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

put forward by

Dipl.-Phys. Manfred Bohn

born in: Marburg

Date of oral examination: 21.04.2010

Modelling of Interphase Chromosomes:

From
Genome Function
to
Spatial Organization

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

Abstract

Genome function in higher eukaryotes involves major changes in the spatial organization of the chromatin fiber. Nevertheless, our understanding of chromatin folding is remarkably limited. Experimental results suggest that chromatin loops not only impact transcriptional regulation but also act as a major epigenetic mechanism, playing a pivotal role in the observed compartmentalization of chromosomes. However, a unified description of chromatin folding comprising various experimental results is still lacking. After showing that the theory of compact polymers is inconsistent with experimental data, we develop a new model for chromatin based on probabilistic formation of loops. This Random-Loop-Model correctly describes folding into a confined sub-space of the nucleus as well as the observed cell-to-cell variation, suggesting a close relation between expression-dependent compaction and local variations in the looping probabilities. We find that formation of loops is highly beneficial for the nucleus to maintain order and to accomplish entropy-driven segregation of chromosomes. A dynamic model is proposed, showing that the formation of loops can be accomplished solely on the basis of diffusional motion without invoking active mechanisms. Such a dynamic model provides a unified explanatory framework of chromatin folding, yielding testable predictions, which for the first time consistently explain many experimental findings.

Zusammenfassung

Die Steuerung der Genexpression in höheren Eukaryonten erfordert größere Veränderungen in der räumlichen Anordnung der Chromatinfiber. Nichtsdestotrotz ist unser Wissen über die Struktur von Chromosomen äußerst begrenzt. Wie experimentelle Resultate zeigen, beinflussen Chromatin-Schleifen nicht nur Genexpression, sondern wirken als epigenetischer Mechanismus, welcher eine entscheidende Rolle bei der Bildung von Chromosomenterritorien spielen. Trotz dieses Wissens gibt es noch kein einheitliches Modell der Chromatinfaltung unter Integration verschiedenster experimenteller Resultate. Wir zeigen, dass Chromosome sich anders verhalten, als es die klassische Polymertheorie vorhersagt. Ausgehend von der Annahme probabilistischer Schleifenbildung wird ein das neue Random-Loop-Modell entwickelt. Diese Modell erklärt sowohl die Faltung von Chromosomen in einen begrenzten Teilraum des Zellkerns als auch die beobachtete hohe Variabilität zwischen Zellen. Die Ergebnisse zeigen eine mögliche Verbindung zwischen expressionsabhängiger Kompaktifizierung und der lokalen Schleifenwahrscheinlichkeit. Wir finden, dass Schleifenbildung ein treibende Kraft im Zellkern ist um sowohl Ordnung als auch Chromosomensegregation aufrechtzuhalten. Ein dynamisches Modell wird präsentiert, in dem Loops sich ohne aktive Mechanismen lediglich aufgrund von Diffusion bilden und welches eine expressionsabhängige Chromatinstruktur postuliert. Es erklärt eine große Anzahl experimenteller Resultate mittels eines einheitlichen Modells und liefert überprüfbare Vorhersagen für weitere Experimente.

Publications related to this thesis

Large parts of this thesis have already been published or are currently submitted. The following list contains the journal references, or the journal to which the paper has been submitted. Papers in preparation are also listed. (Information as of January 29th, 2010)

- **Bohn, M.**, Heermann, D. W. & van Driel, R. Random loop model for long polymers. *Physical Review E* **(2007)**, *76*, 051805

  DOI: 10.1103/PhysRevE.76.051805 preprint: arXiv:0705.1470v3 [cond-mat.soft]

- Langerak, J. M.[1], **Bohn, M.**[1], de Leeuw, W., Giromus, O., Manders, E. M. M., Verschure, P. J., Indemans, M. H. G., Gierman, H. J., Heermann, D. W., van Driel, R. & Goetze, S. Spatially confined folding of chromatin in the interphase nucleus. *Proc. Natl. Acad. Sci. U.S.A.* **(2009)**, *106*, 3812-3817

  DOI: 10.1073/pnas.0809501106

- **Bohn, M.** & Heermann, D. W. Conformational properties of compact polymers. *Journal of Chemical Physics* **(2009)**, *130*, 174901+

  DOI: 10.1063/1.3126651 preprint: arXiv:0905.0798v1 [cond-mat.soft]

- **Bohn, M.** & Heermann, D.W. Topological interactions between ring polymers: Implications for chromatin loops. *Journal of Chemical Physics* **(2010)**, *132 (4)*, 044904

  DOI: 10.1063/1.3302812 preprint: arXiv:1001.4246 [cond-mat.soft]

- **Bohn, M.** & Heermann, D.W. Diffusion-driven looping provides a consistent framework for chromatin organization. *PLoS Comp. Biol. (submitted, under review)*

- **Bohn, M.**, Heermann, D. W., Lourenço, O. & Cordeiro, C. E. On the influence of topological catenation and bonding constraints on ring polymers. *Macromolecules (submitted, under review)*

- **Bohn, M.**[1], Diesinger, P.[1] *et al.* Localization microscopy reveals expression-dependent parameters of chromatin nanostructure. *Biophys. J. (submitted)*

- Heermann, D. W., **Bohn, M.**, Diesinger, P. The Relation between the Gene Network and the Physical Structure of Chromosomes. *Proceedings of the 4th International Conference on High Performance Scientific Computing: Modeling, Simulation and Optimization of Complex Processes. (submitted, under review)* **(2010)**

- **Bohn, M.** & Heermann, D.W. Repulsive Forces between Chromosomes. *(in preparation)*

- Tolhuis B., Kerkhoven R., Pagie L., **Bohn M.**, Teunissen H., Nieuwland M., Simonis M., de Laat W., Heermann D.W., v. Lohuizen M.& v. Steensel B. Interactions among Polycomb Domains are Guided by Chromosome Architecture. *Cell (submitted)*

[1]Shared first authorship due to equal contribution

Contents

1 Scope and Intentions **15**
- 1.1 Introduction . . . 15
- 1.2 Intention of this thesis . . . 17
- 1.3 The structure of this thesis . . . 19

2 The organization of life – Biological Background **21**
- 2.1 The Cell – Building block of life . . . 21
- 2.2 Chromatin – The carrier of information . . . 22
  - 2.2.1 From DNA to the 10 nm chromatin fiber . . . 23
  - 2.2.2 The 30 nm chromatin fiber . . . 24
  - 2.2.3 Higher-order folding in interphase and metaphase . . . 25
- 2.3 The connection between genome folding and function . . . 25
  - 2.3.1 Non-random alignment of active and inactive regions along the 1D genome . . . 25
  - 2.3.2 The histone code . . . 26
  - 2.3.3 3D organization of sub-chromosomal regions dependents on transcriptional activity . . . 26
  - 2.3.4 Folding of chromosomes into sub-nuclear domains . . . 26
  - 2.3.5 The formation of chromosome territories . . . 28
- 2.4 Chromatin loops as a mediator of the folding-function relationship . . . 29
  - 2.4.1 Gene regulation and long-range control . . . 29
  - 2.4.2 Long-range control mediated by chromatin loops . . . 30
  - 2.4.3 Three-dimensional structure of loops: Transcription factories, CTCF and the nuclear matrix . . . 31
  - 2.4.4 Implications for a model of chromatin . . . 32

3 Polymer models of chromatin **33**
- 3.1 Statistical Physics of Polymers . . . 33
  - 3.1.1 The freely jointed chain model . . . 34
  - 3.1.2 The Gaussian Chain model . . . 35
  - 3.1.3 Effects of excluded volume – the self-avoiding walk . . . 37
  - 3.1.4 Effects of the solvent – transition to the globular state . . . 38
  - 3.1.5 Intermingling in polymeric solutions . . . 38
  - 3.1.6 A question of scale – Coarse graining and bending rigidity . . . 39
- 3.2 Models of chromatin . . . 40
  - 3.2.1 Random walk in a confined space . . . 41
  - 3.2.2 The Random Walk / Giant - Loop model . . . 42
  - 3.2.3 The Multi-Loop-Subcompartment model . . . 43

3.2.4 Other models 44

4 The conformational properties of compact polymers 45
4.1 Introduction 46
4.2 Is chromatin organized as a random walk or self-avoiding walk? 48
4.2.1 Scaling of the mean square displacement 48
4.2.2 Distance distributions 49
4.2.3 Moment ratios of the distribution 50
4.3 Monte Carlo simulations of compact polymers 53
4.3.1 The Algorithm 53
4.3.2 Ergodicity and unbiased sampling 54
4.3.3 Autocorrelation times 57
4.4 Conformational properties of compact polymers 57
4.4.1 End-to-end distance statistics 57
4.4.2 Intrachain distance statistics 60
4.4.3 End-point statistics 62
4.4.4 Correlations of intrachain segments 62
4.4.5 Screening of excluded volume in compact polymers 64
4.4.6 The gyration tensor 65
4.5 Is chromatin organized as a compact polymer? 66
4.6 Conclusions 68

5 The Random Loop Model 71
5.1 Introduction 72
5.2 Basic model assumptions 73
5.3 The homogeneous Random Loop Model in the annealed and quenched average 75
5.3.1 The Hamiltonian 75
5.3.2 Quenched or annealed ensemble? 75
5.3.3 The quenched average 76
5.3.4 Results and comparison to experiments 78
5.3.5 The annealed ensemble 81
5.3.6 Limiting cases without disorder 83
5.4 Conclusions 84

6 Molecular Dynamics simulations of the Random Loop Model 87
6.1 Introduction 88
6.2 The Simulational Method 89
6.2.1 Molecular Dynamics 89
6.2.2 Setting up the potentials 90
6.2.3 Details on the implementation 92
6.2.4 Equilibration of the system 93
6.3 Results 94
6.4 Simulations of the annealed ensemble 97
6.5 Conclusions 98

7 A unified view of short and large scale folding 99
7.1 Introduction 100
7.2 First evidence for different local looping probabilities 101
7.3 Integration of short and long length scale data by the RL model 103

7.3.1 The Heterogeneous Random Loop model . . . 103
7.3.2 Results for chromosome 11 . . . 104
7.4 Conclusions . . . 106

8 Effect of topological catenation and bonding constraints on rings and loops 109
8.1 Introduction . . . 110
8.2 The computational model . . . 112
8.2.1 The Monte Carlo algorithm . . . 113
8.2.2 Data acquisition and autocorrelation times . . . 115
8.3 Results . . . 117
8.3.1 Catenation and bonding influences ring dimensions slightly . . . 117
8.3.2 Shape of ring polymers changes due to topological constraints . . . 121
8.3.3 Ring structure induces entropy-driven segregation . . . 124
8.3.4 Alignment of ring polymers . . . 126
8.4 Conclusions . . . 128

9 Topological interactions between ring polymers 131
9.1 Introduction . . . 132
9.2 Simulations and methods . . . 133
9.2.1 Effective potentials . . . 133
9.2.2 Monte-Carlo algorithm . . . 134
9.2.3 Calculation of the effective potential . . . 134
9.2.4 The effective potential of self-avoiding walks . . . 136
9.2.5 Rings have a stronger repulsion than linear polymers . . . 138
9.2.6 Topological constraints induce additional repulsion . . . 139
9.2.7 Dimensions of two-ring conformations . . . 141
9.2.8 Shape of two-ring conformations . . . 143
9.2.9 Orientation and Alignment of rings . . . 144
9.3 Conclusions . . . 146
9.4 Supplementary Information . . . 148
9.4.1 Supplementary Figure 1 . . . 148
9.4.2 Supplementary Figure 2 . . . 149
9.4.3 Supplementary Figure 3 . . . 150

10 Diffusion-driven looping as a consistent framework for chromatin organization 151
10.1 Introduction . . . 152
10.2 The chromatin model . . . 154
10.2.1 General overview . . . 154
10.2.2 Simulational Method . . . 155
10.3 Results . . . 156
10.3.1 Mean square distance between chromatin segments . . . 157
10.3.2 Self-organized formation of large loops . . . 158
10.3.3 Cell-to-cell variation and dynamic fluctuations of the distance distribution . . . 162
10.3.4 The shape of chromosomes . . . 164
10.3.5 The dynamics of looping chromosomes . . . 165
10.3.6 The formation of aspherical chromosome territories . . . 167

10.4 Discussion . . . 170
10.5 Supplementary Information . . . 173
10.5.1 Supplementary Figure 1 . . . 173
10.5.2 Supplementary Figure 2 . . . 174
10.5.3 Supplementary Figure 3 . . . 175
10.5.4 Supplementary Figure 4 . . . 176
10.5.5 Supplementary Figure 5 . . . 176

11 Repulsive forces between chromosomes 177
11.1 Introduction . . . 178
11.2 Methods . . . 179
11.2.1 Computer simulations of chromosomes . . . 179
11.2.2 Calculation of the effective potential . . . 179
11.3 Results . . . 179
11.3.1 The effective repulsion between chromosomes increases strongly with loop number . . . 179
11.3.2 The force between chromosomes can be calculated . . . 182
11.3.3 Looping polymers become aspherically elongated . . . 182
11.3.4 Looping polymers avoid intermingling . . . 184
11.4 Conclusions . . . 186

12 Nuclear nanostructure detected with localization microscopy 187
12.1 Introduction . . . 188
12.2 Experimental Methods and Image Segmenation . . . 189
12.2.1 Specimen preparation . . . 190
12.2.2 SPDM setup . . . 190
12.2.3 Data Acquisition and Evaluation . . . 192
12.2.4 Image segmentation . . . 193
12.3 Radial distribution function reveals nanoscale structure . . . 194
12.3.1 Definition . . . 194
12.3.2 The influence of projections . . . 195
12.3.3 Calculations of $g(r)$ for the data . . . 195
12.3.4 Results . . . 197
12.4 Density distribution analysis reveals large-scale fluctuations . . . 197
12.4.1 Background and Method . . . 197
12.4.2 Results . . . 200
12.5 Relating to polymer models . . . 202
12.5.1 Simulations of the Dynamic Loop model in a melt . . . 203
12.5.2 Results . . . 204
12.6 Conclusions . . . 207

13 Pattern analysis of confocal images 209
13.1 Introduction . . . 210
13.2 Experimental Data and Image Processing . . . 211
13.3 Quantitative Analysis Methods . . . 213
13.3.1 Intensity Distribution and Correlations reveal differences between cell types . . . 213
13.3.2 Active genes are more clustered than inactive genes . . . 214

13.3.3 Repressed genes are more abundant in the nuclear periphery than active genes . . . 216
13.4 Modeling Approaches . . . 216
13.4.1 The Dynamic Loop model . . . 217
13.4.2 The Potts model . . . 218
13.4.3 The Ising model . . . 221
13.5 Conclusions and future work . . . 223

14 Summary & Outlook **225**
14.1 A short summary of the results . . . 225
14.2 Synthesis . . . 228
14.3 Outlook . . . 228

Publications **231**

Conference contributions **233**

References **235**

Acknowledgments **247**

Chapter 1

Scope and Intentions

A short overview over topics and aims

1.1 Introduction

The cell nucleus is a main constituent of eukaryotic organisms and yet its complexity prevents detailed knowledge of its function. The genome content is carried by the chromosomes: compactly folded polymers consisting of DNA and histone proteins. Whereas during mitosis chromosomes are found in an extremely condensed state, the chromatin fiber inside the interphase nucleus has a much more decondensed organization. However, at this stage of the cell cycle, highly coordinated processes such as transcription, replication and DNA repair take place, making a random folding of the chromatin fiber very unlikely. The haploid human genome consists of 23 chromosomes with a total of more than 3 billion basepairs (bp), which constitute the building blocks of the genome and are carrier of information. In diploid cells, about 2 meters of double-stranded DNA have to be densely packed into a nucleus which has a diameter on the order of 10 μm. Folded into a compact globule, these 2 meters of DNA with an average diameter being only about 2 nm could be packed into a sphere of radius 1.8 μm, which indeed is much smaller than the diameter of the nucleus. On the other hand, a typical non-condensed human chromosome of a length of 100 Mb in good solvent would extend to about 130 μm, a scale much larger than the diameter of the nuclear envelope.

The DNA content, which is the same for each cell of the organism, has to be highly accessible by numerous proteins such as transcription factors, polymerases and regulatory elements. Given that gene activity strongly depends on cell differentiation, higher eukaryotes have to provide versatile epigenetic mechanisms for gene regulation. The folding of the genome in the interphase nucleus of eukaryotic cells is done on multiple scales of length and degrees of compaction. The basic filament is the DNA double helix which, in a first stage of compaction, is wrapped in a 1-and-3/4 turn around histone cores. About 147 bp (basepairs) are associated with each histone octamer constituting the nucleosome. The chromatin fiber is a complex of nucleosomes and linker DNA forming a beads-on-a-string type of filament with a diameter of about 11 nm [1]. In-vitro experiments provide evidence that this structure in turn condenses under certain salt conditions to an even more compact structure of 30 nm, but both its regularity and its existence in living cell nuclei are

still under debate [2, 3, 4, 5]. Stunningly, even less is known about the structural organization on the scale above 30 nm. Up to now, experimental techniques have been limited by the resolution of conventional light microscopes of about 200 nm, requiring indirect assays for investigating chromatin folding. Several experimental techniques have been applied: Fluorescent labeling of large parts of a chromosome yields results on structure, shape and position of chromosomal regions [6] or even of entire individual chromosomes [7]. Labeling two loci of a chromosome with a fluorescent marker was successfully used to establish a relationship between the genomic distance g of these markers and its average physical distance in yeast [8], drosophila [9, 10] and human cells [11, 12].

Despite our lack of knowledge on the detailed folding mechanisms of chromatin in interphase nuclei, there is now abundant evidence that genome function is tightly related to chromatin folding on several length scales. The one-dimensional distribution of genes along the chromosome is far from being random: the Human Transcriptome Map [13] reveals a clustering of active genes as well as inactive genes into certain domains, which have been named ridges and anti-ridges [6]. Various experiments have shown that the 3D organization of chromatin depends on transcriptional activity: Active genes tend to be located in the nuclear interior while inactive genes are found more often at the nuclear periphery [6, 14, 15]; the converse behavior is observed in some experiments [16]. Moreover, a change in the transcriptional state of a gene can have direct influence on its positioning inside the nucleus [17, 18]. Transcriptional active regions (ridges) were observed to have a more open structure than inactive regions (anti-ridges) [6]. *Fluorescence in situ hybridization* (FISH) has allowed the positioning of fluorescent markers at specific points along the genome. The relationship between mean square distance (MSD) of two fluorescent markers and their genomic separation has revealed significant differences in compaction between ridges and anti-ridges [12]. Gene-rich and transcriptional active regions are less compact than gene-poor regions. Importantly, above genomic separations of 5-10 Mb, the mean square displacement between two FISH markers does not depend any longer on genomic separation. This study indicated a folding of chromatin regions and entire chromosome arms into a confined sub-space of the nucleus with an average diameter of about 2 μm.

The theory of linear polymers predicts that a semi-dilute solution of chromosomes, as it can be found inside the nucleus, displays a strong intermingling between different chromosomes. Amazingly, experiments reveal that this is not the case. Rather, chromosomes are separated into distinct chromosome territories [7], whose relative positions and ellipsoidal shape varies from cell to cell [19].

Intra-chromosomal as well as inter-chromosomal contacts or loops have been intensively analyzed in the past few years both experimentally and theoretically as a possible mechanism for transcriptional regulation and genome folding. Yet, chromatin loops seem to be an ubiquitous feature of genome organization and genome function. Transcriptional regulation is often controlled by regulatory motifs such as enhancers and silencers. These can be located tens of kb up to several Mb apart from the target gene which they regulate [20, 21]. One possible interaction mechanism is spatial proximity of regulatory element and target gene which requires the looping out of intervening DNA. 3C experiments have demonstrated that this is indeed the case in the β-globin locus [22]. One idea put forward to explain chromatin loops is the existence of transcription factories or active chromatin hubs, where active polymerases cluster and thereby co-locating genes and regulatory elements [23, 24]. 3C and 4C techniques have since then provided evidence that indeed loops up to several Mb exist in interphase cells [20, 25]. However, the detailed mechanisms and driving forces of looping are still under debate.

The understanding of intra-chromatin interactions and chromatin looping on genome structure and organization can be deepened by polymer models which are able to shed light on the effects of specific types of chromatin-chromatin interactions. Coarse-grained models are especially useful since they neglect all molecular details, thus shaping an overall picture of the major driving forces and organizational principles of chromatin folding. Polymer models have been successfully used to explain certain experimental results. The conclusions of these models differ strongly, depending on the experimental data set they refer to. One of the first models that has been proposed was a random walk in the confined geometry of a sphere [26]. This model predicts a random walk behavior for short genomic separations while for large genomic separations the mean square distance shows a leveling-off. The Random-Walk/Giant-Loop Model (RWGL) was proposed to explain data from Yokota et al. [11, 27] and assumes giant loops of several Mb regularly positioned along a random walk backbone. This model exhibits two regimes, both showing a random walk behavior of different slope on the short and large scale. The multi-loop-subcompartment (MLS) model [28, 29] assumes loops in the 100kb-range to form rosette-like structures connected by a backbone. Another study shows that some FISH measurements are compatible with an out-of-equilibrium worm-like chain model [30].

In fact, none of the polymer models proposed so far has succeeded in establishing a detailed connection between genome folding and function. Even worse, they are inconsistent with important aspects of experimental evidence. The folding into a confined sub-space of the nucleus is not observed for the RWGL and the MLS model [11, 28]. Other models impose confinement by means of the boundary conditions [26, 31], thus these models are unable to make predictions about the mechanisms leading to this behavior. Interestingly, the experimental evidence for loops on all scales, many of them being specific and functional, has been neglected in any of the polymer models.

1.2 Intention of this thesis

In this thesis, my intention is to bridge the gap between the abundant experimental evidence for a tight connection between genome folding and function and statistical polymer models. The intention is to present a polymer model which incorporates recent experimental data and provides a consistent explanatory framework of higher-order chromatin organization spanning the length scale from a few hundred kb to the complete chromosome and further explaining the formation of chromosome territories.

In principal, the main aim of building *models* of physical or biological systems is not to have a complete *theory* of the specific system. Rather models aim at explaining the available experimental data by introducing only a minimal set of parameters with which the data can be described and predictions derived in a satisfactory way . The question of what is considered “satisfactory” depends on the application. For example, a simple model of a “real“ gas, say oxygen, is the ideal gas model. It is minimal in the sense that it keeps only the coordinates and impulses of the particles and neglects mutual interactions and the extension of the molecule. This model is good at explaining several experimental observations, but by far not all. Phenomena like phase transitions can only be explained by a model having pair interactions and excluded volume. For an even more detailed description of the system, further parameters like the electron spins are necessary.

The same can be said of the biological system ”cell nucleus“. A lot of experimental information is available. In a simple picture, the nucleus contains 46 DNA polymers and a solvent. However, there is much more to this. DNA is not a homogeneous polymer, rather

it has a specific sequence of basepairs with a specific function. Proteins manipulate and form the DNA, and thermal equilibrium might be disturbed by ATP-consuming agents. Constructing a model explaining all these tiny details would not only be impossible, even worse it would distract the attention from the general laws guiding chromatin organization to very locus-specific properties. Lacking detailed knowledge of all functional mechanisms controlling and maintaining higher eukaryotes, it is reasonable to subsume interactions between chromatin and proteins or solvent only effectively into the model. Here, we are interested in general laws reigning the life of higher eukaryotes. Thus our aim is to present a polymer model based on abundant experimental input, which is in agreement with biological observations both qualitatively as well as quantitatively and has further predictive power to stimulate biological experiments in the near future.

In the following, I list the eight major experimental hallmarks of chromatin organization and its connection to genome function, which, among others, the polymer model presented in this thesis aims at explaining.

- *1. Experimental evidence for loops on all scales.* The 4C and 5C technique has provided evidence that loops exist on all scales ($\curvearrowright$ 2.4.2) spanning genomic separations from several bp to many Mb [25].

- *2. Dynamics of loop formation.* Various studies revealed that loops are closely related to transcriptional regulation (cf. section 2.4), indicating that the nuclear organization is highly dynamic allowing the formation of loops depending on the transcriptional processes required.

- *3. Confined folding of a chromosome into a sub-space of the nucleus.* FISH experiments revealed that human chromosomes 1 and 11 fold into a confined subspace of the nucleus with a diameter of about 2 μm, being far below the diameter of the nucleus ($\curvearrowright$ 2.3.4).

- *4. Large cell-to-cell variation.* Distance measurements using FISH markers reveal a large cell-to-cell variation in the measured distribution of distances. These fluctuations are even larger than one would predict from a random walk polymer model which naturally exhibits large fluctuations due to the conformational degrees of freedom. None of the proposed models can explain such variation, thus it remains a crucial point to develop a polymer model with a fluctuation regime larger than that of a random walk.

- *5. Different levels of compaction of gene-rich and gene-poor regions.* Abundant experimental evidence suggests a strong dependence of chromatin folding on the local transcriptional state. Gene-rich and transcriptional active regions have a more open structure than gene-poor regions [6, 12]. Amongst the polymer models presented in chapter 3 only the MLS model takes chromatin heterogeneity into account.

- *6. Sub-diffusional motion of labeled chromatin loci.* Measurements of interphase chromosome dynamics are rare. Little or no motion has been detected for chromatin domains in the size range of chromosome territories (1 μm) over time scales of several hours [32]. The motion of a specific active gene (GAL) has been measured *in vivo* in yeast nuclei [17]. Cabal and co-workers found anomalous diffusion with $g_1(t) \sim t^{0.4}$. Dynamics are addressed in the model of Rosa *et al.* [30], recovering only the typical dynamics of semi-flexible polymers on short time scales with $g_1(t) \sim t^{0.25}$.

- *7. Formation of chromosome territories.* Compartmentalization of chromosomes in the interphase nucleus into distinct territories has been postulated by early cytologists and was confirmed in the last decades by chromosome painting and fluorescence in situ hybridization approaches [33, 7, 34]. The mechanisms maintaining the segregation have not been fully elucidated up to now, suggestions range from purely entropic effects [35] to results of non-equilibrium dynamics [30].

- *8. Aspherical shape of chromosome territories.* Experimental studies reveal a non-spherical shape of CTs [36, 37]. This is in contrast to a compact globule state of chromatin. Mouse chromosomes exhibit an aspherical shape best approximated by ellipsoids with axis ratios $4.5 : 2.9 : 1$.

1.3 The structure of this thesis

In chapter 2 the necessary biological background information is provided. Rather than being meant as a complete reference of cell biology (cf. Ref. [38]), I will concentrate on introducing basic principles and open questions of chromatin folding and explaining biological terms and experimental techniques referred to in the following chapters. Hopefully, it serves as a reference for unclear terms or background information while reading this thesis.

An introduction to notions and scaling laws of polymer physics is given in chapter 3. Here, the interested reader will find information about simple polymer models like the random walk and self-avoiding walk model as well as a short overview over other chromatin models proposed so far.

In chapter 4 simulations of the *globular state polymer model* are presented. The globular state model or variants thereof have been proposed for chromatin organization in several studies [31, 39]. From our simulational data we derive measures for comparison to experiments and show that neither the globular state model nor the random walk or self-avoiding walk model describe the folding state of chromatin.

The available experimental evidence of loops leads us to propose the *Random Loop (RL) model* in chapter 5. The basic ideas and assumptions of the model are presented. Results for a semi-analytical implementation assuming a homogeneous chromatin fiber are derived and compared to experimental data. Both the confined folding of chromatin as well as the large cell-to-cell variation found in experiments are well reproduced by this model.

In chapter 6 a major short-coming of the analytical Random Loop model is lifted: the negligence of excluded volume. Molecular Dynamics simulations are conducted to obtain sample conformations. We find that excluded volume effects do not alter the qualitative behavior of the model.

Gene-rich and gene-poor chromatin regions display a pronounced difference in the compaction. This *heterogeneity of the fiber* is incorporated in the analytical Random Loop model in chapter 7. We conjecture that local differences in compaction are related to local differences in the looping probability, making chromatin loops the major mediators of a tight connection between genome folding and function. We find excellent agreement with FISH measurements on human chromosome 11 spanning genomic separations from several hundred kb to the complete length of the chromosome arm.

What is the benefit for nature to form rather large-sized loops for transcriptional regulation? This question will be investigated in the following two chapters. In chapter 8 we study a *system of two loops* constrained to either a catenated topology or a bonded,

non-catenated topology. The latter can be used as a toy model for loops attached to transcription hubs [24]. A tendency towards segregation in comparison to similar systems of linear chains is observed, indicating an important role of loop formation in chromosome compartmentalization.

To quantitatively analyze the *effect of ring closure*, i.e. looping, we study the effective potential of mean force between their centers of mass (↷ chapter 9). Indeed, the transition from a linear to a ring polymer induces a strong increase in the entropic repulsion between two rings or loops. Thus, an evolutionary motif for the formation of ring polymers or loops is possibly given by its support in creating distinct polymeric compartments with little intermingling.

After having discussed the general benefits of chromatin looping using simple systems of ring polymers, we turn back to chromatin in chapter 10 to answer the question how large loops can form inside the nucleus. We establish a dynamic Monte Carlo method and demonstrate that large loops form solely on the basis of diffusional motion. This *Dynamic Loop model* provides a unified description of chromatin folding on the basis of dynamic loop formation without invoking active transport mechanisms. We derive testable quantitative predictions on many structural and conformational properties of chromosomes. In fact, for the first time, a huge amount of experimental evidence from FISH distance measurements, 4C/5C data, diffusion measurements up to the formation of chromosome territories and its shape can be explained consistently in the framework of one model.

While the studies on ring polymers served as a simplistic model system to assess the effects of loops, the effective potential between more complex looped polymers, i.e. our model chromosomes, is investigated in chapter 11. We find that chromatin looping strongly increases the repulsive interactions between chromosomes and imposes kind of order on the system, the effects being much stronger than for a system of two rings.

An approach combining novel high-resolution localization microscopy data with statistical methods and the Dynamic Loop model is presented in chapter 12, clearly showing how the theoretical concepts developed in this thesis can be applied to study chromatin structure. Finally, in chapter 13, we present methods to quantitatively analyze images from confocal light microscopy to determine differences between stem cells and differentiated cell types with respect to the distribution of active and inactive genes inside the nucleus.

Chapter 2

The Organization of Life

Biological Background

The work described in this thesis comprises two scientific disciplines, which once were considered being quite distinct: biology and physics. Nowadays, concepts of physics are more and more applied to biological problems. The approach pursued in the following chapters is to consider the biological system from a statistical mechanics point of view: Chromatin fibers are modeled as polymers, i.e. many particle systems subjected to certain bonded and non-bonded interactions. In order for the thesis to be understandable for readers of different disciplines we provide the basic background information on the biological system under consideration in this chapter. The basic concepts of polymer physics and chromatin models are outlined in the next chapter.

2.1 The Cell – Building block of life

Although living species covering our planet are highly diverse, all kinds of life evolve from the same "building block": the cell. While prokaryotic cells (like bacteria) are often independent, eukaryotic cells are usually part of multicellular organisms. Each cell is surrounded by a *plasma membrane* made up of a lipid bilayer. Its purpose is to protect the cell from its surrounding environment and maintain the integrity of the chemical system inside. Both nutrients as well as waste materials have to pass this kind of semi-permeable membrane with its huge number of receptors, channels and pumps for regulating the inflow and outflow. Eukaryotic cells are stabilized by a scaffold called *cytoskeleton*, which amongst others supports cell growth and motility as well as the separation of daughter cells during division. Each cell is a functional unit on its own: It can take up nutrients from the environment, convert these into energy and reproduce itself. Importantly, each cell contains the whole script, i.e. the instructions, for carrying out these functions.

The molecular basis of life is the same independent of the organism. Hereditary information is stored in form of double-stranded polymers (deoxyribonucleic acid, DNA), the linear sequence of bases acting as a code for life (↷ 2.2). This DNA in eukaryotic cells is separated from the cytoplasm by the nuclear envelope, a double layer of membrane. Such a compartmentalization is an important organizational feature found over and over again, which in case of the nucleus ensures a fast catalysis of biochemical reactions by providing and keeping up a high concentration of necessary proteins. One important task of DNA is

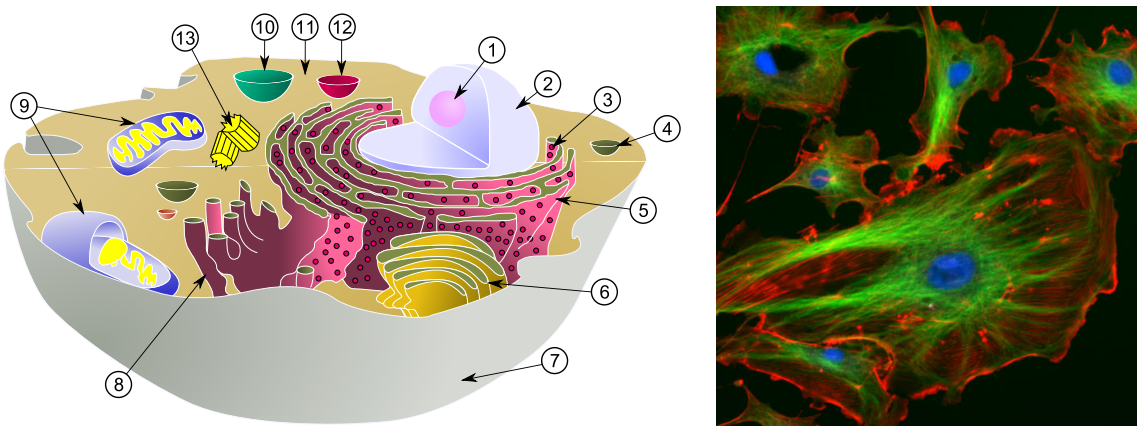


Figure 2.1: The left-hand image shows a sketch of a typical eukaryotic cell with the following subcellular components: (1) nucleolus, (2) nucleus, (3) ribosome, (4) vesicle, (5) rough endoplasmic reticulum (ER), (6) Golgi apparatus, (7) Cytoskeleton, (8) smooth endoplasmic reticulum, (9) mitochondria, (10) vacuole, (11) cytoplasm, (12) lysosome, (13) centrioles within centrosome. The right-hand image shows endothelial cells under the microscope. Nuclei are stained blue with DAPI, microtubules are marked green by an antibody and actin filaments are labeled red with phalloidin. Images are adapted from [40].

to store and provide the sequence building plan for proteins. A *gene* is roughly defined as a segment of DNA corresponding to one protein. Interestingly, the human genome consists of about 98.5% non-coding regions, i.e. regions which do not code for a protein. These regions are often involved in transcriptional regulation, controlling which genes have to be transcribed at a certain time (↷ 2.4.1).

Transcription is the process of reading information stored in the DNA, i.e. the specific sequence of base pairs A, T, G and C. The information is copied into an RNA strand complementary to the DNA sequence, which in the next step is modulated into messenger RNA (mRNA). This mRNA can diffuse through the cell nucleus and pass the nuclear membrane. Once outside the nucleus, the *ribosomes* – large complexes of protein and RNA – use the mRNA code as a template to assemble a specific sequence of amino acids to a protein.

During its lifetime, each cell runs through different stages of development, i.e. the *cell cycle*. This is the mechanism by which an organism is created out of one single cell at birth and by which cells in living organisms are permanently renewed. The cell cycle consists of four distinct phases, which are commonly known as G_1 phase, S phase, G_2 phase and M phase [38]. The G_1, S and G_2 phases together are called *interphase*. During S-phase DNA duplication occurs, i.e. the DNA content of the cell is doubled, while the G_1 and G_2 gap phases provide time for the cell to grow and increase their mass of proteins and other substances. The M phase is divided into mitosis, where the DNA is condensed into chromosomes and cell divisions takes place and cytokinesis, where cytoplasmic division occurs. One stage of the mitosis is the *metaphase*, in which highly condensed chromosomes align in the middle of the cell before being separated into the daughter cells.

2.2 Chromatin – The carrier of information

Life on earth is strongly dependent on the ability to store, retrieve and translate the genetic instructions required to make and maintain a living organism. These instructions are

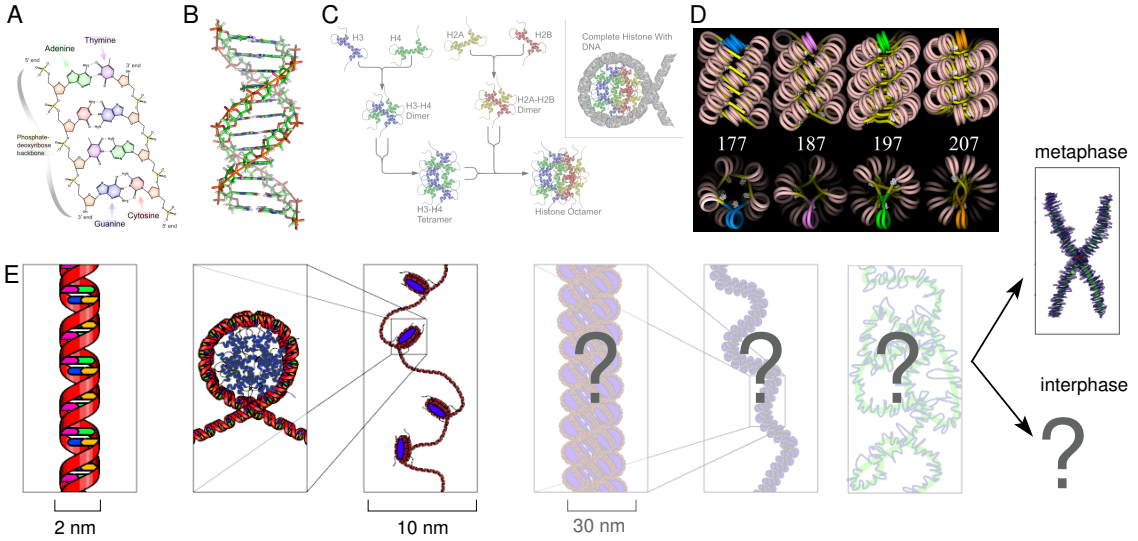


Figure 2.2: Stages of chromatin compaction. **A.** The chemical structure of nucleotides is shown exemplarily for one AT and one GC base pair. **B.** The double helix structure of DNA. **C.** The formation of the nucleosome. First, histone proteins H3 and H4 as well as H2A and H2B form dimers, which are then composed to tetramers and finally to a histone octamer. Together with the DNA wrapped around this octamer, the complex forms the nucleosome core particle. **D.** Proposed structures of the 30 nm chromatin fiber depending on the nucleosome repeat length ranging from 177 to 207 bp. **E.** Schematic view of the stages of chromatin compaction. Starting from a linear polymer, DNA folds around histone octamers forming a beads-on-a string like fiber of diameter 10 nm. Little is known about higher-order structures inside the interphase nucleus. During metaphase chromosomes condense to very compact and stiff objects making them robust for cell division. Images adapted from [40].

stored in the DNA. DNA (deoxyribonucleic acid) is made up of two long polymers, whose building units are called *nucleotides*. Each nucleotide consists of a sugar and phosphate backbone to which a molecule, the *base*, is attached. This base molecule is either adenine (A), cytosine (C), guanine (G) or thymine (T). Typically, in living cells, DNA does not exist as a single-stranded polymer, rather two strands of DNA are aligned to a double helix (Fig. 2.2A and B). This double helix structure of DNA has been discovered in 1953 by Watson and Crick [41] using X-ray diffraction. The two DNA strands run in opposite directions to each other and are therefore aligned anti-parallel. Each base of one strand is bonded to a base on the second strand, which is twined around the first one. The chemical properties of the bases ensure that only A-T and G-C bonds exist, an observation which is called *complementary base pairing*. The diameter of such a DNA strand is about 22 Å, the length scale of a nucleotide unit is 3.3 Å.

2.2.1 From DNA to the 10 nm chromatin fiber

The one-dimensional sequence of A, T, G and C codes the building plans for life. A haploid human genome consists of about 3.2 billion basepairs, resulting in a strand of 2 meters, if unfolded. 23 000 protein-coding genes are encoded in this vast amount of DNA, about 98.5% of the sequences are non-coding [42]. The detailed function of most of the non-coding regions, which amongst others are made up of tandem repeats and transposable elements, is not known. Some non-coding regions, however, contain important regulatory sequences crucial for controlling transcriptional activity. In fact, regulatory sequences exist

for nearly each gene, providing for example binding sites for specific proteins (↷ 2.4.1).

DNA inside the interphase nucleus is highly compacted. This is achieved by proteins that successively coil and fold it into higher levels of organization [38]. The basic filament is the DNA double helix (Fig. 2.2A). In a first step of compaction this double helix is wrapped around histone octamers interconnected by stretches of linker DNA. Histone octamers are constituted of pairs of the four core histones H2A, H2B, H3 and H4, which are evolutionary amongst the most highly conserved proteins. About 146 basepairs of DNA are wrapped around this histone core in a 1.65 left-handed super-helical turn [43]. Eventually, a linker histone H1 is attached at the entry-exit point of the DNA, fixating the structure of this *nucleosome core particle* (Fig. 2.2C). However, such linker histones are sometimes missing [5], allowing for a locally higher flexibility of the fiber. The typical shape of the nucleosome core particle is a cylinder with a radius of ~ 5 nm and a height of ~ 6 nm [44]. The complex of nucleosome core particles and linker DNA forms a beads-on-a-string type of fiber called *chromatin*. This loose structure with a diameter of about 10 nm has been observed using high-resolution X-ray analyses at low salt concentrations *in vitro*, while at physiological salt concentrations the chromatin fiber adopts a more compact state with a diameter of about 30 nm [45].

Although, histones bind non-specifically to DNA, the local regulation of histone binding is supposed to be an important regulatory mechanism (↷ 2.3.2).

2.2.2 The 30 nm chromatin fiber

While the structure of the nucleosome core particle is well-known, higher-order structures are still an open issue of research. Electron microscopy reveals a transition from a beads-on-a-string like structure of 10 nm to a more compacted state of 30 nm on raising the salt concentration towards physiological values of 100 mM *in vitro* [45, 46]. The detailed alignment and regularity of nucleosomes, however, remained highly controversial (Fig. 2.2D). Mainly two competing classes of models have been proposed for the structure of this 30 nm fiber, the solenoid models [47, 45] and zig-zag models [3, 48].

The solenoidal-type of models predict successive nucleosomes to form a helical structure, the axis of the core particles being aligned perpendicular to the solenoidal axis. The entry-exit angle of the DNA wound around the histone octamers is stabilized by linker histones H1 and H5 and faces inward towards the axis of the solenoid. Thus, this model requires the linker DNA to be bent between neighboring nucleosomes, inducing a cost of energy.

The other class of models assumes successive nucleosomes to be located on opposite sites of the fiber, the linkers between them being straight. Such a cross-linker model results in a zig-zag structure of the fiber and allows changes in the compaction level without a change in topology.

A geometrical approach, the two-angle model, was developed by Schiessel [1], where chromatin structure is described by three parameters: (a) the linker length b, (b) the entry-exit angle α of the DNA and (c) rotational angle β between nucleosomes. It was shown by Diesinger *et al.* that such a model allows for a variety of chromatin structures from cross-linker to solenoidal structures [49]. Consistent with experimental observations that linker histone H1 is sometimes missing, the two-angle model was extended to allow for such vacancies [5].

2.2.3 Higher-order folding in interphase and metaphase

The higher-order structures of chromatin above the scale of the chromatin fiber are not well characterized. However, there is extensive evidence that chromatin loops do exist in the interphase nucleus. Various studies have indicated that the chromatin fiber forms loops that at their bases may be attached to a still poorly defined structure that is called nuclear scaffold/matrix [50]. Recent investigations show that the formation of chromatin loops involves specific proteins, including SatB1 [51], CTCF and other insulator binding proteins [52]. Other studies have revealed long-range chromatin-chromatin interactions due to transcription factories in which transcriptionally active genes at different positions on a chromosome and from different chromosomes come together [20]. In the next section (↷ 2.3), we present several experimental results on chromatin folding in the size range of a few kb to the scale of the whole chromosome. In section (↷ 2.4) the importance of loops as a possible mediator for the connection between genome folding and function is pointed out.

During metaphase, chromosomes condense to very compact structures differing vastly from those of interphase. The chromosomes adopt its classical form known from light microscopy images, optimized for stiffness and strength allowing an easy division and transport into the daughter cells. The structure of these condensed chromosomes is assumed to be loops attached to a still poorly defined central scaffold of proteins. In this thesis, however, we concentrate on folding motifs in the interphase nucleus, i.e. where chromatin displays a much more open and flexible structure.

2.3 The connection between genome folding and function

Despite the lack of knowledge on the detailed folding pathways of chromatin inside the interphase nucleus of higher eukaryotes, an emerging opinion is that genome folding is tightly connected to its function. Clearly, chromatin has to be accessible to a variety of proteins to maintain functions like transcription, repair and duplication, while at the same time being quite compact. Live cell imaging is limited currently by the resolution of conventional light microscopy of about 200 nm, making it impossible to track single chromatin fibers inside the cell. Therefore, indirect approaches are necessary to obtain structural information. It should be noted, however, that localization light microscopy allows for an alternative approach on studying nanostructure (↷ chapter 12). Here, we shortly present experimental evidence for function-dependent chromatin folding, all of which is important for creating computational models of chromatin in the course of this work.

2.3.1 Non-random alignment of active and inactive regions along the 1D genome

The arrangement of genes along the one-dimensional genome is far from random. Investigation of the gene activity is conducted using genome-wide messenger RNA expression profiles. Relating this expression data with the location of genes, gene expression profiles have been generated for any chromosomal region. This Human Transcriptome map revealed that highly expressed genes are clustered in distinct regions. These regions of increased gene density and transcriptional activity are called *ridges*. Similarly, gene-sparse regions with little transcriptional activity (*anti-ridges*) are detected [53, 13]. Figure 2.3 displays gene density and transcriptional activity for the q-arms of human chromosome 1

and 11. Markedly active regions (ridges) are labeled green, inactive regions (anti-ridges) red. Clearly, genes are not distributed randomly along the fiber.

2.3.2 The histone code

The three-dimensional folding of chromatin is connected to genome function on several scales. On the smallest scale, modifications of histone tails induce a local change in compaction. It has long been suggested that the histones, by their tight binding to the DNA, are participating in regulatory functions. Post-translational modifications of histones are assumed to carry the information about distinct DNA-templated programs [54, 43], a hypothesis called *histone code*. A pivotal role is taken over by the histone tails, long charged strands protruding from the surface of the chromatin fiber making up about 25-30 % of the individual histones [4], offering a variety of possible interactions with proteins. These tails can be chemically modified by adding or removing several chemical modifications like acetyl, phosphoryl and methyl groups. How can such modifications participate in transcriptional regulation? For example, histones H3 tri-methylated on lysine 27 are predominantly found on gene promoters of repressed genes, while tri-methylated histones H3 on lysine 4 accumulate on promoters of active genes [55]. While certain modifications directly lead to an opening up of the chromatin structure, allowing access of proteins to DNA, specific regulatory proteins also contain distinct binding sites for specific histone modifications with a high binding affinity [56].

2.3.3 3D organization of sub-chromosomal regions dependents on transcriptional activity

On a broader scale of a few mega basepairs, the 3D organization of sub-chromosomal regions has been shown to significantly depend on transcriptional activity. Goetze and co-workers [6] fluorescently labeled stretches of chromosomes belonging to ridges and anti-ridges in the size range of 1-10 Mb and determined shape and volume of such regions. It turned out that ridges are less condensed and more irregularly shaped, while anti-ridges adopt a more compact structure being closer to a sphere. Transcriptional activity also influences the positioning of chromatin inside the interphase nucleus: Gene-rich and transcriptional active regions are in tendency more located towards the nuclear interior while anti-ridges tend to locate at the periphery. Similar results are obtained studying locations of individual genes. Active genes tend to be located in the nuclear interior while inactive genes are found more often at the nuclear periphery [6, 14, 15], the converse behavior is observed in some experiments [16]. Moreover, a change in the transcriptional state of a gene can have direct influence on its positioning inside the nucleus [17, 18].

2.3.4 Folding of chromosomes into sub-nuclear domains

While the folding pathways of chromatin cannot be visualized by light optical methods, labeling two loci of a chromosome with a fluorescent marker was successfully used to establish a relationship between genomic distance g between these markers and its mean square physical distance in yeast [8], drosophila [9, 10] and human cells [11, 12]. This FISH technique has been applied on the q-arms of chromosomes 1 and 11 in primary human fibroblasts, because the Human Transcriptome Map shows that these chromosome arms contain pronounced gene dense and transcriptionally highly active regions, as well as gene-poor areas with low activity (Fig. 2.3A and [13]).

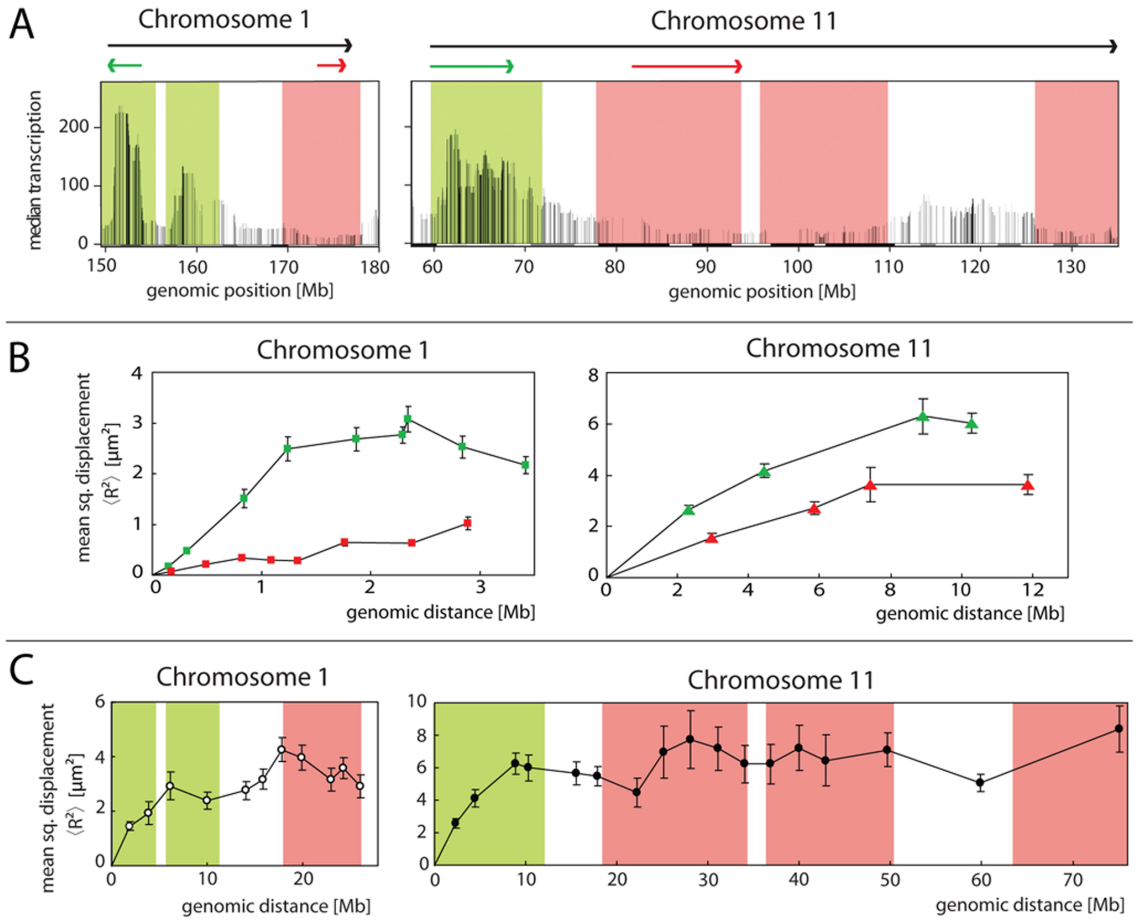


Figure 2.3: Experimental data. **A.** Domains of different transcriptional activity and gene density (ridges and anti-ridges) are shown on the Human Transcriptome Map of the q-arms of chromosomes 1 and 11. Each vertical line in the map represents a specific gene. The length of the line depicts its median transcription over a moving window of 49 genes. Ridges are indicated by green boxes, anti-ridges by red ones. The colored arrows above the map designate the ridge and anti-ridge regions where spatial distances between pairs of BAC probes were measured, using FISH. The tail of each arrow indicates the position of the reference BAC for that set of measurements. Physical distances of the reference BAC to loci at increasing genomic distances in the direction of the arrowhead were measured in 3D using confocal microscopy. **B.** Plots show the mean square physical distances $\langle R^2 \rangle$ as a function of the genomic distance for ridges (green) and anti-ridges (red) on chromosome 1 and 11 in the 0.5 to 10 Mb range. Data points in green and red correspond to the ridges and anti-ridges, respectively. Measurements were made corresponding to the colored arrows shown in panel A **C.** The mean square displacement $\langle R^2 \rangle$ is shown as a function of genomic distance in the 25 to 75 Mb range. Measurements were made corresponding to the black arrows shown in panel A. Error bars represent standard error. Image from Sandra Goetze.

Figure 2.3A shows the transcriptome map of the 1q and 11q areas. The starting points of the arrows above the maps indicate the positions of the reference FISH probes. The arrowheads mark the locations of the FISH probe having the largest genomic distance to the reference probe. All physical distances have been determined with respect to the reference probe. Green arrows and green data points refer to ridges, red ones to anti-ridges. Black arrows in Fig. 2.3A indicate long distance measurements beyond ridge and

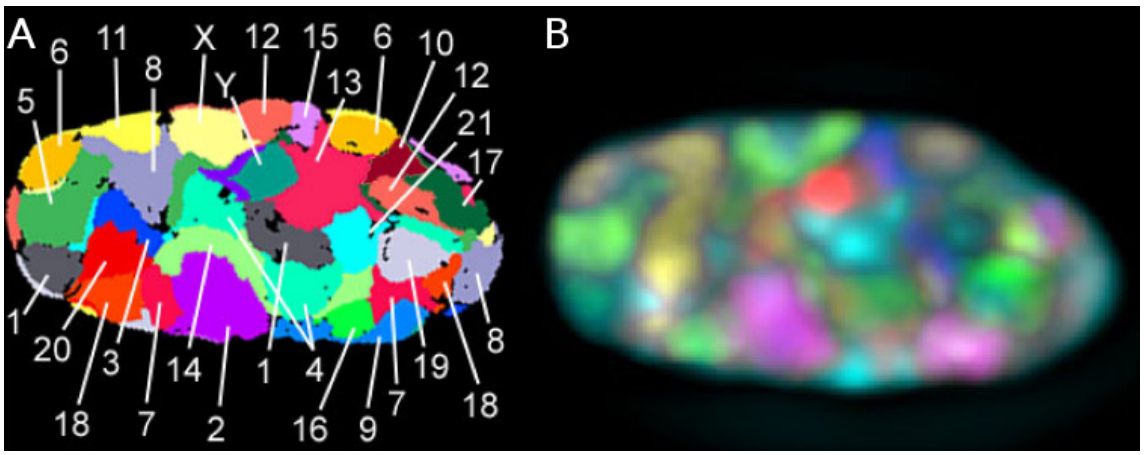


Figure 2.4: **A.** FISH (Fluorescence in situ hybridization) labeling of all 24 different human chromosomes (1–22, X and Y) in a fibroblast nucleus, each with a different combination of in total seven fluorochromes. Shown is a mid-plane of a deconvoluted image stack which was recorded by wide-field microscopy. **B.** False color representation of all chromosome territories visible in this mid-section after computer classification. Images adopted from Ref. [34]

anti-ridge domains. Physical distances were measured in 3D space between the centers of gravity of the 3D FISH signals of the individual BAC probes.

Plots of the mean square distance as a function of the genomic distance, covering a large part of the q-arm of chromosome 1 (27 Mb) and essentially the complete q-arm of chromosome 11 (75 Mb), are shown in Fig. 2.3C. Results show that the average physical distance to the reference probe does not increase at genomic distances beyond 3 to 10 Mb. The maximal distances are in the 1.5 to 2.5 μm range, similar to the size-range of chromosome territories and well below the diameter of the cell nucleus. The observed leveling off is most probably related to the limited space that chromosomes occupy in interphase, i.e. the chromosome territories [7]. Fig. 2.3B shows how the mean square physical distance to the reference FISH probe depends on the genomic distance for the ridge and anti-ridge domains on chromosome 1q and on chromosome 11q. Above about 3 Mb genomic distance the measured physical distances level off, similar as seen for long genomic distances (Fig. 2.3C). Average physical distances for anti-ridges are smaller than observed for ridges, reflecting their different degrees of compaction, agreeing with earlier measurements [6, 57].

2.3.5 The formation of chromosome territories

While chromosomes during metaphase condense into rod-like structures, interphase chromosomes are organized more like a fluffy polymer. Although one would expect from classical polymer theory that polymers in a semi-dilute or dense solution strongly intermingle, the contrary is observed for chromosomes inside the interphase nucleus. Indications for the existence of chromosome territories have been already found in the 1970ies using microbeam UV irradiation techniques [58]. Since the possibility of visualization of whole chromosomes via fluorescence in situ hybridization (FISH), unambiguous evidence was given that indeed chromosomes occupy distinct territories inside the cell [59, 33, 34]. The amount of intermingling between different chromosomes and its connection to transcriptional regulation are still under discussion, however, a recent experimental study suggests an overlap volume of about 20% [60]. Contradictory experimental evidence exists con-

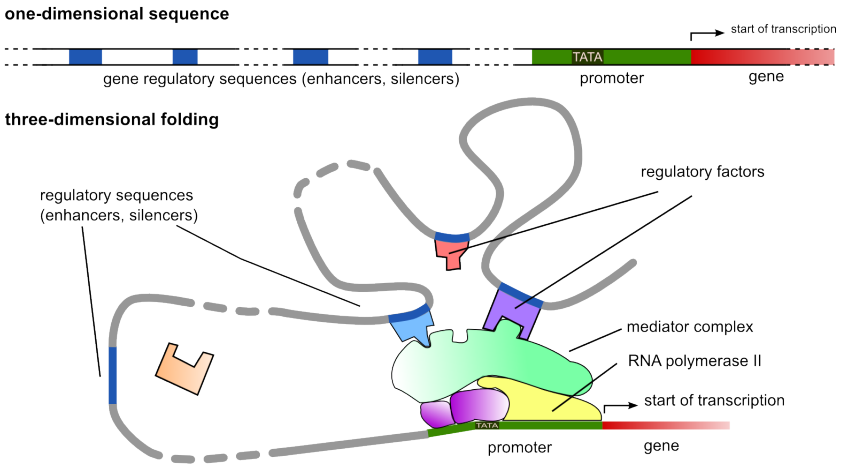


Figure 2.5: A cartoon of gene regulatory mechanisms. Transcription of a gene (red) requires a promoter where RNA polymerase II can bind to, often including the TATA box binding site. Throughout the genome, various enhancers or silencers are positioned controlling the expression level. In this biological model it is assumed that a complex of polymerase and regulatory proteins bind to the promoter, while the enhancers attach to the mediator complex. The intervening DNA thus is looping out.

cerning the regularity of chromosome positions inside the interphase nucleus. Conclusions range from the claim that chromosomes are highly ordered [61] to a random localization of territories [62].

2.4 Chromatin loops as a mediator of the folding-function relationship

Intra-chromatin contacts are presumed to act on the interface between genome folding and function. They will play the pivotal role for the polymer models developed in this thesis and it will be shown that experimental findings from a variety approaches can be consistently explained by models correctly incorporating chromatin looping. Here, evidence for long-range transcriptional control via looping and the corresponding techniques are described.

2.4.1 Gene regulation and long-range control

The smallest part of the human genome actually codes for genes (↷ 2.2.1). In earlier times it has been proposed that these non-coding regions are non-functional; recent years, however, have revealed, that such regions contain lots of regulatory sequences. Mapping of regulatory elements showed that their number is significantly larger than the number of genes [63]. In a simplified picture, five types of regulatory elements can be identified: promoters, enhancers, silencers, insulators and locus control regions. Transcriptional regulation requires the interplay of multiple regulatory elements as well as accessibility to specific genomic regions. Generally, a one-to-one correspondence between regulatory elements and target genes can not be established; rather, genes and regulatory elements form a complex network of interacting genomic elements [22].

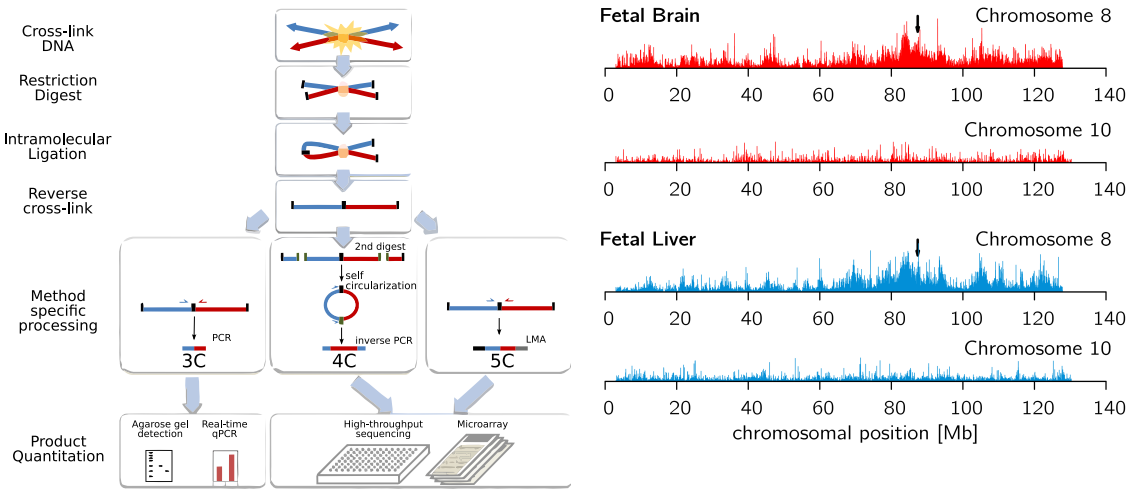


Figure 2.6: **A.** Sketch of 3C, 4C and 5C techniques (adapted from [40]). In the first step, two co-localized DNA regions (blue and red strand) are cross-linked. Using restriction enzymes the cross-linked fragments are separated and then ligated. After reversal of cross-linking further method-specific processing is done, fragments are amplified and analyzed. **B.** Results from 4C experiments [25]. Contact frequencies (y-scale) of genomic segments along chromosome 8 and 10 with the Rad23a locus (black arrow) are measured for liver and brain cells. Rad32a is an ubiquitously expressed DNA repair gene residing in a gene-dense cluster consisting of many housekeeping genes on chromosome 8. The data indicates that both long-range intrachromosomal interactions as well as interchromosomal interactions are abundant. Distinct differences are found between fetal and liver cells, indicating a regulatory function of chromatin contacts. Data has been kindly provided by Wouter de Laat.

Promoters are regions of DNA, which facilitate the transcription of a particular gene by providing binding sites for RNA polymerase II, the enzyme which transcribes a gene producing mRNA. They are located close to the transcription start site, typically about 10 to 100 basepairs upstream. For type II RNA polymerase to function properly, general transcription factors like chromatin remodeling complexes or activator proteins are required to initiate transcription [38].

Binding sites for activator proteins are provided by enhancer regions along the genome. Each gene can have several specific enhancer regions associated with it positioned upstream or downstream of the gene. Enhancers can be located many kb up to several Mb away from the gene they regulate [20] and are even found in gene deserts, regions spanning over 500 kb containing no genes [64]. Surprisingly, binding of an activator protein to an enhancer located distant to the target gene in terms of genomic separation can increase its transcription level. Just as enhancers facilitate gene expression, there exist regions of DNA, where regulatory factors can bind to, which downregulate transcriptional activity. These regions are called silencers.

A sketch of the possible interaction mechanism between gene regulatory sequences, transcription factors (specific proteins) and the promoter-gene region upon transcription is displayed in Figure 2.5.

2.4.2 Long-range control mediated by chromatin loops

The existence of enhancers located far away from the target gene they regulate raises the question, how interactions with this gene are accomplished. It has been suggested that

such a long-range control involves enhancer and promoter regions to come into physical contact [65]. Several novel experimental techniques have since then provided evidence, that indeed transcriptional control involves higher-order folding of the chromatin fiber into 3D structures, bringing regulatory elements into close contact. The intervening DNA then has to loop out.

At the heart of techniques studying long-range chromosomal contacts is what is called *Chromosome Conformation Capture (3C)*. It is a high-throughput molecular biology technique initially used to study the folding of yeast chromosomes [66], but successfully extended to the study of complex loci in mammalian cells [21]. 3C based techniques have become a standard tool to investigate structural properties and spatial organization of chromatin with respect to gene regulation.

The technique involves five experimental steps (for a review and discussion of limitations and problems refer to Ref. [67]). The initial step consists of formaldehyde cross-linking, fixating DNA-protein and protein-protein complexes and thereby cross-linking interacting chromatin segments. In a next step a restriction enzyme is added, which separates non-cross-linked DNA from the cross-linked one. Generally, four- or six-base-cutters are used, ensuring that all relevant regulatory elements are dissected. Then, cells are isolated, diluted and ligation enzymes are added. Thus, ends of cross-linked DNA segments are likely to be ligated. After reversal of the cross-linking, the resulting 3C library consists of linear DNA fragments with specific restriction ends (based on the choice of the enzyme) and a central restriction site corresponding to the site of ligation. In a final step, using polymerase chain reaction (PCR), the fragments of interest are amplified to semi-quantitatively assess the occurrence frequencies of certain restriction fragments.

Several improvements have been applied to the 3C technique. While 3C allows for the determination of interactions between two specific sites, the 4C technology [25] provides a complete interaction map between one specific site (the bait) and all other inter- and intra-chromosomal sites. Interactions in about one million cells can be determined at the same time. The 5C [68] and Hi-C [39] methods allow for establishing a complete two-dimensional map of interactions. A schematic overview of the techniques as well as results from 4C measurements are given in Figure 2.6.

Important findings from these experimental approaches are that intra-chromosomal interactions drop rapidly with increasing genomic separation. The half-width of the curve lying at about 20 kb for yeast [66] and being larger in mammalian cells. 4C experiments revealed that physical chromatin-chromatin interactions can be found on the scale of several tens of Mb [25], the interactions being cell-type specific. Long-range contacts have been confirmed by FISH experiments on the single cell level, showing that only a small fraction of cells (5-15%) actually forms a detected specific contact. Such a large cell-to-cell variation is typical in biological experiments and might be related to dynamic looping interactions, a hypothesis which will be assessed in this thesis. Recently, the Hi-C technique provided evidence that long-range contact probabilities on the scale between 500 kb and 7 Mb decay with a power-law $p_c(l) \sim 1.08$ [39].

2.4.3 Three-dimensional structure of loops: Transcription factories, CTCF and the nuclear matrix

While the connection between chromatin loops and transcriptional regulation could be well-established by 3C/4C/5C and FISH techniques, the detailed three-dimensional structure imposed on chromosomes by gene regulatory mechanisms still remains elusive. The picture of one gene associated with a complex machinery of regulatory elements and pro-

teins (Figure 2.5) is probably an oversimplification.

Several mechanisms have been proposed for the three-dimensional organization of chromatin loops. One recurrent idea is that genes and regulatory elements form so called transcription factories or hubs where regulatory elements and a high concentration of transcription factors assemble to initiate transcription [20]. It was conjectured that such an aggregation is driven by macromolecular crowding in a self-organized manner, the high concentration of transcription factors and genes facilitating gene expression [23]. It has been shown that the RNA polymerase II concentration is not uniformly distributed throughout the nucleus, rather certain foci of high concentration have been detected [69]. The number of these "transcription factories" visible in the cell nucleus is much smaller than the number of genes, rendering it likely that many genes are assembled in one transcription factory [70]. Importantly, such a formation of transcription factories is dynamic. It has been shown that individual genes are transcribed in pulses of production, temporary silenced genes being located far away from the transcription factory [71, 72].

