Dissertation

submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

presented by
Dipl. Phys. Philipp M. Diesinger
born in Kleinblittersdorf, Germany
Oral examination: June 17th, 2009

Genome Folding
at the 30nm Scale

Referees: **Prof. Dr. Dieter W. Heermann**
Prof. Dr. Christoph Cremer

Genomorganisation auf der 30nm Skala

In dieser Arbeit wurde eine grundlegende Frage der Genomorganisation beantwortet, indem gezeigt wurde, dass eine Struktur höherer Ordnung nach dem Nukleosom, d.h. Chromatin, tatsächlich existiert. Ein Chromatin-Modell wurde entwickelt, das die Untersuchung sehr langer Strukturen (im Bereich von Mega-Basenpaaren) erlaubt. Des Weiteren wurde zum ersten Mal die Ablösung von Linker Histonen und ganzen Nukleosomen in ein Chromatin-Modell integriert. Dies erlaubt die Untersuchung des Chromatin-Phasendiagramms. Die darin enthaltenen Strukturen werden vor dem Hintergrund von DNA-Kompaktifizierung und -Zugänglichkeit sowie anderer wichtiger Chromatineigenschaften diskutiert. Die Verteilungen der Modellparameter stammen aus experimentellen Daten [186; 187]. Zusammen mit den Ablösungseffekten zeigen sie, dass jede Chromatinkonformation aus einer Verteilung von verschiedenen Strukturen besteht. Dies erklärt, warum es experimentell so schwierig ist, reguläre 30nm Fasern zu finden. Die Ergebnisse zeigen, dass Histonablösungen die Eigenschaften von Chromatin massiv beeinflussen. Nukleosomablösung kann einerseits zu einem Chromatinkollabs, andererseits aber auch zum Anschwellen von Chromatin führen und der vorhergesagte Bereich optimaler DNA-Kompaktifizierung stimmt exakt mit experimentellen Daten [187] überein. Außerdem zeigt das Modell gute Übereinstimmung mit vielen experimentell bestimmten Chromatineigenschaften. Ein Vergleich mit Daten aus 5C-Experimenten [72] belegt, dass Histonablösung eine wichtige Chromatineigenschaft ist, da nur Fasern mit Ablösungenseffekten die wichtigen physikalischen Kontakte auf der kleinen Längenskala aufweisen. Zufällige Chromatinkontakte werden theoretisch untersucht, um 3C-basierte Technologien dadurch zu verbessern, dass präziser zwischen Zufallskontakten und spezifischen Kontakten der DNA unterschieden werden kann.
Große Teile dieser Arbeit wurden bereits veröffentlicht [63; 64], werden zur Zeit geprüft [65; 66] oder für eine Veröffentlichung vorbereitet [26; 27; 67; 68].

Genome Folding at the 30nm Scale

This thesis addressed and succeeded in answering a fundamental question concerning genome folding at the 30nm scale by proving that a higher-order DNA folding pattern beyond the nucleosome, i.e. chromatin, actually exists. A model for chromatin was developed that allows to study very long fibers (in the range of Mega base pairs). Furthermore, linker histone depletion as well as nucleosome depletion have been included for the first time in a chromatin model. This allowed to study the chromatin phase diagram and the corresponding structures are discussed against the background of compaction and DNA accessibility and other important chromatin features. The basic model parameter distributions come from experimental data [186; 187]. Together with the histone depletion effects they show that every chromatin conformation consists of a distribution of different structures. This explains why regular 30nm fibers are so hard to find experimentally. The results show that histone depletion massively affects the properties of chromatin. Nucleosome depletion can either lead to a collapse or to swelling of chromatin and the predicted regime of optimal DNA condensation coincides with experimental data [187] which proves that histone depletion is used as a regulatory tool for DNA extension. Moreover, the model is in good agreement with many experimentally determined chromatin properties. A comparison of the developed model with 5C experimental data [72] proves that histone depletion is an important chromatin feature because only fibers with depletion allow the important physical contacts on a small length scale. Random chromatin collisions are theoretically studied to improve 3C-based experimental technologies by allowing to distinguish more accurately between specific and random DNA contacts.
Major parts of this work have already been published [63; 64] are currently under revision [65; 66] or respectively under preparation [26; 27; 67; 68].

Preface

The results which will be presented in this work have been developed between February 2006 and March 2009, i.e. the time of my PhD studies at the Institute of Theoretical Physics, University of Heidelberg.

A brief outline of this thesis will be given below. A more detailed overview can be found at the beginning of each chapter.

Outline

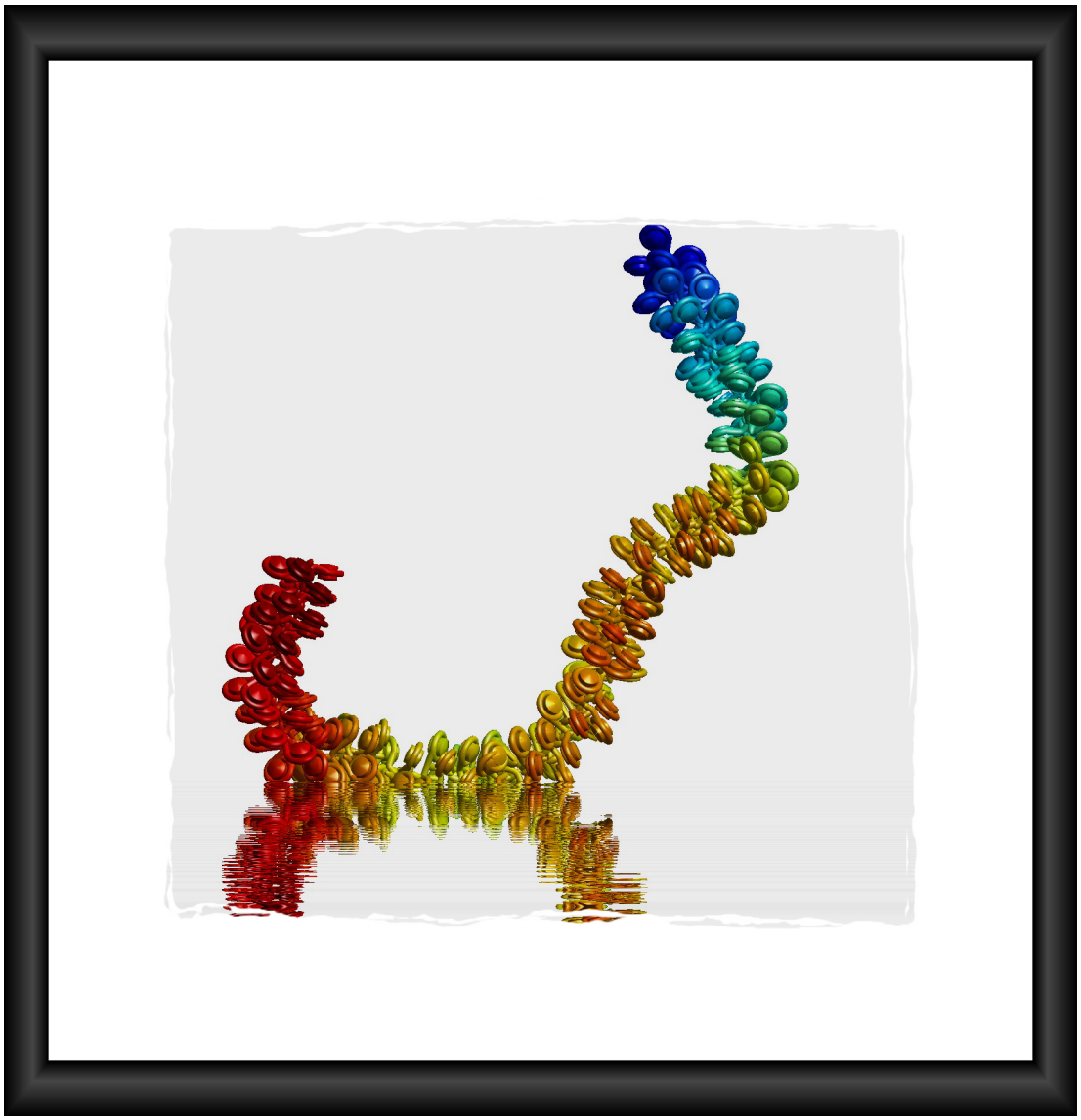
Chapter 1 will provide some basic concepts of the statistical mechanics of polymers and introduce some mathematical tools which will be used frequently in this work. After that, chapter 2 will introduce chromatin and give an overview over the problems and fundamental questions concerning genome folding with focus on the small length scale. In Chapter 3 my part in a collaboration project with the group of Prof. Cremer at the Kirchhoff-Institute for Physics (University of Heidelberg) will be described. The analysis of the localization microscopic data in this section will show that a higher order folding pattern beyond the nucleosome exists.
Chapter 4 will describe the geometric modelling of the 30nm fiber by an improved two-angle model and in the succeeding chapter 5 the chromatin phase diagram will be discussed comprehensively. After that, chapter 6 will again deal with modelling issues. This time the focus will lie on fiber flexibility and the question how histone depletion can be allowed for.
Then the results of the chromatin simulations will be presented in chapters 7 and 8.
After that, chapter 9 will deal with a completely different topic resulting from a collaboration with the group of Dirk Görlich at the MPI in Göttingen namely the nuclear pore complex and the matter whether or not nucleoporins can form hydro-gels in particular.
In the final chapter 10 the connection between polymer loops, the scattering function and the mean Average Crossing Number will be discussed.

Some supplemental figures and tables are only presented in the appendix of this work. They can be found in App. A respectively App. B.

An electronic version of this thesis in pdf format is available in App. I.

Major parts of this thesis have already been published in scientific journals. Furthermore, several results of this work have been presented at scientific conferences and workshops. App. J provides an overview over these publications.

A comprehensive list of all figures as well as a list of the tables is given in App. K respectively in App. L.



"Art is science made clear."
- Wilson Mizner

Contents

1 Basic Polymer Physics

Some basic concepts of the statistical mechanics of polymers will be explained briefly in the following sections. For a more comprehensive approach cf. [114; 207].

1.1 Statistical Mechanics of Long Chain Molecules

Understanding the configuration and dynamics of long polymer chains has been a significant source of problems within statistical physics from the 1950's onwards.
One of the most interesting points is the *universality* of some of the problems. What was considered to be the most fundamental questions turned out to be insensitive to the details of the chemistry of the chains. What is more the governing question turned out to be a diffusion equation since the chain's shape resembles a random diffusive walk in space.

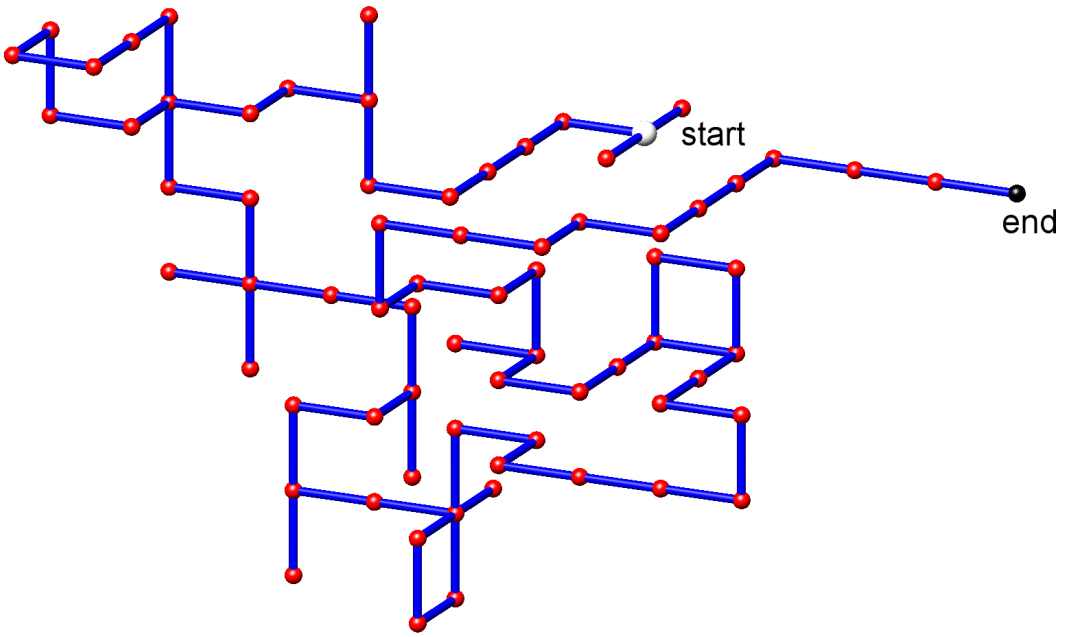


Figure 1: Example of a random walk with 100 steps on a cubic lattice. The lattice constant is 1, the end-to-end distance of the walk is 6.8. The white sphere denotes the start point of the walk and the black one represents the end point. Another example of length 20 can be found in App. A (Fig. 162).

Schrödinger's equation is also a diffusion equation (in imaginary time $t' = it$). Therefore, there are many analogies with quantum mechanics.
Random walks (cf. Fig. 1 and Fig. 2) are an important topic in mathematics, too, particulary when polymer confirmations appear to resemble not just simple random

walks but self-avoiding ones.

1.2 The Freely-Jointed Chain Model

1.2.1 Introduction

The simplest model to describe polymers is the ideal or *freely-jointed chain* (FJC). It considers polymers as random walks and neglects any kind of interactions among monomers. But although it is simple, its generality gives already insights about the physics of polymers.

In this model, monomers are rigid rods of a fixed length b, and their orientation is completely independent of the orientations and positions of neighboring monomers to the extent that two monomers can co-exist at the same place. The contour length of such a polymer is then given by $L = Nb$ were N is the total number of monomers.

In this very simple approach where no interactions between monomers are considered, the energy of the polymer is taken to be independent of its shape which means that at thermodynamic equilibrium all of its shape configurations are equally likely to occur as the polymer fluctuates in time according to the Maxwell-Boltzmann distribution.

While this elementary model is totally unadapted to the description of real-world polymers at the microscopic scale it does show some relevance at the macroscopic scale in the case of a polymer in solution whose monomers form an ideal mix with the solvent (in which case, the interactions between monomer and monomer, solvent molecule and solvent molecule and between monomer and solvent are identical, and the system's energy can be considered constant, validating the hypothesis of the model).

The relevancy of the model is, however, limited, even at the macroscopic scale by the fact that it does not consider any excluded volume for monomers.

Other fluctuating polymer models that consider no interaction between monomers and no excluded volume, like the worm-like chain model (cf. Sec. 1.3), are all asymptotically convergent toward this model at the thermodynamic limit.

For purpose of this analogy a *Kuhn segment* is introduced corresponding to the equivalent monomer length to be considered in the analogous ideal chain. The number of Kuhn segments to be considered in the analogous ideal chain is equal to the total unfolded length of the polymer divided by the length of a Kuhn segment.

1.2.2 Mathematical Description

In the model of the freely-jointed chain the chain is made up of N links, each of length b (cf. Fig. 3). These links can be each thought of as representing one small chemical group of atoms along the polymer which is called a *monomer*.

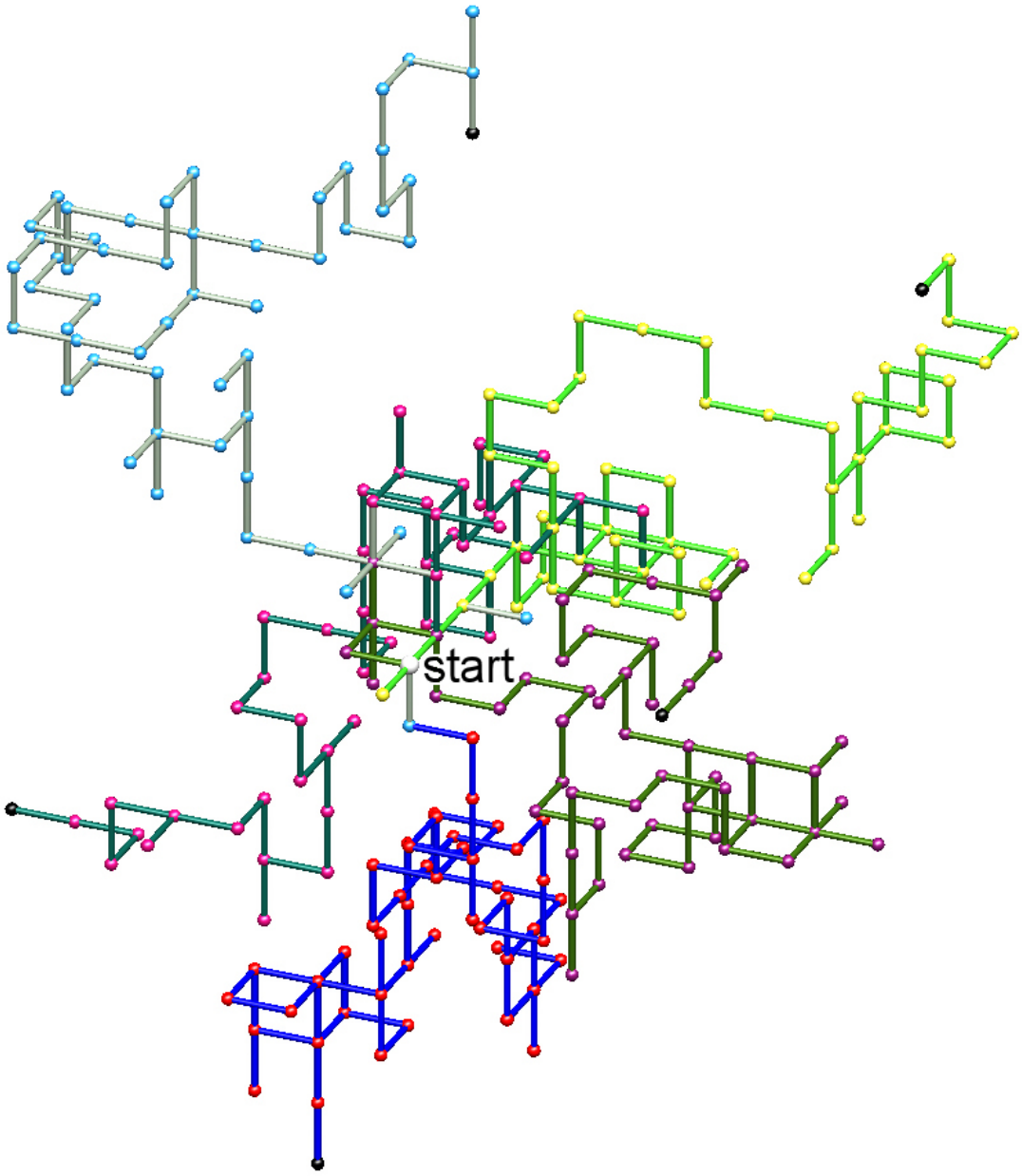


Figure 2: Example of five random walks with 80 steps each on a cubic lattice. The lattice constant is 1. All walks start at the origin (big white sphere). The black spheres denote the end points of the walks.

The joints of the chain are found at positions $R_n \in \mathbb{R}^3$ and are jointed by the link vectors $r_n = R_n - R_{n-1}$. Thus the end-to-end vector (cf. Fig. 5) of the chain is given by

$$R = R_N - R_0 = \sum_{i=1}^{N} r_n.$$

In this model the statistics of the link vectors are determined by

$$\langle r_n \rangle = 0$$

$$\langle r_n \cdot r_m \rangle = b^2 \delta_{nm}$$

Hence the statistics of the end-to-end vector R are given by the following ensemble averages

$$\langle R \rangle = \langle \sum_{n=1}^{N} r_n \rangle = \sum_{n=1}^{N} \langle r_n \rangle = 0 \tag{1}$$

as well as

$$\begin{aligned} \langle R \cdot R \rangle &= \langle \sum_{n=1}^{N} \sum_{m=1}^{N} r_n \cdot r_m \rangle \\ &= \sum_{n=1}^{N} \sum_{m=1}^{N} \langle r_n \cdot r_m \rangle \\ &= \sum_{n=1}^{N} b^2 = Nb^2 \end{aligned}$$

so finally one gets

$$\langle R \cdot R \rangle = Nb^2 \tag{2}$$

respectively

$$\langle R_x^2 \rangle = \langle R_y^2 \rangle = \langle R_z^2 \rangle = N \frac{b^2}{3}$$

and the assertion that the probability distribution of the end-to-end vector is a Gaussian in the $N \to \infty$ limit as a result of the central limit theorem in statistics (cf. [122]).

A quantity frequently used in polymer physics is the *radius of gyration* (cf. Fig. 3):

$$R_g^2 := \frac{1}{N} \sum_{i=1}^{N} (r_i - r_m)^2.$$

Here r_m denotes the center of mass. In this case one gets

$$R_g = \frac{\sqrt{N} b}{\sqrt{6}}.$$

It is worth noting that the above average end-to-end distance which in the case of this simple model is also the typical amplitude of the system's fluctuations becomes

negligible compared to the total unfolded contour length of the polymer at the thermodynamic limit. This result is a general property of statistical systems.

The most general form of a Gaussian probability distribution in three dimensions, that is normalized to $\int p(R)\, dR = 1$, is

$$p(R) = (2\pi\sigma^2)^{-\frac{3}{2}} e^{-\frac{(R-\mu)^2}{2\sigma^2}}.$$

The symmetry condition (1) means that the Gaussian must have its maximum at $r = 0$. Therefore, $\mu = 0$. It turns out that $\sigma^2 = \frac{Nb^2}{3}$ in 2 dimensions so that $\int R \cdot R\, p(R)\, dR = Nb^2$ in order to satisfy condition (2).

This leads to

$$p(R) = \left(\frac{3}{2\pi N b^2}\right)^{\frac{3}{2}} e^{-\frac{3R^2}{2Nb^2}}. \tag{3}$$

The length b is equivalent to the link size within the freely-jointed chain model but may be interpreted differently within other microscopic models.
Since the form of the probability distribution for the polymer is insensitive to the precise value of b, and indeed other aspects of the microscopic description of the chain, it is described as having *universality*.

The fact that this probability distribution reflects only the central limit theorem of statistics means that it is recovered for any microscopic model including the freely-jointed chain but also other models in which e.g. the monomer length distribution or monomer-monomer correlations are different. One such example is the wormlike chain model in which the polymer follows a smooth contour that can bend but not link (cf. Sec. 1.3). Another one is the Gaussian chain model (cf. Sec. 1.5).

One criticism of Eq. (3) is that $|R|$ can be larger than the total contour length of the chain Nb. This reflects the fact that the central limit theorem applies exactly only in the case $N \to \infty$.

Furthermore, Eq. 3 is identical with the distribution function for the displacement of a Brownian particle after N uncorrelated steps.
Eq. 3 will include only one parameter, if one uses Eq. 2 to substitute the mean squarred end-to-end distance: $\langle R^2 \rangle = Nb^2$:

$$p(R) = \left(\frac{3}{2\pi\langle R^2\rangle}\right)^{\frac{3}{2}} e^{-\frac{3R^2}{2\langle R^2\rangle}}. \tag{4}$$

In this thesis $R_0 := \langle R^2 \rangle^{1/2}$ will be used as another measure (besides R_g) for the extension of a polymer chain.

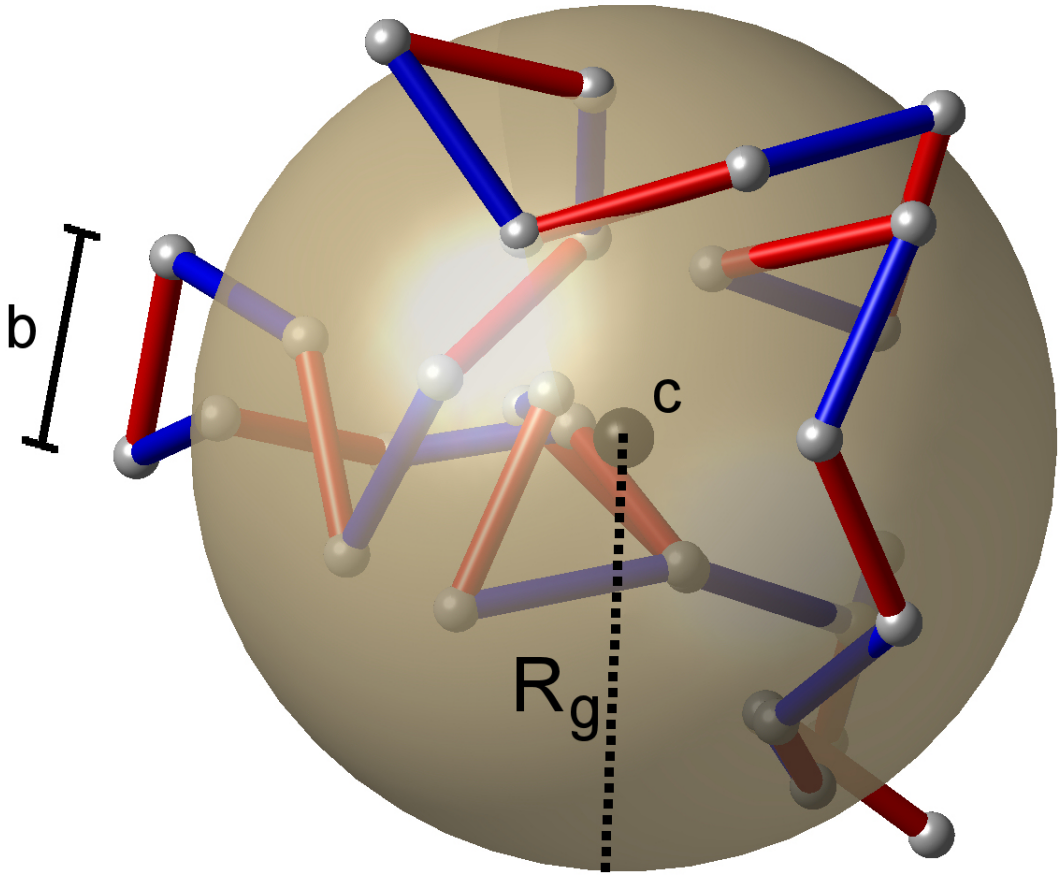


Figure 3: Example conformation of a short freely-jointed chain of $N = 30$ links with alternating color. The monomer length is $b = 100$ and the radius of gyration $R_g = 112$ in this particular case. c denotes the center of mass (black sphere). Further examples with $N = 1000$ (Fig. 159) and $N = 5000$ (Fig. 160) can be found in App. A.

A measure for the flexibility of a polymer chain is provided by the *persistence length*. Generally it can be derived from the orientational correlation function O which describes the correlation between the chain directions at two points with a distance Δl along the chain:

$$O(\Delta l) = \langle \hat{e}(l)\hat{e}(l + \Delta l)\rangle.$$

The unit vectors $\hat{e}$ describe the local chain direction in this case. The average is taken over the whole ensemble of possible chain conformations. Due to the flexibility of the chain the orientational correlations have to vanish for sufficiently large distances Δl i.e.

$$O(\Delta l \to \infty) \to 0.$$

A suitable choice for a parameter which measures the chain stiffness is given by

$$l_p = \int_0^\infty O(\Delta l) d(\Delta l).$$

It is called the *persistence length*.
In the case of a freely-jointed chain the chain segments show no orientational correlations.

If the considered chains are long enough they may be split into subchains of uniform length which has to be much larger than the persistence length. Then the junctions of the subchains are connected by the vectors

$$(p_1, p_2, ..., p_N)^t.$$

One can define a mean subchain diameter by

$$p_s = \sqrt{\langle p_i^2 \rangle}$$

like the mean squared end-to-end distance previously. This leads to

$$R_0 := \sqrt{\langle R^2 \rangle} = p_s N^{1/2},$$

where N is the number of subchains. Different choices of subchains have to lead to the same value of R_0. For two different choices with parameters N, p_s and M, q_s this gives

$$N p_s^2 = R_0 = M q_s^2$$

and therefore

$$\frac{N}{M} = \frac{q_s^2}{p_s^2}.$$

For large chain this equation can also be applied for the number of segments $n_s(i,j)$ and $m_s(i,j)$ between two chain points i and j:

$$\frac{n_s(i,j)}{m_s(i,j)} = \frac{q_s^2}{p_s^2}. \tag{5}$$

Eq. 5 shows the possibility of rescaling a freely-jointed chain. It expresses a basic property of ideal polymer chains namely the concept of *self-similarity*. This means that independent of the chosen length scale an ideal chain always has the same internal structure.

Self-similarity is the basic property of fractal objects and all spatial or temporal self-similar (fractal) structures have power law distributions.
The probability distribution function (pdf) of a self-similar random variable X is denoted by $p(X)$ in the following. Due to self-similarity the functional form of $p(X)$

should be left unchanged within a constant factor when X is multiplied by an arbitrary factor λ which means

$$p(\lambda X) = c(\lambda)p(X),$$

where $c(\lambda)$ is a constant factor depending on λ.
Differentiating this equation with respect to λ and substituting $\lambda = 1$ leads to

$$p'(X)X = c'(1)p(X).$$

The solution of this differential equation is a power law:

$$p(X) \propto X^{c'(1)}.$$

The *fractal dimension* d of an object can be defined as the exponent in the general equation

$$n \propto r^{\nu},$$

between the mass of an object and its diameter.

In the case of a freely-jointed chain d can be derived easily: If one proceeds n_s steps from a random chain point in the interior of a long chain, the corresponding average displacement will be given by

$$r(n_s) \approx p_s n_s^{1/2}.$$

That means the number of monomers n is a sphere of radius r may be estimated by

$$n \propto n_s \propto r^2$$

and thus the fractal dimension of an ideal chain is $d = 2$.

One can remove the arbitrariness in the choice of the segments by taking the chain's contour length C into account as well:

$$C = Np_s.$$

Together with the equation for the chain extension

$$R_0^2 = Np_s^2$$

this leads to

$$p_s = p_K := \frac{R_0^2}{C}.$$

p_K is called the *Kuhn length* and it also describes the stiffness of a polymer chain. A completely stiff chain has a Kuhn length of $p_K = C$.

The Kuhn length is highly correlated with the persistence length [207] by

$$p_K = 2l_p.$$

1.3 The Worm-Like Chain Model

The (Kratky–Porod) worm-like chain (WLC) model in polymer physics describes the behavior of semi-flexible polymers fibers. It envisions an isotropic rod (cf. Fig. 4) that is continuously flexible.
This is in contrast to the freely-jointed chain model which is flexible only between discrete segments.
The WLC model is particularly suited for describing stiffer polymers with successive segments displaying a sort of cooperativity: All pointing in roughly the same direction. At room temperature the polymer adopts a conformational ensemble that is smoothly curved. At $T = 0$K, the polymer adopts a rigid rod conformation.

Several biologically important polymers can be effectively modelled as WLCs including: Double-stranded DNA, unstructured RNA and unstructured polypeptides or proteins. The WLC-model will be used to simulate the chromatin linker DNA and thus will provide the opportunity to allow for nucleosome skips in the E2A-model (cf. Sec. 4.6.2 and Sec. 6.4).

For a polymer of length l the path of the polymer is parameterized by the unit tangent vector $\hat{t}(s)$ with $s \in [0; l]$ and by the position vector $r(s) \in \mathbb{R}^3$ along the chain.

Then

$$\hat{t}(s) = \frac{\partial r(s)}{\partial s}$$

and the end-to-end distance $R \in \mathbb{R}^3$ is given by

$$R = \int_0^l \hat{t}(s) ds \ .$$

One can show that the orientation correlation function for a worm-like chain decays exponentially:

$$\langle \hat{t}(s)\hat{t}(0) \rangle = \langle cos\Theta(s) \rangle = \exp\left(-s/l_P\right).$$

Thus one can estimate the *persistence length* l_P of the polymer.

A useful measure for the spatial extension of a WLC polymer is again the mean square end-to-end distance:

$$\begin{aligned}\langle R^2\rangle &= \langle RR\rangle \\ &= \langle \int_0^l \hat{t}(s)ds \int_0^l \hat{t}(s')ds'\rangle \\ &= \int_0^l ds \int_0^l \langle \hat{t}(s)\hat{t}(s')\rangle ds' \\ &= \int_0^l ds \int_0^l e^{-|s-s'|/l_P} ds' \\ &= 2l_P\, l \left[1 - \frac{l_P}{l}\left(1 - e^{-l/l_P}\right)\right].\end{aligned}$$

In the limit $l \gg l_P$ one gets $\langle R^2\rangle \approx 2l_P\, l$ or more precisely the approximation

$$l_P \approx \frac{l}{2} - \sqrt{\frac{l^2}{4} - \frac{\langle R^2\rangle}{2}}$$

by neglecting the e^{-l/l_P}-term.
This can be used to show that a Kuhn segment is equal to twice the persistence length of a WLC.

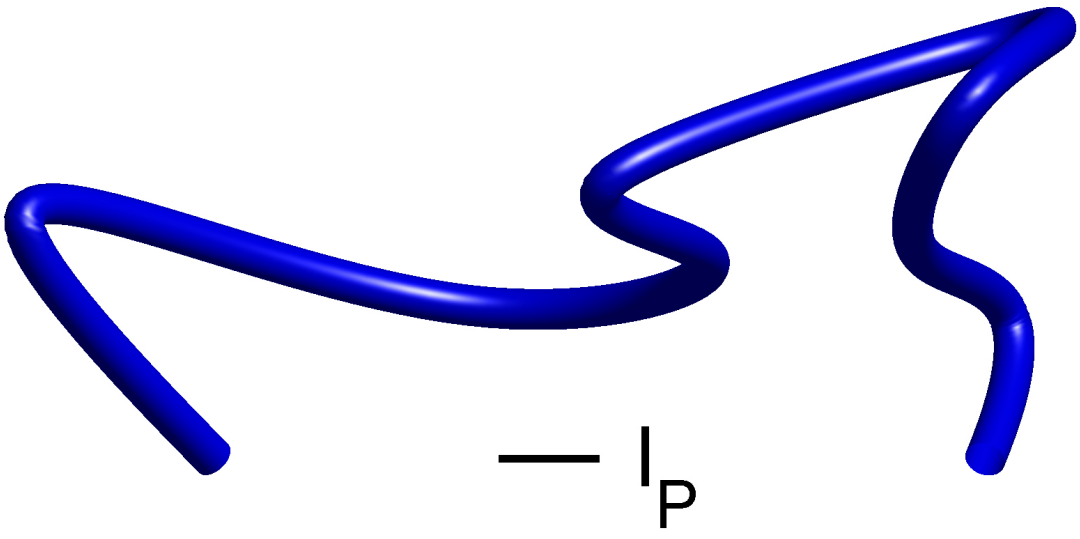


Figure 4: Example of a Worm-like Chain with sketched in persistence length l_P. Another WLC conformation can be found in Fig. 161 in App. A.

1.4 Free Energy of Distorting a Polymer Chain

If the two free ends of an ideal chain are attached to some kind of micro-manipulation device, then the device will experience a force exerted by the polymer. The ideal chain's energy is constant and thus its time-average, the internal energy, is also constant which means that this force necessarily stems from a purely *entropic* effect.

This entropic force is very similar to the pressure experienced by the walls of a box containing an ideal gas.
The internal energy of an ideal gas depends only on its temperature and not on the volume of its containing box. So it is not an energy effect that tends to increase the volume of the box like gas pressure does. This implies that the pressure of an ideal gas has a purely entropic origin.

In its solvent the ideal chain is constantly subject to collisions from moving solvent molecules and each of these collisions shifts the system from its current microscopic state to another, very similar microscopic state.
For an ideal polymer there are more microscopic states compatible with a short end-to-end distance than there are microscopic states compatible with a large end-to-end distance (cf. Eq. 4). Thus, for an ideal chain, maximizing its entropy means reducing the distance between its two free ends. Consequently, a force that tends to collapse the chain is exerted by the ideal chain between its two free ends.

1.4.1 Free Energy of Pulling a Polymer Chain

The free energy cost of pulling the ends of a whole chain until they are apart will be considered in the following.

Let $S = k\log(\Omega)$ be the entropy where Ω is the number of accessible microstates. For a chain with end-to-end vector R the number Ω is proportional to the probability of that microstate (end-to-end vector) being occupied by the principle of equal a priori probability (cf. [97]).
The entropy difference between a chain held with end-to-end distance R and one held with the preferred end-to-end vector of zero is

$$\Delta S(R) = k\log(c\,p(R)) - k\log(c\,p(0)) = k\log\left(\frac{p(R)}{p(0)}\right).$$

The constant c relates the probability p to the number of microstates Ω. With Eq. 4 this leads to

$$\Delta F = -T\Delta S = \frac{3}{2}\frac{kTR^2}{Nb^2}. \tag{6}$$

This can be interpreted as an *entropic spring* with spring constant $k = 3\frac{kT}{Nb^2}$.

1.4.2 Freely-Jointed Chain under Length Constraint

The case of an ideal chain whose two ends are attached to fixed points will be considered in this section.

The vector R joining these two points characterizes the macroscopic state of the freely-jointed chain. To each macro-state corresponds a certain number of micro-states $\Omega(R)$. Since the ideal chain's energy is constant each of these micro-states is equally likely to occur. The entropy associated to a macro-state is thus equal to:

$$S(R) = k \log(\Omega(R)).$$

A precise determination of $\Omega(R)$ would require a quantum model for the ideal chain. However, Eq. 4 gives the probability density associated with the end-to-end vector of the unconstrained ideal chain.

Since all micro-states of the ideal chain are equally likely to occur, $p(R)$ is proportional to $\Omega(R)$. This leads to the following expression for the classical (relative) entropy of the ideal chain:

$$S(R) = k \log(p(R)) + c$$

where c is a fixed constant.

Let F denote the force exerted by the chain on the point to which its end is attached. From the above expression of the entropy one can deduce an expression of this force. Suppose that, instead of being fixed, the positions of the two ends of the ideal chain are now flexible. If R is changed by a tiny amount dR, then the variation of internal energy of the chain is zero since the energy of the chain is constant.

This condition can be written as:

$$0 = dU = \delta W + \delta Q.$$

δW is the amount of mechanical work by changing R and δQ is the heat transferred by the solvent to the ideal chain.

If one assumes now that the transformation imposed on the system is quasi-static (i.e. infinitely slow), then the system's transformation will be time-reversible and one can assume that during its passage from macro-state R to macro-state $R + dR$, the system passes through a series of thermodynamic equilibrium macro-states.

This has two consequences: First, the amount of heat received by the system during the transformation can be tied to the variation of its entropy: $\delta Q = TdS$, where T is the temperature of the chain, and second, in order for the transformation to remain

infinitely slow the mean force exerted on the end points of the chain must balance the mean force exerted by the chain on its end points. Calling f_R the force exerted by the change of R and f_c the force exerted by the chain, we have:

$$\delta W = \langle f_R \rangle dR = -\langle f_c \rangle dR.$$

This leads to

$$\langle f_c \rangle = t\frac{dS}{dR} = \frac{kT}{p(R)}\frac{dp(R)}{dR}$$

$$\langle f_c \rangle = -kT\frac{3R}{Nl^2} \tag{7}$$

Eq. 7 is the equation of state of the ideal chain.
Since the expression depends on the central limit theorem it is only exact in the limit of polymers containing a large number of monomers (i.e. the thermodynamic limit).

1.5 The Gaussian Chain Model

The Gaussian model is inspired by the insensitivity to the microscopic details of the chain.
Consider a chain made up of orientationally uncorrelated (freely-jointed) links, where the length of any link vector r is not constant but has a probability distribution

$$p(r) = \left(\frac{3}{2\pi b^2}\right)^{\frac{3}{2}} e^{-\frac{3r^2}{2b^2}},$$

which leads to

$$\langle r^2 \rangle = b^2$$

and so, since the links are orientationally uncorrelated, both the links and the resulting chain of links have statistics that are also given by the equations above (cf. Sec. **??**).

Furthermore, from the probability distribution for the end-to-end vector one gets

$$\begin{aligned} p(R) &= \prod_{n=1}^{N}\left(\frac{3}{2\pi b^2}\right)^{\frac{3}{2}} e^{-\frac{3r_n^2}{2b^2}} \\ &= \left(\frac{3}{2\pi b^2}\right)^{\frac{3N}{2}} \exp\left(-\sum_{n=1}^{N}\frac{3(R_n - R_{n-1})^2}{2b^2}\right). \end{aligned}$$

One can see from this that this model has the physical analogy of "beads on a string" via the potential

$$U = \frac{3}{2b^2}kT\sum_{N}^{n=1}(R_n - R_{n-1})^2$$

similar to that which would arise by N beads connected by Hookian springs.

Since the mean ACN is scale invariant this means that the mean ACN of equilateral chains is equal to that of Gaussian chains at $T = 0$ when no spring is exited.

At equilibrium the Boltzmann distribution for the potential U above gives the same equilibrium probability distribution as the freely-jointed chain (or any other choice with short range correlations). For the Gaussian model this is an exact result for all R.

1.6 The Self-Avoiding Walk in Polymer Statistics

An excluded volume chain is a chain in which the individual links (or monomers) feel each other's presence and are repelled from one-another.

If nothing else, the monomers in the chain always have steric interactions with each other, this being the name given to hard-wall like interactions (ultimately quantum mechanical in origin) that stop two chemical groups occupying the same volume.
The problem was first discussed by Kuhn and Flory (cf. [83; 123]) who predicted

$$\langle R^2 \rangle \propto N^{2\nu} \quad \text{with } \nu \approx \frac{3}{5},$$

a result which seems to agree with many experiments.

The excluded volume interactions are difficult to include properly because they are non-local in n: The ith and jth monomers might be very close spatially, even if $|i-j|$ is large.

A prototype mean field theory, the Flory theory (cf. [83; 123]), gives ν quite well (but fails in other respects, e.g. higher order moments of the chain density).

The excluded volume interactions are assumed to be driven by some general repulsive monomer-monomer interaction potential $U(R_n - R_m)$. Provided U is short ranged one can make a mean filed estimate of the two body interactions by treating the polymer as a gas of disconnected monomers confined within the same volume V as the polymer coil (of the order of the cube of the root mean squared (r.m.s.) polymer end-to-end distance $V \simeq \bar{R}^3$).
This procedure can be thought of as equivalent to cutting all the bonds between neighboring monomers along the polymer backbone.

The free energy cost of putting N such excluded volume monomers into a box with volume V and hence average concentration $c = \frac{N}{V}$ is approximately

$$\Delta F_\nu = \Delta F_1 + \Delta F_2,$$

where the subscript ν denotes that it is the free energy change due to the effect of excluded volume.

The term ΔF_1 is an unimportant term that arises from the chemical work done by initially preparing the monomers in the box, i.e. it is a constant independent of $\bar{R}$. These might be called one-body terms as they are not identified with any inter-monomer interactions.

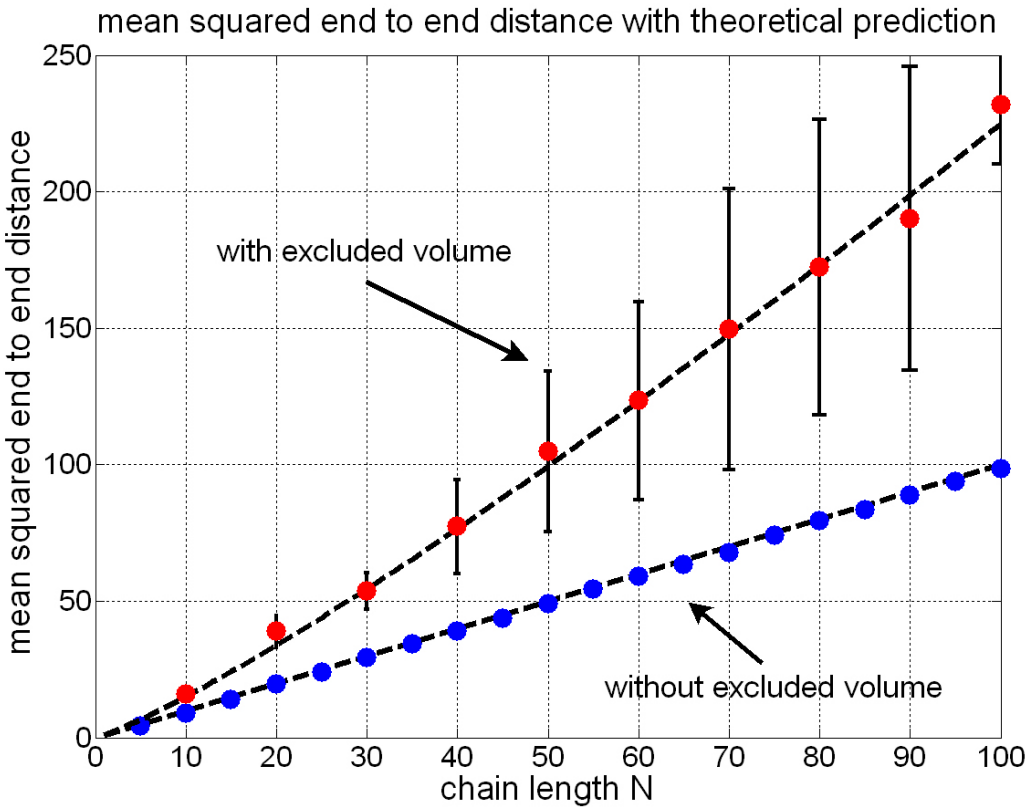


Figure 5: Mean squared end-to-end distance for equilateral chains with and without excluded volume. The dashed black lines show the theoretical predictions for $\nu = 0.5$, $\nu = 0.588$ and $b^2 = 1$. The chains with excluded volume are much more stretched than those ones without. This will play an important role for the mean Average Crossing Number in the following sections. Adapted from [61].

The second term ΔF_2 depends on the concentration squared

$$\Delta F_2 = \nu k T c^2 V. \tag{8}$$

Thus $\Delta F_2 \propto V$ because it is extensive and $\Delta F_2 \propto c^2$ since it corresponds to two-body terms and hence to the effect of interactions between pairs of monomers.
The remaining unknown prefactor has dimensions of energy $\times$ volume and so one can write it as $kT\nu$ where ν is an excluded volume constant.
Eq. (8) is the first non-trivial term in an expansion in powers of c. Higher order terms (such as $\Delta F_3 \sim c^3$) are all negligible at the smallest concentrations of interest and so

one can truncate at second order. The three-body terms, here neglected, would correspond to interactions simultaneously involving three closely approaching monomers. These interactions are simply much more rare at the very low concentrations of interest here.

One can motivate Eq. (8) by considering the work done by the pressure field of an ideal gas of monomers with concentration c that are made non-ideal by being given an effective volume. Mathematically this is done by allowing them to smoothly swell from point particles to particles with some finite volume of the order of ν, the effective volume from which other monomers are excluded due to the repulsive form of the potential $U(r)$. Thus the work done is $dW = -\int p\, dV$, where the ideal gas equation ($pV = NkT$) is used to find the pressure $p = \frac{NkT}{V}$ exerted by the ideal gas of monomers (the minus sign arises from considering the work done on, rather than done by, the system).
Now work is done against this pressure as one forces the monomers to swell from point like objects to objects with volume ν.

The total available volume in the box decreases from V to $V(1 - c\nu)$. Thus

$$dW = -NkT \int_{V}^{V(1-c\nu)} \frac{dV'}{V'} \quad \text{for } c\nu \gg 1$$

which can be shown to give Eq. (8).

From Eq. (8) one can construct a scaling argument for the r.m.s. chain size $\bar{R} \sim N^{\nu}$. The total free energy change is the entropic spring distribution Eq. 6 plus the excluded volume term (cf. Eq. (8)):

$$\delta F = \Delta F + \Delta F_2 \simeq kT \left(\frac{\bar{R}^2}{Nb^2} + \frac{\nu N^2}{\bar{R}^3} \right)$$

The equilibrium state is the state of minimum free energy. Minimizing δF with respect to $\bar{R}$ gives by solution of $\frac{\partial \delta F}{\partial \bar{R}} = 0$:

$$\bar{R} \simeq \left(\frac{\nu}{b^3} \right)^{\frac{1}{5}} b N^{\frac{3}{5}}$$

i.e. $\nu = \frac{3}{5}$.
This result is remarkably close to that obtained from far more sophisticated treatments, including perturbation and renormalization group theories which both yield $\nu \approx 0.588$ (cf. Fig. 5).

1.7 Ideal Chains and the Θ-Temperature

The excluded volume parameter is also called a second virial coefficient and is defined by

$$\nu = \int \left(1 - e^{-\frac{U(r)}{kT}} \right) dV \tag{9}$$

where U is the monomer-monomer interaction potential.

Consider the simple case of a potential U made up of a hard wall, or steric, repulsive term and a weak attractive term $U = V_{\text{hard}} + V_{\text{attr}}$ where V_{hard} would be infinite, if there were overlaps between particles and zero otherwise.

$$\begin{aligned}\nu &= \int \left(1 - \exp\left[-\frac{V_{\text{hard}}}{kT}\right] \exp\left[-\frac{V_{\text{attr}}}{kT}\right]\right)\, dV \\ &\approx \int \left(1 - \exp\left[-\frac{V_{\text{hard}}}{kT}\right] \left(1 - \frac{V_{\text{attr}}}{kT}\right)\right)\, dV\end{aligned}$$

provided $V_{\text{attr}} \ll kT$. One can split the domain of this integral into two parts, volume V_1 lying outside the range of the hard wall potential and V_2 lying inside. Thus

$$\nu = \int_{V_1} dV + \frac{1}{kT}\int_{V_2} V_{\text{attr}}(r)\, dV = A - \frac{B}{T}.$$

Defining $\Theta = \frac{B}{A}$, one can see that

$$\nu = \nu_{\text{hard}}\left(1 - \frac{\Theta}{T}\right).$$

Hence, at $T = \Theta$, the chain is ideal (Gaussian) and $\nu = \frac{1}{2}$.

1.8 Scattering Function and Pair Distribution Function

The *scattering function* $S(\mathbf{q})$ for a system of equal particles is defined by

$$S(\mathbf{q}) = \frac{I(\mathbf{q})}{I_m N_m},$$

where N_m represents the total number of particles in the sample and I_m is the scattering intensity produced by one particle, if placed in the same incident beam. $I(\mathbf{q})$ is the distribution of the intensity depending on the scattering vector $\mathbf{q}$ which is defined as

$$\mathbf{q} = \mathbf{k}' - \mathbf{k},$$

i.e. the difference between the wave vectors of the incident and the scattered plane waves.

If $k' \approx k = \frac{2\pi}{\lambda}$, then q will be related to the Bragg scattering angle by

$$q = \frac{4\pi}{\lambda}\sin(\theta_B).$$

The *scattering amplitude* C is given by

$$C = \sum_{i=1}^{N_m} \exp(i\phi_i),$$

where phases ϕ_i are determined by the particle position and the scattering vector:

$$\phi_i = -\mathbf{q}\mathbf{r}_i.$$

This leads to the following expression for the scattering amplitude:

$$C(\mathbf{q}) = \sum_{i=1}^{N_m} \exp(-i\mathbf{q}\mathbf{r}_i).$$

Since the scattering intensity is proportional to the squared modulus of C, i.e.

$$I(\mathbf{q}) \propto \langle |C(\mathbf{q})|^2 \rangle$$

one gets the following expression for the scattering function $S(\mathbf{q})$:

$$\begin{aligned} S(\mathbf{q}) &= \frac{1}{N_m} \langle |C(\mathbf{q})|^2 \rangle \\ &= \frac{1}{N_m} \sum_{i=1}^{N_m} \langle \exp(-i\mathbf{q}\mathbf{r}_i) \rangle. \end{aligned}$$

Instead of discrete positions r_i one can also use a continuum description by introducing the distribution of the particle density $c_m(\mathbf{r})$ (cf. [207]). Then it can be showed that $S(q)$ is the Fourier transform (cf. App. G) of the space dependent correlation function of the particle density:

$$S(\mathbf{q}) = \frac{1}{\langle c_m \rangle} \int \exp(-i\mathbf{q}\mathbf{r})(\langle c_m(\mathbf{r}) c_m(0) \rangle - \langle c_m \rangle^2) d\mathbf{r}$$

Another important variable which describes results of scattering experiments is the pair distribution function $g(\mathbf{r})$. In this case the product $g(\mathbf{r})d^3\mathbf{r}$ is the probability that starting from a particular particle another one (even the same one) can be found in the volume element $d^3\mathbf{r}$ at a distance $\mathbf{r}$.
In the discrete case the radial pair distribution function is given by

$$g(r) = \frac{1}{4\pi r^2} \frac{V}{N_m - 1} \left(\frac{1}{N_m} \sum_i \sum_{j \neq i} \delta(r - r_{i,j}) \right).$$

A value of one corresponds to the mean particle density of the system. Since the density distribution and the pair distribution are related one can connect the scattering function with the pair distribution as well [207]:

$$S(\mathbf{q}) = \int \exp(-i\mathbf{q}\mathbf{r})(g(\mathbf{r}) - 1) d^3 r.$$

Hereby, the function $h(\mathbf{r}) = g(\mathbf{r}) - 1$ is called the *pair correlation function*. This shows that the scattering function is actually the Fourier transform (cf. App. G) of the pair

distribution function.

The pair distribution function will be used several times in Sec. 7 to describe the chromatin nanostructure.
Furthermore, the two-dimensional pair distribution function will be applied for a comparison with experimental data. It is explained in Sec. 7.3.3.
Moreover, in Sec. 7.3.2 it will be shown what the *radial* pair distribution function looks like and how this changes the relation to the scattering function $S(q)$.

In the case of a freely-jointed chain the pair distribution function is given by

$$g(\mathbf{r}) = \frac{1}{N} \sum_{i=-(N-1)}^{N-1} (N - |m|) \left(\frac{3}{2\pi |m| p_s^2} \right)^{3/2} \exp\left(-\frac{3r^2}{2|m|p_s^2} \right).$$

It can be shown that in the case of small distances within the macromolecule $g(\mathbf{r})$ behaves like

$$g(\mathbf{r}) \propto \frac{1}{r}.$$

This power law behavior comes again from the self-similarity of the chain [207].

2 Background - Introducing Chromatin

In this section a summary of several publications (cf. references below) concerning chromatin will be given to provide an overview of the current state of knowledge.
To explain the basic motivation of this thesis it will furthermore be pointed out, how little is actually known about chromatin and why this is the case.

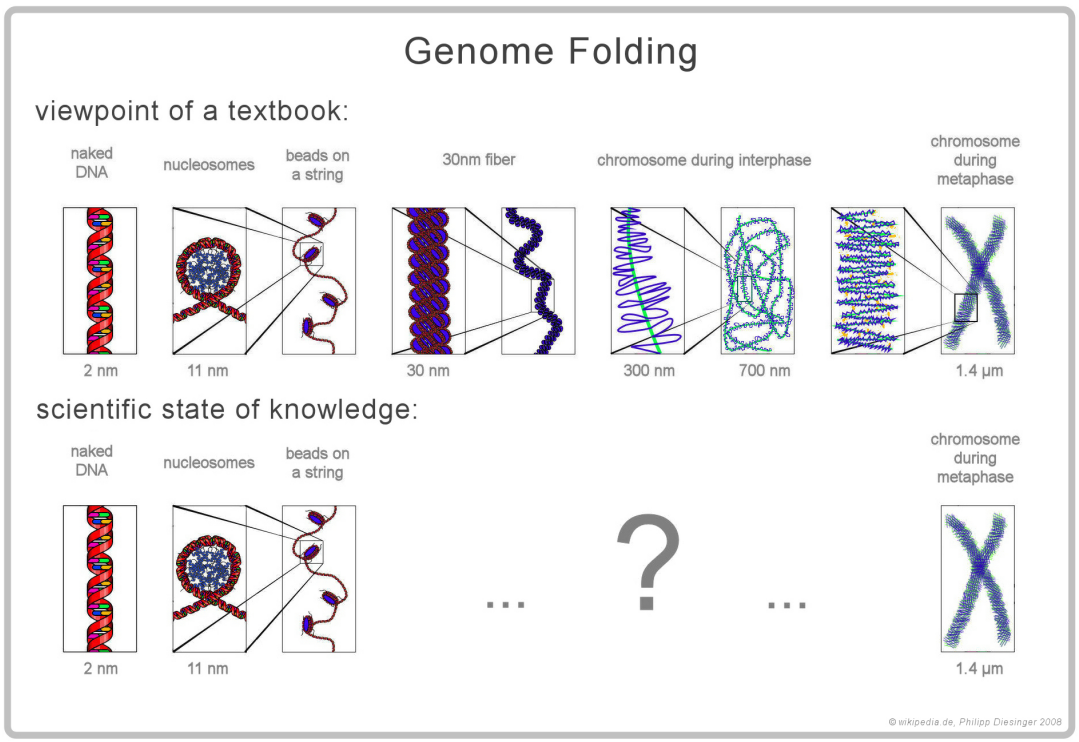


Figure 6: Genome folding at different scales. Comparison between the naive viewpoint of a textbook [231] and the current scientific state of knowledge. A basic motivation of this thesis is to contribute to the understanding of chromatin on the length scale which is marked by the question mark. This figure is partially adapted from [231]. A larger version of this figure can be found in App. A (Fig. 164).

2.1 Introduction

Chromatin is found inside the nuclei of eukaryotic cells, and within the nucleoid in prokaryotic cells. The major proteins involved in chromatin are histone proteins, although many other chromosomal proteins have prominent roles, too. DNA packaging in prokaryotic cells, although also involving architectural proteins, follows quite different rules [43; 44].
The DNA molecules that run the length of each human chromosome are arguably the

longest and most important biomolecules known.
But although the DNA that runs the length of each of the 46 chromosomes has been sequenced, it is still very little known about how DNA is folded in 3D space within the nucleus of a living cell. Common sense suggests there must be some underlying order within the apparent tangle [41].

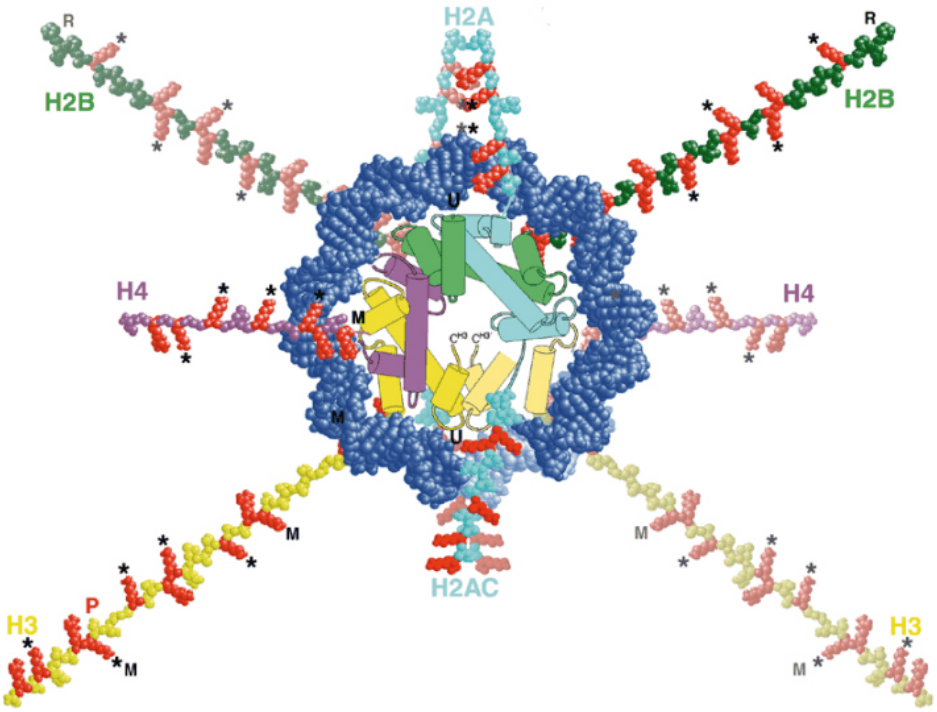


Figure 7: Nucleosome Structure: Although 75–80 % of the histone molecule is incorporated in the core, the N-termini of all the histones extend from the rest of the protein that makes up the nucleosome core. These "tails" are very well conserved, despite the fact that they do not seem to be well structured. If they are removed by trypsin treatment, the nucleosome will remain stable but the DNA becomes more accessible to nucleases: they only protect 106 bp instead of 146 (20 bp at each end is lost). They also lose the ability to form a compacted structure at high salt concentrations indicating that the tails may mediate an interaction with histones in adjacent nucleosomes. Figure adapted from [234].

2.2 Genome Folding - Spatial Organization of DNA

The problem of DNA folding, i.e. the problem of how plant and animal genomes organize themselves into volumes, whose linear dimensions are many orders of magnitude smaller than their contour length, is an important issue in modern biophysics.
For instance, human DNA is billions of base pairs (bp) long ($5 \cdot 10^9\ bp \approx 2m$). This

length of highly-charged ($\lambda_{DNA} \approx \frac{2e}{0.34nm} = 2\frac{e}{bp}$ for naked DNA, cf. [181]) and hard to bend linear polymer (persistence length ≈50nm) must be condensed into chromosomes that fit into cell nuclei, whose characteristic size is a *micron*.

2.2.1 DNA Organization at the 10nm Level

An important part of the condensation process is the compaction of DNA with oppositely-charged protein aggregates (so called *histone*) that have the shape of squat cylinders. These aggregates are octameric complexes consisting of pairs of the four core histones H2A, H2B, H3 and H4 (cf. Fig. 8).

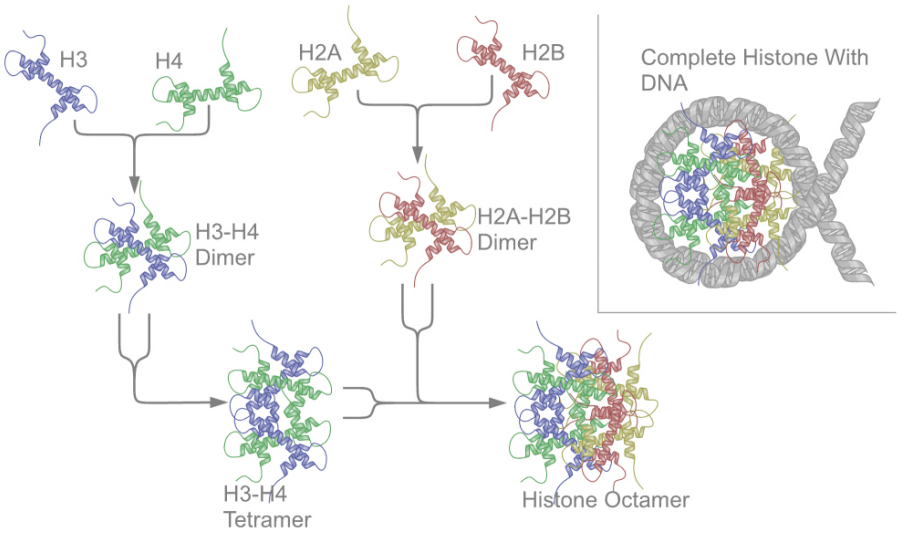


Figure 8: The proteins that make up the nucleosome are called histones. Histones H2A, H2B, H3 and H4 are part of the nucleosome while histone H1 is involved the linker DNA between the two nucleosomes. Adapted from [231].

A DNA stretch of 146bp is wrapped in a left-handed superhelical turn around each histone octamer and is connected via a stretch of linker DNA to the next such protein complex. The whole DNA-histone-complex is called a *nucleosome* (cf. Fig. 7).

Each protein aggregate together with its wrapped DNA comprises a nucleosome core particle with a radius of about 5nm and a height of about 6nm. Together with its linker DNA it is the fundamental chromatin repeating unit. It carries a large electrostatic charge [118].
Whereas the structure of the core particle has been resolved up to high atomic resolution [137], there is still considerable controversy about the nature of the higher-order

structures to which they give rise.

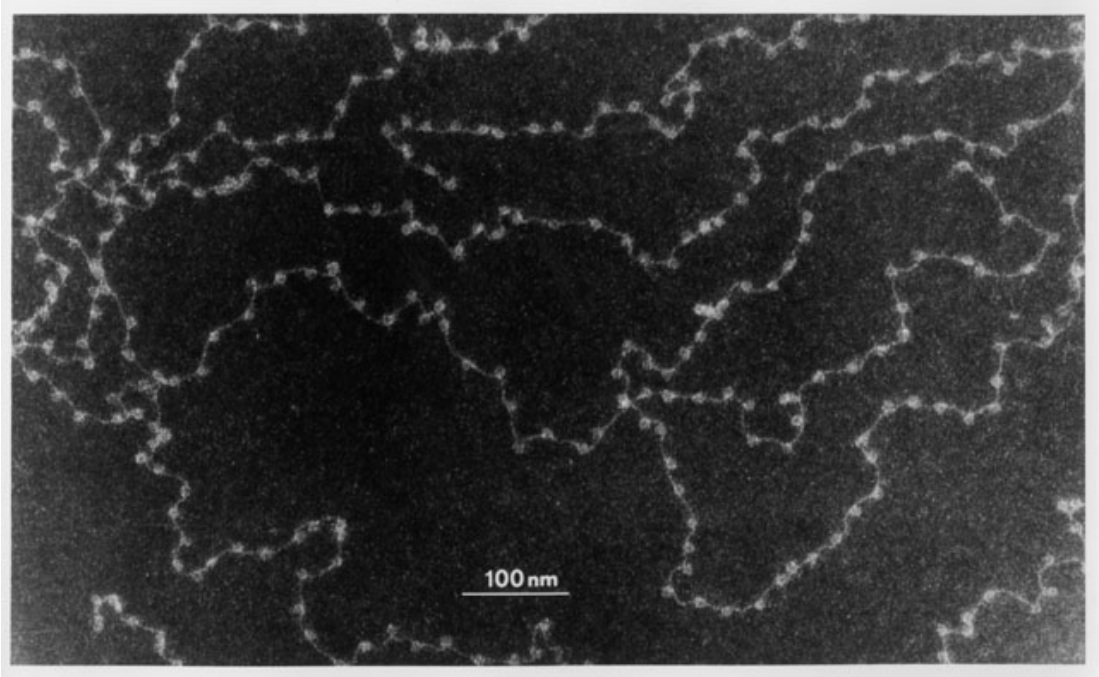


Figure 9: A darkfield electron micrograph of chromatin spilling out of a chicken erythrocyte nucleus in a low ionic strength buffer. Stained with uranyl acetate. The beads are nucleosomes, which are each made up of an octamer of histone proteins (two H2A, two H2B, two H3 and two H4). Adapted from [152].

When stretched the chromatin string looks like beads-on-a-string (cf. Fig. 9).
The recent accumulation of evidence that the chromatin structure above the level of the core particle plays a key role in determining the transcriptional status of genes and genetic loci [78; 81] illustrates the critical importance of understanding the fundamental folding properties of nucleosome arrays: Studies of chromatin compaction in response to changes in the ionic environment [226] have established that the phenomenon can be accounted for by electrostatic interactions between DNA, histone proteins and free ions [40].
Major contributions to these interactions are provided by the N-terminal domains of the core histones which contain roughly half of the basic amino acids of the octamer and the C-terminal domains of the linker histones which contain about $\frac{3}{5}$ of the positive charges in these molecules.

Indeed, chromatin compaction requires the presence of the core histone N-termini and 30nm chromatin fibers are not formed in the absence of linker histones [19].
However, the precise interactions that lead to specific chromatin higher-order structures have remained elusive.

2.2.2 DNA Organization at the 30nm Level

Although the precise structure of the chromatin fiber (also called *30nm-fiber* due to its diameter) in the cell is not known in detail, there are some accepted facts:

1. the nucleosomes lie perpendicular to the axis of the fibre
2. the linker histones lie on the inside of the structure
3. it readily unwinds into the 10nm "beads-on-a-string" fiber.

Longstanding controversy [216; 218; 227] surrounds the structure of the 30nm fiber for which there are mainly two competing classes of models: The solenoid models [80; 211] and the crossed-linker-models [105; 133; 235].

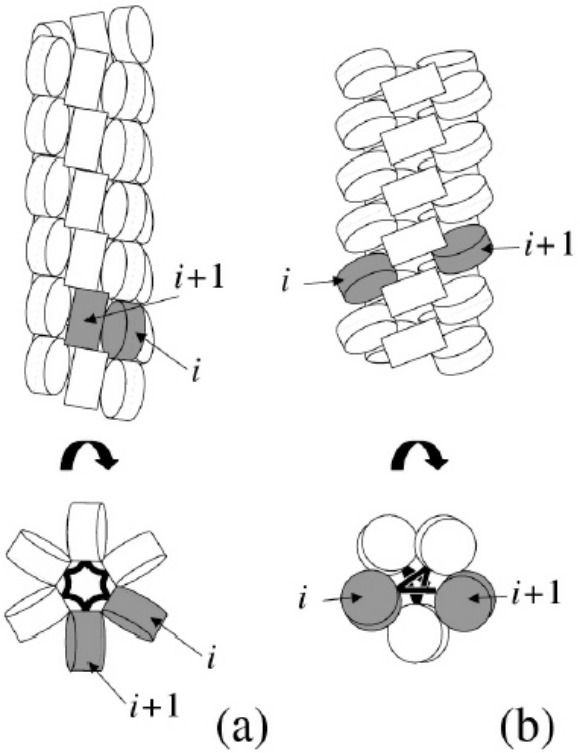


Figure 10: The two competing classes of models for the 30-nm fiber can be distributed into (a) solenoid models and (b) crossed-linker models. For both fiber types the side and the top view are shown. Two nucleosomes that are directly connected via DNA linker are shaded in gray. In the solenoid these nucleosomes are located on the same side of the fiber requiring the linker to be bent. In the crossed-linker case they sit on opposite sides of the fiber and are connected via a straight linker. Adapted from [128].

In the solenoid model (cf. Fig. 10) it is assumed that the chain of nucleosomes forms a helical structure with the axis of the core particles being perpendicular to the solenoidal axis (the axis of an octamer corresponds to the axis of the superhelical path of the DNA that wraps around it).
The DNA entry-exit side faces inward towards the axis of the solenoid. The linker DNA is required to be bent in order to connect neighboring nucleosomes in the solenoid.

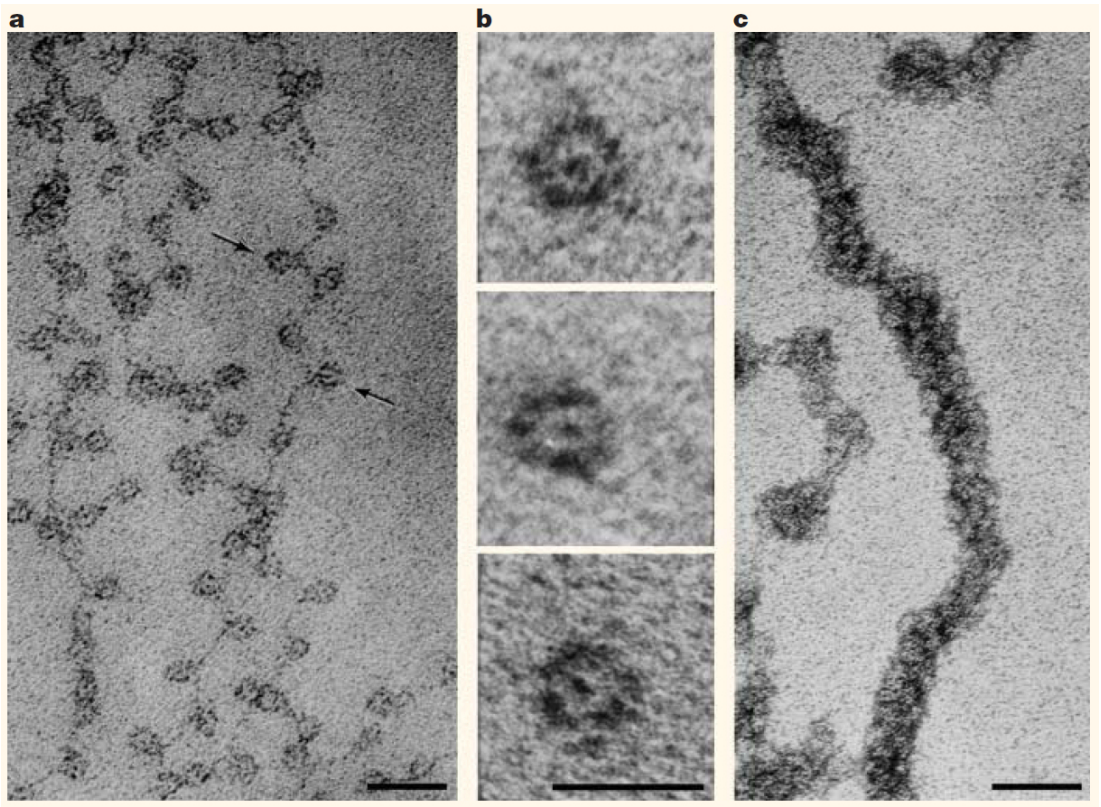


Figure 11: A gallery of electron micrographs of chromatin. a) low ionic-strength chromatin spread, the "beads on a string". Size marker: 30nm. b) Isolated mononucleosomes derived from nuclease-digested chromatin. Size marker: 10nm. c) chromatin spread at a moderate ionic strength to maintain the 30-nm higher-order fiber. Size marker: 50nm. Adapted from [151].

The other class of models assumes straight linkers that connect nucleosomes located on *opposite* sides of the fiber (cf. Fig. 10). This results in a three-dimensional crossed-linker-pattern. Such an arrangement with peripherally arranged nucleosomes and internal linker DNA segments is fully consistent with observations on intact nuclei and also allows dramatic changes in compaction level to occur without a concomitant change in topology.

Images obtained by electron cryo-microscopy (EC-M) should in principle be able to distinguish between the structural features proposed by the different models mentioned above (cf. Fig. 11 and Fig. 12).
The micrographs show a crossed-linker-pattern at lower salt concentrations and they indicate that the chromatin fiber becomes more and more compact when the ionic strength is raised towards the physiological value [19].

However, for these denser fibers it is still not possible to detect the exact linker geometry.

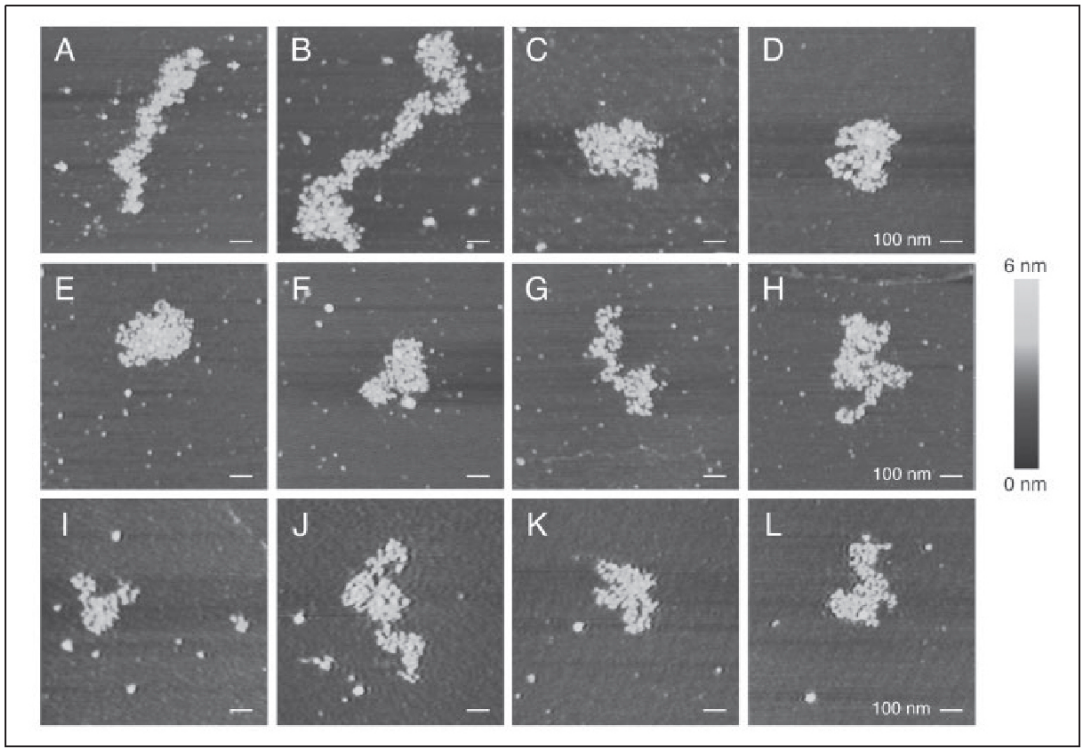


Figure 12: SFM images of chromatin fibers (1x1 μm images). A–D, single chromatin fiber fragments. E–H, chromatin fibers incubated with NAP1. I–L, chromatin fibers depleted of linker histone H1. Image adapted from [116].

A major difficulty encountered in this structural issue comes from the size of the fiber, in between the range of light microscopy (scales larger than a few hundreds nanometers) and cristallography and other experimental tools available at molecular scales (less than a few tens of nm).
An additional difficulty comes from the fact that it is by no means certain that the structures observed in vitro for purified fibers or with reconstituted nucleosome arrays reproduce the actual functional (and possibly dynamic) structure of the fiber in vivo.

Therefore, information should be gained from in vivo imaging.
Furthermore, the fiber experiences major structural reorganizations during the cell cycle. Several experimental studies provided a direct, visual access to mitotic condensation [90; 140; 143] to be compared to in vivo investigations of nuclear organization during transcription [45; 99; 195].

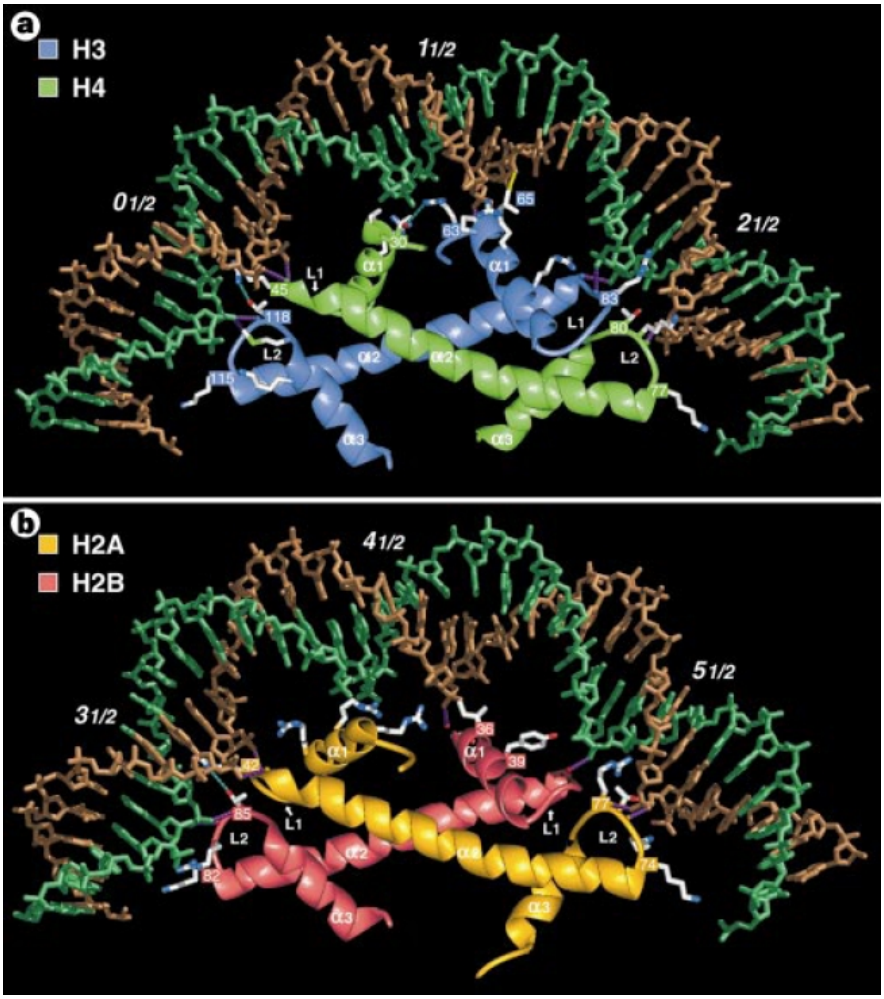


Figure 13: Nucleosomes are the fundamental repeating subunits of all eukaryotic chromatin. They provide the lowest level of compaction and are important in the regulation of transcription by preventing RNA polymerase from unnecessarily accessing the promoter regions of genes which are not needed by the cell. If the requirements of the cell change, enzymes ('remodeling factors') can remove or change the position of the nucleosome to allow access. Adapted from [173].

These observations evidence that different condensed structures are to be considered according to the cell cycle stage and to the functional role of the compaction.

The structure designed to achieve the tight and maximal packing required during mitosis is presumably different from the organized condensed structure framing transcription regulation.

Furthermore, it has been observed that the condensed mitotic structure itself exhibits two levels of compaction (6 nucleosomes per 10nm and 10 nucleosomes per 10nm) [140]. Conversely, the decompaction also occurs in two stages [143].

A further discussion of the fiber structure as well as references about experimental supports and structural reorganizations of the zig-zag crossed-linkers structure can be found in [149].

Experiments on dinucleosomes have been made to check if the nucleosomes collapse upon an increase in ionic strength. A collapse would only occur, if the linkers bent, and an observation of this phenomenon would support the solenoid model.

The experiments by Yao et al. [238] as well as more recent experiments by Butler and Thomas [33] indeed reported a bending of the linkers but do not agree with experiments by Bednar et al. [19] and by others that did not find any evidence for a collapse.

However, the internal structure of the dense 30nm fiber could not be resolved despite enormous experimental efforts, especially X-ray diffraction studies [218]. Even less is known about how chromatin folds into chromosomes on larger scales.

Changes in chromatin structure are affected by chemical modifications of histone proteins such as methylation (DNA and proteins) and acetylation (proteins), and by non-histone, DNA-binding proteins.

It was proposed that an additional conformational transition between two structures of the nucleosome takes place to achieve the mitotic compaction ratio. This transition is called *nucleosome gaping*.

Nucleosome gaping corresponds to the unsticking of both H2A and H2B dimers. Gaping is beneficial insofar as it improves the strength of internucleosomal stacking interactions. It is to note that recent observations of tetranucleosome crystals by the Richmond group support the key role of stacking interactions between nucleosomes [179]. However, the tetranucleosome crystal structure does not contain any linker histones.

The energy balance should also take into account the variation in the electrostatic repulsion of nucleosomal DNA and the change in linker twist energy upon gaping. The detailed estimation leads to the conclusion that the nucleosome gaped conformation is a *metastable* state and that some active process is required to cross irreversibly the energy barrier separating it from the native state [148; 149].

The height of this energy barrier is highly sensitive to the local ionic strength and could be tuned by any local supply of charges: This is a first hint towards the possible role of ionic exchanges in fiber conformational changes and associated regulatory mechanisms.

2.3 The Role of Chromatin for DNA-Protein Interactions

Embedding of DNA within the highly organized and constrained chromatin structure strongly modifies its affinity for any ligand compared to free DNA.
Actually, any way of controlling the mechanical constraints experienced by linker DNA would in turn control its affinity for any protein, whose binding is accompanied by some distorsion of DNA (bend or twist), hence associated with an elastic energy cost in an end-constrained linker DNA.
In a locked chromatin fiber the strength of these constraints is controlled by the linker DNA anchoring onto fixed nucleosomes. Any modification of this anchoring or any weakening of the tight three-dimensional positioning of the nucleosomes would relax these constraints and lower the associated energy barrier. Carrying on the argument allows to connect the degree of chromatin organization and the nature and rate of DNA transactions.
This provides a mechanistic link between the variations of gene expression pattern and the variations of chromatin compaction with the cell type and along the cell cycle. It thus hints at a direct interplay between chromatin fiber conformation (at different scales), cell cycle and cell differentiation.

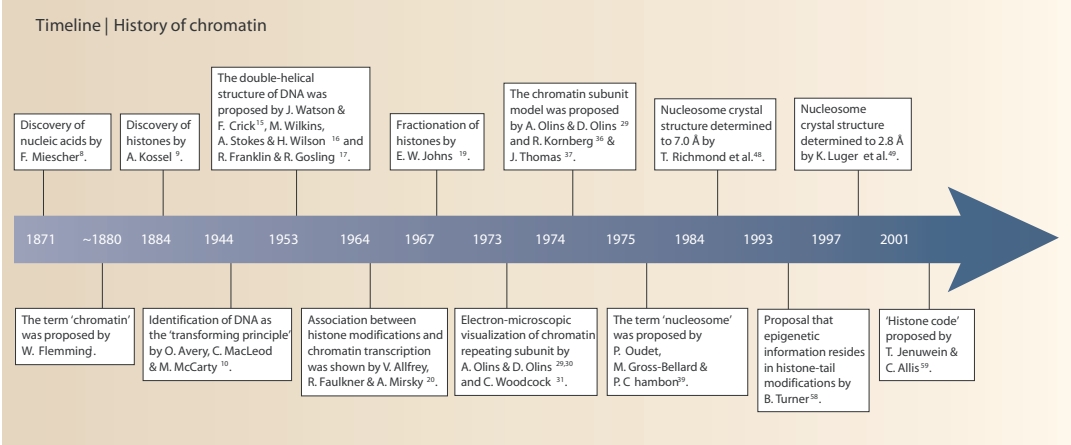


Figure 14: Overview: A brief outline of the history of chromatin. Adapted from [151].

2.4 Epigenetic Regulation

Regulatory mechanisms involving the chromatin structure offer numerous ways of epigenetic control of transcription, beyond the schemes based only on genomic sequences:

1. The chromatin fiber structure is sensitive to the intrinsic variability of mechanical properties of the linker DNA associated with its sequence (natural curvature, local increase of flexibility). In this way, it might amplify direct physical by-products of the genomic sequence giving a physical meaning to the DNA sequence beyond the genetic code and its protein translation.

2. There is an ever growing number of experimental evidences of the interplay between the chromatin fiber structure and the histone post-translational modifications.

   It suggests physical ways of reading these modifications beyond the biochemical ones (e.g. molecular recognition, recruitment of factors and co-activators).

   A first way is the electrostatic landscape modification due to the lowering of histone tail positive charge associated with some histone post-translational modifications as phosphorylation and acetylation. Histone post-translational modifications tune histone tail affinities for DNA, hence control the entry-exit angle α made by the linkers. They also affect the stacking interactions between nucleosomes.

   In consequence, they indirectly control the fiber architecture and the local mechanical constraints on DNA [74].

   Histone tail status finally tunes DNA anchoring onto nucleosomes. It thus controls DNA conformational changes and indirectly the consequences of these conformational changes on the fiber structure and on the DNA binding affinities.

The central regulatory role of the chromatin fiber and the previous remarks lead to the notion of an epigenetic information imprinted in the features of the fiber.
It is mediated by a huge variety of elementary features: DNA methylation, nucleosome gaping, linker histones (H1, H5), non-histone proteins (HMG) and all the histone tail post-translational modifications.

Speaking of information requires to specify how it is read and interpreted, i.e. how it gets a meaning. Actually, several meanings superimpose: the reading of epigenetic information, say, histone tail modifications, might be either chemical (molecular recognition and factor recruitment), physical (mechanical constraints) or kinetic (involved in hypercycles [21; 22]). Involvement of mechanisms of different nature is likely to play a key role in regulation pathways: for instance, physical mechanisms do not compete on the same footing with plain chemical reactions and might be the limiting, controlling step in the kinetic control of a pathway. A noticeable feature of these physical links, in particular mechanical and topological constraints, is their long-range nature connecting distant regions of the genome and propagating constraints, i.e. information, over a large number of base pairs.

This suggests that transcription regulation relies jointly on mechanisms, mainly physical, controlling a global transcriptional potentiation and specific biochemical pathways locally controlling a targeted activation.

2.5 Purpose of the Chromatin Structure

Besides packaging DNA there are further functions of chromatin: to strengthen the DNA, to allow mitosis and meiosis and to serve as a mechanism to control expression

and DNA replication. Numerous facts point towards an adapted functional organization of the chromatin fiber and its regulatory role in cell functions, in particular at the level of transcription.

Furthermore, the structure of chromatin during interphase is assumed to be optimized to allow easy access of transcription and DNA repair factors to the DNA while compacting the DNA into the nucleus. The structure varies depending on the access required to the DNA. Genes that require regular access by RNA polymerase require the looser structure provided by euchromatin.

It should also be noted that during mitosis, while most of the chromatin is tightly compacted, there are small regions that are not as tightly compacted [231].

These regions often correspond to promoter regions of genes that were active in that cell type prior to entry into mitosis. The lack of compaction of these regions is called *bookmarking* which is an epigenetic mechanism believed to be important for transmitting to daughter cells the "memory" of which genes were active prior to entry into mitosis. This bookmarking mechanism is needed to help transmit this memory since transcription ceases during mitosis [231].

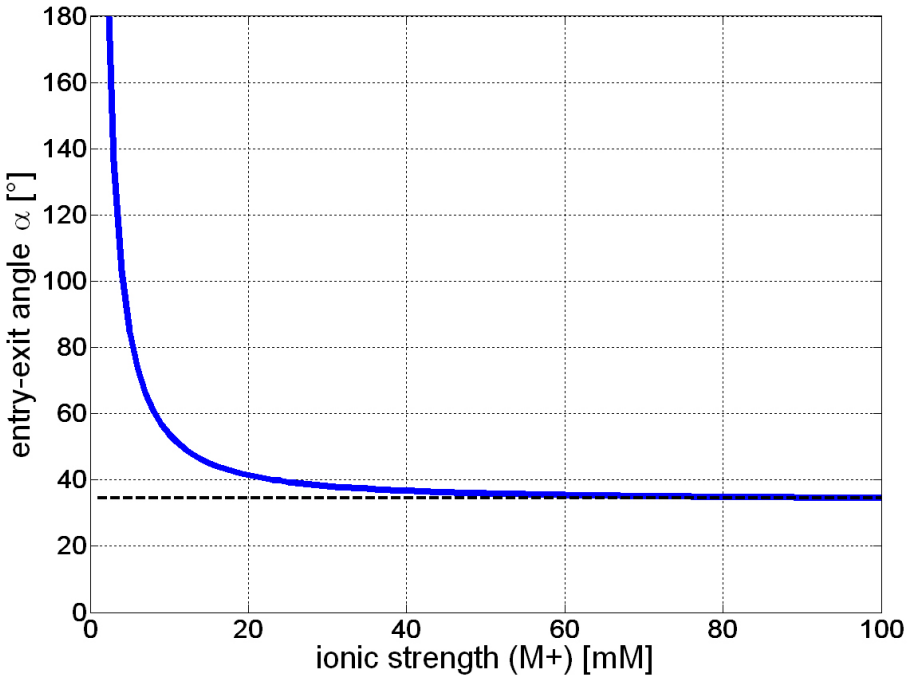


Figure 15: Fit of the mean linker entry-exit angle from EC-M (electron cryomicroscopy) images of chicken erythrocyte chromatin fibers at different ionic strengths (cf. [19]).

2.6 Salt-Concentration and the Entry-Exit-Angle

The beads-on-a-string structure can be seen clearly when chromatin is exposed to very low salt concentrations (cf. Fig. 9) and is known as the 10nm-fiber since the diameter of the core particle is 10nm.
With increasing salt concentration, i.e. heading towards physiological conditions ($c \approx$100mM), this fiber appears to thicken attaining a diameter of 30nm. The absence of the extra linker histones (H1 or H5) leads to more open structures so it is surmised that the linker histones act near the entry-exit point of the DNA. They carry an overall positive charge and bind the two strands together leading to a stem formation [19]. Increasing the salt concentration decreases the entry-exit angle α of the stem as it reduces the electrostatic repulsion between two strands (cf. Fig. 15).

Fig. 15 shows a fit ($\alpha = ax^b + c$) of different entry exit angles obtained by EC-M observations of chicken erythrocyte chromatin at different ionic strengths: The salt-induced compaction by monovalent ions (M+) leads to a strengthening of the histone stem structures and thus to a decrease of the entry-exit-angle and an accordion like reduction of the fiber length.

2.7 DNA Organization at Larger Scales

Chromosomes, the packaged DNA carriers of heredity and instructions for proper cell functioning, undergo dramatic morphological transformations during the cell-division cycle.
In metaphase which includes the alignment of chromosomes before their separation between the two daughter cells the 46 chromosomes in a human cell are condensed to such a degree that they can be observed by light microscopy as clearly separate individual entities (cf. Fig. 17).
In cells that have entered the subsequent interphase, the chromosomes partially decondense into chromosome territories [45].
The information contained in chromosomes is retrieved and acted upon in their (partly) decondensed state.
A precise understanding of how decondensed interphase chromosomes interact is important because close contact between and within chromosomes has implications for such fundamental processes as transcription and DNA damage repair; chromosome association can influence gene expression [154; 197] and misrepair of DNA double-strand breaks can promote genome instability in the form of chromosome translocations [76; 236].

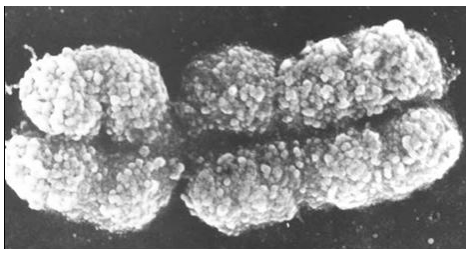


Figure 16: A chromosome during metaphase. The chromatin density is larger than 280 mg/ml. Adapted from [231].

A further key issue in biophysics is to determine if, and if so to what extent, different chromosomes interact in the nucleus.

A number of arrangements are possible. At one end, there is a complete lack of interaction: chromosomes may be contiguous without intermingling or they may be separated by interchromatin domains.
Interchromatin domains are nuclear areas mostly void of chromatin where important chromosomal transactions such as transcription, pre-mRNA splicing, DNA damage repair and DNA replication occur. At the other extreme, chromatin loops of different chromosomes can freely intermingle leading to a situation where the borders between chromosome territories and chromosome subdomains are no longer clearly defined [9].

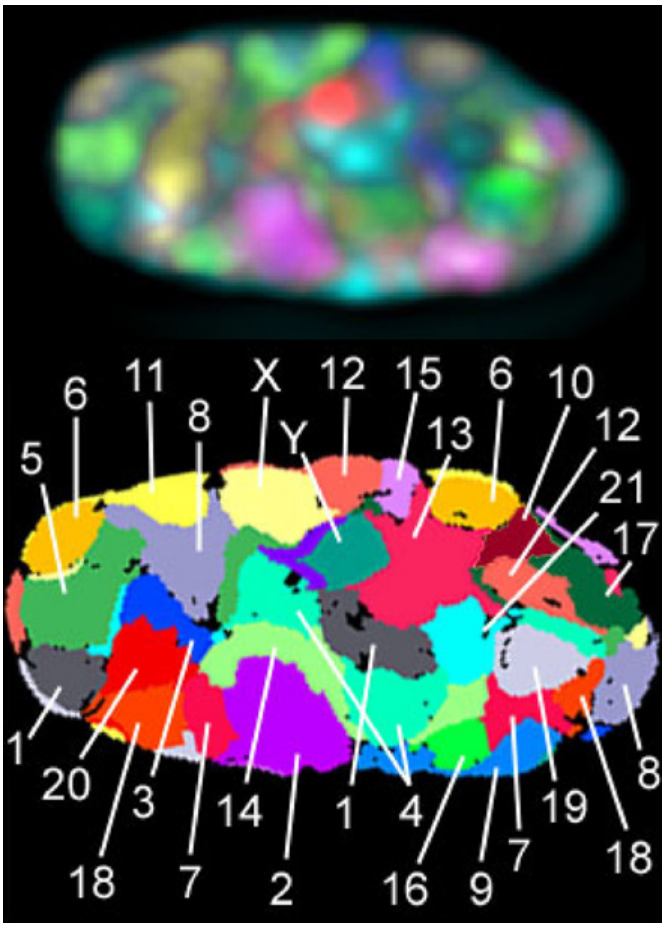


Figure 17: FISH (Fluorescence in situ hybridization) labelling of all 24 different human chromosomes (1 - 22, X, and Y) in a fibroblast nucleus, each with a different combination of in total seven fluorochromes. Shown is a mid-plane of a deconvoluted image stack which was recorded by wide-field microscopy. Bottom: False color representation of all chromosome territories visible in this mid-section after computer classification. Adapted from [28].

A way to deal with DNA organization on large scales up to whole chromosomes is supplied by the random loop model [25]. This model gives a satisfactory explanation

for the behavior of the mean square distance between two FISH markers in relation to their genomic distance. It assumes that chromatin fibers form loops on all length scales.

3 SPDM Nanostructure Examination of Chromatin

3.1 Project Background and Motivation

This chapter describes a collaboration project between the group of Prof. Cremer at the Kirchhoff-Institute for Physics (University of Heidelberg) and the group of Prof. Heermann at the Institute of Theoretical Physics (ITP) at the University of Heidelberg.

Until now it has been impossible to study the nanostructure of chromatin in eukaryotic cells by light optical techniques. While it is well-established that nuclear organization on the scale of the whole nucleus is strongly dependend on cell type [46] it is assumed that the chromatin structure on the nanometer scale is identically.

However, the presumed structure of chromatin is far too small for resolving it via conventional light optical microscopy where the resolution is limited to ≈200 nm. Currently most of the knowledge on chromatin structure using light optical techniques comes from fluorescent in situ hybridisation (FISH) [34; 144; 240] where two markers are positioned along the chromatin fiber with known genomic separation. Thereby the relationship between mean-square displacement between two such markers and the genomic distance can be evaluated.

Figure 18: The main function of fibroblasts is to maintain the structural integrity of connective tissues. Fibroblasts synthesize the extracellular matrix and collagen. They play a very important role for wound healing as well. Fibroblasts have one or two elliptical nuclei. Adapted from [231].

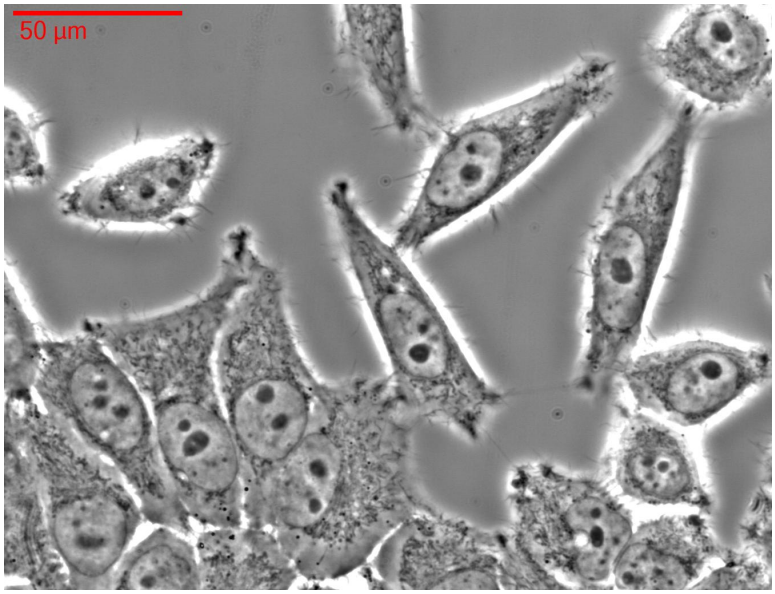


Figure 19: All HeLa cells are descended from the same tumor cells removed from Ms. Henrietta Lacks in 1951. It has been estimated that the total number of HeLa cells that have been propagated in cell culture far exceeds the total number of cells of Ms. Lacks body. Due to the world-wide sale of HeLa cells it has been claimed that Ms. Lacks is the most expensive human being that ever lived [190]. HeLa cells are called immortal since they have the ability to divide an unlimited times as long as fundamental cell survival conditions are met [127]. Adapted from [231].

The group of Prof. Cremer applied high-resolution localization microscopy to study the spatial distribution of previously EmGFP labelled H2B histones in the nuclei of two different human cell lines: Human fibroblasts (cf. Fig. 18) and HeLa cells (cf. Fig. 19). The H2B histones are part of each nucleosome in the nucleus (cf. Sec. 2). Therefore, the nucleosomes themselves have been labelled by EmGFP as well.

The measurements were done on a Spectral Precision Distance Microscopy (SPDM) setup for 2D localization of single fluorescent molecules [113; 131] which allowed to determine the lateral positions of the single molecules with localization accuracy of about 20nm. The experiments have been carried out by Rainer Kaufmann, Alexa von Bassewitz, Paul Lemmer, Manuel Gunkel, Yanina Weiland, Patrick Müller, Michael Hausmann and Roman Amberger.

Afterwards the experimental raw data was analyzed by Manfred Bohn and myself at the ITP. We used two independent approaches to reduce the noise in the data: Manfred Bohn applied a density cut-off to define a mask which eliminated regions where the average density was below the cut-off and the author used a Voronoi-diagram approach to repress the noise (cf. Sec. 3.4).

Later on, we used methods from polymer physics and condensed matter physics to further analyze the nucleosome distribution data. Although we had two very different approaches independent of each other, we came to very similar results concerning the chromatin nanostructure.

In this chapter of my thesis I will give a short overview of the project and describe the work I have done for this collaboration.

Although this project was one of the last ones during my PhD, it will be described here in one of the first chapters since it provides a basic motivation of this work. On the one hand it answers the question whether a higher-order DNA folding pattern beyond the nucleosome, i.e. chromatin, exists at all, which is still a matter of discussion but on the other hand it leaves open the interesting question why the 30nm fiber seems so hard to detect.

After providing evidence for the existence of a 30nm fiber here, a three-dimensional Monte Carlo model for chromatin will be developed in the following chapters and its nanostructure and properties will be discussed and compared with further experimental data. Moreover, some possible reasons will be given that explain why 30nm fibers seem to be that hard to find (cf. Sec. 6).

3.2 Overview

First, a short overview of the experimental setup will be given in Sec. 3.3. Then the noise repression with Voronoi diagrams will be explained (cf. Sec. 3.4) and finally the results of the data analysis will be discussed in Sec. 9.3.

Major parts of the results of this chapter have already been published [26].

Furthermore, an announcement of an invention will be made due to this project: As will be shown in Sec. 9.3 it seems to be possible to distinguish between healthy and cancer cells with this combined method of localization microscopy and theoretical analysis. So far, it is not clear whether this applies generally for all types of cells, but a difference on the chromatin nanostructure is remarkably and might be cancer-related since the spatial organization of the genome plays important roles in regulation of genes and maintenance of genome stability. It was already shown that a lot of diseases (e.g. cancer) show alterations in the spatial genome organization on a large length scale but not on the nanometer scale.

3.3 Experimental Setup

A detailed description of the Spectral Precision Distance Microscopy (SPDM) setup including a schematic overview can be found in App. C. Nevertheless, a short overview will be given below.

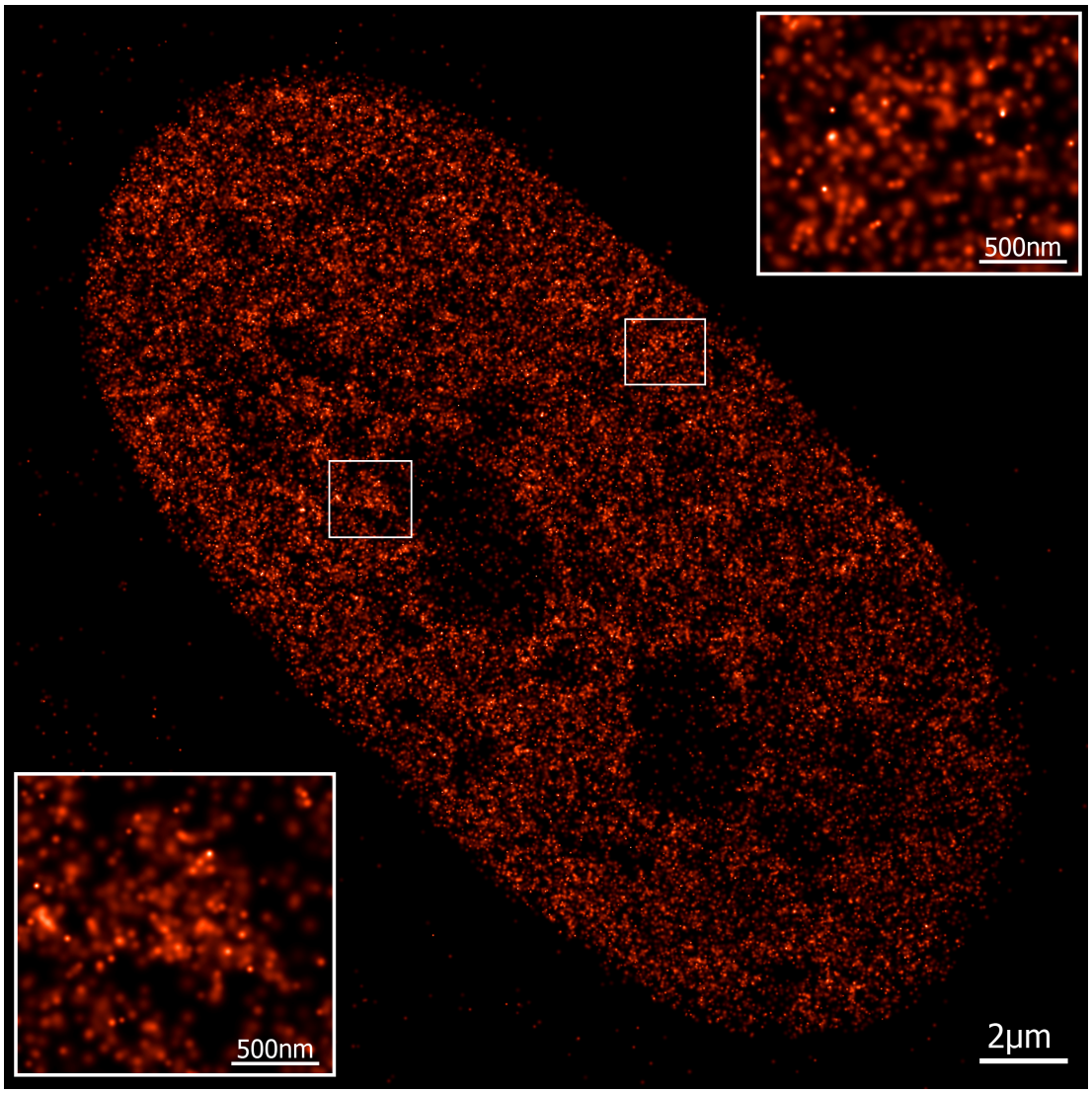


Figure 20: Localization image of histones H2B in the nucleus of a human fibroblast labelled with EmGFP. The localized fluorophores are blurred with a Gaussian corresponding to their individual localization accuracy. The inserts show the marked areas four times magnified. This image was kindly provided by Rainer Kaufmann [113].

Localization microscopy was applied to study the distribution of H2B histones and thus nucleosomes respectively chromatin in two kinds of mammalian nuclei: Human fibroblasts and HeLa cells. All measurements were done on a SPDM setup for two-dimensional localization of single fluorescent molecules [113; 131]. This method is based on a light induced reversibly bleached state of conventional fluorophores. The lateral positions of

single molecules can be determined with an accuracy of about 20 nm regime with this technique (cf. Fig. 20).

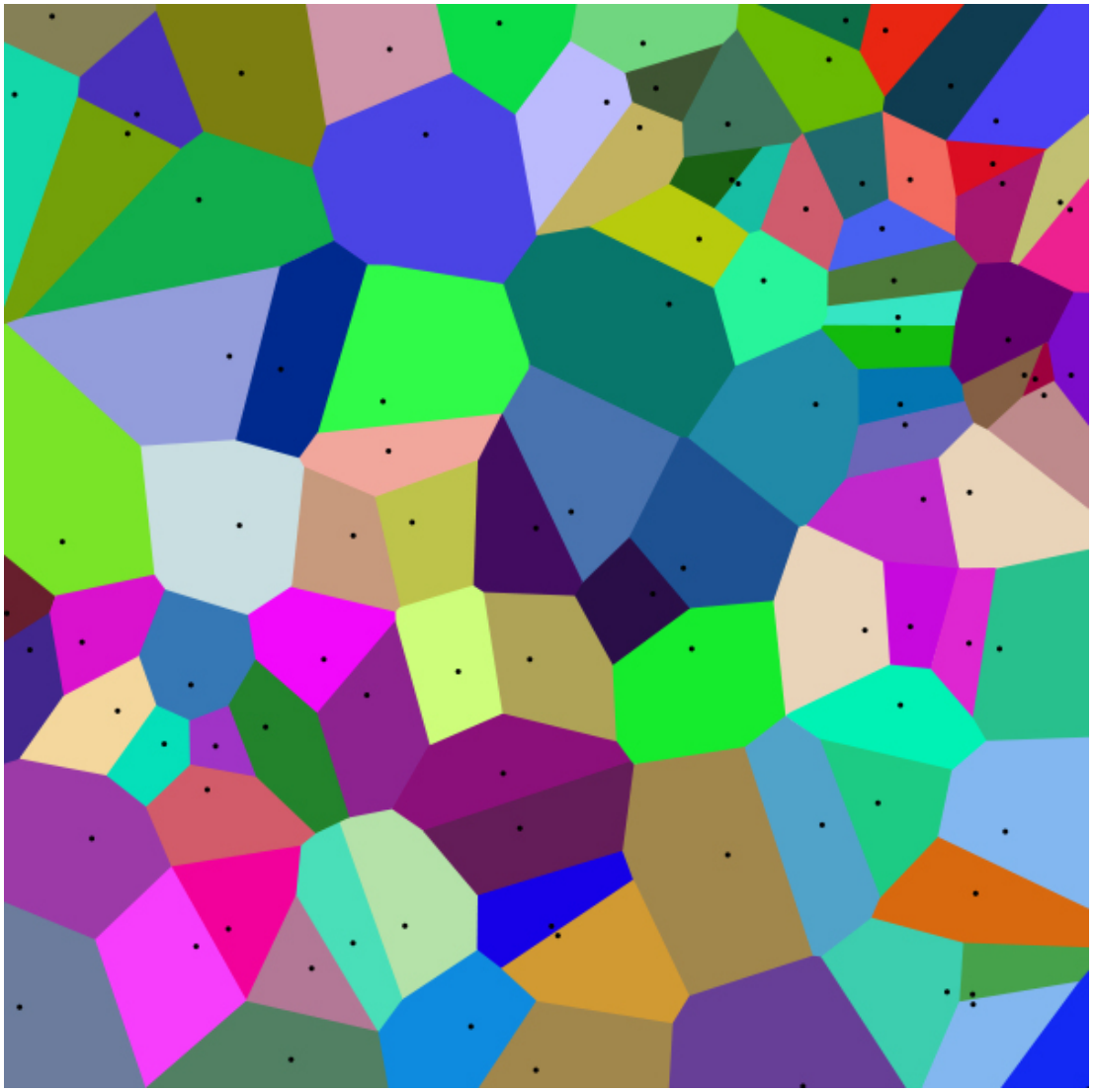


Figure 21: An example of a two-dimensional Voronoi tessellation for a random set of points. Adapted from [231].

To achieve this accuracy, a time stack of 2000 frames was recorded with a frame rate of about 18 fps. After data acquisition the count number at each pixel of the CCD chip was translated into the number of incident photons. For signals with a low signal-to-noise ratio a differential photon stack between the succeeding and the preceding frame was calculated. The high precision localization of the single molecules was done by

fitting a model function to the acquired signals, revealing the positions and localization accuracies of them.

The cultivation of the fibroblast and HeLa cells was done according to standard methods. The histones H2B of both cells lines were labelled with EmGFP by using Organell Lights (Invitrogen, Carlsbad, USA) according to the manufacturers protocol. Furthermore, HeLa cells stably expressing GFP-H2B histone fusion proteins were used. All specimens were prepared onto coverslips, fixed with 4% formaldehyde in PBS and embedded with ProLong Gold antifade reagent (Invitrogen). Altogether 32 fibroblasts and 30 HeLa nuclei have been captured. This corresponds to ≈1.5 million histone locations.

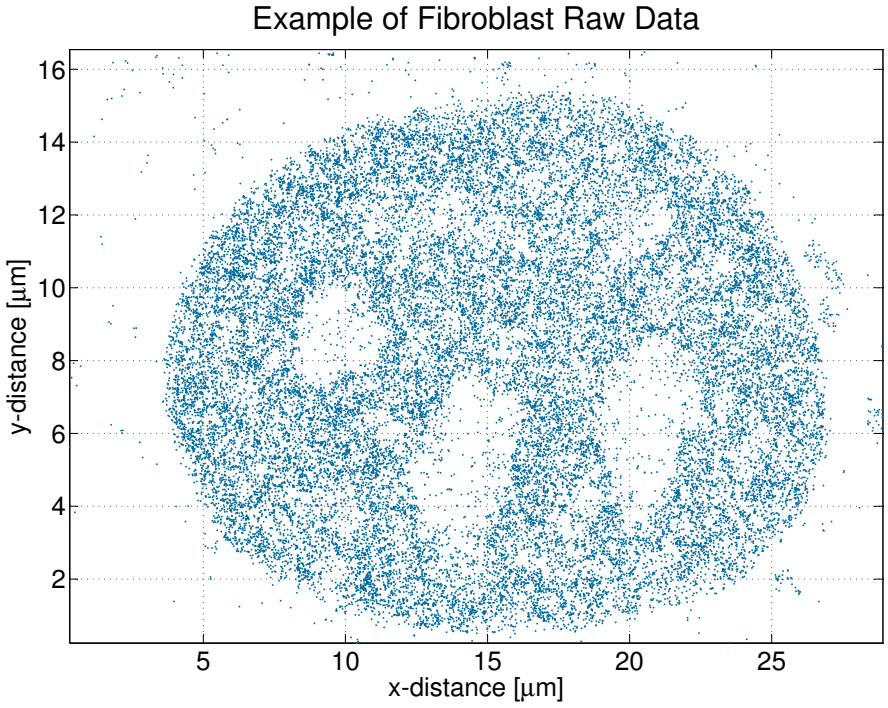


Figure 22: An Example of the raw data provided by the localization microscopy. One can see some data points outside the nucleus which are not of interest here. Furthermore, there are areas inside the nucleus which have a very low density compared to the rest of the data points. An image that shows the raw data for a HeLa nucleus can be found in App. A (cf. Fig. 165).

3.4 Data Analysis

3.4.1 Noise Repression

In this section the Voronoi tessellation approach to eliminate unwanted background signals inside the fibroblasts or HeLa nuclei nuclei is described. Fig. 22 and Fig. 23 show some raw data of Fibroblast nuclei. Images of HeLa nuclei can be found in App. A (Fig. 165 and Fig. 166).

If a set of two-dimensional coordinates is given, a Voronoi cell can be constructed for each of these points. If r is one of the given points, then the corresponding cell C_r will consist of all points in the xy-plane that are closer to r than to any other data point. The *segments* of a Voronoi tessellation are the points that are equidistant to two of the given points. Fig. 21 shows such a Voronoi tessellation for some random coordinates.

The Voronoi tessellation has some very useful properties: Its dual graph corresponds to the Delaunay triangulation of the same set coordinates and points are adjacent on the convex hull if and only if their Voronoi cells share an infinitely long segment. A Voronoi tessellation can be used to solve many frequent geometrical problems like the next neighbor problem (resp. the *all* next neighbors problem) and the largest free sphere problem. Furthermore, Voronoi diagrams are used for the analysis of protein structures [164] and many other problems [29].

The size of a Voronoi cell can be calculated easily with the Gaussian trapeze formula that describes the area of a polygon:

$$A = \frac{1}{2}\left|\sum_{i=1}^{n}(y_i + y_{i+1})(x_i - x_{i+1})\right|.$$

The Voronoi tessellation was calculated with the Qhull algorithm [12] in this work by performing a Delauney triangulation first.

For every localization image first a Voronoi diagram was generated (cf. Fig. 24) in order to eliminate the background noise of the raw data. One can see that points which lie in areas that have a very low density have huge Voronoi cells.
The large scale density fluctuations in the localization data come from regions which were occupied during the capture time for instance by the nucleole. Therefore, the number of data points in the final projection image is smaller in these regions. They will not be allowed for in the block analysis of the data.
Furthermore, one can also see that some data points lie even outside the cell nucleus. They are not of interest here and will be sorted out as well.

After the generation of the Voronoi tessellation all data points the Voronoi cell size of which is larger than a given cut-off value were taken away. This process is repeated iteratively until the convex hull of the nucleus is unveiled. During these calculations the interior of the nucleus is not changed at all.

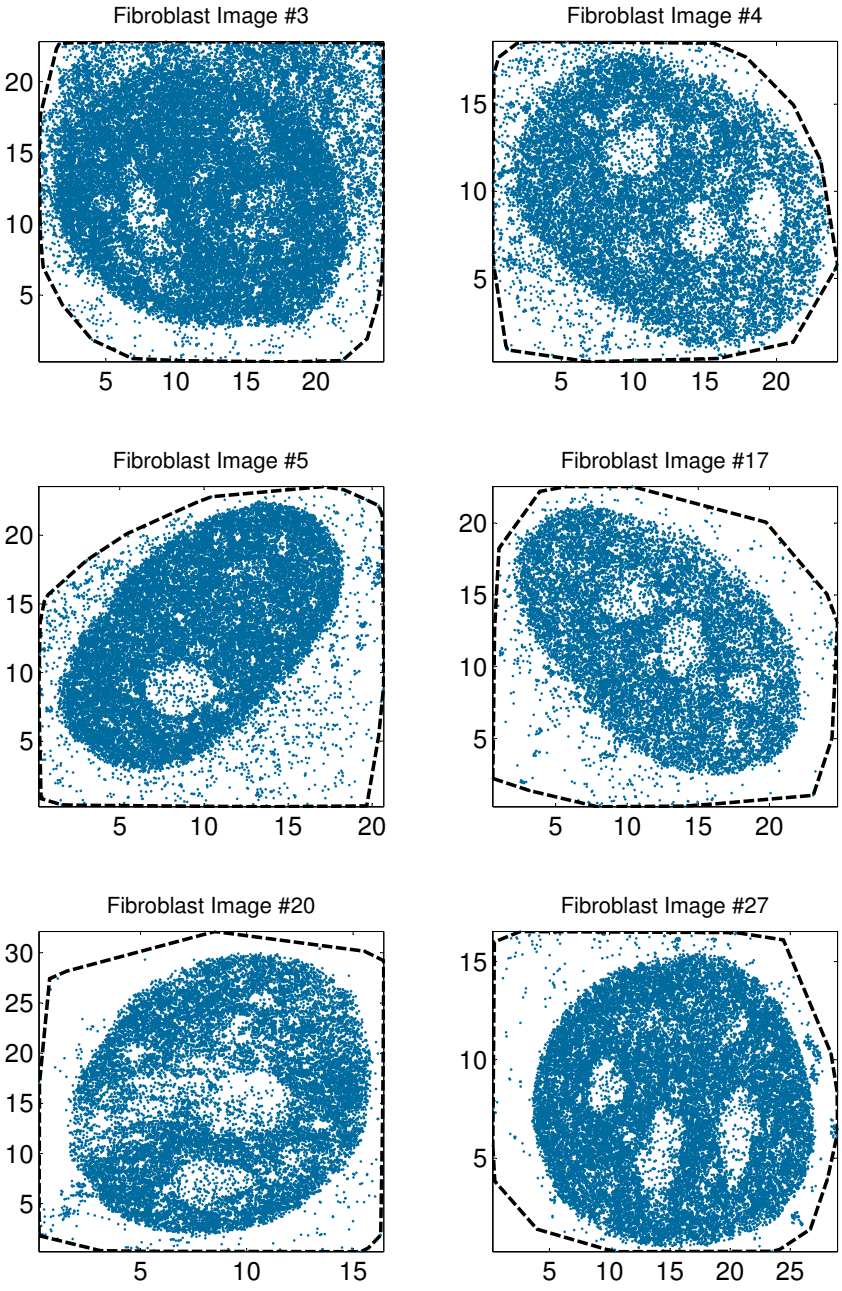


Figure 23: Several localization images of emGFP-marked H2B histones inside Fibroblast nuclei. The dashed black line represents the convex hull of the raw data. The axis units are μm. Several raw data images of HeLa nuclei can be found in App. A (cf. Fig. 166).

The iterative process gives us the advantage that data points which are close to each other (i.e. small Voronoi cell size) are not changed at all, whereas single abandoned points are taken away. Fig. 25 shows an example of the convex hull determination in the case of a fibroblast nucleus.

After the determination of the nucleus's convex hull the same iterative selection due to Voronoi cell size is done with the data points that lie inside the nucleus (cf. Fig. 26). The two steps are independent of each other to have the possibility of applying different cut-off criteria to the exterior and the interior.

With this method the fluorescent markers which are associated with the heterochromatin are sorted out, too. In the following block analysis only euchromatin shall be considered.

At the end of the described process one gets a noise reduced localization image (cf. Fig. 26).

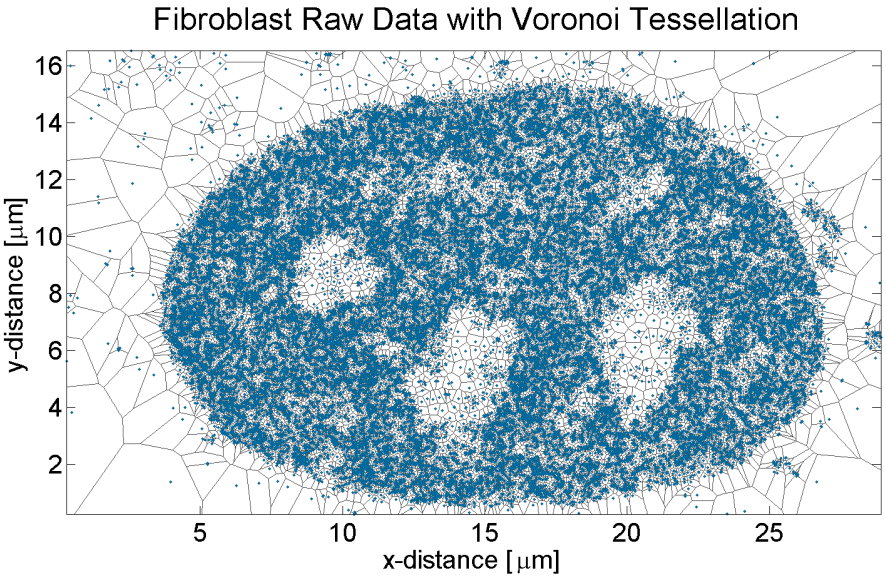


Figure 24: This figure shows a Voronoi diagram of the localization data which was illustrated in Fig. 22. One can see that points that are close to each other have small Voronoi cell sizes, whereas other points that are trailed off have large cells. For this figure Fibroblast #27 was used. A similar HeLa nucleus image can be found in App. A (cf. Fig. 167).

