

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. Jens Odenheimer
born in Karlsruhe, Germany
Oral examination: June 21st, 2006

Dynamic Simulation of active/inactive Chromatin Domains

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

Dynamische Simulation aktiver und inaktiver Chromatin Domänen

Zusammenfassung: In dieser Arbeit wird ein neues Modell vorgestellt, welches mit Hilfe der Polymerphysik zum ersten Mal die Bildung höherer Organisationsstufen von Chromatin beschreibt. Es handelt sich um ein mesoskopisches Block-Copolymer Modell der 30nm Chromatin Fiber. Verschiedene Substanzen, welche eine Kondensierung bewirken, können als ein effektives attraktives Potential bestimmter Kettenglieder modelliert werden. Auf diese Weise beobachtet man die Entstehung von einzelnen 1Mbp Rosetten aus einer linearen Kette. Ferner wurden mehrere Mbp simuliert, bis hin zu einem ganzen Chromosom und schließlich wurde die Simulation eines ganzen Zellkerns von *Drosophila Melanogaster* durchgeführt. Die Simulationsdaten wurden mit Experimenten verglichen und lieferten eine gute Übereinstimmung. Die Ergebnisse wurden unter anderem bereits in den Zeitschriften *Eur. Biophys. J.*, *Int. J. Mod. Phys. C*, *Int. J. Biol. Phys.* und *Biophys. Rev. Lett.* veröffentlicht. Eine detaillierte Liste befindet sich im Anhang. Simulationen wurden unter anderem auch auf dem IBM Blue Gene/L Supercomputer im Forschungszentrum Jülich durchgeführt.

Dynamic Simulation of active/inactive Chromatin Domains

Abstract: In this thesis a new model is presented, which describes the formation of higher order chromatin structures with the help of polymer physics for the first time. It is a block-copolymer model for the compactification of the 30nm Chromatin fiber into higher order structures. The idea is that basically every condensing agent (HMG/SAR, HP1, cohesin, condensin, DNA-DNA interaction...) can be modelled as an effective attractive potential of specific chain segments. This way the formation of individual 1Mbp sized rosettes from a linear chain could be observed. Furthermore several Mbp of fiber were simulated, up to an entire chromosome and finally the entire nucleus of *Drosophila Melanogaster*. The simulation results were compared to experimental data and good agreement was found. The results have been published in the journals *Eur. Biophys. J.*, *Int. J. Mod. Phys. C*, *Int. J. Biol. Phys.* and *Biophys. Rev. Lett.* A detailed list can be found in the appendix. Part of the computation was done on the IBM Blue Gene/L supercomputer at the Forschungszentrum Jülich.

Contents

1	Background	9
1.1	Scale, packing and the microscope problem	9
1.2	Nuclear structure and Chromatin	11
1.3	Gene Expression and Silencing	11
1.4	The Cell Cycle	13
1.5	Replication and Transcription	13
2	Modeling	17
2.1	The Biological model	17
2.2	Computational Model	18
2.2.1	Integration Scheme	21
2.2.2	Dissipative Particle Dynamics	22
2.3	Physical Model	22
3	Polymer Theory	25
3.1	General Chains	25
3.1.1	End-to-End Distance	25
3.1.2	Radius of Gyration	26
3.2	Freely Joint Chains	26
3.2.1	Distribution of the End-to-End Distance	27
3.3	Freely Rotating Chain	27
3.4	Characteristic Quantities of Polymers	28
3.4.1	Persistence Length	28
3.4.2	Kuhn Segment	29
3.5	Excluded Volume	29
3.5.1	Flory Theory	29
3.5.2	Corrections to the Flory Theory	30
3.5.3	Real End-to-End Distances	30
3.6	Finite Size Scaling	31
3.7	Diffusion	32

4	The 1Mbp Domain	35
4.1	Introduction	35
4.2	Simulation	36
4.3	Results	37
4.3.1	Free Energy	43
5	Diffusion	47
5.1	Motivation	47
5.2	Experiment	47
5.3	Simulation	50
5.4	Results	51
5.4.1	Accessibility	54
5.5	Conclusion	57
6	Chromosome Mapping	59
6.1	Motivation	59
6.2	Experiment	59
6.3	Simulation	61
6.4	Results	62
7	Chromosome 22	67
7.1	Motivation	67
7.2	Simulation	67
7.3	Results	69
8	Drosophila	73
8.1	Motivation	73
8.2	Experiment	74
8.3	Simulation	77
8.4	Results	81
9	Scientific Computing	89
9.1	History of DePoSiTo	89
9.2	Parallelization	89
10	Conclusion	93
A	The simulation program DePoSiTo	97
A.1	Setup	97
A.2	DePoSiTo	101
A.3	Analysis	101
B	Publications	103
B.1	Publications	103
B.2	Conferences	104

Chapter 1

Background

The size of a typical animal cell is about $10 - 30\mu m$. A human eye is capable of resolving about $100\mu m$, thus cells could only be investigated after the invention of the light microscope. Robert Hooke coined the term cell in 1660 and in 1683 Leeuwenhoek discovered the first bacteria. Subsequently, the cell was studied in great detail and more and more organelles were discovered. In 1833 Robert Brown concisely described nuclei of epidermal orchid cells. Henceforth living organisms could be divided into eukaryotes, which possess a nuclear membrane and prokaryotes with a lack thereof. The idea that cells emerge from other cells was first postulated by Schleiden and Schwamm in 1838-39 and later manifested by Virchows phrase 'omnis cellula a cellula' [1].

Chromosomes as carriers of information were recognized by van Beneden in 1883 and Sutton realized in 1903 that they were linked to Mendels 'laws' of inheritance, which date back to 1865 and have pretty much taken a back seat for nearly 40 years. The jump to chemistry was performed in 1944 by Avery, MacLeod and McCarty by realizing that DNA was the carrier of genetic information. The next milestone came 9 years later when Watson and Crick discovered the double helical structure of DNA [2, 3].

Of course cell biology and in particular the biology of the nucleus doesn't stop after 1953 and this little historical overview is far from complete, but the gap between chromosomes and the DNA double helical structure with regards to the scale is already apparent. That is exactly the realm of this thesis, trying to bridge the gap between the macroscopic (chromosomes) and the microscopic (DNA), the realm of chromatin and its higher order structures.

1.1 Scale, packing and the microscope problem

Let me give a short overview of the scales involved in the biology of the cell. As I already mentioned, an average animal cell has a diameter of about $10\mu m$, its nucleus about $5\mu m$. A human chromosome ranges from $1\mu m$ to $5\mu m$ in length, the chromatin fiber about $30nm$ and finally the diameter of the DNA

double helix is about $2nm$. If one were to hypothetically stretch out the entire human genome, one would have a string of about $1m$ length. Thus we have $1m$ of double stranded DNA in every nucleus of almost every cell in our body ($\sim 10^{14}$). Obviously there have to be various levels of compactification to explain a packing factor of 10.000 [4].

The necessity of computer simulation arises from the problems and limitations of microscopy in living cells on the scale below about $200nm$ (about half the excitation wavelength of the light microscope). Electron microscopy yields a resolution of about $0.1nm$ but is not possible on living cells. When investigating samples with an EM, these samples have to be fixated somehow. The fixation procedure is by definition highly invasive and it remains questionable whether structures observed by EM are actually unaltered from their native states [5]. Moreover, the capability of EM to label specific structures, e.g. by gold particles, is presently still limited.

However, light microscopy has come a long way and ingenious inventions have been made. The SMI microscopy for example uses so-called point spread function (PSF) engineering methods. It modifies the PSF of a microscope in such a way, that information of an object below the classical resolution limit will be gained. In the case of the SMI microscope this is accomplished by the fact that the illumination intensity is not homogeneous in the object area but is spatially modulated. Two laser beams propagating in opposite directions and interfering in axial direction are used to set up a standing wave field with intensity modulation along the optical axis. The principle of spatially modulated wavefield has been developed in 1993 by Bailey et al. In the SMI microscopy approach in C. Cremers group in Heidelberg, the object is moved in highly precise steps through the wave field. From this an increase in the axial size- and distance-resolution is gained [6, 7, 8, 9, 10, 11, 12].

Fluorescence-Resonance-Energy-Transfer (FRET) allows the determination of distances in biomolecules. It tracks the fluorescence of inserted markers by measuring their resonance energy transfer. Fluorescence in situ hybridization (FISH) allows the direct localization of DNA and RNA sequences on chromosomes, in cells and in tissue. This technology is based on the hybridization between target sequences of the single-strand DNA of chromosomes or cell nuclei with labelled complementary specimens. The signal is intensified by means of specific fluorochrome-labelled antibodies and visualized in the microscope. This technique allows, for example, the localization of genes and also the direct morphological detection of genetic defects causing hereditary diseases.

Using the concept of stimulated emission depletion (STED) Stefan Hell et al. were able to go beyond the classical diffraction barrier. Unlike in the light microscopes, in a STED microscope, the relevant focal fluorescence spot can, in principle, be reduced in size to the size of a molecule ($2 - 5nm$). This is due to the fact that the spot size is no longer subject to Abbe's formula, but to a

new law that differs from Abbe's original formula in a crucial factor:

$$\Delta d = \frac{\lambda}{2n \sin \alpha \sqrt{1 + I/I_{sat}}} \quad . \quad (1.1)$$

The key to this method is that one stops the fluorescence, which is in essence spontaneous emission, by stimulated emission, thus increasing the intensity ratio in 1.1. Stefan Hell et al. have proven this method theoretically as well as experimentally [13, 14, 15, 16].

1.2 Nuclear structure and Chromatin

As mentioned earlier, life is either eukaryotic meaning its cells possess a nuclear membrane and hence a nucleus or prokaryotic with a lack thereof. I will only focus on eukaryotic cells as found in most animals. The main structural component of the cell nucleus is chromatin. Chromatin is a complex of DNA and nucleosomes. The nucleosome is formed by the wrapping of DNA around the histone octamer [17]. The nucleosomes provide the first level of compaction and restrict transcription from unnecessarily accessing unwanted promoter regions. The word chromatin meaning colored material indicates that it received its name from the staining methods used to make it visible. Fig. (1.1) illustrates the different levels of compaction. The DNA nucleosome complex form the so called $11nm$ fiber, which is basically a string with beads. The $11nm$ fiber then forms a zig-zag pattern and compactifies into the $30nm$ fiber. It has been an ongoing dispute for the last 20 years if the $30nm$ fiber is ordered linearly in a solenoidal fashion or in a star shaped zig-zag structure. Current research clearly favors the zig-zag model [18]. The next level of compaction is still actively investigated. This thesis supports the rosette-model [19, 20]. The details will be explained in the chapters to come.

Chromatin is historically divided into two types: euchromatin and heterochromatin. Heterochromatin is densely packed and often found at the nuclear membrane, whereas euchromatin is relatively loose and can be found throughout the nucleus. Indeed it basically fills most of the nuclear volume. Heterochromatin itself is divided into two types: constitutive and facultative heterochromatin. Constitutive HC is never expressed, meaning it contains no active genes, whereas facultative HC is expressed in some cell lines but not in others. In female humans, for example, one copy of the X chromosome is almost entirely facultative HC and transcriptionally inactive, whereas the other copy is euchromatic and transcriptionally active [21].

1.3 Gene Expression and Silencing

There are many mechanisms involved in gene expression and silencing, such as the binding of activator/repressor proteins to specific DNA sequences at the

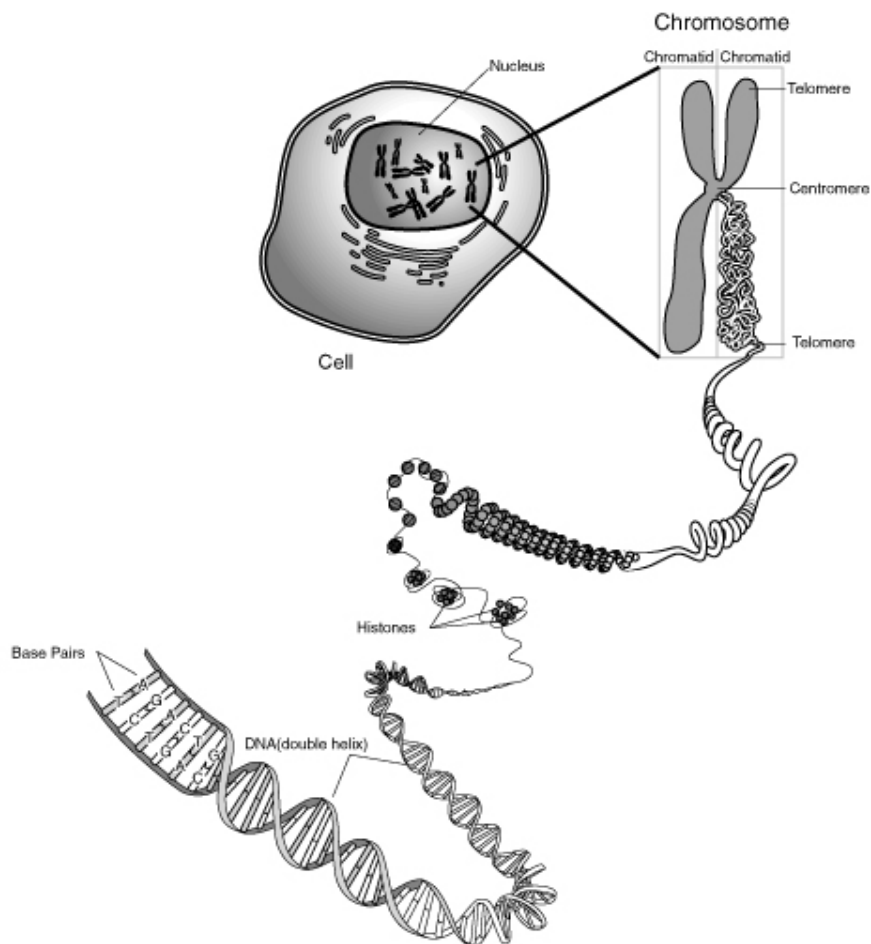


Figure 1.1: The levels of chromatin compaction. Image source: NIH Talking Glossary of Genetics

level of nucleosome [17]. DNA methylation is important for the inactivity of heterochromatin. The methyl group probably prevents the binding of transcription factors and even cause the wrong proteins to bind to the specifically methylated sites [22]. This is therefore directly linked to the gene density and transcriptional activity and will be studied thoroughly in chapters 5 and 6.

In *Drosophila*, however, there is no DNA methylase, so methylation cannot be responsible for the heterochromatinization. Here the polycomb group proteins (PcG) are important for gene regulation, especially in homeotic genes. Mutation in these genes causes the replacement of homologous body parts such as legs and antennae [23].

1.4 The Cell Cycle

Most of the cell's lifetime, the so called interphase, chromatin forms a single dense mass, the so called chromosome territories. Only during the cell division (mitosis) it condenses into chromosomes. A newly born cell rapidly grows (G1) and gets ready for protein synthesis. The so called S phase stands for synthetic, meaning that proteins can be synthesized. After the synthesis the cell prepares (G2) for mitosis. During mitosis (M) the entire genome is duplicated and the cell division is facilitated by the mitotic spindle. However, most adult cells do not cycle, but enter a resting state G0 without replicating their DNA. An illustration of the cell cycle is shown in Fig. (1.2). The possibility of apoptosis, the cellular death is not shown.

1.5 Replication and Transcription

During replication the DNA double strand is separated into two individual strands which are then copied into a new complementary partner. This copying process is the basis for inheritance. An enzyme called DNA polymerase is responsible for this process and many DNA polymerases copy different parts of a chromosome simultaneously. The entire process involves separating parental strands, copying small segments, glueing them together and verifying if everything was done correctly [24]. DNA synthesis is restricted to the S phase of the cell cycle. During S phase the DNA synthesis occurs semiconservatively, meaning that each daughter DNA molecule contains one original and one copied strand. Replication does not start at the end of the chromosome and works its way through, but starts at many sites known as origins, spaced apart about every 50kbp [25]. The replication of many parts occurs simultaneously in large immobile replication factories, where the necessary molecules are concentrated [26, 27]. Until recently it was believed that the DNA polymerase moved along the DNA and copied it as it goes along [28].

While replication copies genetic information, transcription decodes this information into a sequence of amino acid residues in a protein. The principles

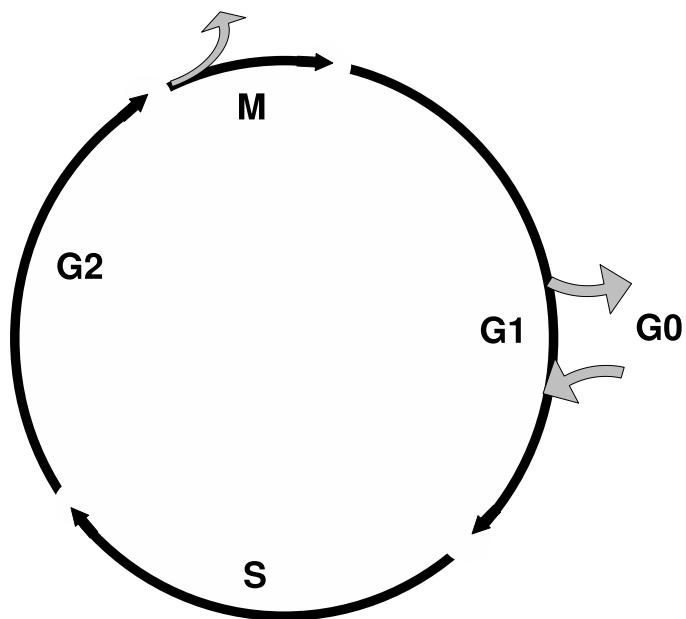


Figure 1.2: The cell cycle of a typical mammalian cell. A new cell can pass through G1, S and G2 before it divides at mitosis (M) or it can exit from the cycle by going into a resting state G0. The possibility of apoptotic death is not shown.

of transcription are in some aspects similar to replication. Transcription is carried out by RNA polymerase. Not the entire DNA is transcribed, only the parts that lie between short control sequences called promoters and termination signals [29]. As in replication there are two competing models: immobile RNA polymerases in large transcription factories or dynamic tracking of RNA polymerase along the DNA [30, 31, 28].

Chapter 2

Modeling

2.1 The Biological model

The genome content of a typical human chromosome is on the order of about 100Mbp (e.g. 245Mbp for Chromosome 1). To handle this amount of data on a computer, course-graining is mandatory. On a large scale one identifies a coiled state of a chromatin fiber as a 1Mbp bead. Experimental data yield a diameter of a 1 Mbp domain of about 300 to 800nm [32]. The aim of this analysis will be to see whether computer simulations of chromatin fibers in interphase yield the known size and assumed structure of such a coil.

On a more detailed level it is interesting to see the inner structure of such a coil. For a pure 30nm chromatin fiber one assumes there are about 40 segments of about 30kbp per bead. One instance of the model would be the ‘10 loop model’, where the segments form a rosette of 10 loops. Each loop consists of 120kbp, so that each segment has 30kbp. The 10 loop domains are interconnected by 120kbp linkers. Thus, every 4 segments there are believed to be attraction sites which couple the segments and thus lead to the formation of a non-random structure. The MLS model [19] and simulations thereof assume a rosette structure from considerations of the bead diameter and weight. It considers the attractive sites to be connected in the center of the rosette at the base segments of the loops.

My model is more general than the MLS-model. I will look into the formation of any possible higher order structure by starting out with a linear chain. This chain has repulsive and attractive segments. The attractive segments correspond to that part of the chromatin fiber which is affected by some condensing agent. I model the chain such that rosette formation is possible, but not a priori assumed and compulsory. Furthermore, I believe it to be too restrictive not to allow the breaking up of bonds between attraction sites. Therefore, I assume a Lennard-Jones potential. Since there is no reliable data for the Kuhn length of the 30nm fiber, I simulate the two extreme cases (150 and 300nm [33]) and see whether there are any significant differences in the results of the interesting

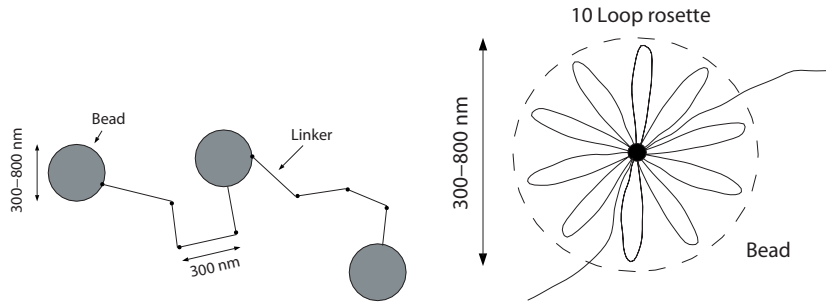


Figure 2.1: Left: Course-grained model of a chromosome. A typical bead is about $300 - 800\text{nm}$ in diameter, the linker segment length is around 300nm and consists of 30kbp . Chromosome 1 has approximately 245 such coils. Right: Detailed structure of a bead. The 10 loop model suggests a rosette like structure.

observables.

The assumption of attractive sites is biologically justified. For G/Q - R bands one has shown that HMG/SAR binding proteins act as mediators of attraction [34, 35]. For hetero/euchromatin the HP1 protein has been associated with chromatin linking [36, 37]. Furthermore, cohesin and condensin play a crucial role in chromosome compaction [38, 39]. The type of condensing agent is not my primary concern, though. My model also holds if the attraction is mediated by DNA-DNA interaction [40, 41]. Thus, I do not claim that the formation of higher order structures can only occur with a certain type of condensing agent, but rather I look at the general aspect of all agents, namely that they make a certain part of the chromatin fiber effectively attractive.

My model extracts the underlying idea in every case, namely that the chromatin fiber can be modeled as a multiblock copolymer. Whether you associate the different polymer blocks with GC rich and AT rich [42] or with other chromatin characteristics is not of primary importance for my model. A multiblock copolymer containing two alternately located types of blocks can form a single-chain string of loop clusters called micelles [43]. A micelle consists of a certain number of loops. The ends of the loops formed by blocks of one type are located close to each other.

2.2 Computational Model

In this section I would like to describe the concept and the realization of course-graining. An outline of the program used for the simulation can be found in chapter 9, as well as in the appendix. Course-graining focuses on the important aspects of the underlying problem and neglects other factors such as atomistic or molecular states. It is absolutely mandatory for all non-atomistic computer

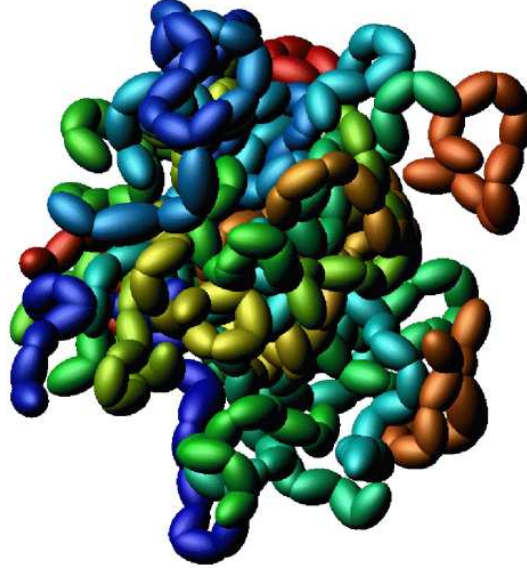


Figure 2.2: After the course graining complex and dense structures may be simulated.

simulations. Atomistic simulations above the level of relatively simple molecules is computationally impossible for all practical reasons. Therefore one lumps together a multitude of atoms in a single entity. This allows the simulation of complex and dense macroscopic structures (see Fig. (2.2)).

In this work the continuous backbone mass model [44] is used. The model interpolates between the united atom model and the bead-spring model. In contrast to these two models it uses non-spherical force fields for the non-bonded interaction. The main idea of this approach with a more general form of the force field is to generalise the united atom model in a way that larger atom groups are combined into one construction unit, but the possible anisotropy of these groups is still taken into account. The simplest anisotropic geometrical object one can think of is an ellipsoid of rotational symmetric form and thus it is considered as the interaction volume of the chemical sequences in my model (see Fig. (2.2)).

As one wants the force field to degenerate into a sphere with increasing distance, I use a confocal force field inside this interaction volume:

$$U_{\text{inter}} = U_{\text{abs}} \left(\frac{d_1^{(p)} + d_2^{(p)}}{2} - c \right), \quad (2.1)$$

where $d_1^{(p)}$ and $d_2^{(p)}$ denote the distance of the point \mathbf{p} to the focal points of the ellipsoid and U_{abs} is the absolute potential. For convenience I use only a

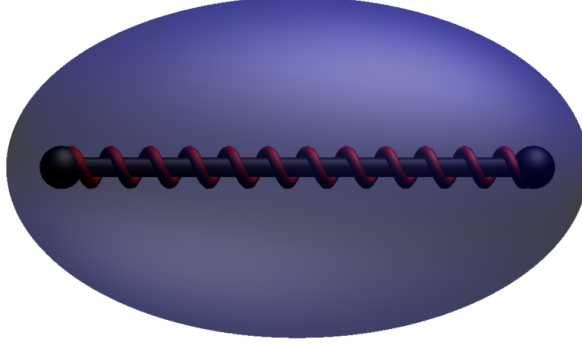


Figure 2.3: An ellipsoid with an intrinsic spring. The spring symbolizes the harmonic bond potential.

repulsive part

$$U_{\text{abs}}(r) \sim r^{-6} \quad . \quad (2.2)$$

The mass of the building units is distributed between the focal points of the ellipsoids in the hard core region of the confocal potential.

The main ingredient of the model is the mass matrix of the rod-chains. In order to construct it one, must first calculate the Lagrangian of a single rod $\mathcal{L}_i = T_i - V_i$ with the kinetic energy T_i and the potential energy V_i . The subindex i marks the position of the rods in the chain. This one-dimensional homogeneous rod i has the length l_i starting at \vec{a}_i and ending at \vec{b}_i . If one supposes that the rods all have the same mass m and that the velocity of the rod mass scales linearly with the position between the boundaries of the rod, the kinetic energy can be written as

$$\begin{aligned} T_i &= \frac{1}{2} \int_0^{l_i} \frac{m}{l_i} \left(\frac{(l_i - x)\dot{\vec{a}}_i + x\dot{\vec{b}}_i}{l_i} \right)^2 dx \\ &= \frac{1}{6} m (\dot{\vec{a}}_i^2 + \dot{\vec{a}}_i \dot{\vec{b}}_i + \dot{\vec{b}}_i^2). \end{aligned}$$

Adding the single terms of the rods building the chain one gets the Lagrangian \mathcal{L} of the whole rod chain. The equations of motion of the chain can be calculated from the Lagrange equations of the second kind. Since the equations of motion separate in each direction, one only has to solve three tridiagonal $(N + 1) \times$

$(N + 1)$ matrices per chain which consist of N rods per time step of the form

$$\mathbf{W}\ddot{\vec{x}} = \vec{F} \quad (2.3)$$

$$\frac{m}{6} \begin{pmatrix} 2 & 1 & 0 & 0 & \dots \\ 1 & 4 & 1 & 0 & \dots \\ 0 & 1 & 4 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} \ddot{x}_0 \\ \ddot{x}_1 \\ \ddot{x}_2 \\ \vdots \end{pmatrix} = \begin{pmatrix} F_{10} \\ F_{11} + F_{21} \\ F_{22} + F_{32} \\ \vdots \end{pmatrix} \quad (2.4)$$

with the force F_{ij} on the coordinate j of the flexible point i of the chain

$$F_{ij} = -\frac{\partial V_i}{\partial j} \quad (2.5)$$

and \ddot{x}_i denote the accelerations of the flexible points of the chain. The flexible points are the linking points of the ellipsoids and the end points of the rod chain. The sub-indices mark the positions in the chain: 0 and $N + 1$ are the end-points of the chain and the numbers between them denote the linking points of rods in the chain.

The bonded interactions between neighboring units are given by harmonic length and angle potentials:

$$U_{bond} = \frac{1}{2}k(r - r_0)^2 \quad (2.6)$$

$$U_{angle} = \frac{1}{2}k_\theta(\cos \theta - \cos \theta_0)^2 \quad (2.7)$$

with the bond lengths r and the bending angles θ . Here r_0 and θ_0 denote the mean values.

The ellipsoidal model therefore maps polymers onto focally joint ellipsoids with a confocal hard-core potential, i.e. the polymers cannot penetrate each other. The half-axis can be altered individually. The ellipsoids have an intrinsic harmonic bond-potential as well as an angle and torsion potential. This allows a chemically and biologically realistic mapping of different substances while at the same time maintaining computational feasibility.