Another structure which has been associated with the formation of chromatin loops is the so-called "nuclear matrix" or "nuclear scaffold". It is defined as the network of fibers found throughout the inside of the cell nucleus comparable to the cytoskeleton. Originally, a scaffold where loops are attached has been detected by electron microscopy of histone-depleted chromosomes [73]. Specific DNA regions called S/MARs and associated proteins are supposed to attach DNA to the scaffold. However, the existence of these structures *in vivo* is discussed controversially [74].

Recent investigations indicate the the formation of chromatin loops involves specific proteins, including SatB1 [51], insulator proteins and CTCF [52]. The CTCF protein [75] is evolutionary highly conserved in higher eukaryotes, indicating its vital importance. CTCF is a CCCTC-binding factor and the human genome contains of about 15 000 binding sites for this protein. It has been shown that CTCF is involved in the regulation of many genes, both concerning activation and repression [76]. More importantly, CTCF seems to play an important role in chromatin looping. Indeed, it has been observed that CTCF is required for looping in the beta-globin cluster [77].

2.4.4 Implications for a model of chromatin

The few sections on the biology of the cell might have provided the reader with a short and rather tiny selection of the abundant experimental data available. Most of the data are very locus-specific, studying for example the influence of certain proteins on specific genes. Clearly, no model of chromatin can capture all the details of such a complex system including amongst others heterogeneous gene content, protein activity or ATP-mediated non-equilibrium effects. In the next chapter (↷ chapter 3) we will argue why a coarse-graining of this system is inevitable. Especially, the abundance of specific and non-specific protein-DNA interactions requires a more abstract view on chromatin folding. The key principles a polymer model should include can be summarized in the following way

- Chromatin loops are an abundant regulatory and folding motif and exist on all scales
- Chromatin looping is dynamic and cell-type specific, i.e. different looping configurations are found from cell to cell
- Chromatin looping is related to transcriptional activity

In this thesis, polymer models will be proposed and compared covering the experimental evidence for loops and specific folding mechanisms.

Chapter 3

Polymer models of chromatin

In this chapter, basic concepts of polymer physics are addressed. The first sections are meant to give the reader an overview over basic concepts in and theoretical foundations of the physics of large polymeric macromolecules. For a comprehensive and detailed review of polymer physics we want to refer the reader to Refs. [78, 79, 80]. The second part of this chapter is dedicated to presenting the most prominent polymer models that have been used to explain higher-order folding motifs of chromatin. We will discuss the assumptions, the results and the short-comings of each of these models. It should become clear that none of the polymer models proposed so far offers a functional and unified framework of chromatin organization comprising all experimental evidence listed in section 1.2. The reader familiar with polymer physics and chromatin models might want to skip this chapter.

3.1 Statistical Physics of Polymers

A polymer is a macromolecule composed of a large number of small chemical units, aligned in a repetitive sequence by chemical interactions. The small chemical units are called "monomers", the number of monomers constituting a polymer is called the "chain length" or degree of polymerization N.

The degree of polymerization can be huge; we have seen in section 2.2 that the degree of polymerization of a typical human chromosome is in the order of $N \approx 10^8$ base pairs. Given this large number of particles, it is evident that concepts of statistical physics have to be applied to study the properties of such polymers. Characteristic measures are given as mean values, displaying an ensemble average over the huge conformational space one single polymer can explore.

We will present the freely-jointed chain model ($\curvearrowright$ 3.1.1) to introduce typically used measures for polymeric complexes. Then, the Gaussian chain model is presented ($\curvearrowright$ 3.1.2), which lies at the heart of the Random Loop Model (chapter 5) and many other chromatin models [26, 11]. The behavior of polymer chains with excluded volume interactions is developed ($\curvearrowright$ 3.1.3), which will be required in chapter 6. The effect of the solvent and the transition to a globular state polymer, which has been proposed for chromatin organization, is introduced ($\curvearrowright$ 3.1.4). Finally, the principles and theory of coarse-graining

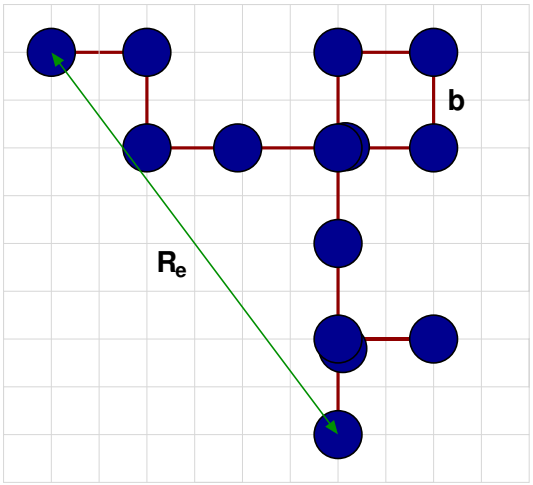


Figure 3.1: A model polymer of the freely jointed chain (FJC) model. The beads are represented by blue circles, which are connected by linkers of fixed width b. The coordinates are constrained to a lattice in this illustration. The FJC model allows beads to overlap, which happens two times in the sample conformation.

and scaling, which are necessary for every large-scale description of a long polymer are emphasized (↷ 3.1.6).

3.1.1 The freely jointed chain model

The simplest polymer model which can be thought of is the *freely jointed chain* model. Let the positions of the monomeric units be denoted by $\mathbf{r}_1, \ldots, \mathbf{r}_N$. Adjacent monomers $\mathbf{r}_i$ and $\mathbf{r}_{i+1}$ are connected by a rigid linker of length b. There are no constraints concerning the angle between two linkers imposed on the system and two monomers are allowed to overlap. An illustration is shown in Fig. 3.1, where, for simplicity, the configurational space is limited to positions on a lattice.

The spatial extend of the polymer is given by its *end-to-end distance*

$$\mathbf{R}_e = \mathbf{r}_N - \mathbf{r}_1 = \sum_{i=1}^{N-1} \mathbf{b}_i \,,$$

where $\mathbf{b}_i = \mathbf{r}_{i+1} - \mathbf{r}_i$ denotes the bond vector between the monomeric units i and $i+1$.

The *mean squared end-to-end distance* $\left\langle \mathbf{R}_e^2 \right\rangle$ can be easily calculated,

$$\left\langle \mathbf{R}_e^2 \right\rangle = \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} \langle \mathbf{b}_i \mathbf{b}_j \rangle = \sum_{i=1}^{N} \left\langle \mathbf{b}_i^2 \right\rangle = Nb^2 \tag{3.1}$$

as the correlations between different bond vectors vanish due to the models assumptions ($\langle \mathbf{b}_i \mathbf{b}_j \rangle = b^2 \langle \cos\theta \rangle = 0$ for $i \neq j$).

The mean squared end-to-end distance represents the typical size of a polymer, and we just found that this size is proportional to $N^{0.5}$. In fact, this does not come as a big surprise as there is a one-to-one connection between the freely jointed chain and Brownian motion. Consider a particle at position $\mathbf{r}_1$ at time $t_1 = 0$. In each time step, the particle moves a certain distance b without memory of its covered path so far (Brownian motion). After N time steps, the particle has moved on average a distance $\langle R^2 \rangle = b^2 N$. If we track the positions of the particle at the intermediate times, $\mathbf{r}_1, \ldots, \mathbf{r}_N$, and connect the path of the particle, the resulting trajectory is a freely jointed chain. Due to this correspondence, polymer chains without excluded volume interactions (↷ 3.1.3) are called *random walks*.

3.1.2 The Gaussian Chain model

In contrast to the assumptions of the freely jointed chain model, chemical bonds possess a certain intrinsic flexibility, resulting in a fluctuating bond distance b. A model, which is often used to model random walk polymers and which overcomes this problem, is the *Gaussian chain* model. This model assumes that each bond vector $\mathbf{b}$ follows a Gaussian distribution,

$$p(\mathbf{b}) = \left(\frac{3}{2\pi b^2}\right)^{3/2} \exp\left(-\frac{3\mathbf{b}^2}{2b^2}\right) . \tag{3.2}$$

The conformational distribution function of a chain of length N is then given by

$$P(\mathbf{r}_1, \ldots, \mathbf{r}_N) = \prod_{i=1}^{N-1} p\left(\mathbf{r}_{i+1} - \mathbf{r}_i\right) . \tag{3.3}$$

The Gaussian distribution of the bond length $p(\mathbf{b})$ can be thought of as a system of beads connected by harmonic springs. Thus, the Gaussian chain is equivalent to a many-body system with interaction energy

$$U(\mathbf{r}_1, \ldots, \mathbf{r}_N) = \frac{1}{2}\kappa \sum_{i=1}^{N-1} \left(\mathbf{r}_{i+1} - \mathbf{r}_i\right)^2 . \tag{3.4}$$

The equilibrium probability distribution of the monomer positions is given by the Boltzmann factor $P(\mathbf{r}_1, \ldots, \mathbf{r}_N) = \exp(-U/k_BT)$, the correspondence to eq. (3.3) is obtained by setting

$$\kappa = \frac{3}{b^2} k_B T .$$

An important measure for any polymer model directly related to FISH distance measurements between two markers (cf. section 2.3.4) is the distribution of the vector $\mathbf{r}_n - \mathbf{r}_m$ between any two beads n and m.

$$P(\mathbf{r}_n - \mathbf{r}_m) = \int d\mathbf{r}_1 \ldots \int d\mathbf{r}_N P(\mathbf{r}_1, \ldots, \mathbf{r}_N) \delta\left(\mathbf{r}_n - \mathbf{r}_m\right) .$$

This quantity can be calculated easily for a Gaussian chain [81],

$$P(\mathbf{r}_n - \mathbf{r}_m) = \left(\frac{3}{2\pi b^2 |n-m|}\right)^{3/2} \exp\left(-\frac{3}{2b^2|n-m|}\left(\mathbf{r}_n - \mathbf{r}_m\right)^2\right) . \tag{3.5}$$

A direct conclusion is that the conformational average is given by

$$\left\langle \left(\mathbf{r}_n - \mathbf{r}_m\right)^2 \right\rangle = |n-m| b^2 . \tag{3.6}$$

Specifically, for the end-to-end distance we find

$$\left\langle R_e^2 \right\rangle = N b^2 . \tag{3.7}$$

Not surprisingly, this result is the same as for the freely jointed chain model [eq. (3.1)]: The detailed form of the short-range interaction potential represented by the bond vector distribution $p(\mathbf{b})$ [eq. (3.2)] does not influence large-scale quantities like the mean squared end-to-end distances. The central limit theorem in statistics [82] guarantees that the distribution of the end-to-end distance $\mathbf{R}_e = \sum \mathbf{b}_i$ converges to a Gaussian distribution for independent bond vectors $\mathbf{b}_i$ in the limit of large N.

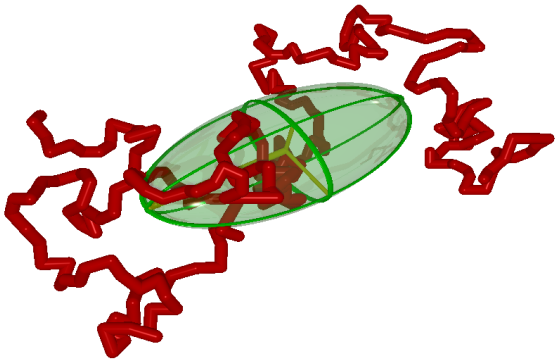


Figure 3.2: The gyration ellipsoid of a polymer. The gyration ellipsoid describes the three-dimensional distribution of monomers in space. Both the typical size and shape of the polymer can be derived from the tensor's invariants.

For looping polymers which we will introduce later, it is convenient to use the *radius of gyration* R_g as a measure of size. It is generally given by

$$\left\langle R_g^2 \right\rangle = \frac{1}{2N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left\langle (\mathbf{r}_i - \mathbf{r}_j)^2 \right\rangle \,. \tag{3.8}$$

For the Gaussian chain we find

$$\left\langle R_g^2 \right\rangle = \frac{1}{2N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} |i-j| b^2 = \frac{1}{6} N b^2 \,. \tag{3.9}$$

In fact, the mean squared radius of gyration is an invariant of the gyration tensor, which measures the distribution of monomers in the system. It is defined by

$$S_{mn} = \frac{1}{N} \sum_{i=1}^{N} x_m^{(i)} x_n^{(i)} \,. \tag{3.10}$$

Here $\mathbf{x}^{(i)}$ is the coordinate vector of the ith monomer with respect to the polymer's center of mass and the subindex denotes its Cartesian components. The eigenvalues $\lambda_1 \leq \lambda_2 \leq \lambda_3$ correspond to the squared lengths of the principal axes of gyration. The squared radius of gyration is the trace of the gyration tensor

$$R_g^2 = \lambda_1 + \lambda_2 + \lambda_3 \,.$$

The ratios of the eigenvalues indicate the deviation from a sphere-like shape of the polymer. It is well-known that the gyration tensor for random walks in good solvent has a pronounced asphericity [83, 84], given by $\langle \lambda_3 \rangle : \langle \lambda_2 \rangle : \langle \lambda_1 \rangle \to 12 : 2.7 : 1$. Figure 3.2 shows an illustration of a polymer chain and its gyration ellipsoid, clearly revealing an anisotropic shape.

Ideal chains, as the universality class of models without excluded volume interactions are often called, exhibit certain scale invariant features, which can be used for comparison to experimental data from biology. First, the proportionality of the characteristic size of a polymer coil to chain length N is independent of the bond length or the specific distribution of bond vectors $p(\mathbf{b})$. Second, the ratio between the mean squared end-to-end distance and the radius of gyration is always

$$R_e^2 / R_g^2 = 6 \,,$$

More importantly, the Gaussian character of the distribution function eq. (3.5) implies for example that

$$\left\langle R^4 \right\rangle / \left\langle R^2 \right\rangle^2 = 5/3,$$

a value independent of any length scale involved in the model. We will make use of these quantities in chapters 4-10.

3.1.3 Effects of excluded volume – the self-avoiding walk

The polymer models presented in the last sections did not take into account interactions other than short-range bonded ones. As a result, the model permits the chain to fold back onto itself, allowing beads to occupy the same position in space. In physical systems, this is impossible as each monomer occupies a finite volume. In a lattice system such as the one presented for the freely jointed chain model in Fig. 3.1, it is easy to take into account excluded volume interactions: We subject any polymer to the condition that each lattice site is allowed to be occupied by maximally one monomeric unit. Such a model is called the self-avoiding walk, and there exist a lot of Monte Carlo algorithms for unbiased sampling of these polymers. The Verdier-Stockmayer [85] algorithm is based on local moves of the monomers and thus one Monte Carlo step can be easily related to physical time scales of the system allowing exploration of the dynamical behavior. More efficient for sampling equilibrium self-avoiding walk conformations of large chains is the pivot algorithm [86, 87], where global conformational changes are induced in one Monte Carlo step.

The size of a chain with excluded volume is larger than that of a corresponding ideal polymer. This can be easily seen: The more compact the polymer is, the larger is the possibility of overlap. Thus, the configurational space is restricted with a bias towards more open conformations compared to ideal chains. In 1949, Flory devised a simple argument [88] to estimate the size of a swollen chain by considering two counteracting forces. Firstly, the free energy contribution of constraining the end-to-end distance vector to $\mathbf{R}$ is given by (according to eq. (3.5))

$$F_1(\mathbf{R}) = -k_B T \ln P(\mathbf{r}_N - \mathbf{r}_1) = k_B T \frac{3}{2Nb^2}\mathbf{R}^2 \,. \tag{3.11}$$

While F_1 penalizes a strong elongation of the chain, excluded volume interactions act against a strong compression of the polymer. In a mean field argument only considering two-body interactions, this free energy contribution F_2 is proportional to the square of the local concentration $c \simeq N/R^3$ and the strength of the excluded volume interaction v_c,

$$F_2(\mathbf{R}) \simeq k_B T v_c c^2 R^3 = k_B T v_c \frac{N^2}{R^3} \,. \tag{3.12}$$

Minimizing the free energy $F = F_1 + F_2$ with respect to R yields a scaling of the average size of the polymer with

$$R \sim N^{3/5} \,.$$

Thus, a general scaling law arises for random walks as well as self-avoiding walks,

$$\left\langle R_e^2 \right\rangle \sim N^{2\nu} \,. \tag{3.13}$$

The exponent $\nu = 0.5$ for a random walk is exact. The exponent for a self-avoiding walk has been estimated by perturbation calculations to $\nu = 0.588(1)$ [89], a value quite close to the crude estimate by Flory.

3.1.4 Effects of the solvent – transition to the globular state

Normally, a polymer is dissolved in some solution, i.e. small molecules compared to the polymer's size surrounding the chain. The models discussed so far do not take into account the effect of the solvent. This assumptions correspond to the *athermal* situation where the interactions between the solvent molecules and the chain molecules can be neglected. Normally this is not the case. DNA, for example, is a highly negatively charged polymer (it carries $2e^-$ per Å [1]) and interacts with counterions from the solvent. Nevertheless, the use of athermal models is often justified. The negative charge in DNA is screened by counterions with a screening length of about 10 nm [1], i.e. on a very small scale compared to the chromatin fiber's extension. To take into account the effect of the solvent, interaction energies between neighboring elements are introduced: $-\epsilon_{pp}$ (polymer segment – polymer segment), $-\epsilon_{ps}$ (solvent molecule – polymer segment) and $-\epsilon_{ss}$ (solvent–solvent interaction). It can be shown [80] that the resulting interaction term in the mean field free energy has the same functional form as in the athermal case, eq. (3.12),

$$F_2 = k_B T v \frac{N^2}{R^3} \,,$$

where the excluded volume parameter v comprises the additional interactions

$$v = v_c(1 - 2\chi), \qquad \chi \simeq k_B T \left[\frac{1}{2}(\epsilon_{pp} + \epsilon_{ss}) - \epsilon_{ps} \right] .$$

Without going into detail, it is clear that the behavior of the system changes at $\chi = \frac{1}{2}$, where the free energy changes sign. Indeed, for large positive χ-values, the monomer-solvent interaction dominates and the system tries to minimize the polymer-solvent surface. A polymer in such a "bad solvent" adopts a sphere-like conformational state where the mean-square end-to-end distance is solely dependent on the size of the system, i.e.

$$\left\langle R_e^2 \right\rangle \sim N^{2/3} \qquad \text{(globular state)}. \tag{3.14}$$

Again, we find the universal scaling law eq. (3.13) with an exponent of $\nu = 1/3$. In case of low χ-values, the solvent-monomer interaction is weak, thus the polymer adopts the swollen state known from the self-avoiding walk, the scaling exponent becoming $\nu \approx 0.588$.

When increasing the temperature, i.e. the interaction parameter χ, the polymer undergoes a transition from the compact globular state ($\nu = 1/3$) to the swollen state ($\nu = 0.588$). The transition temperature for this *coil-globule transition* is $\chi = 1/2$. The transition point is characterized by the chains following the ideal gas law, i.e. an exponent of $\nu = 1/2$. This behavior has been explained by an exact counterbalance of the entropic gain of chain expansion and the energy loss of volume link attraction at the transition point [79].

3.1.5 Intermingling in polymeric solutions

An interesting question concerning chromatin organization is why chromosomes arrange into distinct territories (cf. sections 1.2 and 2.3.5). An enormous amount of information about a concentrated solution of polymers in solution can already be obtained by scaling theory [78] such as phase separation, fluctuations and osmotic pressure. I will not go into the details here, which can be read everywhere [80, 78], rather the main conclusion concerning intermingling is presented.

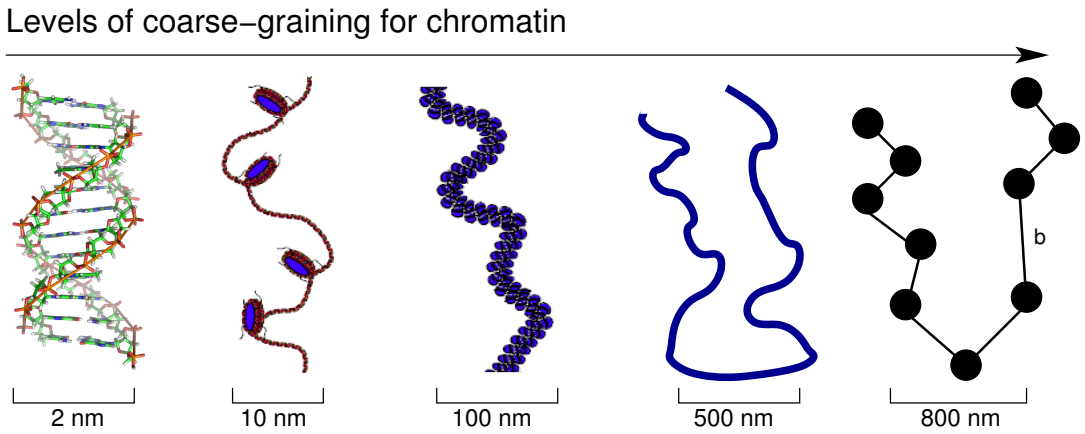


Figure 3.3: Levels of coarse-graining for chromatin. The basic filament is the DNA double helix. Its detailed structure becomes negligible when looking at the chromatin fiber. On a scale above 100 nm, the chromatin fiber itself can be viewed as a flexible string, which in turn is approximated by a bead-linker model on a scale above the persistence length l_p of the flexible string. It should become clear that for large-scale quantities, such a bead-linker model is fully sufficient and molecular details can be effectively subsumed into the average bead length b (partly adapted from [40]).

Consider a number of polymers in solution (good solvent, $\chi < 1/2$), the concentration being given by c. If the concentration is very small, the polymer coils do not overlap and display their ideal behavior (dilute regime). At a certain concentration threshold $c^\star$, the coils begin to overlap. This threshold is reached when the global concentration c becomes comparable to the local concentration inside one polymer, i.e.

$$c^\star = \frac{N}{R^3} \sim N^{1-3\nu} \, .$$

Importantly, for large N, the overlap concentration becomes very small. Consider the case of chromatin: We assume in a coarse-grained model the Kuhn segment length to be 300 nm [30]. In a densely packed 30 nm fiber, this corresponds to a chromatin stretch of about 30 kb [90]. Given the typical size of a human chromosome of 100 Mb, the effective chain length of flexible monomeric units is $N = 3400$. The overlap density is thus

$$c^\star \sim 1.5 \times 10^{-3} \, .$$

Given a concentration of chromatin of about 10% inside the cell nucleus we can see that the chains should indeed be strongly overlapping.

3.1.6 A question of scale – Coarse graining and bending rigidity

Each polymer has a certain rigidity, inducing a cost of energy to bend it strongly on a short scale. The extreme case is a rigid rod, like a ski stick, which can hardly be bent at all (of course it can, luckily, but by far not as strong as a parcel string). The freely jointed chain is basically composed of N connected rigid rods, which at their attachment points are fully flexible. This behavior is seldom found in nature. Rather, on an atomistic level and for small deformations, classical elastical theory can be applied using Hooke's deformation law. This model is called the *worm-like chain model*. While the freely jointed chain model assumes that the correlation between bond vectors vanish, i.e. $\langle \mathbf{b}_i \mathbf{b}_j \rangle = b^2 \langle \cos \theta_{ij} \rangle \ (i \neq j)$,

this is not the case for a worm-like chain. For this model and for many others, the correlations decay exponentially with the contour separation $s = |i - j|$,

$$\langle \cos\theta(s) \rangle = \exp(-s/l_p)\,.$$

The factor l_p characterizes polymer flexibility and is called the *persistence length*.

Double-stranded DNA has a persistence length of about $l_p = 50$ nm. The physical meaning of this length is that below this length scale, the polymer has a memory of the chain direction, while above it does not. Thus, the memory of a direction prevails only for a finite distance on the order of l_p. Polymers which are much larger than the persistence length are flexible. In fact, a long macromolecule contains basically $N_{\text{eff}} \approx L/l_p$ flexible sections. Here L denotes the contour length of the polymer. Imagine you are traveling along the contour of this polymer and put a marker on it every time a distance of $\sim l_p$ is covered. Take these markers as the beads of a polymer, which are connected by rigid linkers. From the considerations of the Gaussian chain we find that

$$\left\langle R_e^2 \right\rangle \sim l_p^2 N_{\text{eff}} \sim (L/l_p) l_p^2 \sim L l_p\,.$$

Thus, each sufficiently long macromolecule can be described by a simple polymer model above the scale where bending rigidity plays a role.

Actually this comes quite handy, as modeling a chromosome with all one hundred million basepairs in detail would not be great fun and probably prolong the time to get a PhD considerably. The method of going from a detailed description to a large-scale description is called *coarse-graining*. In fact, it is due to the general validity of the Gaussian chain model (besides excluded volume interactions) for long polymers that this model is often called the "standard model of polymer theory" [79].

Moreover, it is not even necessary to stop the coarse-graining procedure on the length scale of the persistence length l_p. The presumed persistence length of chromatin is on the order of 40-250 nm [91], thus using a packing density of about 10 nm/kb [90] would still require simulated chain lengths of $N \approx 4000 - 25000$. Gaussian chains display an interesting scaling behavior. Starting from a description with N monomers at positions $\mathbf{r}_1, \ldots, \mathbf{r}_N$ and an average bond length b, we pick out every J^{th} monomer at positions $\mathbf{r}_0, \mathbf{r}_J, \ldots, \mathbf{r}_{nJ}, \ldots, \mathbf{r}_{N^\star J}$ with $N^\star = N/J$. Owing to the central limit theorem these monomers again form a Gaussian chain and the large-scale properties can be described on an equal footing using a rescaled segment length $b^\star$ obtained by setting

$$b^{\star^2} N^\star = b^2 N\,.$$

Figure 3.3 displays the transition from the DNA double helix to a more coarse-grained description of the fiber. Looking at the 10 nm chromatin fiber, it seems irrelevant which structure prevails on the scale of the base pairs. Looking on the fiber from an even wider perspective, the nucleosomes turn out to be irrelevant and the view as a flexible polymer emerges.

3.2 Models of chromatin

Now, we have a basic understanding of polymer physics, and we know that polymer models obey simple scaling laws. In the following sections, more complicated polymer models that have been proposed for explaining chromatin folding are presented. All of them are somehow based on the simple models, i.e. the random walk, self-avoiding walk or globular state model.

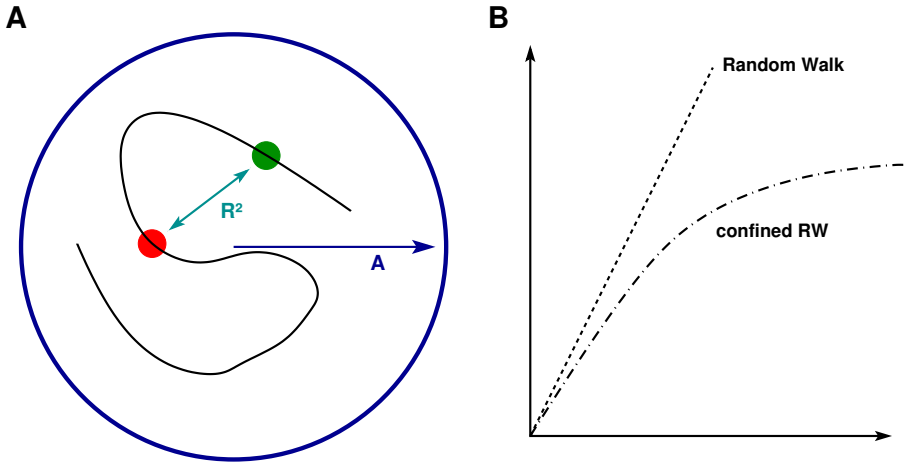


Figure 3.4: Sketch of a random walk in a confined space. **A.** In the model by Hahnfeldt et al. [26], chromatin is modeled as a random walk confined by a sphere of radius A. The distance R^2 between two markers (red and green label) are measured. **B.** Qualitative plot of the results. For distances smaller than the radius of the confining sphere A, the model displays a random walk behavior. When genomic separations get larger and the distance R^2 between two markers is in the order of the sphere radius, the mean square distance adopts a plateau level.

3.2.1 Random walk in a confined space

Fluorescent in situ hybridization (FISH) has allowed the determination of the physical distance between two chosen points on one specific chromosome for a large number of cells. The resulting relation between the mean squared physical distance $\langle R^2 \rangle$ and genomic separation g between the markers can be compared to polymer model predictions, e.g. the universal scaling law eq. (3.13). In 1992, it was proposed that chromatin organization on the scale below 2 mega basepairs (Mb) can be described by a random walk model [92]. The model is based upon 2D-FISH measurements of interphase chromosomes in the size region of 100 kb to 2000 kb.

Based on identical measurements up to 4 Mb, Hahnfeldt et al. [26] proposed another model for the overall organization of chromatin. This model roughly takes into account the geometry of the nucleus by modeling a chromosome as a Gaussian polymer subjected to forces that confine it to a sphere of radius A (see Fig. 3.4A). Such a model predicts that on scales from 0.1 Mb to 1.5 Mb, chromatin folding is equivalent to an unconstrained random walk model. On scales larger than 1.5 Mb, the model predicts a leveling-off as the mean square distance between two FISH probes feels the confinement of the sphere. The asymptotic value of the mean square displacement $\langle R^2 \rangle$ clearly is in the order of the sphere radius A. A sketch of the relation between mean square displacement and contour length in comparison to a random walk is shown in Fig. 3.4B.

A comparison of the model to recent experimental FISH data (Fig. 2.3) shows that the leveling-off is predicted correctly. However, the plateau level observed in experiments is far below the typical diameter of the cell nucleus, whereas the model predicts a leveling off at a physical distance comparable to the nuclear diameter. Furthermore, this model does not take into account experimental evidence for loops and is in disagreement with the observed large cell-to-cell variation [12, 93].

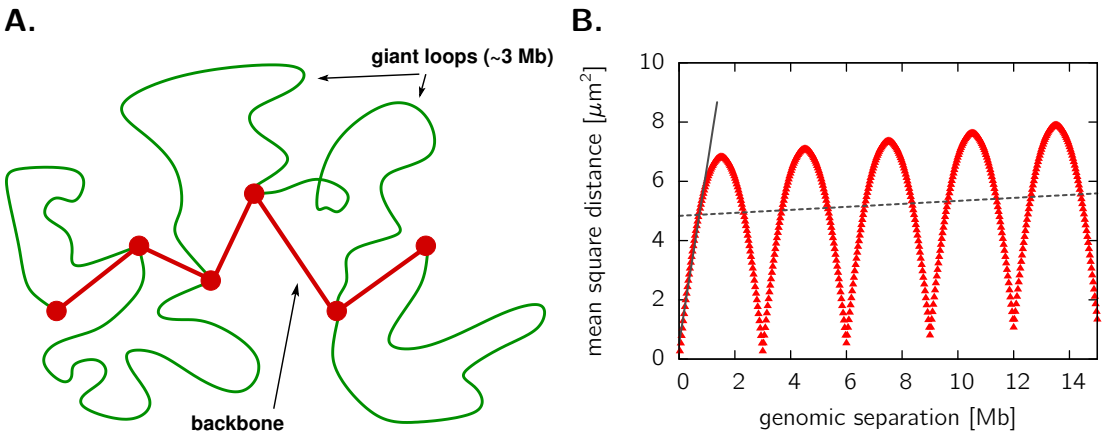


Figure 3.5: Illustration of the Random Walk / Giant-Loop (RWGL) model. **A.** Sketch of the RWGL model. The chromatin fiber (green) is attached to a random walk backbone (red), which might be related to the nuclear matrix/scaffold. In between the attachment points, the chromatin fiber forms loops of about 3 Mb. **B.** Numerical results for the mean square distance $\langle R^2 \rangle$ between two beads separated by genomic separation g. The loop-size is 3 Mb, the Kuhn segment length is assumed to be 300 nm, corresponding to 30 kb.

3.2.2 The Random Walk / Giant - Loop model

Yokota et al. conducted FISH measurements in human G_0/G_1 cells spanning the whole size range of human chromosome 4 [27]. The physical distance between two FISH probes was detected on a small scale ($0.15 - 3.5$ Mb) as well as on the scale of the whole chromosome (up to 200 Mb). These measurements revealed a biphasic behavior with two different regimes of structural chromatin organization, the transition region being at about 2 Mb. On both scales, a random-walk behavior was proposed, whereas the slope, i.e. the effective segment length, is much smaller on the scale above 2 Mb.

To explain the observed folding behavior, Sachs *et al.* [11] proposed a model where flexible chromatin loops of size 3 Mb are attached to a random-walk backbone. This backbone has been related to the hypothesis of a flexible nuclear matrix/scaffold (↷ 2.4.3) to which DNA and chromatin fibers are attached. A sketch of the model is displayed in Fig. 3.5A. Consider a chain of length N, the monomer positions being denoted by $\mathbf{r}_1, \ldots \mathbf{r}_N$. The backbone is given by a Gaussian chain (cf. section 3.1.2), i.e. the interaction potential is set to

$$U_0 = \frac{1}{2}\kappa_0 \sum_{i=1}^{N-1} \| \mathbf{r}_{i+1} - \mathbf{r}_i \|^2 \quad .$$

Giant loops are introduced by harmonic interactions between any two beads nJ and $(n+1)J$ where n is an integer and J determines the loop size. Thus, the complete interaction potential is given by

$$U = U_0 + \frac{1}{2}\kappa \sum_{m=1}^{N/J} \| \mathbf{r}_{mJ} - \mathbf{r}_{(m+1)J} \|^2 \quad .$$

Obviously, such a model predicts a random-walk type of behavior on the scale below the size of a loop J. Typical oscillations are induced by the loops in the mean square distance vs. genomic separation plots (see Fig. 3.5B). They are an artifact from assuming loops of a fixed size of 3 Mb.

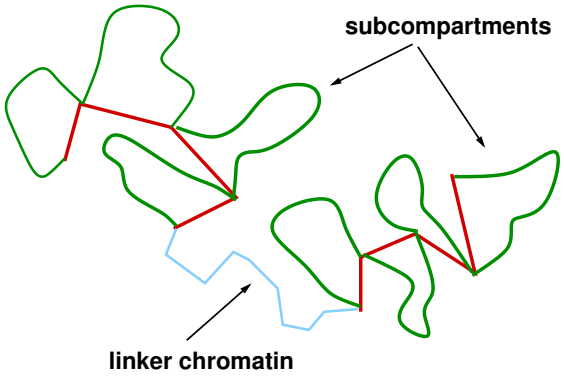


Figure 3.6: Sketch of the Multi-Loop-Subcompartment model. In this model, chromatin is subdivided into subcompartments consistent of 120 kb sized loops. The subcompartments are connected by linker chromatin.

On a large scale, the folding of the chromatin fiber can be described by a simple random walk, which is given by the backbone. Loops perturb this behavior by the typical oscillations, but nevertheless, the mean square distance is on average linear in the genomic separation. Speaking in terms of scaling theory, the random walk has been effectively rescaled with the loops being the effective monomers.

The RWGL model has several short-comings. Amongst others, the already mentioned oscillations are model artefacts. Secondly, loops of fixed size are not found in experimental data, rather 4C experiments revealed loops on all length scales [25]. More importantly, while the model is in agreement with data from 2D FISH, it does not display the strong leveling-off observed in recent 3D FISH experiments [12]. Measurements of Trask and coworkers [26, 11, 27, 94] did not show such leveling off of physical distances at large genomic distances. At least in part, this discrepancy can be explained by the fact that these authors used different cell fixation and FISH labeling methods, which preserve the structure of the nucleus less well than those used in Ref. [12]. Also, most measurements have been carried out two-dimensionally. Together, this is likely to results in systematic distortions of their data sets.

Sachs *et al.* [11] derive a Rayleigh probability density for the mean square distance between two markers. Such a probability distribution predicts the ratio between the standard deviation σ and the distance $\langle R \rangle$ to be a constant: $\sigma / \langle R \rangle \approx 0.94$. In fact, the experimental data show a similar ratio. We will make frequent use of such ratios in later chapters. A result will be that deviations from the random-walk polymer models only show up in the ratios of higher-order moments starting from the 4^{th} moment (cf. section 4.5).

3.2.3 The Multi-Loop-Subcompartment model

Early studies indicated the existence of loops with a size of about 100 kb [95], while giant loop structures as predicted by the RWGL model have not been found [28]. Ongoing research activity focused on the territorial organization of interphase chromosomes. It was shown by microbeam UV radiation [58] and later by fluorescence labeling techniques [7] that chromosomes indeed tend to be segregated into distinct territories (↷ 2.3.5). The Multi-Loop-Subcompartment (MLS) model [28, 29] was proposed to connect these experimental findings with a computational model.

Chromatin is organized into subcompartments according to the MLS model. Each subcompartment consists of several 120 kb sized loops. The subcompartments itself are connected with each other by chromatin fragments in the size range of the intra-compartment loops, i.e. about 120 kb (see Fig. 3.6). It was shown [29] that different subcompartments do not overlap to a large extend, a finding which is in agreement with a FISH study [6] re-

vealing little intermingling between chromatin regions on the same fiber. The MLS model predicts interphase chromosomes to be in a rather compact state on the scale above 10 Mb, i.e. the mean square displacement between two markers scales with genomic distance g as $\left\langle R_g^2 \right\rangle \sim g^{1/3}$. Although it was concluded in Ref. [29] that this state corresponds to a condensed globular state as presented in section 3.1.4, this is not true. We will find in a detailed study in chapter 4 that this scaling law is only valid for end-to-end distances. On a short scale below 10 Mb, the MLS model predicts a random-walk type of folding.

The dynamic formation of rosette-like structures was investigated by Odenheimer *et al.* [96]. In this study, chromatin was modeled as a regular block copolymer with attractive and repulsive sequences. Clearly, starting from a linear chain, attractive sequences arrange into rosette-like structures in this model, the repulsive segments looping out. By variation of the distance between attractive segments, the loop size could be varied. The best agreement with experimental data was found for 120kb-sized loops.

The short-comings of the MLS model are with respect to experimental data from 3D FISH measurements (Fig. 2.3 and Ref. [12]) that it does not predict a leveling-off on the large scale. Furthermore, the evidence for loops on all scales is not incorporated into this model.

3.2.4 Other models

In 1994, Sikorav and co-workers [97] discussed the process of mitotic chromosome condensation. They modeled the transition from interphase chromosomes to condensed mitotic chromosomes as a coil-globule transition of molten polymers. A simple estimate of the disentanglement time τ_d for a typical human chromosome (100 Mb) yields $\tau_d \approx 500$ years. It was concluded that a the enzyme topoisomerase II is required substantially to speed up this process. Rosa and co-workers [30] reversed this argument and suggested that interphase nuclei never equilibrate. They argue based on simulations of long linear chain molecules that chromosomes behave like solutions of unentangled ring polymers, which display a segregation due to topological constraints. The experimental evidence for loops has not been included into this model. Furthermore, the worm-like chain polymers in this study do not display the sub-diffusional regime of experimental data with $g_1(t) \sim t^{0.4}$, rather the typical dynamics of polymers in the melt of $g_1(t) \sim t^{0.25}$ is found.

Emanuel *et al.* [31] proposed recently, that chromatin organization can be explained by a globular state polymer model. Indeed simulations on the scale below 5 Mb are in agreement with experimental data from Fig. 2.3. Naturally, compact polymers display a leveling-off in the mean square displacement due to the assumption of a confined geometry. In chapter 4 we will investigate thoroughly whether the globular state model can explain experimental data.

Grosberg [98] and recently Lieberman-Aiden and co-workers [39] proposed that chromatin is organized as a fractal globule. In contrast to the equilibrium globular state model (↷ 3.1.4), such a model represents a long-lived, non-equilibrium knotless conformation. The average distance between two loci scale as $g^{1/3}$ for a fractal globule, inconsistent with the leveling-off observed on the scale of whole chromosomes.

Chapter 4

The conformational properties of compact polymers

Chromatin and basic polymer models

References

The results presented in this chapter are published in:

- M. Bohn, D. W. Heermann, **2009**. Conformational properties of compact polymers. *The Journal of Chemical Physics* 130 (17), 174901+.

Parts of the analysis are taken from

- J. Mateos-Langerak, M. Bohn *et al.*, **2009**. Spatially confined folding of chromatin in the interphase nucleus. *Proceedings of the National Academy of Sciences of the United States of America* 106 (10), 3812-3817.

Chapter Summary

The three-dimensional organization of chromatin is highly complex and yet its detailed mechanisms are poorly understood. Polymer models like the random walk, self-avoiding walk and globular state have been proposed, although on the scale above 10 Mb folding into a confined space is observed. In this study a careful analysis of data from fluorescence in situ hybridization (FISH) is conducted to investigate whether a random walk or self-avoiding walk polymer model provide an adequate description of chromatin folding on the short scale. Lacking detailed knowledge on the conformational properties of globular state polymers, we perform Monte Carlo simulations to sample compact conformations on a cubic $L \times L \times L$ lattice with different occupancy fractions by modifying a recently proposed algorithm. The system sizes studied extend up to $N = 256\,000$ monomers, going well beyond the limits of older publications on compact polymers. We analyze several conformational properties of these polymers, including segment correlations and screening of excluded volume. Most importantly we propose a scaling law for the end-to-end distance distribution and analyze the moments of this distribution. It shows universality with respect to different occupancy fractions, i.e. system densities. We further analyze the distance distribution between intrachain segments, which turns out to be of great importance for biological experiments. We apply these new findings to the problem of chromatin folding inside interphase nuclei and show that – although chromatin is in a compacted state – the classical theory of compact polymers does not explain recent experimental results.

4.1 Introduction

It is futile to develop a polymer model which explains cellular processes in every detail. Rather polymer models aim at explaining the available experimental data by introducing only a minimal set of parameters with which the data can be described in a satisfactory way. It is exactly the idea behind coarse-graining that for large-scale properties of polymer chains, the molecular and atomistic details can be neglected. Accepting this aim, we can ask what is the simplest polymer model one can think of? Surely, it is the random walk (RW), introduced in section 3.1.2. Its only parameters are the chain length N and the linker length b, no other interactions are considered. The self-avoiding walk (SAW, sec. 3.1.3) is the natural extension of this model, including excluded volume interactions, but nothing else. For both models, we found in chapter 3 that the mean squared end-to-end distance scales like

$$\left\langle R_e^2 \right\rangle \sim N^{2\nu}\,, \tag{4.1}$$

where $\nu = 0.5$ for a random walk and $\nu \approx 0.588$ for a self-avoiding walk.

Referring to Fig. 2.3 we immediately find that a random walk or a self-avoiding walk cannot be a satisfying description of chromatin folding on the scale above 5-10 Mb. The data clearly shows a scaling of $\langle R^2 \rangle (N) \sim \mathcal{O}(1)$, inconsistent with a RW or a SAW polymer model. However we can ask, whether chromatin organization on the short scale is dominated by such a model. Although the RW model seems unlikely in the first place, as chromatin segments have a certain volume and therefore a repulsive interaction, an effective random walk behavior might stem from the effect of the enzyme topoisomerase-II, which is able to let a strand of DNA pass through another one, effectively removing

topological constraints. Indeed, it has been proposed based on older experiments on 2D FISH, that chromatin organization shows a random walk behavior [92] on the scale below 2 Mb.

It was already mentioned that a polymer chain with excluded volume interactions undergoes a transition to a compact polymer, also called globular state, when solvent-monomer interactions are less favorable than monomer-monomer interactions (↷ 3.1.4). A globular polymer is characterized by its dimensions, e.g. the radius of gyration, scaling with $N^{1/3}$, resulting in a nearly uniform monomer concentration inside a globule of a certain radius scaling with $N^{1/3}$. In fact, 2D FISH measurements on chromatin revealed in the regime of genomic separations between 10 Mb and 20 Mb a scaling of the root mean square displacement of about $g^{0.32}$ [29], g being the genomic separation between the FISH markers. However, these measurements were intrachain distances and not end-to-end distances, and we will show below that this difference makes a drastic effect on the scaling.

Polymers in a globular or compact state are quite abundant in nature. Many polymers in living organisms tend to organize in a compact way, the prime example are proteins [99]. A lot of studies have been devoted to the problem of protein folding starting from the sampling of random compact conformations [100, 101].

While for the random walk and self-avoiding walk there exist a lot of studies on their conformational and statistical properties [79, 102, 103], which can be compared to experimental data, this is not the case for globular polymers. However, in the last few years several Monte Carlo algorithms have been proposed to study compact polymers based on the idea of Hamiltonian paths [104, 105, 106]. A Hamiltonian path on some graph G with set of vertices $\mathcal{V}$ and edges $\mathcal{E}$ is defined as a path which visits each vertex $V \in \mathcal{V}$ exactly once. Obviously, Hamiltonian paths studied on a cubic lattice are prime examples of maximally compact polymers where the number of nearest neighbor contacts is maximized. However, the exact enumeration of all possible conformations is not feasible as the computer resources needed grow exponentially with growing system size [107]. Thus such studies are limited to rather small system sizes which probably do not reflect the properties of compact polymers in the limit of large chain lengths N. Therefore, several Monte Carlo techniques have been developed to sample a representative ensemble of Hamiltonian paths on a cubic lattice. One important aspect of such an algorithm is that sampling of conformations is done in an unbiased way. Two algorithms have been shown to fail this test [104, 106]. Recently, Mansfield proposed an algorithm which is shown to produce unbiased samples to a high degree of certainty [105].

The scope of this study is three-fold. First of all, we compare chromatin folding to a random walk and self-avoiding walk polymer model by looking at the scaling of the mean square distance between the FISH markers in relation to genomic separation g, at the distributions of distances at a fixed separation g and its moments. Then, we extend the algorithm proposed by Mansfield [105] in order to study not only maximally compact conformations where all lattice sites on a cubic lattice are occupied, but also conformations in less dense systems with density $\rho \neq 1$. By using a highly parallel system we generate chains of lengths much larger than the ones studied in previous publications. The largest system size for $\rho = 1$ is $L = 55$, the largest chain length studied is $N = 256\,000$ for a density of $\rho = 0.5$. We analyze several statistical and conformational properties of these compact conformations being of general interest for polymer theory. Special interest is on the distance distribution between the end points of the chain as well as the distance distribution between smaller segments of the chain. Stunningly this quantity has not

been analyzed in previous publications, although it is of severe importance for biological applications as the distributions can be compared directly to experimental data [12]. Furthermore, we provide a comparison of compact polymers to a polymer melt of equal density, which has been suggested to behave similarly [79, 106]. Finally, we apply the results of our simulational study of compact polymers to recent experimental data concerning the folding of chromatin inside the interphase nucleus. We show that chromatin on a scale above 150 kb does not organize simply in a compact state as the behavior of the mean square displacement with genomic distance might suggest [12, 31], but shows important hallmarks of a more disordered system.

4.2 Is chromatin organized as a random walk or self-avoiding walk?

4.2.1 Scaling of the mean square displacement

The mean square distance between two segments of a polymer chain is an important measure as it allows direct comparison to experimental data from FISH measurements, where the mean square displacement between two markers is obtained in relation to genomic separation g. As already stated in section 3.1, the mean square distance between the end points of a chain of length N scales as $N^{2\nu}$ with $\nu = 0.5$ (RW) or $\nu = 0.588$ (SAW) independent of the models parameters.

We compare the FISH data to the scaling law by dividing out its leading order term $g^{2\nu}$ (g being the genomic separation between the FISH markers). By using this kind of plot, deviations from the scaling become more visible than with the unscaled plot, where differences between the exponents $\nu = 0.5$ and $\nu = 0.588$ are hardly distinguishable. If one of the models tested would be correct, the data should be on a straight line. The results can be seen in Fig. 4.2 for both the RW and SAW model. Red and green data points are FISH measurements of Roel van Driel's lab in Amsterdam [12], blue data are 2D FISH data of Yokota *et al.* [11]. Obviously, for all data sets the ratio $\langle R^2 \rangle / g^{2\nu}$ displays a slope unequal zero both for the random walk as well as the self-avoiding walk. Thus, neither the random walk nor the self-avoiding walk correctly predicts chromatin organization on the short scale.

In fact, the results might be caused by effects of looking at intrachain distances instead of end-to-end distances. The scaling law eq. (4.1) makes predictions only on end-to-end distances. It can be easily seen that for a random walk, where segments do not feel the existence of each other, the scaling law is the valid for intra-chain distances $n < N$ as well [cf. eq (3.6)]. For a self-avoiding walk, the situation is different: The chain ends have more entropic degrees of freedom and therefore behave differently. To study finite-size and intra-chain effects, we have simulated self-avoiding walks using the pivot algorithm introduced by Kennedy [87] on a 3D cubic lattice. In Fig. 4.1B the scaled mean square intra-chain distance $\langle R_n^2 \rangle / n^{2\nu}$ between two segments is shown versus contour length n. Obviously there are deviations from the scaling law when looking at intrachain distances. On top, the scaling law (4.1) is only asymptotically valid for large N. As can be seen in Fig. 4.1A the asymptotic limit, in which the scaling law becomes valid, starts at $N \approx 500$. In order to exclude the possibility that these approximations invalidate our results, we introduce other quantities, which are also important for the globular state analysis below.

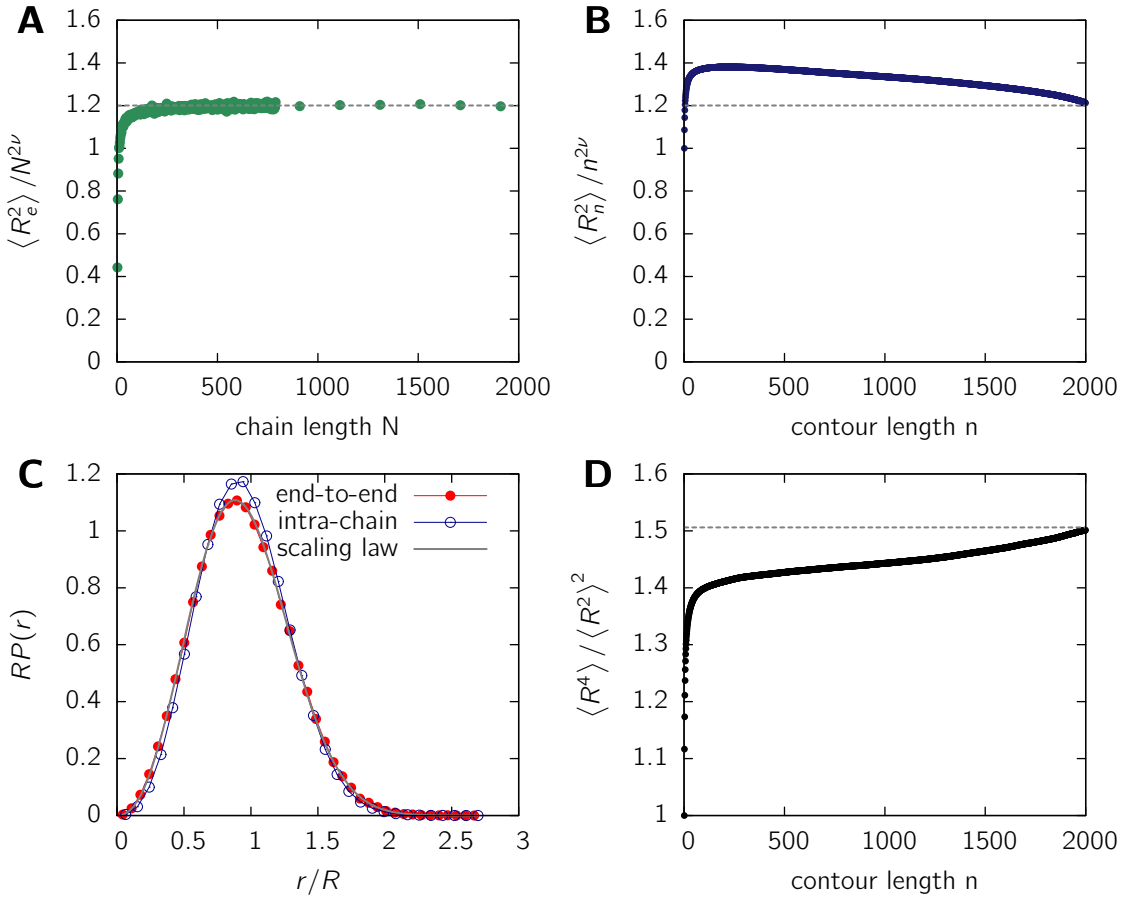


Figure 4.1: Scaling and finite size behavior of self-avoiding walks. Simulations of SAWs have been performed using the fast pivot algorithm [87] on a 3D cubic lattice. **A.** The mean square displacement between the end-points of a chain in dependence of the chain length N. Data is scaled with $N^{2\nu}, \nu = 0.588$ to highlight deviations from the scaling law eq. (4.1) for finite chain length. Starting from about $N = 500$, a the scaling law is approximated well. **B.** Scaling of intrachain distances $\langle R_n^2 \rangle$ with contour length n. Shown is the mean square distance between two monomers separated by n for a chain of length $N = 2000$. The grey line represents the asymptotic behavior for the end-to-end distances. Intra-chain distances display deviations from scaling. **C.** End-to-end distance and intrachain distance distribution $P(r)$ compared to the asymptotic scaling law eq. (4.2). Data is scaled with the root mean squared value of the corresponding distances. For this plot, chains of length $N = 2000$ were used. The distribution of intra-chain distances is measured between two beads separated by $n = 1000$ monomers. Again we find small deviations for intrachain distances from the scaling law. **D.** The moment ratio $c_4 = \langle R^4 \rangle / \langle R^2 \rangle$ for intrachain distances of a chain of length $N = 2000$. The grey line represents the asymptotic value of the end-to-end distance distribution for self-avoiding walks.

4.2.2 Distance distributions

For each genomic separation g, the data from Roel's lab is given as a set of distances R measured in about $50 - 100$ cell nuclei. Instead of looking at the second moment of the distribution $\langle R^2 \rangle$, we can also look at the distance distribution $P(R)$. This distribution then contains all the information available.

The end-to-end distributions for the RW and SAW model are known to follow the

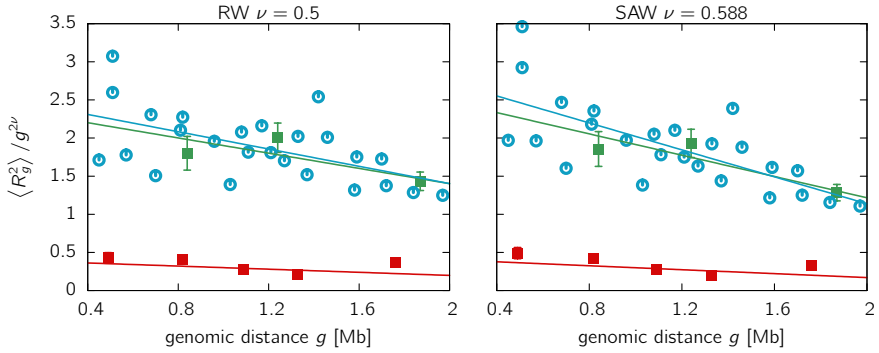


Figure 4.2: Test of the RW and SAW polymer model on experimental data. Shown is the scaled mean square distance $\langle R_g^2 \rangle / N^{2\nu}$ between two FISH markers vs. genomic separation g. Red data points are short distance measurements on chromosome 1 in a anti-ridge region, green data points in a ridge region on the same chromosome [12]. Blue data points are taken from Ref. [27]

scaling law [79]

$$P(r) \sim \frac{1}{R_e} \left(\frac{r}{R_e} \right)^{\mu} \exp \left[-B \left(\frac{r}{R_e} \right)^{\delta} \right] \tag{4.2}$$

where $R_e = \sqrt{\langle R_e^2 \rangle}$ is the root mean squared end-to-end distance.

In case of a random walk, this is just the Gauss distribution averaged over the angles, resulting in $\delta = 1/(1-\nu) = 2$ and $\mu = 2$; for a self-avoiding walk, the exponents are different and can be estimated by scaling arguments [108] to $\delta = 1/(1-\nu) \approx 1/(1-0.588) \approx 2.427$ and $\mu \approx 2.28$ [79].

Again, we first make sure that the distribution of intrachain distances $P(r)$ does not differ strongly from the distance distribution of the end-to-end distances in the SAW model. Fig. 4.1C shows both the end-to-end distance distribution as well as intrachain distance distribution for $n = N/2$ for a chain length of $N = 2000$. The self-avoiding walks in this figure have been simulated using the pivot algorithm introduced by Kennedy [87] on a 3D cubic lattice. The distribution function $P(r)$ is scaled according to eq. (4.2) with the root mean squared end-to-end distance and the root mean squared intrachain distance $\sqrt{\langle R_n^2 \rangle}$ respectively. Although there are deviations for intrachain distances, the curves are still in reasonable agreement.

In Fig. 4.3 the distribution of the data for a genomic separation of $g = 1.87$ Mb is compared to the random walk and the self-avoiding walk distance distribution. Obviously, it is not possible to make statistically significant conclusions about which model fits better. The number of available data points of $50-100$ values are far to small to obtain reasonable histograms. Thus, distribution functions are not a good measure for testing models.

4.2.3 Moment ratios of the distribution

In search for a more precise comparison of experimental data to a polymer model, we had the idea of looking at ratios of higher-order moments of the distribution, especially

$$\text{(a)} \quad c_2 = \frac{\langle R^2 \rangle}{\langle R \rangle^2}, \qquad \text{(b)} \quad c_3 = \frac{\langle R^3 \rangle}{\langle R \rangle^3}, \qquad \text{(c)} \quad c_4 = \frac{\langle R^4 \rangle}{\langle R^2 \rangle^2}. \tag{4.3}$$

Generally speaking, given a probability distribution function $f(x)$, the information about the distribution is stored in all its moments $m_n = \langle f^n(x) \rangle$, $(n = 1, \ldots, \infty)$. While

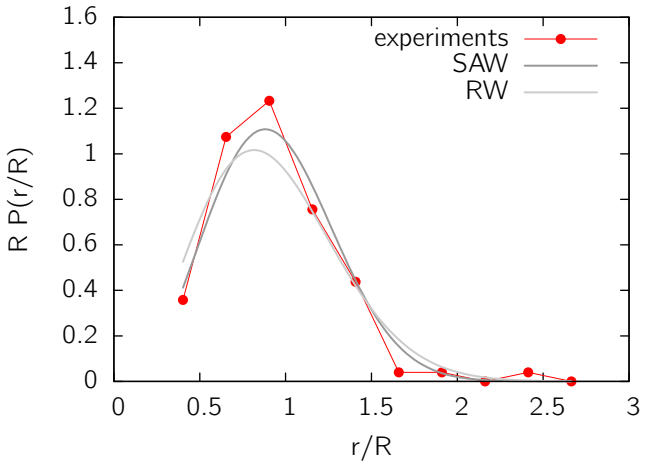


Figure 4.3: Theoretical distance distributions of a RW and SAW according to eq. (4.2) compared to experimental data. Shown are measurements from chromosome 1 at a genomic separation of $g = 1.87$ Mb in a ridge region [12]. The data are scaled with the mean squared distance R such that the curves become independent of the length scales involved.

experimental data is not precise enough to compare the distribution function, i.e. all higher-order moments, to the model predictions (as the standard error of the experimental distribution's moments gets larger the larger the moment), the first few higher-order moments might still yield important information.

The ratios of higher-order moments eq. (4.3) are important from various aspects. First, they are dimensionless, i.e. allow for a quantitative comparison between experiment and model independent of adjustable length scales. Second, these moment ratios carry information about the fluctuations, thus the cell-to-cell variation, in the data.

To calculate the moment ratios of eq. (4.3) for the random walk and self-avoiding walk, we start from the distribution function $P(r)$ of eq. (4.2). We can determine the parameters A and B by the following normalization conditions

$$\int_0^\infty P(r)dr = 1 \qquad \int_0^\infty r^2 P(r)dr = R_e^2 \tag{4.4}$$

and obtain

$$B = \left[\frac{\Gamma\left(\frac{\mu+1}{\delta}\right)}{\Gamma\left(\frac{\mu+3}{\delta}\right)}\right]^{-\delta/2} \qquad A = \delta\,\frac{B^{\frac{1+\mu}{\delta}}}{\Gamma\left(\frac{1+\mu}{\delta}\right)} \tag{4.5}$$

Here $\Gamma(\cdot)$ denotes the Gamma function, which interpolates the factorial function. Substituting the results for A and B in eq. (4.2) we can carry out the moment integrations and obtain

$$\begin{aligned} c_2 &= \frac{\Gamma\left(\frac{3+\mu}{\delta}\right)\Gamma\left(\frac{1+\mu}{\delta}\right)}{\Gamma\left(\frac{2+\mu}{\delta}\right)^2} \\ c_3 &= \frac{\Gamma\left(\frac{\mu+4}{\delta}\right)\Gamma\left(\frac{\mu+1}{\delta}\right)^2}{\Gamma\left(\frac{2+\mu}{\delta}\right)^3} \\ c_4 &= \frac{\Gamma\left(\frac{\mu+1}{\delta}\right)\Gamma\left(\frac{\mu+5}{\delta}\right)}{\Gamma\left(\frac{3+\mu}{\delta}\right)^2} \end{aligned} \tag{4.6}$$

For a random walk ($\mu = 2$, $\delta = 2$), we find

$$c_2 = 1.178, \qquad c_3 = 1.571, \qquad c_4 = 5/3 \tag{4.7}$$

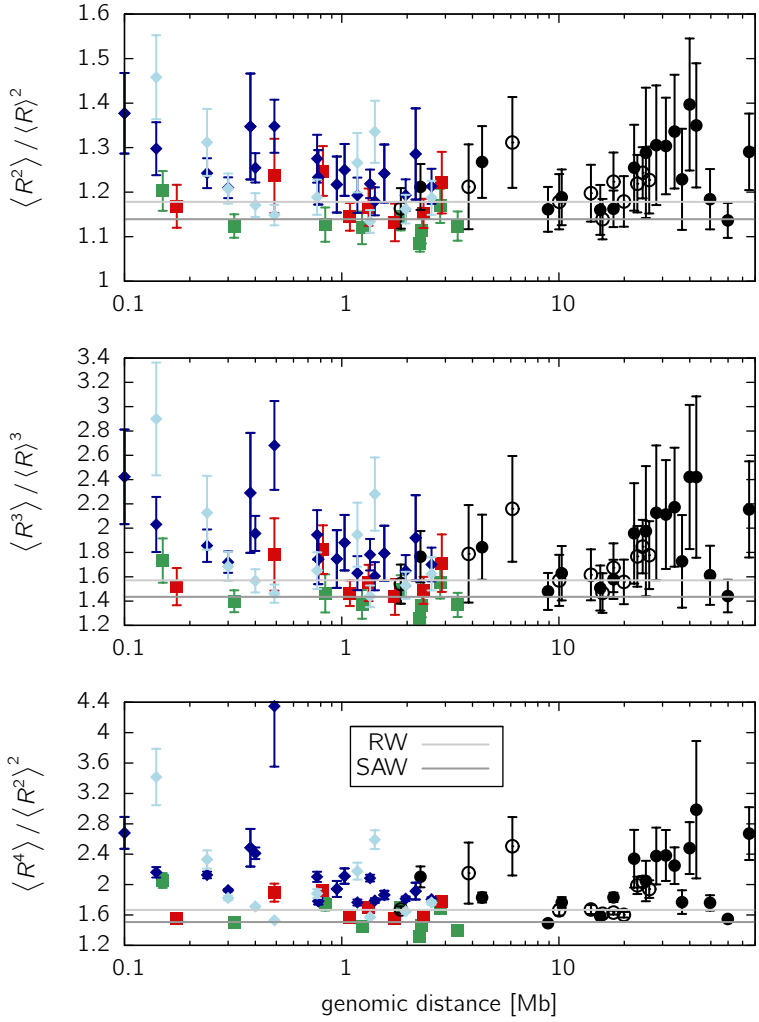


Figure 4.4: Higher-order moment ratios c_2, c_3 and c_4 (cf. equation (4.3)) of the self-avoiding walk and random walk polymer model compared to experimental data. The following experimental data is shown: *Human fibroblasts Chr1* [12]: ■ anti-ridge region, ■ ridge region, ○ long distance measurements; *Human fibroblasts Chr 11* [12]: • long distance measurements; *Murine Igh locus* [109], ♦ pre-pro-B cells, ♦ pro-B cells. The data displays strong deviations towards larger fluctuations in comparison to the random walk (RW), self-avoiding walk (SAW) and globular state (GS) polymer model.

and for a self-avoiding walk ($\mu = 2.28$, $\delta = 2.427$)

$$c_2 = 1.139, \qquad c_3 = 1.435, \qquad c_4 = 1.506\,. \tag{4.8}$$

We test how the ratios change when looking at intrachain distances for the SAW model. Fig. 4.1D shows that the moment ratio c_4 for contour length $n < N$ is always below the value given in eq. (4.8) for the end points of the chain. Thus, fluctuations are larger for the end-points, a result which is expected due to their increased entropic degrees of freedom.

Fig. 4.4 displays a comparison of the ratios for experimental data from human chromosome 1 and 11 [12] and the murine Igh locus [109] with the ratios for the random and self-avoiding walk. Interestingly, the experimental moment ratios are significantly above the values obtained from both polymer models. Especially the higher-order moments c_4

deviate strongly. As the moments of the ratios for intrachain distances n are even below the end-to-end value for a self-avoiding walk, we now have an unambiguous proof that the results from Fig. 4.2 are not due to effects of evaluating distances far away from the chain ends. Clearly neither the random walk nor the self-avoiding walk model fits the experimental data.

4.3 Monte Carlo simulations of compact polymers

4.3.1 The Algorithm

To create conformations of compact polymers on a cubic lattice we use a modified version of the algorithm Mansfield [105] proposed for sampling Hamiltonian paths. Recall that a Hamiltonian Path is defined as a path on a graph G (here a simple cubic lattice), which visits each vertex exactly once. In terms of a polymer on a cubic lattice, this means that each lattice site is occupied. The algorithm of Mansfield is a Metropolis Monte Carlo technique and is one of the few known algorithms except exact enumeration methods which is unbiased, i.e. every allowed conformation is sampled with equal probability. While other proposed algorithms for sampling Hamiltonian paths have been shown to produce biased results [104, 106], Mansfield proved in his paper [105] ergodicity for small lattices by exact enumeration and devised a method providing strong evidence that the algorithm is ergodic for even larger lattices.

Here, we are not only interested in Hamiltonian paths, which only make up a subset of compact polymers but also in compact conformations within the finite temperature regime where $\rho \neq 1$. Therefore, we have to allow vacancies on the simulation lattice. This is done by modifying the algorithm such that it also handles lattices where not all vertices are occupied by introducing a reptation step for the chain ends. We do not consider any interactions between the solvent (i.e. the vacancies on the lattice) and the monomers, therefore having an athermal chain.

Consider a cubic lattice of dimension $L \times L \times L$, each lattice side $\mathbf{r} = (x, y, z)$ is either occupied or unoccupied. The connectivity information of the chain itself is stored in a linked list $\mathcal{L}$ of lattice sites. The algorithm then works as follows:

1. Randomly select one of the two ends of the chain, the coordinates denoted by $\mathbf{r_E}$.
2. Randomly select one of the six neighboring sites of $\mathbf{r_E}$ on the cubic lattice, denoted by $\mathbf{r_N}$.
3. Test if $\mathbf{r_N}$ lies outside the lattice. If so, proceed with (5), otherwise proceed with (4).
4. Test if the lattice site $\mathbf{r_N}$ is occupied. If the lattice site is occupied and $\mathbf{r_E}$ lies at the head of the list $\mathcal{L}$, we reverse the part of the list lying above $\mathbf{r}_N$. If $\mathbf{r_E}$ lies at the tail of the list $\mathcal{L}$, we reverse the part of the list lying below $\mathbf{r_N}$. If the lattice site is unoccupied we do a reptation move, i.e. we append the position $\mathbf{r_N}$ to the head (if $\mathbf{r_E}$ is currently head) or to the tail (if $\mathbf{r_E}$ is currently tail) of the list and remove the other end of the list.
5. Take the new conformation (if or if not it has changed) as the current configuration.