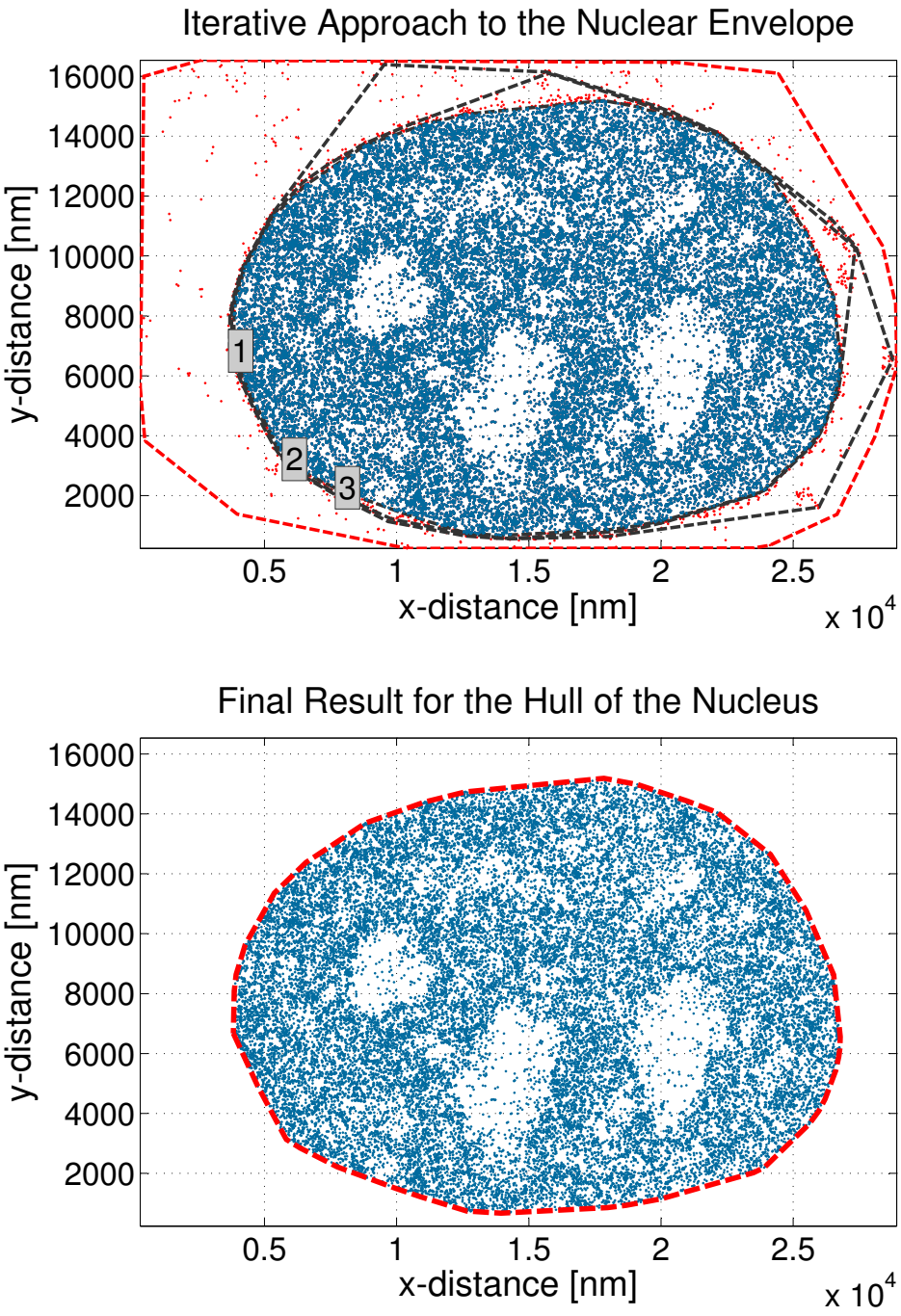


Figure 25: An example of the determination of the convex hull of a Fibroblast nucleus. Iteratively, data points are sorted out which have a too large Voronoi cell size (cf. red dots). The different iterations are marked with the labels 1,2 and 3 (top). The convex hull after every iteration is shown as well (black dashed line). A similar figure for a HeLa nucleus can be found in App. A (cf. Fig. 168).

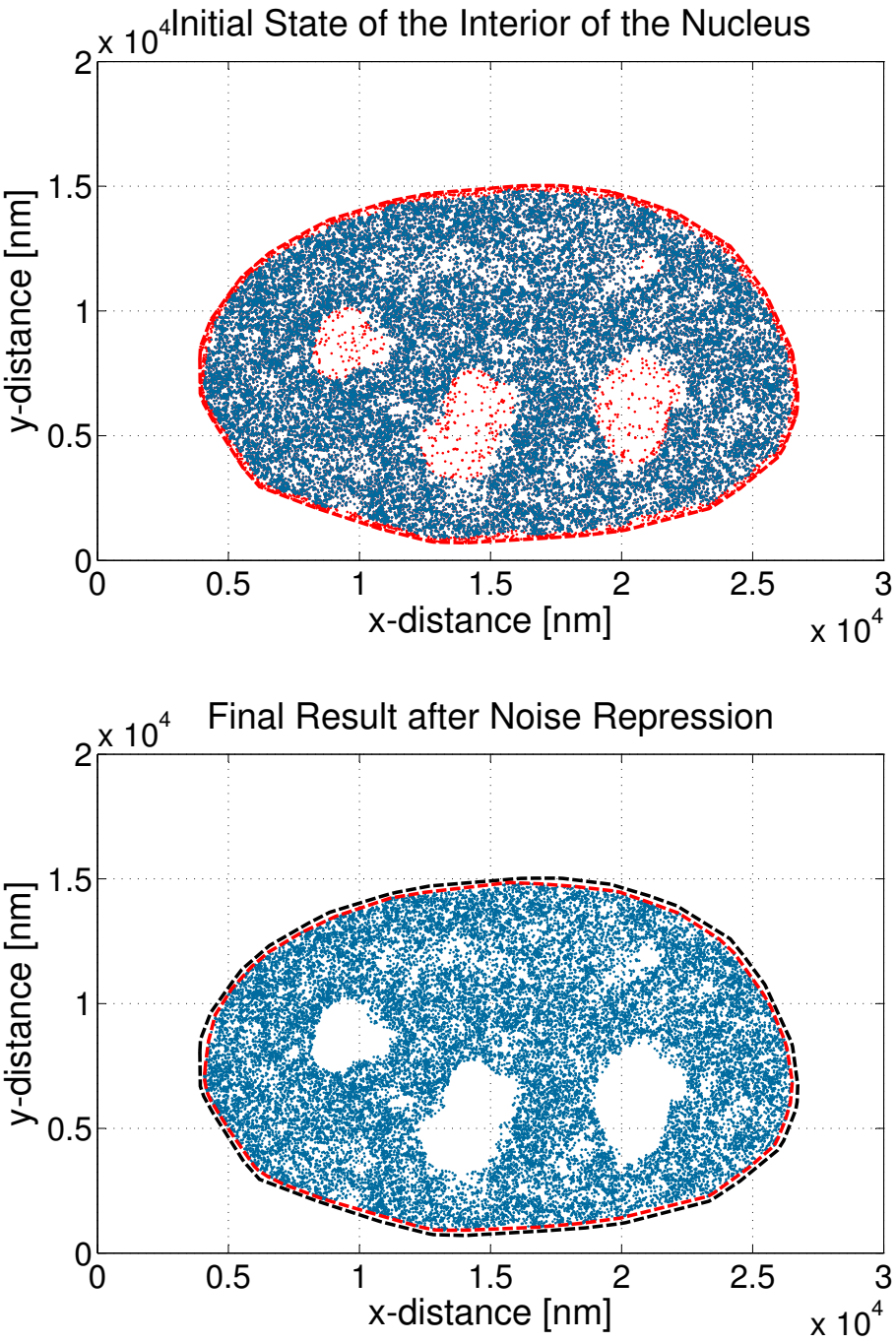


Figure 26: This figure shows the noise reduction in the interior of a Fibroblast nucleus. Again Voronoi diagrams are calculated iteratively and data points with too large cell sizes are sorted out (cf. red dots) just like the heterochromatin. A similar example for a HeLa nucleus can be found in App. A (cf. Fig. 169).

3.4.2 Block Analysis

For the nanostructure examination of the data several steps have been carried out. For every data point of a given nucleus first a square with edge length 800nm that is centered around that very data point will be cut out. Then the mean particle density of that square will be calculated.
If the particle density is equal or larger than the the mean particle density of the nucleus, then a circle of radius 400nm around the considered data point in the center will be cut out. Now, if this circle fulfils the same condition i.e. that its mean particle density is equal or lager than the average density of the nucleus, then all distances between the central point and all other points in the circle will be evaluated. A histogram of these distances is made to calculate the radial two-dimensional pair distribution function later.

The circles are necessary to avoid any bias from data points that lie in the corners of the square-shaped boxes. The density cut-off avoids bias from regions that are partially empty (e.g. by edges or projections during the capture time of the image).
Fig. 27 shows an example of the block analysis. Eventually evaluated data points are marked red. Two complete bock analysis visualizations of Fibroblasts and HeLa nuclei are shown in App. A (cf. Fig. 170 and Fig. 171).

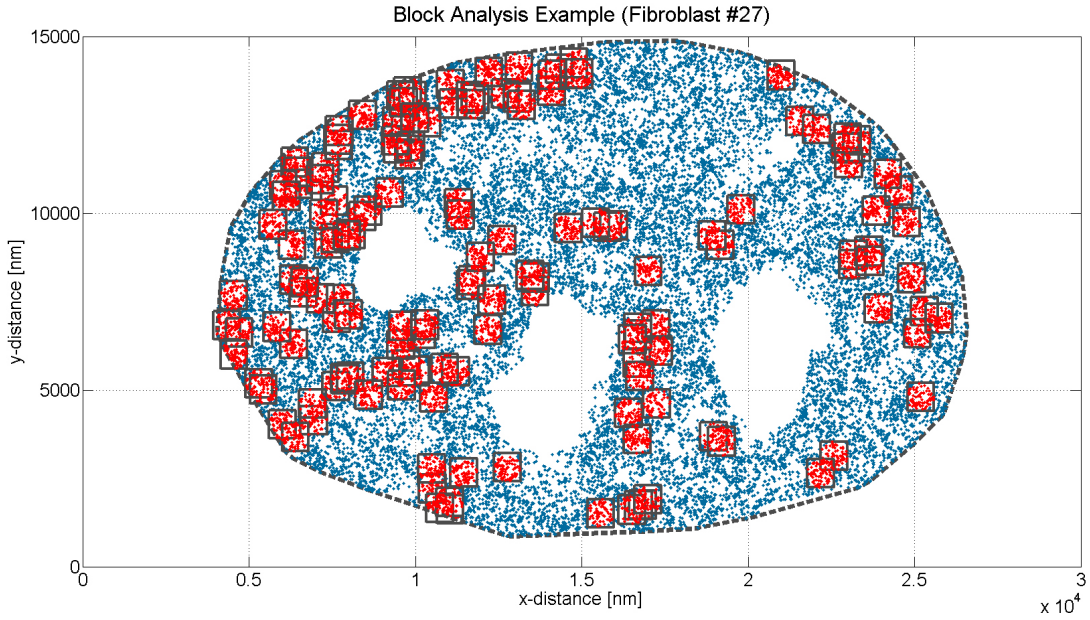


Figure 27: Example of the block analysis in the case of a Fibroblast. First squares and circles are cut out around each data point. If the mean particle density in the circle is equal or lager than the average density of the nucleus, the data points in the circle will be evaluated (red dots). A complete block analysis is shown in Fig. 170 in App. A.

3.5 Results

3.5.1 Histone H2B Density

Fig. 28 and Fig. 29 show the distribution of the number of H2B histones per nm^2 and thus the distribution of the chromatin density for a Fibroblast nucleus. Similar figures for the case of a HeLa Cell can be found in App. A again (cf. Fig. 172 and Fig. 173). One can see that no particular density distributions are visible.

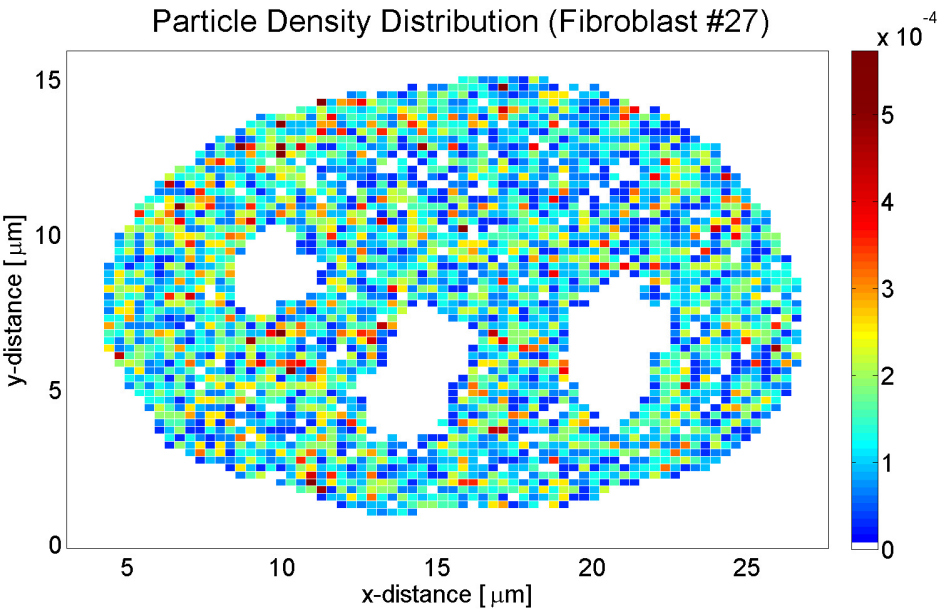


Figure 28: This figure shows the histone H2B and thus the chromatin density of a Fibroblast as a heat map. The units of the colorbar are particles per nm^2.

Fig. 30 shows the average probability density distribution of the Fibroblast nuclei. The density values on the x-axis were averaged over spheres with a radius of 100nm around each data point.

3.5.2 Global Distance Distribution

Fig. 31 shows the global distribution of distances between different markers for several Fibroblast nuclei. One can see that the distributions vary widely since the shape of the nuclei and the positions of the nucleoli varies, too. A similar figure for the case of a HeLa nucleus can be found in App. A (cf. Fig. 174).

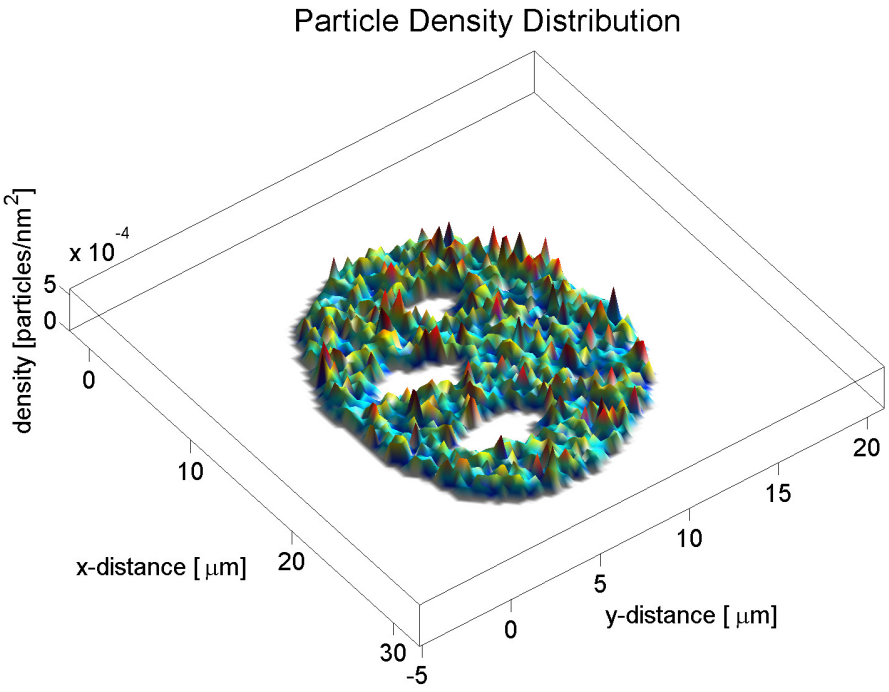


Figure 29: The spatial histone H2B density distribution of a noise-reduced SPDM image in the case of a Fibroblast.

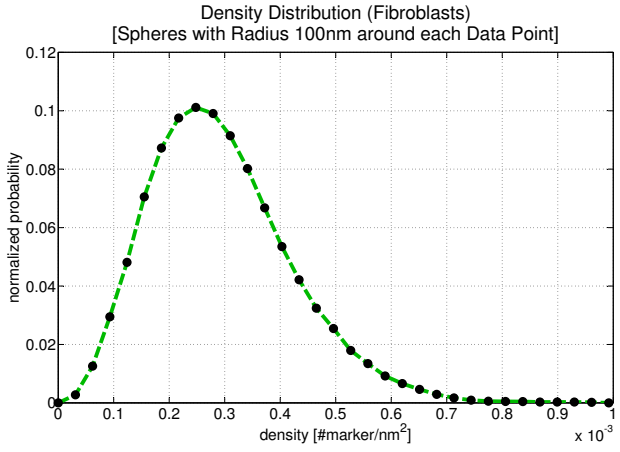


Figure 30: The average probability density distribution of the Fibroblast nuclei.

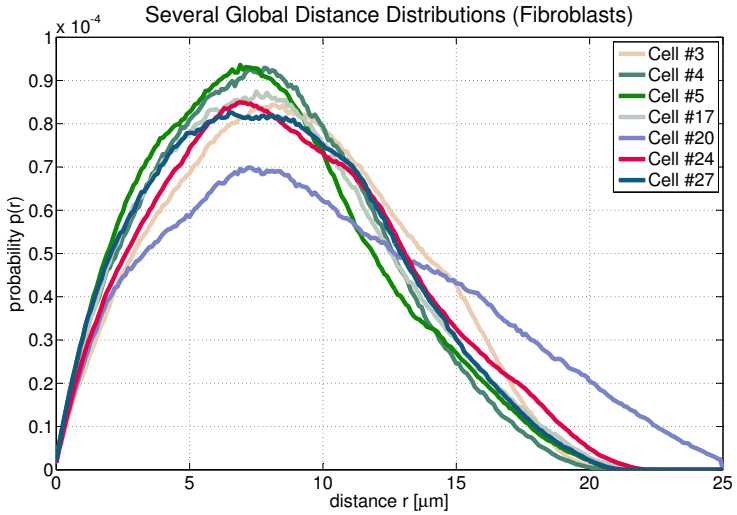


Figure 31: Several distance distributions of different Fibroblast nuclei. The distributions are different since the shape of the nuclei as well as the positions of the nucleoli are different, too.

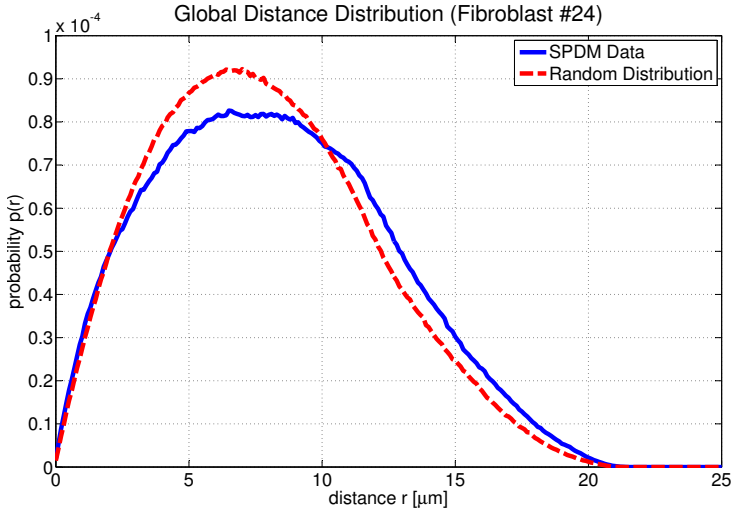


Figure 32: A comparison of a measured H2B distance distribution and a random particle distribution within an identically shaped nucleus. The differences come mainly from the influence of the nucleoli. A similar figure for a HeLa nucleus can be found in App. A (cf. Fig. 175).

Fig. 32 shows a comparison between the measured distance distribution and the distance distributions one will get, if one puts the same number of particles into an identically shaped nucleus by random. The difference between the two curves comes mainly from the influence of the nucleoli.

3.5.3 Chromatin Nanostructure

In order to examine the chromatin nanostructure the radial two-dimensional pair distribution function $g_{2D}(r)$ was used (cf. Sec. 1.8).

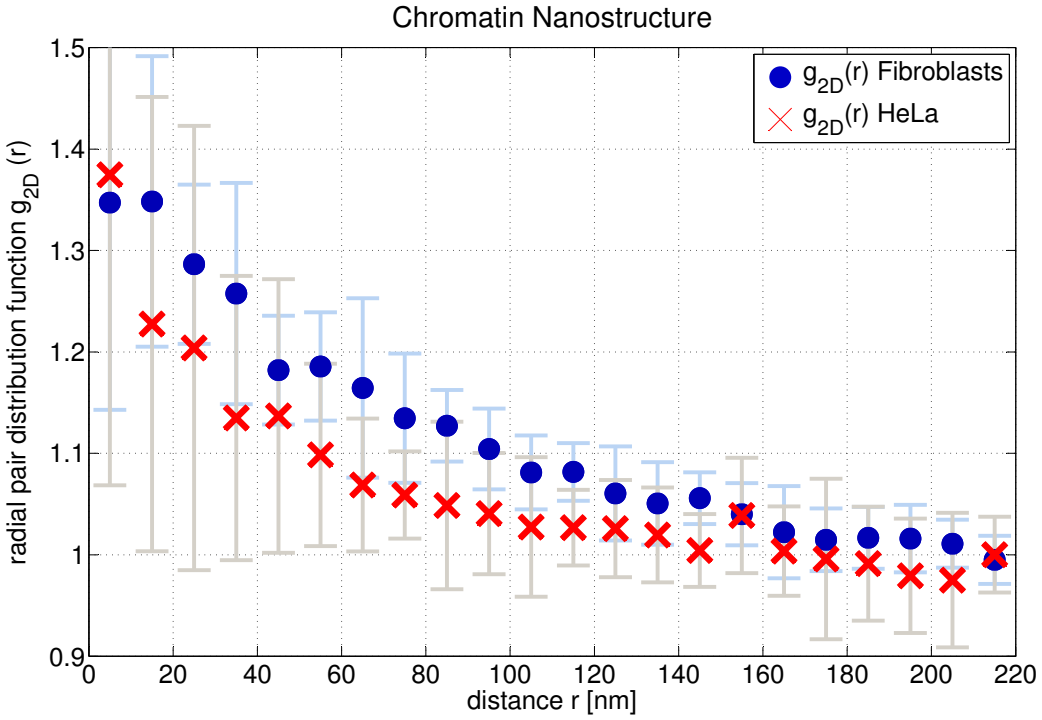


Figure 33: The radial two-dimensional pair distribution function describes the nanostructure of the emGFP-marked nucleosomes. One can clearly see some structure beyond the scale of the nucleosome since $g_{2D}(r)$ is larger than one for small distances. The differences between the different cell types were larger when the H2B histones were marked with another kit. Nevertheless, the two graphs can be clearly separated.

The pair distribution function will be proportional to the conditional probability $p(r)$ of finding a marker at a distance r, if another one is sitting at the origin. It is normalized in a way that a value of one corresponds to the mean particle density of the system. That means unbound H2B histones which are also fluorescently labelled are not expected to contribute to any structural motif since they are assumed to be uniformly distributed all over the nucleus.

For this analysis the formerly described block analysis (cf. Sec. 3.4.2) was used to find and cut out suitable spheres with a radius of 400nm. Then all distances between the center and any other marker in each sphere are evaluated to calculate $p(r)$ and it was averaged over all possible spheres in a particular nucleus (i.e. $g_{2D}(r)$ is calculated by a histogram of $\approx 10^6 - 10^7$ distances). The centers of the spheres themselves were not considered. Thus one gets one distribution function per nucleus.

The two-dimensional pair distribution function is given by

$$g_{2D}(r) = \left[\left(\frac{A}{N-1}\right)\frac{1}{\pi[2r\Delta r + \Delta r^2]}\right]\left[\frac{1}{N}\sum_{i=1}^{N}\sum_{j\neq i}\delta(r - r_{i,j})\right],$$

where Δr=10nm is the binning width for the histogram used to calculate $p(r)$, $(N-1)/A$ is the average particle density of the 400nm-sphere and N is the total number of particles the sphere contains. The factor $\pi[2r\Delta r + \Delta r^2]$ is just the area of a ring of width Δr with a mean radius of $r + \Delta r/2$.

The average pair distribution functions for HeLa and Fibroblast nuclei are shown in Fig. 33. The single distance distribution functions for each nucleus are shown in Fig. 34 and Fig. 35 as scatter plots. The latter show that the cell to cell variation of $g_{2D}(r)$ is large.

First of all, one can clearly see that a chromatin nanostructure beyond the level of the nucleosome exists: $g_{2D}(r)$ is larger than one for distances below 150nm which means that the nucleosomes are organized somehow.

Furthermore, the HeLa curve lies almost completely beneath the curve of the Fibroblasts. This is a hint that their structure on that small length scale is different. Although the nucleosomes might be organized similar inside a 30nm structure, small changes of the salt concentration or the concentration of linker histones can already change the distance distribution within a chromatin strand. This will be examined theoretically in the next chapters of this work (cf. Sec. 5 - Sec. 8).

Moreover, Sec. 7.3.3 and the following Sec. 7.3.4 will provide a direct comparison with chromatin fiber projection data from the E2A-model (cf. Fig. 98).

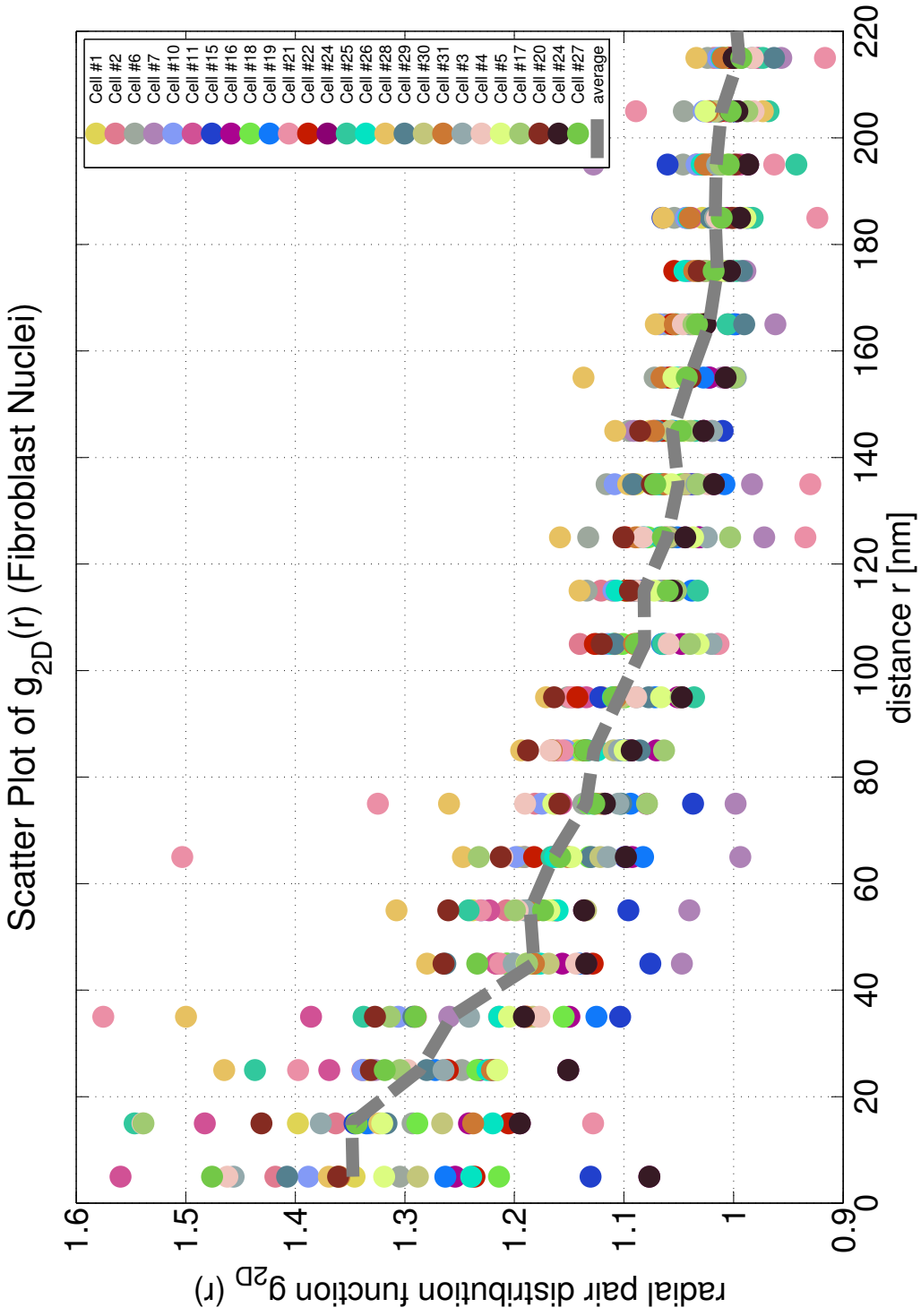


Figure 34: This figure shows all single $g^i_{2D}(r)$ distributions that contribute to the average value $g_{2D}(r)$ of the Fibroblast nuclei. One can see that the cell-to-cell variation is large.

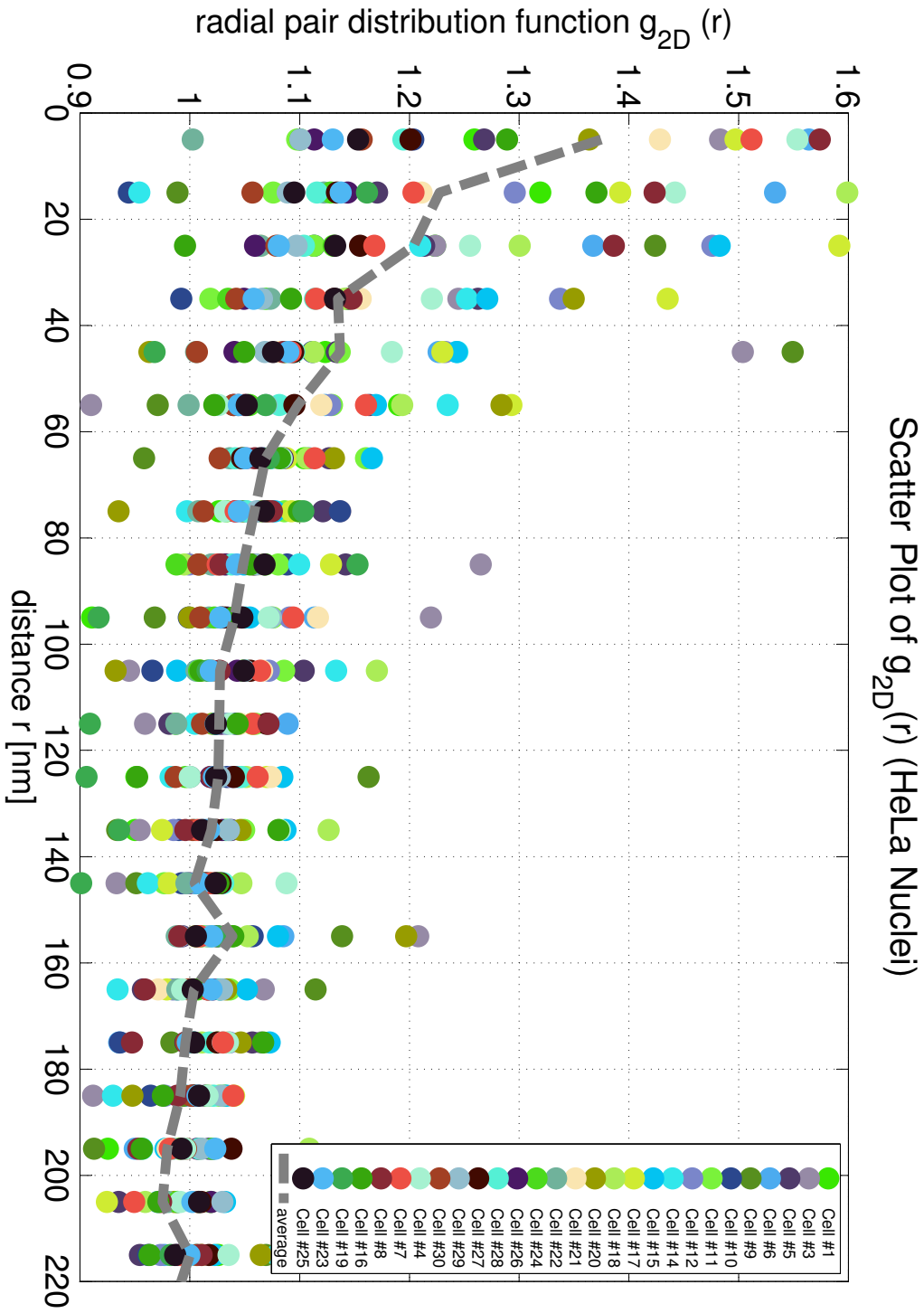


Figure 35: This figure shows all single $g^i_{2D}(r)$ distributions that contribute to the average value $g_{2D}(r)$ of the HeLa nuclei. One can see that the cell-to-cell variation is large.

4 Modelling Chromatin I - Geometry of the 30nm-Fiber

4.1 Overview

In this section the two-angle model for chromatin will be introduced and analytically described (cf. Sec. 4.3). It is the basis of the Extended Two-Angle model (E2A model) which will be used for the chromatin simulation in this work.
The geometry of chromatin in the framework of the E2A model as well as an algorithm for the generation of chromatin fibers within this model are explained in Sec. 4.4.

At the end of this chapter in Sec. 4.6 the modelling of the DNA as well as the nucleosome-nucleosome interactions will be discussed. But they will be replaced by experimental data later on (cf. Sec. 6) since the nucleosome-nucleosome interactions are very complex and in terms of interaction potentials still unknown.

The description of the E2A-model and the algorithm for the generation of chromatin fibers within this model was already published [64]. Furthermore, some calculations concerning the basic two-angle model have been parts of former publications [61; 62].

4.2 Introduction

The two-angle model was introduced by Woodcock et al. [235] to describe the geometry of the 30nm interphase chromatin fiber.
A mathematical framework for the original two-angle model as well as for the E2A-model will be given in this section. This analytical approach is the basis for the computational modelling of the chromatin fiber.

4.3 The Two-Angle Model

To address the folding problem of DNA at the level of the 30nm fiber a mathematical description for the different folding pathways is needed.
At the simplest level this can be done by three parameters: the "entry-exit-angle" α, the "torsion angle" β and the "linker length" b. One can consider these as random variables with probability distributions but for now they shall be constant - such chromatin fibers with constant values of α, β and b will be called "ideal" or "regular" in opposite to fibers with probability distributed α, β and b.

The geometric structure of the 30nm fiber can be obtained from the local geometry of a single nucleosome. The place of a certain nucleosome can be calculated by knowing the locations of the three preceding neighbor nucleosomes.

4.3.1 Basic Definitions of the Two-Angle Model

To see how single-nucleosome properties can control the fiber geometry, remember the fact that DNA is wrapped a fractional number of turns around the nucleosome (146bp $\approx$ 1.8 turns) in the case of no H1 linker. Therefore, the incoming and outgoing linker

make an angle α with respect to each other.

In the presence of the H1 histone the ingoing and outcoming linker are in close contact with each other forming a stem before they diverge (the precise value of α depends on many parameters such as salt concentration (cf. Fig. 15), degree of acetylation of the histones etc.).

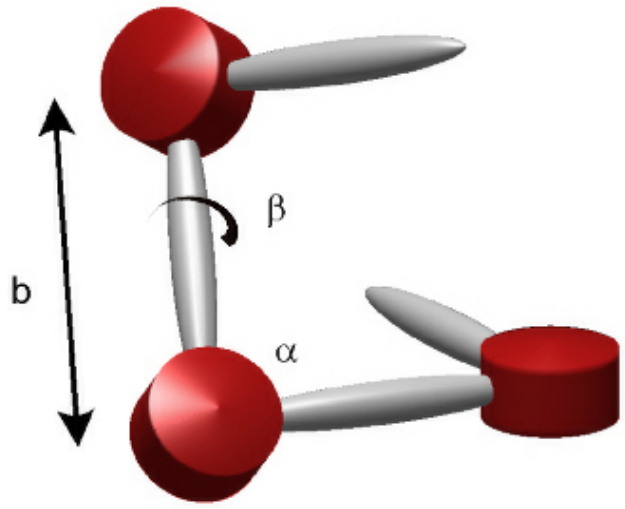


Figure 36: Basic definitions of the two-angle model: The entry-exit angle α, the linker length b, and the rotational angle β. Adapted from [61; 62].

The rotational angle β between the axis of neighboring histone octamers along the necklace (cf. Fig. 36) is a torsion angle. Because nucleosomes are rotationally positioned along the DNA the angle β is a periodic function of the linker length b with the 10bp repeat length of the helical twist of DNA as the period: $\beta(b + 10bp) \approx \beta(b)$.

There is experimental evidence that the linker length b shows a preferential quantization involving a set of values that are related by integer multiples of this helical twist [228] i.e. there is a preferred value of β. The experimental results from [228] for the nucleosome repeat length (i.e. the linker length plus the DNA content which is wrapped around the histone onctamer) will later on be used to model parameter distributions in the E2A-model.

The "linker length" b is in fact the distance between two neighboring nucleosomes which means $b \approx 2\sqrt{r^2 + \left(\frac{l}{2}\right)^2}$ (with r being the radius of the nucleosomes and l being the "true" linker length between two nucleosomes). This leads to typical values of b of about 60bp (in fact the calculated mean value of b is 63bp which comes quite close to experimentally found values of about 20nm≈ 59bp for chicken erythrocyte chromatin

(cf. [219])). The thickness of DNA fiber is $t \approx 2\text{nm} \approx 5.9\text{bp}$.

4.3.2 Mathematical Definition of the Two-Angle Model

Consider four consecutive nucleosomes (cf. Fig. 36): N_0, N_1, N_2 and $N_3 \in \mathbb{R}^3$ within the chain. N_3 is a function of $N_0, .., N_2$ by fulfilling the following conditions:

i) $\sphericalangle((N_0 - N_1), (N_2 - N_1)) = \alpha$;
ii) $\|N_2 - N_1\| = b_2, \|N_0 - N_1\| = b_1, \|N_3 - N_2\| = b_3$, with $b_1, ..., b_3 = b$;
iii) $P := \{r \in \mathbb{R}^3 \mid \exists\, \lambda,\ \mu \in \mathbb{R}, \text{ such that } r = N_1 + \lambda(N_0 - N_1) + \mu(N_2 - N_1)\}$
$P' := \{r \in \mathbb{R}^3 \mid \exists \lambda',\ \mu' \in \mathbb{R}, \text{ such that } r = N_1 + \lambda'(N_2 - N_1) + \mu'(N_3 - N_1)\}$
$\sphericalangle(P, P') = \beta$.

By straightforward considerations this leads to the following expression for N_3:

$$N_3 = \mathcal{R}^{\hat{w}}_{\beta}\, \mathcal{R}^{\hat{v}}_{\pi-\alpha} \left(N_2 + b_3 \cdot \frac{(N_2 - N_1)}{\|N_2 - N_1\|} \right)$$

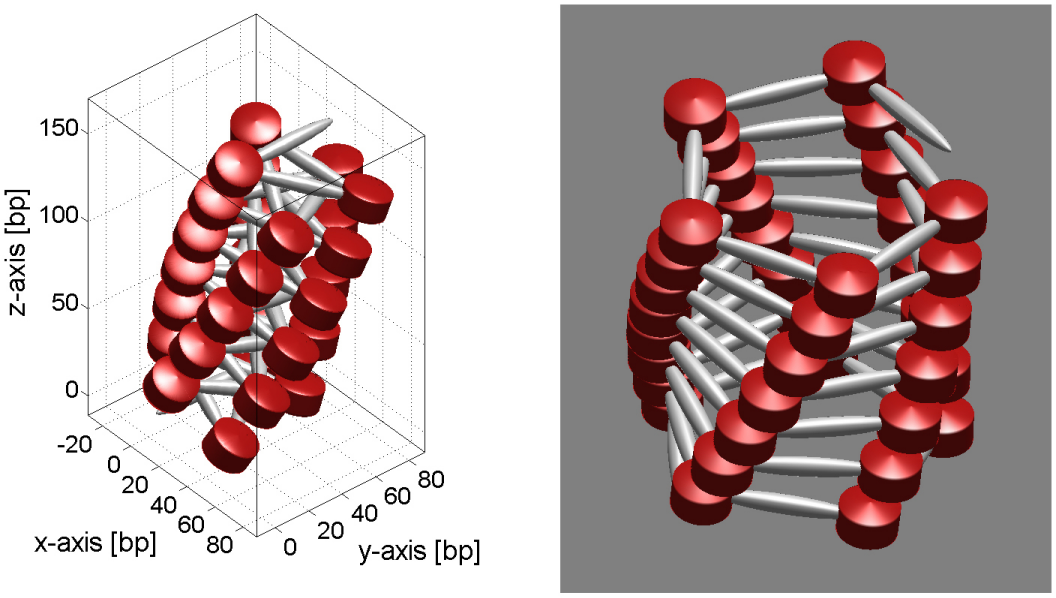
$$\hat{v} := \frac{(N_2 - N_1) \times (N_0 - N_1)}{\|(N_2 - N_1) \times (N_0 - N_1)\|}; \quad \hat{w} := \frac{N_1 - N_2}{\|N_1 - N_2\|};$$

where $\|N_2 - N_1\| = b_2$ and $\mathcal{R}^{v}_{\varphi}$ is the orthogonal rotational transformation matrix defined by the axis $v \in \mathbb{R}^3$ and the rotation angle $\varphi \in [0, 2\pi]$ (with respect to the right-hand-rule).
So the geometrical structure of the necklace is determined entirely by α, β and b.
But this model only describes the linker geometry and does not account for any forms of nucleosome-nucleosome interaction so far.
Furthermore, it assumes *straight* linkers. It is still not clear whether the linkers of the 30nm fiber at high salt concentration are straight or not. But straight linkers are likely.

Nevertheless, the two-angle model is a useful tool for understanding the 30nm structure of chromatin, although it seems to be very simple. Its simplicity is a big advantage from the numerical point of view but it is still complex enough to explain lots of phenomena concerning chromatin fibers which will be shown in detail in the following sections.

4.3.3 Mathematical Properties of the Two-Angle Model

For every ideal chromatin fiber with a certain set of values (α, β, b) it is possible to construct a spiral with radius R and a gradient m so that all the nucleosomes are located on this spiral (cf. Fig. 37).



(a) (b)

Figure 37: Two typical regular chromatin fibers in the two-angle model: (a) shows a typical **crossed-linker-fiber** under physiological conditions and (b) shows a **solenoid** like fiber with a large entry-exit-angle. The white ellipsoids are the linkers between the nucleosomes (red) which are approximately the locations of the linker DNA strings. Adapted from [61; 62].

The nucleosomes are placed along the spiral in such a way that successive nucleosomes have a fixed (Euclidean) distance b from one another (in fact there are many such spirals but the interesting spiral is the one with the largest gradient m). The parametrization of the spiral is given by

$$\gamma(t) := \begin{pmatrix} R \cdot \cos(\frac{a \cdot t}{R}) \\ R \cdot \sin(\frac{a \cdot t}{R}) \\ t \end{pmatrix}, \; t \in \mathbb{R}$$

where R is the radius and $m = \frac{1}{a}$ is the gradient of this master solenoid (which follows from $\dot{r}(0) = (0, a, 1)^T$).

Now assume a given set (α, β, b) with an infinite number of nucleosomes on this spiral. Pick one of these nucleosomes (i.e. mark the beginning of the spiral, t=0) and name its location R_0 $(= (R, 0, 0)^T)$. The locations of the successive nucleosomes are R_1, R_2, R_3, ... the locations of the proceeding nucleosome locations are R_{-1}, R_{-2}, ... The axis of the spiral is the z-axis due to the parametrization above. Now define d as the distance of successive nucleosomes along the axis of the spiral (explicit expressions for

d can be found in a former work [61]). Then R_1 will be given by $R_1 = \gamma(d)$, R_2 will be given by $R_2 = \gamma(2d)$ and so on.

Consider the following linkers between two successive nucleosomes:

$$r_0 := R_1 - R_0 = \begin{pmatrix} R \cdot \left(\cos(\frac{a \cdot d}{R}) - 1\right) \\ R \cdot \sin(\frac{a \cdot d}{R}) \\ d \end{pmatrix}$$
$$r_1 := R_2 - R_1 = \begin{pmatrix} R \cdot \left(\cos(\frac{2a \cdot d}{R}) - \cos(\frac{a \cdot d}{R})\right) \\ R \cdot \left(\sin(\frac{2a \cdot d}{R}) - \sin(\frac{a \cdot d}{R})\right) \\ d \end{pmatrix}$$
$$r_2 := R_0 - R_{-1} = \begin{pmatrix} R \cdot \left(1 - \cos(\frac{a \cdot d}{R})\right) \\ R \cdot \sin(\frac{a \cdot d}{R}) \\ d \end{pmatrix}.$$

Now b follows from the condition of fixed linker length: $\|r_0\| = b$. By straightforward calculus this leads to the following relation:

$$b^2 = 2R^2 \left(1 - \cos\left(\frac{a \cdot d}{R}\right)\right) + d^2. \tag{10}$$

Furthermore, $\cos(\pi - \alpha) = \frac{<r_0|r_2>}{\|r_0^2\|}$ leads to:

$$\cos(\pi - \alpha) = \frac{2R^2 \cos\left(\frac{a \cdot d}{R}\right)\left(1 - \cos\left(\frac{a \cdot d}{R}\right)\right) + d^2}{2R^2\left(1 - \cos\left(\frac{a \cdot d}{R}\right)\right) + d^2}. \tag{11}$$

Finally β can be calculated by evaluating $\cos(\beta) = \frac{<r_0 \times r_1 | r_2 \times r_0>}{\|r_0 \times r_1\| \cdot \|r_2 \times r_0\|}$, as the angle between two successive planes (cf. iii):

$$\cos(\beta) = \frac{d^2 \cos\left(\frac{a \cdot d}{R}\right) + R^2 \sin^2\left(\frac{a \cdot d}{R}\right)}{d^2 + R^2 \sin^2\left(\frac{a \cdot d}{R}\right)}. \tag{12}$$

So equations 10, 11 and 12 relate R, a and d to α, β and b and thus the *global* fiber geometry to these *local* variables. These equations are important for further calculations.

The inverse transformation which relates the local fiber properties to the global spiral geometry (i.e. $(\alpha, \beta, b) \Rightarrow (R, m = \frac{1}{a}, d)$) is very important and useful, too. The inverse transformations are:

$$R = \frac{b \cdot \cos\left(\frac{\alpha}{2}\right)}{2\left(1 - \sin^2\left(\frac{\alpha}{2}\right)\cos^2\left(\frac{\beta}{2}\right)\right)} \tag{13}$$

$$m = \frac{1}{a} = \frac{2\sin\left(\frac{\beta}{2}\right)\sqrt{1-\sin^2\left(\frac{\alpha}{2}\right)\cos^2\left(\frac{\beta}{2}\right)}}{\cot\left(\frac{\alpha}{2}\right)\arccos\left(2\sin^2\left(\frac{\alpha}{2}\right)\cos^2\left(\frac{\beta}{2}\right)-1\right)} \tag{14}$$

$$d = \frac{b\sin\left(\frac{\beta}{2}\right)}{\sqrt{\csc^2\left(\frac{\alpha}{2}\right)-\cos^2\left(\frac{\beta}{2}\right)}} \overset{\beta\ll 1}{\approx} \frac{b\cdot\beta}{2\sqrt{\csc^2(\frac{\alpha}{2})-1}} + o(\beta)^3. \tag{15}$$

These three equations relate the local fiber geometry to the global properties of the associated spiral.

4.4 The E2A Model

In the following the two-angle model will be extended by introducing a parameter (namely d, cf. Fig. 39 for an illustration) for the orthogonal distance between the DNA strands in front of the nucleosome. This accounts for the fact that the DNA gains height along the nucleosome axis while it is wrapped around the histone octamer.

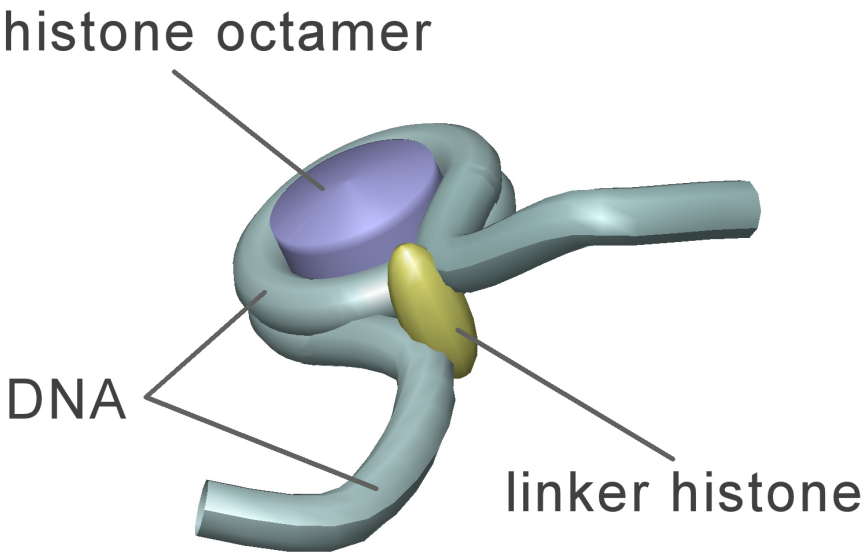


Figure 38: A single nucleosome in the E2A-model: The histone octamer is modelled as a flat cylinder (purple). The light-blue tube represents the DNA and the linker histone that sits in front of the complex is marked yellow. The nucleosome has a diameter of 7nm and a height of 5.6nm. The diameter of the DNA is 2.2nm (cf. Tab. 1). It has been shown that the excluded volume of the histone complex plays an important role for the stiffness of the chromatin fiber [146] and for the topological constraints during condensation/decondensation processes [13].

Furthermore, the cylindrical excluded volume of the nucleosomes as well as the H1 histones which fix the DNA linkers in front of the nucleosomes will be factored in. The H1 histones themselves will later be taken to be random variables to allow for possible missing linker histones. Thus one can later on study H1/H5 skips and their influence on properties of chromatin.
The cylindrical instead of the spherical excluded volume will show some interesting effects in connection with the phase diagram of chromatin.

At first the basic notations for the formulation of the *Extended Two-Angle model* ("**E2A**-model") will be presented. After that an algorithm for the generation of chains within the model will be given.

4.4.1 Basic Notations of the E2A Model

At first some basic notations must be fixed to use for the formulation of the model. The nucleosomes will be characterized by the centers $N_i \in \mathbb{R}^3$ and the orientations $\hat{p}_i \in \mathbb{R}^3$ of the nucleosomes with $i = 0, ..., N$ and $\|\hat{p}_i\| = 1$. N is the length of the fiber. The linkers between the centers of two nucleosomes will be denoted by $b_i := N_i - N_{i-1}$ with $i = 1, ..., N$. The length $\|b_i\|$ of the linkers will be a further input parameter of the model (opposite of the *direction* $b_i \in \mathbb{R}^3$ of the linkers). Furthermore, the entry-exit-angle $\alpha_i \in [0, \pi]$ between two consecutive linkers is defined by $\alpha_i := \sphericalangle(-b_i, b_{i+1})$ with $i = 1, ..., N-1$ and the rotational angle $\beta_i \in [0, \pi]$ between two consecutive orientations is given by $\beta_i := \sphericalangle(p_{i-1}, p_i)$ with $i = 1, ..., N$.

Moreover, h_i represents the distance along the orientational axis $\hat{p}_{i-1}$ from N_{i-1} to N_i due to the spatial discrepancy between in- and outgoing DNA strand. h_i can be expressed by the orthogonal distances d_i which the DNA covers by wrapping up itself around the histone complexes: $h_i = \frac{1}{2}(d_{i-1} + d_i)$ with $i = 1, ..., N$.

Thus a single chromatin strand within the two-angle-model is characterized by the following set of variables

$$(\{\alpha_i\}_{i\in\{1,..,N-1\}}, \{\beta_i\}_{i\in\{1,..,N\}}, \{h_i\}_{i\in\{1,..,N\}}, \{\|b_i\|\}_{i\in\{1,..,N\}}).$$

The general rotational matrix $\mathcal{R}$ around an axis $\hat{a} = (a_1, a_2, a_3)^t$ (with $\|\hat{a}\| = 1$) by an angle γ with respect to the right-hand rule will be used in the following. It is given by:

$$\mathcal{R}^{\gamma}_{\hat{a}} = \begin{pmatrix} \cos\gamma + a_1^2(1-\cos\gamma) & a_1a_2(1-\cos\gamma) - a_3\sin\gamma & a_1a_3(1-\cos\gamma) + a_2\sin\gamma \\ a_2a_1(1-\cos\gamma) + a_3\sin\gamma & \cos\gamma + v_2^2(1-\cos\gamma) & a_2a_3(1-\cos\gamma) - a_1\sin\gamma \\ a_3a_1(1-\cos\gamma) - a_2\sin\gamma & a_3a_2(1-\cos\gamma) + a_1\sin\gamma & \cos\gamma + a_3^2(1-\cos\gamma) \end{pmatrix}.$$

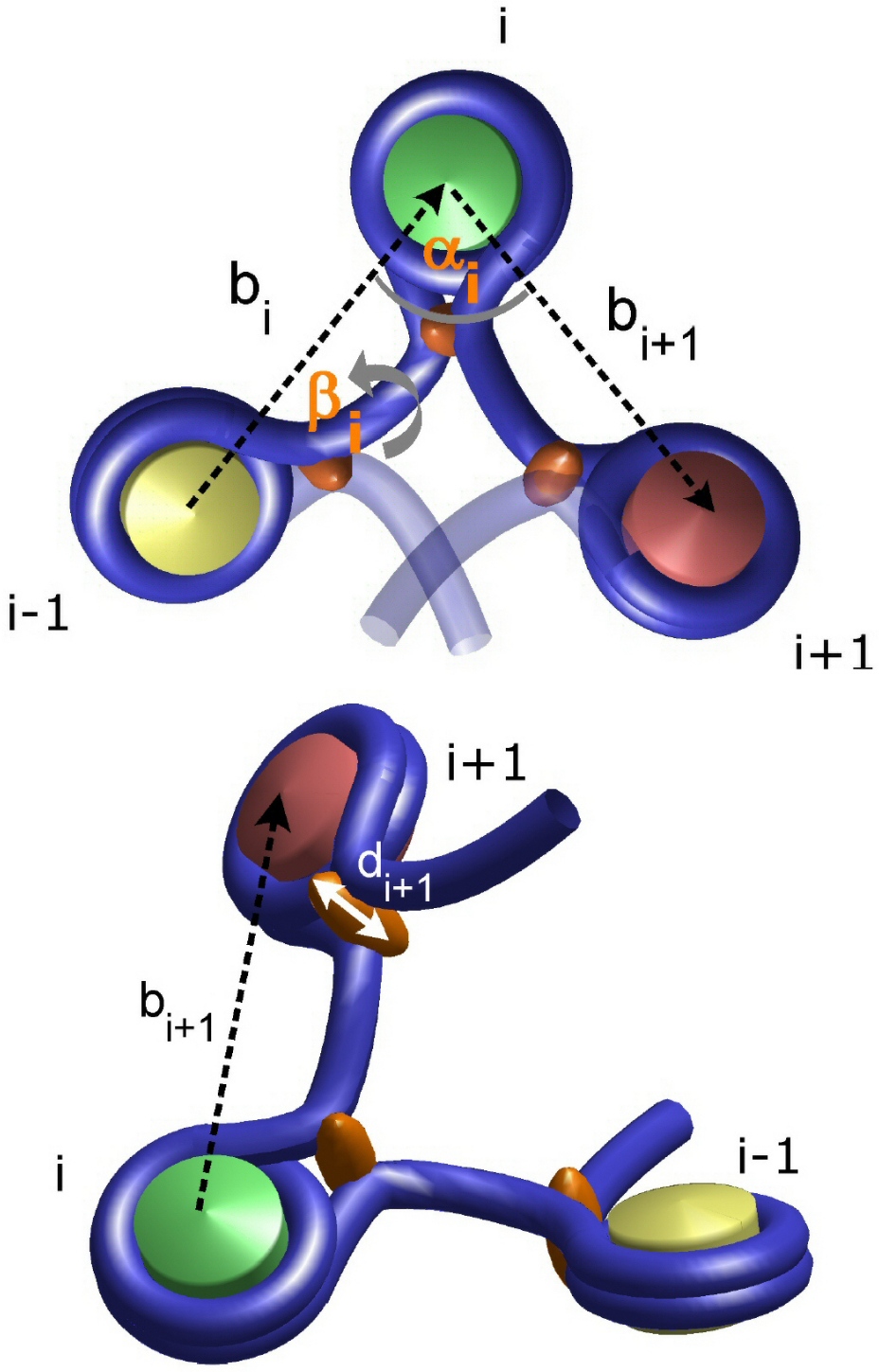


Figure 39: The figure shows the basic parameters of the E2A model: The entry-exit-angle α_i, the rotational angle β_i, the linker length b_i and the pitch d_i. The DNA linkers are fixed in front of the nucleosome by the H1 histone. A large entry-exit-angle was chosen here to make the visualization clear.

4.4.2 Definition of the Two-Angle Model

A chromatin fiber within the framework of the extended two-angle model has to fulfil the following conditions for all $i \in \{1, ..., N\}$:

$$
\begin{aligned}
&i) \quad \sphericalangle(-b_i, b_{i+1}) = \alpha_i \Leftrightarrow \cos(\alpha_i) = \frac{\langle -b_i, b_{i+1}\rangle}{\|b_i\|\|b_{i+1}\|} \\
&ii) \quad \sphericalangle(\hat{p}_{i-1}, \hat{p}_i) = \beta_i \Leftrightarrow \cos(\beta_i) = \langle \hat{p}_{i-1}, \hat{p}_i\rangle \\
&iii) \quad \|N_i - N_{i-1}\| = \|b_i\| \\
&iv) \quad \langle \hat{p}_{i-1}, b_i\rangle \hat{p}_{i-1} = h_i = \frac{1}{2}(d_{i-1} + d_i).
\end{aligned}
$$

These are illustrated in Fig 39.
The first condition adjusts the entry-exit angle of nucleosome i to the given parameter α_i. The second condition does the same for the rotational angle due to the DNA twist from nucleosome $i-1$ to nucleosome i. The third condition fixes the distance of the two consecutive nucleosomes $i-1$ and i. The last condition adjusts the orthogonal distance along the local chromatin axis between the nucleosomes $i-1$ and i.

This set of equations is much more difficult to solve than 4.3.2 since there is one additional boundary condition now. Furthermore, a solution does not necessarily exist in any case (cf. Eq. 23).

4.5 Construction of the Fiber

The construction of the fiber can be done by using an iterative process. A further part of the model is the presence of a H1 histone which is assumed to be present with probability p.

4.5.1 Start of the Iteration

The first nucleosome center and its orientation are arbitrary. One choice would be:

$$
N_0 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad \hat{p}_0 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.
$$

The following vectors fulfil the conditions of the two angle model for the second nucleosome location and its orientation:

$$
N_1 = N_0 + \sqrt{\|b_1\|^2 - h_1^2} \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix} + h_1 \hat{p}_0 = \begin{pmatrix} -\sqrt{\|b_1\|^2 - h_1^2} \\ 0 \\ h_1 \end{pmatrix}
$$

and

$$
\hat{p}_1 = \mathcal{R}_{\hat{a}}^{\beta_1} \hat{p}_0 \quad \text{with} \quad \hat{a} = (1, 0, 0)^t.
$$

4.5.2 Calculation of N_{i+1}

Now one can calculate N_{i+1} and $\hat{p}_{i+1}$ in dependence on N_i, N_{i-1}, $\hat{p}_i$ and $\hat{p}_{i-1}$. We define v_i to be the linker vector that connects N_i with N_{i-1} projected onto a arbitrary plane orthogonal to p_i.

$$v_i := -b_i + \langle \hat{p}_i, -b_i \rangle \hat{p}_i.$$

Then one can construct the projection of the linker starting at N_i to N_{i+1} onto the very same plane.

$$w_i := \mathcal{R}^{\alpha_0}_{\hat{p}_i} \sqrt{b_{i+1}^2 - d_{i+1}^2} \left(\frac{v_i}{\|v_i\|} \right)$$

Thus one gets the projection of the location N_{i+1} which one would like to know. Unfortunately the rotational angle α_0 is still unknown and must yet be calculated (cf. Sec. 4.5.3 and Eq. 22).

With

$$v_i' := \mathcal{R}^{\alpha_0}_{\hat{p}_i} \sqrt{b_{i+1}^2 - d_{i+1}^2} \left(\frac{v_i}{\|v_i\|} \right) + h_{i+1} \hat{p}_i \tag{16}$$

one finally gets the location of nucleosome $i+1$ by

$$\begin{aligned} N_{i+1} =& N_i + v_i' \\ =& N_i + \mathcal{R}^{\alpha_0}_{\hat{p}_i} \sqrt{b_{i+1}^2 - d_{i+1}^2} \left(\frac{-b_i + \langle \hat{p}_i, -b_i \rangle \hat{p}_i}{\| - b_i + \langle \hat{p}_i, -b_i \rangle \hat{p}_i \|} \right) + h_{i+1} \hat{p}_i. \end{aligned}$$

4.5.3 Calculation of α_0 and $\hat{p}_{i+1}$

α_0 is the angle between the projections of b_{i+1} and $-b_{i-1}$ onto an arbitrary plane orthogonal to $\hat{p}_i$. One needs to calculate the dependence of this projected entry-exit-angle α_0 on the actual entry-exit-angle α because at the end one wants to set α by tuning in α_0.
Note that α_0 was used as entry-exit-angle in some other publications [183; 184] but in this work it denotes only the projection of the real entry-exit-angle α.

Using the law of cosine one gets

$$l^2 = b_i^2 + b_{i+1}^2 - 2 b_i b_{i+1} \cos(\alpha) \tag{17}$$

for the distance l between N_{i+1} and N_{i-1}.

Now an affine transformation T to a new coordinate system $(x, y, z) \xrightarrow{T} (x', y', z')$ is applied in order to get a second relation for l. One can shift the origin to N_i and rotate the old coordinate system so that $\hat{p}_i$ corresponds to the new z-axis. Furthermore, the new x-axis has to coincide with the projection of $-b_i$ onto any plane orthogonal to $\hat{p}_i$.

The distance l is not changed by this transformation so obviously,

$$l^2 = \|b_i + v_i'\|^2 = \|b_i' + v_i''\|^2 \tag{18}$$

with

$$b_i \xrightarrow{T} b_i' = \begin{pmatrix} \sqrt{b_i^2 - \langle \hat{p}_i, -b_i \rangle^2} \\ 0 \\ \langle \hat{p}_i, -b_i \rangle \end{pmatrix} \tag{19}$$

and

$$v_i' \xrightarrow{T} v_i'' = \begin{pmatrix} \cos(\alpha_0)\sqrt{b_{i+1}^2 - h_{i+1}^2} \\ \sqrt{b_{i+1}^2 - h_{i+1}^2 - \left(\cos(\alpha_0)\sqrt{b_{i+1}^2 - h_{i+1}^2}\right)^2} \\ h_{i+1} \end{pmatrix}. \tag{20}$$

Equations 18, 19 and 20 lead to

$$l^2 = b_{i+1}^2 + b_i^2 - 2h_{i+1}\langle \hat{p}_i, -b_i \rangle - 2\cos(\alpha_0)\sqrt{b_i^2 - \langle \hat{p}_i, -b_i \rangle^2}\sqrt{b_{i+1}^2 - h_{i+1}^2}. \tag{21}$$

By comparing Eq.17 and Eq.21 one gets eventually

$$\cos(\alpha_0) = \frac{b_i b_{i+1}\cos(\alpha) - h_{i+1}\langle \hat{p}_i, -b_i \rangle}{\sqrt{b_{i+1}^2 - h_{i+1}^2}\sqrt{b_i^2 - \langle \hat{p}_i, -b_i \rangle^2}} \tag{22}$$

with the boundary condition

$$\alpha_0 > \alpha_{min} = \operatorname{acos}\left(\frac{(h_{i+1} + \|\langle \hat{p}_i, b_i \rangle\|)^2 - b_{i+1}^2 - b_i^2}{-2b_i b_{i+1}}\right) \tag{23}$$

due to non-vanishing d_i and d_{i+1}.

The calculation of N_{i+1} is complete since one knows now the dependence of α_0 on α. Therefore, one can use Eq.16 to determine N_{i+1}. But one still has to calculate the orientation p_{i+1} of nucleosome N_{i+1}. Due to the fixation of the in- and outgoing DNA strand by the H1 histones this orientation can be calculated by a rotation around the following normalized axis $\hat{a}$:

$$\hat{a} := \frac{-b_{i+1} - \langle p_i, -b_{i+1} \rangle \hat{p}_i}{\|\hat{a}\|}.$$

$\hat{p}_{i+1}$ then follows by a rotation of $\hat{p}_i$ around this axis:

$$\hat{p}_{i+1} = \mathcal{R}_{\hat{a}}^{\beta_{i+1}} \hat{p}_i.$$

This calculation supplies the basis for an algorithm in order to generate chromatin fibers in the framework of the E2A-model. Note that the four local model parameters (i.e. α_i, β_i, d_i and b_i) still depend on their nucleosome number i. Thus one can even use this iterative construction to generate fibers with parameters which obey a given probability density function. Later on this feature will be used.

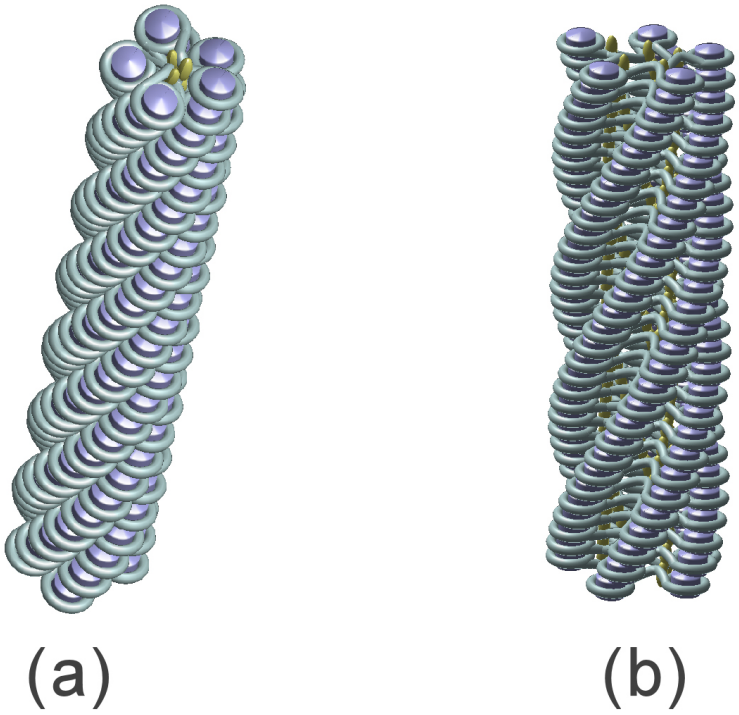


Figure 40: This figure shows two examples of regular fibers generated with the E2A-model: A crossed-linker fiber with a small entry-exit-angle (a) and a solenoid like fiber with a large entry-exit-angle (b). In the latter case one can see that the DNA linkers have to bend which leads to high energies ($\approx$ 10kT) due to the large charge of the DNA. It can be explained if one takes the charged histone tails into account.

4.5.4 DNA Trajectory

Here a short calculation of the DNA trajectory (cf. Fig. 38) around the histone octamer is given. The start of the trajectory is denoted by $s \in \mathbb{R}^3$. An angle $\phi(t)$ will be used for the rotation of the DNA around the cylinder. It is defined by

$$\phi(t) = 2\pi \; 1.65 \; t$$

with $t \in [0, 1]$. Then

$$s = N_i - \frac{d}{e}\hat{p}_i + h$$

which leads to the only unknown variable, namely the auxiliary vector h.

Then one gets the local DNA trajectory $q(t)$ as

$$q(t) = N_i - \frac{d}{2}\hat{p}_i + d\ t\ \hat{p}_i + \mathcal{R}_{\hat{p}_i}^{\phi(t)} h$$

The start and end positions $q(0) = s$ and $q(1)$ can be calculated in dependence on p_{i-1}, p_i, p_{i+1} and N_{i-1}, N_i, N_{i+1}.

4.6 Potentials and Interactions

As mentioned before the nucleosome-nucleosome excluded volume interactions are not the only ones within the chromatin fiber. DNA linkers have a diameter of about 2nm and therefore excluded volume restrictions, too. This is very important especially for all types of crossed-linker fibers. Since DNA linkers have a very strong and screened Coulomb repulsion their excluded volume interactions resemble their Coulomb interactions.

Furthermore, the nucleosome-nucleosome potentials will be discussed in this section.

The DNA-nucleosome interaction (i.e. mainly the interaction between the linker DNA and the histone tails) is not known very well at present. In most of the theoretical work it is deemed negligible compared to DNA-DNA and nucleosome-nucleosome interaction. Furthermore, the E2A-model already allows for the fixation of DNA linkers by H1 and H5 histones. So a part of the interaction between linker DNA and nucleosome is part of the model geometry (note that DNA which is wrapped around a histone octamer is considered as a part of the nucleosome).

4.6.1 Volume Exclusion

Excluded volume potentials play a very important role for biopolymers since they contribute largely to entropic forces and thus to long range interactions (cf. [102] for instance). In this work all excluded volume potentials were modelled by hard-core potentials i.e. $V = \infty$ if excluded volume is violated and $V = 0$ if not.

For the basic two-angle model spheres were applied to approximate the volume of the nucleosomal complex. In the E2A-model stiff cylinders were used to model the volume exclusion. Further parameters which have been exerted for the excluded volume potentials can be found in Table 1.

Remark: In the following sections of this thesis the mean value of the length of a base pair will be used as length unit: $1bp := \langle 1\text{bp} \rangle := 0.34$ nm.

4.6.2 DNA

DNA is modelled as a charged Worm Like Chain (WLC, cf. Sec. 1.3) with a persistence length of 50nm. For the nucleosome skips (cf. Sec. ???) one needs to model blank

Model Parameters		
nucleosome height	16.8 bp	5.6 nm
nucleosome diameter	33.0 bp	7 nm
linker length b	51.0 bp	17.3 nm
vertical displacement d	8.4 bp	2.8 nm
DNA diameter	6.6 bp	2.2 nm
DNA period length	10.2 bp	3.4 nm
DNA major groove	6.6 bp	2.6 nm
DNA minor groove	3.6 bp	1.2 nm
DNA turns ω	1.65	–
length DNA turns	146 bp	49.6 nm

Table 1: Parameters used to describe the geometry and the hard core excluded volume potentials of the chromatin fiber. $\langle 1\text{bp} \rangle \approx 0.34$ nm.

DNA stretches of integer multiples of the nucleosome repeat length (NRL) i.e. of 197 bp (≈70nm).

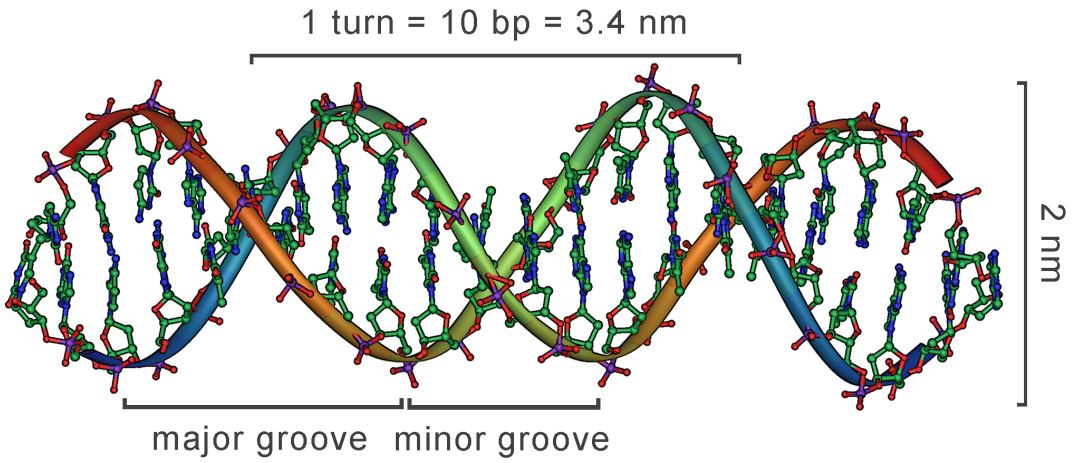


Figure 41: A DNA stretch.

4.6.3 Coulomb Repulsion of DNA Linkers

The DNA has a very high charge density of $\lambda_{DNA} \approx 2e/0.34nm$ but the Coulomb potentials are strongly screened due to the ionic environment of the chromatin fiber in the cell nucleus. The electrostatic potential and charge density are related in the Poisson equation:

$$\triangle \Phi = -\frac{\varrho}{\epsilon_0 \epsilon_r} \tag{24}$$

in this case $\epsilon_r = 81$ (the relative permittivity of water) was used. The concentration of positive and negative charged ions (n^+, n^-) of the point where the potential is Φ may be found from the Boltzmann distribution:

$$n^+ = n_0^+ \exp\left(\frac{-z_+ e\Phi}{kT}\right)$$
$$n^- = n_0^- \exp\left(\frac{-z_- e\Phi}{kT}\right)$$

where z_+ and z_- are the numbers of charges carried by positive and negative ions respectively, n_0^+ and n_0^- the numbers of ions of each type per unit volume, e is the unit electronic charge and k is the Boltzmann constant. The electrical density ϱ at the point where the potential is Φ, is the excess positive or negative electricity per unit volume:

$$\varrho = n^+ z_+ e - n^- z_- e = e\left(z_+ n_0^+ \exp\left(\frac{-z_+ e\Phi}{kT}\right) - z_- n_0^- \exp\left(\frac{-z_- e\Phi}{kT}\right)\right)$$

due to electroneutrality the condition

$$n_0^+ z_+ \stackrel{!}{=} n_0^- z_-$$

must be satisfied. Substituting now the electrical density in the Poisson equation leads to

$$\begin{aligned}\triangle\Phi &= -\frac{e}{\epsilon_0\epsilon_r}(z_+ n_0^+ \exp\left(\frac{-z_+ e\Phi}{kT}\right) - z_- n_0^- \exp\left(\frac{+z_- e\Phi}{kT}\right)) \\ &= -\frac{e}{\epsilon_0\epsilon_r}\left(z_+ n_0^+ [\exp\left(\frac{-z_+ e\Phi}{kT}\right) - 1] - z_- n_0^- [\exp\left(\frac{+z_- e\Phi}{kT}\right) - 1]\right) \\ &\stackrel{z_\pm e\Phi \ll kT}{\approx} \frac{e^2}{\epsilon_0\epsilon_r kT}[z_+^2 n_0^+ + z_-^2 n_0^-]\Phi\end{aligned}$$

and with

$$\kappa^2 = \frac{e^2}{\epsilon_0\epsilon_r kT}[z_+^2 n_0^+ + z_-^2 n_0^-]$$

this results in

$$\triangle\Phi = \kappa^2\Phi.$$

The spherically symmetric solution of this equation (i.e. for a point charge) follows with $\triangle_r = \frac{1}{r}\frac{\partial^2}{\partial r^2}r$ and $\Psi = r\Phi$ since this leads to $(-\frac{d^2}{dr^2} + \kappa^2)\Psi = \Psi$ and thus to $\Psi = c \cdot \exp(\pm\kappa r)$. Within a small sphere of radius ϵ around the point charge there is no screening, this determines the constant c:

$$\Rightarrow\ \Phi(r) = \frac{q}{4\pi\epsilon_0\epsilon_r}\frac{\exp(-\kappa r)}{r}\ .$$

κ is the inverse Debye length which is a measure for the screening of the Coulomb interaction $\kappa = \frac{1}{\lambda}$. Different values of κ in dependence on the salt-concentration are listed in Table 2.

c_S [10^{-2}M]	1	2	3	4	5
κ [nm^{-1}]	0.330	0.467	0.572	0.660	0.738
ν_{eff} [e/nm]	-2.43	-2.96	-3.39	-3.91	-4.15

Table 2: Screening parameters for the Coulomb Repulsion of DNA.

Since the screening of the Coulomb interaction starts at the radius of the DNA strand and due to the condensation of ions along the DNA linkers one has to calculate a correction of the screened potential above by fitting the tail of the Debye-Hückel potential for an infinitely-long cylinder to the Gouy-Chapman potential in the far zone. This calculation can be found in [181; 202] and leads to a corrected linear charge density ν_{eff} which is also given in Table 2 for different levels of monovalent salt concentration.

Now one can calculate the Coulomb repulsion between the DNA segments i and j by evaluating

$$V_{i,j} = \frac{\nu_{\text{eff}}}{c} \int \int \frac{e^{-\kappa r_{i,j}}}{r_{i,j}} \mathrm{d}x_i \mathrm{d}x_j \tag{25}$$

where c is the total dielectric constant of water. These double-integrals were numerically evaluated during some of the chromatin simulations.

4.6.4 Nucleosome Interactions

The charges on the linker histones H1, H5 and the histone tails are under biochemical control and allow the cell to regulate the stem formation and the attractive interaction between nucleosomes (cf. Sec. 2). Therefore, the soft parts of the nucleosome are problematic for all attempts to model chromatin. Less is known about the interaction of the nucleosomes between themselves or with free DNA.

The nucleosome-nucleosome interaction has recently been parameterized by using the surface charge density of the known crystal structure in a point-charge model (cf. [17]). While in that work only electrostatic interactions were considered and the quantitative influence of the histone tails on the interaction potential still remains obscure, simulations based on this potential allowed to predict an ionic-strength dependent structural transition of a 50nm-nucleosome chromatin fragment that occurred at a salt concentration compatible with known experimental data (cf. [18]).

Nucleosome-nucleosome interaction potentials can be calibrated by comparison with the characteristics of liquid crystals of mononucleosomes at high concentrations. Under suitable conditions nucleosome core particles form a hexagonal-columnar phase with a distance of 11.55 ± 1nm between the columns and a mean distance of 7.16 ± 0.65nm between the particles in one column (cf. [129; 130]). These distances may be assumed to correspond to the positions of the minima of an attractive internucleosomal potential. The depth of the interaction potential (i.e. the binding energy per nucleosome) was

estimated in the stretching experiments in [47] to $2.6 - 3.4$kT. A slightly lower potential minimum of 1.25kT is obtained by a comparison of the stability of the nucleosome liquid crystal phase with simulations [223].

In [61] and this work the nucleosome-nucleosome interaction was modelled by a isotropic Morse-potential

$$U_{\text{Morse}}(r) = \epsilon \left(1 - e^{-a(r-r_0)}\right)^2 - \epsilon \tag{26}$$

with the parameters $a = 0.07$, $r_0 = 46$bp and $\epsilon = 3$kT.

The results for the nucleosome-nucleosome interaction will be shown in Fig. 177. It will turn out that the excluded volume borderline is reproduced by the nucleosome-nucleosome potential.
The results of the simulations with interaction potentials will be discussed in Sec. 5.5.1.

5 Chromatin Phase Diagram

5.1 Overview

This section provides a comprehensive discussion of the chromatin phase diagram.

First (a cut through) the phase diagram is presented and all the possible chromatin structures are discussed (cf. Sec. 5.2 and 5.3).

Then it will be shown that the shape of the excluded volume borderline (i.e. the curve that separates the allowed from the forbidden states in the phase diagram) can be understood by geometrical considerations and that the irregularity of its peaks comes from an underlying prime factor dismantling problem (cf. Sec. 5.4). Furthermore, the influence of the pitch d and the cylindrical (instead of spherical) excluded volume of the nucleosomes on this borderline is investigated (cf. Sec. 5.4.4). After that, some properties of the different structures will be discussed (cf. Sec. 5.5).

Finally, it will be shown that parameter distributions lead to a *distribution* of structures in the phase diagram instead of a particular *fixed* structure (cf. Sec. 5.6).

Most of the results of this chapter have already been published [61; 62; 64] or are currently under review [65]. Some of the analytical parts are adapted from [61; 62] for the sake of completeness.

5.2 The Chromatin Phase Diagram

As explained in Sec. 4.5 one can build up *stiff* chromatin fibers by solving the conditions 4.4.2. *Stiff* means that for each nucleosome the four parameters α, β, b and d are equal. Thus one gets perfectly regular fibers without any thermal fluctuations.

Of course, these are very artificial states far from equilibrium but they are helpful to get an impression of the different chromatin structures and they can be interpreted as a first mean-field approach on small length scales.
Furthermore, one can investigate them with respect to interesting chromatin features like line density, fiber extension or compactness and so on.

For every set of parameters $(\alpha, \beta, b, d)_N$ one can generate a certain (stiff) chromatin fiber of length N. That means the chromatin phase diagram in the E2A-model has four dimensions.
In this section the linker length b will be fixed to its expected mean value of 17.3nm (cf. Tab. 1) so that three dimensions remain. During the first considerations of this chapter the pitch d will be set to zero but later its influence on the phase diagram will be considered as well (cf. Sec. 5.5.2).

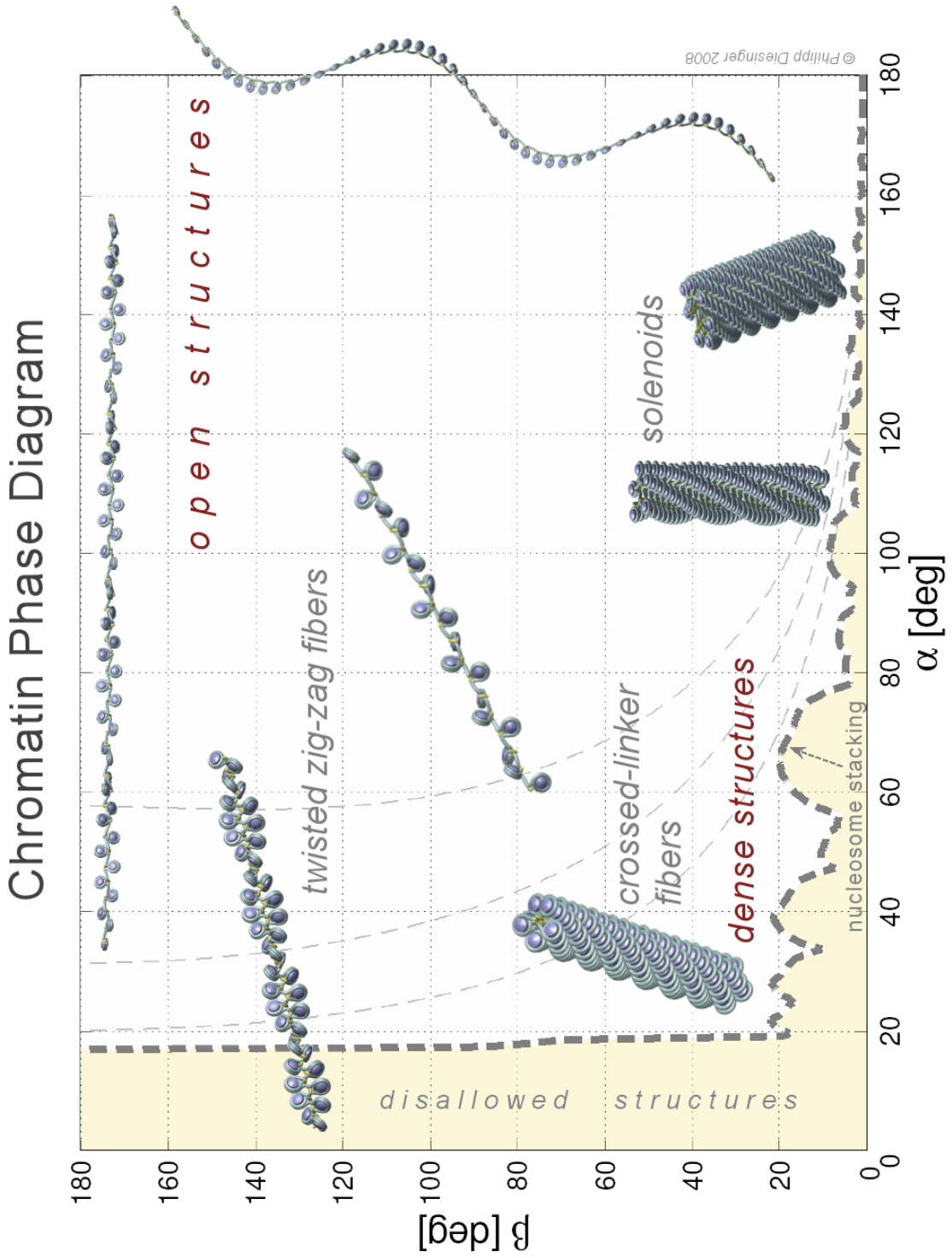


Figure 42: A cut through the actually four-dimensional chromatin phase diagram. The pitch d and the linker length b are fixed here so that only the two angles remain as parameters. A single point in this diagram corresponds to a specific chromatin structure. The forbidden structures lie left and below the dashed line which is the excluded volume borderline [62].

Actually there is a fifth important parameter namely the fiber length N. N denotes the number of nucleosome repeat lengths of which the chromatin fiber consists of.

In this chapter N has to be chosen carefully in the numerical part since in some cases one can separate forbidden and allowed states (with respect to volume exclusion) only when N is large enough.

On the one hand, N should be as small as possibly to reduce the running time of the simulations and on the other hand it should be large enough to obtain the right physical insights. For the theoretical part infinitely long fibers: $N \to \infty$ are always assumed.

A cut through the four-dimensional chromatin phase diagram can be found in Fig. 42. As explained b and d are fixed in this case so only the two angles α and β remain as fiber parameters. Therefore, every single point of this diagram corresponds to a certain chromatin structure.

One can see dense states like the crossed-linker fibers or very open states which resemble beads-on-a-string.

The classical solenoid model of Finch and Klug [80] is found in the large α and small β area in Fig. 42 (but in their case the linkers were bent).

Various other structures with $\alpha \approx 30°$ and different values of β were found by Woodcock et. al. in electron micrographs [235]. These are located on the left side (i.e. small α) of the phase diagram (cf. Fig. 42).

Of course, some of the states in the phase diagram are forbidden due to violation the excluded volume restrictions: These are all states left and below the dashed line. The dashed line will be called *excluded volume borderline* in the following since it separates the allowed from the forbidden states.

Furthermore, the excluded volume borderline indicates the area where the nucleosome stacking takes place. Stacking of nucleosomes is particularly interesting because it leads to very compact chromatin fibers and as pointed out before one important feature of chromatin is DNA condensation (cf. Sec. 2.2)

All the different chromatin states will be discussed in the following sections. They can be divided into *planar* and *three-dimensional* structures. Although the planar states seem not to be of interest, since they lie in the forbidden area, it will turn out that these very structures are of essential importance to understand the shape of the excluded volume borderline.

5.3 Chromatin Structures

In this section and partially in Sec. 5.4 the nucleosomes will be treated as spherical instead of cylindrical objects. This approximation is necessary for the analytical calculation of the excluded volume borderline. The effect of cylindrically shaped nucleosomes as well as the physical chromatin features will be discussed later (cf. Sec. 5.4.4 and Sec. 5.5).

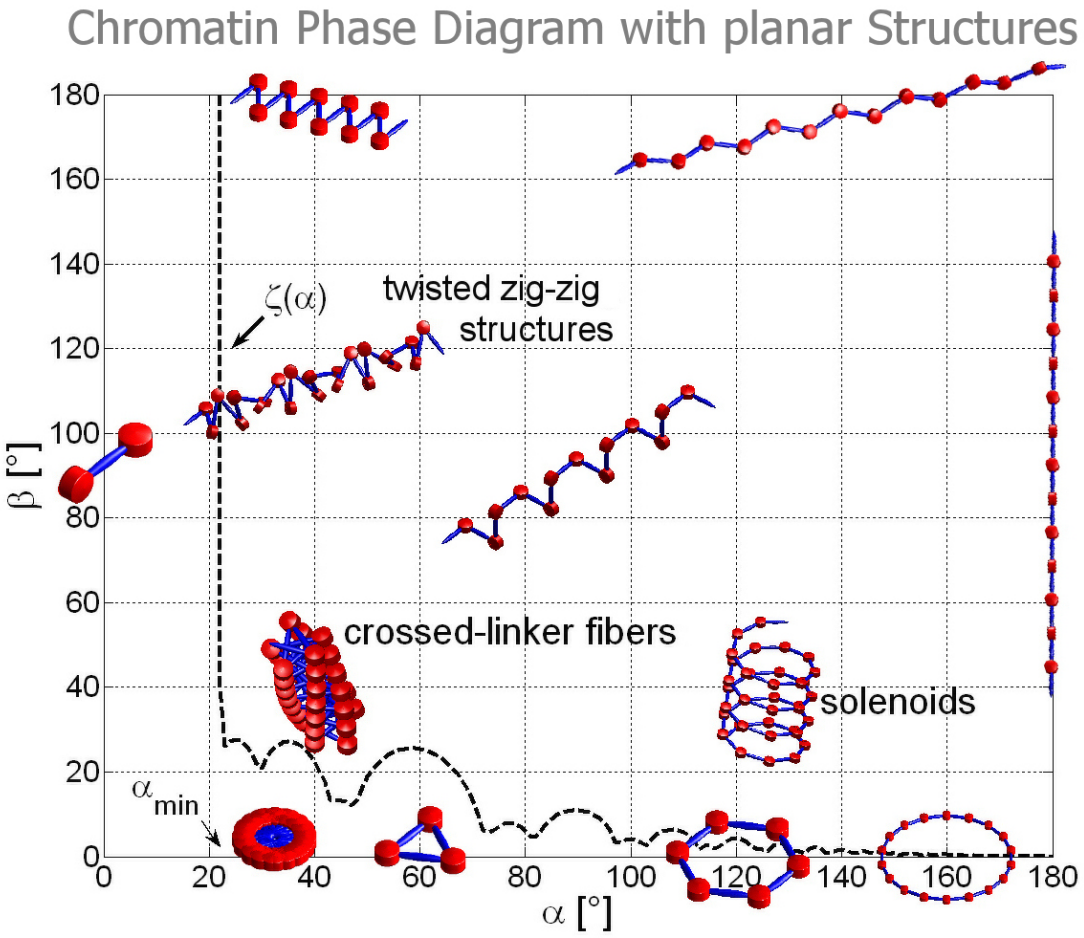


Figure 43: The chromatin phase diagram with some planar structures. The pitch is set to zero ($d = 0$) to facilitate the analytical calculation of the excluded volume borderline. The influence of a non-vanishing pitch will be discussed later (cf. Sec. 5.4.4). Adapted from [61].

5.3.1 Planar Structures

If either one of the angles α or β equals 0 or π the resulting structure will be planar (cf. Fig. 43). Although it seems that the planar structures aren't of much interest from the physical point of view it turned out that they play a very important role for the excluded volume borderline of the phase diagram, and therefore, they will be discussed here in detail.

Case A: $\beta = \pi$, $\alpha \in [0, 2\pi]$

In the case $\beta = \pi$ and α arbitrary the fiber forms 2D zig-zag-like structures, as shown

at the top of Fig. 43. The length of a fiber consisting of N monomers is given by

$$L = d \cdot N = b \cdot \sin\left(\frac{\alpha}{2}\right) N \quad \text{(cf. Eq. 15)}$$

and the diameter is given by

$$D = b \cos\left(\frac{\alpha}{2}\right) \quad \text{(cf. Eq. 13)}.$$

The length of the fiber increases with increasing α.

Case B: $\alpha = \pi$ **or** $\alpha = 0$**,** $\beta \in [0, 2\pi]$
In the case of $\alpha = \pi$ with an arbitrary value of β the fiber is a straight line and for $\alpha = 0$ and $\beta \in [0,\pi]$ corresponds to linkers that go back and forth between two positions (cf. Fig. 43).

Case C: $\beta = 0$**,** $\alpha \in [0, 2\pi]$

Consider now the important case of the edge $\beta = 0$, at the bottom of the phase diagram (cf. Fig. 43):

If $\alpha \approx \pi$ the fiber forms a circle (with radius $R \approx \frac{b}{\pi - \alpha}$ as follows from equation 13). Its radius converges towards infinity for $\alpha \to \pi$.
For some values of α the fiber forms a *regular polygon*. For instance the value $\alpha = \frac{\pi}{2}$ corresponds to a square and $\alpha = \frac{\pi}{3}$ corresponds to the regular triangle (cf. Fig. 42). To characterize these regular polygons one needs two variables: At first the *number of the tips* i and secondly the *order* n of the polygon which represents the number of whitlows or circulations the fiber has to do in order arrive at the starting point again. The α-values that lead to these regular (planar) polygons will be denoted by α_i^n in the following.
To give an example Fig. 44 shows the two possible geometries of a polygon with 5 tips: One with only one loop ($\alpha_5^1 = 108°$) and another one with two loops ($\alpha_5^2 = 36°$).

The values of $\alpha = \alpha_i^n$ that lead to such regular polygons are given by

$$\alpha_i^n = \pi - \left(\frac{n \cdot 2\pi}{i}\right), \text{ with } i, n \in \mathbb{N} \text{ and } i \geq 2n, \text{ such that} \tag{27}$$

$$\nexists\, n', i' \in \mathbb{N} \text{ with } n' < n \text{ and } \alpha_i^n = \alpha_{i'}^{n'} \quad (\triangle).$$

Remarks: The order n of α_i^n is a measure for its contribution to the excluded volume borderline in the chromatin phase diagram which will be shown below (cf. Eq. 35). Therefore, all different values of α_i^n will be sorted by n first and secondly by i (i.e. if $n_1 > n_2$, then $\alpha_i^{n_1}$ is more important than $\alpha_i^{n_2}$ and if $i_1 > i_2$ then $\alpha_{i_1}^n$ is more important than $\alpha_{i_2}^n$ to the excluded volume borderline). In fact, this is what makes condition ($\triangle$)

necessary: For example $\alpha_3^1 = \alpha_6^2 = \alpha_9^3 = \frac{\pi}{3}$ but the order of these three α_i^n is always $n = 1$, and therefore, they are all in the same equivalence class.

The special case $i = 2n$ leads to one-dimensional structures with $\alpha_2^1 = 0$ - this value of α_i^n of highest order in n and i plays an important role in Sec. 5.4.2 and is therefore mentioned here, too.

According to Eq. 27 all α_i^n are irrational numbers .

5.1 Theorem
Condition ($\triangle$) *of Eq. 27* $\Leftrightarrow$ $\nexists\, \alpha, \beta, b \in \mathbb{N}$ with $n = \alpha \cdot b$ and $i = \beta \cdot b$ ($\bigstar$).

($\bigstar$) *indicates that* n *and* i *are* coprime *since* α_i^n *depends only on the relation* $\frac{n}{i}$.

Proof
"$\Rightarrow$"

Let n be an arbitrary element of $\mathbb{N}$. Now assume that $i \in \mathbb{N}$, $i \geq 2n$ and n and i have a common divisor b i.e. $\exists\ \alpha,\ \beta$ and $b \in \mathbb{N}$ such that $i = \alpha \cdot b$ and $n = \beta \cdot b \ \Rightarrow\ \frac{n}{i} = \frac{\beta \cdot b}{\alpha \cdot b} = \frac{\beta}{\alpha} = \frac{\frac{n}{b}}{\frac{i}{b}}$, furthermore $i \geq 2n$ indicates $\frac{i}{b} \geq 2\frac{n}{b} \ \Rightarrow\ \frac{\beta}{\alpha} \leq \frac{1}{2}$ which means $\exists\ n', i' \in \mathbb{N}$ with $i' \geq 2n'$ and $\frac{n'}{i'} = \frac{n}{i}$, namely $n' = \beta$ and $i' = \alpha$. This means $\neg\bigstar$ implies $\neg\triangle$ which is equivalent to $\triangle \Rightarrow \bigstar$.

"$\Leftarrow$"

To proof the other implication above it suffices to show that if n and i are coprime, there will be no numbers m, j with $m < n$ such that $nj = mi$. Dividing this equation by $\gcd(n, m)$ one gets a new equation of the same structure with m' and n' instead of m and n, fulfilling $\gcd(m', n') = 1$. Moreover, $\gcd(n, i) = 1$ leads to $\gcd(n', i) = 1$. Thus $\gcd(m'i, n') = 1 \Rightarrow j = 1$ which contradicts $j > 2n$. □

5.2 Theorem
If two orders n_1 *and* n_2 *are coprime with* $n_1, n_2 > 1$ *, there* never *will be equal values of* $\alpha_i^{n_1}$ *and* $\alpha_{i'}^{n_2}$ *for all possible* $i \geq 2n$, $i' \geq 2n'$ *and* $i, i' \in \mathbb{N}$.

For example n *and* $n+1$ *are always coprime numbers, and therefore, have never common* α_i^n.

Proof
Consider the case that n and n' are coprime with $n' > n$, $i \geq 2n$ and $i' \geq 2n'$ such that $\frac{n}{i}$, $\frac{n'}{i'} \notin \frac{1}{\mathbb{N}}$ (which means that they are not equivalent to any order-1-value of α_i^n). Furthermore, assume that $\frac{n'}{i'} = \frac{n}{i} \Rightarrow\ i' = \left(\frac{n'}{n}\right) \cdot i$, it is clear that $\left(\frac{n'}{n}\right) \notin \mathbb{N}$ and $i \in \mathbb{N}$, therefore $\exists\ \alpha \in \mathbb{N}$ so that $i = \alpha \cdot n$ which contradicts $\left(\frac{i}{n}\right) \notin \mathbb{N}$. □

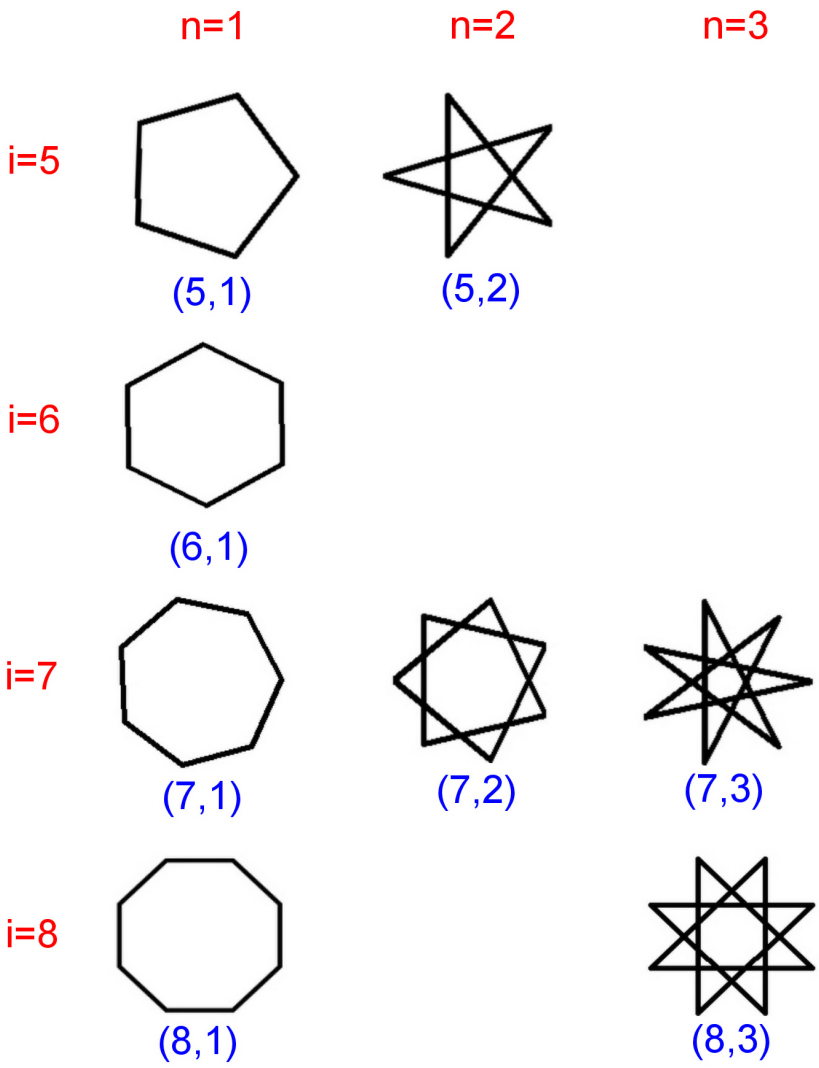


Figure 44: This figure shows some of the regular polygons, their classification (n, i) in blue and their number of tips i as well as their order in red: For $i = 5$, 6, 7 and 8 all possible structures are shown. The maximum number of possible polygons for a given i is max=floor$(i/2)$, achieved for prime numbers. The only possible structures for $i = 3$ and $i = 4$ are the equilateral triangle and the equilateral square. Adapted from [61].

One can learn from Theorem 5.1 and 5.2 that for a given order n the possible values of α_i^n depend on the *prime factor dismantling* of n and thus are very irregular.
If n is a prime number, all $i \geq 2n$ with $i \notin n \cdot \mathbb{N}$ are allowed: for $n = 2$ all $i \geq 4$ with $i \notin 2 \cdot \mathbb{N}$ are allowed and for $n = 3$ all i with $i \geq 2n$, $i \notin 3 \cdot \mathbb{N}$ are allowed (and for $n = 1$ all $i \geq 2$ are allowed). So between two values of α_i^1 there is one of α_i^2 and two of α_i^3 - this will be important in Sec. 5.4.2.

The distances $\Delta_{i_1}^n = \alpha_{i_2}^n - \alpha_{i_1}^n$ between two consecutive $i_2 > i_1$ are given below for the

first three and for prime orders:

$$n = 1\text{: } \alpha_i^1 = \pi - \frac{2\pi}{i},\ (i \geq 2n) \ \Rightarrow \Delta_i^1 = \frac{2\pi}{i(i+1)}$$
$$n = 2\text{: } \alpha_i^2 = \pi - \frac{4\pi}{i},\ (i \geq 2n \text{ and } i \notin 2 \cdot \mathbb{N}) \ \Rightarrow \Delta_i^2 = \frac{8\pi}{i(i+2)}$$
$$n = 3\text{: } \alpha_i^3 = \pi - \frac{6\pi}{i},\ (i \geq 2n \text{ and } i \notin 3 \cdot \mathbb{N}) \ \Rightarrow \Delta_i^3 = \frac{\mod(i,3) \cdot 6\pi}{i(i + \mod(i,3))}$$
$$n \text{ prime: } \alpha_i^n = \pi - \frac{n \cdot 2\pi}{i},\ (i \geq 2n \text{ and } i \notin n \cdot \mathbb{N}) \ \Rightarrow \Delta_i^n = \frac{m \cdot n \cdot 2\pi}{i(i+m)}.$$

$$\text{with } m = \begin{cases} 1 \Leftarrow & \mod(i,n) < n-1 \\ 2 \Leftarrow & \mod(i,n) = n-1 \end{cases}$$

Furthermore, this shows that $\Delta_i^n \overset{n\to\infty}{\to} 0$, because the counter is always $\sim n$ and the leading term of the denominator is at least $\sim n^2$ (as $i > 2n$).

5.3.2 Three-Dimensional Chromatin Fibers

The most interesting structures of the chromatin phase diagram are the following two cases: Solenoid like structures and fibers with crossed-linkers (cf. Sec. 2.2.2). These cases will be discussed here in detail:

a) Solenoids

For small $\beta \ll 1$ and $\alpha \approx \pi$ the chromatin fibers resemble solenoids where the linkers themselves follow closely a helical path. This implies $\left(\frac{d \cdot a}{R}\right) \ll 1$. The lowest order approximation in $\left(\frac{d \cdot a}{R}\right)$ indicates $b \approx d\sqrt{1+a^2}$ (cf. Eq. 10), $\alpha \approx \pi - \frac{a^2 b}{R(1+a^2)}$ (cf. Eq. 11) and $\beta \approx \frac{\pi - \alpha}{a}$ (cf. Eq. 12).
These approximations lead to geometrical features of the solenoid-like fibers: The radius of the fiber is denoted by r, l_N is the length of a fiber of N monomers, $\lambda = N/l_N$ is the corresponding line density and $\varrho = \lambda/\pi r^2$ is the density of the fiber:

$$r = \frac{b\,(\pi - \alpha)}{\beta^2 + (\pi - \alpha)^2} \qquad l_N = \frac{b\,N\,\beta}{\sqrt{\beta^2 + (\pi - \alpha)^2}}$$
$$\lambda = \frac{\sqrt{\beta^2 + (\pi - \alpha)^2}}{b\,\beta} \qquad \varrho = \frac{(\beta^2 + (\pi - \alpha)^2)^{5/2}}{\pi\, b^3\, \beta\, (\pi - \alpha)^2}.$$

The vertical distance δ between two loops plays an important role in the following sections. It can be obtained by $\delta \approx l_{\frac{2\pi}{\pi - \alpha}}$, which leads to

$$\delta \approx \frac{\left(\frac{2\pi}{\pi - \alpha}\right)\, b\,\beta}{\sqrt{\beta^2 + (\pi - \alpha)^2}}. \tag{28}$$

One can calculate the exact value of δ by:

$$\delta = \int_{t=0}^{2\pi r/a} m(\gamma(t)) \cdot |(\gamma_x'(t), \gamma_y'(t))^t| \, dt$$
$$= \frac{4b\pi \sin(\frac{\alpha}{2}) \sin(\frac{\beta}{2})}{\arccos\left(2\cos^2(\frac{\beta}{2})\sin^2(\frac{\alpha}{2}) - 1\right)\sqrt{3 + \cos(\alpha) + (\cos(\alpha) - 1)\cos(\beta)}}$$

For $\beta \ll (\pi - \alpha)$ one finds a very dense spiral ($\delta \ll r$). In the opposite limit $\beta \gg (\pi - \alpha)$ the solenoid has a very open structure ($\delta \gg r$).

b) Crossed-Linker Fibers

Now consider the case $\beta \ll 1$ and $\alpha \ll \pi$: In the previous section (Sec. 5.3.1) the regular polygons have been discussed. For a non-vanishing value of β these regular polymers open up in a accordion-like manner. This leads to three-dimensional fibers with crossed linkers (cf. Fig. 42). Using $\beta \ll 1$ and Eq. 10, 11 and 12 one gets:

$$d^2 = \frac{1}{4}\beta^2(4R^2 - b^2)$$
$$\Rightarrow r = \frac{b}{2\cos(\alpha/2)}\left(1 - \frac{\beta^2}{4}\cot^2(\frac{\pi - \alpha}{2})\right)$$
$$\Rightarrow l_N = \frac{N\,\beta\,b}{2}\cot(\frac{\pi - \alpha}{2}), \quad \lambda = \frac{4}{\alpha\,\beta\,b}$$
$$\Rightarrow \varrho = \frac{16}{\pi\,\beta\,\alpha\,b^3.}$$

Now δ follows again from $\delta \approx l_{\frac{2\pi}{\pi-\alpha}}$:

$$\delta \approx \frac{1}{2}\left(\frac{2\pi}{\pi - \alpha}\right)\,\beta\,b\tan\left(\frac{\alpha}{2}\right). \tag{29}$$

For a rotational angle β close to π, say $\beta = \pi - \varepsilon$ with $\varepsilon \ll 1$ one gets *twisted zig-zag structures* (cf. Fig. 42). In this case monomer $k+1$ is located nearly opposite to the kth monomer but slightly twisted by an angle ε. Monomer $k+2$ is then on the same side as monomer k but slightly twisted by an angle 2ε and so on. The fiber lies inside a cylinder of radius

$$r = \frac{b}{2}\cos\alpha 2\left(1 + \left(\frac{\varepsilon}{2}\right)^2\right)$$

and the length l_N is given by

$$l_N = b\sin(\frac{\alpha}{2})N\left(1 - \frac{\varepsilon^2}{8}\cot^2(\frac{\alpha}{2})\right).$$

The monomers are located at the surface with the linker passing back and forth (approximately) through the middle axis of the cylinder. The monomers n, $n \pm 2$, $n \pm 4$, ... and the monomers n, $n \pm 1$, $n \pm 3$, ... form a double helix that winds around the cylinder. Within each of these two spirals the monomers are not directly linked together, even though monomer k and $k+2$ can come quite close.

5.4 Excluded Volume Restrictions

Not all of the different chromatin structures have the same probability to occur in real chromatin fibers (cf. Sec. 5.6). But there are even some structures which have zero probability because they are forbidden due to excluded volume restrictions. These forbidden areas of the chromatin phase diagram are the subject of this section.
In this numerical and analytical study hard-core potentials were used to model the excluded volume interactions of the fibers.

In the following Γ_N represents the chromatin phase space and $(\alpha, \beta, b, d)^t_N = \gamma_N \in \Gamma_N$ a specific chromatin structure with a length of nN NRLs. The general function η indicates whether a point of the phase diagram belongs to the forbidden area due to excluded volume interactions or not. So the aim of this section is to learn how Γ_N looks like.
It will turn out that there exists a border that separates the non-excluded volume structures from forbidden from those which fulfil the excluded volume conditions. This border is denoted by ζ and called *excluded volume borderline*. Furthermore, it will turn out that ζ is a function of α.

$$\eta : \Gamma_N \to \{0,1\} : \gamma_N \mapsto \begin{cases} 1 \Leftarrow \gamma_N \text{ fulfills the excluded volume conditions} \\ 0 \Leftarrow \gamma_N \text{ does not fulfill the excluded volume conditions} \end{cases}$$

$$\zeta : [0,\pi] \to [0,\pi] : \alpha \mapsto \min\{\beta \mid \eta(\gamma_{n,\alpha,\beta}) = 1, \ \forall\, n > 3\left(\frac{\pi+\alpha_e}{\pi-\alpha_e}\right)\}$$

The excluded volume borderline is already plotted in Fig. 42 and 43.
$n_{min} = 3\left(\frac{2\pi}{\pi-\alpha_e} - 1\right) = 3\left(\frac{\pi+\alpha_e}{\pi-\alpha_e}\right)$ is given by the chromatin geometry (linker length, octamer diameter and height, see below).
As shown in Fig. 42 the interesting part of the phase diagram for the excluded-volume phase-transition is the lower one with $\beta \in [0, 30°]$. This cut-out of the phase diagram is shown in Fig. 45.

5.4.1 Some Basic Estimations

One can explain Fig. 45 by dividing the excluded volume interactions between the nucleosomes in *short-range* and *long-range* interactions: There is never an excluded volume interaction between two consecutive nucleosomes, short-range interactions occur between nucleosome k and nucleosome $k+2$, and long-range interactions are those between nucleosomes with a larger distance than 2 ($\Delta k > 2$).

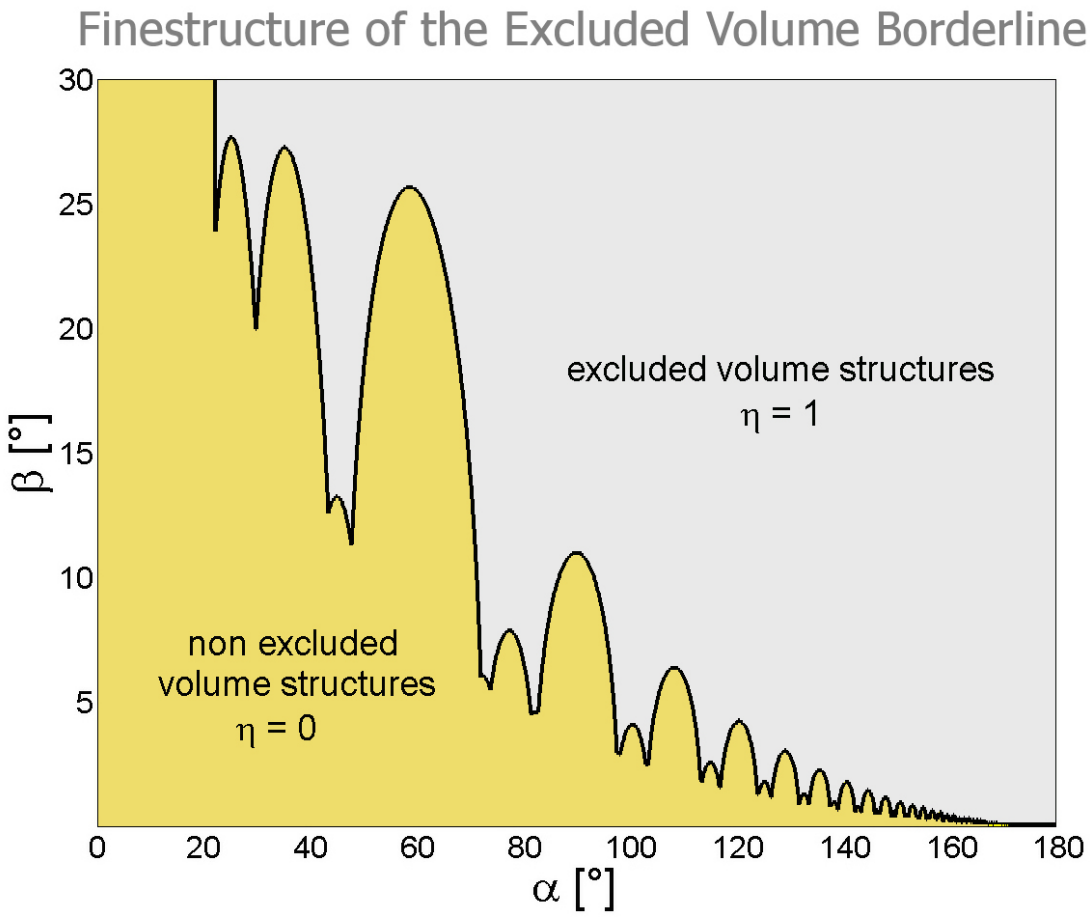


Figure 45: Fine structure of the excluded-volume "phase transition". The chromatin fibers in the upper area fulfil the excluded volume conditions, those in the light-yellow area do not. The borderline is a function of the entry-exit angle α: $\zeta = \zeta(\alpha)$.

The short-range interaction explains the minimal value α_{min} for α to find at least some structures in the allowed regime. The fiber never fulfils the excluded volume conditions for any $\alpha < \alpha_{min}$ (i.e. $\eta(\alpha) = 0\ \forall \alpha < \alpha_{min}$). α_{min} can be calculated easily by simple trigonometry:

$$\sin\left(\frac{\alpha_{min}}{2}\right) = \frac{d}{2b} \Rightarrow \alpha_{min} \approx 22^\circ \tag{30}$$

(d is the octamer diameter and b is the linker length)

This condition excludes a vertical strip at the left side of Fig. 45 (and 42).

For small $\beta > 0$ the long-range excluded-volume interaction comes into play: For a given α, starting with a planar, circular structure (i.e. $\beta = 0$) one has to increase β

above $\zeta(\alpha)$ so that neighboring *loops* do not interact. With Eq. 28 and $\beta \ll \pi - \alpha$ this leads to (cf. Fig. 45)

$$\beta \gtrapprox \frac{r\,(\pi-\alpha)^2}{\pi\; b} \overset{\alpha \lll 1}{\Rightarrow} \beta \gtrapprox \frac{4d}{\pi b} \approx 28^\circ. \tag{31}$$

These two approximations for critical values of α and β are shown in Fig. 46, too.

Another structure explanation of the excluded-volume phase-transition which even allows to derive the explicit progression of ζ and does not need to distinguish between short-range and long-range interactions is presented in the following section (Sec. 5.4.2).

5.4.2 Calculation of the Excluded Volume Borderline $\zeta(\alpha)$

In this section it will be shown that every peak of the excluded volume borderline ζ corresponds to a regular polygon (cf. Sec. 5.3.1): The large peaks correspond to polygons of order 1 and the smaller ones to order 2 and order 3 polygons. Between two order-1-peaks there is one order-2-peak and two order-3-peaks (as was shown in Sec. 5.3.1, too).
The classification of these peaks is shown in Fig. 46: ζ has local maxima at every α_i^n. This is because the planar structures which belong to $(\alpha_i^n, \beta = 0)$ need a large increase of β to arrive in the area of the excluded volume structures, since at $\beta = 0$ nucleosome k and $k + \frac{n \cdot 2\pi}{\pi - \alpha_i^n}$ are located exactly at the same position.

At first, consider only the particular α-values α_i^n: At first order increasing β from $\beta = 0$ to $\beta = \tilde{\beta}$ leads to a vertical movement $\Delta_i^n(\tilde{\beta})$ of the relevant nucleosomes k and $k + \frac{n \cdot 2\pi}{\pi - \alpha_i^n}$ along the chromatin axis . If $\Delta_i^n(\tilde{\beta}) = d \;\Rightarrow\; \zeta(\alpha_i^n) = \tilde{\beta}$. For large α, $\Delta_i^n(\tilde{\beta})$ can be calculated by Eq. 28 and one finds:

$$\Delta_i^n(\tilde{\beta}) = \frac{n \cdot \left(\frac{2\pi}{\pi - \alpha_i^n}\right)\; b\; \tilde{\beta}}{\sqrt{\tilde{\beta}^2 + (\pi - \alpha_i^n)^2}} \tag{32}$$

with $\beta > 0$ this leads to

$$\tilde{\beta}_i^n(\Delta_i^n) = \sqrt{\frac{(\Delta_i^n(\pi - \alpha_i^n))^2}{b^2\; n^2 \left(\frac{2\pi}{\pi - \alpha_i^n}\right)^2 - (\Delta_i^n)^2}} \tag{33}$$

and furthermore $\tilde{\beta}_i^n(d) = \zeta(\alpha_i^n)$ implies

$$\zeta(\alpha_i^n) = \sqrt{\frac{d^2 \cdot (\pi - \alpha_i^n)}{b^2\; n^2 \left(\frac{2\pi}{\pi - \alpha_i^n}\right)^2 - d^2}}. \tag{34}$$

Fig. 47 shows the numerically determined excluded volume borderline and the theoretical predictions for the maxima. As Eq. 32 shows,

$$\Delta_i^n(\tilde{\beta}) \sim n. \tag{35}$$

This means that planar structures at an α_i^n of *higher* order need a *smaller* rise of β to fulfill the excluded volume conditions, and therefore, ζ is decreasing with increasing α. In fact, this is the reason why n is called the *order* of the peaks of the ζ-function. It can be understood easily, if one considers the fact that a higher order means more nucleosomes are located between two overlapping ones. And thus a rise of β has a stronger effect than at lower orders.
This explains the forbidden strip at the left side of the figures as the maximum of highest order $(1, 2)$ with a corresponding one-dimensional structure.

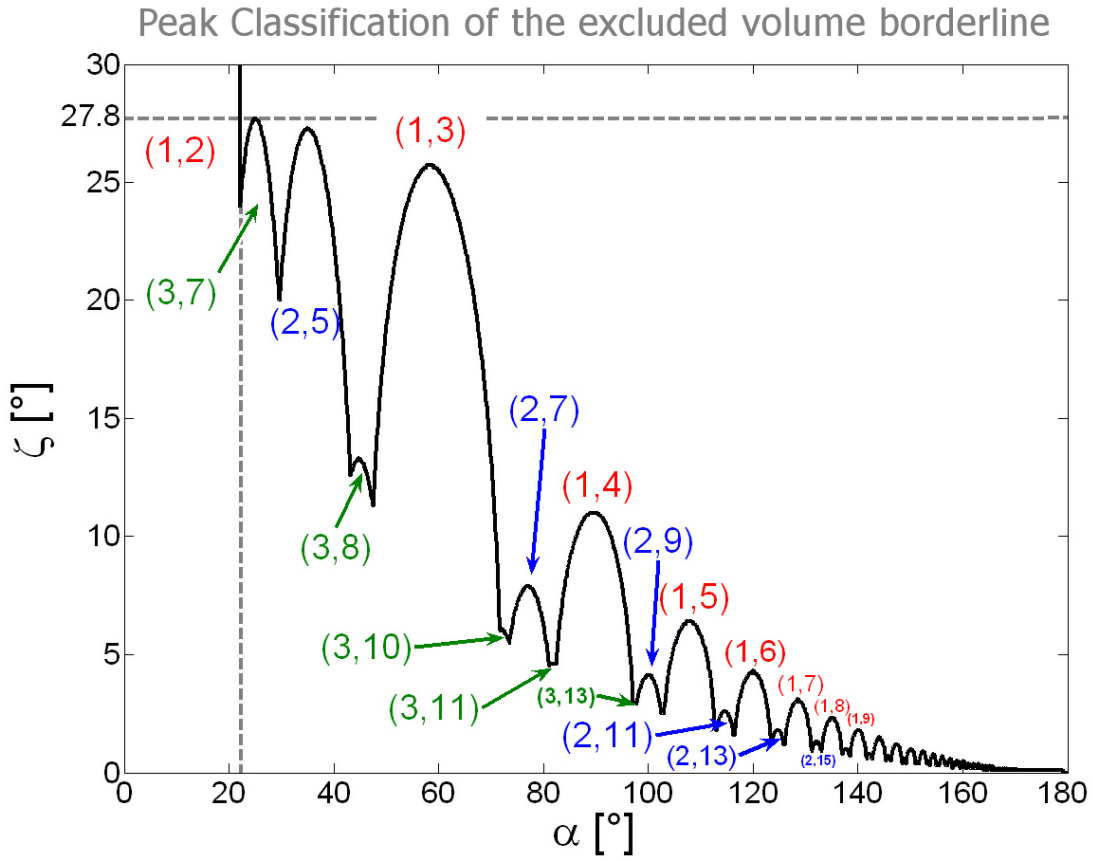


Figure 46: Classification of the peaks which determine the envelope of ζ. The numbers in the brackets are the classification of the maxima (n, i) i.e. the order n and the number of tips i of the corresponding planar structure. Exact values can be found in App. B (Tab. 10 - 14). The dashed lines give the rough estimations which were calculated in Sec. 5.4.1. Adapted from [61].

Because $\Delta_i^n(\tilde{\beta}) \to 0$ for $\alpha_i^n \to \pi$ the maxima of $\zeta(\alpha)$ converge towards 0 for $\alpha_i^n \to \pi$. There are infinitely many α_i^n for $n \to \infty$ and the distance Δ_i^n converges to zero (cf. Sec. 5.3.1), so $\zeta(\alpha)$ has infinitely many local maxima and minima for $\alpha_i^n \to \pi$.

It is clear that $\zeta(\pi) = 0$ as the fiber forms a fully stretched fiber (a circle of radius $r = \infty$).

So far the $\zeta(\alpha)$ is only known at the positions of the maxima $\alpha = \alpha_i^n$. Now consider values of α which are close to an α_i^n, say $\alpha' = \alpha_i^n \pm \Delta\alpha$. *Close* means such $\Delta\alpha$ which lead to a shift $\Delta a < 2r$. This leads to a slight shift Δa of those nucleosomes which are located at the same places namely k and $k + \left(\frac{n \cdot 2\pi}{\pi - \alpha}\right)$.
At $\beta = 0$ this shift Δa is orthogonal to the fiber's axis. This time Δ_i^n still denotes the distance between the nucleosomes k and $k + \left(\frac{n \cdot 2\pi}{\pi - \alpha}\right)$ along the fiber's axis but now they are not located at the same spots but instead slightly shifted. Therefore, their distance Δ is no longer equal to their distance Δ_i^n along the axis when increasing β.
In this case increasing β still leads to a movement along the vertical axis of the fiber but now the distance Δ of nucleosome k and $k + \left(\frac{n \cdot 2\pi}{\pi - \alpha}\right)$ increases like $\Delta^2 \approx \Delta a^2 + (\Delta_i^n)^2$.
The fiber fulfils the excluded volume conditions when $\Delta = 2r$ which means

$$\Delta_i^n = \sqrt{4r^2 - \Delta a^2} < 2r.$$

So the critical value of Δ_i^n, which has to be achieved to fulfil excluded volume decreases with increasing Δa.
Therefore, $\zeta(\alpha)$ has always a *local maximum* at α_i^n: $\zeta(\alpha_i^n \pm \Delta\alpha) < \zeta(\alpha_i^n)$.