2.2.1 Integration Scheme

In order to integrate the equations of motion in molecular dynamics (MD), the simulation program uses the velocity Verlet algorithm. A Taylor expansion of the position at discrete time steps yields

$$r(t + \Delta t) = r(t) + v(t)\Delta t + \frac{f(t)}{2m}\Delta t^2 + \frac{\Delta t^3}{3!}\frac{d^3r}{dt^3} + \mathcal{O}(\Delta t^4) \quad .$$

Doing the same expansion for $r(t - \Delta t)$ and summing these two equations yields

$$r(t + \Delta t) = 2r(t) - r(t - \Delta t) + \frac{f(t)}{m}\Delta t^2 + \mathcal{O}(\Delta t^4) \quad ,$$

therefore only giving an error of the order Δt^4 , Δt being the time step used in the simulation.

2.2.2 Dissipative Particle Dynamics

For the part of this thesis dealing with diffusion, the dissipative particle dynamics (DPD) method was used. It is very similar to MD, only that the force now contains a dissipative and a random component in addition to the conservative component

$$\vec{F}_i = \sum_{j \neq i} \left[\vec{f}_C(\vec{r}_{ij}) + \vec{f}_D(\vec{r}_{ij}, \vec{v}_{ij}) + \vec{f}_R(\vec{r}_{ij}) \right] \quad .$$

The dissipative force corresponds to a frictional force that depends on the positions and relative velocities of the particles:

$$\vec{f}_D(\vec{r}_{ij}, \vec{v}_{ij}) = -\gamma \omega_D(r_{ij}) (\vec{v}_{ij} \cdot \hat{r}_{ij}) \hat{r}_{ij} \quad ,$$

where \hat{r}_{ij} is the unit vector in direction of \vec{r}_{ij} and $\omega_D(r_{ij})$ describes the dependence of the friction coefficient γ with distance. The random force is of the form:

$$\vec{f}_R(\vec{r}_{ij}) = \sigma \omega_R(r_{ij}) \xi_{ij} \hat{r}_{ij} \quad ,$$

where σ determines the magnitude of the random force and $\omega_R(r_{ij})$ its distance dependence. ξ_{ij} is a random variable with a Gaussian distribution. Note that the following relations must hold, in order to guarantee a proper Boltzmann weight:

$$\begin{aligned} \omega_D(r_{ij}) &= [\omega_R(r_{ij})]^2 \quad \text{and} \\ \sigma^2 &= 2k_B T \gamma \quad . \end{aligned}$$

The detailed justification of this method and its superiority over regular Brownian dynamics, especially for block copolymers can be found in [45].

2.3 Physical Model

Micellar structures have been thoroughly studied for diblock copolymers and ionomers [46, 47]. Large multiblock copolymers form single-chain micelles, and small diblock copolymers form multichain micelles. The formation of loops and their organization into micelles are basically an entropically unfavorable process, because the number of possible polymer conformations decreases, but it occurs nonetheless in multiblock copolymers because of the energetically favorable processes of repulsion between unlike monomer units and/or attraction between like monomer units [48, 49]. In an aqueous solution, for example, this means that the hydrophobic parts of the copolymer concentrate in the center of the micelle and the hydrophilic parts form the loops.

With the abstraction to multiblock copolymers this leads to the potentials and parameters I use in my simulation:

- Segment diameter: $30nm$

- Segment length: 2 different simulations
 1. 30kbp = 300nm each
 2. 15kbp = 150nm each
- The harmonic bond potential is taken to be

$$U_{bond}(l) = \frac{kT}{2\delta^2}(l - l_0)^2 \quad (2.8)$$

with $\delta = 0.1$ and $l_0 = 300nm$ at 310.15K.

- The angular and torsional potentials are taken to be 0. On this scale the chain is flexible.
- Repulsive segments potential

$$U_{rep}(r) = \epsilon \left(\frac{\sigma}{r - r_{segment}} \right)^6 \quad (2.9)$$

with $\epsilon = 0.14k_bT$ at body temperature, $\sigma = 15nm$ and $r_{segment} = 15nm$ being the fiber radius.

- Cutoff for the repulsive potential is $r_c = 8nm$ (after the 15nm fiber radius).
- Attraction segments potential

$$U_{attr}(r) = 4\epsilon \left[\left(\frac{\sigma}{r - r_{segment}} \right)^{12} - \left(\frac{\sigma}{r - r_{segment}} \right)^6 \right] \quad (2.10)$$

with $\epsilon = 7k_bT$ at body temperature and $\sigma = 30nm$.

- Cutoff for the Lennard-Jones potential is $r_c = 80nm$ (after the 30nm fiber diameter).

The spring constant $\delta = 0.1$ was chosen, such that the 300nm segment was reasonably stiff and at the same time soft enough to ensure a reasonable integration time step. For the Lennard-Jones potential $\epsilon = 7k_bT$ was chosen because this proved to be the smallest potential depth for which the segments remain attractive at body temperature. The analogous reasoning applies to the $\epsilon = 0.14k_bT$ for the repulsive potential.

Chapter 3

Polymer Theory

In this chapter I would like to address some basic concepts of polymer physics and diffusion. This will be the theoretical foundation of the following chapters on simulation. Starting out with elementary statistical physics of chain molecules I will continue with real chains and conclude with the theory of diffusion.

3.1 General Chains

3.1.1 End-to-End Distance

A chain with n vertices and $n - 1$ segments of the length \vec{l}_i has the vector end-to-end distance of

$$\vec{r} = \sum_{i=1}^n \vec{l}_i \quad (3.1)$$

as well as the square distance

$$r^2 = \vec{r} \cdot \vec{r} = \sum_{i,j=1}^n \vec{l}_i \cdot \vec{l}_j = \sum_{i=1}^n l_i^2 + 2 \sum_{0 < i < j \leq n} \vec{l}_i \cdot \vec{l}_j \quad (3.2)$$

The distance between the two vertices i and j (with $i < j$) is thus

$$r_{ij}^2 = \sum_{i'=i+1}^j l_{i'}^2 + 2 \sum_{i < i' < j' \leq j} \vec{l}_{i'} \cdot \vec{l}_{j'} \quad (3.3)$$

and for the ensemble average

$$\langle r^2 \rangle = \sum_{i=1}^n \langle l_i^2 \rangle + 2 \sum_{i < j} \langle \vec{l}_i \cdot \vec{l}_j \rangle \quad (3.4)$$

Assuming all segments to be of length l , one obtains

$$\langle r^2 \rangle = nl^2 + 2 \sum_{i < j} \langle \vec{l}_i \cdot \vec{l}_j \rangle \quad (3.5)$$

3.1.2 Radius of Gyration

The radius of gyration s^2 is defined by the coordinates \vec{s}_i in the chains center-of-mass system.

$$s^2 = \frac{\sum_{i=0}^n m_i \vec{s}_i^2}{\sum_{i=0}^n m_i} = \frac{m \sum_{i=0}^n \vec{s}_i^2}{m(n+1)} = \frac{1}{n+1} \sum_{i=0}^n \vec{s}_i^2 \quad (3.6)$$

The radius of gyration is thus connected to the end-to-end distance by $\vec{r}_{ij} = \vec{s}_j - \vec{s}_i$. Using (3.6) and (3.3) one obtains

$$s^2 = \frac{1}{(n+1)^2} \sum_{0 < i < j \leq n} \vec{r}_{ij}^2 \quad (3.7)$$

and for the ensemble average

$$\langle s^2 \rangle = \frac{1}{(n+1)^2} \sum_{0 < i < j \leq n} \langle \vec{r}_{ij}^2 \rangle \quad (3.8)$$

3.2 Freely Joint Chains

In the case of freely joint chains, all bond lengths are the same and all bond angles are distributed with equal probability. There is no preferred orientation and every configuration is eventually achieved. For the scalar product of the individual chain segments one thus gets

$$\langle \vec{l}_i \cdot \vec{l}_j \rangle = l^2 \langle \cos \theta_{ij} \rangle = 0 \quad i \neq j \quad (3.9)$$

This simplifies the two quantities to

$$\langle r^2 \rangle = nl^2 \quad (3.10)$$

$$\vec{r}_{ij}^2 = |j - i| \cdot l^2 \quad (3.11)$$

and (3.8) yields

$$\langle s^2 \rangle = \frac{l^2}{(n+1)^2} \sum_{0 < i < j \leq n} (j - i) \quad (3.12)$$

By complete induction one proves that the ratio of radius of gyration to end-to-end distance is

$$\frac{\langle r^2 \rangle}{\langle s^2 \rangle} = \frac{6(n+1)}{n+2} \quad (3.13)$$

for $n \rightarrow \infty$ yielding

$$\left(\frac{\langle r^2 \rangle}{\langle s^2 \rangle} \right)_{\infty} = 6 \quad (3.14)$$

3.2.1 Distribution of the End-to-End Distance

In a freely joint chain, the vector end-to-end distance \vec{R} corresponds to the sum of N independent randomly distributed orientations \vec{u}_i . By the central limit theorem such quantities exhibit a Gaussian distribution.

$$P_N(\vec{R}) = \left(\frac{3}{2\pi Nl^2} \right)^{3/2} \exp \left(-\frac{3\vec{R}^2}{2Nl^2} \right) \quad (3.15)$$

Moments

Using the distribution function one can calculate the moments.

$$\langle (\vec{R}^2)^n \rangle = \int (\vec{R}^2)^n P_N(\vec{R}) d^3R = (Nl^2)^n \left[\frac{(2n+1)!!}{3^n} \right] \quad (3.16)$$

The double factorial $!!$ is the factorial over next to nearest neighbors $n \cdot (n-2) \cdot (n-4) \cdot \dots$. The identity $(2n+1)!! = (2n+1)! / (2^n n!)$ relates the double factorial to the regular factorial. The averaged odd powers of \vec{R} are 0, since the Gaussian distribution is even. Therefore the second moment $\langle \vec{R}^2 \rangle = \int \vec{R}^2 P_N(\vec{R}) d^3R = Nl^2$ is identical to (3.10). Notably one can use (3.16) to calculate the relative magnitude of the fluctuations \vec{R}^2 .

$$\frac{\langle (\vec{R}^2 - \langle \vec{R}^2 \rangle)^2 \rangle}{\langle \vec{R}^2 \rangle^2} = \frac{\langle \vec{R}^4 \rangle - \langle \vec{R}^2 \rangle^2}{\langle \vec{R}^2 \rangle^2} = \frac{2}{3} \quad (3.17)$$

This means that the fluctuations of R^2 are of the same order of magnitude as its mean. A Gaussian chain is thus a strongly fluctuating system.

3.3 Freely Rotating Chain

We now look at a chain with fixed bond angles and a free rotation around this angle. Let θ_{ij} be the bond angle between the segments i and j . We still have

$$\langle r^2 \rangle = \sum_{i=1}^n \langle l_i^2 \rangle + 2 \sum_{i < j} \langle \vec{l}_i \cdot \vec{l}_j \rangle = nl^2 + 2 \sum_{i < j} l^2 \langle \cos \theta_{ij} \rangle \quad (3.18)$$

and for $\langle \cos \theta_{ij} \rangle$ we have

$$\begin{aligned} \langle \cos \theta_{i,i+1} \rangle &= \cos \gamma \\ \langle \cos \theta_{i,i+2} \rangle &= \cos \gamma \cos \gamma = \cos^2 \gamma \end{aligned}$$

For k segments we thus obtain

$$\langle \cos \theta_{i,i+k} \rangle = \cos^k \gamma \quad (3.19)$$

Therefore yielding for (3.18)

$$\begin{aligned}
\langle r^2 \rangle &= nl^2 + 2l^2 \sum_{i=0}^N \sum_{k=0}^N -i \langle \cos \theta_{i,i+k} \rangle \\
&= nl^2 + 2l^2 \sum_{i=0}^N -1 \sum_{k=0}^N -i \cos^k \gamma \\
&= nl^2 + 2l^2 \sum_{i=0}^N -1 \frac{\cos \gamma}{1 - \cos \gamma} \\
&= nl^2 + 2Nl^2 \frac{\cos \gamma}{1 - \cos \gamma}
\end{aligned}$$

Thus the end-to-end distance turns out to be

$$\langle r^2 \rangle = nl^2 \frac{1 + \cos \gamma}{1 - \cos \gamma} \quad (3.20)$$

This is modulo a constant factor identical to the freely joint case. For the radius of gyration and the corresponding ratio a rather lengthy calculation [50] shows that for long chains the Gaussian case is approached. One can show that correlations and orientation decay exponentially. Thus for long chains, the freely joint chain is a justified model.

3.4 Characteristic Quantities of Polymers

3.4.1 Persistence Length

Rewriting (3.19) one obtains

$$\begin{aligned}
\langle \cos(\theta_{i,i+k}) \rangle &= \cos^k \gamma \\
&= \exp(k \ln(\cos \gamma)) \\
&= \exp\left(-\frac{kl}{l/|\ln(\cos \gamma)|}\right) \\
&= \exp\left(-\frac{s}{\tilde{l}}\right)
\end{aligned}$$

with $\tilde{l} = l/|\ln(\cos \gamma)|$ and $s = kl$ as the contour length between two monomers along the chain. As a result one writes:

$$\langle \cos(\theta_{\vec{u}(0), \vec{u}(s)}) \rangle \propto \exp\left(-\frac{s}{\tilde{l}}\right) \quad (3.21)$$

Although the result for the persistence length \tilde{l} was derived for a freely rotating chain, it is valid for all chains due to the exponential decay of the correlations and orientations. For $s \ll \tilde{l}$ the chain is nearly straight, for $s \gg \tilde{l}$ all memory of the orientation of the chain is lost.

3.4.2 Kuhn Segment

For an ideal chain we have $\langle \vec{R}^2 \rangle \sim nl^2$ (see 3.10). The Kuhn Segment is defined as

$$\hat{l} = \frac{\vec{R}^2}{nl^2} \quad (3.22)$$

The advantage of the Kuhn segment l is that is experimentally easier accessible, while the persistence length \tilde{l} has a direct microscopic interpretation. It is always $l \sim \tilde{l}$. For the freely joint chain, one can easily calculate the proportionality.

$$\frac{\hat{l}}{\tilde{l}} = |\ln(\cos \gamma)| \frac{1 + \cos \gamma}{1 - \cos \gamma} \quad (3.23)$$

3.5 Excluded Volume

Previously only Gaussian chains were treated. These chains may penetrate each other freely, therefore possess no excluded volume. This is obviously not the case for real chains. In reality the individual chain segments have an excluded volume, given by their potential (typically repulsive or Lennard-Jones-type) and thus repel each other during collisions. Fortunately the stated equations can be modified to take the excluded volume into account.

3.5.1 Flory Theory

Take a chain with an unknown radius R and an internal monomer concentration

$$c_{int} \cong \frac{N}{R^d} \quad \text{in } d - \text{Dimensionen} \quad (3.24)$$

There is a repulsive Energy due to the monomer-monomer interaction. Let c be the local monomer concentration, then the repulsive energy per unit volume is proportional to the number of pairs, thus c^2 .

$$F_{rep} = \frac{T}{2} v(T) c^2 \quad (3.25)$$

where $v(T)$ has the dimension of a d -dimensional volume. The key to the Flory theory is the *mean field* approach [51, 52]: all correlations between monomers are neglected.

$$\langle c^2 \rangle \longrightarrow \langle c \rangle^2 \sim c_{int}^2 \quad (3.26)$$

After the integration over the entire volume R^d the entire repulsive energy turns out to be

$$F_{rep}^{tot} \cong T v(T) c_{int}^2 R^d = T v \frac{N^2}{R^d} \quad (3.27)$$

Therefore large R are preferred, so the chain swells. However, if R gets too large, it is entropically unfavorable.

Now one takes the elastic energy into account

$$F_{el} \cong T \frac{R^2}{Nl^2}$$

and obtains for the free energy

$$\frac{F}{T} \cong v \frac{N^2}{R^d} \frac{R^2}{Nl^2} \quad (3.28)$$

The free energy is minimal for the radius $R = R_F$. Considering all numerical factors, one gets the following relation

$$R_F^{d+2} \cong v a^2 N^3 \quad (3.29)$$

i.e.

$$R_F \sim N^\nu \quad \text{mit} \quad \nu = \frac{3}{d+2} \quad (d \leq 4) \quad (3.30)$$

Interestingly for $d = 4$ dimensions this yields the ideal Flory exponent $\nu = 1/2$. For dimensions larger than 4 ν is always equal to $1/2$. In 3 dimensions the Flory exponent is $\nu = 3/5$.

3.5.2 Corrections to the Flory Theory

The Flory theory is amazingly precise although it makes strong simplifications. Indeed, two effects nearly compensate each other:

- the repulsive energy is overestimated
- the elastic energy is also overestimated

One can use an analogy to magnetic systems in order to obtain a more precise value for ν [53]. From the theory of second order phase transitions one gets the corresponding formula

$$\nu = 1/2 + \varepsilon(n+2)/[4(n+8)] + \varepsilon^2(n+2)(n^2+23n+60)/[8(n+8)^3] + \dots \quad (3.31)$$

with $\varepsilon = 4 - d = 1$ and $n = 0$ this yields

$$\nu = 1/2 + 1/16 + 15/512 + \dots \approx 0.592 \quad (3.32)$$

3.5.3 Real End-to-End Distances

The end-to-end distance for the free chain has already been calculated previously. The distribution for a chain with excluded volume looks qualitatively different. At short distances a new critical exponent γ appears, which describes the end effects and is independent of ν . At long distances the critical exponent ν governs the behavior.

The size R of the polymer determines its intrinsic scale. Generally the distribution is of the type

$$P_N(\vec{R}) = \frac{1}{R^d} \phi\left(\frac{\vec{R}}{R}\right) \quad (3.33)$$

where R^{-d} is a normalization resulting from $\int P_N(\vec{R}) d^d R = 1$. This is indeed also true for a free chain. (3.15) A stretched real chain then behaves as

$$\phi(x) \sim \exp(-|x|^\delta) \quad \text{mit} \quad \delta = \frac{1}{1-\nu} \quad (3.34)$$

(in three Dimensions $\delta = 5/2$). The distribution thus approaches zero faster for large x than the ideal chain ($\delta = 2$ (3.15)).

When the ends come very close, $|x| \ll 1$. Obviously $P_N(0) = 0$ because the ends cannot touch due to the excluded volume. For small $|x|$, $P_N(x)$ must be small as well because the neighbors of the end points also possess an excluded volume and therefore also decrease the probability of a collision. At the end points the screening of the neighbors acts differently, because there is only one neighbor. Therefore a new critical exponent g is introduced which takes these end effects into account. Traditionally one uses the exponent γ from the theory of self-avoiding random walks, which is connected to g by $\gamma = \nu g + 1$. For the self-avoiding walk one knows that $\gamma \approx 7/6$ in 3 dimensions. It thus follows that $g \approx 5/18 \approx 0.28$. For small probabilities a the probability is

$$P_N(|\vec{R}| \sim a) = R^{-3} \phi(a/R) \sim a^{-3} N^{-3\nu-\gamma-1} \sim N^{-1.97} \quad (3.35)$$

It therefore approaches zero almost like N^{-2} as compared to the ideal chain which goes like $N^{-3/2}$.

3.6 Finite Size Scaling

Most of the analytical results deal with infinitely long chains. Computer simulations cannot calculate infinitely long chains, therefore one must deal with finite chains and consequently the size effects that go along with them. In order to be able to compare the results from computer simulation to theory one has to scale the chain length. The original theory dealt with phase transitions, due to its general nature, however, it can also be applied in this case.

The most simple illustration of finite size scaling (FSS) is done using the percolation problem [54]. Take an infinite lattice whose vertices are occupied with a probability p and remain empty with $1 - p$. Adjacent occupied sites are called a cluster. There exists a critical concentration p_c , such that for $p < p_c$ only finite clusters exist, for $p \geq p_c$, however, one infinite cluster exists. For finite lattices there is a corresponding cluster that spans from one end of the lattice to the other. The quantities of interest are among others the

percolation probability $P_\infty(p)$ which states the probability of a site being part of the percolating cluster, or the percolationsusceptibility

$$\chi = \sum_{l=1}^{\infty} l^2 n_l(p)/p \quad , \quad (3.36)$$

where the largest cluster isn't included in the sum. These quantities possess a singularity in the infinite case. In finite systems, there is also a finite height of the maximum. In order to compare results from different size lattices, one scales with the exponent ν i.e. $1/\nu$. The critical exponent ν operates on the linear dimension L of the lattice. In order to scale other quantities, one must take into account the appropriate powers of ν .

3.7 Diffusion

The terms diffusion and Brownian motion are closely linked. In 1828 Robert Brown discovered the random motion of tiny pollen under a microscope. This is of course due to the constant bombardment of the pollen by very many molecules that make up the stochastic force. The equation of motion for such a particle in a fluid is

$$M \frac{d\vec{v}}{dt} = \mathcal{F}(t) \quad , \quad (3.37)$$

where M is the particle mass, \vec{v} the particle velocity and $\mathcal{F}(t)$ the force acting upon the particle due to the impacts from the fluid molecules. Langevin suggested that $\mathcal{F}(t)$ can be split up in two parts: a viscous drag $-\vec{v}/B$ and a rapidly fluctuating part $\vec{F}(t)$ which averages to zero over long intervals. This leads to

$$M \frac{d\vec{v}}{dt} = -\frac{\vec{v}}{B} + \vec{F}(t) \quad , \quad (3.38)$$

leading to the ensemble average

$$M \frac{d}{dt} \langle \vec{v} \rangle = -\frac{1}{B} \langle \vec{v} \rangle \quad , \quad (3.39)$$

whence

$$\langle \vec{v}(t) \rangle = \vec{v}(0) \exp(-t/\tau) \quad (\tau = MB) \quad . \quad (3.40)$$

Then (3.38) can be rewritten as

$$\frac{d^2}{dt^2} \langle r^2 \rangle + \frac{1}{\tau} \frac{d}{dt} \langle r^2 \rangle = 2 \langle v^2 \rangle \quad . \quad (3.41)$$

If the particle is in equilibrium, then one can use the equipartition theorem and obtain after integration

$$\langle r^2 \rangle = \frac{6kT\tau^2}{M} \left[\frac{t}{\tau} - \left(1 - e^{-t/\tau} \right) \right] \quad . \quad (3.42)$$

For $t \gg \tau$ one obtains

$$\langle r^2 \rangle \simeq \frac{6kT\tau}{M}t = 6BkTt \quad (3.43)$$

yielding

$$D = BkT \quad . \quad (3.44)$$

which is known as the Einstein relation.

Chapter 4

The 1Mbp Domain

4.1 Introduction

The structure formation of chromatin is studied on many different scales [55]. Interesting properties can already be obtained from models of the formation of the $30nm$ fiber [56, 57, 58, 59, 60, 61]. Chromatin structures beyond the level of the linear array of nucleosomes play an essential role in gene regulation, repair processes and pathogenic rearrangements in eukaryotes. Changes in functional activity are assumed to be tightly coupled to changes in the chromatin structure. Thus, a full understanding of genome function is not possible without detailed investigations of the functional chromatin structure and its control, requiring appropriate tools for quantitative analysis.

The exact details of the 3D folding of the chromatin fiber of a chromosome are still controversial. Experiments are highly difficult due to the following limitations. The aggregation is promoted by the high cellular concentration and charge of those genomes. Furthermore, many structures are smaller than the resolution of the light microscope (roughly $200nm$) and therefore can only be seen by electron microscopy, which in turn causes problems associated with preserving structure and recognition of the 3D folding of specific chromatin structures.

It is generally agreed upon that in eukaryotes the double helix is coiled locally around nucleosomes and globally into distinct nuclear territories [62]. The levels in between are still under discussion [21]. There are several models for the different stages of compactification: (i) structures, in which strings are coiled into solenoids (of roughly $30nm$ diameter) or zig-zag tubes [56], which in turn again form higher-order structures [63]; (ii) loops of about 50-150kbp which are attached to the peripheral lamina or other internal structures, such as skeletons / scaffolds [38] or factories [64]; and (iii) combinations of the above for example, of helical coils and radial loops [65] or helical coils and random folding [66].

To investigate the folding and accessibility of virtual active/inactive chro-

matin domains within the nuclear volume, my model assumes attractive sites at some locations along the chain. The results of my simulation favor the ‘Multi Loop Subcompartment (MLS)’ model developed in the group of J. Langowski [19, 20] for the overall structure of chromosome territories.

According to the MLS model, the experimentally observed foci structure of chromosomes (for an overview see [32, 62]) is described by rosettes of several 100kbp loops assuming a $30nm$ chromatin fiber. Adjacent rosettes are connected by chromatin linker segments with the same DNA content as one loop. Approaches based on the isochore model also predict the formation of rosettes of about 1Mbp [42].

It has not yet been understood how these rosettes form dynamically. I have developed a model that not only shows that the higher order structures formed are indeed rosettes, it also explains the formation process and predicts quantitative results for certain quantities of interest such as formation time and rosette diameter. Finally the model predictions will be compared with experimental light optical data.

4.2 Simulation

The first point of interest is the structure formation of a 1Mbp domain in interphase. The final structure turns out to always be a rosette. I take 2 non-reactive linkers at the end and an attraction site after n successive segments. The number n of segments is the varying parameter. The diameter of the rosettes will be analyzed as a function of this parameter. One then gets a starting configuration as shown in Fig. (4.1). The attraction agents are marked as spheres. The attractive sites are of the same length as the repulsive ones. The intermediate and final structure can be seen in the same figure.

The following patterns have been simulated

1. $300nm$ segment
 - attractive site every segment
 - attractive sites every 3 segments
 - attractive sites every 5 segments
 - attractive sites every 7 segments
 - attractive sites every 9 segments
2. $150nm$ segment
 - attractive site every segment
 - attractive sites every 6 segments
 - attractive sites every 10 segments
 - attractive sites every 14 segments
 - attractive sites every 18 segments

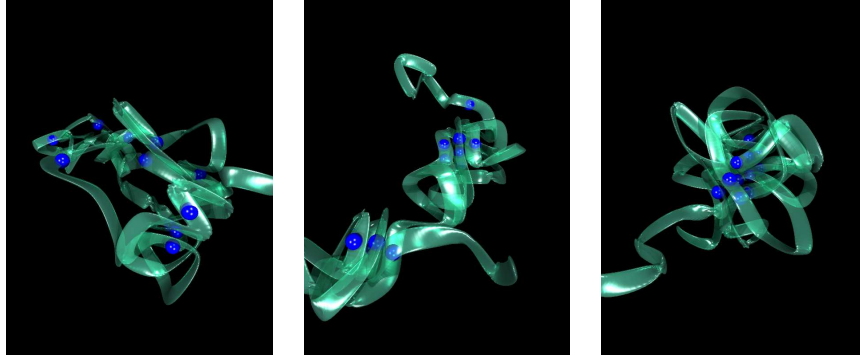


Figure 4.1: Left: The starting configuration of a 60 segment chromatin fiber. The spheres represent the condensing agents. Center: An intermediary configuration. This state of mainly two clusters of approximately equal size turns out to be a metastable state. Right: In the final state all attractive segments are concentrated in the center. A rosette has formed.

4.3 Results

In Fig. (4.2) one sees the average formation time of a rosette. Plotted are the minimum, average and maximum distance of the attractive Lennard-Jones segments. The minimum distance drops almost instantly, implying that two attractive segments immediately find each other. After an initial rise, which is due to the random and hence mostly unphysical starting configuration, the maximum distance decays over a time of about 10,000 MD-Steps. After this amount of time the distance drops no more. Hence all attractive segments have found each other. Therefore the fully equilibrated structure is shown to be a rosette. With the given parameters of ϵ and σ , one observes a formation time of about 48 ms. Note that this is the formation time based on the simulation of a single rosette. It remains to be seen how this time changes when a larger region is analyzed.

A crucial question is obviously how many attractive sites are required to form rosettes of the size observed under a microscope. In order to analyse this I have done simulations for regular patterns of attractive/repulsive segments. I am aware of the fact that a regular pattern is a severe restriction but it allows drawing conclusions with less statistics than for random patterns. Furthermore, it should provide a first insight whether rosettes of the required size are possible in the first place.

In Fig. (4.3) I have plotted the radii of gyration for different regular patterns and different Kuhn lengths. The circles represent the 300nm segment and the squares the 150nm segment. This is the accepted range for the Kuhn length of the 30nm fiber. Thus, by taking the two extreme cases I can study the effect of the stiffness of the chain.