The two types of Monte Carlo moves (list reversal and reptation) are visualized in Fig. 4.5. The algorithm equals the one proposed by Mansfield for compact polymers with $\rho = 1$, as

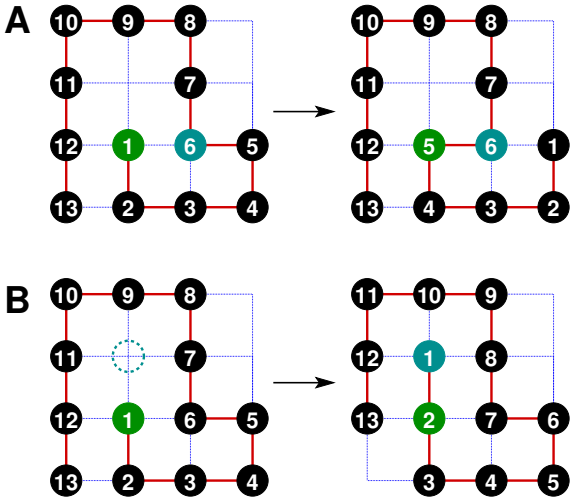


Figure 4.5: The two different Monte Carlo moves for the sampling of compact polymers. Shown is a simplified example on a 2-D lattice with a chain of length $N = 13$. Each monomer carries a number, indicating its position in the linked list $\mathcal{L}$. **A.** Half-list reversal move. The end-monomer 1 is selected and the neighbor with list index 6 is randomly chosen. The list reversal moves the end-point of the chain to a different position. **B.** Reptation move. If the neighboring lattice site is empty, a new monomer is added at the top of the list $\mathcal{L}$ and the last list element is deleted.

in this case reptation steps are not possible. The only change applied to the algorithm is that we allow for a reptation step in (4) whenever the lattice site $\mathbf{r}_N$ is unoccupied.

4.3.2 Ergodicity and unbiased sampling

The algorithm is a Monte Carlo technique, which means that we have to make sure that it produces unbiased samples. *Unbiased* means that

- the algorithm samples the complete configuration space
- each allowed configuration is sampled with equal probability

Unbiased sampling is ensured if the algorithm satisfies *detailed balance* and *ergodicity* [110].

A Markov process is said to satisfy *detailed balance* if the transition rates P_{ij} between each pair of states i and j obey the equation [111]

$$P_{ij}\pi_i = P_{ji}\pi_j$$

P_{ij} denotes the transition probability to get from state i to state j, π_i and π_j denote the equilibrium probabilities of being in state i and j, respectively. In the case of athermal chains, each possible conformation satisfying the excluded volume constraint has equal statistical weight $\pi_i \equiv \pi$, thus the detailed balance criterion reduces to

$$P_{ij} = P_{ji}$$

Obviously, this criterion is satisfied by the algorithm.

The proof of *ergodicity* is much more complicated. Mansfield proved ergodicity for $\rho = 1$ and small lattices, and he devised a method providing strong evidence that the

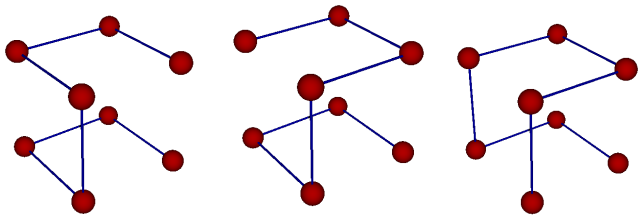


Figure 4.6: For the $2 \times 2 \times 2$ lattice only the three shown Hamiltonian Paths exist after summarizing them with respect to the symmetry operations on the cubic lattice

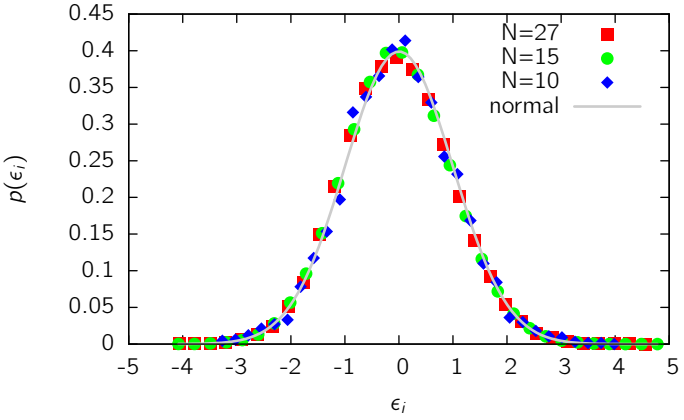


Figure 4.7: Ergodicity test for compact polymers of different densities on a $3 \times 3 \times 3$ lattice. Shown is the probability distribution of ϵ_i, given in eq. (4.9). The distribution $p(\epsilon_i)$ is in good agreement with the standard normal distribution (grey line), being evidence that the algorithm is ergodic.

algorithm is ergodic for any lattice size. However, it is not clear a priori that ergodicity is satisfied even for $\rho < 1$. The reptation algorithm on its own is non-ergodic [111] as a conformation can get trapped in a state where no more reptation moves are possible. However, this problem is resolved here: Whenever there is no free adjacent lattice site to a chain end, the algorithm performs a half-list reversal such that the chain's end moves to another lattice site.

One method to test ergodicity is to exactly enumerate all possible walks on a given lattice with size $L_x \times L_y \times L_z$ and density ρ and to compare the exact probability for one conformation to the probability that the Monte Carlo algorithm samples this specific conformation. As the number of possible conformations increases very rapidly with system size (for a 3^3-lattice we have about 5 million conformations) it is necessary to handle conformations related by symmetry as one and the same. Although this procedure may ignore biases within one symmetry class, this procedure is justified as we only look at properties that do not change by symmetry operations.

In addition to the symmetry operations we also want to drop the distinction between the start and the end of the chain. This is necessary as the algorithm of Mansfield [105] is only ergodic with this condition. Again, this has no influence on the measures we analyze later.

Exact enumeration is done using a recursive back-tracking algorithm: We start from a certain lattice site $\mathbf{r}_s$ and let the chain grow step by step. At some point, we have either reached the target chain length N, then we save this as a valid conformation; or we are stuck at a chain length $n < N$ and there is no possibility for further growth. Then we go one step back to a chain of length $n - 1$ and test the other possible directions. We repeat this procedure until all conformations are enumerated for each start lattice site $\mathbf{r}_s$. Of course, exact enumeration is restricted to very small lattice sizes, on a $2 \times 2 \times 2$ lattice with $\rho = 1$ there are 144 possible conformations, on a $3 \times 3 \times 3$ lattice there are already 4 960 608. After exact enumeration of the conformations, we collect them in certain symmetry classes A_i. One symmetry class summarizes all conformations related by one of the 47 symmetry operations on a cubic lattice, translational shift or path reversal.

Let p_i be the probability for a conformation to fall in equivalence class A_i, determined by exact enumeration. Next, we sample a number K of independent conformations using the Monte Carlo algorithm and determine the relative abundances q_i of a sampled conformation to belong to equivalence class A_i. Under the assumption that the algorithm is

Table 4.1: The probabilities q_i for a Hamiltonian Path generated by the Monte Carlo algorithm on a $2 \times 2 \times 2$ to fall into on of the three symmetry classes (see Fig. 4.6) compared to the resulting probabilities p_i of exact enumeration.

conformation number	1	2	3
p_i (exact enumeration)	1/3	1/3	1/3
q_i (sampling)	0.3333	0.3334	0.33331

indeed unbiased, the probability that k samples fall in equivalence class A_i is given by

$$P_i(k) = \binom{K}{k} p_i^k (1-p_i)^{K-k} .$$

Thus, the random variable

$$\epsilon_i = \frac{q_i - p_i}{\sqrt{\frac{p_i(1-p_i)}{K}}} \tag{4.9}$$

is normal distributed with mean zero and variance equal to unity.

To test our ergodicity calculations, we use the well-known results for $\rho = 1$ on a lattice of dimensions $2 \times 2 \times 2$. Only three symmetry classes A_i exist here, displayed in Fig. 4.6, each of them having the exact probability of $p_i = 1/3$. We samples $K = 10\,000\,000$ conformations using the Monte Carlo algorithm and determined the relative abundances q_i for the conformations falling into symmetry class A_i . The results are shown in table 4.1 and display very good ergodicity.

Fig. 4.7 displays the ergodicity tests on a $3 \times 3 \times 3$ lattice for different monomer densities ρ. There are 51 704 symmetry classes for $N = 27$ and 2 750 classes for $N = 10$ on this lattice. We find that the ϵ_i are in a very good approximation normal distributed for different chain length N. Deviations from the normal distribution most probably are due to the fact that the ϵ_i are not independent. As there is no reason to believe that ergodicity is broken for larger lattice sizes, our modified algorithm most probably satisfies ergodicity.

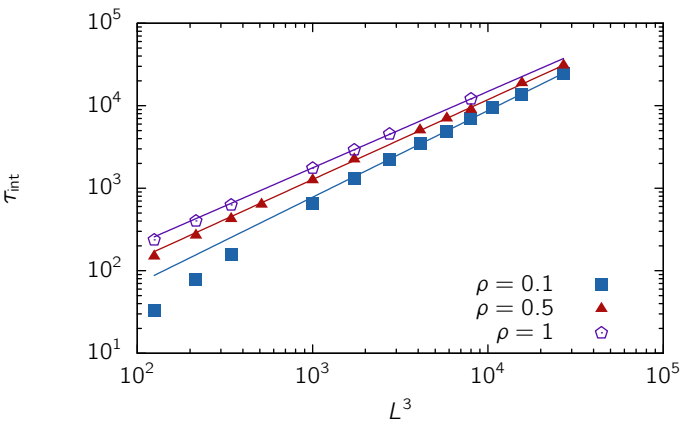


Figure 4.8: Integrated autocorrelation times τ_{int} of compact polymers using the described algorithm. Integrated autocorrelation times show a power-law behavior yielding a linear dependence on N for all densities ρ used in this study.

Table 4.2: Fitting parameters for autocorrelation times. The exponential and integrated autocorrelation times τ_{exp} and τ_{int} were determined for different system sizes L. Then, for each density simulated, the obtained autocorrelation times were fitted to the function $\tau(N) = cN^d$.

ρ	$d(\tau_{exp})$	$c(\tau_{exp})$	$d(\tau_{int})$	$c(\tau_{int})$
0.1	1.07(1)	5.07(36)	1.05(1)	6.2(5)
0.5	0.995(3)	2.50(6)	0.97(2)	3.1(4)
1	0.954(3)	2.32(5)	0.926(4)	2.94(10)

4.3.3 Autocorrelation times

As the algorithm only changes one bond per Monte Carlo step, subsequent conformations are highly correlated. A lot of subsequent Monte Carlo steps have to be performed until conformations get uncorrelated. Following Mansfield [105], we calculate the autocorrelation function $C_{N_x}(t)$ of the observable N_x, which is defined as the number of bonds oriented along the x-direction. We then determine the exponential decay time of the correlations by a fit to the function $f(x) = \exp(-t/\tau_{exp})$ and obtain the exponential autocorrelation time $\tau_{exp}(\rho, N)$. This is the time scale defining how long we have to wait initially before sampling any conformations. On the other hand, the integrated autocorrelation time τ_{int} tells us how many Monte Carlo steps have to be carried out until we have a subsequent independent conformation. From Fig 4.8 one can see that $\tau_{int}(\rho, N)$ as a function of N has a power-law behavior. We fit the correlation times to the function

$$\tau(\rho, N) = c(\rho)N^{d(\rho)}$$

and obtain the results shown in table 4.2. Two conformations can be considered uncorrelated after $2\tau_{int}$ Monte Carlo steps [111]. Based on these results we write out a conformation after $14N$ steps for $\rho = 0.1$ and $6N$ steps for $\rho > 0.1$.

4.4 Conformational properties of compact polymers

In this section we present results on the conformational properties of compact polymers using the algorithm described above. We performed extensive simulations for three different densities $\rho = 0.1, 0.5$ and 1.0. For each density a broad range of system sizes has been studied. For $\rho = 0.1$ we did simulations up to $L = 130$, for $\rho = 0.5$ the largest system size studied is $L = 80$ and for $\rho = 1.0$ we were limited by computing time to $L = 55$. Thus the largest simulated chains are made up of $N = 166\,375$ monomers for $\rho = 0.1$, $N = 256\,000$ for $\rho = 0.5$ and $N = 219\,700$ for $\rho = 0.1$. For each density and system size we sampled between 20 000 up to one million conformations, depending on the system size.

4.4.1 End-to-end distance statistics

One characteristic length scale of a polymer is given by the mean squared end-to-end distance (↷ 3.1.1), i.e. the distance between the two endpoints of the chain, which is often denoted by $\langle R_e^2 \rangle$. Obviously, $\langle R_e^2 \rangle$ depends on the total length of the chain, which is denoted by the number of monomers N and in our case related to the system size L. The

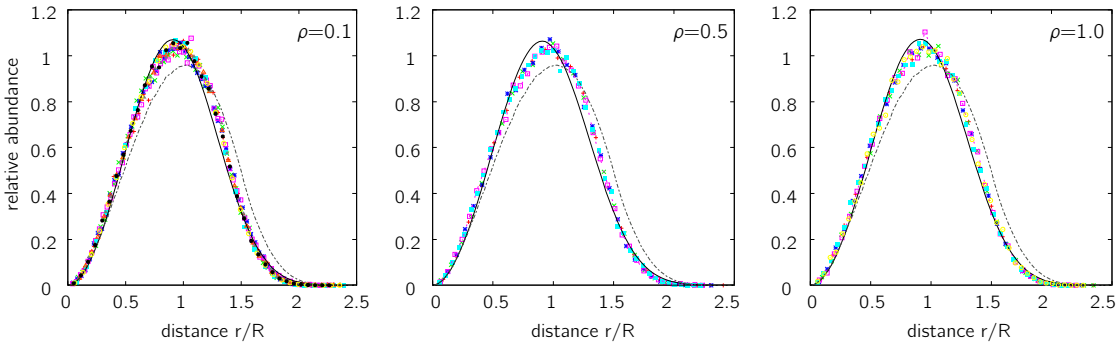


Figure 4.9: Probability distribution of the end-to-end distances for compact polymers. The distributions are scaled with the root mean squared end-to-end distance $R = \sqrt{\langle R^2 \rangle} = aN^{1/3}$ where the parameter a is determined by a fit according to table 4.3. For each density systems of different size fall on top of each other. For $\rho = 0.1$ system sizes vary from $L = 10$ to $L = 130$, for $\rho = 0.5$ we analyzed systems from $L = 10$ to $L = 80$ and for $\rho = 1$ the lattices used range from $L = 10$ to $L = 55$. The black lines represent a two-parameter fit to the empirical distribution function eq. (4.11), the fitting parameters are listed in eq. (4.12). For comparison, the dotted line shows the distribution of two random points on a cubic lattice obeying excluded volume.

relation between the mean squared end-to-end distance and the number of monomers can be written in terms of a scaling law for polymers in good solvents,

$$\left\langle R_e^2 \right\rangle \sim N^{2\nu} \,. \tag{4.10}$$

For random walks, i.e. polymers where excluded volume effects are ignored, it can be shown straightforwardly that the scaling exponent is $\nu = 0.5$ [79]. For self-avoiding walks in good solvents, where excluded volume effects are taken into account, the polymer is more swollen compared to the random walk (↷ 3.1.3). The resulting exponent is not known exactly, but estimated by field theoretical methods to $\nu \approx 0.588$ [89]. A compact polymer on the other hand is characterized by an exponent of $\nu = 1/3$ representing a globular shape with homogeneous density (↷ 3.1.4). The scaling law $\sqrt{\langle R_e^2 \rangle} = bN^{1/3}$ is also valid for the compact conformations studied here. Table 4.3 shows values for the parameter b determined by a fit to the data for different densities ρ.

While a scaling law for the end-to-end distance distribution $P(r)$ for self-avoiding walks has been proposed long ago by Fisher [108], it remains unclear whether there is kind of universal scaling law for the end-to-end distribution of compact polymers as well. Moreover it is not known what is the functional form of this scaling function. As L is the only length scale in our system, which is related to $R = \sqrt{\langle R_e^2 \rangle} = bN^{1/3} = b\rho^{1/3}L$, there is a good chance that the distributions scale with r/R. This leads us to propose a scaling law for the distributions similar to that of a random walk or self-avoiding walk [see also eq. (4.2)],

$$P(r) = \frac{A}{R}\left(\frac{r}{R}\right)^{\mu} \exp\left[-B\left(\frac{r}{R}\right)^{\delta}\right] . \tag{4.11}$$

While for the random walk as well as the self-avoiding walk, the exponents μ and δ are well-known [eqs. (4.7) and (4.8), see for example the book [79]] and we have already concentrated on these two cases in sec. 4.2, we now want to determine these exponents for the compact polymers studied here. The parameters A and B are given by the normalization conditions [eq. (4.4) and eq. (4.5)], thus there are two remaining parameters μ and δ.

ρ	b	a
0.1	1.50(2)	0.696(7)
0.5	0.89(7)	0.712(5)
1	0.71(1)	0.71(1)

Table 4.3: The effective monomer size determined by a fit to the function $\langle R_e^2 \rangle = b^2 N^{2/3} = a^2 L^2$ for each density ρ. Obviously, the effective monomer size decreases with increasing density.

In Fig. 4.9 it is shown that the scaling with R holds very well for different densities ρ. The analytic form of $P(r)$ approximates the data fairly well although deviations from the data are larger than for a self-avoiding walk, as the normalization condition above neglects the fact that end-to-end distances cannot extend beyond $\sqrt{2}L$.

We fit the theoretical distribution function $P(r)$ to the data for three different ρ-values and obtain:

$$\begin{aligned} \rho = 0.1: &\qquad \mu = 1.89(3) \quad \delta = 2.94(4) \\ \rho = 0.5: &\qquad \mu = 1.87(4) \quad \delta = 2.91(6) \\ \rho = 1.0: &\qquad \mu = 1.90(3) \quad \delta = 2.94(5) \end{aligned} \tag{4.12}$$

The errors given here and in the following are the asymptotic standard errors from the least squares fit performed by the program `gnuplot` (version 4.2). Within these fitting errors, the exponents μ and δ are the same for different densities ρ, suggesting that these values show some universal features of compact polymers. On average we obtain from the above data

$$\mu = 1.889(65), \qquad \delta = 2.932(89)\,. \tag{4.13}$$

In a recently published paper Jacobsen [112] determined the scaling exponent μ in the case $\rho = 1.0$ by fitting a power-law function $P(r) = r^{\mu}$ to the end-to-end distance distribution for small r. Although we came across this paper only when ours was already published, he found a similar scaling exponent $\mu = 1.85 \pm 0.02$.

Comparing experimental data to the distribution function $P(r)$ is not always the method of choice, especially when the number of data points is too small for creating reasonable histograms. Here, we analyze dimensionless ratios of moments of the end-to-end distance distribution as given in eq. (4.3), having the advantage that no adjustable parameter is present. The ratios of interest here are

$$c_2 = \frac{\langle R^2 \rangle}{\langle R \rangle^2}, \qquad c_3 = \frac{\langle R^3 \rangle}{\langle R \rangle^3}, \qquad c_4 = \frac{\left\langle R^4 \right\rangle}{\langle R^2 \rangle^2}\,.$$

We have already derived ($\curvearrowright$ 4.2.3) that for a random walk and a self-avoiding walk, the ratios c_i are constants not depending on any model parameters (such as linker length l). Here we show that this is also the case for compact polymers and we determine its values. The ratio plots are shown in Fig. 4.10. A fit yields the values

$$c_2 = 1.1395(5) \qquad c_3 = 1.421(2) \qquad c_4 = 1.458(1)$$

Obviously, there are significant differences from a RW, however differences from a SAW behavior become only visible in the fourth order ratio c_4. We demonstrate in section 4.5 how this information can be used to characterize the behavior of biopolymers.

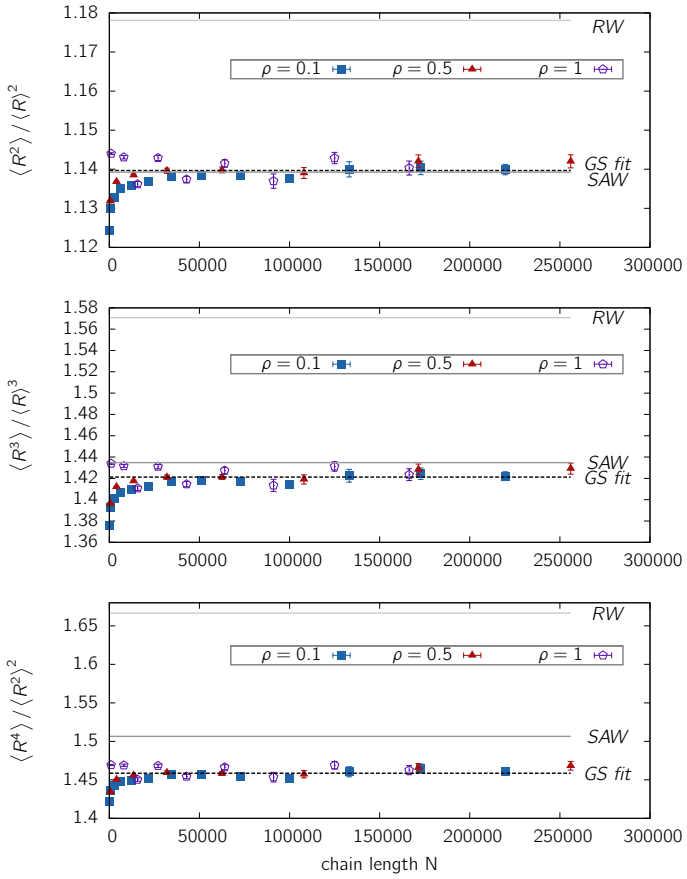


Figure 4.10: Higher-order moment ratios of the end-to-end distance distribution. Shown are the ratios $c_2 = \langle R^2 \rangle / \langle R \rangle^2, c_3 = \langle R^3 \rangle / \langle R \rangle^3$ and $c_4 = \langle R^4 \rangle / \langle R^2 \rangle^2$ in relation to the chain length N. Clearly, simulational results show that the ratios become independent both on system size N as well as system density ρ approaching a constant value. Fluctuations are smaller than for the self-avoiding walk and the random walk model.

Instead of determining the moments by the raw simulational data we can determine the moments by the analytic function $P(r)$ given in eq. (4.11) with the parameters obtained by the fit [eq. (4.13)]. Using the analytical formulas of equation (4.6) we obtain

$$c_2 = 1.140(6) \qquad c_3 = 1.43(2) \qquad c_4 = 1.488(24)$$

The errors are calculated here by evaluating c_i for all values of μ and δ in the range of their fitting errors and determining the maximum deviation from the average value. The values are compatible with the ones calculated directly by a fit to the simulation data within the range of the errors. Deviations become large for larger moments reflecting the approximative character of the scaling function $P(r)$.

4.4.2 Intrachain distance statistics

In the past a lot of effort has been undertaken to study the end-to-end distributions of self-avoiding walks [108]. For random walks, this problem is easily solved analytically. In the last section we studied these distributions for compact polymers which has not been done so far. We found a similar scaling function as for random and self-avoiding walks with exponents which seem to be universal for compact polymers of different densities and chain lengths. However, from the experimental point of view one is not only interested in the distance between end points of a compact polymer, but also in the distance distribution between two arbitrary monomers along the chain which are separated by a certain contour length n. For example, this quantity becomes important in experiments measuring the spatial arrangement of two fluorescently labeled parts of the human genome [12] (Fig. 2.3).

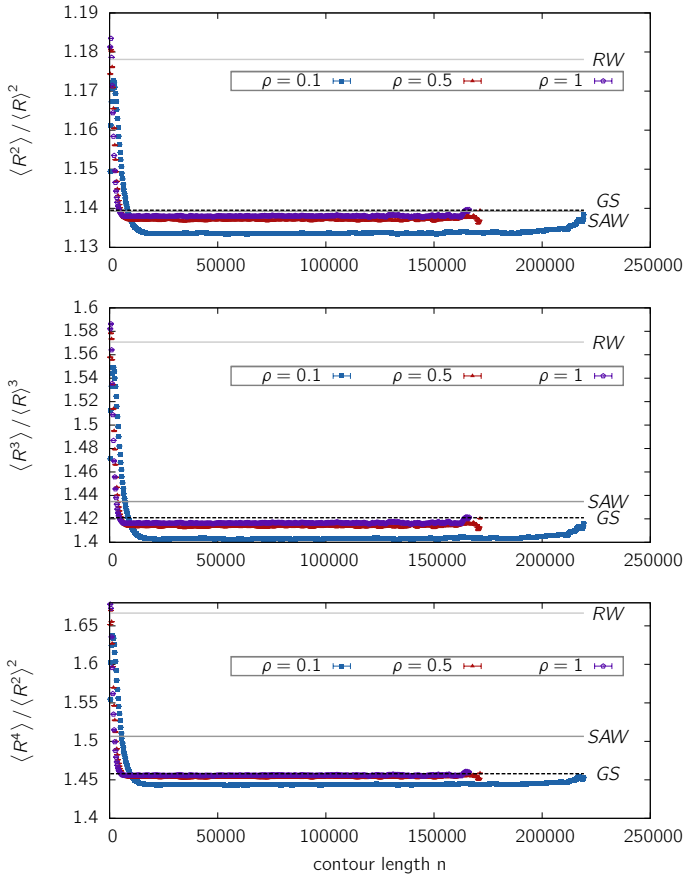


Figure 4.11: Higher-order moment ratios of intrachain distance distribution. Shown are the ratios $c_2 = \langle R^2\rangle / \langle R\rangle^2, c_3 = \langle R^3\rangle / \langle R\rangle^3$ and $c_4 = \langle R^4\rangle / \langle R^2\rangle^2$ in relation to contour length n between the beads. Data is shown for the largest system size simulated for different densities $\rho = 0.1, 0.5$ and 1. For comparison, the end-to-end distance ratios are displayed for the random walk (RW), self-avoiding walk (SAW) and globular state (GS) model. The intrachain ratios c_i show a random walk behavior on the short scale due to the screening effect (↷ 4.4.5). Generally, the fluctuations are smaller than for the end-to-end distances.

Therefore, we have to evaluate how the distribution changes when looking at intrachain segments.

It is most interesting to look at the moment ratios as these are easiest to compare to experimental data, which quite often do not provide enough data points to obtain a complete distribution function. For the largest chains simulated the moment ratios are shown in Fig. 4.11 for various contour length n. The mean values are averages both over different positions along one chain as well as over the set of sampled conformations $\mathcal{C}$. The kth moment is thus evaluated as

$$\left\langle R_n^k \right\rangle = \frac{1}{|\mathcal{C}|} \frac{1}{N-n} \sum_{C\in\mathcal{C}} \sum_{i=1}^{N-n} \parallel \mathbf{r}_{i+n}^C - \mathbf{r}_i^C \parallel^k \ .$$

Here, $\mathbf{r}_i^C$ denotes the position of the ith monomer of the conformation C out of the set of sampled conformations $\mathcal{C}$.

For small contour length n the moment ratios are peaked and reach the value of a random-walk. This is due to the screening effect in compact polymers, which is in detail analyzed in section 4.4.5. However, the ratios pretty fast fall below the ratios for self-avoiding walks and stay mostly constant. There is only a small increase for the ratios where the contour length approaches the chain length N, indicating that the chain ends have more freedom for fluctuations than parts embedded in the middle of the chain. We want to stress here that for sufficiently large contour lengths the moment ratios for intrachain distances are smaller than those for the end-to-end distances.

4.4.3 End-point statistics

It is an open question, raised for example in Ref. [106], whether the positions of the end points of a compact polymer are correlated or not. It was suspected that an entropic cost associated with local rearrangement around the chain's ends might cause some effective attraction or repulsion between them. Let $\mathbf{r}_1 = (x_1, y_1, z_1)$ and $\mathbf{r}_2 = (x_2, y_2, z_2)$ denote the end point vectors with respect to the center of the simulation box. Lua et al. [106] showed then for Hamiltonian paths up to $L = 10$ that the end point correlation coefficient

$$c = \frac{\langle x_1 x_2 \rangle}{\sqrt{\langle x_1^2 \rangle \langle x_2^2 \rangle}} \tag{4.14}$$

is negative for small lattice sizes but pretty fast approaches the correlation between disconnected points only obeying excluded volume and the chess board theorem [106]. The latter theorem states that, if we mark adjacent vertices on the lattice graph with different colors similar to a chess board, then the end points of a chain with even numbers of monomers are sitting on lattice sites with different color while the end points of a chain with an odd number of monomers are positioned on lattice sites of same color. This restriction has to be taken into account when comparing to randomly positioned points, as this is an inherent feature of the lattice model but not of the ensemble of compact polymers in general. However it becomes more and more negligible the larger the lattice size.

Here we study the correlation coefficient c for lattice sizes much larger than in [106]. Note that the coordinates in eq. (4.14) are taken with respect to the center of the simulation cube. Fig. 4.12 shows that there are negative correlations for all densities considered, which approach zero for larger lattice sizes. The only deviations are for $\rho = 1$ (i.e. Hamiltonian paths) and even lattice sizes, which obviously is an effect of the lattice geometry and the chess board theorem and therefore no intrinsic property of compact polymers. These results are in very good agreement with the results by Lua et al. [106], suggesting that there are no end-point correlations in the $N \to \infty$ limit.

We also analyzed the mean square displacement of the endpoints from the center of the cube in order to answer the question whether the polymer tends to arrange such that the endpoints predominantly locate in the center of the cube or at its periphery. Being located in the center of the cube might be disadvantageous due to entropic reasons. Fig. 4.13 shows that for $\rho = 0.5$ and $\rho = 1$ the points are predominantly shifted towards the periphery of the cube, while for $\rho = 0.1$ the points are located more in the interior of the cube.

4.4.4 Correlations of intrachain segments

Consider two arbitrary points on a polymer with coordinates $\mathbf{r}_1 = (x_1, y_1, z_1)$ and $\mathbf{r}_2 = (x_2, y_2, z_2)$. As for the end points of the chain, we can pose the question whether the coordinates of these points are correlated by evaluating the correlation coefficient of eq. (4.14) We assume the coordinates $\mathbf{r}_1$ and $\mathbf{r}_2$ to be given with respect to the center of mass of the polymer. A value of $c = 0$ indicates that there is no correlation between the coordinates $\mathbf{r}_1$ and $\mathbf{r}_2$, i.e. they effectively behave like two randomly chosen points on the cubic lattice. A value $c \neq 0$ indicates an effective attraction or repulsion. Fig. 4.14 shows the correlation coefficient in dependence of the contour length n between the segments. While for short contour length there are high correlations because of the connectivity of the chain, these correlations decay fast and for larger contour length correlations are nearly vanished. This is in stark contrast to the behavior of a self-avoiding walk, where (negative) correlation

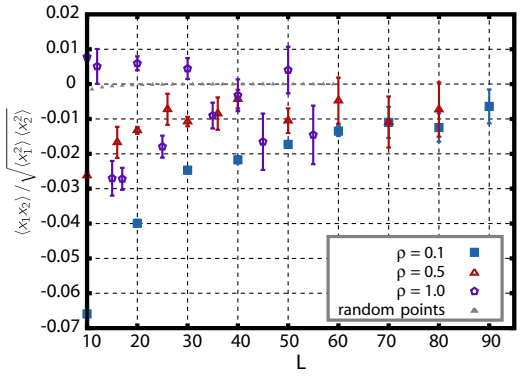


Figure 4.12: End-end correlation coefficients according to eq. (4.14) for compact polymers with different densities ($\rho = 0.1$, $\rho = 0.5$ and $\rho = 1$). Additionally shown are exact results for the correlation coefficient of two randomly positioned points on a cubic lattice only obeying excluded volume and the chess board theorem.

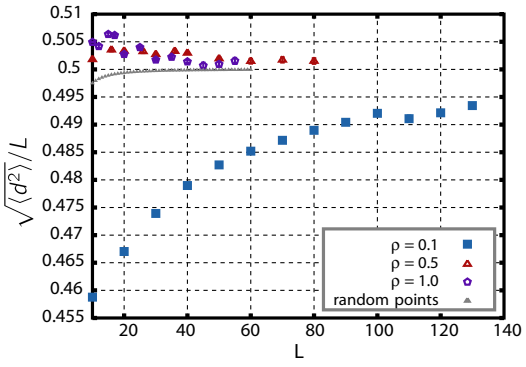


Figure 4.13: Distance of the end points from the center of the cube for compact polymers with different densities. For comparison shown are also the results for random points obeying excluded volume and the chessboard theorem.

effects are dominant even on the length of the whole chain. This result is in perfect agreement with the scaling theory developed by De Gennes [78], which predicts that on the length scale of the compact system, parts of the chain become practically independent. If the scaling theory is correct then the decay length of the correlations n_d, which we define as the length where $c(n_d) \sim 1/e$, should be related to the system size L. We test this prediction by evaluating the ratio $r = \langle R^2(n_d) \rangle / L^2$ where $\langle R^2(n_d) \rangle$ is the mean square displacement between the end points of a segment of length n_d. For the largest system sizes studied we find the values $r \approx 0.31$ for $\rho = 1$, $r \approx 0.31$ for $\rho = 0.5$, $r \approx 0.29$ for $\rho = 0.1$. As r-values are nearly equal for systems of completely different size and density there is strong evidence that the decay of the position correlations is directly related to the system size.

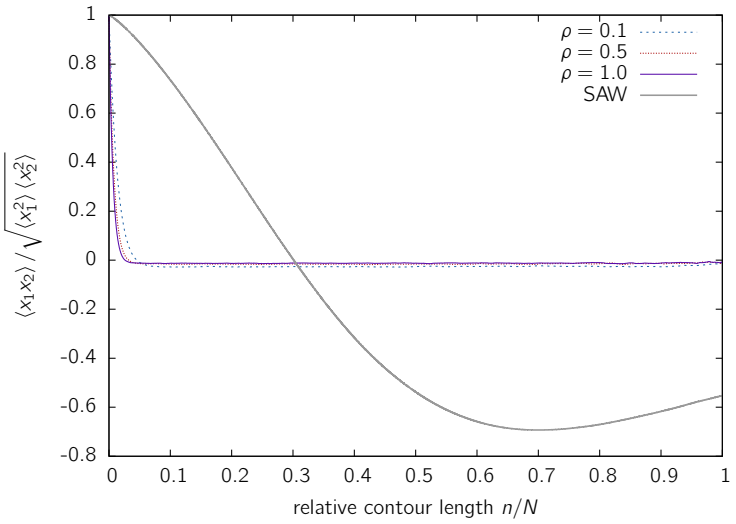


Figure 4.14: Position correlations of monomers. This figure shows the correlation coefficient $c = \langle x_1 x_2 \rangle / \sqrt{\langle x_1^2 \rangle \langle x_2^2 \rangle}$ between two monomers in dependence of their separation along the chain. The data shown are for system sizes $L = 55$ ($\rho = 1$), $L = 70$ ($\rho = 0.5$) and $L = 130$ ($\rho = 0.1$). For comparison, the correlations in a self-avoiding walk chain of length $N = 10\,000$ are also shown.

4.4.5 Screening of excluded volume in compact polymers

Much is known about polymer melts, where a number of polymers with degree of polymerization N is placed in a system with volume fraction Φ. Below a critical concentration (or volume fraction) $\Phi^\star$, the polymers do not feel the existence of the others and basically behave like self-avoiding walks in a good solvent. This critical concentration $\Phi^\star$ is given by the volume fraction where the free coils with extension $\langle R^2 \rangle_0 \sim N^{2\nu}$ begin to overlap [78]. The index 0 indicates here the limit $\Phi \to 0$. The value of the critical concentration scales like

$$\Phi^\star \approx \frac{N}{\langle R^2 \rangle_0} \sim N^{-(3\nu-1)} .$$

At volume fractions above $\Phi^\star$ the polymers begin to feel each other and the system can be described by the correlation length ξ. On scales larger than this correlation length, the chains effectively behave like ideal coils, a theorem most often referred to as Flory theorem. On a scale smaller than ξ excluded volume effects still play a dominant role. By scaling arguments one finds [78] that

$$\xi \sim \Phi^{-\nu/(3\nu-1)} .$$

Here we want to investigate the question whether there is a fundamental difference between a polymer melt with volume fraction Φ and a compact polymer, i.e. a melt with degree of polymerization $N = \Phi L^3$. Consider an arbitrary segment of a compact polymer of length N_m. We impose on N_m the condition that the extent of this segment must be smaller than the system size in order to avoid effects of the confinement to play a role. We now ask whether these segments of length N_m of our compact polymers behave different from a polymer with degree of polymerization N_m in a corresponding melt. The analysis of positional and angular correlation effects which are decaying pretty fast suggests that a part of the chain should not "feel" that it is connected with a part far away.

The theory of polymer solutions predicts that there is a crossover from a self-avoiding walk behavior to a random walk behavior on the short scale [113]. Fig. 4.15 shows that this crossover becomes indeed apparent in the mean square displacement for short contour lengths in compact polymers. While for $\rho = 0.1$ we find a self-avoiding walk type of scaling with exponent $\nu = 0.588$ for contour lengths up to $n \approx 50$, the maximal dense system behaves more like a random walk with $\nu = 0.5$ indicating the screening of excluded volume in this system.

To analyze the screening length we have to look in more detail at the structure function of parts of the chains. The structure function is defined as

$$S(q) = \left\langle \frac{1}{N_m} \left| \sum_{j=0}^{N_m} e^{i\mathbf{q}\cdot\mathbf{r}_j} \right|^2 \right\rangle_q \tag{4.15}$$

The brackets denote a spherical average over all $\mathbf{q}$-vectors of equal magnitude and over all conformations. The sum is over a subchain of length N_m whose position vectors are denoted by $\mathbf{r}_0 \ldots \mathbf{r}_{N_m}$. One expects these subchains to behave like random walks on distances larger than ξ and self-avoiding walks on distances smaller than ξ, i.e.

$$S(q) \sim \begin{cases} q^{-2}, & \langle R^2 \rangle^{1/2} > \frac{2\pi}{q} > \xi \\ q^{-1/\nu}, & \xi > \frac{2\pi}{q} > 1 \end{cases} \tag{4.16}$$

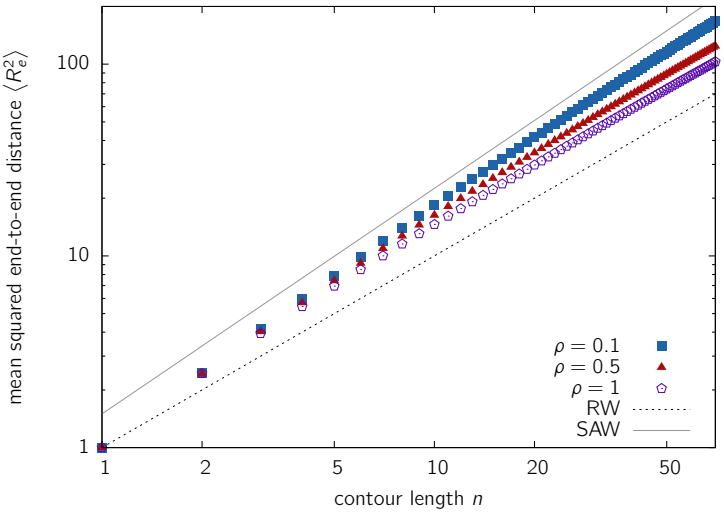


Figure 4.15: Mean squared distances $\langle R_n^2 \rangle$ between monomers separated by a contour length of n for different densities. The compact polymers also show a screening of excluded volume for large densities.

In Fig. 4.16 the structure function is shown for chain segments of length $N_m = 100$. One can immediately see that the system with $\rho = 0.1$ shows a range of q-values, where excluded volume is not screened, extending much beyond the length scale of a single bond. On the other hand for $\rho = 0.5$ excluded volume interactions are screened very fast resulting in ideal chain behavior over a wider range of q-values. We can determine the screening length by performing a linear fit with slope -2 for small q-values (but beyond the scale where asymptotic behavior sets in) and a linear fit with slope $-1/\nu$ for large q-values (but away from the length scale of a single bond). We then extract the value q_ξ where the crossover between the two regimes occurs and obtain the following screening lengths

$$\begin{aligned} \rho = 0.1: \quad & q_\xi \approx 0.6 \quad \rightarrow \quad & \xi = \frac{2\pi}{q_\xi} \approx 10.5 \\ \rho = 0.5: \quad & q_\xi \approx 1.8 \quad \rightarrow \quad & \xi = \frac{2\pi}{q_\xi} \approx 3.5 \\ \rho = 1.0: \quad & & \xi = \frac{2\pi}{q_\xi} \approx 1 \end{aligned}$$

For $\rho = 1$ the system seems to be that dense that excluded volume is shielded on the order of a bond length, therefore the scaling regime where $S(q) \sim q^{-1/\nu}$ does not show off any more. This result does not come as a big surprise as the screening length ξ is related to the average mesh size in the system [78]. This mesh size is – as every lattice site is occupied – approximately equal to unity.

4.4.6 The gyration tensor

The shape of a a polymer is described by its gyration tensor. The gyration tensor is defined as

$$S_{mn} = \frac{1}{N} \sum_{i=1}^{N} r_m^{(i)} r_n^{(i)} \,. \tag{4.17}$$

Here, $\mathbf{r}^{(i)}$ is the coordinate vector of the ith monomer and the subindex denotes its Cartesian components. The eigenvalues $\lambda_1 \leq \lambda_2 \leq \lambda_3$ give the squared lengths of the principal axes of gyration. The ratios of the eigenvalues indicate the deviation from a sphere-like

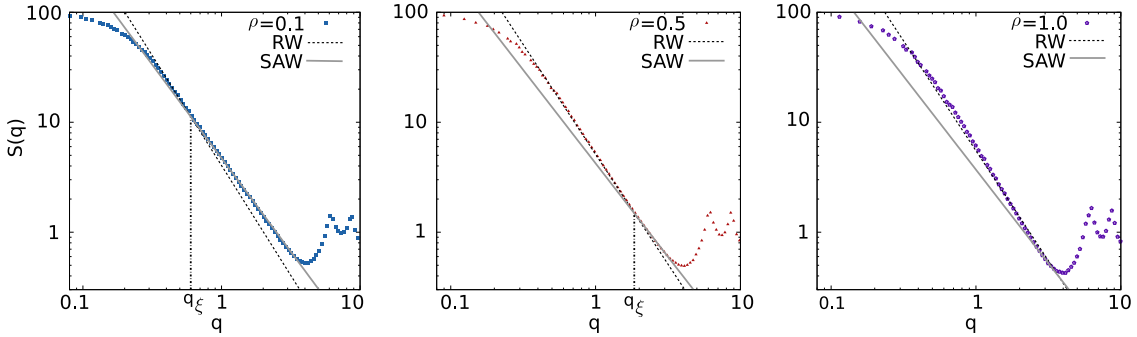


Figure 4.16: Structure function of chain segments of length $N_m = 100$ for compact polymers with different densities. For the calculation the largest simulated system sizes for each density was used. The value of the segment length was chosen such that the radius of gyration of the segments is way below the system size. Shown is also the scaling regimes where $S(q) \sim q^{-2}$ and $S(q) \sim q^{-1/\nu}$. From a fit to the curves one can determine the crossover value q_ξ which determines the screening length.

shape of the polymer. It is well-known, for example, that the gyration tensor for self-avoiding walks and random walks in good solvent has a pronounced asphericity. This asphericity shows up in the asymptotic ratios of the eigenvalues, namely [83, 84]

$$\langle\lambda_3\rangle : \langle\lambda_2\rangle : \langle\lambda_1\rangle \to 12 : 2.7 : 1 \qquad \text{for a RW}$$
$$\langle\lambda_3\rangle : \langle\lambda_2\rangle : \langle\lambda_1\rangle \to 14 : 2.98 : 1 \qquad \text{for a SAW}$$

For Hamiltonian paths it is clear that there can be no asphericity in a symmetric simulation box as every lattice site is occupied. However it is not clear *a priori* that this holds true for less dense systems, where a crossover to the self-avoiding walk behavior might occur. Fig. 4.17 shows the ratio $\langle\lambda_3\rangle : \langle\lambda_1\rangle$. For $\rho = 0.1$ a pronounced deviation from the symmetry shows up for small N, but evidently, this deviation vanishes for large system sizes.

4.5 Is chromatin organized as a compact polymer?

In section 4.2 it was shown using different measures derived from the distance distribution of two FISH markers separated by a certain distance g along the contour of the chain, that

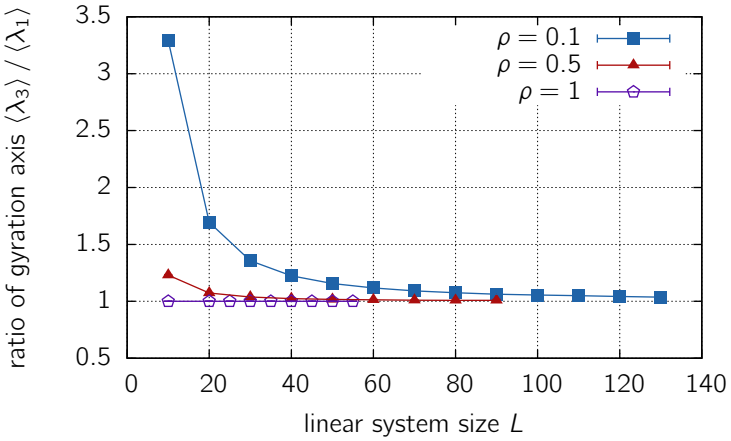


Figure 4.17: The ratio between the largest and the smallest eigenvalue of the gyration tensor. Eigenvalues are the squared lengths of the gyration ellipsoids main axes. While for RW or SAW polymers the tensors show a characteristic asphericity, compact polymers have an almost spherical shape in the asymptotic limit.

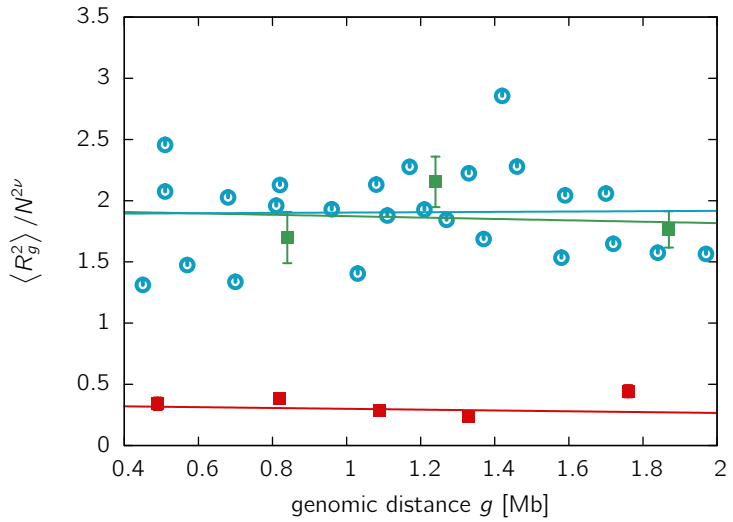


Figure 4.18: Short-distance FISH measurements with the leading order scaling term of the globular state model $N^{2/3}$ divided out (cf. Fig. 4.2). The following experimental data is shown: *Human fibroblasts Chr1 (3D FISH)* [12]: ■ anti-ridge region, ■ ridge region. *Human chromosome 4 (2D FISH)* [11]: ⊙. Note that the scaling with $N^{2/3}$ does not correctly describe intrachain distances for compact polymers, see also Fig. 4.19.

chromatin is not organized as a random walk and self-avoiding walk polymer, neither on the short nor on the long scale. After having studied the properties of compact polymers, we can apply the same measures to see whether chromatin is organized in a globular state.

Following the idea of Fig. 4.2, we first divide out the leading order term $N^{2/3}$ of the scaling law for the squared end-to-end distances of compact polymers. If the data scales with this exponent, the ratio $\langle R^2 \rangle / N^{2/3}$ should be independent of the contour length. We use data up to genomic distances of 2 Mb to keep away from distances at which leveling-off begins. Fig. 4.18 indicates that a scaling with $\nu = 1/3$ is more consistent with experimental data, indicating a considerably more compact state than predicted by the RW or SAW models.

Although in the globular state model, the end-to-end distances scale with an exponent of $\nu = 1/3$, we know from the analysis of the self-avoiding walk (Fig. 4.1) that deviations exist for distances between two intrachain segments in the presence of excluded volume. While these deviations do not exist for the random walk polymer model and are small for the self-avoiding walk model, the situation is completely different for the globular state. Intrachain distances are displayed in Fig. 4.19 for a system with density $\rho = 0.5$ and

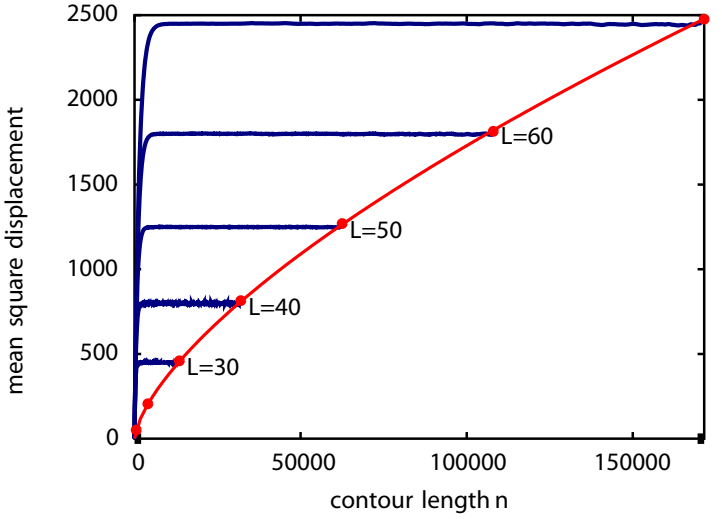


Figure 4.19: The mean square displacement $\langle R_n^2 \rangle$ in relation to contour length n for compact polymers with density $\rho = 0.5$. Data is shown for different chain length. i.e. system sizes. The end-to-end distances are marked by red points, the red line denotes the scaling law for the end-to-end distances $N^{2\nu}$, ν being 1/3. Intrachain distances display a leveling-off.

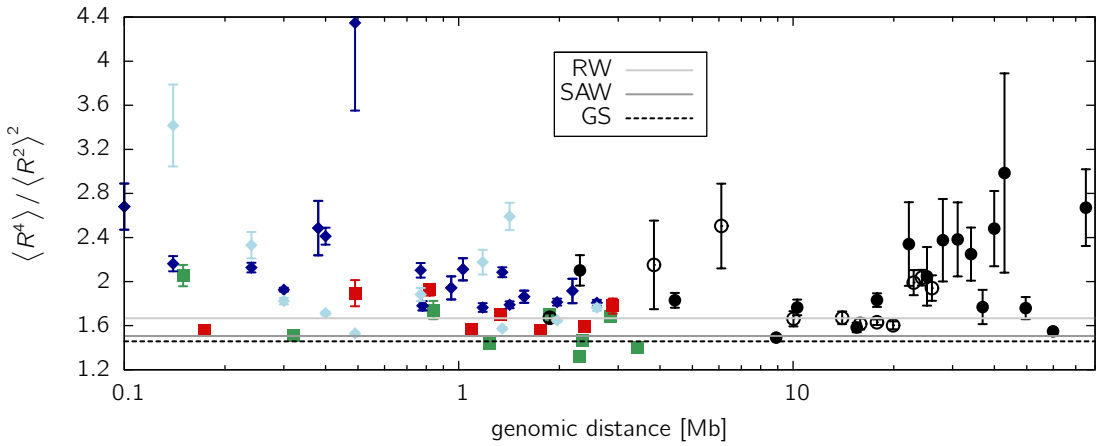


Figure 4.20: Higher-order moment ratio $c_4 = \langle R^4 \rangle / \langle R^2 \rangle^2$ for experimental data and polymer models. Shown is the value of c_4 for the random walk (RW), self-avoiding walk (SAW) and globular state (GS) model. The following experimental data is shown: *Human fibroblasts Chr1* [12]: ■ anti-ridge region, ■ ridge region, ∘ long distance measurements; *Human fibroblast Chr 11* [12]: • long distance measurements; *Murine Igh locus* [109], ♦ pre-pro-B cells, ♦ pro-B cells. Experimental measurements display larger fluctuations, i.e. c_4-values, than the polymer models.

different chain lengths. The mean square displacement shows a leveling-off and adopts a plateau level in the order of the linear system size L. The leveling-off is a direct consequence of the confined space the system is put in. The scaling law $\langle R^2 \rangle \sim N^{2\nu}$, however, is regained for the end-to-end distances marked by the red line in Fig. 4.19.

As both experimental data [12] and the globular state display a leveling-off in the mean square displacement, it has been proposed, that a globular state model can explain chromatin folding [31]. To test this hypothesis, we conduct a more sensitive test by looking at the moment ratios [cf. eq. (4.3)]. It was shown in Fig. 4.4 that the moment ratios display pronounced differences from a random walk or self-avoiding walk. They have been determined in section 4.4.1 for compact polymers. The ratio c_4 is displayed in Fig. 4.20 together with the data from Roel van Driel's lab [12] as well as data from another group [109]. Interestingly, the moment ratios for globular polymers are below the ratios of a self-avoiding walk, thus they perform even worse in comparison to experimental data.

4.6 Conclusions

In this chapter, we compared experimental data from FISH measurements to three fundamental polymer models, the random walk, the self-avoiding walk and the globular state model.

We found that the *distribution of measured distances* $P_g(r)$ for a certain genomic separation g is, due to insufficient data, not a good measure. By dividing out the leading order term of the *scaling law* $\langle R^2 \rangle \sim N^{2\nu}$ we found, that chromatin is not organized as a RW or SAW, a fact even better represented in the *dimensionless ratios of the higher-order moments* of the distance distribution [eq. (4.3)]. Interestingly, the fluctuations in the experimental data are much larger than predicted even by a random walk model, which naturally has quite large fluctuations (Fig. 4.4).

For compact polymers, little information was available on the intra-chain distance distributions, thus we performed extensive simulation runs to evaluate the conformational

properties of compact polymers. The algorithm we use here is basically the one proposed by Mansfield [105] for unbiased sampling Hamiltonian paths on a cubic $L \times L \times L$ lattice. The algorithm was modified by allowing vacancies on the cubic lattice such that we were able to sample compact polymers in the regime where $T \neq 0$ — or in terms of the lattice occupancy fraction $\rho \neq 1$. Using a highly parallel computing system we sampled conformations up to chain lengths of $N = 256\,000$ and three different densities $\rho = 0.1$, $\rho = 0.5$ and $\rho = 1$.

While a lot of studies are devoted to the properties of compact polymers, little attention has been paid to the end-to-end distance distribution as well as to the distribution of intrachain segments within a globular polymer. However, this information is crucial for comparison to experimental data [12]. Importantly, the moment ratios turned out to be independent of the system density ρ and approach a constant value in the limit of infinite chain length. The moment ratios of compact polymers were found to be

$$c_2 = 1.1395(5) \qquad c_3 = 1.421(2) \qquad c_4 = 1.458(1)$$

and thus are smaller than for the random walk or self-avoiding walk. The corresponding moments for intrachain distances were resembling the random walk behaviour on the scale of a few segments due to the screening effect, but pretty fast falling below the values of the endpoint distances (Fig. 4.11).

While the mean square displacement of intrachain distances of compact polymers displays a leveling-off (Fig. 4.19), the moment ratios clearly reveal, that experimental data is not described by a globular state. This falsifies a recent study [31] where it was claimed, based on the same experimental data used here, that chromatin might be explained by a globular state model. Thus, chromatin organization seems to be more complex than simple models would predict. The experimental values for the higher-order moments already guide the way to a more realistic model. The ratios are measures of the fluctuations in the system, and these fluctuations seem to be too small in all models presented so far.

Chapter 5

The Random Loop Model

Introduction and Analytical Approach

References

The results presented in this chapter are published in:

- M. Bohn, D. W. Heermann & R. van Driel, **2007**. Random loop model for long polymers. *Physical Review E* 76 (5), 051805.

Chapter Summary

Genome function in higher eukaryotes involves major changes in the spatial organization of the chromatin fiber. Nevertheless, our understanding of chromatin folding is remarkably limited. Polymer models have been used to describe chromatin folding. However, none of the proposed models gives a satisfactory explanation of experimental data. Particularly, they ignore that each chromosome occupies a confined space, i.e. the chromosome territory. Here, we present a polymer model that is able to describe key properties of chromatin over length scales ranging from 0.5 to 75 Mb. In contrast to other polymer models, this random loop (RL) model takes into account abundant experimental evidence of loops existing in a broad range of sizes and of loop attachment points varying from cell to cell. Assuming a homogeneous fiber with a random walk folding of the polymer backbone, the model defines a probability $\mathcal{P}$ for two monomers to interact, creating loops of a broad size range. Model predictions are compared to systematic measurements of chromatin folding of the q-arms of chromosomes 1 and 11. The experimentally observed folding into a confined space as well as the large cell-to-cell variation is correctly predicted by this RL model.

5.1 Introduction

The chromatin fiber inside the interphase nucleus of higher eukaryotes is folded and compacted on several length scales (↷ 2.2). Lacking high resolution live cell imaging techniques, indirect approaches have been applied to obtain information about chromatin folding on the scale above the detailed level of the chromatin fiber, e.g. fluorescence in situ hybridization (FISH) to measure the relationship between the physical distance between genomic sequence elements (in μm) and their genomic distance (↷ 2.3.4). There have been several attempts to explain the folding of chromatin in the interphase nucleus using polymer models (↷ 3.2). The strength of polymer models is their ability to make predictions on the structure of chromatin by pointing out the driving forces for observed folding motifs. These predictions can then be tested experimentally. However, a polymer model that is able to explain chromatin folding spanning different length scales is still lacking.

Although random walk models have been proposed for chromatin structure below 2 Mb, we have given unambiguous evidence using more sensitive analysis methods than in Refs. [92, 11] based on FISH data from several groups [27, 12], that chromatin folding is neither in a random walk, self-avoiding walk or compact globular state (↷ chapter 4). These results hold true both on the short scale ($< 2-10$ Mb) as well as on the large scale (> 10 Mb).

Folding at larger length scales has been explained using several models (↷ 3.2). Two polymer models have been proposed that introduce loops into the problem of chromatin folding. One is the random-walk/giant-loop (RWGL) model (↷ 3.2.2), which assumes a RW-backbone to which loops of about 3 Mb are attached [11]. A second model, the multi-loop-subcompartment (MLS) model (↷ 3.2.3), proposes rosette-like structures consisting of multiple 120 kb loops [28, 29]. None of these models is able to describe the folding of chromatin at all relevant length scales. All predict that the physical distance between

two FISH markers monotonously increases with the genomic distance. Clearly, this is incorrect at bigger length scales, since the chromatin fiber is geometrically confined by the dimensions of the cell nucleus. More so, individual chromosomes have been shown to occupy subnuclear domains that are much smaller than the nucleus itself, i.e. the chromosome territories with sizes in the range of one to a few micrometers [7]. Evidently, an intrinsic property of the chromatin fiber inside the cell nucleus is that it assumes a compact state that cannot be described by classic polymer models. This raises the fundamental question of what physical principles make chromatin fold in a limited volume.

Importantly, a feature that has been neglected in any of the polymer models is that experimental data displays huge cell-to-cell variations, which cannot be explained by simple polymer models (cf. Fig. 4.20). Intriguingly, these fluctuations are larger than for a random walk, which, due to lack of excluded volume, only has a minimum of conformational constraints, therefore displaying large fluctuations.

The Random Loop model which is introduced in this chapter, amongst others overcomes these limitations of older polymer approaches on chromatin folding. In fact, the objective of this Random Loop model is much broader, aiming at offering a unified explanatory framework for the connection between genome folding and function. Abundant experimental evidence indicates that this connection is maintained by chromatin loops ($\curvearrowright$ 2.3), which seem to constitute a key mechanism for transcriptional regulation [20, 25]. Amongst others, the model predicts a confined folding of chromosomes, the formation of chromosome territories and the large cell-to-cell variation based on probabilistic or dynamic loop formation. Indeed, we want to approach a model explaining all of the experimental observations addressed in section 1.2.

In this chapter, we first introduce the basic assumptions of the Random Loop (RL) model. The model presented in section 5.2 is quite general and allows for several realizations. The first one, assuming a homogeneous fiber without excluded volume interactions, qualifies for a partly analytical solution ($\curvearrowright$ 5.3). Although not explaining each experimental issue, this homogeneous RL model predicts the confined folding and cell-to-cell variation. Heterogeneity of the chromatin fiber, excluded volume interactions and the dynamics of loop formation will be incorporated in later chapters.

5.2 Basic model assumptions

We will now introduce the core ideas of the Random Loop (RL) model. They originate from reconciling experimental approaches on chromatin folding with evidence for its relationship to function. A possible link between genome folding and function is given by experimental data from nC ($n = 3, 4, 5$) experiments ($\curvearrowright$ 2.4.2), where a map of specific chromatin-chromatin contacts is established, showing that

1. chromatin-chromatin contacts do exist on all scales from a few kb to several tens of Mb [25]

2. the number of interactions rapidly decreases with genomic separation [66, 39]

3. loops or contacts are closely related to transcriptional regulation. Contacts found in 4C experiments [25] show significant differences in liver and brain cells, indicating a dependence on the differentiation state of the cell.

On the other hand FISH experiments probing structure of chromatin reveal that

1. chromosomes are confined to a subspace of the nucleus, the mean square distance between two FISH markers does not increase above about 2 μm [12].

2. chromosomes as a whole segregate into distinct territories [7]. Also, small regions of about 10 Mb chromatin on the same chromosome tend to segregate [6].

3. FISH measurements display a strong cell-to-cell variation exceeding the values of normal polymer models

4. compaction is different depending on transcriptional activity [6, 12]

How can we bring these results together? We start our considerations with a linear polymer such as a random walk or a self-avoiding walk. Further interactions are necessary to obtain a leveling-off in the mean-square distance between two markers. These interactions might well be given by the loops, however, chromatin models with loops proposed so far, fail to explain important aspects of experimental data ($\curvearrowright$ 3.2). The key idea is now to introduce disorder or dynamics in the loop formation process, leading to the following model assumptions:

- The backbone of the chain is a linear polymer, either with or without excluded volume, i.e a sequence of N monomers with coordinates $\mathbf{r}_1, \ldots, \mathbf{r}_N$, where adjacent ones are connected by a linker.

- Additionally, the polymer forms intra-chromatin contacts, i.e. monomer i might interact with monomer j in an attractive fashion. This corresponds e.g. to regulatory elements coming into contact with the target sequence they regulate.

- Contacts in the model are not fixed, rather the interactions are probabilistic and/or dynamic. This leads to different loop configurations either in the ensemble space or in time.

- Interactions are allowed on a broad range of scales.

In this general formulation, the RL model incorporates important available experimental evidence on loops: Amongst others it allows for loops on all scales (4C experiments [25]) and does not require loops to exist in all cells at the same time. Playing an important role in transcriptional regulation, loops probably only persist as long as activation/repression of a gene is necessary, allowing the cell to form other regulatory contacts afterwards without running into topological problems. This is confirmed by FISH control experiments [25], showing that abundant and significant contacts revealed by 4C are only found in about 5-10% cells of a population.

The probabilistic loop formation can be achieved by different methods and some of them will be presented in the next few chapters: Analytical results can be obtained to some extend for a Gaussian chain as a backbone with probabilistic Gaussian interactions (see below). Molecular Dynamics simulations of this model with excluded volume interactions are demonstrated in chapter 6. To incorporate transcriptional heterogeneity into the RL model, we extend it by defining looping probabilities locally, resulting in a stunning agreement with experimental data (chapter 7). Finally, in chapter 10, we present a dynamic model where loop formation is achieved without long-range interactions.

5.3 The homogeneous Random Loop Model in the annealed and quenched average

5.3.1 The Hamiltonian

To set up our Random Loop Model in an analytically tractable way, we neglect effects of excluded volume and consider a Gaussian chain (↷ 3.1.4) as backbone,

$$U_{\text{Gaussian}} \equiv U_0 = \frac{\kappa}{2} \sum_{j=1}^{N} \| \mathbf{r}_j - \mathbf{r}_{j-1} \|^2 \ . \tag{5.1}$$

In addition to the random walk backbone, we allow each bead to interact with any other via a harmonic potential, resulting in the complete interaction term

$$U = U_{\text{Gaussian}} + \frac{1}{2} \sum_{\substack{i<j \\ |i-j|>1}}^{N} \kappa_{ij} \| \mathbf{r}_i - \mathbf{r}_j \|^2 \ ,$$

where $\kappa_{ij} = \kappa_{ji}$ are the spring constants between the loop attachment points. This potential has already been proposed by Sachs *et al.* [11]. Here we extend it according to the model's assumptions: The loop attachment points and interactions are not fixed, rather they are probabilistic. This kind of disorder is achieved by choosing the κ_{ij} as Bernoulli distributed random variables, i.e.

$$\kappa_{ij} = \begin{cases} \kappa & \text{with probability } \mathcal{P} \\ 0 & \text{with probability } 1 - \mathcal{P} \end{cases} \tag{5.2}$$

The potential can be rewritten in the form

$$U = \frac{1}{2} \sum_{i<j}^{N} \kappa_{ij} \| \mathbf{r}_i - \mathbf{r}_j \|^2 = \frac{1}{4} \sum_{\substack{i,j=0 \\ j\neq i}}^{N} \kappa_{ij} \| \mathbf{r}_i - \mathbf{r}_j \|^2 \ . \tag{5.3}$$

where $\kappa_{ij} = \kappa$ for $|i - j| = 1$.

Thus, loops are implemented by a Gaussian potential in this model and the loop attachment points are chosen randomly between all pairs of non-adjacent beads i, j. Similar potentials are found in case of random resistor networks [114].

5.3.2 Quenched or annealed ensemble?

The principal aim is to calculate ensemble averages, like the mean square distance $\langle R_n^2 \rangle$ between two arbitrary beads separated by n monomers allowing direct comparison to FISH measurements.

Averages of an observable A in a system with Hamiltonian $\mathcal{H}$ are calculated using the partition sum $\mathcal{Z} = \text{Tr} \exp(-\beta\mathcal{H})$,

$$\langle A \rangle = \frac{1}{\mathcal{Z}} \text{Tr} \left[A \exp(-\beta\mathcal{H}) \right] . \tag{5.4}$$

The Hamiltonian of the Random Loop Model, however, depends on both statistical variables $\mathbf{r}_1, \ldots, \mathbf{r}_N$ representing the conformational degrees of freedom and a set of random variables representing the disorder κ_{ij} (Fig. 5.1). The sequence in which the averaging

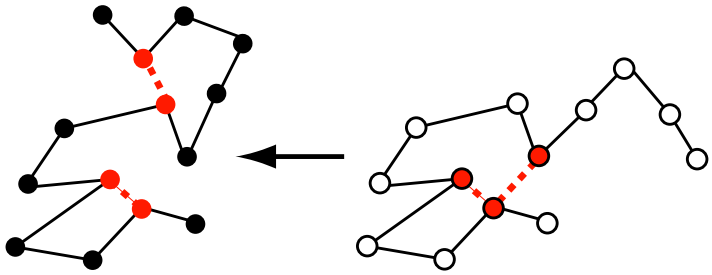


Figure 5.1: The Random Loop model averages over (a) the thermal disorder and (b) over possible configurations of loops. Two sample conformations of the ensemble can be seen in this figure. Loop connections are drawn by red dotted lines.

procedures have to be executed, depends on the typical time scales over which the disorder, i.e. the loop attachment points, (τ_{dis}) and the conformational degrees of freedom (τ_{eq}) change.

If the random variables change fast, i.e. $\tau_{dis} \ll \tau_{eq}$, the disorder average has to be carried out at the same time as the average over the conformational variables, resulting in a free energy

$$F = -kT \ln \langle \mathcal{Z}(\{\mathbf{r}_i\}) \rangle_{dis} \ . \tag{5.5}$$

The braces $\langle \cdot \rangle_{dis}$ denote the average over the set of random variables $\{\kappa_{ij}\}$. This ensemble is called the *annealed ensemble*.