To calculate $\Delta a_{i,n}(\Delta\alpha)$ imagine a planar structure of j nucleosomes with an entry-exit-angle $\alpha' = \alpha_i^n \pm \Delta\alpha$ of two consecutive octamers. The locations of these nucleosomes are denoted by $p_0,\ p_1,\ p_2,\ ...,\ p_{j-1} \in \mathbb{R}^3$. Without loss of generality one can assume

$$p_0 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},\ p_1 = \begin{pmatrix} b & \cos(\frac{\alpha'}{2}) \\ b & \sin(\frac{\alpha'}{2}) \\ & 0 \end{pmatrix} \text{ and}$$

$$\forall k \geq 2\ :\ p_k = p_{k-1} + \mathcal{R} \cdot (p_{k-1} - p_{k-2})$$

$$\text{with } \mathcal{R} := \mathcal{R}^z_{\pi - \alpha'} = \begin{pmatrix} \cos(\pi - \alpha') & \sin(\pi - \alpha') & 0 \\ -\sin(\pi - \alpha') & \cos(\pi - \alpha') & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

the rotational matrix along the z-axis. This leads to

$$p_k = \sum_{m=0}^{k-1} \mathcal{R}^m \cdot p_1 \quad \forall\ k > 0.$$

Now, $\Delta a_{i,n}$ is given by $\Delta a_{i,n} = \|p_{k=i}\|$ which leads to

$$\Delta a_{i,n} = \left\| \sum_{m=1}^{i} \mathcal{R}^{m-1} \cdot p_1 \right\| \tag{36}$$

(were $\mathcal{R}$ and p_1 depend on i, n and $\Delta\alpha$). Now one can use Eq. 33 to calculate ζ around the maximal values α_i^n:

$$\frac{n \cdot \left(\frac{2\pi}{\pi-\alpha'}\right)\ b\,\zeta(\alpha')}{\sqrt{\zeta(\alpha')^2+(\pi-\alpha')^2}} \stackrel{!}{=} \sqrt{4r^2-\Delta a^2}$$

$$\stackrel{\zeta>0}{\Rightarrow}\ \zeta(\alpha'=\alpha_i^n\pm\Delta\alpha)=\sqrt{\frac{(4r^2-\parallel\sum_{m=1}^{i}\mathcal{R}^{m-1}\cdot p_1\parallel^2)(\pi-\alpha')^2}{b^2\,n^2\left(\frac{2\pi}{\pi-\alpha'}\right)^2-4r^2+\parallel\sum_{m=1}^{i}\mathcal{R}^{m-1}\cdot p_1\parallel^2}} \tag{37}$$

Similar to the case of large α one can find equations for ζ of small α: Eq. 29 leads to

$$\Delta_i^n(\tilde{\beta})=\frac{1}{2}n\left(\frac{2\pi}{\pi-\alpha'}\right)\ \tilde{\beta}\ b\ \tan(\frac{\alpha'}{2})$$

with $\alpha'=\alpha_i^n\pm\Delta\alpha$. Now again $\zeta(\alpha')$ follows from $\tilde{\beta}(\Delta_i^n=\sqrt{4r^2-\Delta a_{i,n}^2})=\zeta(\alpha')$:

$$\Rightarrow\ \zeta(\alpha'=\alpha_i^n\pm\Delta\alpha)=\frac{\cot(\frac{\alpha'}{2})}{n\left(\frac{2\pi}{\pi-\alpha'}\right)\cdot b}\sqrt{4r^2-\parallel\sum_{m=1}^{i}\mathcal{R}^{m-1}\cdot p_1\parallel^2} \tag{38}$$

$|\Delta\alpha|<c$ fulfiling $\Delta a(c)=2r$ gives the interval of the allowed α-values: $\alpha'\in[\alpha_i^n-c,\ \alpha_i^n+c]$ for a certain peak (n,i).

The predictions of Eq. 37 and 38 are shown in Fig. 47 together with simulation results.

In these simulations regular chromatin fibers with given model parameters were generated and then it was decided, whether or not the excluded volume conditions were violated. Thus one gets results for the excluded volume borderline which are independent of the approximations that have been made during the analytical calculations of the borderline.

So the numerical results are more accurate than the analytical ones but the analytically achieved results supply a profound understanding of the excluded volume borderline.

One can see in Fig. 47 that the numerically obtained exact curve is in good agreement with the analytically calculated one, although some approximations had to be made.

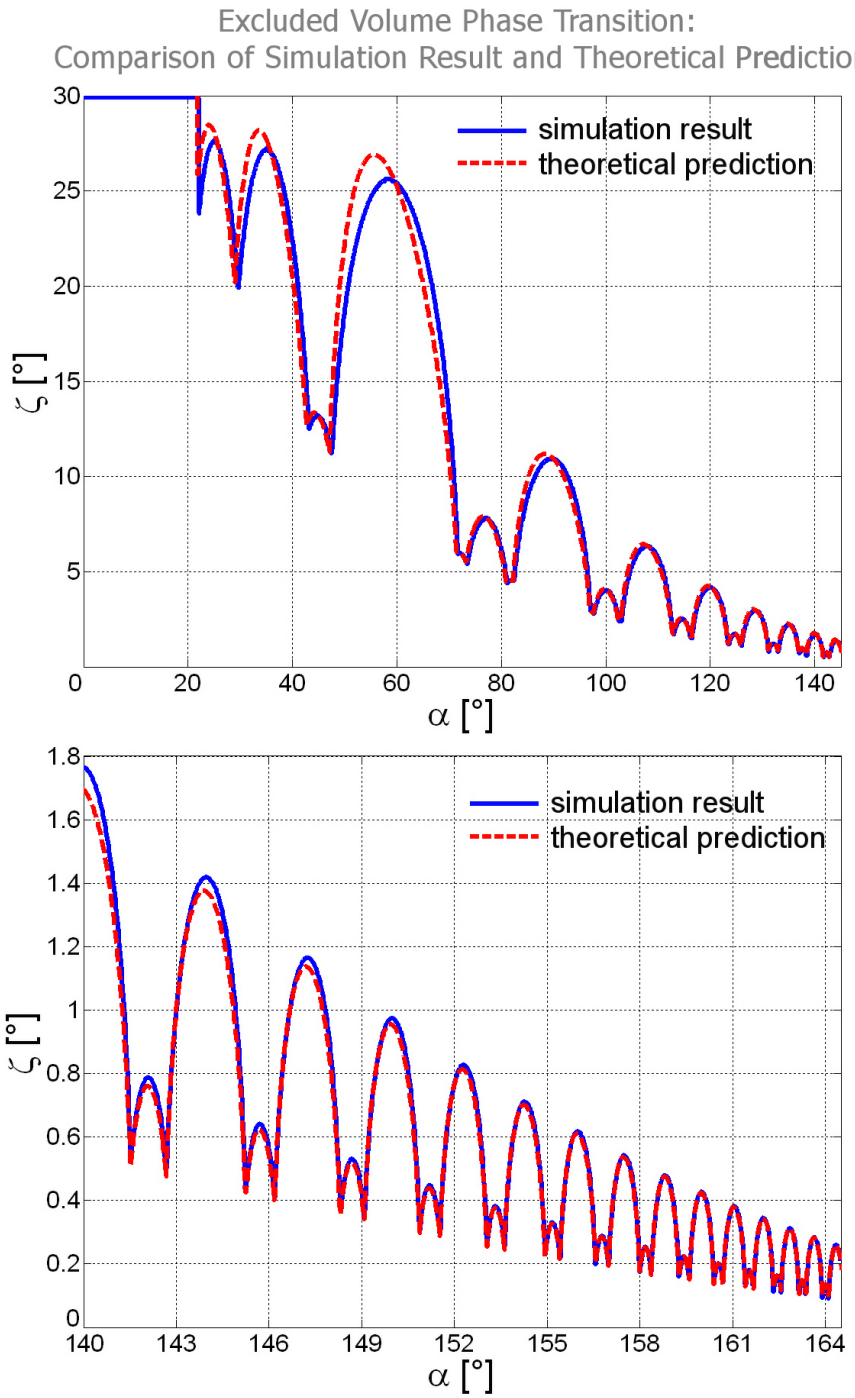


Figure 47: These two figures show the calculated theoretical prediction (dashed red line) and the simulation result (blue line) for the excluded volume borderline $\zeta(\alpha)$. The first figure shows the behavior of ζ for small α up to 140° according Eq. 38. The second one shows the behavior of $\zeta(\alpha)$ for α values larger than 140° according to Eq. 37. The nice agreement of the prediction and the numerical result shows that the assumed estimations lead to suitable expressions for $\zeta(\alpha)$. Adapted from [61].

5.4.3 Remarks

The numerical results and the analytical prediction do not match perfectly (cf. Fig. 47). Some of the reasons for this are explained in this section supplemented with further remarks.

a) The *Effective* Entry-Exit Angle α_e
The *effective* entry-exit angle α_e is the projection of α onto a plane which is orthogonal to the axis of the master solenoid (cf. Sec. 4.3.3). It decreases with increasing twist angle β because the length of the fiber increases with increasing β.
α_e is important for the calculation of the number of linkers which form a closed loop:

$$N_l := \frac{2\pi}{\pi - \alpha_e}$$

is the number of multiples of α_e to get a full loop (i.e. 2π). If $\frac{2\pi}{\pi-\alpha_e}$ is an integer this will be equal to the number of linkers which corresponds to a full loop. If it is not an integer number, then the fractional part will give the fraction of $\pi - \alpha_e$ which is missing for a full loop but the entry-exit angle is fixed here and it is assumed that $N_l = \frac{2\pi}{\pi-\alpha_e}$.

In the calculations above the assumption $\alpha_e = \alpha = \text{const.}$ was made because only small β were considered. α_e will be calculated as a function of α and β below. It converges towards 0 for $\beta \to \pi$ and is shown in Fig. 48 for different values of α and β.
To calculate α_e consider three consecutive nucleosome locations within a given fiber: n_0, n_1 and $n_2 \in \mathbb{R}^3$. Without loss of generality one can assume: $n_0 = p_0$, $n_1 = p_1 + p_0$ and $n_2 = p_2 + p_1 + p_0$ were the p_i are the following linker vectors:

$$p_0 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \; p_1 = \begin{pmatrix} \sqrt{b^2 - d^2} \\ 0 \\ d \end{pmatrix} \text{ and } p_2 = \begin{pmatrix} x_2 \\ y_2 \\ d \end{pmatrix} .$$

This means that the z-axis of the coordinate system is the axis of the fiber (d is given by Eq. 15 and b is the linker length). Now $\|p_2\| = b$ gives

$$y_2 = \sqrt{b^2 - d^2 - x_2^2}$$

and x_2 can be calculated by using

$$\cos(\alpha) = \frac{\langle -p_1 \mid p_2 \rangle}{b^2}$$

$$\Rightarrow x_2 = -\frac{b^2 \cos(\alpha) + d^2}{\sqrt{b^2 - d^2}}$$

$$\text{with } d = \frac{b \sin\left(\frac{\beta}{2}\right)}{\sqrt{\csc^2\left(\frac{\alpha}{2}\right) - \cos^2\left(\frac{\beta}{2}\right)}} \text{ this leads to}$$

$$y_2 = \sqrt{b^2 - d^2 - \frac{(b^2\cos(\alpha) + d^2)^2}{b^2 - d^2}} \overset{d}{=} \sqrt{b^2 \cos^2\left(\frac{\beta}{2}\right) \sin^2(\alpha)}$$

and now α_e follows from the projection onto the x-y-plane:

$$\frac{\sin(\alpha_e)}{2} = \frac{\sqrt{(x_1 + x_2)^2 + y_2^2}}{2\sqrt{x_2^2 + y_2^2}}$$

$$\Rightarrow \sin\left(\frac{\alpha_e}{2}\right) = \sqrt{\frac{\cos\left(\frac{\beta}{2}\right) \sin^2\left(\frac{\alpha}{2}\right)}{2(1 + \cos(\alpha))}}. \tag{39}$$

As Eq. 39 shows α_e decreases with increasing β. Therefore, N_l decreases, too: If $\alpha_e \to 0$ then $N_l(\alpha, \beta) = \frac{2\pi}{\pi - \alpha_e(\alpha,\beta)} \to 2$. So all fibers with high values of beta need only approximately 2 nucleosomes for a complete loop. This is confirmed by Fig. 42 and 43.
Since α_e was assumed to be constant, N_l was also constant in the calculations above. This is a suitable approximation for small β. For larger β the assumed values of N_l were a bit too large, and therefore, the calculated values of ζ were a bit too small but this effect is small compared to the other estimations which were made above. The error of ζ, due to the assumption that N_l is constant, increases with increasing β.

One way to increase the accuracy in the calculations above would be to replace the term $\frac{2\pi}{\pi - \alpha'}$ by $\frac{2\pi}{\pi - \alpha_e(\alpha',\beta)}$, and furthermore, one should use Eq. 15 to derive exact equations in order to replace Eq. 28 and 29. But this can only be done by using numerical methods. Moreover, Fig. 47 shows that the assumed approximations lead to useful equations which describe ζ in a suitable way.

b) The Torsion of the Fiber

Another effect which has been neglected, but not mentioned explicitly before, is the fact that with increasing β the fiber's torsion τ increases, too. This can be seen easily by considering the fact that N_l decreases with increasing β.
Therefore, all the maxima of $\zeta(\alpha)$ are slightly shifted to larger values of $\alpha = \alpha_i^n + \varepsilon$: It was assumed that increasing β leads only to a *vertical* movement of overlapping nucleosomes but in fact there is a *horizontal* one, too. Together both induce the torsion of the fiber.

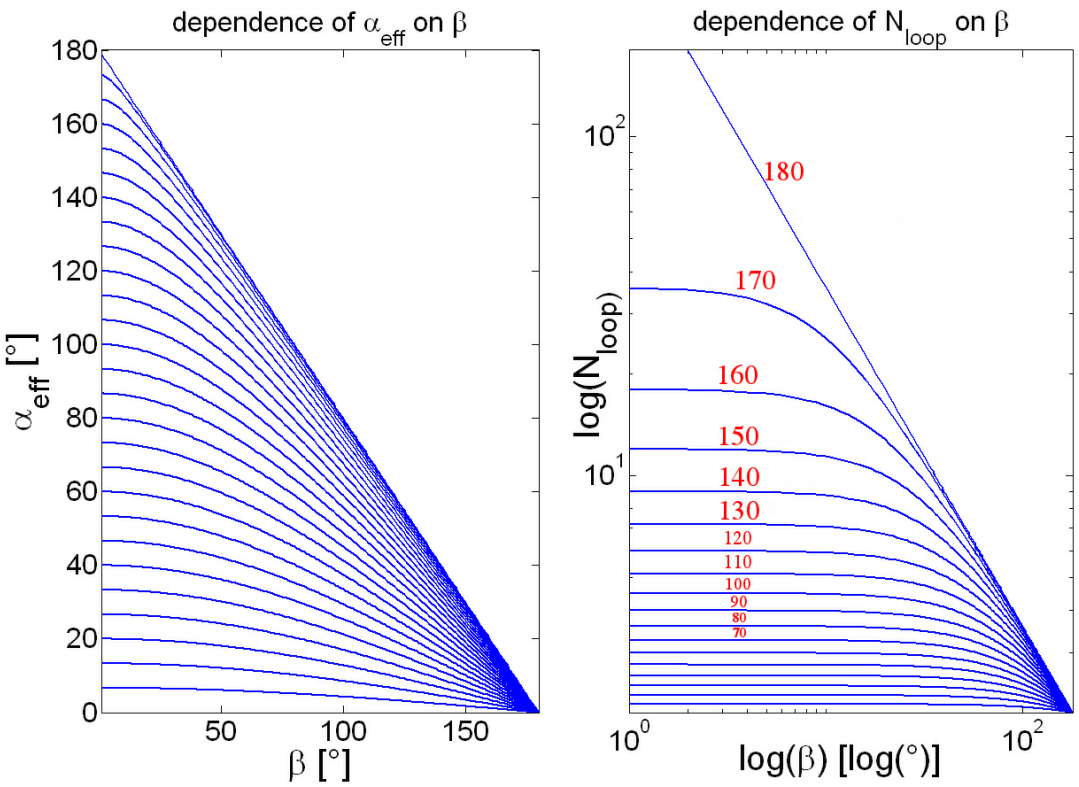


Figure 48: The effective entry-exit angle α_e depends on α and β as shown in the left figure. The values of $\alpha = \alpha_e(\alpha, \beta = 0)$ can be seen at the left side of the figure. Because α_e depends on α and β the number of linkers forming a closed loop N_l shows this dependence, too. It is shown in the right figure as a double-log-plot. The red numbers give α in degrees - increasing from 10° to 180° by steps of 10°. Adapted from [61].

The horizontal movement of the overlapping nucleosomes k and $k+\frac{n\cdot 2\pi}{\pi-\alpha_i^n}$ implies a shift along the rotation direction of the fiber. Therefore, at the points $\alpha_i^n - \Delta\alpha$ increasing β moves the nucleosomes vertically *and* horizontally away from each other, and their distance is increased even more than predicted by the calculation.
This leads to an earlier stepping across the excluded volume borderline i.e. ζ should be decreased.
Otherwise such α' that are a bit larger than α_i^n have a lower Δa because increasing β causes a vertical movement of the overlapping nucleosomes away from each other but a horizontal one towards each other. This effect should increase with the order n of the maxima and with α because the larger the number of nucleosomes between two overlapping ones the stronger the effects which are caused by a rise of β.
So altogether the increase of β rises the torsion τ of the fiber and therefore the maxima of $\zeta(\alpha)$ should be shifted slightly towards higher values of α. But unfortunately the

simulation results do not show this effect. The 60°-peak is even shifted towards lower α.

c) Fractal Dimension of $\zeta(\alpha)$

One can try to use a box counting algorithm to determine the Haussdorff-dimension of $\zeta(\alpha)$. This results in a dimension of one i.e. one finds no fractal dimension at all. The reason for this is that $\zeta(\alpha)$ might show self-similarity only for $\alpha \to \pi$ and these states are very hard to sample (cf. next issue).

From the theoretical point of view there could be a critical value of α so that from this point onwards (i.e. $\alpha > \alpha_{crit}$) only peaks with the same orders $n \in \{1, .., n_{max}\}$ occur. The reason for this is, that the higher order peaks might be covered completely by the lower order peaks. If this was the case, the excluded volume would indeed be self-similar and should have a fractal dimension larger than one.

d) Numerical Restrictions

The simulation time τ_s for chromatin fibers with large α values goes to infinity for $\alpha \to \pi$ since the amount of linkers which form a whole loop converges towards infinity. But to decide whether or not a chromatin fiber violates the excluded volume conditions or to calculate reasonable physical properties of the chromatin fiber it is necessary to generate fibers with multiple complete loops. Because of limited simulation time only fibers up to $\alpha = 179.5°$ were generated.

5.4.4 Influence of Cylindrically Shaped Nucleosomes and Pitch

In this section the influence of the cylindrical (instead of spherical) excluded volume of the nucleosomes and a non-vanishing orthogonal distance between in- and outgoing DNA strand on the phase diagram of chromatin are determined. Both of these parameters have been neglected in chromatin research so far [62; 184].

In order to do so, simulations [24] of regular chromatin fibers have been made and it was checked, whether they fulfil the excluded volume conditions or not. The results were plotted in a Ramachandran-like diagram again (cf. Fig. 49) to find out which states of the whole phase diagram are forbidden by excluded volume interactions and which are not.
The fibers that have been simulated were *regular* again, i.e. all linker lengths, entry-exit-angles, rotational angles and pitches were fixed. The cylinders had a height of $16.8\langle bp\rangle$, and a diameter of $33.0\langle bp\rangle$ according to [233].
They were orientated by using the vectors p_i (cf. Sec. 4). Moreover, a DNA diameter of $6.6\langle bp\rangle$, a twist length of $10.2\langle bp\rangle$, a mean linker length of $63\langle bp\rangle$ and that 1.65 turns of DNA are wrapped around the histone octameres (cf. Tab. 1) was assumed.

The colored lines in Fig. 49 represent the phase transition between allowed and forbidden states for different pitches d. All states below the corresponding line are forbidden, those above it are allowed.
One can interpret Fig. 49 as many cuts through the actually four-dimensional phase diagram which have been piled. The states near the excluded volume borderline are the most interesting of the phase diagram since they are the most compact ones (cf. Sec. 5.5). Therefore, the gaps in the borderline might be used by the fiber to become (at least locally) very dense.

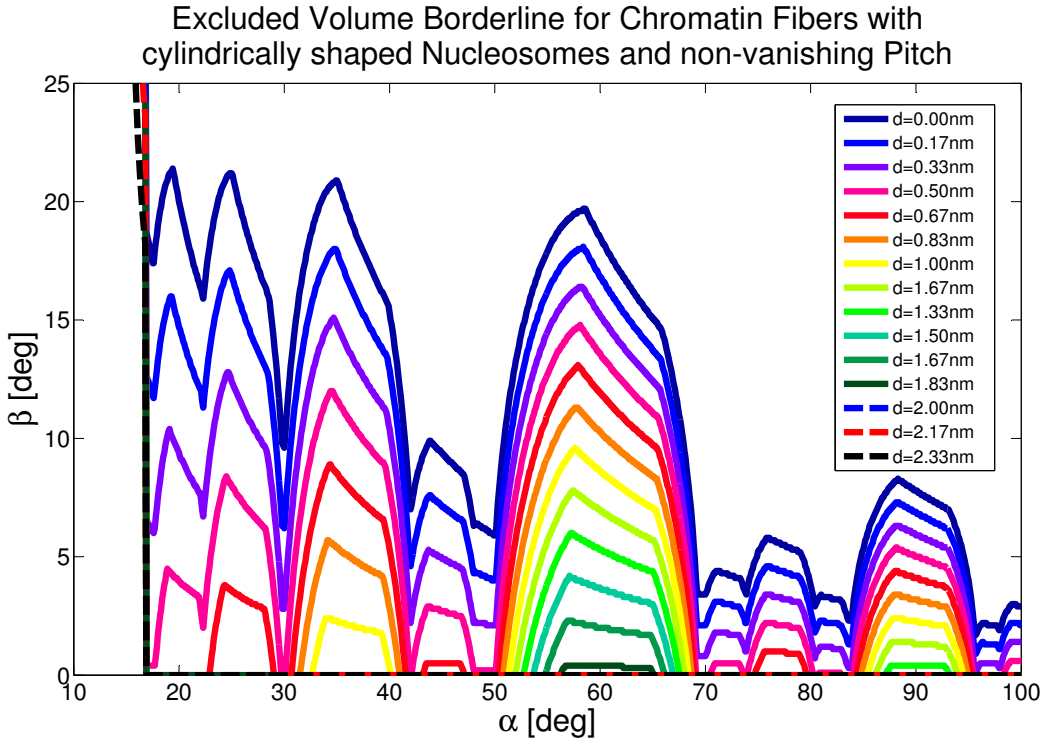


Figure 49: Cut-out of the chromatin phase diagram (for different d). The states below the corresponding lines are forbidden due to excluded volume interactions. With increasing d more and more states become accessible to the fiber.

There is another borderline at the left side of the diagram which prevents α from getting smaller than some minimal value $\alpha_{min}(d)$ that depends on the pitch d. This "arrow-like" structure can be seen best in Fig. 57, 58 and 60. It shifts towards larger $\alpha_{min}(d)$ with increasing d.
The gap in this line is a further consequence of the *cylindrical* excluded volume and can not be seen in the phase diagram for spherical nucleosomes [62] (cf. Fig. 42 and Fig. 45). It will be discussed in Sec. 5.5.2.

Fibers with a small entry-exit angle α and $\beta \ll \pi$ resemble zig-zag-structures. Those with large entry-exit angles and $\beta \approx \pi$ look like solenoids. A comprehensive discussion of the different structures of the phase diagram can be found in [62; 184] or Sec. 5.3 and Sec. 5.5. Here only the changes due to cylindrical (instead of spherical) excluded volume are addressed.

As a consequence of the cylindrical instead of spherical excluded volume of the nucleosomes the shape of the peaks in the phase diagram is changed: Their top shows a wedge-like shape due to the edges of the cylinders. With increasing entry-exit-angle α there is more space between the nucleosomes which leads to a larger variety of allowed rotational angles β and thus to the missing tip at the top of the peaks. This effect gets weaker with increasing d: The edges which cut the peaks become more parallel to the α-axis.

With increasing orthogonal distance between in- and outgoing DNA strand there is also more space between consecutive nucleosomes which leads to a decrease of the borderline. More and more states become accessible and with $d = 5.5\langle\text{bp}\rangle$ the borderline almost vanishes.
The natural mean value of d is approximately $d = 2.8nm = 8.4\langle\text{bp}\rangle$ due to 1.65 turns of DNA with a diameter of $2.2nm$. One can see from Fig. 49 that the excluded volume borderline seems to play a role only for lower pitches.
Lower d-values might occur where the DNA has less turns around the histone complex and, more importantly, when the DNA linkers *bend slightly*. It has been shown that nucleosomes tend to stack on top of each other which would lead to such a bending of the linkers.
During the simulations all DNA linkers were *straight*. In real chromatin fibers (especially with attractive nucleosome-nucleosome interaction) the linkers have the opportunity to bend slightly and thus there probably remain some forbidden regions in the phase diagram, although d is large. The persistence length of blank DNA is about 50nm.

Nevertheless, the simulations show that for *straight* linkers the excluded borderline will vanish almost completely, if d approaches realistic values.

5.5 Chromatin Density and Fiber Accessibility in the Phase Diagram

The problem of DNA-folding was explained in Sec. 2: Inactive chromatin should be packed as densely as possible because of the very large ratio of DNA length to nucleus size. But obviously on the other hand the DNA has to be locally accessible for gene transcription. This optimization problem of simultaneously high compaction and high local accessibility will be discussed in this section with respect to the chromatin phase diagram.
First, chromatin structures without pitch ($d = 0$) and with *spherical* excluded volume of the nucleosomes will be considered (cf. Sec. 5.5.1) and afterwards the cylindrically shaped nucleosomes and different pitches will be factored in (cf. Sec. 5.5.2).

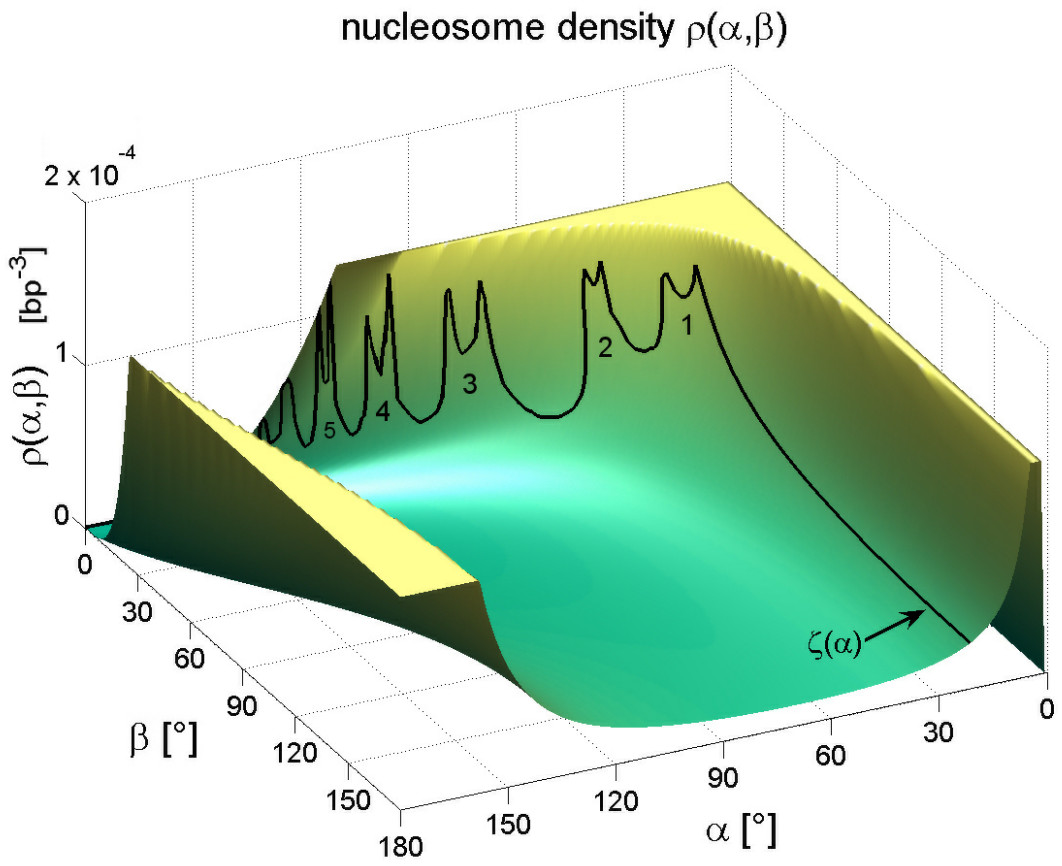


Figure 50: The nucleosome density $\varrho(\alpha, \beta)$ increases towards the excluded volume phase transition which is projected onto the density surface (black line). The density diverges for large α, too, because the radius of the corresponding cylinder converges towards 0, so this is an artefact of the simulation without further physical meaning. There is a cutoff for $\varrho(\alpha, \beta)$ at $1.31 \cdot 10^{-4}$ bp^{-3} $\approx 35.4 \cdot 10^{-4}$nm^{-3} which is the maximum the density achieves within the excluded volume restrictions. But this point of the phase diagram is energetically very unfavorable (cf. Fig. 54). Adapted from [61].

5.5.1 Spherical Nucleosomes without Pitch

In this section chromatin fibers with spherical excluded volume potential for the nucleosomes and no pitch d will be considered. First, the nucleosome density and the nucleosome line density are investigated. The latter one can be measured experimentally. Then the accessability of the fibers is defined and examined.

a) Nucleosome Density and Nucleosome Line Density

The nucleosome density and the line density of the nucleosomes can be used as a measure for the compaction of a 30nm fiber. The basic volume for the 3D-density is a cylinder: Its radius was numerically minimized with the restriction that all nucleosomes have to fit into its interior. The results can be seen in Fig. 50: As expected the nucleosome density achieves its maximal values in the gaps of the excluded volume borderline $\zeta(\alpha)$.

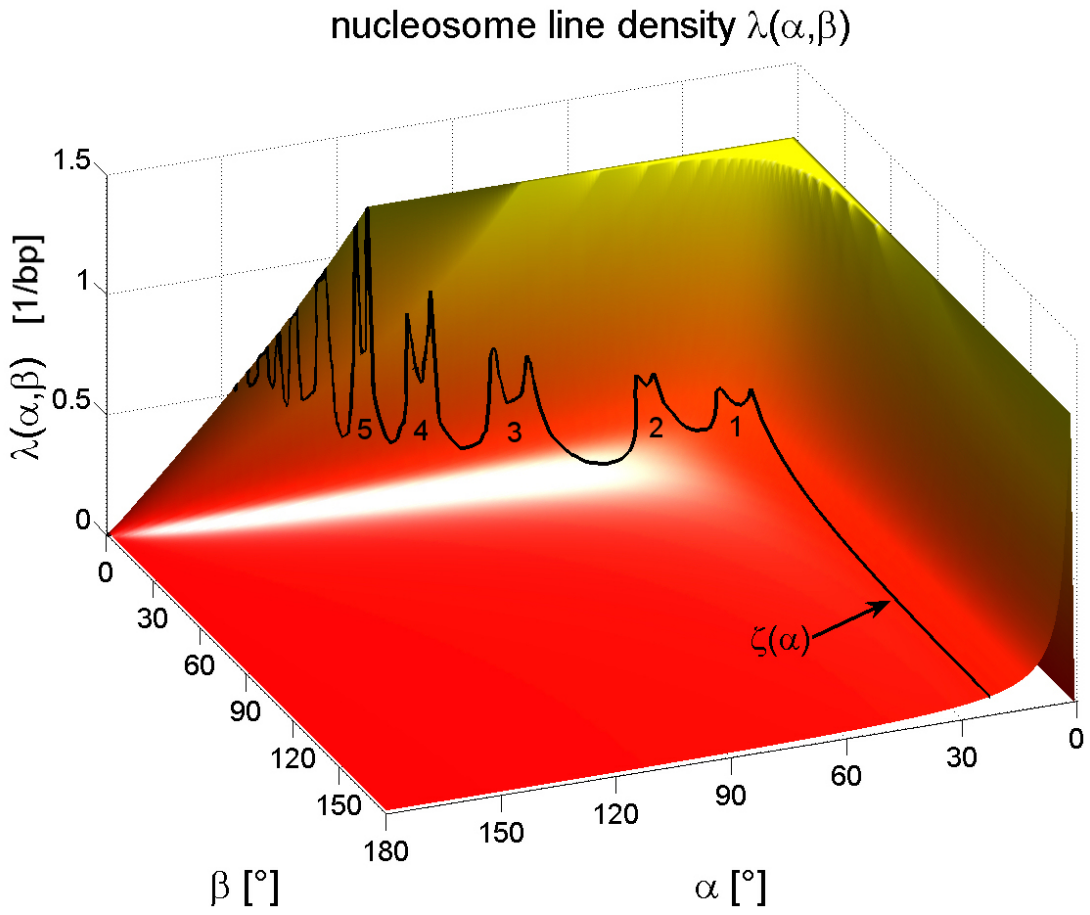


Figure 51: The chromatin line density λ: it diverges for $\alpha \to 0$ because the length of the fiber's axis converges towards zero. λ increases with increasing gap number. There is a cutoff of $\lambda(\alpha, \beta)$ at its maximum value within the excluded volume restrictions. The black line is the projection of the excluded volume borderline $\zeta(\alpha)$ onto the line-density surface. Adapted from [61].

With Eq. 13 and 15 one can give an analytical expression for the 3D-density:

$$\varrho(\alpha,\beta) = \frac{N}{\pi R^2 \cdot L_N} = \frac{4\sqrt{\csc^2(\frac{\alpha}{2}) - \cos^2(\frac{\beta}{2})}\ \left(1 - \sin^2(\frac{\alpha}{2})\cos^2(\frac{\beta}{2})\right)}{\pi b^3 \sin(\frac{\beta}{2})\cos^2(\frac{\alpha}{2})}. \tag{40}$$

Note that the nucleosome density is proportionally to the inverse of the cube of the linker length: $\varrho \propto b^{-3}$.

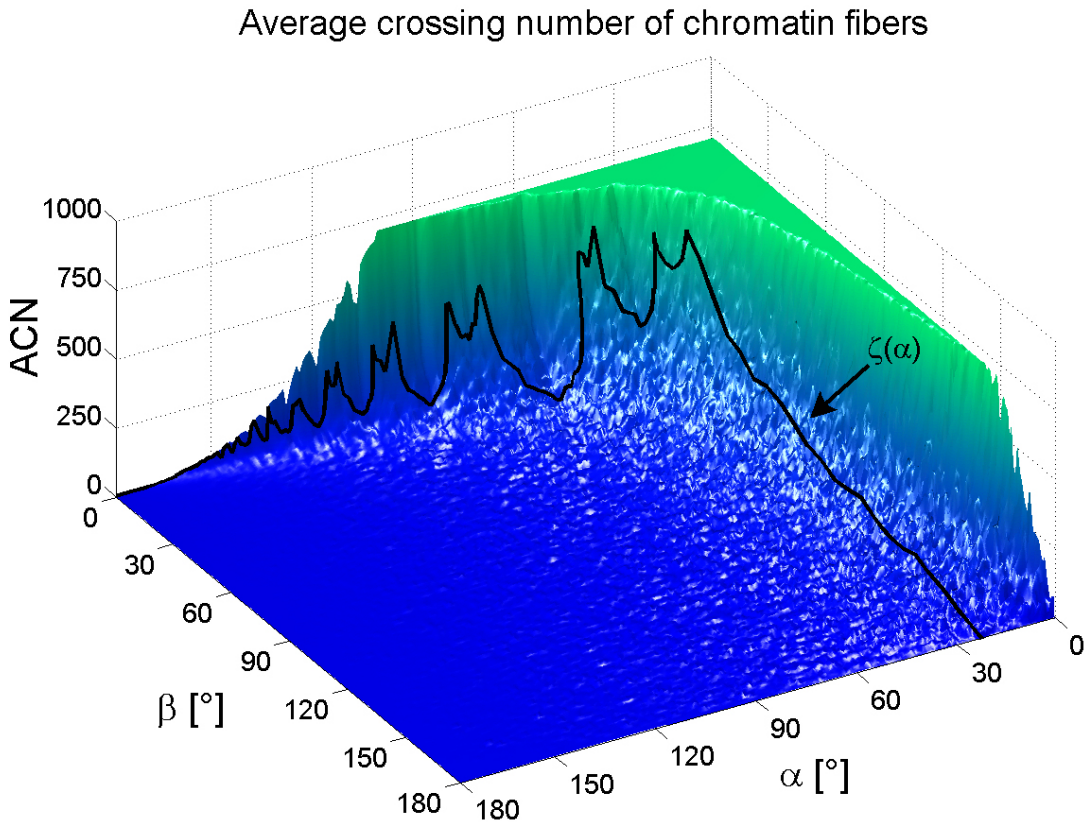


Figure 52: The mean Average Crossing Number (mACN) of different chromatin structures. The black line is again the projection of the excluded volume borderline $\zeta(\alpha)$ onto the ACN-surface. To calculate the average crossing number at a certain point (α, β) 1000 randomly chosen planes were used. But although this number of projections is quite high, the ACN still shows some fluctuations. Nevertheless, it also shows where the interesting dense fibers are located within the phase diagram. There is a cutoff along the z-axis at 700. Adapted from [61].

In order to understand the singularities of the density it is necessary to discuss the behavior of the cylindrical volume $v = \pi r^2 N d$ of the fiber: All structures of either $\alpha = 0$ or $\beta = 0$ are planar, and therefore, possible singularities in the 3D-density.

For $\beta = 0$ the distance d between two consecutive nucleosomes along the chromatin axis is $d_{\beta=0} = 0$ (this can be seen easily from Eq. 15). As $\csc(\alpha) \to \infty$ for $\alpha \to 0$ it follows that $d \overset{\alpha \to 0}{\to} 0$ and $\beta \in [0, \pi]$. For $\alpha = \pi$ Eq. 15 gives

$$d_{\alpha=\pi} = \frac{b \cdot \sin(\frac{\beta}{2})}{\sqrt{1 - \cos^2(\frac{\beta}{2})}} = b.$$

Furthermore, $\beta = \pi$ implies $d_{\beta=\pi} = b \cdot \sin(\frac{\alpha}{2})$. For $\beta = 0$ the radius r of the corresponding spiral structure is given by

$$r_{\beta=0} = \frac{b}{2} \cdot \frac{\cos(\frac{\alpha}{2})}{1 - \sin^2(\frac{\alpha}{2})} = \frac{b}{2} \cdot \csc\left(\frac{\alpha}{2}\right) \overset{\alpha \to \pi}{\to} \infty.$$

If $\alpha = 0$, it will follow from Eq. 13 that $r_{\alpha=0} = \frac{b}{2}$. For $\alpha = \pi$ the radius is $r_{\alpha=\pi} = 0$ ($r(\alpha, \beta)$ is not continuous at $(\pi, 0)$) and for $\beta = \pi$ the corresponding radius is $r_{\beta=\pi} = \frac{b}{2} \cdot \cos(\frac{\alpha}{2})$.

For small α the 3D-density $\varrho = \frac{N}{v}$ goes towards infinity for $\beta \to 0$ because $d \to 0$. With $\beta = 0$ and increasing α, r converges towards infinity but $d = 0$. If α and β are large, the density will diverge because $d \approx b$ and $r \approx 0$.
This is a problem of the chosen basic volume model which is assumed for the fiber and has no further physical meaning. One can see that the most dense fibers are achieved at the border of the excluded volume phase transition, for instance at $\alpha \approx 30°$ (gap #1), $\alpha \approx 45°$ (gap #2) and $\alpha \approx 75°$ (gap #3).
The depth of these gaps is increased by the projection onto the density surface. The maxima of ϱ within the first five gaps are given in Tab. 3.

gap #	ϱ_{max} $[10^{-4}\ bp^{-3}]$	ϱ_{max} $[10^{-4}\ nm^{-3}]$	α_{max} [°]	gap width $\Delta\alpha$ [°]
1	1.1280	30.4560	23	11
2	1.1180	30.1860	43	7
3	1.0726	28.9602	73	12
4	1.1999	32.3973	97	8
5	1.2580	33.9660	113	5

Table 3: Maximal densities in the first five gaps of the excluded volume borderline.

The *line density* λ is the number of nucleosomes of a certain fiber length l divided by the length of the corresponding fiber axis: $\lambda = \frac{N_l}{D_l} = \varrho \cdot \pi R^2 = \frac{1}{d}$. With Eq. 15 this gives:

$$\lambda = \frac{1}{b} \cdot \left(\csc\left(\frac{\beta}{2}\right) \sqrt{\csc^2\left(\frac{\alpha}{2}\right) - \cos^2\left(\frac{\beta}{2}\right)} \right).$$

The results for the line density are shown in Fig. 51. One can see that the line density increases within the excluded volume gaps with increasing α. Furthermore, λ does not

diverge for large α and β like the 3D-density which shows that the divergence of ϱ is an effect due to the vanishing cylindrical volume V of the fiber (i.e. an artefact of the density definition). The line density of the nucleosomes diverges for β towards zero because the length of the fiber's axis goes to zero, too.

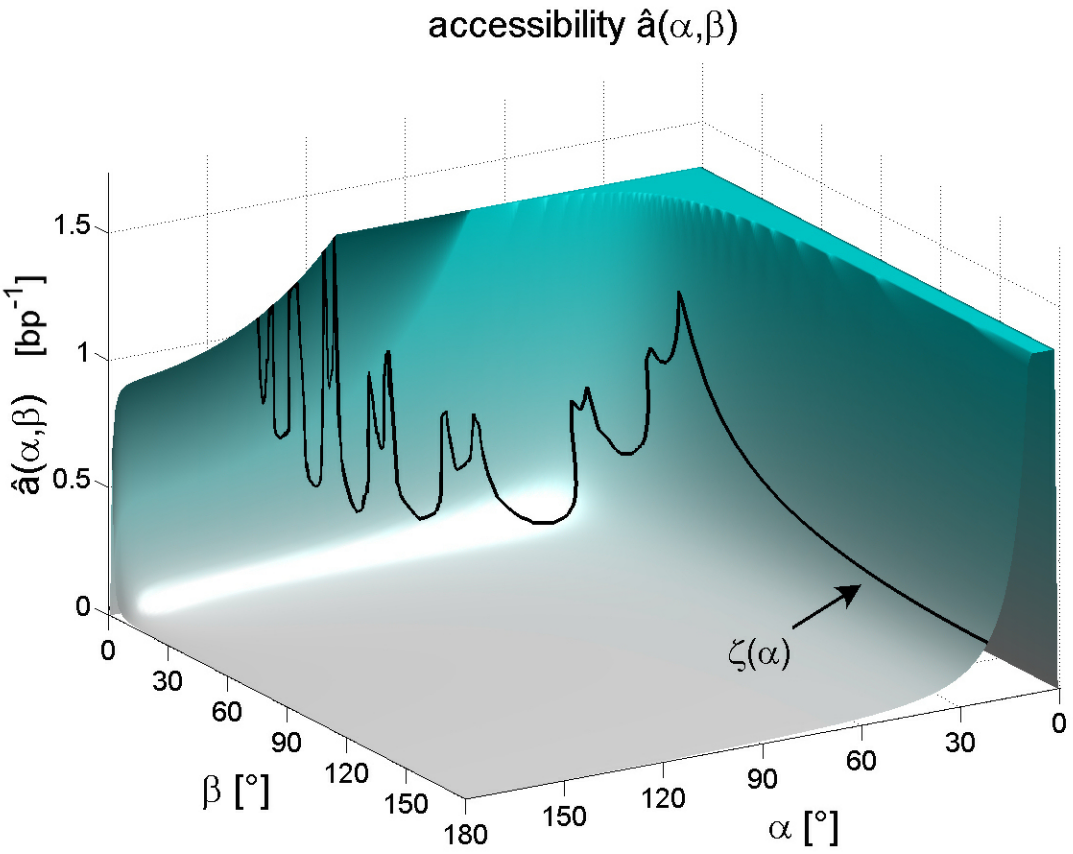


Figure 53: The accessibility of the different chromatin structures. The black line gives the excluded volume borderline. One can see that the fiber accessability achieves its maximum within the allowed surface at the first and the last gaps in the borderline. Adapted from [61].

The critical values of λ for the gaps of the excluded volume borderline can be found in Tab. 4.

The *mean Average Crossing Number* (mACN) (cf. Sec. **??**) is a measure for the entanglement and for the density of a polymer chain as well. Here the mACN is used as another measure for the compactness of the chromatin structures. It brings in some advantages in opposite to the ordinary 3D-density but some disadvantages, too: At first no basic volume is needed like the cylindrical volume for the 3D-density. But the

gap #	λ_{max} $[bp^{-1}]$	λ_{max} $[nm^{-1}]$	α_{max} [°]	gap width $\Delta\alpha$ [°]
1	0.3674	1.081	23	10
2	0.4007	1.179	43	7
3	0.5472	1.609	81	13
4	0.8200	2.412	97	7
5	1.2240	3.600	113	5

Table 4: Maximal line densities in the first five gaps of the excluded volume borderline.

average crossing number is a statistical variable which has statistical fluctuations when calculated numerically.
The average crossing number of chromatin chains is shown in Fig. 52. All chains have a fixed length of 100 NRLs. This lowers the ACN of fibers with large α-values because no complete loops are achieved with only 100 linkers. To calculate the average crossing number at a certain point (α, β) 1000 projections were used. But although this number of projections is quite high the ACN still shows some fluctuations. Nevertheless, it also shows where the interesting dense fibers are located within the phase diagram.

b) DNA Accessibility
As a measure of the *local fiber accessibility* $\tilde{a} := -\frac{\mathrm{d}\lambda}{\mathrm{d}\alpha}$ can be used. It shows how strong the line density λ changes with increasing entry-exit angle α. This is important because the DNA has to be locally accessible for purposes of gene transcription. This local access can for instance be provided by a change of the ionic environment of the DNA because this changes the screening length of the Coulomb repulsion and thus α. The accessability can analytically be calculated by using the equations above:

$$\tilde{a}(\alpha,\beta) = \left(\frac{1}{2b}\right) \frac{\cot(\frac{\alpha}{2})\csc^2(\frac{\alpha}{2})\csc(\frac{\beta}{2})}{\sqrt{\csc^2(\frac{\alpha}{2}) - \cos^2(\frac{\beta}{2})}}. \tag{41}$$

Fig. 53 shows $\tilde{a}$ over the whole chromatin phase diagram. One can see that the highest accessability values in the allowed surface are achieved for small and large values of α and small β. With increasing α the distance of two nucleosomes k and $k+2$ increases. Therefore, the line density is increased by increasing the number of linkers which form a loop but the vertical distance d between two consecutive nucleosomes is only weakly dependent on a change of α. Nevertheless, the vertical distance between two nucleosomes in direction of the fiber axis increases with α as well. It is given by

$$\begin{aligned} d_{vert} &= N_l \cdot d \approx \frac{2\pi d}{\pi - \alpha_e} \\ &= \frac{2\pi b \sin(\frac{\beta}{2})}{\sqrt{\csc^2(\frac{\alpha}{2}) - \cos^2(\frac{\beta}{2})}\left(\pi - 2\arcsin\left(\sqrt{\frac{\cos(\frac{\beta}{2})\sin^2(\frac{\alpha}{2})}{2(1+\cos(\alpha))}}\right)\right)}. \end{aligned}$$

It is interesting to see that the nucleosome density and the accessability are maximized at the same region within the allowed region of the phase diagram.

It is important to mention that the 3D density and the line density of the fiber can not only be changed by changing α or β but also by changing the linker length b (in multiples of 10 bp). A variation in b changes the location of the areas of the phase diagram with the highest accessability and densities. Fibers with smaller values of b can achieve higher 3D densities but have a smaller maximal line density and accessability. From this one might infer that active cells should have larger nucleosome repeat lengths in order to maximize the accessability to their genetic material (cf. [216] for a table of nucleosome repeat lengths in different organisms).

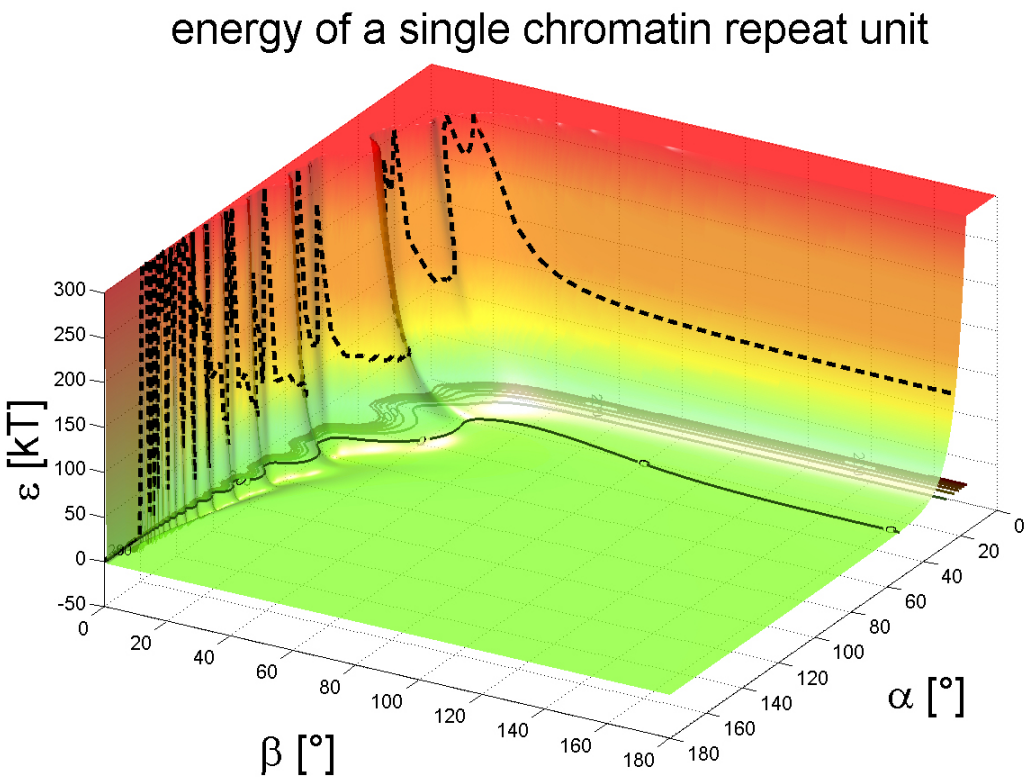


Figure 54: The total energy ϵ of a single chromatin repeat unit. The dashed line gives the excluded volume phase transition and the black lines are the projections of the equipotential lines. Adapted from [61].

b) NRL Energy

In Sec. 4.6 the Coulomb repulsion of the DNA linkers as well as interaction potentials for the nucleosomes have been introduced. Therefore, one can examine the total energy

ϵ of a chromatin repeat unit (NRL), i.e. the Coulomb repulsion and the nucleosome-nucleosome interaction.

The results of chromatin simulations with these interactions are shown in Fig. 54.

One can see that the Coulomb repulsion of the linkers is very high within the gaps of the excluded volume borderline. The repulsion also diverges for the crossed linker fibers when β becomes too small. So the optimal regions in the phase diagram for line density and accessability have a very high Coulomb repulsion, too.

Furthermore, a small potential well occurs along the excluded volume borderline which can be seen best in Fig. 179. Its shape resembles strongly the excluded volume borderline.

Further figures can be found in App. A. The single energy distribution in the case of solely Coulomb repulsions is shown in Fig. 176 and the nucleosome-nucleosome interaction is illustrated in Fig. 177. Furthermore, Fig. 178 provides a top view of the total energy per nucleosome repeat length (NRL) and Fig. 179 shows some cuts through the energy landscape to make the described potential well more visible.

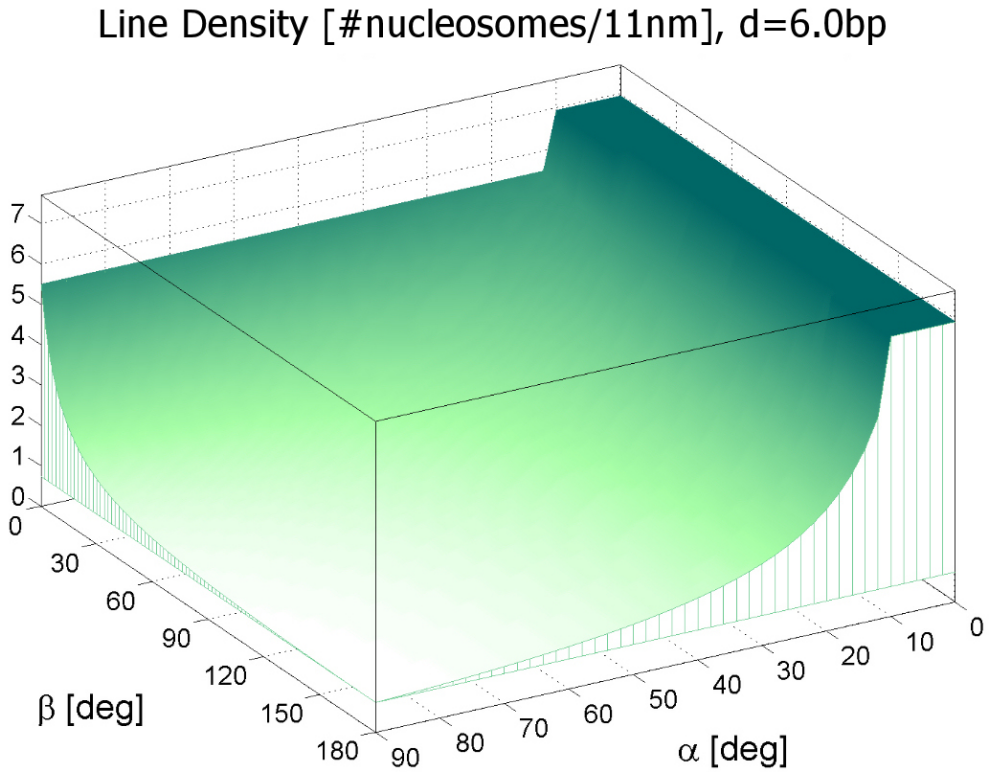


Figure 55: Chromatin line density λ for cylindrical nucleosomes with a pitch d of 6bp (i.e. 2nm). The line densities along the excluded volume borderline coincident with experimental values [34].

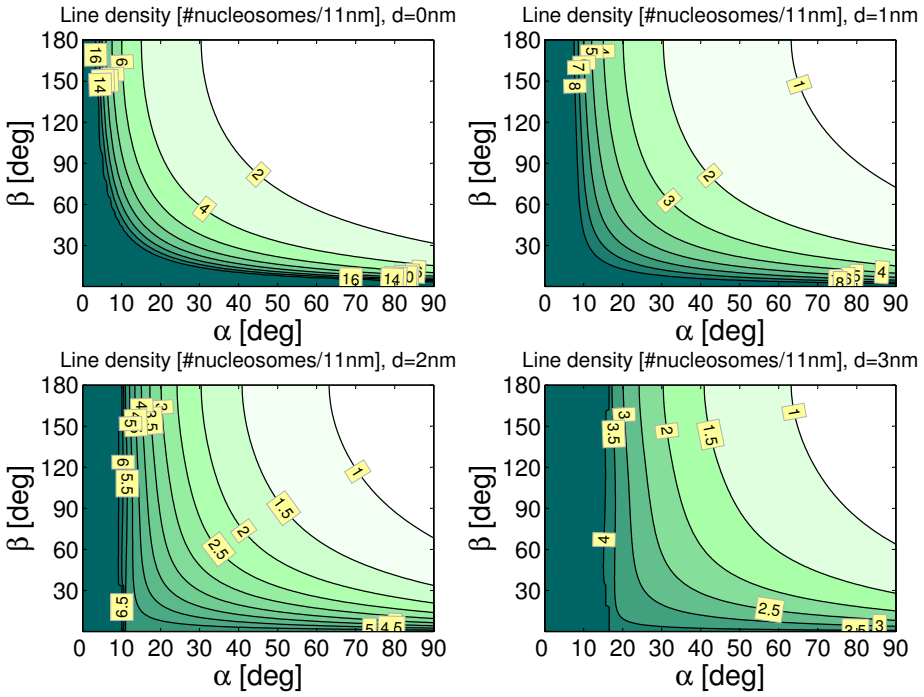


Figure 56: The line density of the nucleosomes within a regular chromatin fiber is large for small β_i. However, the comparison with the phase diagram shows that some of these states are forbidden due to excluded volume interactions. As one can see the line density decreases with increasing pitch d.

5.5.2 Cylindrical Nucleosomes with Pitch

In this section the influence of cylindrically shaped nucleosomes as well as a non-vanishing pitch d between in and outgoing DNA linker will be investigated.
The line density (cf. Fig. 55 and Fig. 56), the radius of gyration (cf. Fig. 57 and Fig. 58) and the mean end-to-end distance (cf. Fig. 59 and Fig. 60) of regular chromatin structures will be investigated along the phase diagram. The line density is historically defined as the number of nucleosomes along a 11nm stretch of the chromatin fiber. The mean radius of gyration R_g as well as the mean end-to-end distance are measures for the spatial extension of the chromatin structures (cf. Sec. 1.2.2). The simulated regular chromatin fibers had a total length of 500 NRLs.

Again the most compact states can be found near the excluded volume borderline (cf. Fig. 56). Furthermore, the line densities which were found in the simulations coincide with experimental values [34]. The line densities at the peaks of the borderline are approximately twice the densities in the gaps. The fiber might use these gaps to compact parts of itself. To do so core histone modifications or changes of the entry-exit angle by changes of the salt concentration might play a role.

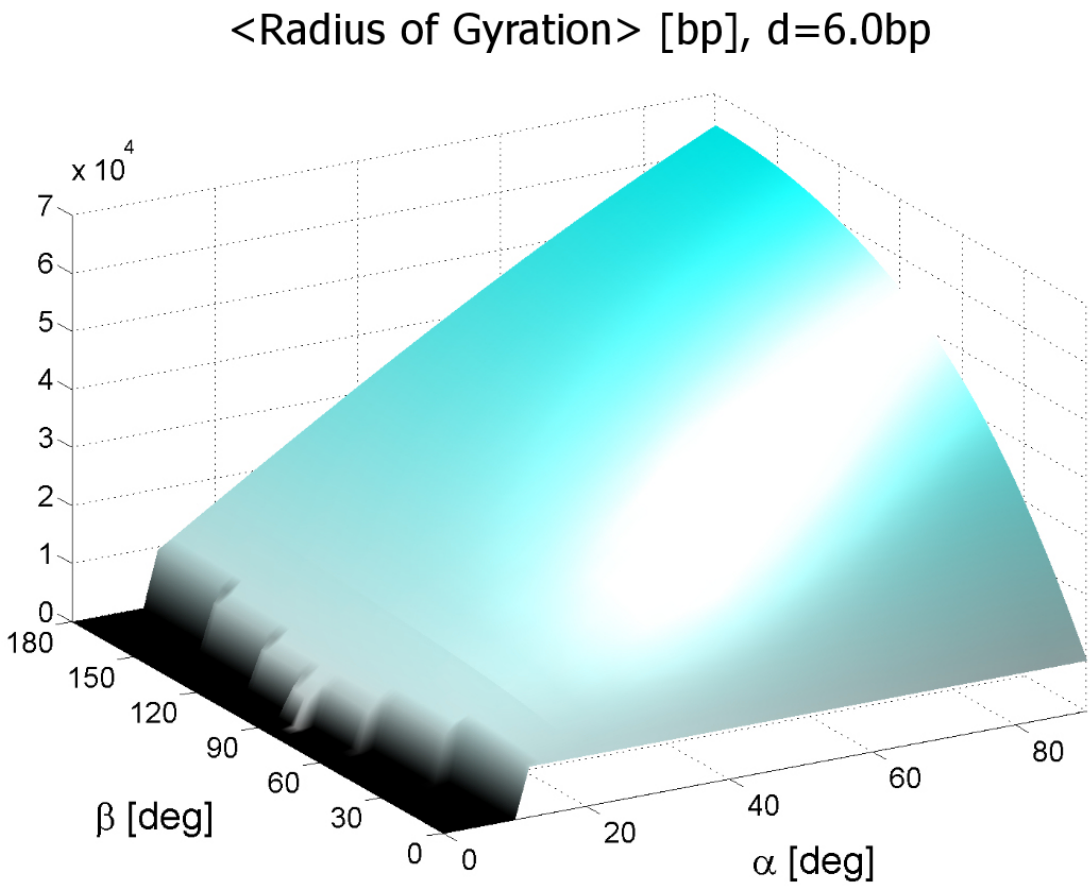


Figure 57: The mean radius of gyration R_g along the chromatin phase diagram. R_g is a measure for the spatial extension of the chromatin fibers. The arrow-like structure at the left side (i.e. small α) is a consequence of the cylindrical excluded volume of the nucleosomes. For the simulations chains with a length of 500 NRLs were used.

Increasing the pitch d decreases the line density (cf. Fig. 56), whereas the mean radius of gyration and the mean end-to-end distance increase (cf. Fig. 56 and Fig. 60).

The borderline at the left side (i.e. small α) of Fig. 57, 58 and 60 that separates allowed from forbidden states due to excluded volume interactions has another shape compared to the previous results for spherical nucleosomes without pitch (Fig. 42).

Previously, there was a cut-off angle α_{min} such that for all $\alpha < \alpha_{min}$ no allowed chromatin structure could be generated. At large this barrier remains but its shape changed: Now there is a arrow-like gap in this part of the excluded volume borderline. The gap in this line is a further consequence of the *cylindrical* excluded volume: If one increases β at $\alpha = \alpha_{min}$, at some point consecutive cylinders will find orientations

where they do not touch each other any more. The most favorable β value is $\beta = 90°$ because then nucleosomes turned by 90° alternate and thus give their neighbors as much free space as possible. This is also the reason why the gap is symmetric with respect to $\beta = 90°$.

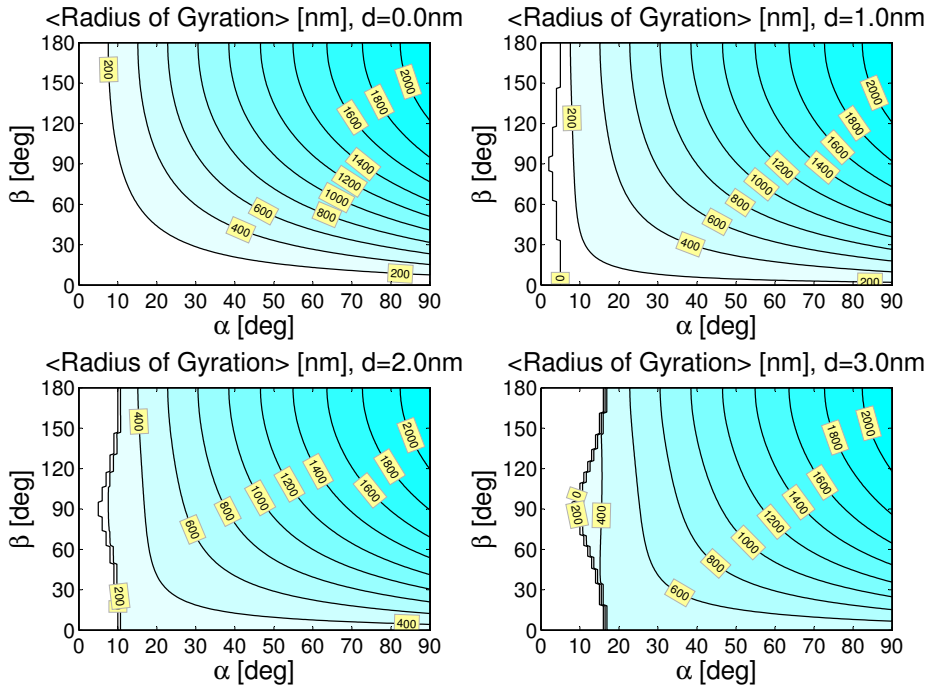


Figure 58: The radius of gyration of regular chromatin fibers along a cut-out of the phase diagram with fixed fiber length of N=500 segments. The compaction of the fibers decreases strongly with increasing d.

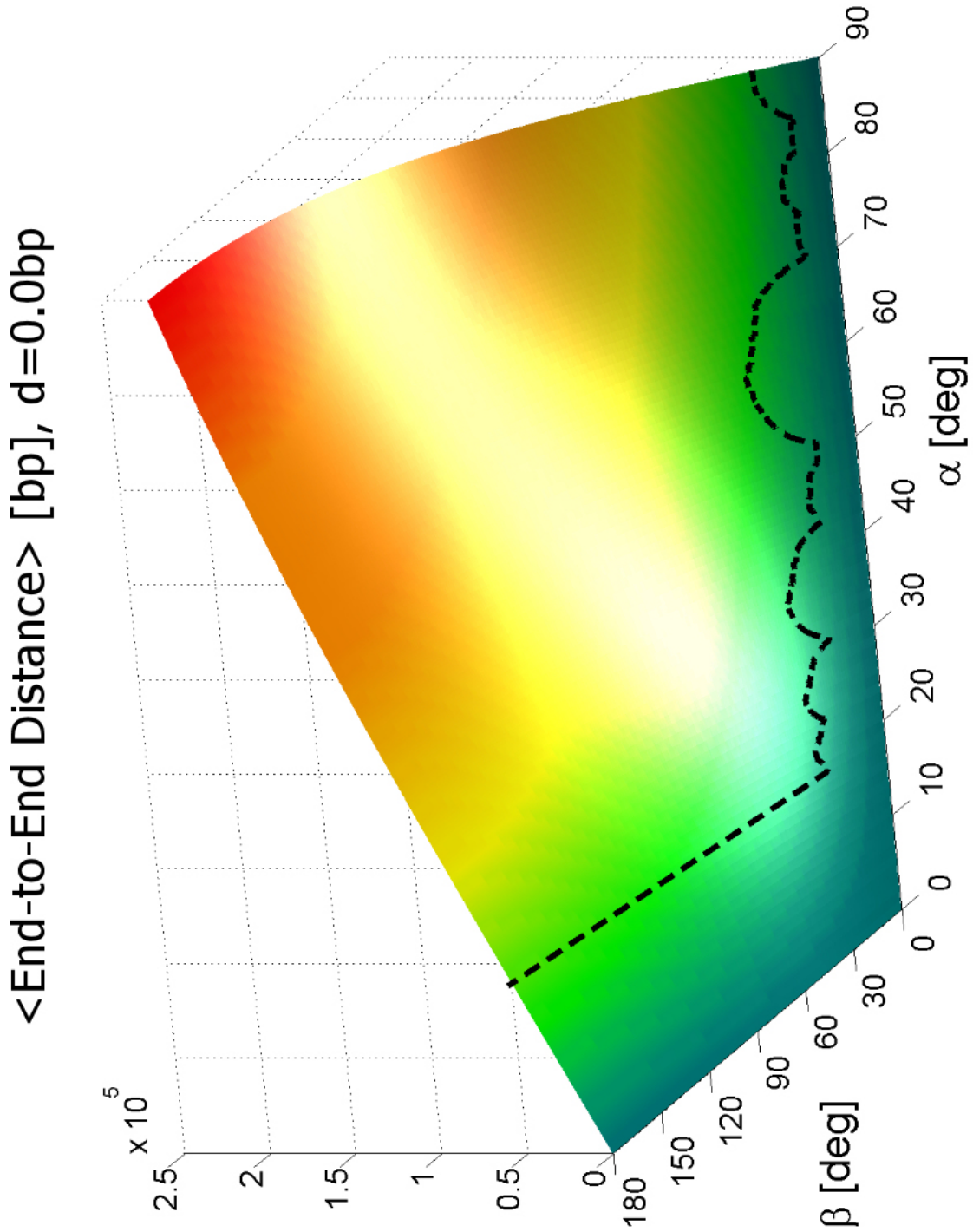


Figure 59: The end-to-end distance R of regular chromatin fibers along a cut-out of the phase diagram with fixed fiber length (N=500 segments) and fixed d=0.0⟨bp⟩. The solid black line represents the corresponding phase transition.

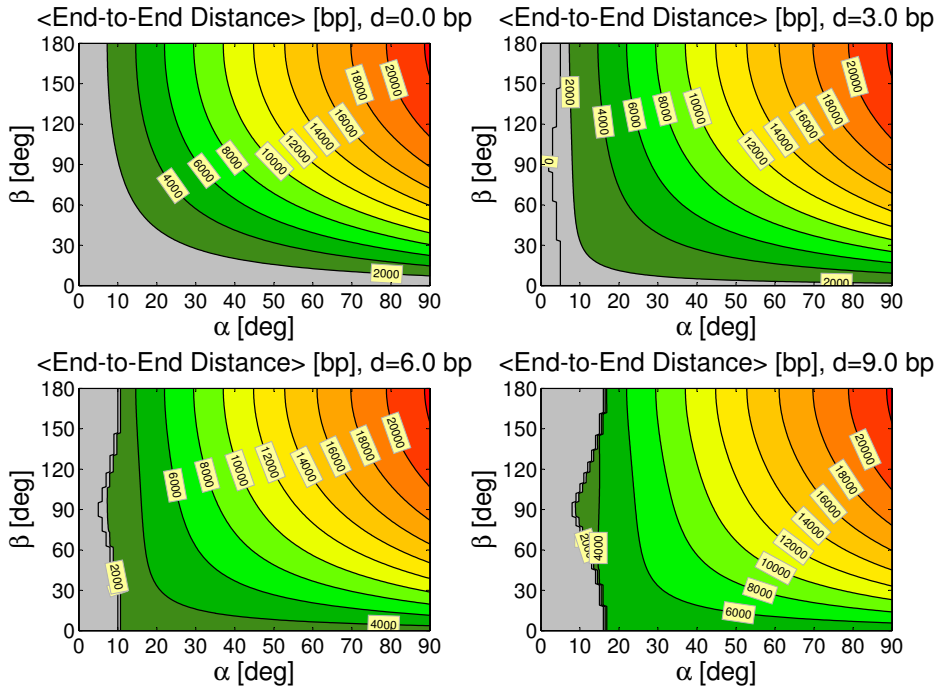


Figure 60: The mean end-to-end distance R is another measure for the spatial extension of the chromatin fibers. It increases with increasing pitch d. The arrow-like shape of the equipotential lines at the left side of the diagram comes again from the cylindrically shaped nucleosomes.

5.6 Distribution of Non-Regular Chromatin Structures

In this chapter (Sec. 5) only perfectly *regular* chromatin fibers have been investigated. These regular fibers are very artificial and unfavorably from the entropic point of view. Nevertheless, they allow a first investigation of the different chromatin structures with respect to fiber features like densities, accessability, excluded volume interactions or topology.

In the next section (Sec. 6) the local chromatin fiber parameters α_i, β_i, d_i and b_i will no longer be fixed. Instead, parameter distributions will be derived for them from experimental data [228]. Thus one gets equilibrated chromatin fibers. These fibers do not have a single fixed topology any more. Instead, they show a distribution of different chromatin structures. If one generates such an equilibrated chromatin fiber with fluctuations, cuts it into small pieces and then marks the structure of every piece in the phase diagram, one will get a structure distribution as illustrated in Fig. 61.

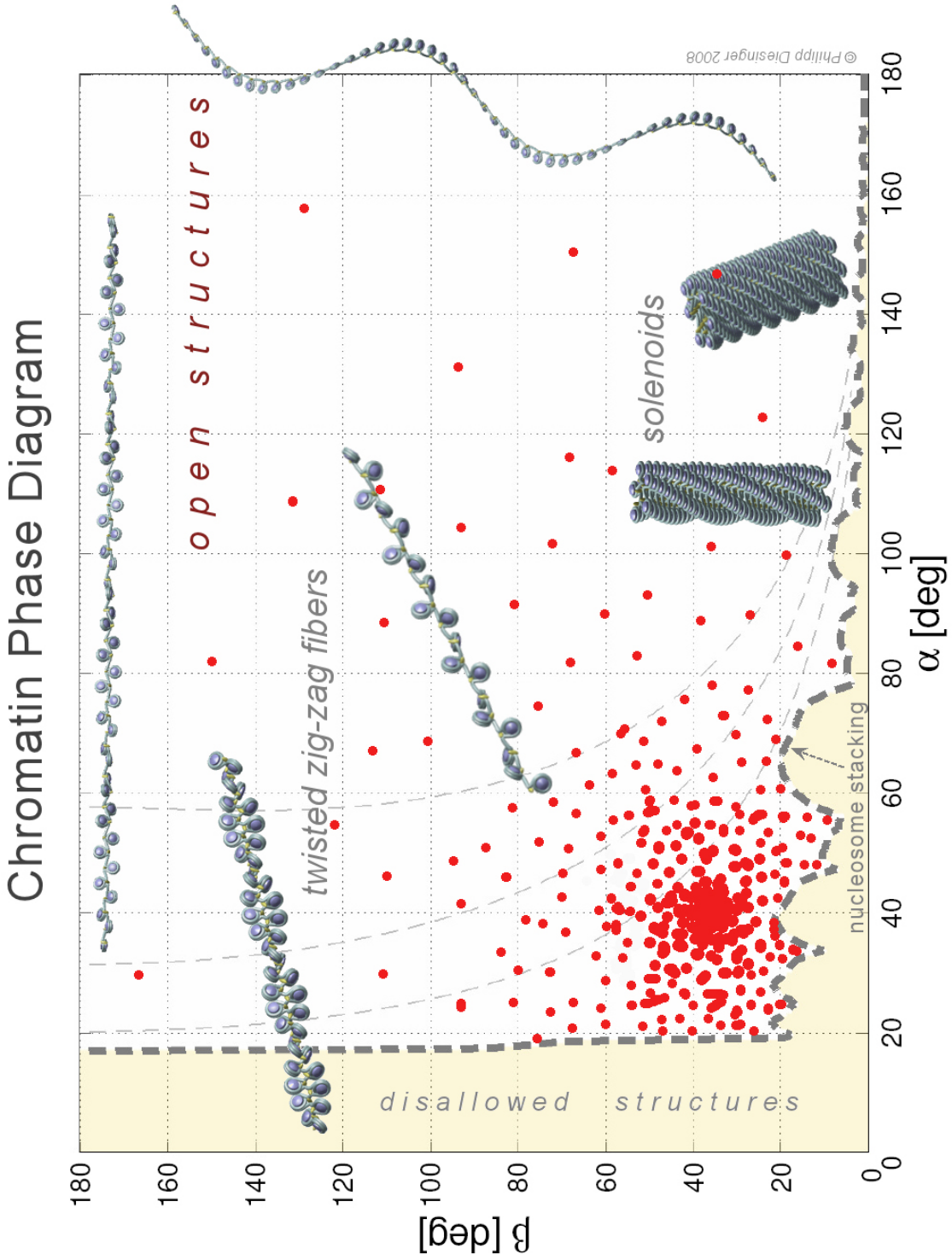


Figure 61: In Sec. 5 only *regular* chromatin fibers were considered. If one generates equilibrated fibers (cf. Sec. 6) that have local parameter fluctuations, one will get a *distribution* of different chromatin structures. This is illustrated by the red dots in this diagram.

6 Modelling Chromatin II - Flexibility and Depletion Effects

6.1 Overview

So far, only *regular* chromatin fibers have been considered (cf. Sec. 4 and Sec. 5), although it was already explained (in Sec. 4.5) how *flexible* fibers in principle could be generated with the knowledge of the right parameter distributions.
In this chapter the chromatin model will be developed further: The local parameter distributions will be obtained from experimental data and put in the chromatin model to generate flexible fibers i.e. chromatin fibers with local fluctuations (cf. Sec. 6.3).
Furthermore, another important feature of the chromatin fiber, namely *histone depletion*, will be taken into account.

There are two kinds of depletion effects that have a strong impact on the chromatin fiber: Linker histone skips (cf. Sec. 6.2) and nucleosomes skips (cf. Sec. 6.4). The linker histones are in a dynamic equilibrium with the surrounding fluid. Therefore, the model should be able to simulate nucleosomes with missing linker histones (i.e. histone H1/H5).

Nucleosome skips are "defects" in the chromatin fiber where a whole nucleosome and not only a linker histone is missing. Therefore, they leave just a long blank stretch of DNA. Nucleosome skips will be implemented into the E2A model as well.
Finally, some first results for the flexible chromatin fibers with and without depletion effects are presented (cf. Sec. 6.5) and some of the new E2A model conformations are compared with electron micrographs (cf. Sec. 6.5.3).

The main parts of this chapter have already been published [64; 65].

6.2 Linker Histone Depletion

The histone H1 (and the variant histone H5 with similar structure and functions) is involved in the packing of the beads on a string structure into the 30nm chromatin structure (cf. Sec. 2). To do so it sits in front of the nucleosome keeping in place the DNA which is wrapped around the histone octamer (cf. Fig. 38) and thus stabilizes the chromatin fiber. Therefore, the degree of chromatin compaction depends on the presence of linker histones [19] (as well as on the salt concentration [211]) and H1 depletion can cause dramatic alterations in the chromatin structure [77].

Linker histones have the ability to dissolve and to be taken up by chromatin fibers again. This is probably happening all the time somewhere in the genome. Furthermore, other processes in the cell nucleus like transcription can remove linker histones as well. Therefore, a dynamic equilibrium between free and bound linker histones exists (although this equilibrium could be near saturation). This means there could be a linker histone skip rate which is constant over a large period of time but locally the linker histones fluctuate to some extent.

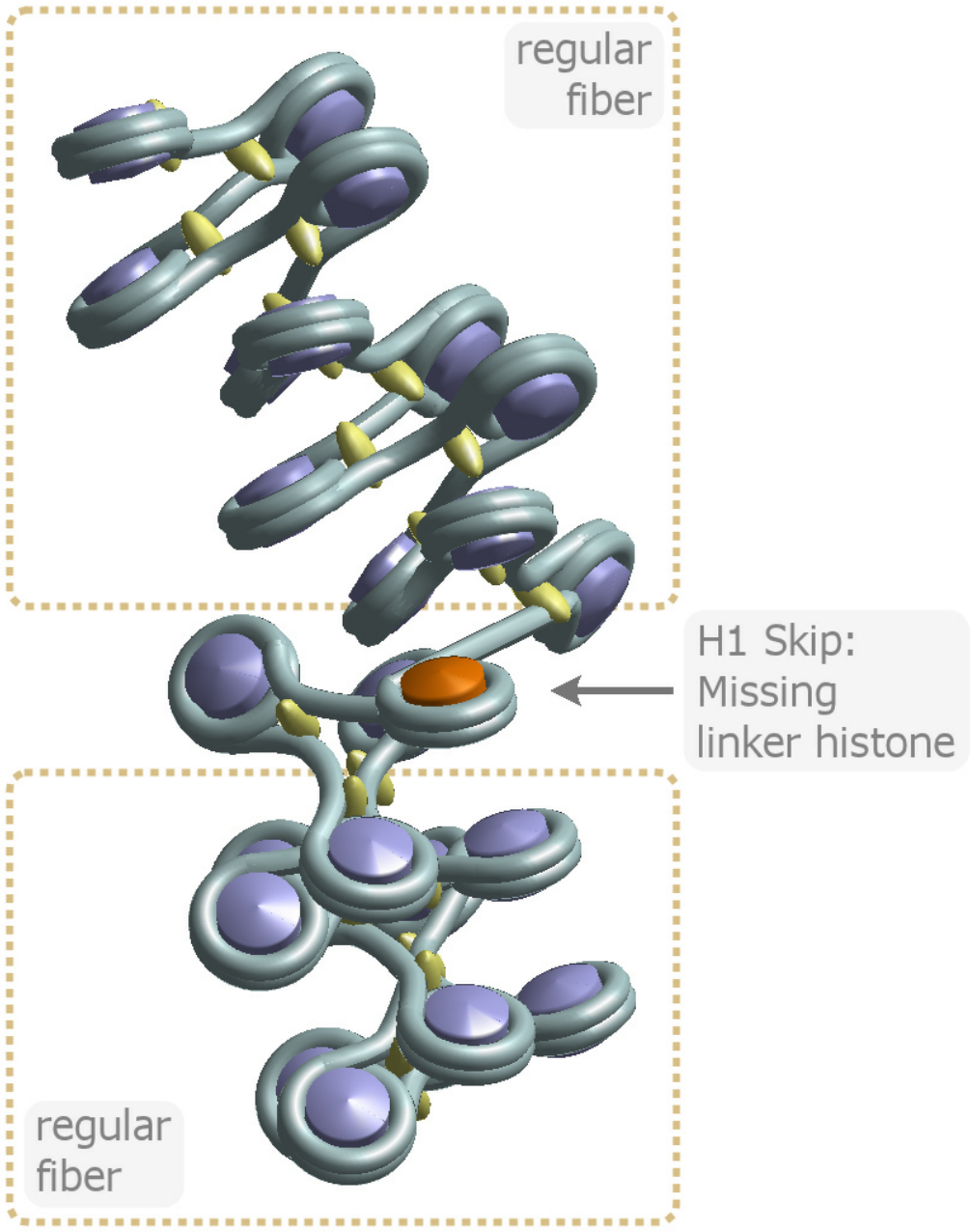


Figure 62: An example conformation of a short chromatin fiber with a missing linker histone. The upper strand and the strand below the defect are regular, i.e. the local fiber parameters are fixed. Please note that the fiber is very open to make the visualization clear. The fiber parameters of the conformation in this figure do not correspond to the values mentioned in Sec. 6.2.1.

6.2.1 Simulation of Linker Histone Skips

In this section it will be described how the linker histone skips are included in the Monte Carlo model.

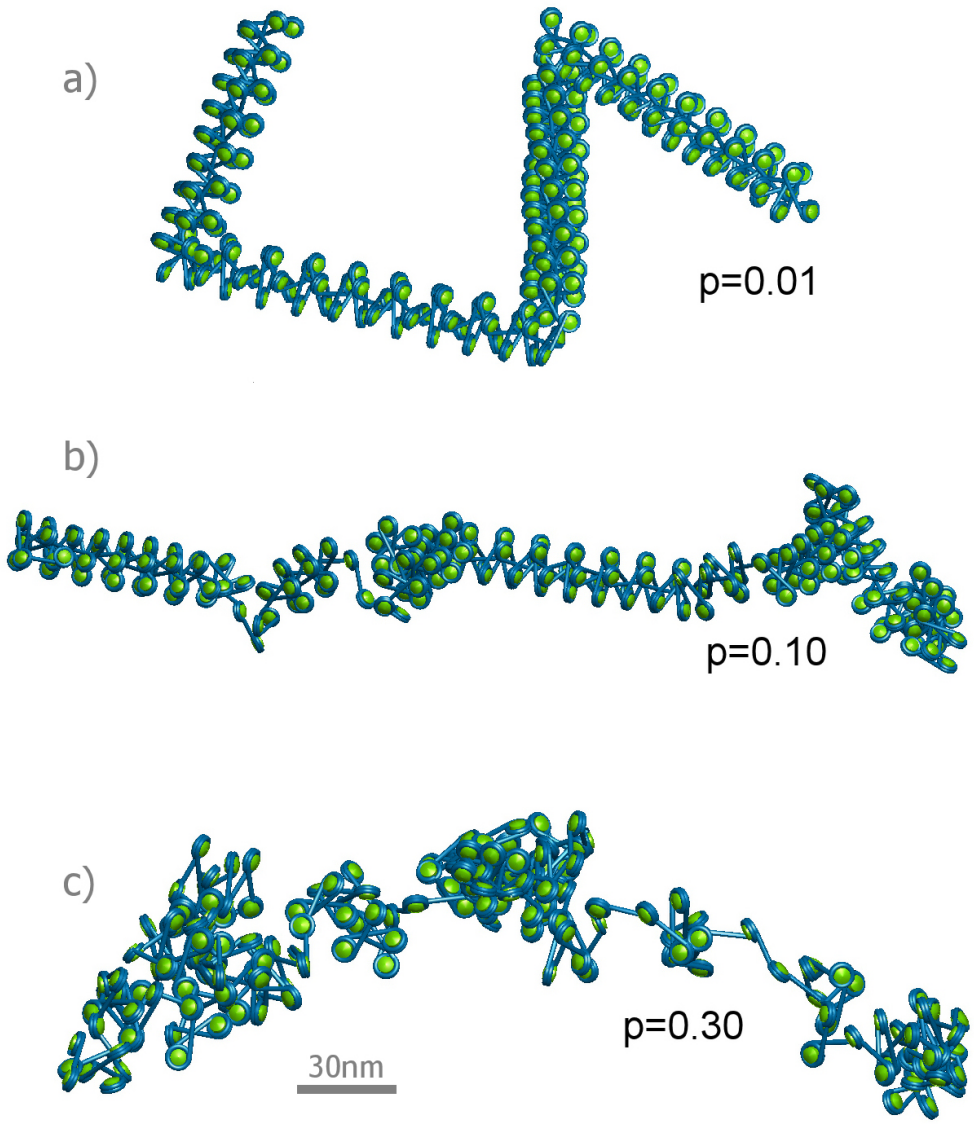


Figure 63: Chromatin fibers with different linker histone skip probabilities p. At $p = 0.30$ the regular structure of the 30nm strand is almost completely collapsed.

Again the two-angle-model was used with some fixed parameters (cf. Tab. 1) which reflect the probable mean values within the cell. For the simulations the entry-exit-angles α_i were fixed to 40 degrees, the rotational angles β_i to 36 degrees and $h_i = 7\langle \text{bp} \rangle$. The only interaction potential is the hard core excluded volume. In this mean field-like

approach [48] thermal fluctuations are neglected to concentrate on the interaction between H1 defects and volume exclusion.

For a certain nucleosome position N_i the *skip probability* (or sometimes "defect probability") p gives the chance of a missing linker histone. If the histone is missing, the in- and outgoing DNA strand will no longer be fixed in front of the nucleosome but instead are arbitrary with respect to the excluded volume interactions of the chromatin strand. In that case the remaining two DNA linkers are pivoted into their new position. A example conformation of a short chromatin fiber with one linker histone skip in its middle is shown in Fig. 62.

Three further chromatin fiber conformations with a fixed linker histone skip rate can be found in Fig. 63. As one can see the formerly stiff fibers collapse already when the skip rate reaches 30%.

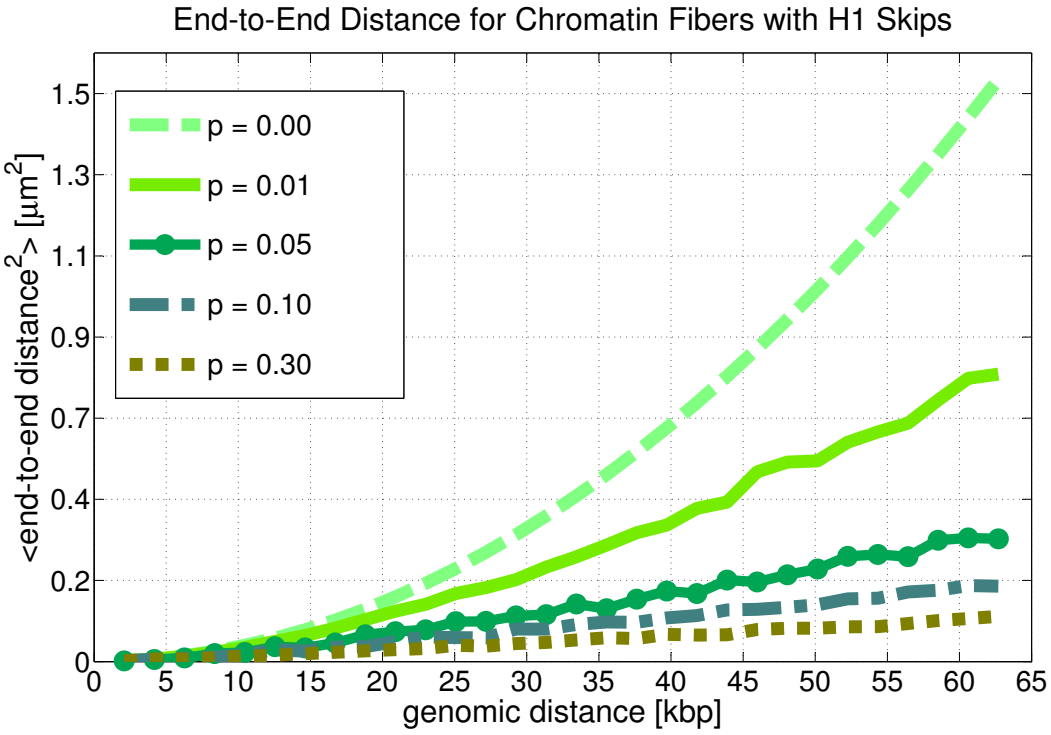


Figure 64: The mean end-to-end distance for chromatin fibers with H1 defects. With increasing defect probability p the length of the fibers decreases rapidly. H1 defects might play a crucial role for chromatin compaction.

6.2.2 Chromatin Fibers with Linker Histone Skips

As described above one can generate chromatin strands with linker histone skips. Thus one can examine fiber properties (cf. Sec. 1) like the the mean end-to-end distance (cf.

Fig. 64) and the mean radius of gyration (cf. Fig. 65) for fibers with various defect probabilities: $p = 0.00$, $p = 0.01$, $p = 0.05$, $p = 0.10$ and $p = 0.30$. One can clearly see that even very small defect rates of some percent have a huge effect on the compaction of chromatin: Both, the mean squared radius of gyration and the mean end-to-end distance will decrease rapidly, if one allows only a few H1 skips.

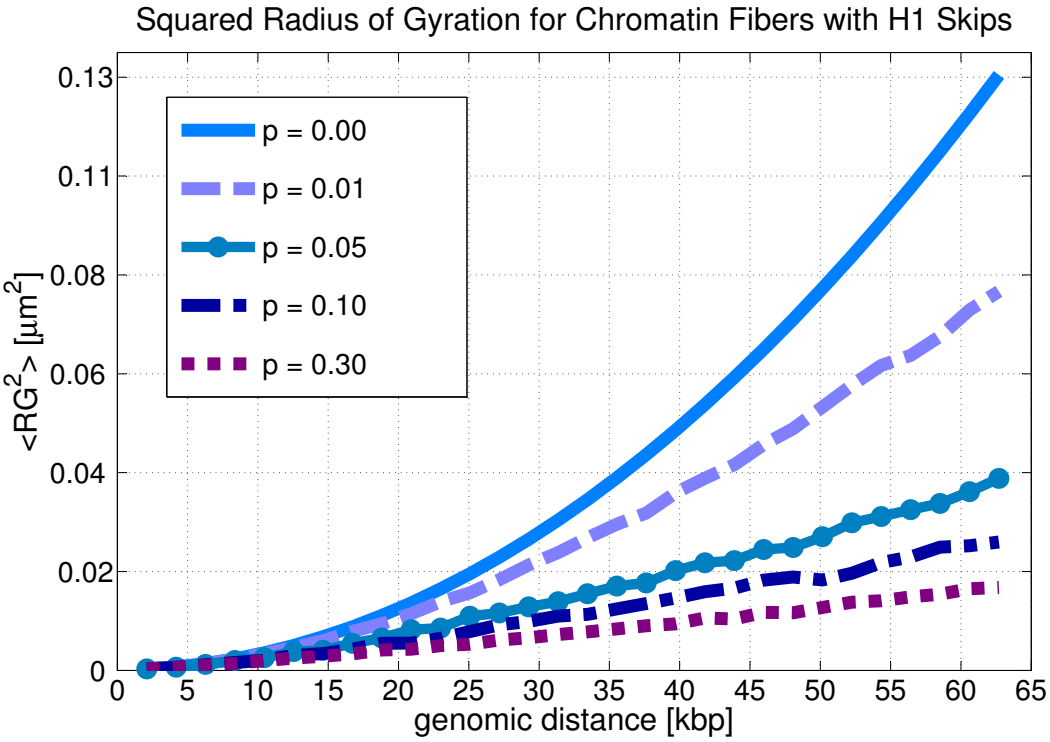


Figure 65: Squared radius of gyration for chromatin fibers with different H1 skip probabilities p. With increasing number of H1 defects the fiber becomes much more compact which could be an important mechanism to compact the chromatin fiber. All skip probability distributions which have been used for these fibers were uniform.

Without H1 defects ($p = 0$) one gets a *regular* (i.e. stiff) chromatin fiber within the restriction of the extended two-angle model. This ideal fiber reflects the properties of the 30nm strand only on small length scales. Therefore, the increase of the compaction due to defects will probably be not as strong as implicated by these results. Nevertheless, missing H1 histones might contribute to chromatin compaction and DNA accessibility for transcription purposes at the same time since one can see from Fig. 63 that although the fiber gets more compact, some very open parts appear. One can also see here the increasing disorder with increasing skip probability p. Fig. 63 shows visual similarities with EC-M images (for instance Figures 1 and 2 in [19]) which show typical chromatin confirmations. This might among others come from H1 defects in the fiber.

So far, only uniform linker histone skip distributions along the chromatin fiber have

been considered. Fig. 66 shows the behavior of the squared radius of gyration in dependence on the fiber length for different defect probability distributions. The probability density functions range from a Gaussian peak in the middle of the fiber to a uniform pdf along the whole chromatin strand (cf. inset Fig. 66 or for a larger version Fig. 180 in App. A). Nevertheless, the effect on the squared radius of gyration is small.

Missing H1 histones might supply a further contribution (cf. Sec. 5) to the compaction of the fiber. These H1 defects might play a crucial role in the task of chromatin compaction and at the same time serve the transcription of the DNA by opening locally the chromatin fiber. Taking fluctuations of the basic model parameters into account the end-to-end distance and the radius of gyration will not decrease that strongly any more with increasing defect probability (cf. Sec. 6.3). But nevertheless, it is plausible that H1 defects play a role in chromatin compaction.

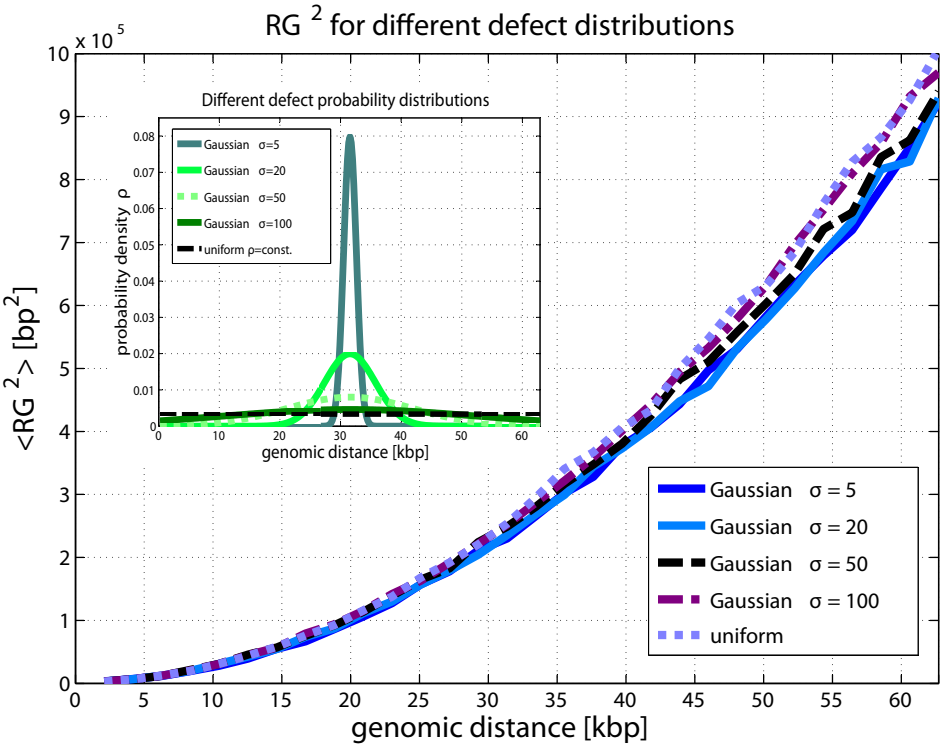


Figure 66: Squared radius of gyration for different linker histone skip probability distributions ρ and different fiber lengths. The pdfs range from a Gaussian peak in the middle of the chromatin fiber to a uniform defect probability distribution along the whole fiber (cf. inset). The corresponding squared radius of gyration is rather independent of the defect distribution. A larger version of the inset can be found in App. A Fig. 180.

6.3 Parameter Distributions

The chromatin fibers in the first chapters of this work (cf. Sec. 4 and Sec. 5) were *regular*, i.e. there were no distributions for the local fiber parameters. On a small length scale these stiff chromatin fibers represent a mean-field approach to the actual chromatin fiber.

The nucleosome-nucleosome as well as the nucleosome-DNA interactions are highly complex and still an area of current research. The problem of avoiding these potentials was solved by using experimental data [228] for the distribution of the nucleosome repeat length (cf. Sec. 6.3.1) and taking advantage of the fact that the local chromatin parameters are not independent. In that case it is possible to partially invert the convolution of the probability distributions (which is given by the experimental data) and thus get information about the individual distributions of our model parameters (cf. Sec. 6.3.2).

Making use of given parameter distributions for the model parameters gives us the advantage of saving computation time that would otherwise be spent for the equilibration of the fibers. The saved computation time can then be used to generate very large fibers up to 10^4 NRLs (i.e. chromatin fibers of the order of magnitude of Mbp). This is the tenfold of comparable chromatin models [201].

Of course, excluded volume potentials for the DNA and the nucleosomes are still taken into account: The DNA has a tube-like shape and the nucleosomes have the excluded volume of flat cylinders.

6.3.1 Nucleosome Repeat Length Distribution

In [228] a large number of measurements of nucleosome repeat lengths were analyzed and found to exhibit preferential quantization to a set of values related by integral multiples of the helical twist of DNA. This implies that the nucleosomal DNA content is preferentially quantized which in turn implies that linker DNA lengths are preferentially quantized.

This study confirms and extends previous observations in the literature [111; 135; 136; 206] that had suggested but not firmly established that linker lengths might be *quantized*. The quantization of repeat lengths applies even for very long repeat lengths.

Nucleosome repeat lengths have been measured in many laboratories, for a large number of cell types, tissue types, and organisms. Van Holde [216] has abstracted 185 such measurements from the literature and presented them in tabular form. The measured (average) nucleosome repeat length obtained in each study is reported and error bars are reported when available.

The analysis in [228] was restricted to only these data. They were considered collectively and also in statistically defined subgroups and it was asked whether the reported nucleosome repeat lengths for all of these cell types, tissue types and organisms are related to one another by some fundamental periodicity.

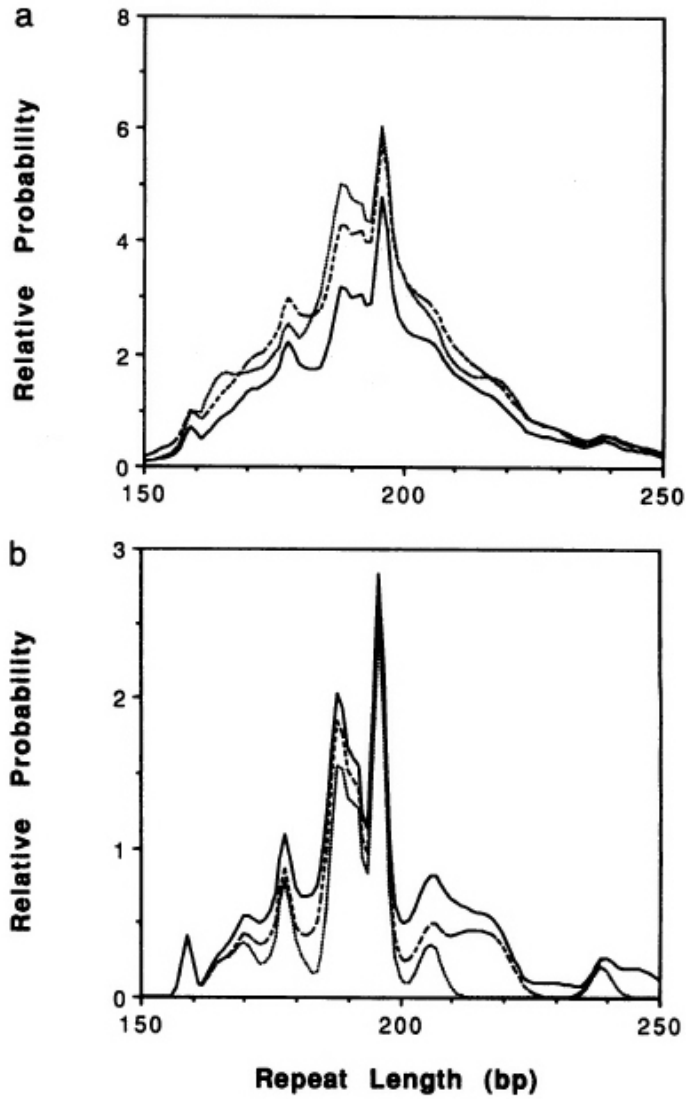


Figure 67: Nucleosome-repeat-length probability distribution functions, $P(L)$, for distinct sets or subsets of the data. (a) $P(L)$ functions obtained using most or nil of the data. Solid lines, data for which error bars are reported (n =134); dotted lines, data (n =185) when error bars are not reported, σ is set equal to 3 bp; dashed lines, data when error bars are not reported, σ set equal to 10 bp. (b) $P(L)$ functions obtained using restricted subsets of the data. Solid lines, data having error bars $\leq$4 bp (n =55); dashed lines, data having error bars $\leq$3 bp (n =43); dotted lines, data having error bars $\leq$2 bp (n =29). Only the range L of 150-250 bp is shown. The probabilities are relative: their integrals (if extended to the range $-\infty \leq L \leq \infty$) are equal to the number of measurements on which they are based. Adapted from [228].

Each measurement was represented as a Gaussian function of probability versus nucleosome repeat length L, having unit integrated probability with a mean equal to the

reported repeat length and a standard deviation σ equal to the reported error bar for that measurement.
When no error bars are given that repeat length measurement was (in separate calculations) either skipped or assigned a σ of 10 bp or 3 bp. The measurements were then combined by summation of their Gaussians to give a probability distribution function for the set of nucleosome repeat length measurements. This is effectively an averaging procedure which suppresses possible errors in individual measurements. Plots of the resulting functions are shown in Fig. 67.

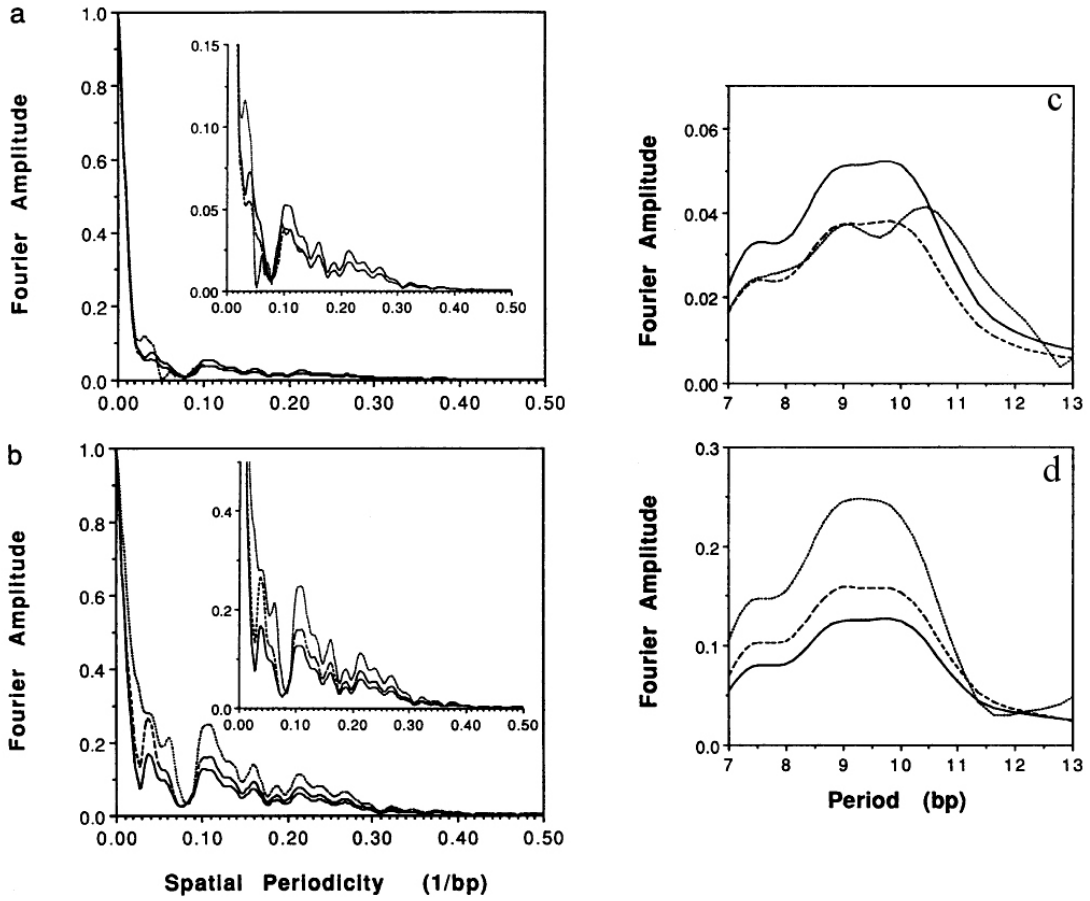


Figure 68: Plots of Fourier transforms of the 6 nucleosome repeat length probability distribution functions $P(L)$. The curve designations in a and b are identical to those in Fig. 67. Only the amplitudes of the Fourier transforms are shown; they have been normalized by dividing by the integral (i.e. sum) of the corresponding $P(L)$. The Insets show the same results plotted on expanded scales.
c and d show the Fourier transforms plotted against a real-space abscissa. The region of the six transforms centered on the peak at 0.10-0.11 bp^{-1} was plotted against the reciprocal of the abscissa. Thus, this figure plots the amplitude of the transforms against spatial periods measured in bp for periods from 7 to 13 bp. The curve designations in c and d are identical to those in a and b and in Fig. 67. Adapted from [228].

The probability distribution functions are broadly centered around 190-200 bp, and they exhibit multiple peaks or shoulders. In Fig. 67a, the curve calculated using all 134 measurements for which error bars were reported shows peaks or shoulders at 159, ≈171, 178, 188, 192, 196, and ≈206 bp repeat lengths.
In Fig. 67b, the curve calculated using all 55 measurements having error bars of ≤4 bp shows peaks or shoulders at 159, 170-171, 178, 188, 196, 206-207, and ≈217-218 bp.
Thus, the probability distribution functions show an imperfect but pronounced quantization with a preference for a set of repeat lengths differing from each other by multiples of ≈10 bp.

To analyze quantitatively the spatial periodicities presented in these plots, Fourier transforms of the distribution functions were calculated. The results for the six calculations are shown in Fig. 68 (in Sec. App. A).
The transforms are dominated by a large peak centered at zero periodicity, arising from the envelope of the repeat length distribution. There are numerous subsidiary maxima; the most pronounced (measured as the height, or integrated area, above the envelope of the transform) is a broad peak occurring at a periodicity of 0.10-0.11 bp^{-1} that is present in all six transforms. The reciprocal of this range of periodicities is ≈9-10 bp. It is this maximum in the Fourier transforms that is responsible for the set of roughly evenly spaced peaks or shoulders detected in the probability distribution functions.

6.3.2 Local Parameter Distributions

In this section it will be shown how the experimental data for the nucleosome repeat length (NRL) distributions [228] can be used to get information about the parameter distributions of the E2A model.
The NRL consists of a stretch of linker DNA b and the DNA content which is wrapped around the nucleosomes ω: NRL $= b + \omega$. Only the 197 bp repeat length peak of the data in [216; 228] will be used here.
In the model b_i is the linker length between two consecutive nucleosomes ($i-1$ and i) and ω_i is the length of the DNA strand which covers nucleosome i.
The distribution for the nucleosome repeat length in [228] will be approximated by a Gaussian standard normal distribution:

$$f_{\text{NRL}}(l) = \mathcal{N}[\mu_{\text{NRL}}, \sigma^2_{\text{NRL}}](l) = \frac{1}{\sigma_{\text{NRL}}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{l-\mu_{\text{NRL}}}{\sigma_{\text{NRL}}}\right)^2}. \tag{42}$$

Furthermore, one knows NRL $= b + \omega$ where NRL, b and ω are random variables. Therefore, the convolution of the probability density functions of b and ω must be the probability density of NRL, i.e.:

$$f_{\text{NRL}}(l) = \int_{-\infty}^{\infty} f_b(l') f_\omega(l - l') dl'.$$

By assuming Gaussian standard normal distributions and statistical independence

$$f_b = \mathcal{N}[b_0, \sigma_b^2]$$
$$f_\omega = \mathcal{N}[\mu_\omega, \sigma_\omega^2]$$

this gives

$$\mu_{\mathrm{NRL}} = b_0 + \mu_\omega \text{ and } \sigma_{\mathrm{NRL}}^2 = \sigma_b^2 + \sigma_\omega^2.$$

In this context σ_ω^2 represents a part of the mobility of the nucleosomes inside the DNA loops. σ_b^2 and σ_ω^2 are chosen proportionally to the DNA content of their corresponding nucleosome repeat unit parts (cf. Tab. 5).

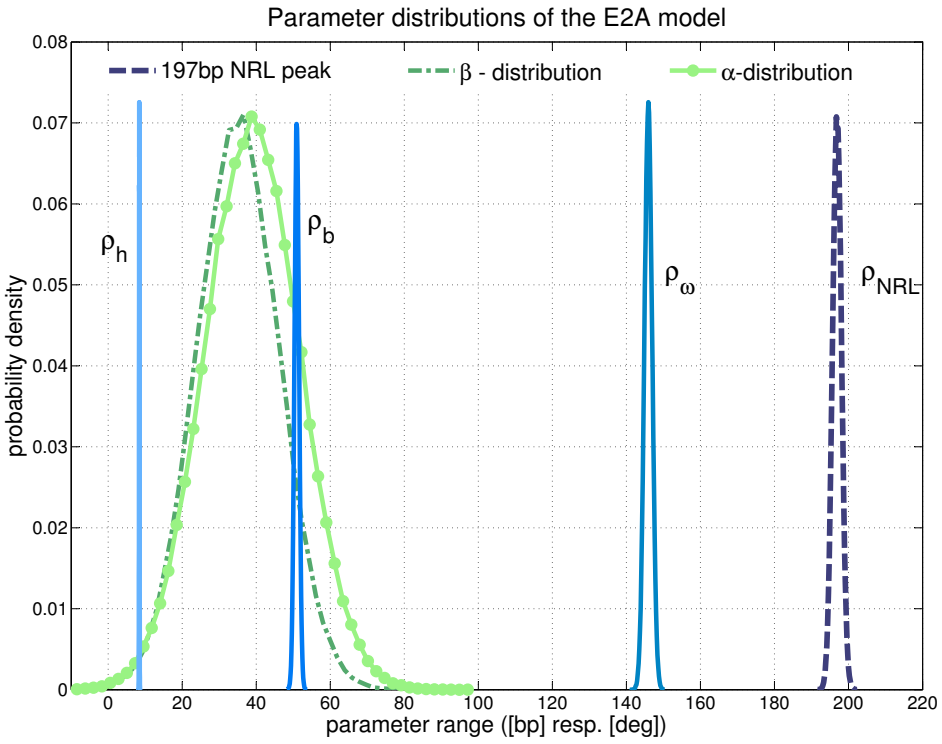


Figure 69: This figure shows the underlying parameter distributions of the E2A model. They come from the NRL distribution of [228] and are used instead of interaction potentials since these are still unknown.

The rotational angle β_i is a periodic function of the linker length b_i because DNA adsorption always begins with the minor groove turned in towards the first histone binding site:

$$\beta(b) = \beta_0 + \frac{2\pi}{10.2bp}(b - b_0).$$

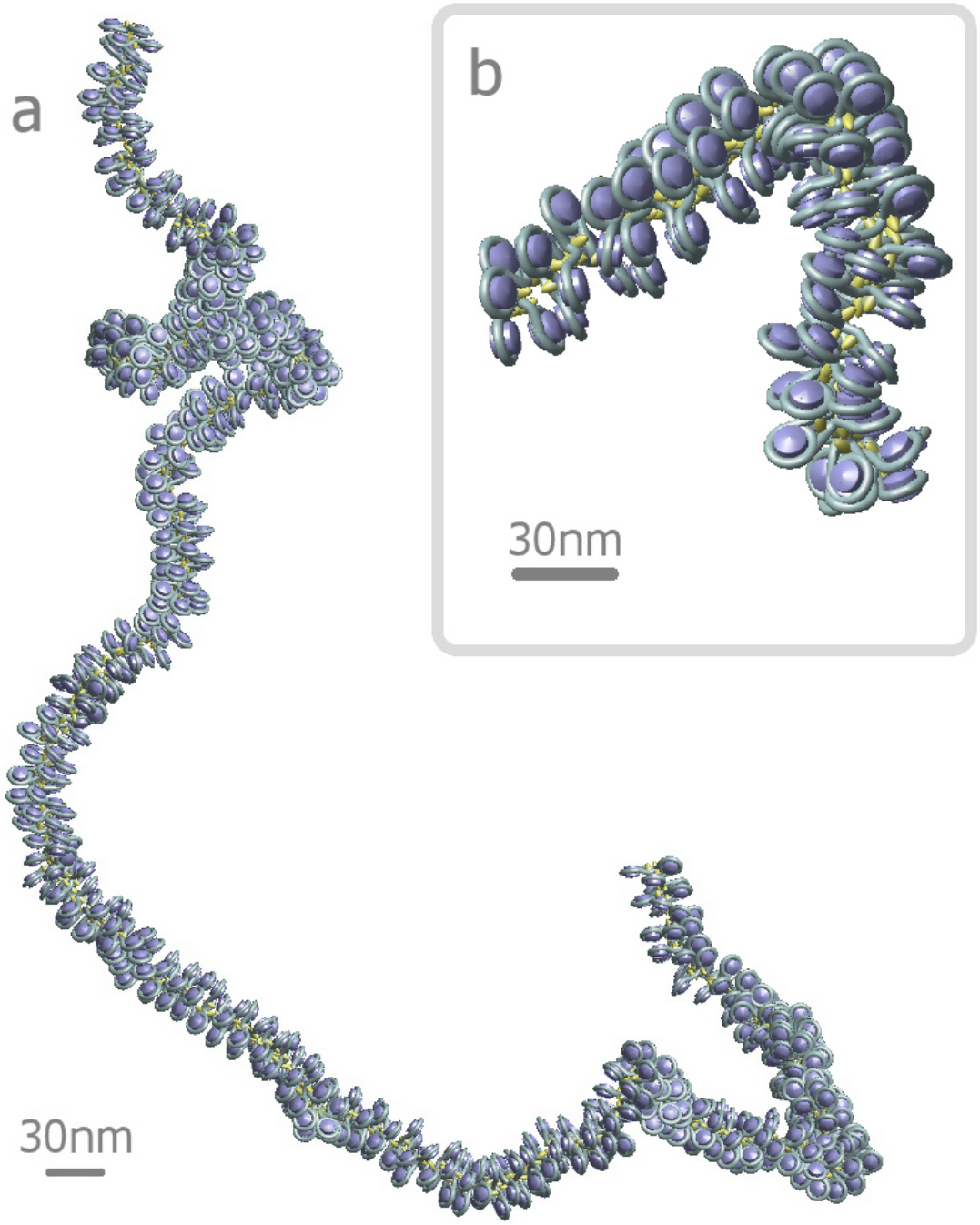


Figure 70: This figure shows example conformations of flexible chromatin strands with the parameters that have been determined in Sec. 6.3.2 (cf. Tab. 5). a) This fiber has a length of 800 nucleosome repeat lengths, i.e. ≈158kbp. The small chromatin fiber in the inset (b) has a length of only 100 NRL, i.e. ≈20kbp. The flexible chromatin fibers have an average diameter of 34.6nm and a persistence length (l_p) of ≈280nm.

The period is the repeat length of the helical twist of the DNA. It is fixed to 10.2 bp in this work. β_0 and b_0 are the mean values of β and b. They can be found in Tab. 5. With the equation above one can calculate the probability density function of β by transforming the density of b:

$$\begin{aligned} f_\beta(\beta') &= f_b(b') \, \frac{db(\beta')}{d\beta} \\ &= \mathcal{N}[b_0, \sigma_b^2](b') \, \left(\frac{10.2\text{bp}}{2\pi}\right) \\ &= \frac{10.2\text{bp}}{2\pi} \, \frac{1}{\sigma_b\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{b_0 + 10.2\text{bp}\left(\frac{\beta'-\beta_0}{2\pi}\right) - b_0}{\sigma_b}\right)^2\right] \\ &= \mathcal{N}\left[\beta_0, \left(\frac{2\pi}{10.2\text{bp}}\sigma_b\right)^2\right](\beta'). \end{aligned}$$

In the E2A model β_i depends directly on b_i. Its distribution was calculated here for the sake of completeness.
Similarly one can calculate the density function of h by the relation

$$h = \frac{1}{2}\left(\frac{8.4 \text{ bp}}{146 \text{ bp}}\omega_{i-1} + \frac{8.4 \text{ bp}}{146 \text{ bp}}\omega_i\right) = \frac{8.4}{292}(\omega_{i-1} + \omega_i).$$

ω_{i-1} and ω_i are statistically independent random variables with the same probability density f_ω. Therefore, one gets the probability density of h by

$$\begin{aligned} f_h(h') &= f_{\omega_{i-1}+\omega_i}(\omega') \, \frac{292}{8.4} \\ &= \mathcal{N}[2\mu_\omega, (2\sigma_\omega)^2](\frac{292}{8.4}h') \, \frac{292}{8.4} \\ &= \mathcal{N}\left[\frac{8.4}{292}2\mu_\omega, \left(\frac{8.4}{292}2\sigma_\omega\right)^2\right](h'). \end{aligned}$$

μ_{NRL}	197 bp	σ^2_{NRL}	1.5 bp^2
b_0	146 bp	σ_b^2	0.4 bp^2
β_0	36 deg	σ^2_{beta}	0.1518
μ_ω	146 bp	σ_ω^2	1.1 bp^2
μ_h	8.4 bp	σ_h^2	0.0036
α_0	40 deg	σ_α^2	0.05

Table 5: Local Parameter Distribution of the E2A Model

The results for the parameter distributions in the chromatin model are shown in Fig. 69 and in Tab. 5. Two example conformations with these parameter distributions are illustrated in Fig. 70 and Fig. 71. Another example conformation is shown in Fig. 181

(App. A).

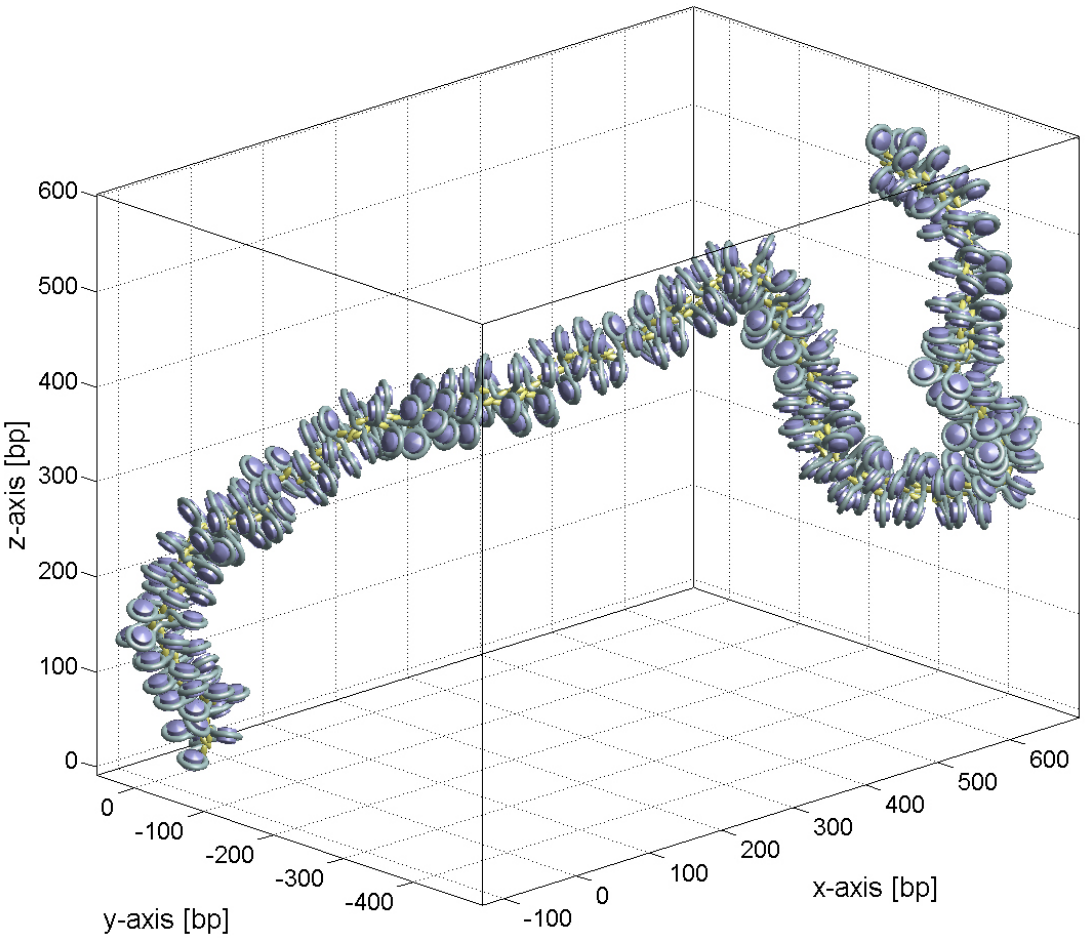


Figure 71: Another example conformation of a flexile chromatin fiber of length 300 NRL (≈60kbp).

6.3.3 Scaling of R^2 and R_g^2 for Flexible Fibers

The scaling of R^2 and R_G^2 for random walks was calculated in Sec. 1 and can be found in [207]. In the case of self-avoiding walks one gets $R^2 \propto N^{2\nu}$ with $\nu \approx 0.589$ [24; 207]. With the parameter distributions of Tab. 5 the chromatin fibers become flexible like random walks but the short range interactions due to the excluded volume of the DNA and the nucleosomes lead to long range correlations and that prevents them from behaving like random walks. Instead, the end-to-end distance and the radius of gyration comply the scaling laws of a self-avoiding walk (on large length scales). On smaller length scales (up to ≈5000 NRLs) the end-to-end distance and the radius of gyration show strong finite size effects. This is illustrated in Fig. 72.
Fig. 182 and Fig. 183 in App. A show the end-to-end distance and the radius of gyration,

too, but here the leading order $N^{2\nu}$ is divided out. Therefore, the agreement with the scaling of the self-avoiding walk is easier to see in these figures.