As expected one sees a clear increase in the radius of gyration with larger

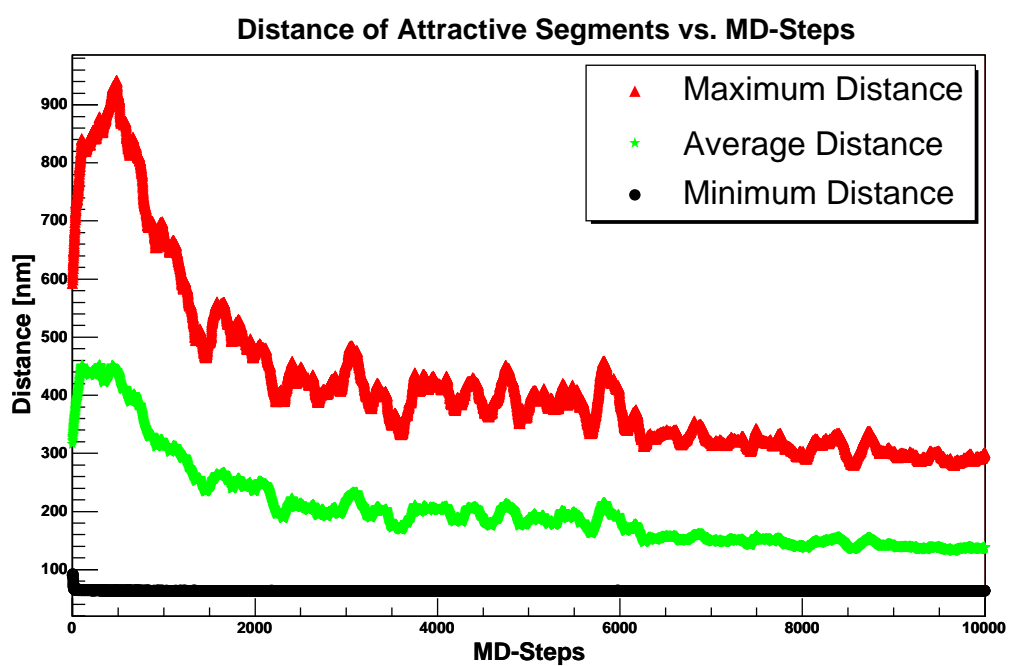


Figure 4.2: Shown is the average formation time of a rosette. From the minimal distance one can conclude that two segments ‘snap’ very fast. After about 10,000 MD Steps a complete rosette is formed.

$L_{segment}$	$N_{attractive}$	$N_{intermediary}$	Loop Size
300	7 to 19	7 to 3	90 to 210kbp
150	7 to 19	14 to 6	90 to 210kbp

Table 4.1: Loop Sizes for different Kuhn lengths

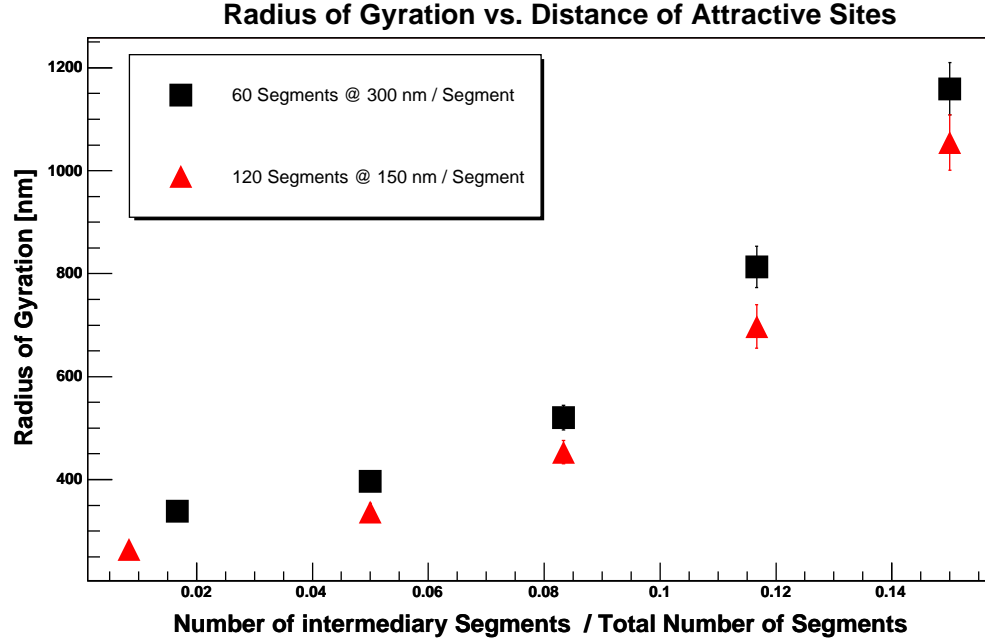


Figure 4.3: The diameter of the rosettes depends on the number of attractive sites. The different Kuhn lengths are only of minor importance.

loop sizes. For example, for the $300nm$ Kuhn length with a loop size of 7 segments (with a total number of 60 segments per rosette one gets a ratio of about 0.12) one obtains a radius of gyration of about $800nm$. One will see later that the radius of gyration corresponds very well to the actual size. So for a given experimental size of 300 to $800nm$ [67] one needs 7 to 19 attractive segments, i.e. every 3^{rd} to every 8^{th} segment, for the $300nm$ Kuhn segment. Therefore one has loop sizes of 90 to 210kbp (i.e. 3 to 7 segments), with the optimal value around 120kbp (4 segments) per loop. This result is well supported by the literature [64, 68, 69, 70]. The different Kuhn lengths produce only a slight difference in the size. The main parameter is clearly the number of attractive segments. Table 4.1 illustrates the optimal values for rosette sizes of 300 to $800nm$. $L_{segment}$ indicates the segment length, $N_{attractive}$ the total number of attractive segments per rosette and $N_{intermediary}$ the number of repulsive segments between the attractive ones.

Plotting the two-dimensional average monomer concentration yields Fig.

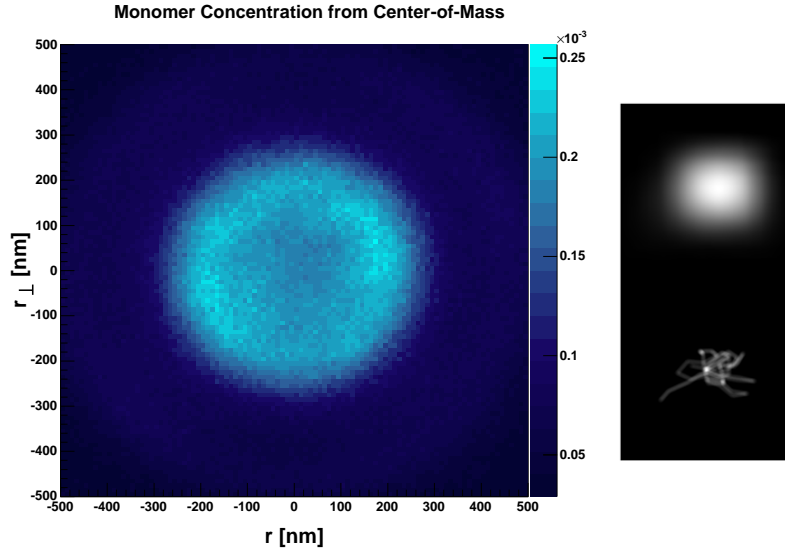


Figure 4.4: Left: Projection of the average monomer concentration of a rosette with a diameter of about $600nm$. Right: Projections of virtual microscopy image data stacks of one simulated rosette (before convolution with a measured confocal point spread function (bottom) and after convolution (top)).

(4.4). The average is taken from 10.000 uncorrelated configurations. The data in the figure corresponds to the $300nm$ segment with an attractive site every 5 segments. One sees that the 1Mbp domain has a diameter of about $600nm$. It is clear that the center cannot be the most likely position for the monomers since one has excluded volume effects. Furthermore, the attractive sites are all within each other's potential minimum so that there is a smeared out center where the attractive sites are. Moving farther out to the periphery one observes a higher monomer concentration. This is the region dominated by the loops.

To compare the modeled configurations with the outcome of optical light microscopy, in Fig. (4.4 top) projections of virtual microscopy image data stacks are shown. This approach consisted of a digitization of the $30nm$ thick segments (about 10 points per 30kbp sized segment were digitized) by a grid of $4.9 \times 4.9 \times 10.5nm$ voxel spacing and a convolution of the digitized stacks with a measured confocal Point Spread Function (PSF) (with a full width at half maximum (FWHM): $FWHM_x=279nm$, $FWHM_y=254nm$, $FWHM_z=642nm$).

In order to determine the sizes of gene-like structures, David Baddeley of C. Cremers group has convoluted my simulation data of 180kbp segments with a Spatially Modulated Illumination Microscope PSF (point spread function) and then performed a line scan in axial direction. By determining the modulation contrast one can then calculate the axial extension of the 180kbp segment. Fig. (4.5) shows the 180kbp segment as well as the PSF fit and Fig. (4.6) shows the virtual SMI image and the size distribution. Therefore the typical size of a

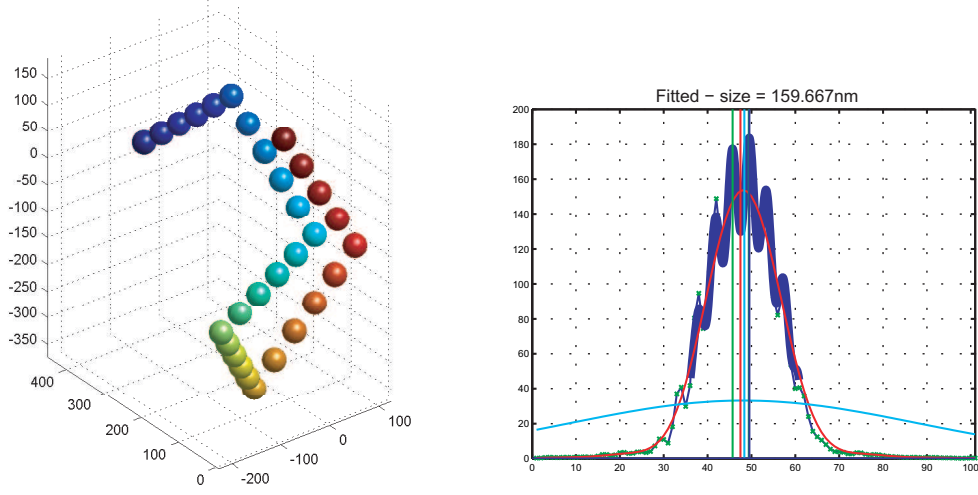


Figure 4.5: Size determination of a 180kbp gene-like segment. The left shows the simulation data and the right side the SMI-PSF fit.

$N_{attractive}$	a	b
5	177.74 ± 0.86	0.1535 ± 0.0014
7	148.05 ± 0.76	0.2075 ± 0.0015
11	106.97 ± 0.68	0.3031 ± 0.0019
19	66.60 ± 0.53	0.4425 ± 0.0024
50	46.06 ± 0.38	0.5509 ± 0.0024

Table 4.2: Fit Parameters for $y = ax^b$

180kbp gene-like segment is about $180 \pm 20nm$.

To further narrow the number of possible attraction sites per 1Mbp domain I have plotted the spatial versus the genomic distance in Fig. (4.7). The data are fitted against the power law $y = ax^b$ for the $300nm/segment$ chains, where y is the spatial distance, x the genomic distance and a and b the fit parameters. The results of the fits are shown in table 4.2.

The exponents yielded by the fits show that about 11-12 attractive segments per 1Mbp domain are needed in order to exhibit a non-random walk behavior with an exponent of about $1/3$. Therefore, the number of attractive sites suggested by the evaluation of the diameter, namely about 11 per 1Mbp domain, is consistent with the number proposed by the exponent. The exponent of $0.32(2)$ was obtained by Münkkel [19] based on the experiments by Yokota et al. [71] for genomic distances of several Mbp. The parameter range yielded by my simulation results is therefore in good agreement with experiment.

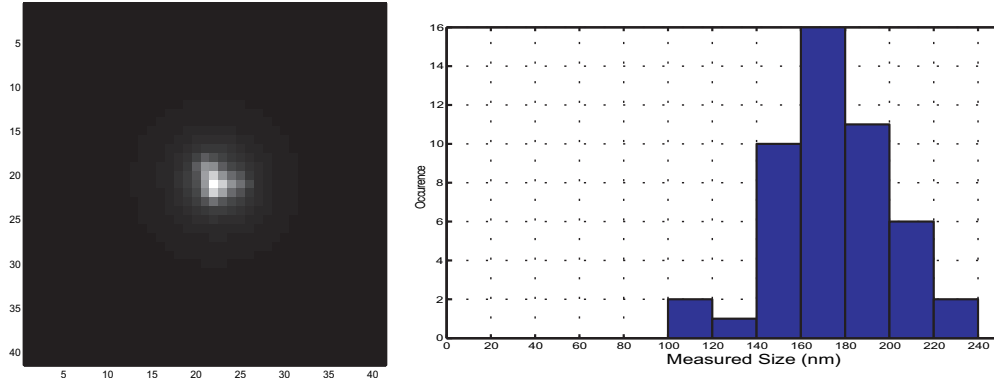


Figure 4.6: A virtual SMI image is shown on the left, the units are in 10nm steps. The right hand side shows the axial extension histogram.

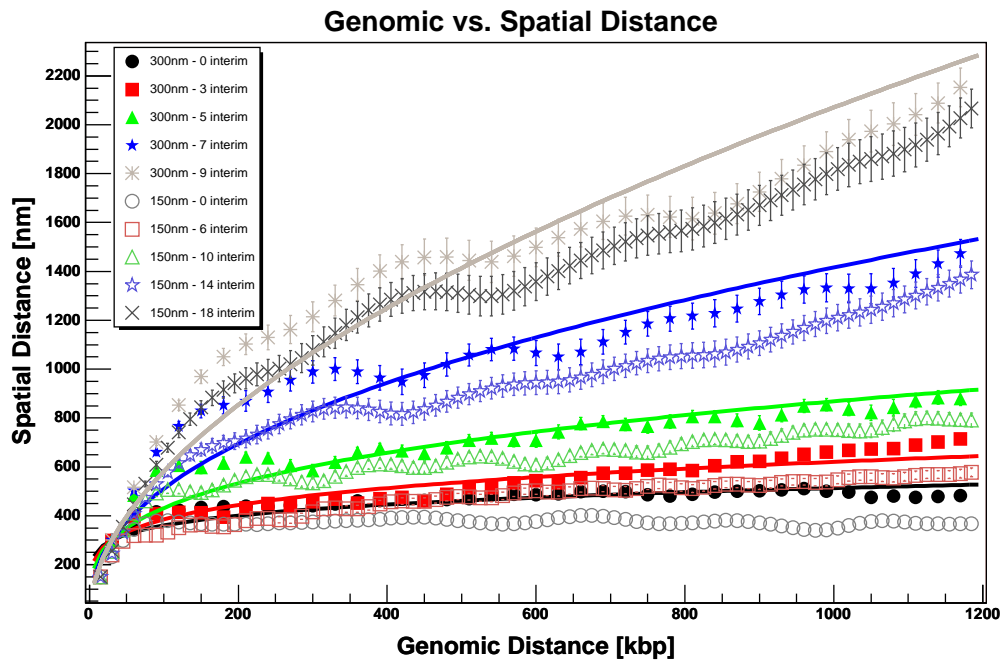


Figure 4.7: Shown is the spatial vs. the genomic distance of a 1Mbp domain. The exponents yielded by the fits clearly favor about 11 attractive sites per 1Mbp domain. The oscillations in the data are due to the rosette nature of the 1Mbp domain.

4.3.1 Free Energy

In order to see if the observed structures are stable it is necessary to look at the free energy of the system. The calculation of free energies of biomolecules is usually no simple task and many elaborate methods have been devised [45]. In this case, however, I use the straight forward approach and use the definition of statistical mechanics.

$$F = -k_B \ln Z = -k_B \ln \sum_r \exp[-\beta E_r(x)] \quad (4.1)$$

where $\beta = 1/k_b T$ and Z is the canonical partition function.

First one evaluate the total energy with respect to time. I took 100 uncorrelated starting configurations and let them equilibrate for the same amount of time. This is shown in Fig. (4.8). Here one can already make the interesting observation that there seems to be a metastable state with a relatively long lifetime before a rosette is finally formed. Then for each time interval one gets an energy distribution which is shown in Fig. (4.9). Finally one only needs to sum over all times for every energy interval and take the negative logarithm to arrive at the free energy. The result is shown in Fig. (4.10). One may observe that there are several metastable states. The one metastable state with a particular long lifetime is a state of two clusters of roughly the same size. That is because the more loops there are on a cluster the larger the entropy barrier that must be overcome in order to form a final cluster.

This method for calculating the free energy is obviously computationally very intensive because a large number of uncorrelated starting configurations must be equilibrated for the same amount of time. This can only be done for relatively short chains. However, this method uses only the statistical definitions and is therefore not susceptible for any bias. However, one should observe that the strict definition of the free energy as used above only holds for equilibrium processes. Since one integrates over time for each energy, one obviously also includes all transients of non-equilibrium mechanics. These transients, however, become negligible for longer and longer times. The true free energy then becomes infinitely deep. Formally this is the only valid domain for term free energy. Nevertheless, I find that the usage of the term free energy is justified since it represents the main idea of the process.

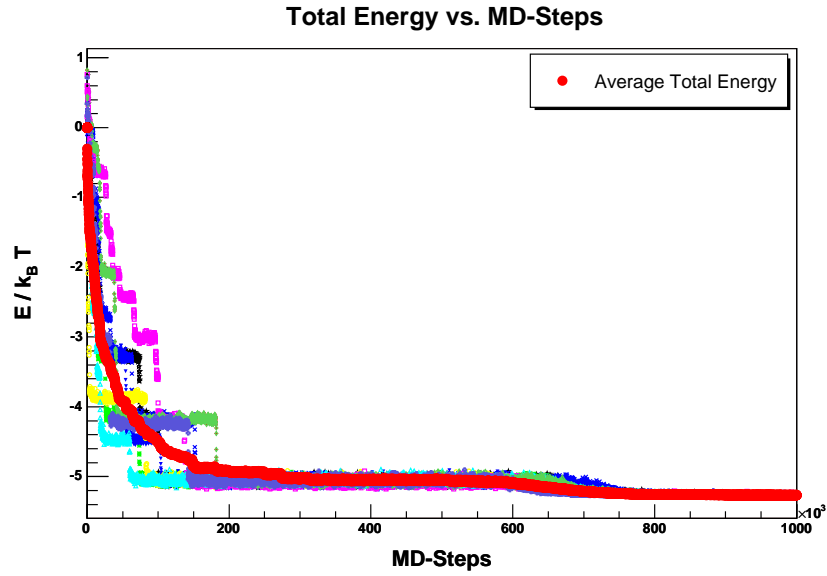


Figure 4.8: Shown is the total energy with respect to the number of simulation steps. The red line represents the average of 100 runs. Only a few individual runs are shown for clarity.

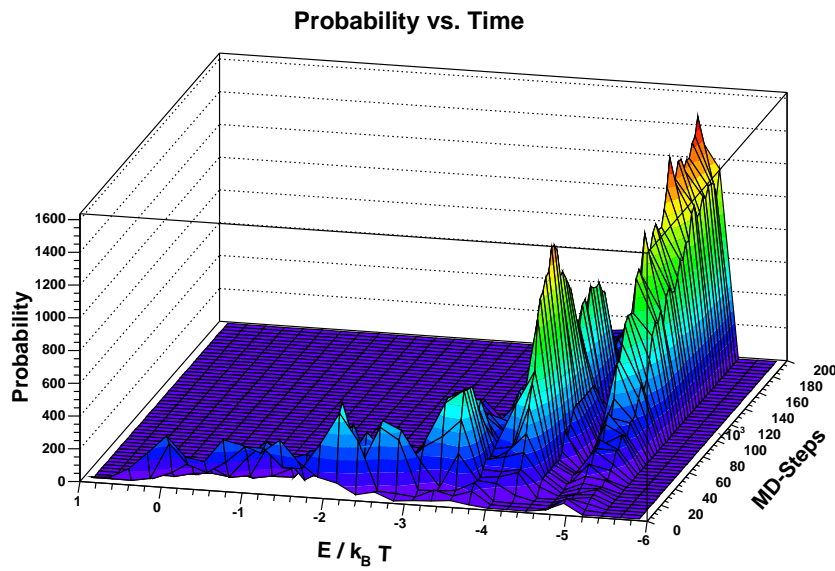


Figure 4.9: A two-dimensional probability distribution of the total energy with respect to time.

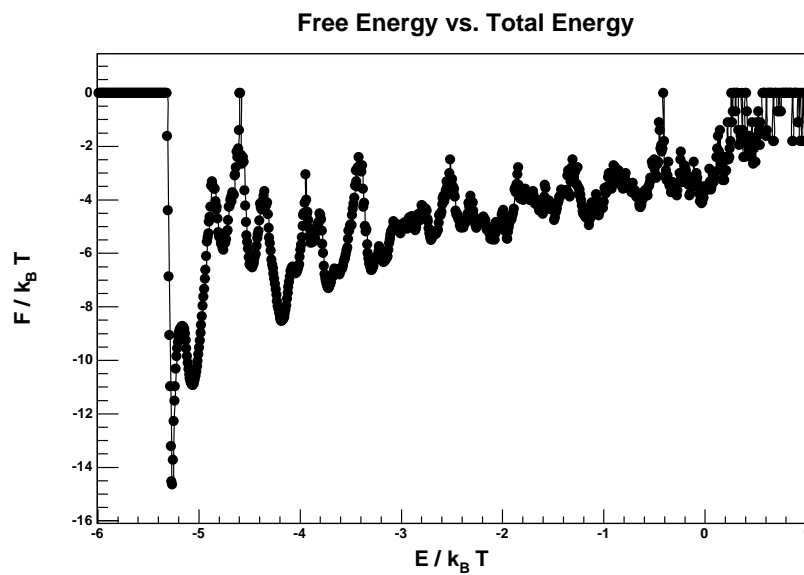


Figure 4.10: Shown is the free energy with respect to the number of simulation steps. One sees a distinct metastable state and the final state of lowest free energy.

Chapter 5

Diffusion

5.1 Motivation

In this chapter I will investigate the diffusive properties of various substances into the 1Mbp chromatin domains. This is particularly interesting for replication and transcription. In essence the detailed dynamics of replication are still unknown. Modern theories no longer cling to the idea of a sequential reading and duplicating of the genetic information by a zipper-like machine from start to finish. Moreover it is believed that replication starts at multiple sites within one chromosome simultaneously[72]. At these sites replication machinery is assembled of various different components. Thereafter, the secondary chromatin structure, i.e. the rosettes, is locally decondensed. After the local replication, the structure recondenses and the replication procedure starts over elsewhere. Hence, these components need to be able to access these sites. It is therefore of great importance to know whether certain substances are able to arrive at these sites by regular diffusion, or if some kind of driving mechanism needs to exist in order for these substances to reach the replication site.

Replication is not the only process that requires accessibility for certain molecules. For transcription the process is very similar. Transcription Factor Complexes (TFC's) need to access the genome, i.e. the chromatin fiber, in order to read the genes and produce proteins. The active genes must not necessarily lie on easily accessible regions of the chromosome[21, 73]. Hence, the question remains the same as for replication.

5.2 Experiment

The ultimate goal of computer simulation in biological physics is to produce results which are in agreement with actual experiments and to be able to make useful and verifiable predictions. I was fortunate enough to be able to collaborate with Cristina Cardoso and her group at the Max-Delbrück Centrum in Berlin. They have done outstanding state-of-the-art experiments which were

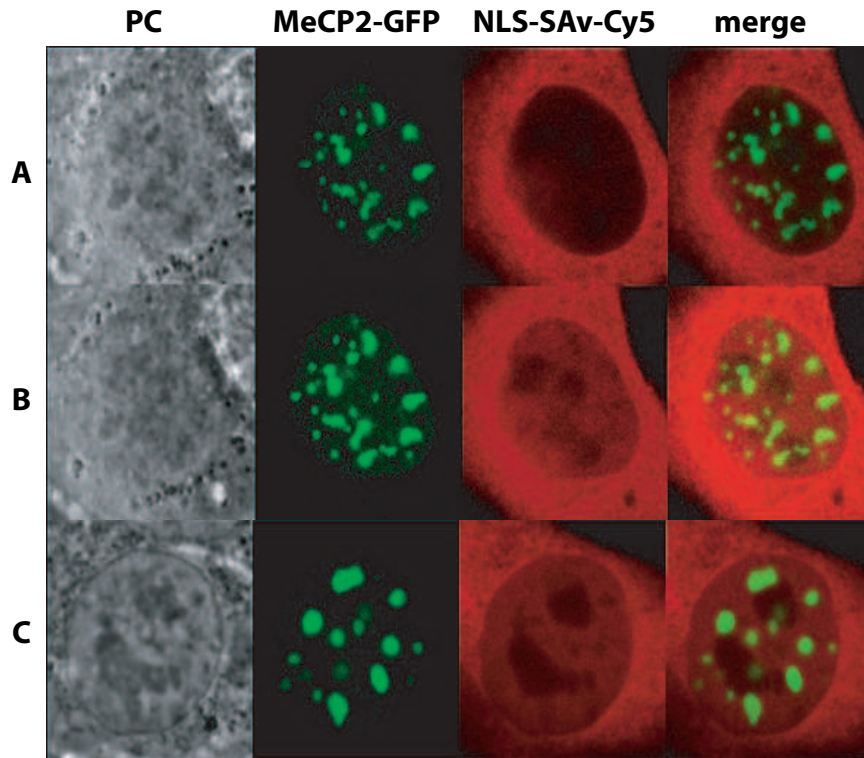


Figure 5.1: Cy5 fluorescently labelled complexes of Streptavidin (SAv) were microinjected into the cytoplasm. (A) A cell immediately after microinjection, (B) The same cell as in A) but 12 min later, the complexes have been transported into the nucleus, (C) A cell with aggregations of MeCP2 labelled centromeric heterochromatin structures (C. Cardoso, unpublished data).

compared to the simulation results.

The Cardoso-lab has performed the NLS-Streptavidin-Cy5 steady state distribution in mouse cells expressing MeCP2-GFP to label heterochromatin. The results of the experiments can be seen in Fig. (5.1) and Fig. (5.2).

The experimental procedures were as follows:

- 10 nuclei were selected showing different FI of NLS-SA v-Cy5 in cyto/nucleoplasm
- MeCP2 labelled structures were selected according to GFP fluorescence label
- nucleoli were selected according to phase contrast (PC) images
- determination of intensity range and background level in individual images

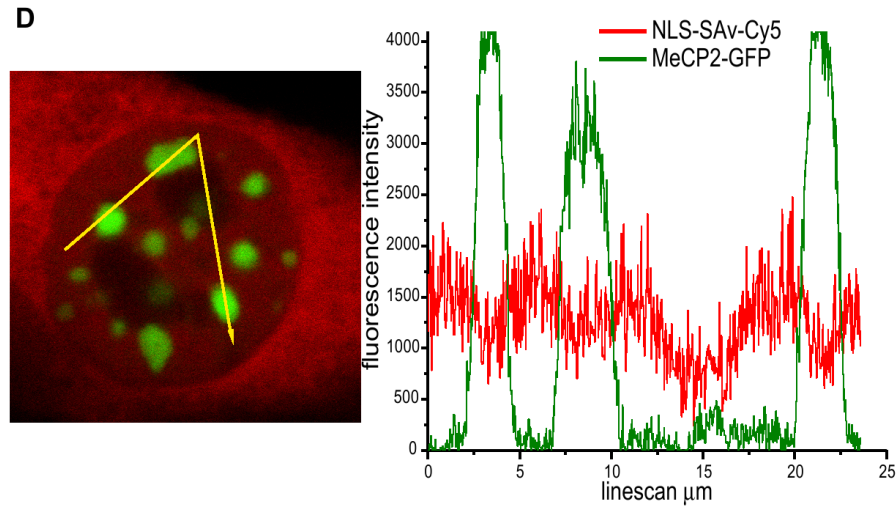


Figure 5.2: The linescan analysis (D) shows that the concentration of the NLS-SAv-Cy5 complexes is decreased in the nucleoli and that some decrease of the complex concentration can also be seen in MeCP2 labelled heterochromatin (C. Cardoso, unpublished data).