In the opposite case, when the time scale over which the loop attachment points change is large compared to the equilibration time of the polymer, i.e. $\tau_{dis} \gg \tau_{eq}$, the average over the disorder has to be evaluated after all other averages. In this *quenched average* we calculate the free energy as

$$F = -\langle kT \ln \mathcal{Z}(\{\mathbf{r}_i\}) \rangle_{dis} \ . \tag{5.6}$$

Most of our work in this section concentrates on the calculation of the quenched average, assuming that once loops have formed, the cell has enough time to adopt at least a local equilibrium-like state. If this was not the case, it would be hard for the regulatory mechanisms to behave in a controlled way. However, the question of time scales has not been answered yet. The life-time of loops is unknown as both FISH as well as 3C/4C/5C techniques are performed in a fixated environment, unable to evaluate any dynamics. Recently, a computer simulational study [30] suggested that chromatin is never in an equilibrated state on a global scale. However, this study neglected the effect of topoisomerase II, which was shown to greatly reduce the disentanglement time of polymers [97].

5.3.3 The quenched average

Here we want to calculate the mean square distance $\langle r_{IJ}^2 \rangle$ between two beads I and J, allowing for a direct comparison to FISH measurements (Fig. 2.3). To calculate the quenched average, we choose a fixed set of random variables $\{\kappa_{ij}\}$ and first average over the thermal fluctuations. The probability density for a bead conformation $(\mathbf{r}_0, \ldots, \mathbf{r}_N)$ in the canonical ensemble is given by the Boltzmann factor

$$P(\mathbf{r}_0, \ldots \mathbf{r}_N) = C \left(-\frac{U}{k_B T} \right) , \tag{5.7}$$

where C is a normalization constant and $U = U(\mathbf{r}_0, \ldots \mathbf{r}_N)$ is the total potential energy of the chain.

We now eliminate the degrees of freedom stemming from the translational invariance of the problem by setting $\mathbf{r}_0 = \mathbf{0}$ (the absolute position of the chain in space is irrelevant for distances between beads).

Due to the Gaussian character of the probability density P, the spatial dimensions factorize,

$$P(\mathbf{r}_1, \ldots, \mathbf{r}_N) = P_1(x_1, \ldots, x_N) \cdot P_1(y_1, \ldots, y_N) \cdot P_1(z_1, \ldots, z_N)\,,$$

and we can concentrate on the one-dimensional density function P_1. By an easy calculation omitted here, P_1 can be rewritten,

$$P_1(x_1, \ldots, x_N) = C_1 \exp(-\frac{1}{2}\mathbf{X}^T K \mathbf{X})\,, \tag{5.8}$$

where $\mathbf{X} = (x_1, \ldots, x_N)^T$ and

$$K = \begin{pmatrix} \sum_{\substack{j=0 \\ j\neq 1}}^{N} \kappa_{1j} & -\kappa_{12} & \ldots & -\kappa_{1N} \\ -\kappa_{21} & \sum_{\substack{j=0 \\ j\neq 2}}^{N} \kappa_{2j} & \ldots & -\kappa_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ -\kappa_{N1} & -\kappa_{N2} & \ldots & \sum_{\substack{j=0 \\ j\neq N}}^{N} \kappa_{Nj} \end{pmatrix}. \tag{5.9}$$

From now on, we assume the κ_{ij}'s to be given in units of $k_B T$. We have not made any assumptions yet concerning the spring constants (i.e. basically the matrix entries) κ_{ij}. In the following we only assume that K is a *symmetric* and *regular* matrix, so that P_1 in eq. (5.8) turns out to be a multivariate normal distribution with mean $\mu = 0$ and covariance matrix $\Sigma = K^{-1}$. The marginal distribution for two arbitrary beads I and J,

$$P(x_I, x_J) = \int \ldots \int \prod_{\substack{i=1 \\ i\neq I,J}}^{N} P(x_1, \ldots, x_N)\,, \tag{5.10}$$

can be evaluated by standard methods for normal distributions. Going back to three dimensions we obtain after some basic integral evaluations the joint probability density for the distance between two beads I and J,

$$P(\parallel \mathbf{r}_I - \mathbf{r}_J \parallel) \equiv P(r_{IJ}) = \tilde{C}\, r_{IJ}^2 \, \exp\left[-\frac{1}{2}\frac{1}{\sigma_{JJ} + \sigma_{II} - 2\sigma_{IJ}} r_{IJ}^2\right]\,.$$

Here

$$\Sigma = K^{-1} = (\sigma_{ij})_{i,j}$$

and $\tilde{C}$ is the normalization constant. Using

$$\Gamma = \frac{1}{2}\frac{1}{\sigma_{JJ} + \sigma_{II} - 2\sigma_{IJ}}$$

and calculating the correct normalization we obtain

$$P(r_{IJ}) = \frac{4}{\sqrt{\pi}}\, \Gamma^{\frac{3}{2}}\, r_{IJ}^2 \, \exp\left[-\Gamma r_{IJ}^2\right] \tag{5.11}$$

and finally

$$\left\langle r_{IJ}^2 \right\rangle_{\text{thermal}} = \int r_{IJ}^2 P(r_{IJ}) dr_{IJ} = \frac{3}{2\Gamma} = 3(\sigma_{JJ} + \sigma_{II} - 2\sigma_{IJ})\,.$$

The bracket delimiters here denote the average over the thermal ensemble of $N+1$ beads interacting via a given, but fixed harmonic potential.

We have now derived a general formula for the mean square distance between two arbitrary beads of the chain where each bead may interact with any other via an harmonic potential. This quantity turned out to depend only on the matrix K, or more accurately speaking, on its inverse. Now we want to specify the matrix K, which contains all information about the interactions. Our model assumes the chromatin fiber to have a random walk backbone, meaning that we have to set $\kappa_{ij} = \kappa$ with $|i-j| = 1$. Furthermore, chromatin forms loops whose size and positions are randomly distributed along the chain. In a first approach, we choose the κ_{ij} to be Bernoulli distributed random variables, as given in eq. (5.2). Only being a constant scaling factor, κ can be set to unity. Thus, our model has two adjustable parameters, namely the chain length N and the probability $\mathcal{P}$.

The resulting matrices $\{K\}$ represent an ensemble of diagonally dominated band random matrices and each matrix of this ensemble represents a loop configuration. This ensemble of random matrices has been investigated recently [115]. We are interested in the ensemble average of the mean square distance, i.e. in the quantity

$$\left\langle r_{IJ}^2 \right\rangle = \left\langle \left\langle r_{IJ}^2 \right\rangle_{\text{thermal}} \right\rangle_{\text{dis}} = 3\left(\langle \sigma_{JJ} \rangle_{\text{dis}} + \langle \sigma_{II} \rangle_{\text{dis}} - 2\langle \sigma_{IJ} \rangle_{\text{dis}}\right) , \tag{5.12}$$

where the average over the ensemble of loops is performed as a quenched average and is equivalent to averaging over the ensemble of random matrices given by the above constraints.

Our simple model presented here assumes homogeneity of the chromatin fiber, i.e. translational invariance. Thus, we are not interested in the mean square distance between two specific beads i and j, rather in an averaged distance between two beads separated by n monomers,

$$\left\langle R_n^2 \right\rangle = \frac{1}{N-n+1} \sum_{i=0}^{N-n} \left\langle r_{i,i+n}^2 \right\rangle . \tag{5.13}$$

The average over the ensemble $\{K\}$ of random matrices cannot be performed analytically. To obtain approximate results for $\langle R_n^2 \rangle$, we randomly select a subset of matrices K and calculate its inverse using the C library CLAPACK [116]. From this inverse matrix, the thermal average eq. (5.12) can be obtained. We then average over the resulting mean square distances between any two beads and several realizations of the matrix K according to eq. (5.13).

5.3.4 Results and comparison to experiments

Our polymer model makes use of coarse-graining, since it is impossible to model a long fiber like a chromosome in detail. Restrictions are given by computing time, which basically depends on the size of the matrix K. Choosing a matrix size (chain length) of $N = 1000$ is a good compromise between required computing time and the level of coarse-graining. Using a coarse-graining approach implies that we neglect details on a scale below the effective segment length being 150 kb in the following figures. Therefore we cannot resolve those loops that have been investigated in some gene-expression systems like the β-globin locus. This, however, does not affect the possibility of confined folding on the large scale, rather it leads to a rescaling of the effective Kuhn length. Us being interested in large scale chromatin organization, it is thus justified to neglect these loops on the short scale.

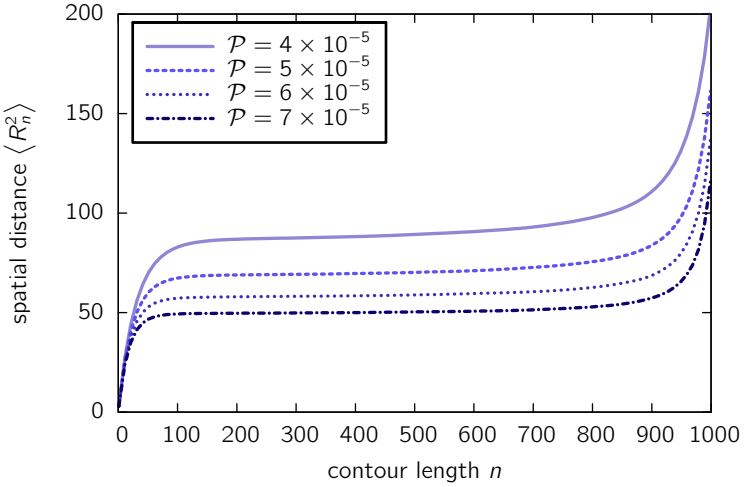


Figure 5.2: Mean square distance $\langle R_n^2 \rangle$ between two chain segments in relation to their contour length separation n for the Random Loop model with loops on all scales. The chain length is $N = 1000$. With this homogeneous Random Loop model, a leveling off to $\langle R_n^2 \rangle \sim \mathcal{O}(1)$ can already be achieved by a small number of loops. Looping probabilities $\mathcal{P}$ control the height of the plateau level.

The mean square distance

We evaluate the mean square distance for our model, where we allow loops on all scales. The spatial distance $\langle R_n^2 \rangle$ between two beads separated by n monomers is shown in Fig. 5.2. At short contour lengths, $\langle R_n^2 \rangle$ grows similar to a random walk, but soon a leveling-off can be observed being fairly $\sim O(1)$ due to the attractive long-range interactions. While the contour length approaches N, the mean square distance again rises to a random-walk like behavior. This is a chain-end-effect which is not of interest to us, as experiments only measure intra-chain distances. Firstly, this effect is due to the construction of the loops, as the probability for having a loop with a larger size becomes increasingly small. Secondly, the entropic degrees of freedom for the chain ends are larger than for intrachain segments.

Thus, adding long-range interactions like loops yields completely different traits than a simple random walk or self-avoiding walk model. No scaling behavior where $\langle R^2 \rangle \sim N^{2\nu}$

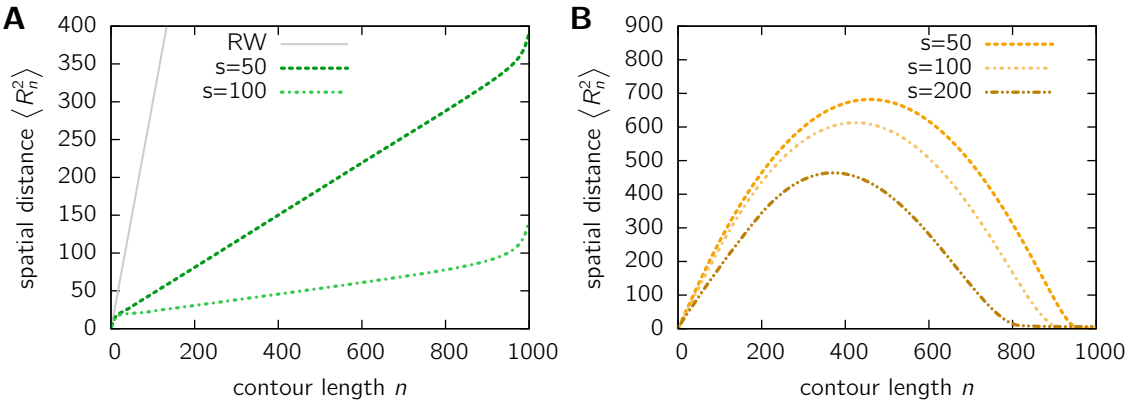


Figure 5.3: Mean square distance $\langle R_n^2 \rangle$ between two chain segments in relation to their contour length separation n for the Random Loop model with loops restricted to certain sizes. The chain length is $N = 1000$. In **A.** only loops in the interval of sizes $[l_1, l_2] = [2, s]$ are formed. The resulting scaling behavior is still the random walk type $\langle R_n^2 \rangle \sim n$, but with a rescaled segment length. **B** shows the model where only loops of sizes ℓ in the range $[N - s, N]$ are admitted. While large loops seem to be responsible for the collapse of the chain, they alone cannot explain the experimental data. In both figures, the looping probability $\mathcal{P}$ is chosen such that the mean number of loops per configuration is 100.

can be recovered any more.

It is an interesting question, whether it is really necessary to have loops on all scales to obtain such a leveling-off. As already mentioned, in all polymer models presented so far, including the RWGL (↷ 3.2.2) as well as the MLS model (↷ 3.2.3) the loop sizes were basically fixed. One might ask, whether small loops (in the order of 100 kb to 1 Mb) already lead to the leveling-off, or whether loops on all scales up to 80 Mb are necessary. In order to restrict the loop sizes ℓ to lengths between $[l_1, l_2]$, we redefine the random variables κ_{ij} by setting

$$\kappa_{ij} = \begin{cases} \kappa & \text{with probability } \mathcal{P} \\ 0 & \text{with probability } 1-\mathcal{P} \end{cases}, \qquad \text{if } l_1 \leq |i-j| \leq l_2$$

$$\kappa_{ij} = 0 \qquad \text{otherwise}$$

First, the loop sizes are restricted to not extending a certain size: $\ell \in [1, s]$. Results in Fig. 5.3A show that introducing small loops still results in a scaling $\langle R_n^2 \rangle \sim n$, not yielding the experimentally observed $\langle R_n^2 \rangle \sim O(1)$, although the slope becomes smaller, meaning that the effective segment length has been rescaled. In Fig. 5.3B we analyzed the ensemble where only large loops in the range $[N-s, N]$ were allowed. Obviously, large loops are responsible for forcing the chain to collapse, but the behavior on intermediate lengths does not fit experimental data. Therefore loops on all scales are needed to obtain the leveling-off observed in experiment.

Comparison to experimental data

In Fig. 5.4 the model is compared to the experimental data for different values of $\mathcal{P}$. Two new scaling parameters have to be introduced, the segment length in physical units (e.g. nm) and the segment length in base pairs. The data is shown for a segment length of 300 nm and 150 kb. The latter is the size of the fluorescent markers used in experiments, rendering it unnecessary to model on a more detailed scale. The model quite well explains the leveling-off at genomic distances above a few Mb as well as the increase at small genomic distances. As we have shown in section 4.2 that for small genomic distances a random walk behaviour is incorrect, this random-walk-based model does not yield perfect results here.

On the large scale, adding long-range interactions forcing the polymer to form loops yields completely different traits than a simple random walk or self-avoiding walk model. Note that the probabilities $\mathcal{P}$ are chosen very small, meaning that a few loops suffice to obtain this leveling-off. The number of independent randomly chosen entries κ_{ij} is $\mathcal{C} = (N-1)(N-2)/2$ for a $N \times N$-matrix and therefore the average number of loops per configuration is given by $\mathcal{C} \cdot \mathcal{P}$. With $\mathcal{P} = 4 \times 10^{-5}$ and $N = 1000$ one has an average of about 20 loops.

It was already mentioned in section 4.2 that plotting $\langle R_n^2 \rangle$ versus n is not a very sensitive method to check for the correctness of a model. Dimensionless higher-order moments were introduced [eq. (4.3)], being a measure of the fluctuations in the ensemble. They give much stronger evidence for the correctness of a model, as these ratios are related to the distribution of the distances and not only its average value. Here we focus on the fourth order ratio $c_4 = \left\langle R^4 \right\rangle / \langle R^2 \rangle^2$, being plotted in Fig. 5.5 for different looping probabilities. On the scale of a few polymer segments, fluctuations are similar to the random walk value due to the random walk backbone. However, the fluctuations increase quickly to values around $c_4 \approx 2$. These values are significantly larger than the random

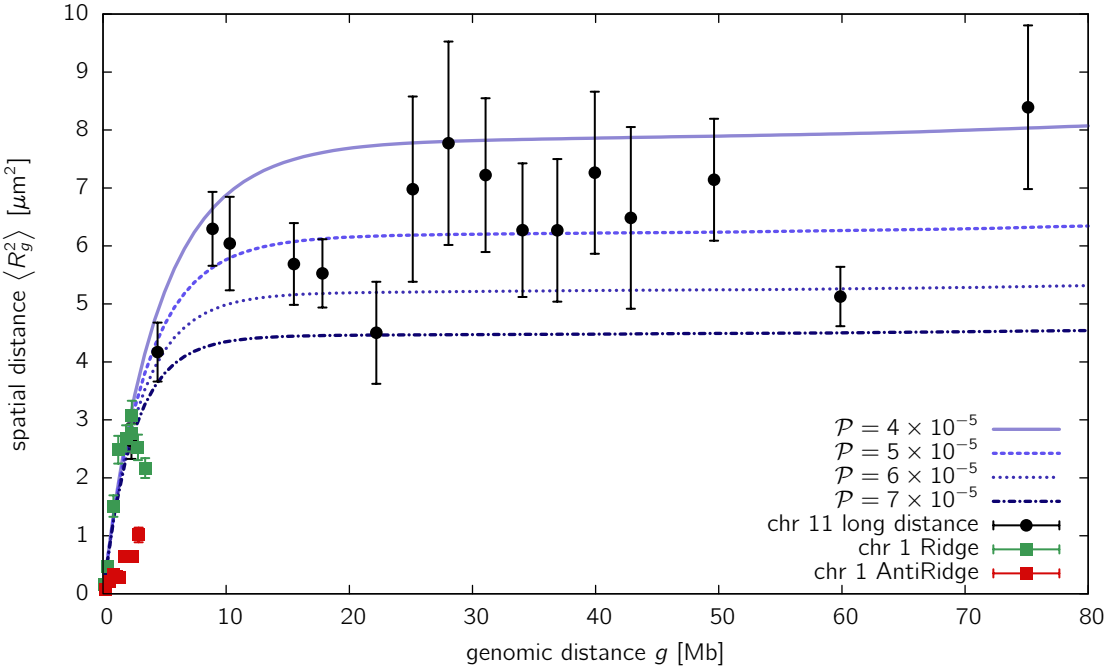


Figure 5.4: Experimental data from FISH measurements compared to the Random Loop model with loops on all scales. The data is taken from Ref. [12] and includes short and long distance sets in human interphase cells. The results of the Random Loop model are shown for $N = 1000$ and different values of $\mathcal{P}$, assuming a segment length of 150 kb.

walk value and already quite close to the data. The Random Loop model thus shows a good agreement with experimental data, performing much better than previous models. The large values of c_4 arise in our model from the disorder average. Importantly, if we do not carry out the disorder average, but calculate c_4 for a fixed loop configuration over the distribution function in eq. (5.11), we exactly obtain the random walk value $c_4 = 5.0/3.0$. Thus, probabilistic or dynamic loop formation seems to be an indispensable property of chromatin organization.

5.3.5 The annealed ensemble

The calculations presented in section 5.3.3 resemble a quenched average procedure: First, the average over the thermal ensemble, i.e. the coordinates $\mathbf{r}_i$ are carried out, and only in the end, the observable is disorder-averaged. Here we want to document some calculations for the annealed ensemble, where the average over the set of random variables $\{\kappa_{ij}\}$ is taken simultaneously with the average over the statistical variables $\{\mathbf{r}_i\}$.

The partition sum for one configuration of random variables $\{\kappa_{ij}\}$ is given by

$$\mathcal{Z}\{\kappa_{ij}\} = \int d\mathbf{r}_1 \dots d\mathbf{r}_N \exp(-U(\mathbf{r}_1, \dots, \mathbf{r}_N)) \,. \tag{5.14}$$

The annealed average over the disorder can then be carried out

$$\langle \mathcal{Z} \rangle_{\text{dis}} = \langle \mathcal{Z}(\{\kappa_{ij}\}) \rangle_{\text{dis}} = \sum_{\{\kappa_{ij}\}} \mathcal{Z}(\{\kappa_{ij}\}) p(\{\kappa_{ij}\}) \,. \tag{5.15}$$

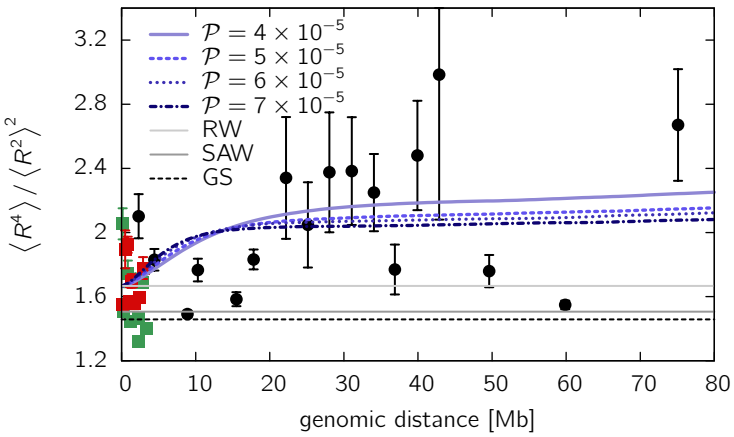


Figure 5.5: The moment ratio $c_4 = \langle R^4 \rangle / \langle R^2 \rangle^2$ for FISH distance measurements (cf. Fig. 5.4) compared to the results of the Random Loop model. For reference, the results for the RW, SAW and GS model are shown (cf. Fig. 4.20). The Random Loop model predicts larger fluctuation than does the random walk model, in agreement with experimental data.

The last sum is over all possible sets of κ_{ij}, p being the probability measure acting on the space of $\{\kappa_{ij}\}$. Using the definition of κ_{ij} from eq. (5.2), we find

$$\langle \mathcal{Z} \rangle_{\text{dis}} = \int d\mathbf{r}_1 \dots d\mathbf{r}_N \exp(-U_0) \times \prod_{i<j-1} \left[\mathcal{P} \left(V_{ij} - 1 \right) + 1 \right] , \tag{5.16}$$

where

$$V_{ij} = \exp\left(-\frac{1}{2} \hat{\kappa} \parallel \mathbf{r}_i - \mathbf{r}_j \parallel^2 \right) \tag{5.17}$$

and U_0 is the Gaussian chain part of the potential, eq. (5.1). Introducing an effective potential

$$U_{\text{eff}} = U_0 - \sum_{i<j-1} \log \left[1 + \mathcal{P} \left(V_{ij} - 1 \right) \right] , \tag{5.18}$$

the annealed average can be written as

$$\mathcal{Z}_{\text{ann}} = \langle \mathcal{Z} \rangle_{\text{dis}} = \int d\mathbf{r}_1 \dots d\mathbf{r}_N \exp\left(-U_{\text{eff}} \right) . \tag{5.19}$$

Adjacent beads are interacting with the normal harmonic potential, while non-adjacent beads are interacting via the potential

$$V(r) = -\log \left[1 + \mathcal{P} \left(\exp\left(-\frac{1}{2} \kappa \, r^2 \right) - 1 \right) \right] . \tag{5.20}$$

The effective potential has two parts: Adjacent beads with $|i - j| = 1$ keep their attractive harmonic potential, while *all* non-adjacent beads interact via a pairwise attractive potential $V(r)$. The partition sum in (5.19) cannot be evaluated analytically and therefore we do not obtain an expression for the mean square distance in the annealed case. This potential $V(r)$, displayed in Fig. 5.6, is characterized by a minimum at $r = 0$, while for large r it reaches a plateau at $V(r \to \infty) = -\log(1 - \mathcal{P})$.

It is interesting to perform a low temperature approximation; a series expansion around $r = 0$ up to second order gives

$$V(r) = \frac{1}{2} \mathcal{P} \kappa r^2 \tag{5.21}$$

– a harmonic potential with effective spring constant $\mathcal{P}\kappa$.

Although we cannot present results for the mean square distance between to two beads here, in chapter 6 the effective potential is implemented via Molecular Dynamics simulations to obtain these results.

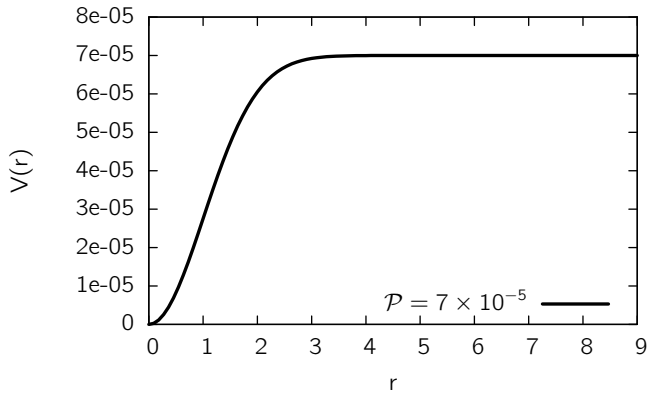


Figure 5.6: The effective potential $V(r)$ between two non-adjacent beads in the annealed ensemble. At short separations there is a soft attractive interaction, while for large distances r it reaches a constant value of $-\log(1-\mathcal{P})$.

5.3.6 Limiting cases without disorder

Although neglecting excluded volume interactions, the RL model presented above does not allow for a rigorous analytical solution. In the quenched ensemble it is impossible to calculate the disorder average while the annealed ensemble renders the evaluation of the disorder-averaged partition sum futile. In the quenched case, sample averages can be calculated using numerical matrix inversion techniques (↷ 5.3.3). Two special cases exist where the model can be solved analytically: The limiting cases where no disorder is present. $\mathcal{P} = 0$ is the situation of a normal Gaussian Chain with spring constant κ. It is well-known (↷ 3.1.4) that the mean square distance between two beads separated by n monomers is given by $\langle R_n^2 \rangle = \frac{3}{\kappa} n$. The other limit, $\mathcal{P} = 1$, corresponds to a fully connected network of beads. Assuming that all beads interact with spring constant κ, we can solve this problem analytically. Here, we basically do not deal with a linear chain any more. The interaction matrix $K = (k_{ij})_{i,j}$ in this case writes

$$k_{ij} = \begin{cases} N\kappa & \text{for} \quad i = j \\ -\kappa & \text{for} \quad i \neq j \end{cases} \tag{5.22}$$

By an easy calculation one can show that the inverse matrix is given by

$$\sigma_{ij} = \begin{cases} \frac{2}{(N+1)\kappa} & \text{for} \quad i = j \\ \frac{1}{(N+1)\kappa} & \text{for} \quad i \neq j \end{cases} \tag{5.23}$$

Recall our definition of the chain at the beginning of section 5.3.3: Although we have an $N \times N$-matrix our chain has $N+1$ beads, as we set $\mathbf{r}_0 \equiv 0$. Inserting into eq. (5.12) yields

$$\left\langle R_n^2 \right\rangle \equiv \left\langle r_{ij}^2 \right\rangle = \frac{3}{(N+1)\kappa/2} \tag{5.24}$$

Within this system two beads are interacting with an effective harmonic potential with $\kappa_{\text{eff}} = (N+1)\kappa/2$.

Of major interest is the case where $\mathcal{P} = 1$, but where adjacent beads interact with a different spring constant than loops, i.e. $\kappa_{ij} = \kappa$ for $|i-j| = 1$ and $\kappa_{ij} = \hat{\kappa}$ for $|i-j| > 1$. We were not able to solve this case analytically. One might take this system as a model for the low-temperature limit of the annealed case in eq. (5.21) where κ is replaced by an effective interaction $\hat{\kappa} = \mathcal{P}\kappa$. On a more general footing this case might also be regarded as a model for a system where the random attraction with probability $\mathcal{P}$

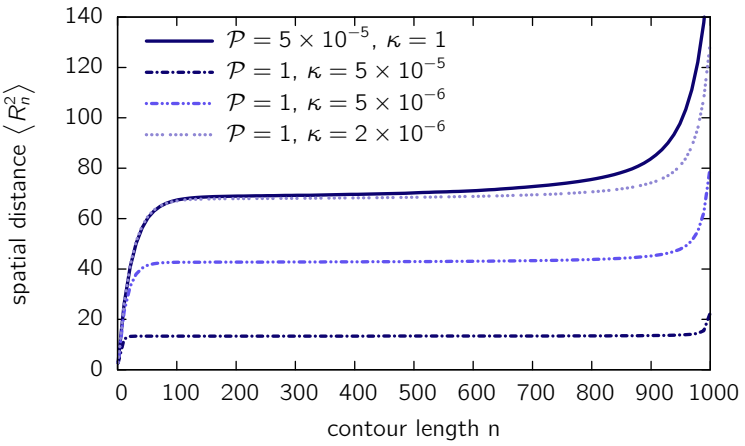


Figure 5.7: The Random Loop model compared to a system where the random attraction (setting $\kappa_{ij} = 1$ with probability $\mathcal{P}$) has been replaced by an average attraction (setting all $\kappa_{ij} = \mathcal{P}\kappa$ for $|i - j| > 1$). Shown are the RL model reference curve for $\mathcal{P} = 5 \times 10^{-5}$ (grey line), the corresponding system with average attraction (black line) and two systems with smaller average attraction.

and loop spring constant κ has been replaced by an average attraction with probability $\mathcal{P} = 1$ and loop spring constant $\mathcal{P}\kappa$. It is clear a priori that such a potential will lead to a collapse of the chain, as all beads are interconnected. In Fig. 5.7 we chose $\hat{\kappa} = \kappa = 1$ and $\mathcal{P} = 4 \times 10^{-5}$ as the reference curve. In comparison with the case of average attraction ($\mathcal{P} = 1, \kappa = 1, \hat{\kappa} = \mathcal{P}$) the leveling-off is much less pronounced. Of course it is possible to come into close agreement with the reference curve by choosing another interaction constant. For our reference curve one would have to lower $\hat{\kappa}$ by about one order of magnitude, corresponding to $\mathcal{P} \sim 2 \times 10^{-6}$ (< 1 loop per chain!). Although one could fit the data with these averaged attraction potential concerning the mean square distance, this would not be true for the higher-order moment ratios. And, even more important, such a potential surely will not exist in the cell.

5.4 Conclusions

In this chapter, a polymer model was presented which lifts major short-comings of the ones having been proposed so far. These models, either assuming no loops or loops of a fixed size, do not satisfactorily explain both the folding into a confined sub-space [12] of the nucleus as well as a huge cell-to-cell variation. Experiments show that the chromatin fiber inside the human cell nucleus is ruled by two different folding regimes: At small genomic distances, i.e. below $\sim 3-10$ Mb the data increases monotonicly with genomic separation. At larger genomic distances there is a leveling-off to a scaling behavior $\langle R^2 \rangle \sim O(1)$ (Fig. 2.3), indicating the presence of long-range interactions.

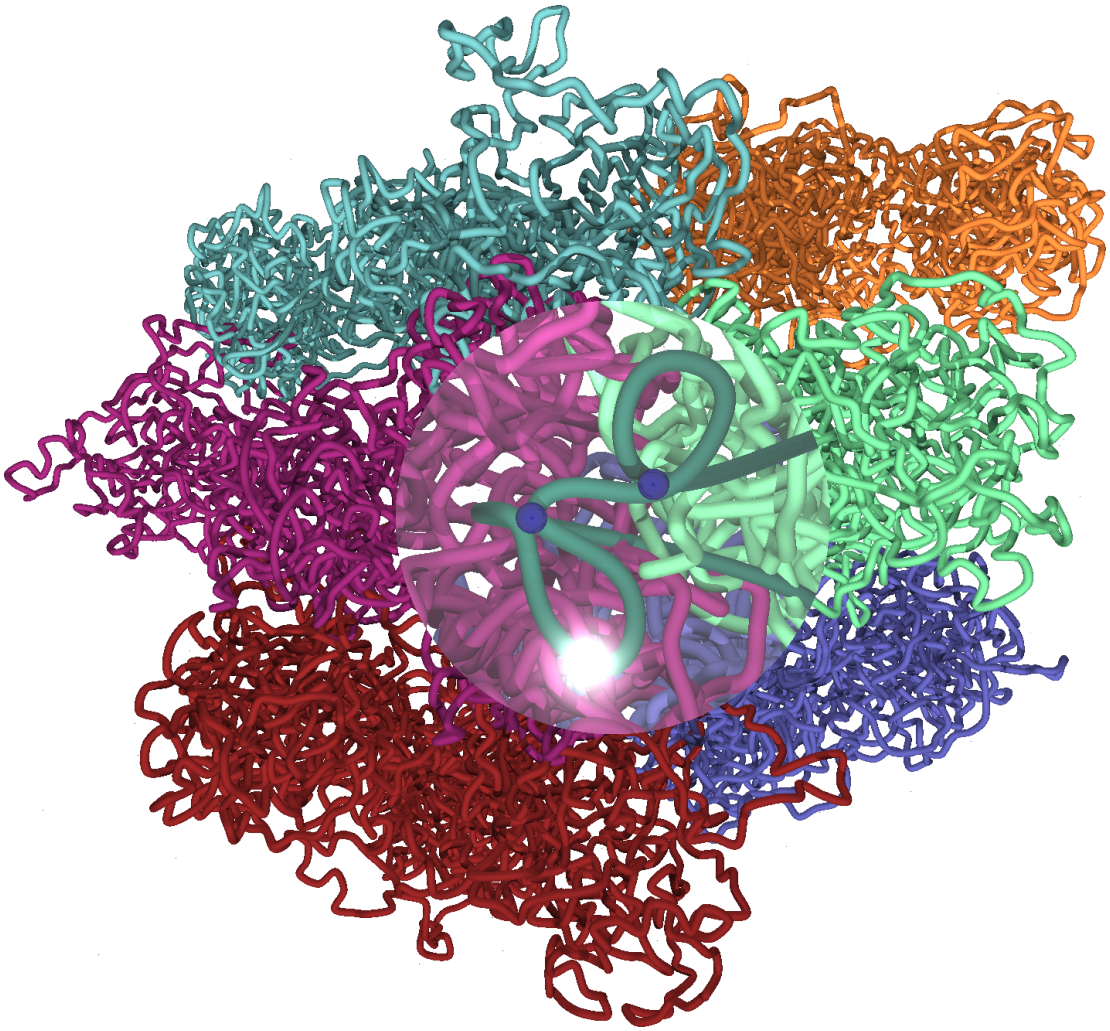
The Random Loop model introduced here incorporates the experimental evidence for the existence of loops of a broad range of sizes, which are not fixed but varying from cell-to-cell or over time. We have presented the general assumptions ($\curvearrowright$ 5.2) of the RL model, which can be implemented in several ways. For example, excluded volume can be integrated and the pathways of loop formation can be varied. In this chapter, a possible implementation was proposed, which can be solved partly analytically by neglecting excluded volume interactions and restricting interactions to harmonic potentials. Thus, the backbone is a Gaussian chain and the probabilistic loops are introduced by random harmonic interactions. Using a homogeneous model, where the looping probability $\mathcal{P}$ is constant along the contour of the chain, we have shown that $\langle R_n^2 \rangle$ only depends on the interaction matrix K [eq. (5.9)]. Each non-diagonal matrix entry, corresponding to one random loop, was chosen to be $-\kappa$ with probability $\mathcal{P}$ or 0 otherwise. The average over

the ensemble of loop configurations then turned out to be equivalent to averaging over the specific ensemble of random matrices with fixed $\mathcal{P}$. We found that random loops on all length scales explain the leveling-off observed in experiment, while restricting the loop sizes to only large loops (in the order of 50 - 100 Mb) or small loops (100 kb - 10 Mb) does not fit the data.

In contrast to the Random-Walk/Giant-Loop model [11] we do not assume fixed-size and regularly placed loops. Importantly, the average over the ensemble of different loop configurations turns out to be necessary to obtain the leveling-off observed in experiment. Its importance also becomes apparent when looking at higher-order moments (Fig. 5.5): Here the RWGL model would yield the same result as the random walk, $c_4^{\mathrm{RWGL}} = 5/3$, whereas the disorder averaging leads to larger fluctuations.

We have already shown ($\curvearrowright$ 4.5) that on short genomic distances a scaling with $\nu = 1/3$ fits the experimental data best compared to a self-avoiding and random walk. In this section, we did not focus further on the regime below 5 Mb, as this homogeneous model by design cannot explain differences in compaction of gene-rich and gene-poor regions (ridges and anti-ridges). The short distance regime will be covered by extending the model to include heterogeneity in chapter 7.

In the present study we have neglected the effect of excluded volume. On genomic distances above a few Mb the existence of a small number of loops can explain the leveling-off without the need for introducing excluded volume interactions. The role of excluded volume inside the cell nucleus on the spatial chromatin properties remains an open question, because of the impact of an enzyme called topoisomerase II, which is capable of cutting the DNA double strand in order to let another strand pass through it. This event might well give rise to a random-walk-like statistics, depending on the frequency of these events. In the next chapter, we investigate the effects of excluded volume using Molecular Dynamics simulations.



A schematic view on the three-dimensional organization of chromosomes inside the cell nucleus according to the Random Loop model.

Chapter 6

Molecular Dynamics simulations of the Random Loop Model

References

The results presented in this chapter are published in:

- J. Mateos-Langerak, M. Bohn *et al.*, **2009**. Spatially confined folding of chromatin in the interphase nucleus. *Proceedings of the National Academy of Sciences of the United States of America* 106 (10), 3812-3817.

Chapter Summary

Remarkably little is known about the higher-order folding motifs of the chromatin fiber inside the cell nucleus. Folding depends amongst others on local gene density and transcriptional activity and plays an important role in gene regulation. Strikingly, at fiber lengths above 5 to 10 Mb the measured mean square distance $\langle R^2 \rangle$ between any two points on the chromatin fiber is independent of polymer length. In chapter 5 the Random Loop model was introduced, explaining the confined folding of chromatin into a sub-space of the nucleus by introducing loops of a broad size range. Here, we extend the analytical model by incorporating excluded volume interactions. We study the impact of such interactions, which possibly have a major impact on the confined folding, by setting up Molecular Dynamics simulations to create sample conformations. We find that the quantitative behavior of the model, i.e. the folding into a confined sub-space of the nucleus, is not affected by excluded volume interactions. The results are in good agreement with experimental data from FISH measurements on human chromosome 1 and 11.

6.1 Introduction

The Random Loop (RL) model introduces two important features that have not been addressed by polymer models for chromatin up to now. First, it takes into account that intra-chromosomal interactions, i.e. loops, vary from cell to cell and therefore measurements are an average over an extended ensemble larger than the purely thermodynamic one. This ensemble is represented in the model by assigning a probability for looping, resulting in a huge number of different sets of loop attachment points (disorder average). Secondly, in contrast to the Random-Walk/Giant-Loop (↷ 3.2.2) and Multi-Loop-Subcompartment (↷ 3.2.3) model, it allows for a broad range of loop sizes. The assumption of fixed-sized loops in the Random-Walk/Giant-Loop model, for example, leads to a random walk behavior on a scale larger than the loop size, with the loops playing the role of "effective monomers".

In a first approach, the RL model assumed that the probability $\mathcal{P}$ for two monomers to interact is the same for any pair of monomers [117]. Such model allows for a semi-analytical calculation of the mean square distance (↷ 5.3.3) between two markers, which rapidly becomes independent of their genomic separation. The RL model ignored excluded volume interactions for reasons of mathematical tractability. This homogeneous Random Loop model without excluded volume already offers an explanation for the confined folding of chromatin in the interphase nucleus and, more importantly, yields large distance fluctuations similar to the experimental data. Evidently, an intrinsic property of the chromatin fiber inside the cell nucleus is that it assumes a state that cannot be described by a random walk or self-avoiding walk polymer model. The incorporation of looping disorder in the system, mimicking the cell-to-cell variation in loop configurations, effects such large fluctuations compared to a random walk model. Importantly, the model offers a framework for the connection between genome folding and function.

The Random Loop model presented so far is, despite its success in explaining large-scale data, lacking explanation and detail concerning the following points:

1. The assumption of a phantom chain used in chapter 5 could be contested. Although topoisomerase II might play an important role in resolving topological constraints,

surely, chromatin is matter and therefore occupies a certain volume in space. It is not clear, whether the leveling-off still shows up once excluded volume is introduced.

2. Experimental data (Fig. 2.3) shows different levels of compaction for gene-rich and transcriptional active regions (ridges) and gene-poor and transcriptional inactive regions (anti-ridges). Unavoidably, the RL model so far is unable to give an explanation for this behavior.

3. Although highlighting the importance of loops on genome organization, possible mechanisms of loop formation and their dynamics are not captured by the calculations in chapter 5.

In the course of this thesis, these short-comings will be integrated into the model, resulting in a unified picture of chromatin folding based on the simple principle of dynamic looping (↷ chapter 10). To test, which experimental observations can be explained with a minimal set of assumptions and parameters, in a first step, we keep the model as simple as possible and analyze the effect of excluded volume. Because this may have a major impact on the behavior of the model, we have analyzed how the predictions of the model will change if we lift the limitation of phantom chains. We ask whether the consideration of excluded volume changes the property of confined folding, i.e. whether the mean square distance $\langle R^2 \rangle$ does not become independent of contour length any more. In principle, one could expect such a breakdown of confinement. Phantom chains, as they were considered in section 5.3, in principle could collapse onto one single point. Real polymers do not have this possibility and at least occupy the self-volume of its constituents.

Excluded volume interactions require performing computer simulations, as analytical calculations become intractable. We use Molecular Dynamics simulations for two reasons: Firstly, the long-range looping interactions make Monte Carlo simulations ineffective. Secondly, a lot of tested software packages exist, like Amber [118], CHARMM [119], GROMACS [120] or ESPResSo [121], which allow implementing the particles and interactions in an efficient way.

Even more than for phantom chain calculations, simulations with excluded volume necessitate coarse-graining of the polymer. The existence of two averaging procedures – over the thermal ensemble as well as the ensemble of loops – require a huge amount of independent conformations. In this study, we use a chain length of $N = 300$; therefore statements should always be regarded as qualitative – not quantitative – ones.

In section 6.2 the setup of the Molecular Dynamics simulations is explained in detail, in a way that the reader should in principle be able to redo the simulations. Results are presented in section 6.3. Finally, the annealed ensemble (↷ 5.3.5) for a phantom chain is investigated by Molecular Dynamics simulations.

6.2 The Simulational Method

6.2.1 Molecular Dynamics

This section gives a short and general overview over Molecular Dynamics (MD) simulations, addressing the reader not so familiar with MD. A recommendable review of Molecular Dynamics can be found in the book of Rapaport [122].

Molecular Dynamics aims at predicting and verifying properties of assemblies of molecules both from a structural as well as a dynamical point of view. While the Monte Carlo approach [111] is suited finding the equilibrium statistical properties (the equilibrium

state) of a system, Molecular Dynamics yields the trajectory of the system in phase space to reach this equilibrium state. (Note: Monte Carlo methods in polymer physics where only local moves are applied, can also be mapped to dynamics [110]. This is not true for Monte Carlo in general). Consider a system of N particles, whose coordinates and momenta are given by $\mathbf{r}_i$ and $\mathbf{p}_i$ ($i = 1, \ldots, N$) respectively. Let $U(\{\mathbf{r}_i\})$ be the interaction potential the system is subjected to.

In principle, Molecular Dynamics is nothing more than a numerical integration of the classical Newtonian equations of motion

$$\frac{d}{dt}\mathbf{p}_i = \mathbf{F}_i \qquad \text{and} \qquad \frac{d}{dt}\mathbf{r}_i = \mathbf{p}_i/m_i$$

where $\mathbf{F}_i = -\nabla_i U(\{\mathbf{r}_i\})$ is the force acting on particle i derived from the potential U.

In numerical integration, time is propagated in small steps Δt. Given the momenta and impulses at time t, step-wise integration has to be performed to obtain the momenta and impulses at time $t + \Delta t$. One possibility of doing this is the *velocity Verlet algorithm* [122], which follows the scheme [123]

1. Calculate the momenta at the intermediate time $t + \frac{1}{2}\Delta t$
$$\mathbf{p}_i(t + \frac{1}{2}\Delta t) = \mathbf{p}_i(t) + \frac{1}{2}\mathbf{F}_i(t)\Delta t\,.$$
2. Integrate the particle coordinates at time $t + \Delta t$ using the momenta calculated in step 1,
$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \mathbf{p}_i(t + \frac{1}{2}\Delta t)/m_i\ \Delta t\,.$$
3. Calculate the forces $\mathbf{F}_i(t + \Delta t)$ from the interaction potential U using the updated coordinates $\mathbf{r}_i(t + \Delta t)$.
4. Calculate the momenta at time $t + \Delta t$ using the updated forces
$$\mathbf{p}_i(t + \Delta t) = \mathbf{p}_i(t + \frac{1}{2}\Delta t) + \mathbf{F}_i(t + \Delta t)\frac{1}{2}\Delta t\,.$$

Solving Newton's equations of motions results in a constant energy ensemble, i.e. the microcanonical or *NVE* ensemble. However, we are interested in a system at constant temperature, i.e. the canonical ensemble, which resembles the biological system more than the microcanonical. There are several possibilities to implement constant temperature in MD simulations. One method is to draw random velocities from a Maxwell-Boltzmann distribution from time to time, corresponding to an occasional random coupling with a thermal bath [124]. It is also possible to introduce an extra variable representing the thermal reservoir [123] into the dynamical equations.

For simulating chromosomes we make use of a pre-written software package developed at the Max-Plank-Institute in Mainz [121]. This package, ESPResSo, allows to easily set up and control the Molecular Dynamics process using the TCL scripting language.

6.2.2 Setting up the potentials

Recall that for the Random Loop Model without excluded volume and homogeneous looping probability $\mathcal{P}$ the Hamiltonian read

$$U = \frac{1}{2}\sum_{i<j}^{N} \kappa_{ij} \parallel \mathbf{r}_i - \mathbf{r}_j \parallel^2 = \frac{1}{4}\sum_{\substack{i,j=0 \\ j\neq i}}^{N} \kappa_{ij} \parallel \mathbf{r}_i - \mathbf{r}_j \parallel^2 \ .$$

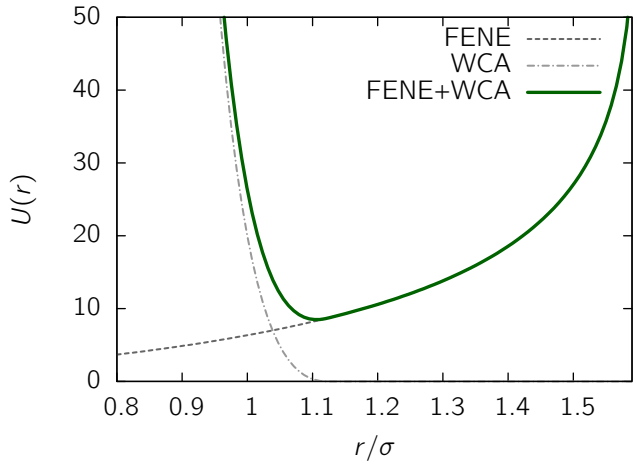


Figure 6.1: Combined FENE + WCA potential for the interaction between two neighboring beads. The WCA potential models the excluded volume interactions by a hard-core repulsive potential of radius σ. The FENE potential resembles a harmonic interaction on the short scale, but diverges at $R_0 = 1.6\sigma$ to ensure that no bond-crossings can occur during the simulation run.

We want to set up MD simulations for the Random Loop Model with excluded volume interactions, introducing as little changes as possible to the Hamiltonian. In the following we introduce the potentials used and explain the necessary modifications to the potential.

Excluded volume interactions

Excluded volume interactions are simulated by a so-called WCA (Weeks-Chandler-Andersen) potential [125], which was designed to model excluded volume interactions by a short-range repulsive force. It has been used in several other MD studies on polymers [126]. The WCA potential is basically a truncated and shifted Lennard-Jones potential with the following functional form,

$$U_{\mathrm{WCA}}(r) = \begin{cases} 4\epsilon \left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} + c_{shift} \right) & r < r_{cut} \\ 0 & r \geq r_{cut} \end{cases} \tag{6.1}$$

Here $r_{cut} = \sqrt[6]{2}$ and $c_{shift} = \frac{1}{4}$ are chosen such that the minimum of the potential is $U_{\mathrm{WCA}}(r_{min}) = 0$, the attractive part of the Lennard-Jones interaction being cut off. The WCA potential has two parameters ϵ and σ. σ defines the radius of the monomers' hard core. ϵ controls the energy penalty of another monomer penetrating this hard core.

Backbone potential

Simulating polymers with excluded volume interactions renders the use of a harmonic potential for the backbone potential as in eq. (6.2.2) impossible. A harmonic backbone potential in principle allows two adjacent beads to adopt a huge separation larger than their hard-core diameter σ, which would result in the possibility of bond crossings. To circumvent this problem, it is convenient to use the finitely extensible nonlinear elastic model (FENE) potential.

$$U_{\mathrm{FENE}}(r) = \begin{cases} -\frac{1}{2}\, k_{\mathrm{FENE}}\, R_0^2\, \log(1 - (r/R_0)^2) & r < R_0 \\ +\infty & r \geq R_0 \end{cases} \tag{6.2}$$

It is similar to the harmonic potential but grows to infinity at a predefined distance R_0. The pair potential between two beads (FENE + WCA) is displayed in Fig. 6.1.

Table 6.1: Exponential and integrated autocorrelation times for the energies of different conformations for a chain of length $N = 300$ and different probabilities $\mathcal{P}$.

$\mathcal{P}$	τ_{exp}	τ_{int}
0.0001	270.37	288.22
0.0005	300.9	313.66
0.001	340.05	427.33

Loop potential

The looping potential is chosen to be the same as in the original model, i.e. a Gaussian with Bernoulli-distributed random variables,

$$U_{\text{loops}} = \frac{1}{2} \sum_{\substack{i<j \\ |i-j|>1}}^{N} \kappa_{ij} \parallel \mathbf{x}_i - \mathbf{x}_j \parallel^2 .$$

Here, the parameters are the looping probability $\mathcal{P}$ and the interaction strength κ_{loops} (the κ_{ij} being either this value or zero).

Parameters

The following parameters are chosen for the simulation runs:

$R_0 = 1.6\sigma$	$\kappa_{loops} = 2.0$
$k_{\text{FENE}} = 10.0$	temperature $T = 1.0$
$\sigma = 1.0$	friction $\Gamma = 0.5$
$\epsilon = 20.0$	timestep $t = 0.006$

Special care is required for the relation between R_0 and σ. If R_0 is too large, other parts of the chain may pass through the gap between two monomers. Setting $R_0 = 1.6\sigma$ is a reasonable choice to prevent from such bond crossings [126].

6.2.3 Details on the implementation

As we are simulating quite dense systems, entanglement effects during the simulation can lead to very large correlation times and we cannot ensure that the whole configuration space is sampled in an unbiased way. As we are only interested in static (and not dynamic) properties of the polymers, we pursue the following approach: for each conformation we begin with a new random start configuration which is then equilibrated. Unfortunately, for our simulations we cannot just put the chain inside the simulation box, randomly select loops and then start the simulation. This is because a random start conformation might have large distances between loop-attachment points. The harmonic loop potential will accelerate these beads that much that the FENE bonds will break (meaning that a monomer is moved in one timestep that far, that it exceeds the maximum distance R_0 to its neighboring bead, resulting in an infinite potential). Therefore we have to slowly increase the strength of the potentials.

Averaging is required both over the conformational ensemble, the set of $\{\kappa_{ij}\}$ being fixed, and over the random variables $\{\kappa_{ij}\}$. To mimic the quenched situation, the values of the random variables (i.e. the loop attachment points) are chosen in advance of the actual MD simulations and kept fixed during the run. In our simulations we have generated

500 different sequences of random variables $\{\kappa_{ij}\}$ and for every sequence of $\{\kappa_{ij}\}$ we have equilibrated 300 conformations. Thus, in total we have generated 150 000 conformations for each parameter $\mathcal{P}$. The necessity for this large number of conformations arises from the averaging over the thermal ensemble as well as over the disorder, rendering the simulation of chain lengths larger than $N = 300$ impossible. The scheme of the algorithm is as follows:

1. Looping pairs (i, j) are chosen randomly according to the model. This means that each pair (i, j), $(|i - j| > 1)$ forms a loop ($\kappa_{ij} = \kappa_{loops}$) with probability $\mathcal{P}$.

2. 300 conformations are simulated according to the following steps:

   (a) *A random walk conformation with bond length 1.2 is put into the simulation box.* The bond length is chosen such that it is near the minimum of the combined FENE+WCA-potential (see Fig. 6.1).

   (b) *The excluded volume (WCA) is switched on slowly.* This is necessary to ensure the validity of the FENE bonds. Placing the monomers randomly might result in overlap of the particles. A zero distance would render the WCA potential singular; a small distance $r < \sigma$ might result in a high repulsive potential, causing a high velocity of the monomers. This results in the breakage of the FENE bond, whenever the velocity causes the particle to travel larger than the FENE cutoff distance R_0 in one timestep. The method of slowly increasing the excluded volume potential is an established method [121] and is done by introducing a cutoff U_{cut} to the WCA potential such that we effectively simulate excluded volume with the potential $U(r) = \min\{U_{\mathrm{WCA}}(r), U_{cut}\}$. The cutoff is increased slowly, finally reaching full excluded volume interactions.

   (c) *Looping potential is switched on slowly*, i.e. the interaction constant is increased up to its final value κ_{loops}. This prevents the monomers to speed up extremely causing breaks of the FENE bonds.

   (d) After excluded volume and looping interaction is fully present, the simulation is run for one million MD steps (this is more than 10 times the autocorrelation time of the radius of gyration). Then the simulation run is proceeded until the energy distribution approaches a Gaussian shape (χ^2-test, see below).

3. This sequence is repeated from step 1 until 500 sequences $\{\kappa_{ij}\}$ have been generated.

6.2.4 Equilibration of the system

To ensure that conformations are carefully equilibrated in step 2d) we pursue two tests: One is the determination of the autocorrelation time of the radius of gyration, the other is the distribution of energies during the simulation run.

The exponential as well as the integrated autocorrelation time of the radius of gyration are on the order of $\tau \sim 100\,000$ for all looping probabilities $\mathcal{P}$ studied here. It is commonly agreed that conformations have relaxated after 10τ MD steps. Therefore we perform at least one million MD steps before writing out the conformation.

As a second test on equilibration we analyze the distribution of the energies of the system. In thermodynamic equilibrium, they should obey a Gaussian distribution. If the system is still in the progress of drifting towards equilibrium, another distribution should become apparent. After one million MD steps are performed, the current energy of the system is written out every 500 MD steps. The correlations between energies in the system

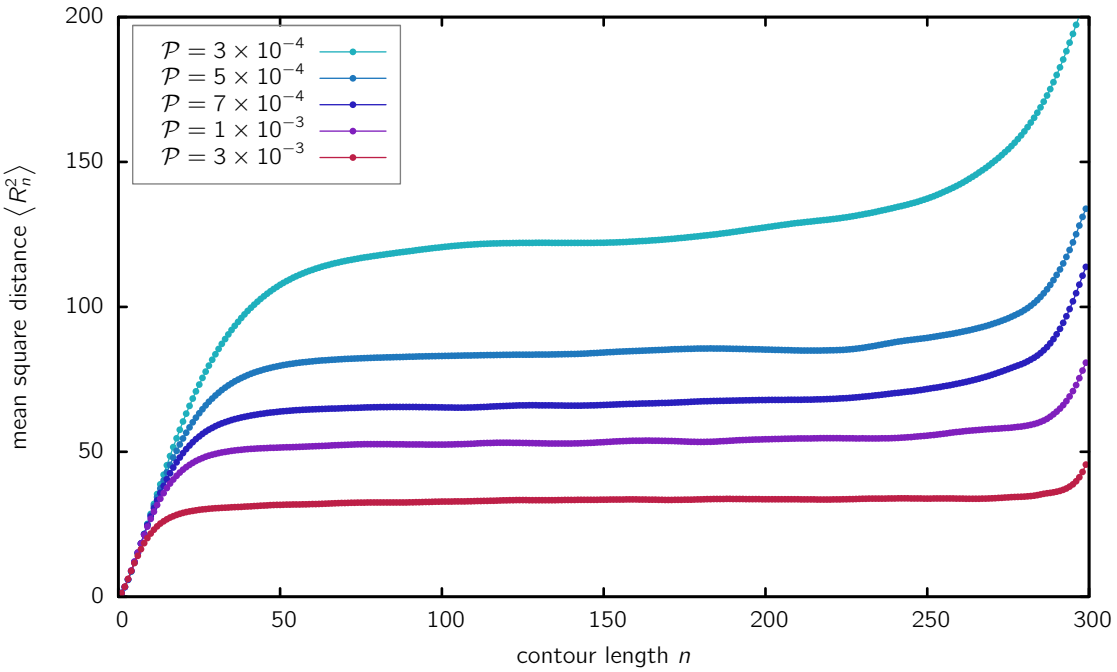


Figure 6.2: Molecular Dynamics simulations of a polymer with randomly positioned loops. The relationship between the mean square distance $\langle R_n^2 \rangle$ between two monomers and their contour length n is shown for different values of $\mathcal{P}$. $\mathcal{P}$ denotes the probability that a pair of monomers interacts. Looping probabilities used induce an average number of loops from 13 ($\mathcal{P} = 3 \times 10^{-4}$) to 133 ($\mathcal{P} = 3 \times 10^{-3}$) per chain. The chain length is $N = 300$ monomers. The increase in mean square distance at $N_m \gtrsim 250$ is due an increased freedom of the chain ends.

drop rapidly, the exponential autocorrelation time being about 200-300 MD integration steps (see table 6.1). After having sampled 200 independent energy values, a χ^2-test with five degrees of freedom is performed on a Gaussian distribution. The hypothesis of a Gaussian distribution is accepted if the χ^2-value is below 3.0, giving a high confidence that the distribution is indeed Gaussian. The conformation is written out once a Gaussian distribution is reached.

6.3 Results

We have used Molecular Dynamics (MD) simulations to obtain chain conformations and to introduce excluded volume interactions in the model. As two averaging processes have to be performed, i.e. over the thermal disorder and over the ensemble of loop configurations, simulations are very time-consuming. Since here we are only interested in large-scale behavior, a coarse-graining approach can be used. In our simulations we equilibrate polymers of length $N = 300$. Figure 6.2 shows the results of simulations for different looping probabilities $\mathcal{P}$. In contrast to classical polymer models, the mean square distance becomes independent of the contour length at intermediate length scales, resulting in a spatially confined polymer structure. Interestingly, already a small number of loops results in an almost complete independency of the mean square distance of the genomic distance, without any additional assumptions. It is stressed that loops on all length scales are necessary to make the spatial distance independent of contour length (see also section 5.3.4 and

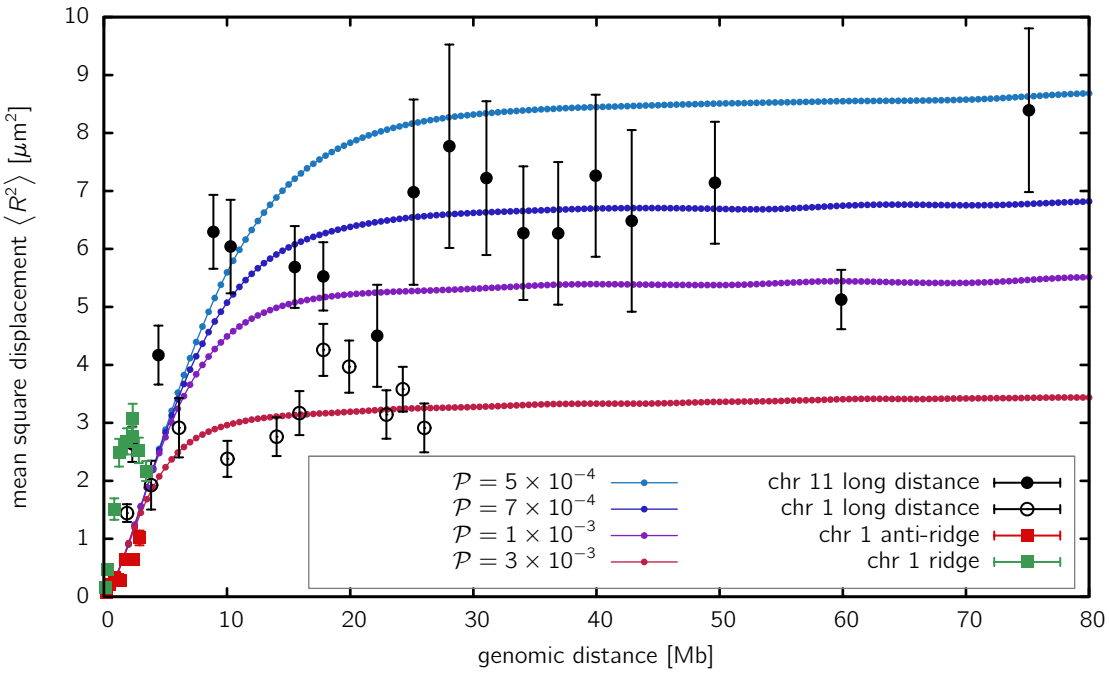


Figure 6.3: Comparison of the mean square distance between two markers in relation to genomic separation g between the model and experimental data. Simulations have been performed using Molecular Dynamics for various looping probabilities $\mathcal{P}$ (range from 5×10^{-4} to 3×10^{-3}). The polymer chain length is $N = 300$ monomers and a coarse-grained monomer is equivalent to 500 kb. Experimental data shows FISH distance measurements on human chromosome 1 and 11 [12]. At this scaling the RL model correctly predicts the leveling off at genomic distances above about 10 Mb.

Ref. [117]). Looping probabilities $\mathcal{P}$ in Fig. 6.2 range from 3×10^{-4} to 3×10^{-3}. A looping probability of $\mathcal{P}$ implies and average of $\mathcal{P}(N-1)(N-2)/2$ loops per conformation; shown results thus correspond to 13 up to 133 loops per $N = 300$ polymer. As expected, the plateau value of $\langle R^2 \rangle$ rapidly decreases, as the number of loops increases and therefore the polymer becomes more compact. For $\mathcal{P}$ smaller than 10^{-4} leveling-off becomes less pronounced, becoming a normal SAW model as P approaches zero. Notably, qualitatively the same behavior is observed for the RL model ignoring excluded volume interactions [117]. We therefore conclude that at bigger length scales excluded volume interactions contribute only to a limited extend to the behavior of the RL model.

The RL model proposes that large-scale chromatin folding is driven by chromatin looping. The prediction of a leveling-off in the mean square distance is in agreement with the experimental data. How can we bring theory and experiment together? The simulations use a polymer with a length $N = 300$. By mapping a coarse-grained monomer to 500 kb chromatin we obtain a chain of an effective length of 150 Mb, i.e. the size range of a human chromosome. In the model the mean square distance is a complex function of the chain length N, separation between monomers N_m and looping probability P: $\langle R^2 \rangle = f_N(N_m, \mathcal{P})$. In this context the single variable parameter is $\mathcal{P}$, as N is fixed to 300. To compare our simulation results to the experimental data we have to introduce a scaling factor for the $\langle R^2 \rangle$-axis. This factor is somewhat arbitrary and on this level of coarsening strongly depends on monomer geometry and does not reflect biological parameters in a simple manner [117]. In Fig. 6.3 we have scaled the results of the simulations to

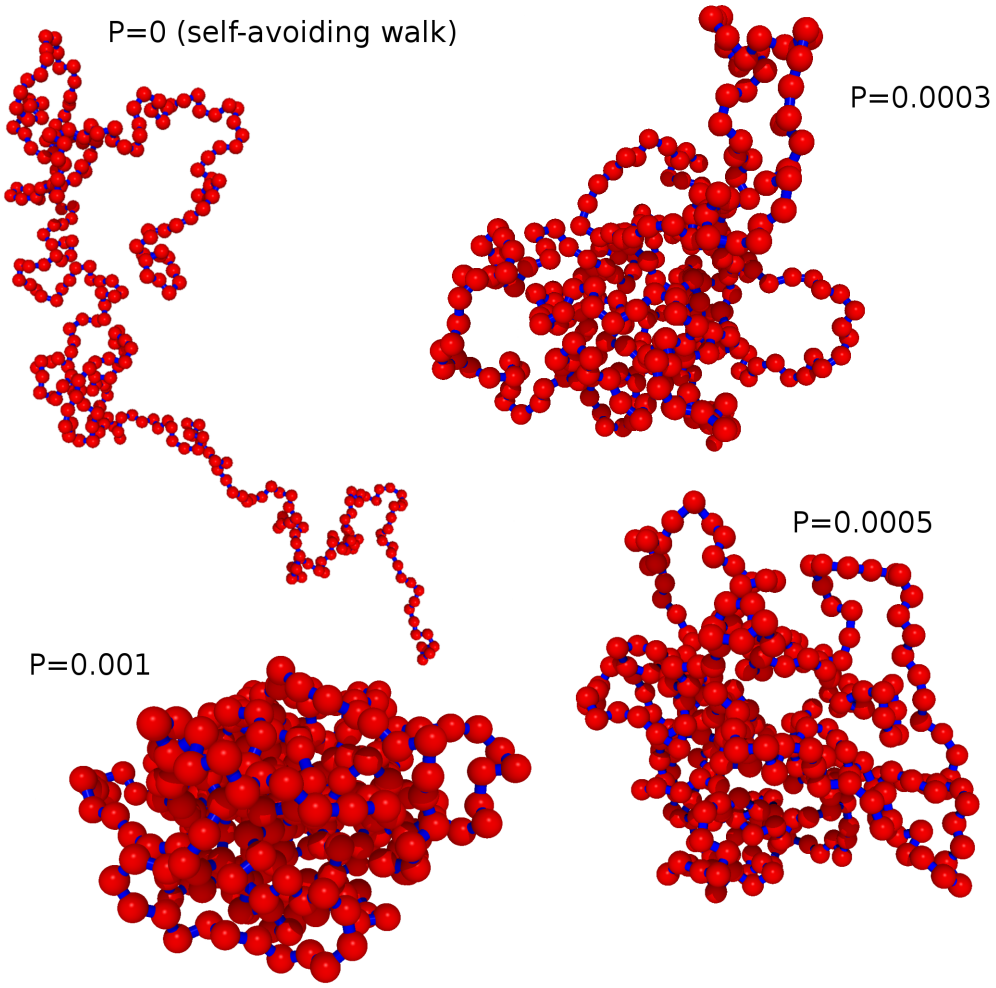


Figure 6.4: Example conformations of Random Loop model polymers obtained by Molecular Dynamics simulations. Increasing the $\mathcal{P}$-value results in a strong compaction of the polymer. The self-avoiding walk polymer ($\mathcal{P} = 0$) has a very open structure. Further conformations are shown for $\mathcal{P} = 0.0003$, $\mathcal{P} = 0.0005$ and $\mathcal{P} = 0.001$. The number of loops is 15, 24 and 40 respectively in case of this $N = 300$ polymers.

the experimental data, using 320 nm per coarse-grained monomer. This number has been determined such that the model fits to the plateau level of the experimental data. Figure 6.3 shows that the RL model is able to qualitatively describe the large-scale genomic distance data quite well. This is remarkable because we do not include information about the positions along the chromatin fiber where loops are actually formed.

Example conformations for different looping probabilities are visualized in Fig. 6.4. Obviously, introducing loops results in a strong compaction of the polymer. Note that due to coarse-graining, each monomer represents a folded 500-kb-segment of chromatin. As loops probably persist on the short scale below the coarse-graining length, resulting in a compacted state of chromatin, the spherical geometry of this coarse-grained model is justified.