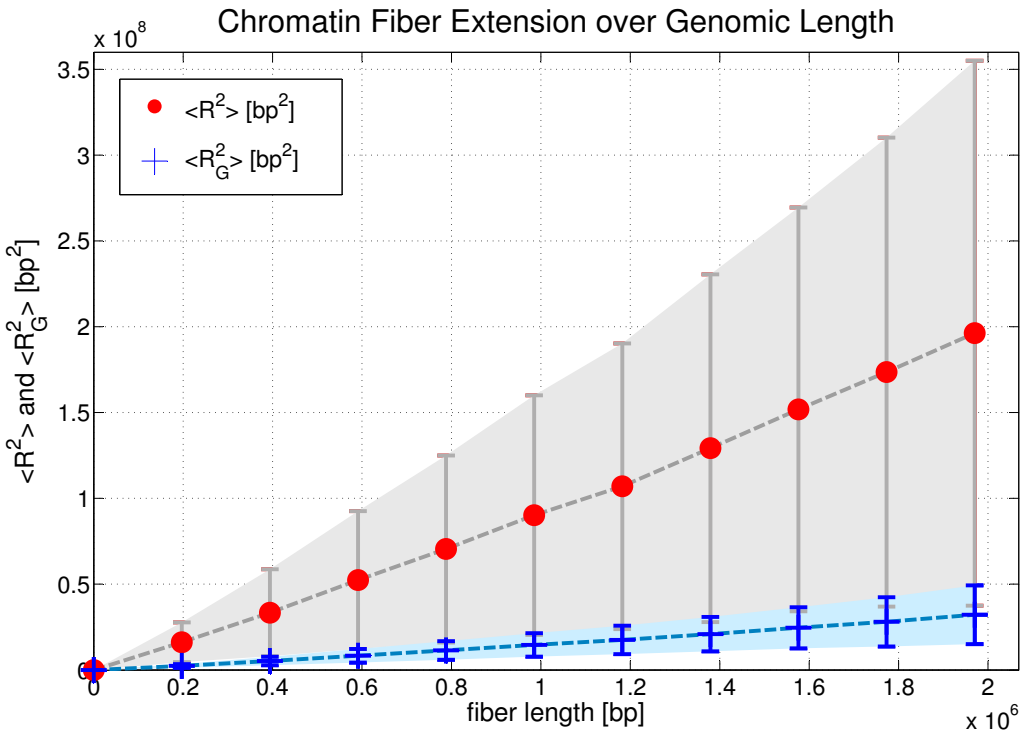


Figure 72: Since the chromatin fibers are flexible and have excluded volume they behave like self-avoiding walks on large length scales. The mean squared end-to-end distance and the mean squared radius of gyration scale with $N^{2\nu}$ in this case ($\nu \approx 0.589$). The error bars give the standard deviation of R^2 and R_G^2.

6.4 Nucleosome Depletion

Another new feature of the E2A model that will be explained in this chapter is the possibility for nucleosomes to dissolve entirely so that only blank stretches of DNA remain.

The DNA content of a nucleosome is 147bp. Its DNA is sharply bent and tightly wrapped around the histone protein octamer [173]. This bending occurs at every DNA helical repeat (i.e. ≈10 bp) when the major DNA groove faces towards the histone octamer and again ≈5 bp away with opposite direction when the major groove faces outwards. Specific dinucleotides facilitate the bends of each direction [178; 229]. The linker DNA that connects consecutive nucleosomes has a length of approximately 50bp. Thus, approximately 75% of genomic DNA will be wrapped in nucleosomes if the chromatin fiber is entirely saturated with them.
Access to DNA wrapped in a nucleosome is occluded [173] for polymerase, regulatory, repair and recombination complexes, yet nucleosomes also recruit other proteins

through interactions with their histone tail domains [110]. Thus, the detailed locations of nucleosomes along the DNA may have important inhibitory or facilitatory roles in regulating gene expression [120; 237].

Since DNA sequences differ in their ability to bend sharply [178; 213; 229] the ability of the histone octamer to wrap different DNA sequences into nucleosomes is highly dependent on the DNA sequence [3; 188]. In-vitro studies show this range of affinities to be 1000-fold or greater [209]. Thus, nucleosomes have substantial DNA sequence preferences which results in a non-regular arrangement of the nucleosomes along the DNA. Furthermore, nucleosomes can dissolve entirely by unwrapping the DNA, leaving blank DNA stretches behind and later on they can reform again. This means that nucleosomes are in a dynamic equilibrium with the chromatin fiber. These effects lead to an average nucleosome occupation (i.e. the probability that a bp is covered by a nucleosome) of *less* than 75%. Therefore, it is necessary to allow for nucleosome depletion in a particular chromatin model.

6.4.1 Modelling of Nucleosome Skips

When nucleosomes dissolve entirely they leave blank DNA strands behind. These DNA stretches are modeled as worm-like chains (cf. Sec. 1.3) with a diameter of 2.2nm and a persistence length of 50nm as illustrated in Fig. 73 and Fig. 184 (App. A).

The nucleosome detachment itself is not modelled by the E2A model. That would require a more detailed approach from a smaller length scale (e.g. based on model structures at atomic resolution). And even then it is very hard to do [116]. Furthermore, we are not interested in the dynamics of the chromatin strand but want to sample conformations directly.

The blank DNA stretches have a tube-like excluded volume potential with the DNA diameter of 2.2nm.

In this work linker histone depletion and nucleosome depletion are treated statistically independent to investigate the two different kinds of chromatin "defects" independently from each other. In fact, linker histone depletion *facilitates* nucleosome depletion because if the glue particle that keeps the in- and outgoing DNA strand together is missing, it will be much easier for the DNA to unroll from the histone octamer. The linker histone corresponds to an energy barrier which inhibits the unrolling of the DNA arms.

To estimate the average rate of nucleosome skips in a chromatin fiber data for the average nucleosome occupancy per bp [186] was used. This data was obtained by experiments combined with a probabilistic prediction model (cf. Sec. 6.4.2). For the E2A model a recent prediction of the nucleosome occupancy for the whole yeast genome [187] was used.

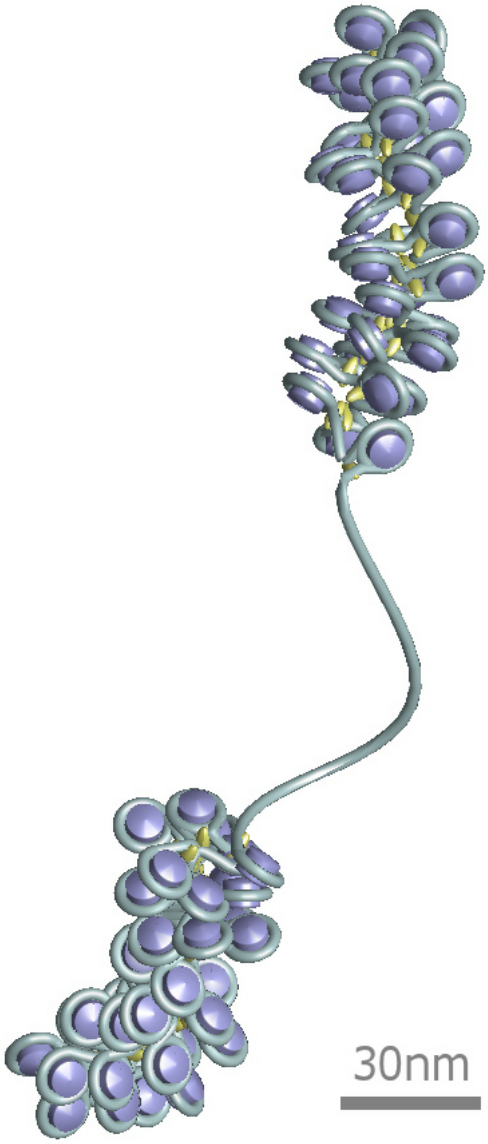


Figure 73: An example of a single nucleosome skip. If a nucleosome is dissolved, a blank stretch of DNA will remain. The naked DNA stretches have lengths of multiple integers of the nucleosome repeat length plus once the length of a DNA linker and can either lead to a collapse or to a swelling of the chromatin fiber (cf. Sec. 7). In both cases they increase the flexibility of the chromatin chain massively.

6.4.2 The Average Nucleosome Occupancy

In [186] the average nucleosome occupancy was partially determined by a combined experimental and computational approach. They looked for DNA sequence preferences of nucleosomes and the intrinsic nucleosome organization of the genome that these preferences dictate and demonstrated that eukaryotic genomes use a nucleosome positioning

code (cf. Fig. 74 and Fig. 185 in App. A).

Segal et. al. isolated nucleosome-bound sequences at high resolution from yeast and used these sequences in a new computational approach to construct and validate experimentally a nucleosome–DNA interaction model and to predict the genome-wide organization of nucleosomes. Their results demonstrate that genomes encode an intrinsic nucleosome organization and that this intrinsic organization can explain ≈50% of the *in vivo* nucleosome positions. This nucleosome positioning code may facilitate specific chromosome functions including transcription factor binding, transcription initiation and even remodelling of the nucleosomes themselves.

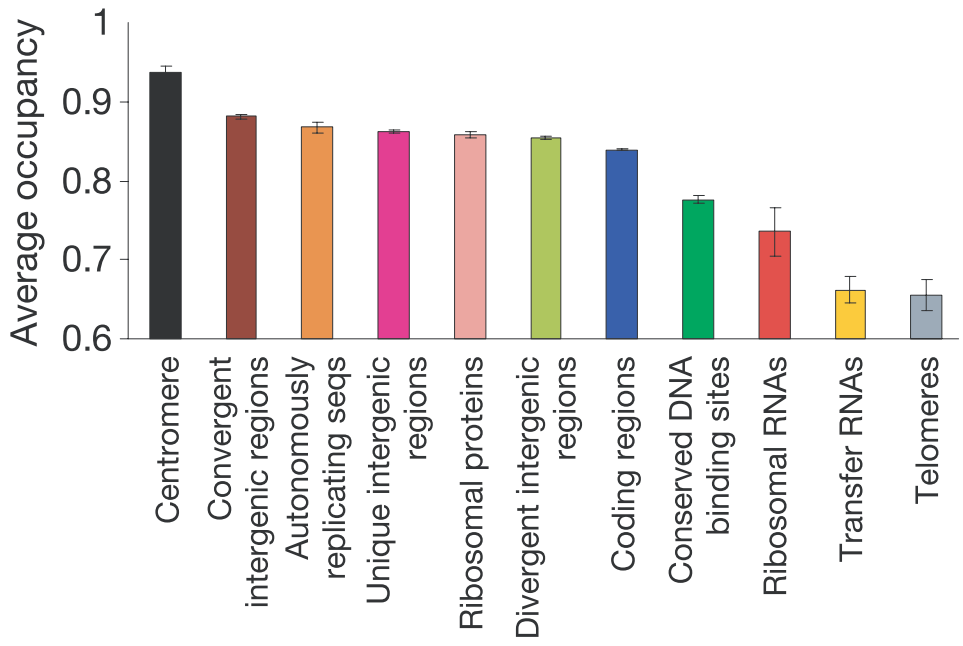


Figure 74: Average nucleosome occupancies and standard errors for different types of genomic regions. Adapted from [186].

Segal et al. extended their model in 2008 to make a prediction for the whole yeast genome [187] and found an average nucleosome occupancy of 68% (cf. Fig. 75), i.e. not all nucleosome locations are actually occupied by a histone octamer. Instead, one gets strong depletion effects that affect the chromatin fiber properties massively as will be shown below (cf. Sec. 6.5 and Sec. 7). It will turn out that increasing the amount of nucleosome skips (i.e. the nucleosome depletion) can either lead to a collapse or to swelling of chromatin fibers.

The prediction of the mean average nucleosome occupancy in [187] will be used as an average reference value for the E2A model.

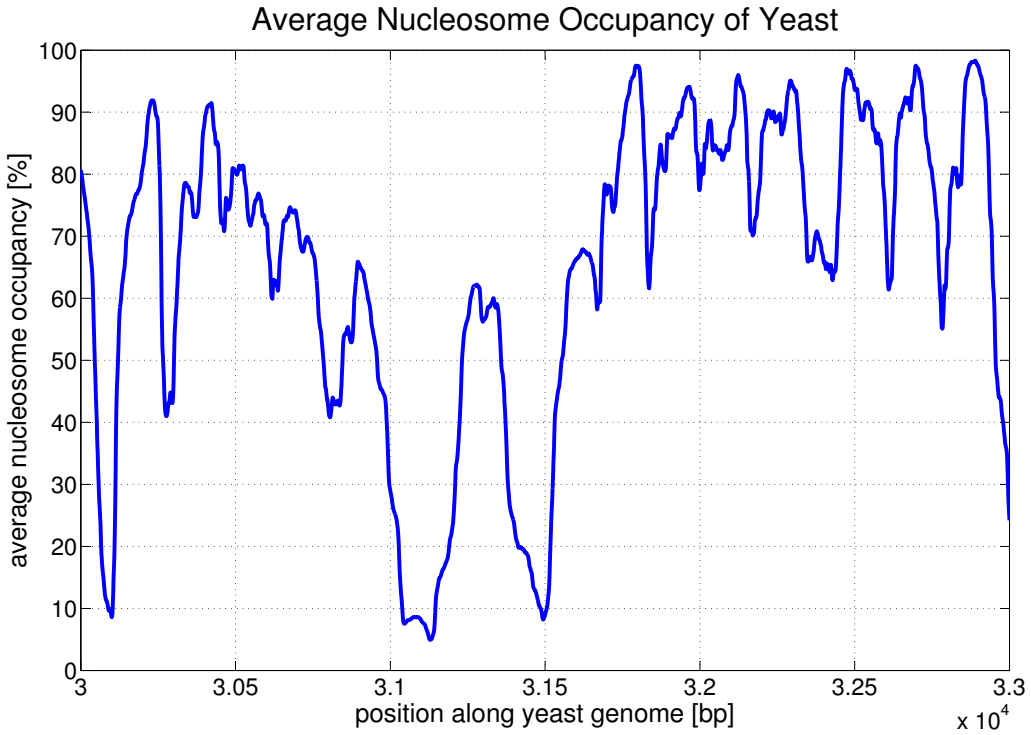


Figure 75: A small cut-out of the average nucleosome occupancy of the whole yeast genome [187]. The average value for the entire genome (12.1 Mbp) is 68%.

The transformation between the average nucleosome occupancy n in [186; 187] and the nucleosome skip probability p_{nuc} in this model is

$$p_{\text{nuc}} = 1 - n\frac{197\text{bp}}{147\text{bp}} = 1 - 1.34n.$$

A negative value for the nucleosome skip probability means that even more base pairs are covered by nucleosomes than expected by a regular, ideal chromatin strand. In that case the linker lengths between the nucleosomes have to be adjusted (i.e. narrowed) to get the given nucleosome occupancy n. With a linker length of approximately 50bp and ≈150bp wrapped around each histone complex one would expect an average nucleosome occupancy of 75% for a perfectly regular chromatin strand.

The predicted nucleosome occupancy for the yeast genome [187] had a mean value of ≈ 68% which means that one should expect a nucleosome skip rate of 8% for an average chromatin fiber.

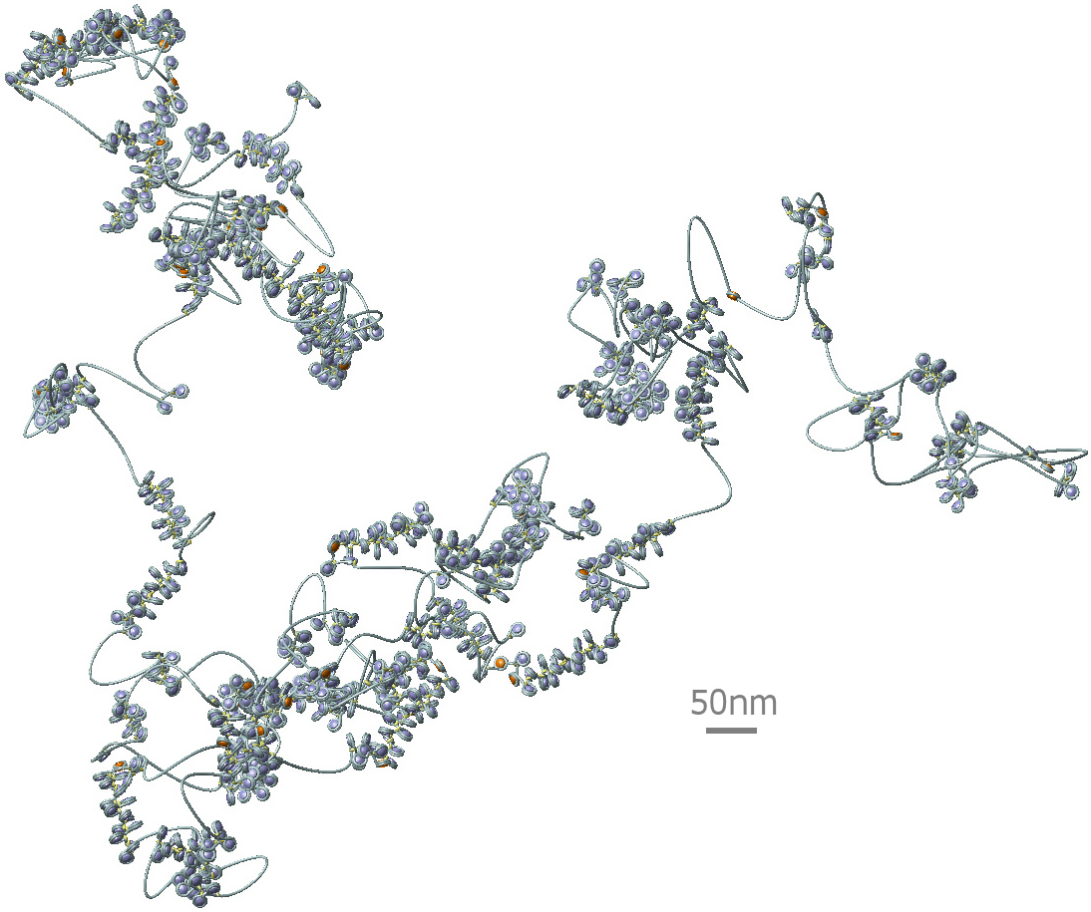


Figure 76: Example conformation of a chromatin fiber with depletion effects: The linker histone skip rate is 6% and the nucleosome skip rate is 8%. The linker histone skips are marked orange. One can see that the concept of a regular 30nm fiber does not hold anymore. Instead, one gets very flexible coil-like structures of compact regions which are separated by blank DNA stretches. The fiber has a total length of 394kbp.

Two conformations for chromatin fibers with a nucleosome skip rate of 8% and an estimated linker histone skip rate of 6% are shown in Fig. 76 and Fig. 77. Later on (cf. Sec. 7), the skip rates are treated as parameters in our model, i.e. they are varied over a large range of possible values to see how depletion effects affect chromatin properties. Fig. 76 and Fig. 77 will be discussed in Sec. 6.5.2.

6.5 Some First Results

Now that one can generate flexible instead of only stiff fibers and thus one can ask for properties that describe this flexibility: Chromatin strands with the parameter distributions that have been determined in Sec. 6.3.2 (cf. Tab. 5) have a persistence length (l_p) of ≈280nm, a line density (λ) of 5.8 nucleosomes per 11nm and an average

diameter of 34.6nm. These values are in excellent agreement with experimental values from the literature [34]. Furthermore, they agree with other chromatin models [116; 223].

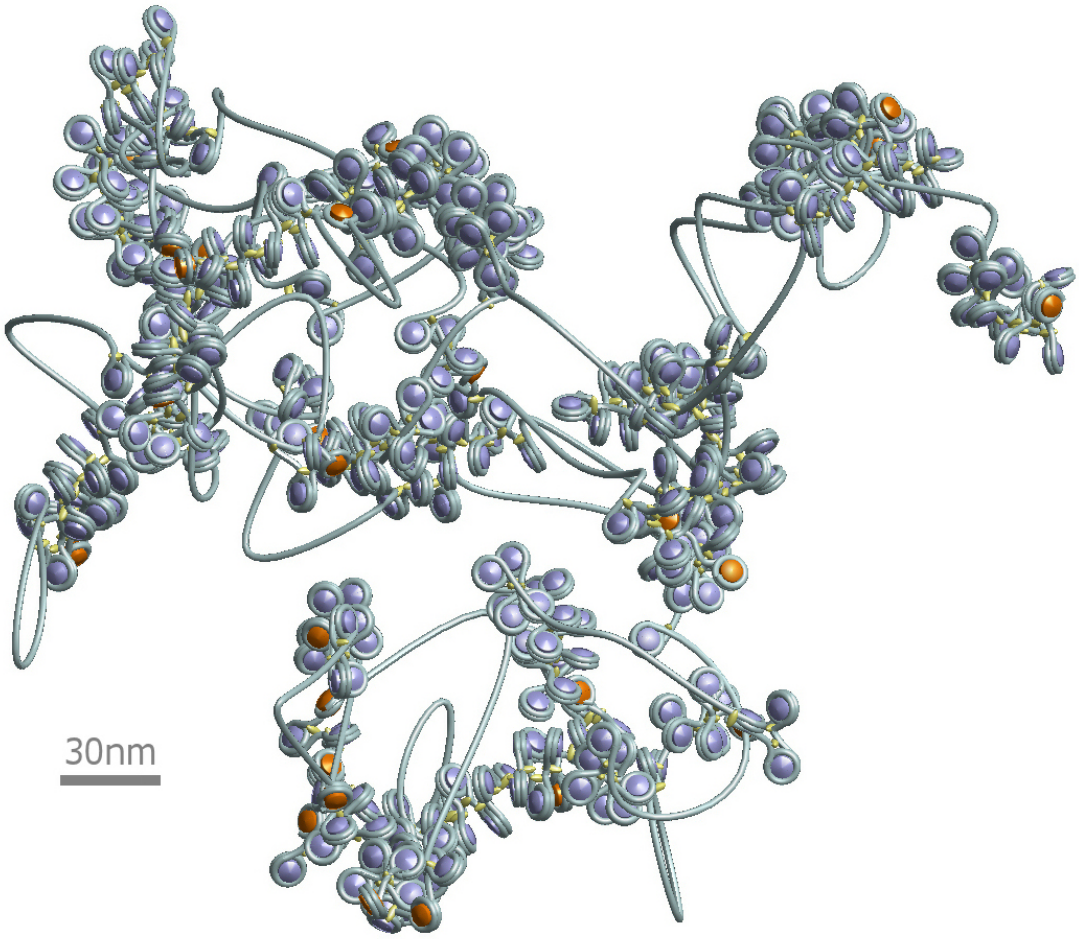


Figure 77: Another example conformation of a chromatin fiber with depletion effects: The linker histone skip rate is again 6% and the nucleosome skip rate is 8%. This shorter fiber has a total length of 98.5kbp. The linker histone skips are marked orange again.

Experimental values of the persistence length of chromatin depend on the technique which was used to determine it. Analysis of the distances between genetic markers in nuclei from human fibroblasts obtained from experiments using fluorescence in situ hybridization [214; 239] using a Flory model results in a value of 100-140 nm and an analysis using the Porod-Kratky wormlike chain model in 70-110 nm [145].
Furthermore, the persistence length was determined by an analysis of the end-to-end distances of chromatin fibers bound to a mica surface and measured with scanning force microscopy. This results in a value of 30-50 nm [37] as cited in [106].

It is not clear whether or not the binding of the fiber to mica causes artifacts. Experiments stretching single chromatin fibers from chicken erythrocytes with laser tweezers resulted in a value of 30nm [47] at low salt concentrations. Unfortunately, no data for the persistence length was given at physiological salt.

6.5.1 Flexible Chromatin Fibers without Depletion Effects

The scaling of the end-to-end distance and the radius of gyration in the case of flexible chromatin fibers without depletion effects was already discussed in Sec. 6.3.3.
The chromatin fibers which are generated with the probability distributions from Sec. 6.3.2 do not have a fixed structure (or topology) like the regular ones that have been examined in the last chapter (cf. Sec. 5). Instead, one gets a probability *distribution of structures* which is shown in the background of Fig. 78. This was already indicated in Sec. 5.6. One can see that the main part of a chromatin strand is still a crossed-linker fiber but for instance some smaller parts can be in more open beads-on-a-string-like or solenoidal states (cf. Fig. 82 for an example conformation).

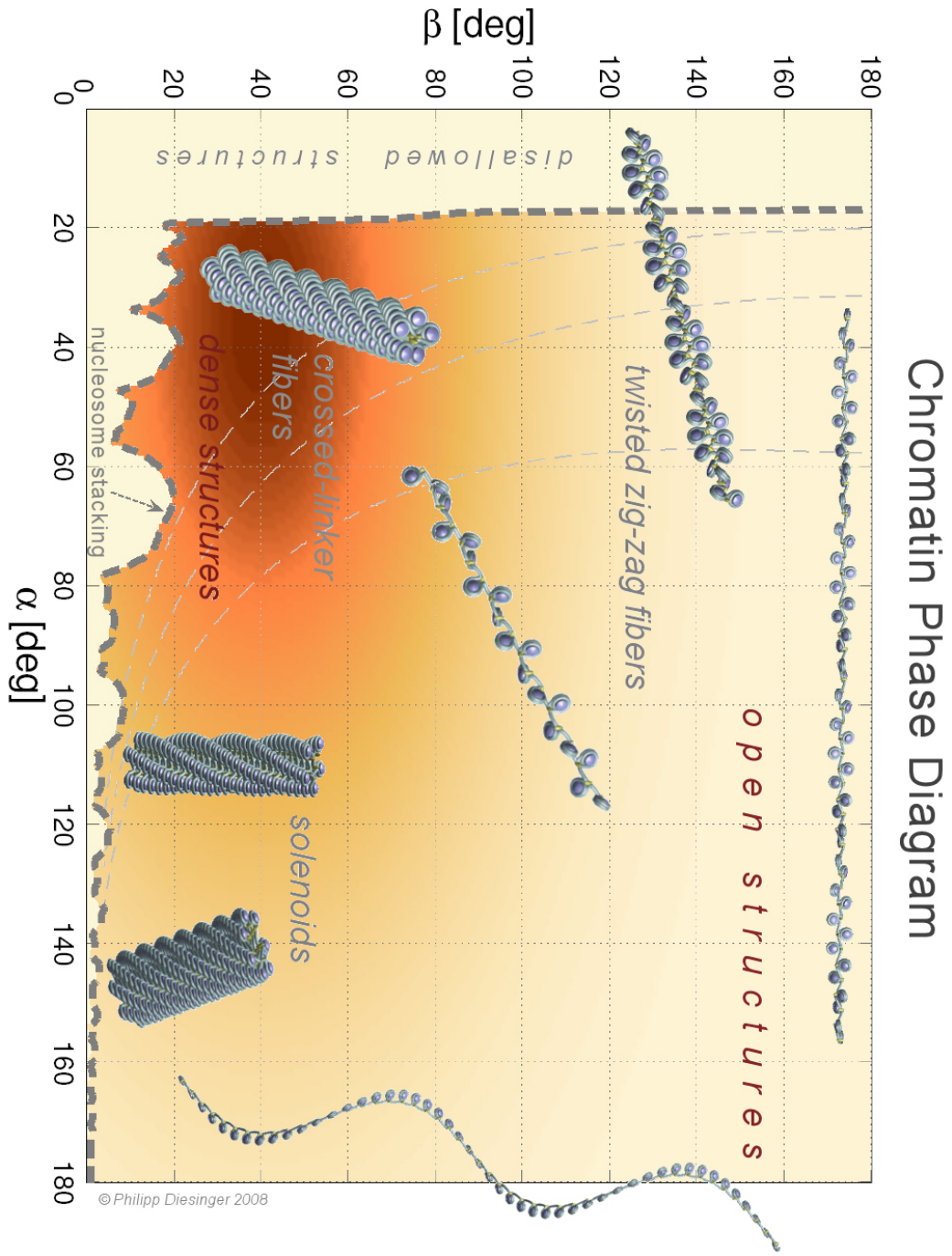


Figure 78: A cut-out of the actually four-dimensional chromatin phase diagram. The pitch d and the linker length b are fixed again. A single point in this diagram corresponds to a specific chromatin structure. The forbidden structures lie left and below the dashed line which is the excluded volume borderline [62]. Due to the parameter distributions of the model one should not expect a specific chromatin structure but instead a distribution of structures in the phase diagram. This probability distribution is shown in the back of the figure.

6.5.2 Flexible Chromatin Fibers with Depletion Effects

The depletion rates which have been applied in order to generate the fibers that are discussed in this section were 6% for the linker histone skips and 8% for the nucleosome skips. The chromatin chains with these depletion rates have a persistence length of ≈140nm, i.e. approximately half the value for the flexible fibers without any skips.

The flexible chromatin chains with depletion effects comply the scaling laws of a self-avoiding walk as well but since the linker histone skips and the nucleosome depletion increase their flexibility the convergence to the $N^{2\nu}$-power law is much faster (approximately one order of magnitude on the length scale). This can be seen below in Fig. 79 as well as in Fig. 186 and Fig. 187 where the leading order was divided out again. The scaling of R^2 and R_G^2 for flexible fiber without any depletion effects was discussed in Sec. 6.3.3.

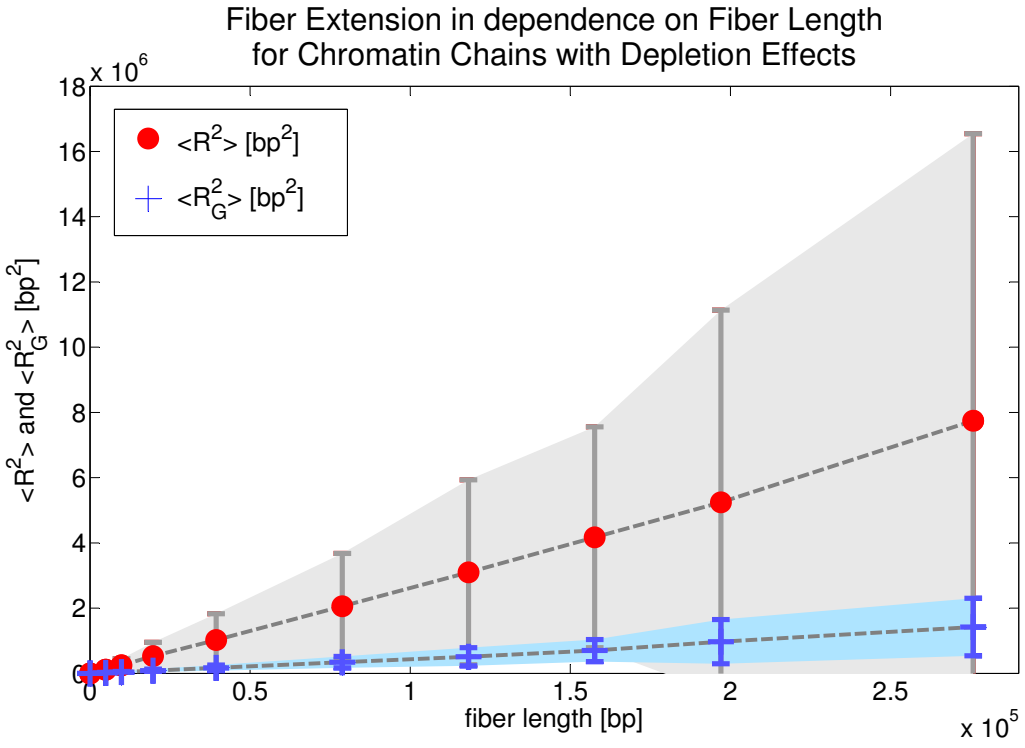


Figure 79: The mean squared end-to-end distance and the mean squared radius of gyration scale with $N^{2\nu}$ like self-avoiding walks ($\nu \approx 0.589$). The error bars give the standard deviation of R^2 and R_G^2.

By the comparison of Fig. 70 (i.e. a chromatin fiber without nucleosome or linker histone depletion) with fiber conformations that allow for depletion effects (Fig. 76 and Fig. 77) one can easily see that an expectation of a regular 30nm chromatin fiber is not justified anymore. Instead, one gets a coil-like structure of some more or less regular

parts separated by very flexible parts which mainly consist of blank DNA stretches. The linker histone depletion does not destroy the picture of a regular 30 nm fiber as much as the nucleosome depletion. Moreover, one has to keep in mind that the nucleosome depletion rate which was used to generate the conformations in Fig. 76 and Fig. 77) is only the *average* depletion rate (of the yeast genome). Some parts of the genome (for instance telomeres [186; 187]) have even larger nucleosome depletion rates, and therefore, are much more coiled. Furthermore, small nucleosome skip rates already induce very coiled structures (cf. Sec. 7).

Fig. 76 and Fig. 77 suggest that the concept of a specific regular 30nm chromatin fiber possibly has to be adjusted towards a distribution of structures (cf. Fig. 78) which coexist in one chromatin fiber. Furthermore, depletion effects which obviously occur in vivo lead to much more coiled conformations even at small skip rates than one would expect from existing chromatin models [19; 38; 62; 64; 71; 80; 179; 183; 211; 216; 217; 218; 230; 235].
These coils surely show small stretches of 30nm fibers that are separated by large pieces of naked DNA but a perfectly regular 30nm fiber seems to be very unlikely since it needs to be almost completely saturated with nucleosomes. Even from a thermodynamic point of view this seems unlikely: Nucleosomes have the ability to dissolve and form again so a dynamic equilibrium resulting in a certain nucleosome skip rate seems very plausible. Maybe these regular fibers exist only under lab conditions where they are built up at high histone concentrations, and therefore, are entirely saturated with nucleosomes and linker histones.

6.5.3 Comparison with Electron Micrographs

At the end of this chapter some example conformations that can be found, if one takes the parameter distributions as well as the depletion effects into account, are compared with electron micrographs.
Fig. 80 shows a comparison of a 30nm fiber obtained by electron microscopy with a 30nm crossed-linker fiber from the E2A model whereas Fig. 81 shows a more open conformation with some nucleosome and linker histone skips. In this case one can clearly identify a beads-on-a-string structure in the EM-image and one can even distinguish single nucleosomes. Finally, Fig. 82 shows that even single nucleosomes separated by blank DNA stretches can be found. The EM images were adapted from [151].
Another figure with scales for the model conformations and without the fuzzy background of Fig. 80 - Fig. 82 can be found in App. A, Fig. 188.

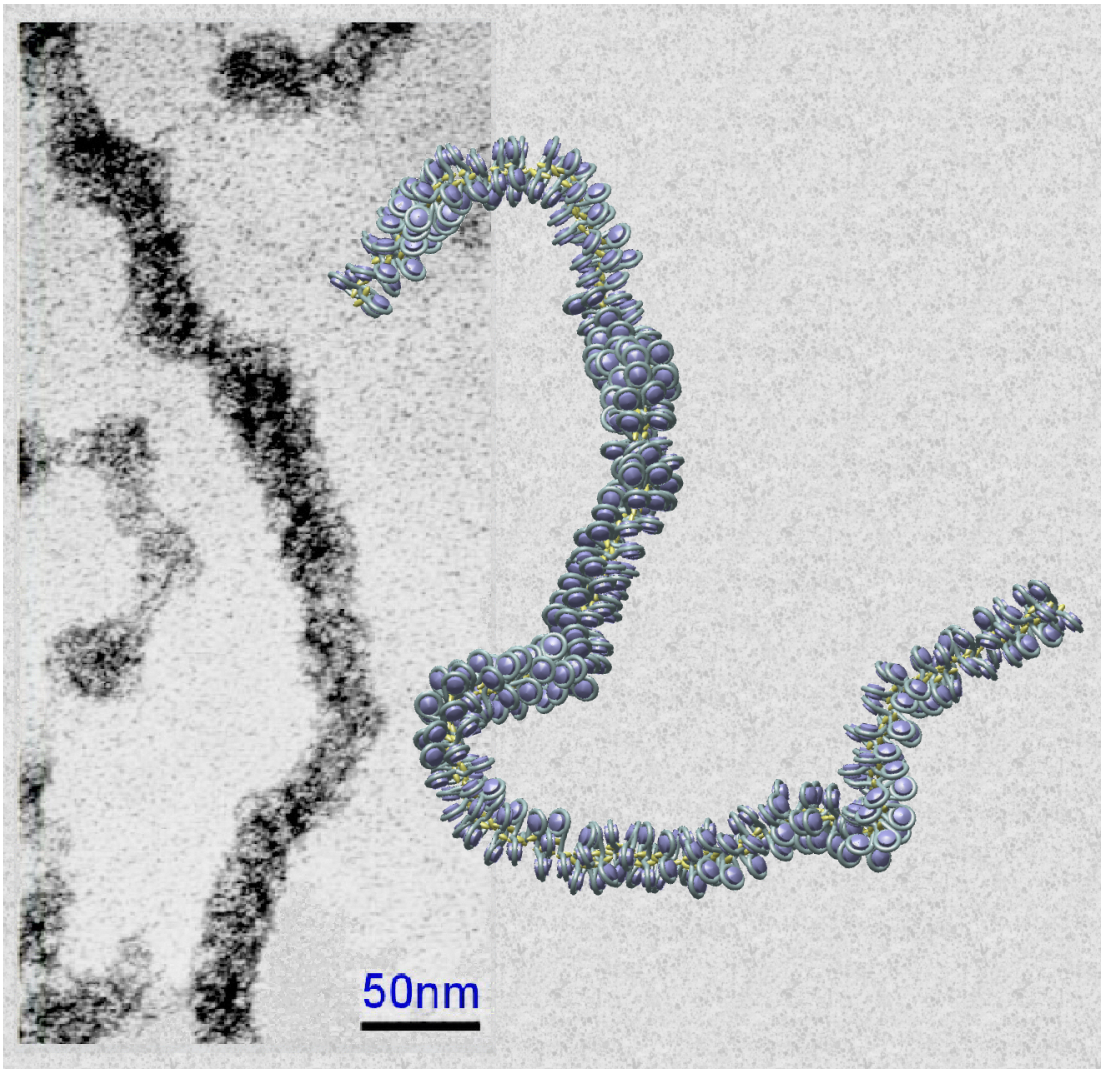


Figure 80: This figure shows an example conformation of a chromatin strand of length 40kbp compared to a similar electron micrograph [151]. The light blue tubes represent the DNA, the histone octamers are modelled as purple cylinders and the linker histones are marked pale yellow.

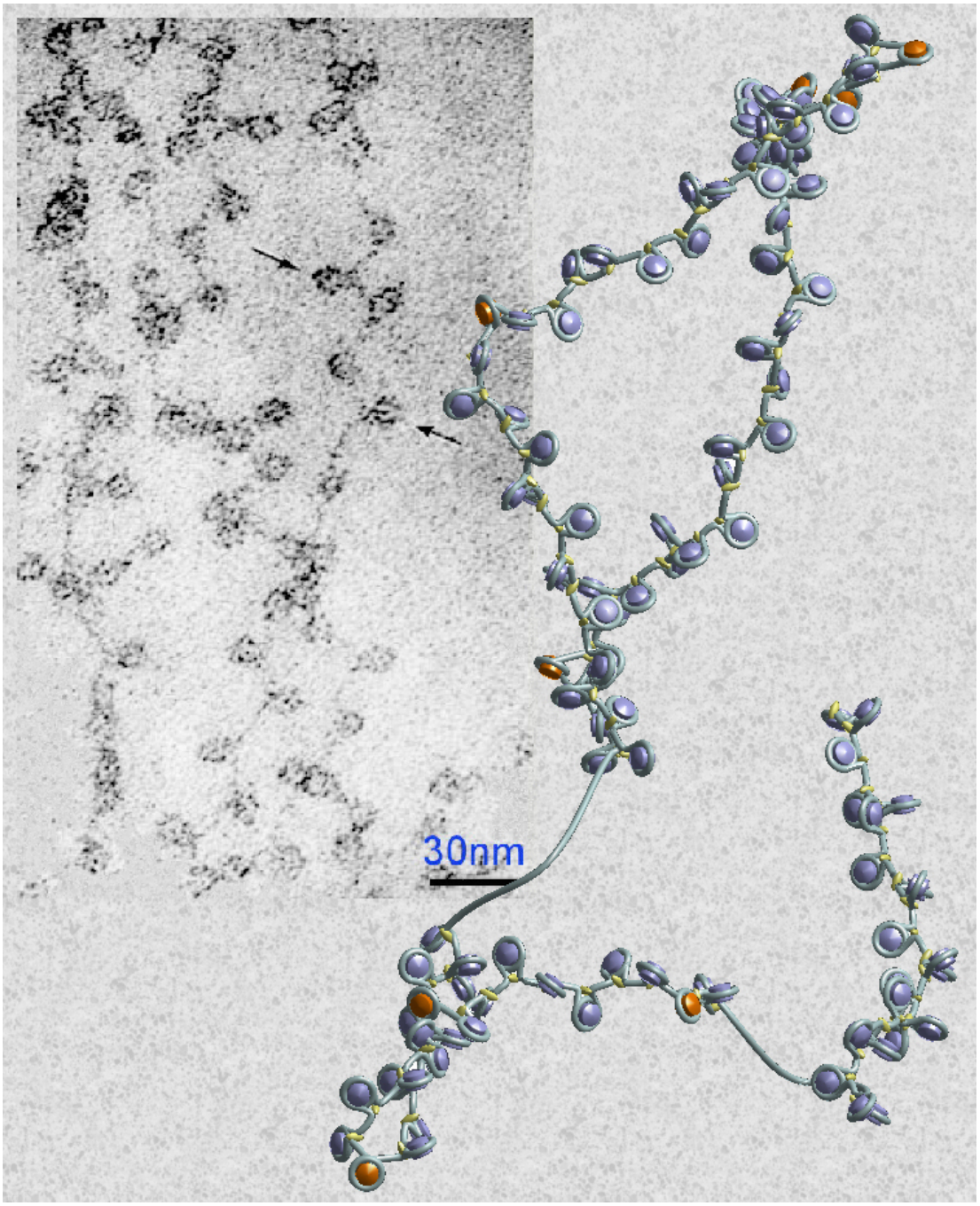


Figure 81: This figure shows a beads-on-a-string-like structure with some nucleosome and linker histone skips. The latter are marked orange. On the left side of the figure a similar electron micrograph is shown [151].

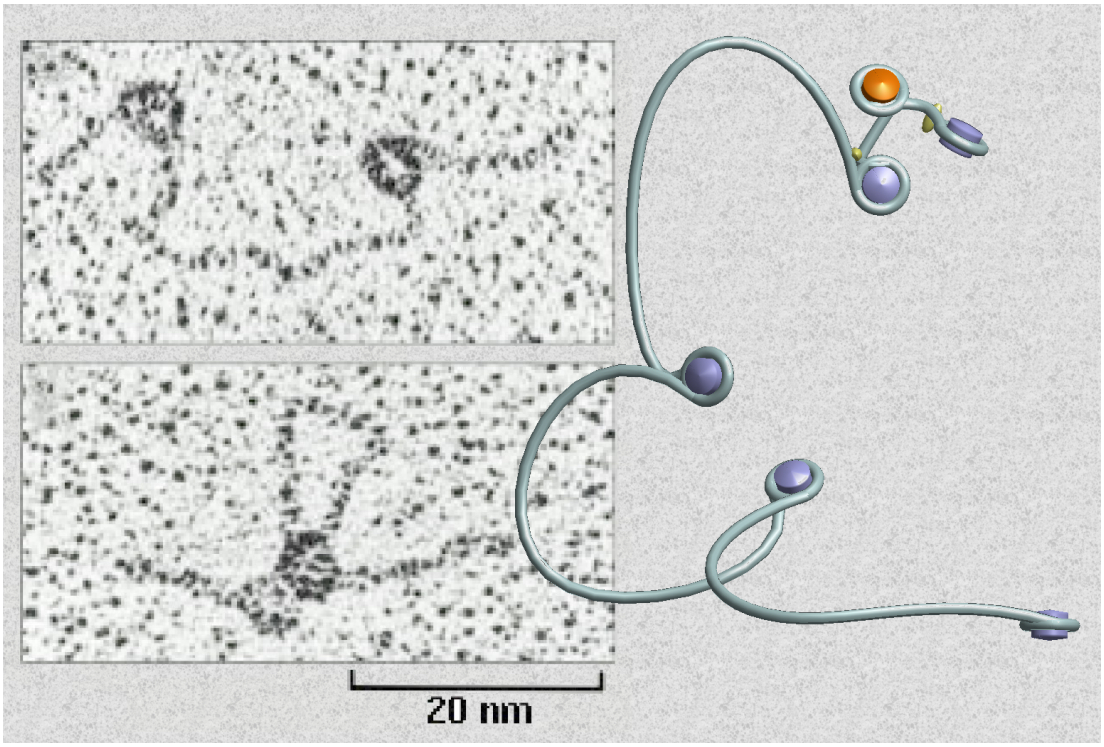


Figure 82: Another comparison of a model conformation with an electron micrograph from [151]. The conformations consist of single nucleosomes connected by blank DNA stretches.

7 Influence of Histone Depletion on Chromatin Properties

7.1 Overview

In Sec. 6 linker histone depletion as well as nucleosome depletion have been introduced and integrated into the chromatin model. Some first depletion effects have been discussed in Sec. 6.5.
In this chapter various chromatin features will be discussed against the background of nucleosome and linker histone depletion: For every one of the considered chromatin properties we will show interest in the differences between chromatin fibers with histone depletion and fibers without histone depletion.

It will turn out that depletion of linker histones and nucleosomes affects massively the flexibility and the extension of chromatin fibers (cf. Sec. 7.2). Increasing the amount of nucleosome skips (i.e. the nucleosome depletion) can either lead to a collapse or to swelling of chromatin fibers. These oppositional effects will be discussed in Sec. 7.2.2 and it will be shown that histone depletion may even contribute to DNA condensation. Furthermore, it will be shown that predictions from experimental data [187] for the average nucleosome skip rate lie exactly in the regime of maximum DNA condensation. This could be a hint that cells are in fact using the plateau regime in Fig. 86 to compact DNA or even regulate its extension.
Transcription factors increase the nucleosome skip rate locally and according to the results of this section (cf. Sec. 7.2.2) this leads automatically to a swelling of the chromatin strand. This could explain why genome parts that are transcribed very frequently tend to lie more inside of the nucleus [45].
This is a hint that nucleosome and linker histone depletion may not only be some sort of defect in the chromatin fiber instead they might even be *functional*.

In Sec. 7.3 the chromatin nanostructure will be discussed with respect to radial nucleosome distributions (cf. Sec. 7.3.1). Again the two cases of chromatin fibers with and without histone depletion will be discussed and the relation between the spatial and the genomic distances in the found peak structure will be explained.
Furthermore, it turns out that the peak structure in the radial nucleosome distribution and the nucleosome pair distribution function is even preserved in the case of fibers with depletion effects (cf. Sec. 7.3.2). This peak structure reflects the local order of the nucleosomes in the chromatin fiber.
In principal, the radial pair distribution function could be determined by scattering experiments.

After the discussion of the radial pair distribution function in the case of three-dimensional chromatin fibers, the radial distance distribution $p(r)$ as well as $g(r)$ are determined in the case of previously projected and thus two-dimensional fibers. This is done to provide a direct comparison with the SPDM data in Sec. 3. The comparison is presented in Sec. 7.3.4 and although the application of $g(r)$ is problematic in the case of single fibers instead of melts one can see a good agreement.

Finally, in Sec. 7.4 the simulation results for the nucleosome distributions are compared with FISH measurements [34].

The main results of this section have already been published [65].

7.2 Chromatin Extension and Flexibility

In this section the impact of histone depletion on DNA condensation and on chromatin flexibility will be investigated. It will be shown that nucleosome depletion may even contribute to the compaction of chromatin.

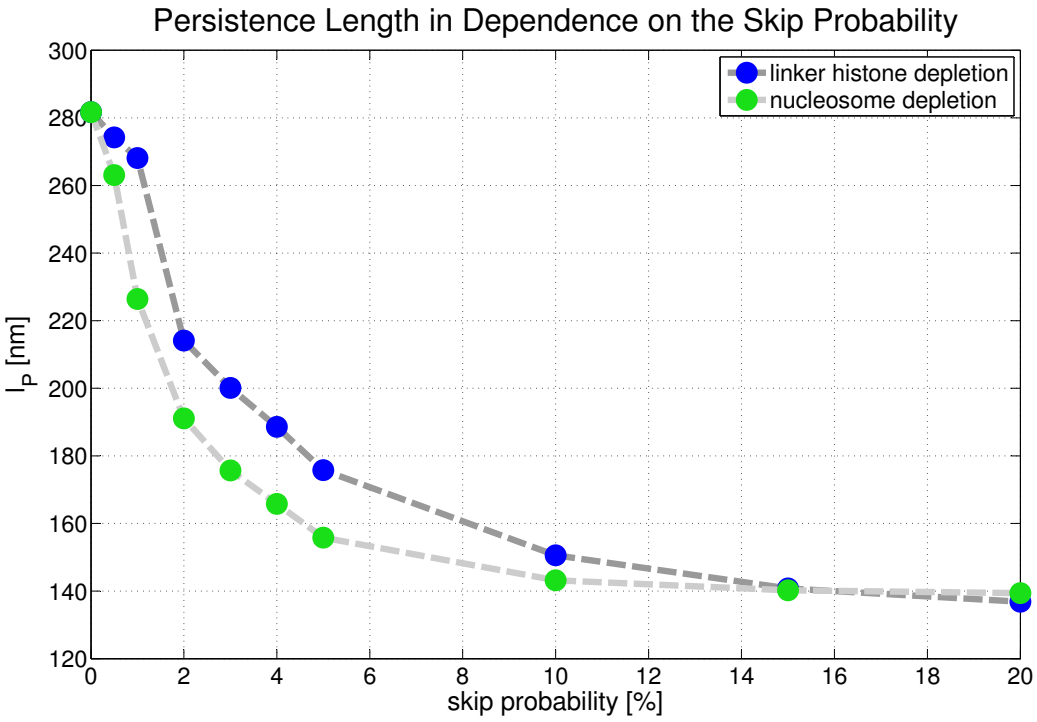


Figure 83: This figure shows how depletion effects decrease the chromatin persistence length l_P. Here are either linker histone or nucleosome depletion effects considered. The persistence length decreases very similar for both cases. Nucleosome depletion has a slightly larger effect for small skip rates.

Fig. 83 shows how either linker histone or nucleosome depletion affect the flexibility of chromatin fibers: Without any depletion effects (i.e. zero skip rate) the fiber has a persistence length of about 280nm. If the skip rate for either linker histones or nucleosomes increases, the persistence length will drop very fast in both cases, i.e. the fiber will get much more flexible.

Fig. 84 shows the persistence length in dependence on the skip rates but this time *combinations* of the two kinds of histone depletions are considered as well.

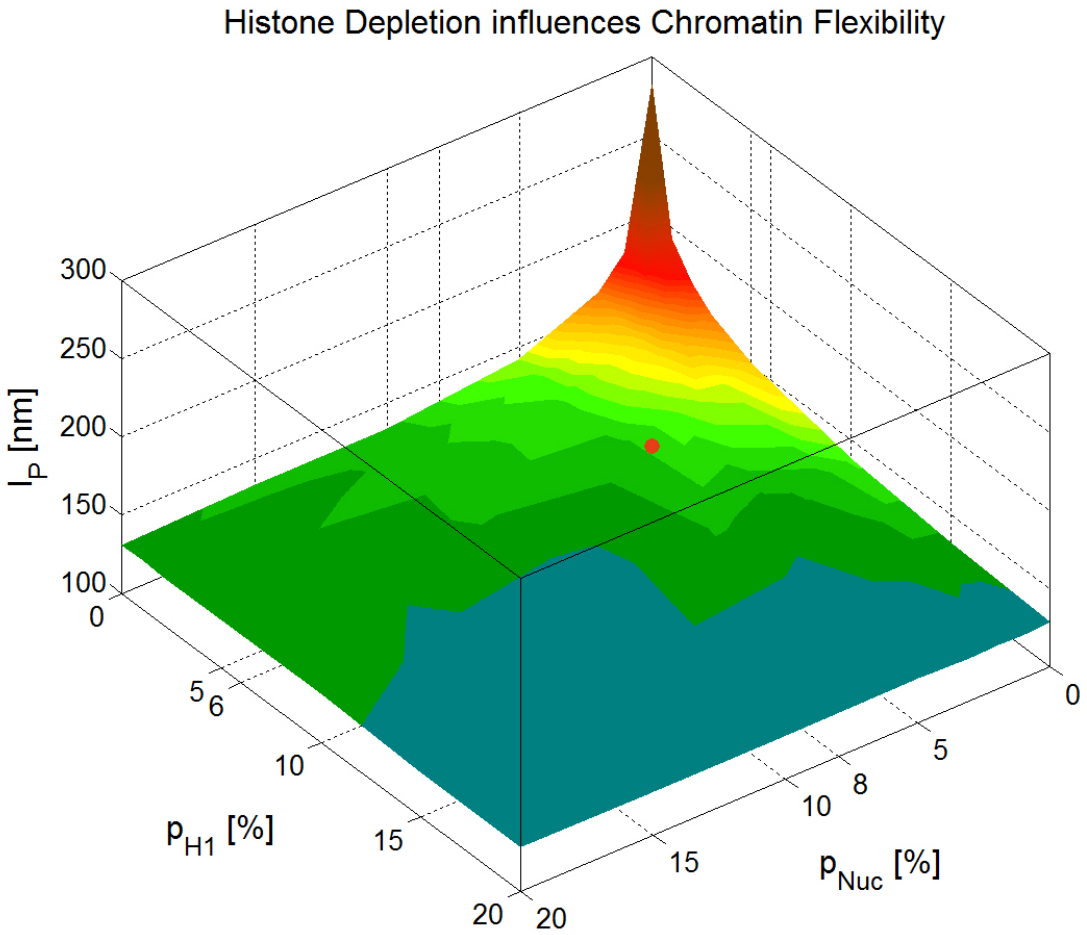


Figure 84: This figure shows how depletion effects decrease the chromatin persistence length l_P. In this case combinations of linker histone and nucleosome depletion effects are considered. The red spot marks the expected average rates for the two depletion effects [187].

The radius of gyration as well as the mean squared end-to-end distance are used as a measure for the fiber extension. Furthermore, it is interesting to see how they are changed by the depletion effects. In Fig. 85 and Fig. 86 one can see how either linker histone depletion or nucleosome depletion affect the fiber extension and Fig. 87 and Fig. 88 again show the same results for the *combinations* of the two types of histone depletions.

7.2.1 Linker Histone Depletion

In the case of linker histone skips the fiber extension decreases with increasing skip rate until a skip rate of ≈10% (cf. Fig. 85 and Fig. 86). In terms of DNA condensation that means it is even necessary to have a linker histone depletion rate of about 10% or, vice versa: It is not efficient to have more than 90% of all nucleosomes saturated with linker histones because at some point the fiber gets too stiff.

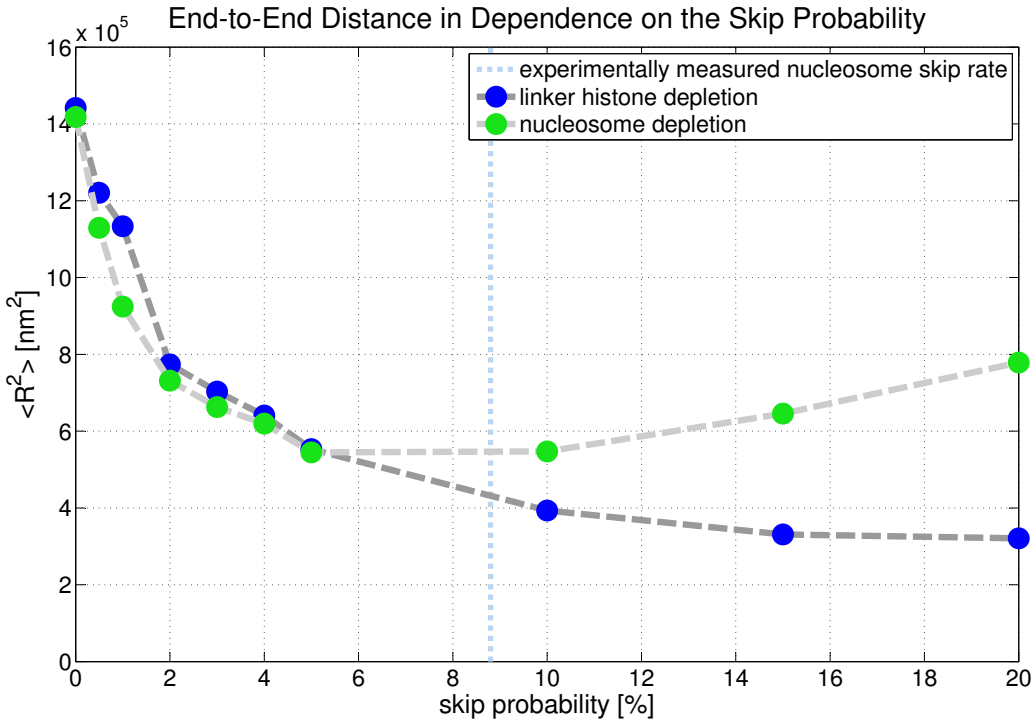


Figure 85: The mean squared end-to-end distance is used as a measure for the extension of the chromatin fiber. With increasing linker histone depletion (blue) it decreases, whereas with increasing nucleosome depletion (green) there is a plateau regime after a initial decrease and at the end the fiber extension even increases. These effects come on the one hand from the gain of flexibility that the nucleosome skips bring in and on the other hand from the spatial need of blank DNA in opposite to the compact wrapped DNA. In the plateau regime theses two effects level off. This is the regime where DNA condensation is optimal. As one can see the experimental values [187] for the nucleosome skip rate with an average of 8% lie just in this region of optimal condensation.

From experiments [19; 77] one would actually expect that chromatin fibers will become more compact, if one adds linker histones. One has to keep in mind that in this model linker histone depletion and nucleosome depletion are statistically independent but in experiments they are highly correlated since linker histone depletion *facilitates*

nucleosome depletion.
Furthermore, if one increased the linker histone skip rate up to 100%, one would probably see that the fiber DNA condensation decreases again (i.e. R^2 and R_G^2 increase). This is because at some point the gain of flexibility does not lead to denser states any more due to the excluded volume interactions of the fiber. So maybe one finds in the experiments only this last increase of the fiber extension and not the former decrease at the smaller skip rates.
Other chromatin models with variations of the nucleosome linker length show similar effects [223].

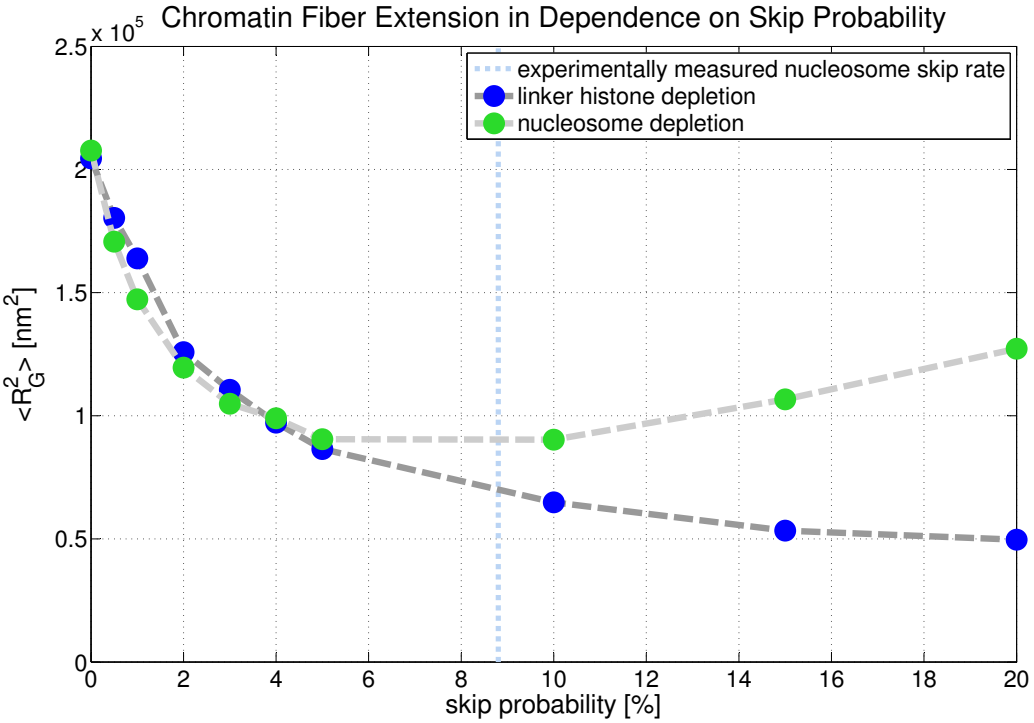


Figure 86: The mean squared radius of gyration as another measure for the extension of the chromatin fiber. It shows a similar behavior as the mean squared radius of gyration (cf. Fig. 85). The experimentally determined nucleosome skip rate is adapted from [187].

7.2.2 Nucleosome Depletion

Increasing the nucleosome skip rate has two effects on the chromatin fiber: First, it gets much more flexible due to the blank DNA stretches which connect more compact regions. Secondly, the naked DNA is much less condensed and thus needs more space due to entropic forces than the DNA which is wrapped up around histone octamers. In Fig. 85 and Fig. 86 one can see that the fiber extension first decreases with increasing

nucleosome depletion ($p_{nuc} \approx 0\%$ - 5%), then there is a plateau regime where it is almost constant (5% - 9%) and finally it increases again ($p_{Nuc} > 9\%$).

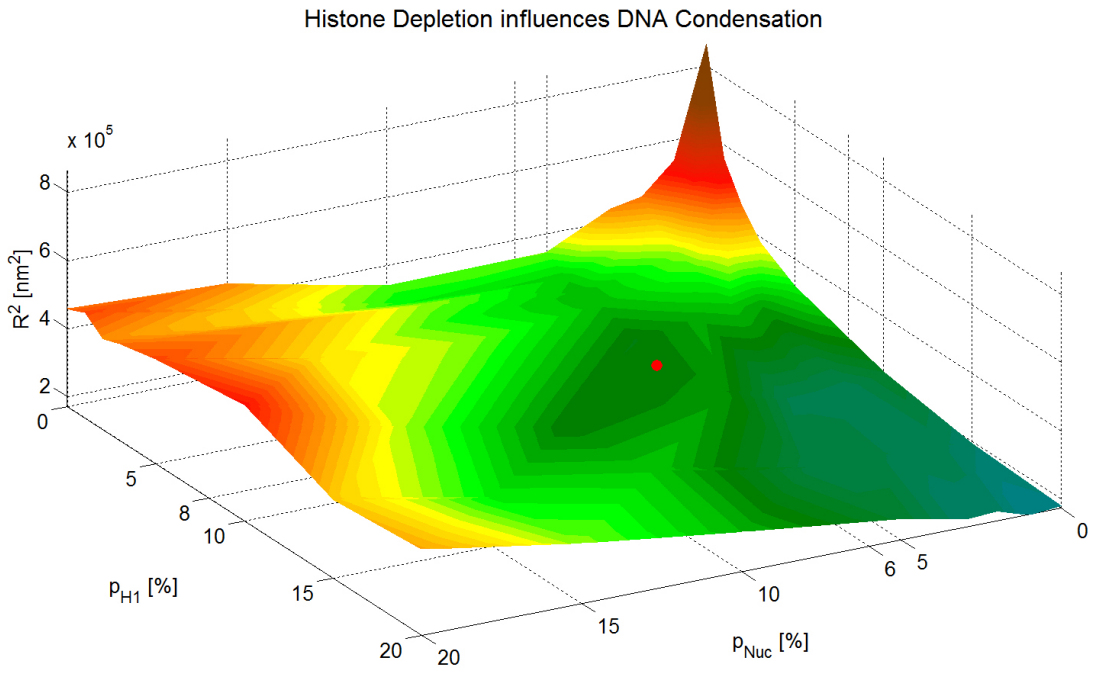


Figure 87: The mean squared end-to-end distance in dependence on linker histone and nucleosome depletion. Here combinations of linker histone and nucleosome depletion effects are considered. The red spot marks the expected average values that come from experimental data [187].

The first decrease comes from the increase of flexibility which the blank DNA stretches bring in. The increase of the extension at the end ($p_{Nuc} > 9\%$) stems from the spacial need of the blank DNA and in the plateau regime in the middle the two effects level off.

One can also consider this from the view of DNA condensation: At a nucleosome skip rate of 100% one would have the whole genome as a naked stretch of DNA. Adding nucleosomes at this point means that one *decreases* the nucleosome skip rate. Then, at the plateau regime the DNA condensation is optimal because a further increase will make the chromatin fiber too stiff.

One has to keep in mind that the experimental data for the average nucleosome occupancy [187] suggested an average nucleosome skip rate of 8% which lies just in this plateau regime of optimal DNA condensation. That means the experimentally determined average nucleosome occupancy is *optimal* in terms of DNA condensation.

This is nice because it shows that this plateau regime might actually be used by the cell

to compact the DNA: Cells could use it to regulate the chromatin fiber extension *locally*.

Transcription will increase the nucleosome skip rate in a genome region and thus (according to Fig. 85 and Fig. 86) lead automatically to swelling of the chromatin strand. This could explain why parts of the genome which are transcribed very frequently tend to lie more inside the nucleus [45].

The nucleosome skips are much more important for this effect than the linker histone skips (cf. Fig. 86 and Fig. 85).

In this context, linker histone and nucleosome skips may not just be *defects* in the chromatin fiber but instead they may play a very important role for DNA condensation and support a tool to locally regulate the chromatin fiber extension as well as its flexibility within the cell nucleus.

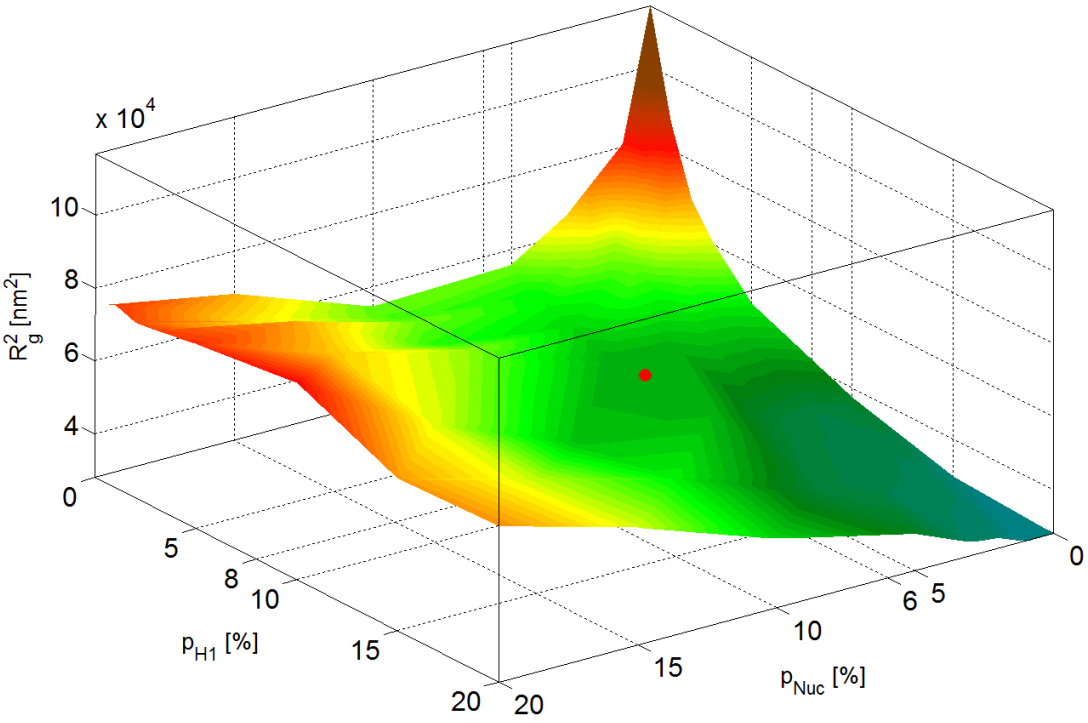


Figure 88: The mean squared radius of gyration as a measure for the extension of the chromatin fibers in dependence on histone depletion. Here again combinations of linker histone and nucleosome depletion effects are examined. The red spot marks again the expected average values that come from experimental data [187].

In Fig. 87 and Fig. 88 one can see that only in the case of linker histone skip rates less than ≈10% one will get a plateau regime in the fiber extension, if one considers different nucleosome skip rates. If the linker histone skip rate is larger, one will only find an increase in R^2 and R_G^2 with increasing nucleosome skip rate.

7.3 Chromatin Nanostructure

In this section the chromatin nanostructure will be examined. First, the radial nucleosome distribution is investigated (cf. Sec. 7.3.1), then the nucleosome pair distribution function is discussed (cf. Sec. 7.3.2) and it will be shown that even in the case of fibers with depletion effects its main dominant peaks can still be identified.

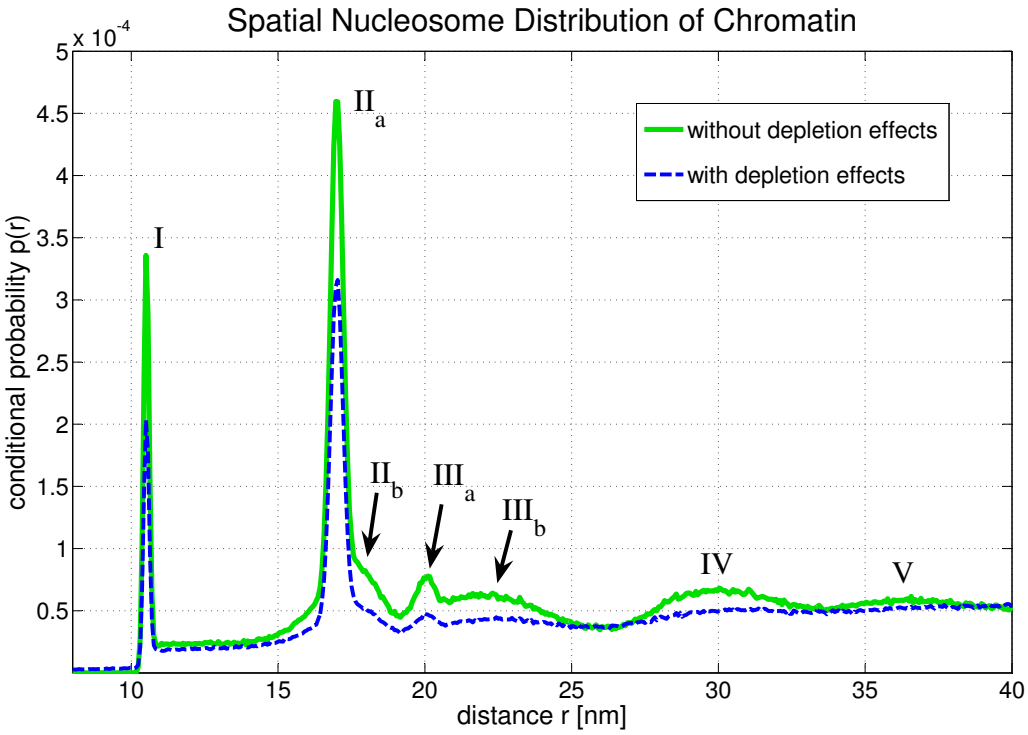


Figure 89: $p(r)$ is the conditional probability of finding a nucleosome at a spatial distance r, if another nucleosome is located at the origin. One can see a peak structure which comes from the local order of the nucleosomes in the chromatin fiber. The peaks r_Δ can be associated with certain genomic nucleosome-nucleosome distances Δ. Here it was averaged over approximately 10^{14} nucleosome distances.

7.3.1 Radial Nucleosome Distribution

Fig. 89 shows the conditional probability $p(r)$ to find a nucleosome at a distance r, if another nucleosome is located at the origin:

$$p(r) = \left\langle \sum_{i=1}^{N} \sum_{j \neq i} \delta(r - r_{i,j}) \right\rangle$$
$$= \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \delta(r - r_{i,j}).$$

Two different cases are considered: fiber with and without histone depletion. Is was averaged over 10^4 chromatin fibers of length 160kbp, i.e. over approximately 10^{14} nucleosome-nucleosome distances altogether.

As one can see $p(r)$ shows some very dominant peaks which are labelled I - V. They express the local nucleosome order in the chromatin fiber since they represent very frequent spatial nucleosome distances. They can be associated with certain genomic nucleosome-nucleosome distances.

The genomic distances are given in multiples of the nucleosome repeat length, i.e. they are integer numbers. The corresponding spatial nucleosome-nucleosome distance to a genomic distance Δ is denoted by r_Δ. Some of the genomic distances r_Δ are illustrated in Fig. 90.

The $r = 0$ peak is not shown in Fig. 89. It has approximately the same height as peak II.

$p(r)$ is almost zero until r is approximately larger than 10nm which is the diameter of the reference nucleosome.

Some of the peaks in Fig. 89 superimpose of several single r_Δ-distributions. This is indicated in some cases by subscript letters (cf. peak II and peak III). The allocation of the peaks to genomic nucleosome-nucleosome distances can be found in Tab. 6. For instance, the second peak (II) is a superimposition of the distribution of r_1 and the distribution of r_3.

peak #	I	IIa	IIb	IIIa	IIIb	IV	V
genomic distance Δ [NRL]	2	1	3	4	5	6,7	8,9,10
genomic distance Δ [bp]	394	197	591	788	985	1182 - 1379	1576 - 1970

Table 6: Allocation of peak number to particular genomic nucleosome-nucleosome distances for the first five peaks in $p(r)$ and the pair distribution function $g(r)$. The peak locations can be found in Tab. 15: They are determined by the most frequent values (mf_Δ) of the r_Δ-distributions.

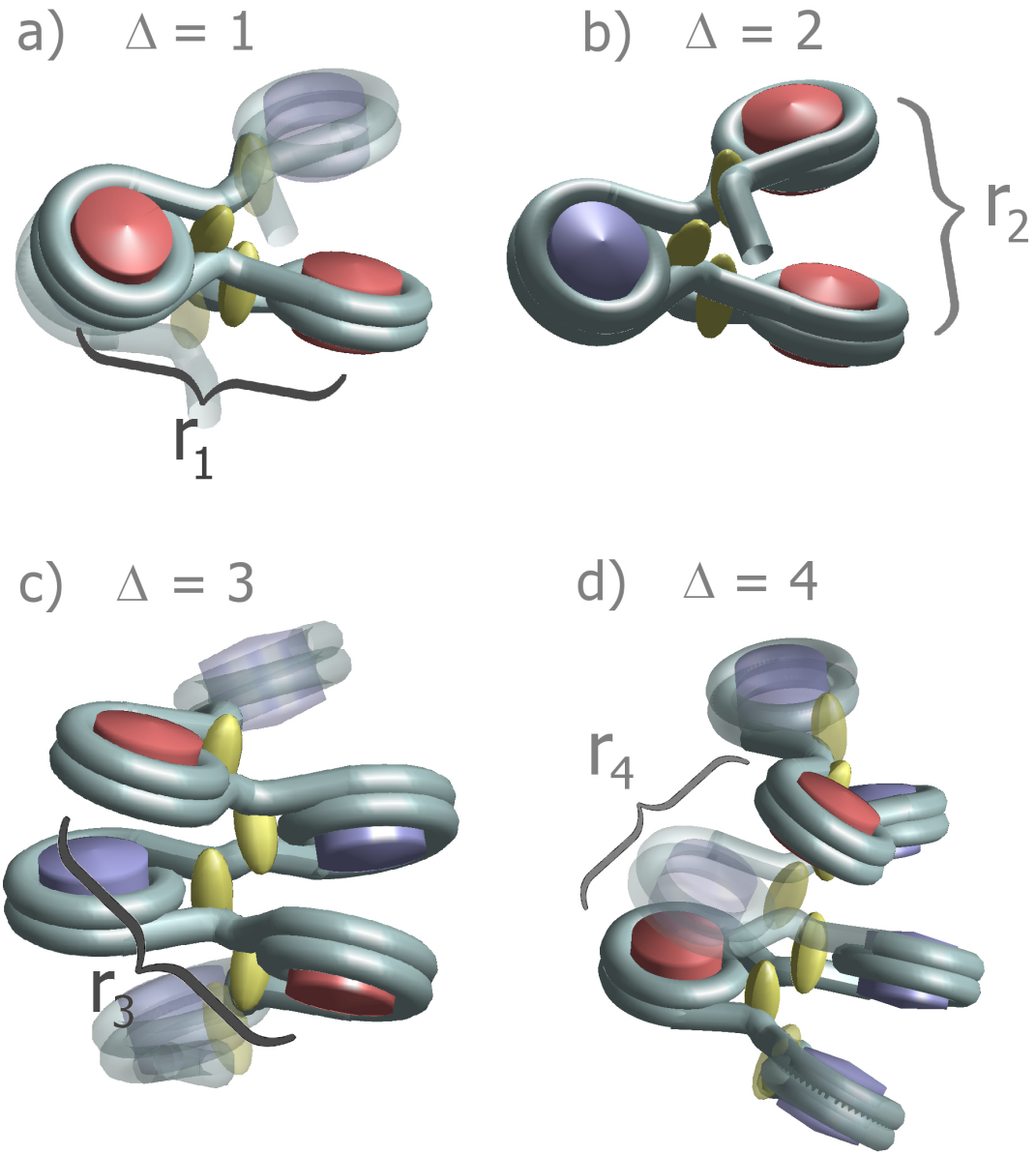


Figure 90: This figure illustrates the connection between the genomic distance Δ and the spatial distance r_Δ for the first four cases. The reference nucleosomes which have the genomic distance Δ (given in NRLs) are marked red.

The first five peaks of $p(r)$ can even be identified, if one allows for depletion effects (cf. Fig. 89), although in this case $p(r)$ is decreased in comparison to the conditional probability without depletion effects. This is because linker histone and nucleosome depletion destroy the local order at some points within the chromatin strand so that these spots do not contribute to the dominant peaks any longer.

Furthermore, the distance distribution with depletion effects is a bit shorter (cf. Fig.

92) since the fibers are more flexible (cf. Fig. 83). Nevertheless, the first peaks can still be identified clearly.
On the long scale the overall shape of $p(r)$ is very similar, although not identical, to the distribution of the end-to-end distance of a random walk (cf. Eq. 3).
The average depletion rates of 8% (for nucleosome skips) and 6% (for linker histone skips) were applied again in this section.

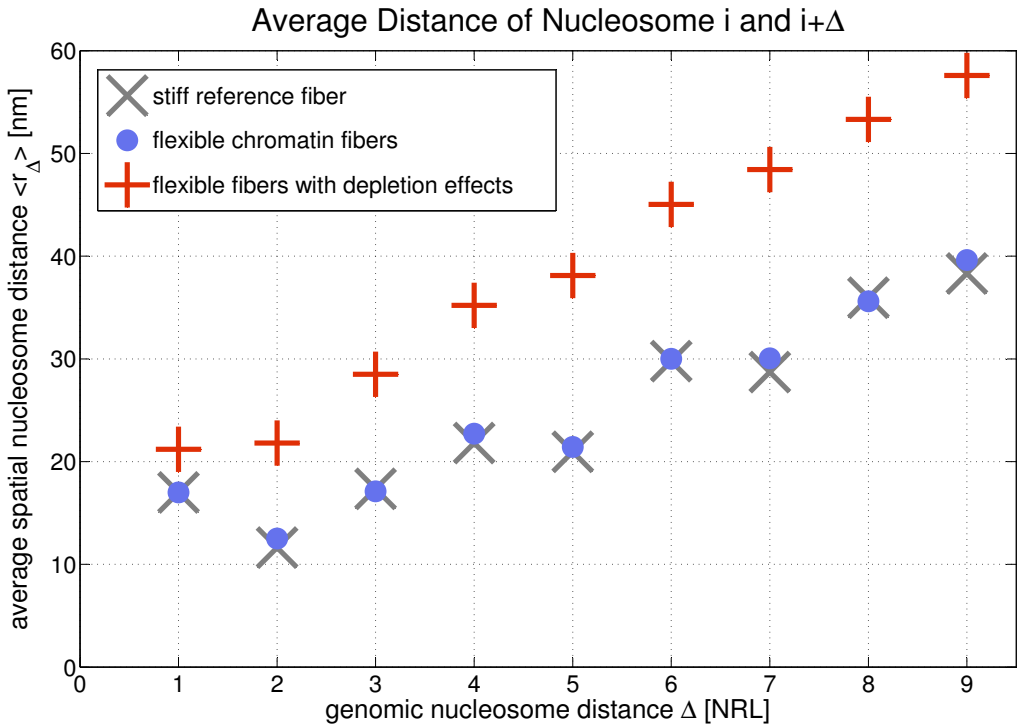


Figure 91: The mean spatial nucleosome distance $\langle r_\Delta \rangle$ in dependence on the genomic distance Δ for stiff, flexible and flexible fibers with histone depletion. Some of the r_Δ-distributions are very asymmetric which shifts the mean value far away from the reference value of the regular fiber.

Fig. 91 shows the dependence of the average spatial nucleosome distance on the genomic nucleosome distance for three different cases: Stiff fibers, flexible fibers and flexible fibers with depletion effects. One can see that the depletion effects have a great influence on r_Δ. This is because the depletion effects destroy the local nucleosome order and the average of r_Δ was also taken over possibly existing nucleosome or linker histone skips. Furthermore, the flexibility and the depletion effects shift the mean values of r_Δ because their probability distributions are very asymmetric in some cases. This can be seen in Tab. 15: The mean value of r_Δ and the most frequent value which is responsible for the peak location differ greatly in some cases.
The dependence of the mean spatial nucleosome distance on the corresponding genomic

distance on larger scales can be seen in Fig. 191 in App. A.

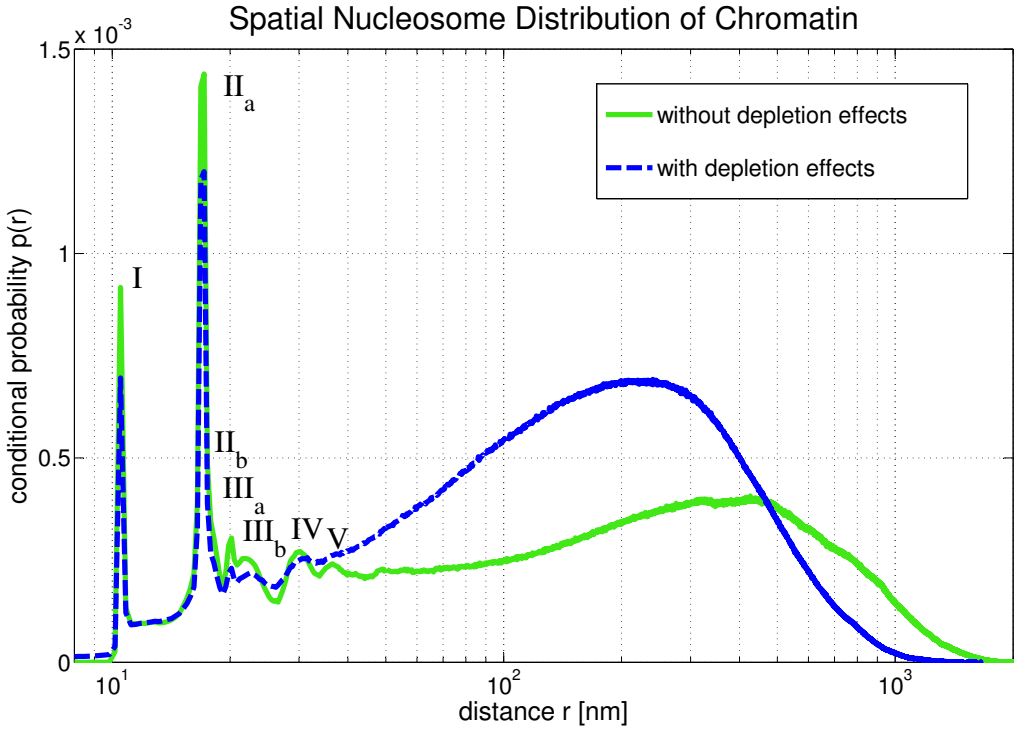


Figure 92: The conditional probability $p(r)$ on the large scale. One can see that the distribution for fibers with histone depletion is much shorter due to the gain of flexibility that the linker histone skips and the nucleosome skips bring in. The short scale is dominated by the peaks that come from the local nucleosome order, whereas on the large scale many r_Δ-distributions superimpose, and therefore, no peak structure is visible any more.

Moreover, the single distributions of the first ten r_Δ are plotted in Fig. 189 and Fig. 190 in App. A.

The conditional probability $p(r)$ is the superimposition of all r_Δ-distributions. This is illustrated in Fig. 93 for the flexible fibers and in Fig. 94 for the flexible fibers with depletion effects. With the help of these two figures one can determine which peak corresponds to which r_Δ-distribution (cf. Tab. 6).

With increasing Δ the distributions of r_Δ become broader and finally the superimposition of these distributions, i.e. $p(r)$ does no longer show a peaks structure which can also be seen in Fig. 92.

	stiff fibers	flexible fibers			stiff fibers with depletion effects		
Δ	r_Δ	$\langle r_\Delta \rangle$	σ_Δ	mf_Δ	$\langle r_\Delta \rangle$	σ_Δ	mf_Δ
1.00	17.00	17.00	0.21	16.85	21.20	10.24	45.51
2.00	11.63	12.52	2.39	10.44	21.81	18.55	10.40
3.00	17.35	17.11	1.33	17.46	28.51	20.76	18.47
4.00	21.81	22.71	2.39	23.34	35.22	22.13	19.69
5.00	20.96	21.42	2.76	22.10	38.12	25.17	2.53
6.00	29.77	30.01	2.03	29.98	45.04	24.98	28.14
7.00	28.71	30.06	3.71	27.59	48.43	26.96	30.76
8.00	35.95	35.60	2.65	19.61	53.32	27.51	1.11
9.00	38.31	39.62	3.63	37.57	57.59	28.35	2.79
10.00	41.82	41.69	3.85	42.73	61.16	29.54	40.67

Table 7: This table shows the mean value ($\langle r_\Delta \rangle$), the standard deviation (σ_Δ) and the most frequent value (mf_Δ) of the first ten r_Δ-distributions. In the first case a stiff reference fiber was generated. The following three columns show the values for flexible fibers and the last three columns show the results for flexible fibers with depletion effects. Here, the nucleosome skip rate was again fixed to 8% and the linker histone skip rate was fixed to 6%. The table shows that some of the r_Δ-distribution are very asymmetric since the mean value differs greatly from the most frequent value (e.g. the distribution of r_2). All spatial distances are given in nm (a similar table can be found in App. B with the values in bp). The genomic distances are given in multiples of NRL again.

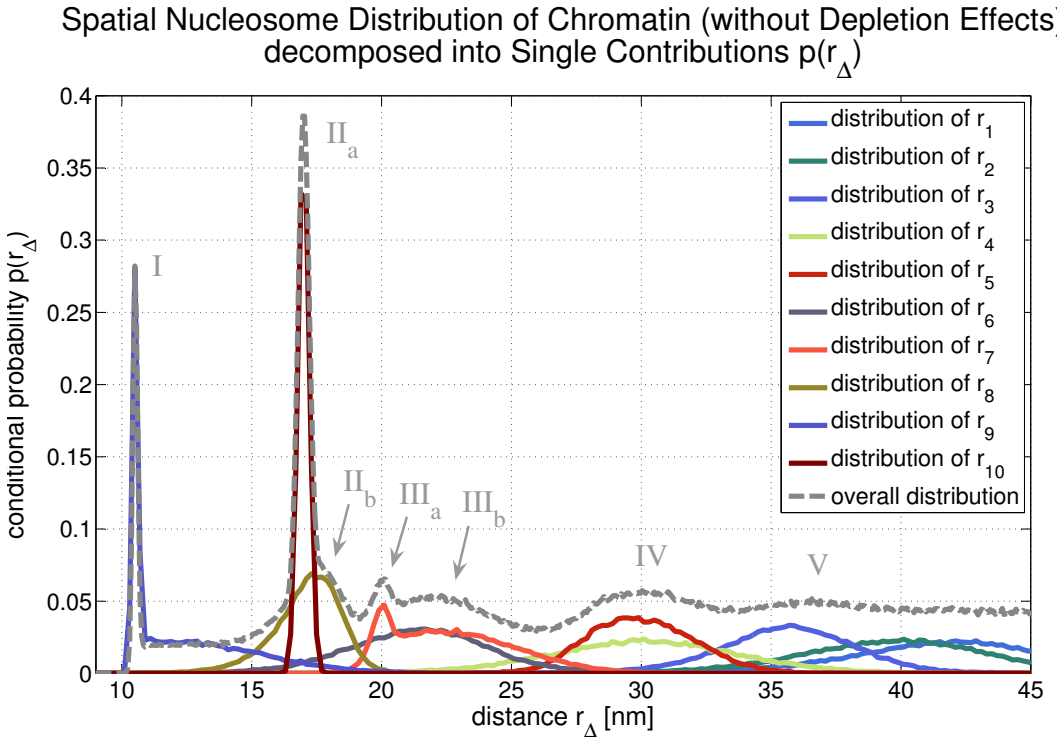


Figure 93: The overall probability distribution $p(r)$ is a superimposition of the single r_Δ-distributions which is illustrated here for the case of flexible chromatin fibers. One can see that some distributions are very asymmetric (e.g. r_1 and r_7).

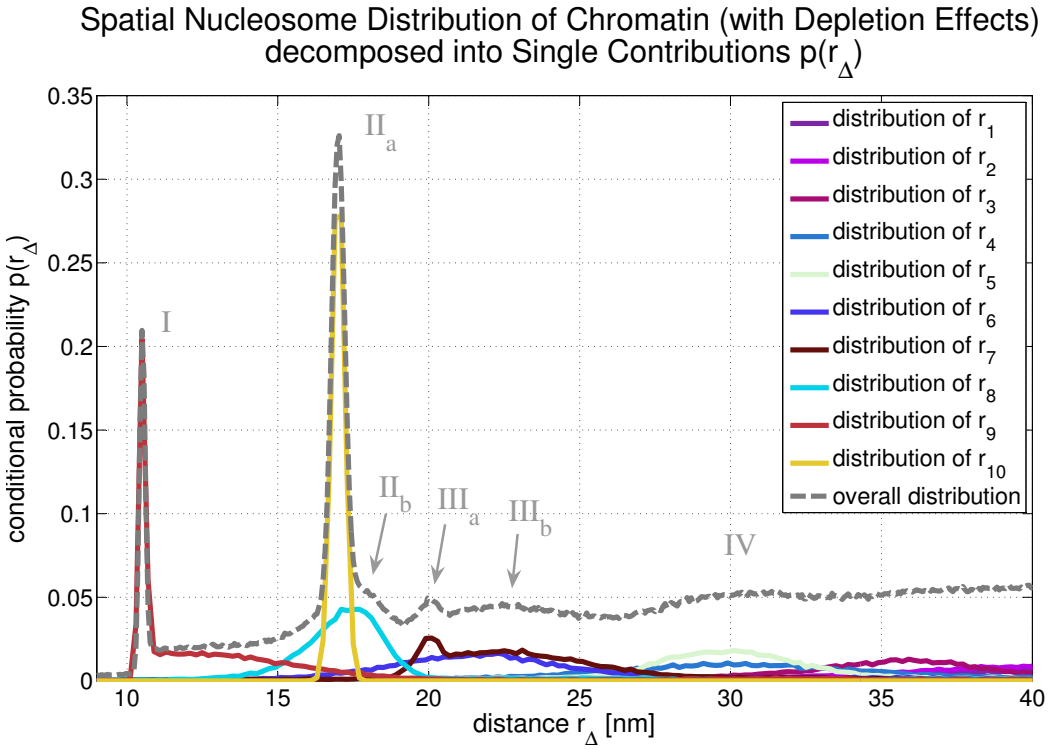


Figure 94: This figure shows how the different r_Δ-distributions contribute to the overall distribution $p(r)$ in the case of flexible chromatin fibers with histone depletion.

7.3.2 The Nucleosome Pair Distribution Function

Fig. 96 shows the (radial) pair distribution function $g(r)$ [207] of nucleosomes in a chromatin strand. The pair distribution function is a major descriptor for the atomic structure of solids, amorphic materials and liquids and was described partially in Sec. 1.8.

In this case one can apply this mathematical tool only for small distances because one does not have a chromatin melt but instead only a *single* fiber at a time. Therefore, the distance cut-off for the following structure analysis was set to a small value, namely 40nm. That means that spheres with this radius around each nucleosome are analyzed by looking for very frequent spatial distances (cf. Fig. 95 for an illustration).

In this context it is important to keep in mind that the nucleosomes sit at the edge of the chromatin fiber, and furthermore, the fiber itself has only a diameter of about 35nm. Therefore, the main part of the 40nm-sphere is empty which leads to a decrease of the mean nucleosome density (cf. Fig. 96).

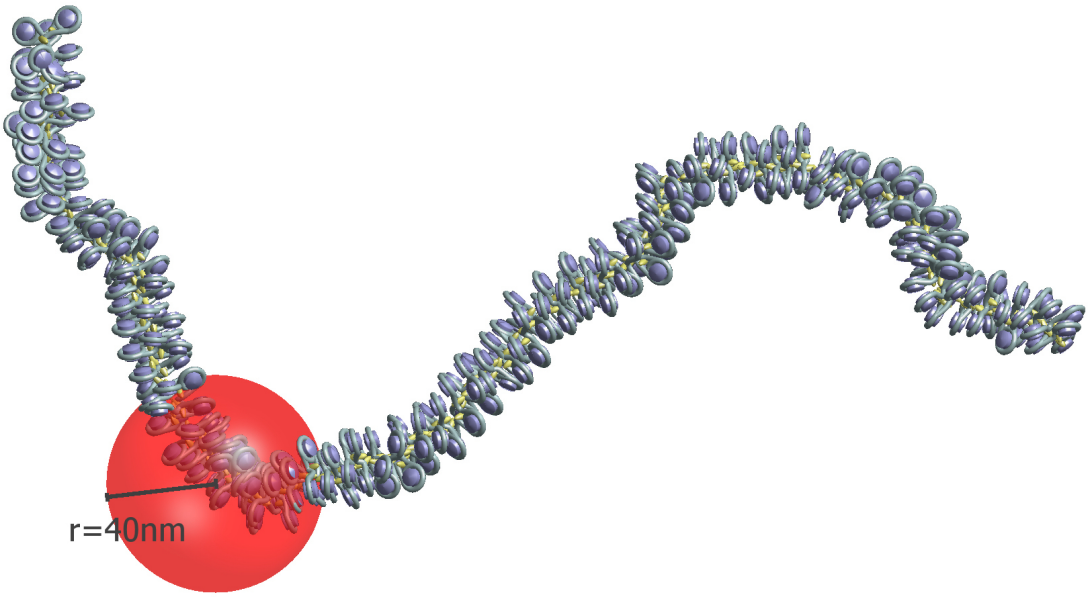


Figure 95: Illustration of the system size V for the structure analysis with the pair distribution function: The red sphere has a radius of 40nm and covers only a part of the chromatin fiber.

The pair distribution function will be proportional to the conditional probability $p(r)$ of finding a nucleosome at a distance r, if another nucleosome sits at the origin. It is normalized in a way that a value of one corresponds to the mean nucleosome density of the considered system (i.e. in this case of a 40nm sphere):

$$\begin{aligned} g(r) &= \frac{1}{4\pi r^2}\frac{V}{N-1}\left(\frac{1}{N}\sum_i\sum_{j\neq i}\delta(r-r_{i,j})\right) \\ &= \frac{1}{4\pi r^2}\frac{V}{N-1}\left(\frac{N-1}{2}p(r)\right). \end{aligned}$$

The pair *correlation* function is given by $g(\mathbf{r})-1$ and the Fourier transform of it is the scattering function $S(\mathbf{q})$ which could theoretically be determined by scattering experiments:

$$S(\mathbf{q}) = \int_V e^{-i\mathbf{q}\,\mathbf{r}}(g(\mathbf{r})-1)d^3\mathbf{r}.$$

In the case of an isotropic system (like the one that is considered here) with a radial pair distribution function the scattering function is also isotropic ($S(\mathbf{q}) = S(q)$) and one gets

$$S(q) = \int_{r=0}^{\infty}\frac{\sin(qr)}{qr}4\pi r^2(g(r)-1)dr.$$

$g(r)$ shows some very dominant nucleosome-nucleosome distances which are labelled I - V again (cf. Fig. 96). These peaks that represent frequent spatial distances r_Δ can be associated with the same genomic distances as in the case of $p(r)$ (cf. Tab. 6) and again $g(r)$ is the (normalized) superimposition of all r_Δ-distributions.
The first four peaks of the pair distribution function can even be identified, if one allows for depletion effects (cf. Fig. 96), although in this case the pair distribution function is again decreased in comparison to $g(r)$ without depletion effects because the skips reduce the number of nucleosomes that have a certain genomic distance.

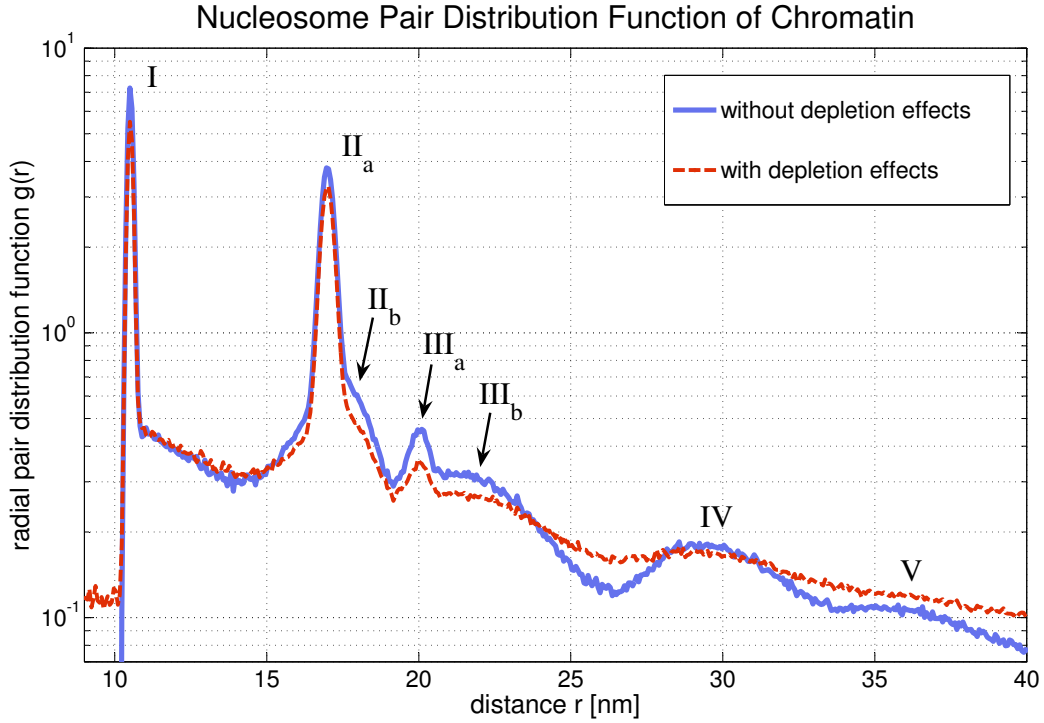


Figure 96: The pair distribution function of nucleosomes in a chromatin fiber (with and without depletion effects). $g(r)$ is proportional to the probability of finding a nucleosome in distance r to a reference nucleosome. The labelled peaks can be associated with particular genomic nucleosome-nucleosome distances Δ (cf. Tab. 6). Even in the case of fibers with depletion effects (i.e. a linker histone skip rate of 6% and a nucleosome skip rate of 8%) the first four peaks can clearly be identified. The large δ-peak at $r = 0$ has been eliminated here.

7.3.3 Pair Distribution Function of 2D Chromatin Fibers

In Sec. 3.5.3 the two-dimensional radial pair distribution function of HeLa and Fibroblast nuclei was examined. In this section this experimental data will be compared with results from the E2A model.

Fig. 97 shows the conditional probability $p(r)$ (cf. Sec. 7.3.1) of previously projected i.e. two-dimensional chromatin fibers. Again, the two cases of chromatin fibers with and without histone depletion are considered.
Interestingly, there are distances below 10nm now. These distances do not occur in the 3D chromatin structure and are an artefact due to the projection of the fibers: Some formerly larger distances have been foreshortened by the projection. A comparison with Fig. 89 shows that all peaks are smeared out towards smaller distances by the projections. This is clear, since projecting vectors can only shorten them but never increases their length.
Nevertheless, the first four peaks in the distance distribution function can still be identified in the case of fibers without defects. In the other case of fibers with histone depletion only the first two peaks can yet be clearly identified.

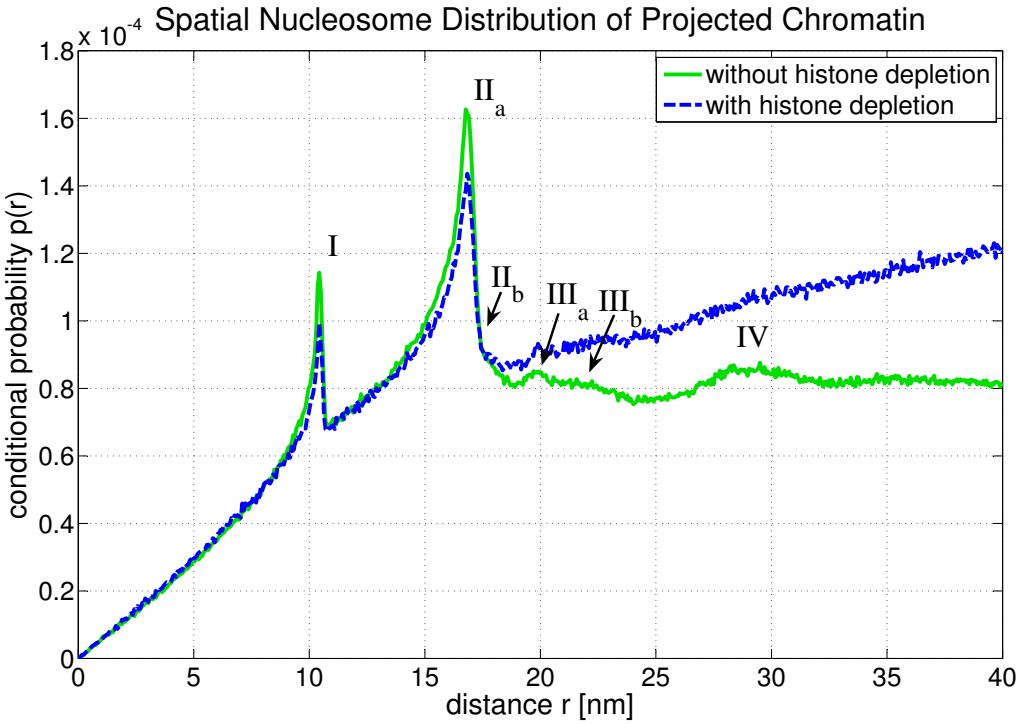


Figure 97: The radial nucleosome distance distribution in the case of 2D chromatin fibers. Some of the peaks can still be identified in the case of fibers without depletion effects. Furthermore, $p(r)$ does not vanish any more for distances smaller than 10nm. This is an artifact of the projections.

Fig. 98 shows the two-dimensional radial pair distribution function that corresponds to $p(r)$ in Fig. 97. Moreover, the Fibroblast and HeLa data from Fig. 33 in Sec. 3.5.3 are shown as well.
For the calculation of the two-dimensional pair distribution function a 2D-sphere with

radius 40nm was used again as the limiting system size (cf. Sec. 7.3.2). The connection of the conditional probability $p(r)$ and the two-dimensional pair distribution function is in this case given by:

$$\begin{aligned} g_{2D}(r) &= \frac{1}{A(r)} \frac{A}{N-1} \left(\frac{1}{N} \sum_i \sum_{j \neq i} \delta(r - r_{i,j}) \right) \\ &= \frac{1}{2\pi r \Delta r} \frac{\pi R^2}{N-1} \left(\frac{N-1}{2} p(r) \right) \\ &= \left(\frac{(40\text{nm})^2}{4\Delta r} \right) \frac{p(r)}{r}. \end{aligned}$$

Δr is the binning parameter and was set to 0.2 bp during the calculations of $g_{2D}(r)$ and R denotes the 40nm radius that limits the system size.

7.3.4 Comparison with SPDM Data

Fig. 98 shows the two-dimensional radial pair distribution function compared with the Fibroblast and HeLa nuclei data from Sec. 3.

One can identify only the first two peaks in the case of chromatin fibers with histone depletion. The fibers without histone depletion still have the first four peaks in the distance distribution.
One has to be very careful with the comparison of the model data with the SPDM data since the model data comes from a single chromatin fiber of finite size. It was tried to compensate this by choosing the small system size of a 40nm-sphere so that $g(r)$ is at least matched correctly at the small scale below $\approx$ .30nm. For a comprehensive examination of $g(r)$ on larger scales it would be necessary to simulate several chromatin fibers or chromatin melts.
It could be possible that the Fibroblast and HeLa data points below 10nm are artefacts of the projection as well as in the case of the model data. Furthermore, one has to keep in mind that their localization accuracy in x- and y-direction is fairly large (20nm).

7.4 Comparison with Experimental Distance Measurements

In Fig. 99 one can see the average spatial nucleosome distance in dependence on the genomic distance for the three different cases of chromatin fibers on a larger scale (cf. Fig. 91 for the small scale). Furthermore, FISH measurements from [34] are shown, too.

In [34] high-resolution fluorescence in situ hybridization data from intervals of 14-100 kb along single chromatids were reconciled with measurements of whole chromosome arms (122–623 kbp in length), monitored in intact cells through the targeted binding of bacterial repressors fused to GFP derivatives. The results are interpreted with a flexible polymer model and suggest that interphase chromatin exists in a compact

higher-order conformation with a persistence length of 170–220 nm and a mass density of ≈110–150 bp per nm. These values are equivalent to 7–10 nucleosomes per 11nm within a 30nm-like fiber structure (cf. Sec. 6.5).

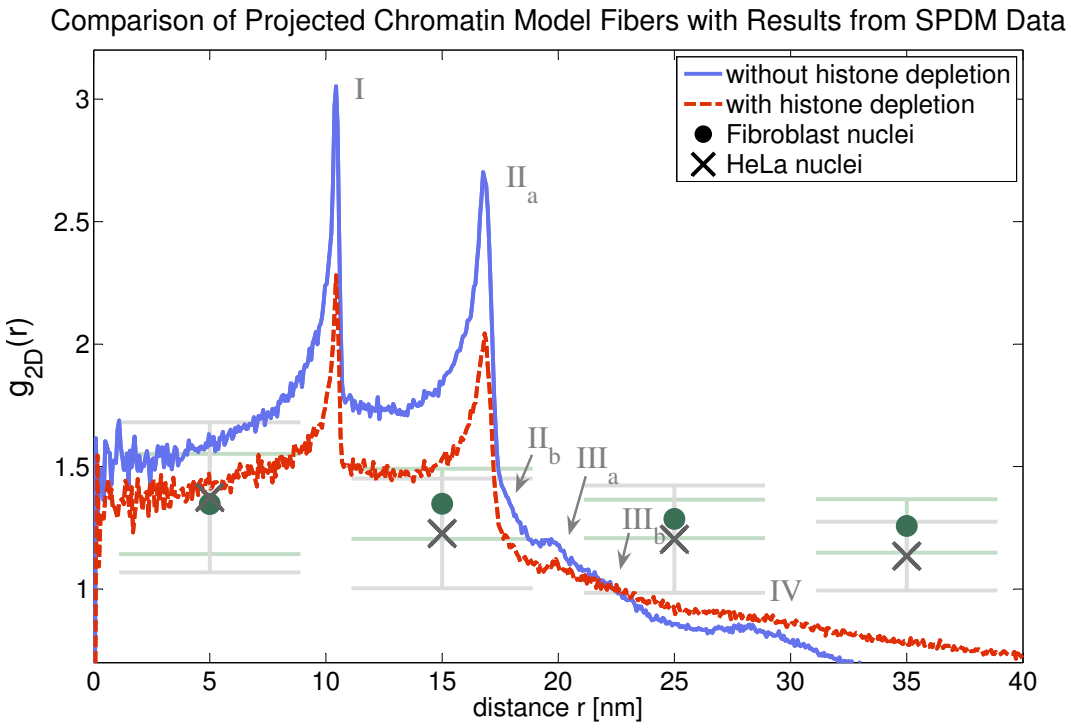


Figure 98: Comparison of the SPDM data from Sec. 3.5.3 with the distance distribution of projected E2A model chromatin fibers. The system size had to be limited to 40nm since only single model fibers can be evaluated and thus the application of $g(r)$ is problematic for larger length scales.

To analyze chromatin compaction ratios in interphase nuclei, laboratories have generally applied fluorescence in situ hybridization (FISH) using differentially derivatized probes. Gasser et al determined [34], end-to-end distances for a range of genomic intervals by using unique techniques for high-resolution FISH [94; 103] and live GFP-fusion imaging based on repressor binding to chromosomally integrated, non-amplified *lac* or *tet* operator arrays [205].
This process combined with immunofluorescence (IF) allowed them to examine chromatin folding over small distances in intact yeast cells.

They also compared arm length measurements with distances separating different repressor array insertions. The polymer chain model (WLC) then allowed them to determine both the persistence length and mass density of chromatin from these end-to-end distance values.

As one can see in Fig. 99 the FISH-data lies even beneath the curve of the flexible fibers with histone depletion. This might be due to the confinement in the nucleus that has not been considered in the simulations.

Another very important effect has still been neglected: Flory showed [83] that polymers in melts behave as if their monomers have zero volume. That means that the actual distance graph for chromatin *melts* would definitely lie beneath the curves presented in Fig. 99.
In that sense, the graphs of Fig. 99 provide only an upper bound to the actual behavior of chromatin in the cell nucleus.

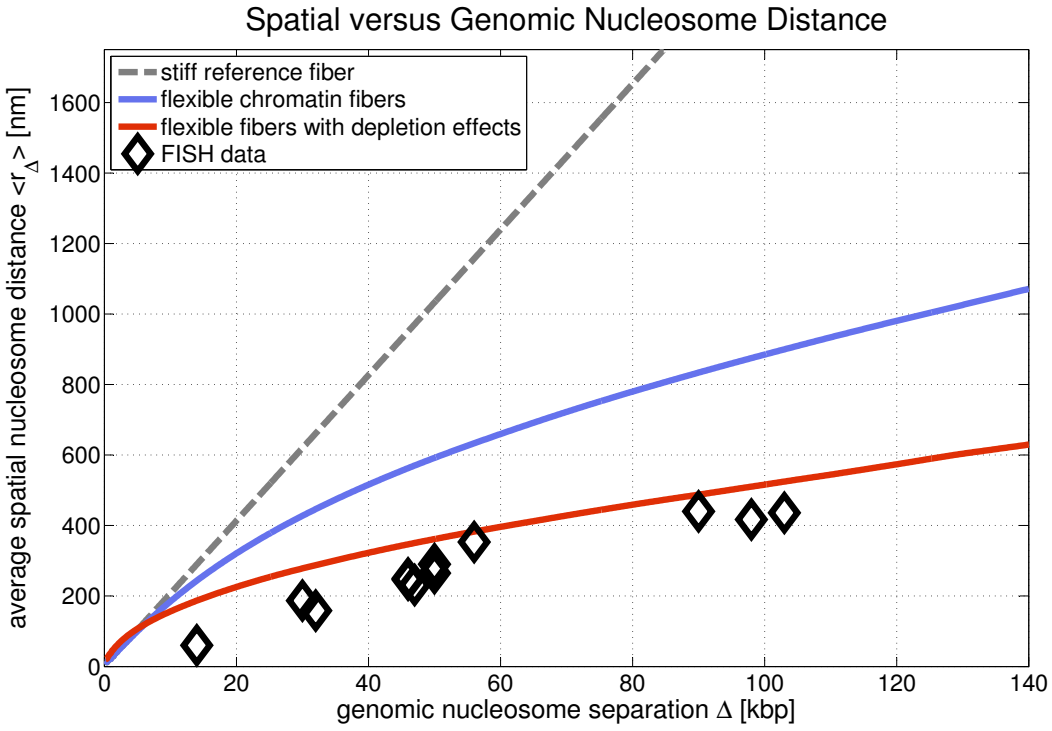


Figure 99: Shown is the dependence of the spatial nucleosome distance on the corresponding genomic distance for the different types of chromatin fibers (stiff reference fiber, flexible fibers and flexible fibers with depletion effects). Furthermore, one can see a comparison with two-color FISH measurements of interphase chromatin in budding yeast cells [34].

Furthermore, one has to keep in mind that only the *average* values for the histone depletion rates have been used. If one applies larger histone skip rates, one will get a further decrease of the spatial distances.
Moreover, FISH measurements on this small length scale unfortunately change the investigated system and therefore, have to be interpreted very carefully.

Nevertheless, it seems that one has to allow for histone depletion effects in order to get suitable predictions for spatial chromatin distributions. The results for chromatin fibers without any defects are much too stiff and hence the curve increases too rapidly.

8 Chromatin Loops

8.1 Overview

In this section the focus will lie on chromatin conformations with loops i.e. with direct physical DNA contacts over large genomic distances.

First, Sec. 8.2 will provide some basic motivation for the examination of chromatin loops by explaining which role they play for transcription and other biological processes that are crucial for several cell functions (cf. Sec. 8.2.1). Furthermore, the role of random chromatin contacts in 3C-based experiments will be discussed (cf. Sec. 8.2.2 and Sec. 8.2.3) and it will be explained how a better understanding of these random collision would improve 3C-based technologies by allowing to distinguish more accurately between specific and random DNA contacts.

In Sec. 8.3 the simulation results concerning chromatin loops are presented. It will be shown that fibers with histone depletion and fibers without histone depletion have very different loop statistics. Especially the loop number distribution and the loop length distribution are qualitatively as well as quantitatively very diverse.

Experimental 3C-based methods are very important for the investigation of DNA interactions and chromatin loops. All the current 3C-based methods are described comprehensively in [53; 192]. But a brief introductory overview will be given in Sec. 8.4.1 as well.

After that, Sec. 8.4 will provide a comparison between the simulation results and actual experimental data from 5C experiments.

It will turn out that only the chromatin fibers with histone depletion match the experimental data, because only they have physical interactions on a small genomic length scale. Furthermore, only chromatin with histone depletion shows qualitatively the same behavior of the interaction probability as the experimental data.

In Sec. 8.5 the loop entropy of chromatin loops will be investigated. It will be proven that in the case of fibers with histone depletion it is favorable for a chromatin conformation to have some loops. A further increase or decrease of the loop number leads to a decreases of the loop entropy. This is remarkable because chromatin loop formations and thus DNA contacts are important for biological processes as mentioned before. Most of the fibers without histone depletion have no loop at all and in this case every loop formation leads to a loss of entropy.

The results of this section are in preparation for publication [67].

8.2 Introduction

Loop conformations play a very important role in chromatin physics: A large number of molecules permanently access the DNA and thus change the chromatin structure. As experiments have shown this leads to permanent building and dissolving of loops mediated by transcription factories [25; 42; 109; 132].
The genomic sequence for different proteins which have to work together, and therefore find each other, has to be somehow spatially clustered for the time the genetic code is being transcribed, while some time later in cell evolution another clustering of genes might be necessary for the production of other proteins [126].
The fact that transcriptional regulatory DNA elements loop towards their target genes to regulate the expression has been revealed by the 3C-based technologies (cf. Sec. 8.4) which contributed enormously to the understanding of the intricate folding of gene loci [192].

8.2.1 Loop Formations during Transcription

The transcription is a complex mechanism [124]. For the initiation of the transcription specific proteins have to assemble at the promoter, that is the DNA region which identifies the beginning of a gene. The RNA polymerase II (cf. Fig. 100) depends on other proteins, the general transcription factors, to be enabled to transcribe the gene's DNA sequence into mRNA. One interesting problem that is not yet finally solved is how the DNA packing is negotiated by the transcription complex.
In fact, in most of the eukaryotes only a small fraction of the whole genome is dedicated to encode for protein production, for example approximately two percent in human cells. Some of the remaining DNA contains binding sites for regulatory proteins that determine, for instance, how often a referring gene is transcribed and thus affect the synthesis rate of a specific protein.

Binding sequences are quite frequently situated about 100 to 200 bp upstream from the promoter and, hence, denoted as upstream elements. But in particular in eucaryotic cells, regulatory proteins can bind to referring sequences, for instance enhancers, thousands of base-pairs away from the transcription start and still influence the transcription rate. These long-range interactions between regulatory proteins and the transcription complex are facilitated by bending of the intervening chromatin.

Competition or synergy between proteins, regulating the transcription of the same gene, constitute the basis for a complex gene-regulation network. The efficiency of a regulatory protein depends on its global concentration in the cell nucleus, but also on its local concentration in proximity of the promoter according to the probability of intervening chromatin loops. So, the positioning of a regulatory sequence with respect to the promoter is an important factor determining the impact of the referring enhancer or repressor protein. If the intervening chromatin segment is too short, bending might be impossible. But if it is too long, polymer dynamic effects causing chromatin bending

will become irrelevant.

Another example for loop formations in chromatin is chromatin looping in centromeres [124].

8.2.2 Chromatin Interaction Mapping

Intense efforts are made to map genes and regulatory elements throughout the whole human genome [72; 210]. These studies are expected to identify many different types of elements, including those involved in gene regulation and DNA replication. Analysis of only one percent of the human genome has already revealed that genes are surrounded by a large number of supposed regulatory elements [210].

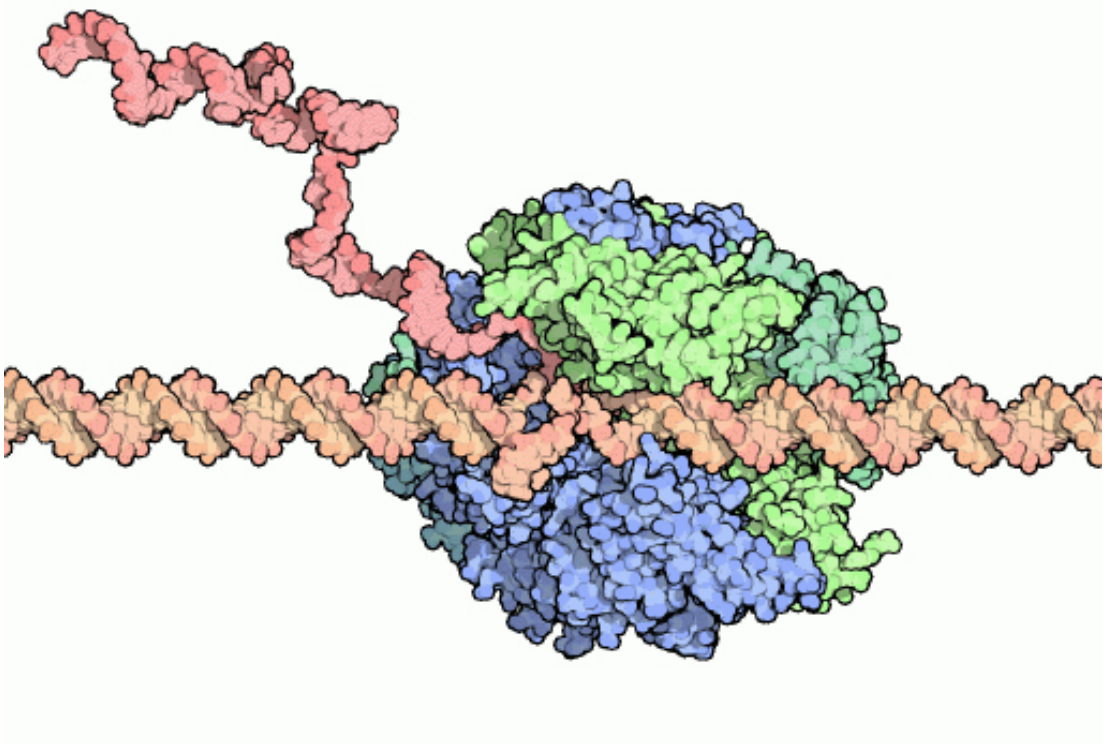


Figure 100: An illustration of RNA polymerase II, an enzyme that produces RNA chains from DNA genes. Adapted from [231].

In order to fully annotate the human genome and to understand its regulation it is important to map all genes and functional elements and also to determine all relationships between them. This endeavor is complicated by the fact that the genomic positions of genes and elements do not provide direct information about functional relationships between them. A well known example is provided by enhancers that can regulate multiple target genes that are located at large genomic distances or even on different chromosomes without affecting genes immediately next to them [197; 225].

Recent evidence indicates that regulatory elements can act over large genomic distances by engaging in direct physical interactions, i.e. loop formation with their target genes or other elements [39; 49; 52; 225]. These observations indicate that the genome may be organized as a complex three-dimensional network that is determined by physical interactions between genes and elements. Therefore, mapping of chromatin interactions by 3C-based technologies is very important.

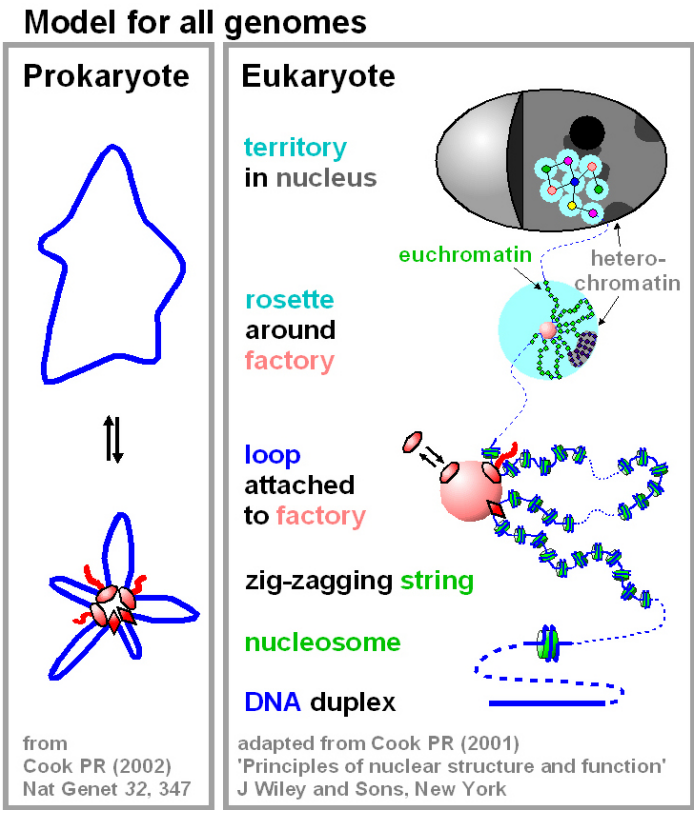


Figure 101: Illustration of loop formations. Left-hand side: In bacteria, transcription of the circular chromosome, followed by clustering of polymerases (ovals), transcriptional activators/repressors (diamonds), and transcripts (wavy red lines), generates loops. Only one of a number of rosettes is shown, and the path of the DNA is expected to be complicated. Right-hand side: In eukaryotes , DNA is coiled around the histone octamer, and runs of nucleosomes form a zig-zagging string. At the intermediate level in the hierarchy, this string is organized into loops (average contour length 86 kbp; range 5-200 kbp) by attachment to transcriptional activators/repressors (diamonds) and engaged RNA polymerases (ovals) bound to a factory. Adapted from [41]

8.2.3 The Role of Random Chromatin Interactions

As discussed in Sec. 8.2.2 it is very important to distinguish between non-functional physical contacts by intrinsic, random collisions and specific, functional contacts. The

term "non-functional" is very common but maybe misleading: As pointed out in Sec. 8.2.1 most of the promoters sit only 100 to 200 bp away from a corresponding binding site. Since this is the length scale where random collisions dominate physical chromatin interactions might be very important for transcription purposes, too.