- determination of mean fluorescence intensity (FI) for selected structures in each image
- background subtraction for individual images and transfer FI value data in percentage values
- calculation of mean percent values for all cells and standard deviation for NLS-SAv-Cy5 fluorescence intensity in MeCP2 labelled heterochromatin, nucleoli and remaining nuclear regions

and from Fig. (5.2) one can see that:

1. NLS-SAv-Cy5 fluorescence decreases about 30% in MeCP2 labelled heterochromatin
2. NLS-SAv-Cy5 fluorescence decreases 50% in nucleolar regions

The conclusion of this experiment is that the concentration of the NLS-SAv-Cy5 complexes is drastically decreased in the nucleoli and the concentration in the GFP labelled heterochromatin is also decreased, but to a lesser extent. Thus Streptavidin should easily penetrate most of the volume of the chromatin rosettes. The question is if larger substances, such as GFP PCNA Trimer, RNA Polymerase II and Ribosomes can also penetrate the rosettes and thus the heterochromatin.

The Cardoso lab has also done Single-Molecule-Tracking (SMT) and Fluorescence-Correlation-Spectroscopy (FCS) experiments with Streptavidin.

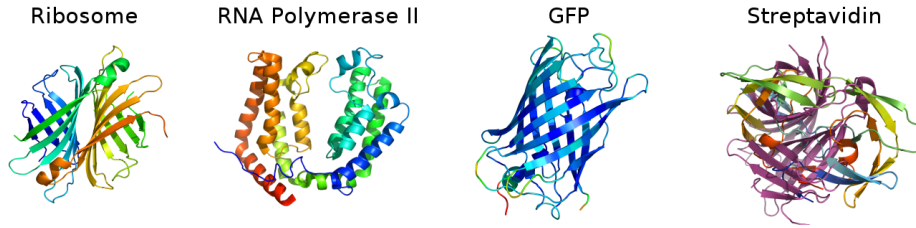


Figure 5.3: The simulated substances shown are not drawn to scale.

They have measured a diffusion coefficient of $80 \pm 5 \mu\text{m}^2/\text{s}$ with SMT and $87 \pm 10 \mu\text{m}^2/\text{s}$ with FCS. These results are in good agreement with the theoretical prediction for the diffusion in aqueous solution, which yields a value of $81 \mu\text{m}^2/\text{s}$.

5.3 Simulation

I have studied the diffusive properties of four different substances: Streptavidin, GFP PCNA Trimer, RNA Polymerase II and Ribosomes. The masses and dimensions as well as the force fields and their respective cutoffs are given in Table 5.1 and Table 5.2. For the chromatin fiber the same mapping was used as in chapter 4, especially the loop size was chosen to be 120kbp and the rosette size was 1.2Mbp, which corresponded best with the required genetic vs. spatial distance exponent of [19, 71] (see chapter 4). Also four different intra-rosette-spacings were investigated, i.e. the distances of the attractive sites. This corresponds to 4 different σ in the Lennard-Jones force-field mappings, ranging from 2.5nm to 27.3nm .

The 1.2Mbp chromatin rosette was put in a simulation box of $1 \times 1 \times 1 \mu\text{m}^3$ with periodic boundary conditions. The chromatin rosette had a diameter of about 800nm . Thus the average distance between rosettes was about 200nm . A closer packing of rosettes, i.e. a smaller simulation box was not possible because of the necessity to generate a feasible starting configuration. The starting configuration was generated as follows. First an already condensed rosette was moved to the center of the simulation box. Then a spherical shell with a radius of 900nm was tessellated into a 1024-on. At each vertex of the 1024-on a particle was placed. This starting configuration was then equilibrated with Dissipative Particle Dynamics until a random distribution was achieved. This configuration was then the new starting configuration for one DPD run. The prior configurations were discarded. This was the only feasible way to generate a working starting configuration. Other approaches were initially tried and then discarded. For instance, randomly distributing the substances lead to very close distances either between the substances themselves or between the substance and the chromatin fiber. In turn this automatically produced large forces and

Substance	Mass	Ellipsoid half-axis (hard core) [nm^3]
Ribosome	4.200 kDa	30 x 20 x 20
RNA Polymerase II	500 kDa	26 x 14 x 14
GFP PCNA Trimer	180 kDa	11 x 11 x 11
Streptavidin	53 kDa	5 x 4 x 4
Chromatin segment	19.500 kDa	300 x 30 x 30

Table 5.1: Mass and dimensions of the simulated substances. 1 Da = 1 u (in SI-Units).

Substance	Potential	Cutoff [nm]
Ribosome	r^{-6}	$7nm$
RNA Polymerase II	r^{-6}	$4nm$
GFP PCNA Trimer	r^{-6}	$3nm$
Streptavidin	r^{-6}	$1nm$
Chromatin segment	r^{-6}	$8nm$

Table 5.2: Parameters for the substances. The potential starts at the hard core and ends after the cutoff.

the simulation was not stable. Setting all force fields initially to zero and then gradually increasing them until their final state lead to a decondensation of the rosette, since the Lennard-Jones force-fields were also decreased and the rosette thus lost coherence.

One simulation consists of one substances at one intra-rosette spacing. Ten runs were computed for each simulation, each run consists of 7.000 uncorrelated configurations. Simulations were done for all substances at all intra-rosette-spacings.

5.4 Results

First of all the diffusion of Streptavidin throughout rosettes of different intra-rosette-spacings is looked at. Simulations were done for spacings of $2.5nm$, $10.0nm$, $21.1nm$, and $27.3nm$. Fig. (5.4) shows the behavior of Streptavidin for the different spacings and the calibration run for the aqueous solution. Qualitatively the straight lines describe regular diffusion for all spacings. Furthermore the diffusion constants differ only slightly for the different spacings. This behavior was observed for all of the simulated substances, hence the plot for Streptavidin is exemplary and the ones for the other substances are omitted. The different diffusion constants for the different spacings for all substances are shown in Tab. 5.3.

Fig. (5.5) shows the diffusive behavior for all of the simulated substances. Again one observes regular diffusion only. Obviously the Streptavidin diffuses much faster than the Ribosomes, but even the large Ribosomes diffuse regu-

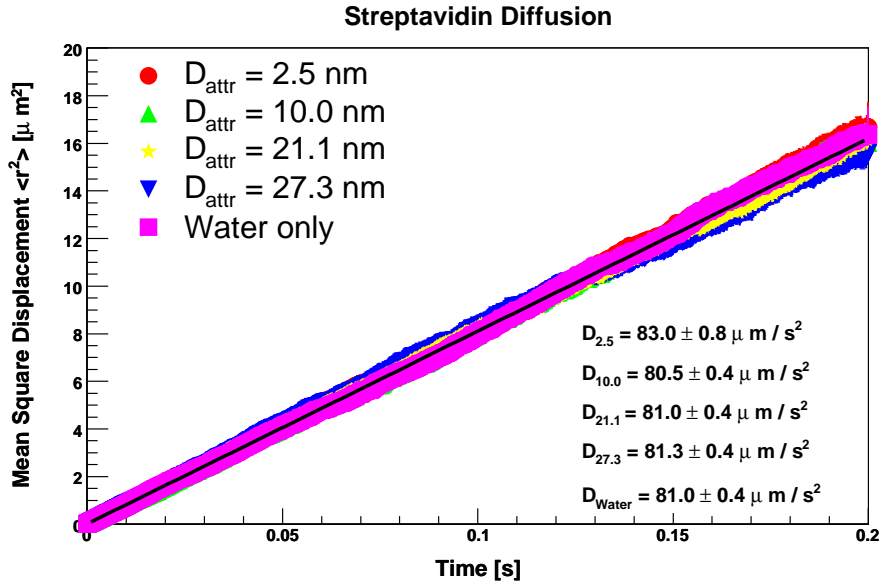


Figure 5.4: The diffusion constants of all the substances vary only slightly with respect to the intra-rosette-spacing. Shown is the diffusion of Streptavidin.

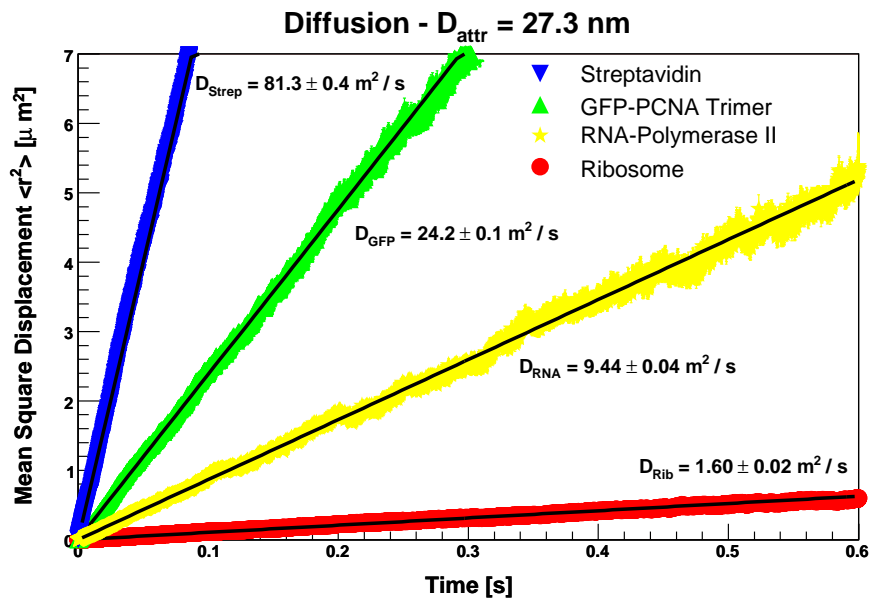


Figure 5.5: All substances show regular diffusion. The diffusion constants are stated for a intra-rosette-spacing of 27.3 nm .

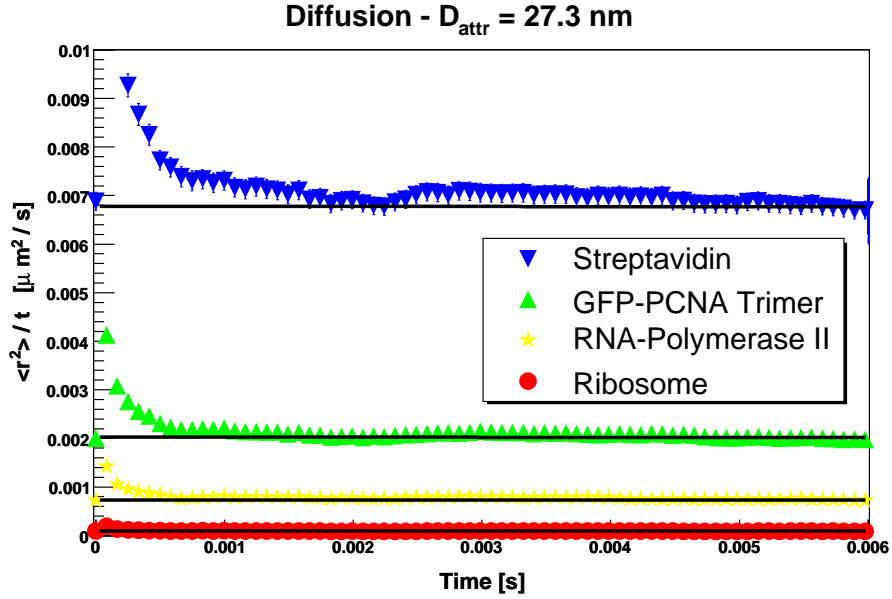


Figure 5.6: Mean square displacement divided by t for the first 0.006 seconds. Note that the time scale is 1/100 of the total simulated time.

Substance	27.3nm $D[\mu\text{m}^2/\text{s}]$	21.1nm $D[\mu\text{m}^2/\text{s}]$	10.0nm $D[\mu\text{m}^2/\text{s}]$	2.5nm $D[\mu\text{m}^2/\text{s}]$
Ribosome	1.60 ± 0.02	1.61 ± 0.02	1.64 ± 0.02	1.69 ± 0.05
RNA Polymerase II	9.44 ± 0.04	9.46 ± 0.04	9.51 ± 0.05	9.73 ± 0.17
GFP PCNA Trimer	24.2 ± 0.1	24.3 ± 0.1	24.4 ± 0.1	24.4 ± 0.3
Streptavidin	81.3 ± 0.4	81.0 ± 0.4	80.5 ± 0.4	83.0 ± 0.8

Table 5.3: Diffusion constants for different intra-rosette-spacings.

larly. A complete overview of all the diffusion constants for all substances and spacings is shown in Tab. 5.3. The table shows that for the same substance and different spacings the diffusion constant do not differ much, within errors they are essentially the same. The errors for the 2.5nm spacing is larger because the sample size is smaller.

In order to see whether the diffusion is truly regular over the entire scale, Fig. (5.6) shows the MSD divided by t for the first 0.006 seconds. One can observe that the regular diffusion is quickly attained. The first approximately 10 points grow increasingly larger, for smaller times. This is due to the fact that at this time scale one can see the distinct DPD steps that only decorrelate after a certain number of steps. Therefore one can conclude that all particles diffuse regularly throughout the entire time scale.

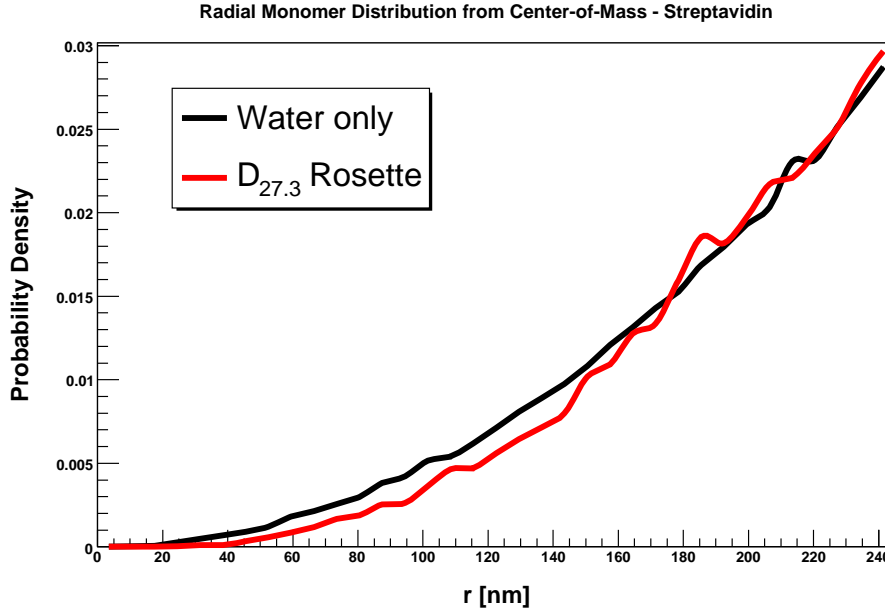


Figure 5.7: The probability density of the Streptavidin from the center-of-mass. In the inner core of the rosette the density is lower than that of the water-only simulation. Beyond about 180nm the density is equal to the bulk density.

5.4.1 Accessibility

Another interesting point to look into is the accessibility of the rosettes. To shed light on this matter I looked at the density of the substance at different z -slices around the core of the rosette. Tab. 5.4 shows these slices for the two extreme intra-rosette-spacings of 2.5nm on the left-hand side and 27.3nm on the right-hand side. Slices at $z = -100\text{nm}$, $z = 0\text{nm}$ and $z = 100\text{nm}$ from the core are shown. One can observe that the core of the small spacing remains compact while the one of the large spacing is more smeared out. Thus the core of the larger spacing is a bit larger, but its boundaries less well defined than the one for the smaller spacing.

A more quantitative analysis is done by looking at the radial monomer distribution from the center of the rosette in Fig. (5.7). In this plot the distribution for the 27.3nm spacing is compared to the regular behavior for the water-only run. Up until about 180nm the density is significantly lower than in the aqueous solution, after that the density is the same as the bulk density.

The relative density of the Streptavidin particles in the 27.3nm rosette with respect to the water-only run is shown in Fig. (5.9). As one approaches the center of the rosette it becomes less and less accessible. The innermost part is absolutely inaccessible. If one fits the density with

$$\rho/\rho_{bulk} = 1 - e^{-\frac{r-\Delta r}{r_0}} \quad (5.1)$$

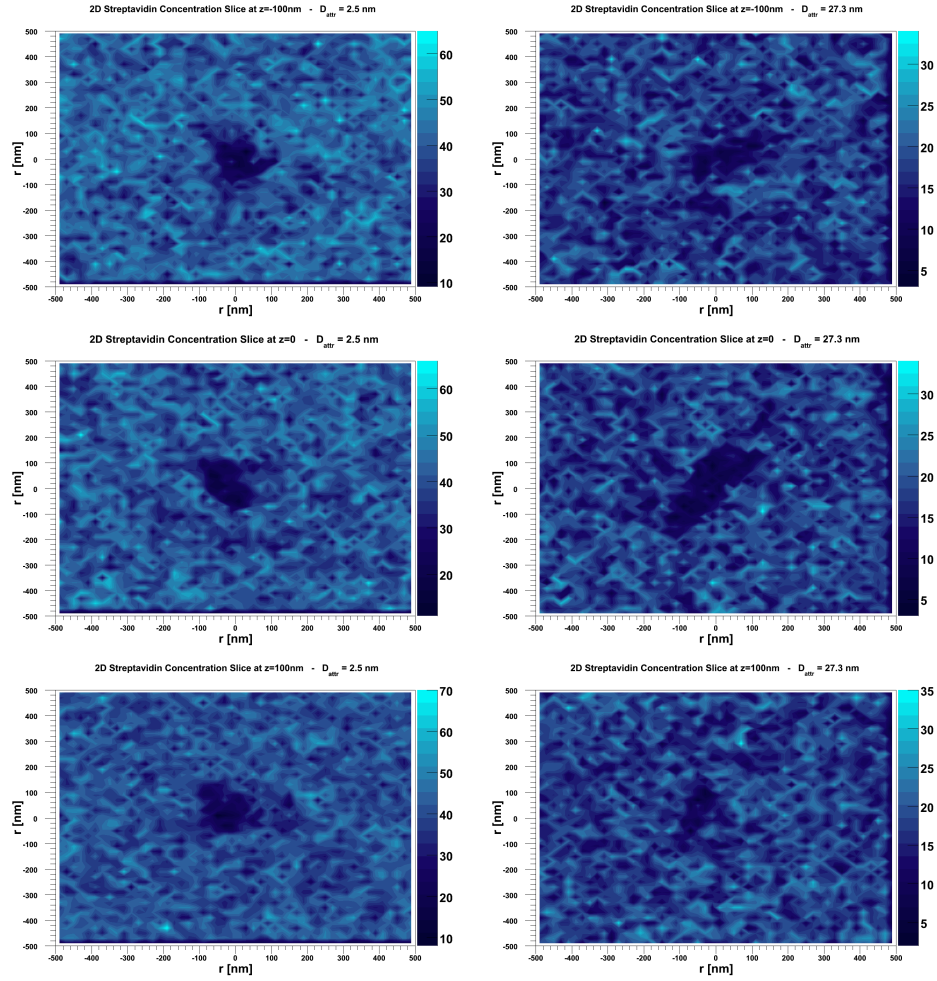


Table 5.4: 2 dimensional density plot. The images on the left show a compact 2.5nm core while the ones on the right show a 27.3nm core. For larger intra-rosette-spacings the rosette-core is smeared out and more penetrable.

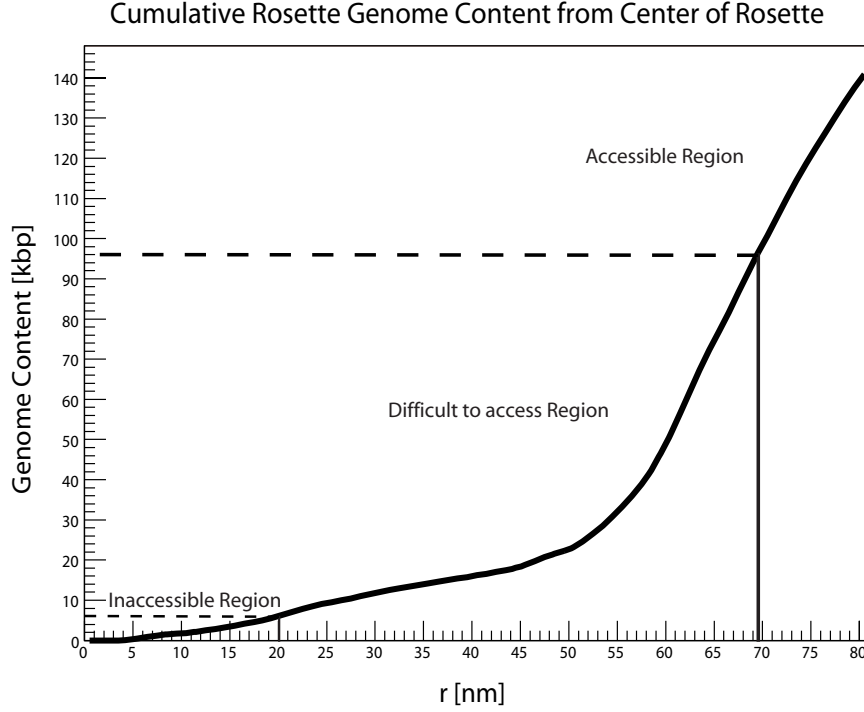


Figure 5.8: The cumulative genome content with respect to the center of the rosette. One can observe 3 regions of different accessibility.

one obtains a fit value of $\Delta r = 20.1 \pm 3nm$ and $r_0 = 69.0 \pm 6nm$. Therefore the excluded volume in the center of the rosette is a sphere of about $40nm$ diameter. This volume comprises $6 \pm 2kbp$ of the rosette-fiber. If one associates r_0 with an average penetration depth, then the DNA between $20nm$ and $69nm$ from the core is difficult to access. This corresponds to $79 \pm 8kbp$ of DNA which is difficult to access, as Fig. (5.8) illustrates.

It is obviously also interesting to see how ribosome-like proteins can penetrate the rosette. All the values of r_0 and Δr for Streptavidin and Ribosomes are listed in Tab. 5.6. There is one trend independent of the substance. For the larger core the particles seem to be able to penetrate into the deep core more easily while the average penetrability is less than for the dense core. This means that a larger core has a greater extension but is less dense and thus particles can occasionally reach further into the core but are in general further away from it. A dense core is basically concentrated at the center and no particles can penetrate deep inside the core. For the larger Ribosomes the r_0 and Δr obviously also increase. Therefore for the $D_{27.3nm}$ rosette approximately $12 \pm 3kbp$ are totally inaccessible to Ribosomes, while $190 \pm 8kbp$ of DNA is difficult to access.

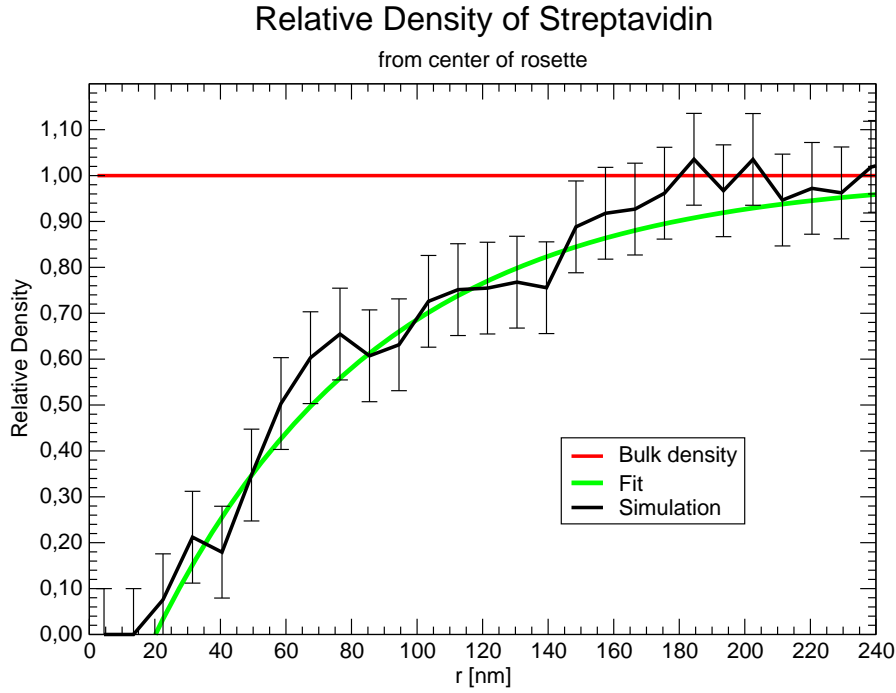


Figure 5.9: The relative density from the center of the $D_{27.3nm}$ rosette. As one approaches the inner core, the density continually decreases. The fit function and values are stated in the text.

The average density drop-offs (a drop-off of e.g. 0.3 means that the intensity under a microscope would drop from 1.0 to 0.7) is shown in Tab. 5.5. These results are in excellent agreement with the experiments.

5.5 Conclusion

The simulations have revealed several noteworthy results. First of all, regular diffusion was observed for all substances and for all intra-rosette-spacings. If particles diffuse anomalously in the nucleus then it is not due to the physical restriction of the rosettes and their cores. At least not on the scale of a single rosette. Even the relatively large Ribosomes access nearly as much of the rosette

Intra-Rosette-Spacing [nm]	Average Density Drop-off
27.3	0.33 ± 0.05
21.1	0.38 ± 0.05
10.0	0.30 ± 0.05
2.5	0.23 ± 0.04

Table 5.5: Average density drop-off for the different intra-rosette-spacings.

Substance	$D_{27.3nm} \Delta r$	$D_{27.3nm} r_0$	$D_{2.5nm} \Delta r$	$D_{2.5nm} r_0$
Streptavidin	$20.1 \pm 3nm$	$69.0 \pm 6nm$	$26.21 \pm 3nm$	$60.9 \pm 5nm$
Ribosome	$27 \pm 5nm$	$143 \pm 6nm$	$32 \pm 5nm$	$110 \pm 6nm$

Table 5.6: Δr and r_0 for Streptavidin and Ribosome-like proteins for the $D_{27.3nm}$ and $D_{2.5nm}$ rosettes.

as the smaller particles. This is supported experimentally by [74], although they measured anomalous diffusion on a larger scale. Furthermore the size of the spacing was basically irrelevant for the diffusion constants.

The second part of the analysis focused on the accessibility of the rosette. I have shown that the center of the rosette is inaccessible for all spacings. Experimentally Streptavidin decreases in the heterochromatic regions (see Fig. (5.2)) and from the simulation results one can conclude that intra-rosette-spacings from $2.5nm$ to $27.3nm$ yield density drop-offs between 23% and 38%, which is good agreement with the experiments.

Even though the most part of the rosette is accessible even to large molecules such as Ribosomes, a small region in the center of the rosette is not. A promoter region is 3kbp, which is comparable to the inaccessible genome in the center of the rosette. Therefore if exactly this promoter region is inaccessible, the RNA polymerase may not bind and the entire gene may not be transcribed. Therefore, unlike [75], my results show that the silencing of genes in the 1Mbp Chromatin domains could possibly be caused solely by physical inaccessibility.

Chapter 6

Chromosome Mapping

6.1 Motivation

Two chapters back the formation of 1Mbp domains with a regular A_nB_m pattern was observed. The obvious next step is to investigate longer sequences of randomly distributed A and B with a certain probability. That is what is done in this chapter. In addition I compare the simulational results to experimental ones. This serves as a real test for the model and it will be shown here that it holds really well.

This chapter deals with the human chromosome 1, in both simulation and experiment. It is the largest of all human chromosomes with 246Mbp [76]. Along with a multitude of syndromes, it is associated with Alzheimer's disease, Leukemia, Glaucoma and Cataracts. It is obvious that, at this stage, no polymer-type model can explain the occurrences of these diseases, however, the analysis of structure-formation may be fundamental for further understanding.

Fig. (6.1) shows the average transcription vs. genomic distance for the human chromosome 1. The blue areas indicate little transcriptional activity, so called ridges, while the pink ones represent vigorously transcribed regions, so called anti-ridges. Ridges are characterized by high gene density and high GC content, anti-ridges exhibit fully opposite characteristics [77, 78]. The experiment and the simulation deal with about 3Mbp of the chromosome, marked by the red and green arrows. Genomic vs. spatial distance measurements are performed in vitro and in silico and compared to one another.

