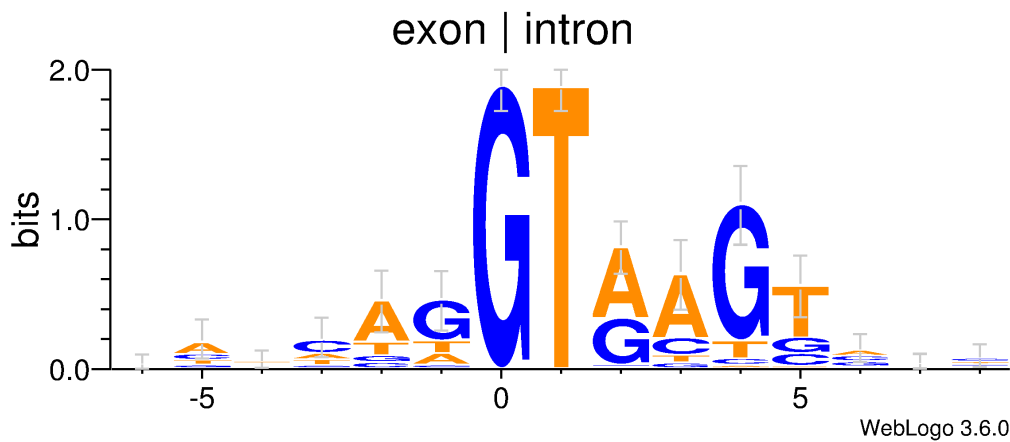
**Figure 3.7:** *An example of a sequence logo. The logo is constructed from sequence-alignments of intron-exon boundaries in human DNA.*

### 3.2.3   Sequence of the Linker DNA

A sequence logo gave further insight in which nucleotides were preferred at specific positions in the linker DNA of the 1-start and 2-start (figure 3.7). The sequence logo was constructed by aligning the last 10.000 MMC-steps. The internet application Weblogo 3 was used for the plotting [33]. Figure 3.7 shows an example of such a plot, downloaded from the website. The data represents 99 different sequences of intron-exon boundaries [34]. The total height of the stacked letters illustrate the Shannon entropy in bits, to which we refer as the conservation. A maximum conservation of 2 bits means the nucleotides do not change at that position. The relative sizes of the letters indicate their frequency in the alignment.