$\mathcal{P}$	τ_{exp}	τ_{int}
0.007	15032	15982
0.01	17875	19231
0.05	17730	17494

Table 6.2: Exponential and integrated autocorrelation times τ_{exp} and τ_{int} for the squared radius of gyration R_g^2. The underlying system is a chain of length $N = 50$ in the annealed ensemble of the Random Loop model, i.e. a chain subjected to the potential given in eq. (6.3).

6.4 Simulations of the annealed ensemble

Up to now, both the calculations of the mean square distance for chains without excluded volume (↷ 5.3.3) and the simulations with excluded volume (↷ 6.3) were done in the quenched ensemble. For the annealed ensemble, it was shown in section 5.3.5 that the average over the set of random variables $\{\kappa_{ij}\}$ can be carried out before evaluating the configurational integrals, yielding an effective potential for the chain, given by

$$U_{\text{eff}}(\mathbf{r}) = \frac{1}{2}\kappa \sum_{i=0}^{N-1} r_{i,i+1}^2 - \sum_{i<j-1} \log\left[1 + \mathcal{P}\left(\exp^{-\frac{1}{2}\hat{\kappa}\|\mathbf{r}_i - \mathbf{r}_j\|^2} - 1\right)\right] . \tag{6.3}$$

However, an analytical calculation of the partition sum with this disorder-averaged effective potential is not possible. Here we perform Molecular Dynamics simulations with particles subjected to this effective potential U_{eff}. The ESPResSo software package [121] can be easily extended to include other forces. To allow a comparison between the annealed and the quenched average, the effective potential U_{eff} has been implemented in form of TCL script. The interaction constant is chosen to be $\kappa = \hat{\kappa} = 1$ both for the interactions between adjacent beads and for the looping interactions. Note that excluded volume interactions are not taken into account here. The chain length studied is $N = 50$.

The autocorrelation times were calculated for the squared radius of gyration $R_g^2(t)$ for different values of the looping probability $\mathcal{P}$. Both the integrated as well as the exponential autocorrelation times are listed in table 6.2. The relaxation times are below 20 000 MD integration steps using a time interval of $\Delta t = 0.005$. Independent conformations are obtained by writing out samples during the simulation run every 100 000 MD integration steps.

We have evaluated the mean square distance between two beads in relation to their contour length n for different looping probabilities ranging from $\mathcal{P} = 0.007$ up to $\mathcal{P} = 0.1$.

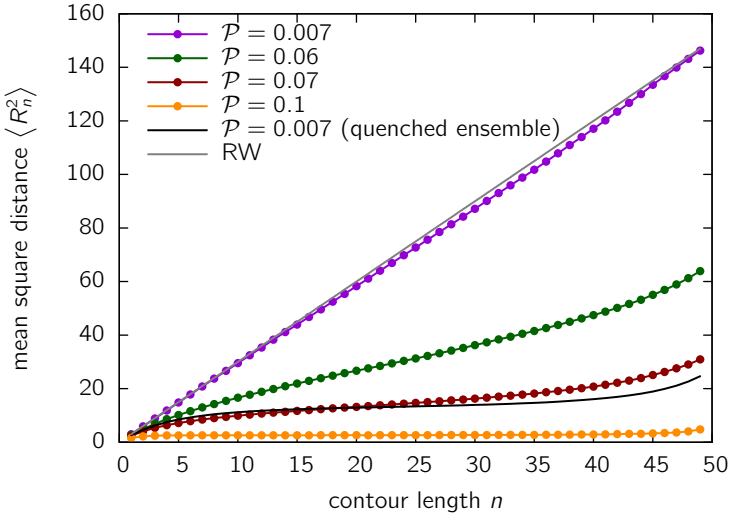


Figure 6.5: Mean square distance $\langle R_n^2 \rangle$ vs. contour length n for the homogeneous RL model without excluded volume in the annealed ensemble. Shown is the data from MD simulations for a chain of length $N = 50$ and different looping probabilities $\mathcal{P}$. Although a confined folding is observed in the annealed ensemble, the effect is markedly weaker compared to the quenched ensemble (compare the results for $\mathcal{P} = 0.007$).

The results are shown in Figure 6.5. A polymer with looping probability of $\mathcal{P} = 0.007$, which would correspond to an average of 8 loops per chain, displays only slight deviations from a random-walk behavior (grey line) in the annealed ensemble. For comparison we have added the results for the quenched average calculations for this $\mathcal{P}$-value (black line), which already shows a pronounced leveling-off in the mean square distance. Only at a 10-fold higher looping probability of $\mathcal{P} = 0.07$ (corresponding to an average of 80 loops per chain), the level of compaction is similar to that of the quenched ensemble at a $\mathcal{P}$-value of 0.007. Thus, the interaction in the annealed ensemble is much weaker than in the quenched ensemble, and the leveling-off is only obtained for loop numbers which are practically impossible to realize in a cell due to the topological constraints. The reason for this behavior is that the probabilistic interactions in the annealed ensemble effectively boil down to a mean-field attraction between all monomers of the chain, due to the short lifetime of the loops, which has to be much smaller than the equilibration time of the chain for the annealed ensemble to be a valid approximation.

As already mentioned, the annealed approximation is probably not valid for the system of the cell nucleus. Here, loops are tightly connected to genome function and transcriptional regulation. Thus, loops have to persist a while for the nucleus being able to transcribe the genes.

6.5 Conclusions

In this chapter, we lifted one of the major limitations of the homogeneous Random Loop model (↷ chapter 5), i.e. the negligence of excluded volume interactions. This random loop (RL) model predicts that loop formation is the major driving force for chromatin compaction [117]. The RL model assumes that the measured observables, e.g. the mean square distance between two markers, are derived from an ensemble of loop configurations formed by interactions between different parts of the polymer with a certain probability $\mathcal{P}$. A major characteristic of the RL model without excluded volume interactions is that the mean square distance $\langle R^2 \rangle$ becomes independent of the contour length at longer distances. While the RL model without excluded volume already explains the major characteristics of experimental FISH measurements (see Figs. 5.4 and 5.5), the negligence of excluded volume might have a drastic effect on the folding into a confined space. Here we extended the original RL model beyond the limitations of its original formulation (↷ 5.3). We performed extensive MD simulations to establish the effect of excluded volume on the behavior of the RL model. It turns out that the introduction of excluded volume does not alter the model's main properties.

Using Molecular Dynamics simulations, we derived results for the homogeneous Random Loop model in the annealed ensemble (↷ 5.3.2). While in this ensemble, a folding into a confined sub-space of the nucleus can still be observed, the required looping interaction is much stronger compared to the quenched ensemble.

Chapter 7

A unified view of short and large scale folding

The heterogeneous Random Loop Model

References

The results presented in this chapter are published in:

- J. Mateos-Langerak, M. Bohn *et al.*, **2009**. Spatially confined folding of chromatin in the interphase nucleus. *Proceedings of the National Academy of Sciences of the United States of America* 106 (10), 3812-3817.

Chapter Summary

The folding of chromatin on the scale above 10 Mb, where experiments measuring the mean square distance between two FISH markers reveal that chromosomes fold into a confined sub-space of the nucleus, has been explained by the Random Loop model in recent chapters. However, genome function in higher eukaryotes involves major changes in the spatial organization of the chromatin fiber, resulting in different levels of compaction dependent on gene activity on the scale below 10 Mb. In this regime, the physical distance between two markers increases monotonously with genomic separation. This increase has been fitted to a power-law behavior $\langle R^2 \rangle \sim N^{2\nu}$ in the past. The obtained scaling exponents ν displayed a huge variety strongly dependent on the chromatin region studied. In this chapter, the aspect of different levels of chromatin compaction in different subchromosomal domains on the scale below 10 Mb is incorporated into the Random Loop model. We show that effective scaling exponents ν strongly depend on the local looping probability, suggesting a tight connection between the local state of compaction and loop formation. Simulations of human chromosome 11 with different local looping probabilities for gene-rich and gene-poor regions reveal a surprising agreement with experimental data. Thus, the heterogeneous Random Loop model creates a quantitative and predictive framework for the three-dimensional organization of the chromatin fiber, spanning different scales of length and different degrees of compaction.

7.1 Introduction

The chromatin fiber inside the interphase nucleus of higher eukaryotes is folded and compacted on several length scales. Evidently, neither DNA nor chromatin are folded and arranged homogeneously along the contour of the fiber. Different bases (A, T, G and C) are distributed in a non-random sequence along the fiber, defining genes and non-coding sequences. The positioning of the nucleosomes along the fiber follows a non-regular pattern [127]. Proteins called "chromatin remodeling complexes" induce changes in the structure of the nucleosomes such that DNA temporarily becomes less tightly packed, making it accessible for repair units or gene expression. Chromatin remodellers are possibly attracted by modifications in the histone tails, which can further stabilize or destabilize chromatin structure locally ($\curvearrowright$ 2.3.2). On the length scale of the 10 nm fiber, it is expected that nucleosomal skips as well as depletion of linker histone H1 strongly influences compaction and decompaction [5]. Genes are not distributed randomly along the one-dimensional genome, rather gene deserts containing no or little gene content and transcriptional activity can be observed [128]. The Human Transcriptome Map ($\curvearrowright$ 2.3.1), mapping gene content and transcriptional activity, distinguishes between regions of increased gene density and transcriptional activity (ridges) and gene-poor regions with little transcriptional activity (anti-ridges) [13]. This one-dimensional sequence of genes and non-coding elements itself influences the three-dimensional structure of chromatin. Gene deserts or anti-ridges show distinct patterns of organization, trying to cluster, while the intervening gene rich regions have to loop out [129]. Shape and size parameters of ridges and anti-ridges were shown to be markedly different [6]. Ridges are in tendency located more at the nuclear periphery while anti-ridges are positioned more towards the nuclear interior [6, 14]. Finally, FISH

distance measurements (Figure 2.3 and Ref. [12]) reveal that ridges show different patterns of compaction than anti-ridges, resulting in a more compact state, i.e. the mean square distance $\langle R^2 \rangle$ increases more slowly with increasing genomic separation g.

It is partly due to the different local folding state of chromatin, that several studies show pronounced differences in compaction on the scale below 10 Mb. For example, different scaling exponents for the mean square distance $\langle R^2 \rangle \sim N^{2\nu}$ have been determined and models for the overall chromatin folding have been proposed based on these local results. Based on data from Yokota *et al.* [27], it was shown that folding on the scale below 10 Mb is best explained by a random walk model, the scaling exponent being $\nu = 0.5$ [29]. Findings for the scale between 10 Mb to 20 Mb range from a random walk type of behavior, which has been explained by the formation of giant loops [11] to a scaling with $\nu = 0.32$ for which the multi-loop subcompartment model has provided an explanation [28, 29]. Recently, Jhunjhunwala *et al.* [109] measured exponents of $\nu \approx 0.1 - 0.2$, a value being much smaller than the globular state value (↷ chapter 4).

A unified view of chromatin folding explaining these different exponents is still missing. The Random Loop (RL) Model presented in chapter 5 does not comprise heterogeneity of the fiber, therefore cannot explain local differences in compaction. Any two chromatin segments were assigned the same probability to interact with each other, i.e. the looping probability has been chosen translationally invariant. While this model correctly predicted the folding state above 10 Mb, where a leveling-off in the mean square distance $\langle R^2 \rangle \sim \mathcal{O}(1)$ between two chromatin segments was encountered, the different folding of ridges and anti-ridges on the scale below 10 Mb still lacks a valid model description.

In this chapter the Random Loop model is extended such that it predicts the folding state on the scale of 100 kb up to the complete chromosome arm on the basis of random loops. The major modification to the homogeneous RL model is that we allow for different local looping probabilities related to the transcriptional state of local chromatin regions. We find that these different local looping probabilities can lead to different local states of compaction. This heterogeneous model is compared to experimental data by a direct mapping of ridge and anti-ridge regions of chromosome 11 according to the Human Transcriptome Map and by assigning different local looping probabilities to these regions. We show that the model correctly predicts distance measurements between three reference BAC probes (for ridges, anti-ridges and long distances) and a second chromatin segment on chromosome 11.

7.2 First evidence for different local looping probabilities

When measuring the mean square distance $\langle R^2 \rangle$ between two loci separated by genomic distance g, the scaling exponents found vary between $\nu = 0.1$ [109] and $\nu = 0.5$ [27] on the scale below 10 Mb. Amazingly, the compaction found in the murine Igh locus by Jhunjhunwala and co-workers is far below the predictions of the globular state model ($\nu = 1/3$) which already exhibits a very compact organization. These findings have been explained by different polymer models with the choice of the models depending on which fits the experimentally determined scaling exponent ν best [11, 26, 28]. However, a unified model for chromatin folding is still missing.

So far, the short scale behavior of the homogeneous Random Loop model proposed in section 5.3 has not been analyzed. Although a scaling exponent of $\nu = 1/3$ fits the experimental data from Fig. 2.3 best on the scale below 2 Mb (see Figure 4.18), one should be aware of the fact that the globular state model is only valid for end-to-end distances of

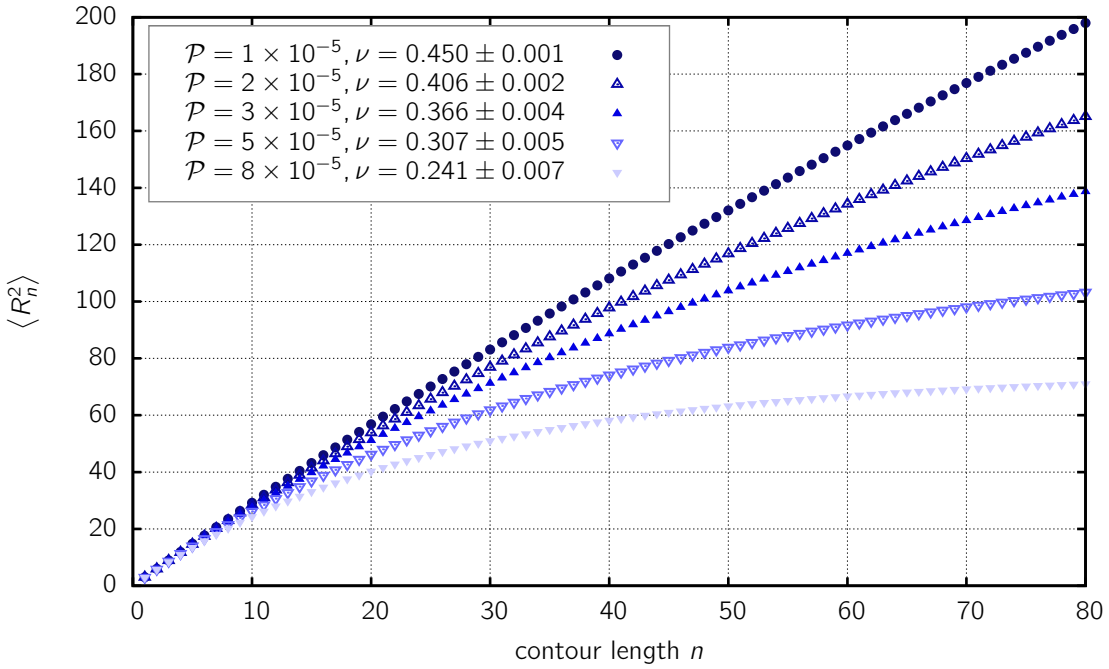


Figure 7.1: The relationship between the mean square distance between two monomers and their contour distance is shown for different values of the looping probability P. In the short length scale regime the mean square distance follows effectively a scaling law where $\langle R^2 \rangle \sim N^{2\nu}$. The scaling exponent ν varies over a broad range of values, depending on the looping probability P. The figure shows data from the homogeneous model without excluded volume and for a chain length of $N = 600$.

a polymer, whereas here we deal with intra-chain distances (↷ 4.5). Fitting the RL model to the experimental data shows that such a value of ν can only be valid in a narrow range of genomic distances before a plateau level is reached. Thus, the interpretation of the data in terms of one of the classical polymer models would be an extreme oversimplification.

The behavior of the homogeneous Random Loop model for short genomic separations is displayed in Figure 7.1. Calculations are based on a chain of length $N = 600$, the data is shown up to genomic separations of $n = 80$ segments in order to keep away from the region where a plateau level is reached. As excluded volume interactions do not alter the overall properties of the model (↷ chapter 6), calculations are based on the model from section 5.3 without excluded volume, as much better statistics can be gained here. Clearly, we find that the compaction is strongly dependent on the looping probability $\mathcal{P}$, which has been chosen between $\mathcal{P} = 1 \cdot 10^{-5}$ and $\mathcal{P} = 8 \cdot 10^{-5}$ corresponding to 4 to 28 loops per coarse-grained polymer. To see whether different scaling exponents found in experiments on the short scale below a few Mb can be explained by different looping probabilities, the mean square distance $\langle R_n^2 \rangle$ has been fitted to the scaling law $\langle R_n^2 \rangle = an^{2\nu}$ for different looping probabilities $\mathcal{P}$ for contour lengths $n < 80$, i.e. below the scale of the plateau level. Interestingly, the scaling exponent ν varies strongly with looping probability $\mathcal{P}$. In fact, by adjusting the $\mathcal{P}$-value, nearly all exponents ν below the random walk value of $\nu = 0.5$ can be obtained. Effective scaling exponents ν can adopt values far below the one of the globular state $\nu = 1/3$, rendering a very compact organization possible in agreement with measurements in the murine Igh locus [109].

These findings indicate that the looping probability might be a cellular mechanism for

switching between different states of compaction. Different scaling exponents ν determined in experiments thus are merely representing more than a cross-over behavior on short genomic separations, where the plateau level is not yet reached.

7.3 Integration of short and long length scale data by the RL model

7.3.1 The Heterogeneous Random Loop model

How do we integrate different local looping probabilities into the Random Loop model? In principal, the RL model offers a lot of choice concerning the random variables κ_{ij}. Recall that the potential function of the general Random Loop model ($\curvearrowright$ 5.2) was given by

$$U = U_{\text{Gaussian}} + \frac{1}{2} \sum_{\substack{i<j \\ |i-j|>1}}^{N} \kappa_{ij} \parallel \mathbf{x}_i - \mathbf{x}_j \parallel^2 .$$

The first potential term represents the Gaussian backbone potential [eq. (3.4)], while the second term is responsible for loop formation. As shown in section 5.3, the thermal average can be calculated analytically only depending on the random variables κ_{ij}

$$\left\langle r_{IJ}^2 \right\rangle_{\text{thermal}} = \frac{3}{2\Gamma} = 3(\sigma_{JJ} + \sigma_{II} - 2\sigma_{IJ}) , \tag{7.1}$$

where

$$\sigma^{-1} = K = \begin{pmatrix} \sum_{\substack{j=0 \\ j\neq 1}}^{N} \kappa_{1j} & -\kappa_{12} & \dots & -\kappa_{1N} \\ -\kappa_{21} & \sum_{\substack{j=0 \\ j\neq 2}}^{N} \kappa_{2j} & \dots & -\kappa_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ -\kappa_{N1} & -\kappa_{N2} & \dots & \sum_{\substack{j=0 \\ j\neq N}}^{N} \kappa_{Nj} \end{pmatrix} \tag{7.2}$$

The spring constants κ_{ij} have been chosen to be Bernoulli distributed random variables with probability $\mathcal{P}$ in the homogeneous RL model,

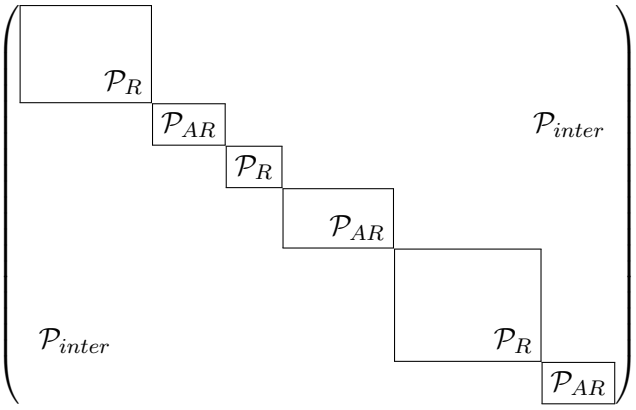
$$\kappa_{ij} = \begin{cases} 1, & \text{with probability } \mathcal{P} \\ 0, & \text{with probability } 1 - \mathcal{P} \end{cases} \tag{7.3}$$

Here, we extend the RL model to incorporate local differences in $\mathcal{P}$-values, thus making the polymer heterogeneous. Lacking experimental data on looping probabilities, in a first approach, we assign different looping probabilities for different regions based on the distribution of ridges and anti-ridges in the Human Transcriptome Map as shown in Fig. 2.3A. We divide the polymer into ridge and anti-ridge regions corresponding to the Human Transcriptome Map and define three different looping probabilities, i.e. $\mathcal{P}_R$, defining loop formation in ridge regions, $\mathcal{P}_{AR}$ for anti-ridges and $\mathcal{P}_{inter}$ for the interaction between such regions.

The interaction random variables κ_{ij} are thus drawn from a Bernoulli distribution with a probability $\mathcal{P}(i,j)$, now depending on i and j. Here

$$\mathcal{P}(i,j) = \begin{cases} \mathcal{P}_R & \text{if } i \text{ and } j \text{ are inside the same ridge domain} \\ \mathcal{P}_{ar} & \text{if } i \text{ and } j \text{ are inside the same anti-ridge domain} \\ \mathcal{P}_{inter} & \text{if } i \text{ and } j \text{ are not inside the same domain} \end{cases} \tag{7.4}$$

This results in a block structure of the chromosome and the matrix K:



(7.5)

Along the diagonal, blocks are distributed according to the distribution of ridges and anti-ridges along the one-dimensional Transcriptome. Inside each of the diagonal blocks, interactions are distributed according to a Bernoulli distribution with the same probability $\mathcal{P}_R$ or $\mathcal{P}_{AR}$ (depending on the property of being ridge or anti-ridge). Outside these blocks, elements interact with the looping probability $\mathcal{P}_{inter}$.

7.3.2 Results for chromosome 11

Calculations are performed for chromosome 11 as distance measurements (Figure 2.3) span the whole q-arm comprising about 80 Mb. To map chromosome 11 on our polymer model, we use a coarse-grained segment length of 75 kb. The q-arm of chromosome 11 is thus represented by $N = 1120$ polymer segments. The Human Transcriptome Map [13] data of chromosome 11 reveals distinct regions where genes cluster. These ridge as well as the anti-ridge regions of the entire chromosome arm are mapped onto the chain of length N as depicted by the Human Transcriptome Map. The mapping of ridge and anti-ridge regions to the model polymer is shown in table 7.1 for chromosome 11. Clearly, the figure in table 7.1 indicates that there are regions where it is not clear whether to assign it to a ridge or anti-ridge. In order not to incorporate a fourth parameter, we crudely assign these regions either the ridge or anti-ridge property. Especially the large region ranging from genomic position 110 Mb to 127 Mb is mapped onto an anti-ridge region.

After mapping chromosome 11 to our model polymer, the mean square distance between two chromatin segments can be calculated. Calculations are done averaging over different realizations of the random matrix K, i.e. different configurations of loop attachment points. The mean square distance is evaluated differently than in the homogeneous case. For the latter, the mean square distance for a certain contour length n was averaged over all different pairs of beads (i, j) with $n = |i - j|$,

$$\left\langle R_n^2 \right\rangle = \frac{1}{|\{K\}|} \frac{1}{N-n} \sum_{K \in \{K\}} \sum_{i=0}^{i=N-n-1} \left\langle r_{i,i+n}^2 \right\rangle_{\text{thermal}} \tag{7.6}$$

where $\{K\}$ is a sample of random matrices randomly drawn out of the underlying ensemble.

As the polymer is no longer homogeneous, distances have to be measured by fixing the reference segment i to the monomer which corresponds to the position of the reference BAC probe used in experiments (marked by the starting point of the arrows at the top of the figure in table 7.1). Thus, the mean square distance $\langle R_n^2 \rangle$ is calculated by

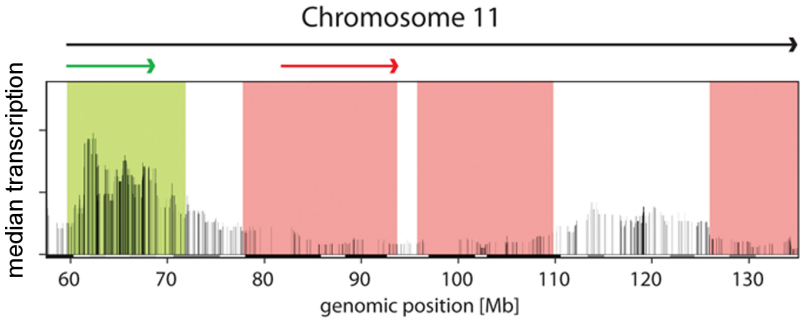
$$\left\langle R_n^2 \right\rangle = \frac{1}{|\{K\}|} \sum_{K \in \{K\}} \left\langle r_{i,i+n}^2 \right\rangle_{\text{thermal}} \tag{7.7}$$

Table 7.1: Mapping of the q-arm of chromosome 11 to the heterogeneous Random Loop model. The q-arm is mapped onto a model polymer with segment length 75 kb, corresponding to a chain of length $N = 1120$. To keep calculations efficient, not the entire chromosome is mapped, but only a region slightly larger than the length scale of the measurements, i.e. one chromosome arm. The Human Transcriptome Map of the q-arm of chromosome 11 is shown at the bottom. Red areas indicate anti-ridge regions, green areas ridge regions. The starting points of the arrows above the maps indicate the positions of the reference FISH probes. The arrowheads marks the locations of the FISH BAC probe that has the largest genomic distance to the reference probe. All physical distances are determined with respect to this reference probe. The table gives an overview over the regions and the corresponding polymer segments in the calculations. The positions along the contour of the chromosome are given according to the nomenclature in the excerpt from the Human Transcriptome Map shown at the bottom. The q-arm is subdivided into gene-poor and transcriptional inactive regions (anti-ridges) as well as gene-rich and transcriptional active regions (ridges). In order not to introduce too many parameters, the regions marked white – neither showing clear ridge or anti-ridge behavior – are roughly mapped onto either ridges or anti-ridges. The large region ranging from chromosomal position position 110 Mb to 127 Mb is considered to have the same looping probability as anti-ridges. Distances between polymer segments are measured from the same reference points where the FISH BAC probes are positioned.

Mapping of Chromosome 11

region on chromosome	beads in polymer model	type of region
50 Mb – 55 Mb	0 – 67	• anti-ridge
55 Mb – 75 Mb	68 – 333	• ridge
75 Mb – 110 Mb	334 – 800	• anti-ridge
110 Mb – 127 Mb	801 – 1026	◦ (anti-ridge)
127 Mb – 134 Mb	1027 – 1120	• anti-ridge

reference position ridge: 59.5 Mb (monomer number 126)
reference position anti-ridge: 81.6 Mb (monomer number 421)
reference position long distance: 59.5 Mb (monomer number 126)



The reference segment i, of course, is different for the measurements in the ridge and anti-ridge region. This is depicted in table 7.1. The reference BAC probe for the ridge region as well as the long distance measurements is centered at chromosomal position 59.5 Mb, corresponding to monomer number $i = 126$ using our mapping, which starts at

chromosomal position 50 Mb. The reference BAC probe for measurements in the anti-ridge region is at chromosomal position 81.6 Mb, corresponding to $i = 421$.

We have performed calculations of the mean square distance for various parameter sets. It turned out that the data is described best by looping probabilities being $\mathcal{P}_R = 3\times10^{-5}, \mathcal{P}_{AR} = 7\times10^{-5}$ and $\mathcal{P}_{inter} = 1\times10^{-5}$. The average physical distance between the reference FISH probe and a second marker inside ridge and anti-ridge regions is displayed in Figure 7.2. Interestingly, the RL model with these looping probabilities describes the folding of the ridge and anti-ridge data of chromosome 11 remarkably well. Importantly, the observed large scale behavior, i.e. the leveling-off in the mean square distance, is still a prominent feature of the heterogeneous model. Distance measurements on the scale of the entire chromosomal arm are depicted in Figure 7.3 both for experimental data as well as the heterogeneous RL model. This model captures the fluctuations on the large scale even better than the homogeneous model presented in section 5.3.

7.4 Conclusions

While the homogeneous Random Loop model explained basic principles of large-scale chromatin folding, i.e. the folding into a confined sub-space of the nucleus and the large cell-to-cell variation, a detailed description of folding principles on the scale below 10

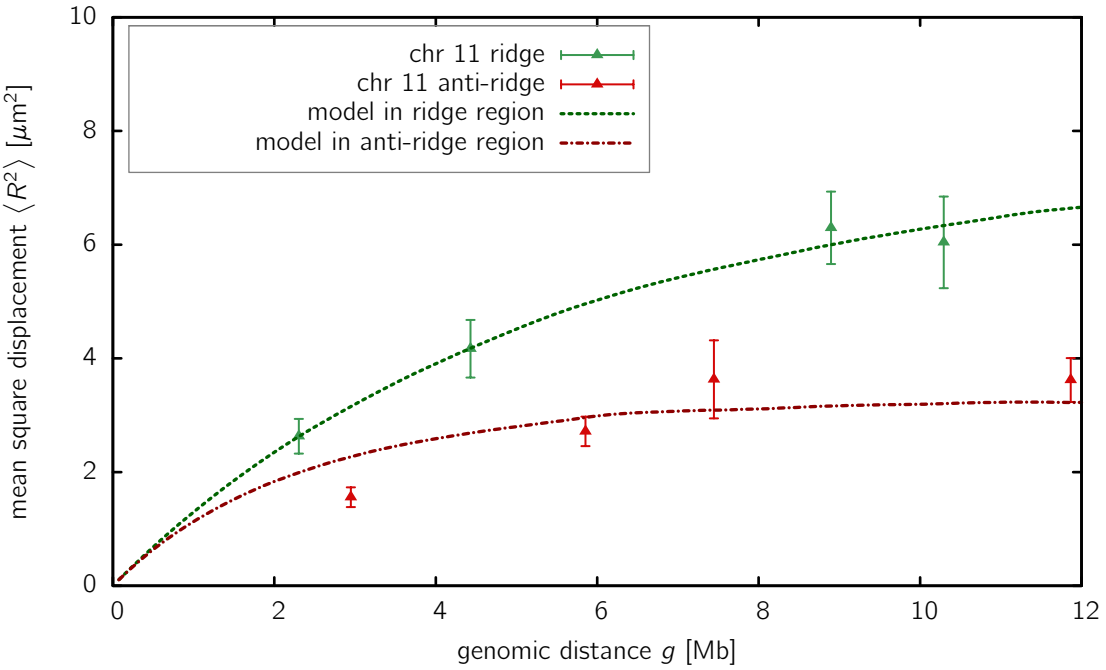


Figure 7.2: Comparison of FISH distance measurements on human chromosome 11 [12] to the heterogeneous Random Loop model on the short scale below 12 Mb. Chromosome 11 is mapped onto a model polymer of length $N = 1120$, the coarse-grained segment length is set to 75 kb. The chromosome is divided into transcriptionally active (ridges) and inactive (anti-ridges) sections according to the Human Transcriptome Map (table 7.1). Different looping probabilities are assigned to ridge and anti-ridge sections. The model assumes that differences in local compaction arise due to different local looping probabilities. Results are shown for $\mathcal{P}_{\mathrm{R}} = 3 \times 10^{-5}$ (interaction inside ridge sections), $\mathcal{P}_{\mathrm{AR}} = 7 \times 10^{-5}$ (interaction inside anti-ridge sections) and $\mathcal{P}_{\mathrm{inter}} = 1 \times 10^{-5}$ (interactions between different sections).

Mb has been lacking. Several polymer models have been proposed, trying to explain experimental data from FISH distance measurements. Such experiments yielded scaling exponents for the mean square physical distance $\langle R^2 \rangle \sim N^{2\nu}$ ranging from $\nu \approx 0.1 - 0.5$. A similar variety can be found for the proposed models. In this chapter, we have extended the homogeneous RL model ($\curvearrowright$ 5.3) to present a unified view of chromatin folding, spanning different length scales and states of compaction. Such local differences in chromatin compaction are found for instance in ridges and anti-ridges along the q-arms of chromosomes 1 and 11 (Fig. 2.3A). The key idea is that local differences in compaction are related to different local looping probabilities along the contour of the polymer. Instead of developing different polymer models for differently compacted regions, such a model would offer an explanation comprising different states of compaction, i.e. the complete chromosome.

We presented a simple model, where local looping probabilities are assigned dependent on transcriptional activity and gene density as depicted by the Human Transcriptome

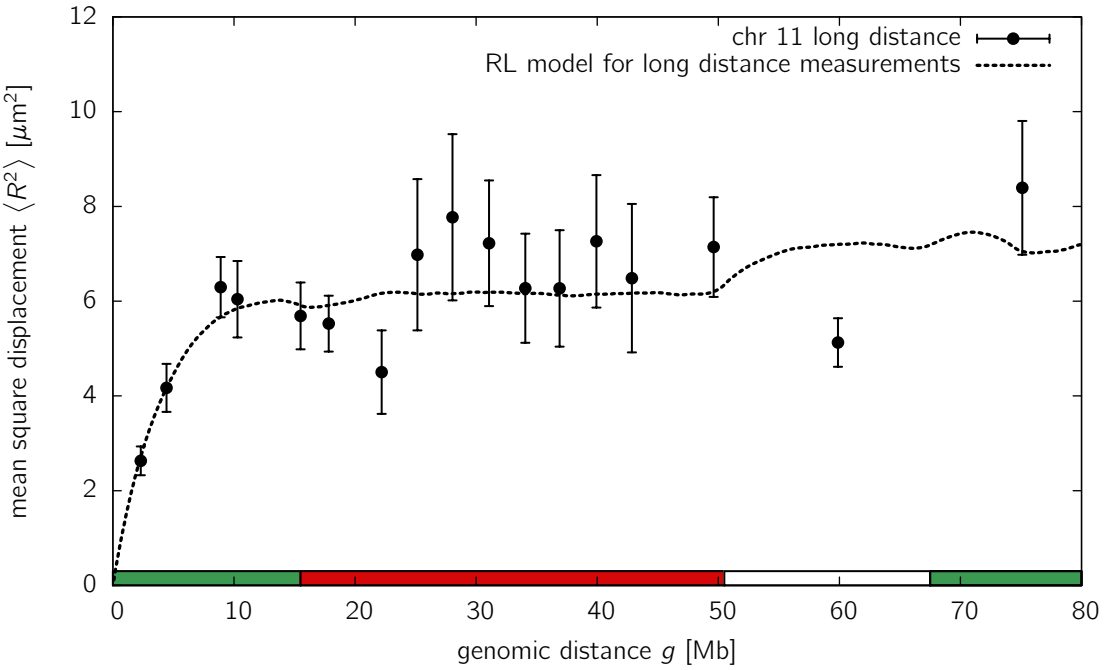


Figure 7.3: Comparison of FISH distance measurements on human chromosome 11 [12] to the heterogeneous Random Loop model on the length scale of the whole q-arm. We assign each region either the ridge or anti-ridge property according to Fig. 2.3A. In difference to Fig. 2.3A, however, in this first approach we assign only the anti-ridge and ridge property here in order to keep the number of freely adjustable parameters small. Parameters and calculations are the same as in Figure 7.2. The mean square distance is plotted against the genomic separation between two FISH probes, where the one FISH probe is held fixed at a specific reference point (starting point of the black arrow, cf. 7.1). The red, green and white rectangles indicate whether the position of the second marker is within a ridge or anti-ridge or within the large intergenic region, which is here assigned the same looping probability $\mathcal{P}_{AR}$ as the anti-ridge regions (see table 7.1). The data shows very good agreement with the model except for two data points. The reason for this deviation might be that we only distinguish between two kinds of regions here, namely ridge and anti-ridge, being a rather rough classification of transcriptional activity. However, the heterogeneous model with this rough distinction already leads to a good overall agreement with the data, without introducing unnecessarily many parameters.

Map [13]. In order not to introduce unnecessarily many parameters, different regions are assigned either one of the two properties: gene-rich and transcriptional active or gene-poor and transcriptional inactive. Although still highly simplifying, this explains remarkably well the difference in compaction of ridges and anti-ridges, assuming a 2.5-fold difference in looping probability for the studied region on human chromosome 11 (Fig. 7.2). There is abundant experimental evidence for heterogeneous chromatin looping along the chromatin fiber. For instance, loops with sizes in the 10 kb range have been observed in the beta-globin locus, where gene activity is correlated with loop formation that brings together different regulatory elements of the locus [130]. Another example are loops between promoter and enhancer sequences, which span a broad genomic length scale in the 1 and 1000 kb range [131]. Even larger loops are associated with transcription factories, where transcriptionally active genes from different parts of a chromosome as well as from different chromosomes are assembled [70].

Thus, the heterogeneous RL model allows for a unified description of the folding of the chromatin fiber inside the interphase nucleus over different length scales and explains different levels of compaction by assuming different looping probabilities, related for instance to local differences in transcription level and gene density. The heterogeneous RL model creates a basis for explaining the formation of chromosome territories, not requiring a scaffold or other physical confinement. While there is a lot of evidence that chromatin-chromatin interactions play a crucial role in genome function (e.g. see [130, 70]), our study proposes that it also plays an important role in chromatin organization inside the interphase nucleus on the scale of the whole chromosome as well as on that of subchromosomal domains in the size range of a few Mb. Importantly, various aspects of the RL model can be experimentally verified, e.g. by perturbing chromatin-chromatin interactions and analyzing its effect on chromatin folding. Although experimental data on loop distributions are not yet available, experimental techniques such as the 4C technology [25, 68] will allow the measuring of looping probabilities and loop size distribution along the length of complete chromosomes. These and other experimental parameters can be incorporated into the RL model, moving towards a stepwise more realistic polymer model for chromatin folding in higher eukaryotes.

Chapter 8

Effect of topological catenation and bonding constraints on rings and loops

References

The results presented in this chapter are published in:

- M. Bohn, D. W. Heermann, O. Lourenço and C. E. Cordeiro, **2010**. On the influence of topological catenation and bonding constraints on ring polymers. *revised manuscript submitted to Macromolecules*.

We thank Prof. H. Horner for fruitful discussions on this topic.

Chapter Summary

The existence of a ring structure for certain polymers in nature like proteins and DNA indicates a benefit compared to the linear form. Transcriptional regulation in higher eukaryotes is maintained amongst others by the formation of chromatin loops. Experimental studies revealed that different chromosomes as well as chromatin regions on one single chromosome tend to be segregated into distinct territories. Here, we study a system of two rings in both catenane and bonded topology as a toy model for the influence of loops and topological constraints on the polymers' conformational properties. Athermal Monte Carlo simulations reveal that the mean square radius of gyration $\left\langle R_{gyr}^2 \right\rangle$ *of catenated or bonded rings follows a scaling exponent similar to that of isolated rings, which in turn is close to the value of* $\nu \approx 0.588$ *for a self-avoiding walk. However, the effective segment length is larger for catenated rings, reflecting the swelling of the polymers. The shape of catenated and bonded rings, in contrast, shows pronounced differences even in the limit of infinite chain length. We observe a strong tendency towards segregation for the bonded topology in comparison to a similar ring-linear and linear-linear system. The orientation of the rings' gyration ellipsoids is slightly perpendicular, trying to minimize the overlap area. These findings indicate that loops might play an important role in the entropy-driven segregation of chromatin.*

8.1 Introduction

In this chapter, we take the focus away from modeling chromatin with all its complexity and investigate the effect of loops on the polymers' conformational properties separated from the detailed environment of the nucleus. A system of two ring polymers is used as an ideal toy model to study the effect of such topological interactions.

In the recent decades, ring polymers received considerable attention. Ring polymers differ from linear polymers in that the endpoints are connected to a cyclic form. The investigation of these polymers is not of pure theoretical interest. In fact, they are abundant in nature. First evidence for certain DNA molecules to occur in a ring shape has been found as early as 1962 by ultracentrifugation [132]. One year later, experimental results indicated that the double-stranded DNA of the polyoma virus exists in a closed cyclic structure [133, 134]. Nowadays it is clear that cyclic ring polymers are quite abundant in organisms and the ring structure seems to play an important role in the genomic function of organisms. Amongst others, circular DNA is found in bacteria, viruses as well as the mitochondrial DNA of eukaryotic cells. Last but not least, we tried to convince the reader in the last chapters that loops are an ubiquitous feature of chromatin organization in higher eukaryotes. These loops in turn can be viewed as ring polymers, although they are not isolated from the rest of the chain.

After the early findings of cyclic DNA one succeeded to artificially produce ring polymers with materials like polystyrene [135, 136, 137]. These synthetic ring polymers are created by anionic polymerization and have the advantage that their molecular weight can be controlled, allowing quantitative investigations on the static and dynamic properties. A lot of attention has been given to the dynamic properties in the melt, where a reptation-like movement was observed for ring polymers. While reptation for linear polymers is

driven by the effect of the chain ends, it was proposed [138] that the observations might be explained by a higher entanglement length of rings.

The conformational properties of isolated ring polymers have been extensively studied in the past by theory and simulations [139, 140, 141] both with and without excluded volume interactions. It was shown that self-avoiding polymer rings follow the same scaling behavior as isolated linear polymers in the limit of infinite chain length $N \to \infty$ concerning its static properties [142]. A lot of studies are devoted to the questions how the scaling behaviour of the radius of gyration $R_{gyr}(N)$ changes, when isolated ring polymers are constrained to a fixed topology, i.e. knot type. It has been proposed in an analytical argument [143] that topological constraints alone induce the same scaling as excluded volume interactions, i.e.

$$R_{gyr}(N) \sim N^{2\nu} . \tag{8.1}$$

Later, it was shown [144] that the size of a ring polymer without excluded volume in a quenched topology displays a crossover from a random-walk type of scaling to a self-avoiding walk type of scaling for larger chain lengths. This result was later confirmed by numerical studies for trivially knotted ring polymers [145].

A system of several rings can display a variety of different topological states. Rings can be either non-catenated or catenated, the degree of catenation varying. Brown *et al.* [146] studied the influence of topological constraints of isolated rings and rings in a melt with excluded volume and came up with the conclusion that topological constraints influence the isolated chain behavior only marginally while rings in a melt exhibit dramatic changes by turning on topological constraints. This is in agreement with a study by Müller [147] showing that non-catenated rings in a melt become compact with a scaling exponent of $\nu \approx 0.39$. This compactification, not detected for e.g. linear polymers, already indicates a strong effect of topology. The scaling exponent found is consistent with a Flory-type of argument [148], where the entropy loss due to the non-catenation constraint is linearly dependent on the number of neighboring rings, yielding a free energy contribution of $F_1 \sim kTR^3/N$ in three dimensions. This free energy contribution counteracts the entropy loss given by $F_2 \sim kTN/R^2$ arising from compaction. Minimizing the free energy $F = F_1 + F_2$ yields a scaling exponent of $\nu = 2/5$. Further computational studies led to the proposal that the scaling exponent is a cross-over effect asymptotically moving towards a behavior similar to compact lattice animals in a self-consistent network of topological constraints [149]. Summarizing these results, a general conclusion emerging is that topological constraints introduced by the cyclic structure into the problem of polymeric conformations alter conformational properties – in a melt of rings quite drastically.

While a lot of studies had their focus predominantly on the topological effects of the non-catenation constraint in a melt, the study of catenated or bonded rings can yield further insight into their topological interactions. Such rings are held relatively close together by means of the constraint, therefore they can be used as an ideal model for examining interactions between them. The catenane topology is common in nature. The equilibrium distribution of topological states of DNA was determined by random cyclization experiments of linear DNA, resulting in a broad distribution of linking numbers [150]. Later, it was shown that this equilibrium distribution is greatly modified by the activity of type II DNA topoisomerase, an enzyme which allows to release topological constraints [151]. Recently, the effect of catenation on protein folding stability has been studied [152]. Although chromatin is not circular inside the human interphase nucleus, there is now abundant experimental evidence that transcriptional regulation is accomplished by the formation of polymeric loops [20, 25]. Transcription factors are assumed to have the possibility to

aggregate inside the nucleus, building transcription hubs or factories [24] (↷ 2.4.3), resembling a topology of bonded rings (or loops). Such looping has recently been proposed to have a major impact on the observed segregation of chromosomes [35].

The scope of this study is to deepen the understanding of the influence of ring structure accompanied by the induced topological state on the conformational properties of the polymers involved. To achieve this, we study both isolated ring polymers as well as two-ring conformations with a fixed topology. The topological state of two ring polymers with paths C_1 and C_2 can be characterized by the Gauss linking number [153]

$$\Phi = \frac{1}{4\pi} \oint_{C_1} \oint_{C_2} \frac{\langle d\mathbf{r}_1 \times d\mathbf{r}_2, \mathbf{r}_1 - \mathbf{r}_2 \rangle}{|\mathbf{r}_1 - \mathbf{r}_2|^3} \,. \tag{8.2}$$

It has to be noted that a one-to-one correspondence between the Gauss link invariant and the topological state cannot be established. Three different topologies are investigated exemplarily: simple catenated rings (the Hopf link, $\Phi = 1$), double catenated rings ($\Phi = 2$) and rings which are non-catenated ($\Phi = 0$) but bonded together at one monomer. While the first two systems are interesting for DNA and protein catenanes, the latter resembles chromatin loops assembled in transcription hubs. Effects of catenation have been analyzed concerning dimensionality by Sikorski [154] for rings up to $N = 800$. He conjectured that the dimensions and shape of a catenated polymer ring approaches the values of an isolated ring in the limit of infinite chain length. However, a stringent analysis of the changes in dimensionality as well as shape has not been undertaken. In this study we focus on three points: Firstly, the changes in the dimensions characterized by the radius of gyration and the mean square ring diameter induced by the existence of a second ring are investigated. We carefully analyze whether the scaling exponent ν is different in the limit of infinite chain length $N \to \infty$. Secondly, we ask whether rings in a catenated or bonded topology display a different shape than their linear counterparts. In contrast to Sikorski's conjecture, we find pronounced differences in the shape of isolated rings compared to catenated or bonded rings even in the limit of large chain length. Finally, especially with respect to the biological systems, we ask the question of whether such a two-ring system is able intermingle freely or whether there is a tendency to segregate. Comparison with corresponding linear counterparts provides us with evidence that indeed topological constraints and looping lead to a more ordered and compartmentalized state.

This chapter is organized as follows: In the next section (↷ 8.2) we describe the simulation method as well as the sampling procedure. We then present results for the dimensions of isolated rings as well as the two-ring topologies studied in section 8.3.1. Afterwards, the influence of the topological constraints on the shape of the rings is investigated (↷ 8.3.2). In the last two sections, we elaborate on their tendency to intermingle or segregate (↷ 8.3.3) as well as their mutual alignment (↷ 8.3.4).

8.2 The computational model

To investigate the influence of topology on the properties of ring polymers, a model system of two ring conformations is simulated in a fixed topological state with excluded volume interactions. The topologies studied are displayed in Figure 8.1. Dimensionality and shape of isolated rings (Fig. 8.1A) are calculated as a reference. The catenane topology is studied using the two simplest link types: First, the Hopf link with a Gauss linking number $\Phi = 1$, i.e. chain 1 passes through the surface of chain 2 exactly once (Fig. 8.1B). Secondly, double catenated rings with a Gauss linking number $\Phi = 2$ (Fig. 8.1C). It is to be stressed at this

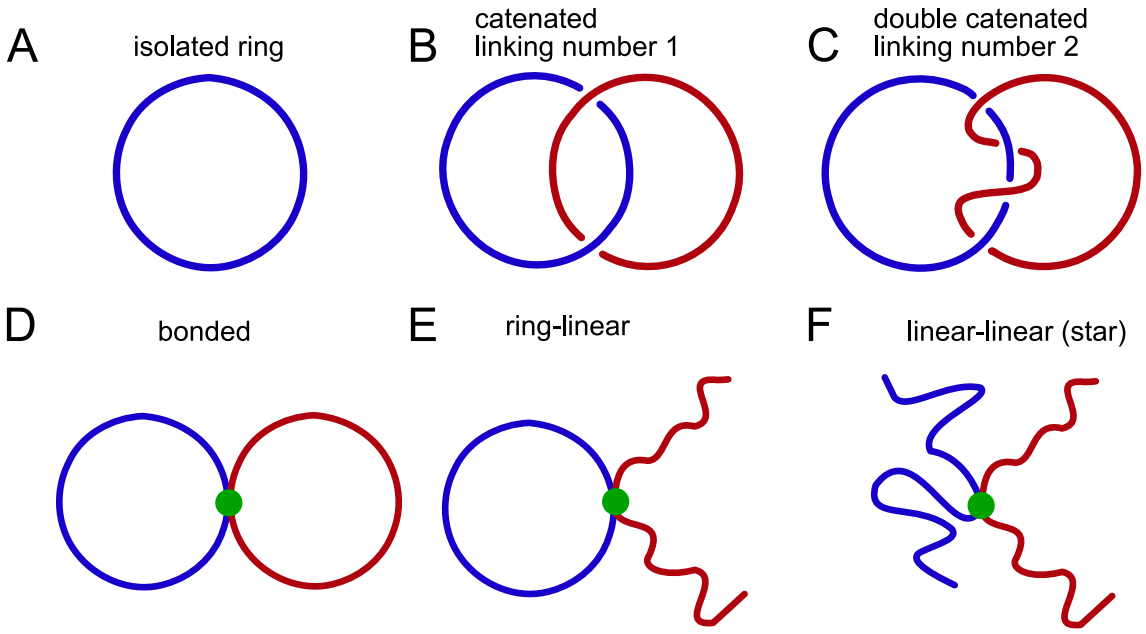


Figure 8.1: Different ring topologies used in this study. Isolated rings (A) are used as the reference ensemble. Two types of linked two-ring conformations are studied: Catenated rings with a Hopf link (B), i.e. the simplest non-trivial link type with Gauss linking number $\Phi = 1$, and double catenated rings with $\Phi = 2$ (C). Rings without catenation, but with two monomers bonded to each other are studied (D). Ring-linear (E) and star-shaped linear-linear (F) conformations are simulated for purposes of comparison.

point that we do not sample the complete ensemble of rings with Gauss linking number one or two, as the correspondence of linking number and link topology is not one-to-one. Rather we sample the specific link types shown in Fig. 8.1 as we are interested in the influence of the topological constraints rather than the properties of the Gauss linking number. Non-catenated rings with $\Phi = 0$ are set up such that they are connected at one point, this connection persisting during the time of simulations (Fig. 8.1D). To deepen the insight into entropic and topological effects of ring closure, we study the behavior of bonded non-catenated chains where one or both rings are cut to make a ring-linear or linear-linear star-shaped system (Fig. 8.1E and F).

8.2.1 The Monte Carlo algorithm

To simulate ring polymers we use a coarse-grained lattice Monte Carlo method [111]. Simulations are performed on a lattice in order to simplify the handling of excluded volume. Calculation of excluded volume interactions is thereby reduced to checking whether one lattice site is already occupied or not. In general, our lattice algorithm should fulfill the following three criteria: (i) it produces unbiased results, i.e. each possible conformation out of the ensemble is sampled with equal probability, (ii) it takes into account excluded volume interactions, i.e. two monomers are not allowed to occupy the same region in space and (iii) using some restrictions on the moves and bond vectors it ensures that no bond crossings can occur during a Monte Carlo step, i.e. it preserves the topological state of the conformation.

Several lattice algorithms have been used in the past to study polymer conformations. The Verdier-Stockmayer algorithm [85] assumes each monomer to reside on a vertex of a

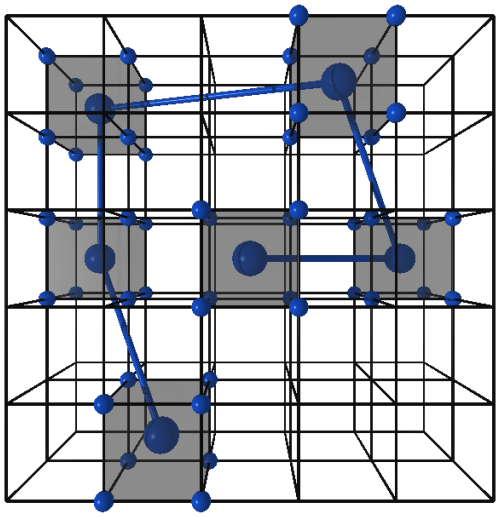


Figure 8.2: Illustration of the bond fluctuation model. Each monomer sits in the center of a unit cube (large sphere), rendering all of its eight vertices occupied (small spheres). Bond vectors are allowed to fluctuate with a maximum distance of $b = \sqrt{10}$.

simple cubic lattice. Beads along the chain are connected by a linker with a fixed length b of one lattice unit. Several local moves can be applied, including kink, crankshaft and end rotation moves. The Verdier-Stockmayer algorithm has several drawbacks. Amongst others, he is non-ergodic [110] and angles between neighboring bonds are restricted to 90° or 180°. To overcome these problems and at the same time maintaining the advantages of a lattice model, the bond fluctuation method has been introduced by Carmesin [155]. It has been successfully applied to several studies on the static and dynamical properties of polymer systems [156, 113, 147, 149].

In the bond fluctuation algorithm, each monomer resides in the center of a unit cube, blocking all its eight vertices to be occupied by other monomers (Figure 8.2). Thus, the implementation of excluded volume effects reduces to checking whether one of the vertices belonging to a monomer position is already occupied or not. Clearly, the minimum bond length is $b = 2$ lattice units. However, the bond length is allowed to fluctuate in a certain range $2 \leq b \leq b_{\max}$. Care has to be taken concerning the preservation of topology, which is important here as we would like to investigate effects of topological constraints. However, a fixed topology is accomplished automatically by putting restrictions on the allowed bond vectors and moves. In order to resemble the dynamics of real polymers, only local motion is applied where one monomer is moved at a time step maximally to one of its six nearest neighbors, i.e. the change in distance is $\delta r \in \{0, \pm 1\}$. To ensure that no bond-crossings can occur during one local move without checking it manually every time, certain constraints are imposed on the allowed bond vectors. Indeed, it can be shown [156] that using an allowed set $\mathcal{B}$ of 108 bond vectors topology is automatically preserved. The set of vectors we use in our simulation are

$$\mathcal{B} = \{(2,0,0), (2,1,0), (2,1,1), (2,2,1), (3,0,0), (3,1,0) + \text{permutations} + \text{sign combinations}\}\,. \quad (8.3)$$

Clearly, using this set of vectors, the allowed bond length is restricted to

$$b \in \left\{2, \sqrt{5}, \sqrt{6}, 3, \sqrt{10}\right\}\,. \quad (8.4)$$

The chain is propagated via the Metropolis Monte Carlo algorithm [111], i.e. a new conformation is proposed and accepted with a probability of $\min\{1, \exp(-\beta\,\delta U)\}$. The proposal of a new conformation is based on randomly selecting one monomer, followed

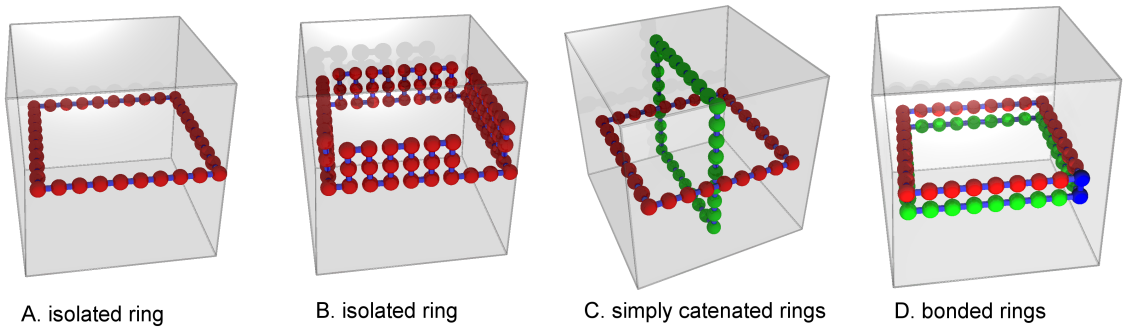


Figure 8.3: Initial conformations for the Monte Carlo algorithm. Isolated rings (A) are positioned as a square in the xy-plane, if the chain is too large to fit into the given lattice, small humps in the z-direction are created (B). Simple catenated rings (C) are positioned perpendicular to each other such that one ring goes through the surface of the path created by the second ring. Bonded rings (D) are set up both as rectangles in the xy-plane, the z-coordinate differing by 2. Between two monomers (marked blue), a bond $\mathbf{b} = (0, 0, 2)$ is created.

by randomly selecting one of the nearest neighbors' positions on the lattice. The energy calculation of our athermal simulations is quite easy: If the lattice site is already occupied, or one of the two bond vectors does not belong to the set $\mathcal{B}$ of allowed bond vectors, the change in energy is $\delta U = +\infty$ and the proposed conformation is rejected, otherwise the change in energy is $\delta U = 0$, resulting in an acceptance of the new conformation. In the following, one Monte Carlo step (1 MCS) is defined as the number of attempted moves, where on average each monomer has performed one trial move.

The basic source template for the program has been kindly provided by Wolfgang Paul written in Fortran. During my PhD I have rewritten the program using C++, allowing for a high degree of flexibility. Especially, bonds between two monomers can be arbitrarily added and removed during the simulation run, each monomer can have several bonds attached to it and bonds can exist between any two monomers even on different chains. In this chapter, bonds are fixed by the start configuration and not modified during the simulation. For isolated rings (Fig. 8.1A), each monomer has two bonds attaching it to its nearest neighbors along the chain. The bonded topology (Fig. 8.1D) is implemented by each monomer being connected to its neighbors along the chain. On top, a bond is created between two monomers residing on different chains. Clearly, this bond obeys the same rules as any other, i.e. it is restricted to the set $\mathcal{B}$ of bond vectors.

8.2.2 Data acquisition and autocorrelation times

The linear dimension of the simulation box is chosen between $L = 256$ and $L = 800$, ensuring L to be larger than the typical size of the polymer, $L \gg \sqrt{\langle R_{gyr}^2 \rangle}$. Although periodic boundary conditions are used, the algorithm keeps track of unfolded coordinates so that the monomer can diffuse freely through space. The choice of the size of the simulation box is guided by the attempt to minimize the probability that monomers which are far apart in terms of unfolded coordinates touch each other inside the simulation box, while at the same time keeping the effort in memory usage during the simulation reasonable. By using this method, the chains do not feel the simulation box nor the periodic boundaries.

As the topology cannot be changed during the simulation run, the initial conformation

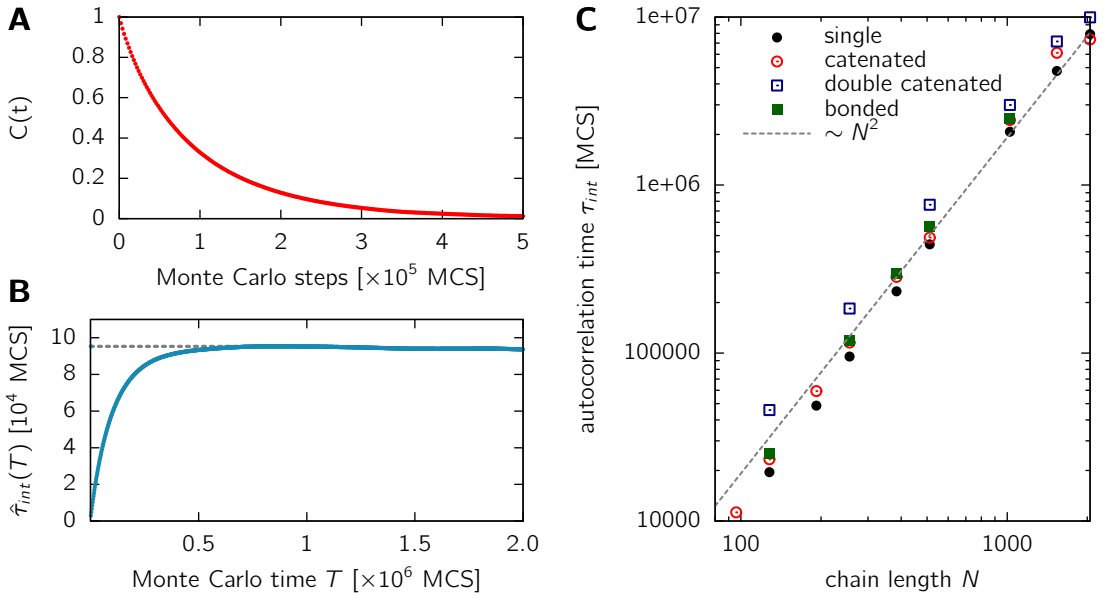


Figure 8.4: Autocorrelation analysis of ring polymers. **A.** This panel shows the autocorrelation function $C(t)$ [eq. (8.6)] for isolated ring polymers of length $N = 256$. **B.** The integrated autocorrelation time $\hat{\tau}_{int}(T)$ depending on the cut-off time T of the autocorrelation function $C(t)$ in panel A. The grey line marks the autocorrelation time τ_{int} where the break condition eq. (8.9) is fulfilled. **C.** This panel displays the integrated autocorrelation times τ_{int} vs. chain length N for the ring topologies studied. The data is shown in a log-log plot to highlight the scaling behavior $\tau_{int} \sim N^2$ for large chain lengths N.

has to be set up with the correct topology. We demand isolated rings to be unknotted, while simple catenated rings have to pass through each others surface spanned by their paths exactly once. The initial conformation for isolated chains is set up as a rectangle in the xy-plane as shown in Figure 8.3A; if the chain length is too large with respect to the lattice size, small humps in the z-direction of two monomers are created (Fig. 8.3B). For the simple catenated ring conformations, the second ring is positioned perpendicularly to the first one and shifted by 2 lattice units, as shown in Figure 8.3C. Similarly, double catenated rings are created. In case of non-catenated bonded rings, two rings are set up comparable to the isolated case at different z-values and connected at one site (Fig. 8.3D).

As the simulation method applied here only implements local moves, one conformation emerging from another by propagating it one Monte Carlo step differs only slightly; in other words, both conformations are highly correlated. In order to obtain uncorrelated samples of ring conformations, a certain number of Monte Carlo steps have to be performed, given by the autocorrelation or relaxation time τ_{int} [110]. We approximate τ by calculating the autocorrelation function $C(t)$ of the squared radius of gyration $A(t) = R^2_{gyr}(t)$ from the simulation data,

$$R(t) = \frac{1}{K-t}\sum_{i=0}^{K-t}(A(i+t) - \langle A\rangle)(A(i) - \langle A\rangle) \tag{8.5}$$

$$C(t) = \frac{R(t)}{R(0)} \tag{8.6}$$

Here K denotes the number of Monte Carlo steps performed, $C(t)$ gives the correlation factor of two conformations being separated by t time steps along the Markov chain gen-

erated by the Monte Carlo algorithm. From the autocorrelation function, the integrated autocorrelation time τ_{int} is calculated. It is generally defined as

$$\tau_{int} = \frac{1}{2} + \sum_{t=1}^{\infty} C(t) \tag{8.7}$$

As simulations do not run up to infinite time, an approximation becomes necessary. Here, we apply Sokal's windowing procedure [157]: First, the integrated autocorrelation time is calculated up to a certain cutoff T,

$$\hat{\tau}_{int}(T) = \frac{1}{2} + \sum_{t=1}^{T} C(t) \tag{8.8}$$

Then, the smallest integer $T^\star$ is determined which fulfills the criterion

$$T \geq 10\ \hat{\tau}_{int}(T) \tag{8.9}$$

Finally, the integrated autocorrelation time for the observable $A(t)$ is defined as

$$\tau_{int} = \hat{\tau}_{int}(T^\star) \tag{8.10}$$

For this method to be applicable, sufficient data has to be available, meaning $K \gtrsim 1000\tau_{int}$. Although, in principle, the integrated autocorrelation time has to be calculated for the (unknown) slowest relaxation mode in the system [110], we use the radius of gyration as a measure of relaxation time following the approach from various publications [147, 149].

The integrated autocorrelation time is related to the statistical error of measurements, $2\tau_{int}$ being the factor by which the variance of the observable $A(t)$ is larger than it would be if the conformations were completely decorrelated. For other observables $B(t)$, such a simple relation can not be established. In what follows, the statistical errors are calculated based on considering two subsequent conformations as independent after $5\tau_{int}$ MCS.

The autocorrelation function $C(t)$ is shown exemplarily for isolated rings of length $N = 256$ in Figure 8.4A. Clearly, the autocorrelation function displays an exponential decay. The function of integrated autocorrelation times $\hat{\tau}_{int}(T)$ with cutoff T [eq. (8.8)], is shown in Figure 8.4B. The grey line represents the integrated autocorrelation time τ_{int}, determined by the break condition eq. (8.9). The autocorrelation times are displayed in panel C for various chain lengths and topologies in a log-log plot. Autocorrelation times increase quadraticly in the chain length, $\tau \approx aN^2$, the scaling constant being $a \approx 2.01$ for isolated rings and $a \approx 2.54$ for simple catenated rings.

Simulations are performed with chain lengths ranging from $N = 32$ up to $N = 2048$. For each chain length and topology, 8 000 to 80 000 independent conformations were created.

8.3 Results

8.3.1 Catenation and bonding influences ring dimensions slightly

While the dimensions of linear polymers are often characterized by the mean squared end-to-end distance $\langle R_e^2 \rangle$, which exhibits the well-known scaling law $\langle R_e^2 \rangle \sim N^{2\nu}$, this is obviously not a good measure for the size of a ring polymer, where end points are not defined at all. Instead, one can characterize the dimensions of a ring polymer by the mean

Table 8.1: Scaling exponents ν for the mean squared radius of gyration and ring diameter. The scaling exponents have been determined by a fit to $\langle R_{gyr}^2 \rangle = b^2 N^{2\nu_1}$ and $\langle d^2 \rangle = b^2 N^{2\nu_2}$ using the least-squares algorithm. Errors represent the asymptotic standard error of the fit.

type	ν_1	ν_2
isolated	0.601(1)	0.596(1)
catenated	0.592(1)	0.5871(3)
double catenated	0.594(1)	0.594(1)
bonded	0.598(1)	0.595(1)

square ring diameter $\langle d^2 \rangle$ which is defined by the average squared distance between two monomers separated by $N/2$ monomers along the contour of the chain

$$\left\langle d^2 \right\rangle = \frac{1}{|\mathcal{C}|} \frac{1}{N} \sum_{c \in \mathcal{C}} \sum_{i=1}^{N} \parallel \mathbf{r}_{i+N/2}^c - \mathbf{r}_i^c \parallel^2 \ . \tag{8.11}$$

Here $\mathcal{C}$ denotes the set of sampled conformations, and $\mathbf{r}_i^c$ denote the coordinates of the i^{th} monomer in conformation $c \in \mathcal{C}$. Furthermore, the radius of gyration $\left\langle R_{gyr}^2 \right\rangle$ can be used as a characterization of the length scale, both for rings as well as self-avoiding walks.

It has been shown in several studies, that isolated self-avoiding ring polymers follow a similar scaling law as linear polymers both for the radius of gyration $\left\langle R_{gyr}^2 \right\rangle$ as well as for the ring diameter $\langle d^2 \rangle$,

$$\left\langle R_{gyr}^2 \right\rangle_i \sim N^{2\nu} \ , \qquad \left\langle d^2 \right\rangle_i \sim N^{2\nu} \ .$$

The index i denotes an average over the ensemble of isolated chains. Sikorski [154] determined the scaling exponent for the radius of gyration to $\nu = 0.587$. Similar results were found by Brown et al. [146] ($\nu = 0.585$) and Müller et al. [147] ($\nu = 0.595$). The exponents for the mean square ring diameter are differing only slightly in these studies: $\nu = 0.59$ (Sikorski), $\nu = 0.585$ (Brown), $\nu = 0.605$ (Müller). Within the errors of the simulations, these results show that isolated polymer rings have the same scaling exponent as linear self-avoiding polymers.

Furthermore, Sikorski in his paper [154], analyzed these quantities for one ring concatenated with a second ring. His data shows that even catenated rings ($\Phi = 1$) obey a scaling law with scaling exponents $\nu = 0.591 \pm 0.002$ for the radius of gyration and $\nu = 0.587 \pm 0.002$ for the ring diameter, suggesting that the scaling exponents are equal within the statistical errors. However, an analysis for the limit $N \to \infty$ has not been performed.