For small distances the contacts due to random collisions prevail due to the intrinsic close proximity of neighboring fragments [51; 72]. The frequent random interactions between adjacent genomic fragments are likely dependent on local physical properties of the chromatin fiber and limit the ability to detect specific looping interactions between elements separated by small genomic distances [72; 92]. Hence, in 3C-based experiments often a gene desert region is measured as a control sequence to determine the influence of random contacts on the experimental results [72].

The E2A model does not allow for specific chromatin interactions which supplies the advantage that all physical chromatin interactions in this model are random collisions. Therefore, one can determine and examine the amount of non-specific contacts in order to improve 3C-based experiments for example.
In order to model the specific chromatin interactions on the large length scale one needs much more coarse-grained approaches like the concept of random loops [25]. In the latter case the biological environment of the cell nucleus that affects the chromatin fiber is treated as a feature of the modelled polymer itself.
As discussed above the chromatin flexibility plays a very important role not only for its structural properties but also for its *functional* features. The chromatin flexibility has already been discussed in Sec. 7.2. Therefore, chromatin loop statistics will be the focus of the next section (i.e. Sec. 8.3).

8.3 Chromatin Loop Statistics

8.3.1 Modelling

The chromatin loops that are examined in this section come from mere *random* chromatin collisions in opposite to the situation in the real cell nucleus where specific DNA interactions induce loop formations, too. This is a problem in 3C-based experiments where one is explicitly looking for the specific interactions (cf. Sec. 8.4) and thus wants to avoid the random chromatin contacts.
It will turn out that flexible chromatin fibers without depletion effects and fibers with histone depletion have very different loop statistics.

Since there are no other molecules than chromatin involved in the E2A model one has to define what a loop is in the first place. For the considerations of this whole chapter a part of a chromatin fiber is called a *loop*, when its end points come closer to each other than 105 bp (i.e. 35nm). Other studies use similar values [124].

Furthermore, there was a coarse-graining for the fibers without histone depletion to save computation time: Six nucleosomes were pooled to one coarse-grained chain segment in the loop analysis. The fibers with histone depletion were not coarse-grained.

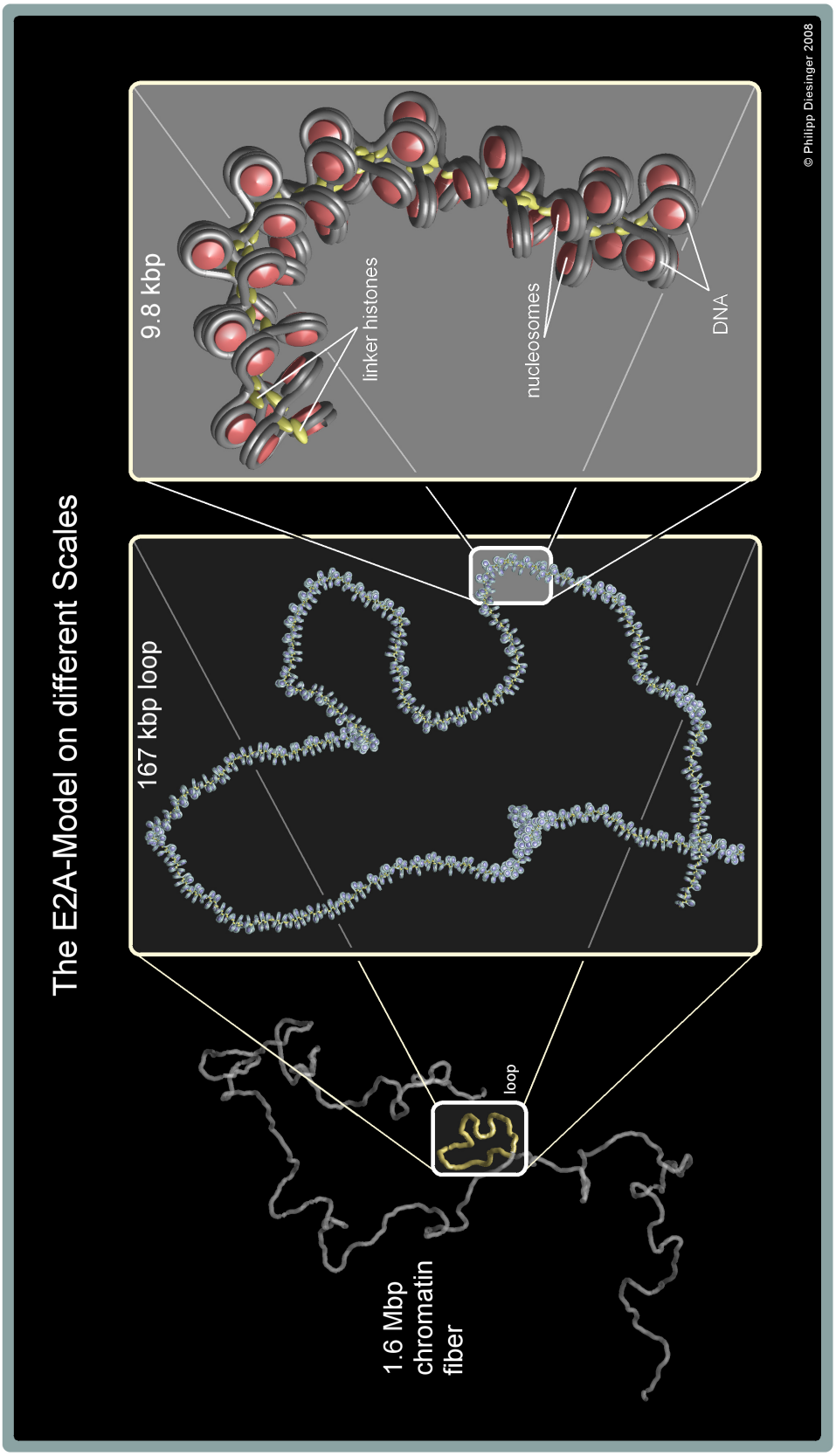


Figure 102: This figure shows a 1.6Mbp long chromatin fiber conformation (without histone depletion) with a large loop of length 167kbp (middle panel).

If two nucleosomes come closer than the spatial cut-off length of 105 bp they will have to be more than five nucleosome repeat lengths away from each other along the chromatin fiber. This additional condition makes sure that only real loops count. This leads to a second, but this time genomic, cut-off length of 5.91kbp. The cut-off lengths are equal in the two cases of chromatin fibers.

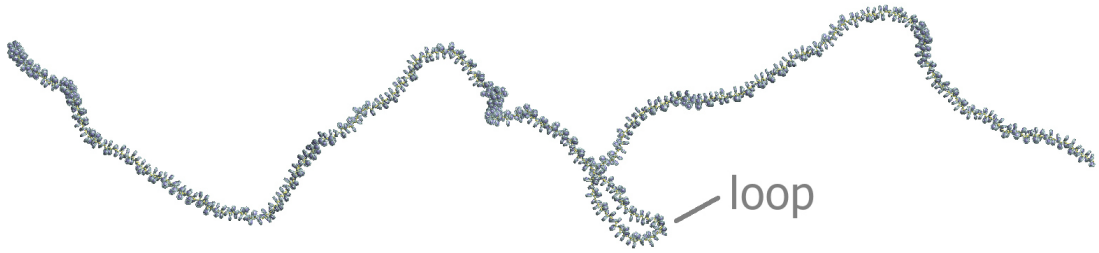


Figure 103: Another example of a chromatin conformation with a small loop. The chain has a total length of 160kbp. The loop length is ≈10kbp.

The fibers without depletion effects have lengths of 1.6Mbp, 800kbp and 160kbp (i.e. 8000NRLs, 4000NRLs resp. 800NRLs) and the fibers with depletion effects have contour lengths of 160kbp and 80kbp. In the case of fibers with depletion effects smaller fiber lengths are already sufficient to resolve the loop distributions as will be shown below. Furthermore, they are harder to sample and in the following analysis sometimes conditional probabilities are needed. Unfortunately, these require huge sample sizes (e.g. to evaluate the conditional probability distribution of the loop end position over the set of all fibers with a loop that starts a given position along the fiber). It will be shown that the total length of a fiber does not much change the loop statistics as long as it is above the persistence length.
For every fiber type the sample consists of at least $3 * 10^6$ chromatin conformations.

The concept of the persistence length has to be applied very carefully to the fibers with H1 and nucleosome depletion: Although the persistence length is still of the order of hundred nm, fiber parts can come quite close together due to the nucleosome depletion, although they are only a few NRLs apart along the fiber axis, i.e. they are still in a region where one would actually expect that the fiber was too stiff to bend.
Moreover, chromatin fibers in the cell nucleus which are larger than some hundred kbp are not unconstrained any more and start to feel the presence of the biological environment. The large chains without depletion effects have been sampled to exclude that the differences of the loop distributions which will be discussed in the following will disappear, if one changes the fiber length.
For the fibers with histone depletion the average skip rates of 8% (nucleosome skips) and 6% (H1 skips) are applied again.

Fig. 102 shows a 1.6Mbp chromatin conformation with a large 167kbp long loop.
Most of the given probabilities in this chapter are under the condition that the chromatin fiber has at least one loop. These probabilities are denoted as *conditional* prob-

abilities in the following. All unlooped fibers are left out in these cases.

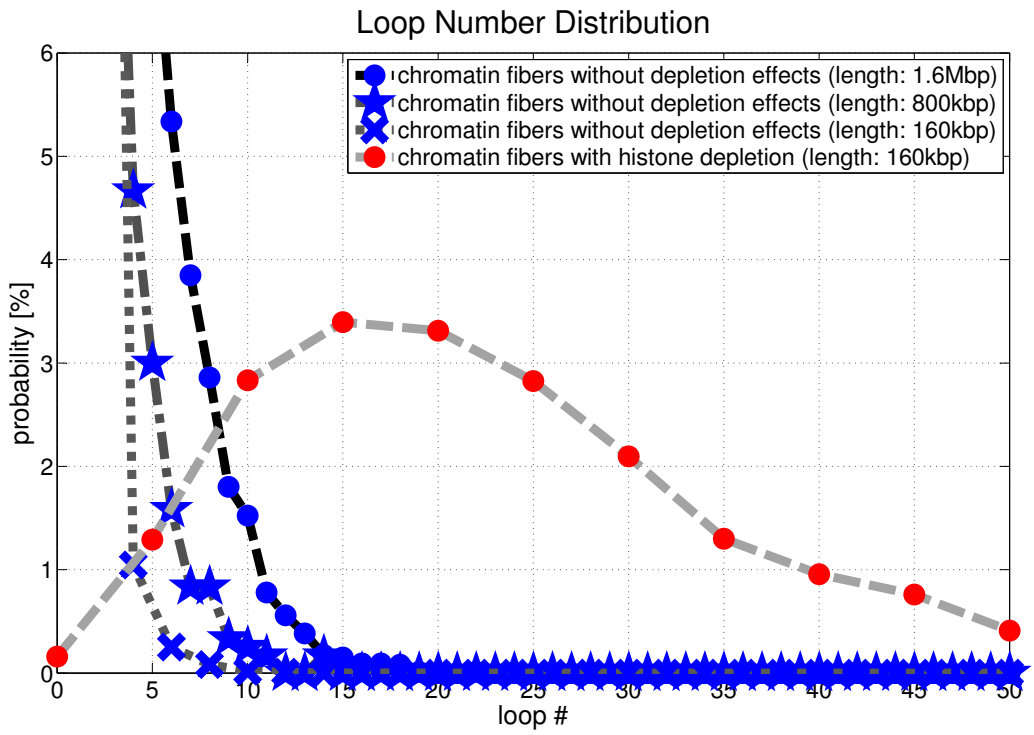


Figure 104: The distribution of the loop number for the two cases of chromatin fibers with histone depletion and fibers without depletion effects. The average loop number is 2.96 in the first case and 25.5 in the second case.

8.3.2 Results

One can see (cf. Fig. 104) that chromatin fibers with and without histone depletion have completely different loop number distributions.
First of all, the chance that an arbitrary fiber without histone depletion of length 160kbp does not have a single loop is almost 50%, whereas the chance of finding a fiber of the same length without a loop but *with* histone depletion is almost zero. The average loop numbers for these two types of fibers are 0.31 for fibers without histone depletion and 25.5 for fibers with histone depletion.
Furthermore, the shapes of the loop number distributions are very different, too. The fibers with depletion effects show the behavior one would expect from a self-avoiding random walk (cf. Fig. 192 in App. A). In the case of chromatin strands without histone depletion only the tail of the distribution is visible, i.e. they are (even with huge fiber length) too stiff with respect to their persistence length.

The distribution of the chromatin loop length is illustrated in Fig. 105. Shown are four

cases: Fibers with and without histone depletion and with different contour lengths. One can see that the fibers without histone depletion have a completely different loop length distribution than the fibers that allow for depletion effects.

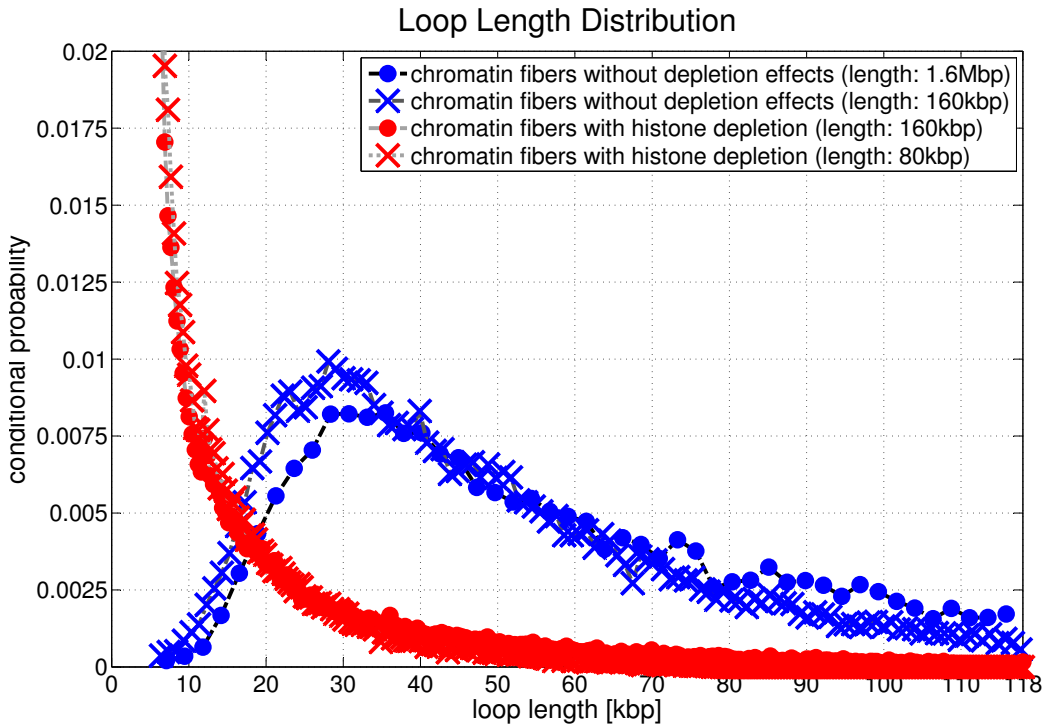


Figure 105: The loop length distributions of chromatin fibers with and without depletion effects for different fiber lengths. In the case of fibers without depletion effects small loops are very unlikely, whereas in the case of fibers that allow for histone depletion these very loops are dominating. The differences of distributions with different fiber length come from the normalization.

The distribution for fibers without depletion effects resembles the loop distribution of an ordinary (self-avoiding) random walk (cf. Fig. 193). Small loops are unlikely because the fiber is too stiff to bend on the short length scales. The persistence length of 280nm corresponds to a fiber contour length of about ≈13.5 kbp. One can see that this is approximately the region where the loop probability starts to grow. Large loops are unlikely, too, since the random walk is transient in three dimensions.
The loop length distribution of a random walk follows the scaling law

$$P(l) \propto l^{-c},$$

where $c = d/2$ for a random walk and in the case of a self-avoiding walk $c = d\nu - \sigma_4$ i.e. 2.68 in the 2D case and 2.22 in the 3D case [215].
The loop distribution of the chromatin fibers with depletion effects looks completely

different: Small loops are very likely in this case. The depletion effects give the fiber a lot more flexibility. Especially the nucleosome depletion allows the fibers to bend even on small length scales.

The fiber length does not have a recognizable effect on the loop distribution on this small length scale. The difference of the 160kbp-fiber without depletion effects and the corresponding 10-fold larger fiber comes from the normalization since the fiber length limits the maximal loop length.
Furthermore, one has to keep in mind that there is a cut-off for the minimal loop length at about 5.9kbp.
Another plot of the loop length distribution for more different fiber sizes is supplied by Fig. 105 in App. A.

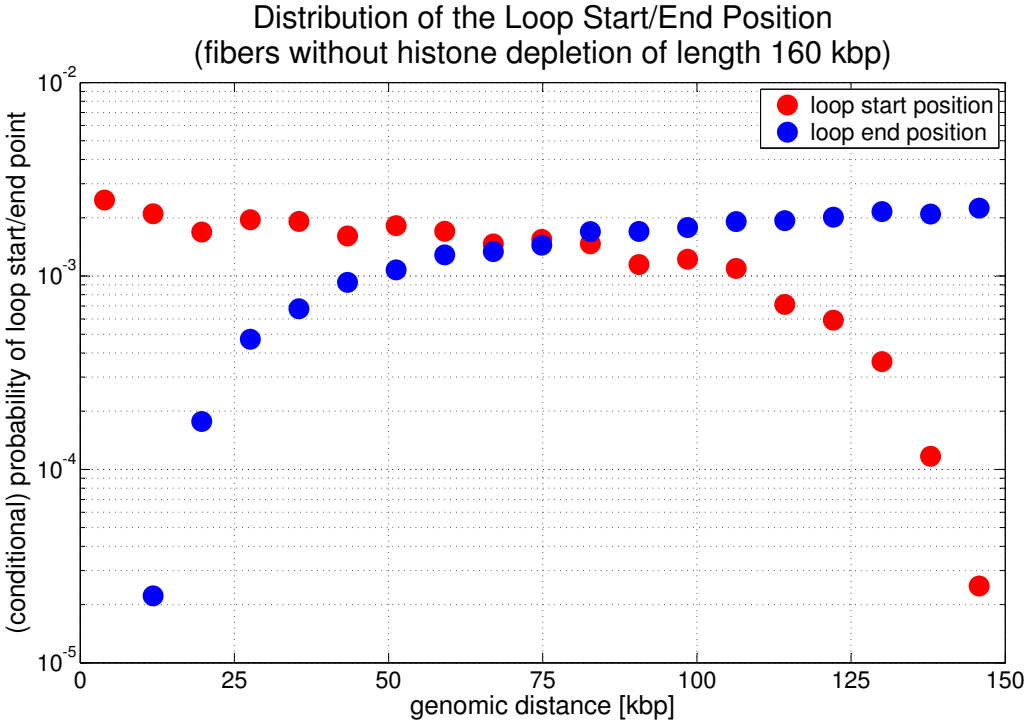


Figure 106: The distribution of the loop start respectively end position for fibers without histone depletion and a length of 160kbp.

Fig. 106 shows the distribution of the loop start and end positions for chromatin fibers without nucleosome depletion and a length of 160kbp. A similar figure for the case of a random walk can be found in App. A (cf. Fig. 195).
The loop start probability is largest at the beginning of the fiber since there are simply more possibly end points for the loop. The two distributions are mirror images of each other since there is no difference between the two fiber ends.

If the loops were independent of each other and one did not have any finite size effects, the loop start/end position distribution would simply be uniform. That means the two random variables S (loop start position) and E (loop end position) would have uniform distributions: $f_S = f_E = 1/l$. Then the loop length $L = |S - E|$ has the probability distribution $f_L = f_{\text{abs}} \circ f_{S-E}$.

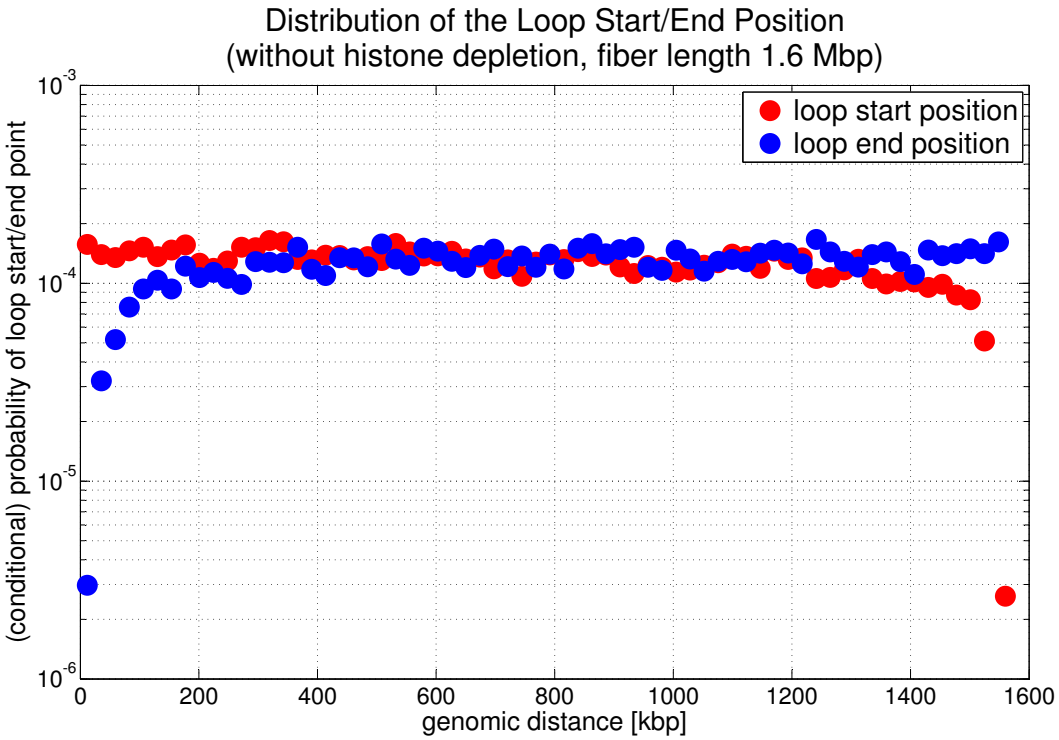


Figure 107: This figure shows the distribution of the loop start respectively end position for larger fibers (1.6Mbp contour length) without histone depletion as a semi-logarithmic plot.

The latter is simply a triangle function

$$f_{S-E}(x) = \begin{cases} 1/l + x/l^2 \Leftarrow x < 0 \\ 1/l - x/l^2 \Leftarrow x \geq 0 \end{cases}$$

with $l = 160kbp$. The composition of the two probability densities $f_{\text{abs}} \circ f_{S-E}$ can be calculated by evaluating the convolution of them (cf. App. H). The density of the absolute value is a simple boxcar-function on the positive x-axis. Therefore, the convolution of the densities is simply the half of the triangle on the positive x-axis:

$$f_L(x) = \begin{cases} 0 & \Leftarrow x < 0 \\ 2/l - 2x/l^2 & \Leftarrow x \geq 0 \end{cases}$$

where l is again the fiber length. But since the loop start and end positions are not uniformly distributed along the contour length of the fiber one does not get a simple linear decrease of the loop length. Nevertheless, some models for genome folding on larger scales assume independent and uniformly distributed chromatin loops [25] since this assumption gets better with increasing fiber length (cf. Fig. 107).

Fig. 108 shows the loop length in dependence on the loop end position as a scatter plot. One can see that for a given end position of the loop all possible loop lengths occur (but not equally distributed of course).
The offset with respect to the x-axis comes from the minimal loop-length condition defined in Sec. 8.3.1. The same applies for the offset in y-direction.
The dependence of the loop length on the loop start position looks like the mirror image of Fig. 108 i.e. the range of possible loop lengths is decreasing with increasing start position.

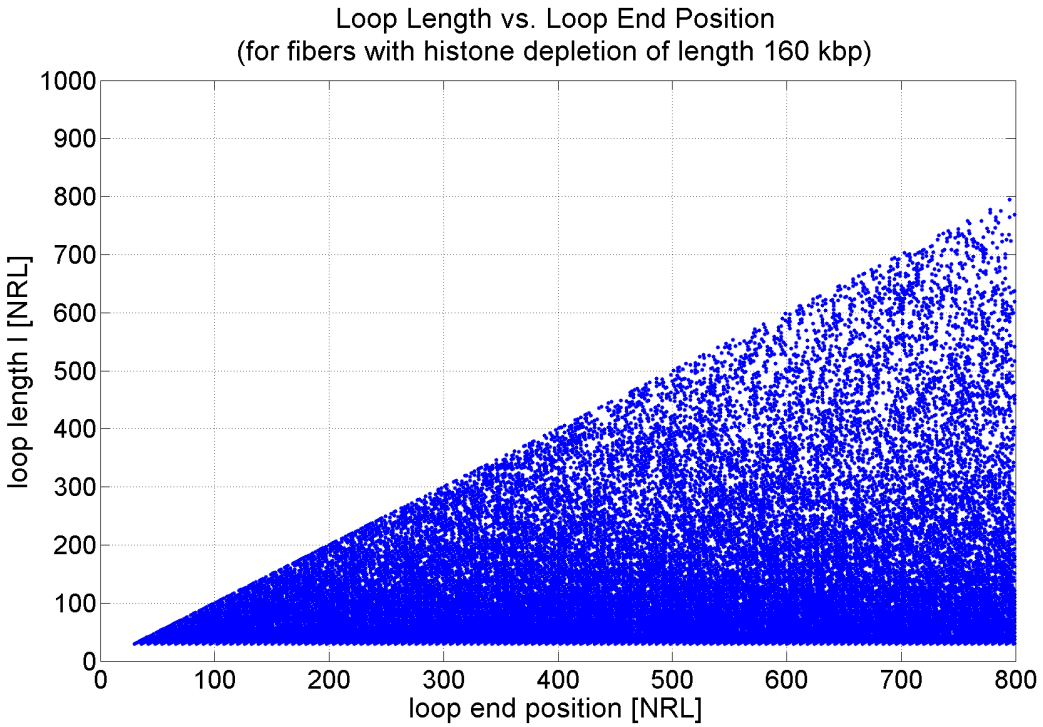


Figure 108: The loop length in dependence on the loop end position as a scatter plot. One can see that for a given end position all possible loop lengths exist.

The dependence of the loop length on the loop center is shown in Fig. 196 in App. A again as a scatter plot. Hereby, the loop center C is defined as follows: $C = (S + E)/2$. As expected the loop length distribution is broadest for loops around the center of the fiber.

8.4 Comparison with 5C Experiments

Experimental 3C-based methods are very important for the investigation of DNA interactions. All the current 3C-based methods are described in [53; 192]. Nevertheless, a short summary will be given in the next section (i.e. Sec. 8.4.1).
After that the simulation results are compared to results of a 5C experiment [72].

8.4.1 3C-Based Methods to Capture DNA Interactions

The shape of the genome is thought to play an important part in the coordination of transcription and other DNA-metabolic processes. Chromosome conformation capture (3C) technology allows to analyze the folding of chromatin in the native cellular state at a resolution beyond that provided by current microscopy techniques, although it brings in some other difficulties [52].
It has been used, for example, to show that regulatory DNA elements communicate with distant target genes through direct physical interactions that loop out the intervening chromatin fiber.

3C technology has become a standard research tool for studying the relationship between nuclear organization and transcription in the native cellular state.
Other technologies based on the 3C principle have been developed that aim to increase the throughput: 4C technology allows for an unbiased genome-wide screen for interactions with a locus of choice, whereas 5C technology permits parallel analysis of interactions between many selected DNA fragments. Furthermore, Chip-loop methodology combines 3C with chromatin immunoprecipitation to analyze interactions between specific protein-bound DNA sequences.

The 3C procedure involves five experimental steps (cf. Fig. 109). First, cells are fixed with formaldehyde, which cross-links proteins to other proteins and to DNA segments that are in close proximity in the nuclear space. Second, the cross-linked chromatin is digested with an excess of restriction enzyme, separating cross-linked from non-cross-linked DNA fragments. Third, DNA ends are ligated under conditions that favor junctions between cross-linked DNA fragments. Fourth, cross-links are reversed. Finally, ligation events between selected pairs of restriction fragments are quantified by PCR, using primers specific for the fragments being studied.

The technique allows the identification of physical interactions between distant DNA segments and of chromatin loops that are formed as a consequence of these interactions, for example between transcriptional regulatory elements and distant target genes [134; 150; 156; 196; 212].

3C technology is particularly suited to study the conformation of genomic regions that range roughly from five to several hundred kilobases in size. The smallest region studied so far by 3C technology spans 6,700 base pairs (bp) [155], whereas the largest region analyzed spans ≈600kb [194].

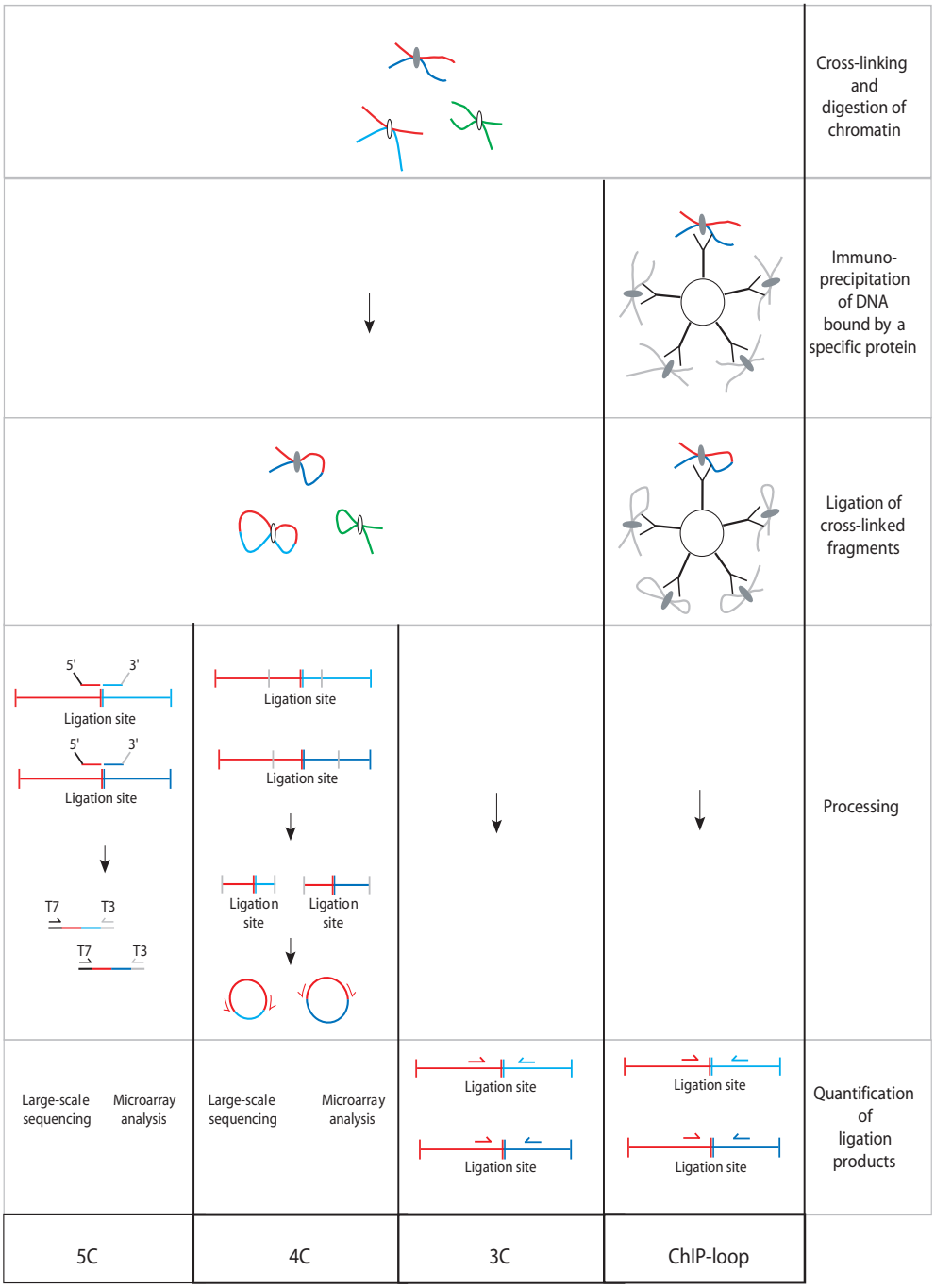


Figure 109: Schematic representation of 3C-based methods. In 3C, 4C and 5C methods DNA interactions are captured by formaldehyde treatment, DNA digestion with a restriction enzyme and ligation of cross-linked fragments and then ligation frequencies are measured. Adapted from [192].

It is important to note that because of the flexibility of the chromatin fiber DNA segments on the same fiber are engaged in random collisions (cf. Sec. 8.2.3). Therefore, the mere detection of a ligation product does not necessarily reveal a specific interaction. To ascertain that an interaction is specific requires the demonstration that two DNA sites interact more frequently with each other than with neighboring DNA sequences. To do so it is important to know the amount of random chromatin collisions which is the subject of this chapter.

3C and 3C-based technologies provide information about the frequency but not the functionality of DNA interactions. Thus, additional, often genetic, experiments are required to address whether an interaction identified by 3C-based technologies is functionally meaningful. For example, many of the interactions identified by 4C technology [191] between genomic regions far apart on the same chromosome or on other chromosomes may well be nonfunctional and merely the consequence of general folding patterns of chromosomes [50].

Furthermore, a meaningful (that is, quantitative) 3C analysis must be performed on a DNA template that represents many genome equivalents and DNA interactions can be quantified accurately only, if they occur in a substantial proportion of the cells. Sites separated by large genomic distances or present on different chromosomes often do not form enough ligation products for accurate quantification, even if microscopy studies suggest that they come together in a substantial proportion of cells. To study such long-range interactions one should use high throughput 4C technology.

Large-scale mapping of several hundred chromatin interactions using standard 3C is time-consuming and difficult. The introduction of the 3C–carbon copy (5C) method generates the possibility of such large-scale locus-wide analysis [72; 73]. The method uses a multiplex ligation-mediated amplification step to amplify selected ligation junctions, thereby generating a quantitative carbon copy of a part of the initial 3C library which is subsequently analyzed by microarray detection or high-throughput sequencing (cf. Fig. 109).

8.4.2 Comparison of Loop Occurrence with 5C Measurements

In this section the chromatin loop statistics (cf. Sec. 8.3) obtained by simulations with the E2A model are compared to actual data sets from 5C experiments [72]. The experimental data which will be used is compiled in Tab. 17, Tab. 18, Tab. 19 and Tab. 20.

Dekker et al. [72] verified the 5C technology at a previously by 3C experiments investigated region of 400kbp length around the human β-globin locus. Furthermore, they investigated a 100kb large gene desert on chromosome 16 as a reference system to measure the physical contacts which come from random chromatin collisions and to normalize the data sets obtained by the different experiments.
Moreover, two different cell types have been used in this study: The erythroleukemia

cell line K562 where the β-globin locus is expressed ("On") and the lymphoblastoid cell line GM06990 where the locus is not expressed ("Off").

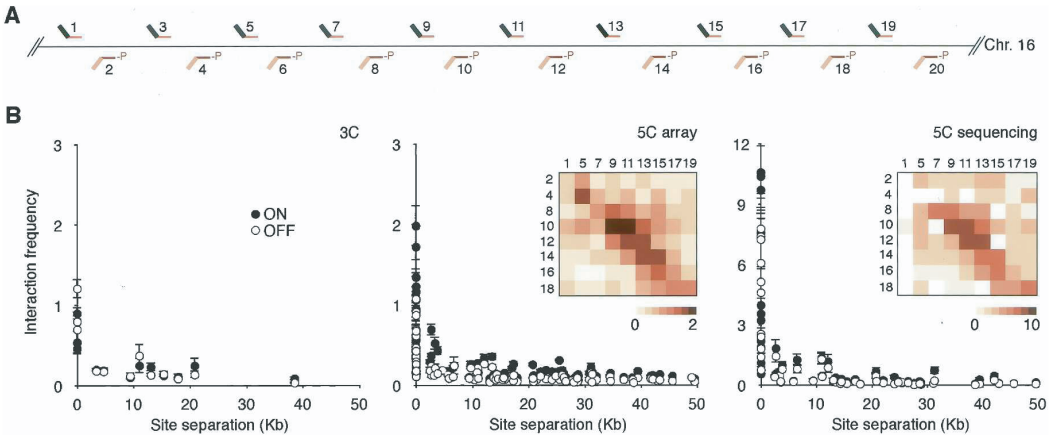


Figure 110: Analysis of the conformation of the gene desert control region. (A) Positions of alternating 5C forward (top) and reverse (bottom) primers throughout the gene desert control region. (B) Chromatin interaction frequencies of the gene desert region as determined by conventional 3C (left panel), by 5C and microarrays (middle panel) and by 5C and quantitative sequencing (right panel). Interaction frequencies are plotted vs. genomic distance between the interacting fragments. Bars represent the S.E.M. Interaction frequencies detected in K562 cells are also presented as two-dimensional heat maps in which the color of each square is a measure of the interaction frequency. Note: Primers 3, 6, and 20 recognize repetitive sequences. These primers were included to generate 5C libraries, but data obtained with these primers are not included in B. Adapted from [72].

Since we are only interested in the statistics of purely *random* contacts here we will focus on the part of the 5C data which concerns the 100kbp gene desert. This region has only the strong interactions between nearby sites but (apparently) no functional long-range looping contacts [51; 53; 172]. Hence, the data should reflect a rather unconstrained chromatin fiber (cf. Fig. 110) which shows the random coil behavior of chromatin.
The β-globin region data looks very different from the data of the gene desert and shows some strong long-range looping interactions [72].
The locations of the primers which where used in [72] to receive the data for the gene desert region can be found in Tab. 16 (App. B).

Fig. 111 shows a comparison between the interaction frequencies obtained by the E2A model simulations and the 5C-data from [72]. This plot is very similar to Fig. 110 and the shape of the loop length distribution has already been discussed in Sec. 9.3. Fig. 112 shows the comparison between the simulations results and the 5C-data, too, but for more different fiber lengths.

One can clearly see that the fibers without histone depletion are much too stiff to allow the looping on the small scale which was found in the experiments. Furthermore, a lot of comparable 30nm chromatin models (cf. Sec. 2) have persistence lengths around 300nm, too, which corresponds to a contour length of about 15kbp along the fiber. That means these models are not able to explain the chromatin loops on the small scale below this value.

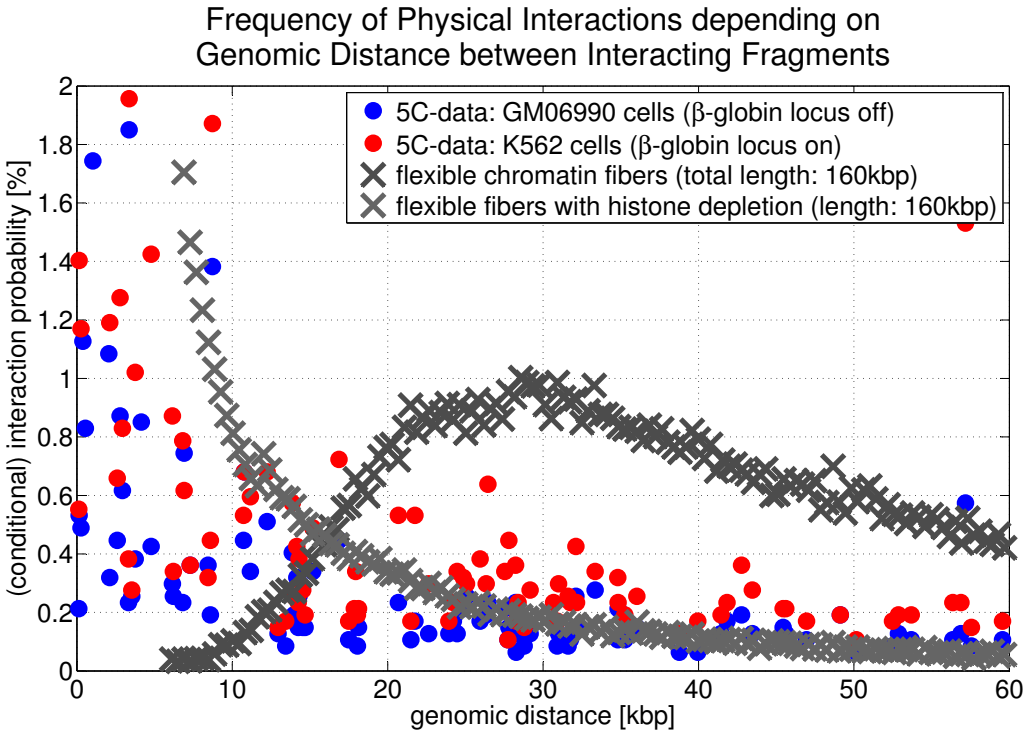


Figure 111: The comparison with the 5C-data [72] shows that the chromatin fibers which do not allow for histone depletion are much too stiff on the small scale. With a persistence length of 280nm, i.e. 13.5kbp along the fiber they don't have any significant looping interactions below genomic distances of this value.

The graph for the fibers with depletion effects do not fit perfectly but show at least the same qualitative behavior as the 5C-data: The loop probability increases when the loop length decreases (cf. Fig. 110 and Fig. 111), especially in the region below 15 kbp. This can only be explained by allowing for depletion effects, nucleosome depletion in particular.
This is a clear hint that histone depletion plays a significant role for the looping properties of chromatin. The depletion effects might even be crucial for transcription purposes as only they allow physical contacts on a small scale.

Due to the cut-off for the loop size formerly defined (cf. Sec. 8.3.1) we do not get smaller

loops than 5.9kbp.

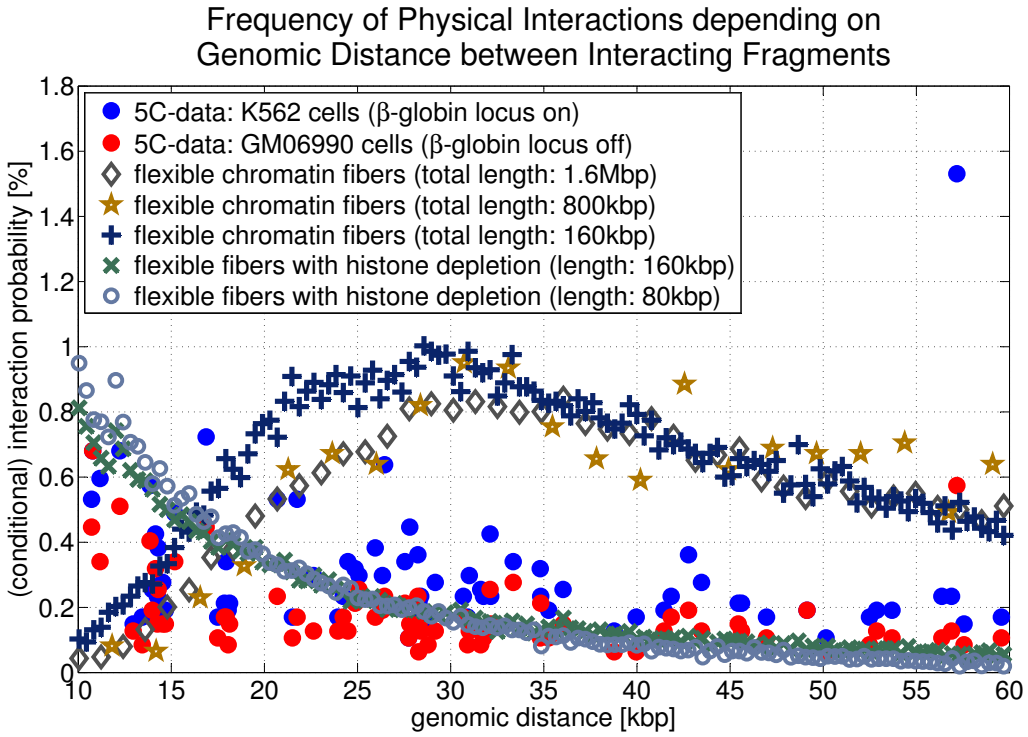


Figure 112: The interaction frequency depending on the corresponding genomic distance for different fiber lengths compared with 5C-data [72]. The main difference in the different distributions of the same fiber type come from the normalization since the fiber length limits the maximal loop length. One can see that only fibers with histone depletion and nucleosome depletion in particular explain the loops on the small scale.

The graphs in Fig. 111 and Fig. 112 do not reveal where along chromosome 16 the found interactions occur. To better illustrate this chromatin interaction maps are shown in Fig. 113 (5C-data), Fig. 114 (simulated chromatin fibers of length 160kbp) and Fig. 115 (simulated fibers without histone depletion of length 1.6Mbp).
Furthermore, the interaction frequencies are also illustrated as heat maps in Fig. 116 and Fig. 117 for the case of 5C-data and Fig. 118 and Fig. 119 and Fig. 120 for the case of chromatin fibers generated by the E2A model. Fig. 197 and Fig. 198 show the 5C-data in the among biologists popular way of heat maps with equidistant spacing. They can be found in App. A. The color of each square is an indication of the interaction frequency between the considered fiber parts (cf. colorbar).

One can see again that the most frequent interactions occur between chromatin parts that have a small genomic distance.
The large gap in the case of fibers without histone depletion (cf. Fig. 114, top) comes

from the local fiber stiffness: Points that are very close to each other along the chromatin fiber can't have a small spatial distance because the fiber is not able to bend that tightly. The gap occurs in the case of chromatin with histone depletion as well, but its width is much smaller (cf. Fig. 114). The gaps in Fig. 114 (top) and Fig. 115 are equally broad.

The gap widths can be compared in Fig. 121 where a cut through the previous 3D plots (i.e. Fig. 113, Fig. 114 and Fig. 115) is presented. In the case of the experimental data two cuts have been made (denoted by 'I' and 'II') for each cell type.

All cuts go through the central interaction peak and therefore the graphs represent the interaction frequency of a particular fiber part with the center of the fiber. The distance from the fiber center is shown at the x-axis. The sign of the distance corresponds to either one fiber end or the other.

One can see that the interaction frequency of chromatin without histone depletion is much too small in order to explain the behavior of the 5C-data. The graph for the fibers with depletion effects comes much closer although some data points seem to lie within the central zero interaction gap.

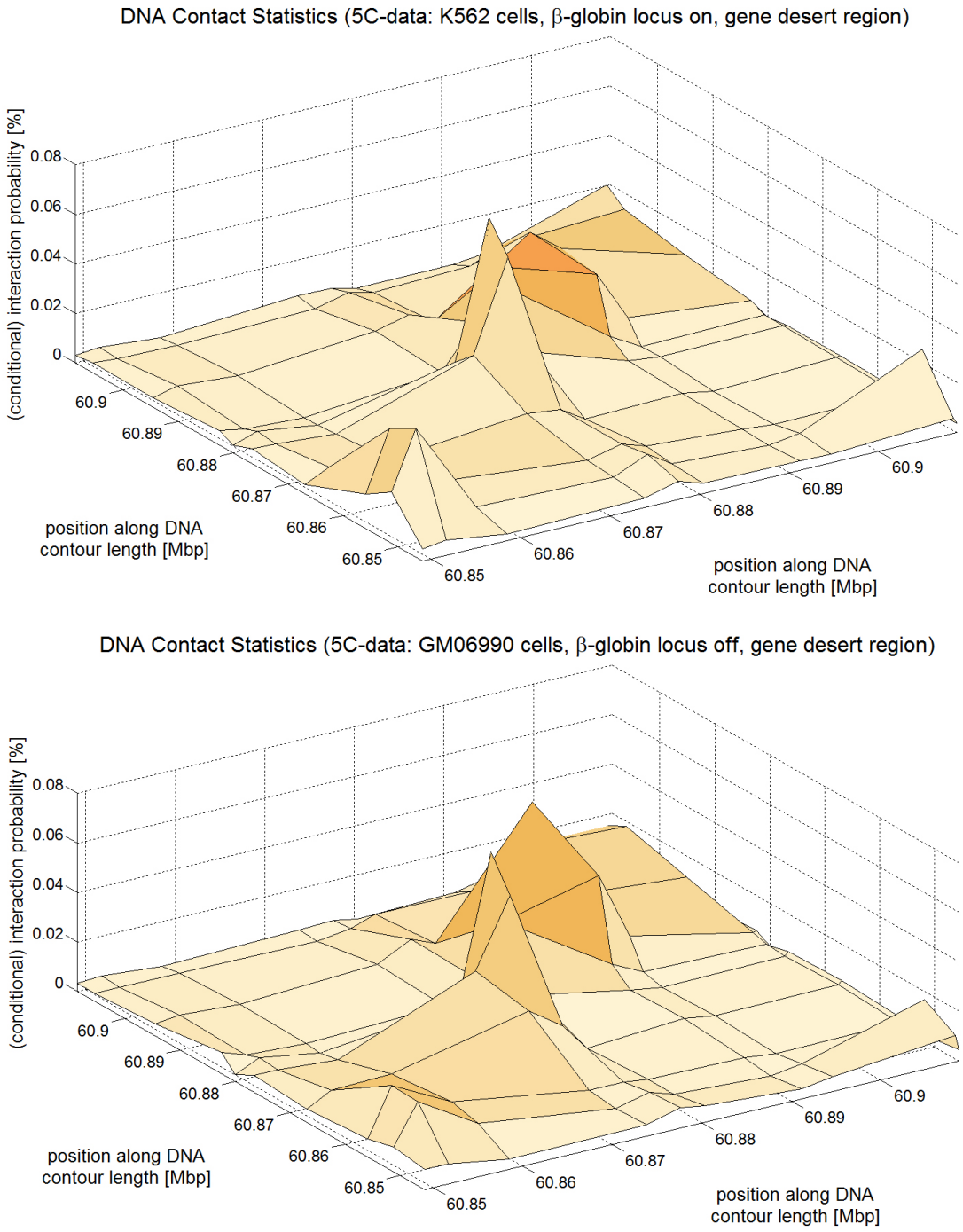


Figure 113: This figure shows the interaction frequency which was measured by 5C experiments [72] on chromosome 16. One can see that chromatin parts with a small genomic distance interact most frequently.

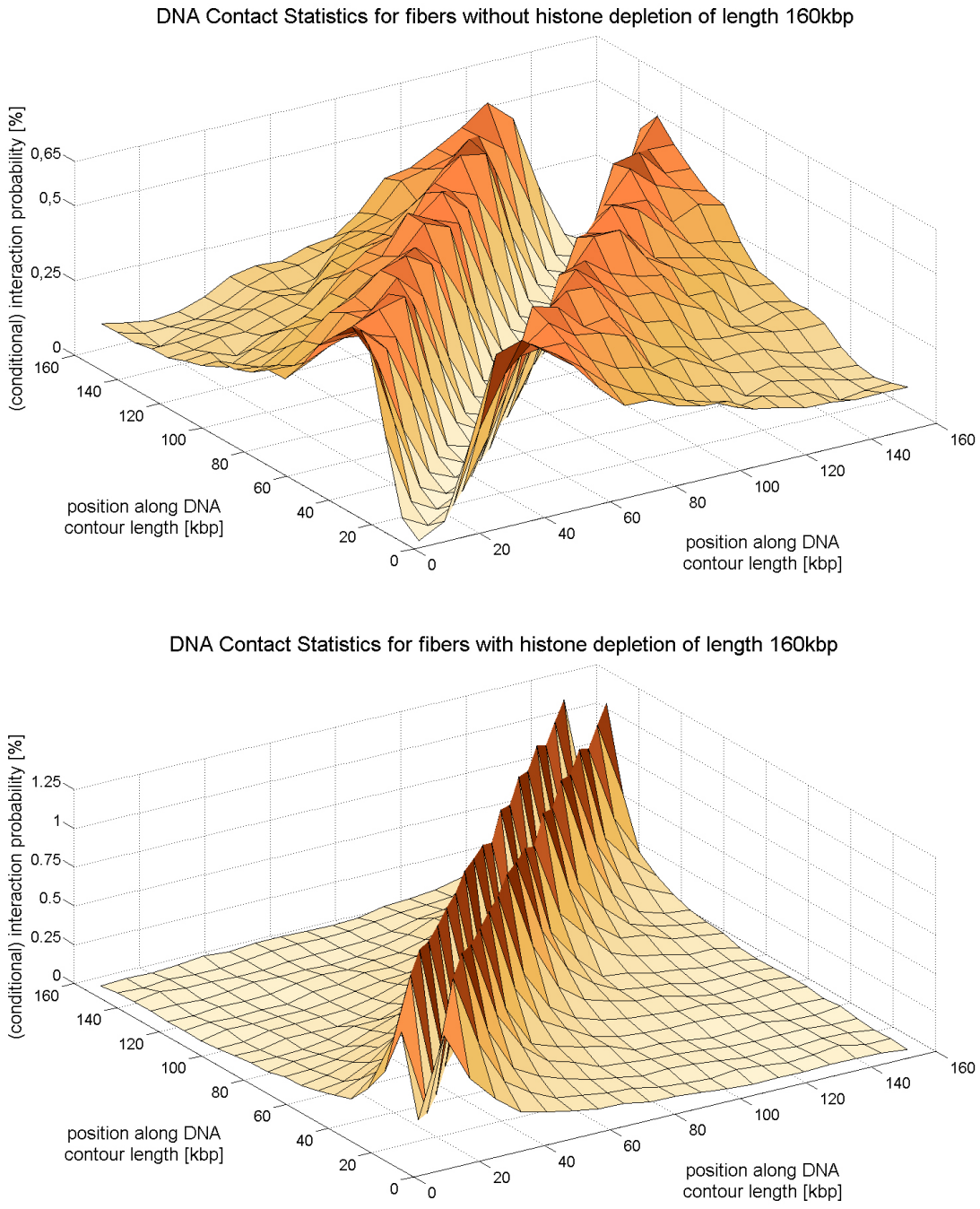


Figure 114: A plot of the interaction frequency found by chromatin simulations with and without histone depletion. The fiber length is fixed to 160kbp in this case. The central gap comes from the local stiffness of the fibers. One can see that parts of chromatin with histone depletion can even interact when their genomic distance is small.

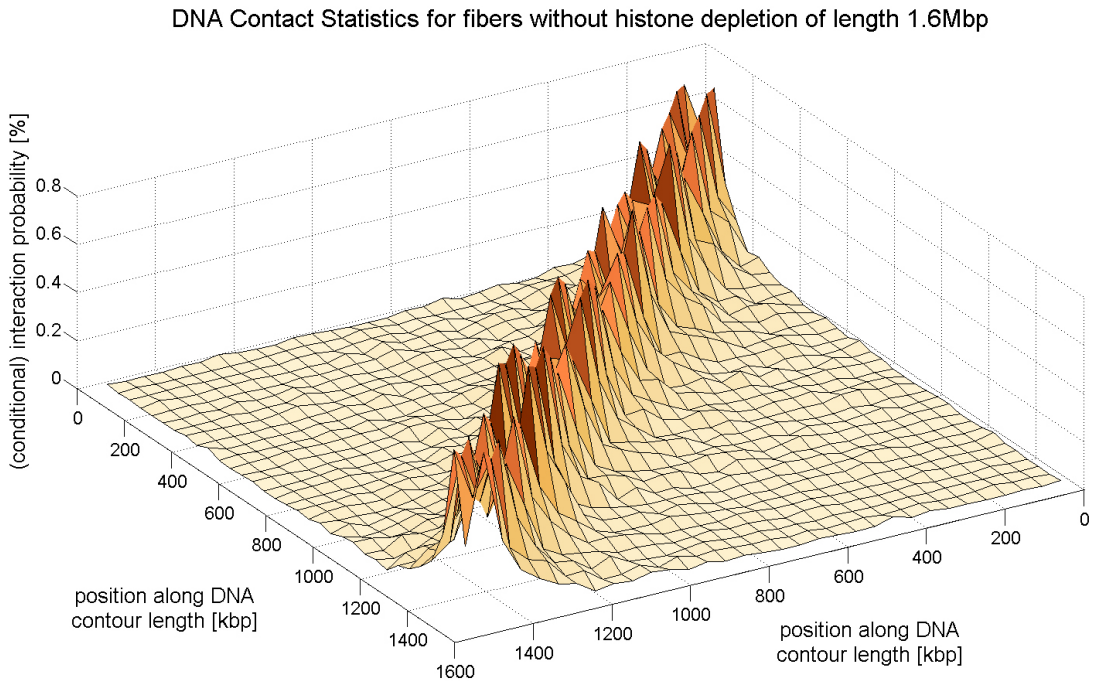


Figure 115: This figure shows the interaction frequency of chromatin fibers without histone depletion of length 1.6Mbp.

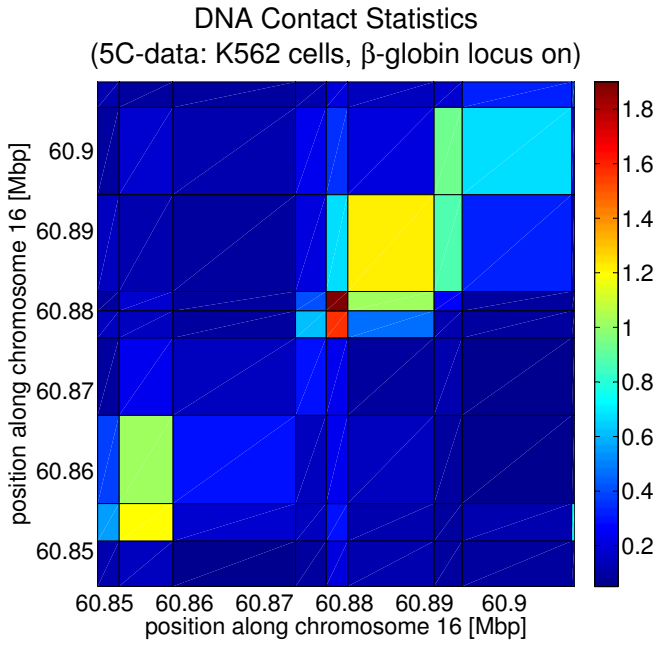


Figure 116: Heat map illustration of 5C-data (K562 cells) [72].

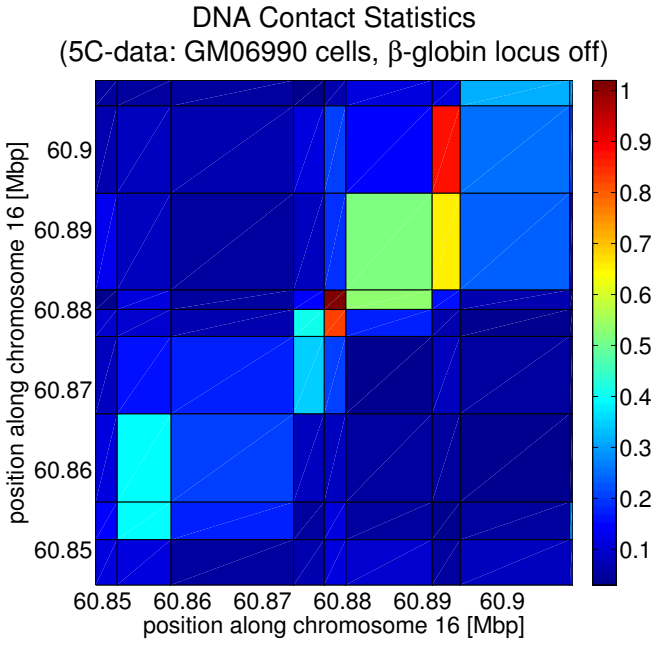


Figure 117: Heat map illustration of 5C-data (GM06990 cells) [72].

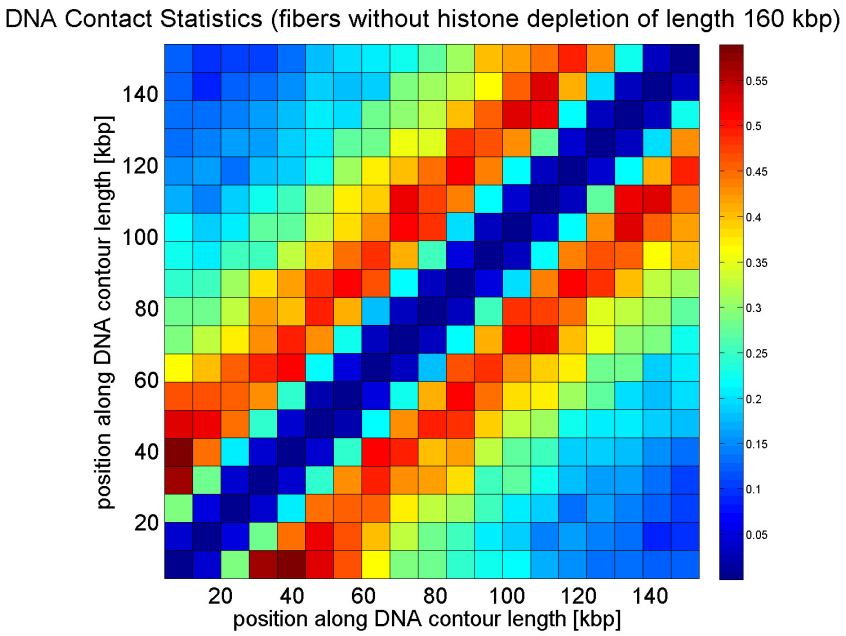


Figure 118: Heat map illustration of fibers without histone depletion (160kbp).

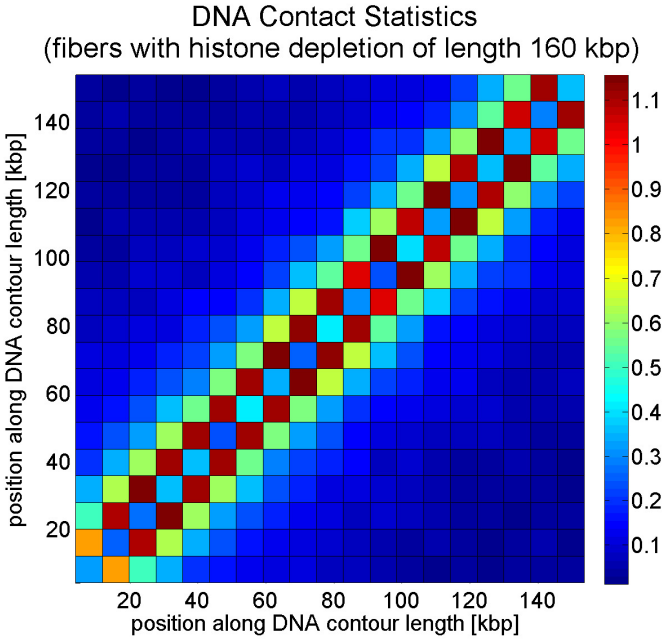


Figure 119: Heat map illustration of fibers with histone depletion (160kbp).

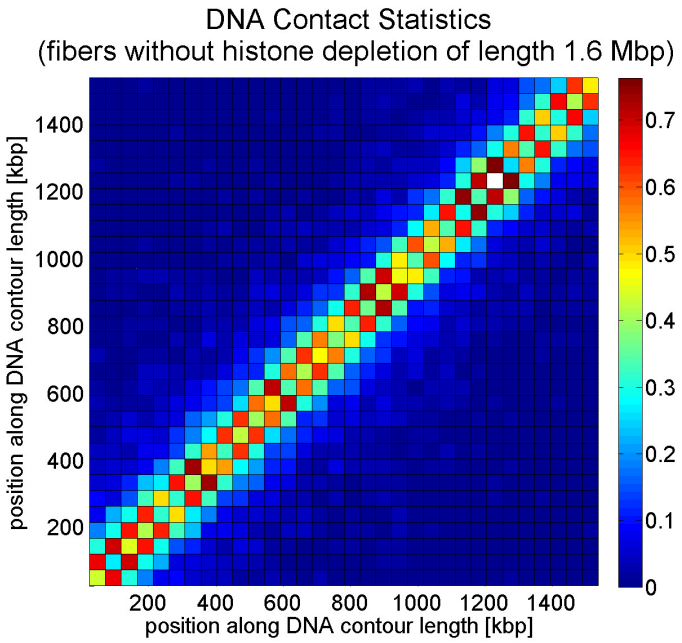


Figure 120: Heat map illustration of fibers without histone depletion (1.6Mbp).

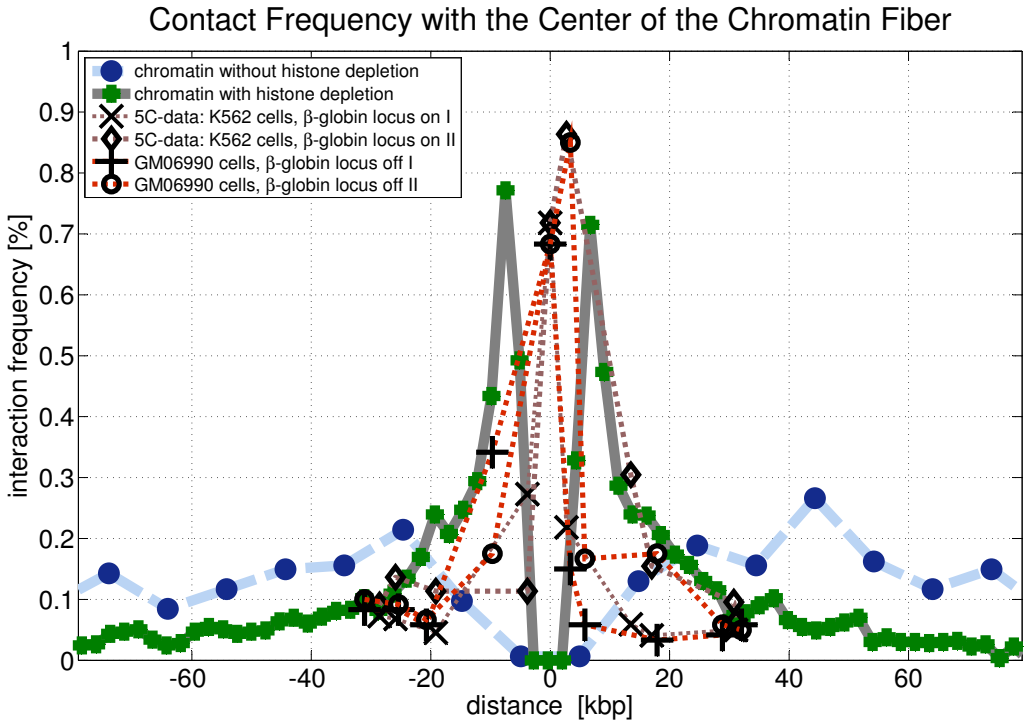


Figure 121: This figure shows the interaction frequency of different fiber parts with the center of the chromatin fiber. Shown are simulation results of chromatin with and without histone depletion as well as results from 5C-experiments on chromosome 16 [72].

8.5 Chromatin Loop Entropy

In systems which are in thermal equilibrium with their environment, i.e. which have a constant temperature, the free energy F is minimized. In this section the entropy S which contributes to the free energy is examined with respect to chromatin loops. Of course the entropy is only one side of the medal since $F = U - TS$ determines the system's behavior.

8.5.1 Loop Number Entropy

Fig. 122 shows the loop number entropy difference $\Delta S_{k,0}$ between chromatin fibers which have k loops and fibers without loops:

$$\Delta S_{k,0} = S_k - S_0 = k_B \ln\left(\frac{p(k)}{p(0)}\right),$$

where $p(k)$ is the probability for a chromatin fiber to have exactly k loops (cf. Fig. 104). Here it is assumed that $p(k) \propto \Omega_k$, the number of microstates with exactly k loops. The loop entropy number difference of other fiber with several different lengths can be

found in App. A (cf. Fig. 199).

One can see that the entropy difference of the fibers without depletion effects is always negative. This is because the probability to have no loop at all $p(0)$ is larger than for any other loop number $p(k > 0)$.
The entropy difference in the case of fibers with histone depletion first increases since $p(0)$ is very small in this case and it is unlikely to have only a few loops. Then it decreases because $p(k)$ decreases, too (due to the fact that the chain length and the transience of the fiber limits the loop distribution). The entropy is positive in this case since it is very likely that chromatin fibers with histone depletions have several loops.

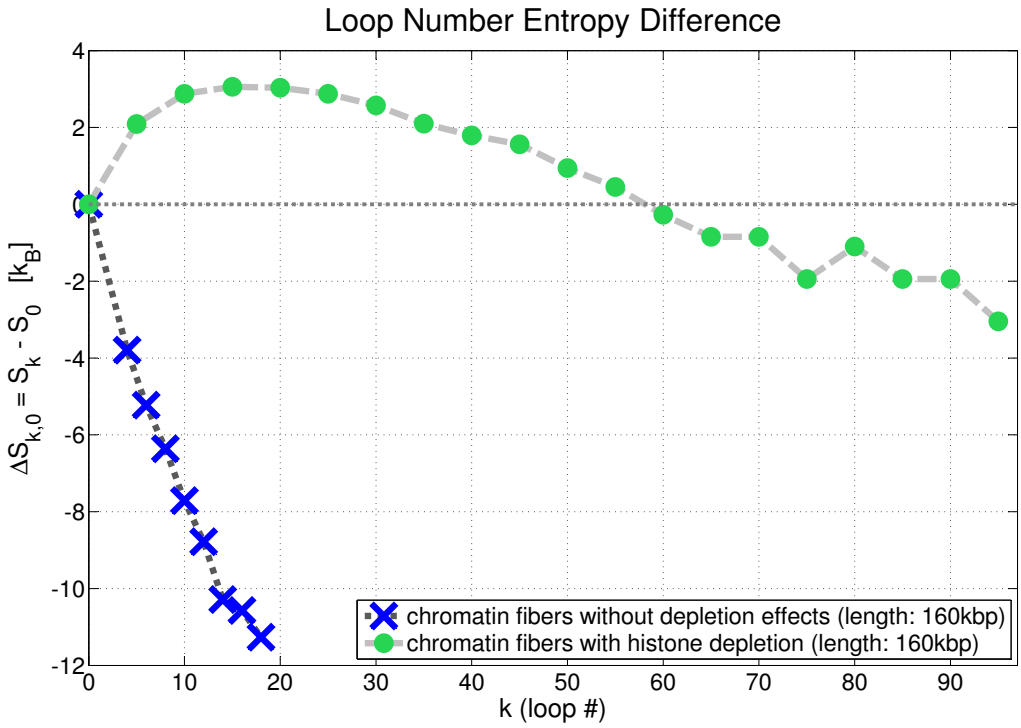


Figure 122: This figure shows the loop number entropy difference between the chromatin fibers without loops (S_0) an the fibers with k loops. The entropy difference in the case of fibers without histone depletion is always smaller than zero because here the largest loop number probability is $p(0)$.

8.5.2 Loop Entropy

Entropy loss due to loop formation was studied for the case of disconnected loops in [84; 180; 189]. In their seminal article, Poland and Scheraga [162], and Wang and Uhlenbeck [222], considered coupled Gaussian loops.

Loop entropy is defined as the entropy lost upon bringing together two segments of a

polymer. For a fiber conformation with only one single loop the entropy loss is given by

$$\Delta S = \alpha \, k_B \, \ln(l),$$

where l is the loop length and α is a polymer type specific coefficient.

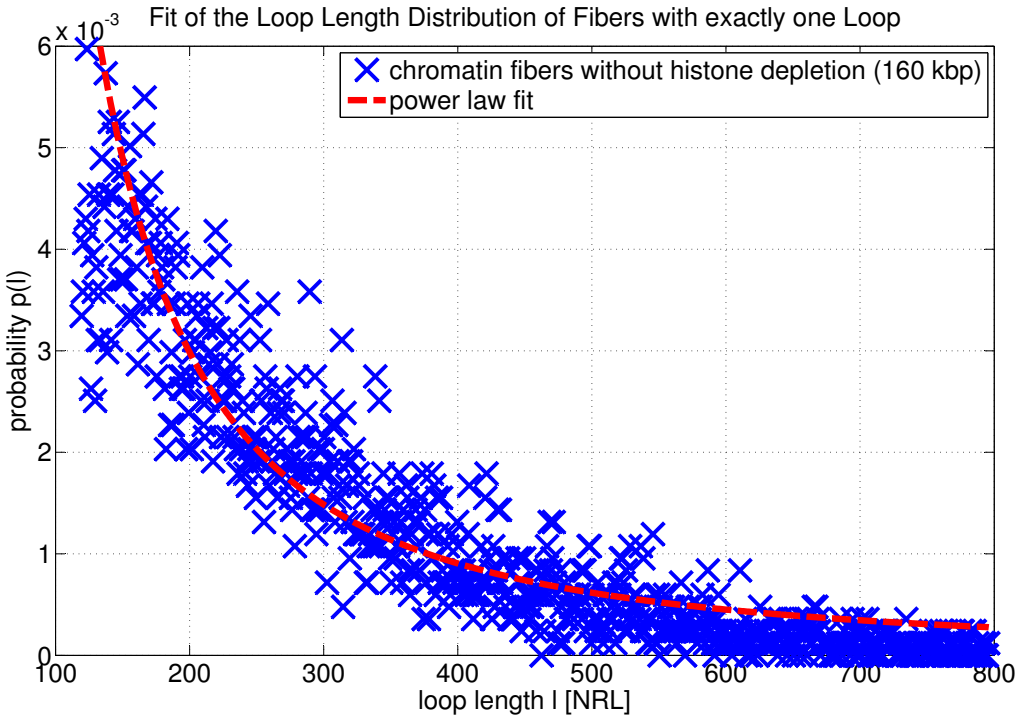


Figure 123: Power law fit of the loop length distribution in the case of chromatin fibers without histone depletion. The goodness of the fit is SSE=0.0001155 (Sum of Squares due to Error) and R-square = 0.8667 (coefficient of multiple determination). The exponent of the fit is -1.725($= -\alpha_1$).

This expression for the entropy comes from a power-law probability distribution of the loop length: $p(l) \propto l^{-\alpha}$. In the case of a three-dimensional random walk $\alpha = 3/2$ and in the case of a self-avoiding walk $\alpha \approx 2.2$ [215].
Furthermore, the loop entropy varies with the loop position since contacts near the ends of a polymer are more likely than contacts in the middle due to excluded volume effects. But the influence of the loop position on the entropy will not be considered here. It was already discussed in Sec. 9.3.

The loop length distribution was fitted with a power law in order to estimate α in the two cases of chromatin fibers with and without histone depletion.

One needs independent loops for this loop length distribution. Therefore, in the case of fibers without histone depletion only such fibers were considered which have exactly one loop.
In the other case of fibers with histone depletion it is very hard to find conformations with only one loop. Hence, the used loop length distribution contains all fibers that have less than ten loops in that case assuming that the loops are not influencing each other very much.

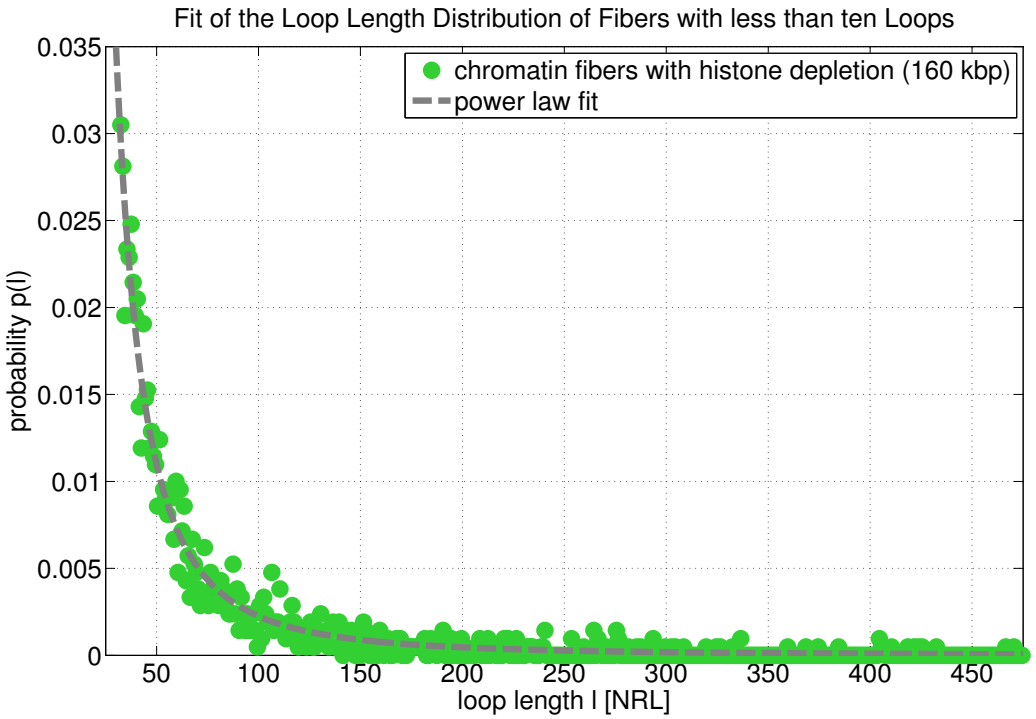


Figure 124: Power law fit of the loop length distribution in the case of chromatin fibers with histone depletion. The goodness of the fit is SSE=0.0001589 (Sum of Squares due to Error) and R-square = 0.8809 (coefficient of multiple determination). The exponent of the fit is -2.231($= -\alpha_2$).

The results of the fits are shown in Fig. 123 and Fig. 124. The determined coefficients are $\alpha_1 = 1.725$ in the case without histone depletion and $\alpha_2 = 2.231$ in the case of fibers with histone depletion. The chromatin fiber length was fixed to 160kbp for both types of chromatin fibers.

α_1 is smaller than the theoretical prediction for self-avoiding random walks. It seems that the fibers are too short to already feel the long-range excluded volume interactions at full extent. On the other hand, the fit is strongly influenced by the scatter of the data points with small loop lengths. Therefore, the fit does not match the tail of the distribution very well. If this was the case, the coefficient would be larger (abs. val.)

and thus fit better the theoretical prediction.
In the case of fibers with histone depletion the coefficient matches the prediction of 2.2 very well.
Fig. 125 shows the distribution of the loop entropy $\Delta S_i = \alpha_i k_B \ln(l)$ in the cases of fibers with one respectively less than ten loops. One can see that the distributions resemble the distribution of the loop lengths (cf. Fig. 105).

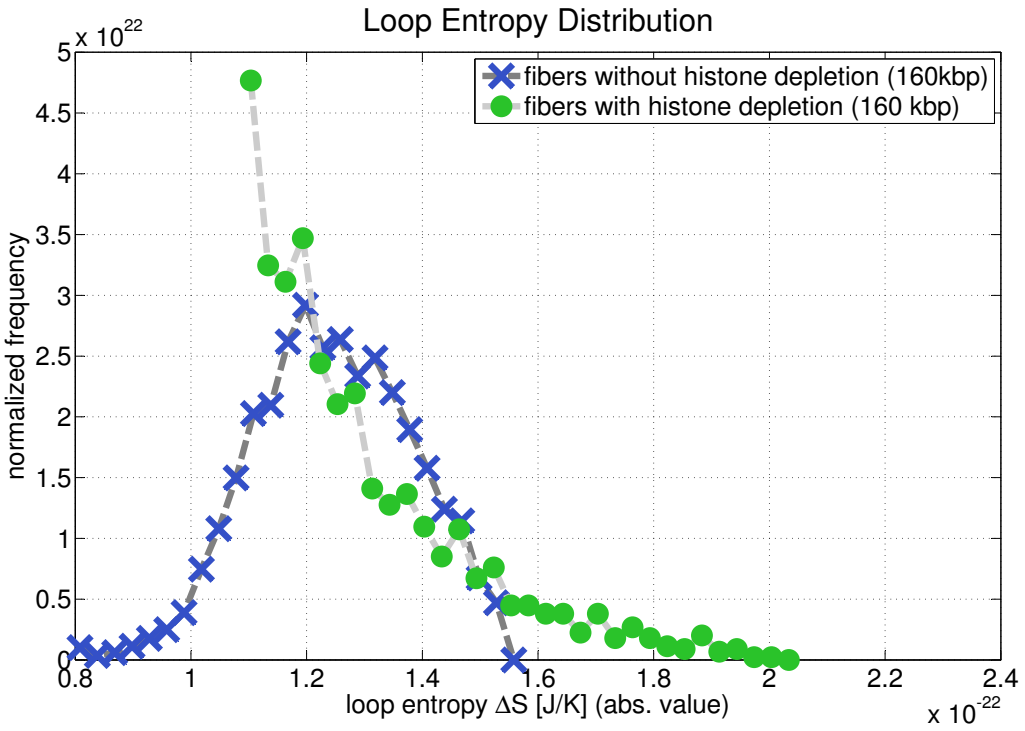


Figure 125: The loop entropy distribution of chromatin fibers with one respectively less than ten loops.

8.5.3 Wang-Uhlenbeck Entropy

The loop entropy becomes more complicated in the case of multiple loops within the one fiber. In this case the loop entropy is determined by a matrix method which was developed by Wang and Uhlenbeck [222].

If a fiber has k loops then the corresponding Wang-Uhlenbeck matrix is a $l \times l$ symmetric, real matrix. The elements W_{ij} of the matrix equal the number of common monomers of loop i and j i.e. the diagonal elements are simply the loop lengths. The entropy of making the particular contacts is then given by

$$\Delta S = \alpha \; k_B \; \ln(\det(W)).$$

As an example consider a fiber with $k = 3$ loops with the starting points $s = (82, 44, 62)^t$ and the loop end positions $e = (95, 56, 91)^t$ in multiples of the nucleosome repeat length. Loop one and loop three overlap which means that the corresponding Wang-Uhlenbeck matrix is

$$W = \begin{pmatrix} 13 & 0 & 9 \\ 0 & 12 & 0 \\ 9 & 0 & 29 \end{pmatrix}.$$

Fig. 126 shows the distribution of the Wang-Uhlenbeck loop entropy for all chromatin fibers without histone depletion and chromatin fibers with histone depletion. All kinds of loops are considered here and the fiber length was fixed to 160kbp in both cases. The mean values of the entropy distributions are about $440k_B$ (without histone depletion) and $\approx 200k_B$ (with histone depletion) but due to the results shown in Fig. 122 the Wang-Uhlenbeck entropy is only conditionally applicable in the latter case.

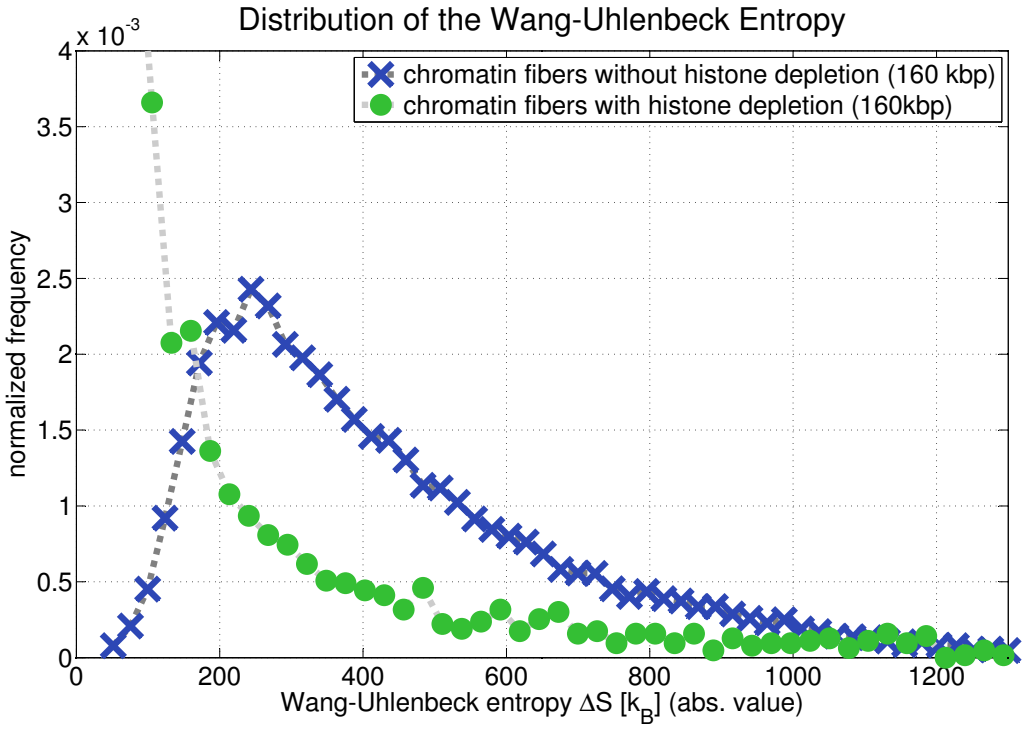


Figure 126: This figure shows the Wang-Uhlenbeck entropy distribution of chromatin fibers with multiple loops.

9 Gelation of Nucleoporins

In this chapter it will be examined whether hydrophobic parts of FG-rich nucleoporins are the reason for their ability to form a hydro-gel [87]. Thereby, the N-terminal fsFG-domain of the essential yeast nucleoporin Nsp1p [107] will serve as a nucleoporin model system (cf. Sec. 9.1).

There is a competition between two kinds of entropic forces in the system: The excluded volume interactions and the hydrophobic parts of the nucleoporin strands. Therefore, it is not a priori clear whether the system percolates at a biologically realistic density. A Metropolis Monte Carlo model was developed to address this question. It will be described in Sec. 9.2 together with the percolation theoretical approach that will be used to examine the connectivity properties of the model system.

Finally, a sol-gel phase transition will be found in the system at a critical density of 42mg/ml (cf. Sec. 9.3). This may be considered as a hint that hydrophobic nucleoporin parts are key for the formation of gels in the nuclear pore complex (cf. Sec. 9.4).

This project was inspired by Prof. Dirk Görlich at the MPI for Biophysical Chemistry in Göttingen who funded the main part of my PhD.
The main results of this chapter have already been submitted to a scientific journal [66].

9.1 Introduction

All the proteins of the cell nucleus are imported from the cytoplasm. On the other hand, the nucleus supplies the cytoplasm with all kinds of nuclear products such as messenger RNA or transfer RNA. Since, during interphase, the eucaryotic cell nucleus is surrounded by a double membrane called the nuclear envelope (NE, cf. Fig. 127 and Fig. 129), transport between the cytoplasm and the nucleus has to proceed through specialized pores (cf. Fig. 205, Fig. 128 and Fig. 206 in App. A).

Such pores are built by so-called nuclear pores complexes (NPCs, cf. Fig. 130), large proteinaceous channels distributed across the NE. About 2000 of such nuclear pore complexes exist on average within the nuclear envelope of a vertebrate cell. Their exact number varies depending on cell type and throughout the cellular life cycle.
The central aqueous channel within NPCs is ≈40-60nm in length and has a diameter of ≈30-50nm depending on the organism (cf. Fig. 132). This channel is guarded by a permeability barrier [93; 138] that prevents uncontrolled intermixing of cytoplasmic and nuclear contents by effectively blocking diffusion of particles larger than ≈30-40kD in size. At the same time, however, this barrier allows efficient passage of molecules which are in complex with nuclear transport receptors (NTRs) [93; 159; 224]. This facilitated mode of NPC passage is reliant on interactions of NTRs with the FG-rich nucleoporin repeats [15; 16; 20; 108; 167].

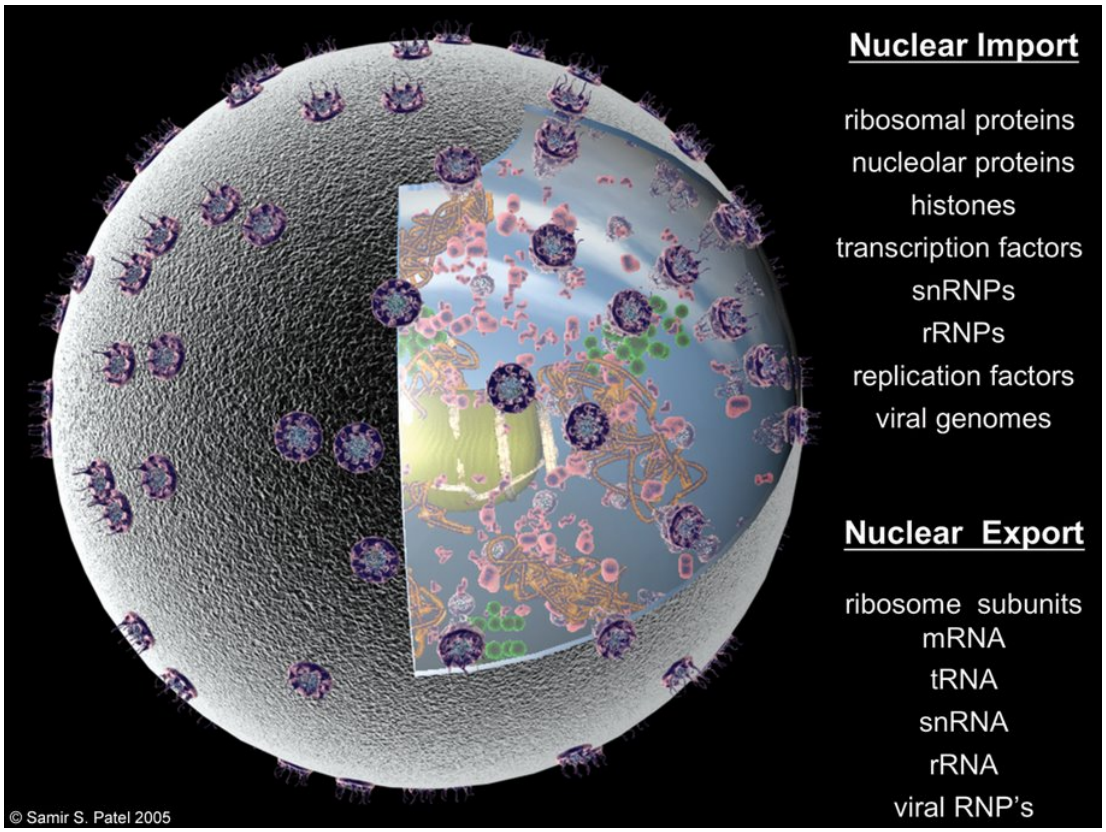


Figure 127: Nuclear pores are large protein complexes that are located at the nuclear envelope. In an average vertebrate cell a nuclear envelope has approximately 2000 nuclear pores. Adapted from [157].

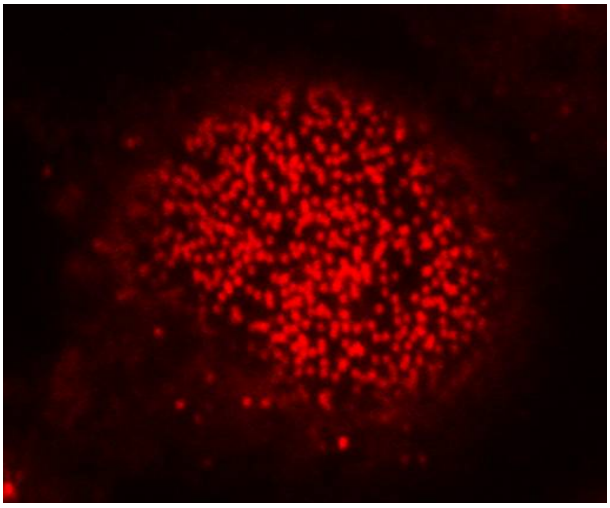


Figure 128: Fluorescently labelled NPCs on the surface of a HeLa nucleus. Adapted from [231].

Each NPC is built-up from multiple copies of a set of about 30 different proteins collectively called nucleoporins (cf. Fig. 130 in App. A). About half of these nucleoporins predominantly consist of folded domains often displaying alpha solenoid- or beta-propeller folds. The second half also contains domains that show structural characteristics typical for intrinsically unfolded proteins [54], i.e. highly flexible elements lacking ordered secondary structure [54]. Such unstructured domains typically consist of up to 50 Phenylalanine-Glycine (FG) dipeptide motives [160; 174] spaced by hydrophilic amino acid stretches and are therefore called FG-rich nucleoporin repeat domains (cf. Fig. 133).

Although it is commonly accepted that the permeability barrier consists of FG-rich nucleoporin repeat domains, several different models exist that try to explain how such repeats can form an efficient barrier.

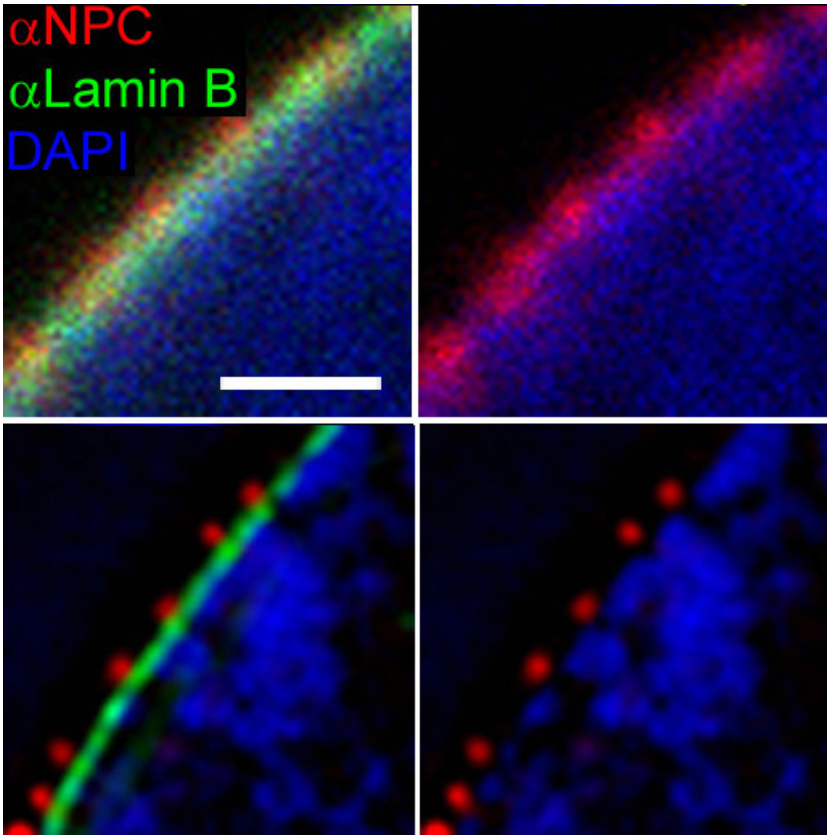


Figure 129: Shown are details of a nuclear envelope. Nuclear pores (anti-NPC) red, nuclear envelope (anti-Lamin) green, chromatin (DAPI-staining) blue. Scale bars: $1\mu m$. The images have been obtained by confocal laser scanning microscopy (top) and 3D structured illumination microscopy (3D-SIM-Microscopy, bottom). Adapted from [182].

These models [138; 170; 171; 175] differ foremost in the question of whether the permeability barrier is tightened by inter-repeat interactions or not. The "selective-phase" model [170; 171] assumes the barrier to be a sievelike structure formed by hydrophobic interactions between the FG-repeats' hydrophobic clusters. Then the sieve's mesh size predetermines a size limit for passive exclusion. NTRs could bypass this size restriction since their binding to hydrophobic clusters competes the noncovalent inter-repeat cross-links and, thereby, opens adjacent meshes of the sieve.

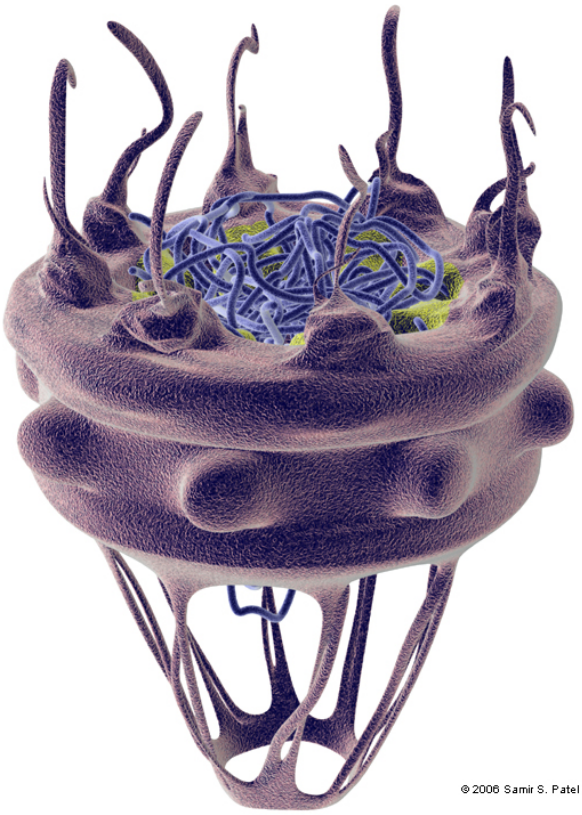


Figure 130: An illustration of a nuclear pore complex. One can see the nuclear envelope and the basket as well as the filaments of the pore complex. Adapted from [157].

All other models have in common that they neglect interactions between FG-repeat domains. Recent attempts to experimentally discriminate between the proposed models were based on a simple assumption: Without inter-repeat interactions aqueous solutions of FG-repeat domains should behave like viscous fluids. In contrast, if cross-linking prevented a free sliding between the linear polymers, an elastic hydrogel should form. Görlich et al. [87] tested these predictions for the N-terminal fsFG-repeat domain of the essential yeast nucleoporin Nsp1p which comprises 18 regular FSFG-repeats and [107] less regular FG-repeats and showed that indeed a macroscopic gel is formed when the protein concentration exceeds 8-10mg/ml of repeat domains [88] (cf. Fig. 131).

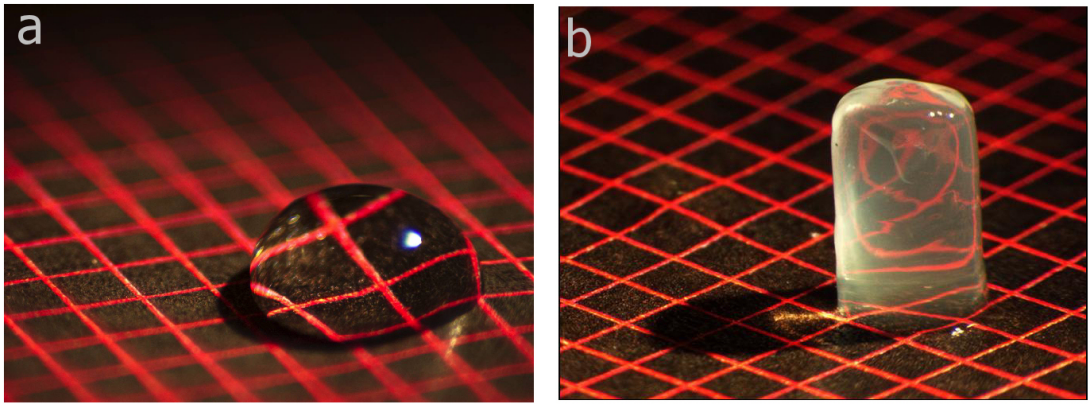


Figure 131: FG repeats can form an elastic hydrogel in aqueous solution. (b) An aqueous solution with 26 mg/ml wildtype fsFG-repeat domain from Nsp1p (400 mM) was filled into a silicon tubing, where it completed gelling. The formed gel was pushed out of the tubing by gentle pressure, placed onto a patterned support (1 square = 1.4 mm), and photographed. (a) The F→S mutated repeat domain remained liquid after identical treatment as b). This shows that nucleoporin FG-domains can form a hydrogel under native conditions and that Phenylalanine side chains are crucial for gel formation. Adapted from [87].

The same repeat domain (cf. App. F) will be considered here as a FG-rich model system. We want to address the question whether the interactions of the hydrophobic FG-rich repeats are actually strong enough to form a hydrogel. These attractive cluster-forming interactions are in strong competition with the repelling (entropic) forces caused by excluded volume of the nucleoporins and their side chains so that it is not a priori clear, whether the system shows a transition from the sol to the gel state [83] at a biologically realistic density. In contrast to the colloids sol state, gel formation would argue for a substantially dilute crosslinked system.

Answering this question will give important information about possible implications of hydrophobic interactions for the gelation of nucleoporin FG-repeat domains. Furthermore, one can examine whether the system is in the 3D percolation universality class. Eventually, a finite size scaling will supply the order of magnitude of the critical density at which the phase transition lies since we are interested in modeling the main effect here and do not want to go to much into details in order to keep the model fast.

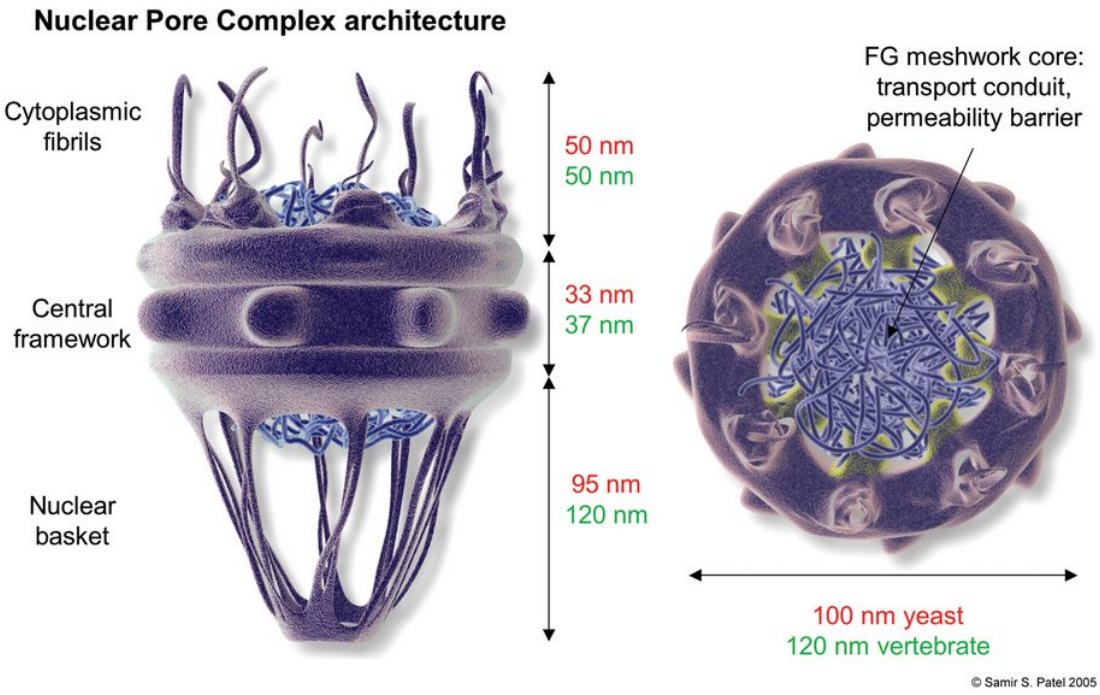


Figure 132: This figure illustrates the geometry of nuclear pores. Adapted from [157].

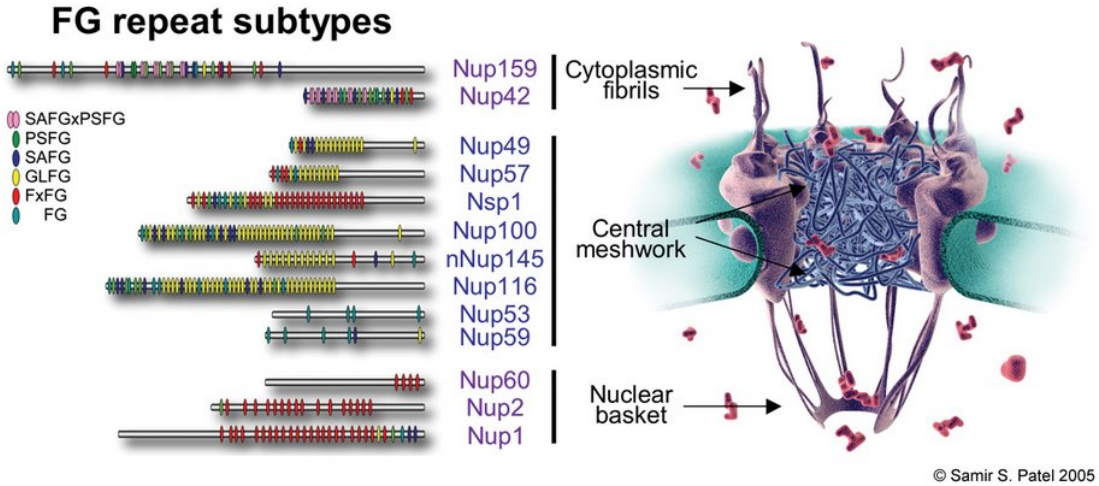


Figure 133: The nuclear pore consists of proteins which are called nucleoporins. About half of the nucleoporins in a pore are natively unfolded i.e. they lack of ordered secondary structure. These unfolded peptides are called FG nucleoporins because their amino acid sequence is rich on phenylalanine—glycine subunits. Adapted from [157].

9.2 Hydrophobicity induced Gelation of FG-rich Nucleoporins

In this chapter the question whether hydrophobic parts of FG-rich nucleoporins can be the reason for their ability to form a hydro-gel [87] is addressed. Thereby, the N-terminal fsFG-domain of the essential yeast nucleoporin Nsp1p [107] will serve as a nucleoporin model system.
It comprises 18 regular FSFG-repeats and 16 less regular FG-repeats. This domain is modeled and equilibrated ensembles of peptide networks were generated by a Metropolis Monte-Carlo algorithm which then were analyzed by percolation theoretical methods. The excluded volume of the protein backbone and all side chains which are at least medium-sized (starting with Glu / E) as well as the hydrophobic clusters of the amino acid sequence (cf. App. F) will be taken into account.

There is a competition between two kinds of entropic forces in the system: The excluded volume interactions and the hydrophobic parts of the nucleoporin strands. Therefore, it is not a priori clear, whether the system percolates at a biologically realistic density. Nevertheless, a sol-gel phase transition was found in the system at a critical density of 42 mg/ml. This may be considered as a hint that hydrophobic nucleoporin parts are key for the formation of gels in the nuclear pore complex.

9.2.1 Sol-Gel Transition as a Percolation Problem

FG-rich parts of nucleoporins in water have the ability to cluster together due to their strong hydrophobicity [11]. Therefore, two or more different peptid chains can coalesce via hydrophobic bonds to a *cluster* cf. Fig. 134 and Fig. 138. On the other hand, repelling forces as the excluded volume of the peptide backbone or the side chains counteract this effect. The mixture of peptide chains with the water molecules and possibly already existing small peptide clusters represents the sol phase. In sol systems [176; 200] it happens that as the mass density c of the monomers (or even polymers) increases, larger and larger clusters are formed.

The interesting question is whether the hydrophobic clustering can become so strong that it leads to a phase transition of the system. If this is the case, the system will develop macroscopic properties like a finite elasticity or shear modulus [176; 200] when the density exceeds a critical value c_{crit}.

At the phase transition a large cluster will exist that represents the gel network and often coexists with single unbound peptide chains that may even be trapped in the interior of the gel cf. Fig. 138. Further increasing of the density will let the sol phase disappear completely and eventually all peptides contribute to one large cluster i.e. the gel network.

In the case of such a sol-gel phase transition the critical density signals a connectivity transition of the system. This model of polymerization and gelation was first invented by Flory and Stockmayer [82; 83; 203] who were interested in the formation of large branched polymers.

Physical gels, in contrast to chemical gels, are formed when no permanent chemical reaction takes place between the monomers (or polymers). Instead only a reversible association links the particles to each other. Examples are silica aerogels and structures formed during the gelation of silica particles in pure water or NaCl solution [176].
In the following model hydrophobic bonds between peptides will have the ability to be dissolved again by thermic fluctuations (cf. Sec. E). In all these examples the connectivity properties of the system play a key role for their physical understanding.

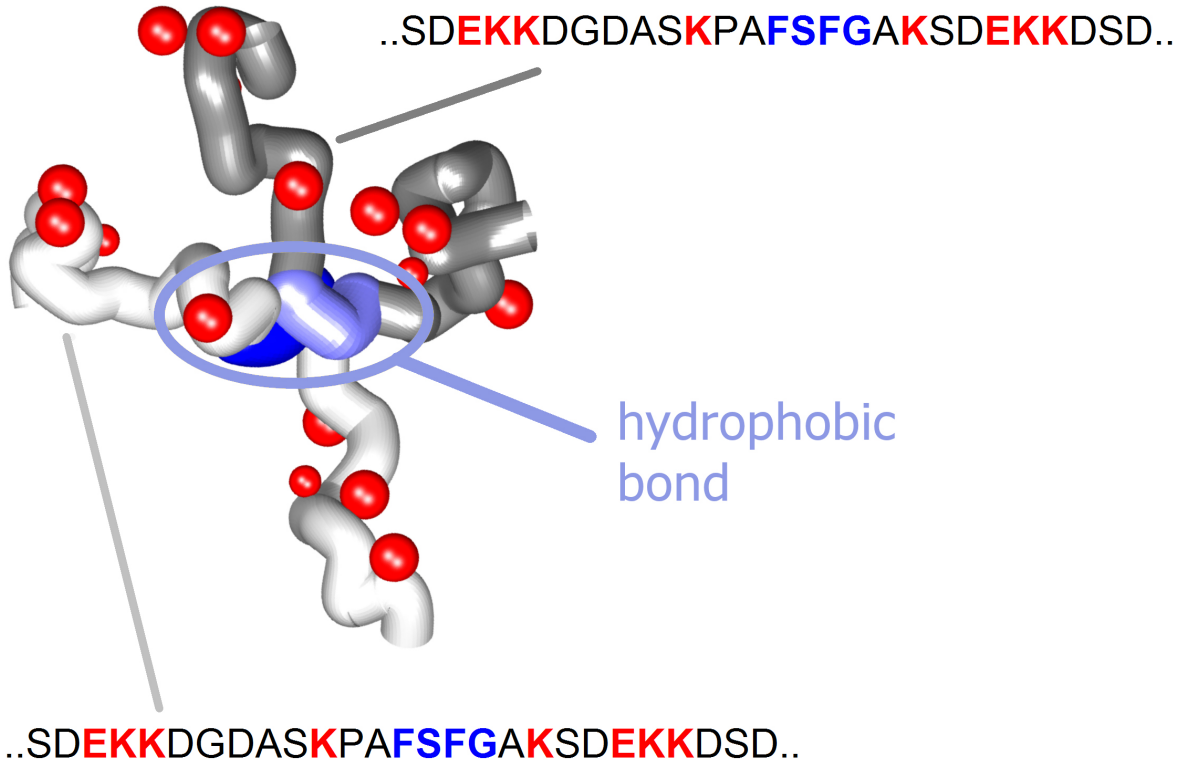


Figure 134: Two parts (colored white and grey) of the N-terminal fsFG-domain of the essential yeast nucleoporin Nsp1p coalesce a hydrophobic bond and thus form a *cluster*. The blue parts of the peptide backbone mark hydrophobic regions. The red spheres illustrate the positions of the side chains (that are at least medium sized). The corresponding amino acid sequences of the illustrated nucleoporin parts are shown, too, and have the same color code as in the cluster visualization.

9.2.2 Percolation Theoretical Methods

If a transition from a sol phase with only a few clusters to a gel phase with cross-linked peptides exists, increasing the mass density c of the system over some critical density c_{crit} will push the system from the colloidal sol state to a gel state [200]. To determine whether the system has such a phase transition from a sol to a gel state a model is

developed and then percolation theoretical methods [83; 200] will be applied to analyze the data.

Here, one has to examine the hydrophobic cross-linking between peptide chains. Two important properties in this context are the correlation length $\xi(c)$ and the percolation probability $P_\infty(c)$ of the system. In the case that the linear dimension of the system l is larger than the correlation length, ξ is a measure of the largest hole in the largest cluster and it decreases as c is increased above c_{crit}. P_∞ is the probability that a given peptide belongs to the largest cluster of the system.

Near the percolation threshold of infinitely large systems, i.e. c approaches c_{crit}, most percolation quantities obey *scaling laws* that are largely insensitive to the structure and the microscopic details of the system. For example the correlation length and P_∞ scale as

$$\xi \propto |c - c_{crit}|^{-\nu} \text{ and } P_\infty \propto (c - c_{crit})^{\beta}.$$

Where ν and β are critical exponents that are completely universal, i.e. they are independent of the microscopic details of the system and depend only on the dimensionality. Even long but finite range interactions do not change this universality, although they may change the value of c_{crit}.
The prefactors are non-universal (although certain ratios of them are universal again) and therefore not mentioned here. Although percolation is universal, c_{crit} is very difficult to measure and depends on the microscopic details of the system. The main question in this chapter is, does such a critical density and thus a phase transition exist at all?

For an infinite system P_∞ behaves like $P_\infty \propto (c - c_{crit})^{\beta}$ if the density approaches the critical value. As the mass density is decreased to the critical value P_∞ decreases to zero. For $c < c_{crit}$ P_∞ is zero. In a finite system, as c_{crit} is approached, ξ eventually becomes comparable to the linear size of the network. Therefore, the variation of any property X of a system of linear size L is given by

$$X \propto L^{-a} f(x),$$

where $x = L^{1/\nu}(c - c_{crit})$ and $f(0)$ is nonsingular [79; 200]. One can use this equation either to determine critical exponents or to estimate the critical density itself. The latter is done here.

Near c_{crit} and in the limit $L \to \infty$ one has $P_\infty \propto (c - c_{crit})^{\beta}$ [200]. This leads to $a = \frac{\beta}{\nu}$. Therefore, all different system sizes which have been simulated fulfil the condition

$$P_\infty L^{\beta/\nu} \propto f(L^{1/\nu}(c - c_{crit})). \tag{43}$$

This means variation of c_{crit} until all curves f for the different system sizes collapse to a master curve supplies the critical mass density at which the percolation threshold lies.

9.2.3 Description of the Nucleoporin Model

In the model we take into account all FG-rich hydrophobic clusters of the nucleoporin strands cf. App. F for details.
An individual Nsp1p peptide chain is shown in Fig. 135. Every single chain consists of 606 amino acids, has a mass of about 62kDa and consists of 37 hydrophobic clusters i.e. 110 hydrophobic amino acids.
The hydrophobicity is modeled similar to the hydrophobic-polar protein folding model [60]: If two hydrophobic parts of the peptide chains come closer than 0.7 nm the total energy of the system will be reduced by 0.04336 eV [11; 125].

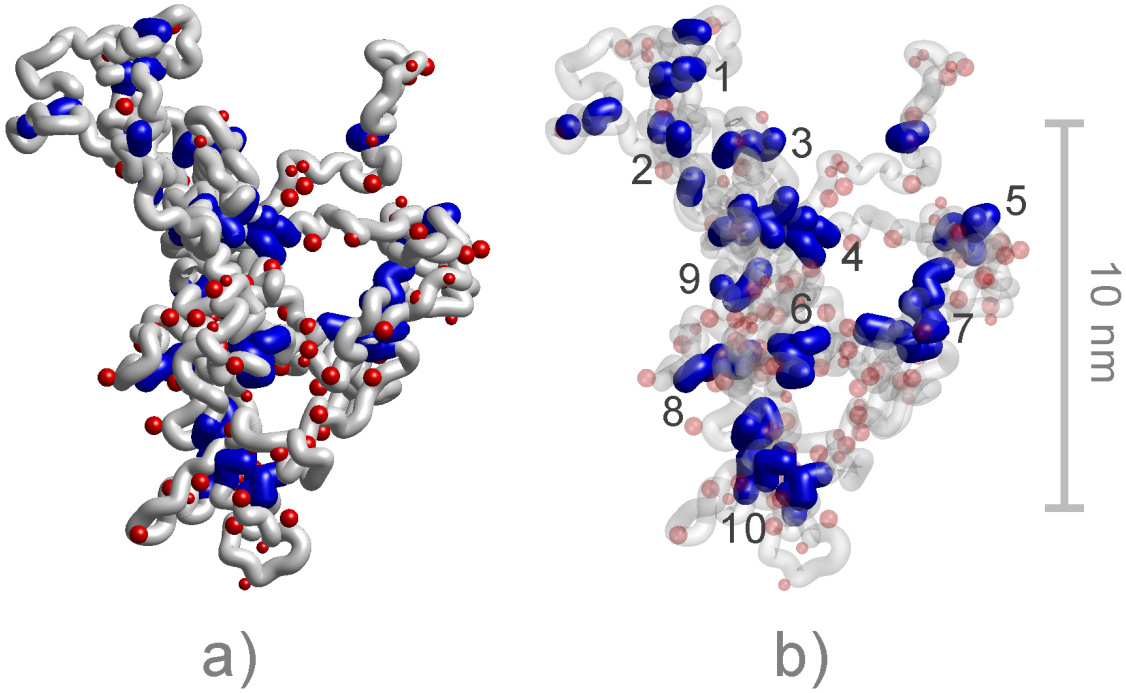


Figure 135: a) One conformation (out of the whole ensemble) of a single nucleoporin strand. The red spheres represent side chains of large and medium sized amino acids (starting with Glu / E). The FG-rich hydrophobic parts of the nucleoporin are marked blue. They have been enlarged by 30 percent for this visualization. b) Same as a) but transparent backbone and amino acid side chains to give a better view at the hydrophobic parts. One can see small hydrophobic clusters (numbered 1-10). Since there is no other peptide chain they are intra-bonds (cf. Fig. 141), i.e. bonds that the peptide coalesces with itself.

The excluded volume of every amino acid is factored in just like the side chains of all large and medium sized amino acids (starting at Glu / E). All smaller side chains were

neglected.
The peptide backbone is modeled as a tube with side chains and hard-core excluded volume potential.
Furthermore, all chains are trapped in a cubic box with side length l (cf. Fig. 137 and 136 for an illustration).

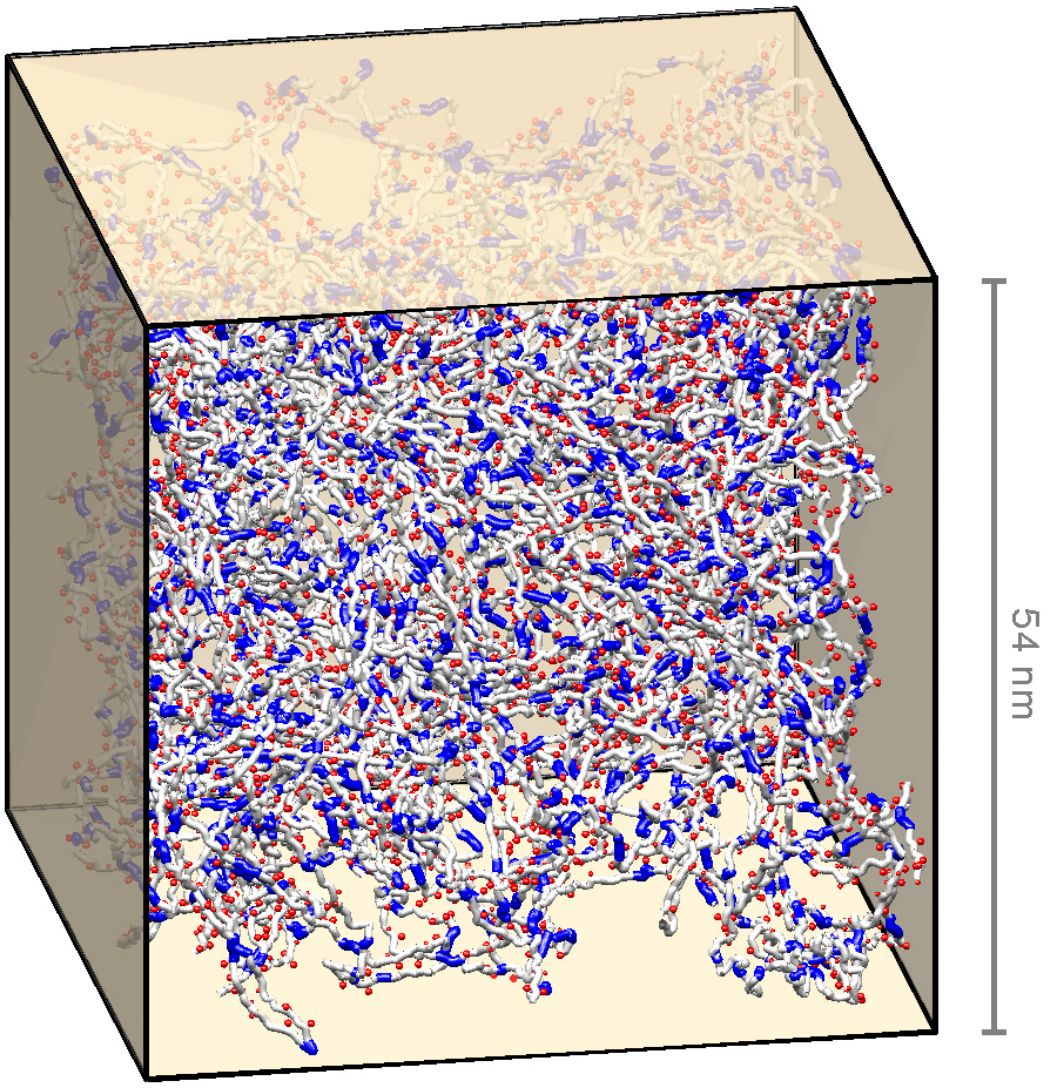


Figure 136: An example peptide system of $N = 40$ Nsp1p strands with density $c = 25$ mg/ml. There is one large cluster in the system that consists of 72.5% of all peptides. The rest of the peptides are free. The total binding energy is -213 eV ($-8.2\ 10^3$ kT).

The formation of hydrogen bonds as well as other hydrophobic or hydrophilic interac-

tions than that of the FG-repeats are neglected in this model.
Moreover, the model is not based on a underlying spacial lattice.

The rotational freedom (Φ and Ψ) of the peptide bonds between sp^3 and sp^2 hybridization of each amino acid is modeled by pivoting the nucleoporin strand with respect to the excluded volume potentials. Every amino acid contributes to the peptide flexibility with one joint of this kind.

A Metropolis-Hastings Monte-Carlo algorithm was used in order to generate many equilibrated configurations of the described nucleoporin model system (the N-terminal fsFG-domain of the essential yeast nucleoporin Nsp1p, cf. App. F) at various densities to analyze their connectivity properties.
The Metropolis algorithm is a rejection sampling algorithm (cf. App. E) used to generate a sequence of samples from a probability distribution that is difficult to sample from directly. It was used to generate equilibrated ensembles of peptide networks of a given density and temperature with Maxwell-Boltzmann distributed energy.
The start configuration for every peptide chain is a self-avoiding random walk. The algorithm uses pivot moves to alter the peptide chains during the iterations as described above. Randomly a number of peptide chains is chosen and pivoted (with respect to the excluded volume interactions). Then the energy difference ΔE between the new state after and the old state before the pivot moves is calculated. At the end of the step a random variable decides with respect to the Maxwell-Boltzmann distribution whether the new state is accepted or rejected. Hereby, the acceptance probability is

$$p_A = \min\left(1, \exp(-\frac{\Delta E}{kT})\right),$$

where k is the Boltzmann constant. This last step allows thermal fluctuations to break bonds and dissolve clusters with a certain probability. Thus one gets equilibrated ensembles of system configurations at a certain density c and a temperature T. The latter was fixed to $T = 300K$ (i.e. lab conditions).