In our figure, figure 3.8, we choose to display the nucleotides of the linker between the 4th and 5th nucleosome. If there was no preference the conservation was 0. For positions where the nucleotides did mutate, these letters were stacked on top of each other. Letters were stacked in order of frequency, the highest on top. All conservation were lower than 0.5 bits. There was a clear preference for certain nucleotides at some positions though. For the 167 NRL assembly these positions showed up in the first half. In the 197 NRL assembly, one position for both the 1-start and 2-start showed a conservation of approximately 0.4 bits. The data indicated that for both the 1-start and 2-start conformation there are sequence preferences.
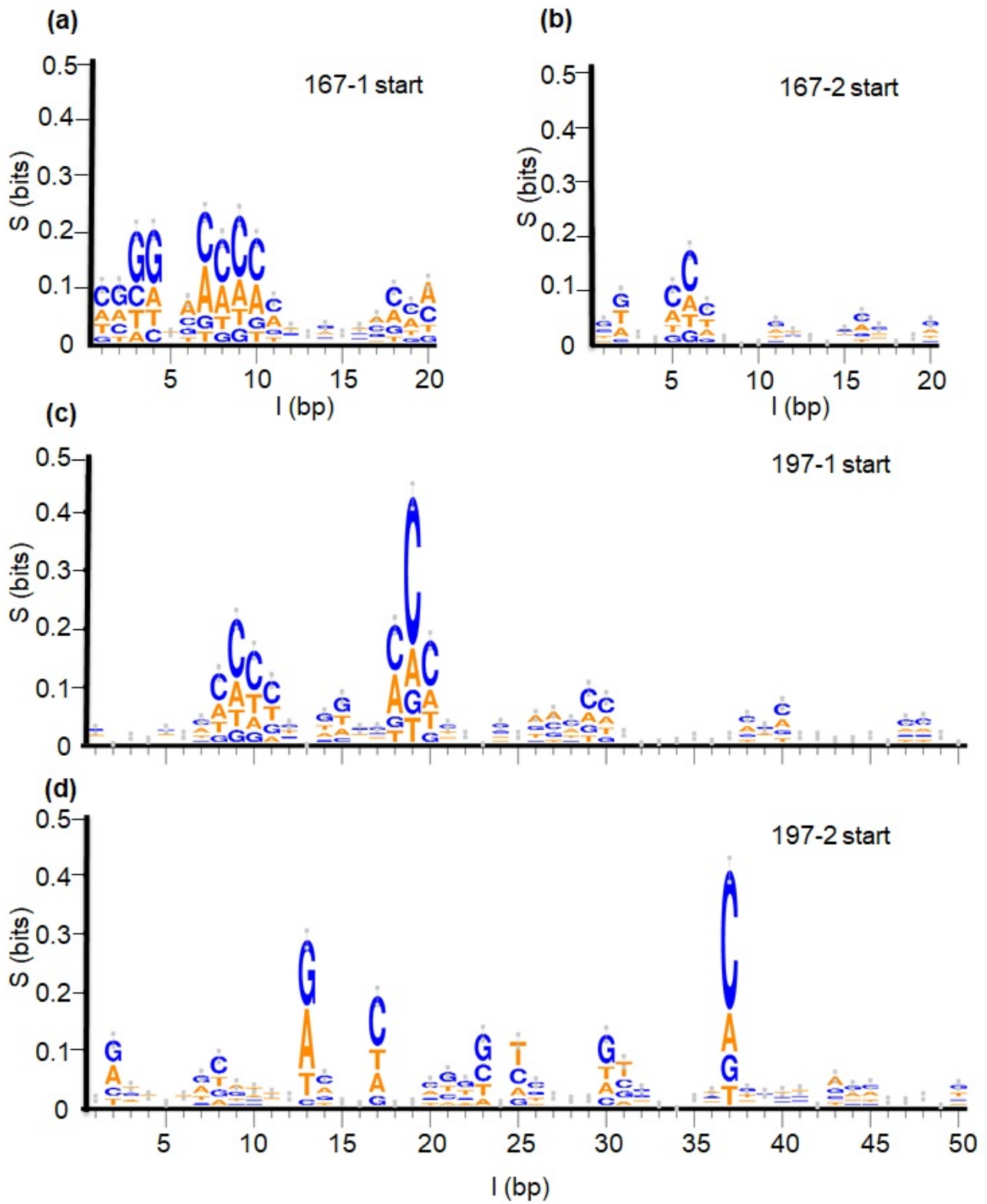
**(a)**



**(b)**



**(c)**



**(d)**

**Figure 3.8** (previous page)**:** *Linker DNA for stacked fiber configurations showed sequence preference at specific positions. A sequence logo of the linker nucleotide positions between the 4th and the 5th nucleosome of the 1-start and 2-start for the 167 NRL fiber and 197 NRL fiber. Errors indicate an approximate Bayesian 95% confidence interval. All of the errors were lower than 0.05 bit. For the 1-start conformation of the 167 NRL fiber (a) the conservation was higher at the first half of the linker nucleotides, than at the second half. At positions 6-10 there was a preference for nucleotides 'C' and 'A' during the MMC-simulations. At positions 3-4 there was a preference for nucleotide 'G'. Position 6 showed a preference for the nucleotide 'C'. For the 2-start conformation (b) of this fiber the difference between the first and second half of the linker nucleotides was even more striking. Overall, the conservations of the 2-start conformations were lower than those for the 1-start conformations. The peak-positions in the 197 NRL fibers (c, d) were around two times higher then the peak-positions in the 167 NRL fibers. For the 1-start (c) of the 197 NRL fiber the peak-position of around 0.4 bits appeared at position 19. Also at the positions 9-10, 18, 20 there was a higher frequency for the nucleotide 'C'. For positions 9 and 18 also the frequency of the nucleotide 'A' was higher, while for position 10 the frequency of nucleotide 'T' was also higher than the respective other two. In the 2-start conformation (d) of the 197 NRL fiber the highest conservation appeared at positions 13 ($\approx$0.28), 17 ($\approx$0.2) and 37 ($\approx$0.4 bit). For position 13 the highest conservation were for nucleotides 'G' and 'A', while for position 17 and 37 the highest conservation was for nucleotide 'C'.*