Here, we determine the scaling exponents ν for the above mentioned topologies and carry out a finite-size analysis to see whether the scaling exponents differ in the limit of infinite chain length N. Figures 8.5A and B show the radius of gyration $\left\langle R_{gyr}^2 \right\rangle$ and the ring diameter $\langle d^2 \rangle$ for chains of length up to $N = 2048$ in a logarithmic plot. Consistent with other studies, we find exponents of $\nu = 0.601(1)$ for the radius of gyration and $\nu = 0.596(1)$ for the mean square ring diameter of single rings. For simple catenated rings parameter fitting yields $\nu = 0.592(1)$ for the radius of gyration and $\nu = 0.5871(3)$ for the ring diameter. Error estimates of the fitting parameters are based on the asymptotic standard errors resulting from a least-squares fit using the gnuplot software (version 4.2). Scaling exponents including other topologies are summarized in table 8.1.

It has been shown [149] that the topological constraint of non-catenation induces a change in the scaling law from the self-avoiding walk to a globular state in a melt of

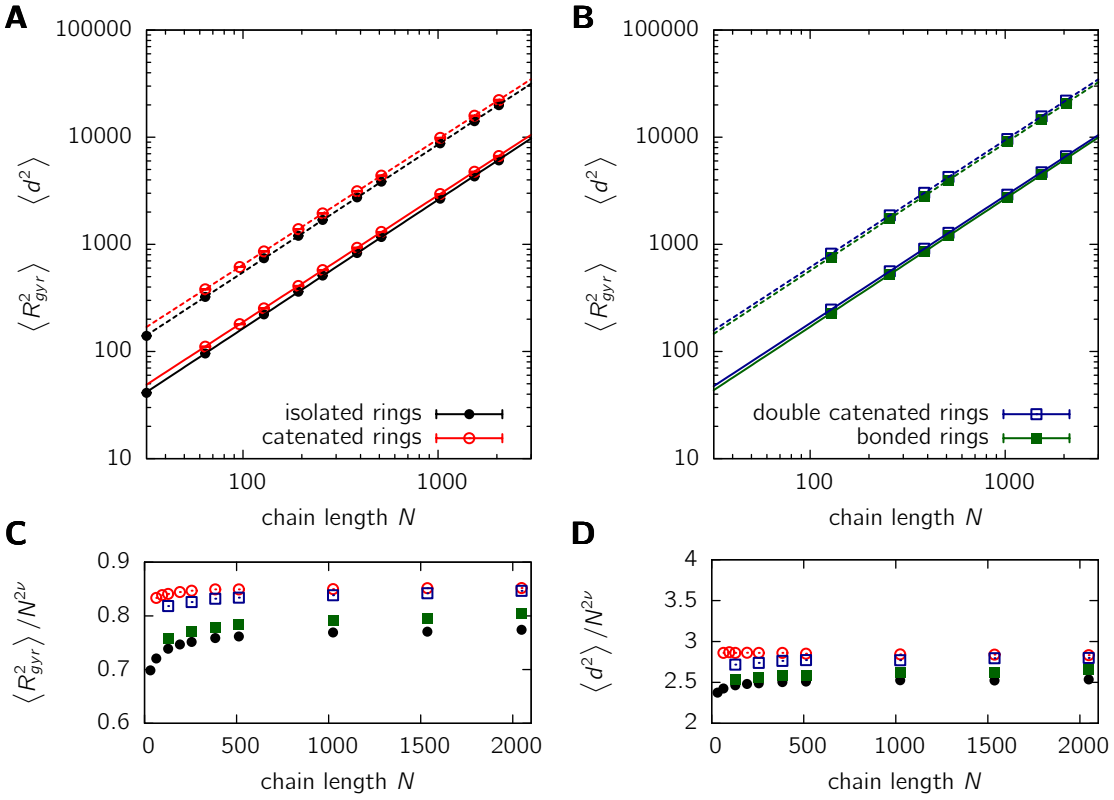


Figure 8.5: Dimensions of ring polymers. A. and B. Log-log plots of the mean squared radius of gyration $\langle R_{gyr}^2 \rangle$ (solid line) and the mean square ring diameter $\langle d^2 \rangle$ (dashed line) in dependence of chain length N. Results are presented for isolated rings (solid circles), single catenated rings (open circles), double catenated rings (open squares) and non-catenated bonded rings (solid squares); separation in two figures is done for reasons of readability. Lines represent a fit to the scaling law $\langle \cdot \rangle = b^2 N^{2\nu}$. **C. and D.** The mean squared radius of gyration $\langle R_{gyr}^2 \rangle$ and ring diameter $\langle d^2 \rangle$ with the leading order term $N^{2\nu}$ of the self-avoiding walk behavior $\nu = 0.588$ divided out. In all cases standard errors are smaller than the point size.

polymers. To analyze whether the scaling exponent of a ring linked or bonded to a second ring is actually equal in the limit of large chains, we analyze the swelling factor s,

$$s = \left\langle R_{gyr}^2 \right\rangle_t / \left\langle R_{gyr}^2 \right\rangle_i . \tag{8.12}$$

Here $\langle \cdot \rangle_i$ denotes an average over the ensemble of isolated rings, whereas $\langle \cdot \rangle_t$ designates the ensemble average over one of the two-ring topologies, i.e. catenated ($\Phi = 1$), double catenated ($\Phi = 2$) or bonded ($\Phi = 0$). In case of the scaling exponents being unequal, the swelling factor would scale with $s \sim N^{\delta\nu}$, $\delta\nu \neq 0$, thus displaying either convergence to zero or divergence to infinity in the limit $N \to \infty$. The swelling factor is shown in Figure 8.6 versus the inverse chain length to allow for extrapolation. We find that for all two-ring topologies studied, the swelling factor s adopts a constant and finite value in the limit of large chain lengths. Clearly, no divergence to infinity or convergence to zero can be found. To obtain an estimate for the swelling factor in the large chain limit, we perform a crude linear extrapolation assuming that in first order finite-size corrections scale as $s \sim \mathcal{O}(1/N)$. We find that catenated and double catenated rings are swollen by a factor of about 10% compared to isolated chains (catenated rings: $s = 1.105(2)$, double

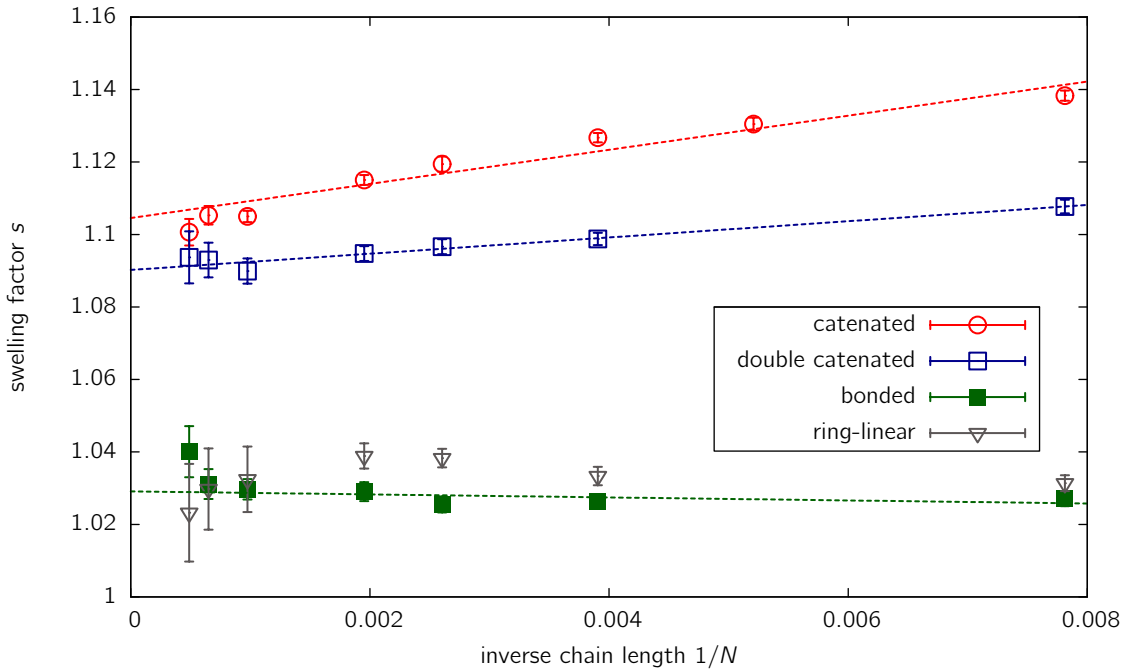


Figure 8.6: Swelling of ring polymers in the presence of a second one.. Shown is the swelling factor $s = \langle R_{gyr}^2 \rangle_t / \langle R_{gyr}^2 \rangle_i$ between the mean squared radius of gyration of a ring topologically constrained to a second ring (index t) and isolated rings (index i). The figure reveals that catenated rings forming the Hopf link and double catenated rings are more swollen compared to isolated chains by about 10%, the effect being smaller but existent for bonded rings. Data is plotted versus the inverse chain length $1/N$ to allow for an extrapolation to $N \to \infty$. Such an extrapolation results in finite values of the swelling factor s, indicating that the scaling exponents of isolated rings and rings within a two-chain system display the same scaling exponent.

catenated rings: $s = 1.090(1)$). Bonded rings are smaller, their size being only about 3% larger than isolated rings ($s = 1.028(2)$).

We conclude that changes in dimensionality are not as drastic by the introduction of topological constraints as a different scaling exponent would induce. Nevertheless the effective segment length b (the prefactor of the scaling law) becomes larger in the presence of a second ring, thus the ring is swollen by a constant factor in comparison to isolated chains. Comparison of the swelling factor s for bonded rings and the corresponding ring-linear system indicates that the swelling is due to the extra material in vicinity rather than topological constraints.

Scaling considerations suggest that both the radius of gyration and the mean square ring diameter have to follow the same scaling law: There is only one length-scale involved in the system, which is parameterized by the chain length and the bond length. Therefore, the scaling exponent ν has to be equal for both quantities. This can be validated by looking at the ratio of radius of gyration to mean square ring diameter $\mathcal{R} \equiv \left\langle R_{gyr}^2 \right\rangle / \langle d^2 \rangle$. The complementary quantities for linear chains are the radius of gyration and the end-to-end distance, here the ratio $\left\langle R_{gyr}^2 \right\rangle / \langle R_e^2 \rangle$ is a constant, which turns out to be equal to $\frac{1}{6}$ for both the random walk as well as the self-avoiding walk polymer model [79]. The ratios $\mathcal{R}$ for single and catenated rings are shown in Figure 8.7. Deviations from a constant value are most probably due to finite-size effects, as we find $\mathcal{R}(N) \to \text{const.}$ $(N \to \infty)$ for both single and catenated rings. Although we do not know how to correct this quantity for

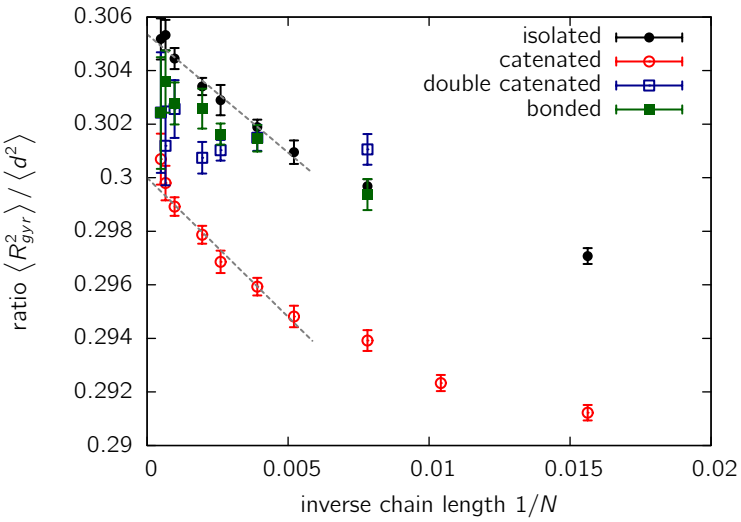


Figure 8.7: Ratio $\langle R_{gyr}^2 \rangle / \langle d^2 \rangle$ between radius of gyration and the mean square ring diameter for the ring topologies under study. A linear extrapolation to $N \to \infty$ has been conducted for isolated and simple catenated rings (grey lines).

finite-size effects, for large chain length ($N > 256$) corrections to the limiting case seem to be of the order of $\mathcal{O}(1/N)$. Linear regression yields the ratios $\mathcal{R}(N \to \infty) = 0.3053(2)$ for isolated rings and $\mathcal{R}(N \to \infty) = 0.2995(3)$ for catenated rings. Extrapolation for double catenated and bonded rings is more difficult, since the data shows more fluctuations, the ratios are, however, bounded by the values for isolated and catenated rings.

8.3.2 Shape of ring polymers changes due to topological constraints

In the last section there has been given evidence that the dimensions of a ring polymer constrained to another using the topologies in Fig. 8.1 change only with respect to the effective segment length b and not to the scaling exponent ν compared to an isolated ring, leading only to a mild swelling of the polymer (by a factor of up to 1.10) even in the limit of very large chains. Here, we want to analyze how the existence of a second ring catenated or bonded to it influences the shape of this polymer. It is common to describe the shape properties of a ring by means of the gyration tensor [158, 159]. This quantity, represented as a 3×3 matrix, describes the distribution of points in space and is defined by

$$S_{mn} = \frac{1}{N} \sum_{i=1}^{N} r_m^{(i)} r_n^{(i)} \tag{8.13}$$

Here $\mathbf{r}^{(i)}$ is the coordinate vector of the ith monomer and the subindex denotes its Cartesian components. The matrix $\mathbf{S}$ is symmetric and positive semi-definite, thus it can be transformed to a diagonal matrix where the three eigenvalues $\lambda_1 \leq \lambda_2 \leq \lambda_3$ give the squared lengths of the principal axes of gyration of the associated gyration ellipsoid. The shape of the ellipsoid resembles the distribution of masses of the polymer containing the monomer coordinates only in a summarized way, therefore allowing comparison between different conformations.

The ratios of the eigenvalues $\langle \lambda_3 \rangle / \langle \lambda_1 \rangle$ and $\langle \lambda_2 \rangle / \langle \lambda_1 \rangle$ indicate the deviation from a sphere-like shape of the polymer, both having a value of unity for a sphere. It is well-known that individual linear polymer chains display a pronounced asphericity, which shows up in

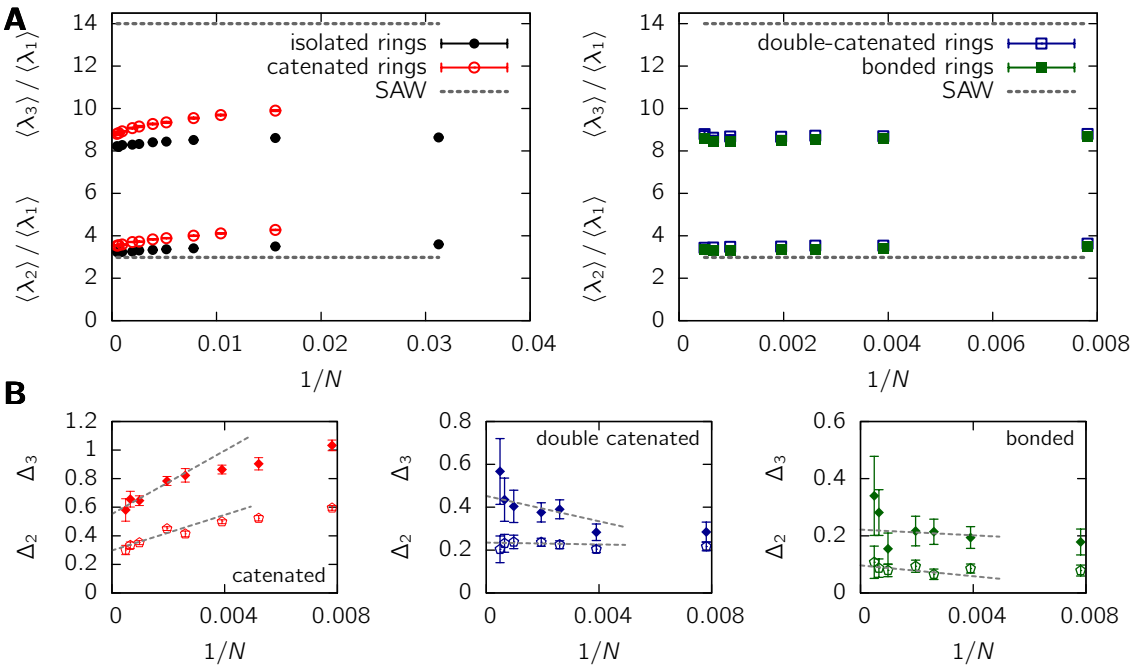


Figure 8.8: Changes in shape of two-ring conformations. A. The ratios of the average eigenvalues $\langle\lambda_3\rangle / \langle\lambda_1\rangle$ and $\langle\lambda_2\rangle / \langle\lambda_1\rangle$ of the gyration tensor for single (closed circles) rings, catenated rings (open circles, linking number $\Phi = 1$), double catenated rings (open squares, linking number $\Phi = 2$) and rings bonded to each other (solid squares). Ratios show distinct differences from linear chains (self-avoiding walk, dashed grey line), however differences between isolated rings and rings catenated or bonded to a second one are more subtle. **B.** This panel shows the difference between the eigenvalue ratios of a ring in a two-chain conformation and isolated rings $\Delta_k = \langle\lambda_k\rangle_t / \langle\lambda_1\rangle_t - \langle\lambda_k\rangle_i / \langle\lambda_1\rangle_i$ $(k = 2, 3)$. Δ_3 is represented by solid diamonds, Δ_2 by open diamonds. The plots reveal that the shape of isolated and catenated or bonded rings is different even in the limit of infinite chains.

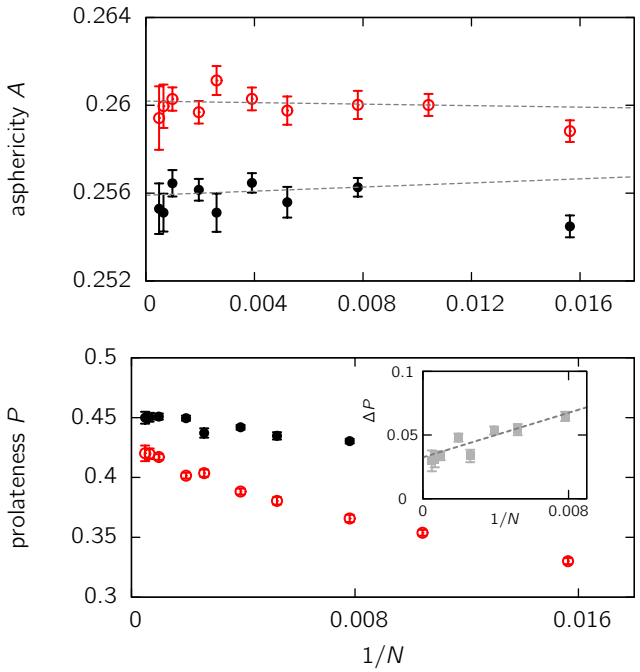


Figure 8.9: Asphericity A and prolateness P for single and catenated ring polymers. Data is plotted against the inverse chain length $1/N$ to allow for extracting the asymptotic limit $N \to \infty$. Isolated rings are represented by solid black circles, simple catenated rings by open red circles. The inset shows the difference in the prolateness P between single and catenated rings in relation to inverse chain length $1/N$. The difference does not vanish for infinite chain length, thus prolateness between isolated and catenated rings is systematically different.

Table 8.2: Shape of ring polymers. The shape is characterized by the asphericity A, prolateness P and the eigenvalue ratios $\langle\lambda_3\rangle : \langle\lambda_2\rangle : \langle\lambda_1\rangle$ of the gyration tensor extrapolated to $N \to \infty$. Extrapolation has been conducted by a linear fit to the data displayed in Fig. 8.8A for chains longer than $N = 128$. Fitting was done using the least-squares algorithm in gnuplot (version 4.2), errors indicate the asymptotic standard error of the fit.

type	$\langle\lambda_3\rangle : \langle\lambda_2\rangle : \langle\lambda_1\rangle$	A	P
isolated	8.23(2) : 3.22(2) : 1	0.2559(3)	0.451(2)
catenated	8.87(3) : 3.56(2) : 1	0.2602(3)	0.416(3)
double catenated	8.66(2) : 3.46(1) : 1	0.258(1)	0.424(3)
bonded	8.46(2) : 3.31(1) : 1	0.2592(3)	0.416(2)

the asymptotic ratios of the eigenvalues, namely [83]

$$\langle\lambda_3\rangle : \langle\lambda_2\rangle : \langle\lambda_1\rangle \to 12 : 2.7 : 1 \qquad \text{(random walk)}$$
$$\langle\lambda_3\rangle : \langle\lambda_2\rangle : \langle\lambda_1\rangle \to 14 : 2.98 : 1 \qquad \text{(self-avoiding walk)}$$

We determined the eigenvalue ratios for isolated, simple and double catenated ring polymers and bonded rings in dependency of their chain length. Results in Figure 8.8A show that isolated ring polymers are more spherical than their linear counterparts, in agreement with simulations in Ref. [139]. In fact, this result does not come as a big surprise: One would recover the linear chain result if the ring was collapsed to a linear polymer with two strands aligned parallel to each other. However, this is only one possible conformation and the ring has much more entropic degrees of freedom, all of them having a less pronounced asymmetry. While the deviations between self-avoiding walks and rings are large concerning the elongation of the polymer, differences in shape of isolated rings and the two-ring topologies studied are more subtle. In fact, we find differences between isolated rings compared to their counterparts in a two-ring system; the question of whether these differences remain in the infinite chain limit, however, cannot be answered from Figure 8.8A. Therefore a thorough extrapolation to the infinite chain limit is conducted in Figure 8.8B for the three topologies under investigation. The figures show the difference in the eigenvalue ratios compared to the isolated case,

$$\Delta_k \equiv \langle\lambda_k\rangle_t \,/\, \langle\lambda_1\rangle_t - \langle\lambda_k\rangle_i \,/\, \langle\lambda_1\rangle_i \qquad (k = 2, 3). \tag{8.14}$$

Data is plotted against the inverse chain length $1/N$ to allow for an extrapolation to the infinite chain limit. Linear extrapolation to $N \to \infty$ indeed indicates that the difference does not vanish for either of the topologies studied, hence the shape of an isolated ring is significantly different to that of a ring topologically constrained to another.

A crude linear extrapolation of the eigenvalue ratios for single rings yields

$$\langle\lambda_3\rangle \,/\, \langle\lambda_1\rangle \to 8.23 \pm 0.02 \quad (N \to \infty) \tag{8.15}$$
$$\langle\lambda_2\rangle \,/\, \langle\lambda_1\rangle \to 3.22 \pm 0.02 \quad (N \to \infty) \tag{8.16}$$

These results differ slightly from data of Bishop *et al.* [139], where ratios of 7.76 : 3.10 : 1.00 were reported for isolated rings with excluded volume. Although this study applied off-lattice simulations, the maximum chain length $N = 64$ might still be in a regime where finite-size effects are observable.

For catenated rings with linking number $\Phi = 1$ we obtain the ratios

$$\langle\lambda_3\rangle : \langle\lambda_2\rangle : \langle\lambda_1\rangle \rightarrow 8.87(3) : 3.56(2) : 1 \tag{8.17}$$

A list with the ratios for double catenated rings and bonded rings is given in table 8.2.

Other publications [160, 158] use different measures for the shape of the gyration tensor. Two of these measures are the asphericity A and the prolateness P. The asphericity value displays deviations from the sphere-like shape of the polymer and is defined by

$$A(\lambda_1, \lambda_2, \lambda_3) = \frac{(\lambda_1 - \lambda_2)^2 + (\lambda_1 - \lambda_3)^2 + (\lambda_2 - \lambda_3)^2}{2(\lambda_1 + \lambda_2 + \lambda_3)^2}. \tag{8.18}$$

For a rod-like shape of the gyration ellipsoid we find $A = 1$, while for a sphere-like shape $A = 0$. The prolateness is given by

$$P(\lambda_1, \lambda_2, \lambda_3) = \frac{(2\lambda_1 - \lambda_2 - \lambda_3)(2\lambda_2 - \lambda_1 - \lambda_3)(2\lambda_3 - \lambda_1 - \lambda_2)}{2(\lambda_1^2 + \lambda_2^2 + \lambda_3^2 - \lambda_1\lambda_2 - \lambda_1\lambda_3 - \lambda_2\lambda_3)^{3/2}}. \tag{8.19}$$

A positive value of P indicates a prolate shape of the gyration ellipsoid, a negative value an oblate shape.

Prolateness and asphericity for isolated and simple catenated rings with up to $N = 2048$ monomers are shown in Figure 8.9. We find that the asphericity is larger for catenated rings than for isolated rings, even in the limit of infinite chain length. Extrapolation to $N \rightarrow \infty$ yields $A = 0.2559(3)$ for single rings and $A = 0.2602(3)$ for catenated rings. The result for single rings is in agreement with earlier studies [160, 158]. The prolateness is smaller for catenated rings, but the inset of Figure 8.9 shows that differences remain even in the limit $N \rightarrow \infty$. Linear extrapolation to $N \rightarrow \infty$ yields $P = 0.451(2)$ for isolated rings and $P = 0.416(3)$ for catenated rings. The values for other topologies are displayed in table 8.2.

8.3.3 Ring structure induces entropy-driven segregation

Of major interest concerning the benefit of loop formation in proteins, DNA or chromatin is the influence of the imposed topological constraint on the relative positioning of the two polymers, i.e. the entropic effect of the ring structure. Catenation, for example, sets a constraint on the maximal distance between the centers of mass of the two rings. Intuitively, a large center-of-mass separation strongly restricts the number of accessible conformations and therefore seems entropically unfavorable. A complete mixing of both rings might be possible, which would also agree with the finding that catenated rings are swollen compared to isolated ones. Similar behavior can be expected in case of non-catenated rings bonded to each other.

Information about the segregation of the two ring polymers is given by the average distance between the centers of mass of both rings. Let $R_{gyr} = \sqrt{\left\langle R_{gyr}^2 \right\rangle}$ be the root mean squared radius of gyration of ring 1, measuring the average distance between two monomers on ring 1. If the chains intermingle completely, the average distance between the centers of mass Δ_{CM} given in units of the radius of gyration should be equal to unity. However, if the rings tend to segregate, this quantity should be above unity. We first looked at this quantity with respect to the constraint of catenation both for the case of the Hopf link ($\Phi = 1$) and the case of linking number $\Phi = 2$. We find from Figure 8.10A that the center of mass $\mathbf{r}_{CM}^1$ of ring 1 is indeed positioned more distant than one radius

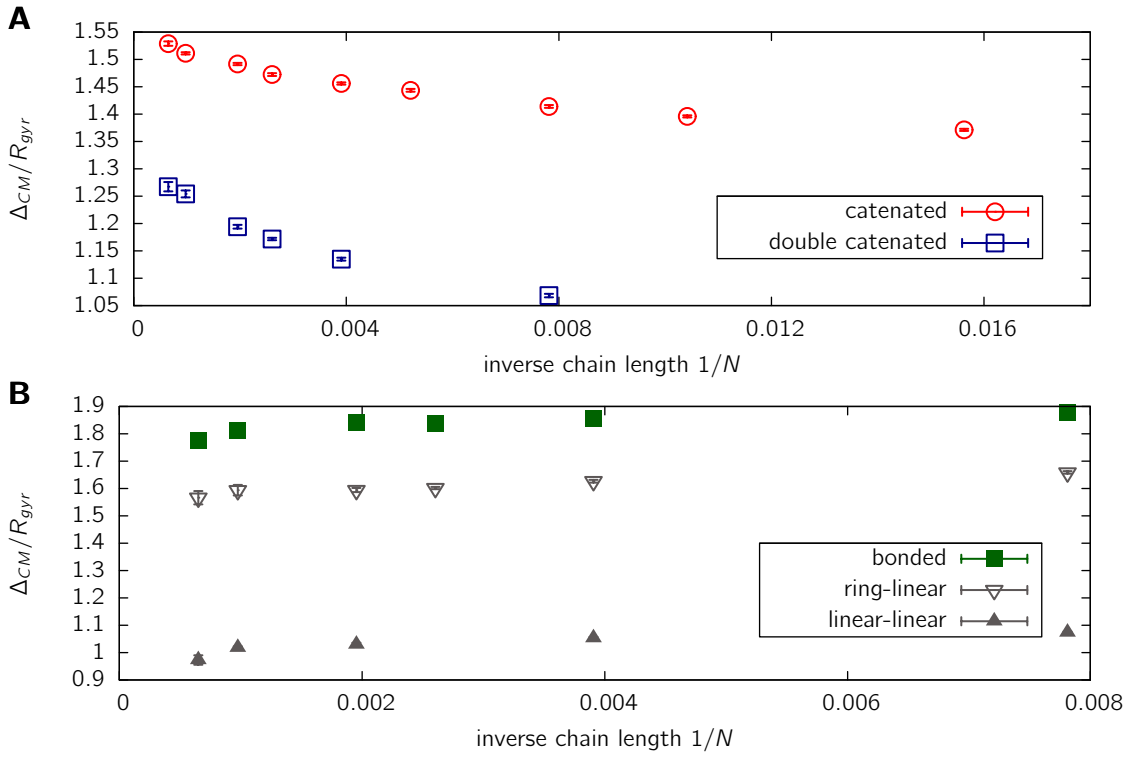


Figure 8.10: Segregation of ring polymers. A. Influence of the catenation constraints with linking number $\Phi = 1, 2$ on the average distance Δ_{CM} between the centers of mass. Shown is data for rings with a single link and a double link versus inverse chain length. The distances are given relative to the radius of gyration $R_{gyr} = \sqrt{\langle R_{gyr}^2 \rangle}$ of one ring. Data shows that the two rings have a tendency to segregate, which becomes stronger for larger chains. **B.** Influence of the ring shape on non-catenated polymers bonded to each other. Shown is the average distance between the centers of mass in relation to inverse chain length for a system consisting of two rings (solid square), a ring bonded to a linear polymer (open triangles) and two linear chains(solid triangles). Data is scaled with the radius of gyration of one ring (or one linear chain in case of the linear-linear system). Segregation is markedly stronger for the system consisting of two rings.

of gyration R_{gyr} from that of ring 2. Thus, the two rings are not completely overlapping but segregate to some extend, the tendency becoming even stronger the larger the chain length N. The question of whether the normalized distance approaches a constant value or diverges to infinity cannot be answered from the data. Increasing the linking number between two ring polymers forces them to be closer together, thus double catenated rings display a smaller center-of-mass separation in relation to size (Fig. 8.10A), which however shows a steep ascent in the long chain limit. These findings can well be explained by fixed catenation topology: Consider two phantom ring polymers, which are brought in close proximity. The probability of having a large linking number increases with chain length. Thus, the probability for linking number 1 and 2 decreases, making short center-of-mass separations unlikely in the ensemble constrained to simple catenation or double catenation.

The biological problem of chromatin folding and the effect of chromatin loops on segregation can be investigated most clearly by looking at the non-catenated bonded topology (Fig. 8.1D). In order to highlight the effect of loops, we compare this system to both the case of a linear polymer bonded to a ring (Fig. 8.1E) and two linear polymers bonded to each other at the central monomer (Fig. 8.1F). The center-of-mass separation is shown in

Fig. 8.10B. As in the case of catenated rings, the center-of-mass distance is scaled with the radius of gyration R_{gyr} of one single ring, in case of the linear-linear system with the radius of gyration of one linear chain. We find that the segregation of the ring-ring system is much stronger than the ring-linear and linear-linear system. The two last mentioned systems do not have the non-catenation constraints, i.e. the accessible configurational space is larger at short center-of-mass separations. In fact, the linear-linear system shows almost complete intermingling, the center-of-mass separation being close to the radius of gyration of a single chain. A general conclusion emerging from these results is that topological constraints imposed by looping, e.g. the non-catenation constraint, play a dominant role in driving the segregation of close-by loops. Thus, a kind of order is induced in the system, which cannot be accomplished by linear polymers.

By projecting the monomers to the line connecting the centers of mass we can determine the density distribution of monomers of both chains along this line, displaying the degree of intermingling or overlap. Figure 8.11A shows the projected line density of monomers for a system of two catenated ring polymers along the vector connecting the centers of mass. The data is evaluated in units of the center of mass distance, the origin being halfway between the centers of mass. The figure shows that the polymer rings are indeed well separated. Interestingly, as the curves for $N = 128$ and $N = 1024$ show, the overlap area decreases with chain length. This overlap area of the two curves in Figure 8.11A is plotted against the chain length in Figure 8.11B for the catenated systems studied. We find that the overlap area decreases with N and reaches a constant value in the limit of $N \to \infty$ both for rings with one or two links.

Importantly, such a separation of density clouds is mediated by the looping. This can be seen most clearly comparing two bonded rings with the corresponding ring-linear and linear-linear systems (Fig. 8.11C). We find that the overlap area of the monomer density distributions is significantly smaller in case of two bonded ring polymers compared to systems with linear polymers.

8.3.4 Alignment of ring polymers

Of interest concerning the segregation of monomers is not only the center-of-mass distance, but also the alignment of the two rings with respect to each other. Again, we turn to the gyration ellipsoid, which yields direct information about the distribution of masses in space. Especially, we study the orientation of the longest principal axes of the gyration tensors with respect to each other. For this purpose we analyze the average angle $\langle \cos\theta \rangle$ between these axes. As they are bi-directional, they force the angle θ to be in the interval $[0, \pi/2]$. We find that the gyration tensors are not independently oriented with respect to each other, which would correspond to an average angle between the main axes of $\langle \cos\theta \rangle = 0.5$. In fact, the gyration ellipsoids tend to collocate in a more perpendicular orientation (Figure 8.12) for all two-ring topologies studied. Linear extrapolation to $N \to \infty$ yields $\langle \cos\theta \rangle = 0.466(1)$ ($\theta \approx 62.2°$) for catenated rings, $\langle \cos\theta \rangle = 0.477(1)$ ($\theta \approx 61.5°$) for double catenated rings and $\langle \cos\theta \rangle = 0.491(2)$ ($\theta \approx 60.6°$) for bonded rings. Intuitively, one could expect that if there is a repulsion between the two rings due to excluded volume or topological constraints, the gyration ellipsoids would try to align in the same orientation and separate their centers of mass such that the gyration ellipsoids do not overlap. However, it seems that this is entropically very unfavorable due to the catenation or bonding constraint, which prohibits a large separation of the centers of mass. Therefore, if there is an entropic barrier preventing segregation and forcing the gyration tensors to overlap, the perpendicular orientation minimizes the overlap area. This is

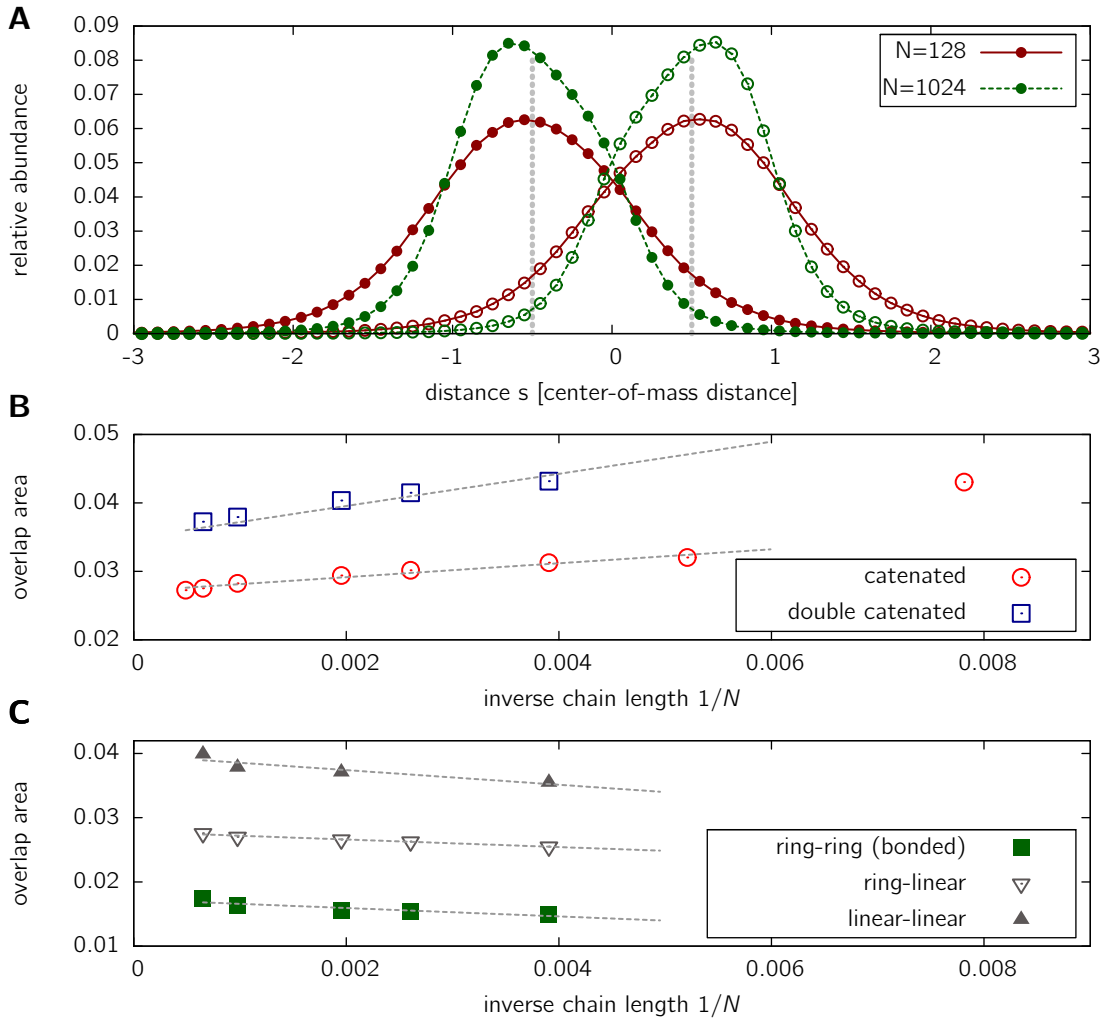


Figure 8.11: Topology and chain length dependent overlap of rings.A. Distributions of monomers projected onto the line connecting the centers of mass of two catenated rings (linking number 1). The data shows the relative abundance of projected monomer positions of both rings with respect to the axis between the centers of mass. The scale on the x-axis is given in units of the center-of-mass distance. The origin corresponds to the point in between the centers of mass, the grey vertical lines represent the positions of the centers of mass. Data shows that monomer clouds of both rings are rather separated. **B**. This figure shows the overlap area in the graphs of the projected monomer distributions as displayed in A for catenated (linking number 1) and double catenated rings (linking number 2). This overlap area decreases approximately linear with inverse chain length reaching asymptotically a constant value. **C.** Overlap area of projected monomer distributions for a system of two non-catenated polymers which at their centers are bonded. The existence of rings markedly decreases the overlap.

exactly what we observe. Furthermore, the more perpendicular orientation makes the complete system more sphere-shaped and thus more symmetric. In agreement with this is our finding that the angle θ_{CM} between the vector connecting the centers of mass and the largest principal axis of the gyration tensor of a ring is slightly smaller than expected from a random orientation of both vectors. Extrapolation of $\langle \cos \theta_{\mathrm{CM}} \rangle$ yields values of $0.580(1), 0.586(1)$ and $0.566(1)$ for simple catenated, double catenated and bonded rings, respectively. Figure 8.12B shows a typical conformation out of the ensemble of simple

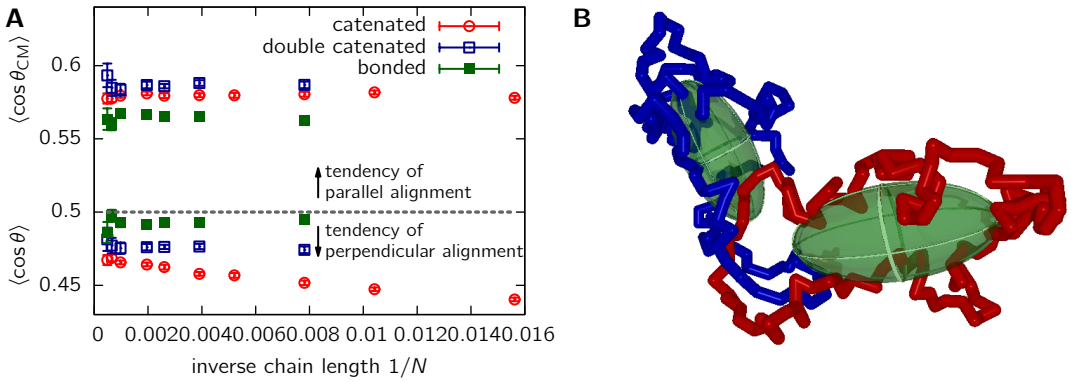


Figure 8.12: Alignment of the gyration ellipsoids with respect to each other. A. Data at the bottom displays the average angle $\langle \cos\theta \rangle$ between the two largest principal axes of the gyration ellipsoid. Data at the top represents the average angle $\langle \cos\theta_{\mathrm{CM}} \rangle$ between the largest principal axis of one ring and the vector connecting the centers of mass. The grey line corresponds to the average orientation in the case of both vectors having a random orientation. The gyration ellipsoids of all topologies studied align on average more perpendicularly than expected by random orientation, while their alignment w.r.t the center of mass connecting vector tends to be parallel **B.** One conformation consisting of two catenated rings and its gyration ellipsoids drawn out of the ensemble. The example conformation visualizes (i) the segregation of the two rings (the ellipsoids are well separated), (ii) the aspherical shape of the gyration ellipsoids and (iii) their perpendicular alignment.

catenated two-ring conformations. Both the alignment of the gyration ellipsoids as well as the aspherical shape are visible.

To analyze the orientations of the gyration tensor in more detail, we investigated its orientation in dependence of the center of mass separation. The average angle $\langle \cos\theta \rangle$ between the two largest principal axes of the gyration ellipsoids is shown in Figure 8.13 for simple and double catenated rings. The closer the centers of mass are the more perpendicular the ellipsoids. For large separations, the orientation approaches a parallel alignment: The centers of mass are separated to such an extend that the gyration tensors have to align in a row in the same direction due to the catenation constraint. For double catenated rings, the tendency is not so clear due to insufficient statistics: a linking number of two becomes very unlikely for large separations. Simple catenated rings display a point at intermediate distances $r \approx 2R_g$ where the two driving forces – minimization of overlap and catenation constraint – are in balance such that the average angle is close to the random orientation value.

8.4 Conclusions

In this chapter we have studied the influence of topological catenation and bonding constraints on the conformational properties of ring polymers using three different topologies: simple catenated (linking number 1), double catenated (linking number 2) and non-catenated bonded rings. In particular, we were interested in the question of whether the dimensions and the shape change compared to isolated rings. More importantly, we investigated the positioning and alignment of the two rings with respect to each other to detect whether there is an entropic repulsion which can lead to a segregation of the polymers.

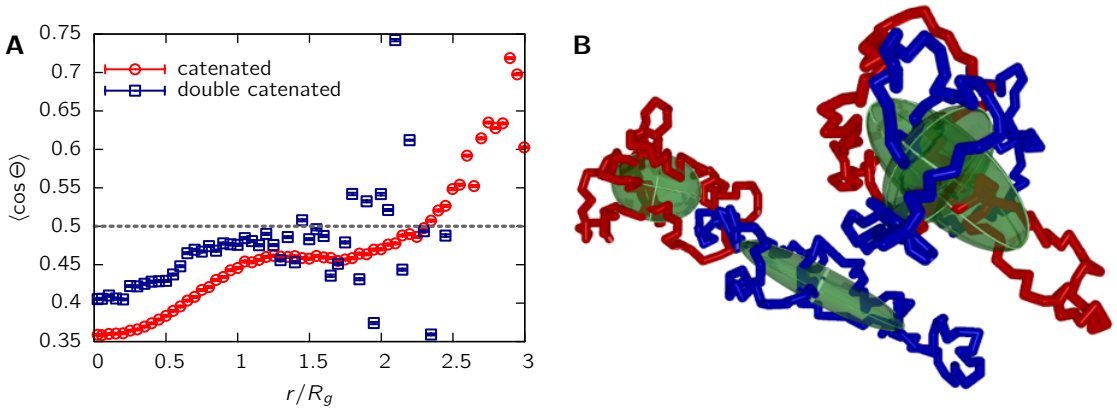


Figure 8.13: Orientation of the gyration ellipsoids in dependence of CM separation. **A.** Shown is the average angle $\langle \cos\theta \rangle$ between the largest principal axes of the gyration ellipsoid of both simple catenated rings (linking number 1) and double catenated rings (linking number 2) in dependence of their center-of-mass distance. The chain length is $N = 1024$. At small separations, the alignment is preferably perpendicular, while for large separations, alignment becomes more and more parallel. **B.** Two conformations with different center-of-mass separation and simple catenation. The left catenated rings have a CM separation of $r = 3.11R_g$, the right ones of $r = 0.13R_g$.

We find that the size of a ring topologically constrained to another ring is more swollen compared to isolated rings, the swelling factor in the infinite chain limit $N \to \infty$ being about 1.10 for catenanes and 1.03 for bonded rings (Figure 8.6). However no indication was found in the regime of chain lengths studied here that the scaling exponents are different from the case isolated rings, where the scaling exponent is close to the self-avoiding walk value of $\nu \approx 0.588$.

While dimensions of the ring polymers only change by a constant factor, the effective bond length b, the shape of catenated or bonded rings differs substantially, a finding which is in contrast to the conjecture put forward by Sikorski [154]. This is shown for the ratios of the gyration tensors eigenvalues (Figure 8.8). These ratios are generally smaller for rings compared to their linear counterparts, the shape being still markedly prolate. However, the prolongation is stronger for a catenated or bonded ring than for isolated rings (table 8.2). This difference also shows up in the measures of asphericity A and prolateness P, which are commonly used shape parameters [158].

The change in shape of one ring in the system of two topologically constrained rings already indicates an entropic effect of the fixed topology. We have analyzed whether rings tend to segregate, arranging themselves into distinct spatial territories, or if they can intermingle freely. We found that for catenanes with small linking numbers, there is a strong tendency towards segregation, which shows up in the separation of the centers of mass compared to the radius of gyration of a single chain (Figure 8.10A) and in the density distribution of monomers along the center of mass (Figure 8.11A). The segregation is even more pronounced for longer chains driven by the tendency to create more complicated links at short separations. Interestingly, non-catenated bonded rings show a markedly stronger segregation than bonded linear chains, revealing the importance of the ring closure on segregation. These effects also become visible in the alignment of the gyration ellipsoids. We found a tendency to align more perpendicular than expected from a random orientation (8.12). Presumably, such an orientation minimizes the overlap

between the monomers of different chains and bypasses the prohibited formation of more complicated links. As a large center-of-mass separation is entropically unfavorable and therefore the chains have to stay quite close, the perpendicular alignment has a smaller overlap volume than a parallel alignment.

Our findings have several biological implications for catenated DNA [151], proteins [152] and chromatin folding [12]. Generally speaking, simple and double catenated polymers adopt a more aligned and segregated structure. Ring polymers in proximity – which is achieved here by creating a bond between them – have a strong tendency to segregate, inducing a kind of order in the system which cannot be accomplished by linear polymers. It is well-known that compartmentalization is important for maintaining life. Both the plasma membrane as well as the nuclear membrane provide such compartments. The tendency of two ring polymers to minimize the overlap volume gives a simple explanation for experimental findings showing that whole chromosomes as well as regions within chromosomes of eukaryotic cells segregate into distinct compartments of the nucleus [6]. While it has been shown experimentally that chromatin loops are an ubiquitous feature of chromatin organization, this study gives evidence that loops not only play a dominant role in transcriptional regulation but are also key factors in keeping chromatin domains and whole chromosomes segregated by means of entropic and topological interactions.

The findings of this study have several biological implications for catenated DNA [151] and proteins [152]. Similar effects might be expected for chromatin folding, although the chromatin fiber has a much more complex topology [117, 12]. Generally speaking, simple and double catenated polymers adopt a more aligned and segregated structure. Ring polymers in proximity – which is achieved here by creating a bond between them – have a strong tendency to segregate, inducing a kind of order in the system which cannot be accomplished by linear polymers. It is well-known that compartmentalization is important for maintaining life in higher eukaryotes, both the plasma membrane as well as the nuclear membrane providing such compartments. Although the detailed topology of chromatin is not known, chromatin loops have been shown to play a dominant role in transcriptional regulation [20]. Several studies have indicated a strong effect of looping on the observed segregation of chromosomes, partly studied by a model of ring polymers [35, 161, 162]. While we would expect the influence of looping to become stronger in presence of multiple loops as found in transcription factories [70], it remains to be shown to what extend the segregation and alignment of ring polymers observed in the present study also apply to a more complex model of chromatin. Expecting that the effects are even stronger in the presence of several loops, the results would imply that loops not only play a dominant role in transcriptional regulation but are also key factors in keeping chromatin domains and whole chromosomes segregated by means of entropic and topological interactions.

Chapter 9

Topological interactions between ring polymers

Implications for chromatin loops

References

The results presented in this chapter are published in:

- M. Bohn & D. W. Heermann, **2010**. Topological interactions between ring polymers: Implications for chromatin loops. *J. Chem. Phys.*, 132 (4), 044904.

Chapter Summary

Chromatin looping is a major epigenetic regulatory mechanism in higher eukaryotes. Besides its role in transcriptional regulation, chromatin loops have been proposed to play a pivotal role in the segregation of entire chromosomes. The detailed topological and entropic forces between loops still remain elusive. Here, we quantitatively determine the potential of mean force between the centers of mass of two ring polymers, i.e. loops. We find that the transition from a linear to a ring polymer induces a strong increase in the entropic repulsion between these two polymers. On top, topological interactions such as the non-catenation constraint further reduce the number of accessible conformations of close-by ring polymers by about 50%, resulting in an additional effective repulsion. Furthermore, the transition from linear to ring polymers displays changes in the conformational and structural properties of the system. In fact, ring polymers adopt a markedly more ordered and aligned state than linear ones. The forces and accompanying changes in shape and alignment between ring polymers suggest an important regulatory function of such a topology in biopolymers. We conjecture that dynamic loop formation in chromatin might act as a versatile control mechanism regulating and maintaining different local states of compaction and order.

9.1 Introduction

In chapter 8 we have investigated the conformational properties of a ring polymer which is either catenated or bonded to another one. This system served us as a toy model to study the effects of two rings being held close together by means of a topological constraint. Among others, we found that such a system displays a pronounced segregation due to the ring structure. Here, we want to conduct a quantitative analysis of the repulsive forces acting between two ring polymers brought in close proximity.

The experimental observation of chromosome segregation has been attributed to different mechanisms. In principle, active ATP-consuming mechanisms might be responsible for maintaining segregation. Rosa and Everaers [30] proposed that chromosomes are aggregating into distinct compartments because of their large relaxation time, making it impossible for them to intermingle during the time of one cell cycle. However, this study does not explain segregation of smaller regions within chromosomes as found by Goetze and co-workers [6]. Vettorel *et al.* [163] suggested that the collapsed state might be responsible for segregation: Polymer gels, where single polymers are in a collapsed state, show no reptation, therefore intermingling between different polymers becomes practically impossible. Recently, evidence from polymer models has indicated that chromatin loops force such a compartmentalization already by virtue of purely entropic forces [35, 93].

While the mean squared radius of gyration $\left\langle R_g^2 \right\rangle$ of isolated ring polymers and a system of two catenated rings [164] displays a scaling similar to that of a self-avoiding walk, where $\left\langle R_g^2 \right\rangle \sim N^{2\nu}$ with $\nu \approx 0.588$ [89], non-catenated rings in a melt become compact with a scaling exponent of $\nu = 1/3$ [147, 149, 163]. In fact, chromatin organization is not consistent with a simple melt of ring polymers. Scaling exponents found in experiments differ a lot, ranging from $\nu \approx 0.1 - 0.2$ [109] to $\nu = 0.5$ [92], the level of compaction

depending on gene activity [12]. Such a behaviour, howere, could well be explained by different looping probabilities in the framework of the Random Loop model (↷ chapter 7).

The purpose of this study is to deepen the understanding of the topological interactions ring polymers exert on each other as well as the changes in their conformational properties induced by such forces. While topological effects of rings have been considered in several studies [165, 154, 147], the forces generated have not been quantitatively assessed. A system of two ring polymers can be viewed as a toy system for the influence of loop formation on chromatin folding. While the biological system is beyond doubt much more complicated – multiple loops and the dense system of polymers playing a dominant role – the effect of topological forces and the advantage of the ring shape for biopolymers can be highlighted best by studying it apart from other influences. We evaluate the potential of mean force between two ring polymers dependent on their mutual topological state to derive a quantitative measure for the entropic repulsion two rings or loops exert on each other. Therefore, Monte Carlo simulations of ring polymers with lengths of up to $N = 2048$ monomers are conducted. We separate the entropic effects due to excluded volume interactions from the topological interactions arising from the non-catenation constraint. The ring polymers' conformational properties and their mutual alignment subjected to the topological and entropic forces are investigated. In fact, we find that ring polymers display a markedly stronger repulsion and adopt a more ordered state compared to linear polymers. The findings indicate a natural benefit of rings over linear ones, leading to the conjecture that chromatin loops not only facilitate entropy-driven segregation of chromosomes or intrachromosomal regions, but act as a versatile control mechanism regulating and maintaining different states of compaction and order.

9.2 Simulations and methods

9.2.1 Effective potentials

The principle aim of this study is to investigate the potential of mean force exerted between the centers of mass of ring polymers, partly under the constraint of non-catenation. Therefore we introduce the concept of the effective potential [166], which will be outlined in the following paragraphs.

Consider two polymer chains. We want to measure the force which is needed to keep the centers of mass of the two polymer chains at a certain distance $\mathbf{r}_0$. This force could in principle be measured by installing a virtual spring with equilibrium elongation r_0 connecting the centers of mass $\mathbf{R}_1$ and $\mathbf{R}_2$ of both polymers. The force acting between the centers of mass can then be obtained from the average elongation of the spring.

A much more efficient way to calculate the forces between polymers is accomplished via the notion of effective interactions. The method has been used for a variety of polymer systems, and for some even analytical approximations have been found [167, 168]. A recommendable review is given in Ref. [166]. Here, we only present a short overview of the underlying statistical mechanics as well as the actual algorithm applied in this study.

Let the monomer positions and momenta of the two polymers be denoted by $\mathbf{r}_i^\alpha$ and $\mathbf{p}_i^\alpha$ ($i = 1, \ldots, N, \alpha = 1, 2$), respectively. α denotes the polymer chain index and i indexes subsequently the monomers of the polymer. The properties of the system are given by the complete Hamiltonian

$$\mathcal{H}(\{\mathbf{p}_i^\alpha\}, \{\mathbf{r}_i^\alpha\}) = \sum_{\alpha=1,2} \sum_{i=1}^{N} \frac{\mathbf{p}_i^{\alpha 2}}{2m} + V(\{\mathbf{r}_i^\alpha\})\,. \tag{9.1}$$

For the problem of ring polymers investigated here, the interaction potential $V(\{\mathbf{r}_i^\alpha\})$ is given by an excluded volume term and eventually by a term describing the topological constraints.

The partition sum of the complete system is then given by

$$Z = \mathrm{Tr}\left[\exp(-\beta H)\right] = \int \cdots \int \prod_{\alpha=1,2} \prod_{i=1}^{N} d\mathbf{r}_i^\alpha d\mathbf{p}_i^\alpha \exp\left[-\beta \mathcal{H}\right] . \tag{9.2}$$

How do we obtain the potential of mean force between the centers of mass? We have to find an expression independent of the single monomer coordinates $\mathbf{r}_i$ of the system. The way of eliminating degrees of freedom mathematically is to trace out those degrees of freedom one is not interested in – here the monomer coordinates $\mathbf{r}_i^\alpha$ –, resulting in the constrained partition sum with a fixed center of mass (CM) separation $\mathbf{R}$,

$$Z_c(\mathbf{R}) = \mathrm{Tr}\left[\exp(-\beta\mathcal{H})\ \delta\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{r}_i^{(1)} - \frac{1}{N}\sum_{i=1}^{N}\mathbf{r}_i^{(2)} - \mathbf{R}\right)\right] . \tag{9.3}$$

The trace denotes the integral over all momenta and monomer coordinates. $Z_c(\mathbf{R})$ represents the partition sum of the monomers in an external field generated by the constraint of keeping the center-of-mass distance fixed to $\mathbf{R}$.

The effective potential is now defined by the relation

$$\exp\left(-\beta U_{\mathrm{eff}}(\mathbf{R})\right) = Z_c(\mathbf{R}) \tag{9.4}$$

The definition of the effective potential $U_{\mathrm{eff}}(\mathbf{R})$ is such that the non-constrained partition sum is regained by a trace over the remaining degrees of freedom.

For purposes of calculation, however, the following representation [166, 168] of the effective potential is more useful

$$U_{\mathrm{eff}}(R) = -k_B T \ln \frac{Z_c(R)}{Z_1^2} \tag{9.5}$$

where Z_1 is the partition sum of a single polymer. Equivalently, Z_1^2 may be imagined as the partiton sum of a two-chain system whose centers of mass are infinitely far apart.

9.2.2 Monte-Carlo algorithm

The results presented in this chapter use simulations of isolated ring polymers from chapter 8 with chain lengths ranging from $N = 32$ to $N = 2048$. For details on the Monte Carlo algorithm we refer to section 8.2.

9.2.3 Calculation of the effective potential

Single ring conformations are used to calculate the average interaction energy between two polymers by means of equation (9.5). This procedure is well-established and has been applied in several studies [167, 168]. At first, the set of sampled conformations $\mathcal{C}$ is split into two parts of equal size. Then, we randomly pick two conformations, one from each subset. These two ring conformations are then positioned such that their centers of mass have a certain distance r, the angular positioning in doing so is chosen randomly. In order to ensure all monomers residing on lattice sites after shifting, a maximum distance inaccuracy of $\delta r \approx 0.87$ lattice units is accepted. In the next step, the interaction energy $E_i(r)$ of the

two-ring conformation and the corresponding statistical weight $W_i(r) = \exp(-E_i/k_BT)$ are calculated.

Repeating this operation for a number K of randomly selected pairs of conformations, the effective potential at a center-of-mass distance r is then given by

$$U_{\text{eff}}(r) = -k_BT \ln \frac{\sum_{i=1}^{K} W_i(r)}{K} \tag{9.6}$$

The standard error of the effective potential is determined by randomly subdividing the K two-chain conformations into $M = 50$ smaller subsets and calculating the effective potential $U_m(r)$ for these subsets [167]. The standard error is then calculated by

$$\Delta U_{\text{eff}}(r) = \sqrt{\frac{1}{M} \sum_{m=1}^{M} (U_m(r) - U_{\text{eff}}(r))^2} \tag{9.7}$$

In a first step, only taking into account excluded volume interactions, the statistical weight $W_i(r)$ of a two-chain conformation is set to $W_i = 0$ in case the excluded volume condition is violated, i.e. two monomers occupy the same lattice site, and $W_i = 1$ in case it is not.

However, care has to be taken using the bond fluctuation model. During the simulation run it is ensured that no bond crossings can occur; this, however, is not guaranteed if we just shift two conformations inside each other. There are two possible combinations of bond vectors, where the excluded volume condition is not violated, but where the bonds come into contact. Consider the two bonds spanned by the following four monomers $\{(0,0,0) \to (3,1,0), (1,2,0) \to (2,-1,0)\}$. Both bond vectors $(3,1,0) \in \mathcal{B}$ and $(1,-3,0) \in \mathcal{B}$ are valid and the excluded volume condition is satisfied. The second problematic case is $\{(0,0,1) \to (2,2,0), (0,2,0) \to (2,0,1)\}$. As can be easily seen, the participating beads are stuck in these positions, there is no valid move for any of the beads using the restrictions of bond vectors the algorithm is subjected to. Therefore, during a simulation run such a situation can never happen, assuming the start configuration is chosen properly. Here, we test each two-chain conformation, after positioning them with their centers of mass a given distance r apart, on such a situation. If a bond crossing occurs, a weight of $W_i = 0$ is assigned to this conformation.

In this study, we are furthermore interested in the topological state of a two-ring conformation. We calculate the weight factor differently whether we consider a preserved topological state or not. In the case that we force a certain topological state, we have to add another criterion for accepting a two-chain conformation (i.e. assigning a weight $W_i \neq 0$). Two rings that are positioned at a center-of-mass distance r might be non-catenated or catenated. Furthermore, the degree of catenation might vary (see illustrations in Fig. 8.1B and C). We determine the topological state by a topological invariant: the "Gauss linking number" [153]. This invariant has been used in analytical studies on interlinked rings to keep track of a certain topology [153, 169]. It is defined by

$$\Phi(C_1, C_2) = \frac{1}{4\pi} \oint_{C_1} \oint_{C_2} \frac{\langle d\mathbf{r}_1 \times d\mathbf{r}_2, \mathbf{r}_1 - \mathbf{r}_2 \rangle}{|\mathbf{r}_1 - \mathbf{r}_2|^3} \tag{9.8}$$

The closed line integrals are evaluated along the contours of the two rings, denoted by C_k $(k = 1, 2)$. The vector function $\mathbf{r}_k = \mathbf{r}_k(s)$ denotes the three-dimensional coordinates of the ring polymers, parameterized e.g. by the contour length s. Roughly speaking, this integral counts the number of times ring 1 passes through the surface created by the closed path of ring 2.

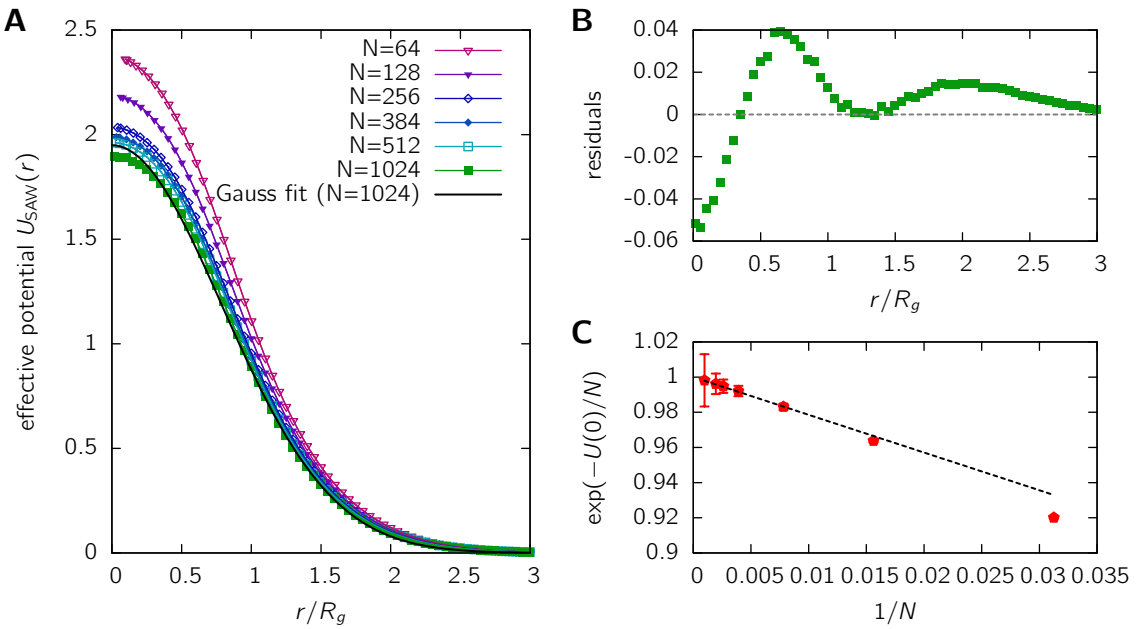


Figure 9.1: Effective potential acting between the centers of mass of two self-avoiding walks. **A**. This panel displays the effective potential $U_{\mathrm{SAW}}(r)$ for different chain lengths. The x-axis is scaled by the root mean squared radius of gyration R_g of isolated chains. Errors are smaller than the point size and therefore not shown. The potential approximately has a Gaussian shape, as expected by mean-field theory. The fit for $N = 1024$ is done with one adjustable parameter, the amplitude. Residuals of the fit are shown in **B** and display small systematic deviations. **C**. Scaling plot of the effective potential at full overlap. The effective potential $U_{\mathrm{SAW}}(0)$ reaches a finite value for large chain length as expected from scaling theory (cf. equation (9.10)).

The Gauss link invariant has the advantage that it is easy to evaluate, although it has the disadvantage that it is not one-to-one. The "Whitehead link" [153] is one example for a two-ring conformation having the same linking number as non-catenated rings. The result of the integral for the Gauss link invariant is an integer for non-overlapping chains, especially $\Phi = 0$ for non-catenated (unlinked) chains and $\Phi = 1$ for simple catenated chains (Fig. 8.1B). We numerically evaluate this integral to classify the topological state of a two-ring conformation and assign a weight $W_i = 0$ for all two-ring conformation not in the topological state we are interested in. Thereby we are able to calculate the effective potential separately for any given topological state.

9.2.4 The effective potential of self-avoiding walks

We first recapitulate theoretical results for the effective potential between the centers of mass of two self-avoiding walks. The study of self-avoiding walks allows an easy-to-understand introduction to the mean field arguments used to approximate effective potentials of polymer coils. Moreover, the algorithm used in this study can be validated by comparison to results from other publications. Furthermore, we will apply the results to relate the strength and form of the potential for linear polymers to their ring counterparts.

A simple mean field argument can be devised for the entropic repulsion of two polymer coils with excluded volume at full overlap, i.e. with zero center-of-mass separation. It has been first proposed by Flory and Krickbaum [170] and later been corrected in other studies [171, 172]. Consider two polymer coils which are completely overlapping, such

that the overlap volume is $V \sim R_g^3$ where R_g is the radius of gyration of one polymer coil. Following the arguments above, the effective potential is given by the logarithm of the fraction of accepted two-chain conformations. The probability that both coils touch each other can be estimated in the mean field picture as follows: Assume that the monomers of both polymer coils are randomly smeared out over the volume V. Consider an arbitrary monomer, say with index k, of chain A. The probability that it is in contact with any monomer from chain B is given by $p_1 = \frac{Na^3}{V} = na^3$, where a^3 is the volume occupied by one monomer. The probability that chain A does not interfere with chain B is then given by $p = (1 - p_1)^N = (1 - na^3)^N$ in this mean field picture. Thus, the effective potential (setting $k_BT = 1$) at full overlap is estimated to

$$U_{\mathrm{FK}}(r = 0) = -N \ln(1 - na^3) = -N \ln(1 - bN^{1-3\nu}) \tag{9.9}$$

In fact, this simple argument is wrong [172] for the reason that monomers are not distributed randomly throughout the volume. Rather the connectivity of the chains induces a depletion of monomers from chain B around one monomer from chain A. Scaling arguments [171] show that the probability p_1 of one monomer of chain A being in contact with any monomer of B is given by $p_1 = (na^3)^{\frac{1}{3\nu-1}}$, thus $p = (1 - p_1)^N = (1 - \frac{b'}{N})^N$ and

$$U(r = 0) = -N \ln\left(1 - \frac{b'}{N}\right) \tag{9.10}$$

In the limit of infinite chain $N \to \infty$, the effective potential at full overlap approaches a constant value in the order of k_BT. Consequently, linear polymer coils with excluded volume have a rather soft potential, allowing for a high level of mutual penetration.