Then one can average over such an ensemble to determine system properties like P_∞ or the radius of gyration (cf. Sec. 1.2.2) which is a measure for the extension of the system. The radius of gyration is defined by

$$R_g^2 = \frac{1}{M}\int d^3r\rho(\overrightarrow{r})|\overrightarrow{r} - \overrightarrow{r}_S|^2,$$

where ρ is the mass density, M the mass and $\overrightarrow{r}_S$ the center of mass. The radius of gyration will be used later in this chapter.

For every simulated system size (N = 10, 25, 40, 50 and 60 peptides) we generated an ensemble of at least 10^5 configurations. Some example configurations are shown in Fig. 135, Fig. 137 and Fig. 136 (as well as in Fig. 200, Fig. 201, Fig. 202, Fig. 203 and Fig. 204 in App. A). The red spheres represent amino acid side chains, whereas the blue tubes picture hydrophobic regions.

The box surrounding the peptide system has free boundary conditions and it's size l is determined by the density c.

Fig. 135 (as well as Fig. 201 in App. A) shows only a single simulated N-terminal Nsp1p fsFG-domain. One can see that some hydrophobic clusters are formed, whereas other hydrophobic parts remain unbound.

Three different example configurations of peptide systems with 40 peptides are shown in Fig. 137 (a larger version of this figure is available in App. A, Fig. 200). The configurations have different peptide mass densities. At a low density (at the left) only a few peptides coalesce bonds. But increasing the density leads to an increase of these bonds and thus an increase of peptide clusters in the system. The box on the right of Fig. 137 is enlarged in Fig. 136. Here one can already see a configuration where all peptides contribute to one large cluster. Hydrophobic bonds can again be dissolved by thermal fluctuations of the system.

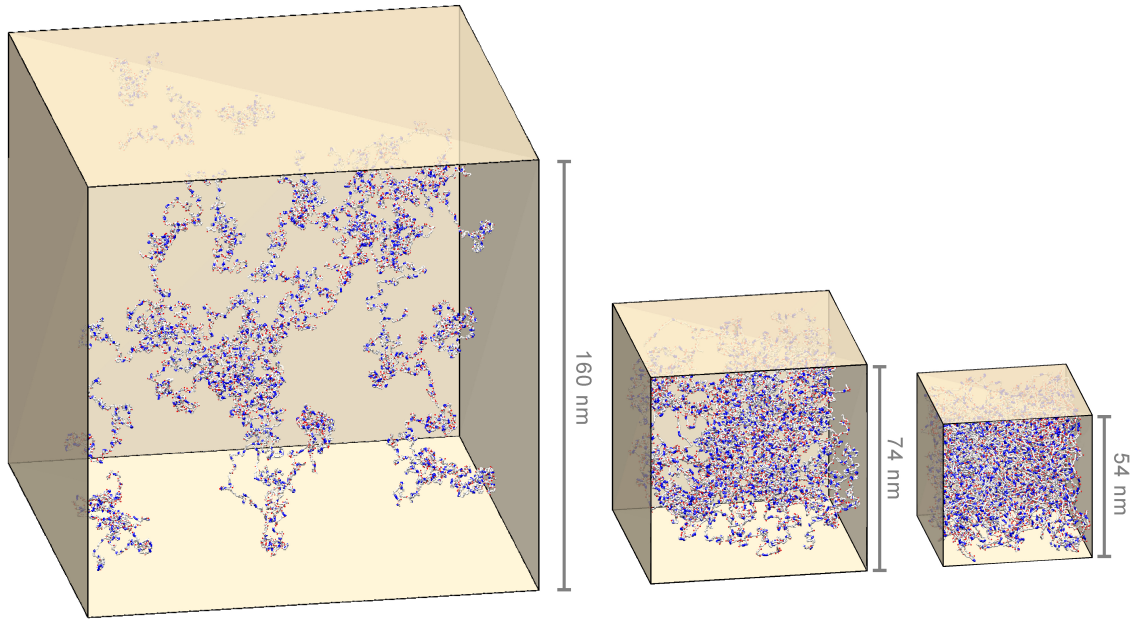


Figure 137: Three example configurations of a system with 40 peptides with different box sizes and thus different mass densities (1 mg/ml, 10 mg/ml and 25 mg/ml). The clustering of the peptide chains increases (the fraction of clustered peptides are 10%, 37.5% and 72.5% with decreasing density (resp. decreasing box volume). A larger version of this figure can be found in App. A (Fig. 200).

9.3 Results

Fig. 139 and 143 show the increase of clustering in the system: With increasing mass density the peptid network exceeds the percolation threshold and eventually all chains contribute to one large cluster. Finite size effects lead to a rounding of the phase transition [200]. This shows that the attractive cluster-forming hydrophobic interactions of the N-terminal Nsp1 fsFG-domains are strong enough to compete with the entropic forces by the excluded volume of the peptid strands. This may be a hint that indeed FG-rich hydrophobic parts of nucleoporins are key for the percolation and thus the formation of gels in the nuclear pore complex.

Fig. 140 shows the radius of gyration R_g of the largest cluster. For small densities R_g is simply the extension of a single peptide chain (about 10nm) because the chains interact very weakly with each other and therefore collapse by forming hydrophobic clusters. With increasing density the average largest cluster grows and thus the correlation length increases until one large cluster is formed. Eventually at high densities (i.e. near c_{crit}) the box volume limits extension of the cluster, and therefore, R_g decreases again as the box size l decreases with increasing density c (as $c \propto l^{-1/3}$).

The average number of hydrophobic bonds which an individual peptide chain coalesces with itself ("intra-bonds") to form hydrophobic clusters is relatively independent of the mass density (cf. Fig. 141). The average number of intra-bonds of a single Nsp1p peptide chain is of the order of 100. However, the average number of hydrophobic bonds which an arbitrary peptide chain coalesces with other peptide chains ("inter-bonds") rapidly increases with the mass density of the system and the system size (cf. Fig. 141).

The finite size scaling of P_∞ is shown in Fig. 142. c_{crit} was varied in integer intervals until the data collapse to the master curve was best. We used the sum of the squared distances between the curves for different system sizes as a measure for the extend of the data collapse and varied c_{crit} until it was minimal. The final result for the critical density was $c_{crit} = 42$ mg/ml.
The critical exponents used were: $\beta = 0.41$ and $\nu = 0.88$ i.e. the critical exponents of 3D percolation. Other critical exponents give a very bad fit.
One can see in Fig. 142 that the collapse to a master curve is not perfect but one has to keep in mind that P_∞ has strong fluctuations (cf. Fig. 139). The peptide systems were difficult to sample and we decided to simulate different system sizes instead of absolutely minimizing fluctuations of the measured properties for only one number of peptides. At the end, this is what makes the finite size scaling possible. The estimation of the critical density is a byproduct in this work. Our main goal was to answer whether there is a phase transition to a gel phase at all.

9.4 Conclusions

It was not a priori clear whether the examined peptide system would be able to form a gel by hydrophobic FG-rich clusters at biologically realistic densities. In the model

system of the Nsp1p fsFG-domain we found a gel-sol phase transition at a critical density of 42 mg/ml which is not far away from biologically realistic densities in nuclear pore systems: Experiments on Nsp1p systems show that a macroscopic gel is formed above densities of 8-10 mg/ml. The gel was identified in this case by visual inspection [87].

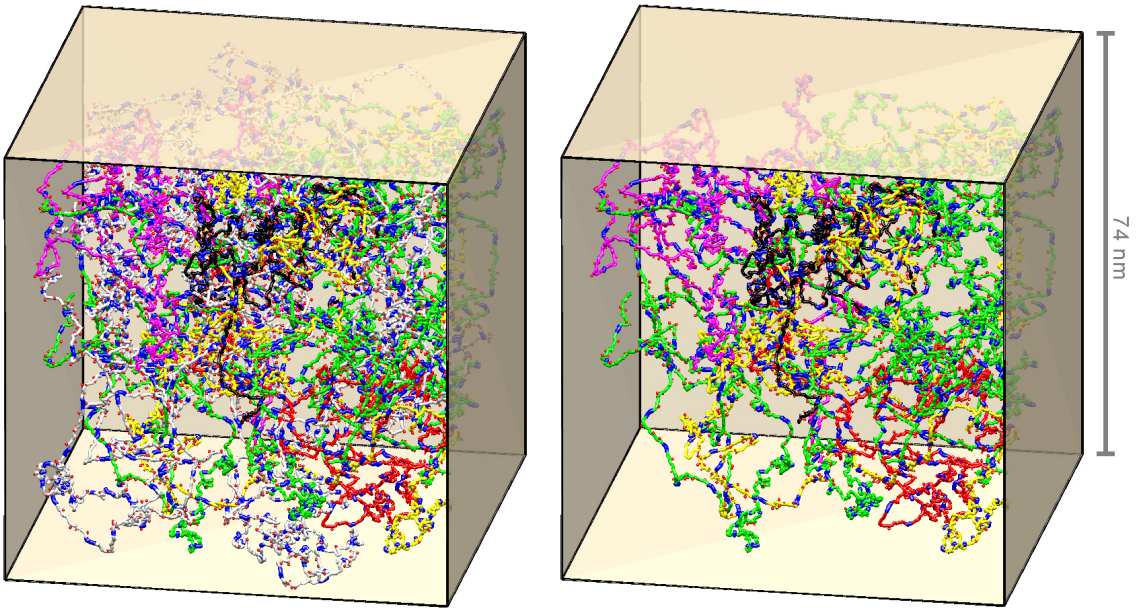


Figure 138: Example configuration of 40 peptides (N-terminal Nsp1p fsFG-domains cf. App. F) with colored clusters at a mass density of 10 mg/ml. Altogether, there are five clusters (colored green (12 peptides), yellow (5 peptides), red (2 peptides), black (2 peptides) and pink (3 peptides)) in this configuration. White peptides are single, unbound nucleoporins that have no hydrophobic bond with another peptide and thus do not contribute to any cluster. The box on the right shows the same configuration as on the left with the difference that the single unbound peptide chains are not shown. 60% of all nucleoporins in this configuration have hydrophobic bonds, i.e. they contribute to some of the five clusters. The other 40% are unbound. A larger version of this figure can be found in App. A (Fig. 202).

This shows that the attractive cluster-forming hydrophobic interactions of the N-terminal Nsp1 fsFG-domains are strong enough to compete with the entropic forces by the excluded volume of the peptid strands. This may be a hint that indeed FG-rich hydrophobic parts of nucleoporins are key for the percolation and thus the formation of gels in the nuclear pore complex.

We neglected hydrogen bonds in our model as well as hydrophilic interactions and fo-

cused instead on modeling the most dominant effects namely the FG-rich hydrophobic clusters. Hydrogen bonds would support attractive forces whereas hydrophilic parts would lead to an increase of the repulsion in the system and thus to more open structures. These opposite effects might slightly shift the critical density of the system. Exact predictions of critical densities in such complex systems are very difficult even in the easier case of Ising-models.

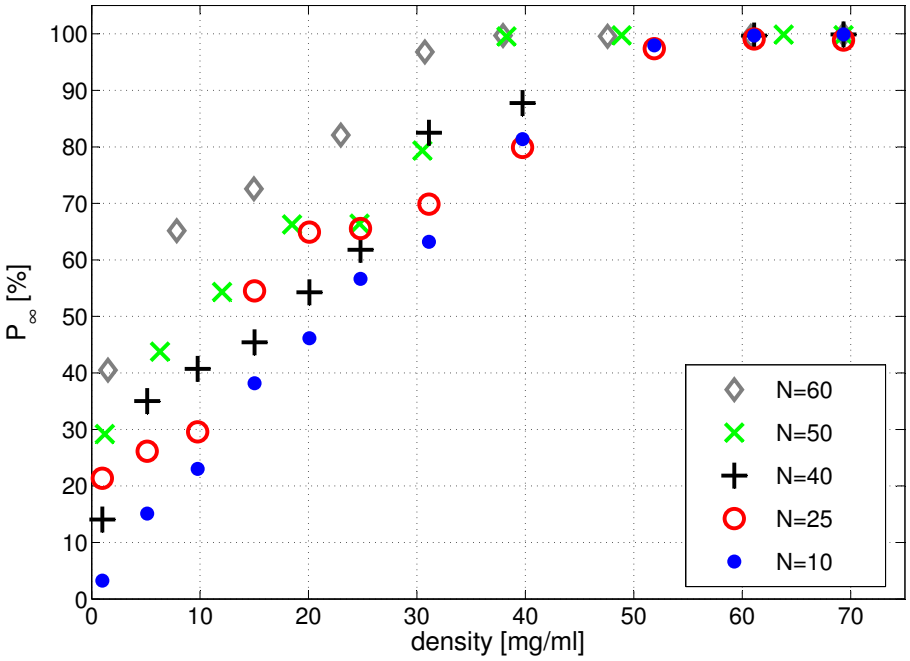


Figure 139: This figure shows the average percentage of the largest cluster in dependence on the mass density of the system. One can see that the largest cluster grows with increasing density until all peptides are linked. In this case the system gets macroscopic properties like a finite elasticity or shear modulus and thus a gel is formed.

The small number of inter-bonds (compared to the number of intra-bonds) may give the peptid network the possibility to open up easily which might be a biologically advantage to facilitate transport through the nuclear pore complex.

Although we already invested a huge amount of computer time, more simulations especially of larger systems would be nice. One could reduce the finite size effects and have a more detailed view at the system. But we would need one order of magnitude more computation time in order to achieve this goal which is not possible with current

machines. Nevertheless, the data presented is already sufficient to show that the Nsp1p peptide network is a percolating system and that the percolation threshold lies at the order of magnitude of 40 mg/ml which is close to experimental results [87].

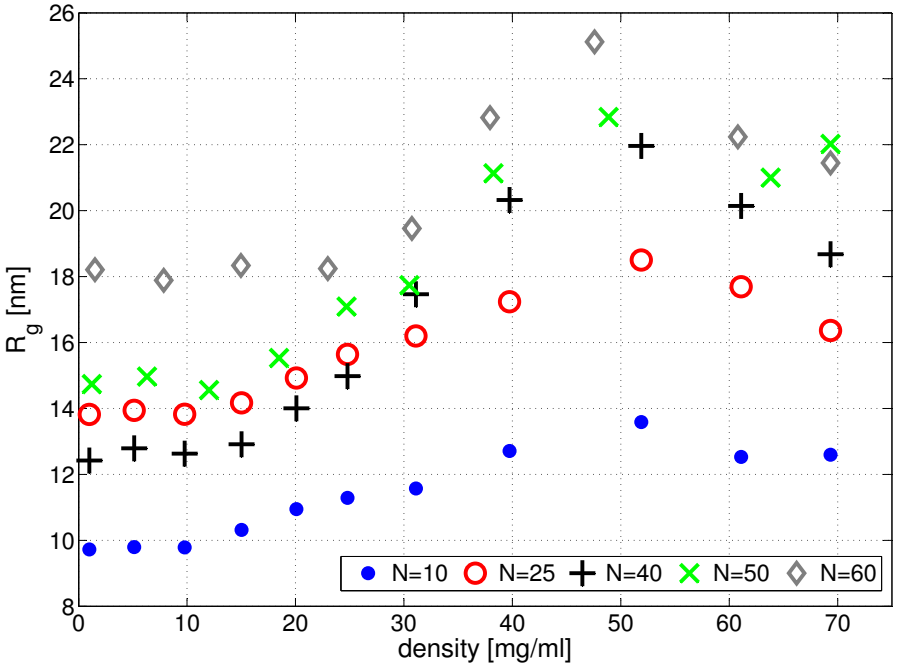


Figure 140: This figure shows the mean radius of gyration R_g of the largest cluster of the nucleoporin system as a measure for the system's correlation length in dependence on the mass density. For very small densities the correlation length is just the average extension of a single peptid (around 10 nm). With increasing density the largest cluster grows and thus the correlation length increases. In the case of large densities ($c > c_{crit}$) the cluster size is eventually limited by the box size l and thus R_g is about the linear dimension of the bounding box which decreases with increasing density c ($c \propto l^{-1/3}$).

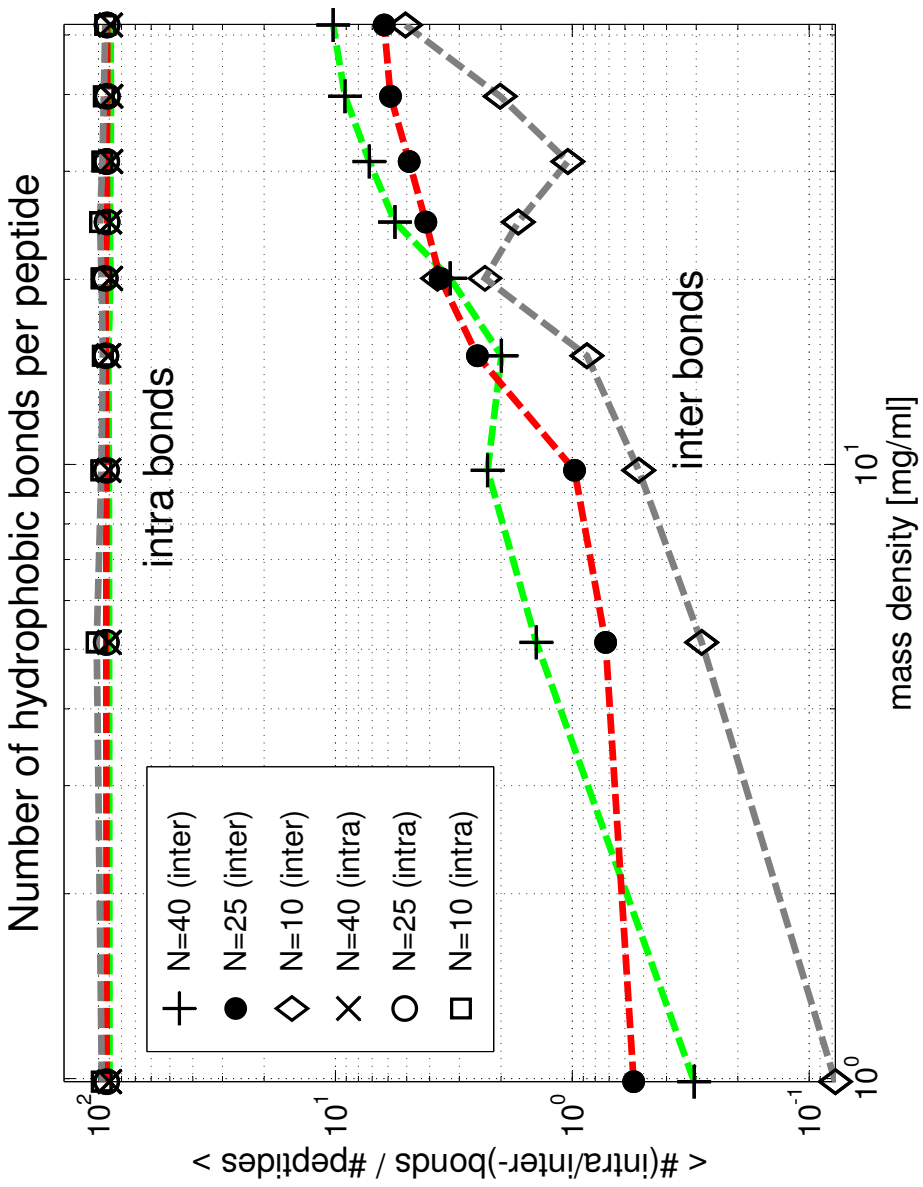


Figure 141: This Figure shows the average number of bonds between different peptides (inter) and the average number of bonds an individual peptide coalesces with itself for different system sizes: N denotes the number of peptides of the system. One can clearly see that the number of inter-bonds increases with the mass density whereas the number of intra-bonds is density independent.

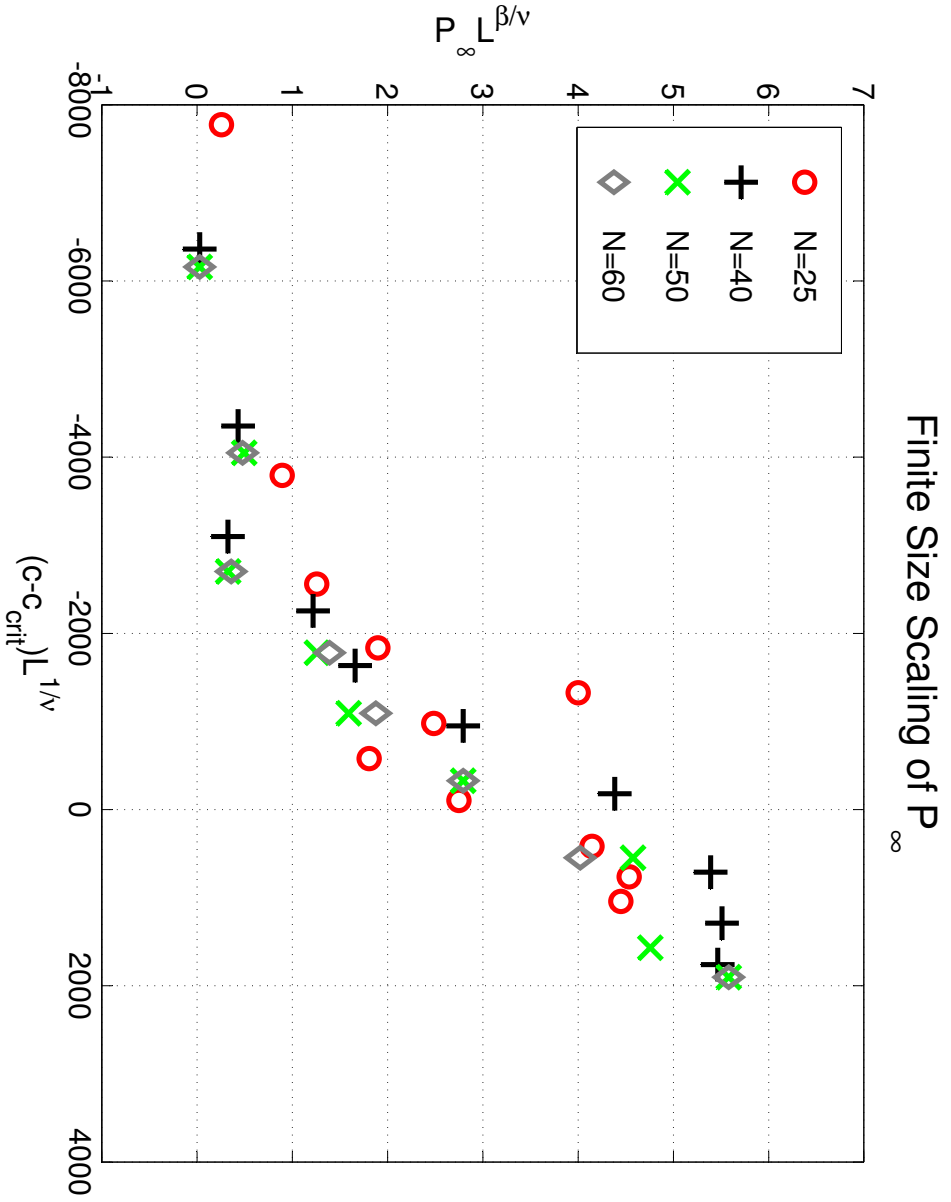


Figure 142: This figure shows the finite size scaling of P_∞ which was used to determine the critical density c_{crit}. The latter was considered as a parameter here and varied in integer steps of 1 mg/ml until the curves for different system size fit best upon one another (cf. Eq. 43). The so found critical density is $c_{crit} = 42$ mg/ml. Our focus on the FG-rich hydrophobic clusters and thus the neglection of hydrogen bonds and other hydrophobic or hydrophilic interactions leave a margin for this value but in principle the order of magnitude should stay and we proved that gelation is possible. The critical exponents used here were: $\beta = 0.41$ and $\nu = 0.88$.

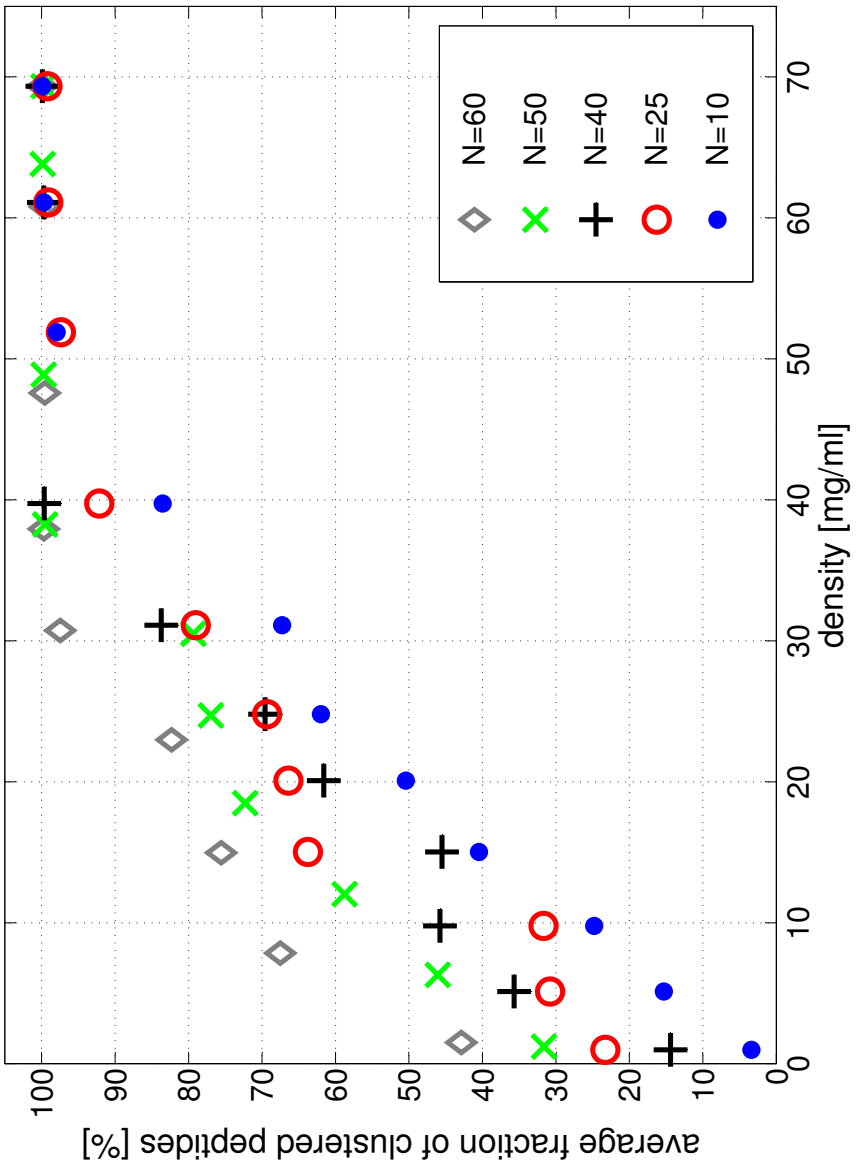


Figure 143: This figure shows the average fraction of clustered peptides in dependence on the mass density of the system. With increasing mass density at some point all peptides contribute to one large cluster and thus a gel is formed.

10 The Mean Average Crossing Number

In this section it will be described how an analytical polymer physics tool, namely the mean Average Crossing Number (mACN), has been developed further during my PhD.

Roughly speaking the mACN is a measure for the entanglement of polymer chains.

One basic motivation was to use the mACN to analyze two-dimensional experimental chromatin projections with respect to their loop content (cf. Sec. 10.6) and then use these results to improve the random loop model [25].

First, an overview over some possible applications of the mACN will be given in Sec. 10.1. After that Sec. 10.2 provides some basic analytical facts about the mean Average Crossing Number together with a description of the simulation methods that were applied in the numerical part of this chapter.

In Sec. 10.3 the influence of excluded volume on the mACN will be examined. This was not done before and it will be shown here that several approximations that are crucial for many mathematical proofs of properties of the mACN still apply in the case of chains with excluded volume.

Moreover, it turned out that the excluded volume interactions have a strong impact on the behavior of the local crossing number of two chain segments at very small distances but only a weak one at large distances.

Furthermore, it will be shown that the data is compatible with a $N\ln(N)$-behavior for the mACN, even in the case with excluded volume.

After that the connection between loops in polymer conformations and the mACN will be investigated for different kinds of random walks (cf. Sec. 10.4) and it will turn out that there is a strong correlation between these two variables. This makes the mACN a potential tool for the estimation of loop numbers in polymer systems and it is the basic motivation of Sec. 10.6 where a short outline of an idea for the estimation of the loop number of two-dimensional experimental data is presented.

In Sec. 10.5 an analytical connection between the mACN and the two-point correlation function $\Phi(\vec{r}, \vec{r'})$ and thus the scattering function $S(\vec{q})$ will be presented.

The major parts of Sec. 10.3.1 are results from my diploma thesis [61]. During my PhD I made further simulations concerning the influence of excluded volume effects on the mACN and the results have been published [63]. The remaining results of this chapter are currently prepared for a further publication [68].

10.1 Introduction

10.1.1 Polymer Entanglement

Random walks are frequently used to model the behavior of polymers at thermodynamic equilibrium [48; 70; 83].
The simplest but also the most fundamental type of random walks is represented by chains composed of freely jointed segments of equal length (equilateral) where the individual segments have no thickness. Such random walks are known as ideal random walks and are used to model the behavior of polymers under so-called theta conditions where polymer segments neither attract nor repel each other [163].

The behavior of ideal random walks is thoroughly researched and it is well established, for example, that such measures of overall dimensions of ideal random walks like the average end-to-end distance or the average radius of gyration scale with the number of segments N as N^{ν} where $\nu = 0.5$ [48; 70; 83] (cf. Sec. 1).
Although the overall dimensions provide important information about the modelled polymers, frequently additional characteristics of polymers are of interest. Among these characteristics are some that can be used to measure the entanglement of the polymers like the mean Average Crossing Number.

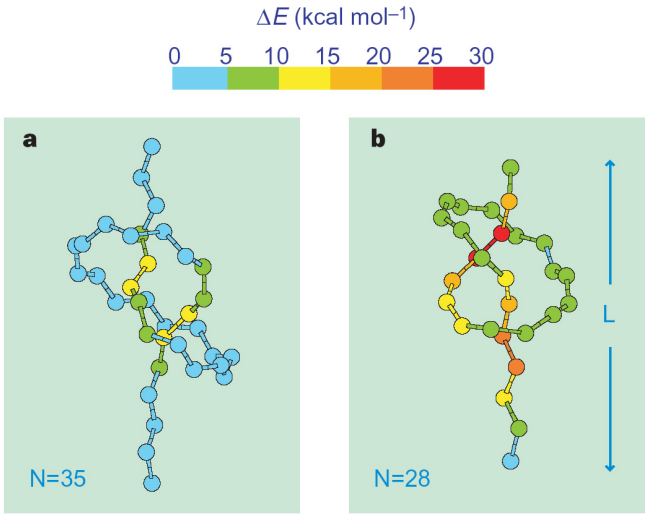


Figure 144: Strain energy distribution in a polymer strand with a trefoil knot. Shown are distributions in chains of 35 (a) and 28 (b) carbon atoms taken from constrained classical MD simulations. When the knot is sufficiently tightened the strain energy localizes most on the bonds immediately outside its entrance points. Adapted from [177].

The influence of knots on the properties of polymers has become of great interest, in part because of their influence on mechanical properties [14].

Knot theory applied to the topology of macromolecules indicates that the simple trefoil knot is likely to be present in any long polymer strand [10; 85; 89; 112]. Many experiments have been done to determine the relative strengths of different knots and these show that the break in a knotted rope or polymer almost invariably occurs at the point just outside the entrance of the knot [177; 220] (cf. Fig. 144).

In the case of circular polymers their knot type can be rigorously determined [1; 121]. However, the information about the knot type gives only the minimal crossing number that one could see upon elimination of all nugatory crossings. The determination of the knot type does not tell us, for example, how many nugatory crossings there are in the analyzed polymer conformations. In addition, in the case of linear polymers it is difficult to define their knottedness [142; 208] and the knot type is only a mathematical property that describes the topology of a polymer and not the polymer conformation itself.

There are no difficulties, however, with defining a directional crossing number that corresponds to the perceived number of crossings that can be observed on an orthogonal projection of a given non perturbed trajectory in open or circular form. If a given conformation or of a random walk is orthogonally projected onto a plane along a given direction, one can count the number of crossings that are visible in this particular projection of the conformation.

To be independent of the choice of a particular projection the average crossing number (ACN) is defined as the average of directional crossing numbers over all possible orthogonal projections (which is equivalent to the unit sphere) of a given conformation. Furthermore, to be independent of the particular conformation it will be averaged over all possible polymer conformations as well. Thus, one gets the *mean* average crossing number (mACN).
For the first mean value one can imagine a sphere which contains the whole polymer chain (resp. random walk). Then the tangential planes of this sphere are the possible projection planes for this particular conformation.
The second mean value is the average over all polymers (or random walks) of length N.

In contrast to the minimum crossing number the average crossing number is a more natural geometric measure of polymer entanglement as it refers to the actual number of crossings that can be perceived while observing a non-perturbed trajectory of a given polymer. Of course, the mACN corresponds also to the time averaged crossing number of randomly fluctuating polymers in solution.

10.1.2 The mean Average Crossing Number

The mean average crossing number (mACN) was introduced in [153; 220] as a simplified version of the writhe of a random walk.
For a given linear closed polymer conformation the crossing number associated with

a particular projection of the random walk is the number of crossings one observes when the polymer is projected to a plane under the given projection direction (cf. Sec. 10.1.1). The *average* crossing number (ACN) of the polymer is then defined as the average of this crossing number over all possible projection directions [56]. The *mean* average crossing number (mACN) is the average of the ACN over all possible random walks of a certain length N.

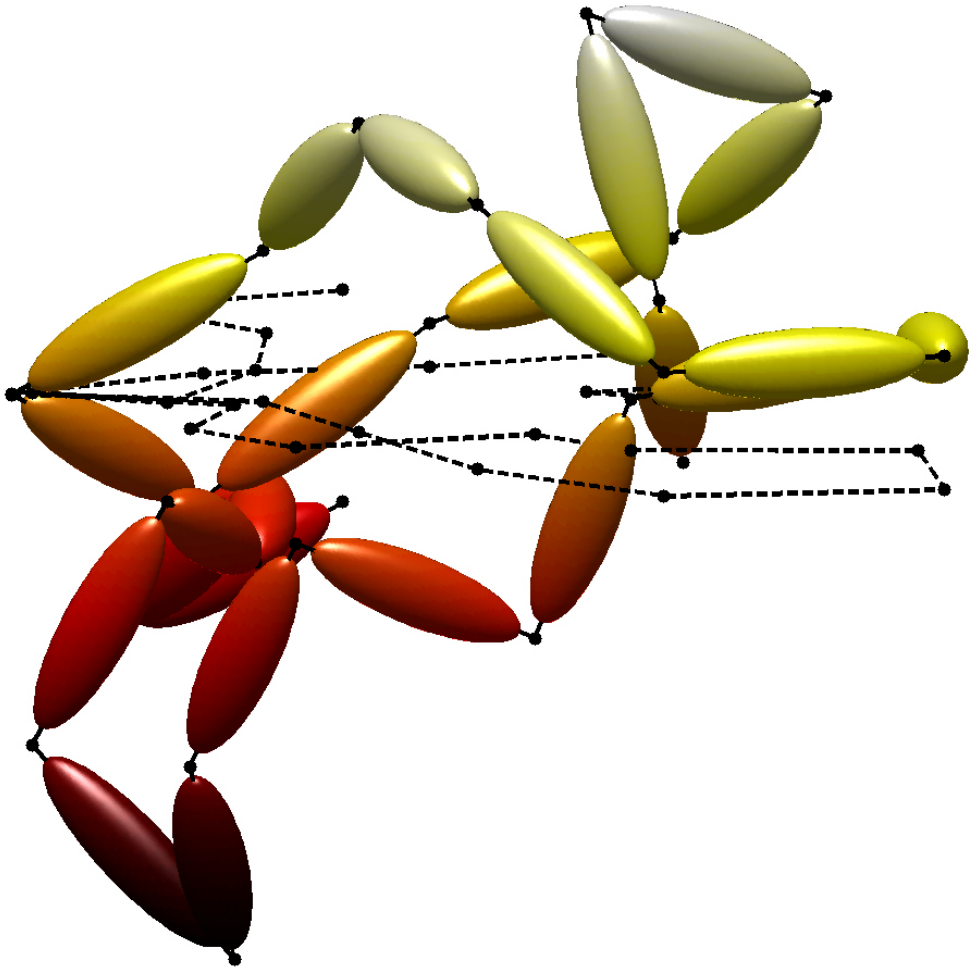


Figure 145: An example for a Gaussian chain of length $N = 20$ with excluded volume. The dashed line is the projection of the chain onto the xy-plane and shows only one crossing. The average crossing number of this particular chain is ACN= 1.61 and the mean average crossing number of all Gaussian chains of length 20 is ≈1.96. The color corresponds to the height of the ellipsoids.

Diao suggested to use the mACN which was called 'entanglement complexity' in a previous work [117], as a measure of entanglement to determine whether a (closed) polymer chain is highly or weakly knotted [55].
Grassberger [96] showed that the mACN is mainly a measure for the 'opacity' of a

random walk.
For equilateral and for Gaussian random walks without excluded volume Diao [56; 57] succeeded in showing that the mean average crossing number behaves $c_1 N \ln N + c_2 N$. Furthermore, in [153; 220] it was shown that the mACN satisfies mACN $\propto N^{\alpha}$ for $N \to \infty$ with $\alpha \approx 1.2$ [6].

10.2 Basic Properties

In this section some basic mathematical properties of the mean average crossing number will be given. The proofs of the following theorems can be found in [56; 57; 61].

10.2.1 Basic Definitions and Lemmata

10.1 Definition

1. A *Gaussian Random vector* $\overrightarrow{v} = (x, y, z)$ is a random point with coordinates x,y and z which are independent standard normal variables (with mean = 0 and variance= 1). The probability density function (pdf) of $\overrightarrow{v}$ is the joint pdf of x, y and z which is

$$\begin{aligned} p(\overrightarrow{v}) &= \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\right)\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}\right)\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\right) \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^3 e^{-\frac{x^2+y^2+z^2}{2}} \\ &= (\frac{1}{\sqrt{2\pi}})^3 \cdot e^{-\frac{|\overrightarrow{v}|^2}{2}}. \end{aligned}$$

2. *A Gaussian random walk (or chain)* of n steps consists of $n+1$ points $X_{\{0\,\ldots\,n\}}$ such that $X_{k+1} - X_k$ $(k = 0,\ 1,\ \ldots,\ n-1)$ are Gaussian random vectors and $X_0 = (0,0,0)^t$. The joint pdf for all the vertices is

$$p(X_1, X_2, ..., X_n) = \left(\frac{1}{\sqrt{2\pi}}\right)^{3n} e^{-\frac{1}{2}(|X_1|^2 + |X_2 - X_1|^2 + \ldots + |X_n - X_{n-1}|^2)}.$$

10.2 Lemma
Let $\overrightarrow{v} = (x, y, z)^t$ be a given Gaussian random vector then x, y, and z are independent variables and the spherical coordinates (r, θ, ϕ) (or spherical random variables) of $\overrightarrow{v}$ are independent, too. Furthermore, one can show that the probability density functions of r, θ and ϕ are given by

$$\begin{aligned} p_1(r) &= \sqrt{\frac{2}{\pi}} r^2 e^{-\frac{r^2}{2}} \\ p_2(\theta) &= \frac{1}{2}\sin(\theta), \text{ and} \\ p_3(\phi) &= \frac{1}{2\pi}. \end{aligned}$$

The proof of Lemma 10.2 is straightforward and can be found in [122] for instance.

10.3 Lemma

1. *The pdf of the kth chain point X_k of a Gaussian random walk is given by*

$$g_k(X_k) = \int\int \ ... \ \int\int p(X_1)p(X_2 - X_1) \ ... \ p(X_k - X_k - 1) \ dX_1 \ ... \ dX_{k-1},$$

which leads to

$$g_k(X_k) = \left(\frac{1}{\sqrt{2\pi k}}\right)^3 e^{-\frac{|X_k|^2}{2k}}. \tag{44}$$

2. *As a consequence of this the pdf of the random vector $X_{j+k} - X_j$ (where X_{j+k} and X_j are vertices from the same Gaussian random walk of length n) is given by*

$$\begin{aligned} g_k(X_{j+k} - X_k) = &\int\int \ ... \ \int p(X_{j+1} - X_j)p(X_{j+2} - X_{j+1}) \ ... \\ &... \ p(X_{j+k} - X_{j+k-1} \ dX_{j+1}dX_{j+2} \ ...dX_{j+k-1} \\ = &\left(\frac{1}{\sqrt{2\pi k}}\right)^3 e^{-\frac{|X_{j+k}-X_j|^2}{2k}}. \end{aligned}$$

The proof of Lemma 10.3 can be obtained easily by using [95].

10.4 Definition

1. A vector $U = (u, v, w)^t$ will be called *uniformly distributed on the unit sphere* S^2, if the probability density function of U is

$$\varphi(U) = \begin{cases} \frac{1}{4\pi} & \Leftarrow |U| = 1 \\ 0 & \text{otherwise} \end{cases}.$$

2. If $U_1, \ U_2, \ ..., U_n$ are n independent random vectors uniformly distributed on S^2, then an *equilateral random walk (or chain)* will be defined as the sequence of the points $X_0 = O, \ X_k = \sum_{i=1}^{k} U_i, \quad k = 1, 2, ..., n$ in $\mathbb{R}^3$. Each X_k is called a *vertex* and the line segment joining X_k and X_{k+1} is called an *edge* of the equilateral random walk.

10.5 Lemma

1. *Let U be a uniformly distributed random vector on the unit sphere. Then the spherical coordinates (θ, ϕ) of U are independent random variables with the following probability density functions:*

$$\begin{aligned} p(\phi) &= \frac{1}{2\pi} \\ p(\theta) &= \frac{1}{2}\sin(\theta). \end{aligned}$$

2. *The joint probability density function $p(X_1, X_2, ..., X_n)$ of the vertices of an equilateral random walk is*

$$\begin{aligned} p(X_1, X_2, ..., X_n) &= \varphi(U_1)\varphi(U_2)...\varphi(U_n) \\ &= \varphi(X_1 - X_0)\varphi(X_2 - X_1)...\varphi(X_n - X_{n-1}). \end{aligned}$$

Let X_k be the kth vertex of an equilateral random walk $(1 < k \leq n)$. Then its density function is given by

$$p_k(X_k) = \int\int \ ... \ \int \varphi(X_1 - X_0)\varphi(X_2 - X_1)...\varphi(X_k - X_{k-1})dX_1dX_2...dX_{k-1}$$

and it has the closed form

$$p_k(X_k) = \frac{1}{2\pi^2 r}\int_0^\infty x \sin rx \left(\frac{\sin x}{x}\right)^k dx. \tag{45}$$

3. *The following estimation shows that for large values of k $p_k(X_k)$ is approximately Gaussian:*

$$\left| f_k(X_k) - \left(\sqrt{\frac{3}{2\pi k}}\right)^3 \exp\left(-\frac{3|X_k|^2}{2k}\right)\right| < \frac{1}{2k^{2.5}} \quad \text{for } k \geq 10. \tag{46}$$

The proof of the first point can be done easily by a transformation of the unit sphere into spherical coordinates or can be found in [122]. The proofs of the second and third point can be found in [55; 58; 59; 168].

Note that the probability density function $p_k(X_k)$ depends only on $|X_k|$ so it can be written as $p_k(|X_k|)$. Furthermore, if one defines $r_k = |X_k|$ (which is also a random variable), the probability density function of r_k is $4\pi r_k^2 p_k(r_k)$. It follows that the mean diameter of an equilateral walk of length n is of order $\sqrt{n}$.

10.2.2 ACN of Two Non-Consecutive Line Segments

The following theorem will give the ACN of two line segments within a *Gaussian* random walk. The behavior of the ACN of two line segments of a equilateral walk will be given in theorem 10.7.

10.6 Theorem (ACN of Two Segments of a Gaussian Chain)
Let P and Q be two fixed points in $\mathbb{R}^3$

1. *If P_1 and Q_1 are also* fixed *points in $\mathbb{R}^3$ and $l_1 := PP_1$, $l_2 := QQ_2$ are the two corresponding line segments, then the average crossing number $a(l_1, l_2)$ is given by*

$$a(l_1, l_2) = \frac{1}{2\pi}\int_t\int_s \frac{\dot{\gamma}_1(t) \cdot (\dot{\gamma}_2(s) \times (\gamma_1(t) - \gamma_2(s)))}{\|\gamma_1(t) - \gamma_2(s)\|^3} \, dtds \tag{47}$$

where γ_1 and γ_2 are the arclength parameterizations of l_1 and l_2.

2. *Now let P_1 and Q_1 be* random variables *such that $P_1 - P$ and $Q_1 - Q$ are two independent Gaussian random vectors. Then the expected value of $a(l_1, l_2)$ is given by*

$$\mathbb{E}(a(l_1, l_2)) = \mathbb{E}(a(d)) = \frac{1}{2\pi d^2} + O\left(\frac{1}{d^{2.5}}\right) \tag{48}$$

for large d (means $\sqrt{d} > \|\gamma_1\|, \|\gamma_2\|$) with $d = \|P - Q\|$.

10.7 Theorem (ACN of Two Segments of an Equilateral Chain)
Let P and Q be two fixed points in $\mathbb{R}^3$

1. *If P_1 and Q_1 are also* fixed *points in $\mathbb{R}^3$ and $l_1 := PP_1$, $l_2 := QQ_2$ are the two corresponding line segments, then the average crossing number $a(l_1, l_2)$ is given by*

$$a(l_1, l_2) = \frac{1}{2\pi} \int_t \int_s \frac{\dot{\gamma}_1(t) \cdot (\dot{\gamma}_2(s) \times (\gamma_1(t) - \gamma_2(s)))}{\|\gamma_1(t) - \gamma_2(s)\|^3} \, dtds \tag{49}$$

where γ_1 and γ_2 are the arclength parameterizations of l_1 and l_2

2. *Now let $d = |P - Q|$ fulfil $d > 4$ and P_1 and Q_1 be* random variables *such that $U_1 = P_1 - P$ and $U_2 = Q_1 - Q$ are two random vectors which are uniformly distributed on the unit sphere S^2. Then the expected value of $a(l_1, l_2)$ is given by*

$$\mathbb{E}(a(l_1, l_2)) = \mathbb{E}(a(d)) = \frac{1}{16d^2} + O\left(\frac{1}{d^3}\right). \tag{50}$$

10.2.3 mACN of Gaussian and Equilateral Random Walks

In this section the means of the crossing number of a random walk over all possible projections will be called the average crossing number and the means over all projections *and* all possible chain confirmations will be called the *mean* average crossing number.

10.8 Theorem
Let $\bar{c}_n$ be the average crossing number of a Gaussian random walk of n steps. Then the mean average crossing number is

$$\mathbb{E}(\bar{c}_n) = \frac{1}{2\pi} n \log(n) + O(n). \tag{51}$$

10.9 Theorem
Let $\bar{c}_n$ be the average crossing number of an equilateral random walk of n steps. Then the mean average crossing number is

$$\mathbb{E}(\bar{c}_n) = \frac{3}{16} n \log(n) + O(n). \tag{52}$$

10.2.4 Simulation Methods

The average crossing number of a random walk can be calculated by using the Gauss formula

$$a(l_1, l_2) = \frac{1}{2\pi} \int_t \int_s \frac{|\dot{\gamma}_1(t), \dot{\gamma}_2(s), \gamma_1(t) - \gamma_2(s)|}{\|\gamma_1(t) - \gamma_2(s)\|^3} \, dtds,$$

where γ is the arclength parametrization of the random walk. But using this formula leads to problems when some non-consecutive segments get close to each other.
To avoid these problems the average crossing number was calculated numerically by counting the crossings in numerous projections of γ and taking the average of all these crossing numbers. For each ACN calculation (at least) 1000 projections onto randomly chosen planes have been made to get a good approximation of the actual ACN value. Unfortunately, the crossing number depends discontinuously on the projection plane, and therefore, a large amount of projections is needed.

Furthermore, the number of possible random walk conformations increases dramatically with the length N of the walk. Therefore, a large sample has to be obtained for every considered chain length. In most of the cases this is the limiting factor for the chain length N due to the limited computation time.

10.3 Excluded Volume Effects

The main results of this section were obtained during former studies [61]. The contents of this section has already been published [63].

In this section the influence of excluded volume interactions on the behavior of the mean average crossing number (mACN) for random off-lattice walks will be studied. Gaussian and equilateral off-lattice random walks with and without ellipsoidal excluded volume up to chain lengths of $N = 1500$ will be investigated as well as equilateral random walks on a cubic lattice up to $N = 20000$.
It will turn out that the excluded volume interactions have a strong influence on the behavior of the local crossing number $\langle a(l_1, l_2) \rangle$ (cf. Sec. 10.2.2) at very short distances but only a weak one at large distances. This behavior is the basis of the proof in [57; 56] for the dependence of the mean average crossing number on the chain length N. It will be shown that the data is compatible with an $N\ln(N)$-behavior for the mACN, even in the case with excluded volume.

We start by investigating the effects of excluded volume interactions on the mean average crossing number (mACN) of equilateral and Gaussian random walks which

represent polymer classes.
A hard-core ellipsoidal potential was used to simulate the excluded volume interactions of the chain segments since many polymers can be modelled by such a potential (cf. Fig. 145).

To understand the effect of the excluded volume interaction, let us focus our attention initially on the scale invariant

$$a(l_1, l_2) = \frac{1}{2\pi} \int_{\gamma_1} \int_{\gamma_2} \frac{|\dot{\gamma}_1(t), \dot{\gamma}_2(s), \gamma_1(t) - \gamma_2(s)|}{\|\gamma_1(t) - \gamma_2(s)\|^3} \, dt \, ds \tag{53}$$

which is the basis for the prediction [56; 57]

$$\begin{aligned} \mathrm{mACN}_{\mathrm{Gaussian}}(N) &= \tfrac{1}{2\pi} N \, ln(N) + O(N) && (54)\\ &\approx \tfrac{1}{2\pi} N \, ln(N) + c_1 \cdot N && (55)\\ \mathrm{mACN}_{\mathrm{equilateral}}(N) &= \tfrac{3}{16} N \, ln(N) + O(N) && (56)\\ &\approx \tfrac{3}{16} N \, ln(N) + c_2 \cdot N && (57) \end{aligned}$$

for the phantom and equilateral chains of length N without excluded volume. In the equations above l_1 and l_2 are segments of the chain and γ is the arc-length parametrization (the above invariant can be considered as the first element in a hierarchy [98; 232]). In [56; 57] it was shown that for two chain segments l_1, l_2 on average $a(l_1, l_2)$ behaves as

$$\langle a(l_1, l_2) \rangle = \frac{1}{2\pi d^2} + O\left(\frac{1}{d^{2.5}}\right) \tag{58}$$

for Gaussian phantom and

$$\langle a(l_1, l_2) \rangle = \frac{1}{16 d^2} + O\left(\frac{1}{d^3}\right) \tag{59}$$

for equilateral phantom chains where d is the (fixed) distance between the two considered chain segments l_1 and l_2 and $a(l_1, l_2)$ is averaged over all possible orientations of l_1 and l_2 (cf. Sec. 10.2.2). So far, no prediction has been derived for the case of chains with excluded volume interaction. Hence it is important to ask how far the estimate also applies to the case of excluded volume interaction and how differences manifest themselves.

Using Monte Carlo simulations [24; 101] the average crossing number was calculated by counting the crossings in numerous projections of γ and taking the average over all these crossing numbers. For every calculated average crossing number it was averaged over 1000 randomly chosen planes to obtain a good estimate of the actual ACN value.

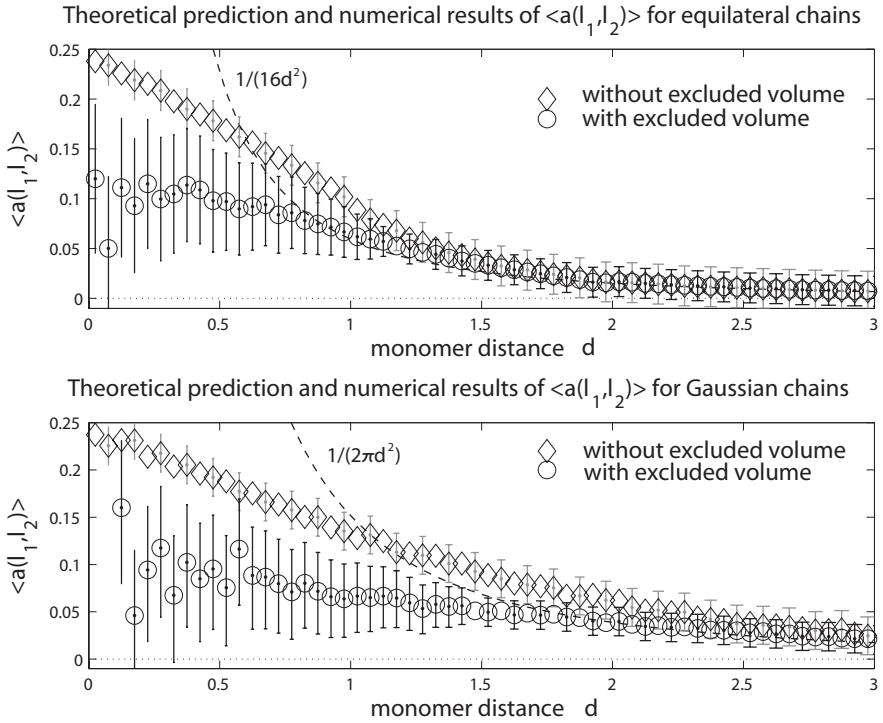


Figure 146: The leading order term for the theoretical prediction of the large d case [56; 57] and the numerical results for $\langle a(l_1, l_2)\rangle$ of Gaussian and equilateral chains for small values of d. One can roughly see that $\langle a(l_1, l_2)\rangle$ is discontinuous at $d = 0$. Every data point represents an average over at least 10^4 values.

Four types of chains have been investigated: Gaussian and equilateral chains with and without excluded volume. All chains are open and start at the origin. The excluded volume chains were generated with an (off-lattice) Pivot-Algorithm which for example can be found in [115; 139] (cf. Sec. D for details). A hard-core excluded volume potential was used in order to speed up the simulations. Between two consecutive chain points there is an ellipsoidal hard-core excluded volume (cf. Figure 145). When generating chains with excluded volume interactions the Pivot-Algorithm simply rejects all chain conformations which have at least two overlapping ellipsoids.

The chains without excluded volume are off-lattice random walks, since a lattice would make it problematic to determine the behavior of the average crossing number because only orthogonal orientations of the line segments would be allowed. The equilateral random walks have a fixed step length of 1 and the orientation of the next line segment is randomly chosen by using the given probability distribution of Lemma 10.5. A consecutive point X_k in a Gaussian random walk was determined by using a Gaussian probability distribution (cf. Lemma 10.2) for the cartesian coordinates $(x_k, y_k, z_k)^t$ of this point and adding the last vertex X_{k-1}.

Roughly speaking the mACN depends on two quantities: The behavior of $\langle a(l_1, l_2)\rangle$ in dependance on d and the probability distribution (pdf) of d. To understand the effects on the mACN due to excluded volume it is necessary to understand how excluded volume interactions affect these two parameters.

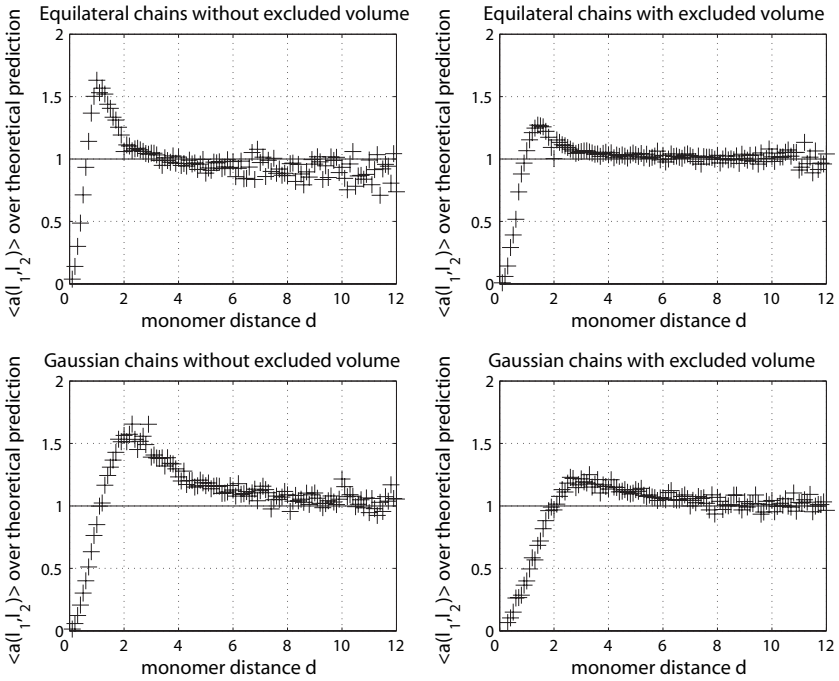


Figure 147: Shown are the ratios between the prediction for the leading order term for the invariant $\langle a(l_1, l_2)\rangle$ and the simulation results. While the prediction pertains to chains without excluded volume the results show the agreement with the theoretical predictions are excellent for distances larger than 10 for chains with and without excluded volume. The strong fluctuations at the end for the larger distances are a consequence of the plotted ratio. A single point in the figure represents at least an average over 10^4 simulation results.

10.3.1 Influence of Excluded Volume on $\langle a(l_1, l_2)\rangle$

For $\langle a(l_1, l_2)\rangle$ it was found that the excluded volume interactions do not have any measurable effect for distances larger than two (equilateral case) or three (Gaussian case) chain segments (cf. Fig. 146). A calculation of correlation lengths for the orientation of the chain segments leads to similar results, namely $d_c \approx 1.39$ for equilateral and $d_c \approx 2.81$ for Gaussian chains.
Each data point of the Figures 146, 147 and 148 for $\langle a(l_1, l_2)\rangle$ is an average over at least 10^4 values. The dashed lines in this figure show the leading order term for the

theoretical prediction of [56; 57] (cf. Eq. 58 and 59).

Although these predictions have been calculated to describe the behavior of $\langle a(l_1, l_2)\rangle$ for *large* d, they already fit very well for $d > 1$ for chains with and without excluded volume. Therefore, they might be used to derive some estimations of the mACN behavior for chains *with* excluded volume, too (only the *large* d case is important here because these chains contribute mostly to the mACN).

There seems to be a discontinuity at $d = 0$, since $\langle a(l_1, l_2)\rangle$ should vanish at the origin according to the Gauss invariant [56; 57]. Fig. 147 shows the comparison of the numerical results with the theoretical prediction of the papers cited above. One can see that the predicted behavior is not only good in the case of phantom chains but also in the case of chains with excluded volume.

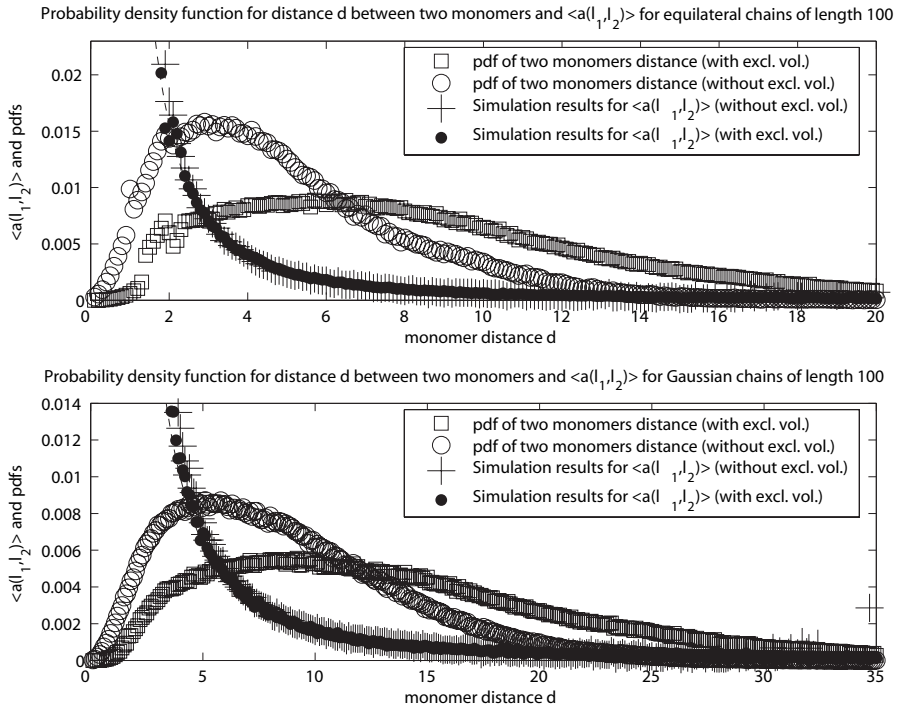


Figure 148: Shown are the distance probability density functions (pdf) for Gaussian and equilateral chains with and without excluded volume. In both cases the distribution is rather different but yield the same behavior on average (cf. Fig. 147).

10.3.2 Influence of Excluded Volume on the Distribution of d

In contrast to $\langle a(l_1, l_2)\rangle$ there is a huge effect of the excluded volume interactions on the probability distributions of d which can be seen in Fig. 148 for equilateral and Gaussian

chains of length $N = 100$ (together with the numerical results for $\langle a(l_1, l_2)\rangle$ on a larger scale in the background of the Figure).

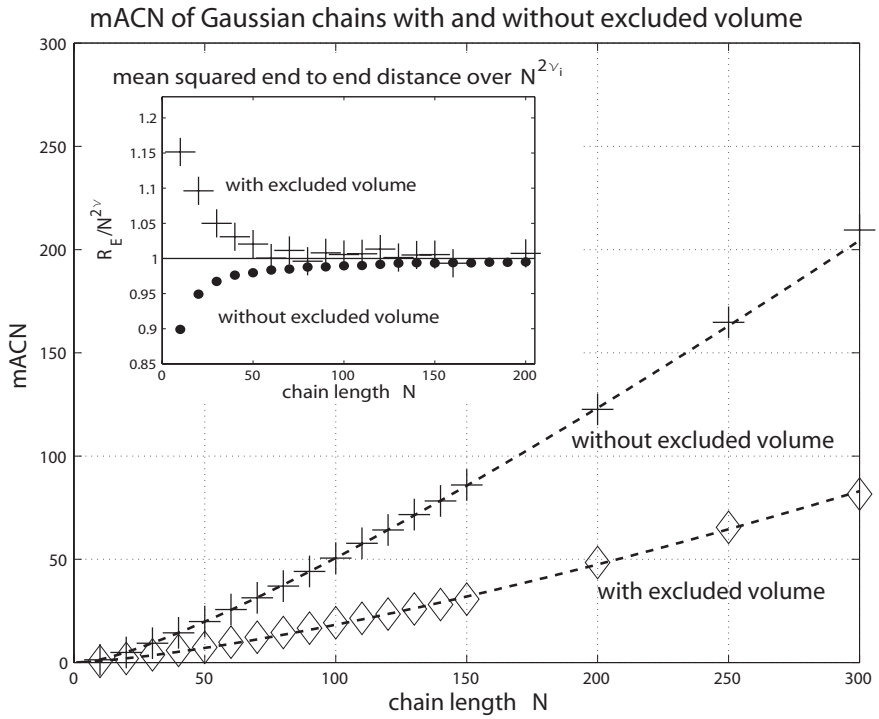


Figure 149: Shown are the mACN for Gaussian chains with and without excluded volume and the best fit for each one (cf. tables). The inset shows the end-to-end distance for the equilateral chains over its theoretical prediction $N^{2\nu_i}$ ($\nu_1 = 0.5$ and $\nu_2 = 0.588$). As this ratio converges to 1 one can see that the end-to-end distance follows the expected behavior. Every data point of the inset represents an average over at least 10^4 points and every point of the main figure is an average of at least 10^5 points.

One can see that the chains with excluded volume are much more stretched as expected due to the enhanced value of ν for the radius of gyration than those without. The Gaussian chains are longer than the equilateral ones. This is a consequence of the Gaussian probability distribution since the mean length of Gaussian chain segments is about 1.6 and the length of all equilateral chain segments is normalized.

The probability density functions of the equilateral chains show peaks at $d \approx 2$. This is a feature of the equilateral chains since the distance of the endpoints of two consecutive line segments is always larger than 0 and smaller than 2. If d exceeds 2, the probability to find monomers with a distance d will drop immediately. In the case of the equilateral chains with excluded volume the end-to-end distance of two consecutive line segments

is always larger than

$$d_{\min} = \sin\left(\frac{\alpha_{\min}}{2}\right) = 0.66 \quad \text{with } \alpha_{\min} = 83.62 \text{ degrees},$$

where the minimal value of the angle α between two consecutive chain segments is determined by the ellipsoidal excluded volume model. The probability density function of the equilateral chains *with* excluded volume has two discontinuity points (at $d_{\min} = 0.66$ and $d = 2$, cf. Fig. 148). The one *without* excluded volume shows only the one at $d = 2$ as explained above.

10.3.3 Influence of Excluded Volume on the mACN

Figure 149 and Figure 150 show that the total mACN for chains with excluded volume is much lower than for phantom chains. As pointed out above this is mainly a consequence of the much broader distance probability density function for chains with excluded volume. The orientation effects (i.e. the changes of $\langle a(l_1, l_2)\rangle$) play only a minor role.

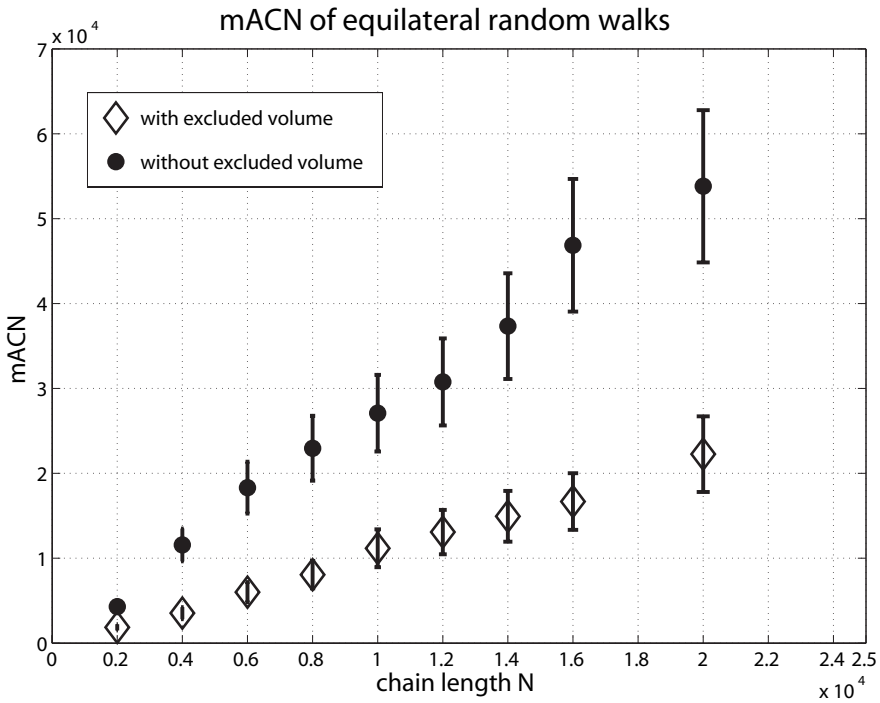


Figure 150: The mACN of equilateral random walks on a cubic lattice with and without excluded volume interactions.

As these random walks are a model for equilibrated polymers in solution and as these polymers do have excluded volume interactions one should expect a much lower ACN

for these polymers than predicted by [56; 57]. Furthermore, a linear correction term to the predicted $N\ lnN$-behavior (cf. Fig. 148) seems to worsen the fits in the case of the chains with excluded volume as early points are all above the dashed line and late points are below the line. Each data point in the figures which show the mACN is an average over at least 10^5 points.

The prediction for the mACN by Diao et. al. [56; 57] as stated in Eq. 55 and 57 are in good agreement with our simulations. Our simulation results confirm these results and we calculated a factor of $c_2 = -0.3051$ for the equilateral chains and a factor of $c_1 = -0.2265$ for the Gaussian ones.

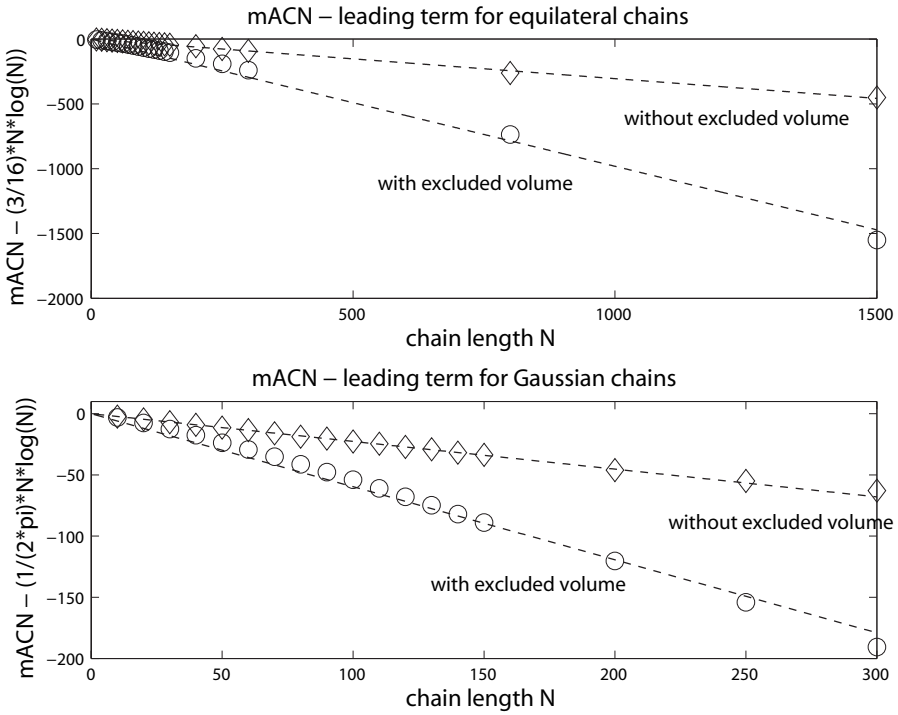


Figure 151: The error term of the theoretical prediction of the mean average crossing number is negative and much smaller for the chains with excluded volume. Every data point represents again an average over at least 10^5 points.

The fit results are complied in the following two tables (cf. Tab. 8 and Tab. 9). Please note that the data for the equilateral chains with excluded volume reaches up to $N = 1500$ for off-lattice chains and up to $N = 20000$ for chains on a cubic lattice, whereas the mACN for the Gaussian chains reaches to $N = 300$ (off-lattice). The summed square of residuals (sse) and the coefficient of multiple determination (rsquare) are also given in the tables below. The power law fits with $b > 1$ will beat $n\ ln(n)$ for large n. They are shown here to give an approximation of the mACN for smaller chains.

Table 8: Data Fits

chains without excluded volume				
data	fit	a	sse	rsquare
equilateral	$(3/16)N\ln(N)+aN$	-0.3051	445.2	0.9998
Gaussian	$(1/2\pi)N\ln(N)+aN$	-0.2265	61.4	1.0000

Table 9: Fitting Parameters

Gaussian chains with excluded volume				
fit	a	b	sse	rsquare
aN^b	0.03239	1.376	10.07	0.9988
$(1/2\pi)N\ln(N)+aN$	-0.5968	—	537.53	0.9349
$aN\ln(N)+bN$	0.07468	-0.1553	16.3505	0.9980
$(1/2\pi)N\ log(N)^a$	0.2426	—	310.4	0.9624
equilateral chains with excluded volume				
aN^b	0.06382	1.232	1306.1	0.9952
$(3/16)N\ln(N)+aN$	-0.9812	—	29987	0.8896
$aN\ln(N)+bN$	0.03914	0.05466	241.58	0.9991
$(3/16)N\ log(N)^a$	0.2902	—	1010.1	0.9963

10.3.4 Conclusions

The results in Sec. 10.3.1 show that the topological invariant and the behavior of $\langle a(l_1,l_2)\rangle$ which are the basis for the proof by Diao [56; 57] and co-workers are weakly influenced by excluded volume interactions. Hence it is still unclear especially in the light of the inconclusive simulation data, whether the proven law $c_1N\ln N+c_2N$ for the non-excluded volume case changes to a power law as suggested by [153; 220].

Much larger chains are needed to clearly discern between the possibilities. A rough estimate shows that we need chains that are at least ten times longer. Here the problem is that despite the well-developed Pivot algorithm the computation of the excluded volume interaction is so time-consuming that for now it seems to be possible to do such a calculation only expending a truly fair amount of computing resources.

10.4 mACN of Looped Polymers

In this section the connection between polymer loops and the mACN will be examined. The importance of loops in biopolymers like DNA or chromatin has already been explained in Sec. 8.

All chains of a particular polymer model with length N will be divided into subsets of chains with a fixed loop number k. The average over the ACNs of such a subset is denoted by $\mathrm{mACN}_N(k)$ and the average over all these $\mathrm{mACN}_N(k)$ weighted by the

probability of the number of loops $p(k)$ again gives the mean average crossing number mACN_N which was discussed in the previous sections.

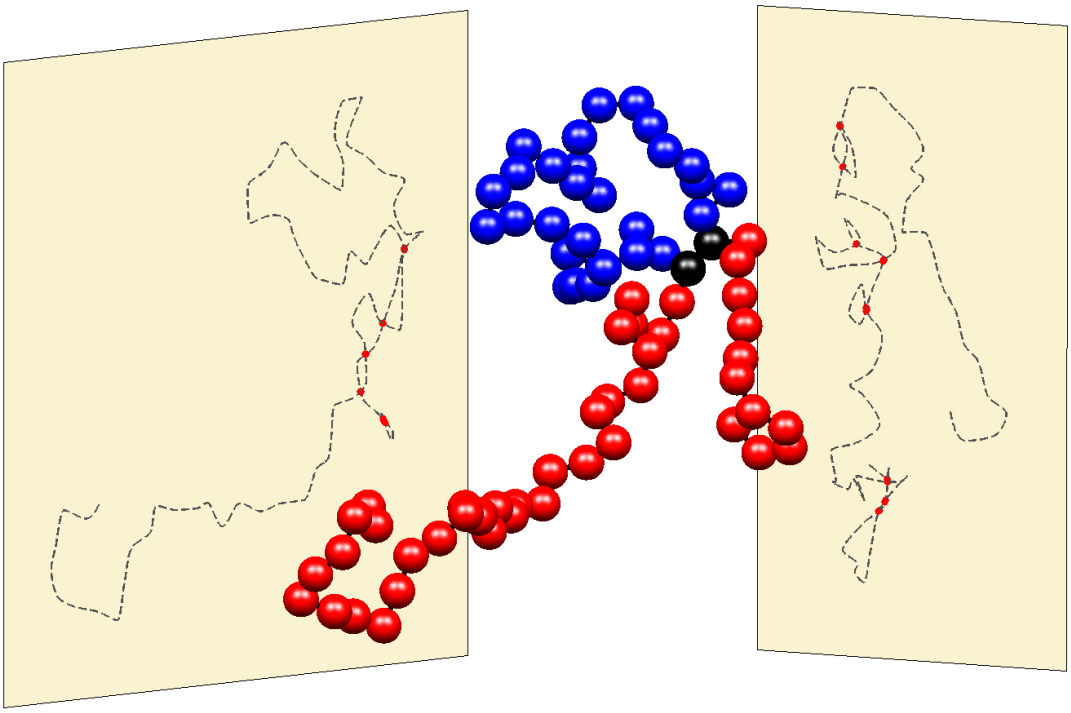


Figure 152: Illustration of an off-lattice self-avoiding random walk with spherical excluded volume of length $N = 75$. The dashed lines in the background are projections of the chain onto the $xz-$ and the yz-plane with five respectively eight crossings (marked as red spots). The chain has one loop ($k = 1$) which is marked blue. The loop length is 28. The start as well as the end point of the loop are marked colored black. A larger version of this figure can be found in App. A (cf. Fig. 207).

Altogether, four different kinds of random walks will be investigated: Simple random walks (RAW) on a cubic lattice and in continuum as well as self-avoiding walks (SAW) on a cubic lattice and in continuum.

The excluded volume of the off-lattice self-avoiding walks is illustrated in Fig. 152: The walk resembles beads-on-a-string because every segment has spherical excluded volume. In the background of Fig. 152 one can see two projections of the shown chain (dashed lines). Their crossings are marked with red spots. To calculate the average crossing number of a single walk conformation 10^3 randomly chosen planes have been used again. Furthermore, one can already see a loop within the chain conformation of Fig. 152 which is marked blue. The black spheres represent the start and the end point of the loop.
The walks have mainly been generated with the FP-Algorithm (cf. Sec. D).

In the case of random walks of a cubic lattice a coarse-graining was applied: In fact every single monomer consists of a random walk of 15 steps itself. This was done to avoid permanent overlapping of monomers and thus improve the loop definition in this case: Two overlapping non-consecutive monomers define the start respectively the end point of a loop.

To determine the number of loops of a self-avoiding walk on the cubic lattice the walk monomers between non-consecutive next neighbor monomers have been identified with loops.

In the off-lattice case two non-consecutive monomers will define a loop, if their spatial distance is smaller than a critical distance d_{crit} defined by $d_{crit} = 0.1D$ were D is the diameter of the spheres that were used to define the excluded volume of the monomers.

According to Eq. 47 it is obvious that the mACN_N should increase at least linearly with the number of loops k in the case that the loops are independent on each other (i.e. random walks). Here it will only be averaged over all chain conformations with a particular loop number k (denoted by $\mathrm{mACN}_N(k)$) and not over all possible conformations like in the previous sections.

Fig. 153 shows the behavior of the mACN_N in dependency on the number of loops k in the case of random walks without self-avoidance. Fig. 154 and Fig. 155 show the same curves for the self-avoiding walks.

Furthermore, Fig. 154 and Fig. 155 show the probability distribution $p(k)$ of the loop numbers for fixed chain length as well. One can see that $\mathrm{mACN}_N(k)$ is only given for chains the loop numbers of which actually occur. The number of loops is only limited above by the chain length but for large walks it is very unlikely to have only a few loops.

Since the distributions of k are almost symmetric the central value of a $\mathrm{mACN}_N(k)$ curve corresponds to the $\mathrm{mACN}(N)$ value, i.e. the mACN averaged over all possible chain conformations and thus over all possible k-values.

In the case of random walks on a cubic lattice one gets much less loops due to the previously made coarse-graining.

If one divides out the linear part, one will find that the $\mathrm{mACN}_N(k)$ of the random walks becomes almost a constant for long walks (cf. Fig. 156). In the case of self-avoiding walks (cf. Fig. 157) one gets a decreasing curve for $\mathrm{mACN}_N(k)/k$ which means that in this case $\mathrm{mACN}_N(k) \propto k^c$ with $c < 1$. But the decrease itself decreases with increasing chain length (as in the case of the random walks).

Maybe in the case of self-avoiding walks one needs longer random walks to actually decide on the scaling. Furthermore, one could try to use self-similarity arguments and a rescaling of the underlying loop distributions to learn about the actual scaling of $\mathrm{mACN}_N(k)$.

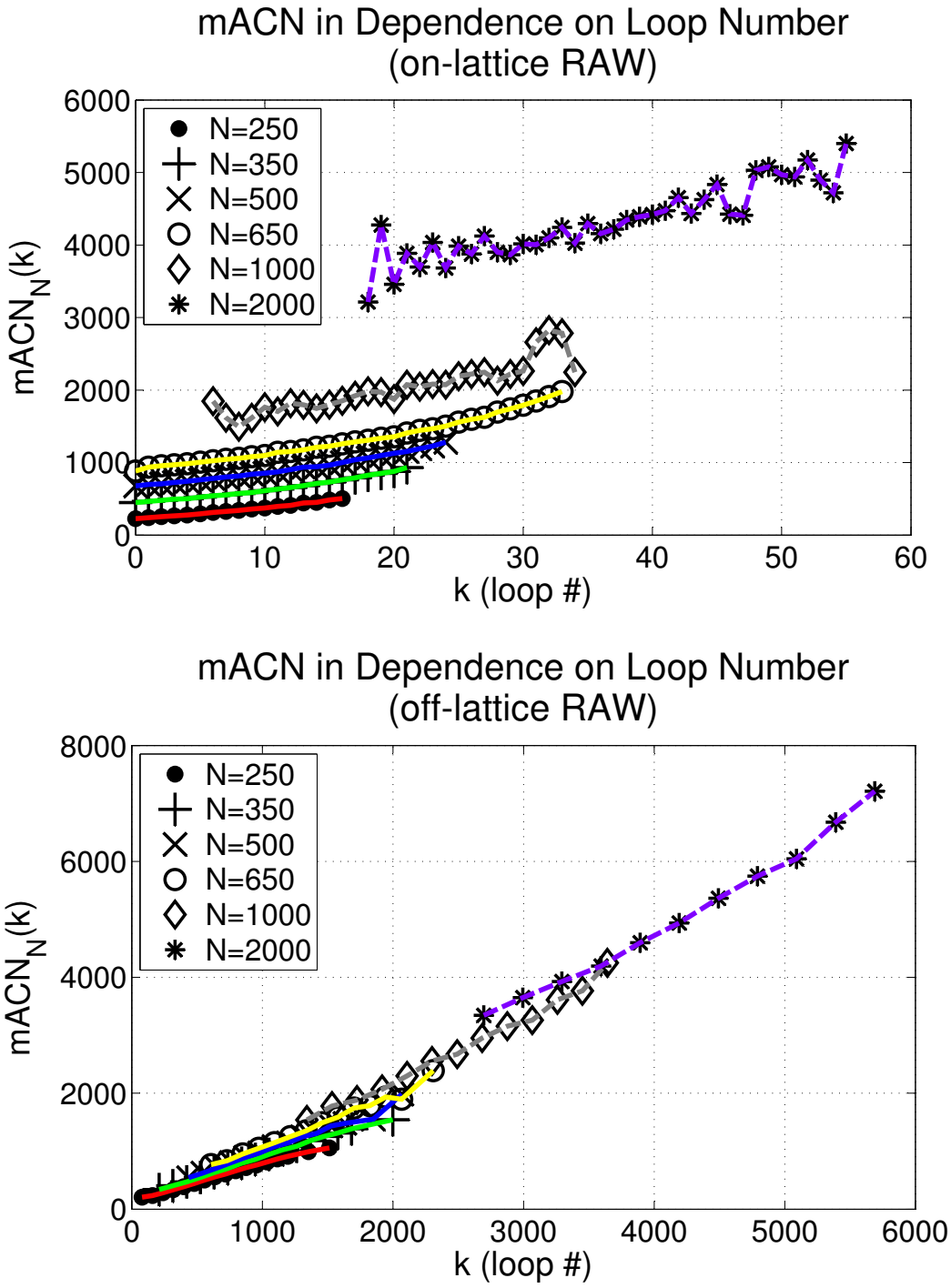


Figure 153: This figure shows the dependence of mACN_N on the number of loops k for random walks on a cubic lattice and in continuum.

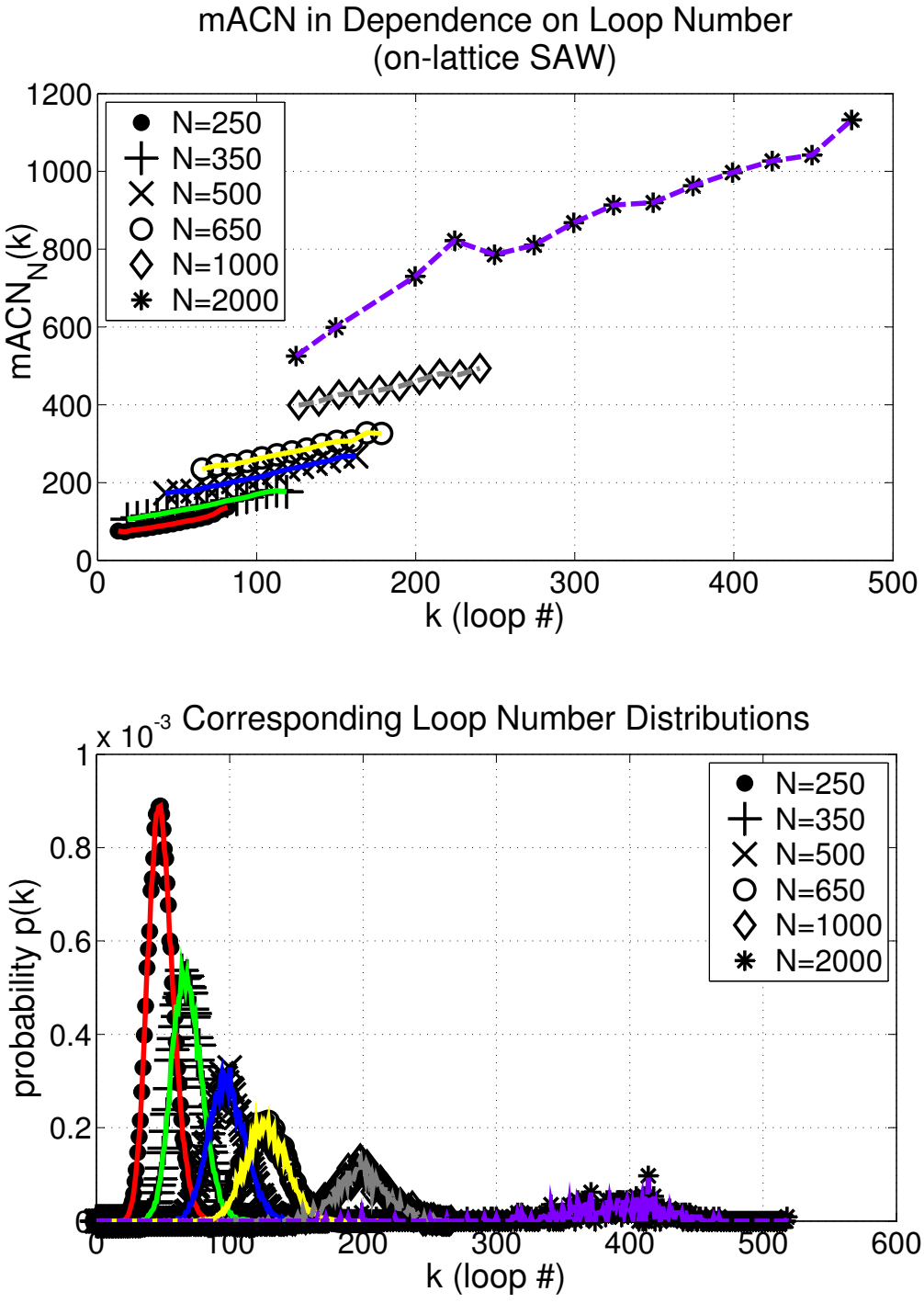


Figure 154: The top figure shows the dependency of mACN_N on the number of loops k in the case of self-avoiding walks on a cubic lattice. Furthermore, the corresponding probability distributions of the loop numbers are shown as well (bottom).

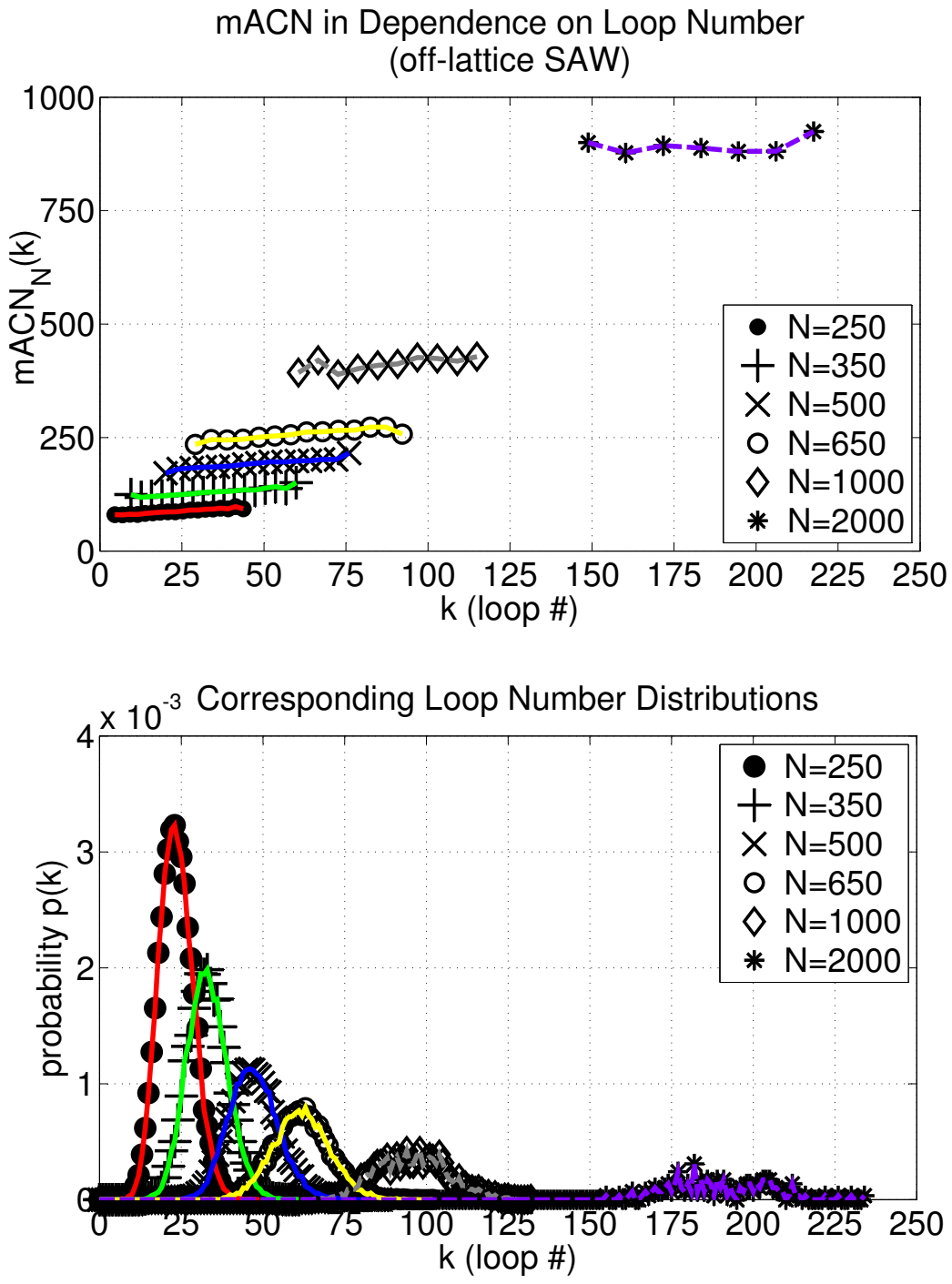


Figure 155: This figure shows the dependency of mACN_N on the number of loops k (top) in the case of self-avoiding walks (off-lattice). Furthermore, the corresponding probability distributions of the loop numbers are shown as well (bottom).

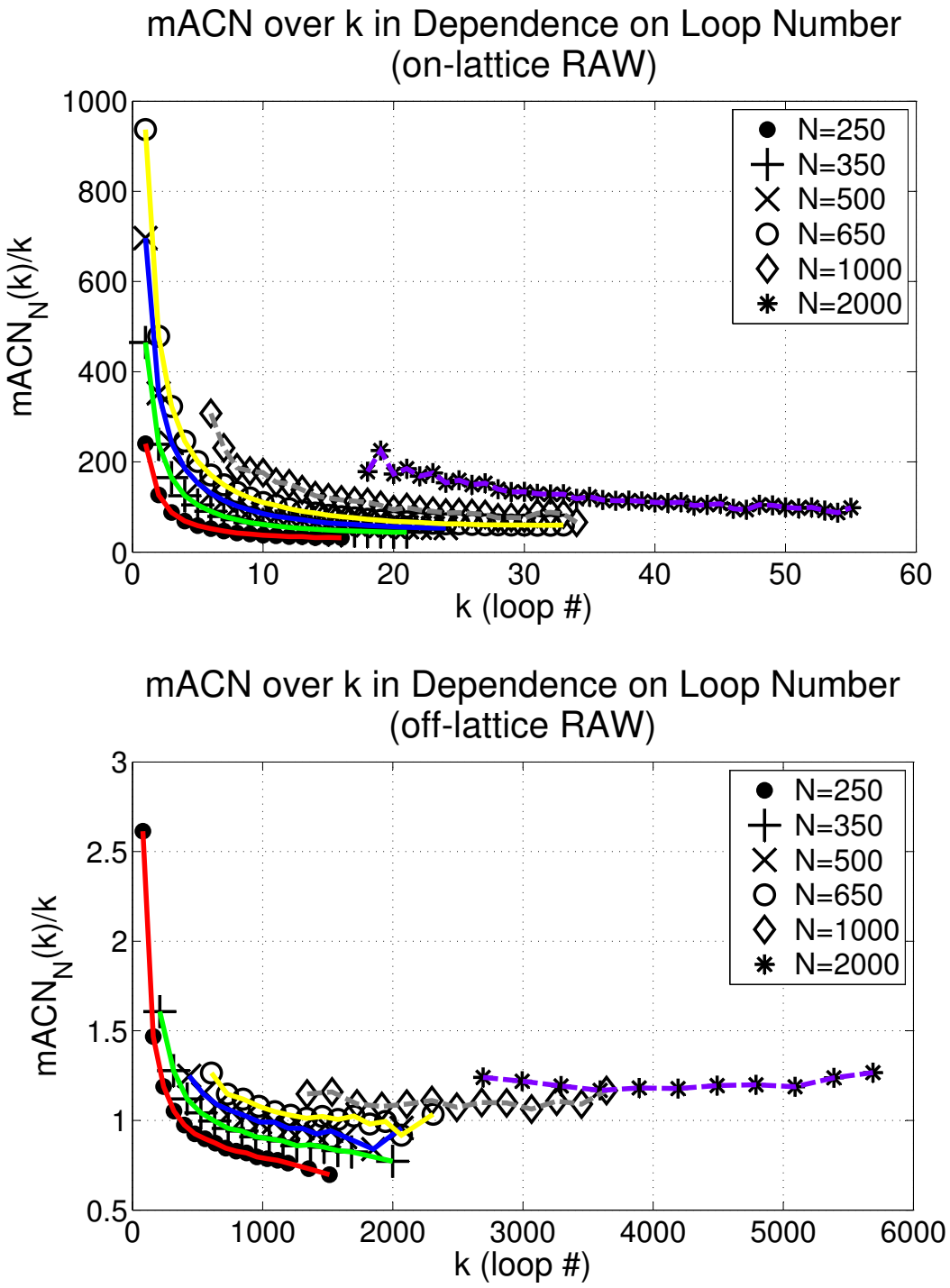


Figure 156: This figure shows $\mathrm{mACN}_N(k)$ over k for random walks. One can see that the long chains get nearly independent of the loop number k.

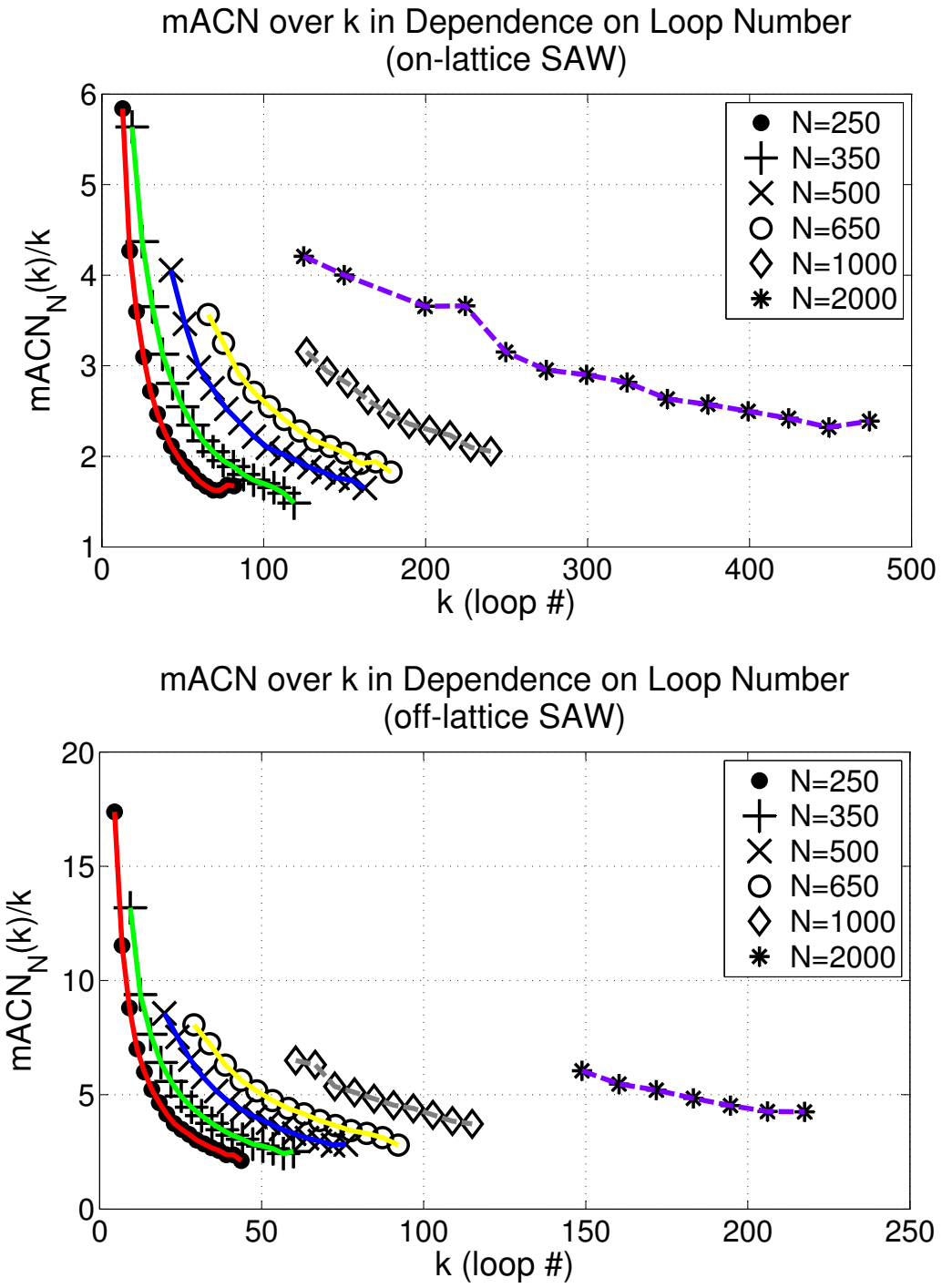


Figure 157: This figure shows $\mathrm{mACN}_N(k)$ over k for Self-Avoiding Walks. One can see that the curves decrease even in the case of long chains.

10.5 mACN and Scattering Function

In this sections an idea will be given how the mACN can analytically be connected with the two-point correlation function $\Phi(\overrightarrow{r}, \overrightarrow{r'})$ and thus with the scattering function $S(\overrightarrow{q})$.

The crossing number (cn) of a chain conformation $\{\overrightarrow{r}_i\}$ on a cubic lattice with respect to a projection onto the xy-plane is given by

$$cn_z\{\overrightarrow{r}_i\} = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \delta_y(y_i - y_j)\delta_z(z_i - z_j) \tag{60}$$

where

$$\delta_{y,z}(a-b) = \begin{cases} 1 & a = b \\ 0 & a \neq b \end{cases} . \tag{61}$$

For an isotropic distribution $P_N(\{\overrightarrow{r_i}\})$ of the chain points $\{\overrightarrow{r_i}\}$ (for instance polymers without an external field) one gets the mean Average Crossing Number (mACN) by the ensemble average of cn_z that means the average is taken over all possible chain conformations $\{\overrightarrow{r_i}\}$ (respectively $\{\overrightarrow{r(t)}\}$ for the continuum case):

$$mACN(N) = \langle cn_z \rangle = \int d\overrightarrow{r_1} \ ... \ \int d\overrightarrow{r_N} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \delta_y(y_i - y_j)\delta_z(z_i - z_j) \ P_N(\{\overrightarrow{r_i}\}). \tag{62}$$

This is because in the case of an isotropic probability distribution for the chain conformations all conformations which are in fact rotations of one underlying conformation are equally likely. Therefore, one will average over all possible projections, too, if one averages over the whole ensemble of possible conformations.

The mACN for a polymer chain can be calculated with the help of the correlation function

$$\Phi_0(\overrightarrow{r}, \overrightarrow{r}') := \frac{1}{N} \sum_i \sum_j \langle \delta(\overrightarrow{r} - \overrightarrow{r_i})\delta(\overrightarrow{r}' - \overrightarrow{r_j}) \rangle \tag{63}$$

respectively

$$\Phi_0(\overrightarrow{r}, \overrightarrow{r}') = \frac{1}{\mathcal{N}} \int ds \int dt \langle \delta(\overrightarrow{r} - \overrightarrow{r(s)})\delta(\overrightarrow{r}' - \overrightarrow{r(t)}) \rangle \tag{64}$$

in the continuum case.

Here, the average is taken again over all possible chain conformations $\{\overrightarrow{r_i}\}$ respectively $\{\overrightarrow{r(t)}\}$. $\Phi_0(\overrightarrow{r}, \overrightarrow{r}')$ is proportional to the probability of finding segments at the locations $\overrightarrow{r}$ and $\overrightarrow{r}'$ simultaneously i.e. they have to be part of the same polymer chain.

In the following always continuous chains will be considered with a chain parameter s or $t \in [0, 1]$ instead of an index i or j.

$$
\begin{aligned}
mACN(N) &= \frac{1}{Z_0} \sum_{\{r(t)\}\in\Gamma_0} cn_z \\
&= \int d\vec{r_o} \ldots \int d\vec{r_N} \int_0^1 ds \int_s^1 dt\ \delta(x(t)-x(s))\delta(y(t)-y(s))P(\{\vec{r_i}\}) \\
&\propto \int_V d\vec{r}\,' \int_V d\vec{r}\ \Phi_0(\vec{r},\vec{r}\,')\delta_y(y,y')\delta_z(z,z')
\end{aligned}
$$

where $\delta_y(y-y')$ and $\delta_z(z-z')$ are the one-dimensional delta-functions and $d\vec{r} = dx\ dy\ dz$ as well as $d\vec{r}\,' = dx'\ dy'\ dz'$.

If there is no external inhomogeneous field imposed on the chain, the center of mass on the chain can be located equally anywhere in the system [114]. With this translational invariance one gets:

$$\Phi_0(\vec{r},\vec{r}\,') = \frac{1}{V}\Phi_0(\vec{r}-\vec{r}\,') = \frac{1}{V}\Phi_0(\vec{q}) \tag{65}$$

where V is the system volume which arises from the translational degree of freedom of the center of mass of the chain which can be anywhere in the volume V. The *Wiener-Khinchin relation* [114] describes the connection between the two-point correlation function and the scattering function $S(\vec{q})$ (i.e. the scattering intensity in direction $\vec{q}$)

$$S(\vec{q}) = N(\mathcal{F}\Phi_0)(\vec{q}) - (2\pi)^3 V\Phi_0^2\delta(\vec{q}) \tag{66}$$

leads to

$$\Phi_0(\vec{q}) = \mathcal{F}^{-1}\left(\frac{S(\vec{q})+(2\pi)^3V\Phi_0^2\delta(\vec{q})}{N}\right). \tag{67}$$

With this equation we get a mean field relation between the mACN an the scattering intensity of the specimen-chain:

$$mACN \propto \frac{1}{N}\int_V d\vec{r} \int_V d\vec{r}\,'\ (\mathcal{F}^{-1}S(\vec{q}))\delta_y(y-y')\delta_z(z-z') \tag{68}$$

$$+\ \frac{(2\pi)^3V\Phi_0^2}{N}\int_V d\vec{r} \int_V d\vec{r}\,'\ (\mathcal{F}^{-1}\delta(\vec{q}))\delta_y(y-y')\delta_z(z-z'). \tag{69}$$

Hereby, Φ_0 is the average segment density and depends on the phase space of the chain. V is defined above and represents the translational degree of freedom.

10.6 Loop Number Estimation

A further possible application of the mACN might be the analysis of two-dimensional experimental data (e.g. FISH measurements, or localization microscopy data (cf. Sec. 3)).

The chain conformations are projected during the calculation of the mACN anyway and furthermore the mACN shows a strong correlation with the number of loops within polymer chains (cf. Sec. 10.4). Therefore, two-dimensional data sets, i.e. projections, of underlying three-dimensional polymer chain configurations might be used to estimate the average number of loops in the original 3D conformation.

One problem in this context is that the connectivity data is lost in most of the cases in which experimental data is available. Only single locations of chain points are given but it is not clear, which points are directly connected in the original polymer chain.

But for the estimation of the loop number with the mACN one has to apply a polymer model anyway (cf. Sec. 10.4) which can be used as an advantage in this case. One can do simulations of polymer chains, project them onto a plane and then randomly reconnect the data points to a chain. Since one knows the initial 3D conformation on can calculate the actual mACN and compare it to the mean value of the crossings of the projected and randomly connected chains.
These simulations have been done but are not shown here. Unfortunately, they consume a large amount of computation time. But in the end one gets the ratio of the actual mACN and the average number of crossings of projected and randomly connected chains.

So one has to decide first which polymer models ought to be applied, determine the explained ratios and then one can use them to calculate the mACN and thus the average number of loops in the original polymer chains.
One could imagine for example to apply random walks as well as self-avoiding walks to estimate an upper and lower bound for the loop number. This was done in the case of FISH data sets of chromatin. Unfortunately the data was limited to four cells, i.e. four images, which did not allow to make any significant statements.

11 Appendix

A Supplemental Figures

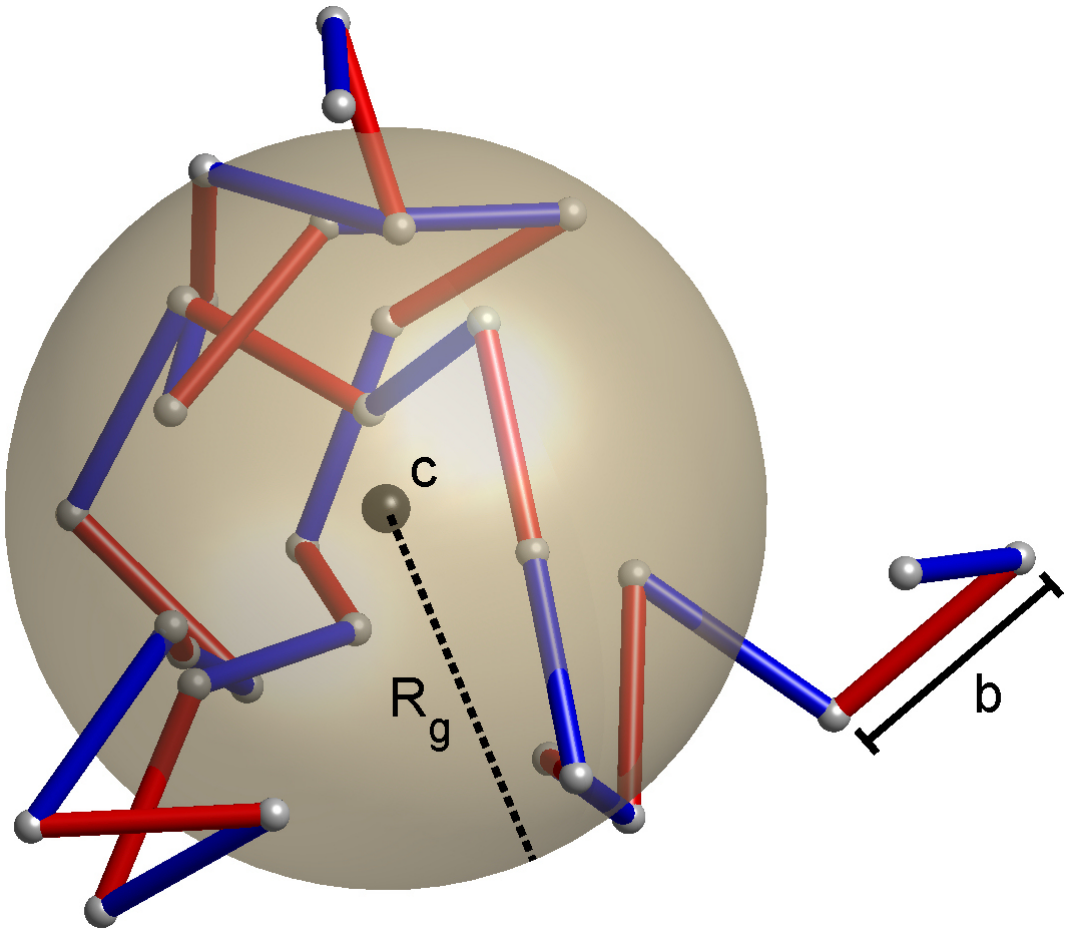


Figure 158: Example for a short freely jointed chain of $N = 30$ links with alternating color. The monomer length is $b = 100$ and the radius of gyration $R_g = 155$. c denotes the center of mass (black sphere).

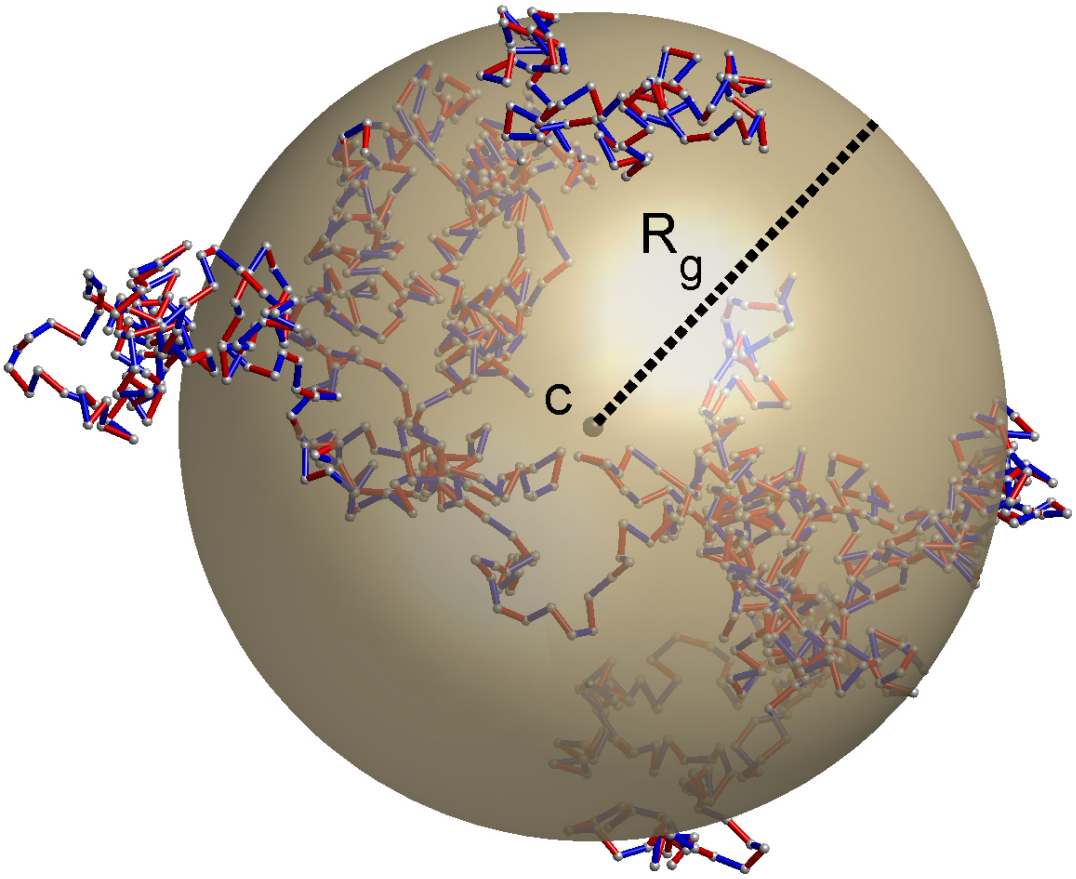


Figure 159: Example of a freely jointed chain of $N = 100$ links with alternating color. The monomer length is $b = 100$ and the radius of gyration $R_g = 1186$. c denotes the center of mass.

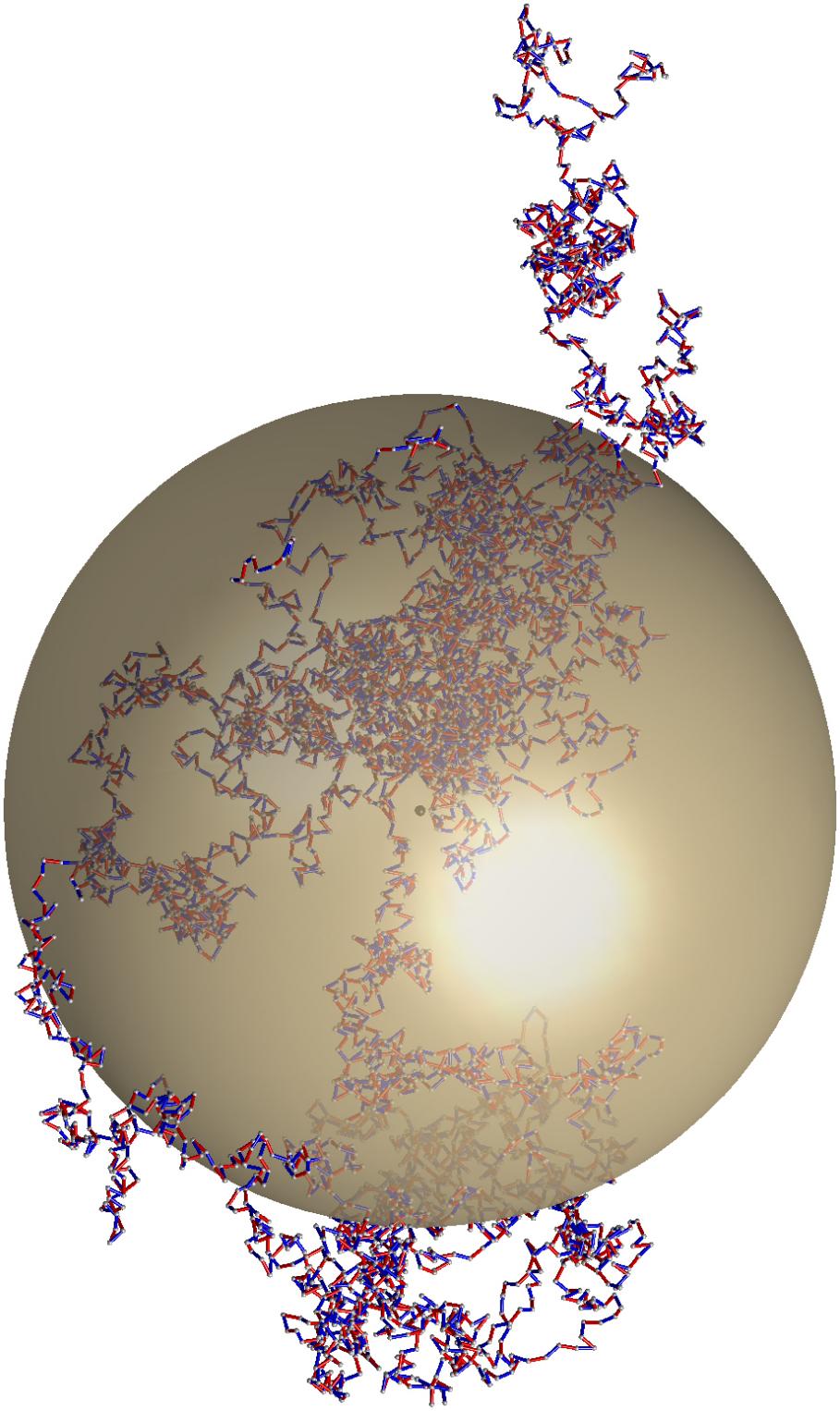


Figure 160: Example of a freely jointed chain of $N = 5000$ links with alternating color. The monomer length is $b = 100$ and the radius of gyration $R_g = 2323$.

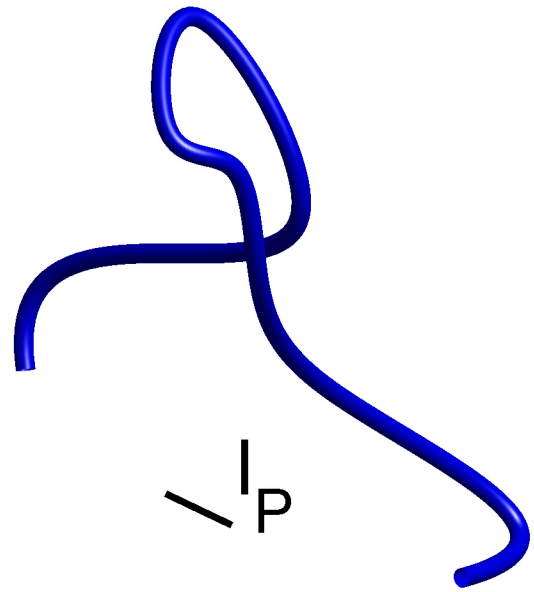


Figure 161: Example of a Worm-like Chain with sketched in persistence length l_P.

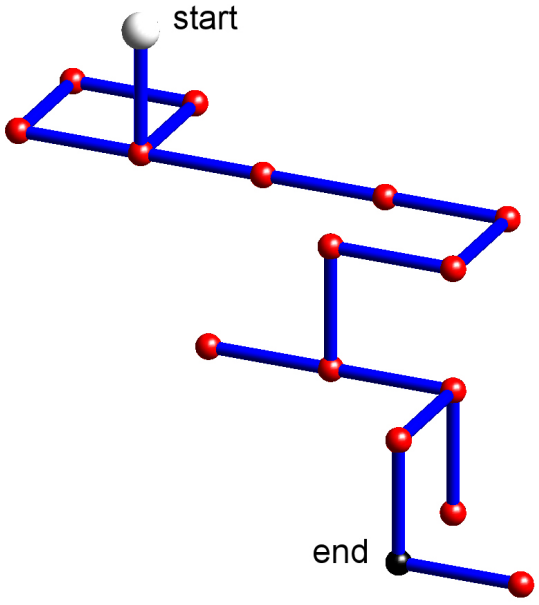


Figure 162: Example of a random walk of 20 steps on a cubic lattice. The lattice constant is 1, the end-to-end distance of the walk is 4.7. The white sphere denotes the start point of the walk and the black one represents the end point.

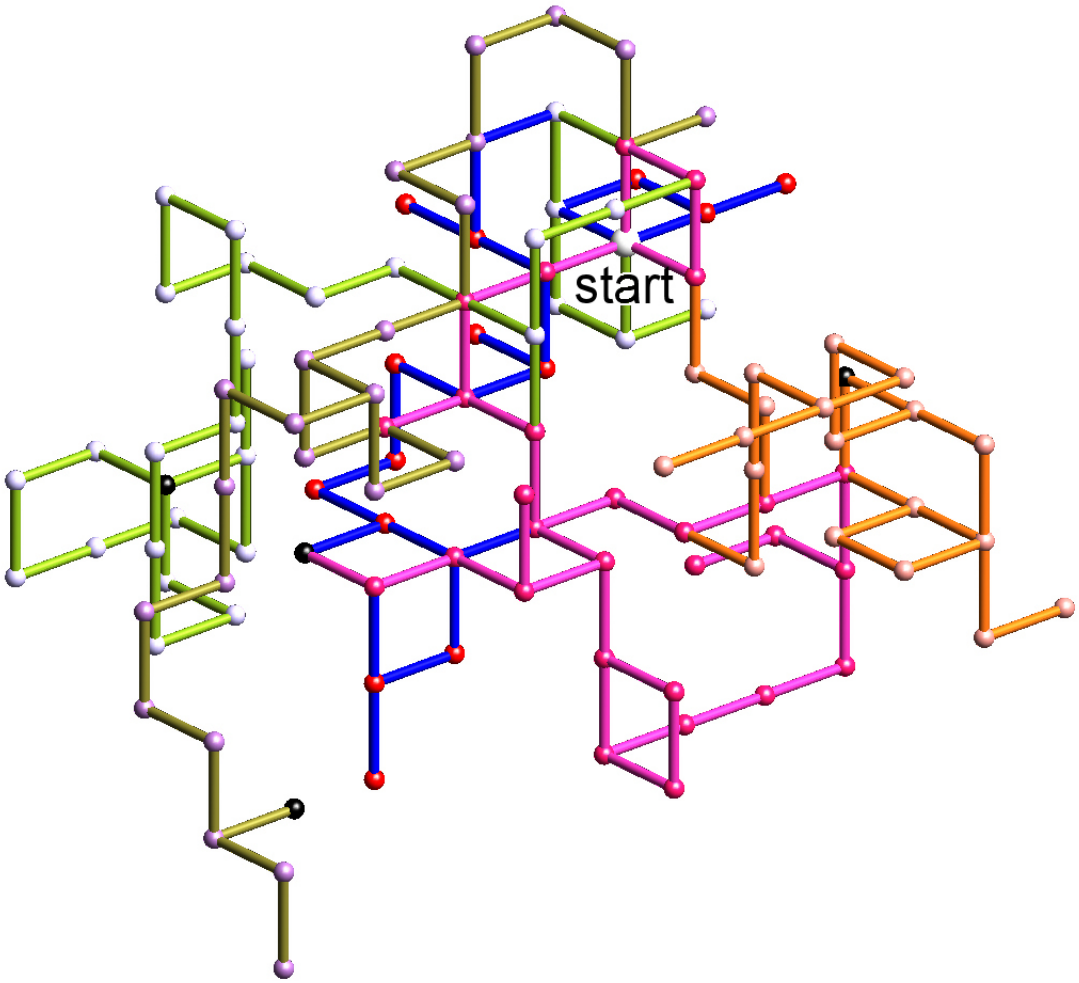


Figure 163: Example of five random walks with 40 steps each on a cubic lattice. The lattice constant is 1. All walks start at the origin (big white sphere). Black spheres denote end points.