# Chapter 4

# Discussion

Monte Carlo simulations of non-stacking and stacking fibers showed a change in structure for increasing NRL. In the non-stacking fibers the orientation of nucleosomes followed the periodicity of the added linker DNA. Upon stacking, the configurations found an optimum between stacking nucleosomes and curving of linker DNA. Strains in the linker DNA occurred predominantly for the rotational degrees of freedom at specific positions with a regular spacing of 5 bps. The total average energy in the linker DNA of the 1-start conformation showed a periodicity of approximately 10 bps for increasing NRL. This pattern was predominantly induced by the same periodicity in the 1-start. For the 2-start conformation a similar pattern was seen in the wrapping energy. The total average energy of the 1-start conformation was higher than the 2-start conformation for all NRL, although standard deviations of the total-energies were larger than this difference. Mutation Monte Carlo simulations showed sequence preference in the linker DNA for NRL=167 and NRL=197. At specific positions in the linker-DNA some nucleotides were preferred above others.

The B-DNA helix has a periodicity of 10.4 bps. For the twist parameter this means a twist of 36 (deg) for each added bp. Our results show the significance of this feature for the higher order structure of nucleosome-arrays. However, the mechanical properties of bps are sequence-dependent. Therefore, we also performed MMC, which showed a sequence preference upon stacking. Clearly the 2-start was more sensitive for MMC than the 1-start. And the sequence-dependence was larger for a higher NRL.

The higher-order structure of 30 nm fibers remains an open topic in the discussion about chromatin. In vivo experiments indicated both the existence of solenoid as well as zig-zag fibers [19, 21]. Data from single-

molecule force spectroscopy suggested the existence of the 1-start and 2-start configurations for different NRL [25]. In magnetic tweezer experiments the linker length seemed to be an important feature for nucleosome-stacking [26]. Our results imply that the linker DNA is important for the regulation of gene-expression and in other genomic processes. To understand the effect of the linker DNA, we need to better understand the effect of the mechanics of individual bps. The ChromatinMC-code relies heavily on the mechanical properties of bps. The dynamics of base pairs and nucleosomes were based on experimental data, and the code assumes the rigid basepair model. Previous force-extension simulations for NRL=167 and NRL=197 were largely in agreement with experiments [27].

Here we did zero-force simulations for chromatin fibers with increasing NRL. Linker DNA in the 1-start needs to bend much more than for the 2-start. This induces strains in the DNA, which smooth out for increasing NRL. Although we obtained higher average energies for the 1-start, the large standard deviations do not yield a clearly energetically favoured stacking-configuration. We expected large deviations since equipartition imposes thermal fluctuations of $\frac{1}{2}k_BT$ for each degree of freedom, so the total energy scales with the linker length. Thus, especially for the NRL corresponding to minima of the 1-start conformation, also the zig-zag configuration is possible.

The periodicity of DNA reoccurred in the average twist-energy of the linker DNA of the 1-start conformation over increasing NRL. Also the specific locations of rotational strains showed up at regular intervals in the linker DNA. In the 2-start conformation this effect was not visible. The linker DNA in the zig-zag configurations appeared to be less strained than the linker DNA in solenoid configurations. Zig-zag configurations were able to change the orientation of two nucleosome-pairs in opposite direction to each other to relieve part of the strain on the linker DNA. An energetic optimum was thus easier to find, since an individual nucleosome only needed to account for half of the change in angle due to the addition of linker DNA. The effect was NRL-dependent, because we saw a periodic pattern in the stacking energy of the 2-start conformation.