6.2 Experiment

The experiments were conducted by Sandra Goetze in R. van Driels Lab at the Swammerdam Institute for Life Sciences at the University of Amsterdam [79]. In order to understand what kind of treatment is involved experimentally, a short synopsis of the experimental procedures is given here, for details please see [79].

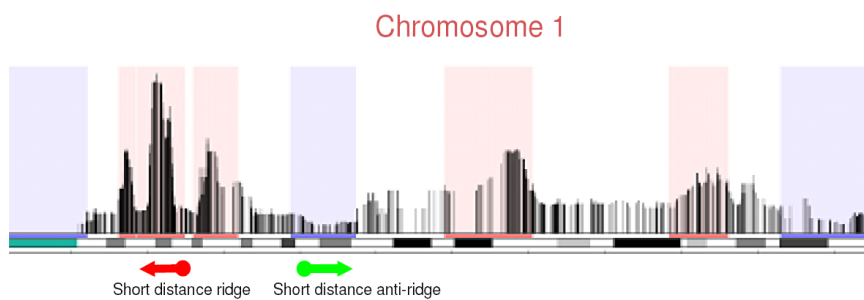


Figure 6.1: Transcriptional activity of chromosome 1. The abscissa shows the genomic distance and the ordinate the transcriptional activity. The ridges are areas of high gene density whereas the anti-ridges are such of low gene densities. (Courtesy of J. Koster and R. Versteeg, Academic Medical Center, University of Amsterdam, unpublished data)

As the cell culture, human female primary fibroblasts were used. For the Fluorescence in situ hybridization (FISH) markers were selected with an average distance of 400kbp between their centers. This resulted in a contiguous and homogeneous set of probes with an average gap-distance of approximately 220kbp. In short, cells were incubated with a 30 min pulse of 25 μ M BrdU to label replicating DNA prior to fixation. Denaturation was carried out at 78°C. Hybridization was allowed to proceed overnight at 37°C. Post-hybridization washes were then performed at 45°C. All incubations for probe detection were performed at room temperature. Conjugated antibodies and conjugated streptavidin were used to visualize the signals.

All experiments were performed in duplicate. For each experiment 40 to 60 nuclei were imaged. S phase, as well as cells in G2 were excluded. Twelve-bit images were recorded using an LSM 510 confocal laser-scanning microscope equipped with a 63x/1.4 NA Apochromat objective. An Ar laser at 488nm, a He/Ne laser at 543nm and a He/Ne laser at 633nm were used to excite the fluorochromes simultaneously. Fluorescence was detected with the appropriate bandpass filters. Images were scanned as 512x512x32 voxel images with a sampling rate of 50x50x100nm.

Image analysis was carried out with Argos software. To identify and quantitatively analyse FISH-labelled areas, 3D images were treated with a bandpass filter and subsequently segmented using a fixed threshold. The center of gravity was calculated for each signal using the original unfiltered data set.

6.3 Simulation

The chapter on the 1Mbp domain indicated that a Kuhn length of 150nm is better suited for modeling the 30nm chromatin fiber. Thus I now use a 150nm, i.e. 15kbp Kuhn segment for my simulation. The potential parameters for the repulsive and attractive segments remain unchanged:

- Segment diameter: 30nm
- Segment length: 15kbp = 150nm each
- Total chain length: 3Mbp = 200 15kbp segments
- The harmonic bond potential is taken to be

$$U_{bond}(l) = \frac{kT}{2\delta^2}(l - l_0)^2 \quad (6.1)$$

with $\delta = 0.1$ and $l_0 = 300nm$ at 310.15K.

- The angular and torsional potentials are taken to be 0. On this scale the chain is flexible.

- Repulsive segments potential

$$U_{rep}(r) = \epsilon \left(\frac{\sigma}{r - r_{segment}} \right)^6 \quad (6.2)$$

with $\epsilon = 0.14k_bT$ at body temperature, $\sigma = 15nm$ and $r_{segment} = 15nm$ being the fiber radius.

- Cutoff for the repulsive potential is $r_c = 8nm$ (after the $30nm$ fiber diameter).
- Attraction segments potential

$$U_{attr}(r) = 4\epsilon \left[\left(\frac{\sigma}{r - r_{segment}} \right)^{12} - \left(\frac{\sigma}{r - r_{segment}} \right)^6 \right] \quad (6.3)$$

with $\epsilon = 7k_bT$ at body temperature and $\sigma = 30nm$.

- Cutoff for the Lennard-Jones potential is $r_c = 80nm$ (after the $30nm$ fiber diameter).

The difference to the 1Mbp domain simulation is that one now has a chain three times as long and the block-copolymer pattern is no longer of the regular A_nB_m type, but randomly distributed. I have simulated 5 different N_A/N_B ratios, N_A being the number of repulsive segments and N_B the number of attractive segments: 0, 0.2, 0.4, 0.6 and 0.8. For example, an AB sequence for a ratio of 0.4 would be:

```
AAABABAABAABBAABAAABBAABBAABBAABBAABAAAAAABBBBABBAABAA
BAABBAABAAABABABAAABBAABAAABABABAAAAABABBAABAAAAAABAABAAA
AABBBBBBBAABAAABABBAABAAABABAAABABAAAAAABBBBAABAAAAAB
AAAAAAAAAABBABAAAAAA
```

For each ratio, 150-200 different random AB distributions were equilibrated, in order to make statistically sound statements. The equilibrated structures were then analyzed.

6.4 Results

At first, the experimental data for chromosome 1 were compared to the 1Mbp domains with 30kbp Kuhn segments of the previous chapters. The result can be seen in Fig. (6.2). I would like to point out that these were the rosettes with a regular A_nB_m pattern. The slightly oscillatory behavior of the 210kbp loop line represents the regular loop structure. The oscillations are not so pronounced because for the shorter distances because the spatial distance was averaged over all possible segments. Looking closely, one can still make out 6 maxima in the 210kbp line representing the 6 210kbp loops totaling to the 1.3Mbp rosette.

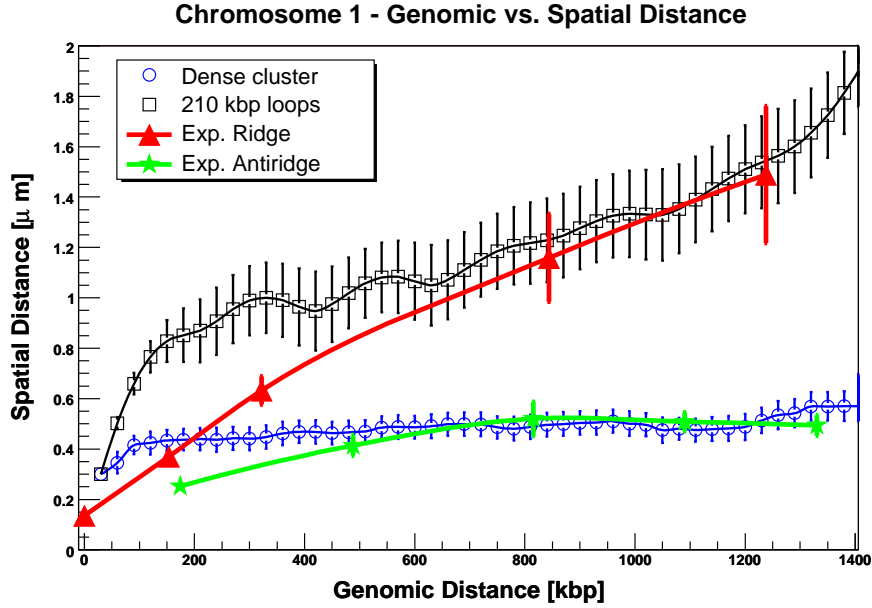


Figure 6.2: Comparison with the simulated rosette sizes for the genomic vs. spatial distance measurements for chromosome1.

The ridge data is in good agreement with the dense cluster line (i.e. only attractive sites) from about 500kbp on. The anti-ridge data fits the simulation from about 800kbp on. For small distances, the simulation data doesn't hold up very well. Furthermore, since roughly 1Mbp domains were simulated, it is impossible to predict a trend for the congruency for larger distances. It is clear that a regular loop rosette was a good starting point for simulations, but doesn't hold up to the entire range of the experimental data. This lead to the idea of randomly distributed attractive sites and longer chains.

In Fig. (6.3) the results are displayed for the different N_A/N_B ratios. From a ratio of about 60% on, there is no further compaction, which puts a natural limit on the possible number of attractive sites, since more do not produce a more compact structure. Furthermore one can observe that 20% attractive sites already reduce the spatial distance by a factor of 0.5, compared to the chain without attractive segments.

The simulation results only reveal their full significance when compared to the experimental data, as shown in Fig. (6.4). The experimental ridge is in excellent agreement with a free chain (in the sense of free from attractive sites, excluded volume is still present). The experimental anti-ridge data is bounded by the 20% and 40% attractive lines, suggesting good compatibility to a 30% attractive line. Furthermore for both lines the congruency is excellent down to the very small genomic distances, the difference observed for the regular patterns is gone. This proves that the modifications to the model were justified

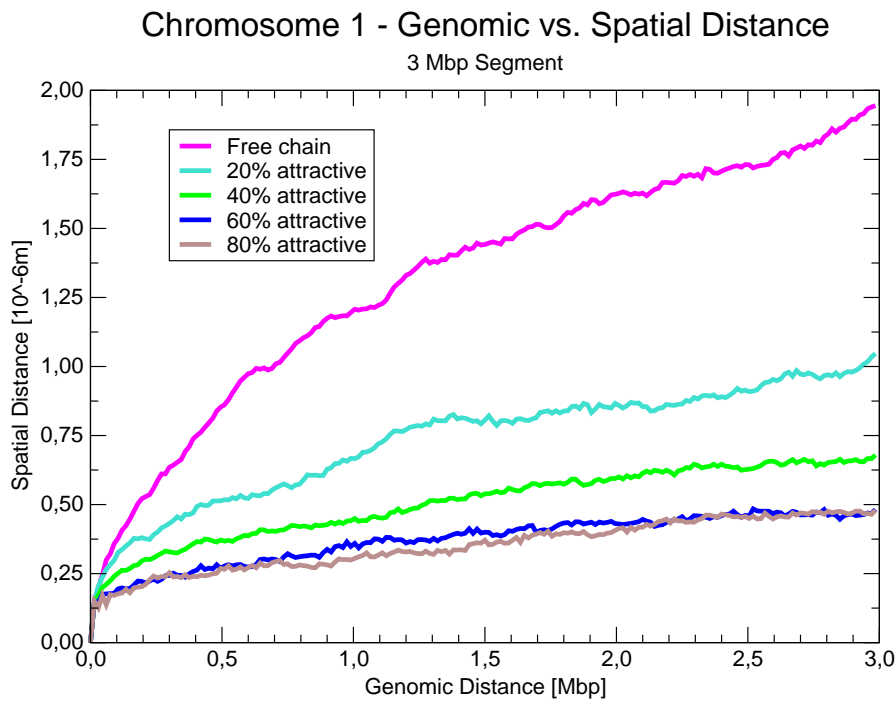


Figure 6.3: Simulation of the genomic vs. spatial distance for different numbers of attractive sites for chromosome 1. The errors are omitted for clarity, they are approximately 10%.

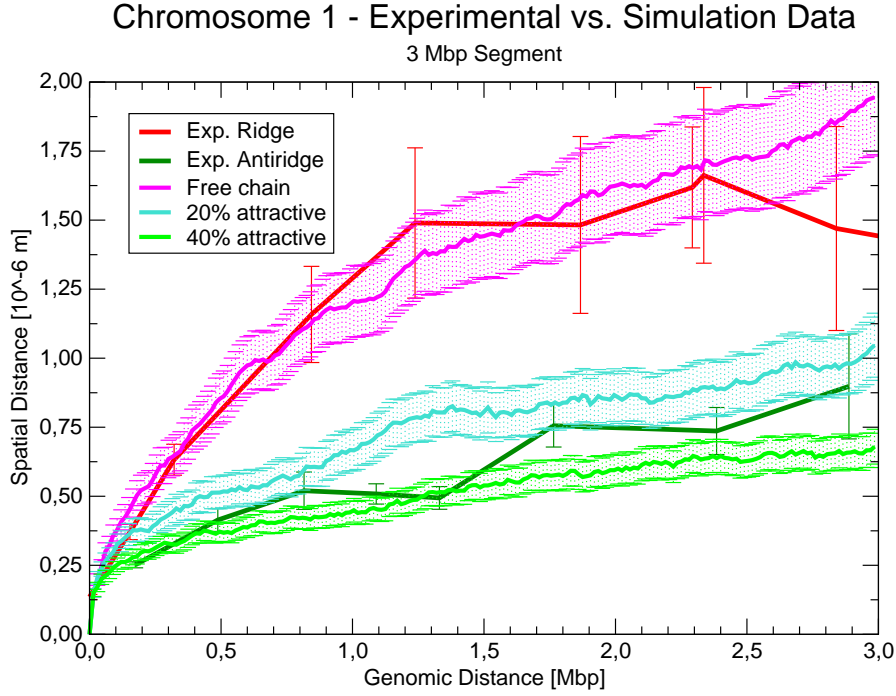


Figure 6.4: Genomic vs. spatial distance measurements and simulations for chromosome 1.

and a more realistic model was obtained.

In summary one can conclude that areas of low transcriptional activity (anti-ridges), i.e. with low gene density have a N_A/N_B ratio of about 30%. Therefore the higher compaction of these areas could be explained by a higher concentration of condensing agents. The ridges, i.e. highly transcribed areas basically behave like a free real polymer chain (with excluded volume). I find this result very intuitive, because the actively transcribed regions need to be more readily accessible. However, the anti-ridges are not entirely compact, as the ratio of 30% emphasizes. Recall that the highest compaction starts at about 60%. Thus there is still room for transcription factors to access these areas. Overall I find the agreement of simulation and experiment striking and believe that comprises strong evidence for the claim of being a realistic model.

It would perhaps be interesting if future studies could correlate this model to a metaspale model based on the Peyrard-Bishop model [80]. In their model they investigate the statistical mechanics of the denaturation of two DNA strands given by the Hamiltonian

$$H = \sum_n \frac{1}{2} m (\dot{u}_n^2 + \dot{v}_n^2) + \frac{1}{2} k \left[(u_n - u_{n-1})^2 + (v_n - v_{n-1})^2 \right] + V(u_n - v_n) \quad ,$$

where

$$V(u_n - v_n) = D (\exp [-a (u_n - v_n)] - 1)^2$$

i.e. a Morse potential is used as a pair potential.

Chapter 7

Chromosome 22

7.1 Motivation

In the previous chapter I have analyzed a 3Mbp region of the human chromosome 1. The next logical step is to simulate an entire chromosome. Obviously one then chooses the smallest, the chromosome 22. It has 49Mbp of DNA and is thus one of the smallest of all chromosomes. Interestingly it was the first human chromosome to be sequenced. It was completed in December 1999 [81]. Associated genetic defects include the Cat Eye Syndrome and the 11/22 Translocation. In this defect a marker chromosome is made up of the upper (p section) and part of the lower (q section) arms of chromosome 22, and a small portion of the lower (q section) arm of chromosome 11.

As before my model doesn't claim to resolve any medical issues. The aim is to see if the experimentally determined size of chromosome 22 can be reproduced by the model without any further assumptions. This would already be a great success for the model because it then holds on many different scales.

7.2 Simulation

For the simulation of the chromosome I am still using the same block-copolymer model as before, only that I now resort back to a regular distribution of attractive sites. This is done because of the computational complexity of simulating an entire chromosome. If one were to take random distributions one would need to take many different distributions to obtain proper statistics. This is not possible because the equilibration of one run took about 3 Months on one Athlon64 processor at 2 GHz. For this analysis 10 different starting configurations were equilibrated and 4.000 configurations were sampled for each starting configuration.

Using the similar parameters as in the previous section, we now simulate the chromosome 22. With a segment length of $300nm$ we obtain 1667 segments. A visualization is shown in Fig. (7.1). The detailed parameters are as follows:

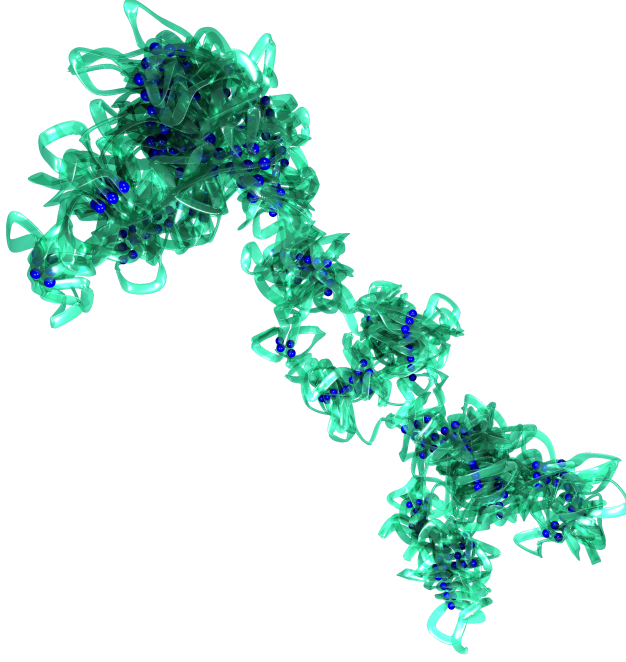


Figure 7.1: The chromatin fiber condensed into the human chromosome 22. The condensing agents are shown as blue spheres.

- Segment diameter: $30nm$
- Segment length: $30.0kbp = 300nm$ each
- Total chain length: $49Mbp = 1667 \text{ } 30.0kbp$ segments
- The harmonic bond potential is taken to be

$$U_{bond}(l) = \frac{kT}{2\delta^2}(l - l_0)^2 \quad (7.1)$$

with $\delta = 0.1$ and $l_0 = 300nm$.

- The angular and torsional potentials are taken to be 0. On this scale the chain is flexible.
- Repulsive segments potential

$$U_{rep}(r) = \epsilon \left(\frac{\sigma}{r - r_{segment}} \right)^6 \quad (7.2)$$

with $\epsilon = 0.14k_bT$, $\sigma = 15nm$ and $r_{segment} = 15nm$ being the fiber radius.

- Cutoff for the repulsive potential is $r_c = 8nm$ (after the $30nm$ fiber diameter).

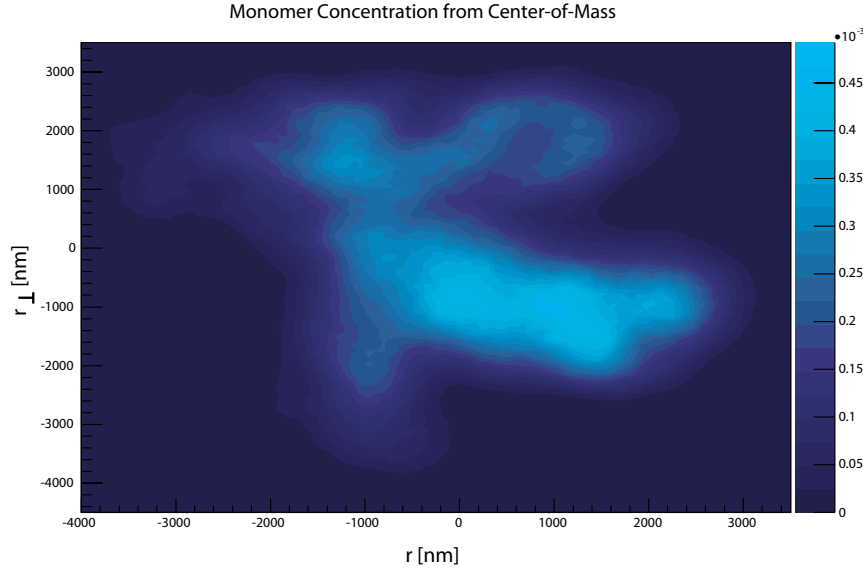


Figure 7.2: A virtual light nanoscopy image of a the simulated chromosome. Plotted is the two-dimensional projection of the monomer concentration.

The attractive sites are positioned at every 5th monomer, this yields a ratio of 20% attractive sites. This was the optimal value yielded by the simulation of the 1Mbp domain. The individual rosettes then have loop sizes of about 120Kbp. As can be seen in Fig. (7.1) the rosettes are no longer restricted to 1Mbp domains, some are larger, some smaller. The different sizes of the rosettes is a natural consequence of the interplay between the attraction of the sites and the repulsion due to the unfavorable decrease in entropy.

7.3 Results

Applying the same analysis techniques as in the previous section we obtain the monomer concentration plot shown in Fig. (7.2). This is merely a concentration plot of one configuration, namely the one depicted in Fig. (7.1). In order to obtain the diameter I have analyzed the radius of gyration.

The distribution of the radius of gyration of the 40.000 sampled configurations is shown in Fig. (7.3). The simulation yields an average radius of gyration of $3272 \pm 360nm$. The experimentally observed value for the diameter is $2000nm$ (M. Branco, A. Pombo, MRC London, unpublished results). Therefore the size predicted by the simulation is off by a factor of 1.6. However, one has to recall that no further restrictions were imposed, especially no external pressure. Thus a chain in 'free' space without interference from other objects was simulated. In particular no nucleus and no other chromosomes were simulated. In a real nucleus, the presence of 45 other chromosomes are a major

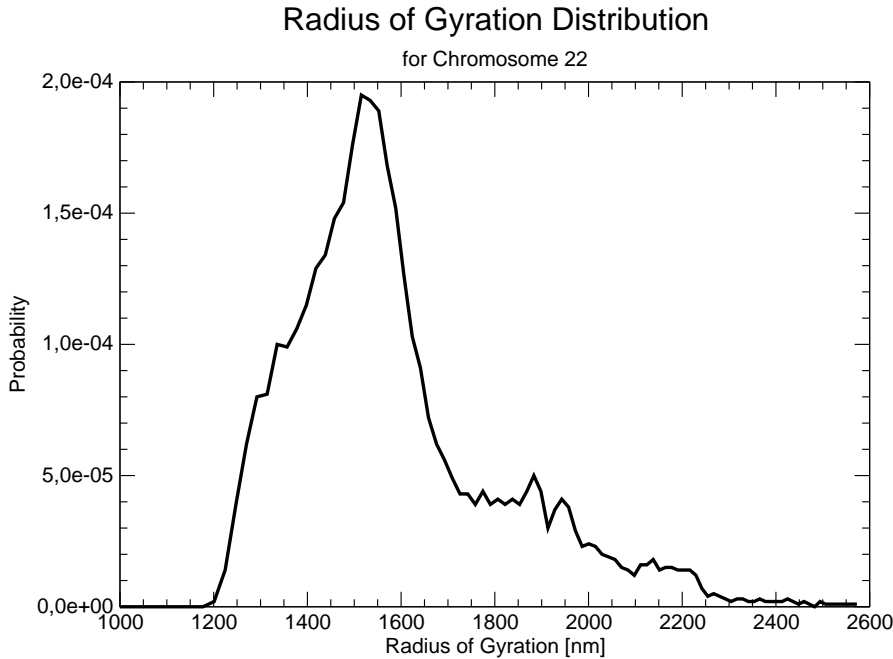


Figure 7.3: The distribution of the radius of gyration of the chromosome 22.

constraint and result in an external pressure on the nucleus but also on the individual chromosomes. Therefore a difference by a factor of 1.6 is acceptable. Most of the compaction was achieved by the attractive segments, an additional factor of 1.6 can easily be facilitated by the external pressure due to the limited space.

Another point of interest is to see at what temperature the chromosome condenses. Having attractive Lennard-Jones potentials, one expects to see a phase transition from a free state to a collapsed state at a certain temperature. This transition is shown in Fig. (7.4). The radius of gyration is chosen as the indicator of the size. Clearly at low temperatures one observes the collapsed state, while for higher temperatures we see the free state. The transition between the two states is very sharp, as expected for the Lennard Jones system.

In summary for the chromosome 22 the simulations yield approximately the correct size just by the assumption of having attractive condensing agents, which were thought only to be important for the formation of the rosettes. Clearly the role of the condensing agents is far greater.

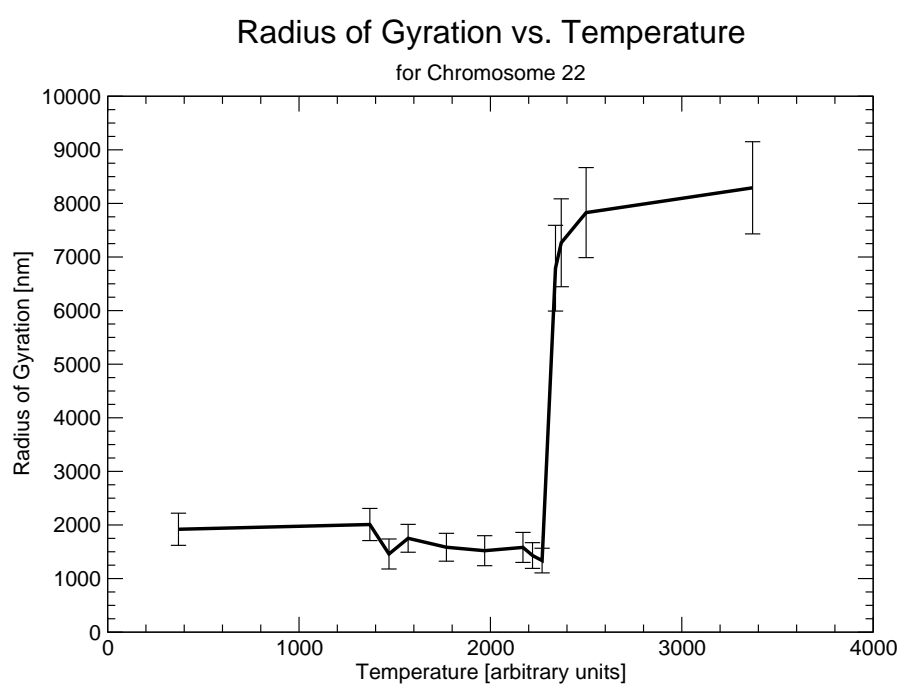


Figure 7.4: The temperature dependence of the radius of gyration. At a temperature of about 2300 one sees the transition from the condensed to the de-condensed state.

Chapter 8

Drosophila

8.1 Motivation

The process of gene silencing is a crucial building block for the picture geneticists developed during the last decade about the functioning of genes within all kinds of organisms. Gene silencing is a highly complex area of research, and several mechanisms have been identified that inhibit gene expression within the nucleus.

The simplest molecular model to explain gene silencing postulates that specific repressors regulate the onset of transcription, by binding directly to specific DNA sequences and counteracting the action of activators and of the transcriptional machinery. A second possibility is that repressors, bound at specific sequences called silencing elements, might act by regulating the structure of the chromatin. The structure of the chromatin reduces the accessibility of DNA to the transcriptional machinery, and repressors might prevent transcription by stabilizing the binding of histones to DNA, or the folding of nucleosomes in compact higher-order chromatin structures.

Beside these levels of regulation, another level exists: namely the three-dimensional organization of chromosomal domains in the cell nucleus during cellular differentiation and development [82, 83]. In several cases, gene silencing has been correlated with relocation of chromosomal domains. In most of the published studies, gene silencing correlated with gene positioning close to heterochromatic compartments. Heterochromatin represents a highly compact region of chromatin where genes are stably repressed.

Another case of gene silencing that shares common features with heterochromatin silencing involves the proteins of the Polycomb group (PcG) [84]. PcG proteins are highly conserved regulatory factors that are responsible for the maintenance of the silent state of important developmental genes, such as homeotic genes. In *Drosophila melanogaster*, PcG proteins form multimeric complexes and regulate their genes through binding to chromosomal regulatory elements named PcG response elements (PREs). This silencing involves repres-

sive modifications on the target chromatin. In addition, it has been observed that silencing via PcG proteins and PREs is enhanced by the presence of multiple copies of PRE-containing elements in the nucleus. These copies may, but do not have to be on the same chromosome. Long-distance pairing between these two loci, which brings them closer together than they would usually be, leads to strong repression of the genes they control (Bantignies et al [84]). This type of regulation represents silencing by geometrical closeness, established in interphase nuclei (see Fig. 8.3).

The goal of this part of my thesis is to build a predictive model for proximity and interactions of chromosomal domains. Within the model, chromosomes were assigned a Rabl configuration [85] in the nucleus, a situation which is present in *Drosophila* embryonic nuclei. I calculated the expected distance distribution of the investigated loci and compared this distribution to experimental results.