As early as 1949, Flory and Krickbaum [170] devised a mean field argument for the effective potential at any separation r. We briefly review this argument here adjusted to our simulations. One major assumption is that the density distribution of monomers from chain i around its center of mass $\mathbf{q}_i$ is given by some Gaussian distribution

$$\rho_i(\mathbf{r}) = N \left(\frac{d^2}{\pi}\right)^{\frac{3}{2}} \exp\left[-d^2 \left(\mathbf{r} - \mathbf{q}_i\right)^2\right] \tag{9.11}$$

The effective potential for a fixed center of mass distance r in an athermal solvent is then given by

$$U(r) = k_BTN^2 \left(\frac{d^2}{2\pi}\right)^{\frac{3}{2}} v \exp\left(-\frac{d^2}{2} r^2\right) \tag{9.12}$$

In the following we set $k_BT = 1$ as our simulations are athermal. For a Gaussian chain without excluded volume, $d^2 = \frac{3}{2R_g^2}$, a value used both in Refs. [170, 173]. For chains with excluded volume, this might not be the best estimate, thus we use the value of d obtained by a fit to the density distribution. Of course, this approximation suffers from the same short-comings as the approximation for the effective potential at full overlap: The polymeric coil is thought of as a cloud of independent monomers without connectivity and thus the additional repulsion by the connectivity is not taken into account.

Results for the self-avoiding walk are displayed in Fig. 9.1. The effective potential $U_{\mathrm{SAW}}(r)$ in Fig. 9.1A reveals a decrease with growing chain length, inconsistent with the Flory approximation in eq. (9.9) but consistent with eq. (9.10). For a chain length of $N = 1024$, the data is fitted to a Gaussian distribution. The parameter d in eq. (9.12) is obtained by a one-parameter fit to the monomer distribution around the center of mass

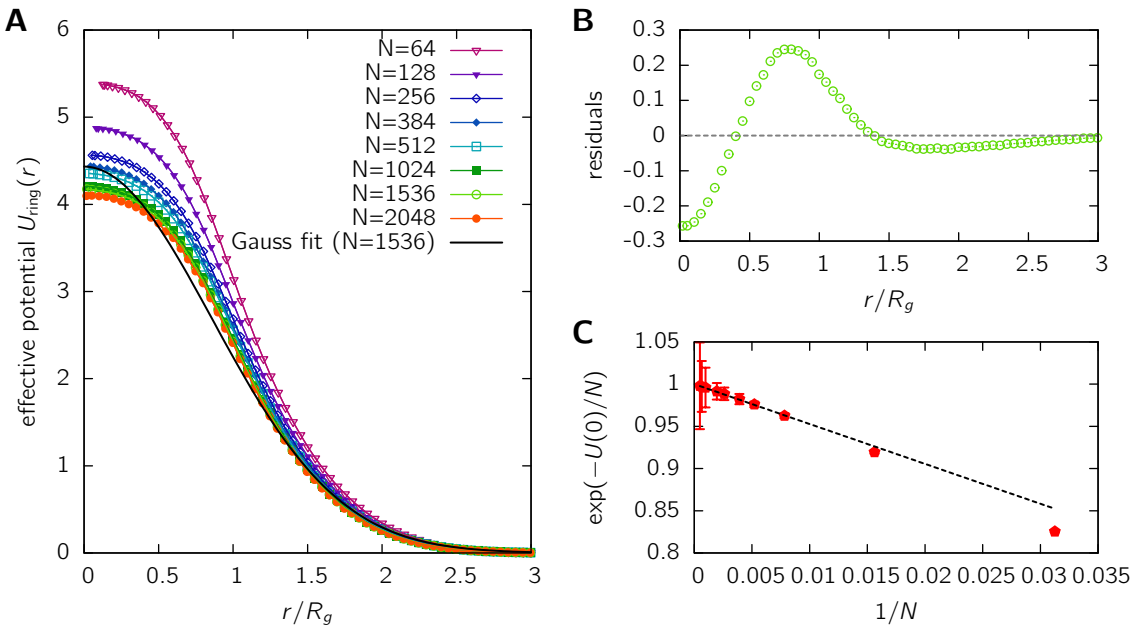


Figure 9.2: Effective potential $U_{\text{ring}}(r)$ of ring polymers disregarding their mutual topological state. **A.** This panel displays the effective potential in units of k_BT for several chain lengths. Errorbars are not shown as they are smaller than the point size. A one-parameter Gaussian fit for $N = 1536$ is performed, showing that predictions of Flory-Krickbaum theory are wrong for ring polymers. Residuals indicate (**B**) that the fit is much worse than the one for linear chains (Fig. 9.1). **C.** The scaling law for the effective potential at full overlap follows the theory from eq. (9.10) allowing for a linear extrapolation to the infinite chain limit when plotting $e^{-U(0)/N}$ vs $1/N$. For the extrapolation, only chain lengths larger than $N = 256$ are used.

(see Fig. 9.3) and the only fit parameter left is the amplitude. The Gaussian distribution is a good approximation, however, the residuals (Fig. 9.1B) of the fit display small systematic deviations. The effective potential at full overlap is shown in a scaling plot (cf. Ref. [173]) in Fig. 9.1C. Data points are on a straight line as expected, deviations are most likely due to lattice effects: When the two polymer coils are brought to a certain distance, the nearest lattice site is chosen, such that the distance can deviate from the target distance $r = 0$. These deviations are more pronounced the smaller N.

9.2.5 Rings have a stronger repulsion than linear polymers

Having studied the effective potential of linear self-avoiding walk polymers, we now turn to the main subject of our investigation: ring polymers. First, we neglect topological constraints between the two chains when bringing them in proximity. However, each single ring obeys the topological constraint of unknottedness by virtue of the simulational method. Thus the effective potential is given by the logarithm of the probability that the ring conformations do not occupy at least one lattice site in common. In this case, mean field theory predictions for ring polymers are the same as for linear polymers, as the connectivity of the chains is neglected.

Figure 9.2A shows the effective potential $U_{\text{ring}}(r)$ for ring polymers without topological constraints. Fitting the data to a Gaussian (solid black line, fit to $N = 1536$) yields more pronounced deviations from the model curve than observed for a self-avoiding walk. Again a one-parameter fit has been performed where the width of the distribution is obtained

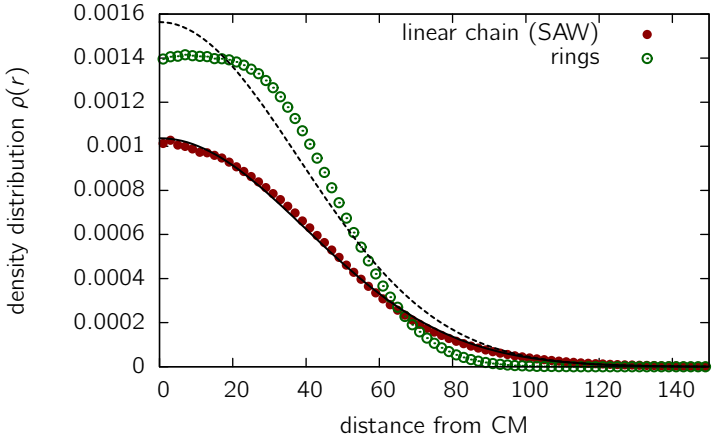


Figure 9.3: The density distribution of monomers around the center of mass for linear chains (SAW, solid circles) and ring polymers (open circles) of chain length $N = 1024$. A fit to a Gaussian distribution $\rho(r) = 4\pi r^2 (d^2/\pi)^{3/2} \exp(-d^2 r^2)$ (see also eq. (9.11)) is reasonable for linear polymers (solid black line), but not for ring polymers (dashed black line).

from the monomer density distribution around the center of mass $\rho(\mathbf{r})$ (eq. (9.11)). Residuals of the fit are displayed in Fig. 9.2B and are larger by a factor of about 5 compared to the self-avoiding walk. Thus we find that the mean field argument of Flory fails at some point in the description of ring polymers. One reason is that the assumption of a Gaussian density distribution $\rho(\mathbf{r})$ is wrong both for linear polymers with excluded volume as well as ring polymers. The density distribution for $N = 1024$ is shown in Fig. 9.3. While the Gaussian approximation is reasonable for linear polymers, it fails strongly for ring polymers. In agreement with the assumption that the underlying density distribution is the reason for deviations from mean field theory, the effective potential at full overlap (Fig. 9.2C) still follows the scaling law in eq. (9.10) from scaling theory. In its derivation the contact probability was estimated by scaling theory without using the assumption of a Gaussian monomer distribution.

Interestingly, the repulsive interactions between two ring polymers are much stronger than for their linear counterparts of equal chain length N. The ratio $U_{\mathrm{ring}}(r)/U_{\mathrm{SAW}}(r)$ is displayed in Fig. 9.4 and has a value of about 3 for center of mass (CM) separations below $2R_g$.

9.2.6 Topological constraints induce additional repulsion

In the last section we have shown that the repulsive forces acting between two ring polymers whose centers of mass are brought in close proximity are much stronger than for linear

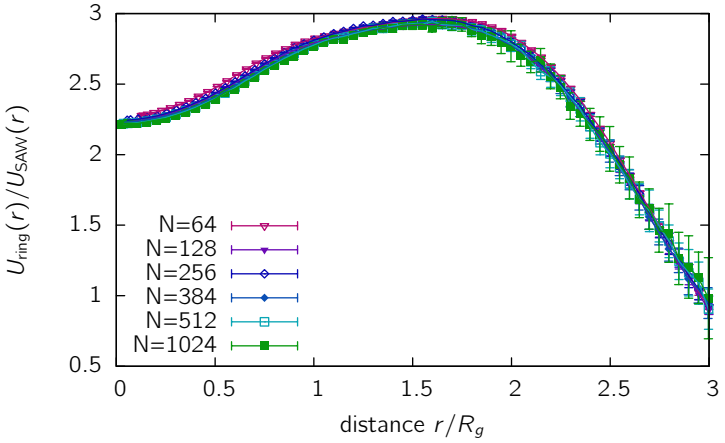


Figure 9.4: Comparison of the effective potentials for ring polymers and self-avoiding walks. The ratio $U_{\mathrm{ring}}(r)/U_{\mathrm{SAW}}(r)$ is plotted against the center-of-mass distance r. The x-axis is scaled by the root mean squared radius of gyration R_g of isolated chains for reasons of comparison.

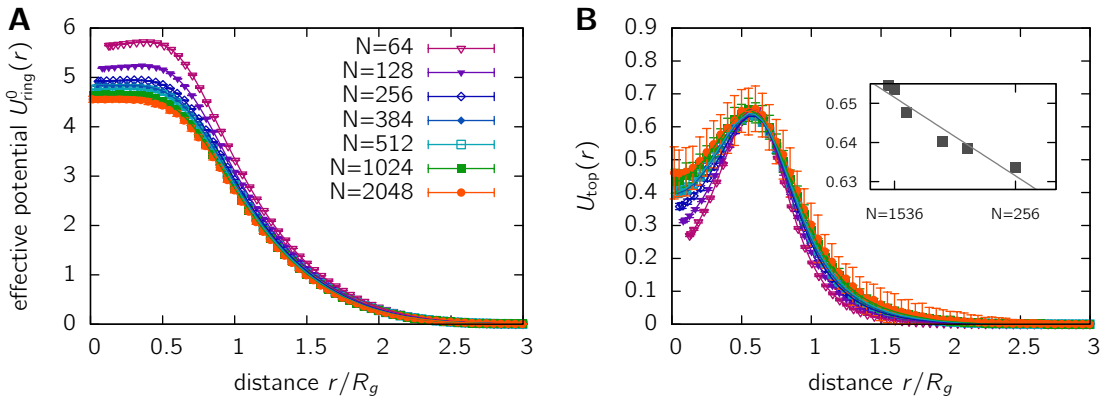


Figure 9.5: **A.** The effective potential $U^0_{\rm ring}(r)$ for non-catenated rings. **B.** The topological potential $U_{\rm top}(r)$ for non-catenated rings is obtained as the difference between $U^0_{\rm ring}(r)$ and $U_{\rm ring}(r)$. It displays the potential arising purely from the topological non-catenation constraint. The inset shows the maximum of the topological potential depending on the chain length N. The maxima are obtained by a fit of $U_{\rm ring}(r)$ to a third-order polynomial in the interval $[0.3 : 0.7]$.

self-avoiding walks, the potential being about 3 times larger. This rise in strength results from the increased density of monomers induced by the ring structure. Here we focus on the additional effect topological constraints induce on the potential of mean force between the polymers. Much attention has been paid to the influence of topological constraints on ring polymers. In a recent study [164] we have analyzed the behavior of two catenated rings. It turned out that while the dimensions of catenated rings show the same scaling behavior than their linear counterparts, the shape changes significantly. Studies on non-catenated and unknotted ring polymers in the melt revealed that their dimensions change dramatically [147, 149, 163]: They behave like compact polymers with a radius of gyration scaling with $N^{1/3}$. The dominant entropic force driving this compaction stems from the topological constraint of non-catenation. To our knowledge, a quantitative study of this topological potential has not been considered yet. In the framework of effective interactions, the topological non-catenation constraint can be included easily. As explained in the section 9.2.3, we determine the topological state of a two-ring conformation by means of its Gauss linking number Φ (eq. (9.8)). A Gauss linking number of zero indicates the non-catenation of the two ring polymers. To obtain the potential of mean force under the non-catenation constraint, we calculate the fraction p of two-ring conformations with both zero linking number and fulfilled excluded volume constraint. The effective potential is then given by $U^0_{\rm ring}(r) = -\ln p$ (the index 0 indicates the Gauss linking number $\Phi = 0$). The resulting potential $U^0_{\rm ring}(r)$ is shown in Fig. 9.5A. We find pronounced deviations from a Gaussian shape for small CM separations $r \lesssim 1R_g$.

From the effective potential $U^0_{\rm ring}(r)$ of non-catenated rings and $U_{\rm ring}(r)$ of rings without topological constraints, the topological potential $U_{\rm top}(r)$, i.e. the part of $U^0_{\rm ring}(r)$ stemming purely from the non-catenation constraint, can be obtained,

$$U_{\rm top}(r) = U^0_{\rm ring}(r) - U_{\rm ring}(r)$$

The topological potential is shown in Fig. 9.5B. It has a maximum at about $r/R_g \approx 0.55$, a value which basically does not change with chain length N. For larger CM separations $U_{\rm top}(r)$ drops to zero, as less and less two-ring conformations will be catenated when increasing r. There is also a decrease in the topological potential when going to small

CM separations $r < 0.5R_g$. This decrease becomes smaller for larger chain length N. The reason for this behavior is that for small r (and small N), two-ring conformations not satisfying the non-catenation constraint, often do not satisfy the excluded volume constraint either, such that the conformations are not counted for the purely topological interaction U_{top}. However, more interesting is the maximum repulsion exerted by the non-catenation constraint. We find that the topological potential is asymptotically

$$U_{\text{top}}^{\max} \to 0.655(2)_B T \quad (N \to \infty)$$

at $r/R_g \approx 0.55$ (see inset of Fig. 9.5B). In numbers, this means that only about 48% ($\exp(-0.65) \approx 0.52$) of all two-ring conformations with fixed center-of-mass distance which would be allowed with respect to excluded volume are actually allowed with respect to topological constraints. In other words, only half of the two-ring conformations satisfying the excluded volume constraint are actually accessible when the non-catenation constraint is considered. Thus, the non-catenation constraint decreases the number of accessible conformations at small CM separations considerably.

Interestingly, the topological interaction is of the order of magnitude of $1\,k_BT$. This value was used for a crude estimate of the free energy of rings in a melt by Cates and Deutsch [148], where they assumed that the free energy cost for a contact between two rings is about $1\,k_BT$, hence in a three-dimensional melt the topological contribution to the free energy is $F \approx k_BTR^3/N$. Although highly simplistic, this scaling exponent $\nu = 2/5$ was later reproduced by simulational results [147]; more recent results, however, suggest that this behavior is only a cross-over to the behavior of compact lattice animals [149].

9.2.7 Dimensions of two-ring conformations

When two polymers are brought closer and closer together, lowering the distance between the centers of mass, not only the repulsive interaction increases dramatically. Also the conformational properties of the polymers are affected [167] in an effort of minimizing the free energy. Here we investigate how the dimensions of a two-ring conformation change compared to a self-avoiding walk when approaching isolated polymers.

There are two measures of the dimensions of the chain, which, however, exhibit the same scaling: the mean squared radius of gyration $\left\langle R_g^2 \right\rangle$ and the mean squared ring diameter $\left\langle d^2 \right\rangle$ (see Ref. [164]). Figure 9.6A shows the root mean squared radius of gyration $R_g(r) = \sqrt{\langle R_g(r)^2 \rangle}$ for self-avoiding walks (SAW), ring polymers without mutual topological interactions (arbitrary linking number Φ) and non-catenated rings (linking number $\Phi = 0$) in dependence of the scaled center-of-mass separation $s = r/R_g$. The data are scaled with the radius of gyration for isolated polymers $R_g = R_g(r = \infty)$ (i.e. the radius of gyration of a polymer at infinite distance away from the second one) for reasons of comparison. For small CM separations ($r < R_g$) the radius of gyration strongly increases and its dimensions are more than 10% larger than in the isolated case. This is consistent with the results for self-avoiding walks by Dautenhahn and Hall [167]. The increase in dimension at separations $r \ll R_g$ originates from the need of the chains to create space for the monomers of the second chain. At intermediate separations of about $r/R_g \approx 1.5 - 2$, the radius of gyration attains a smaller value, thus the polymer effectively undergoes compaction. Both the compaction at intermediate separations as well as the swelling at high overlap remain in the limit of infinite chain length. An extrapolation of the radius of gyration at full overlap ($s = 0$) to $N \to \infty$ shows that $R_g(r = 0)/R_g \to 1.109(1)$ for rings

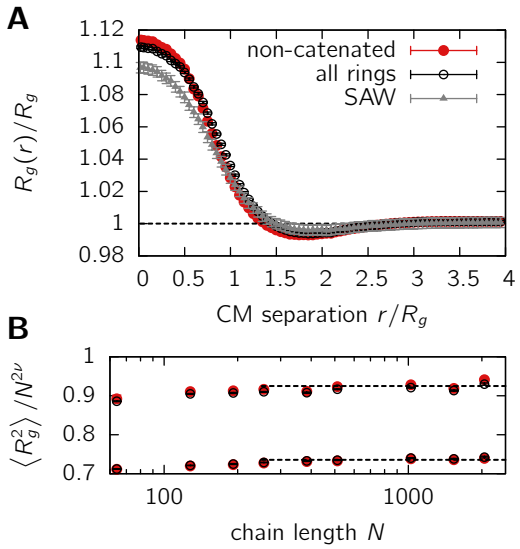


Figure 9.6: **A.** Mean squared radius of gyration for self-avoiding walks (SAW), ring polymers without mutual topological interactions and non-catenated rings vs. distance between the centers of mass. Radii of gyration are scaled by the corresponding values R_g of the isolated chains. **B.** Scaling of the radius of gyration with chain length. The leading term $N^{2\nu}$ of a self-avoiding walk ($\nu = 0.588$) is divided out to see deviations from the linear chain behavior.

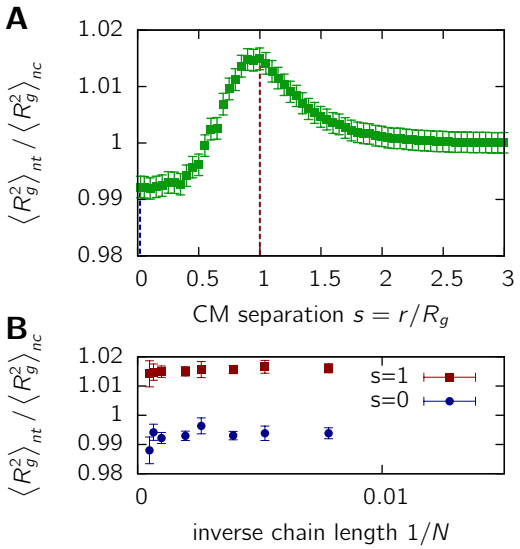


Figure 9.7: **A.** Ratio between the mean squared radii of gyration of rings without mutual topological interactions (NT) and non-catenated rings (NC) vs. scaled distance r/R_g between the centers of mass. Results are shown for a chain of length $N = 1024$. **B.** The ratio is plotted against the inverse chain length allowing for an extrapolation to the infinite chain limit $N \to \infty$. Data is shown for two CM separations: $s = 0$ and $s = 1$ to analyze whether deviations remain in the limit of large chain length.

without topological interactions and $R_g(r = 0)/R_g \to 1.111(1)$ for non-catenated rings (see Supplementary Figure 1 on page 148).

One might ask, whether this change in dimensionality is accompanied by a change in the scaling law. We know from several studies that isolated ring polymers follow a similar scaling law as linear polymers $\left\langle R_g^2 \right\rangle \sim N^{2\nu}$ where $\nu \approx 0.59$ [154, 146, 164]. However, non-catenated rings in a melt behave like compact polymers with a scaling exponent

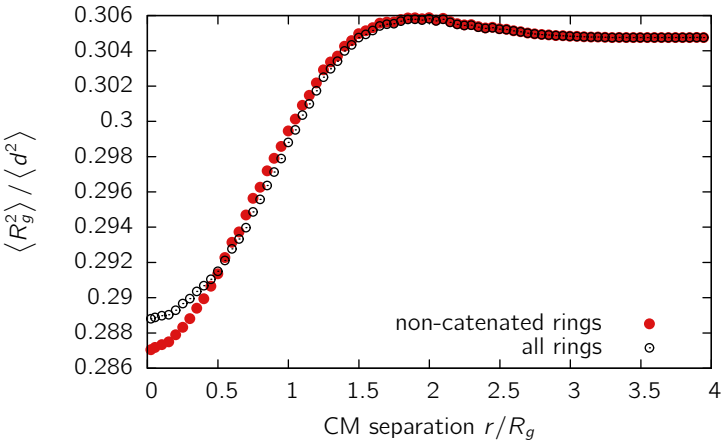


Figure 9.8: Ratio between the mean squared radius of gyration and mean squared ring diameter vs center-of-mass separation r. Data is shown for a chain length of $N = 1024$. Results are shown for two-ring conformations without mutual topological interactions (open circles) and non-catenated ring conformations (solid red circles).

$\nu = 1/3$ [149, 163]. Figure 9.6B shows the scaling of the radius of gyration with chain length in a log-log plot for two center-of-mass separations: $r = 0$ (full overlap, region where the chain is swollen) and $r = 1.7R_g$ (region where the chain is compacted). The leading order term $N^{2\nu}$ of the self-avoiding walk behavior ($\nu = 0.588$ [89]) is divided out. The data adopt a constant value in the limit of large chain length N independent of CM separation, indicating that the scaling behavior is independent of separation. As this also applies to rings with the non-catenation constraint, the compaction found in a polymer melt of non-catenated rings seems to require several chains, exerting forces from all sides on the polymer coil.

Our results show that differences between the radius of gyration $\left\langle R_g^2 \right\rangle_{nt}$ of rings without topological constraints and $\left\langle R_g^2 \right\rangle_{nc}$ of rings obeying non-catenation are rather subtle. The ratio of the radii of gyration $\left\langle R_g^2 \right\rangle_{nt} / \left\langle R_g^2 \right\rangle_{nc}$ with respect to the CM separation is shown in Fig. 9.7A for a chain length of $N = 1024$. For small separations $r < 0.5R_g$ non-catenated rings are smaller, while for intermediate separations ($r \sim 1R_g$) they are larger. The dependence of these results on the chain length N is given in Fig. 9.7B for two different CM separations. The differences are asymptotically stable both in the regime of small as well as intermediate separations. At full overlap ($s = 0$), non-catenated rings are larger in size by about 0.8% in the asymptotic limit $N \to \infty$, for intermediate separations ($s = 1$), rings without topological constraints are larger by about 1.5%. Similar results are found for the mean square ring diameter (data not shown).

As isolated ring polymers as well as self-avoiding walks only have one length scale, different measures of dimension should yield a constant ratio. For a random walk, the ratio between the mean squared radius of gyration and the end-to-end distances is 1/6. For isolated ring polymers, the ratio $\left\langle R_g^2 \right\rangle / \left\langle d^2 \right\rangle$ extrapolates to 0.3053(2) in the asymptotic limit, while catenated rings display a ratio of 0.2995(3) [154, 164]. For the case studied here, the ratio between the radius of gyration and ring diameter $\left\langle R_g^2 \right\rangle / \left\langle d^2 \right\rangle$ changes significantly for different CM separations (see Fig. 9.8). From its isolated chain value the ratio decreases when bringing the centers of mass closer together. Extrapolation to infinite chain length $N \to \infty$ shows that the differences remain even in this limit (Supplementary Figure 2 on page 149). At full overlap ($s = 0$) the ratio extrapolates to $\left\langle R_g^2 \right\rangle / \left\langle d^2 \right\rangle = 0.2874(3)$ for non-catenated rings and $\left\langle R_g^2 \right\rangle / \left\langle d^2 \right\rangle = 0.2891(3)$ for rings without topological interactions.

9.2.8 Shape of two-ring conformations

A typical measure of the shape of a polymer coil is its gyration tensor [83, 158]. It is defined by

$$S_{mn} = \frac{1}{N} \sum_{i=1}^{N} r_m^{(i)} r_n^{(i)} \tag{9.13}$$

Here $\mathbf{r}^{(i)}$ is the coordinate vector of the i^{th} monomer and the subindex denotes its cartesian components. The matrix $\mathbf{S}$ is symmetric and positive semi-definite, thus it can be transformed to a diagonal matrix where the three eigenvalues $\lambda_1 \leq \lambda_2 \leq \lambda_3$ give the squared lengths of the principal axes of gyration of the associated gyration ellipsoid. In fact, the sum of the eigenvalues gives the squared radius of gyration $R_g^2 = \lambda_1 + \lambda_2 + \lambda_3$.

In a recent study [164] on catenated rings we have found that the shape of catenated rings differs significantly from the shape of isolated ring polymers. This shows up in the

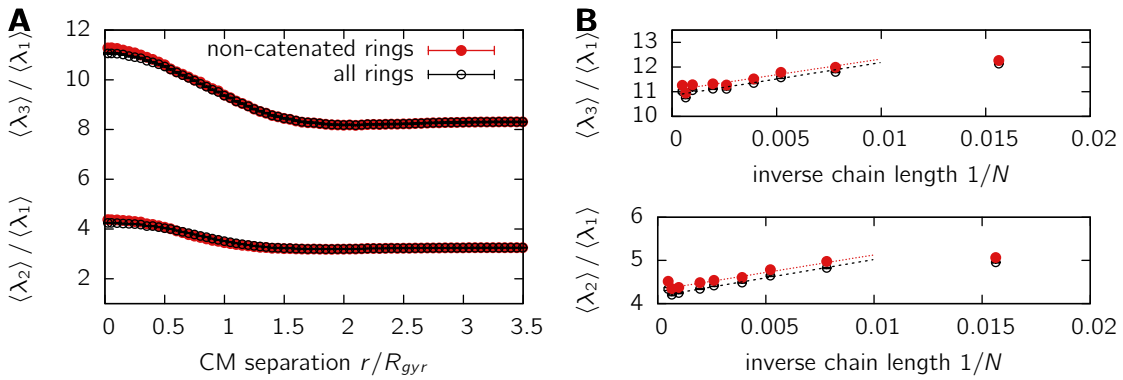


Figure 9.9: **A.** Ratios between the gyration tensor's eigenvalues $\lambda_1 \leq \lambda_2 \leq \lambda_3$ in relation to the center of mass distance $s = r/R_g$. Data is shown for a system with chain length $N = 1024$ for two-ring conformations without mutual topological interactions (open circles) and non-catenated ring conformations (solid red circles). **B.** Finite-size behavior of the ratios $\langle\lambda_3\rangle / \langle\lambda_1\rangle$ and $\langle\lambda_2\rangle / \langle\lambda_1\rangle$ for a center-of-mass separation of $s = 0$.

ratios of the averaged eigenvalues $\langle\lambda_3\rangle : \langle\lambda_2\rangle : \langle\lambda_1\rangle$, which are obtained from extrapolation to the asymptotic limit as $8.23(2) : 3.22(2) : 1$ for isolated rings and $8.87(3) : 3.56(2) : 1$ for catenated rings.

We investigate how the shape of rings changes when approaching the centers of mass, both for rings without topological interactions as well as for conformations, which obey the non-catenation constraint. Figure 9.9A displays the results for a chain of length $N = 1024$. The polymers get strongly elongated for small center-of-mass separations $s = r/R_g$. This elongation at full overlap $s = 0$ remains in the limit of large chain length $N \to \infty$ as is displayed in Fig. 9.9B. For non-catenated rings, the effect of elongation is slightly more pronounced than for rings, where topological effects are neglected. At full overlap we find for non-catenated rings that the ratios $\langle\lambda_3\rangle : \langle\lambda_2\rangle : \langle\lambda_1\rangle$ are asymptotically $11.05(8) : 4.32(3) : 1$, while for rings without mutual topological interactions we have $10.84(6) : 4.17(2) : 1$.

While the analysis of the radius of gyration indicated that the spatial extent of rings becomes larger, we have found here that this opening up does not lead to a complete mixing of the two rings, which would result in a more spherical structure; rather rings elongate strongly in one direction.

9.2.9 Orientation and Alignment of rings

To investigate the interactions between the two rings at full overlap in more detail, we look at the mutual alignment and orientation. The alignment of the gyration ellipsoids of both rings with respect to each other can be investigated via two measures: Firstly, the average angle $\langle\cos\theta_1\rangle$ between the largest main axes of the ellipsoids, secondly the average angle $\langle\cos\theta_2\rangle$ between the vector connecting the centers of mass of both rings and the largest main axis of one ellipsoid. Its behavior has been observed for catenated rings [164]. It was found that the main axes of the ellipsoids tend to align perpendicular for short CM separations, while for large CM separations, the alignment becomes more and more parallel.

The results for two-ring conformations without mutual topological constraints and non-catenated ring conformations as well as SAWs are shown in Fig. 9.10. The black

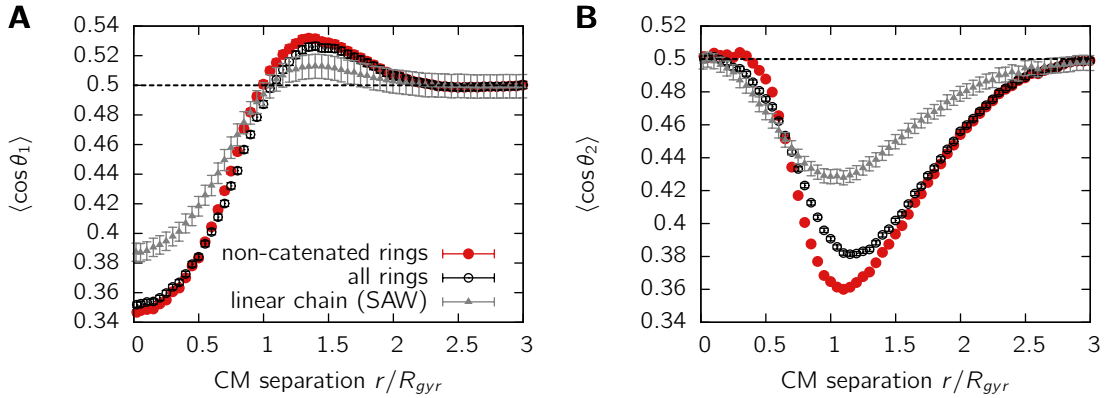


Figure 9.10: **A.** Average angle $\langle \cos\theta_1 \rangle$ between the largest principal axes of the gyration ellipsoids of a two-chain conformation in relation to its center of mass separations. The black line corresponds to a random orientation of the two axes. Results are shown for self-avoiding walks (grey triangles), non-catenated rings (solid red circles) and rings without mutual interactions (open black circles). **B.** Average angle $\langle \cos\theta_2 \rangle$ between the vector connecting the centers of mass and the largest principal axis of one polymer coil in relation of the CM separation.

line at $\langle \cos\theta_i \rangle = 0.5$ corresponds to the average angle for a random orientation of the vectors (Note that the angle θ_i can adopt only values between zero and 90 degrees). The average angle $\langle \cos\theta_1 \rangle$ between the gyration tensors main axes (cf. Fig. 9.10A) displays a pronounced perpendicular orientation for small CM separations $r < R_g$. This effect is markedly stronger for ring polymers than for linear ones. At intermediate separations of $r \approx 1 - 2R_g$, however, a regime is found where a slightly parallel alignment of the rings is preferred compared to a random orientation. For large center of mass separations, a random orientation is approached; in this regime, the rings are nearly independent and therefore do not influence each other. An extrapolation to the infinite chain length regime at full overlap ($s = 0$) yields an average angle $\langle \cos\theta_1 \rangle$ of 0.3543(7) for rings without topological interactions, 0.3492(6) for non-catenated rings and 0.3897(2) for self-avoiding walks. Details of the extrapolation are shown in the Supplementary Figure 3 on page 150.

The angle $\langle \cos\theta_2 \rangle$ (see Fig. 9.10B) shows a nearly random alignment for CM separations of $r \approx 0$. For intermediate $r \approx 1 - 2R_g$, the alignment becomes pronouncedly more perpendicular compared to a random alignment. This effect is much stronger for non-catenated rings than for rings without mutual topological constraints. Clearly, ring polymers display a much stronger tendency to align non-randomly than self-avoiding walks. In the regime of perpendicular alignment at $s = 1.2$ the asymptotic values of 0.3836(5) (rings without topological interactions), 0.3661(4) (non-catenated rings) and 0.4327(3) (SAWs) are found for the average angle $\langle \cos\theta_2 \rangle$ (Supplementary Figure 3)

A recurrant motif in the analysis of shape, dimensions and orientation is a change in structure from polymers at full overlap to polymers at intermediate separations ($r \approx 1-2R_g$). While they are strongly elongated and aligned perpendicular at short separations, at intermediate separations compaction and parallel alignment is observed. Clearly, this results from a tendency to minimize the overlap area between both rings. At full overlap, this is accomplished best by a strong elongation and perpendicular alignment (Fig. 9.11A). When rings are separated further apart, the gyration ellipsoids can avoid intermingling by aligning in parallel (Fig. 9.11B). The restricted conformational space due to the presence of the other ring results in the observed compaction of the radius of gyration compared to isolated chains.

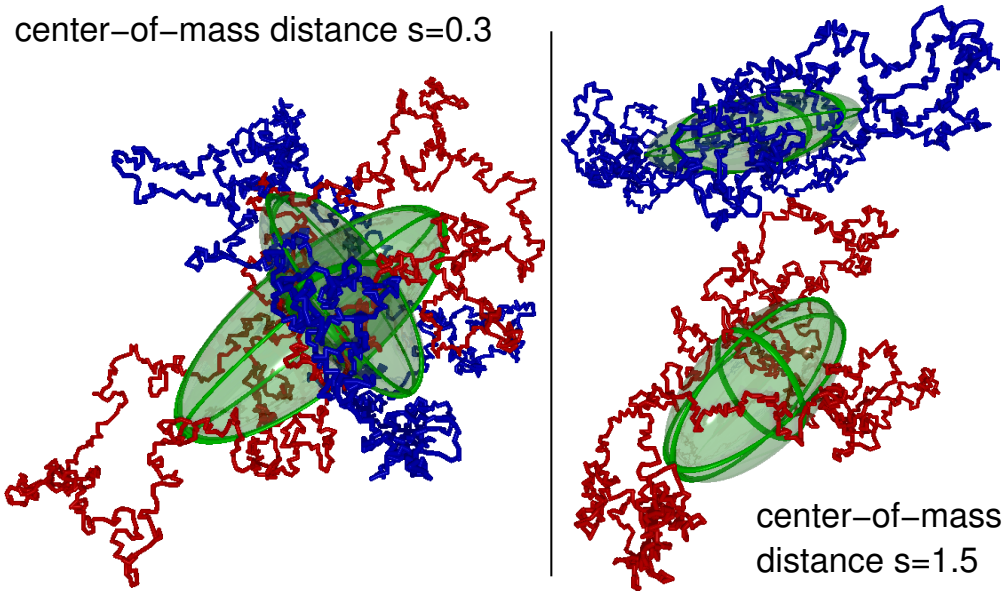


Figure 9.11: Sample conformations of non-catenated rings forced to a certain center-of-mass distance. The left-hand figure displays a conformation of two non-catenated rings at nearly full overlap ($s = r/R_g = 0.3$). The two rings markedly align perpendicularly. The right-hand figure shows a two-ring conformation of non-catenated rings at separation $s = r/R_g = 1.5$ displaying a tendency for parallel alignment.

Both the topological constraints and the ring connectivity of the chain have a strong influence on how the polymers are aligning when brought close together. In general, the view emerges that rings have a more aligned and ordered state than linear chains, an effect which is even amplified by the non-catenation constraint.

9.3 Conclusions

In this study we investigated the effect of the non-catenation constraint on a system with two ring polymers. Our main focus was on the quantitative analysis of the strength of the interactions between ring polymers, including their topological interactions. For this purpose, we sampled conformations of isolated rings using a well-established Monte Carlo method. We evaluated the potential of mean force using the idea of effective interactions [166] following the method introduced by Dautenhahn and Hall [167]. The topological state of a two-ring conformation is analyzed by means of the Gauss linking number.

We found that the effective potential at full overlap adopts a constant value both for linear chains and ring polymers (Figs. 9.1 and 9.2). However, the repulsive interaction between the centers of mass is about 3 times larger for ring polymers than the effective potential for corresponding linear chains (Fig. 9.4) at small center-of-mass separations r.

If the ring polymers have to stay in the fixed mutual topological state of non-catenation, the space of accessible conformations is further reduced, thus the effective potential increases. We have evaluated the strength of these interactions – the topological potential $U_{\mathrm{top}}(r)$ – resulting purely from topological constraints. We find that the non-catenation constraint further increases the total repulsive interaction by about 10%. The strength of the potential is of the order of $1\,k_BT$ at small separations r. The number of rejected con-

formations increases by over 50% compared to only taking into account excluded volume interactions.

The analysis of conformational properties of polymers brought close together reveals effects of the ring structure both on size and shape. Ring polymers in proximity become swollen (Fig. 9.6) and strongly elongated (Fig. 9.9) compared to the situation of isolated or far apart ones. While both effects are found for linear polymers, too, the effects are more pronounced for their ring counterparts. Additional effects from the non-catenation constraint concerning size and shape are visible, but small, indicating that the changes are mainly induced by the presence of additional material due to the rings being more compact than linear polymers.

Remarkable effects of the ring structure are found concerning the alignment of ring polymers in proximity. There is a strong tendency to align perpendicular for short center-of-mass separations, an effect which is much more pronounced than for linear chains. At intermediate separations ($r \sim 1 - 2R_g$) a slightly parallel alignment is encountered.

The transition from a linear to a ring polymer thus induces strong changes in the conformational and structural properties of a system. In fact, ring polymers adopt a much more ordered and regular state, showing less intermingling due to the increased repulsive interactions. The mutual alignment of ring polymers becomes much more different from a random state than found for linear polymers.

Our findings suggest, that the ring topology of certain biopolymers like DNA and proteins, displays a benefit compared to the linear organization form. The formation of chromatin loops in higher eukaryotes therefore might play a dominant role concerning overall nuclear organizational principles. We have shown that loop structures lead to a strong effective repulsion and for a first time quantitatively analyzed the strength of the resulting interactions. Surely, the formation of multiple loops in the system of chromatin which has been proposed in several models [98, 117, 12, 39] induces even stronger effects and therefore might be responsible for maintaining the segregated state of chromosomes found in several experiments [7]. Indeed, studies of looping polymers have revealed an effect on the abundance of inter-chromosomal contacts [35, 93]. Whereas it has already been proposed that the ring structure prevents chromosomes from entangling [163, 30], such a topology also induces a much more ordered and aligned state than linear polymers. It remains to be shown, to which extend similar results are found for chromatin models with loops [98, 12], an approach which will be pursued in chapter 11. We expect that both the local behaviour (e.g. inside of transcription factories or heterochromatic regions) and the global behaviour (segregation of chromosomes) might well be dominated by entropic and topological forces of chromatin loops. Contrary to a melt of ring polymers, which adopts a compact state, the dynamic formation of loops [93] might provide a versatile mechanism for the cell to switch between and maintain different local levels of compaction by facilitating or repressing loop formation.

9.4 Supplementary Information

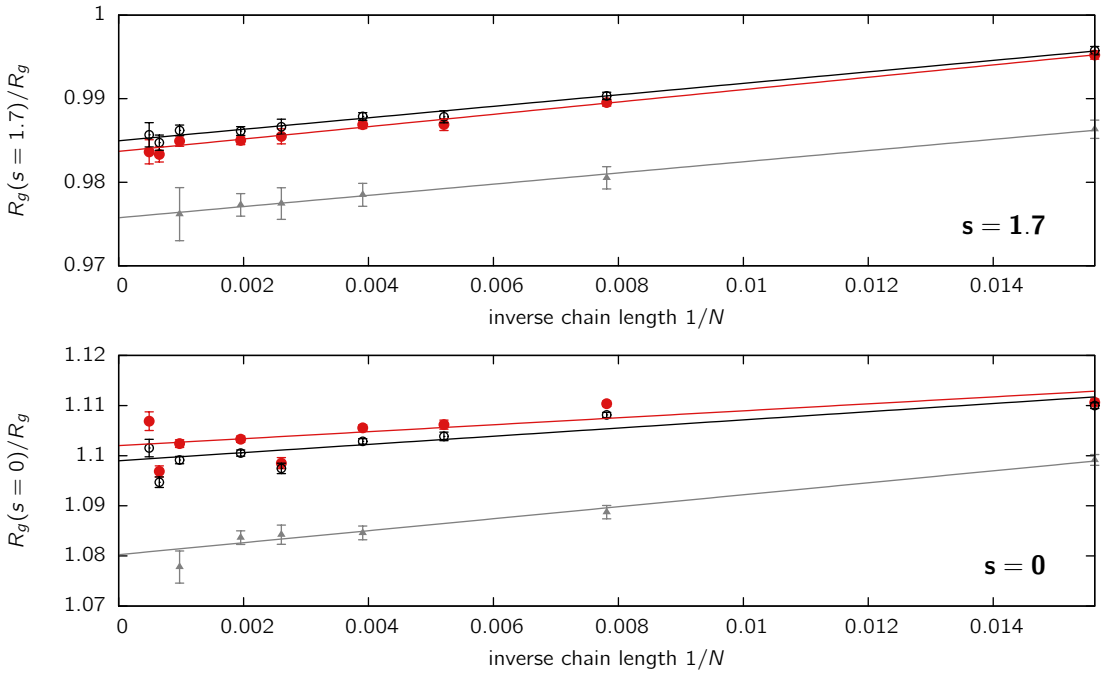
9.4.1 Supplementary Figure 1

The dimensions of two-ring conformations in relation to chain length N. Show is the radius of gyration $R_g(s)$ of a polymer in a system of two identical polymers with a center-of-mass separation $s = r/R_g$. The upper figure shows results for $s = 1.7$, the regime where the polymer is compacted compared to the isolated case. The bottom figure displays results for chains at full overlap ($s = 0$). The data are scaled with the radius of gyration of an isolated chain R_g. Results are shown for self-avoiding walks (grey triangles), non-catenated rings (solid red circles) and rings without mutual topological interactions (open black circles). A linear extrapolation is conducted to obtain the dimensions in the limit of infinite chains $N \to \infty$.
In the region where the chain is compacted ($s = 1.7$), we find the following extrapolated values:

$R_g(s)/R_g$	$\to 0.9757(2)$	self-avoiding walk
$R_g(s)/R_g$	$\to 0.9836(1)$	non-catenated rings
$R_g(s)/R_g$	$\to 0.9848(1)$	rings w/o topological interactions

At full overlap ($s = 0$), linear extrapolation yields

$R_g(s)/R_g$	$\to 1.080(1)$	self-avoiding walk
$R_g(s)/R_g$	$\to 1.111(1$	non-catenated rings
$R_g(s)/R_g$	$\to 1.109(5)$	rings w/o topological interactions



9.4.2 Supplementary Figure 2

The ratio between the radius of gyration and the mean square ring diameter $\langle R_g^2 \rangle / \langle d^2 \rangle$. This figure displays the finite-size behavior of the ratio $\langle R_g^2 \rangle / \langle d^2 \rangle$ of a single polymer in a system of two polymers with center-of-mass distance $s = r/R_g$. The ratio is plotted against the inverse chain length to obtain an extrapolated value for $N \to \infty$. Results for three different center-of-mass separations s are shown. The upper-most figure shows the results at full overlap ($s = 0$), the middle figure shows the results for $s = 1.7$, i.e. polymers separated by 1.7 times the radius of gyration R_g of the corresponding isolated chains. The bottom-most figure shows the results for $s = 3$. Black open circles represent rings without mutual topological interactions, red solid circles non-catenated rings.

Linear extrapolation to $N \to \infty$ yields

- $s = 0$ (full overlap)

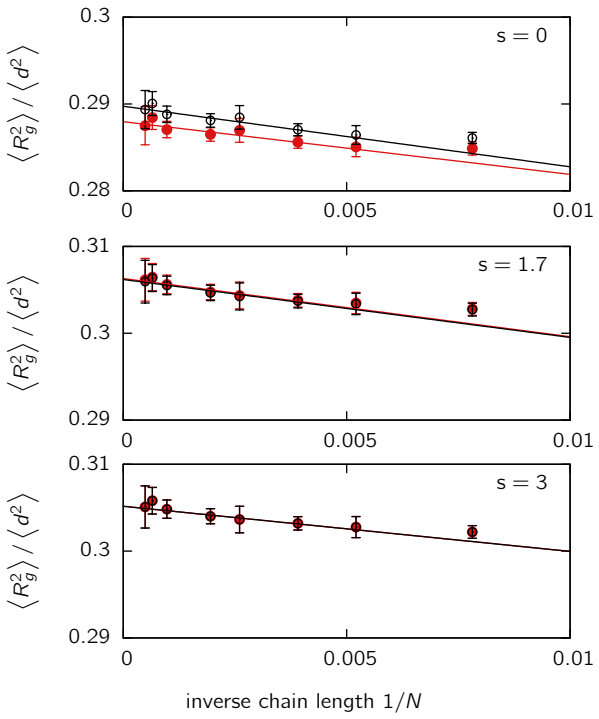
$$\langle R_g^2 \rangle / \langle d^2 \rangle \quad \to 0.2874(3) \qquad \text{non-catenated rings}$$
$$\langle R_g^2 \rangle / \langle d^2 \rangle \quad \to 0.2891(3) \qquad \text{rings w/o mutual topological interactions}$$

- $s = 1.7$

$$\langle R_g^2 \rangle / \langle d^2 \rangle \quad \to 0.3063(3) \qquad \text{non-catenated rings}$$
$$\langle R_g^2 \rangle / \langle d^2 \rangle \quad \to 0.3062(3) \qquad \text{rings w/o mutual topological interactions}$$

- $s = 3$

$$\langle R_g^2 \rangle / \langle d^2 \rangle \quad \to 0.3051(2) \qquad \text{non-catenated rings}$$
$$\langle R_g^2 \rangle / \langle d^2 \rangle \quad \to 0.3051(2) \qquad \text{rings w/o mutual topological interactions}$$



9.4.3 Supplementary Figure 3

Alignment of the gyration tensors of two-chain conformations..
A. This figure shows an extrapolation of the angle $\langle \cos\theta_1 \rangle$ between the largest principal axis of the gyration tensors of a two-chain conformation in relation to chain length N for several center-of-mass separations $s = r/R_g$. At full overlap ($s = 0$), there is a tendency to perpendicular alignment, which remains in the limit of infinite chain length. At intermediate separations ($s = 1.4$), chains tend to align parallel, again this behaviour remains in the limit $N \to \infty$. At large CM separations ($s = 3$), the orientation becomes nearly independent of each other, resulting in the average value $\langle \cos\theta_1 \rangle = 0.5$. Results are shown for self-avoiding walks (grey triangles), non-catenated rings (solid red circles) and rings without mutual topological interactions (open black circles).

Extrapolation yields:

- $s = 0$ (full overlap)

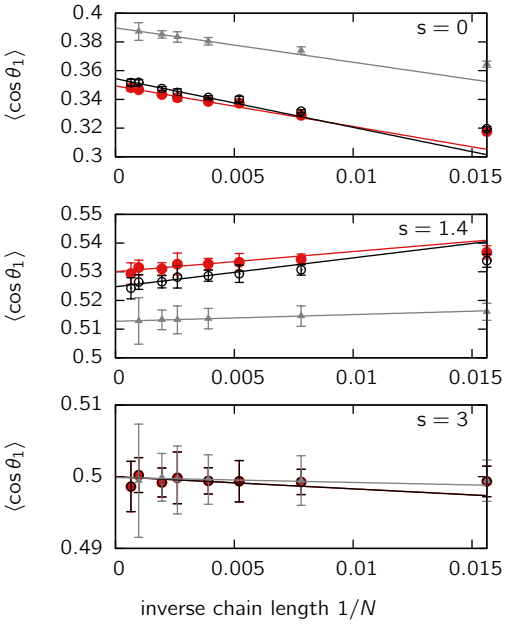
  $\langle \cos\theta_1 \rangle \to 0.3897(2)$ self-avoiding walks
  $\langle \cos\theta_1 \rangle \to 0.3492(6)$ non-catenated rings
  $\langle \cos\theta_1 \rangle \to 0.3543(7)$ rings w/o mutual topological int.

- $s = 1.4$

  $\langle \cos\theta_1 \rangle \to 0.5127(1)$ self-avoiding walks
  $\langle \cos\theta_1 \rangle \to 0.5301(7)$ non-catenated rings
  $\langle \cos\theta_1 \rangle \to 0.5251(8)$ rings w/o mutual topological int.

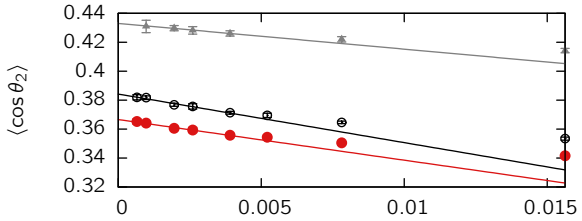
- $s = 3.0$

  $\langle \cos\theta_1 \rangle \to 0.4999(6)$ self-avoiding walks
  $\langle \cos\theta_1 \rangle \to 0.4999(6)$ non-catenated rings
  $\langle \cos\theta_1 \rangle \to 0.4999(3)$ rings w/o mutual topological int.



B. This figure shows an extrapolation of the angle $\langle \cos\theta_2 \rangle$ between the largest principal axes of one polymer and the vector connecting the centers of mass of a two-chain conformation. Data is shown for a center-of-mass separation of $s = 1.2$, i.e. the regime where a pronounced perpendicular orientation is observed. This effects remain in the limit of infinite chain length $N \to \infty$. Linear extrapolation yields for $s = 1.2$:

$\langle \cos\theta_1 \rangle \to 0.4327(3)$ self-avoiding walks
$\langle \cos\theta_1 \rangle \to 0.3661(4)$ non-catenated rings
$\langle \cos\theta_1 \rangle \to 0.3836(5)$ rings w/o mutual topological interactions



Chapter 10

Diffusion-driven looping as a consistent framework for chromatin organization

The Dynamic Loop model

References

The results presented in this chapter are published in:

- M. Bohn & D. W. Heermann, **2009**. Diffusion-driven looping provides a consistent framework for chromatin organization. *revised manuscript submitted to PLoS Comp. Biol.*

Chapter Summary

Experimental results suggest that higher-order chromatin folding is tightly connected to genome function. The building blocks for a consistent overall description of chromatin folding have to be gathered from several experimental techniques. FISH measurements reveal a confined folding of chromosomes into a sub-space of the nucleus, the mean square distance between two markers becoming independent of genomic separation but at the same time revealing a huge cell-to-cell variation. 4C data provides evidence that loops exist on all scales and it has been shown that these take part in transcriptional regulation. Sub-diffusive motion of an active gene has been detected in yeast and aspherical chromosome territories are observed. Chromatin looping seems to play a dominant role both in transcriptional regulation as well as in chromatin organization and has been assumed to be mediated by long-range interactions in many polymer models. The Random Loop model correctly predicts a folding into a confined space and the cell-to-cell variation based on probabilistic looping. However, it remains a crucial question which mechanisms are necessary to make two chromatin regions become co-located, i.e. have them in spatial proximity. We demonstrate that the formation of loops can be accomplished solely on the basis of diffusional motion. The probabilistic nature of temporary contacts mimics the effects of proteins, e.g. transcription factors, in the solvent. Our Dynamic Loop model presented here provides a unified description of chromatin folding on the basis of dynamic loop formation without invoking active transport mechanisms. Testable predictions are derived for a variety of observables. In fact, for the first time, a huge amount of experimental evidence from FISH distance measurements, 4C/5C data, diffusion measurements up to the formation of chromosome territories and its shape can be explained consistently in the framework of one model.

10.1 Introduction

The cell nucleus is a main constituent of eukaryotic organisms and yet its complexity prevents detailed knowledge of its function. The genome content is carried by the chromosomes: compactly folded polymers consisting of DNA and histone proteins. While during mitosis chromosomes are found in an extremely condensed state, the chromatin fiber inside the interphase nucleus has a much more decondensed organization. However, at this stage of the cell cycle, highly coordinated processes such as transcription, replication and DNA repair take place, making a random folding of the chromatin fiber very unlikely. A pivotal question is the connection between genome organization and function, which could not be answered in a satisfying way up to now (↷ 2.3).

Experimental results are available from a variety of approaches, highlighting a tight connection between genome folding and function. The Human Transcriptome map reveals that the one-dimensional distribution of genes along the chromosome is far from being random: Active and inactive genes tend to cluster into certain domains [13]. Various experiments have shown that the 3D organization of chromatin depends on transcriptional activity: Active genes tend to be located in the nuclear interior while inactive genes are found more often at the nuclear periphery [6, 14, 15], the converse behavior is observed in some experiments [16]. Transcriptional active regions (ridges) were observed to have

a more open structure than inactive regions (anti-ridges) [6]. Fluorescence in situ hybridization (FISH) distance experiments displayed a leveling-off in the MSD for genomic separations above 10 Mb (mega basepairs), the plateau level being in the size range of 2 μm (Figure 2.3). Importantly, the fluctuations of these FISH measurements are larger than expected from a random walk polymer model. On the scale of the nucleus, chromosomes are separated into distinct chromosome territories [7], whose relative positions and ellipsoidal shape varies from cell to cell [19].

It has been emphasized in this thesis at several points that intra-chromosomal as well as inter-chromosomal contacts or loops have been intensively analyzed in the past few years both experimentally and theoretically as a possible mechanism for transcriptional regulation and genome folding. Yet, chromatin loops seem to be an ubiquitous feature of genome organization and genome function. One idea put forward to explain chromatin loops is the existence of transcription factories or active chromatin hubs, where active polymerases cluster and thereby co-locate genes and regulatory elements [23, 24]. 3C and 4C techniques have since then provided evidence that indeed loops up to several Mb exist in interphase cells [20, 25]. However, the detailed mechanisms and driving forces of looping are still under debate.

Given such a variety of experimental data, polymer models can help in understanding the functional folding motifs of chromatin in the interphase nucleus and its connection to gene regulation. The Random Loop model proposed in chapters 5 and 7 correctly predicts the folding into a confined sub-space of the nucleus as well as the large cell-to-cell variation observed in FISH measurements. This model connects genome folding and function using experimental results on intra-chromosomal loops. Results yielded evidence that probabilistic loop formation might be a major driving force of chromatin organization.

However, it remains a crucial question which mechanisms are necessary to make two chromatin regions become co-located, i.e. have them in spatial proximity. The RL model assumed a harmonic potential between the loop attachment points ($\curvearrowright$ 5.3). Thus long-range forces were postulated to form and maintain the loops, which is highly questionable. Here, we ask what mechanisms of loop formation are necessary to obtain correspondence with experimental data. While the cell in principal could use ATP to produce energy for actively co-locating chromatin segments, we are going to shed light on the question, whether this is really necessary. We present a polymer model, the Dynamic Loop model, where functional loops are formed solely on the basis of diffusional motion. Importantly, loops are assumed to be dynamic and the sets of loop attachment points change during time. Thus, our loop model is minimal, meaning that we do not assume a priori long-range forces and active transport mechanisms. Besides the new motif of dynamic loop formation, this is a major advance with respect to other chromatin models with loops [11, 28, 117, 109, 12, 35]. Various other polymer models have been proposed [26, 31] which do not take into account chromatin looping. The assumptions of our model arise from biological evidence: 4C experiments clearly show that loops exist on length scales from several thousand basepairs to tens of Mb [25]. Surely, if looping is related to functional processes like transcriptional regulation and the formation of transcription factories, the cell must be able to control this looping dynamically. Large cell-to-cell variations in FISH distance measurements [12, 159] render such a dynamics a necessary feature of any polymer model.

Our model makes testable predictions on a variety of observable quantities. We predict that chromosomes fold into a confined space and display a different fluctuation regime than non-dynamic looping polymers or linear chains. Importantly, the formation of large

loops can be accomplished hierarchically mediated by many loops on the short scale without the assumption of long-range interactions. We demonstrate that the beads of the polymer display sub-diffusive behavior in agreement with experimental data [17] and that chromosome territories are constituted driven by looping; the overlap between different chromosome territories (CTs) depends on the local looping probabilities.

10.2 The chromatin model

10.2.1 General overview

Our model starts by initially assuming chromatin to consist of a coarse-grained linear polymer chain. Loop formation is achieved on the basis of diffusional motion of the monomers in the following way: Whenever two segments co-localize by diffusional motion, a chromatin loop is formed with a certain probability p between these two sites. A certain lifetime is assigned to each loop, thus loop attachment points dissolve again during the course of time. Lacking experimental knowledge on the time scales over which chromatin segments remain co-localized, e.g. in transcription factories, different looping lifetimes are considered. Details are described in the Methods section ($\curvearrowright$ 10.2.2).

The stochastic nature of loop formation provides a method to effectively incorporate protein-chromatin and chromatin-chromatin interactions. Looping is often thought to be mediated by DNA-binding factors such as CTCF [76], Sat1B [174] or PcG [10] or by regions of increased polymerase concentration, i.e. transcription factories [24]. The probabilistic creation of functional chromatin contacts mimics the effect of protein concentration (there being either proteins binding DNA sites or not) and binding affinity. In the following we denote by "loop" a functional interaction between two parts of a chromatin fiber existing for a certain time as created by the algorithm. In contrast, a "contact" denotes two parts of the chromatin fiber close together by thermal fluctuations without necessarily being an interaction.

A typical human chromosome has a length of about 100 mega basepairs (Mb), rendering a detailed description on the molecular level computationally impossible. Typically, coarse-grained approaches are used, where a long stretch of chromatin is modeled as an effective monomer ($\curvearrowright$ 3.1.6). Polymer scaling theory [78] tells us that for linear polymers such an approach is well justified above the scale where bending rigidity plays a role. This length scale is established by the persistence length l_p, defining the transition from a rod-like to a flexbile polymer. Estimates for the persistence length of chromatin are in the range of $l_p = 40 - 250$ nm [91] but are often based on crude approximations by fitting data to linear chain models [26, 8]. Thus, it is reasonable to conduct computer simulations on a coarse-grained scale where it can be securely assumed that the fiber is flexible.

To study the impact of diffusion-based loop formation on the conformational properties isolated from effects of the presence of other chromosomes, we simulate single chromosomes in a dilute solution. In fact, it has been argued that the disentanglement time for the transition from interphase to metaphase chromosomes of size 100 Mb is in the order of 500 years [97, 30], thus requiring the activity of DNA topoisomerase II. Rosa *et al.* reversed the argument proposing that interphase chromosomes never equilibrate [30]. We ask whether the observed confined folding already arises from the experimentally confirmed loop formation without invoking rather unprecise knowledge of time and length scales. If loop formation turns out to cause confined folding, then the presence of other chromosomes should not alter the conformational properties drastically. That is why we focus first on isolated chains. In a coarse-grained approach we study chain lengths of size $N =$

64, 128, 256 and 512. We use Monte-Carlo simulations on a cubic lattice employing the well-established bond fluctuation algorithm [155]. The lattice size is chosen to be $L = 256$. By using periodic boundary conditions and keeping track of unfolded coordinates we avoid forcing the polymers into a confined space.

While simulations of diluted chromosomes can be used to study the effect of looping isolated from the presence of other chains, simulations of polymers in a dense system are necessary to study the formation of chromosome territories and to answer the question whether density-related effects are observable. Thus, it is a natural next step to perform simulations in a system with many chromosomes. For our simulations we choose a linear simulation box of width $L = 64$ and a density of $\rho = 12.5\%$, which is similar to the conditions in interphase nuclei. A total of 4096 monomers was thus studied.

Most often, simulational studies map coarse-grained monomers to physical length scales, e.g. by assuming a certain persistence length [30, 31]. Thus, a parameter-dependent comparison between the physical distance of two markers with experimental data from FISH measurements can be conducted. To obtain testable quantitative predictions we follow another, more universal, approach. We derive quantities which are independent on the detailed mapping of the model fiber to the biological chromosome, but can be easily evaluated both from simulational data as well as from experimental data. Such quantities comprise dimensionless higher-order moment ratios of the distance distributions as well as scaling exponents. We show that these quantities do not depend on the chosen level of coarse-graining, i.e. the chain length. Thus, without assuming unknown time and length scales, a sensitive comparison between theory and experiment is possible.

10.2.2 Simulational Method

The biological model is implemented using Monte Carlo simulations [111]. These Monte Carlo simulations are performed on a lattice in order to simplify the handling of excluded volume. Calculation of excluded volume interactions thereby is reduced to checking whether one lattice site is already occupied or not. Instead of using a simple local-move algorithm on a cubic lattice we employ the bond-fluctuation method introduced by Carmesin [155], which has already been used for the simulation of ring polymers in chapters 8 and 9. The bond-fluctuation model is especially suited for dense and compact systems where a lattice algorithm would no longer be feasible due to high rejection rates during the Monte Carlo process. The algorithm conducts only local moves in order to resemble the dynamics of real polymers [155]. Using a coarse-grained lattice approach is reasonable as we are only interested in features of looping chromatin independent of local structure. Coarse-graining allows us to abstract from the complex environment and highlight the main driving forces and effects of chromatin folding.

Simulations for single polymers are performed on a lattice of size $256 \times 256 \times 256$. Periodic boundary conditions are used, but the algorithm always keeps track of unfolded coordinates, such that the polymer does not feel any confined volume. The lattice size $L = 256$ is chosen larger than the radius of gyration $\left\langle R_g^2 \right\rangle$ of the chains studied such that effects of the backfolding are negligible.

A dense system of model chromosomes is simulated in a system of size $L = 64$ and chain length $N = 128$. The total number of monomers is 4096, the density $\rho = 12.5\%$.

In order to obtain thermodynamical equilibrated conformations we perform the Metropolis Monte Carlo method. Chromosomes are initially equilibrated as self-avoiding walks using local moves of a monomer to one of the nearest neighbors on the lattice. After the initial equilibration steps, the Monte Carlo algorithm allows for the formation of loops.

After each Monte Carlo trial move, one monomer is selected at random. It is then checked whether another monomer on the same chain is in the neighborhood, i.e. co-localized. The co-localization condition is fulfilled whenever the distance between the monomers is less than 3 lattice units. If the two monomers are co-localized, then a loop is formed with a certain probability p. If the loop $i \leftrightarrow j$ is created, it is assigned a certain lifetime t_{loop} which is drawn from a Poissonian distribution

$$P(t_{loop};\tau) = \frac{\tau^{t_{loop}}}{t_{loop}!} e^{-\tau} \,, \tag{10.1}$$

where the parameter τ determines the average lifetime of the loops. In the simulations we use three different values of τ:

$$\tau_1 = 0.01\tau_{\text{int}} \,, \qquad \tau_2 = 1\tau_{\text{int}} \,, \qquad \tau_3 = 1000\tau_{\text{int}} \,, \tag{10.2}$$

where τ_{int} is the integrated autocorrelation time (see below) of the squared radius of gyration for the corresponding self-avoiding walk system. Loop lifetimes are chosen relative to the relaxation time to make results for different values of the other parameters comparable. In the algorithm, looping is implemented by creating a bond between monomer i and j. The bond is subjected to the same restrictions as any other, i.e. it has to be element of the allowed set of bond vectors $\mathcal{B}$ [see eq. (8.3)]. Any move of a monomer resulting in a change of the bond vector $\mathbf{b} \to \mathbf{b}'$ from $\mathbf{b} \in \mathcal{B}$ to $\mathbf{b} \notin \mathcal{B}$ is rejected. The bond and therefore the restriction is removed after the lifetime of the loop is exceeded.

The integrated autocorrelation time τ_{int} is calculated for each set of parameters (chain length N, looping probability p and lifetime of loops τ) according to the method explained in section 8.2.2. Another criterion to ensure the uncorrelatedness of subsequent conformations is given by the motion of the center of mass. This method has been used for example by Mueller *et al.* in his study on ring polymers[147]. Here we determine the function

$$g_3(t) = \left\langle \left(\mathbf{R}_{CM}(t) - \mathbf{R}_{CM}(0)\right)^2 \right\rangle \,.$$

The time of interest, τ_{rep}, after which the center of mass has moved at about one radius of gyration, is defined by

$$g_3(\tau_{rep}) / \left\langle R_g^2 \right\rangle = 1 \,. \tag{10.3}$$

We consider two subsequent conformations as uncorrelated after $5\tau_{int}$ Monte Carlo steps. Actually, for each set of parameters considered here, we found that after this time the center of mass has moved on average at least by one radius of gyration, i.e. $5\tau_{int} > \tau_{rep}$.

Simulations of even small polymers are very time-consuming due to the looping interaction and the resulting compactness of the polymers. Furthermore, simulation runs have to be quite long to capture the dynamics of loop formation. For each set of parameters (N, p, τ_i) we created 10 000-100 000 independent conformations. We study polymers of lengths $N = 64, 128, 256$ and $N = 512$. The looping probabilities are chosen such that the average number of loops in the resulting conformational ensemble is between zero and N. The lifetimes of the loops are chosen from the set given in eq. (10.2).

10.3 Results

Modeling chromosomes with complex interaction patterns results in the need to dramatic simplifications in order to allow sufficient relaxation of the fiber within a feasible computational effort. Therefore, we study the looping dynamics for isolated coarse-grained

chromatin fibers first. Although such conditions are not found in *in-vivo* experiments, the formation of loops and its influence on the chromosome properties can be studied isolated from density-related effects. In the next step, we present the results of simulations of a system of several chromosomes at biological densities. Since the looping results in confined structures, as will be shown below, density-related effects are only minor and, consequently, the formation of chromosome territories is observed.

10.3.1 Mean square distance between chromatin segments

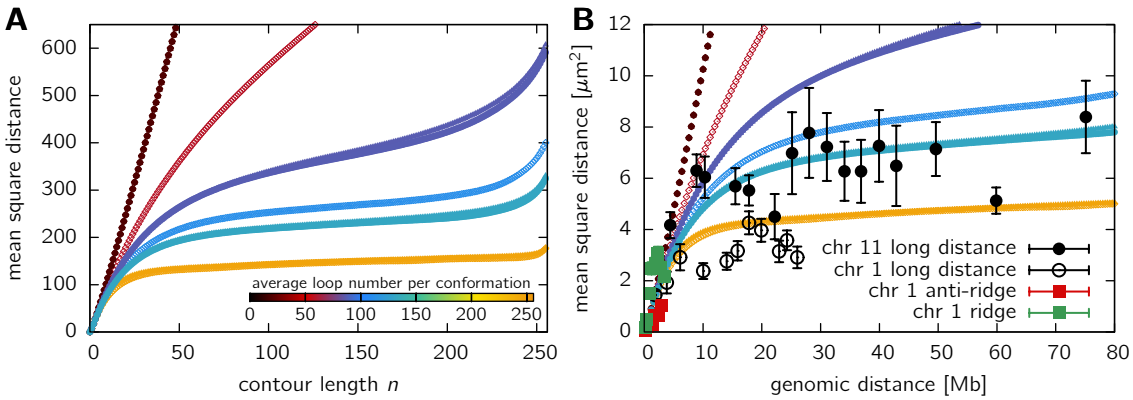


Figure 10.1: Mean square distance $\langle R^2 \rangle$ in relation to contour length for an isolated fiber. A. The mean square distance $\langle R_n^2 \rangle$ between two beads separated by contour length n. Isolated polymers of length $N = 256$ have been fully equilibrated for various looping probabilities p. These probabilities are plotted with different colors depending on the resulting average number of loops per conformation. Simulations have been performed using various lifetimes of loops, the results are displayed by different symbols (triangles [▲] for $\tau = \tau_1$, open diamonds [◊] for $\tau = \tau_2$ and filled circles [•] for $\tau = \tau_3$). The mean square distance displays a leveling off for average loop numbers beginning at a number of 80. **B.** Comparison of the mean square distance to experimental data taken from Mateos-Langerak et al. [12]. Measurements have been performed on the q-arm of human interphase chromosome 1 and 11. Each bead of the model fiber represents a 400 kb stretch of chromatin with an average extension of 480 nm.