The mechanical properties of DNA are sequence-dependent. The role of sequence was already recognised for nucleosome wrapping and predicted for nucleosome-nucleosome interactions [32]. We showed that the linker DNA sequence is important for both the solenoid an zig-zag fiber in MMC-simulations. Mutation-steps in the linker DNA were accepted in a remarkable asymmetric fashion. Previous studies showed asymmet-

ric nucleosome-breathing due to the asymmetry in the nucleosomal DNA [35, 36]. Also the pattern we measured in the linker DNA could be due to the asymmetry in the used 601-sequence. The highest contrast was seen at the ends of the linker DNA. The asymmetry could also be related to the stacking. In the simulations we assumed independent harmonic potentials for all 6 stacking parameters. However, stacking is primarily ascribed to interactions of NCP-units with H4-histone tails [15, 17, 18] and the problem may be less harmonic than we simulated.

In the linker DNA we saw predominantly sequential bias for the nucleotide 'C' and 'A' in both start configurations. The bias was only for some positions in the linker DNA, dependent upon the length and the start configuration. Probably, nucleosome-arrays reach lower energy states when these sequence preferences are used. As such, it could also be possible that 30 nm fibers in vivo exist in both 1-start and 2-start configurations, where the local configurations depends upon the specific sequence and the linker length. However, more simulations in combination with experimental work would be needed to clarify this.

For a better understanding of the role of sequence in the linker-DNA, we would also need to calculate the energies and compare them with values obtained in our sequence-averaged MC-simulations. We used the fits of the probability distribution for each sequence dependent step parameter to calculate the stiffness matrix $k$ and the average step parameters $p_0$. With these we calculated the energies of specific conformations of a chromatin fiber as in formula 2.4. However bp-energies were unreasonably high, of the order of 10.000 $k_B T$. It seemed that the stiffness matrix was not well defined, while entries of 300 $k_B T$ appeared. We did not have enough time to assess this further.

# Chapter 5

# Conclusion

We conclude that the higher order structure of 30 nm nucleosome-arrays is heavily affected by the mechanical properties of the linker DNA, which in turn are sequence-dependent. Upon stacking, linker DNA strives to preserve the 10.4 bps periodicity of DNA. This underlines the importance of the linkerlength. For the 1-start conformation it is harder to compensate for the additional twist of an extra linker-bp. The 2-start conformation, stacks are more flexible in their orientation. These features are sequence-dependent, and specific nucleotides in the linker DNA may be important indicators for what start-configurations can be reached in thermodynamic equilibrium.

For future work it would be advantageous to include sequence dependent energy-calculations. With our code it is also possible to determine linker DNA sequences for experiments. For instance, MMC force extension simulations could shed light on preferred sequences in the linker DNA. These sequences could then be used in Magnetic Tweezers experiments. As such, our work forms a bridge between theory and experiments.

# Chapter 6

# Acknowledgements

This project is made possible by the help of many people. First and foremost, I want to thank professor John van Noort and Thomas Brouwer as my supervisors. Apart from all that I learned about biophysics and bioinformatics, I also learned a lot about doing research itself. I want to thank them for the interesting discussions, for bringing up new ideas and for their help in the problem solving. I also want to thank the second supervisor professor Helmut Schiessel. Thanks also to Daan van de Velde, Thijs de Buck and Kirsten Martens for their occasional help and for providing a nice work atmosphere. I want to thank the whole group, also Babette de Jong and Redmar Vlieg, for the interesting coffee-table discussions.