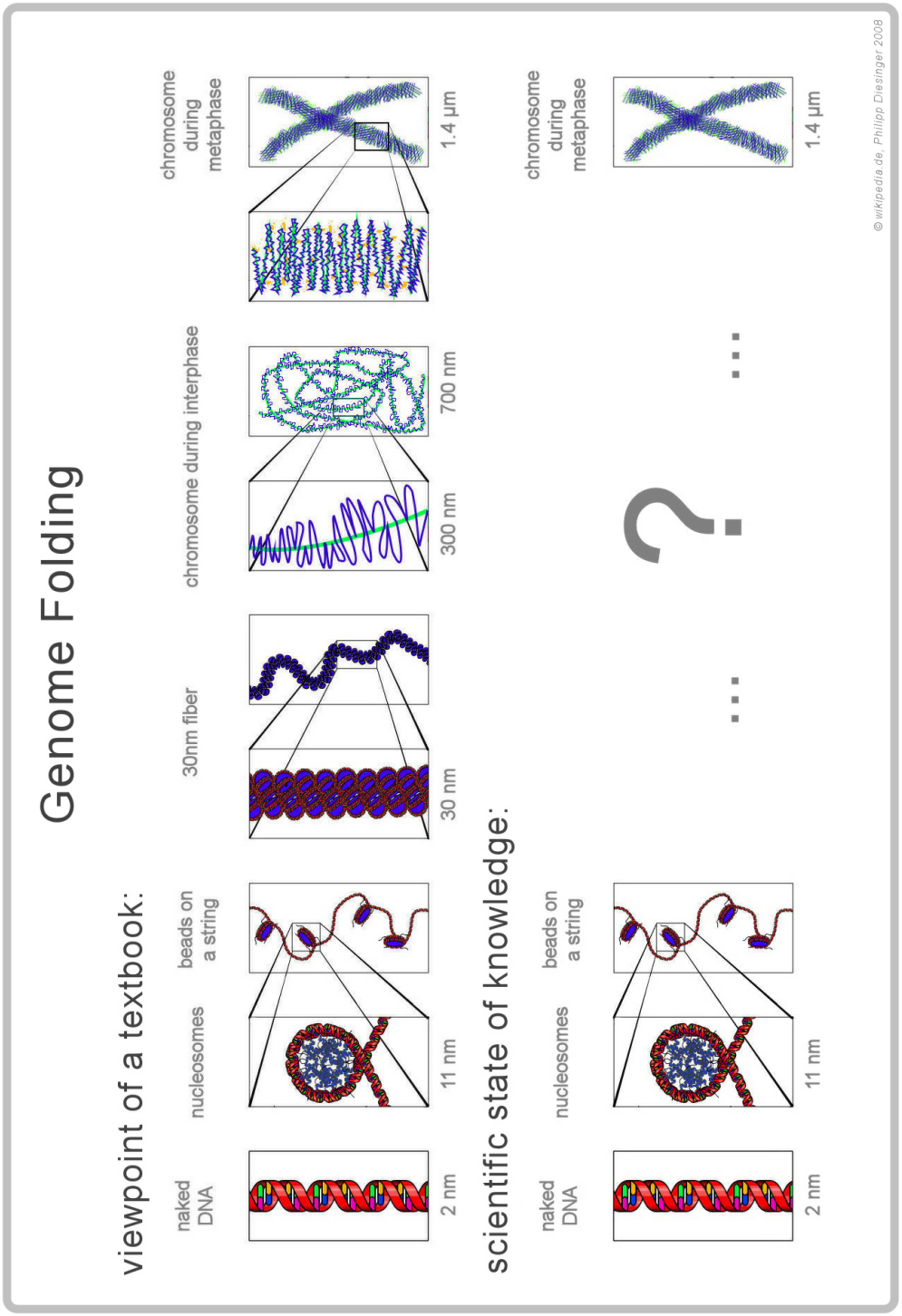


Figure 164: Genome folding at different scales. Comparison between the naive viewpoint of a textbook and the current scientific state of knowledge.

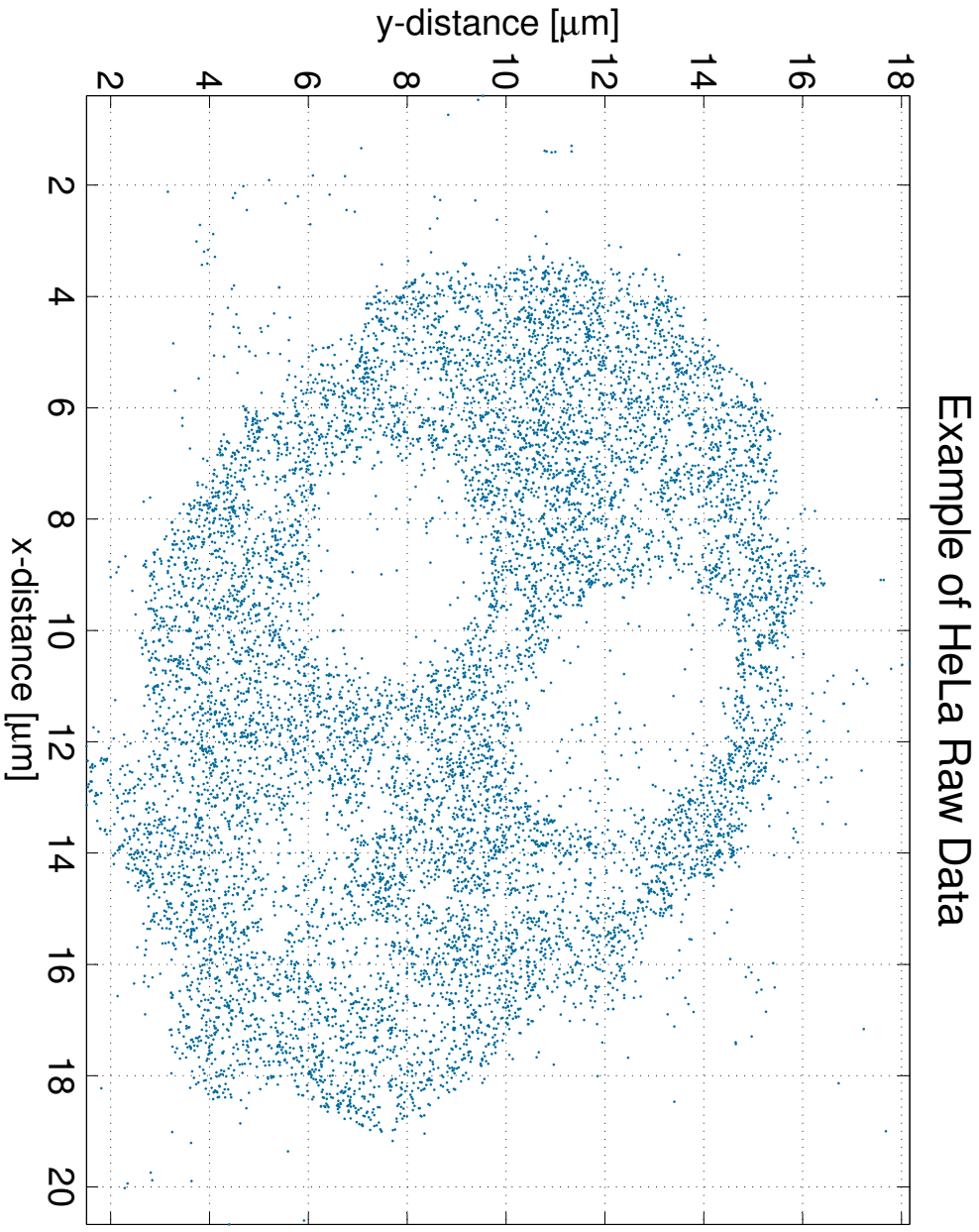


Figure 165: An example of the raw data of a HeLa nucleus.

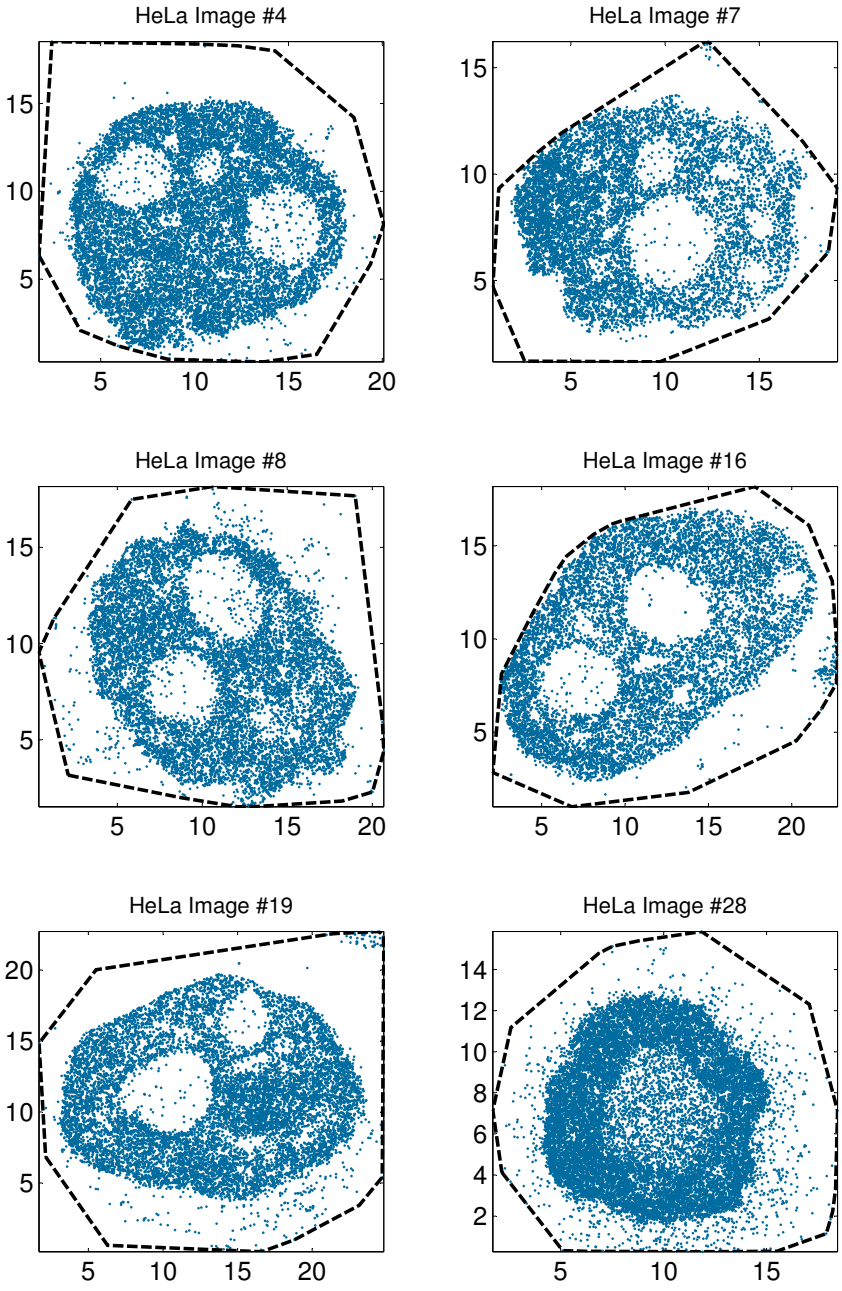


Figure 166: Several localization microscopy images of HeLa nuclei. The black dashed lines represent the convex hull of the raw data. All distances are given in μm.

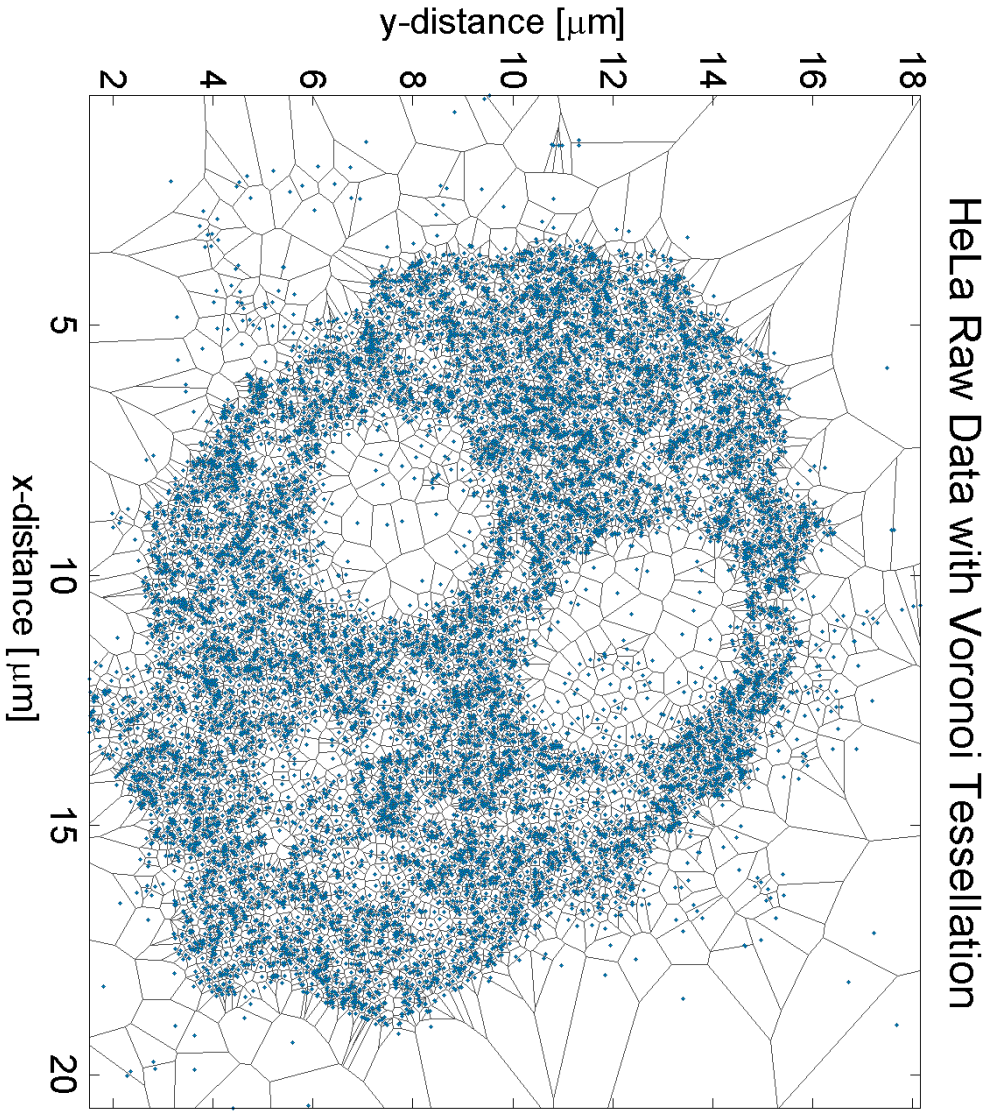


Figure 167: This figure shows a HeLa nucleus image together with the corresponding Voronoi tesselation. In this case HeLa cell #8 was used.

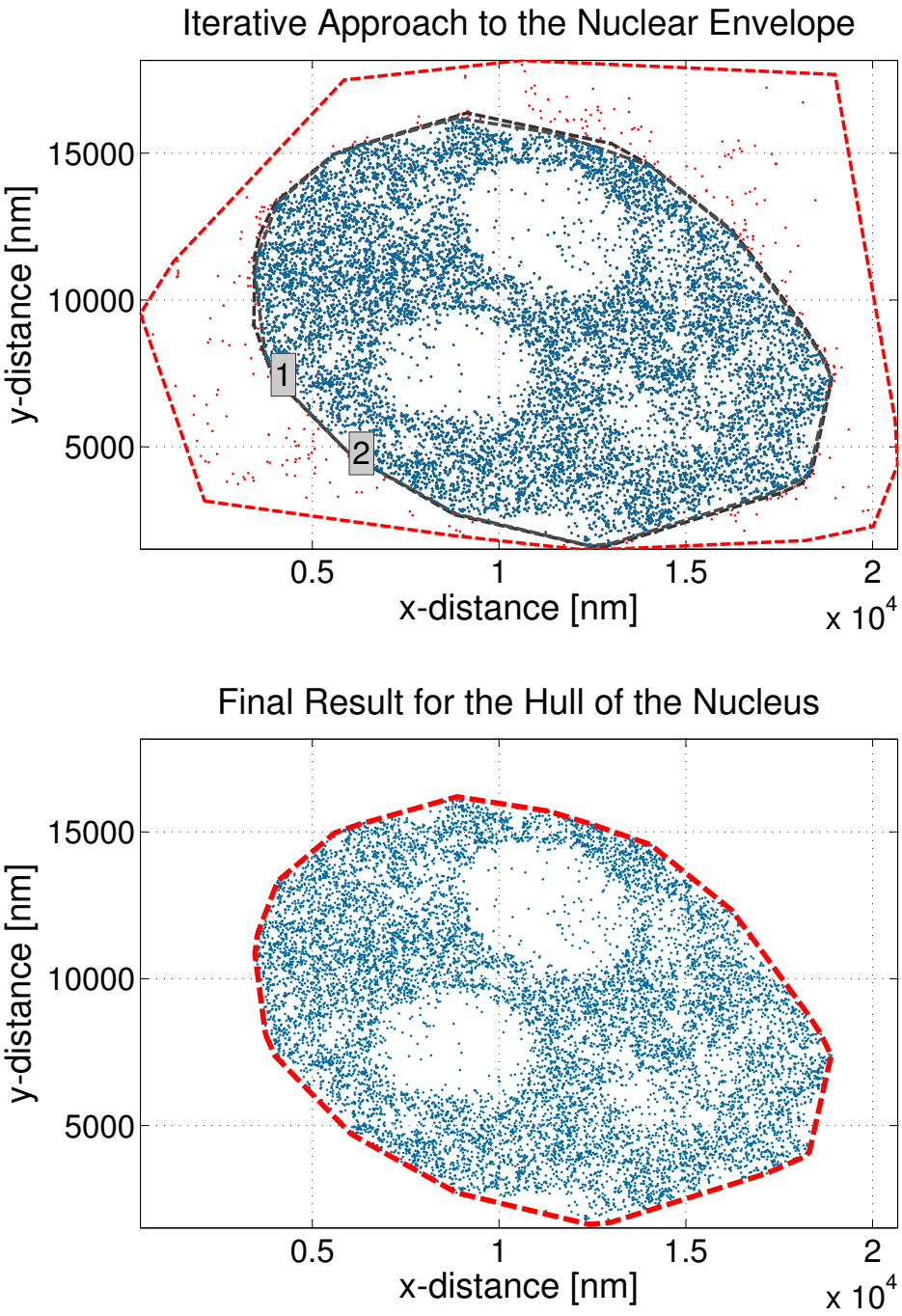


Figure 168: An example for the noise repression in the case of a HeLa nucleus. Here the convex hull is determined by iteratively applied Voronoi tessellations.

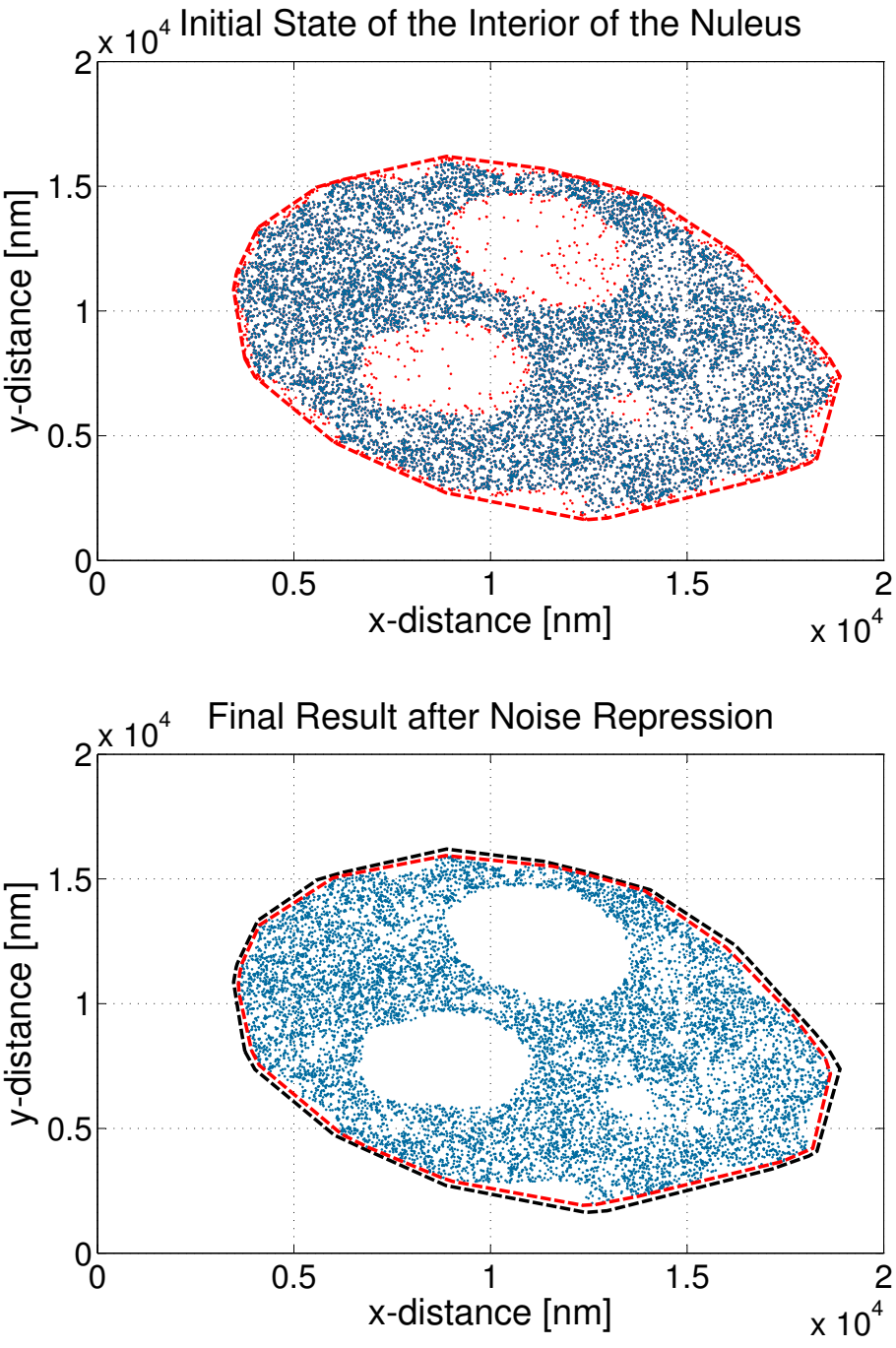


Figure 169: An example of the noise reduction in the interior of a HeLa cell nucleus.

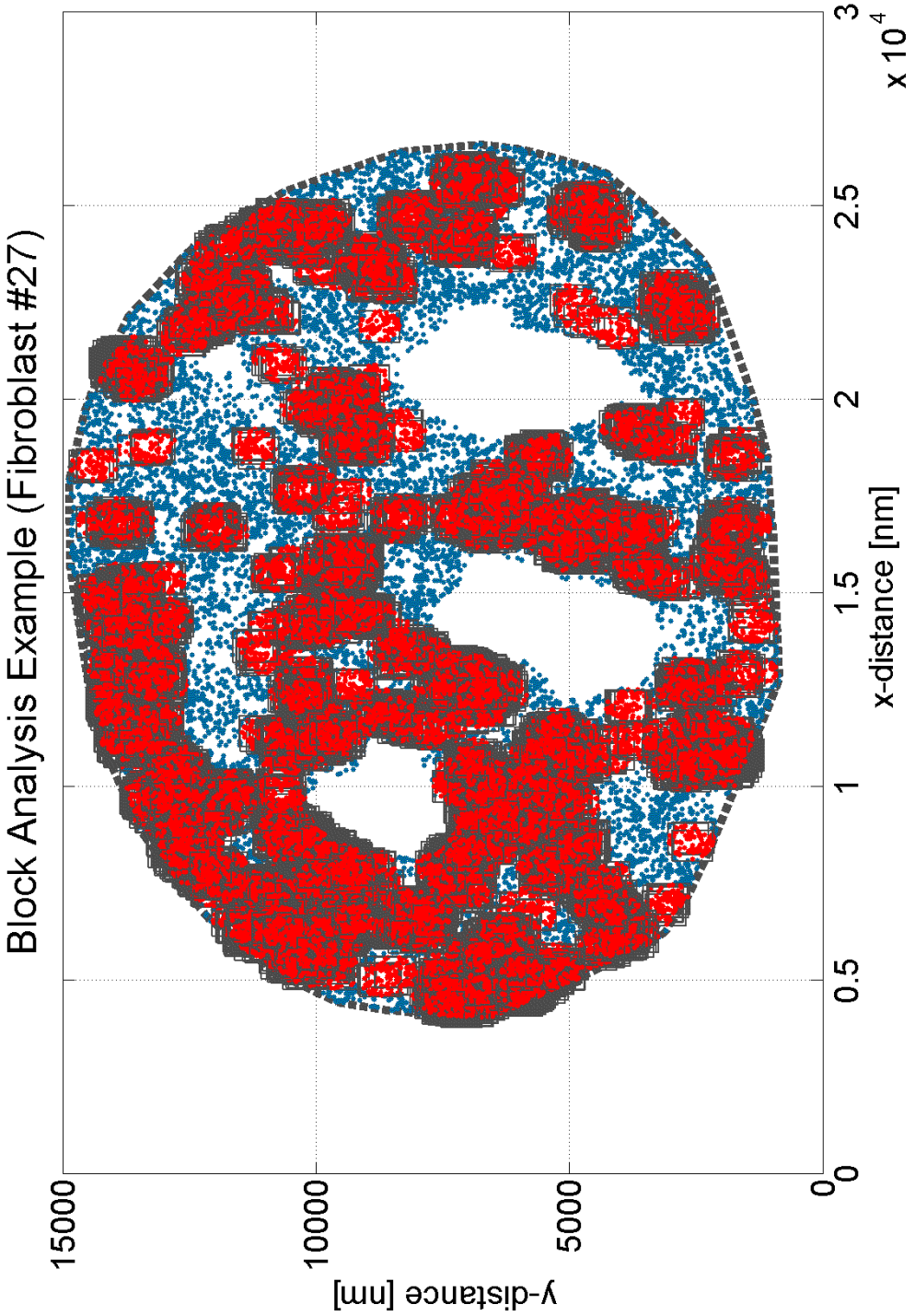


Figure 170: An example of a complete block analysis for the evaluation of the local distance distribution for a Fibroblast. A similar example for a HeLa nucleus is shown in Fig. 171.

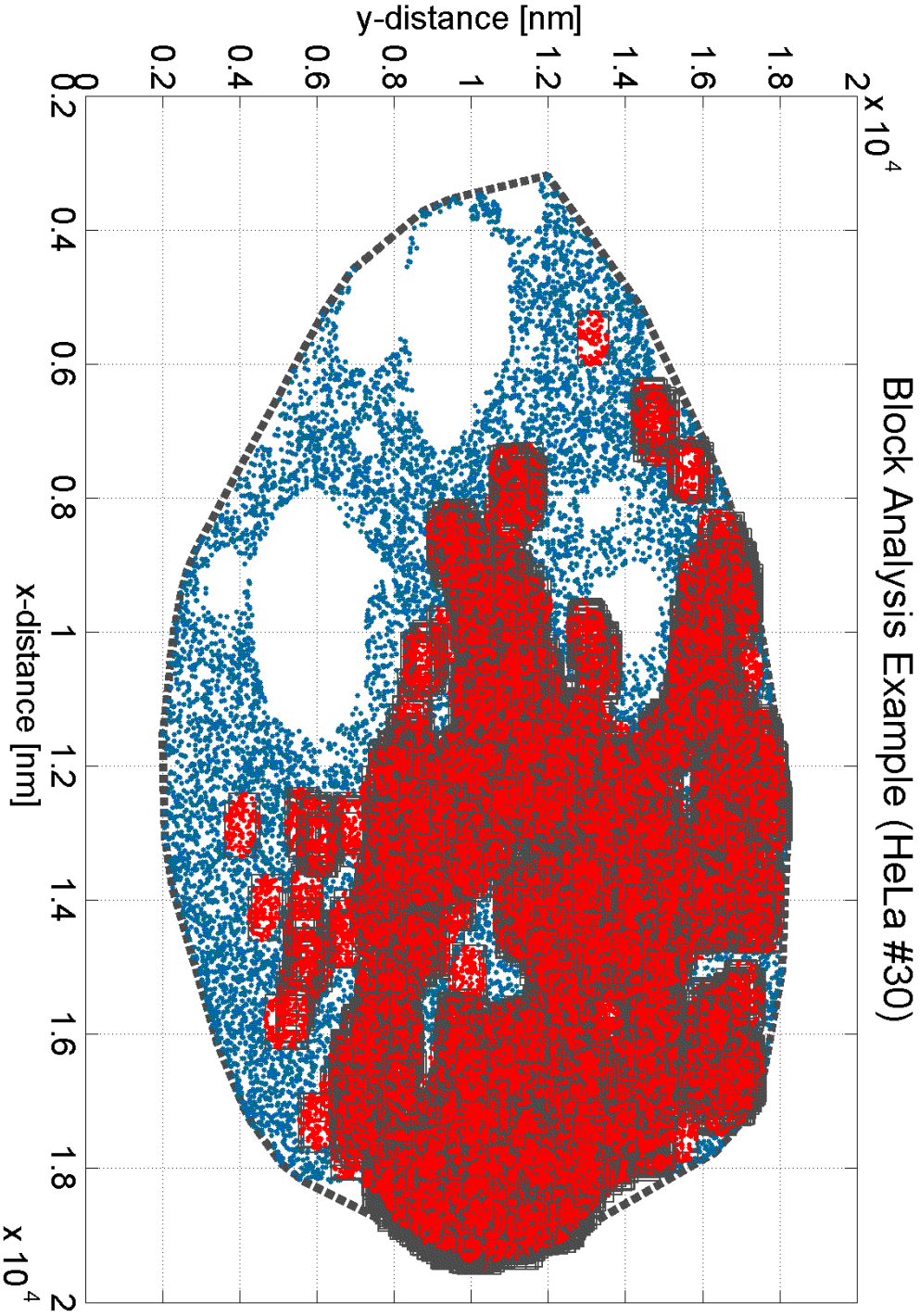


Figure 171: An example of a complete block analysis in the case of a HeLa nucleus.

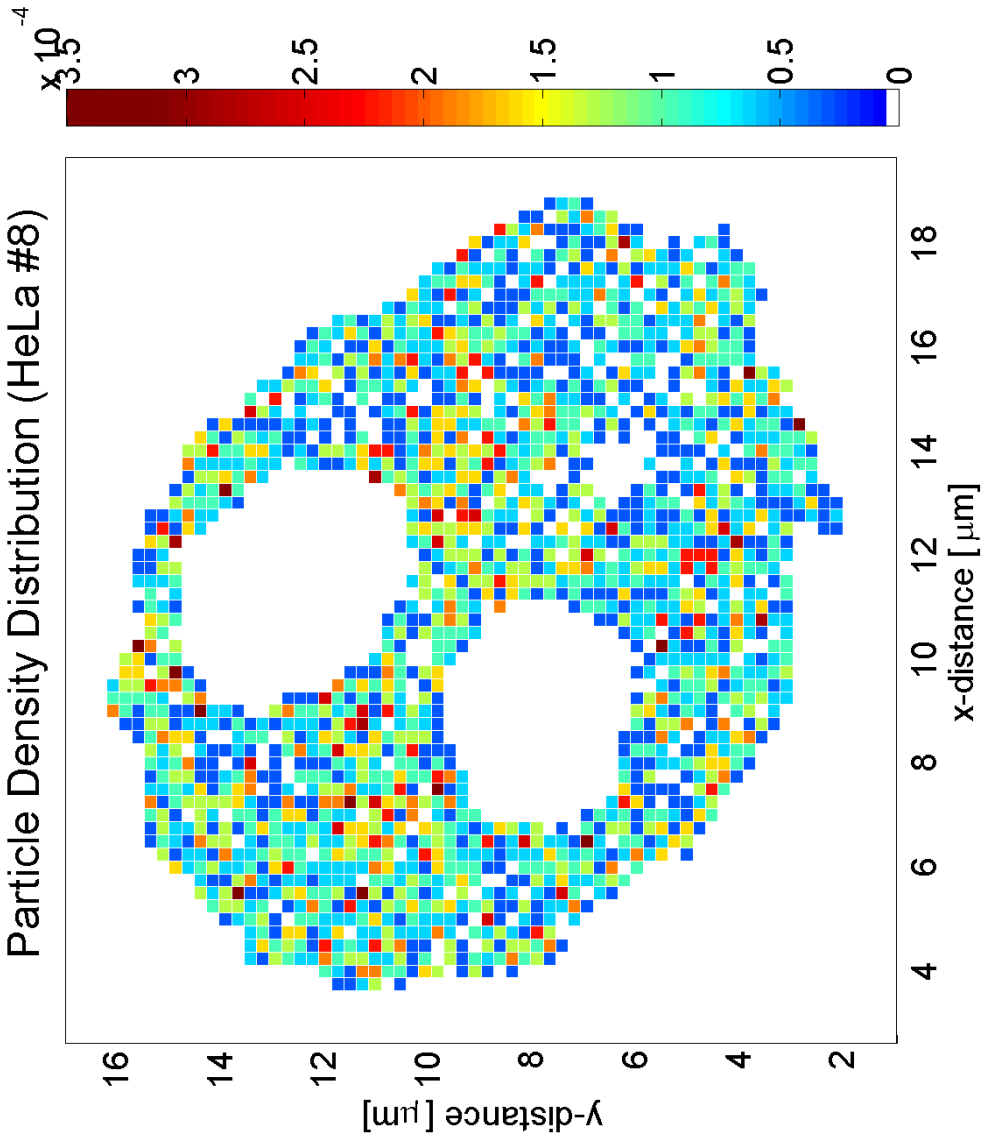


Figure 172: This figure shows the histone H2B density of a HeLa nucleus as a heat map. The units of the colorbar are particles per nm^2.

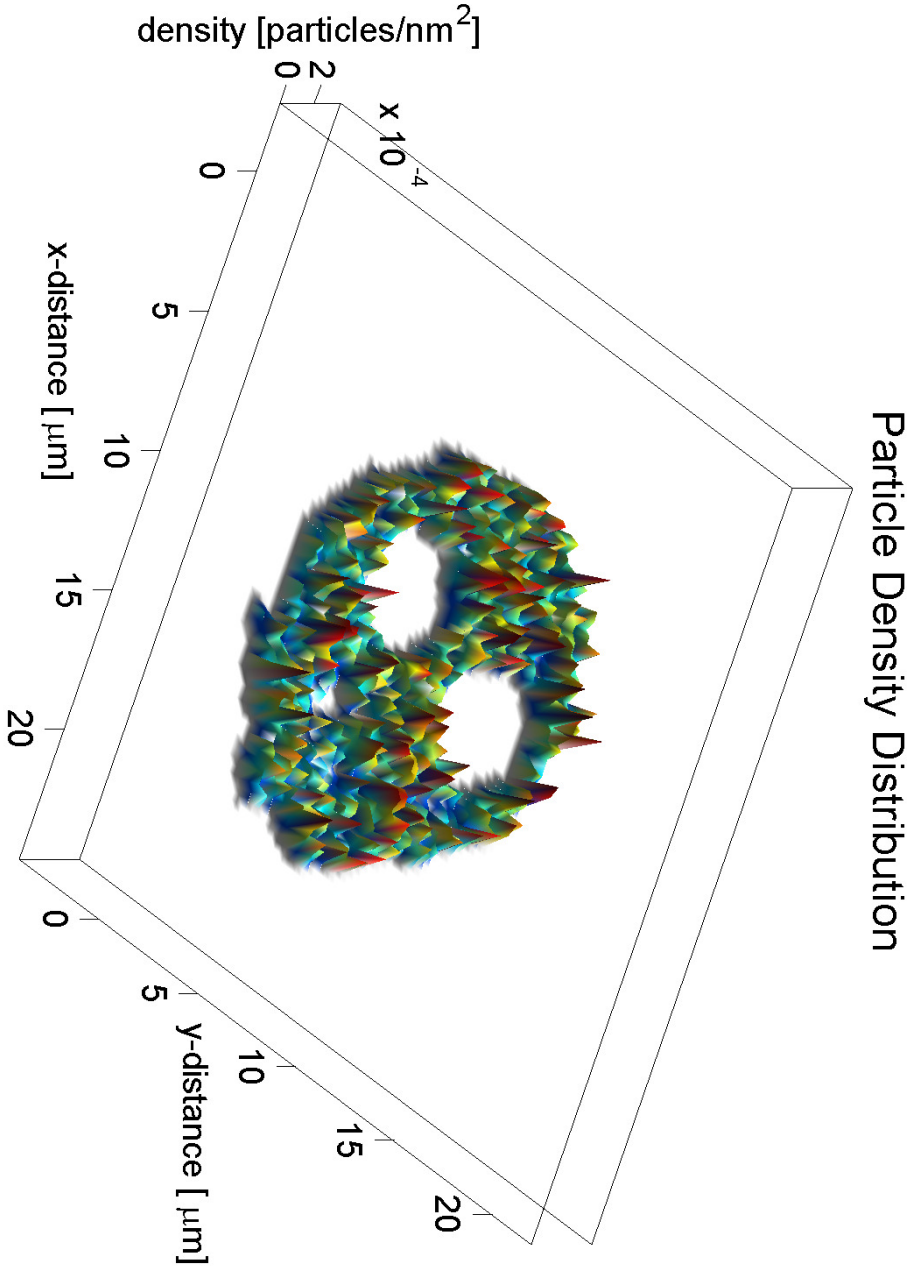


Figure 173: The histone H2B and thus chromatin density of a noise-reduced SPDM image in the case of a HeLa cell.

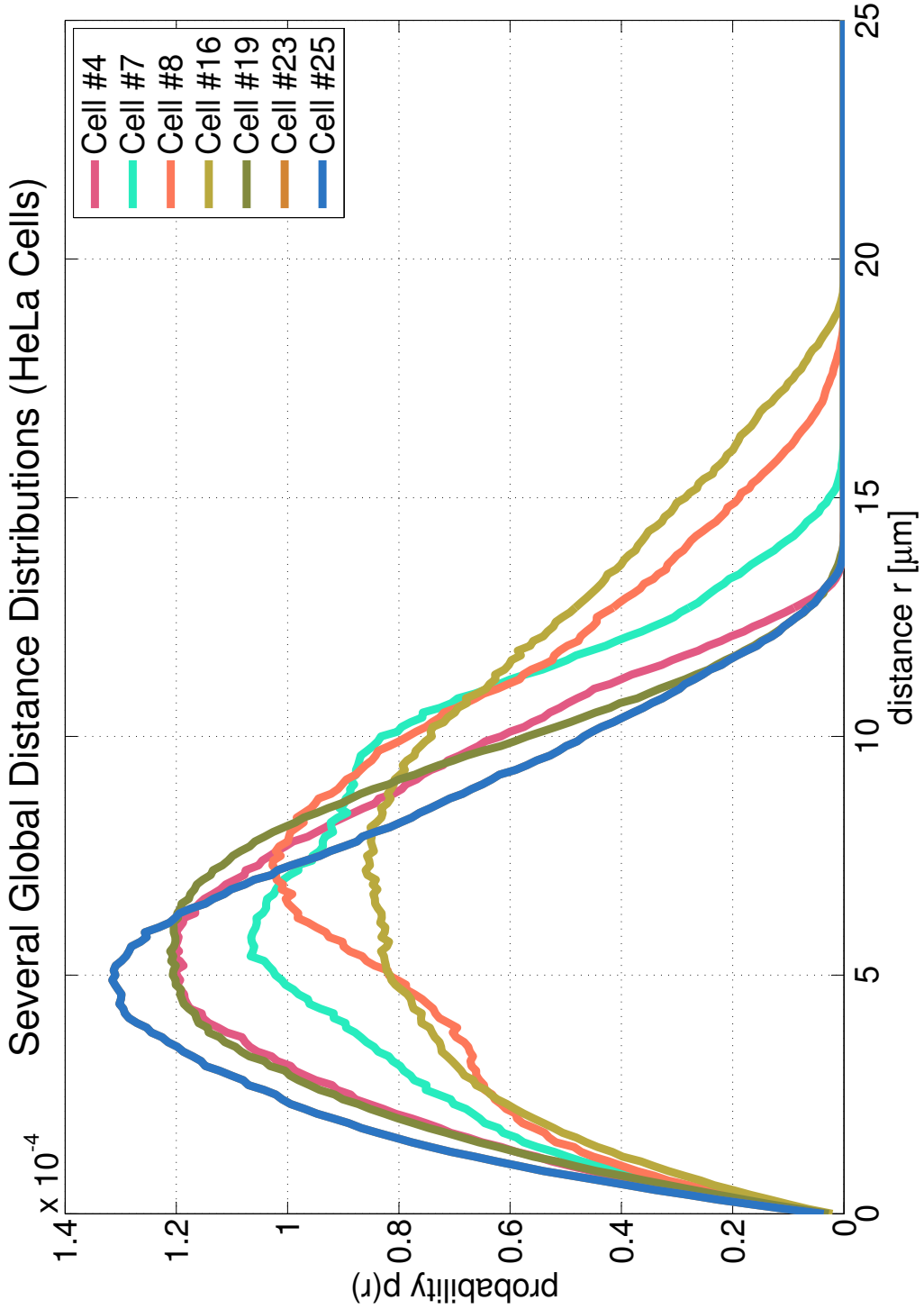


Figure 174: Several distance distributions within different HeLa nuclei. The distributions are different since the shape of the nuclei as well as the positions of the nucleoli are different, too.

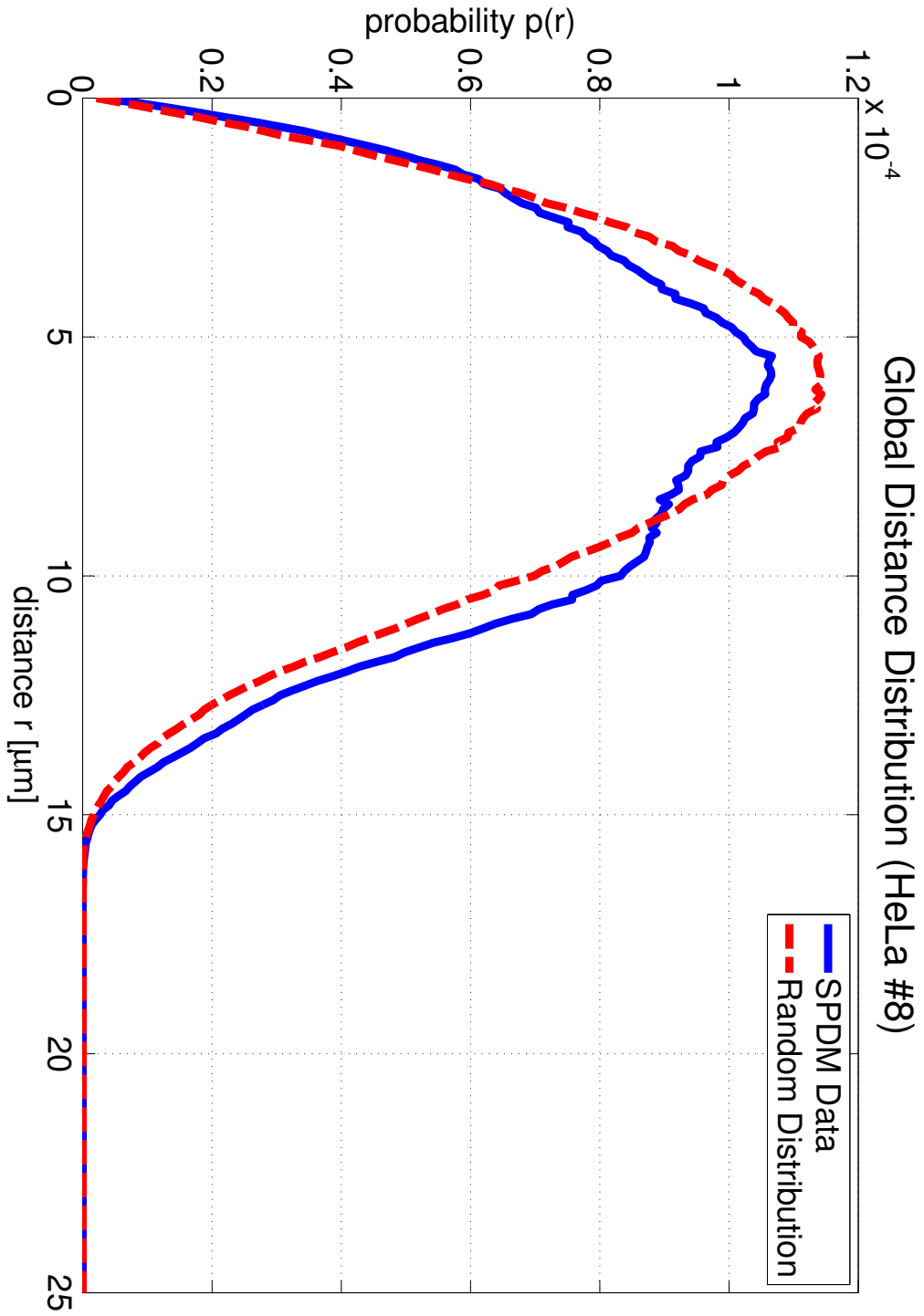


Figure 175: A comparison of a measured H2B distance distribution and a random particle distribution within an identically shaped nucleus. The differences come mainly from the influence of the nucleoli.

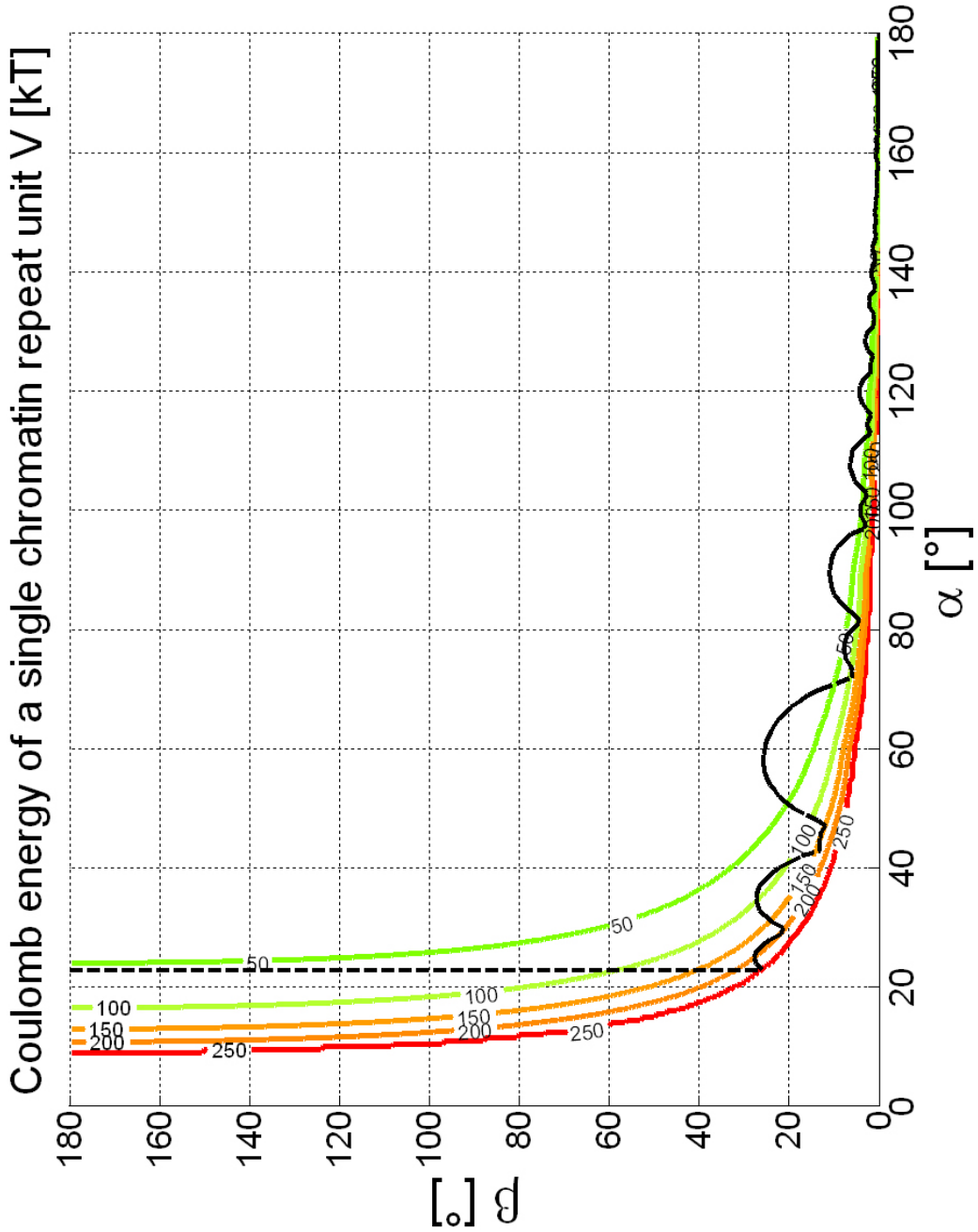


Figure 176: The Coulomb repulsion of the DNA linkers near the excluded volume borderline (black). The colored lines are equipotential paths in the phase diagram.

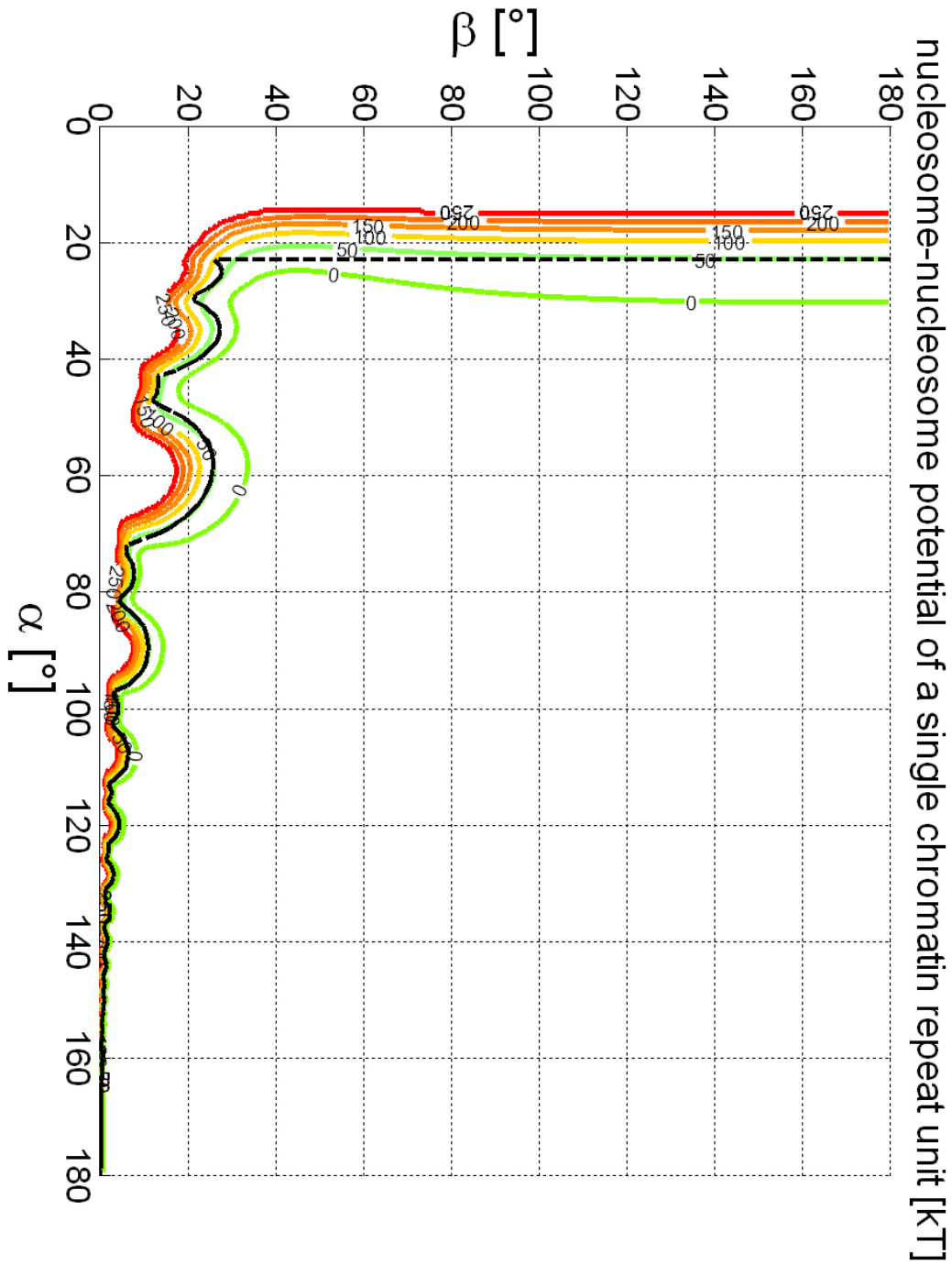


Figure 177: The nucleosome-nucleosome interaction of the chromatin fibers. One can see that the excluded volume borderline (black) is reproduced by the equipotential lines and thus confirmed.

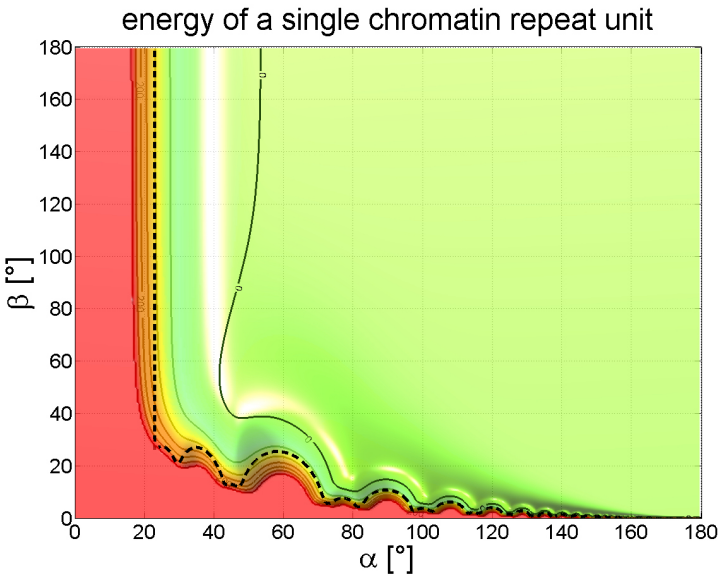


Figure 178: Projection of the total energy ϵ of a single chromatin repeat unit. The dashed line gives the excluded volume phase transition and the black lines are the projections of the equipotential lines.

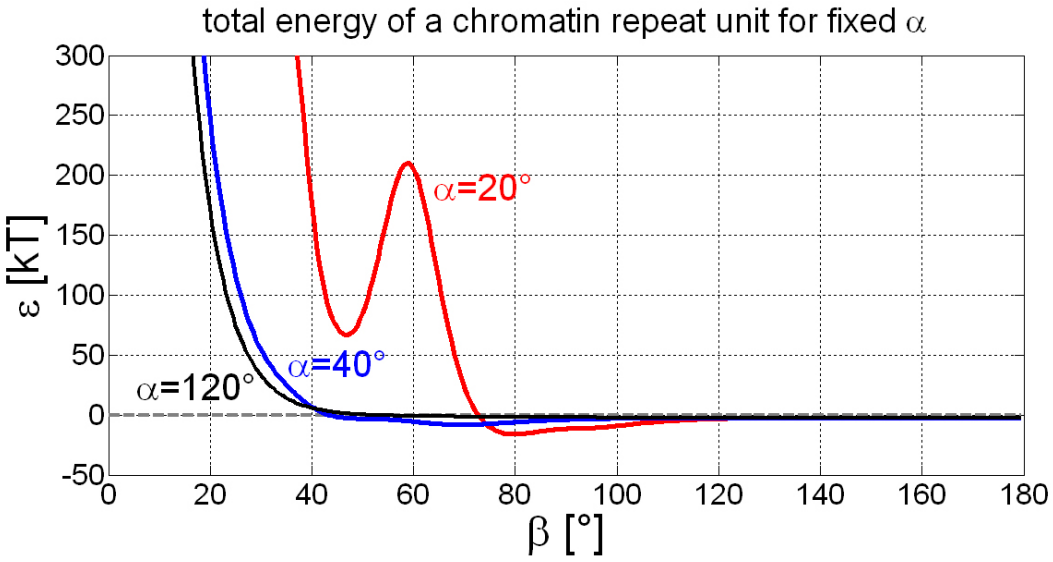


Figure 179: A cut through the figures above. One can see a potential well which has the shape of the excluded volume borderline. The potential at $\alpha = 20°$ is completely forbidden by the excluded volume borderline.

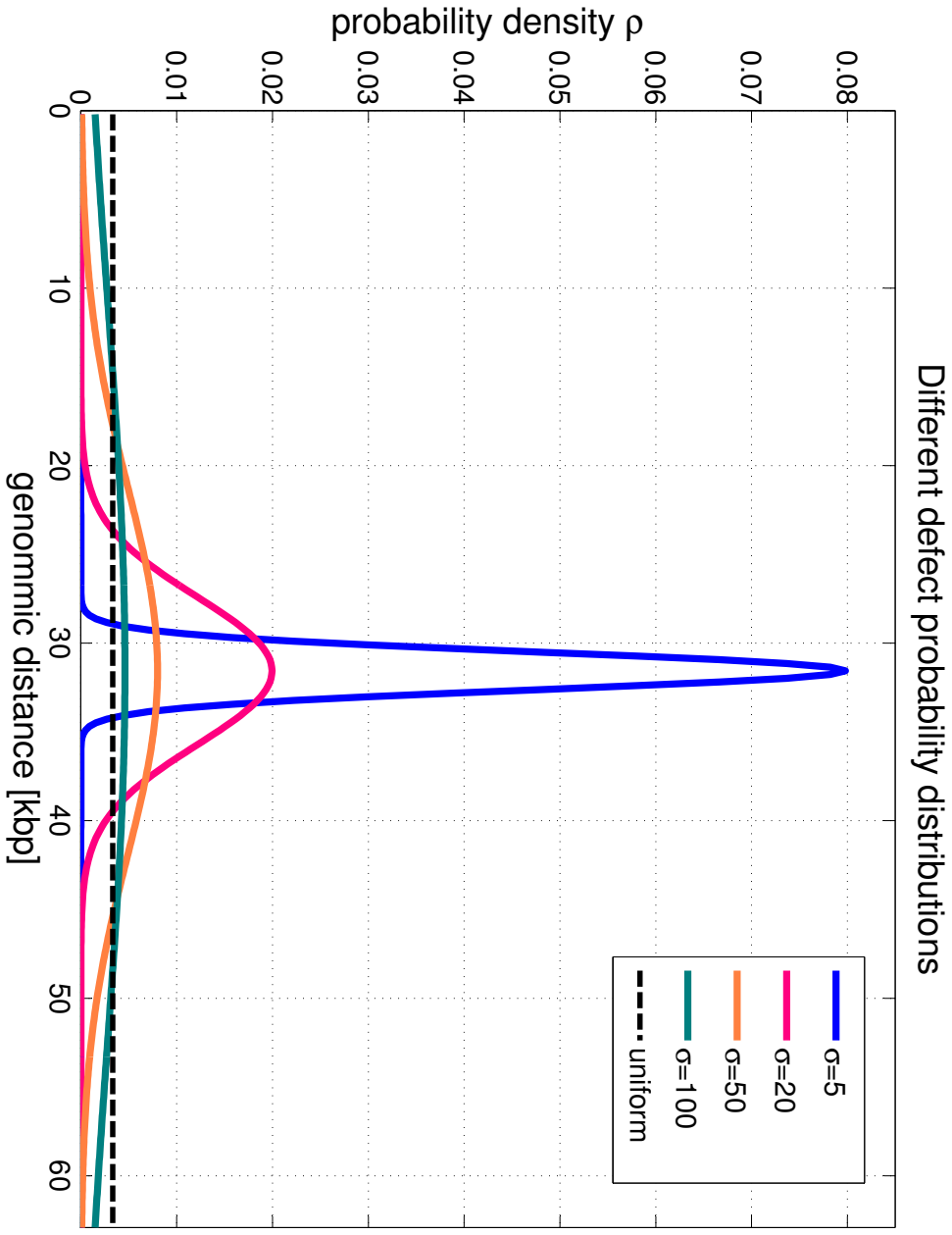


Figure 180: The different probability density functions that have been used for the distribution of the linker histone skips in Sec. 6.2.2. The probability distributions are plotted along the maximum fiber length used in Fig. 66.

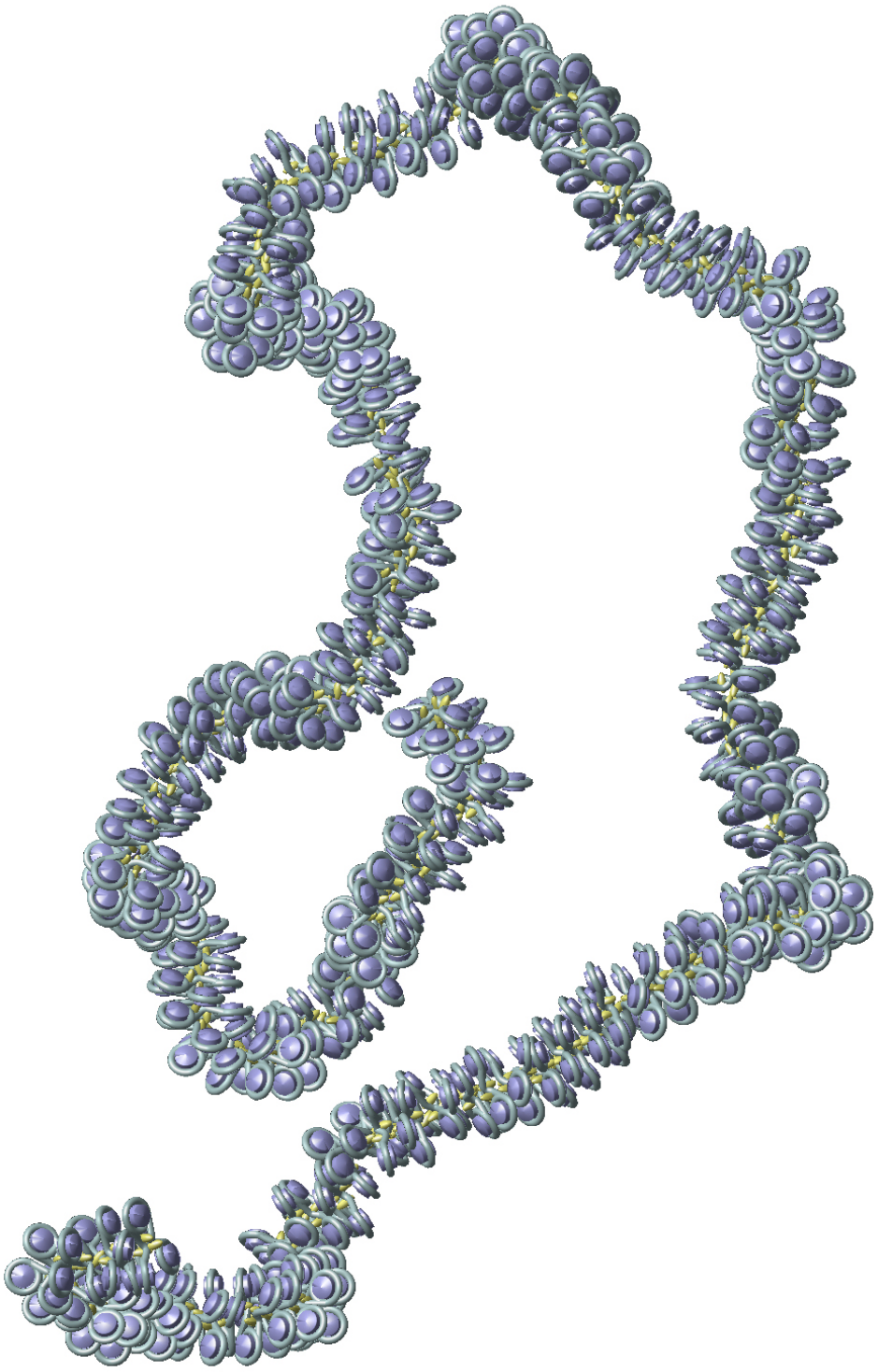


Figure 181: An example conformation of a flexible chromatin fiber (cf. Sec. 6.3). The fiber has a length of 800 nucleosome repeat lengths, i.e. ≈ 158kbp.

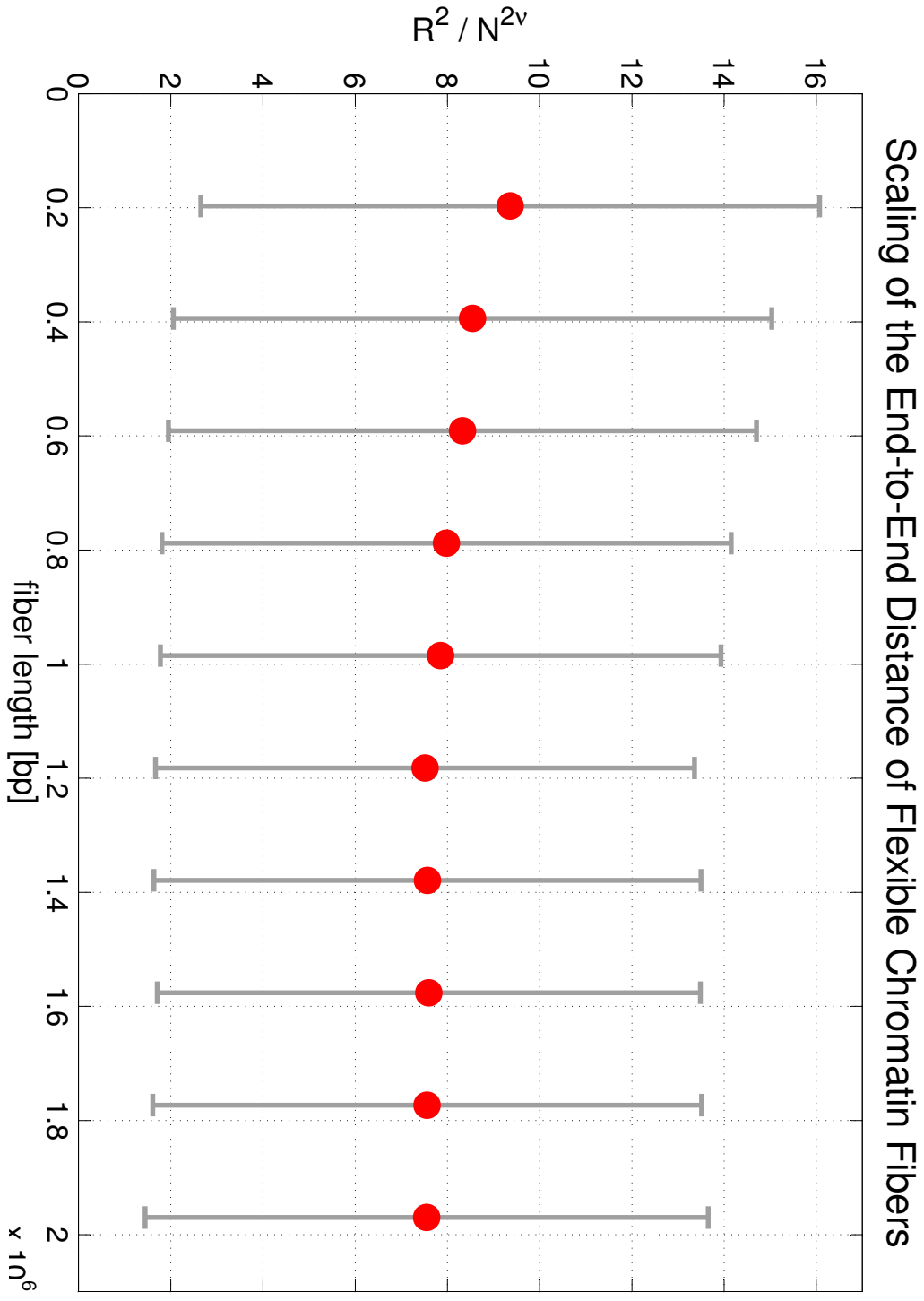


Figure 182: This figure shows the mean squared end-to-end distance for flexible chromatin fibers. Here the factor $N^{2\nu}$ ($\nu \approx 0.589$) is divided out. One can see that on large length scales the chromatin fibers behave like a self-avoiding random walk.

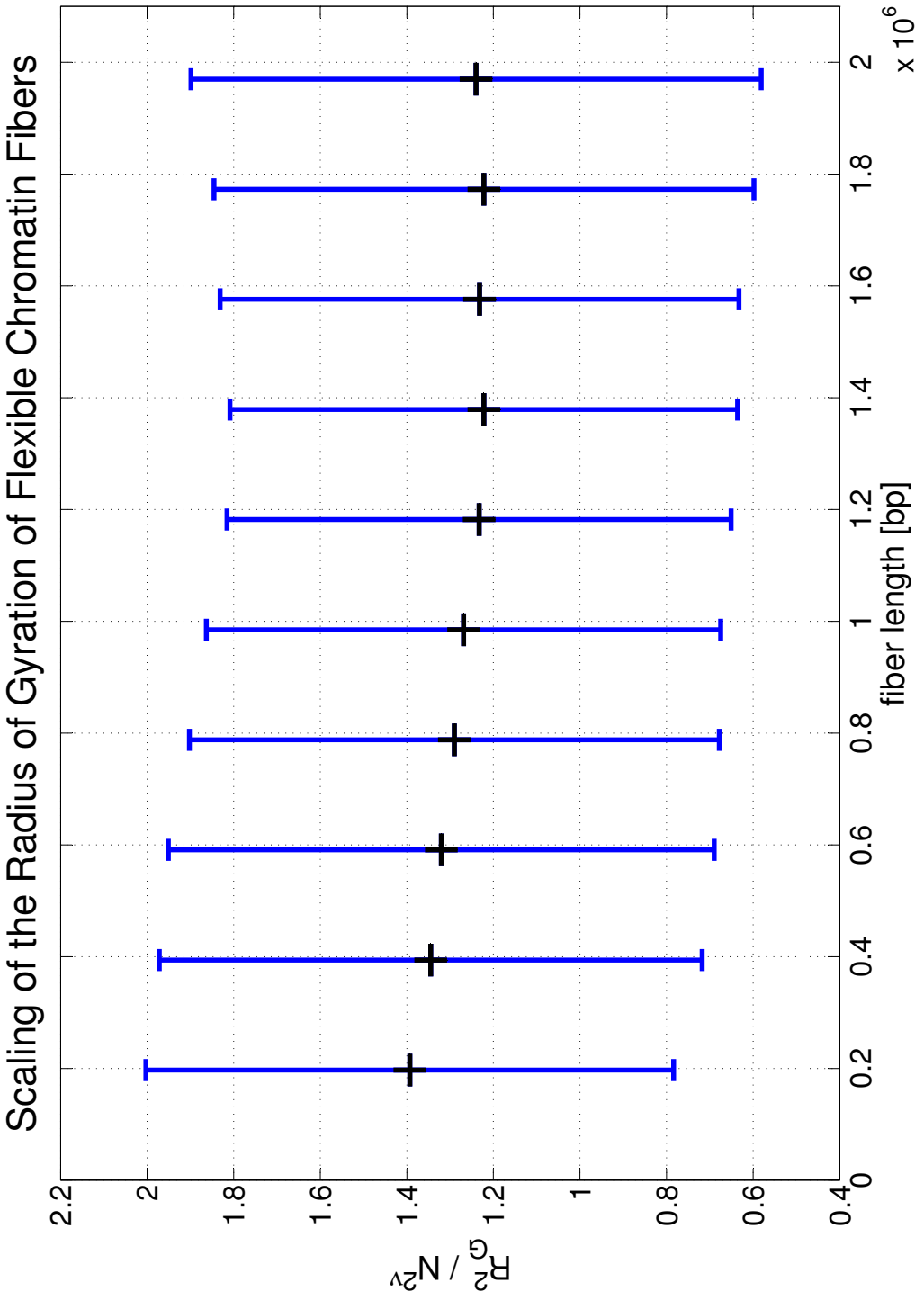


Figure 183: This figure shows the mean squared radius of gyration for flexible chromatin fibers. Here the factor $N^{2\nu}$ ($\nu \approx 0.589$) is divided out. One can see that on large length scales the chromatin fibers behave like a self-avoiding random walk.

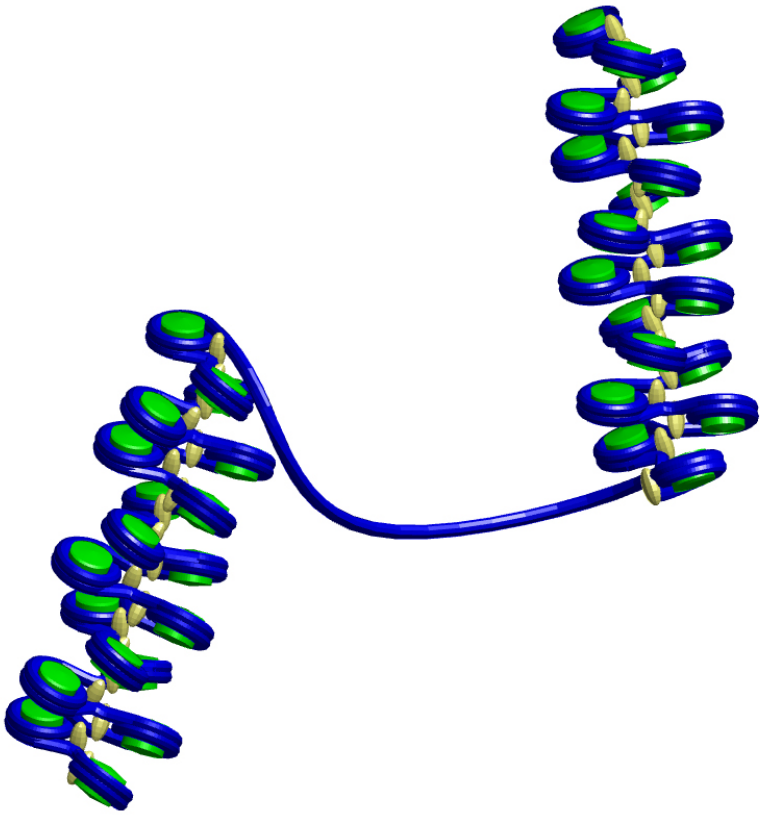


Figure 184: Another example of a single nucleosome skip (cf. Sec. 6.4).

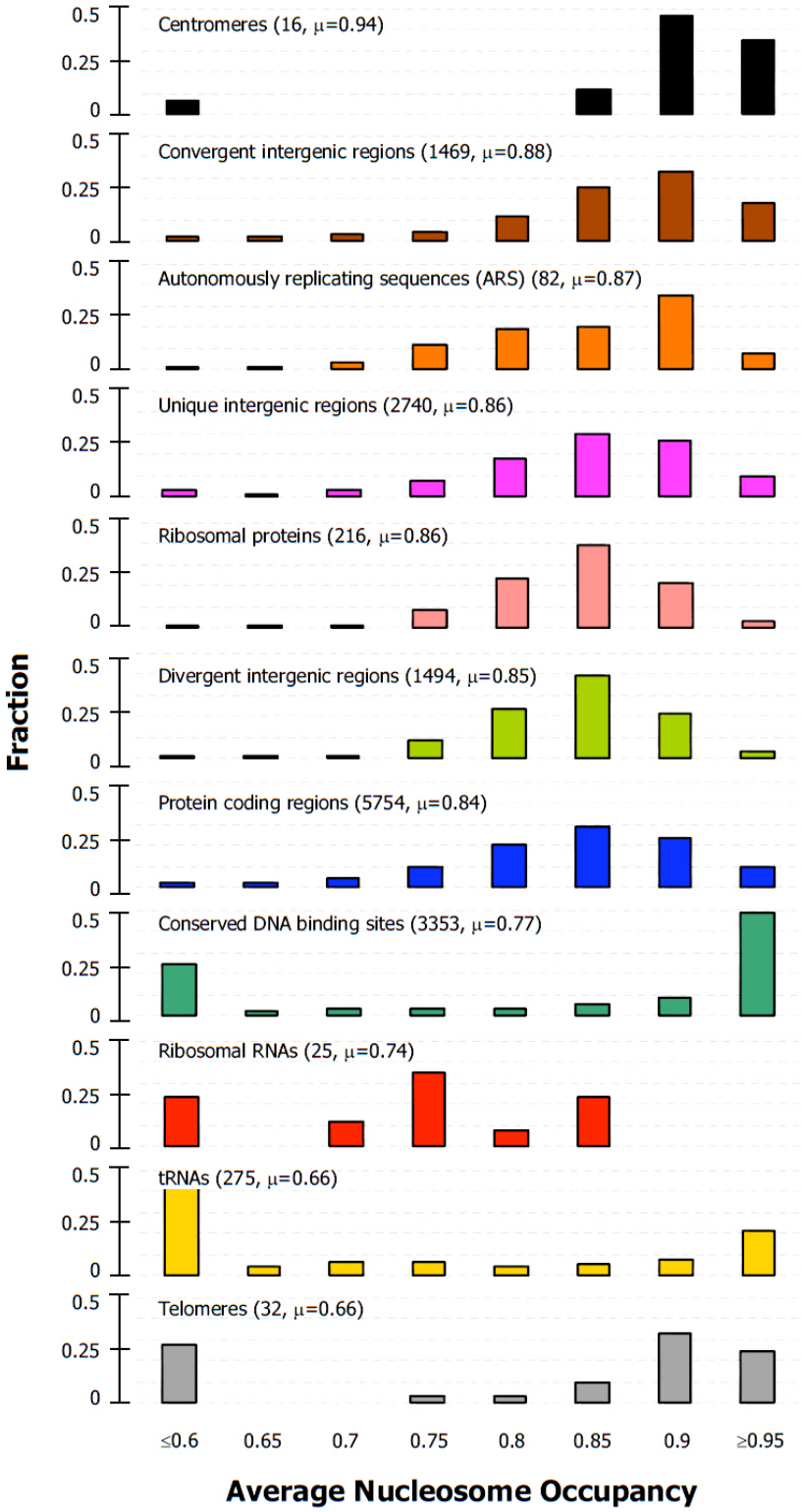


Figure 185: Intrinsic nucleosome occupancy varies with genomic location type. Histogram of average predicted occupancies at different genomic locations. The leftmost bin includes all elements with occupancy below 0.6. The number of elements and average occupancy for each type of genomic region are indicated. Adapted from [186].

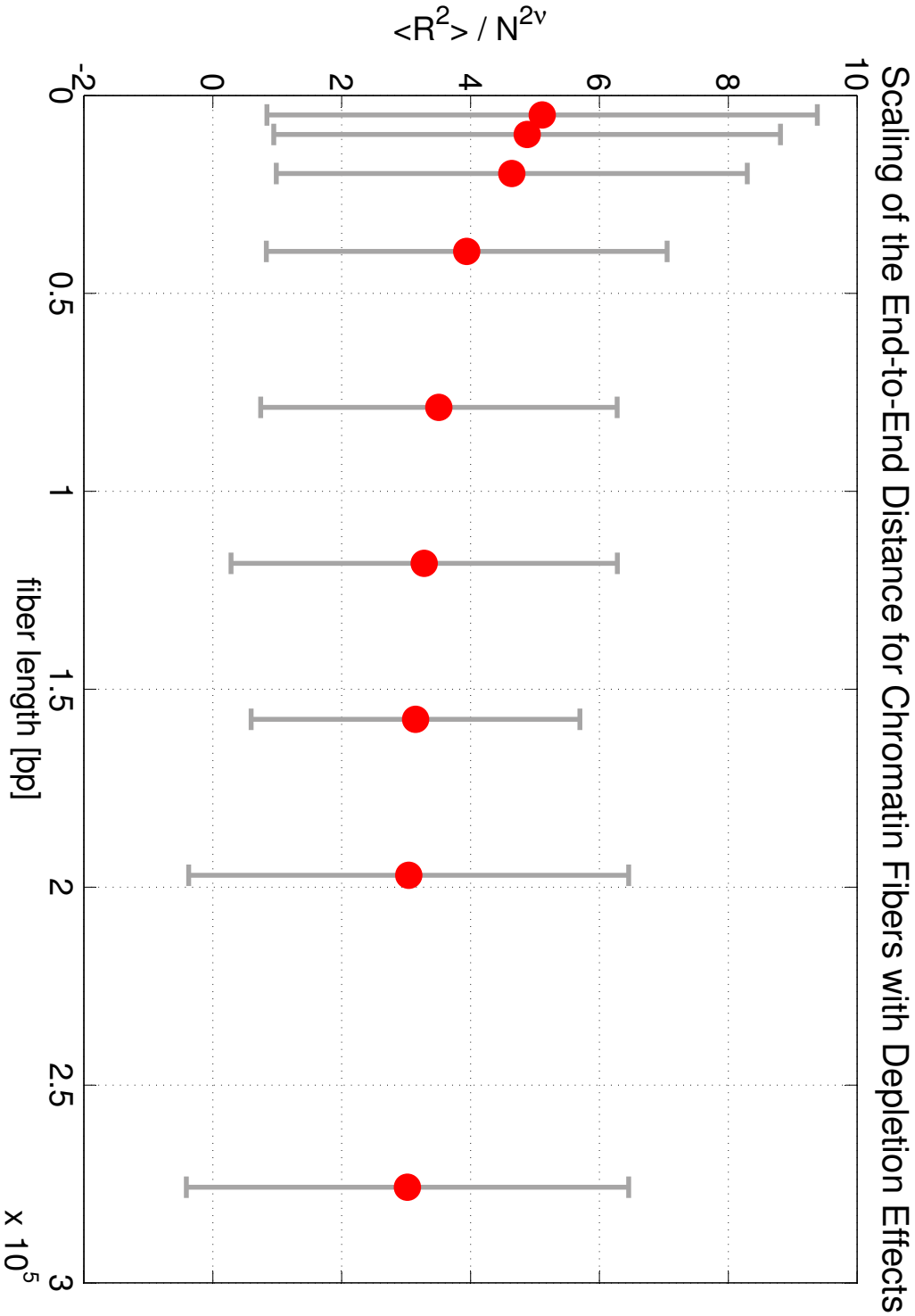


Figure 186: This figure shows the mean squared end-to-end distance for flexible chromatin fibers with depletion effects. Here the factor $N^{2\nu}$ ($\nu \approx 0.589$) is divided out. One can see that on large length scales the chromatin fibers behave like a self-avoiding random walk.

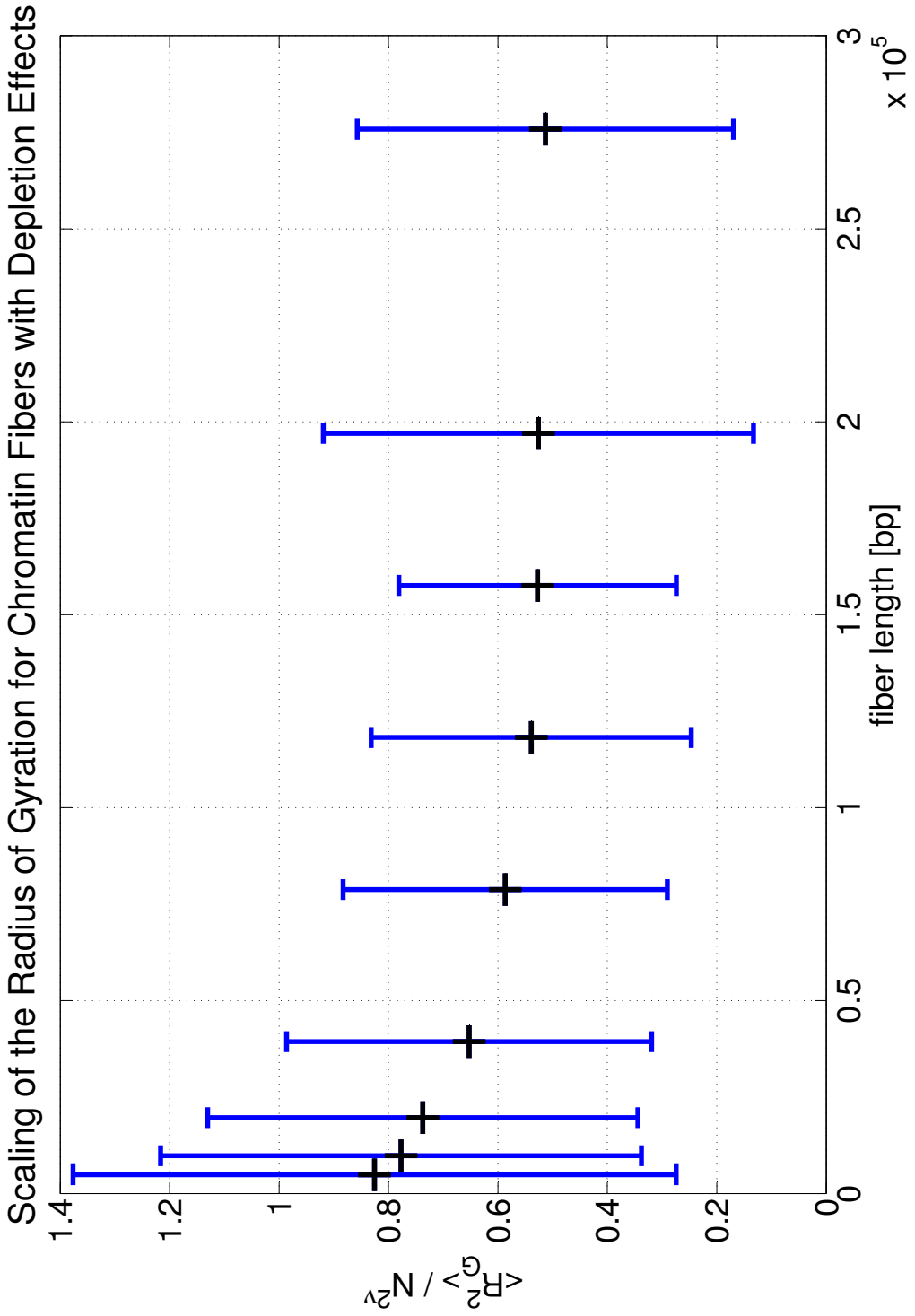


Figure 187: This figure shows the mean squared radius of gyration for flexible chromatin fibers with depletion effects. Here the factor $N^{2\nu}$ ($\nu \approx 0.589$) is divided out. One can see that on large length scales the chromatin fibers behave like a self-avoiding random walk.

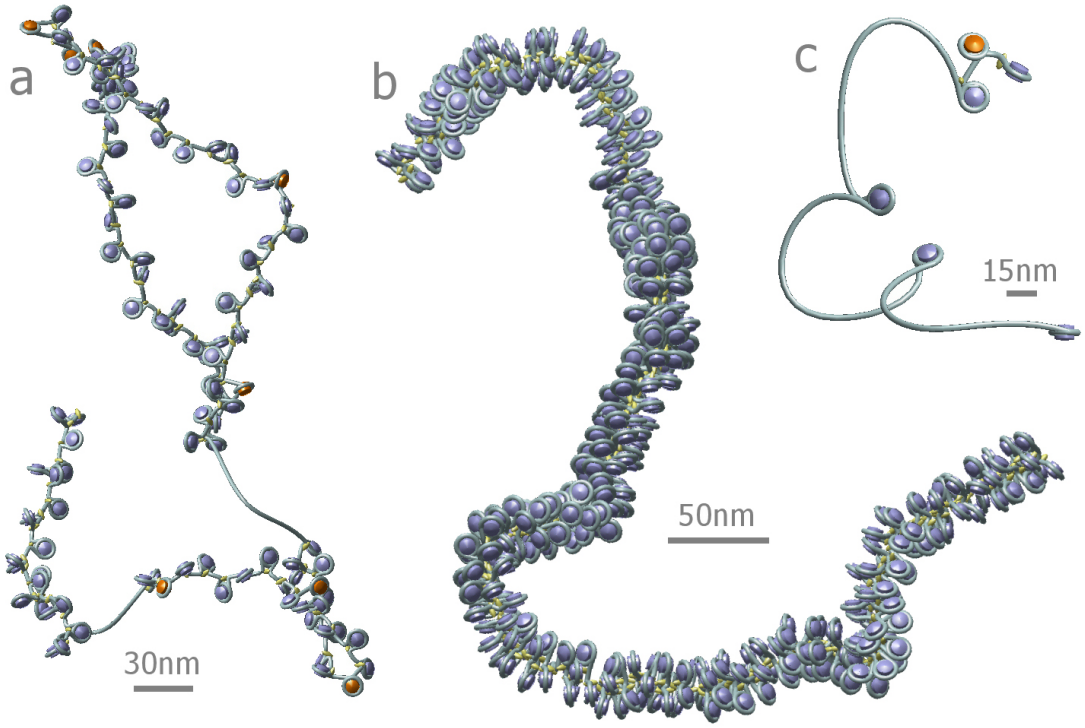


Figure 188: Some example conformations that can be found in the model, if one takes the parameter distributions as well as the depletion effects into account. Similar electron micrographs can be found in [**?**]. a) Shows a beads-on-a-string-like structure with some nucleosome and linker histone skips. The latter are marked orange. b) Example conformation of a chromatin strand of length 40kbp. The light blue tubes represent the DNA, the histone octamers are modelled as purple cylinders and the linker histones are marked pale yellow. This chromatin conformation has no depletion effects and a diameter of about 34nm. c) One can even get single nucleosomes connected by blank DNA stretches, which is shown here.

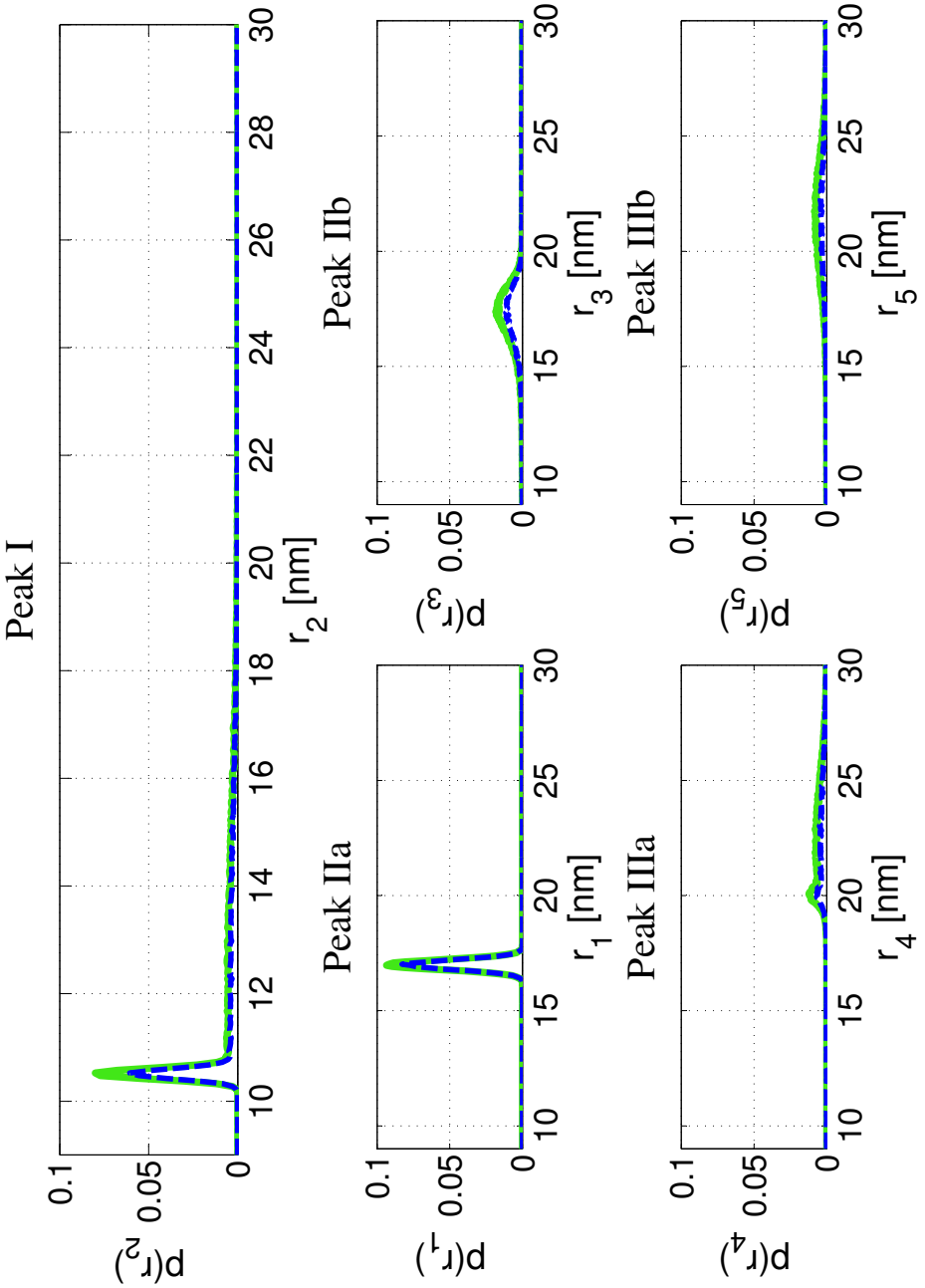


Figure 189: The r_Δ-distributions of the first three peaks of the radial nucleosome distance distribution function (cf. Sec. **??**). One can see that some of the distributions are very asymmetric.

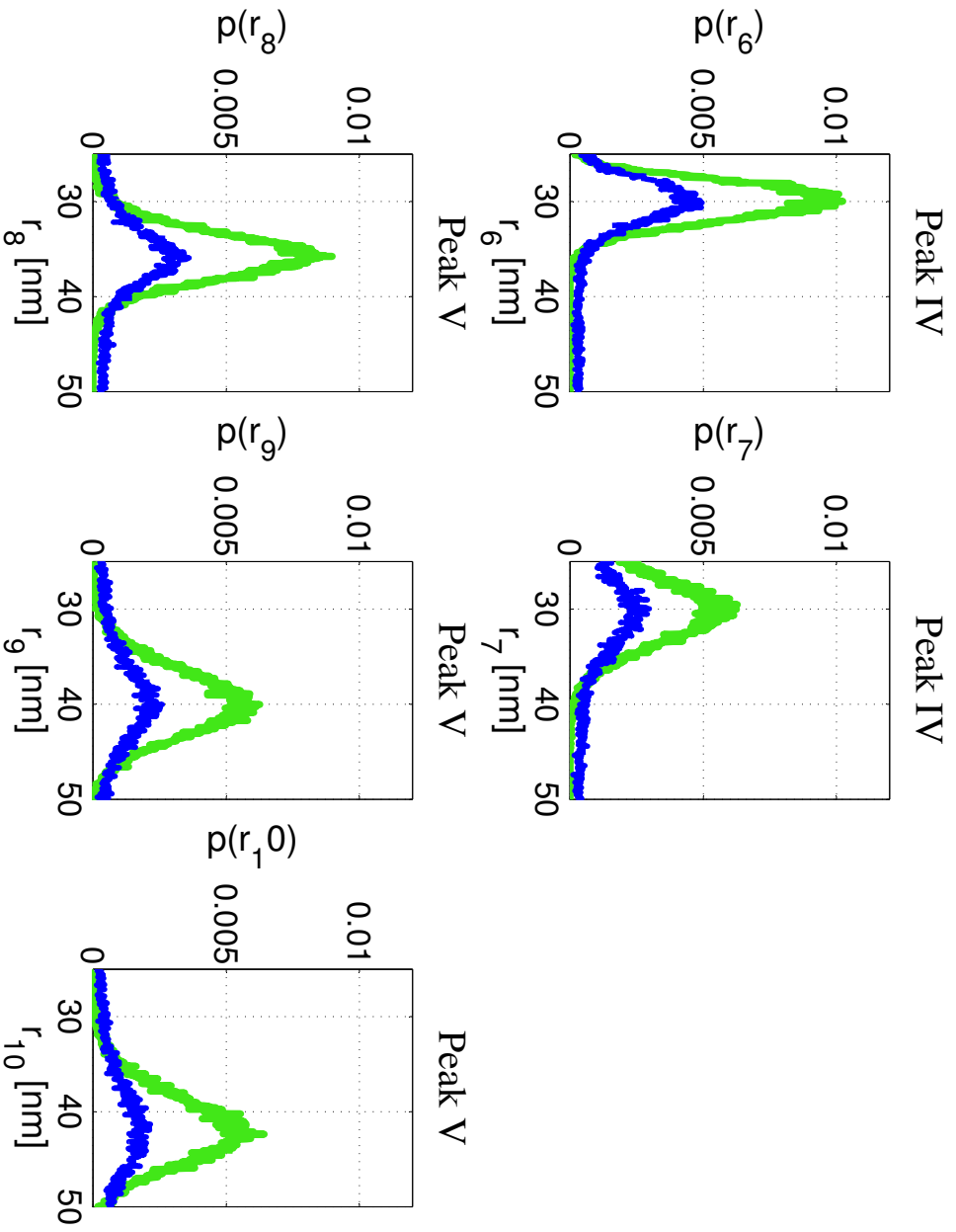


Figure 190: This figure shows the r_Δ-distributions that contribute to the fourth and fifth peak of the radial nucleosome distance distribution function (cf. Sec. **??**) and the nucleosome pair distribution function $g(r)$.

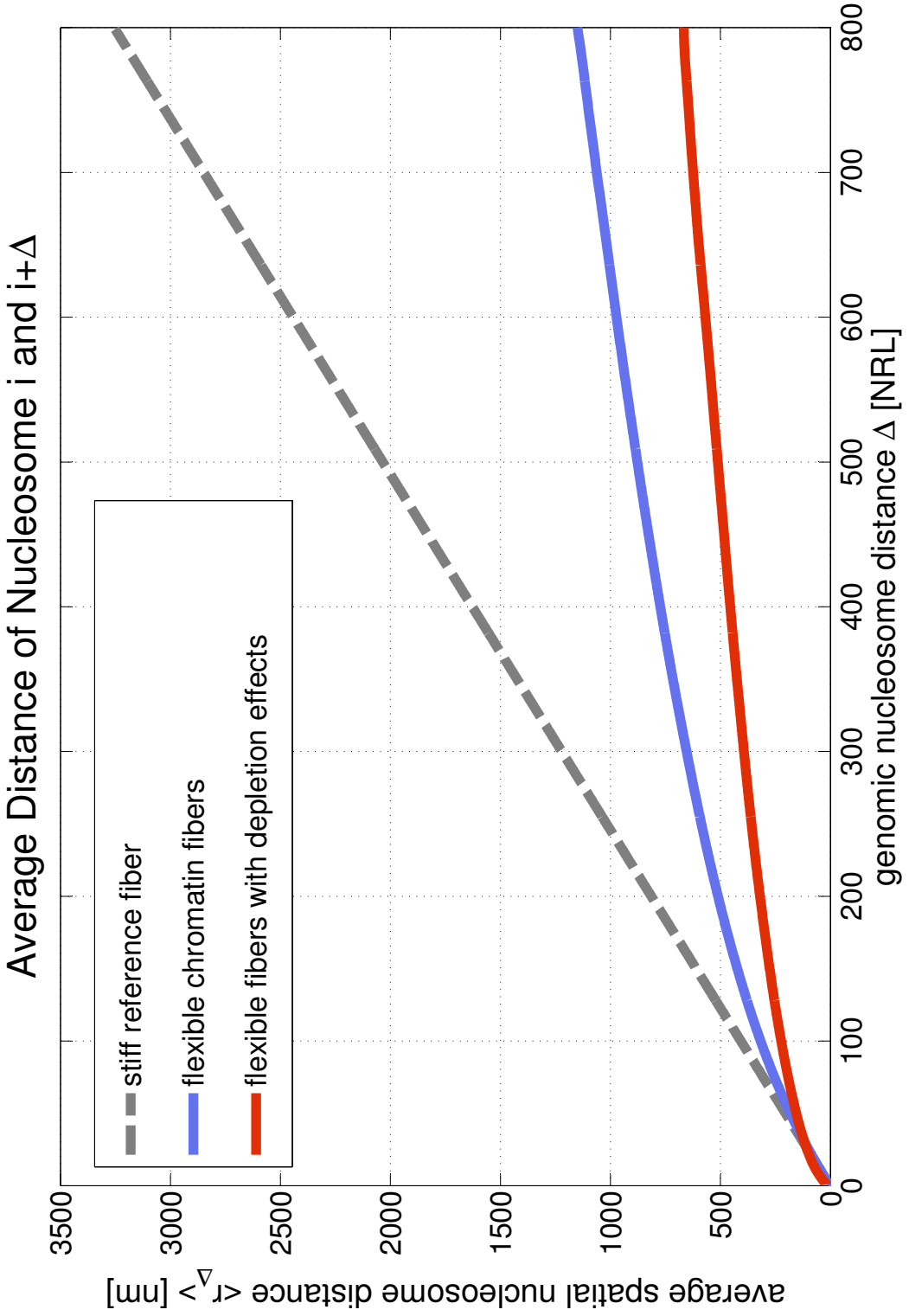


Figure 191: The mean spatial nucleosome distance $\langle r_\Delta \rangle$ in dependence on the genomic distance Δ for stiff, flexible fibers as well as flexible fibers with histone depletion on the large scale.

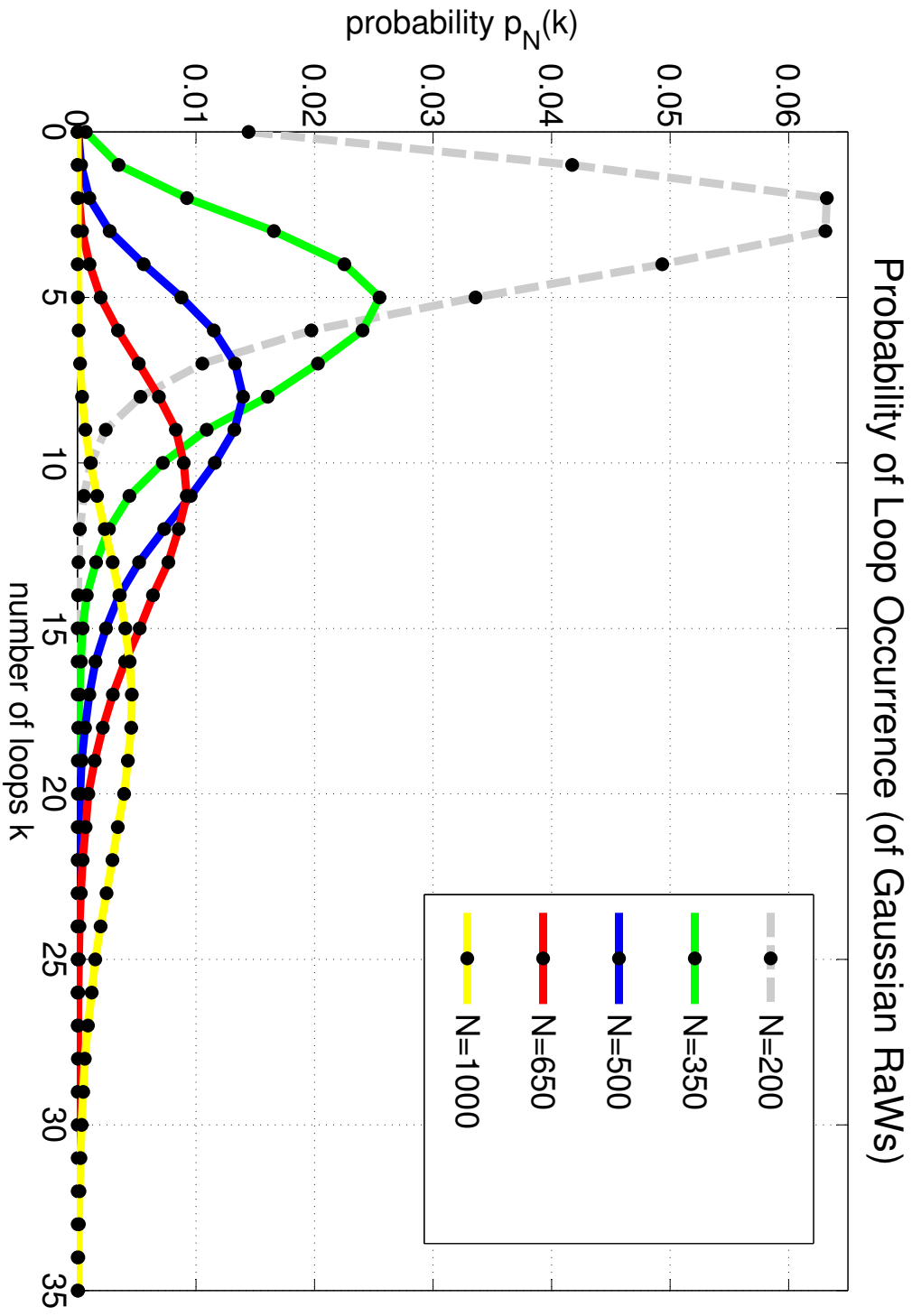


Figure 192: This figures shows how the distribution of the loop number in a random walk will change, if its length N is increased. One will see a similar effect if one decreases the persistence length of the walk.

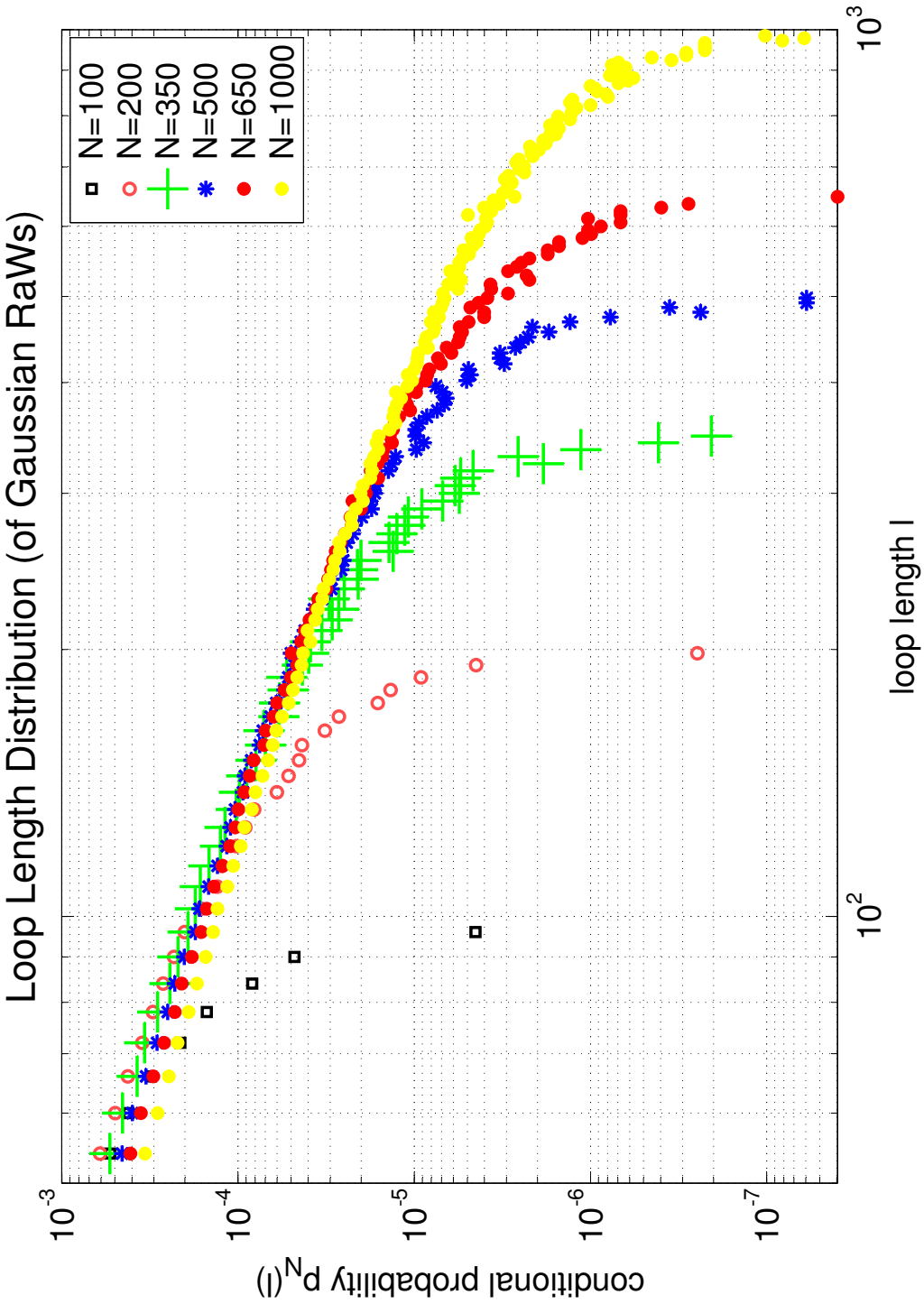


Figure 193: Shown is the distribution of the loop length l of several Gaussian random walks. The loop length probability follows the scaling law $P(l) \propto l^{-c}$ with $c = d/2 = 3/2$ in the case of a 3D random walk. For a self-avoiding walk the exponent would be $c = d\nu - \sigma_4$, i.e. 2.68 in the 2D case and 2.22 in the 3D case [215].

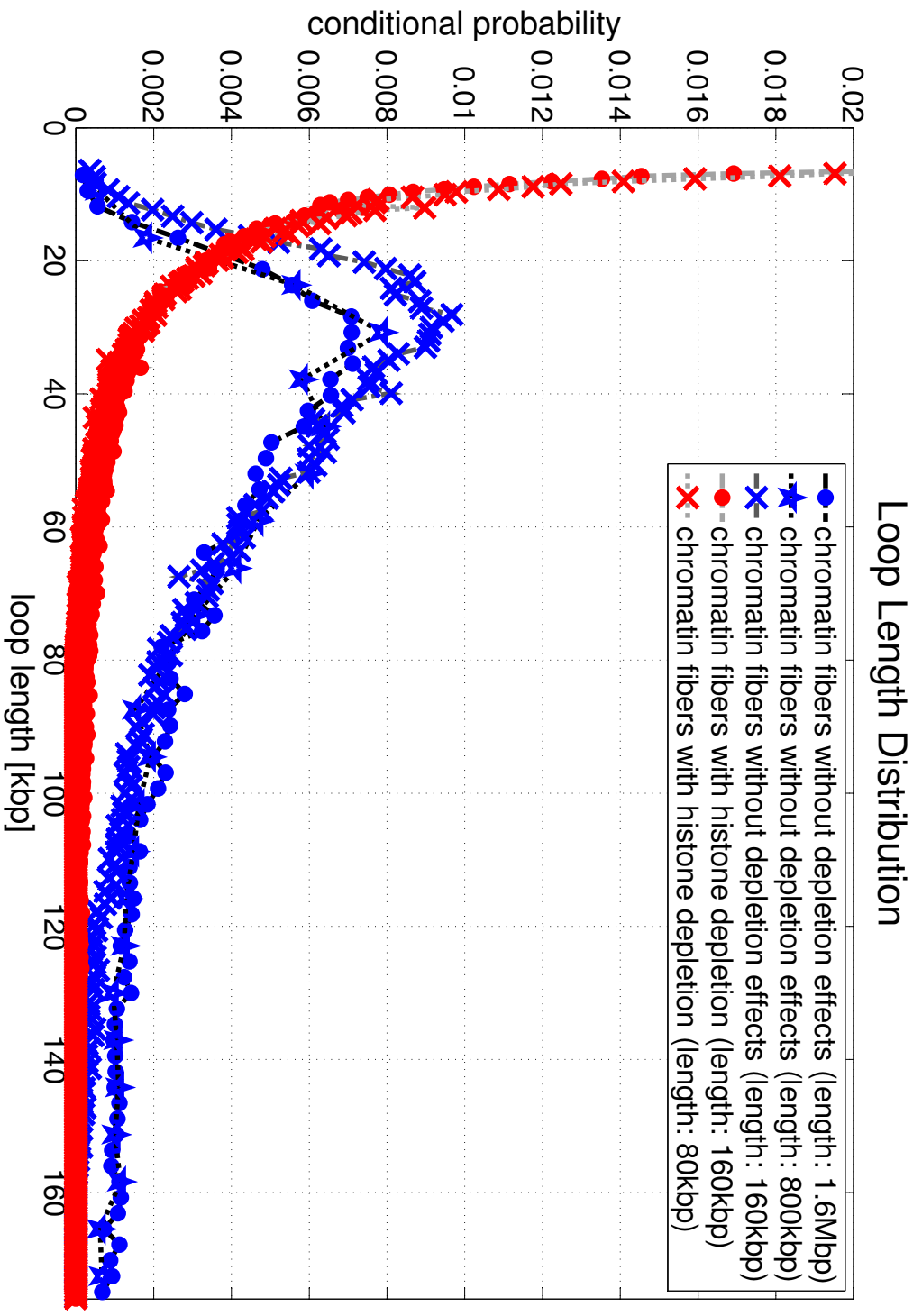


Figure 194: This figure shows the loop length distribution for chromatin strands with and without histone depletion. The fibers without depletion have lengths of 1.6Mbp, 800kbp and 160 kbp. The differences of the corresponding loop distributions come from the normalization of the curves since the fiber length limits the maximal loop length.

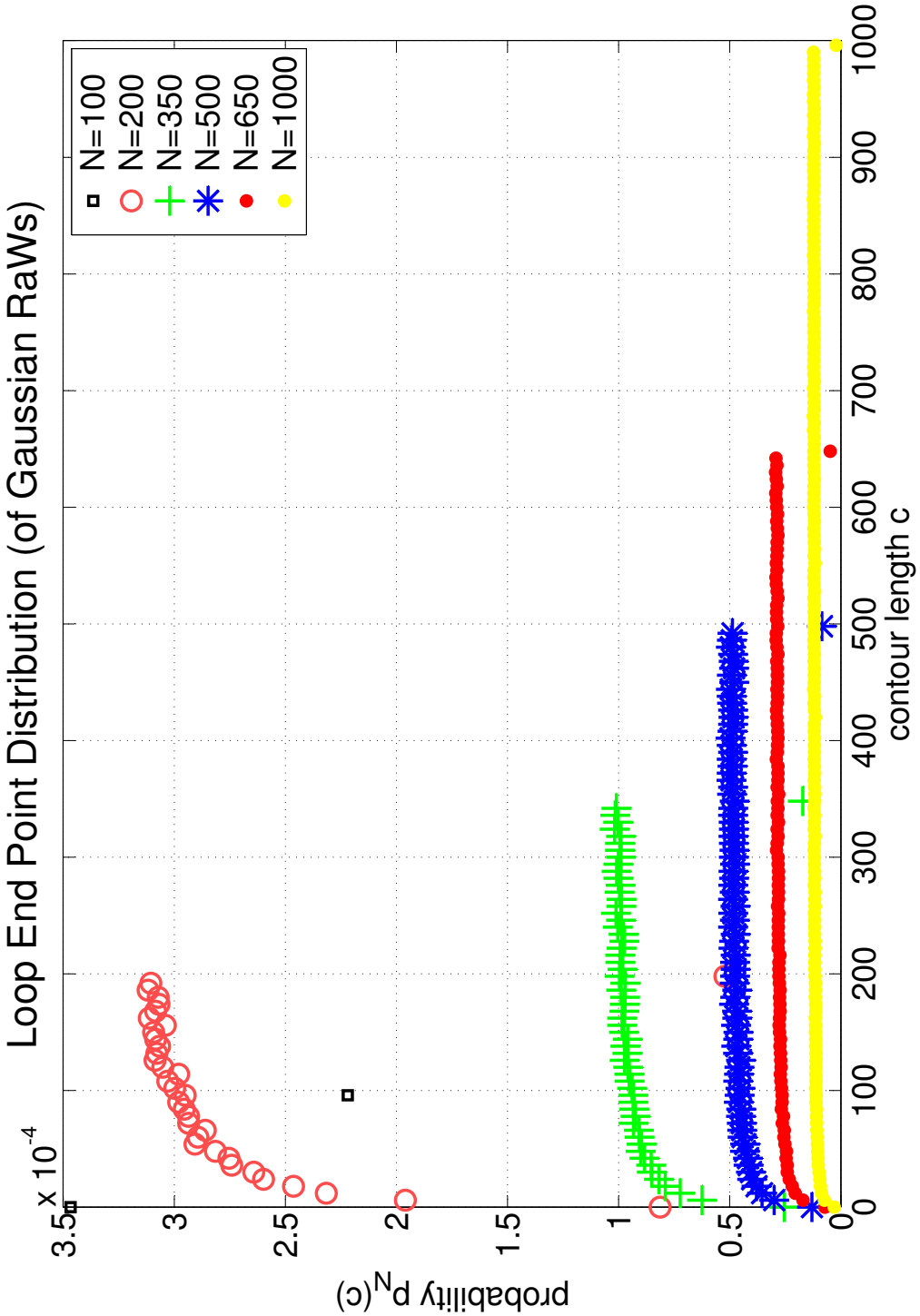


Figure 195: One can see that the distribution of the loop end position is almost uniform in the case of a random walk, i.e. non-interacting loops.

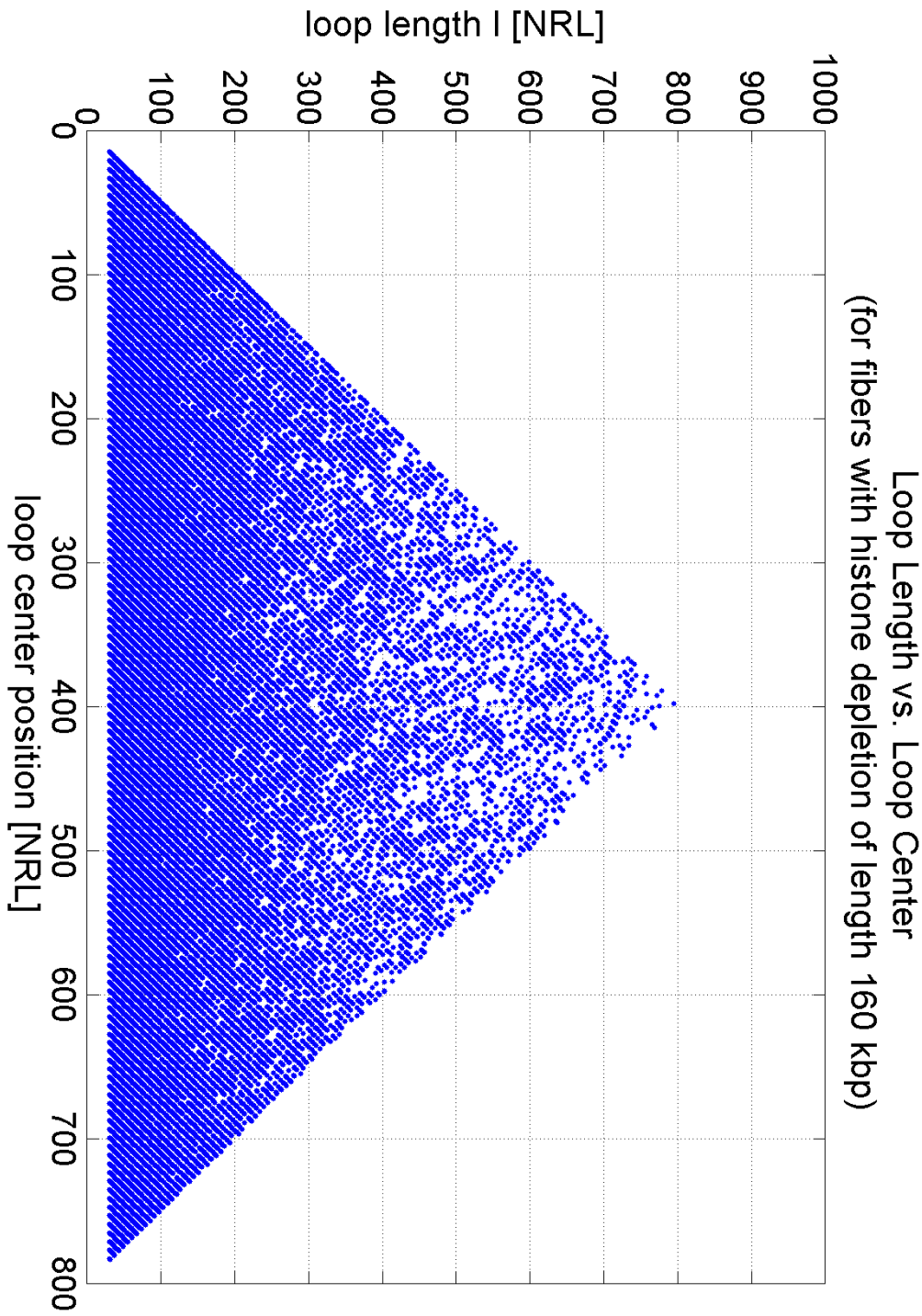


Figure 196: . This figure shows the dependence of the loop length on the loop center as a scatter plot. This is basically the sum of the distributions for the loop length in dependence on the loop start position and the loop end position.

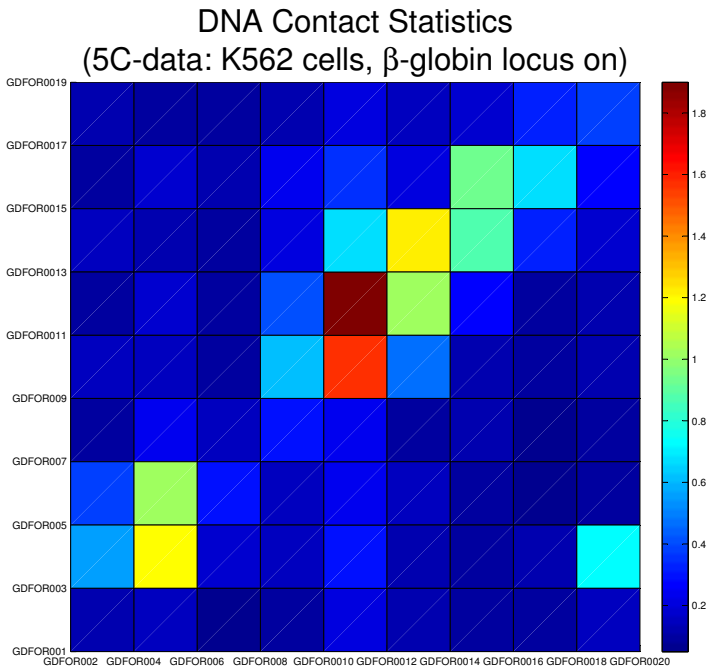


Figure 197: Heat map illustration of 5C-data (K562 cells) [72] with equidistant spacing.

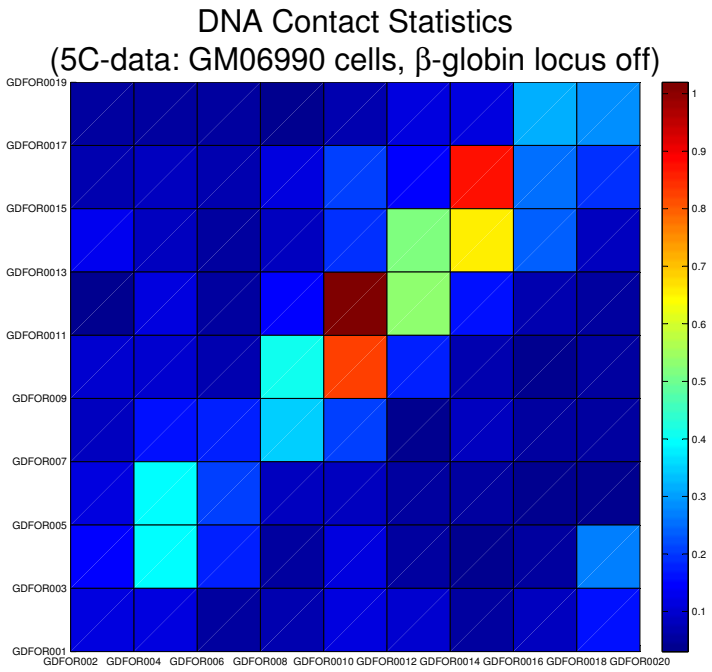


Figure 198: Heat map illustration of 5C-data (GM06990 cells) [72] with equidistant spacing.