8.2 Experiment

Frédéric Bantignies, Virginie Roure and Giacomo Cavalli from the Cavalli Labs in Montpellier have performed the experiments on the movement of a chromosomal domain induced by PREs and PcG proteins. I would like to briefly explain their experiment.

A double transgenic Fab-X,2L; Fab-7' line, which carries two transgenic Fab-7 and no endogenous copy is used to study the movement of a chromosomal domain in the nucleus. In this line, one Fab-7 copy is located in the sd locus, at cytological position 13F, in the middle of the X chromosomal arm, the other one is located in the 38 F locus, cytologically close to heterochromatin, on the left arm of chromosome 2 (chr.2L). Preliminary observations using 3D FISH indicated that, while the 38F locus stays close to the heterochromatin compartment in the double transgenic line, the sd locus changes dramatically its nuclear position and becomes relocalized closer to heterochromatin and to 38F in a large fraction of nuclei. Now it is interesting to analyse if the sd movement toward heterochromatin influences the mobility of other loci located along the chr.X. Besides the sd probe, two additional probes along the chr.X have been produced, 2Mbp apart from the sd locus. One is located at cytological position 12C, more distal to the centromere, the other one is located at cytological position 16D, closer to the centromere. Another probe is a peri-centromeric probe that corresponds to the histone gene cluster locus, located at position 39D-E, above the 38F locus in chr. 2L. Three-color 3D FISH experiments on whole mount embryos were performed, and the relative 3D distances between sd, 12C or 16D loci, and the peri-centromeric histone locus were measured. These distances were measured in the double transgenic line, and compared to a control line containing only one Fab-7 element at the 38F locus (Fab-2L; Fab-7' line).

Measurements were done in both lines, the Fab-X2L; Fab7' and the Fab-2L;

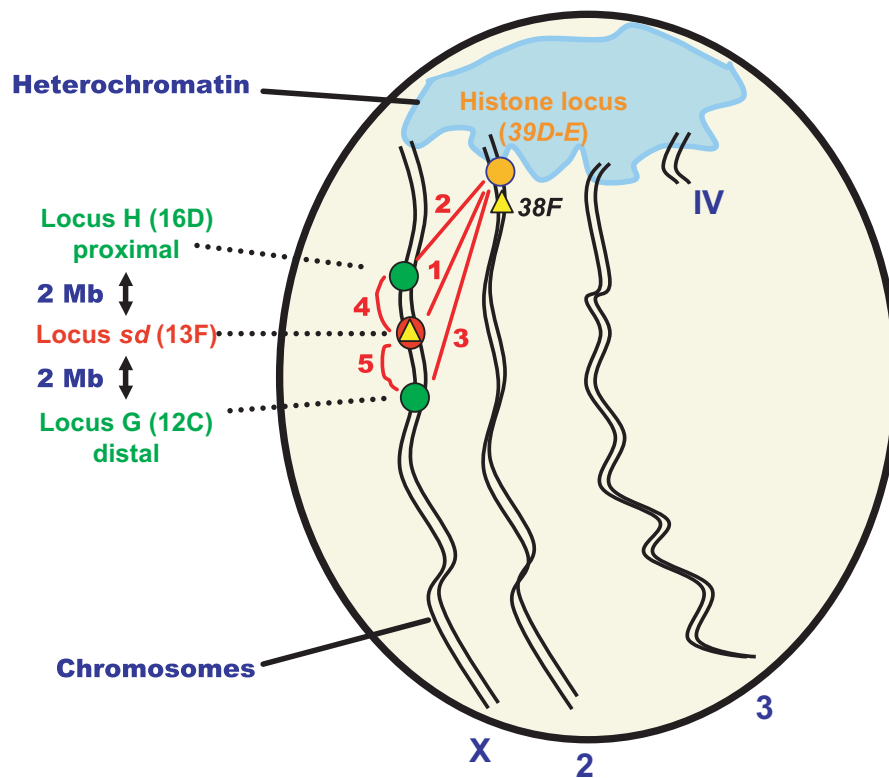


Figure 8.1: The yellow rectangles represent the position of the Fab-7 element: a) in the Fab-X2L; Fab7' line, there is two copies of the Fab-7 element: one at the *sd* locus (13F), one at the 38F locus. b) in the control Fab-2L; Fab7' line, there is only one copy of the Fab-7 element: the one at the 38F locus.

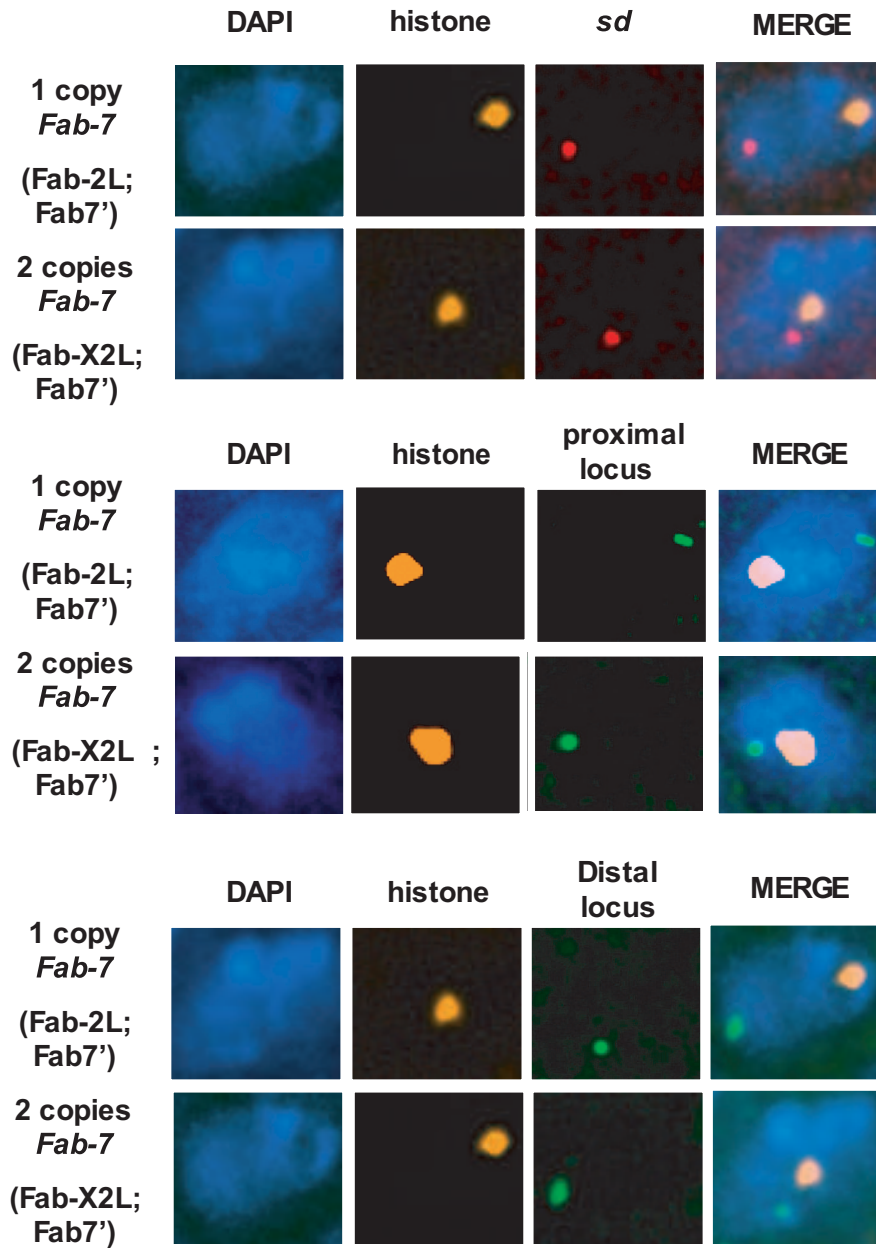


Figure 8.2: 3D-FISH experiment in the *Fab-2L*; *Fab-7'* control line (top), and the *Fab-X2L*; *Fab-7'* line (bottom) mapping the distance between the different loci (top: *sd*, middle: *H*, bottom: *G*) and the pericentric region of chromosome 2 (indicated by 'histone'), since a probe from the histone locus was used. (Roure, Bantignies and Cavalli, unpublished data)

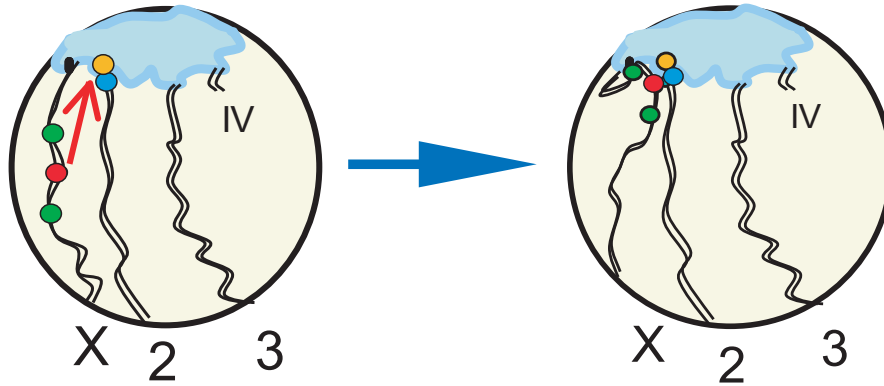


Figure 8.3: The different loci are repositioned close to the heterochromatin when two Fab-7 transgenes are present at the genomic loci.

Fab7' control line, 5 different distances in 210 nuclei:

1. Distance sd (13F) locus - Histone locus (39D-E)
2. Distance H (16D) proximal locus - Histone locus (39D-E)
3. Distance G (12C) distal locus - Histone locus (39D-E)
4. Distance sd (13F) locus - H (16D) proximal locus
5. Distance sd (13F) locus - G (12C) distal locus

It is observed that the distance between sd and the peri-centromeric histone locus is significantly lower in the line with two Fab-7 copies than in the control line, which confirm that the sd locus move toward the heterochromatin region in the Fab-X2L; Fab-7' line. The same phenomena was observed for both the distal and the proximal loci, indicating that these loci move also toward the heterochromatin region in the Fab-X2L; Fab-7' line. However, the relative 3D distances between sd and the distal locus, or sd and the proximal locus do not significantly change in the Fab-X2L; Fab-7' line compared to the Fab-2L; Fab-7' control line. All together, these results demonstrate that the movement of the chr.X, dependent on the presence of two Fab-7 sequences, is not restricted to the sd locus, but involves a large chromosomal domain of at least 4Mbp. Experiments are now in progress to define if the movement is restricted to 4Mbp of the chr.X, or if the entire chromosome is changing its nuclear position.

8.3 Simulation

The simulation of an entire genome inside a nucleus involves a greater level of course-graining than for smaller systems, in order to limit the computational

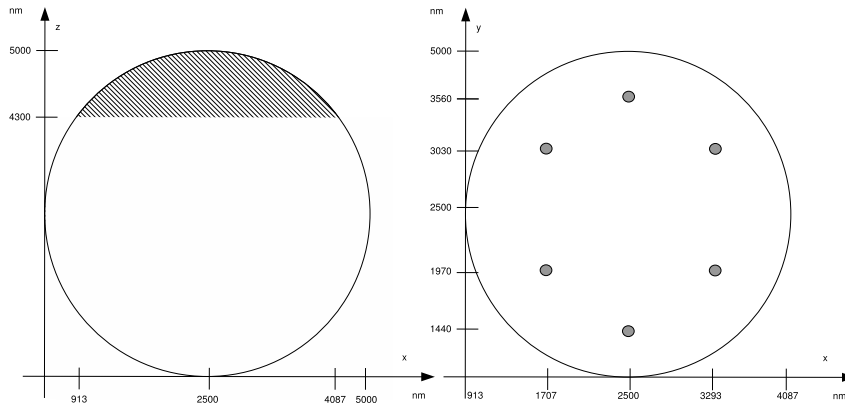


Figure 8.4: Geometry of the simulation. On the left one sees the impenetrable heterochromatin as a shaded area. On the right one sees the location of the grafting points of the individual chromatids.

complexity to a reasonable level. The drosophila nucleus was modelled as follows. The nucleus has a diameter of $5\mu m$. The heterochromatin occupies an impenetrable spherical cap of $700nm$ height. The complete geometry of the simulation is depicted in Fig. (8.4). The four chromosomes have the following characteristics:

- Chromosome X: One arm, 22Mbp
- Chromosome 2: Two arms, 2L (Left arm) 23Mbp and 2R (Right arm) 21.4Mbp
- Chromosome 3: Two arms, 3L of 24.4Mbp, 3R of 28Mbp
- Chromosome 4: One very short arm of 1.2Mbp

The ends of the chromosomes are anchored at the heterochromatin, which is modeled as a short range repulsive wall. The grafting points of the chromosomes are illustrated in Fig. (8.4). The chromosomes themselves are modelled as chromatin fibers. If one considers the DNA to be naked:, then 1 nucleotide = $0.33nm$, 60kbp (60.000 nucleotides) will then represent approximately $20,000nm$, 20Mbp will represent 6.600 micrometers. However, the DNA is not naked in the interphasic nucleus since one has the histones that form the nucleosomes that are the functional unit of the chromatin. So considering the $11nm$ chromatin fiber then 200 nucleotides = $66nm$ of naked DNA, but in this case, 60kbp will represent approximately $3300nm$, the degree of compaction thus being 6 to 11. If one considers the $30nm$ chromatin fiber, then 60kbp will represent approximately $1000nm$, the degree of compaction then being 35.

Therefore I will model the chromosomes as a $30nm$ chromatin fiber, then 20Mbp will measure 333 micrometers backbone length. The segment length is

taken to be 1 micrometer. This is about a factor of 6 too large compared to measurements of the persistence length of the $30nm$ fiber, but allows simulation within a reasonable time span. The potential parameters for the simulated $30nm$ fiber in this case are as follows:

- Segment diameter: $30nm$
- Segment length: $60kbp = 1000nm$ each
- The harmonic bond potential is taken to be

$$U_s(l) = \frac{kT}{2\delta^2}(l - l_0)^2 \quad (8.1)$$

with $\delta = 0.1$ and $l_0 = 300nm$ at $310.15K$.

- The angular and torsional potentials are taken to be 0. On this scale the chain is flexible.
- Repulsive segments potential

$$U_{rep}(r) = \epsilon \left(\frac{\sigma}{r - r_{segment}} \right)^6 \quad (8.2)$$

with $\epsilon = 0.14k_bT$ at body temperature, $\sigma = 15nm$ and $r_{segment} = 15nm$ being the fiber radius.

- Cutoff for the repulsive potential is $r_c = 8nm$ (after the $30nm$ fiber diameter).

As a first step only repulsive segmets are used, therefore producing fibers which only interact via excluded volume effects. At this first step no attractive segments are taken into account because the larger cutoff for the Lennard-Jones segments dramatically augments the computation time. The equilibration of a starting configuration was quite cumbersome. Fig. (8.5) shows such an equilibrated starting configuration. It is clear that there are a lot of interacting fibers inside the spherical nucleus. Randomly distributing grafted chains inside such a spherical volume turned out to be impractical. With approximately 2.000 segments inside the sphere, a random starting configuration always produced several monomers that were extremely close or even intersecting. No matter how slowly the potentials were turned on, the simulation was not stable. Therefore I resorted to the following procedure. First the chromosomes were set up as straight chains, orthogonal to the grafting sites. These straight chains were then equilibrated without the presence of a constraining nuclear envelope. After they were equilibrated, the nuclear envelope was slowly added. The nucleus is modelled as a radial force field, which acts inward with a cutoff at $R_0 = 2500nm$ from the center, i.e. inside this radius there is no inward pushing force. Outside of this radius there is a force proportional to $R - R_0$, allowing

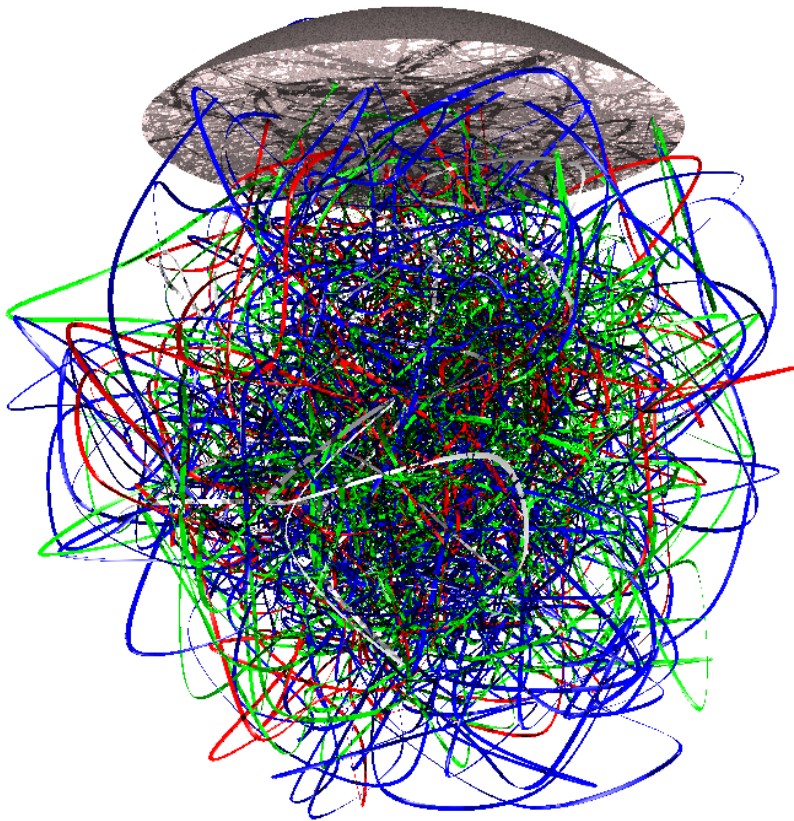


Figure 8.5: The simulated drosophila nucleus. The different colors indicate the different chromosomes. The heterochromatin is shown as the solid spherical cap, the nuclear envelope is not drawn.

a little elasticity of the nucleus. After every chromosome was contained inside the nucleus, the obtained configuration was used as a starting configuration for the simulation runs. With this procedure about 500 independent configurations were produced.

The genomic contents of the individual chromatids is as follows:

- Chromosome X: 22Mbp
- Chromosome 2L = 23Mbp
- Chromosome 2R = 21.4Mbp
- Chromosome 3L = 24.4Mbp
- Chromosome 3R = 28Mbp
- Chromosome 4 = 1.2Mbp

where L and R indicate the left and right homologues.

The experimental distance markers are then implemented at the following locations:

- Locus sd (13F): at 6.35Mbp from the grafting point of chromosome X (approximately one third of the chromosomal length from grafting point)
- Locus H (16D): at 4.35Mbp from the grafting point of chromosome X
- Locus G (12DC): at 8.35Mbp from the grafting point of chromosome X
- Histone Locus (38F): 4Mbp from the grafting point of chromosome 2 (approximately one sixth of the chromosomal length from the grafting point)

8.4 Results

The goal of this chapter is to see how the model compares to the experimental distance measurements. As a first step only excluded volume interaction was taken into account, therefore the model is extremely simplified. Nonetheless it serves as a reasonable starting ground, in order to see what effects can be explained just by excluded volume interaction. Since the simulational procedures are quite complex, I have compared the excluded volume simulation to a simulation without excluded volume, in order to see how prominent the excluded volume effect is in this constrained volume. Fig. (8.6) shows the excluded volume and the phantom chain case. As expected the excluded volume produces a swelling, as can be seen by the shift to the right.

Obviously, one is interested how the simulated distance distribution of the excluded volume case compares to the experiment. In Fig. (8.7) to Fig. (8.11)

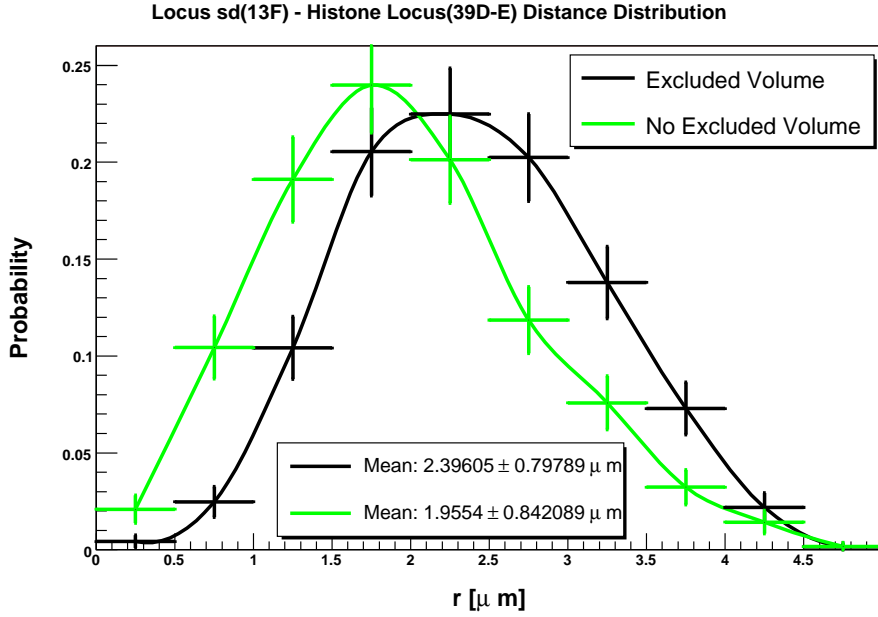


Figure 8.6: Simulation of the Locus-Histone distance with and without excluded volume.

one sees the aforementioned distances 1) - 5) both the experimental and the simulational ones. The errors in the r direction are due to the binning. In order to compare the simulation with the experiment, I have used the same number of bins. Therefore the error in r direction of $0.5\mu m$ is caused by using 10 bins. The plots also include the case with two copies of Fab-7, which causes a contraction as shown by the experimental distances. The green (experiment) and black (simulation) curves of Fig. (8.7) to Fig. (8.9) are already quite close to each other in width, height and mean. This means that the organization of the chromosomes in the drosophila nucleus is almost like a random chain with excluded volume and therefore has no significant higher order structure. In principle all interchromosomal distances are compatible with random packing. The slight offset to the left could be explained by the formation of very loose chromosome territories. This could be achieved by a sparse distribution of attractive sites along the chromosomes.

Fig. (8.10) and Fig. (8.11) show the intrachromosomal distances. Here one finds a quantitatively and qualitatively significant difference. The experimental distance distribution is much more compact than the one yielded by the simulation. However, one must keep in mind that the coarse graining for these simulations was quite extensive. In order to be able to calculate the entire nucleus on a small Linux cluster (25 nodes), I set the Kuhn length of the chromatin fiber to $1000nm$. Therefore the resulting fiber is quite stiff. A more realistic Kuhn length would be a factor of 10 smaller.

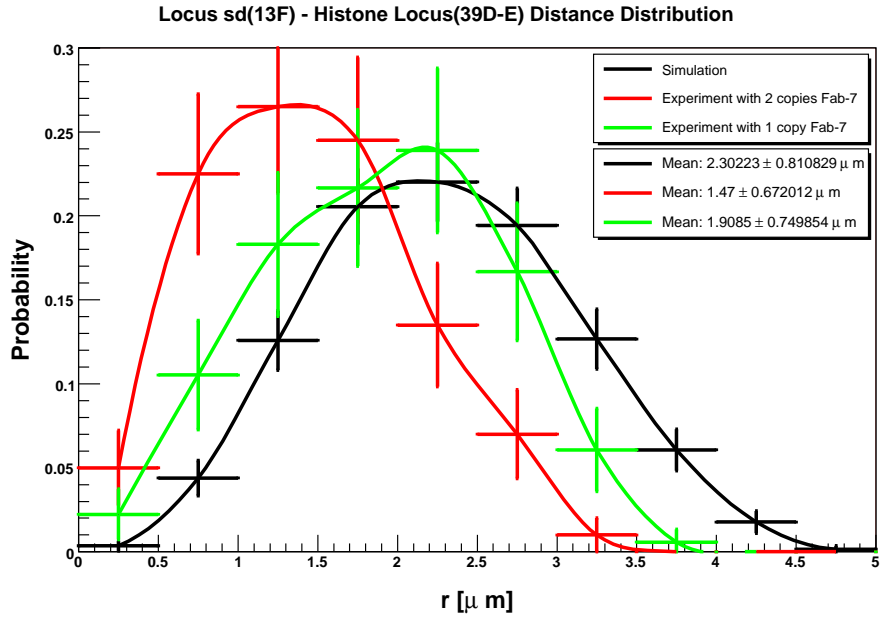


Figure 8.7: Interchromosomal Histone Locus - Locus sd experimental and simulational distances with excluded volume.

Another possibility for the intrachromosomal compaction is the formation of chromosome territories. As already mentioned, this could be achieved by a sparse distribution of attractive segments along the chromatids. If these attractive sites are specific to the individual chromosomes, then the attraction would be mutually exclusive, resulting in the formation of individual chromosome territories. The size and interpenetrability of these territories would be governed by the number and distribution of the attractive sites, much like in chapter 3.

If one closely examines the plot of the excluded volume runs and the phantom chains (Fig. (8.6)) and compares it to Fig. (8.7), one notices a possible agreement between the experimental run with one copy Fab-7 and the phantom chain run. In order to investigate this further, I have plotted all experimental distances against the phantom chain runs. These can be seen in Fig. (8.12) to Fig. (8.16). Indeed the quality of the agreement is striking for all interchromosomal distances. If one thinks about this seriously, then the only possible explanation is that topoisomerases do play a significant role in this stage of the cell cycle [86].

Topoisomerases possess the amazing feature of being able to cut and mend the chromatin fiber so that topological inconveniences can be corrected. If it is not only used for transcription and replication, but actually functional throughout the entire cell cycle, The interpenetration of the chromosomes could be far greater than expected. This would also facilitate the attractive process of the experiment with two copies of Fab-7, since the interchromosomal mobility would

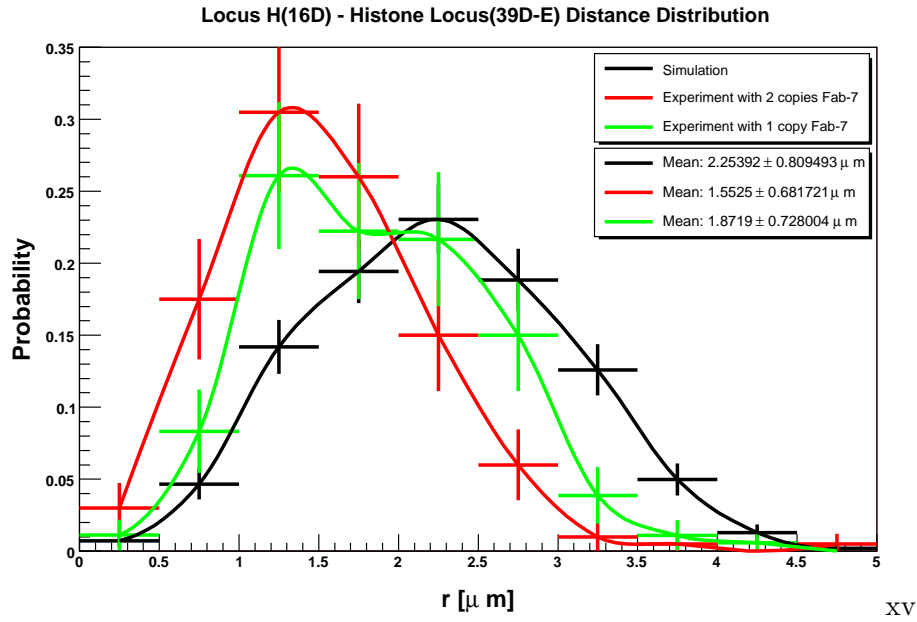


Figure 8.8: Interchromosomal Histone Locus- Locus H experimental and simulation distances with excluded volume.