# Chapter 7

## Appendix

## 7.1 Cluster Usage

Each user needs to request a username and password to access the cluster. You can request access by sending an e-mail to Carlo Beenakker. See for further documentation the webpage of the Maris cluster [31]. Apart from accessing the Maris cluster via a computer that is connected to the Lorentz Network, it is possible to connect with the Lorentz server via a styx connection. Accessing the cluster via an ssh-tunnel:

- Use FireFox browser

- Install foxyproxy

- Add proxy with 'IP address = localhost' and 'port = portvalue'

- Run bash line: ssh –D portvalue user@ssh.lorentz.leidenuniv.nl

For the use of Jupyterhub and Dask.distributed it is necessary to follow the steps below.
First at the following line to .bashrc for jupyterhub:

- export CONDA_ENVS_PATH=/marisdata/$USER/.conda/envs

With conda it is possible to create a python 2.7 environment. The following commands need to be applied in the terminal:

- export PATH=/marisdata/MARISHUB/bin:$PATH

- conda create –n py27 python=2.7 activate the environment:

- source activate py27

41

- conda install ipyparallel

- conda install dask

- conda install dask-jobqueue

- conda install ipykernel

- conda install matplotlib

- conda install lmfit

- conda install pandas==0.16.2

- conda install distributed

- conda install xlrd

- pip install opencv-python

- pip install openpyxl==1.8.6

- pip install helixmc

Running simulation on the Jupyterhub is then simple:

- Upload the ChromatinMC package

- Adjust the RunMC_parallel file

Downloading multiple files at once is convenient by using the following steps. Go to the directions of the folder. Open a new Notebook from within this folder and run the following command:

- !tar cvfz allfiles.tar.gz *

The tar.gz file contains all the files from within the folder in a concise format.

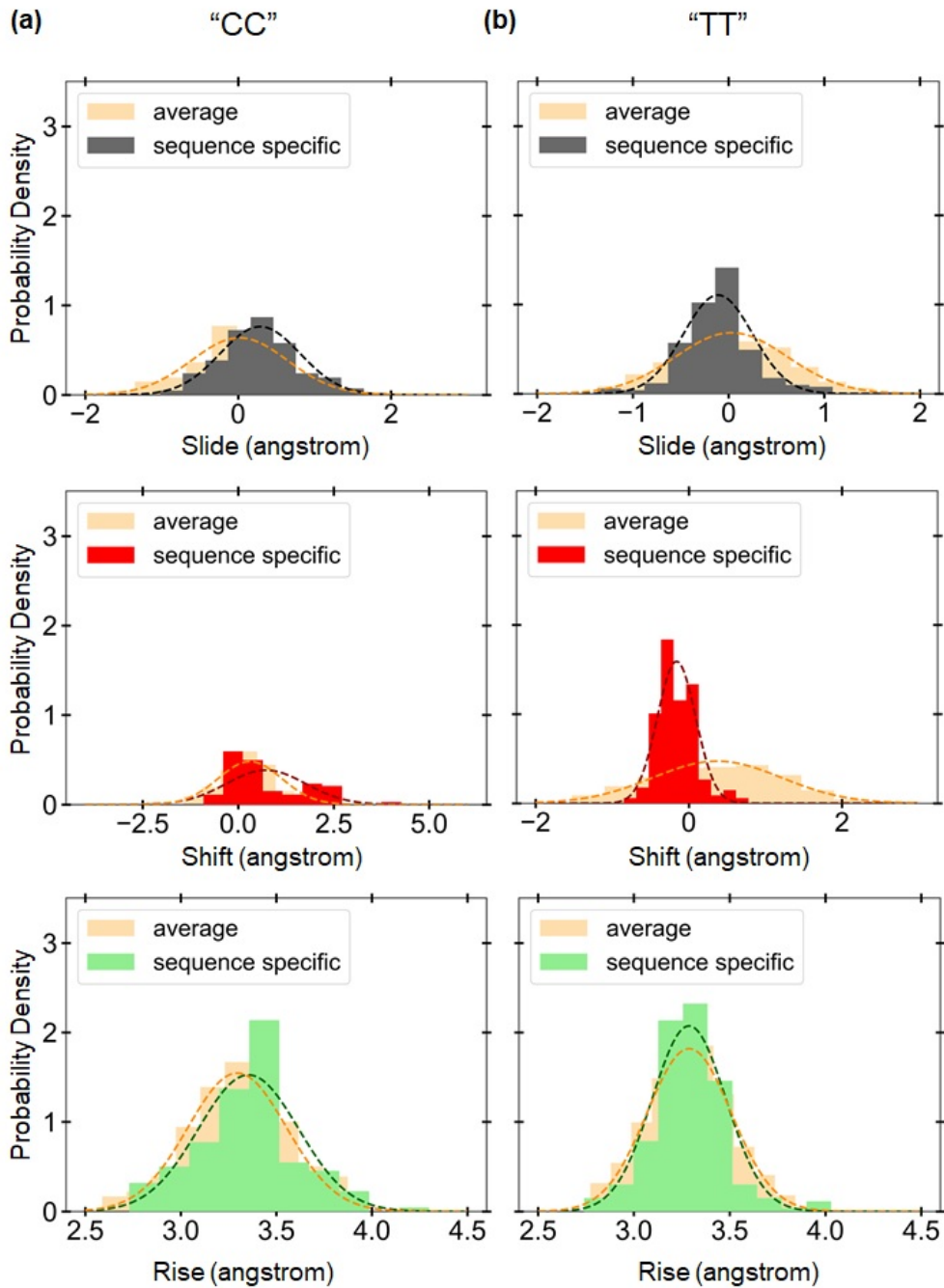# 7.2 Translational Stepparameter Distributions