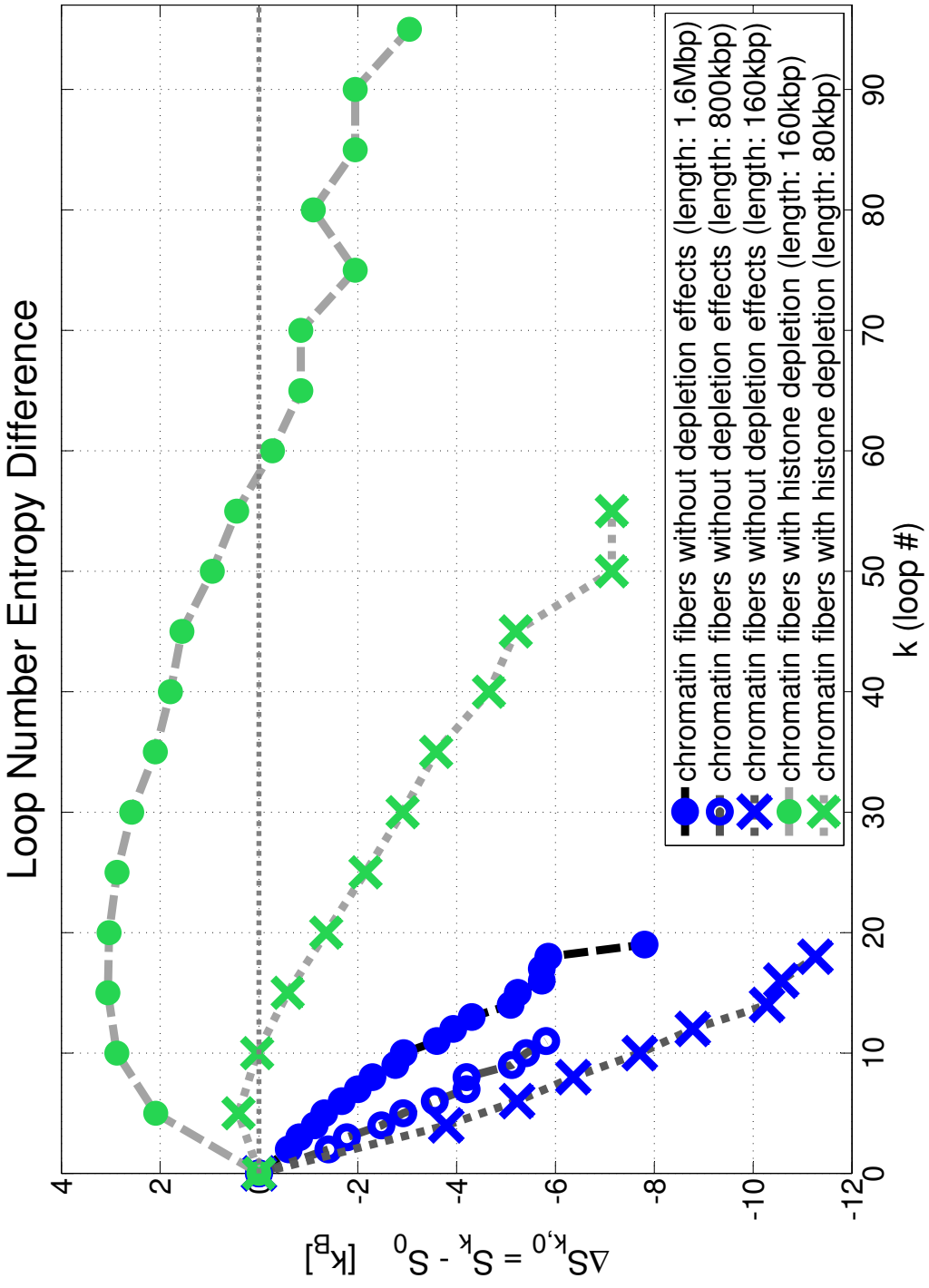


Figure 199: This figure shows the loop number entropy $\Delta S_{k,0}$ difference between the chromatin fibers without loops (S_0) an the fibers with k loops for five different fiber types.

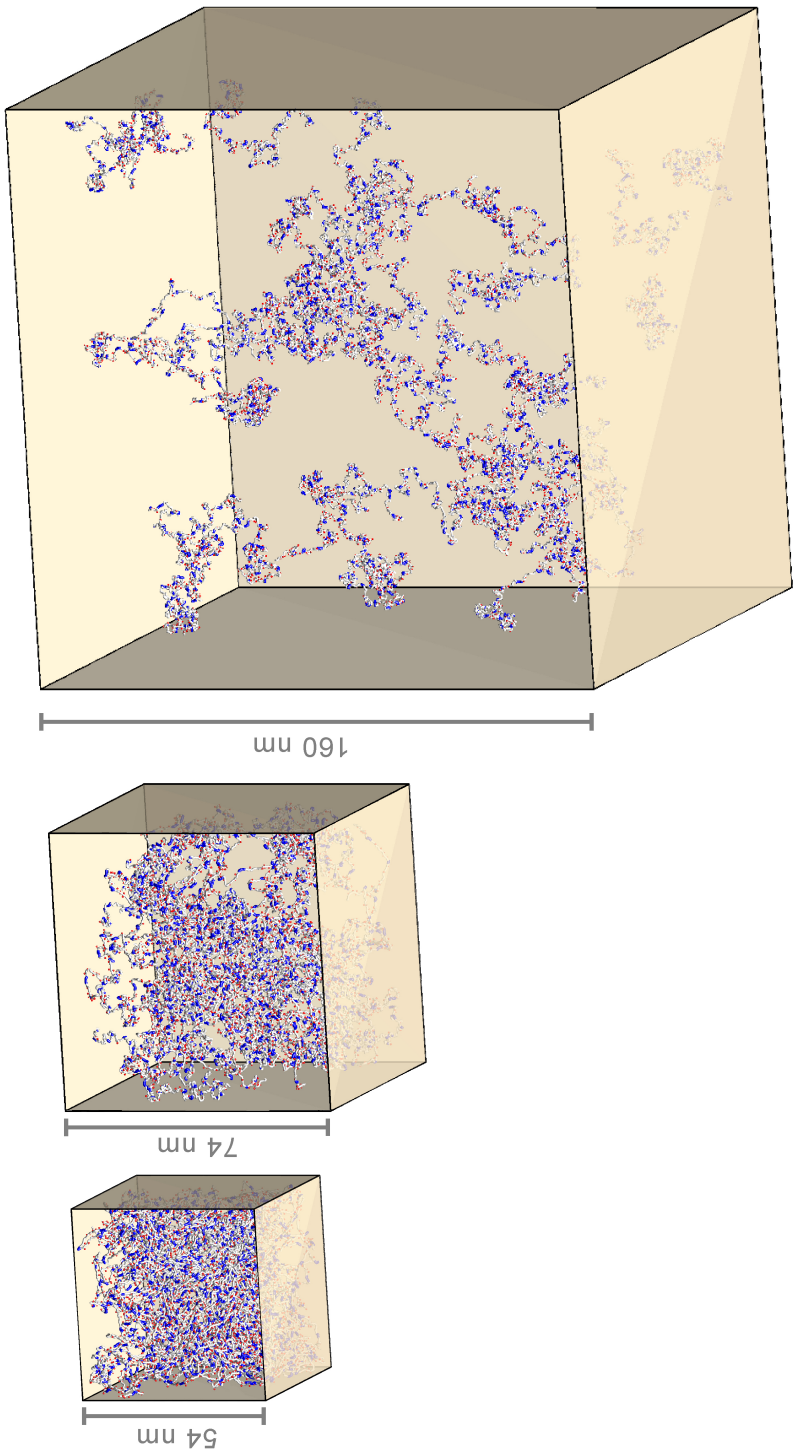


Figure 200: Three example configurations of a system with 40 peptides with different box sizes and thus different mass densities (1 mg/ml, 10 mg/ml and 25 mg/ml). The clustering of the peptide chains increases (the fraction of clustered peptides are 10%, 37.5% and 72.5% with decreasing density (resp. decreasing box volume).

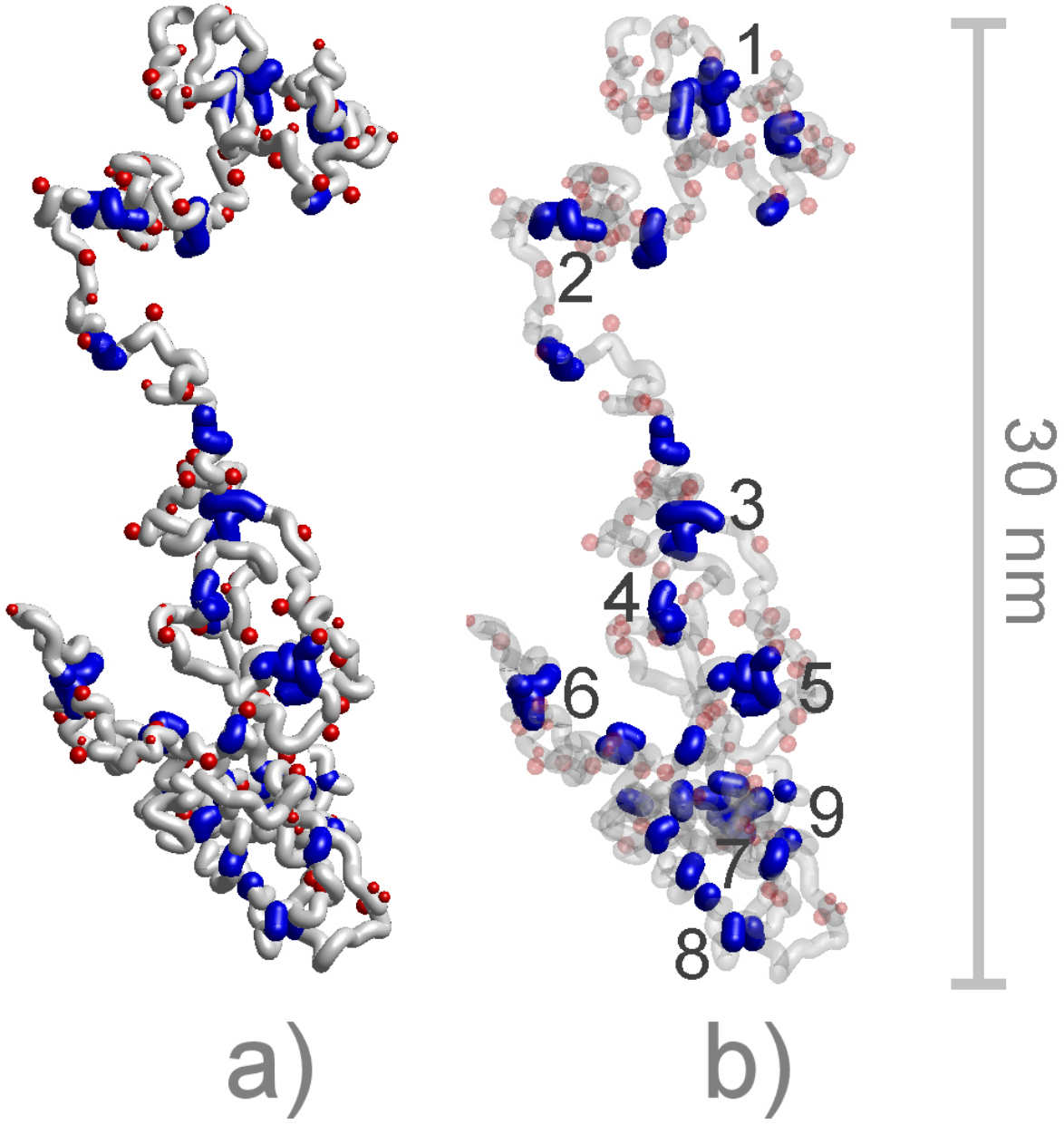


Figure 201: a) One conformation (out of the whole ensemble) of a single nucleoporin strand. The red spheres represent side chains of large and medium sized amino acids (starting with Glu / E). The FG-rich hydrophobic parts of the nucleoporin are marked blue. They have been enlarged by 30 percent for this visualization. b) Same as a) but transparent backbone and amino acid side chains to give a better view at the hydrophobic parts. One can see small hydrophobic clusters (numbered 1-9). Since there is no other peptide chain they are intra-bonds (cf. Fig. 141), i.e. bonds that the peptide coalesces with itself.

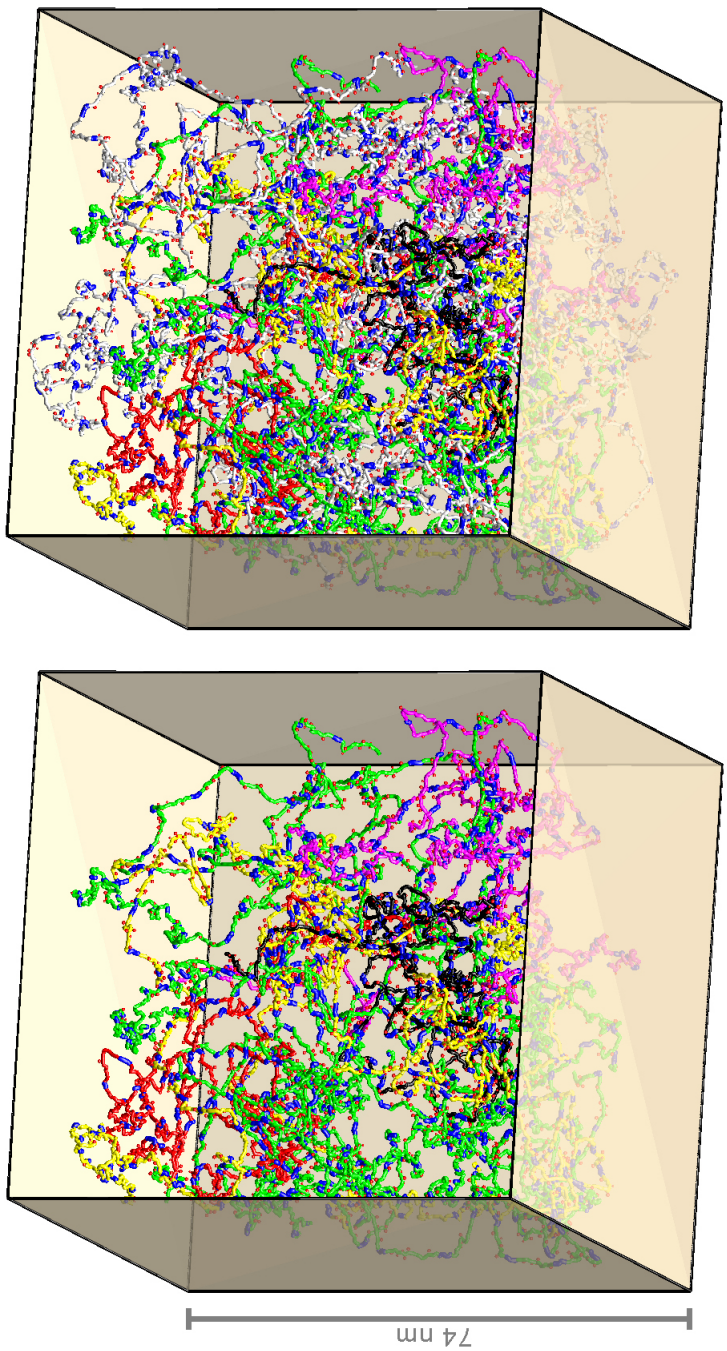


Figure 202: Example configuration of 40 peptides (N-terminal Nsp1p fsFG-domains cf. App. F) with colored clusters at a mass density of 10 mg/ml. Altogether, there are five clusters (colored green (12 peptides), yellow (5 peptides), red (2 peptides), black (2 peptides) and pink (3 peptides)) in this configuration. White peptides are single, unbound nucleoporins that have no hydrophobic bond with another peptide and thus do not contribute to any cluster. The box on the right shows the same configuration as on the left with the difference that the single unbound peptide chains are not shown. 60% of all nucleoporins in this configuration have hydrophobic bonds, i.e. they contribute to some of the five clusters. The other 40% are unbound.

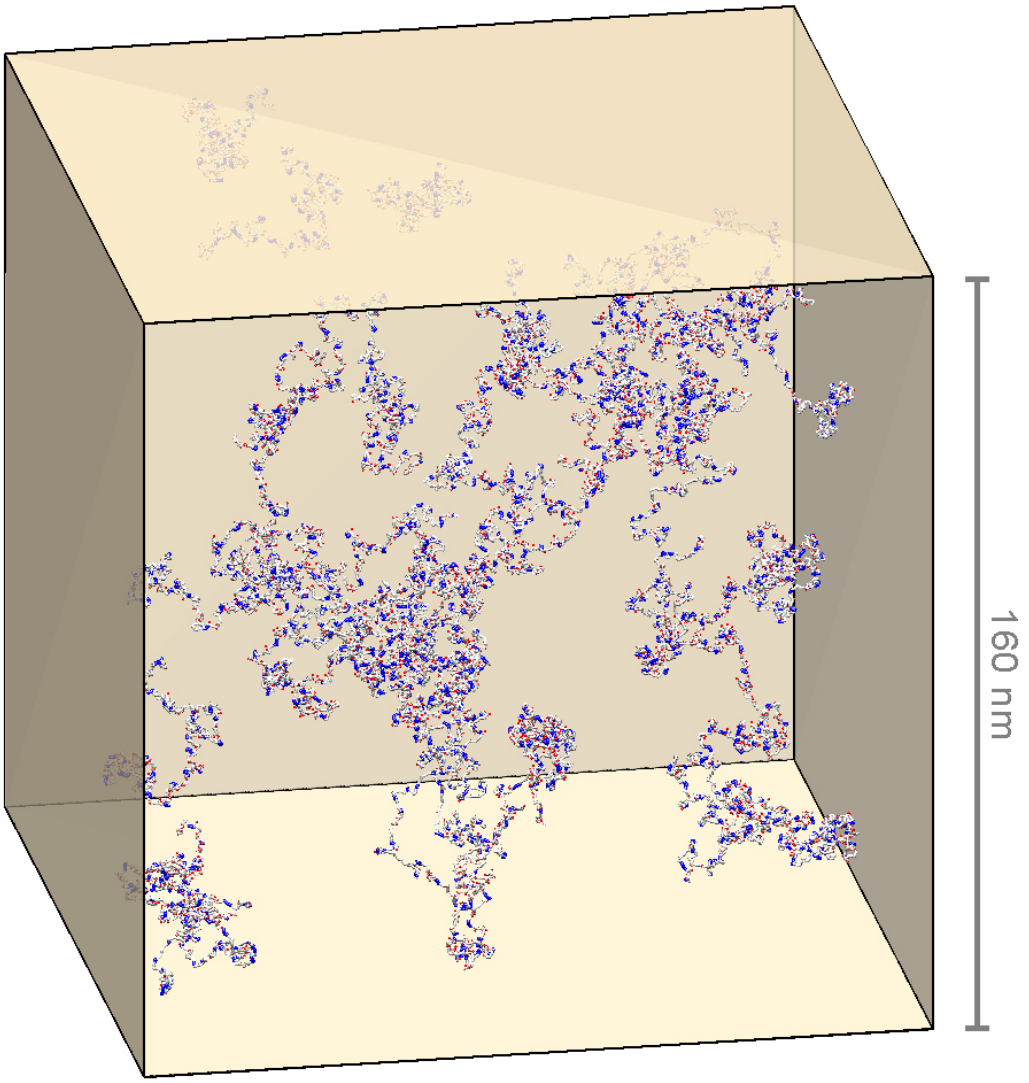


Figure 203: An example peptide system of $N = 40$ Nsp1p strands with density $c =$ 1mg/ml.

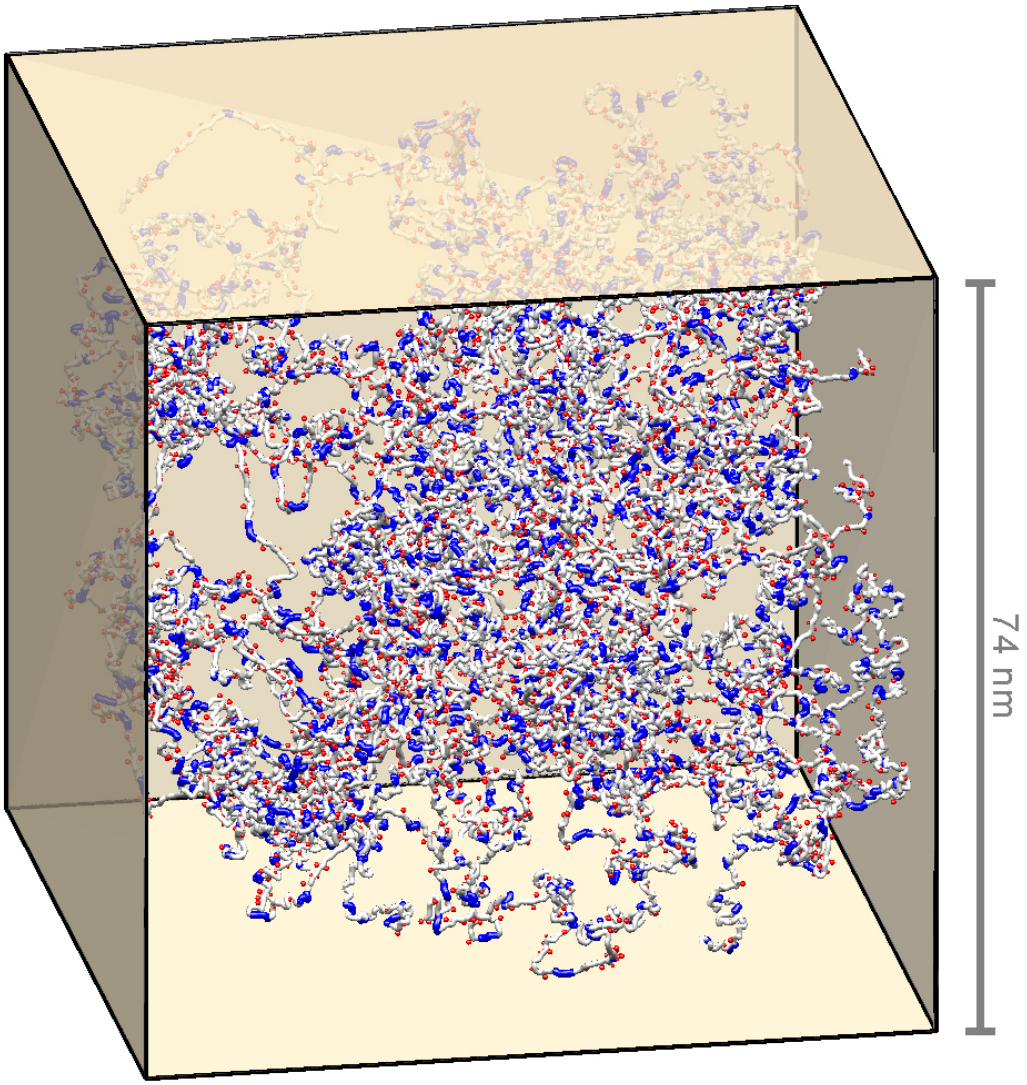


Figure 204: An example peptide system of $N = 40$ Nsp1p strands with density $c =$ 10mg/ml.

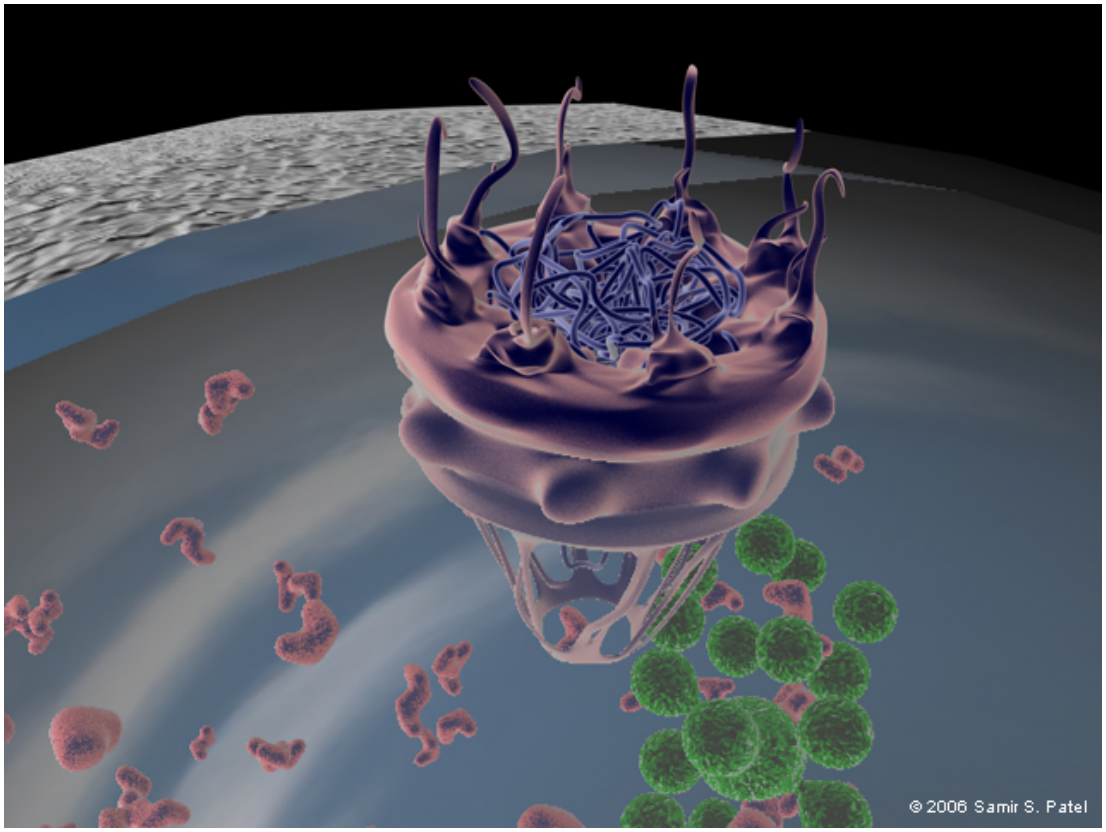


Figure 205: An illustration of a nuclear pore sitting with the nuclear envelope. Adapted from [157].

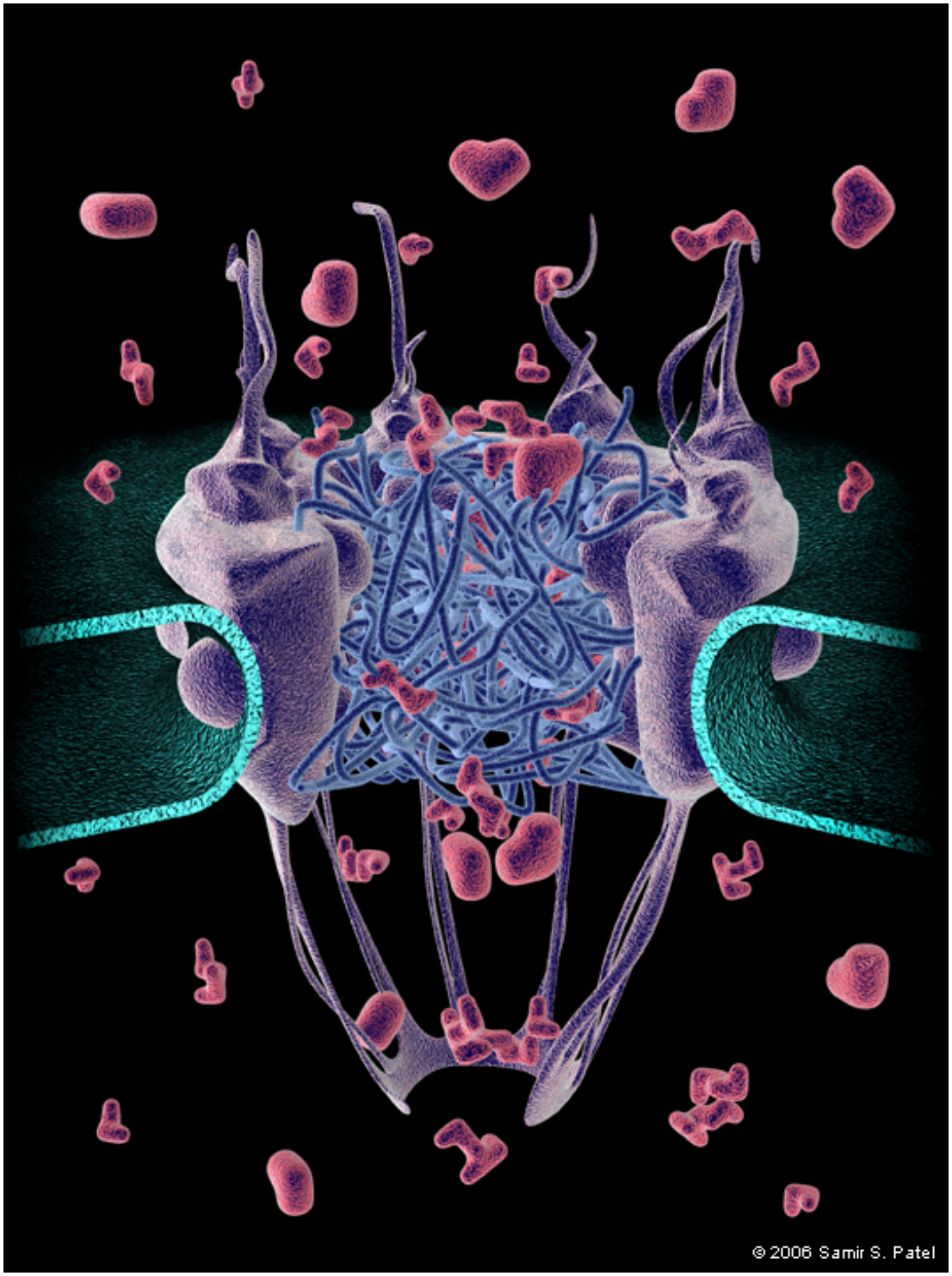


Figure 206: This figure shows an illustration of transport through a nuclear pore complex. Adapted from [157].

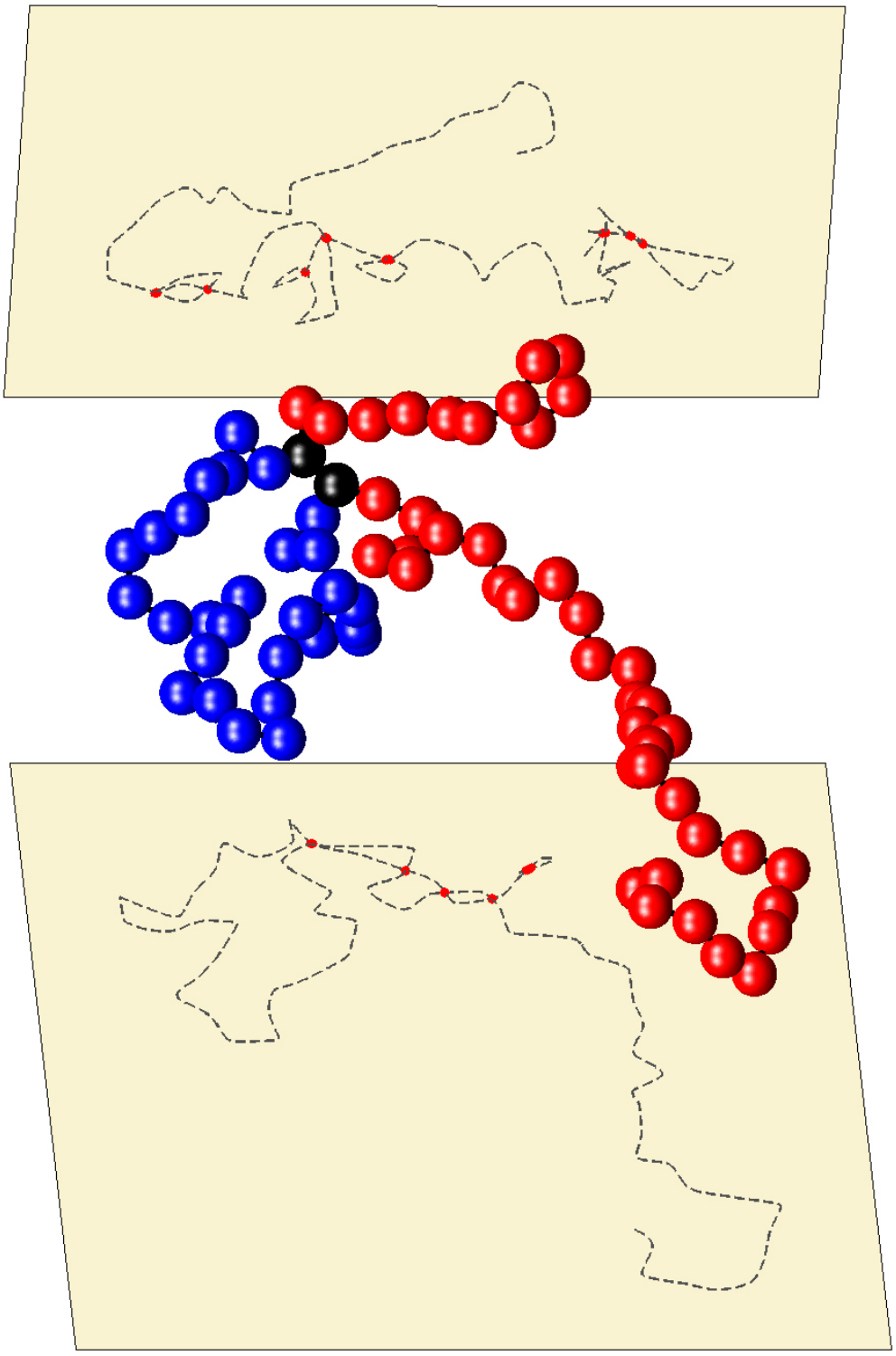


Figure 207: Illustration of an off-lattice self-avoiding random walk with spherical excluded volume. Two projections (onto the xz- and the yz-plane are shown in the background. Furthermore, a loop within the chain conformation is marked blue. The black spheres represent the start and end point of the loop. The whole chain has a length of $N = 75$ and the loop length is 28.

B Supplemental Tables

Different Values of α_i^n
These tables show different values of α_i^n, i.e. the peak locations of the excluded volume borderline (cf. Sec. 5), each with an accuracy of 10^{-4} degrees. Adapted from [61].

n=1										
i	3	4	5	6	7	8	9	10	11	12
α_i^1 [°]	60	90	108	120	128.5714	135	140	144	147.2727	150
i	13	14	15	16	17	18	19	20	21	22
α_i^1 [°]	152.3077	154.2857	156	157.5	158.8235	160	161.0526	162	162.8571	163.6364

Table 10: Different values of α_i^n that define the peak locations of the excluded volume borderline in the chromatin phase diagram.

n=2								
i	5	7	9	11	13	15	17	19
α_i^2 [°]	36	77.1429	100	114.5455	124.6157	132	137.6471	142.1053
i	21	23	25	27	29	31	33	35
α_i^2 [°]	145.7143	148.6957	151.2	153.3333	155.1724	156.7742	158.1818	159.4286

Table 11: Different values of α_i^n that define the peak locations of the excluded volume borderline in the chromatin phase diagram.

n=3								
i	7	8	10	11	13	14	16	17
α_i^3 [°]	25.7143	45	72	81.8182	96.9231	102.8571	112.5	116.4706
i	19	20	22	23	25	26	28	29
α_i^3 [°]	123.1579	126	130.9091	133.0435	136.8	138.4615	141.4286	142.7586

Table 12: Different values of α_i^n that define the peak locations of the excluded volume borderline in the chromatin phase diagram.

n=4								
i	9	11	13	15	17	19	21	23
α_i^4 [°]	20	49.0909	69.2308	84	95.2941	104.2105	111.4286	117.3913
i	25	27	29	31	33	35	37	39
α_i^4 [°]	122.4	126.6667	130.3448	133.5484	136.5484	138.8571	141.0811	143.0769

Table 13: Different values of α_i^n that define the peak locations of the excluded volume borderline in the chromatin phase diagram.

n=5								
i	11	12	13	14	16	17	18	19
α_i^5 [°]	16.3636	30	41.5385	51.4286	67.5	74.1176	80	85.2632
i	21	22	23	24	26	27	28	29
α_i^5 [°]	94.2857	98.1818	101.7391	105	110.7692	113.3333	115.7143	117.9310

Table 14: Different values of α_i^n that define the peak locations of the excluded volume borderline in the chromatin phase diagram.

	stiff fibers	flexible fibers			stiff fibers with depletion effects		
Δ	r_Δ	$\langle r_\Delta \rangle$	σ_Δ	mf_Δ	$\langle r_\Delta \rangle$	σ_Δ	mf_Δ
1.00	51.00	51.00	0.63	50.56	63.61	30.72	136.53
2.00	34.89	37.55	7.16	31.32	65.42	55.66	31.21
3.00	52.05	51.33	4.00	52.37	85.53	62.28	55.42
4.00	65.42	68.14	7.17	70.02	105.65	66.40	59.08
5.00	62.88	64.27	8.27	66.29	114.37	75.52	7.58
6.00	89.32	90.04	6.08	89.94	135.13	74.94	84.41
7.00	86.12	90.17	11.13	82.77	145.28	80.88	92.27
8.00	107.86	106.81	7.94	58.82	159.95	82.53	3.34
9.00	114.92	118.87	10.88	112.72	172.76	85.04	8.36
10.00	125.45	125.06	11.55	128.19	183.47	88.61	122.00

Table 15: This table shows the mean value ($\langle r_\Delta \rangle$), the standard deviation (σ_Δ) and the most frequent value (mf_Δ) of the first ten r_Δ-distributions. In the first case a stiff reference fiber was generated. The following three columns show the values for flexible fibers and the last three columns show the results for flexible fibers with depletion effects. Here, the nucleosome skip rate was again fixed to 8% and the linker histone skip rate was fixed to 6%. The table shows that some of the r_Δ-distribution are very asymmetric since the mean value differs greatly from the most frequent value (e.g. the distribution of r_2). All spatial distances are given in bp and the genomic distances are given in NRL again.

Name	Sequence (5'-3')	Primer position	
		Start	End
GDFOR001	CTGTCCAACTAATACGACTCACTATAGCCTTTATGGGCAATTCCCATAAAGGGAA	60845555	60845580
GDREV002	TTCTTCAAGTTTAAGTTTCCTGTTTCCCTTTAGTGAGGGTTAATAGTCGGACTC	60849095	60849118
GDFOR003	CTGTCCAACTAATACGACTCACTATAGCCTCACTTTACCCTATGGACTTGCCCTGAA	60851214	60851241
GDREV004	TTCCATAAGACATCATGAATTGTTTTTCTTCCCTTTAGTGAGGGTTAATAGTCGGACTC	60851766	60851794
GDFOR005	CTGTCCAACTAATACGACTCACTATAGCCTATGGTCACTATAACCTCAGCTATTTAGAA	60855909	60855938
GDREV006	TTCAGTATGATGTTGGCTGTGGACTTGTCATAATCCCTTTAGTGAGGGTTAATAGTCGGACTC	60858502	60858534
GDFOR007	CTGTCCAACTAATACGACTCACTATAGCCTCAACGTAGTTTGAAGGTTGAGTGAGAA	60866944	60866971
GDREV008	TTCAAATATTAGCTTTTTTAGCAGGCTTCCCTTTAGTGAGGGTTAATAGTCGGACTC	60873850	60873876
GDFOR009	CTGTCCAACTAATACGACTCACTATAGCCTCTGGGTCTGGCTTCATTCTCTGATGGAGAA	60876630	60876660
GDREV010	TTCTATTGTCTCTGCATTGATTGGAAAGTTTCCCTTTAGTGAGGGTTAATAGTCGGACTC	60877663	60877692
GDFOR011	CTGTCCAACTAATACGACTCACTATAGCCTTACCTAGCAATCTGACAAATAGTAGGAA	60880001	60880029
GDREV012	TTCTATTCACTTTATTAAGGTTGTGGTCCCTTTAGTGAGGGTTAATAGTCGGACTC	60880385	60880410
GDFOR013	CTGTCCAACTAATACGACTCACTATAGCCAAGCATACAGCATTACAAGTAAAATGGAA	60882444	60882472
GDREV014	TTCTATAAAGTTAATTAATCACTATTTCCCTTTAGTGAGGGTTAATAGTCGGACTC	60891170	60891195
GDFOR015	CTGTCCAACTAATACGACTCACTATAGCCGTTACATAACAACTAGATGTAATGATGGAA	60894533	60894562
GDREV016	TTCTAGTTCAGGAGAATAAAAGTTATATCCCTTTAGTGAGGGTTAATAGTCGGACTC	60894683	60894709
GDFOR017	CTGTCCAACTAATACGACTCACTATAGCCGCCCCAAAAGCAGTGGAAGGGGAA	60905463	60905486
GDREV018	TTCTACCGTGGGCATAGAGGAAGGGCCTCCCTTTAGTGAGGGTTAATAGTCGGACTC	60908394	60908420
GDFOR019	CTGTCCAACTAATACGACTCACTATAGCCTCAAAGAGCTGAAACAATTGTCTTCCAGAA	60908658	60908687
GDREV020	TTCCACTTATTTATTATAATTATTTATCCCTTTAGTGAGGGTTAATAGTCGGACTC	60908781	60908806

Table 16: Summary of 5C primer sequences used to generate 5C libraries. "Primer position" indicates the position of the genomic primer homology region (genome assembly hg17). Adapted from [72].

K562 cells I	GDFOR002	GDFOR004	GDFOR006	GDFOR008	GDFOR010
FOR001	0.13	0.16	0.07	0.1	0.2
FOR003	0.56	1.19	0.17	0.14	0.3
FOR005	0.37	1.03	0.31	0.16	0.25
FOR007	0.1	0.23	0.15	0.29	0.25
FOR009	0.16	0.15	0.1	0.6	1.58
FOR011	0.08	0.17	0.08	0.41	1.9
FOR013	0.16	0.11	0.08	0.21	0.67
FOR015	0.1	0.17	0.12	0.25	0.34
FOR017	0.11	0.09	0.08	0.12	0.21
FOR019	0.08	0.11	0.05	0.15	0.14

Table 17: Interaction data of K562 cells (β-globin locus on, part I). Adapted from [72].

K562 cells II	GDFOR012	GDFOR014	GDFOR016	GDFOR018	GDFOR020
FOR001	0.11	0.1	0.09	0.16	0.1
FOR003	0.13	0.08	0.13	0.72	0.07
FOR005	0.16	0.09	0.06	0.08	0.09
FOR007	0.08	0.11	0.05	0.09	0.11
FOR009	0.48	0.13	0.09	0.11	0.11
FOR011	1.03	0.28	0.09	0.11	0.07
FOR013	1.21	0.88	0.32	0.18	0.14
FOR015	0.2	0.92	0.66	0.27	0.11
FOR017	0.14	0.18	0.32	0.39	0.18
FOR019	0.08	0.08	0.12	0.55	0.26

Table 18: Interaction data of K562 cells (β-globin locus on, part II). Adapted from [72].

GM06990 cells I	GDFOR012	GDFOR014	GDFOR016	GDFOR018	GDFOR020
FOR001	0.12	0.12	0.06	0.07	0.12
FOR003	0.15	0.39	0.17	0.06	0.11
FOR005	0.11	0.4	0.21	0.08	0.08
FOR007	0.08	0.16	0.17	0.35	0.21
FOR009	0.1	0.1	0.07	0.41	0.82
FOR011	0.04	0.11	0.05	0.14	1.02
FOR013	0.13	0.08	0.06	0.09	0.2
FOR015	0.07	0.09	0.07	0.11	0.21
FOR017	0.05	0.05	0.05	0.04	0.07
FOR019	0.05	0.06	0.03	0.05	0.06

Table 19: Interaction data of GM06990 cells (β-globin locus off, part I). Adapted from [72].

GM06990 cells II	GDFOR002	GDFOR004	GDFOR006	GDFOR008	GDFOR010
FOR001	0.1	0.06	0.09	0.16	0.07
FOR003	0.06	0.03	0.06	0.27	0.04
FOR005	0.06	0.05	0.03	0.04	0.06
FOR007	0.04	0.08	0.05	0.06	0.08
FOR009	0.18	0.07	0.04	0.05	0.07
FOR011	0.53	0.16	0.07	0.06	0.04
FOR013	0.51	0.65	0.24	0.08	0.09
FOR015	0.15	0.87	0.25	0.19	0.07
FOR017	0.12	0.12	0.32	0.29	0.11
FOR019	0.03	0.05	0.09	0.23	0.1

Table 20: Interaction data of GM06990 cells (β-globin locus off, part II). Adapted from [72].

C SPDM Setup

This section depicts the experiments that were done within the framework of the collaboration with Prof. Cremer (cf. Sec. 3). The main parts of the following description of the experimental setup for the localization microscopy were provided by Rainer Kaufmann (Kirchhoff-Institute of Physics). This applies also to the figures in this section (i.e. App. C).

Experimental Setup

The measurements were done on a SPDM setup [113; 131] for 2D localization of single fluorescent molecules. This method is based on a light induced reversibly bleached state of conventional fluorophores [104; 158; 193]. By starting illumination of the sample with an excitation intensity in the 10 kW/cm^2 to 1 MW/cm^2 range, some molecules are bleached irreversibly ($M_{fl} \longrightarrow M_{ibl}$), another amount is transferred into a reversibly bleached state ($M_{fl} \longrightarrow M_{rbl}$). The statistical recovery of fluorophores from this state ($M_{fl} \longleftarrow M_{rbl}$) can be used for optical isolation of the single fluorescent molecules. This allows the SPDM setup single molecule detection and localization of conventional fluorophores specified as $SPDM_{PHYMOD}$.

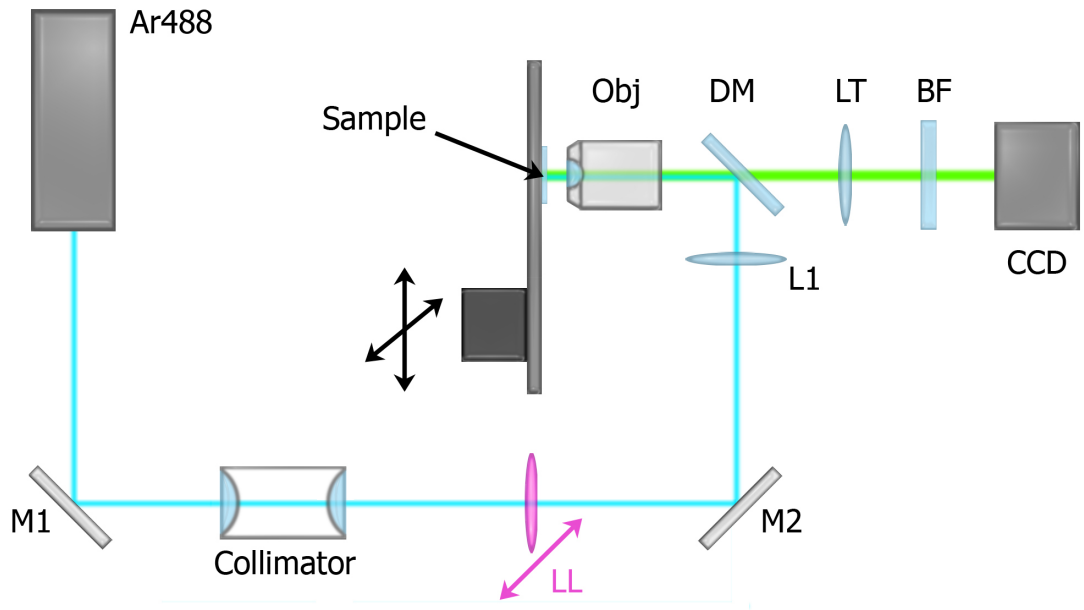


Figure 208: Schematic overview of the microscope setup used for the SPDM measurements

Microscope Setup

For the experiments only one laser source was needed at an excitation wavelength of $\lambda = 488$ nm [Ar488] (Lexel 95-4, Lexel Laser, USA). After deflection at the mirror

[M1] the laser beam is expanded by a factor of 10 using a collimator builded up of two achromates with focal length of 10 mm and 100 mm (Linos Photonics, Göttingen, Germany). Then a lens [LL] can be mounted in the optical pathway, leading to a more focused spot in the object region so one can achieve the high laser intensity, which is necessary for the localization mode. Via a mirror [M2], a dichroic mirror [DM] (AHF Analysetechnik AG, Tübingen, Germany) and a lens [L1] the beam is focused into the back focal plane of an oil immersion objective lens (x100, NA = 1.4, Leica, Bensheim, Germany).

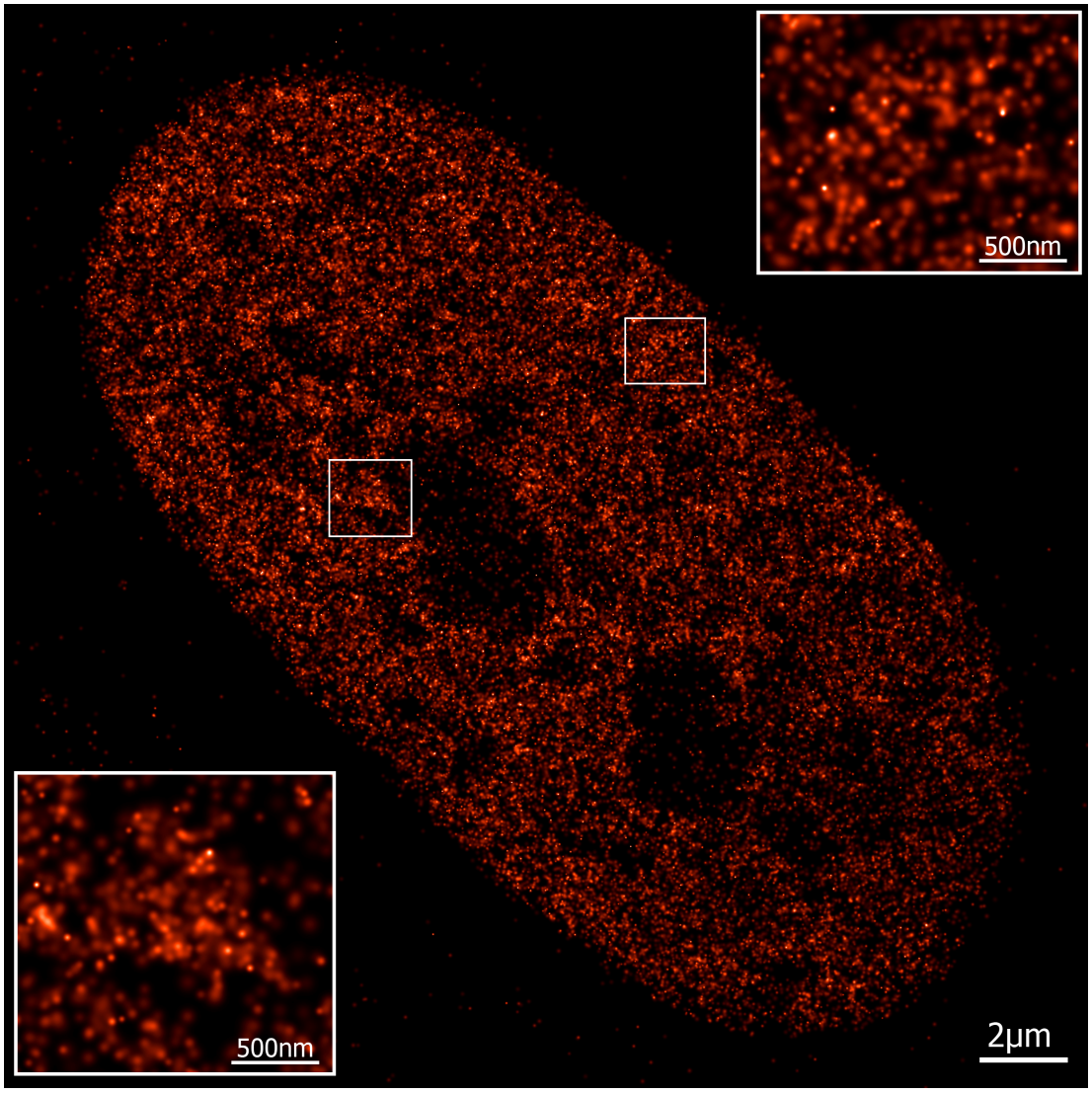


Figure 209: Localization image of histones H2B in the nucleus of a human fibroblast cell labeled with EmGFP. The localized fluorophores are blurred with a Gaussian corresponding to their individual localization accuracy. The inserts show the marked areas four times magnified.

The detection pathway is a conventional epi-fluorescence setup. Emitted fluorescence light passes the dichroic mirror [DM] and is focused by a tube lens [TL] (x1.0, Leica, Bensheim, Germany) onto the CCD chip of a highly sensitive 12 bit black and white camera [CCD] (SensiCam QE, PCO Imaging, Kehlheim, Germany). A blocking filter [BF], which is mounted in front of the CCD chip, reduces the background signal of the excitation light.

Data Acquisition and Evaluation
For data acquisition a time stack consisting of 2000 frames was recorded with a frame rate of about 18 fps. The evaluation was done by algorithms implemented in MATLAB (7.4.0, The MathWorks, Natick, USA).

In a first step, the number of photons for each pixel of the CCD chip is estimated. The count number is translated into the number of incident photons by the multiplication of a factor dependent on the conversion rate of the CCD sensor:

$$N_{Photons}(x, y, t) = c_{conv} \cdot N_{Counts}(x, y, t),$$

where $N_{Photons}(x, y, t)$ is the number of incident photons as a function of the pixel position (x, y) and the time t denoting the individual frames of the image stack obtained by the CCD camera at times $t = t_1, t_2, ..., t_k, ...t_n$. c_{conv} is the conversion factor between the number of counts and the number of photoelectrons. $N_{Counts}(x, y, t)$ gives the number of counts (raw data) as a function of pixel position (x, y) and the time t. For example, using $c_{conv} = 2/count$ (low light mode) a photon number of $N_{Photons}(x, y, t) = 2N_{Counts}(x, y, t)$ is estimated.

In a second step, for signals with a low signal-to-noise ratio (high background and photo bleaching effects active during several succeeding frames) an additional computing step is required to segment signals originating from single molecules only. A differential photon stack $D_{Photons}(x, y, t')$ between the succeeding $(t = t_{k+1})$ and the preceding frame (t_k) is calculated:

$$D_{Photons}(x, y, t') = N_{Photons}(x, y, t_{k+1}) - N_{Photons}(x, y, t_k).$$

The error σ in the photon number produced by the Poisson statistics of the incident photons and the noise σ_{CCD} of the detection at the CCD chip (approx. 4 counts per pixel) was estimated by the Gaussian law of error propagation:

$$\sigma[D_{Photons}(x, y, t')] = [N_{Photons}(x, y, t_{k+1})^2 + N_{Photons}(x, y, t_k)^2 + 2\sigma_{CCD}^2]^{1/2}.$$

In the last step, the data stack $N_{Photons}(x, y, t)$ (in case of low background) or $D_{Photons}(x, y, t')$ (in case of high background) is used for a lateral high precision localization of single molecules. To reduce computing efforts, Regions of Interest (ROIs) of typically 8x8

pixels were used. A model function is fitted to the acquired signals in the object plane using the Levenberg-Marquardt algorithm with a Gaussian distribution:

$$f(x,y) = Ae^{-\frac{(x-x_0)^2+(y-y_0)^2}{2\sigma^2}} + B_0 + B_1(x_0 - x) + B_2(y_0 - y).$$

x_0 and y_0 are start parameters for the position corresponding to the determined center of the segmented signals. A is the amplitude and σ the standard deviation of the Gaussian distribution. B_0, B_1 and B_2 represent a linear background.

In this manner the lateral positions of the single molecules can be determined with a localization accuracy (σ) in the 20 nm regime [131].

Fig. 3 shows a localization image for histones H2B labeled with EmGFP in human fibroblast cells. The localized fluorophores give the positions of the single histone proteins with their individual localization accuracy displayed by a corresponding Gaussian blur. In Fig. 3 the histogram for the localization accuracy of the 43,821 fluorophores in Fig. 2 is shown. The wide-field image in Fig. 4 was recorded with a conventional epi-fluorescence microscope setup.

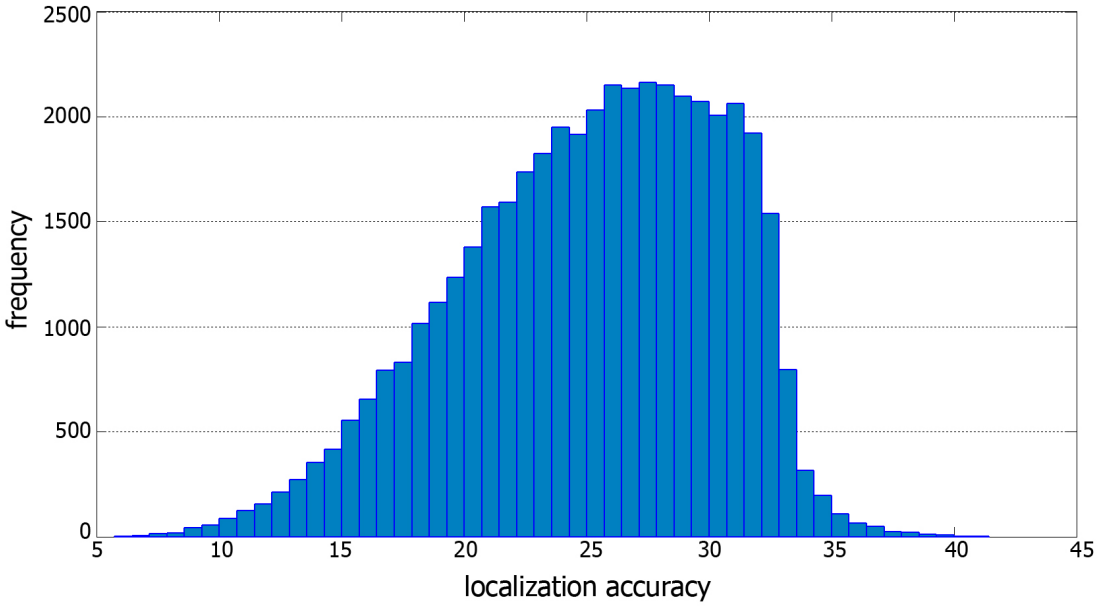


Figure 210: Histogram of the localization accuracy. The pair-wise mean values of x and y localization accuracy are shown for the EmGFP-labeled histones in Fig. 2. The overall mean value of the localization accuracy for the 43,821 localized fluorophores was 25.2 nm with a standard deviation of 5.3 nm

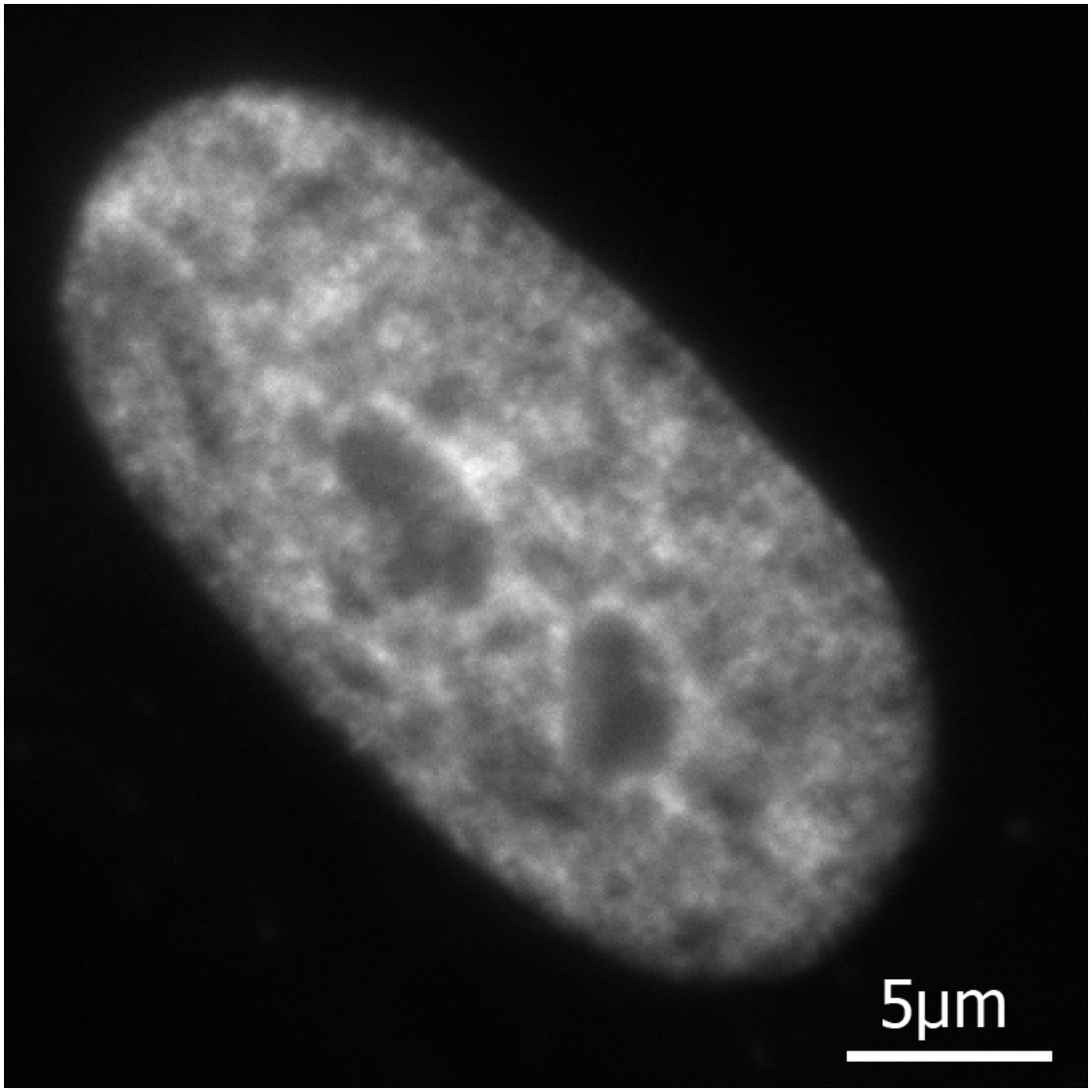


Figure 211: A conventional wide-field image. The image was acquired with a conventional epi-fluorescence microscope setup. The same nucleus as in the localization image in Fig. 2 is shown.

Specimen Preparation

a) Fibroblast Cells
The VH7 diploid human fibroblast cells (kindly provided by Prof. Dr. Beauchamp from DKFZ, Heidelberg), used for the measurements, were cultivated in Dulbecco's Modified Eagle medium supplemented with 10% FCS, 1% L-glutamine, 1% penicillin/streptomycin in a standard CO_2-incubator. After seeding the cells onto cover slips, they were allowed to attach and grow over night. Using Organelle Lights (Invitrogen, Carlsbad, USA)

according to the manufacturer's protocol, EmGFP conjugated histone proteins H2B were expressed in the fibroblast nuclei. The cells were fixed with 4% formaldehyde in PBS and embedded with ProLong Gold antifade reagent (Invitrogen) 20 hours after the transfection.

b) HeLa Cells
HeLa cells were cultivated in RPMI medium supplemented with 10% FCS, 1% L-glutamine, 1% penicillin/Streptomycin in a standard CO_2-incubator.
HeLa cells, whose H2B histones had been stably transfected with GFP, were seeded onto coverslips, and were allowed to attach and grow over night. Afterwards the cells were fixed using 4% formaldehyde in PBS and embedded with ProLong Gold antifade reagent (Invitrogen).
Those HeLa-cells not being stably transfected were first seeded onto coverslips and allowed to attach and grow over night. After using Organelle Lights (Invitrogen, Carlsbad, USA) according to the manufactures protocol, EmGFP conjugated histone proteins H2B were expressed in HeLa-cell-nuclei. The cells were then fixed with 4% formaldehyde in PBS and embedded with ProLong Gold antifade reagent (Invitrogen) 20 hours after transfection.

D The Pivot Algorithm

The pivot algorithm is a dynamic Monte Carlo algorithm which generates SAWs with one endpoint fixed at the origin and the other endpoint free. The state space is L_N, and the invariant probability measure is the standard equal-weight SAW distribution ($\pi_\omega = \frac{1}{c_N}$).

The elementary move of the pivot algorithm is as follows:

At first *choose at random* a pivot point k along the walk ($0 < k \leq N-1$). Then *choose at random* an element of the symmetry group of the lattice (rotation or reflection or a combination thereof). Now apply the symmetry-group element to $X_{k+1}, ..., X_N$ using X_k as a temporary origin (pivot). If the resulting walk is self-avoiding, it will be accepted, otherwise it will be rejected.

Remarks

1. To keep the algorithm ergodic it is necessary that the probability to choose the pivot element is strictly positive for all points of the chain. (In most cases a uniform distribution is used but in some situations other choices may be advantageous).

2. Let G be the group of orthogonal transformations (about the origin) that leave the lattice invariant. Then the symmetry operation $g \, \epsilon \, G$ can be chosen according to any probability distribution $(p_g)_{g \epsilon G}$ that satisfies $p_g = p_{g^{-1}}$ for all g (and has enough nonzero entries to ensure ergocity). The condition $p_g = p_{g^{-1}}$ is both necessary *and* sufficient to ensure detailed balance with respect to the equal-weight distribution π.

3. The computational work can be reduced (roughly by a factor 2) by applying the symmetry operation always to the shorter of the two segments of the walk, whichever it may be.

Acceptance Fraction

The acceptance fraction f of the pivot algorithm is about $f \approx N^p$ (for $d = 2$, p is roughly 0.19 and for $d = 3$, p is about 0.11, p depends on the specific symmetry operation). This seems to slow down the algorithm a lot but remember that the pivot moves are very radical and therefore the chain confirmation will be a completely different after a few accepted pivot moves (in fact this is what makes the pivot algorithm very efficient). The autocorrelation time τ is of order $\frac{N}{f}$.

History

The pivot algorithm has a curious history: It was invented by Lal in 1969, and then completely forgotten by almost everyone. In 1985 it was reinvented by MacDonald et al. and reinvented a short time later by Madras. This third reinvention led to a comprehensive analytical and numerical study by Madras and Sokal [139] which showed that the algorithm is by far the most efficient algorithm yet invented for simulating SAWs. After that applications took off.

Faster Implementation of the Pivot Algorithm

A faster implementation of the pivot algorithm for self-avoiding walks which takes advantage of the underlying grid symmetries was developed in 2004 [115].

Past implementations of the algorithm required a time $O(N)$ per accepted pivot where N is the number of steps in the walk. In [115] the algorithm is implemented so that the time required per accepted pivot is $O(N^q)$ with $q < 1$. q is less than 0.57 in two dimensions, and less than 0.85 in three dimensions. Corrections to the $O(N^q)$ make an accurate estimate of q impossible. They also imply that the asymptotic behavior of $O(N^q)$ cannot be seen for walk lengths which can be simulated.

In simulations the effective q is around 0.7 in two dimensions and 0.9 in three dimensions. Comparisons with simulations that use the standard implementation of the pivot algorithm using a hash table indicate that our implementation is faster by as much as a factor of 80 in two dimensions and as much as a factor of 7 in three dimensions. The fast pivot method does not require the use of a hash table.

E Metropolis Monte-Carlo Simulations

Detailed information about Metropolis Monte Carlo simulations can be found in [24]. Nevertheless, a short overview is given below.

All possible system configurations ω lie in the phase space Ω. The occurrence probability of a certain state ω is denoted by $p(\omega)$.
Generally, one is interested in mean values of configuration properties X (e.g. the radius of gyration R_G) averaged over the whole phase space Ω:

$$\langle X \rangle = \int_{\Omega} p(\omega) X(\omega) d\omega.$$

With the Boltzmann distribution this leads to

$$\langle X \rangle = \frac{1}{Z} \int_{\Omega} \exp\left(\frac{-H(\omega)}{kT}\right) X(\omega) d\omega,$$

where Z is the corresponding partition function:

$$Z = \int_{\Omega} \exp\left(\frac{-H(\omega)}{kT}\right) d\omega.$$

The sample size obtained by computer simulations is always finite, i.e. only a small part Ω' of Ω can be sampled. The Metropolis algorithm provides a method to find a finite set of states $\omega \in \Omega$ so that the average value of X over this limited set Ω' is a good approximation to the actual value of $\langle X \rangle_{\Omega}$. This method is called *importance sampling* in opposite to *simple sampling* where the subset of Ω is chosen by random.

In the case of simple sampling the average value of X is given by

$$\langle X \rangle = \frac{\sum_{\Omega'} p(\omega) X(\omega)}{\sum_{\Omega'} p(\omega)}.$$

If a certain state is generated during the simulations with a probability $w(\omega)$, then the average value of X will be [100; 198]

$$\langle X \rangle = \frac{\sum_{\Omega'} p(\omega) X(\omega)/w(\omega)}{\sum_{\Omega'} p(\omega)/w(\omega)}.$$

If $w(\omega) = p(\omega)$, a simple arithmetic mean will remains.

To do use importance sampling one starts with a certain configuration ω_0 and then generates a Markov chain of system states [100; 147]. That means a system configuration ω_i is generated only from the preceding configuration ω_{i-1}. The transition probability between these two states is denoted by $T_{\omega_i, \omega_{i-1}}$.

The Markovian master equation is

$$\frac{\partial w(\omega_i,t)}{\partial t} = \sum_{\omega_i} T_{\omega_i,\omega_{i-1}} w(\omega_{i-1},t) - T_{\omega_{i-1}\omega_i} w(\omega_i,t).$$

In the case of detailed balance [23], i.e.

$$T_{\omega_i,\omega_{i-1}} w(\omega_{i-1},t) = T_{\omega_{i-1}\omega_i} w(\omega_i,t)$$

one gets $\frac{\partial w(\omega_i,t)}{\partial t} = 0$ which means that $w = w_{\text{eq.}}$, i.e. the equilibrium probability which is given by the Boltzmann distribution.
Furthermore, this leads to $\lim_{i\to\infty} w(\omega_i) = w_{\text{eq.}}(\omega_i)$.

This gives

$$\frac{T_{\omega_i,\omega_{i-1}}}{T_{\omega_{i-1},\omega_i}} = \frac{w_{\text{eq}}(\omega_i,t)}{w_{\text{eq}}(\omega_{i-1},t)} = \exp\left(\frac{H(\omega_i) - H(\omega_{i-1})}{kT}\right).$$

The latter condition will easily be fulfilled, if one chooses the following transition probability:

$$T_{\omega_{i-1},\omega_i} = \begin{cases} \exp(\frac{-\Delta H}{kT}) & \Leftarrow \Delta H > 0 \\ 1 & \Leftarrow \Delta H \leq 0 \end{cases}$$

with $\Delta H = H(\omega_i) - H(\omega_{i-1})$. It is called the Metropolis-function. This is, however, not the only possible choice for a transition probability [24].

F Sequence of the N-terminal fsFG-Domain of Nsp1p

Amino acid sequence of the N-terminal fsFG-domain of the essential yeast nucleoporin Nsp1p. The hydrophobic clusters which have been considered are marked blue whereas the amino acids the side chains of which were allowed for are marked red.

MNFNTPQQ
NKTPFSFGTANNNSNTTNQNSST
GAGAFGT
GQSTFGFNNSAPNNTNNANSS
ITPAFGSNNT
GNTAFGNSNP
TSNVFGSNNS
TTNTFGSNSA
GTSLFGSSSAQQTKSNGTA
GGNTFGSSSLFNNSTNSNT
TKPAFGGLNFGGGNNTTPSSTGNANT
SNNLFGATANANK
PAFSFGATTNDDKKTEPDK
PAFSFNSSVGNKTDAQAPT
TGFSFGSQLGGNKTVNEAAK
PSLSFGSGSAGANPAGASQPEPTTNEPAK
PALSFGTATSDNKTTNTT
PSFSFGAKSDENKAGATSK
PAFSFGAKPEEKKDDNSSK
PAFSFGAKSNEDKQDGTAK
PAFSFGAKPAEKNNNETSK
PAFSFGAKSDEKKDGDASK
PAFSFGAKPDENKASATSK
PAFSFGAKPEEKKDDNSSK
PAFSFGAKSNEDKQDGTAK
PAFSFGAKPAEKNNNETSK
PAFSFGAKSDEKKDGDASK
PAFSFGAKSDEKKDSDSSK
PAFSFGTKSNEKKDSGSSK
PAFSFGAKPDEKKNDEVSK
PAFSFGAKANEKKESDESK
SAFSFGSKPTGKEEGDGAK
AAISFGAKPEEQKSSDTSK
PAFTFGAQKDNEKKTEESSTG

G Some Basic Properties of the Fourier Transform

The Fourier transform in its standard form (i.e. with the Fourier parameters $a = 0$ and $b = 1$) is

$$\mathcal{F}f(\xi) = \sqrt{\frac{|b|}{(2\pi)^{1-a}}} \int_{-\infty}^{\infty} f(x) e^{ib\xi x}\, dx \overset{\substack{a=0\\b=1}}{=} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{i\xi x} f(x)\, dx \quad .$$

The Fourier transform is linear and if $n > 0$ is an integer and $\lim_{|x|\to\infty} f^{(r)}(x) = 0$ for $r = 0, 1, ..., n-1$, then

$$\mathcal{F}\left[f^{(n)}(x)\right](\xi) = (-i\xi)^n \mathcal{F}f(\xi) \quad \text{(transform of a derivative)}.$$

The convolution $f \star g$ of two functions $f(x)$ and $g(x)$ is defined by the integral

$$f \star g(x) = \int_{-\infty}^{\infty} f(x-\xi) g(\xi)\, d\xi$$

and has the property $f \star g = g \star f$, and $f \star (g \star h) = (f \star g) \star h$. The Fourier transform of a convolution can be calculated by using

$$\mathcal{F}[f \star g](\xi) = \sqrt{2\pi} \cdot \mathcal{F}f(\xi) \cdot \mathcal{F}g(\xi) \quad \text{(convolution theorem)}.$$

Furthermore, for $a > 0$ the Fourier transform fulfils

$$\mathcal{F}e^{-a^2x^2} = (a\sqrt{2})^{-1} e^{-\frac{\xi^2}{4a^2}} \quad .$$

H Some Basic Properties of the Convolution

The convolution $f \star g$ of two functions $f(x)$ and $g(x)$ is defined by the integral

$$f \star g(x) = \int_{-\infty}^{\infty} f(x-\xi)g(\xi)\, d\xi$$

and has the properties

$$f \star g = g \star f$$

and

$$D(f \star g) = Df \star g = f \star DG$$

as well as

$$f \star (g \star h) = (f \star g) \star h.$$

The Fourier transform of a convolution can be calculated by using

$$\mathcal{F}[f \star g](\xi) = \sqrt{2\pi} \cdot \mathcal{F}f(\xi) \cdot \mathcal{F}g(\xi) \quad \text{(convolution theorem)}.$$

The convolution of two Gaussians

$$f(x) = \frac{1}{\sigma_1\sqrt{2\pi}}\exp\left(\frac{-(x-\mu_1)^2}{2\sigma_1^2}\right)$$
$$g(x) = \frac{1}{\sigma_2\sqrt{2\pi}}\exp\left(\frac{-(x-\mu_2)^2}{2\sigma_2^2}\right)$$

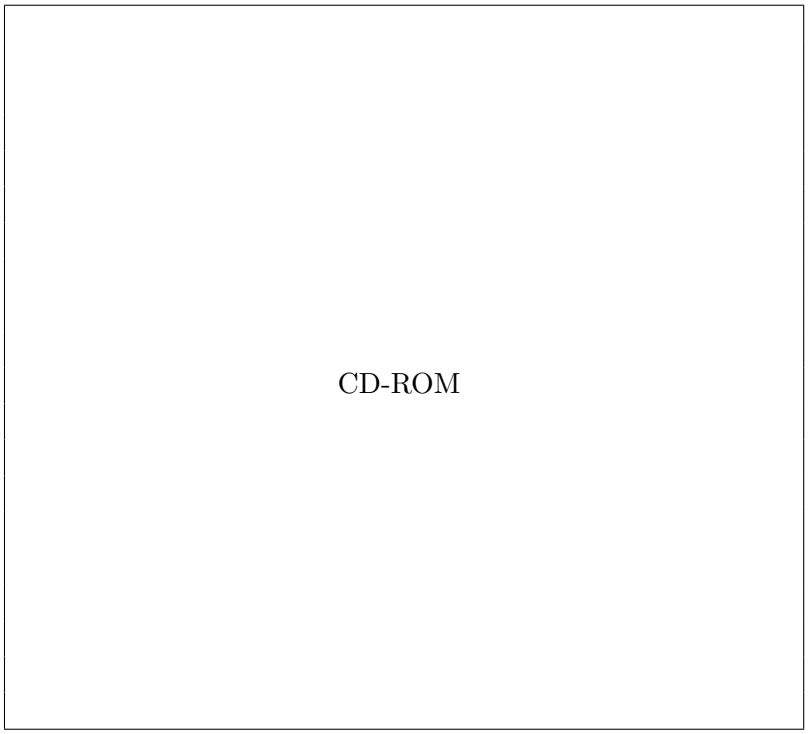
is simply another Gaussian:

$$(f \star g)(x) = \frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}}\exp\left[-\frac{(x-(\mu_1+\mu_2))^2}{2(\sigma_1^2+\sigma_2^2)}\right].$$

I Electronic Version of this Thesis

The CD-ROM below contains this thesis in pdf format and all figures of this work.

The electronic version has some advantages in comparison with the print version: In the pdf all references (Fig., Sec., Eq., headlines, citations and so on) are hyper-links to the associated content. Furthermore, most of the figures are supplied as vector graphics, i.e. they can be enlarged without loss of quality in the pdf version.



The pdf version of this work is also available at

http://www.philipp-diesinger.de/

and

http://www.tphys.uni-heidelberg.de/~diesinger.

J Publications

Major parts of this thesis have already been published in scientific journals, are currently under revision or are in preparation to be published. Furthermore, several results of this work have already been presented at scientific conferences and workshops.

A complete list of my publications as well as a list of the attended conferences can be found on the next page. They include publications and conference participation during the time of my diploma thesis (publications before 2006).

The details of the papers which are not yet submitted are only provisional and may still change.

List of Publications

Diesinger, P.M. 2005. *Excluded Volume Effects in Biopolymers.* (diploma thesis) [61]

Brill, M. P.M. Diesinger and D.W. Heermann. 2005. *Knots in Macromolecules in Constraint Space.* arXiv:cond-mat/0507020

Diesinger, P.M. 2005. *Two-Angle Model and Phase Diagram for Chromatin.* European Biophysics Journal (Abstract), Volume 34 6

Diesinger, P.M. and D.W. Heermann. 2006. *Two-Angle Model and Phase Diagram for Chromatin.* Physical Review E 74, 1 [62]

Diesinger, P.M. and D.W. Heermann. 2008. *Average Crossing Number of Gaussian and Equilateral Chains with and without Excluded Volume.* European Journal of Physics B, Volume 62, Number 2, 209-214 [63]

Diesinger, P.M. and D.W. Heermann. 2007. *The Influence of the Cylindrical Shape of the Nucleosomes and H1 Defects on the Properties of Chromatin.* Biophys. J. doi:10.1529/biophysj.107.113902 [64]

Diesinger, P.M. and D.W. Heermann. 2009. *Depletion Effects massively change Chromatin Properties and influence Genome Folding.* Submitted to Biophysical Journal (status: under review) [65]

Diesinger, P.M. and D.W. Heermann. 2009. *Hydrophobicity as a possible Reason for Gelation of FG-rich Nucleoporins.* Submitted to European Biophysics Journal (status: under review) [66]

Diesinger, P.M. M. Bohn, R. Kaufmann, A. von Bassewitz, P. Lemmer, M. Gunkel, Y. Weiland, P. Müller, M. Hausmann, C. Cremer and D.W. Heermann. 2009. *Differences in Chromatin Nano-Structure of Mammalian Cells.* In preparation (to be submitted to Science Letters) [26]

Bohn, M. Diesinger, P.M. R. Kaufmann, A. von Bassewitz, P. Lemmer, M. Gunkel, Y. Weiland, P. Müller, M. Hausmann, C. Cremer and D.W. Heermann. 2009. *Differences in local chromatin structure of Fibroblasts and HeLa cells.* In preparation [27]

Diesinger, P.M. S. Kunkel, J. Langowski and D.W. Heermann. 2009. *The Influence of Histone Depletion on Chromatin Loop Statistics.* In preparation [68]

Diesinger, P.M. and D.W. Heermann. 2009. *The Connection of the mACN with the Scattering Function and the Loop Number of Polymer Chains.* In preparation [67]

Conference Participation

April 1-3, 2004	Workshop on Statistical Physics of Molecular and Cell Biological Systems and Networks, Heidelberg, Germany;
August 27-31, 2005	15^{th} IUPAB & 5^{th} EBSA International Biophysics Congress, Montpellier, France; poster presentation on "Chromatin within in the Framework of" the Two-Angle Model";
June 5-9, 2006	XX SITGES CONFERENCE on STATISTICAL MECHANICS Physical Biology: from Molecular Interactions to Cellular Behavior, Sitges, Barcelona, Spain; poster presentation on "Knot Statistics and Entanglement for Biomolecules";
September 20-22, 2006	MMS06: Methods of Molecular Simulation 2006, Interdisciplinary Center for Scientific Computing (IWR), Heidelberg, Germany;
March 14-16, 2007	Workshop on computer simulations of soft matter and biosystems, BioQuant, Heidelberg, Germany;
July 9-13, 2007	StatPhys 23, the XXIII IUPAP International Conference on Statistical Physics,; Genova, Italy; poster presentation
January 18-19, 2008	Workshop "Monte-Carlo-Methoden" Fakultät für Physik und Astronomie, Heidelberg, Germany;
September 26, 2008	HICCup DKFZ, Heidelberg, Germany;
October 20-23, 2008	Lorentz Center Workshop The physics of genome folding and function Leiden University, Leiden, Netherlands;
November 20, 2008	HGS MathComp Modellierungstag Interdisciplinary Center for Scientific Computing (IWR), Heidelberg, Germany;
December 5, 2008	HICCup Villa Bosch Studio, Heidelberg; Invited Talk: "Genome Folding at the 30nm Scale";

January 22-24, 2009	Workshop on Monte Carlo Methods IWR, Heidelberg;
February 4-6, 2009	Biophysics of Chromatin Workshop Villa Bosch Studio, Heidelberg; Contributed Talk: "The Effects of Nucleosome and Linker Histone Depletion on Properties of Chromatin: A Monte Carlo Model";

K List of Tables

List of Tables

L List of Figures

List of Figures

References

[1] Adams C. C. The Knot Book W. H. Freeman and Company, New York, 1994.

[2] Alberts, B. A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. 2002. Molecular Biology of the Cell. Garland Science, New York and London.

[3] Anderson, J. D. and Widom, J. 2001. Poly(dA-dT) promoter elements increase the equilibrium accessibility of nucleosomal DNA target sites. Mol. Cell Biol. 21, 3830–3839.

[4] Arteca G. A. 1993. Biopolymers 33:1829-1841.

[5] Arteca G. A. 1994. Phys. Rev. E 2417-2428.

[6] Arteca G. A. 1995. Pys. Rev. E 51:2600-2610.

[7] Arteca G. A. 1997. Phys. Rev. E 56:4516-4520.

[8] Arteca G. A. 2002. J. Chem. Inf. Comp. Sci. 42:326-330.

[9] Aten, J.A. and Roland Kanaar. 2006. Chromosomal Organization: Mingling with the Neighbors. PLoS Biol. 2006 May; 4(5): e155.

[10] Atiyha M. F. 1990 The Geometry and Physics of Knots, Cambridge Univ. Press.

[11] Aryeh, B. Hydrophobic Interaction. Plenum Press, New York (ISBN 0-306-40222-X).

[12] Barber, C. B., D.P. Dobkin, and H.T. Huhdanpaa, "The Quickhull Algorithm for Convex Hulls," ACM Transactions on Mathematical Software, Vol. 22, No. 4, Dec. 1996, p. 469-483. Available in PDF format at http://www.acm.org/pubs/citations/journals/toms/1996-22-4/p469-barber/ and http://www.qhull.org/.

[13] Barbi et al. 2004. How the chromatin fiber deals with topological constraints. Phys. Rev. E 71, 031910.

[14] Bayer, R. K. 1994. Colloid. Polym. Sci. 272:910-932.

[15] Bayliss et al., J. Mol. Biol. 293, 579 (1999).

[16] Bayliss, R., T. Littlewood, M. Stewart, Cell 102, 99 (2000).

[17] Beard D. A. and Schlick T. 2001. Biopolymers 58:106-115.

[18] Beard D. A. and Schlick T. 2001. Structure 9:105-114.

[19] Bednar, J., Horowitz, R.A., Grigoryev, S.A., Carruthers, L.M., Hansen, J.C., Koster, A.J. and Woodcock, C.L. 1998. Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. Proc. Natl. Acad. Sci. USA. 95:14173-14178.

[20] Bednenko, J., G. Cingolani, L. Gerace, J. Cell Biol. 162, 391 (2003).

[21] Benecke, A 2003a. Chromatin dynamics are a hyperstructure of nuclear organization. in Proceedings Modelling and Simulation of Biological Processes in the Context of Genomics, Amar P., Kepes F., Norris V., and Tracqui, P. (eds.) Platypus Press, pp. 31-40.

[22] Benecke, A. 2003b. Genomic plasticity and information processing by transcriptional coregulators, ComPlexUs 1, 65-76.

[23] Bhanot, G. 1988. The Metropolis algoritm, Rept. Prog. Phys. 51:429

[24] Binder, K. and D. W. Heermann. 2002. The Monte Carlo Method in Statistical Physics, 4th Edition. Springer Series in Solid-State Sciences.

[25] Bohn, M. D.W. Heermann and R. van Driel. 2007. A Random Loop Model for Long Polymers. Phys. Rev. E 76, 051805 (2007)

[26] Bohn, M. P.M. Diesinger, R. Kaufmann, A. von Bassewitz, P. Lemmer, M. Gunkel, Y. Weiland, P. Müller, M. Hausmann, C. Cremer and D.W. Heermann. 2009. Differences in chromatin nano-structure of mammalian cells. In preparation (Science Letters).

[27] Bohn, M. P.M. Diesinger, R. Kaufmann, A. von Bassewitz, P. Lemmer, M. Gunkel, Y. Weiland, P. Müller, M. Hausmann, C. Cremer and D.W. Heermann. 2009. Differences in local chromatin structure of Fibroblasts and HeLa cells. In preparation.

[28] Bolzer, A. Gregor Kreth, Irina Solovei, Daniela Koehler, Kaan Saracoglu, Christine Fauth, Stefan Müller, Roland Eils, Christoph Cremer, Michael R. Speicher, Thomas Cremer. Three-Dimensional Maps of All Chromosomes in Human Male Fibroblast Nuclei and Prometaphase Rosettes. 2005. PLoS Biol 3(5): e157 DOI: 10.1371/journal.pbio.0030157

[29] Barry Boots, Atsuyuki Okabe, Kokichi Sugihara and Sung Nok Chiu. 2000. Spatial Tessellations: Concepts and Applications of Voronoi Diagrams (Wiley Series in Probability & Mathematical Statistics) Wiley & Sons, Second Edition.

[30] Bronstein, Semendjajew, Musoil, Mühlig Taschenbuch der Mathematik, Verlag Harri Deutsch.

[31] Buck G. 1998. Nature 392:238-239.

[32] Buck G. and Simon J. 1999. Topology and its Applications 69:1-12.

[33] Butler P. J. G., and Thomas J. O. 1998 J.Mol. Biol. 281:401-407.

[34] Bystricky, K., P. Heun, L. Gehlen, J. Langowski and S. M. Gasser. 2004. Long-range compaction and flexibility of interphase chromatin in budding yeast analyzed by high-resolution imaging techniques. Proc. Natl. Acad. Sci. U S A. 2004 November 23; 101(47): 16495–16500.

[35] Cantarella J., Kusner R. B. and Sullivan J. M. 1998. Nature 392:237-238.

[36] Cantarella J., Kusner R. B. and Sullivan J. M. 2002. Invent. math. 150:257-286.

[37] Castro, C. 1994. Measurement of the elasticity of single chromatin fibers: the effect of histone H1. Ph.D. thesis. University of Oregon, Eugene.

[38] Chakravarthy S, Park YJ, Chodaparambil J, Edayathumangalam RS, Luger K, "Structure and dynamic properties of nucleosome core particles", FEBS Letters. (2005) Feb 7; 579 (4): 895-898.

[39] Chambeyron, S. and Bickmore, W.A. 2004. Does looping and clustering in the nucleus regulate gene expression? Curr. Opin. Cell Biol. 16: 256–262.

[40] Clark D. J. and Kimura T. 1990. J. Mol. Biol. 211:883-896.

[41] Cook, P. Principles of Nuclear Structure and Function. Wiley-Liss; 1st Edition (April 24, 2001).

[42] Cook, P. Nat Genet 32, 347 2002. http://dx.doi.org/10.1038/ng1102-347.

[43] Corces, V. G. 1995. Chromatin insulators. Keeping enhancers under control. Nature 376:462-463.

[44] Cremer, T. 1985. Von der Zellenlehre zur Chromosomentheorie: Naturwissenschaftliche Erkenntnis und Theorienwechsel in der frühen Zell- und Vererbungsforschung, Veröffentlichungen aus der Forschungsstelle für Theoretische Pathologie der Heidelberger Akademie der Wissenschaften. Springer-Vlg., Berlin, Heidelberg.

[45] Cremer, T. and Cremer, C. 2001. Chromosome territories, nuclear architecture and gene regulation in mammalian cells, Nature Rev. Genetics 2, 292-301.

[46] Cremer, M. Katrin Küpper, Babett Wagler, Leah Wizelman, Johann von Hase, Yanina Weiland, Ludwika Kreja, Joachim Diebold, Michael R Speicher and Thomas Cremer. Inheritance of gene density-related higher order chromatin arrangements in normal and tumor cell nuclei. J Cell Biol, 162(5):809–820, Sep 2003.

[47] Cui, Y., and C. Bustamante. 2000. Pulling a single chromatin fiber reveals the forces that maintain its higher-order structure. Proc. Nat. Acad. Sci. U. S. A. 97:127–132.

[48] de Gennes, P.G. 1979. Scaling concepts in polymer physics. Cornell University Press, Ithaca, NY.

[49] de Laat, W. and Grosveld, F. 2003. Spatial organization of gene expression: The active chromatin hub. Chromosome Res. 11: 447–459.

[50] de Laat, W. & Grosveld, F. 2007. Inter-chromosomal gene regulation in the mammalian cell nucleus. Curr. Opin. Genet. Dev., published online (doi:10.1016/j.gde.2007.07.009).

[51] Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. 2002. Capturing chromosome conformation. Science 295: 1306–1311.

[52] Dekker, J. 2003. A closer look at long-range chromosomal interactions. Trends Biochem. Sci. 28: 277–280.

[53] Dekker, J. The three 'C's of chromosome conformation capture: controls, controls, controls. Nat. Methods 3, 17–21 (2006).

[54] Denning D., Patel S., Uversky V., Fink A. and Rexach M. (2003). "Disorder in the nuclear pore complex: the FG repeat regions of nucleoporins are natively unfolded". Proc Natl Acad Sci U S A 100 (5): 2450–5. doi:10.1073/pnas.0437902100.

[55] Diao Y., 1995. Journal of Knot Theory and its Ramifications, 4:189-196.

[56] Diao Y., Dobay A., Knuser R. B., Millett K., Stasiak A. 2003 J. Phys. A: Math. Gen. 36:11561-11574.

[57] Diao Y., Ernst C. The Average Crossing Number of Gaussian Random Walks and Polygons - preprint.

[58] Diao Y., Nardo J. and Sun Y. 2001. J. Knot Ramifications 10:597-607.

[59] Diao Y., Pippenger N. and Sumners D. W. 1994. J. Knot Ramifications 3:419-429.

[60] Dill K.A. (1985). "Theory for the folding and stability of globular proteins". Biochemistry 24 (1501): 1501. doi:10.1021/bi00327a032

[61] Diesinger, P.M. 2005. Excluded Volume Effects in Biopolymers (diploma thesis), http://www.tphys.uni-heidelberg.de/~diesinger/data/Download/Diplomarbeit_DIESINGER.pdf or alternatively http://www.philipp-diesinger.de/data/Download/Diplomarbeit_DIESINGER.pdf.

[62] Diesinger, P. M., Heermann, D. W. 2006. Two-angle model and phase diagram for chromatin. Phys. Rev. E 74, 031904.

[63] Diesinger, P.M. and D.W. Heermann. Nov. 2008. Average crossing number of Gaussian and equilateral chains with and without excluded volume. European Journal of Physics B, Vol 62, No. 2, 209-214

[64] Diesinger, P.M. and D.W. Heermann, 2008. The influence of the cylindrical shape of the nucleosomes and H1 defects on the properties of chromatin Biophys. J. doi:10.1529 biophysj.107.113902

[65] Diesinger, P.M. and D.W. Heermann. 2009. Depletion Effects massively change Chromatin Properties and influence Genome Folding. Submitted to Biophysical Journal (status: under review).

[66] Diesinger, P.M. and D.W. Heermann. 2009. Hydrophobicity as a possible Reason for Gelation of FG-rich Nucleoporins. Submitted to European Biophysics Journal (status: under review).

[67] Diesinger, P.M. S. Kunkel, J. Langowski and D.W. Heermann. 2009. The Influence of Histone Depletion on Chromatin Loop Statistics. In preparation.

[68] Diesinger, P.M. and D.W. Heermann. 2009. The Connection of the mACN with the Scattering Function and the Loop Number of Polymer Chains. In preparation.

[69] Dobay A., Diao Y., Dubochet J. and Stasiak A. Scaling of the average crossing number in equilateral random walks, knots and proteins - preprint.

[70] Doi M. and Edwards S. F. 1986. The Theory of Polymer Dynamics. Oxford University Press, Oxford.

[71] Dorigo et al. Science 306, 1571.

[72] Dostie, J., Richmond, T.A., Arnaout, R.A., Selzer, R.R., Lee, W.L., Honan, T.A., Rubio, E.D., Krumm, A., Lamb, J., Nusbaum, C., Green, R.D. and Dekker, J. 2006. Chromosome Conformation Capture Carbon Copy (5C): A massively parallel solution for mapping interactions between genomic elements. Genome Research, 16(10): 1299-1309.

[73] Dostie, J. & Dekker, J. Mapping networks of physical interactions between genomic elements using 5C technology. Nat. Protoc. 2, 988–1002 (2007).

[74] Eberharter, A. and Becker, P.B. (2002) Histone acetylation: a switch between repressive and permissive chromatin, EMBO Reports 3, 224-229.

[75] Edwards S. F. 1965 Proc. Phys. Soc. Vol.85.

[76] Elliott B, Jasin M (2002) Double-strand breaks and translocations in cancer. Cell Mol Life Sci 59:373–385.

[77] Fan, Y., T. Nikitina, J. Zhao, T. S. Fleury, R. Bhattacharyya, E. E. Bouhassira, A. Stein, C. L. Woodcock and A. I. Skoultchi. 2005. Histone H1 Depletion in Mammals Alters Global Chromatin Structure but Causes Specific Changes in Gene Regulation. DOI 110.1016/J. Cell. 2005.10.028.

[78] Felsenfeld G. 1992 Nature (London) 355:219-224.

[79] Fisher, M. E. 1971. Critical Phenomena: Enrico Fermi Summer School, New York, Academimc Press.

[80] Finch, Klug 1976 Proc. Natl. Acad. Sci. USA 73:1897-1901.

[81] Fletcher, T. M. and Hansen J. C. 1996 Crit. Rev. Eucaryotic Gene Expression 6:149-188.

[82] Flory, P. J. 1941. J. Am. Chem. Soc. 63:3083.

[83] Flory, P.J. 1953. Principles of Polymer Chemistry. University Press, Ithaca, NY. ISBN-10: 0801401348. ISBN-13: 978-0801401343.

[84] Flory, P. J. 1956. Theory of elastic mechanisms in fibrous proteins. J. Am. Chem. Soc. 78:5222–5235.

[85] Frank-Kamenetskii M. D., Lukashin A. V. & Vologodskii 1975, Nature 258:398-402.

[86] Freedman M. H. and He Z. 1991 Annals of Math. 134:189-229.

[87] Frey, S., Ralf P. Richter, Dirk Görlich. 2006. FG-Rich Repeats of Nuclear Pore Proteins Form a Three-Dimensional Meshwork with Hydrogel-Like Properties. Science Vol 314 3 Nov 2006.

[88] Frey, S. Private communication

[89] Frisch H. L. & Wasserman 1961, J. Am. Chem. Soc. 83, 3789-3795.

[90] Gassmann, R., Vagnarelli, P., Hudson, D., and Eamshaw, W.C. (2004) Mitotic chromosome formation and condensin paradox, Exper. Cell Res. 296, 35-42.

[91] Gasquet C. and Witomski P. Fourier Analysis and Applications, Springer, Texts in Applied Mathematics 30.

[92] Gheldof, N., Tabuchi, T.M., and Dekker, J. 2006. The active FMR1 promoter is associated with a large domain of altered chromatin conformation with embedded local histone modifications. Proc. Natl. Acad. Sci. 103: 12463–12468.

[93] Görlich, D., U. Kutay, Annu. Rev. Cell Dev. Biol. 15, 607 (1999).

[94] Gotta, M., Laroche, T., Formenton, A., Maillet, L., Scherthan, H. & Gasser, S. M. 1996. J. Cell Biol. 134, 1349–1363.

[95] Gradshteyn and Ryzhik Table of Integrals, Series and Products, Academic Press.

[96] Grassberger P. 2001 J. Phys. A: Math. Gen. 34:9959-9963.

[97] Greiner W., Neise L., and Stöcker H. Theoretische Physik 9 Thermodynamik und Statistische Dynamik, Verlag Harri Deutsch.

[98] Guadagnini, E. 1993 The Link Invariants of the Chern-Simons Field Theory. Walter de Gruyter, Berlin.

[99] Hansen, J.C. (2002) Conformational dynamics of the chromatin fiber in solution: determinants, mechanisms, and functions, Annu. Rev. Biophys. Biomol. Struct. 31, 361-392.

[100] Hastings, W.K. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. In: Biometrika. 57, 1970, S. 97-109.

[101] Heermann, D.W. 1990 Computer Simulation Methods in Theoretical Physics 2nd ed., Springer Verlag, Heidelberg

[102] Hermans, J. 1982. Excluded-volume theory of polymer—protein interactions based on polymer chain statistics. J. Chem. Phys. 77:2193.

[103] Heun, P., Laroche, T., Raghuraman, M. K. & Gasser, S. M. 2001. J. Cell Biol. 152, 385–400.

[104] Hendrix, J. C. Flors, P. Dedecker, J. Hofkens, and Y. Engelborghs. Dark States in Monomeric Red Fluorescent Proteins Studied by Fluorescence Correlation and Single Molecule Spectroscopy. Biophysical Journal, 94:4103-4113, 2008.

[105] Horowitz et al. 1994 J. Cell. Biol. 125:1-10.

[106] Houchmandzadeh, B., J. F. Marko, D. Chatenay, and A. Libchaber. 1997. Elasticity and structure of eukaryote chromosomes studied by micromanipulation and micropipette aspiration. J. Cell Biol. 139:1–12.

[107] Hurt, E. C. EMBO J. 7, 4323 (1988).

[108] Iovine, M.K., J. L. Watkins, S. R. Wente, J. Cell Biol. 131, 1699 (1995).

[109] Jackson, D.A. S. J. McCready, and P. R. Cook, J Cell Sci Suppl 1, 59 (1984).

[110] Jenuwein, T. & Allis, C. D. Translating the histone code. Science 293, 1074—1080 (2001).

[111] Karpov, V. L., Bavykin, S. G., Preobrazhenskaya, 0. V., Belyavsky, A. V. and Mirzabekov, A. D. 1982. Nucleic Acids Res. 10, 4321-4337.

[112] Katritch V. et al. 1996 Nature 384:142-145.

[113] Kaufmann, R. Paul Lemmera, Manuel Gunkela, Yanina Weilanda, Patrick Müllerb, Michael Hausmannb, David Baddeleya,c, Roman Ambergera, Christoph Cremer. 2008. SPDM – Single Molecule Superresolution of Cellular Nanostructures. Applied Physics B (2008) 93: 1-12, DOI 10.1007/s00340-008-3152-x

[114] Kawakatsu T. Statistical Physics of Polymers, Springer.

[115] Kennedy T. 2004. A Faster Implementation of the Pivot Algorithm for Self-Avoiding Walks. J. Stat. Phys. 106:407-429.
C++ code of the fast pivot algorithm can be found here: http://math.arizona.edu/~tgk/.

[116] Kepert, J.F. Jacek Mazurkiewicz, Gerrit L. Heuvelman, Katalin Fejes Toth, and Karsten Rippe (2005) NAP1 Modulates Binding of Linker Histone H1 to Chromatin and Induces an Extended Chromatin Fiber Conformation JBC Papers in Press, August 16, 2005, DOI 10.1074/jbc.M507322200

[117] Kholodenko, A.L. and D. P. Rolfsen 1996 J. Phys. A: Math. Gen. 29:5677

[118] Khrapunov S.N. Dragan A. I., Sivolob A. V. and Zagariya A. M. 1997 Biochim. Biophys. Acta. 1351:213-222.

[119] Koolmann, Röhm Taschenatlas der Biochemie, Thieme.

[120] Kornberg, R. D. and Lorch, Y. Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. Cell 98, 285–294 (1999).

[121] Krasnow M. A., Stasiak A., Spengler S. J., Dean F., Koller T. and Cozzarelli N. R. 1993 Nature 304:559-560.

[122] Krickeberg K. and Ziezold H. Stochastische Methoden, Springer-Lehrbuch.

[123] Kuhn W. 1934 Kolloid-Z., 68 2.

[124] Kunkel, S. Joŕg Langowski. 2007. The Ring-Closure Probability of the 30-nm Chromatin Fiber (Diplomarbeit).

[125] Kyte, J., R. F. Doolittle: A simple method for displaying the hydropathic character of a protein.. In: Journal of Molecular Biology. 157, Nr. 1, 1982, S. 105–132.

[126] Lanctôt, C. T. Cheutin, M. Cremer, G. Cavalli, and T. Cremer, Nat Rev Genet 8, 104 (2007), http://dx.doi.org/10.1038/nrg2041.

[127] Landecker, H. (2000): Immortality, In Vitro: A History of the HeLa Cell Line. Biotechnology and Culture: Bodies, Anxieties, Ethics, ed. Paul Brodwin, Indiana University Press: 53-74.

[128] Langowski J. and Schiessel H. Chromatin Structure: State of the Art, eds. S. Leuba & J. Zlatanova, 2004, Elsevier, Amsterdam, pp. 397-420.

[129] Leforestier A. and Livolant F. 1997 Biophys. J. 73:1771-1776.

[130] Leforestier A., Dubochet J. and Livolant F. 2001 Biophys. J. 81:2414-2421.

[131] Lemmer, P. Manuel Gunkel, David Baddeley, Rainer Kaufmann, Alexander Urich, Yanina Weiland, Jürgen Reymann, Patrick Müller, Michael Hausmann, and Christoph Cremer. SPDM: Light Microscopy with Single Molecular Resolution at the Nanoscale. Applied Physics B, 93:1-12, 2008.

[132] Lercher, M.J. A. O. Urrutia, and L. D. Hurst, Nat Genet 31, 180 (2002), ISSN 1061-4036, http://dx.doi.org/10.1038/ng887.

[133] Leuba et. al. 1994 Proc. Natl. Acad. Sci. USA 91:11621-11625.

[134] Liu, Z. & Garrard, W.T. Long-range interactions between three transcriptional enhancers, active V_K gene promoters, and a 3' boundary sequence spanning 46 kilobases. Mol. Cell. Biol. 25, 3220.3231 (2005).

[135] Lohr, D. and van Holde, K. E. 1979. Proc. Natl. Acad. Sci. USA 76, 6326-6330.

[136] Lohr, D. 1986. J. Biol. Chem. 261, 9904-9914.

[137] Luger et al. 1997 Nature. 389:251-260.

[138] Macara, I.G., Microbiol. Mol. Biol. Rev. 65, 570 (2001).

[139] Madras N. and Sokal A. J. Stat. Phys. 50:109.

[140] Maeshima, K. and Laemmli, U.K. (2003) A two step scafolding model for mitotic chromosome assembly, Developmental Cell 4, 467-480.

[141] Maeshima, K. and Mikhail Eltsov. 2008. Packaging the Genome: the Structure of Mitotic Chromosomes. Journal of Biochemistry 2008 143(2):145-153; doi:10.1093/jb/mvm214

[142] Mansfield M. L. 1994 Nat. Struct. Biol. 1:213-214.

[143] Manders, E.M.M., Visser, A.E., Koppen, A., de Leeuw, W.C., van Liere, R., Brakenhof, G.J., and van Driel, R. (2003). Four dimensional imaging of chromatin dynamics during the assembly of the interphase nucleus, Chromosome research 11, 537-547.

[144] Julio Mateos-Langerak, Manfred Bohn, Wim de Leeuw, Osdilly Giromus, Erik M. M. Manders, Pernette J. Verschure, Mireille H. G. Indemans, Hinco J. Gierman, Dieter W. Heermann, Roel van Driel, and Sandra Goetze. Spatially confined folding of chromatin in the interphase nucleus. Proceedings of the National Academy of Sciences. February 2009.

[145] Mehring, C. 1998. Modellierung der Dynamik von Interphasechromosomen. Diplomarbeit, Universität Heidelberg.

[146] Mergell, B., R. Everaers and H. Schiessel. 2004. Nucleosome interactions in chromatin: fiber stiffening and hairpin formation. Phys. Rev. E, 70, 011915-1-9.

[147] Metropolis, N. A. Rosenbluth, M. Rosenbluth, A. Teller und E. Teller: Equation of State Calculations by Fast Computing Machines. Journal of Chemical Physics. 21, 1953, S. 1087-1092 (doi:10.1063/1.1699114)

[148] Mozziconacci, J. and Victor, J.M. (2003) Nucleosome gaping supports a functional structure for the 30nm chromatin fiber, J. Struct. Biol. 143, 72-76.

[149] Mozziconacci, J., Lavelle, C., Barbi, M., Lesne, A., and Victor, J.M. (2005) A unified theory of DNA condensation and decondensation in eukaryotic chromosomes. Submitted to FEBS Letters.

[150] Murrell, A., Heeson, S. & Reik, W. Interaction between differentially methylated regions partitions the imprinted genes Igf2 and H19 into parentspecific chromatin loops. Nat. Genet. 36, 889.893 (2004).

[151] Olins, D.E. and Ada L. Olins. 2003. Chromatin history: our view from the bridge. Nat Rev Mol Cell Biol 4: 809-14.

[152] Olins' Lab: http://www.bowdoin.edu/faculty/a/aolins/olinslab.shtml.

[153] Orlandini, E. M. C. Tesi, S. G. Whittington, D. W. Summers and E. J. Janse van Rensburg 1994 J. Phys. A.:Math. Gen. 27 L333

[154] Osborne CS, Chakalova L, Brown KE, Carter D, Horton A, et al. 2004. Active genes dynamically colocalize to shared sites of ongoing transcription. Nat Genet 36:1065–1071.

[155] O'Sullivan, J.M. et al. 2004. Gene loops juxtapose promoters and terminators in yeast. Nat. Genet. 36, 1014–1018.

[156] Palstra, R.J. et al. The β-globin nuclear compartment in development and erythroid differentiation. Nat. Genet. 35, 190.194 (2003).

[157] Patel, Samir S. 2001-2006. Nuclear Transport Illustrations. http://sspatel.googlepages.com/nuclearporecomplex2

[158] Patterson, G.H. and J. Lippincott-Schwartz. A Photoactivatable GFP for Selective Photolabeling of Proteins and Cells. Proceedings of the National Academy of Sciences of the United States of America, 95:13573-13578, 1998.

[159] Pemberton, L.F., B. M. Paschal, Traffic 6, 187 (2005).

[160] Peters R. (2006). "Introduction to nucleocytoplasmic transport: molecules and mechanisms". Methods Mol Biol 322: 235–58.

[161] Plattner, Hentschel Zellbiologie, Thieme.

[162] Poland, D. C., and H. A. Scheraga. 1965. Statistical mechanics of noncovalent bonds in polyamino acids. VIII. Covalent loops in protein. Biopolymers. 3:379–399.

[163] Poland D. and Scheraga H. A. 1970 Theory of Helix-Coil Transitions in Biopolymers Academic Press, NY.

[164] Poupon, A. (2004). Voronoi and Voronoi-related tessellations in studies of protein structure and interaction. Current Opinion in Structural Biology, 14(2), 233-241.

[165] Pury P. A. 2003. arXiv:cond-mat/0305178v1.

[166] Pury P. A. and Cáceres M. O. 2003 J.Phys A 36:2695.

[167] Radu, A., M. S. Moore, G. Blobel, Cell 81, 215 (1995).

[168] Rayleigh L. 1919 Phil. Mag. S. 6 Vol. 37 No. 220, 321-347.

[169] Routh, A., Sandin, S., Rhodes, D. 2008. Nucleosome repeat length and linker histone stoichiometry determine chromatin fiber structure. Proc. Natl. Acad. Sci. vol. 105 no. 26 8872-8877

[170] Ribbeck, K., D. Görlich, EMBO J. 20, 1320 (2001).

[171] Ribbeck, K., D. Görlich, EMBO J. 21, 2664 (2002).

[172] Rippe, K. 2001. Making contacts on a nucleic acid polymer. Trends Biochem. Sci. 26: 733–740.

[173] Richmond, T. J. and Davey, C. A. 2003. The structure of DNA in the nucleosome core. Nature 423, 145–150.

[174] Rout, M.P., S. R. Wente, Trends Cell Biol. 4, 357 (1994).

[175] Rout et al., J. Cell Biol. 148, 635 (2000).

[176] Sahimi, M. Applications of Percolation Theory. Taylor & Francis.

[177] Saitta A. M. and Soper P. D. 1999. Nature 399:46-48.

[178] Satchwell, S. C., Drew, H. R. and Travers, A. A. 1986. Sequence periodicities in chicken nucleosome core DNA. J. Mol. Biol. 191, 659–675.

[179] Schalch, T., Duda, S., Sargent, D.F., and Richmond T.J. 2005. X-ray structure of a tetranucleosome and its implications for the chromatin fibre, Nature 436, 138-141.

[180] Schellmann, J. A. editor. 1955. The stability of hydrogen-bonded peptide structures in aqueous solution. C. R. Trav. Lab. Carlsberg Ser. Chim. 29:230–259.

[181] Schellman J. and Stigter D. 1977 Biopolymers. 16:1415-1434.

[182] Schermelleh L, Carlton PM, Haase S, Shao L, Winoto L, Kner P, Burke B, Cardoso MC, Agard DA, Gustafsson MG, Leonhardt H, Sedat JW. 2008. "Subdiffraction multicolor imaging of the nuclear periphery with 3D structured illumination microscopy". Science 320 (5881): 1332–6. DOI:10.1126/science.1156947. PMID 18535242.

[183] Schiessel, H., W. M. Gelbart and R. Bruinsma. 2001. DNA Folding: Structural and Mechanical Properties of the Two-Angle Model for Chromatin. Biophys. J., April 2001, p. 1940-1956, Vol. 80, No. 4.

[184] Schiessel, H. 2003. Topical Review: The physics of chromatin. J. Phys.: Condens. Matter 15, R699-R774.

[185] Schlick, T. et. al. 2000. Computing, Science and Engeneering Vol.2, No6 pp. 38-51.

[186] Segal, E. Yvonne Fondufe-Mittendorf, Lingyi Chen, AnnChristine Thaström, Yair Field, Irene K. Moore, Ji-Ping Z. Wang and Jonathan Widom.

2006. A genomic code for nucleosome positioning. Vol 442-17 August 2006 doi:10.1038/nature04979.

[187] Segal et al. 2008. Average nucleosome occupancy for the whole yeast genome. http://genie.weizmann.ac.il/pubs/nucleosomes06/segal06_data.html.

[188] Sekinger, E. A., Moqtaderi, Z. and Struhl, K. 2005. Intrinsic histone-DNA interactions and low nucleosome density are important for preferential accessibility of promoter regions in yeast. Mol. Cell 18, 735–748.

[189] Semlyen, J. A. 1997. Large Ring Molecules. John Wiley and Sons, Chichester, West Sussex, England.

[190] Sharrer, T. The Scientist Vol. 20, No. 7 July 2006, p. 22 HeLa Herself.

[191] Simonis, M. et al. 2006. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat. Genet. 38, 1348–1354.

[192] Simonis, M. Juŕgen Kooren & Wouter de Laat. 2007. An evaluation of 3C-based methods to capture DNA interactions. Nature Methods. Vol.4 No.11 Nov. 2007-895

[193] Sinnecker, D. P. Voigt, N. Hellwig, and M. Schaefer. 2005. Reversible Photobleaching of Enhanced Green Fluorescent Proteins. Biochemistry, 44:7085-7094.

[194] Skok, J.A. et al. 2007. Reversible contraction by looping of the Tcra and Tcrb loci in rearranging thymocytes. Nat. Immunol. 8, 378–387.

[195] Spector, D.L. 2003. The dynamics of chromosome; organization and gene regulation, Annu. Rev. Biochem. 72, 573-608.

[196] Spilianakis, C.G. & Flavell, 2004. R.A. Long-range intrachromosomal interactions in the T helper type 2 cytokine locus. Nat. Immunol. 5, 1017.1027.

[197] Spilianakis CG, Lalioti MD, Town T, Lee GR, Flavell RA. 2005. Interchromosomal associations between alternatively expressed loci. Nature 435:637–645.

[198] Srinivasan, R. 2002. Importance sampling - Applications in communications and detection. Springer-Verlag, Berlin 2002, ISBN 978-3-540-43420-7.

[199] Stasiak A., Katritch V., Bednar J., Michoud D. and Dubochet J. 1996 Nature 384:122.

[200] Stauffer, D. and Amnon Aharony. Introduction to Percolation Systems. Taylor & Francis.

[201] Stehr, R. Nick Kepper, Ramona Ettig, Karsten Rippe and Gero Wedemann. 2009. Computer simulations of chromatin fibers. Biophysical Journal, Volume 96, Issue 3, 55a, 1 February 2009 doi:10.1016/j.bpj.2008.12.180

[202] Stigter D. 1977. Biopolymers 16:1435-1448.

[203] Stockmayer, W. H. 1943. J. chem. Phys. 11:45.

[204] Stöcker, Taschenbuch der Physik, Verlag Harri Deutsch.

[205] Straight, A. F., Belmont, A. S., Robinett, C. C. & Murray, A. W. 1996. Curr. Biol. 6, 1599–1608.

[206] Strauss, F. and Prunell, A. 1983. EMBO J. 2, 51-56.

[207] Strobl, G. Textbook: "The Physics of Polymers: Concepts for Understanding Their Structures and Behavior". Third Edition. Springer.

[208] Taylor W. R. 2000 Nature 406:916-919.

[209] Thaström, A. et al. 1999. Sequence motifs and free energies of selected natural and non-natural nucleosome positioning DNA sequences. J. Mol. Biol. 288, 213–229.

[210] The ENCODE (ENCyclopedia Of DNA Elements) Project. http://www.sciencemag.org/cgi/content/abstract/306/5696/636.

[211] Thoma F., Koller T. and Klug A. J. Cell. Biol. 83:403-427.

[212] Tolhuis, B., Palstra, R.J., Splinter, E., Grosveld, F. & de Laat, W. 2002. Looping and interaction between hypersensitive sites in the active â-globin locus. Mol. Cell 10, 1453–1465.

[213] Trifonov, E. N. 1980. Sequence-dependent deformational anisotropy of chromatin DNA. Nucleic Acids Res. 8, 4041–4053.

[214] van den Engh, G., R. Sachs, and B. Trask. 1992. Estimating genomic distance from DNA sequence location in cell nuclei by a random walk model. Science. 257:1410 –1412.

[215] Vanderzande, C. 1996. Lattice Models of Polymers, Cambridge University Press.

[216] van Holde K. E. 1989. Chromatin. Springer Verlag, New York.

[217] van Holde, K. and J. Zlatanova, Proc. Natl. Acad. Sci. USA 93, 10548.

[218] van Holde and Zlatanova. 1995. J. Biol. Chem. 93:8373-8378. 1996.

[219] van Holde K. and Zlatanova. J. 1996. Proc. Natl. Acad. 93:10548:10555.

[220] van Rensburg E. J., Sumners D. A. W., Wasserman E. & Whittington. 1992. J. Phys. A 25:6557-6566.

[221] Vologodskii A., Crisona N., Laurie B., Pieranski P., Katritch V., Dubouchet J. and Stasiak A. 1998. J. Mol. Biol. 278:1-3.

[222] Wang M.C. and G.E. Uhlenbeck. 1945. On the theory of Brownian motion II. Rev. Mod. Phys., 17, 323.

[223] Wedemann G. and Langowski J. 2002. Computer Simulation of the 30-Nanometer Chromatin Fiber. Biophysical Journal, Volume 82, Issue 6, 2847-2859, doi:10.1016/S0006-3495(02)75627-0.

[224] Weis, K., Curr. Opin. Cell Biol. 14, 328 (2002).

[225] West, A.G. and Fraser, P. 2005. Remote control of gene transcription. Hum. Mol. Genet. 14: R101–R111.

[226] Widom J. 1986 J. Mol. Biol. 190:411-424.

[227] Widom J. 1989 Annu. Rev. Biophys. Chem. 18:365-395.

[228] Widom, J. 1992. A relationship between the helical twist of DNA and the ordered positioning of nucleosomes in all eukaryotic cells. Proc. Natl. Acad. Sci. USA 89:1095-1099.

[229] Widom, J. 2001. Role of DNA sequence in nucleosome stability and dynamics. Q. Rev. Biophys. 34, 269–324.

[230] Widom, J. and A. Klug, Annu. Rev. Biophys. Chem. 43, 207.

[231] Wikipedia (The Free Encyclopedia) - http://de.wikipedia.org/.

[232] Witten, E. 1998 Commun. Math. Phys. 121 351

[233] Wolffe A. Chromatin - Structure & Function, Academic Press.

[234] Wolffe, A.P. and J. J. Hayes. 1999. Chromatin disruption and modification. - Nucleic Acid Research, 27, 711-720.

[235] Woodcock et al. 1993. Proc. Natl. Acad. Sci. USA. 90:9021-9025.

[236] Wyman C., Ristic D., Kanaar R. 2004. Homologous recombination-mediated double-strand break repair. DNA Repair (Amst) 3:827–833.

[237] Wyrick, J. J. et al. 1999. Chromosomal landscape of nucleosome-dependent gene expression and silencing in yeast. Nature 402, 418–421.

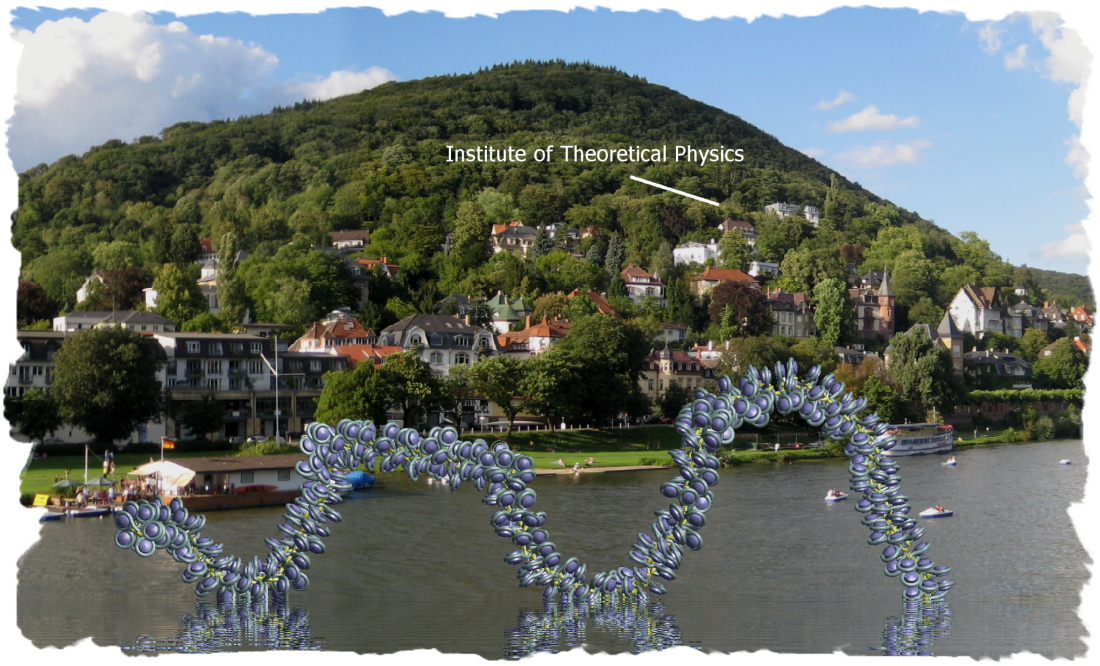
[238] Yao J., Lowary P. T. and Widom J. 1990 Proc. Natl. Acad. Sci. USA 87:7603-7607.

[239] Yokota, H., G. van den Engh, J. E. Hearst, R. K. Sachs, and B. J. Trask. 1995. Evidence for the organization of chromatin in megabase pair-sized loops arranged along a random walk path in the human G0/G1 interphase nucleus. J. Cell Biol. 130:1239–1249.

[240] Yokota, H. M.J. Singer, G.J. van den Engh, and B.J. Trask. 1997. Regional differences in the compaction of chromatin in human G_0/G_1 interphase nuclei. Chromosome Res., 5:157–166.

Acknowledgments

I would like to thank all the people, who contributed to this work and made my time at the institute as pleasant as it was. I learned a lot during my work and enjoyed my time at the Institute of Theoretical Physics very much.



First of all, I would like to thank my professor in charge, **Dieter W. Heermann**, for the effort he put in the supervision of this work. Without his large knowledge and helpful pieces of advice in both physical and technical topics, this work would not have been possible. Many thanks also to my co-advisor **Prof. Christoph Cremer** for many fruitful discussions especially during our collaboration.

This work was mainly funded by a Max-Planck Fellowship of the Max-Planck-Institut für Biophysikalische Chemie in Göttingen. Therefore, I want to thank

Dirk Görlich and **Steffen Frey** for their support.
Furthermore, I want to thank the Heidelberg Graduate School of Mathematical and Computational Methods for the Sciences as well as the Institute of Theoretical Physics (ITP) of the University of Heidelberg for partial funding.

I would also like to thank my lab mates **Manfred Bohn**, **Miriam Fritsche**, **Benoît Knecht**, **Timo Fischer**, **Michael Bach** and **Wadim Smirnow** as well as **Rainer Kaufmann** from the group of Prof. Cremer for many discussions and a pleasant time at the ITP. As the picture above shows my DNA had a wonderful time in Heidelberg.

Furthermore, I thank **Lars Omlor**, **Melinda Szabados**, **Miriam Fritsche** and **Michael Brill** for proofreading my thesis.

Last but not least I would like to thank my family especially my parents **Aloys** and **Margit Diesinger** for their unconditioned support.