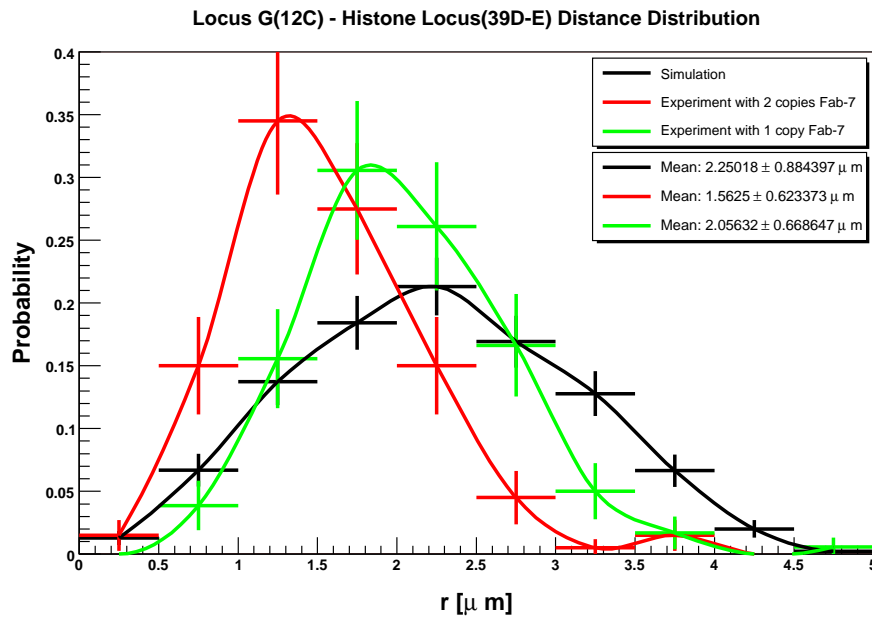


Figure 8.9: Interchromosomal Histone Locus- Locus G experimental and simulation distances with excluded volume.

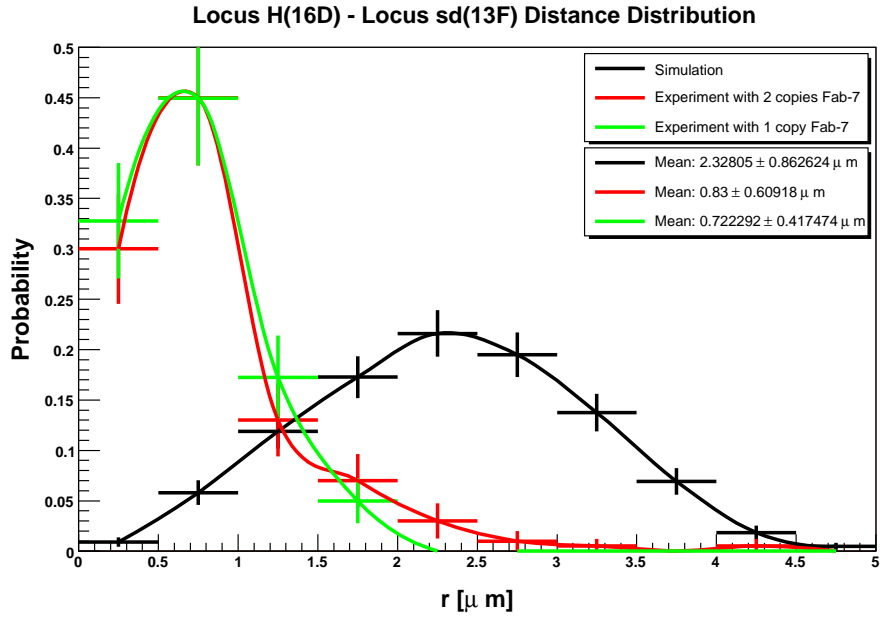


Figure 8.10: Intrachromosomal Locus H - Locus sd experimental and simulation distances with excluded volume.

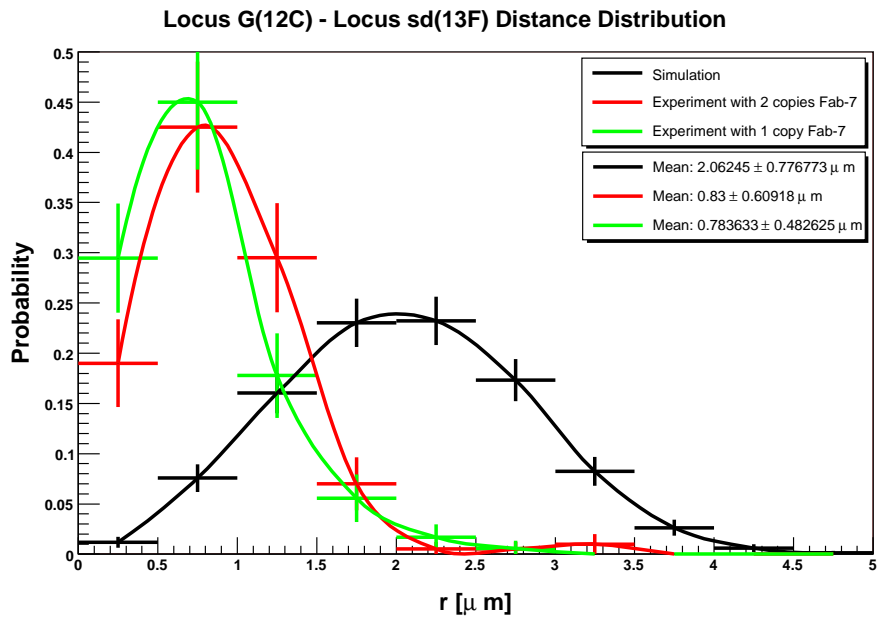


Figure 8.11: Intrachromosomal Locus G - Locus sd experimental and simulation distances with excluded volume.

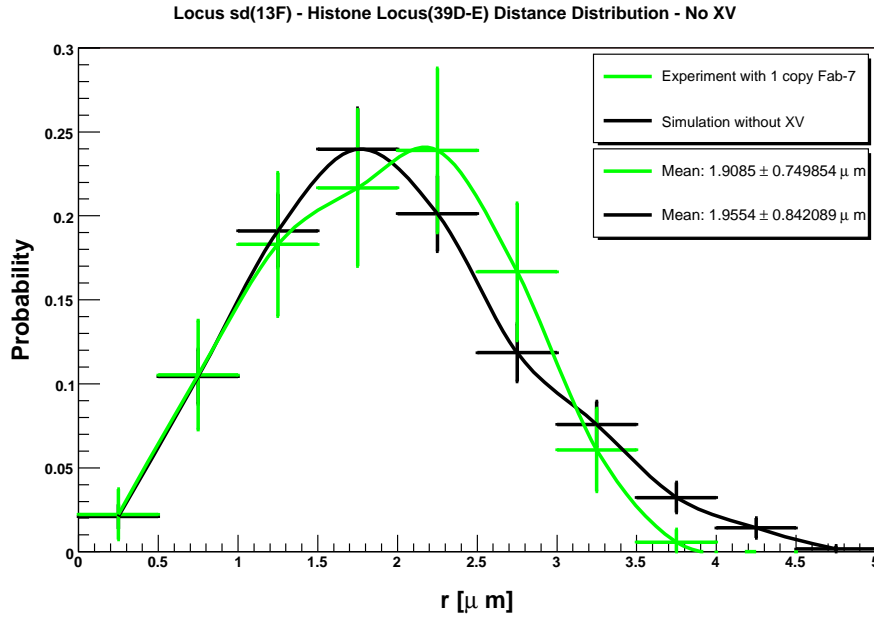


Figure 8.12: Interchromosomal Histone Locus - Locus sd experimental and simulational distances without excluded volume interaction.

be far greater. The intrachromosomal distances, however do not improve significantly, suggesting that territory formation is indeed a possible explanation. This is not a contradiction, because individual chromosomes could form territories and at the same time penetrate territories from other chromosomes more easily.

In summary I would conclude that the phantom chain simulation is indeed realistic, since the agreement with the experimental one copy of Fab-7 run is excellent within errors for all interchromosomal distances. As stated this could be explained by Topoisomerases. This would also explain why the Locus sd approaches the histone locus relatively easy in the experiment with two copies of Fab-7. Since shutting off the excluded volume causes the chromosomes to penetrate each other but does not soften the chain, for it does not reduce the segment length, it is not surprising that the intrachromosomal distances do not improve. Therefore one can conclude that active Topoisomerases could be a major factor in the organization of the nuclear structure in drosophila.

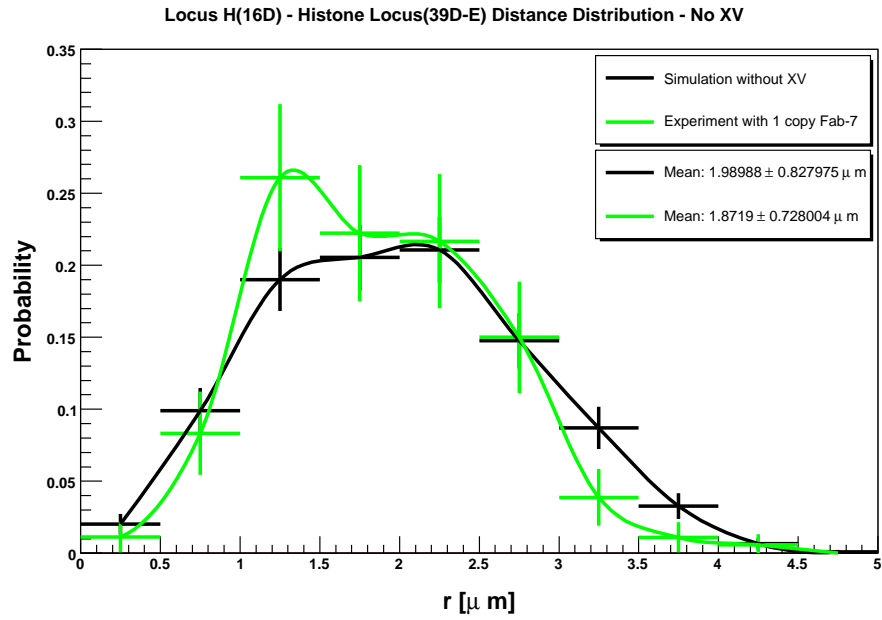


Figure 8.13: Interchromosomal Histone Locus - Locus H experimental and simulational distances without excluded volume interaction.

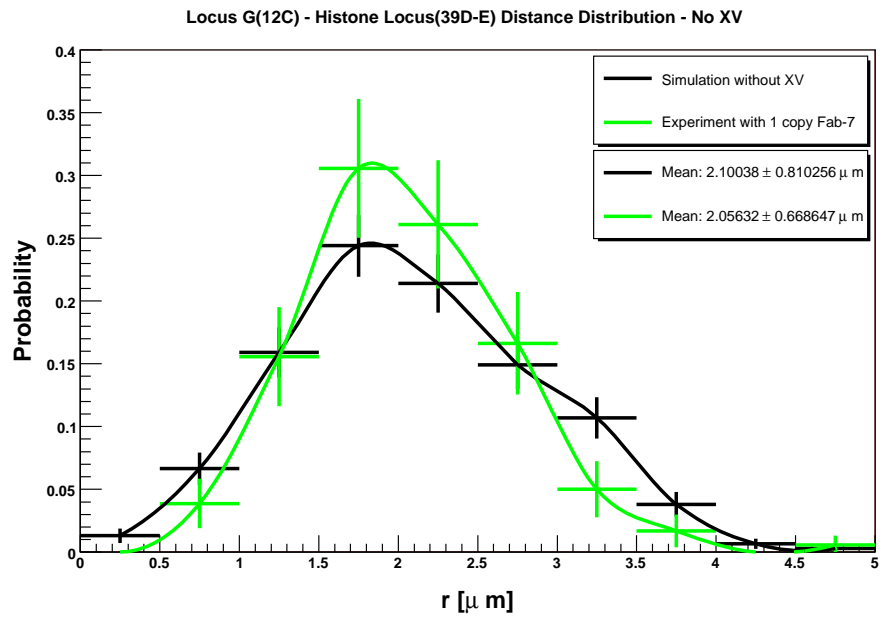


Figure 8.14: Interchromosomal Histone Locus - Locus G experimental and simulational distances without excluded volume interaction.

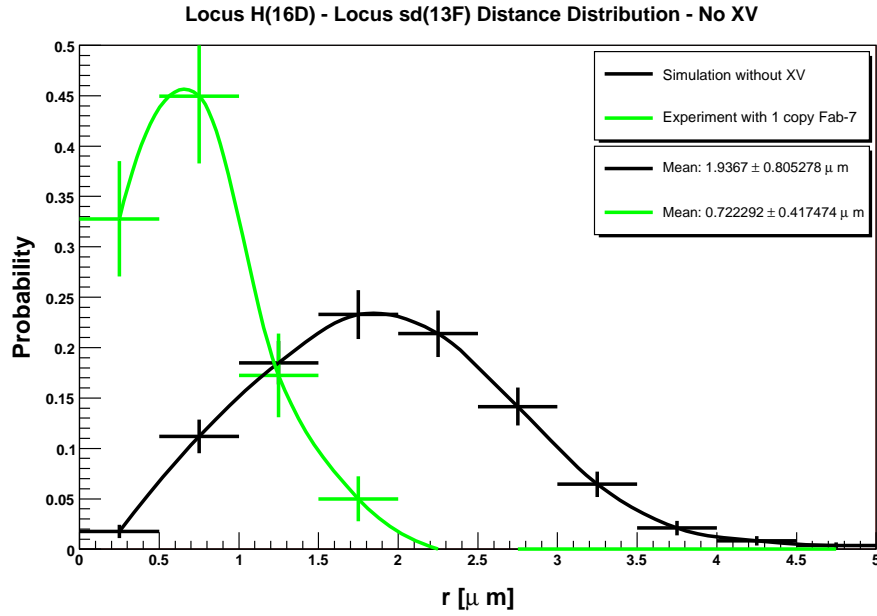


Figure 8.15: Intrachromosomal Locus H - Locus sd experimental and simulation distances without excluded volume interaction.

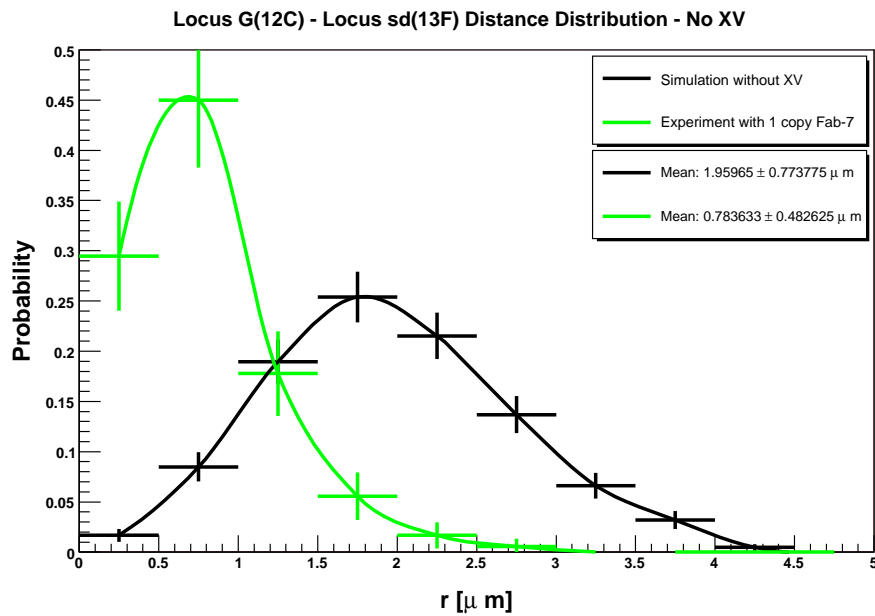


Figure 8.16: Intrachromosomal Locus G - Locus sd experimental and simulation distances without excluded volume interaction.

Chapter 9

Scientific Computing

I would like to briefly address the simulation program in the main part of this dissertation. An short overview, I feel is appropriate since all of the simulations were done with it. Most of the details can be found in the appendix.

9.1 History of DePoSiTo

The program DePoSiTo (Dense Polymer Simulation Tool) already has a reputed history and is still extensively used for current simulations. The original version was based on the programs PolyHMC (1994) of Prof. D. W. Heermann and DePoSiTo 1.0 (1995) of Dr. K. Zimmer. It has since been modified and extended by Dipl. Phys. T. Hapke, Dr. A. Linke, Dr. G. Pätzold, Dipl. Phys. T. Wang, Dipl. Phys. N. Goncalves, Dipl. Phys. M. Brill and Dr. Gunther Schöppe in their respective thesis.[87, 88, 89, 90] Those extensions were among others, the possibility of scratching polymer surfaces and quenching simulations.

All of my predecessors have used this program to simulate dense polymer systems in the field of material science. In this thesis I have used the program to simulate biopolymers. Furthermore I have added several functions, such as the two types of Lennard-Jones Potentials, the confinement to a spherical volume and the grafting to a surface.

9.2 Parallelization

As the computational demands grew due to increasingly larger systems it became necessary to use high performance parallel computers. The technical parallelization of the code was done by R. Reilen-Russ and M. Schwind in G. Rüngers group at the TU Chemnitz.

To develop efficient parallel applications which are adaptable to the computing power and communication performance of a specific parallel platform we use a two level model (TwoL model). It requires an application to be partitioned into separate modules which can be executed consecutively or concurrently on

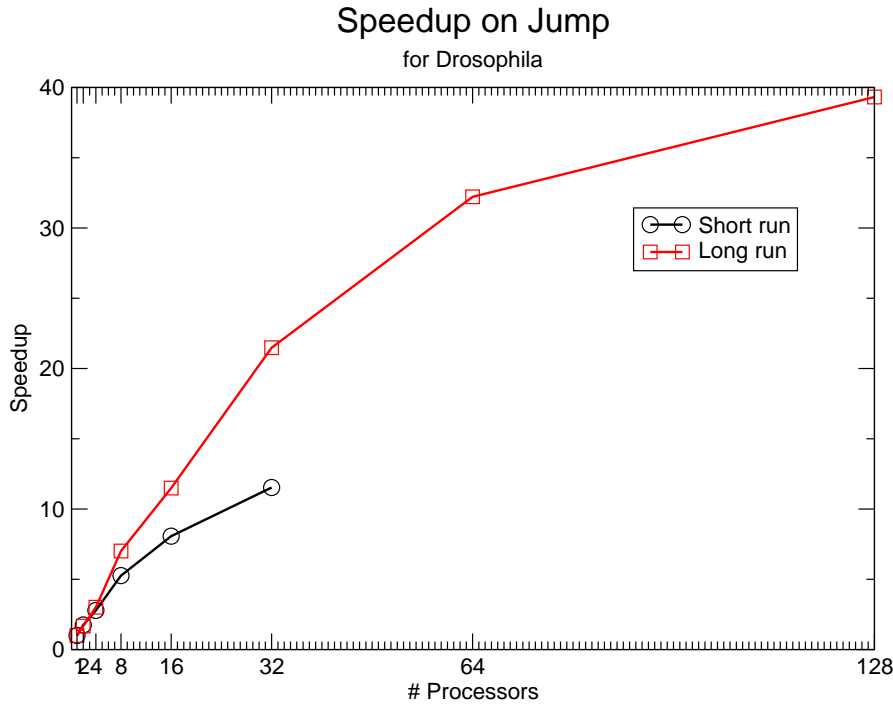


Figure 9.1: Speedup on Jump

groups of processors. On the upper model level dependence relations between modules are specified. Such a specification is transformed stepwise into a coordination code which controls the execution of module calls. This also includes the construction of processor groups as well as organizing data copying and redistributions. The lower model level provides different module versions to support different data distributions and kinds of parallelism. To assure the derivation of an efficient coordination code the transformation process is based on cost models.

The model is currently implemented as a compiler tool suite consisting of several components. Each component realizes a specific model part and provides well-defined interfaces for tool and user interactions. It supports the generation of different coordination codes for a specified algorithm to exploit the features of different parallel platforms.

The first platform the program was tested on was the IBM p690-Cluster Jump at the Forschungszentrum Jülich. Fig. (9.1) shows the speedup for up to 128 Processors. The difference in the two runs is mainly the computational time allotted for each processor before an I/O task. Obviously the less I/O tasks the more efficient the parallel computing becomes. For the Jump runs, a configuration from the drosophila simulation was used with about 2.000 monomers.

Fortunately we also succeeded in obtaining computation time on the IBM Blue Gene/L supercomputer at Jülich with a total of 8192 Processors. Due

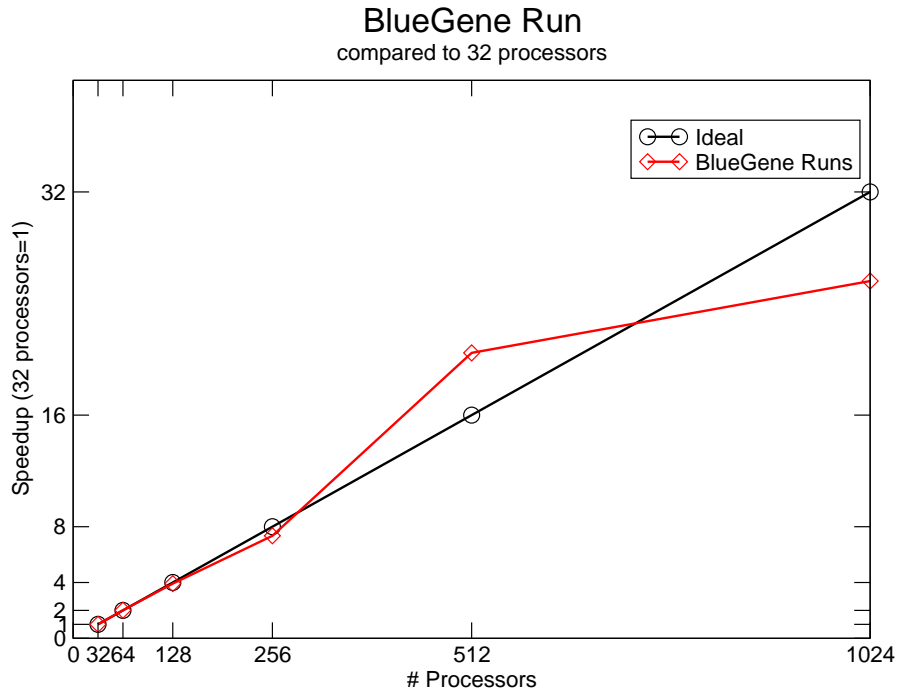


Figure 9.2: Speedup on BlueGene compared to 32 processors

to the improved technical design and infrastructure, a much better scaling was achieved (see Fig. (9.2)). The superlinear speedup visible at 512 processors is probably due to cache-effects. Processor caches have much less latency and higher bandwidth than ordinary memory. With an increasing number of processors more local data fits into the caches and no slow memory access is necessary. Also 512 processors make up one complete rack and a complete rack has a different geometry. Using 512 processors the geometry is a torus, anything below 512 processors is implemented as a mesh. For the Blue Gene runs, a configuration from the drosophila simulation was used with about 20.000 monomers, a factor of 10 larger than on the Jump. Not only is the scaling much better on the Blue Gene, but also due to the increased number of monomers we were able to bring down the simulational persistence length to match the experimental persistence length.

Chapter 10

Conclusion

The aim of this thesis was to see how far a simple polymer model of chromatin would be valid. Using only an A_nB_m block copolymer model with attractive A and repulsive B segments I was able to show that this leads to secondary structure formation in the form of rosettes. These rosettes were thoroughly investigated. The number of attractive sites was analyzed with respect to the rosette diameter and it was found that 6 to 16 loops of 80kbp to 200kbp produced 1Mbp rosettes of about $300 - 800nm$. Diffusion was also investigated in rosettes of these sizes and it was found that all of the simulated substances diffuse regularly, so the rosette-structure cannot cause anomalous diffusion. Furthermore the accessibility was studied and I found out that $20 - 27nm$, i.e. 6kbp to 12kbp of DNA are absolutely inaccessible, depending on the substance. This is direct evidence that structural properties can regulate gene expression and silencing.

The next step was to simulate larger structures. Comparing longer simulated domains to two regions of about 3Mbp of the human chromosome 1 yielded astounding agreement. I have shown that the ridge data corresponds to a free polymer chain without any higher order structure beyond the $30nm$ chromatin fiber. The anti-ridge data corresponded to a ratio of attractive segments of about 30% and thus loop sizes of roughly 120kbp. This is exactly the loop size suggested in the first part of my thesis.

Using the same principles as before, namely only the A_nB_m block copolymer model, I then simulated an entire human chromosome. The chromosome condensed to about $3.2\mu m$, which is a factor of 1.6 too large compared to experimental values. However, the condensation occurred just by the presence of the attractive sites. No external pressure was imposed on the simulated chromosome. I therefore find it reasonable to assume a greater compaction when the chromosome is surrounded by other chromosomes and bound by a nucleus. Thus the factor of 1.6 may be explained by the lack of external pressure.

Finally I have simulated the entire genome of *drosophila melanogaster*. The most outstanding result of this simulation was that the experimentally mea-

sured distances were in good agreement with a simulation without excluded volume interaction. Therefore Topoisomerase, which cuts and mends the chromatin fiber, could be a major factor in chromosomal interactions.

In summary I conclude that the simple polymer model is justified and it explained the structure of 1Mbp domains and the accessibility thereof as well as the compaction of larger sequences up to an entire chromosome. The second major discovery is that the role of Topoisomerase may have to be redefined from merely a helpful tool for unknotting fibers to a major player in chromosomal organization in *drosophila*.

Future simulations could use the mapping I have provided, relating the transcriptional activity to the percentage of attractive sites. In this way, an entire human chromosome, indeed even the entire human genome could be properly mapped. It would be extremely interesting to see how this mapping relates to experimentally measured chromosome territories. Furthermore, if excluded volume effects are negligible in dense biological systems due to Topoisomerases, then much more efficient code could be implemented for simulating them. The present approach originates from polymer physics, where the substances can be isolated and are not as interdependent of their surroundings. In living systems there is an inherent intelligence which one needs to incorporate into ones models. The astounding result is, that this could make simulations simpler.

Acknowledgements

This dissertation could not have been written without Prof. Dr. D. W. Heermann and Prof. Dr. C. Cremer who not only served as my supervisors but also encouraged me throughout my academic program. Prof. Heermann showed a great level of understanding and support for personal and professional matters, never accepting less than my best efforts. Prof. Cremer proved to be a master of motivation by emphasizing the biological impact of my work. I would also like to thank Dr. Gregor Kreth, who patiently guided me through the dissertation process and my sister for having proof-read the manuscript. Above all I thank my wife and kids for having endured this time of little luxury. And of course also my parents, who made me who I am today. I thank you all.

Appendix A

The simulation program DePoSiTo

In this appendix I would like to mention the technical details of the simulations so that they can be reproduced. Since the program is not open source, please contact Prof. Heermann for permission to use or access the program. The deposito simulation framework consists of three parts, which are briefly explained in the following sections.