*Figure 7.1* **(previous page)**: *Sequence dependence of the translational step parameters affected in particular the slide and shift parameters. The curated distributions were taken from the HelixMC-dataset. The distributions for the rotational steps (angstrom) are shown for the dinucleotides 'CC' and 'TT'. Without mutation a mean Gaussian distribution was used. The distributions of 'CC' (a) and 'TT' (b) were compared with the average, by construction same size distributions (in navajowhite) based on the multivariate Gaussian. Both were plotted as probability density, and fitted with a single Gaussian which is dark orange for the average distribution. Step parameters for the dinucleotide 'TT' had smaller fluctuations than those for the dinucleotide 'CC'.*

## 7.3   Python Code

The ChromatinMC-code will be updated with a new version and shared online.

# Bibliography

[1] H. Schiessel, *Biophysics for beginners: a journey through the cell nucleus*, CRC Press, 2013.

[2] P. Nelson, *Biological physics*, WH Freeman New York, 2004.

[3] M. Noll, *Subunit structure of chromatin*, Nature **251**, 249 (1974).

[4] W. et al., *The Structure and Function of Chromatin Creative Diagnostics Blog*, 2017.

[5] R. D. Kornberg, *Structure of chromatin*, Annual review of biochemistry **46**, 931 (1977).

[6] S. A. Grigoryev and C. L. Woodcock, *Chromatin organization—the 30 nm fiber*, Experimental cell research **318**, 1448 (2012).

[7] K. Maeshima, S. Hihara, and M. Eltsov, *Chromatin structure: does the 30-nm fibre exist in vivo?*, Current opinion in cell biology **22**, 291 (2010).

[8] 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).

[9] E. Y. Chua, V. K. Vogirala, O. Inian, A. S. Wong, L. Nordenskiöld, J. M. Plitzko, R. Danev, and S. Sandin, *3.9 Å structure of the nucleosome core particle determined by phase-plate cryo-EM*, Nucleic acids research **44**, 8013 (2016).

[10] R. Ettig, N. Kepper, R. Stehr, G. Wedemann, and K. Rippe, *Dissecting DNA-histone interactions in the nucleosome by molecular dynamics simulations of DNA unwrapping*, Biophysical journal **101**, 1999 (2011).

[11] K. Voltz, J. Trylska, N. Calimet, J. C. Smith, and J. Langowski, *Unwrapping of nucleosomal DNA ends: a multiscale molecular dynamics study*, Biophysical journal **102**, 849 (2012).

[12] L. de Bruin, M. Tompitak, B. Eslami-Mossallam, and H. Schiessel, *Why do nucleosomes unwrap asymmetrically?*, The Journal of Physical Chemistry B **120**, 5855 (2016).

[13] B. Eslami-Mossallam, R. D. Schram, M. Tompitak, J. van Noort, and H. Schiessel, *Multiplexing genetic and nucleosome positioning codes: a computational approach*, PLoS One **11**, e0156905 (2016).

[14] M. Tompitak, G. T. Barkema, and H. Schiessel, *Benchmarking and refining probability-based models for nucleosome-DNA interaction*, BMC bioinformatics **18**, 157 (2017).

[15] Y. Shimamoto, S. Tamura, H. Masumoto, and K. Maeshima, *Nucleosome–nucleosome interactions via histone tails and linker DNA regulate nuclear rigidity*, Molecular biology of the cell **28**, 1580 (2017).

[16] S. Venkatesh and J. L. Workman, *Histone exchange, chromatin structure and the regulation of transcription*, Nature reviews Molecular cell biology **16**, 178 (2015).

[17] N. Korolev, A. P. Lyubartsev, and L. Nordenskiöld, *A systematic analysis of nucleosome core particle and nucleosome-nucleosome stacking structure*, Scientific reports **8**, 1543 (2018).

[18] N. Korolev, Y. Fan, A. P. Lyubartsev, and L. Nordenskiöld, *Modelling chromatin structure and dynamics: status and prospects*, Current opinion in structural biology **22**, 151 (2012).

[19] P. J. Robinson, L. Fairall, V. A. 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).

[20] 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).

[21] 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).

[22] 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).

[23] C. Bouchiat, M. D. 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).

[24] 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).

[25] 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).

[26] T. B. Brouwer, A. Kaczmarczyk, N. Hermans, M. Botto, and J. van Noort, *Linker DNA Length Defines the Structure of Chromatin Fibers*, Biophysical Journal **114**, 256a (2018).

[27] B. E. de Jong, T. B. Brouwer, A. Kaczmarczyk, B. Visscher, and J. van Noort, *Rigid Basepair Monte Carlo Simulations of One-Start and Two-Start Chromatin Fiber Unfolding by Force*, Biophysical journal **115**, 1848 (2018).

[28] W. et al., *MC DNA Help Molecular Modeling and Bioinformatics Group*, 2019.

[29] F.-C. Chou, J. Lipfert, and R. Das, *Blind predictions of DNA and RNA tweezers experiments with force and torque*, PLoS computational biology **10**, e1003756 (2014).

[30] W. R. Gilks, S. Richardson, and D. Spiegelhalter, *Markov chain Monte Carlo in practice*, Chapman and Hall/CRC, 1995.

[31] B. et al, *Computer Documentation Wiki Maris Cluster*, 2019.

[32] B. Eslami-Mossallam, H. Schiessel, and J. van Noort, *Nucleosome dynamics: Sequence matters*, Advances in colloid and interface science **232**, 101 (2016).

[33] G. E. Crooks, G. Hon, J.-M. Chandonia, and S. E. Brenner, *WebLogo: a sequence logo generator*, Genome research **14**, 1188 (2004).

[34] R. M. Stephens and T. D. Schneider, *Features of spliceosome evolution and function inferred from an analysis of the information at human splice sites*, Journal of molecular biology **228**, 1124 (1992).

[35] A. W. Mauney, J. M. Tokuda, L. M. Gloss, O. Gonzalez, and L. Pollack, *Local DNA sequence controls asymmetry of DNA unwrapping from nucleosome core particles*, Biophysical journal **115**, 773 (2018).

[36] H. Schiessel, *Telling left from right in breathing nucleosomes*, Biophysical journal **115**, 749 (2018).