A.1 Setup

The setup program of the deposito framework requires 2 ASCII files: mapping and ini files. The mapping files contain substance specific parameterizations. The example shown here is from the drosophila simulation. The first one is for repulsive segments, the latter for attractive segments.

```
#mapping for drosophila - Date: 21.07.2005   Time: 10:50 Uhr
```

```
MOVEPOINTS
```

```
NUMBER 2
```

```
0  CLU 000
```

```
1  CLU 000
```

```
INTERPARAM
```

#					sigma	Epsilon	Rcoeff	rel_hc
#	ABSF	FORCE	UNIT	INDICES	kcal/mol	Angstroem	Angstroem	
#					-----	-----	-----	-----
ELLIPSOID	R6	0	1	1.0	1.0	1.5333	0.6522	1.0
							0	GAUSS
								10

```
BOND
```

```
#          E = 0.5 * K2 * (R - R0)^2
```

```
#           R0           K2
#           -----
0 1 HARMONIC 1000.0 100.0
```

ANGLE

```
#           E = 0.5 * K2 * (cos(Theta) - cos(Theta0))^2
#           cos(Theta0)           K2
#           -----
```

TORSION

```
#           E = SUM(n=1,3) { V(n) * [ 1 + cos(n*Phi - Phi0(n)) ] }
#           Phi0 = 1 or -1 {for 0 or 180 degree)
#           V1           Phi0           V2           Phi0
#           -----
```

PREVMONOMERBOND

```
#           E = 0.5 * K2 * (R - R0)^2
#           R0           K2
#           -----
```

PREVMONOMERANGLE

```
#           E = 0.5 * K2 * (cos(Theta) - cos(Theta0))^2
#           cos(Theta0)           K2
#           -----
```

PREVMONOMERTORSION

```
#           E = 0.5 * K2 * (cos(Theta) - cos(Theta0))^2
#           cos(Theta0)           K2
#           -----
```

```
#end
```

```
#mapping for drosophila - Date: 21.07.2005   Time: 10:50 Uhr
```

MOVEPOINTS

```
NUMBER 2
0 CLU 000
1 CLU 000
```

INTERPARAM

```
#           sigma           Epsilon           Rcoeff           rel_hc
#           ABSFORCE UNITINDICES kcal/mol      Angstroem      Angstroem
#           -----
ELLIPSOID  LJ  0 1  2.0 500.0 153.33 0.006522 1.0 0 GAUSS 10
```

```

BOND
#          E = 0.5 * K2 * (R - R0)^2
#          R0          K2
#          -----
0 1 HARMONIC 1000.0 100.0

ANGLE
#          E = 0.5 * K2 * (cos(Theta) - cos(Theta0))^2
#          cos(Theta0)      K2
#          -----

TORSION
#          E = SUM(n=1,3) { V(n) * [ 1 + cos(n*Phi - Phi0(n)) ] }
#          Phi0 = 1 or -1 {for 0 or 180 degree}
#          V1      Phi0      V2      Phi0
#          -----
PREVMONOMERBOND
#          E = 0.5 * K2 * (R - R0)^2
#          R0          K2
#          -----

PREVMONOMERANGLE
#          E = 0.5 * K2 * (cos(Theta) - cos(Theta0))^2
#          cos(Theta0)      K2
#          -----

PREVMONOMERTORSION
#          E = 0.5 * K2 * (cos(Theta) - cos(Theta0))^2
#          cos(Theta0)      K2
#          -----

#end

```

The ini file contains information for the simulation infrastructure, such as simulation box dimensions, temperature, confinement and many more. The following ini file from the drosophila configuration is somewhat typical.

```

Program_name    deposito
Format          DePoSiTo2
Creation_date    18-Jan-06_16.24

```

```

Substance_num    1
Substance_type0  rep
Substance_prob0  1

```

```
Number_chains      6
Longest_chain      450
Number_monomers    1920
Number_movepoints  2
Branching          0
Phi                0
Side_x             5000
Side_y             5000
Side_z             5000
Tool_type          None
Support            0
Trough             0
Plates             0

Microarray         1
Microarray_cutoff  1
Microarray_epsilon 1
Microarray_sigma   1
Microarray_bottom  700
Microarray_displacement 0
Toparray           0
Confinement        1
Confinement_strength 1e-06
Confinement_SQradius 6.25e+06
Confinement_init   0

Seed               531477613
Algorithm          MD_VS
Brown_Drag         0
Brown_Rand         0
Temperature        3000
Pressure           0
Thermostat_Q       0
Pressure_Mass       0
Steps_prev         0
Steps_MD           2000
Steps_MC           10000
Steps_NPT          1
Ana_sweeps         1
Save_sweeps        1
New_veloc          1
Int_stepsize       0.01
Verlet_skin        0.1
Relax_cutoff       0.1
Boxsize            5000
ENDPARAM
```

The setup program takes these two files and generates a starting configuration. If the simulated system is very complex, a manual construction of a starting configuration may be necessary.

A.2 DePoSiTo

The `deposito2005` program is the heart of the simulation package. It allows molecular dynamics and dissipative particle dynamics simulations. The dissipative particle dynamics was used for the diffusion of the transcription factors. It produces files like the ini files but with $6N$ additional numbers for the location and momentum of the N simulated particles.

A.3 Analysis

The analysis was done using C-routines that I have implemented myself. I have added them to the existing `deposito` analysis framework. The plots were done with `xmgrace` and ‘ROOT’, developed by CERN and available at <http://root.cern.ch>.

Appendix B

Publications

Major parts of this thesis have already been published in scientific journals and presented at international conferences. Below is a list of publications as well as the conferences at which the research was presented.

B.1 Publications

- S. Ritter, J. Odenheimer, D. W. Heermann, F. Bantignies, C. Grimaud and G. Cavalli,
Modeling of Polycomb-dependent chromosomal interactions involved in Drosophila gene silencing,
To appear in Biophys. Rev. Lett. Vol. 1, No. 2 (2006)
- J. Odenheimer, G. Kreth and D. W. Heermann,
Dynamic Simulation of Active/Inactive Chromatin Domains,
Int. J. Biol. Phys. 31, No. 3-4 (2005)
- J. Odenheimer, D. W. Heermann, and M. Brill,
Forces by and on a polymer grafted to a repulsive wall,
Int. J. Mod. Phys. C Vol. 16, No. 10 (2005)
- J. Odenheimer, G. Kreth, D.W. Heermann, R. Martin, C. Cardoso,
Accessibility and diffusion of transcription factor complexes into 1Mbp chromatin domains,
Eur. Biophys. J., 34, 637 (2005) (Abstract)
- J. Odenheimer, G. Kreth and D. W. Heermann,
Dynamic Simulation of Active/Inactive Chromatin Domains,
Europhysics Conference Abstracts, Vol. 28D (2004) (Abstract)

B.2 Conferences

- 09/2005: International Biophysics Congress, Montpellier, France
Poster presentation on '*Accessibility and Diffusion of Transcription Factor Complexes in 1Mbp Chromatin Domains*'
- 09/2004: Conference on Computational Physics, Genoa, Italy
Poster presentation on '*Dynamic Simulation of active/inactive Chromatin Domains*'
- 08/2004: International Conference on Biological Physics, Gothenburg, Sweden
Talk on '*Dynamic Simulation of active/inactive Chromatin Domains*'

List of Figures

1.1	The levels of chromatin compaction. Image source: NIH Talking Glossary of Genetics	12
1.2	The cell cycle of a typical mammalian cell. A new cell can pass through G1, S and G2 before it divides at mitosis (M) or it can exit from the cycle by going into a resting state G0. The possibility of apoptotic death is not shown.	14
2.1	Left: Course-grained model of a chromosome. A typical bead is about $300 - 800nm$ in diameter, the linker segment length is around $300nm$ and consists of 30kbp. Chromosome 1 has approximately 245 such coils. Right: Detailed structure of a bead. The 10 loop model suggests a rosette like structure.	18
2.2	After the course graining complex and dense structures may be simulated.	19
2.3	An ellipsoid with an intrinsic spring. The spring symbolizes the harmonic bond potential.	20
4.1	Left: The starting configuration of a 60 segment chromatin fiber. The spheres represent the condensing agents. Center: An intermediary configuration. This state of mainly two clusters of approximately equal size turns out to be a metastable state. Right: In the final state all attractive segments are concentrated in the center. A rosette has formed.	37
4.2	Shown is the average formation time of a rosette. From the minimal distance one can conclude that two segments ‘snap’ very fast. After about 10,000 MD Steps a complete rosette is formed.	38
4.3	The diameter of the rosettes depends on the number of attractive sites. The different Kuhn lengths are only of minor importance.	39
4.4	Left: Projection of the average monomer concentration of a rosette with a diameter of about $600nm$. Right: Projections of virtual microscopy image data stacks of one simulated rosette (before convolution with a measured confocal point spread function (bottom) and after convolution (top)).	40

4.5	Size determination of a 180kbp gene-like segment. The left shows the simulation data and the right side the SMI-PSF fit.	41
4.6	A virtual SMI image is shown on the left, the units are in $10nm$ steps. The right hand side shows the axial extension histogram. .	42
4.7	Shown is the spatial vs. the genomic distance of a 1Mbp domain. The exponents yielded by the fits clearly favor about 11 attractive sites per 1Mbp domain. The oscillations in the data are due to the rosette nature of the 1Mbp domain.	42
4.8	Shown is the total energy with respect to the number of simulation steps. The red line represents the average of 100 runs. Only a few individual runs are shown for clarity.	44
4.9	A two-dimensional probability distribution of the total energy with respect to time.	44
4.10	Shown is the free energy with respect to the number of simulation steps. One sees a distinct metastable state and the final state of lowest free energy.	45
5.1	Cy5 fluorescently labelled complexes of Streptavidin (SAv) were microinjected into the cytoplasm. (A) A cell immediately after microinjection, (B) The same cell as in A) but 12 min later, the complexes have been transported into the nucleus, (C) A cell with aggregations of MeCP2 labelled centromeric heterochromatin structures (C. Cardoso, unpublished data).	48
5.2	The linescan analysis (D) shows that the concentration of the NLS-SAv-Cy5 complexes is decreased in the nucleoli and that some decrease of the complex concentration can also be seen in MeCP2 labelled heterochromatin (C. Cardoso, unpublished data). .	49
5.3	The simulated substances shown are not drawn to scale.	50
5.4	The diffusion constants of all the substances vary only slightly with respect to the intra-rosette-spacing. Shown is the diffusion of Streptavidin.	52
5.5	All substances show regular diffusion. The diffusion constants are stated for a intra-rosette-spacing of $27.3nm$	52
5.6	Mean square displacement divided by t for the first 0.006 seconds. Note that the time scale is $1/100$ of the total simulated time.	53
5.7	The probability density of the Streptavidin from the center-of-mass. In the inner core of the rosette the density is lower than that of the water-only simulation. Beyond about $180nm$ the density is equal to the bulk density.	54
5.8	The cumulative genome content with respect to the center of the rosette. One can observe 3 regions of different accessibility. . . .	56

5.9	The relative density from the center of the $D_{27.3nm}$ rosette. As one approaches the inner core, the density continually decreases. The fit function and values are stated in the text.	57
6.1	Transcriptional activity of chromosome 1. The abscissa shows the genomic distance and the ordinate the transcriptional activity. The ridges are areas of high gene density whereas the anti-ridges are such of low gene densities. (Courtesy of J. Koster and R. Versteeg, Academic Medical Center, University of Amsterdam, unpublished data)	60
6.2	Comparison with the simulated rosette sizes for the genomic vs. spatial distance measurements for chromosome1.	63
6.3	Simulation of the genomic vs. spatial distance for different numbers of attractive sites for chromosome 1. The errors are omitted for clarity, they are approximately 10%.	64
6.4	Genomic vs. spatial distance measurements and simulations for chromosome 1.	65
7.1	The chromatin fiber condensed into the human chromosome 22. The condensing agents are shown as blue spheres.	68
7.2	A virtual light nanoscopy image of a the simulated chromosome. Plotted is the two-dimensional projection of the monomer concentration.	69
7.3	The distribution of the radius of gyration of the chromosome 22.	70
7.4	The temperature dependence of the radius of gyration. At a temperature of about 2300 one sees the transition form the condensed to the decondensed state.	71
8.1	The yellow rectangles represent the position of the Fab-7 element: a) in the Fab-X2L; Fab7' line, there is two copies of the Fab-7 element: one at the sd locus (13F), one at the 38F locus. b) in the control Fab-2L; Fab7' line, there is only one copy of the Fab-7 element: the one at the 38F locus.	75
8.2	3D-FISH experiment in the Fab-2L; Fab-7' control line (top), and the Fab-X2L; Fab-7' line (bottom) mapping the distance between the different loci (top: sd, middle: H, bottom: G) and the pericentric region of chromosome 2 (indicated by 'histone'), since a probe from the histone locus was used.(Roure, Bantignies and Cavalli, unpublished data)	76
8.3	The different loci are repositioned close to the heterochromatin when two Fab-7 transgenes are present at the genomic loci. . . .	77
8.4	Geometry of the simulation. On the left one sees the impenetrable heterochromatin as a shaded area. On the right one sees the location of the grafting points of the individual chromatids. . . .	78

8.5	The simulated drosophila nucleus. The different colors indicate the different chromosomes. The heterochromatin is shown as the solid spherical cap, the nuclear envelope is not drawn.	80
8.6	Simulation of the Locus-Histone distance with and without excluded volume.	82
8.7	Interchromosomal Histone Locus - Locus sd experimental and simulational distances with excluded volume.	83
8.8	Interchromosomal Histone Locus- Locus H experimental and simulational distances with excluded volume.	84
8.9	Interchromosomal Histone Locus- Locus G experimental and simulational distances with excluded volume.	84
8.10	Intrachromosomal Locus H - Locus sd experimental and simulational distances with excluded volume.	85
8.11	Intrachromosomal Locus G - Locus sd experimental and simulational distances with excluded volume.	85
8.12	Interchromosomal Histone Locus - Locus sd experimental and simulational distances without excluded volume interaction. . . .	86
8.13	Interchromosomal Histone Locus - Locus H experimental and simulational distances without excluded volume interaction. . . .	87
8.14	Interchromosomal Histone Locus - Locus G experimental and simulational distances without excluded volume interaction. . . .	87
8.15	Intrachromosomal Locus H - Locus sd experimental and simulational distances without excluded volume interaction.	88
8.16	Intrachromosomal Locus G - Locus sd experimental and simulational distances without excluded volume interaction.	88
9.1	Speedup on Jump	90
9.2	Speedup on BlueGene compared to 32 processors	91

Bibliography

- [1] H. Harris. *The birth of the cell*. Yale University Press, 1999.
- [2] J.D. Watson and F.H.C. Crick. Molecular structure of nucleic acids: a structure of deoxyribonucleic acid. *Nature*, 171:737–738, 1953.
- [3] J.D. Watson and F.H.C. Crick. Genetic implications of the structure of deoxyribonucleic acid. *Nature*, 171:964–967, 1953.
- [4] D.S. Goodsell. *The machinery of life*. Springer-Verlag, 1993.
- [5] R. Oldfield. *Light Microscopy: An Illustrated Guide*. Wolfe, 1994.
- [6] B. Albrecht, A.V. Failla, A. Schweitzer, and C. Cremer. Spatially modulated illumination microscopy: A new approach to biological nanostructure analysis. *G.I.T. Imaging and Microscopy*, 2:40, 2001.
- [7] B. Albrecht, A. Schweitzer, A.V. Failla, P. Edelmann, and C. Cremer. Spatially modulated illumination (smi) microscopy allows axial distance resolution in the nanometer range. *Applied Optics*, 41:80, 2002.
- [8] A.V. Failla, A. Cavallo, and C. Cremer. Subwavelength size determination using smi virtual microscopy. *Applied Optics*, 41:6651, 2002.
- [9] A. V. Failla, B. Albrecht, U. Spoeri, A. Kroll, and C. Cremer. Nanosizing of fluorescent objects by spatially modulated illumination microscopy. *Applied Optics*, 41:7275, 2002.
- [10] S. Martin, A.V. Failla, U. Spoeri, C. Cremer, and A. Pombo. Measuring the size of biological nanostructures with spatially modulated illumination microscopy. *Molecular Biology of the Cell*, 15:2449, 2004.
- [11] C. Wagner, U. Spoeri, and C. Cremer. High-precision smi microscopy size measurements by simultaneous frequency domain reconstruction of the axial point spread function. *Optik*, 116:15, 2005.
- [12] G. Hildenbrand, A. Rapp, U. Spoeri, C. Wagner, C. Cremer, and M. Hausmann. Nano-sizing of specific gene domains in intact human cell nuclei by spatially modulated illumination (smi) light microscopy. *Biophys. J.*, 88:4312, 2005.

- [13] S. W. Hell. Strategy for far-field optical imaging and writing without diffraction limit. *Phys. Lett. A*, 326:140, 2004.
- [14] V. Westphal and S. W. Hell. Nanoscale resolution in the focal plane of an optical microscope. *Phys. Rev. Lett.*, 94:143903, 2005.
- [15] V. Westphal, J. Seeger, T. Salditt, and S. W. Hell. Stimulated emission depletion microscopy on lithographic nanostructures. *J. Phys. B: At. Mol. Opt. Phys.*, 38:695, 2005.
- [16] L. Kastrop, H. Blom, C. Eggeling, and S. W. Hell. Fluorescence fluctuation spectroscopy in subdiffraction focal volumes. *Phys. Rev. Lett.*, 94:178104, 2005.
- [17] R. D. Kornberg and Y. Lorch. Nucleosome positioning. *Nucl. Acids Mol. Biol.*, 7:217–225, 1993.
- [18] B. Dorigo, T. Schalch, A. Kulangara, S. Duda, R. R. Schroeder, and T. J. Richmond. Nucleosome arrays reveal the two-start organization of the chromatin fiber. *Science*, 306:1571, 2004.
- [19] C. Münkler and J. Langowski. Chromosome structure predicted by a polymer model. *Phys. Rev. E*, 57:5888–5896, 1998.
- [20] C. Münkler, R. Eils, S. Dietzel, D. Zink, C. Mehring, G. Wedemann, T. Cremer, and J. Langowski. Compartmentalization of interphase chromosomes observed in simulation and experiment. *J. Mol. Biol.*, 285:1053–1065, 1999.
- [21] P. Cook. *Principles of Nuclear Structure and Function*. Wiley-Liss, 2001.
- [22] A. Bird and A. P. Wolffe. Methylation induced repression - belts, braces and chromatin. *Cell*, 99:451–454, 1999.
- [23] V. Pirotta. Polycomb the genome: Pcg, trxg and chromatin silencing. *Cell*, 93:333–336, 1998.
- [24] S. C. West. DNA helicases: new breeds of translocating motors and molecular pumps. *Cell*, 86:334, 1995.
- [25] D. A. Jackson and A. Pombo. Replicon clusters are stable units of chromosome structure: evidence that nuclear organization contributes to efficient activation and propagation of s phase in human cells. *J. Cell Biol.*, 140:1285, 1998.
- [26] H. Nakamura, T. Morita, and C. Sato. Structural organisation of replicon domains during DNA synthetic phase in the mammalian nucleus. *Exp. Cell Res.*, 165:291, 1986.

-
- [27] P. R. Cook. The organization of replication and transcription. *Science*, 284:1790, 1999.
- [28] D. A. Jackson and P. R. Cook. Replication occurs at a nucleoskeleton. *EMBO J.*, 5:1403, 1986.
- [29] W. F. Marzluff and R. C. C. Huang. *Transcription and Translation: A Practical Approach.*, chapter Transcription of RNA in isolated nuclei. IRL Press, Oxford, 1984.
- [30] D. A. Jackson and P. R. Cook. Transcription occurs at a nucleoskeleton. *EMBO J.*, 4:919, 1985.
- [31] P. H. von Hippel. An integrated model of the transcription complex in elongation, termination and editing. *Science*, 281:660, 1998.
- [32] T. Cremer, G. Kreth, H. Koester, R. H. A. Fink, R. Heintzmann, M. Cremer, I. Solovei, D. Zink, and C. Cremer. Chromosome territories, interchromatin domain compartment and nuclear matrix: an integrated view of the functional nuclear architecture. *Crit. Rev. Eukaryotic Gene Expression*, 12:179–212, 2000.
- [33] K. Rippe. Making contacts on a nucleic acid polymer. *TRENDS in Biochemical Sciences*, 26:733–740, 2001.
- [34] F. Girard, B. Bello, U. K. Laemmli, and W. J. Gehring. In vivo analysis of scaffold-associated regions in drosophila: a synthetic high-affinity SAR binding protein suppresses position effect variegation. *The EMBO Journal*, 17:2079–2085, 1998.
- [35] C. M. Hart and U. K. Laemmli. Facilitation of chromatin dynamics by SARs. *Current Opinion in Genetics & Development*, 8:519–525, 1998.
- [36] C. Maison and G. Almouzni. Hp1 and the dynamics of chromatin maintenance. *Nature Reviews Molecular Cell Biology*, 5:296–305, 2004.
- [37] K. Ishii and U. K. Laemmli. Structural and dynamic functions establish chromatin domains. *Molecular Cell*, 11:237–248, 2003.
- [38] K. Maeshima and U. K. Laemmli. A two-step scaffolding model for mitotic chromosome assembly. *Developmental Cell*, 4:467–480, 2003.
- [39] Y. Blat and N. Kleckner. Cohesins bind to preferential sites along yeast chromosome iii, with differential regulation along arms versus the centric region. *Cell*, 98:249–59, 1999.
- [40] A. G. Cherstvy, A. A. Kornyshev, and S. Leikin. Temperature-dependent DNA condensation triggered by rearrangement of adsorbed cations. *J. Phys. Chem. B*, 106:13362–13369, 2002.

- [41] A. G. Cherstvy, A. A. Kornyshev, and S. Leikin. Torsional deformation of double helix in interaction and aggregation of DNA. *J. Phys. Chem. B*, 108:6508–6518, 2004.
- [42] J. Ostashevsky. A polymer model for the structural organization of chromatin loops and minibands in interphase chromosomes. *Molecular Biology of the Cell*, 9:3031–3040, 1998.
- [43] A. Halperin. On the collapse of multiblock copolymers. *Macromol.*, 24:1418–1419, 1991.
- [44] G. Schöppe and D. W. Heermann. Alternative off-lattice model with continuous backbone mass for polymers.
- [45] D. Frenkel and B. J. Smit. *Understanding Molecular Simulation*. Academic Press, 2 edition, 2001.
- [46] A. N. Semenov, J.-F. Joanny, and A. R. Khokhlov. Associating polymers: equilibrium and linear viscoelasticity. *Macromol.*, 28:1066–1075, 1995.
- [47] A. N. Semenov, I.A. Nyrkova, and A. R. Khokhlov. *Ionomers: Characterization, Theory and Applications.*, chapter Statistics and dynamic of ionomer systems, pages 251–279. CRC Press, Boca Raton, FL, 1996.
- [48] P.-G. De Gennes. *Scaling Concepts in Polymers Physics*. Cornell University Press, Ithaca, 1979.
- [49] A. Y. Grosberg and A. R. Khokhlov. *Statistical Physics of Macromolecules*. AIP Press, New York, 1994.
- [50] U. Buchenau. *Ferienkurs ‘81, Physik der Polymere.*, chapter Konformation I. Kernforschungsanlage Jülich GmbH, 1981.
- [51] P. Flory. *Principles of Polymer Chemistry*. Cornell University Press, 1953.
- [52] Pierre-Gilles de Gennes. *Scaling Concepts in Polymer Physics*. Cornell University Press, 1979.
- [53] A.Yu. Grossberg and A.R. Khokhlov. *Statistical Physics of Macromolecules*. AIP Press, 1994.
- [54] K. Binder and D.W. Heermann. *Monte Carlo Simulation in Statistical Physics*. Springer, 2002.
- [55] H. Schiessel. The physics of chromatin. *J. Phys. Cond. Mat.*, 15:699–774, 2003.
- [56] G. Wedemann and J. Langowski. Computer simulation of the 30-nanometer chromatin fiber. *Biophysical Journal*, 82:2847–2859, 2002.

- [57] D. A. Beard and T. Schlick. Computational modeling predicts the structure and dynamics of chromatin fiber. *Structure*, 9:105–114, 2001.
- [58] J. Bednar, R. A. Horowitz, S. A. Grigoryev, L. M. Carruthers, J. C. Hansen, A. J. Koster, and C. L. Woodcock. 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, 1998.
- [59] H. Schiessel. How short-ranged electrostatics controls the chromatin structure on much larger scales. *Europhys. Lett.*, 58:140–146, 2002.
- [60] H. Schiessel. DNA folding: structural and mechanical properties of the two-angle model for chromatin. *Biophys. J.*, 80:1940–1956, 2001.
- [61] H. Schiessel. Theory and computer modeling of the 30nm chromatin fiber. *New Comprehensive Biochemistry*, 39:397–420, 2004.
- [62] T. Cremer and C. Cremer. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nature Reviews Genetics*, 2:292–301, 2001.
- [63] J. Sedat and L. Manuelidis. A direct approach to the structure of eukaryotic chromosomes. *Cold Spring Harb. Symp. Quant. Biol.*, 42:331–350, 1978.
- [64] P. R. Cook. A chromomeric model for nuclear and chromosome structure. *J. Cell Sci.*, 108:2927–2935, 1995.
- [65] L. Manuelidis. A view of interphase chromosomes. *Science*, 250:1533–1540, 1990.
- [66] G. Li, G. Sudlow, and A. S. Belmont. Interphase cell cycle dynamics of a late-replicating, heterochromatic homogeneously staining region: precise choreography of condensation/decondensation and nuclear positioning. *J. Cell Biol.*, 140:975–989, 1998.
- [67] H. Bornfleth, P. Edelmann, D. Zink, T. Cremer, and C. Cremer. Quantitative motion analysis of subchromosomal foci in living cells using four-dimensional microscopy. *Biophys. J.*, 77:2871–2886, 1999.
- [68] A. Wolffe. *Chromatin: Structure and Function*. Academic Press Inc., San Diego, 2 edition, 1995.
- [69] Y. Saitoh and U. K. Laemmli. Metaphase chromosome structure: bands arise from a differential folding path of the highly AT-rich scaffold. *Cell*, 76:609–622, 1994.

- [70] W. A. Bickmore and K. Oghene. Visualizing the spatial relationships between defined DNA sequences and the axial region of extracted metaphase chromosomes. *Cell*, 84:95–104, 1996.
- [71] H. Yokota, G. van den Engh, J. E. Hearst, R. K. Sachs, and B. J. Trask. 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, 1995.
- [72] A. T. Sumner. *Chromosomes: Organization and Function*. Blackwell Publishing, 2003.
- [73] R. van Driel and A. P. Otte. *Nuclear Organization, Chromatin Structure, And Gene Expression*. Oxford University Press, 1997.
- [74] J. C. Politz, R. A. Tuft, and T. Pederson. Diffusion-based transport of nascent ribosomes in the nucleus. *Molecular Biology of the Cell*, 2003.
- [75] P. J. Verschure, I. van der Kraan, E. M. M. Manders, D. Hoogstraten, A. B. Houtsmuller, and R. van Driel. Condensed chromatin domains in the mammalian nucleus are accessible to large macromolecules. *EMBO reports*, 2003.
- [76] NCBI Web Page. Ncbi human genome resources. <http://www.ncbi.nlm.nih.gov/genome/guide/human/>, Feb. 2006.
- [77] R. Versteeg, B. D.C. van Schaik, M. F. van Batenburg, M. Roos, R. Monajemi, H. Caron, H. J. Bussemaker, and A. H.C. van Kampen. The human transcriptome map reveals extremes in gene density, intron length, gc content, and repeat pattern for domains of highly and weakly expressed genes. *Genome Res.*, 13:1998, 2003.
- [78] H. Caron, B. van Schaik, M. van der Meer, F. Baas, G. Riggins, P. van Sluis, M. Hermus, R. van Asperen, K. Boon, P. A. Voute, S. Heisterkamp, A. van Kampen, and R. Versteeg. The human transcriptome map: Clustering of highly expressed genes in chromosomal domains. *Science*, 291:1289, 2001.
- [79] S. Goetze, J. Mateos-Langerak, W. de Leeuw, O. Giromus, H. Gierman, M. Indemanns, K. Koster, R. Versteeg, and R. van Driel. The three-dimensional structure of human interphase chromosomes is related to the transcriptome map. *in preparation*, 2006.
- [80] M. Peyrard and A. R. Bishop. Statistical mechanics of a nonlinear model for DNA denaturation. *Phys. Rev. Lett.*, 62:2755, 1989.
- [81] I. Dunham et al. The DNA sequence of human chromosome 22. *Nature*, 402:489, 1999.

-
- [82] D. L. Spector. The dynamics of chromosome organization and gene regulation. *Annu. Rev. Biochem.*, 72:573, 2003.
- [83] A. Taddei, F. Hediger, F.R. Neumann, and S.M. Gasser. The function of nuclear architecture: a genetic approach. *Annu. Rev. Genet.*, 38:305, 2004.
- [84] F. Bantignies, C. Grimaud, S. Lavrov, M. Gabut, and G. Cavalli. Inheritance of polycomb-dependent chromosomal interactions in drosophila. *Genes Dev*, 17:2406, 2003.
- [85] C. Rabl. Über Zellteilung. *Morphologisches Jahrbuch*, page 214, 1885.
- [86] M. J. Gemkow, J. Dichter, and D. J. Arndt-Jovin. Developmental regulation of DNA-topoisomerases during drosophila embryogenesis. *Exp Cell Res.*, 262(2):114, 2001.
- [87] K. M. Zimmer. *Einführung und Überprüfung eines neuen Modells mit nichtsphärischen Wechselwirkungen für die Computersimulation dichter Polymersysteme*. PhD thesis, Universität Heidelberg, 1997.
- [88] G. Schöppe. Simulation polymerer systeme mit dem ellipsoidmodell. Master's thesis, Universität Heidelberg, 1998.
- [89] G. Schöppe. *Untersuchungen an einem anisotropen Polymermodell und Kratzsimulationen an amorphen Polymeroberflächen*. PhD thesis, Universität Heidelberg, 1998.
- [90] T. Hapke. *Simulation und Untersuchung von Grenzflächen amorpher Polmersysteme*. PhD thesis, Universität Heidelberg, 1998.