We first show that the Dynamic Loop model is in agreement with experimental data from FISH measurements [27, 12], which provide information about the relative physical distance between two target sites. The mean squared distance value $\langle R^2 \rangle$ between those target sites in relation to genomic distance g between them can be compared to polymer models. The random walk (RW) and self-avoiding walk (SAW) polymer models predict this mean squared distance to increase monotonically with the distance between two FISH markers,

$$\left\langle R^2 \right\rangle \sim g^{2\nu}\,, \tag{10.4}$$

where ν is a model-dependent parameter [79]. In principle, such a scaling is only valid for the end-to-end distances, however, we want to stress that in the absence of other interactions, equation (10.4) is approximately valid for genomic separations g of interest much larger than the persistence length l_p of the chromatin fiber.

The confined space of the nucleus renders the random walk and self-avoiding walk polymer model inadequate. A 100 Mb chromosome with assumed Kuhn length of approximately 300 nm [8] in a 30 nm fiber packing (300 nm $\simeq$ 30 kb) would extend on average to 17.3 μm in the random walk case, whereas the average diameter of a nucleus is of the

order of 10 μm. The globular state model fails for other reasons [159] (cf. also Fig. 10.4). Recent experiments [12], however, clearly revealed that while the mean square distance increases monotonically with genomic separation on short distances up to a few Mb, a leveling-off is observed for larger genomic separations. This confined folding is observed on a scale of about 2 μm, far below the diameter of the nucleus but consistent with the estimated size of chromosome territories [7]. The random loop model [117, 12] explains the behavior by the formation of random loops, without invoking a confined geometry a priori.

We first considered the mean square distance between two beads in the DL model for isolated chains. Given a chain of length N with monomer positions denoted by $\mathbf{r}_i \quad (i = 1, \ldots, N)$, the average is calculated over a set $\mathcal{C}$ of independent conformations as well as over different reference points inside the chain

$$\left\langle R_n^2 \right\rangle = \frac{1}{|\mathcal{C}|} \frac{1}{N-n} \sum_{C \in \mathcal{C}} \sum_{i=1}^{N-n} \parallel \mathbf{r}_{i+n}^C - \mathbf{r}_i^C \parallel^2 \ . \tag{10.5}$$

Figure 10.1 shows the results of the model for a chain of length $N = 256$. The looping probability p is varied such that different values for the average number of loops are obtained. Lacking knowledge of the biological lifetime of the loops, results are shown for three different values of τ depending on the relaxation time of the chromosomes τ_{int} (triangles [▲] for $\tau = 0.01\tau_{int} = \tau_1$, open diamonds [◊] for $\tau = \tau_{int} = \tau_2$ and filled circles [•] for $\tau = 100\tau_{int} = \tau_3$, see section 10.2.2). The model displays a cross-over from self-avoiding walk behavior (small number of loops) to a leveling-off in the mean square distance. Such a plateau level is recovered if the average number of loops on a coarse-grained chromosome is larger than about 80. This result is independent of the lifetime of the loops as long as the average number of loops remains the same, indicating that the lifetime has no direct influence on the statistical equilibrium properties. These findings clearly show that no long-range interactions are necessary for forcing the polymer to collapse but a purely diffusional motion together with chromatin-chromatin binding affinity suffices to achieve this.

To quantitatively compare the model to experimental data, we assume each bead to represent a 400 kb segment of chromatin with an average extension of 480 nm (in agreement with experimental data [12]). To ensure that the qualitative results are not dependent on chain length, we studied the mean square distance for $N = 128$ and $N = 512$ (Supplementary Information in section 10.5.1 on page 173). In all three cases a leveling-off is observed, indicating that the observed results are independent on scale and the applied coarse-graining is justified.

10.3.2 Self-organized formation of large loops

Loop formation is a central process for the transcriptional regulation in higher eukaryotes. Several studies indicated that co-localization of chromatin segments results in activation or repression of genes [20]. Hypotheses of loop formation range from the attachment to a structure called nuclear matrix [50] to the formation of transcription factories [24], in which transcriptionally active genes come together, forcing the intervening DNA to loop out. Recently it has been shown [35] that loops can promote territory formation with a simple model using fixed loops. However, such a kind of looping does not yield a correct description for the relative positioning of two markers [12]. Rather, it has been shown by 4C experiments [25], that loops exist on scales up to several Mb. 3C/4C/5C and

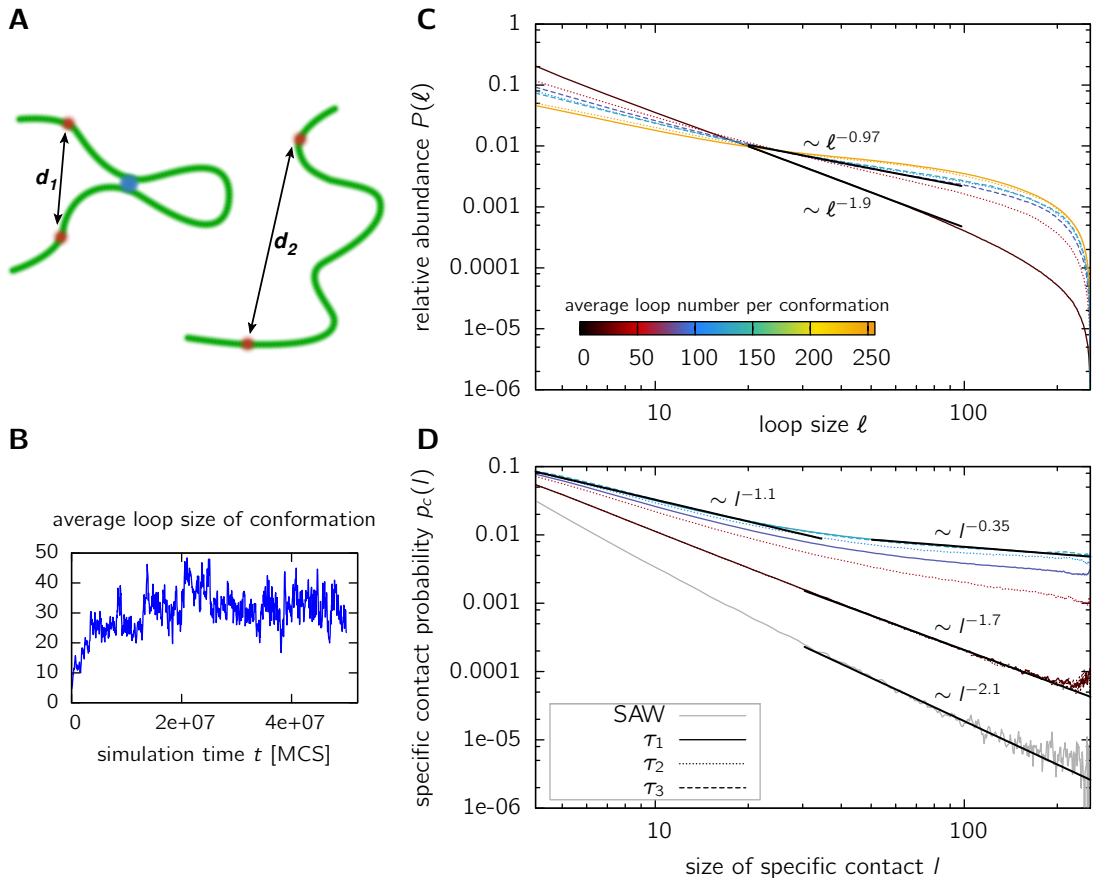


Figure 10.2: Size distribution of loops and random contacts. A. A sketch displaying the facilitated formation of large-loops under the existence of small loops. For a linear polymer, the probability that two chromatin segments marked by red dots co-located by random diffusion is small (right image). Once a small loop has formed in the model (blue dot, left image), the co-localization of the red markers becomes much more frequent. The reason is that the formation of a loop decreases the average distance d_1 between the red markers compared to the linear case d_2. **B.** This figure displays the average loop size of a conformation during the run. Starting from an equilibrated self-avoiding walk conformation ($t = 0$), small loops form by random collisions. This enhances the probability of segments further apart to come into contact, thus the average loop-size increases, allowing even loop-sizes of the length of the chain. **C.** Shown is the size distribution of functional chromatin loops of model polymers with $N = 256$ beads. Model polymers were fully equilibrated and the loop size distribution was determined for various looping probabilities p (for reasons of comparison the average number of loops per conformation is displayed by a color code) and lifetimes τ of the functional loops ($\tau = \tau_1$ solid line, $\tau = \tau_2$ dotted line, $\tau = \tau_3$ dashed line). Looping lifetimes are chosen relative to the relaxation time, see eq. 10.2. In an intermediate region, away from the chain ends, the curve can be roughly fitted to a power-law $P(\ell) \sim \ell^{-\beta}$. Increasing the loop number results in a markedly smaller exponent, leading to a high probability for large loops. **D.** The contact probability $p_c(l)$ for two specific sites with genomic separation (contour length) l to become co-localized. Shown are the results for equilibrated model polymers with $N = 256$ beads and various looping probabilities p. The contact probability decreases as a power-law $l^{-\beta}$ with genomic separation for separations $n \gtrsim 30$, the exponent strongly depending on looping probability. The grey line represents the self-avoiding walk. Again, the co-localization probability is strongly increasing due to diffusion-based looping.

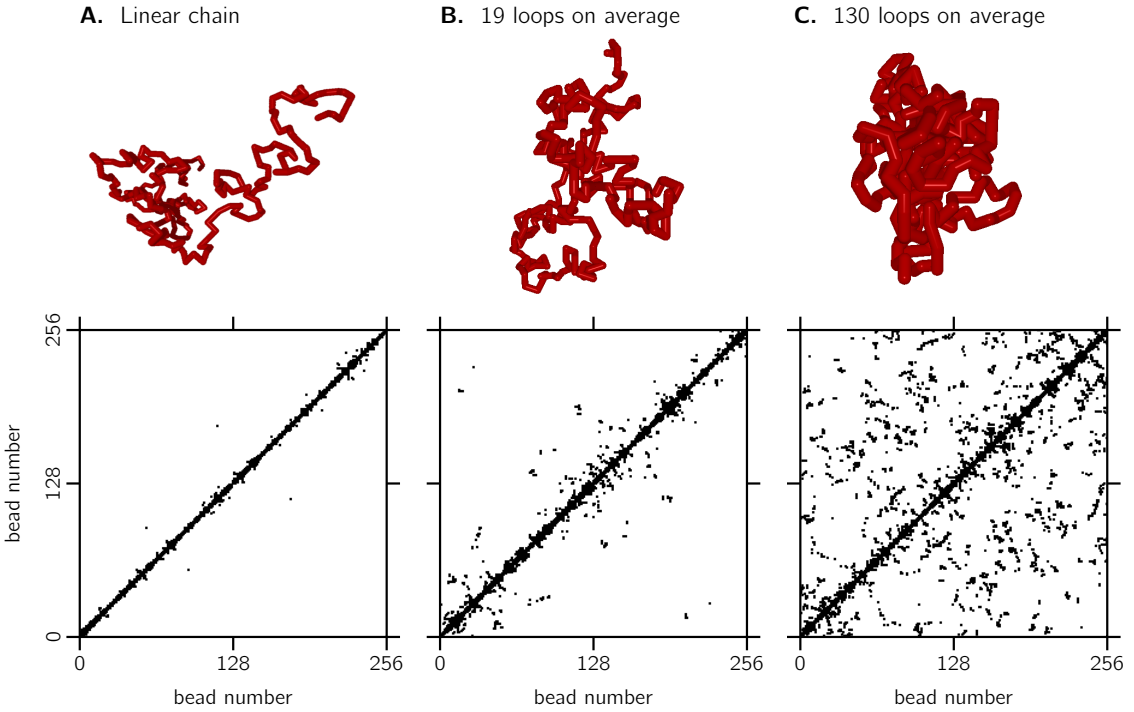


Figure 10.3: Intra-chromosomal contacts of isolated model polymers. Shown are the results for equilibrated fibers of length $N = 256$ with different looping probabilities. For each parameters set (A. linear chains (no loops), B. on average 19 loops per conformation and C. on average 130 loops per conformation) co-localized beads were determined and marked by a black square. For each image, the contacts of 4 independent polymer conformations are plotted. Linear chains (A) have not so many contacts between beads which are widely separated along the contour of the polymer. Increasing the probability of functional loops (B and C) results in a boost of contacts both between close-by segments as well as between segments having a large genomic separation.

the newly developed Hi-C [39] techniques provide an experimental method to measure loop probabilities and distributions. Therefore, we next investigated how the model alters the distribution and frequency of genes to become co-located. Again, we favor measures that do not depend on the level of coarse-graining and parameters like persistence length. One such measure is the decay of the contact probability and abundance with genomic separation g. Consider a random walk polymer chain. Clearly, the probability that two beads n_1 and n_2 come into contact decreases with the separation $|n_2 - n_1|$. More precisely, we obtain a power-law behavior [78]

$$p_c(|n_2 - n_1|) \simeq |n_2 - n_1|^{-3/2}. \tag{10.6}$$

Consider two genes separated by 10 Mb. Assuming a Kuhn segment length of 300 nm [30] consisting of 30 kb chromatin, the probability of co-localization is in the order of $p_c(10\,\mathrm{Mb}) \simeq 333^{-1.5} \sim 10^{-4}$. How, then, does the cell nucleus manage to co-locate different chromatin segments in a reasonable time? To answer this question, we look at the formation of functional loops in our model and its size distribution $P(\ell)$. Interestingly, the diffusional pathway to loop formation results in a size distribution of functional loops $P(\ell)$ which is quite different from the small random contact probabilities of a RW or SAW model (Figure 10.2C). Strikingly the probability of having a loop in the size-range of the

chain length is enhanced by over two orders of magnitude. The increase in probability for large-scale loops in contrast to small-scale loops can be explained on an intuitive basis: Starting from a linear chain, the diffusional process will bring monomers close together which are not so far away along the contour of the chain. Loop formation will be dominated by small-sized loops as equation (10.6) still holds. However, as more and more small loops form, even parts of the polymer located further apart come closer together (Fig. 10.2A), thus enhancing the probability of contact. Figure 10.2B visualizes for one simulation run the average loop size along the simulation time. We find that this average loop size increases fast and then fluctuates around an equilibrium value. Therefore diffusional looping seems to be a quite fast and effective method of large loop formation.

To allow a comparison to experimental data from 4C and 5C experiments we determine two measures. Firstly, the size-distribution $h(l)$ of random contacts (as nC experiments do not only measure functional contacts) between two chromatin segments. Secondly, the specific contact probability $p_c(|n_1 - n_2|)$ that two segments at position n_1 and n_2 are in contact. From eq. (10.6) we know that for a random walk the specific contact probability has a power-law behavior depending on the length $l = |n_1 - n_2|$ given by $p_c(l) \simeq l^{-\beta}, \quad \beta = 1.5$. A power-law behavior is also found for the self-avoiding walk, where the exponent is determined in Fig. 10.2D to $\beta = 2.10$. In fact, scaling theory predicts [78] for self-avoiding or random walk polymers that the contact probability of the end-points of a polymer scales as $p_c(N) \simeq N^{-3\nu} \approx N^{-1.76}$. Our analysis suggests that the contact probability for intra-chain segments decreases more strongly. This is somehow expected, as intra-chain segments have less entropic degrees of freedom and are surrounded by a higher density of adjacent beads than the end points, making contacts with beads further away less likely. Our polymer model as well displays a power-law behavior of the co-localization probability $p_c(l)$ (Fig. 10.2D). However, two different regimes have to be distinguished. For genomic separations l in the size range of the whole chromosome a different exponent is found as in the size range below about 15% of the fiber length. In the regime of probability-values p where leveling-off in the mean square distance is observed, we find exponents of about $\beta_1 \approx 0.8 - 1.1$ for smaller genomic separations in the order of 10 Mb and $\beta_2 \approx 0.35 - 0.70$ for large genomic separations in the order of 100 Mb (table 10.1). Intriguingly, the probability of specific contacts between far-apart chromatin segments is increased by over two orders of magnitude compared to the self-avoiding walk. Increasing the looping probability and thus the average number of loops per chain results in smaller exponents β. Interestingly, this result is in close agreement with recent results from Hi-C data [39] where an exponent of $\beta_1 \approx 1.08$ has been observed in a region between 500 kb and 7 Mb. The overall exponent found is even smaller on the size scale of the whole genome, consistent with our model. Similar results are found for other chain lengths (see Supplementary Information in sections 10.5.2 and 10.5.3 on pages 174 and 175).

5C data provides a detailed map of interactions between chromatin segments without a fixed reference point. Thus, it is more natural to look at the relative abundance of contacts $h(l)$ of size l, encompassing all fragments of a certain length l found in the data independent on their position on the genome. A crude power-law fit $h(l) \simeq l^{-\alpha}$ can be conducted here, too (see Supplementary Information in section 10.5.4 on page 176). We find power-law exponents of $\alpha \approx 0.7 - 1.2$ in the range where leveling-off in the mean square distance is observed (cf. Fig. 10.1). The exponents both for the specific contact probability $p_c(l)$ as well as the size-distribution $h(l)$ are listed in table 10.1.

Fig. 10.3 shows contact maps similar to those obtained by 5C for a $N = 256$ polymer with different looping probabilities. Contacts between any two beads are marked by a

Table 10.1: Decay exponents of the random contact probabilities with genomic separation for direct comparison to 4C and 5C experiments. Shown are the resulting exponents α of a power-law fit to the size distribution of random contacts $h(l) \sim l^{-\alpha}$. A fit to the specific contact probability $p_c(l) \sim l^{-\beta}$ has been performed both in the region of small genomic separations ($l \lesssim 10Mb$) and in the region of genomic separations up to the complete chromosome ($l \gtrsim 10$ Mb), yielding different exponents β_1 and β_2 respectively. These exponents can be compared to results from 5C and 4C experiments. Data is displayed for equilibrated chains of length $N = 256$ for various looping probabilities p, corresponding to different average numbers of loops, and different lifetimes τ of functional loops. For comparison with Fig. 10.1, the corresponding symbols are listed.

number of loops	loop lifetime τ	symbol	exponent α	exponent β_1	exponent β_2
19.0	τ_2	◇	2.01	1.78	1.66
19.0	τ_1	▲	2.00	1.76	1.68
19.1	τ_3	●	2.05	1.79	1.64
59.2	τ_2	◇	1.19	1.24	0.70
87.0	τ_1	▲	0.95	1.09	0.43
87.2	τ_3	●	0.92	1.11	0.38
112	τ_2	◇	0.84	1.00	0.30
131	τ_1	▲	0.81	0.94	0.35
247	τ_1	▲	0.70	0.79	0.35

black square. For better visibility, in each map contacts of 4 equilibrated conformations are plotted. Clearly, the self-avoiding walk polymer model (Fig. 10.3A) only has a few contacts between beads located far apart. Increasing the looping probability (Figs 10.3B and C) results in a strong increase of both the number of loops as well as the abundance of large loops.

10.3.3 Cell-to-cell variation and dynamic fluctuations of the distance distribution

While FISH measurements have been used to establish a connection between the mean square distance of two markers and genomic separation [27, 8, 12], a direct comparison to polymer models requires parameters to map one model bead to physical units like nanometers and base pairs. As these parameters are unknown or based on crude estimates [26, 8], it is desirable to introduce dimensionless quantities not dependent on length scale parameters.

For the the random walk (RW), self-avoiding walk (SAW) and the globular state (GS) model, the following higher-order moments of the distance distribution between two markers turned out to be basically independent of genomic separation [159].

$$c_2 = \frac{\langle R^2 \rangle}{\langle R \rangle^2}, \qquad c_3 = \frac{\langle R^3 \rangle}{\langle R \rangle^3}, \qquad c_4 = \frac{\left\langle R^4 \right\rangle}{\langle R^2 \rangle^2}. \tag{10.7}$$

An intrinsic advantage of these measures is that they are dimensionless, i.e. both experiments and models yield a numeric value. Even more important, the ratios carry information about the fluctuations, i.e. the cell-to-cell variation of the measurements.

One prominent feature of FISH measurements in interphase chromatin is that the fluctuations of the distance distributions are larger than expected from a random walk or self-avoiding walk polymer model [117]. Recently it has been shown that this holds true for

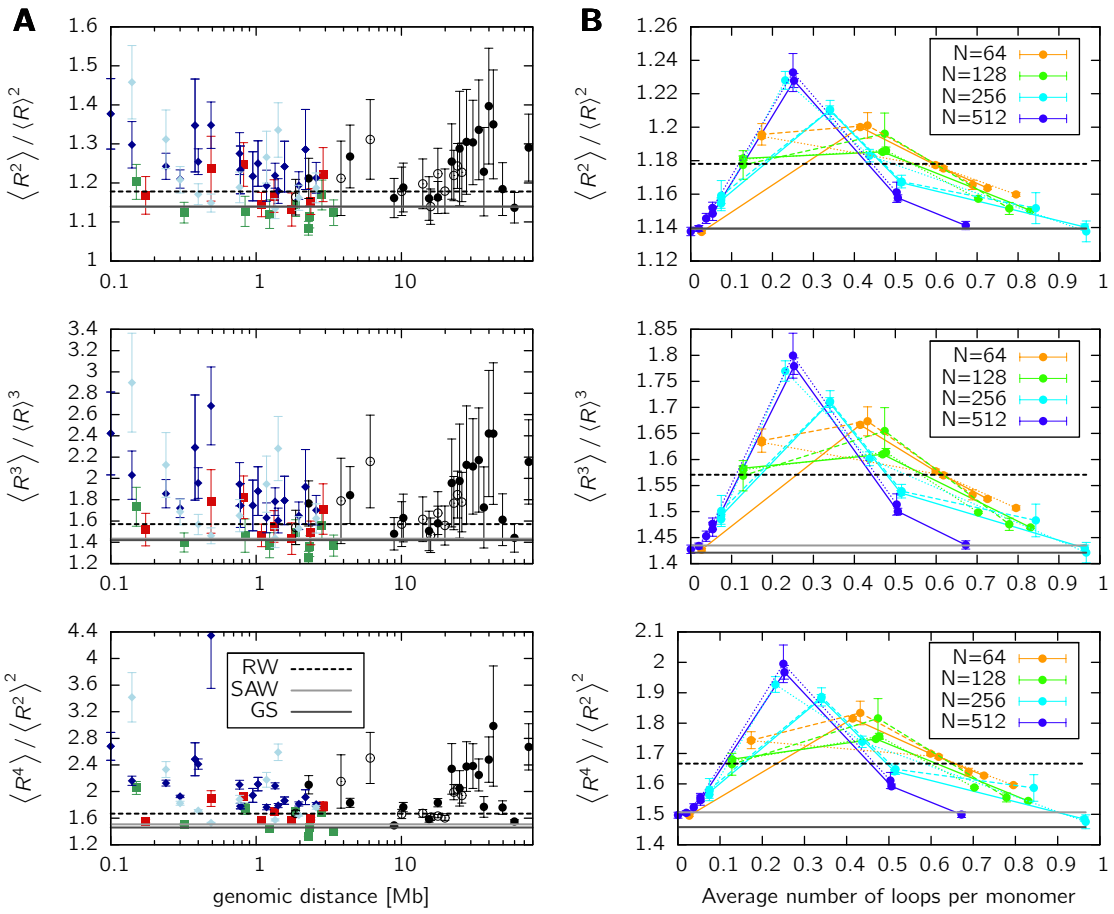


Figure 10.4: Higher-order moments of the distance distributions for experimental data (A) and for the chromatin model (B) according to eq. (10.7)**. A.** The following experimental data is shown: *Human fibroblasts Chr1* [12]: ■ anti-ridge region, ■ ridge region, ○ long distance measurements; *Human fibroblast Chr 11* [12]: • long distance measurements; *Murine Igh locus* [109]: ♦ pre-pro-B cells, ♦ pro-B cells. The data displays strong deviations towards larger fluctuations in comparison to the random walk (RW), self-avoiding walk (SAW) and globular state (GS) polymer model. **B.** Results are shown for simulated polymers of various length ($N = 64, 128, 256$ and 512) in relation to the average number of loops per monomer, which is related to the looping probability but allows for a better comparison. Although incorporating full excluded volume interactions, fluctuations exceed the random walk value due to probabilistic looping.

the case of compact polymers as well [159], where the fluctuations are even smaller. The ratios given in eq. (10.7) for experimental data sets from Mateos-Langerak *et al.* [12] as well as Jhunjhunwala et al [109] are presented in Fig. 10.4A. The figure contains FISH data from human chromosomes 1 and 11 [12], separately measured for ridges (green squares) and anti-ridges (red squares) as well as data from the murine Igh locus [109], which was kindly provided by K. Murre.

The results for the model treated in this paper are shown in Fig. 10.4B. Model polymers of different length ($N = 64, 128, 256$ and 512) have been equilibrated and averaged over a huge ensemble of conformations encompassing various configurations of loop attachment points. The data is plotted against the average number of loops per monomer to allow for a comparison between different chain lengths. For small looping probabilities, i.e. small average number of loops, the self-avoiding walk behavior is recovered, whereas increasing

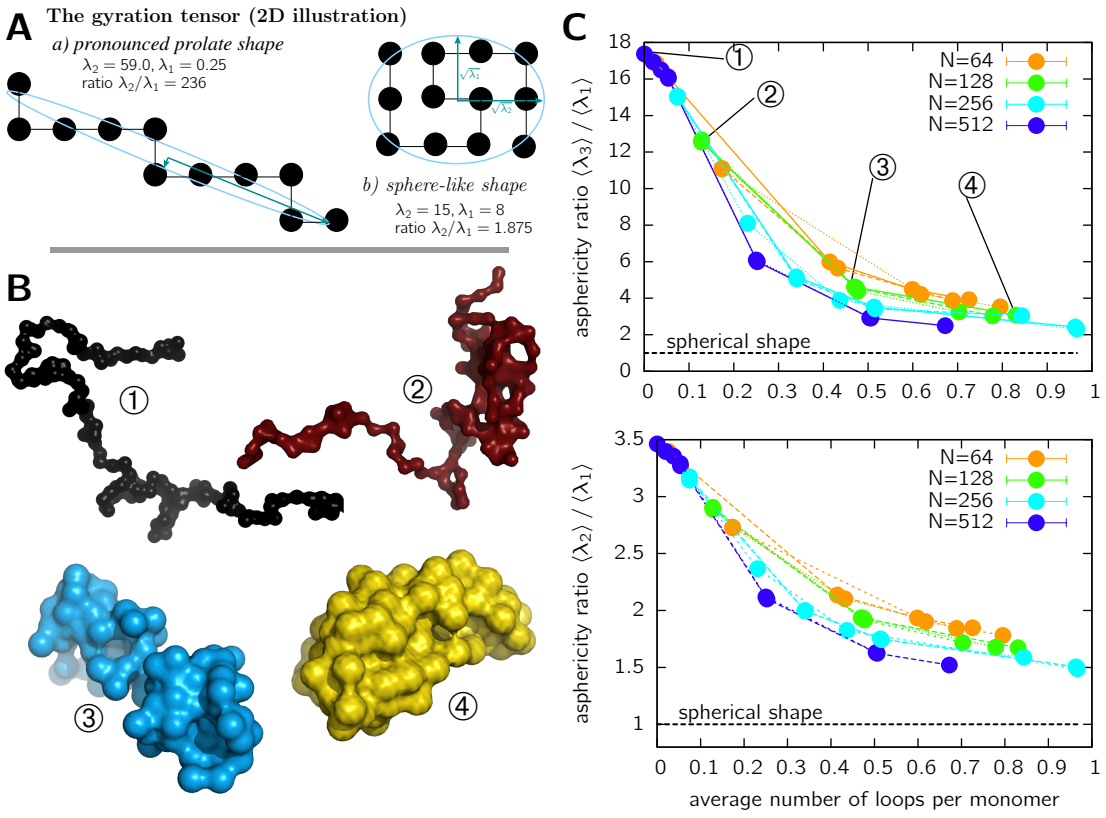


Figure 10.5: Elongated shape of the chromatin model polymers. **A.** Illustration of the gyration tensor. The gyration ellipsoid is shown for an elongated and a compact polymer conformations in two dimensions. The ratio λ_2/λ_1 is large for the elongated polymer, indicating strong devations from a sphere-like shape. **B.** Example conformations for a chain of length $N = 128$ and loop lifetime τ_1 (see eq. 10.2) for different looping probabilities. The shown conformations are one sample of the ensemble of conformations belonging to the data point marked in figure C. **C.** The ratios between the gyration tensor's main axes. The upper graph shows the ratio between the largest and smallest main axis, the lower graph the ratio between the second largest and smallest main axis. The data is shown for chain length up to $N = 512$, different lifetimes of the loops ($\tau = \tau_1$ solid line, $\tau = \tau_2$ dotted line, $\tau = \tau_3$ dashed line) and different looping probabilities p.

the looping probability leads to a strong increase in the fluctuations of the system. The higher-order moment ratios markedly exceed the random walk value in the range of loop numbers between $0.15N$ and $0.65N$. One would expect for a random-walk polymer to have larger fluctuations than a polymer constraint to excluded volume interactions and topological constraints. In our model, the large fluctuations are induced by the dynamic formation of loops, which thus seems to be an important characteristics of chromatin organization. However, it has to be noted that the fluctuations of the model are still too small to explain the moment ratios of most of the experimental data. We will discuss this in more detail in the discussion section.

10.3.4 The shape of chromosomes

Flourescent labeling of whole chromosomes [7, 19, 34, 58] or extended regions of chromosomes [6] can be used to determine the shape of chromosomes. Goetze et al. [6] quan-

titatively found that chromosomal regions show pronounced deviations from a spherical shape and are correlated to transcriptional activity. Using a wavelet method, Khalil et al. [36] found that the shape of a chromosome territory in mouse is highly nonspherical and can be best approximated by an ellipsoid with average axis length ratios of $4.5 : 2.9 : 1$. However, strong differences in shape and positioning were found depending on cell type [175, 176, 177] and chromosome state [37].

A way to characterize the shape of a polymer is the gyration tensor. It is defined by

$$S_{mn} = \frac{1}{N} \sum_{i=1}^{N} r_m^{(i)} r_n^{(i)} \,. \tag{10.8}$$

Here, $\mathbf{r}^{(i)}$ is the coordinate vector of the ith monomer and the subindex denotes its cartesian components. The eigenvalues $\lambda_1 \leq \lambda_2 \leq \lambda_3$ give the squared lengths of the principal axes of gyration. The ratios of the eigenvalues indicate the deviation from a sphere-like shape of the polymer. The gyration tensor is illustrated in a 2D sketch in Fig. 10.5A for an elongated as well as a compact polymer, yielding a pronounced difference in the ratio of the tensor's principal axes. While in an averaged sense, polymers display an isotropic behavior, single conformations are markedly aspherical [84, 83]. The self-avoiding walk, for example, has averaged eigenvalue ratios of $\langle\lambda_3\rangle : \langle\lambda_2\rangle : \langle\lambda_1\rangle = 14 : 2.98 : 1$.

In Fig. 10.5C the ratios $\langle\lambda_3\rangle / \langle\lambda_1\rangle$ and $\langle\lambda_2\rangle / \langle\lambda_1\rangle$ between the principal axes of the chromatin model are shown. Clearly, in the range of loop numbers where leveling-off occurs, the shape of the polymer is such that it is more elongated in one direction by a factor of $\sqrt{2}$ to $\sqrt{5}$. As for the distance fluctuations, this is in contrast to compact globular polymers that have a spherical shape, but in agreement with experimental [36] and simulational studies [30, 35]. Typical chromosome conformations are shown in Fig. 10.5B for different looping probabilities.

10.3.5 The dynamics of looping chromosomes

Finally we study the dynamics of the looping chromatin fibers. The center-of-mass motion of a polymer is measured by $g_3(t) = \left\langle (\mathbf{R}_{CM}(t) - \mathbf{R}_{CM}(0))^2 \right\rangle$. For a self-avoiding walk polymer it shows normal diffusion behavior, i.e. $g_3(t) \sim t$. As can be seen in Fig. 10.6 the chromatin model shows subdiffusive motion $g_3(t) \sim t^\alpha$ $(0 \leq \alpha < 1)$ on time scales smaller than the relaxation time of the polymer. The actual diffusion exponent α depends on the looping probability p. For times larger than the relaxation time one recovers diffusive motion, i.e. $g_3(t) \sim t$, however, this motion is slower than for a normal self-avoiding walk (see Supplementary Information in section 10.5.5 on page 176). This is consistent with experimental results showing that chromosome territories do not move significantly [32].

It has to be noted that the regime of large times is not very sensitive for a comparison to experimental data as here the confinement by other polymers comes into play which is not incorporated into the simulations of a single polymer. It is more instructive to look at the motion of the central monomers of a chain on short time scales. The mean square displacement $g_1(t) = \left\langle \left(\mathbf{r}_{N/2}(t) - \mathbf{r}_{N/2}(0) \right)^2 \right\rangle$ displays a distinct behavior for three different time regimes, which are related to the relaxation time τ_{int} of a chromosome. For $t \ll \tau_{int}$ there is a pronounced subdiffusive behavior. The anomalous diffusion exponents range from $\alpha \approx 0.22 - 0.4$ in the regime where leveling-off is observed for the mean square displacement (cf. Fig. 10.1). For $t \approx \tau_{int}$ the predictions of classical polymer dynamics become valid again and we find $g_1(t) \sim t^{0.59}$ similar to the self-avoiding walk. On large

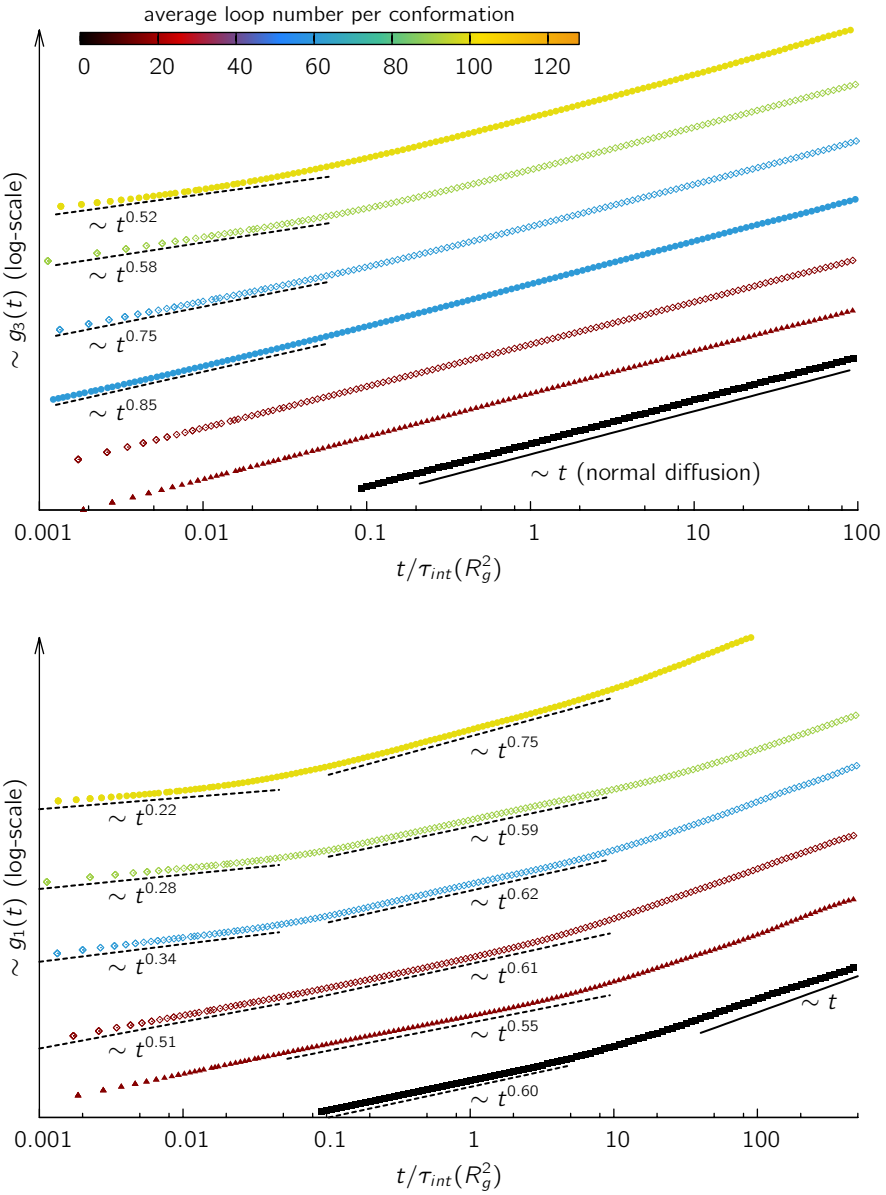


Figure 10.6: Dynamics of the center of mass and the central monomers. The upper figure shows the motion of the center of mass $g_3(t)$ for different parameters for a chain of length $N = 128$. The movement of the polymers' central monomer $g_1(t)$ is displayed in the lower figure. The color indicates the average number of loops per chain (see color bar), the point type indicates the loop lifetime (triangles [▲] for $\tau = \tau_1$, open diamonds [◊] for $\tau = \tau_2$ and filled circles [•] for $\tau = \tau_3$). We find subdiffusive behavior with different exponents dependent on looping probability for time scales below the relaxation time of the polymer. For reasons of readability curves are shifted along the y-axis relative to each other.

time scales ($t \gg \tau_{int}$), the monomer motion follows the motion of the center of mass, displaying normal Brownian motion with $g_1(t) \sim t$.

While at intermediate and large time scales the motion can be described by classical polymer theory, i.e. Rouse dynamics [79], the scaling exponents on the short time scale $t \ll \tau_{int}$ are unexpected. Following the argument in Refs. [97, 30], this time scale is the

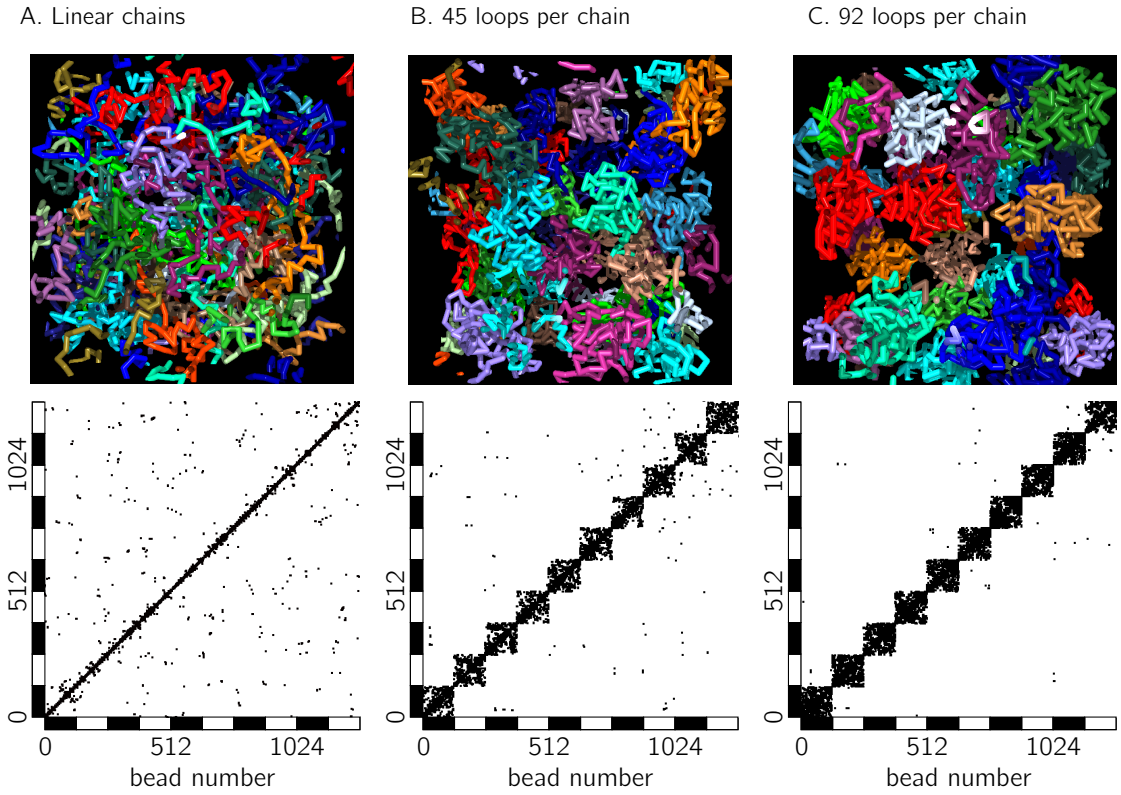


Figure 10.7: Contact maps and illustrations of chromosomes with different looping probabilities. Simulations were performed in a system with density $\rho = 12.5\%$ and chains with a coarse-grained length $N = 128$. Any contact between two beads is represented by a square in the contact map. Statistics is taken over 5 independent conformations. Not the complete contact map is shown, but only contacts between 10 chains. Linear chains (A) display a lot of intermingling and have abundant contacts with other polymers. The fraction of inter-chromosomal contacts is 15.4%. Increasing the loop-size (B and C) results in more and more confined structures, which are depleted of inter-chromosomal interactions. In (B) chains have on average 45 functional loops (symbol ▲ in Fig. 10.8A), the fraction of inter-chromosomal contacts is reduced to 1.8%. This value decreases even more for chains with an average of 92 loops per conformation ($< 1\%$, symbol ▲ in Fig. 10.8A).

prevailing one concerning interphase chromosomes. Clearly, such exponents arise due to the constraints induced by looping, which temporarily slows down the motion of chromatin segments at the loop attachment points. Although experimental data is rare, this is consistent with findings of Cabal et al. [17] in yeast. This study showed that the motion of a labelled spot scales like $g_1(t) \sim t^{0.41}$ up to $g_1(t) \sim t^{0.47}$. Interestingly they found the exponent to depend on the transcriptional state of the *GAL* genes. This is in support of our conjecture put forward in another publication [12] that the local looping probability may be related to transcriptional activity.

10.3.6 The formation of aspherical chromosome territories

Polymer theory predicts that equilibrated polymers with a large molecular weight in a semi-dilute solution are strongly intermingling [78]. Various studies, however, indicate that chromosomes occupy discrete functional domains [7, 178, 179]. It was shown above that

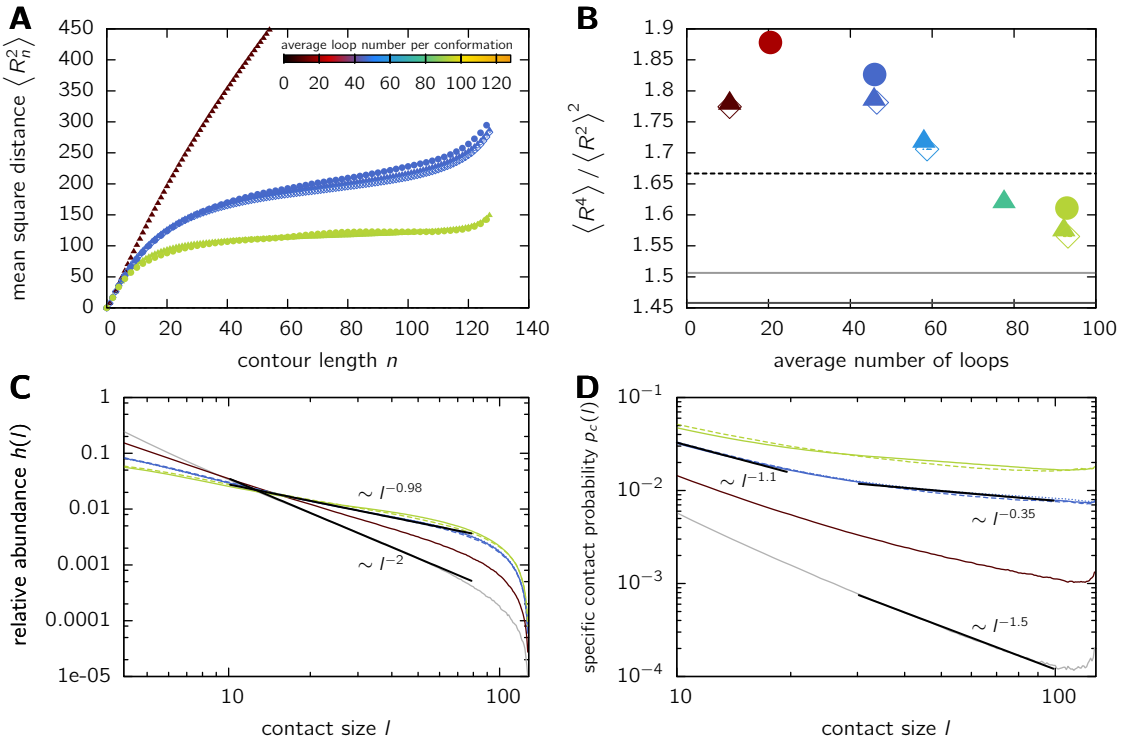


Figure 10.8: Properties of looping polymers in a dense system. Coarse-grained polymers of length $N = 128$ are equilibrated in a system of density $\rho = 12.5\%$. Results are shown for various looping probabilities, which are indicated by the color-coded average number of loops. **A.** Relationship between mean square distance $\langle R^2 \rangle$ and genomic separation (contour length n). Polymers with small looping probabilities (dark red curve) show a continuous increase of mean distance between two markers with their separation n. Thus, these polymers do not form discrete territories but intermingle strongly. If the average number of functional loops exceeds 40-50 loops per monomer, a leveling-off is observed and the chromosomes fold into a confined space. **B.** The ratio between higher-order moments $c_r = \langle R^4 \rangle / \langle R^2 \rangle^2$ indicates a regime of larger fluctuations than in the random walk case for polymers with p-values in the range where leveling-off occurs. The values found are in the size range of $c_4 \approx 1.75 - 1.85$. Such large fluctuations, i.e. cell-to-cell variation, are an intrinsic feature of chromatin organization (Fig. 10.4A), represented in our model by the dynamic formation of probabilistic loops. **C.** The size distribution of random contacts $h(l)$ demonstrates that diffusion-based looping facilitates the formation of large loops. Instead of decreasing with l^{-2} as in the case of linear chains, looping polymers in the parameter range where leveling-off is observed (cf. A) show a power-law behavior of approx. $h(l) \sim l^{-1}$. **D.** The probability $p_c(l)$ that specific loci on one chromosome co-localize as measured in 4C experiments displays approximately a biphasic power-law behavior. On the scale of the whole chromosome, the contact probability decreases with $p_c(l) \sim l^{-0.35}$, the exact exponent depending on the looping probability. On intermediate length scales a power-law of $p_c(l) \sim l^{-1.1}$ is found in agreement with experimental data [39]. Again, the co-localization probability is greatly enhanced by the formation of functional loops.

our model polymers adopt a confined structure by virtue of dynamic looping. Amazingly, this result was obtained without subjecting the system to a confined space (in contrast to Refs. [26, 31]) and without introducing long-range interactions (in contrast the polymer models in Refs. [11, 29, 117]).

Surely, simulating isolated chromatin fibers does not yield complete information about the folding in a dense system as in the nucleus, e.g. the formation of chromosome territories

(CTs). To investigate whether probabilistic loops are the reason for the formation of chromosome territories, we set up simulations of chromosomes in a box of width $L = 64$ lattice units and a length of $N = 128$. The density of the system $\rho = 0.125$ was chosen close to the estimates of chromatin in cell nuclei. Similar values were used in other publications [35].

An established measure of territory formation is the number of contacts displayed in the contact map [35]. Figure 10.7 shows such contact maps for different looping probabilities p. Each map displays contacts between the beads of a subset of 10 model chromosomes out of the system. The beads are numbered consecutively, i.e. bead 0 to 127 belong to chain 1, bead 128 to 256 belong to chain 2 etc. Subsequent chains are alternatingly marked by black and white bars. We find that linear self-avoiding walks (Fig. 10.7A) without loops display a relatively large number of inter-chromosomal in comparison to intra-chromosomal contacts: 15.4 % of the contacts are found to be with other chromosomes. With increasing looping probability p, the percentage of contacts between different chromosomes decreases. Fig. 10.7B displays chains with an average number of 45 loops (symbol ▲ in Fig. 10.8A). Here we find that only 1.8 % of the contacts are inter-chromosomal and in Fig. 10.7C (92 loops per chain on average, symbol ▲ in Fig. 10.8A) this value reduces further to $< 1\%$. Thus, the level of intermingling between CT's strongly depends on the local looping probabilities. As different local looping probabilities seem to play a dominant role in chromatin organization, this finding could explain different levels of intermingling found in several studies [60, 34]. Thus a disentangling of the fibers, which has been estimated to require a huge amount of time or the action of topoisomerase II [97] is not necessary. In mitosis, chromatin adopts a compact state, where different chromosomes are unentangled and well-separated. At the onset of interphase, the loop formation forces the chromosomes to a more open, but confined structure, which results in the formation of CTs without requiring the assumption of unequilibrated polymers [30].

We find that the predictions from the study of isolated model chromosomes are still valid for a dense system of chromatin. Amongst others, this is a direct consequence of loop-based segregation observed in Fig. 10.7. Fig. 10.8A shows the mean square distance between two model beads in relation to contour length n (in biological terms: genomic distance g). Similar to the results of Fig. 10.1, the mean square distance displays a leveling-off for average loop numbers larger than about 45 loops per coarse-grained monomer. Obviously, for small looping probabilities p (red curve) or self-avoiding walks ($p = 0$, not shown), polymers do not level off, thus they do not form separate territories. The behavior of territory formation and segregation is a distinct result of loop formation.

While the mean square distance $\langle R_n^2 \rangle$ displays a leveling-off for several polymer models (e.g. globular state [159, 31], random walk in a confined space [26], etc.), a more sensitive measure are again the dimensionless ratios of higher order moments given by eq. (10.7). As the fluctuation regime could possibly change under the transition from isolated polymers to a dense system, we investigate the ratio $c_4 = \left\langle R^4 \right\rangle / \langle R^2 \rangle^2$. Fig. 10.8B shows that fluctuations are larger than predicted by the random walk, self-avoiding walk or globular state model. In the regime where a leveling-off is obtained in Fig. 10.8A, i.e. the average loop number is larger than about 45, the moment ratios are approximately in the range $c_4 \approx 1.75 - 1.85$.

The relative abundance of contacts $h(l)$ is displayed in Fig. 10.8C for polymers in a dense system. Again, the co-localization frequency is greatly enhanced by the formation of functional loops. A crude power-law fit $h(l) \sim l^{-\alpha}$ results in exponents of $\alpha \approx 1$ and smaller in the region where a leveling-off is observed in the mean square distance. Similar

Table 10.2: Shape parameters of simulated chromosomes. Shown are the results for equilibrated coarse-grained polymers of length $N = 128$ in a melt of density $\rho = 12.5\%$. Results have been calculated using various looping probabilities and lifetimes of the loops. Correspondence to Fig. 10.8A is established via the symbol, which is shown in the third column. The shape is parameterized by the ratios of the eigenvalues of the gyration ellipsoid, corresponding to the squares of its axis lengths. The axis ratios $a : b : c$ are listed for comparison with other studies (e.g. Cook et al. [35])

number of loops	loop lifetime	symbol	eigenvalue ratios	axis ratios
10.5	τ_1	▲	10.2 : 2.6 : 1	3.2 : 1.6 : 1
45.9	τ_1	▲	4.9 : 2.0 : 1	2.2 : 1.4 : 1
45.9	τ_3	●	5.0 : 2.0 : 1	2.2 : 1.4 : 1
92.1	τ_1	▲	3.3 : 1.7 : 1	1.8 : 1.3 : 1
93.0	τ_3	●	3.2 : 1.7 : 1	1.8 : 1.3 : 1

results are found for the specific contact probability $p_c(l)$ (Fig. 10.8C), which display a biphasic behaviour already observed in the case of isolated chromosomes (Fig. 10.2). In the size range of large genomic separations in the order of the entire chromosome, the contact probability decreases with a power-law $p_c(l) \sim l^{-\beta}$ with exponents starting from $\beta = 1.5$ in the self-avoiding walk model to $\beta = 0.35$ in the parameter range where leveling-off is observed. On intermediate scales ($\lesssim 10$ Mb), for biologically relevant looping probabilities, an exponent of $\beta = 1.1$ is found. Amazingly, a similar value of $\beta = 1.08$ has been recently found by Hi-C experiments [39] on a scale between 500 kb to 7 Mb.

We found that isolated model chromosomes displayed a pronounced conformational asphericity (Fig. 10.5). A similar behavior is observed for chromosomes in a dense system. In fact, deviations from a sphere-like shape are expected for the self-avoiding walk as well as the random walk model [83], however, not for a compact globular state polymer [159]. Whereas looping polymers can adopt a highly compacted state, their properties differ clearly from a globular state. Indeed, the shape of simulated chromosomes territories is not spherical as one would expect for compact polymers, rather we find that the gyration ellipsoid has a prolate shape. The ratios of the gyration tensor's eigenvalues are listed in table 10.2. In the parameter range where the mean square distance displays a leveling off, we find ratios of the eigenvalues $\langle\lambda_3\rangle : \langle\lambda_2\rangle : \langle\lambda_1\rangle$ in the regime between 5 : 2.0 : 1 and 3.2 : 1.7 : 1. These values are smaller than those of Rosa *et al.* [30] for ring polymers and consistent with those of looping polymers [35]. A non-spherical shape of CTs has also been found in experimental studies [36, 37]. Mouse chromosomes exhibit an aspherical shape approximated by ellipsoids with axis ratios 4.5 : 2.9 : 1. A one-to-one correspondence of these numbers with results from the shape of the gyration ellipsoid, however, can not be established.

10.4 Discussion

In this study, a polymer model was presented where loops form dynamically on the basis of diffusional collisions. We use Monte Carlo simulations to demonstrate the effect of such a kind of loop formation. While loops have been recognized as an ubiquitous feature in transcriptional regulation, the pathways of its formation remained unclear and most polymer models proposing loops so far did not explain the transport mechanisms by which

two parts of chromatin become co-located. Our results suggest that even large loops can arise without active transport mechanisms. Our model neither assumes a confined geometry nor any long-range interactions. Loop formation is based on the diffusional motion of the fiber. Collisions lead to a probabilistic chromatin-chromatin interaction which forces the participating regions to be co-located for a certain time. The probabilistic nature of the interactions is meant to mimic the effect of chromatin binding factors on chromatin-chromatin interactions. Although this Dynamic Loop model is kept minimal, it reproduces many experimental results quantitatively, highlighting the possibility that chromatin folding is tightly related to function through the loop formation process.

One of our major results is that dynamic loop formation drives chromosomes into an entropically segregated state. Indeed, linear polymers intermingle freely (Fig. 10.7A) in agreement with polymer theory [78]. Looping polymers, in contrast, fold into a confined space (Fig. 10.1). Such a confinement is also observed for the globular state polymer model, which, however, displays a markedly different fluctuation regime than the experimental data [159]. The importance of looping on the formation of chromosome territories has been investigated recently by Cook *et al.* [35]. In their qualitative study, rosette-structures with fixed loop attachment points are used. While this model can be used as a simple model for studying entropic effects of looping, it does not explain fluctuations in FISH data [117, 12].

The second important result of this study concerns the pathway of the formation of large loops. Results from 3C/4C/5C experiments reveal that loops are abundant on the short scale [66]. Nevertheless, functional loops on the scale of several mega basepairs have been detected in 4C experiments [25]. While the probability of specific random contacts $p_c(l)$ decreases strongly with site separation l for linear polymers (given by a power-law behavior $l^{-2.10}$), the contact probability is increased by over two orders of magnitude when introducing loops (Fig. 10.2). Obviously, small functional loops which can easily co-localize by diffusional motion strongly support the formation of long-distance contacts. The contact distribution $h(l) \sim l^{-\alpha}$ displays power-law exponents of $\alpha = 0.7 - 1.2$, the contact probability $p_c(l) \sim l^{-\beta}$ exponents in the size range between $\beta \approx 0.3 - 0.7$ on the scale of the whole chromosome and $\beta \approx 0.7 - 1.3$ on intermediate scales. This is in agreement with recent experimental data by Lieberman-Aiden *et al.* [39], however, their interpretation in terms of a fractional globule differs from ours. Clearly, a fractional globule, where the physical distance between two loci scales with $s^{1/3}$ [98], is in contrast to experimental findings from FISH data [12].

The impossibility to perform Monte-Carlo simulations on a very detailed scale requires a coarse-graining procedure. Looking at large-scale features above the persistence length l_p, such an approach is well-justified [78]. For linear polymers, scaling laws provide a simple way of rescaling a polymer. For a model with loops, the connection between chain length, bond length and looping probability p is non-trivial. In fact, even for chromatin models using linear chains (see Refs. [26, 30, 31]), the establishment of a correspondence between simulational units and biological units requires the knowledge of the persistence length of chromatin. The latter has been estimated by fitting a random walk model [26] or a worm-like chain model [8] to FISH data. Estimates on the persistence length range from 40-220 nm [91]. For a quantitative comparison to experimental data we derived measures independent on both the level of coarsening and unknown biological parameters. These can be easily evaluated both for experimental data as well as polymer models. Amongst others, these measures comprise the power-law exponent of the contact distribution (Fig. 10.2 and table 10.1), the dimensionless higher-order moment ratios of the distance distribution between two FISH markers (Fig. 10.4), the asphericity of chromosomes (Fig. 10.5) and

finally the diffusion exponents (Fig. 10.6).

The DL model studied here displays a pronounced aspherical elongated shape (Figure 10.5) which is also found in experiments [6]. Consistent with experimental data in yeast, the motion of single monomers is subdiffusive (Fig. 10.6); the actual subdiffusion exponent depends on the looping probability, which was suggested to be closely related to transcriptional activity [12]. A good agreement with experimental data is obtained for the higher-order moments of the distance distribution. Surprisingly, the moment ratios (given in eq. (10.7)), which display fluctuations of the distance distributions, exceed the random walk value for looping probabilities p in the range where a leveling off in the mean square distance is observed. This is not necessarily expected for a model with excluded volume which restricts the degrees of freedom and therefore shows less fluctuations. The increase of the fluctuations with respect to the self-avoiding walk is due to the dynamic formation of loops. However, several independent experiments [109, 12] consistently show even larger fluctuations. We suspect two major reasons for this: First, the chromatin fiber is not a homogeneous polymer and there is evidence that looping probabilities vary depending on the transcriptional state [12]. Secondly, inside the nucleus, topoisomerase-II might effectively counteract excluded volume interactions, resulting in an underestimate of the fluctuations in our model.

While our model suggests that chromosome segregation might be driven by the diffusional formation of loops, Rosa and Everaers suggested [30] that segregation is a consequence of large entanglement times. The entanglement times, however, might be strongly reduced by the effect of topoisomerase-II [154]. Notwithstanding that time-scales play an important role, this study reveals that loop formation provides a complementary and fully sufficient mechanism for CT formation.

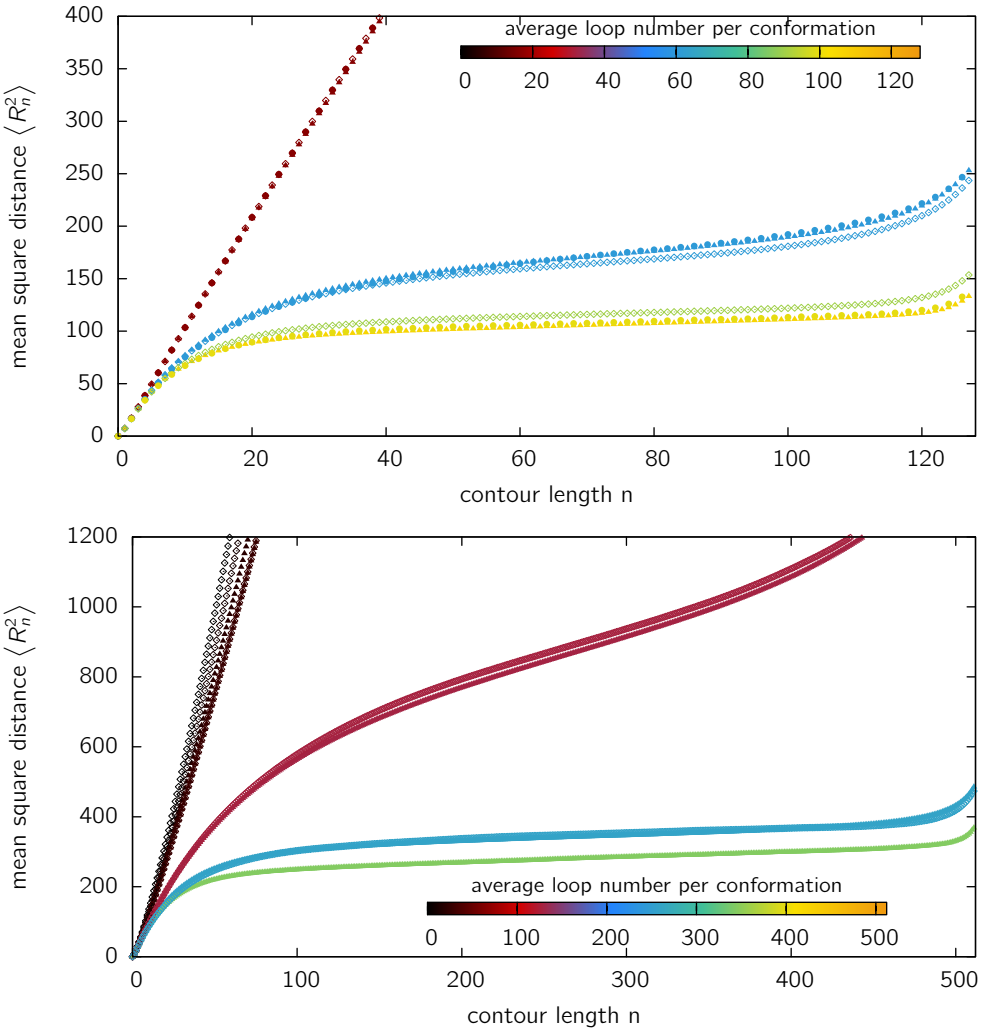
Clearly, the chromatin model proposed here does not capture all details of the complex nuclear organization. First of all, we neglect the heterogeneity of the chromatin fiber and assume the same looping probability and chromatin affinity along the complete chromosome. Thus, regional differences in compaction are not explained by this model. Such differences are clearly found in experiments, showing that gene rich regions have a more open structure than gene poor regions [6, 12]. In principle, such information could be incorporated into the model by assuming different looping probabilities for different regions. Such an analysis has been performed for the Random Loop model (↷ chapter 7) without excluded volume interactions. A similar approach could be pursued with the Dynamic Loop model, requiring, however, a huge amount of computer time.

The direct interaction of DNA-binding factors with DNA and the effect of such proteins on maintaining genomic loops is not included in detail into the model. A natural extension of the model would be to put proteins inside the simulation box that bind to DNA and mediate the loop formation. Then a loop between two chromatin segments can only form if a DNA-binding protein is in proximity. In a next step, the specificity of binding sites (e.g. CTCF binding sites) could be modelled. However, a detailed account on protein-DNA interaction is not possible at the current state of scientific knowledge as by far not all possible interactions and binding sites are known. Therefore, in our model, we have taken the effects of the solvent into account on an effective basis: The probabilistic formation of loops mimics the effect of DNA-binding proteins being there or not.

10.5 Supplementary Information

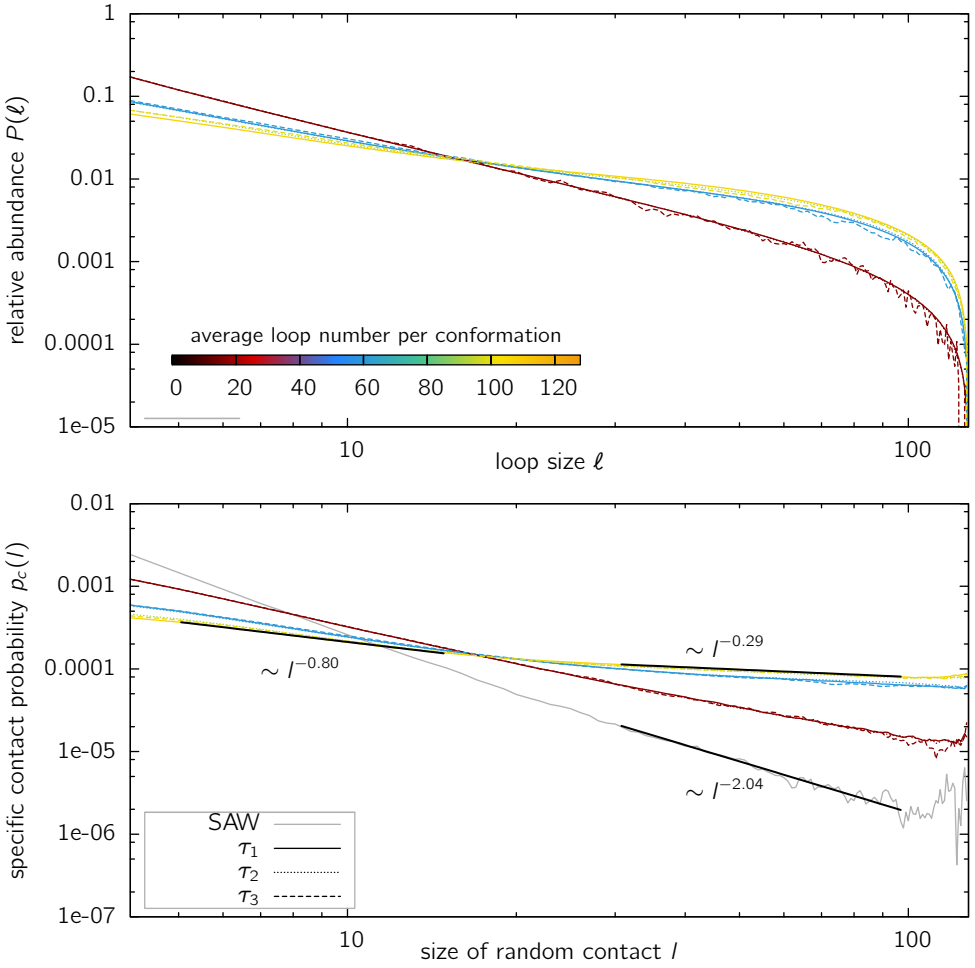
10.5.1 Supplementary Figure 1

Mean square distance $\langle R_n^2 \rangle$ in relation to contour length for an isolated fiber. Both figures show the mean square distance between two beads separated by contour length n. In the upper-most figure, results for a chain length of $N = 128$ are shown. The bottom-most figure shows results for $N = 512$. Isolated polymers have been fully equilibrated for various looping probabilities p. The p-values are plotted with different colors depending on the resulting average number of loops per conformation. Simulations have been performed using various lifetimes of loops, which are chosen relative to the relaxation time τ_{int} of the polymer. The results are displayed by different symbols (triangles [▲] for $\tau = \tau_1 = 0.01\tau_{int}$, open diamonds [◊] for $\tau = \tau_2 = \tau_{int}$ and filled circles [•] for $\tau = \tau_3 = 100\tau_{int}$). The mean square distance displays a leveling-off for all chain lengths studied, inducing a confined folding of chromosomes based on dynamic looping.



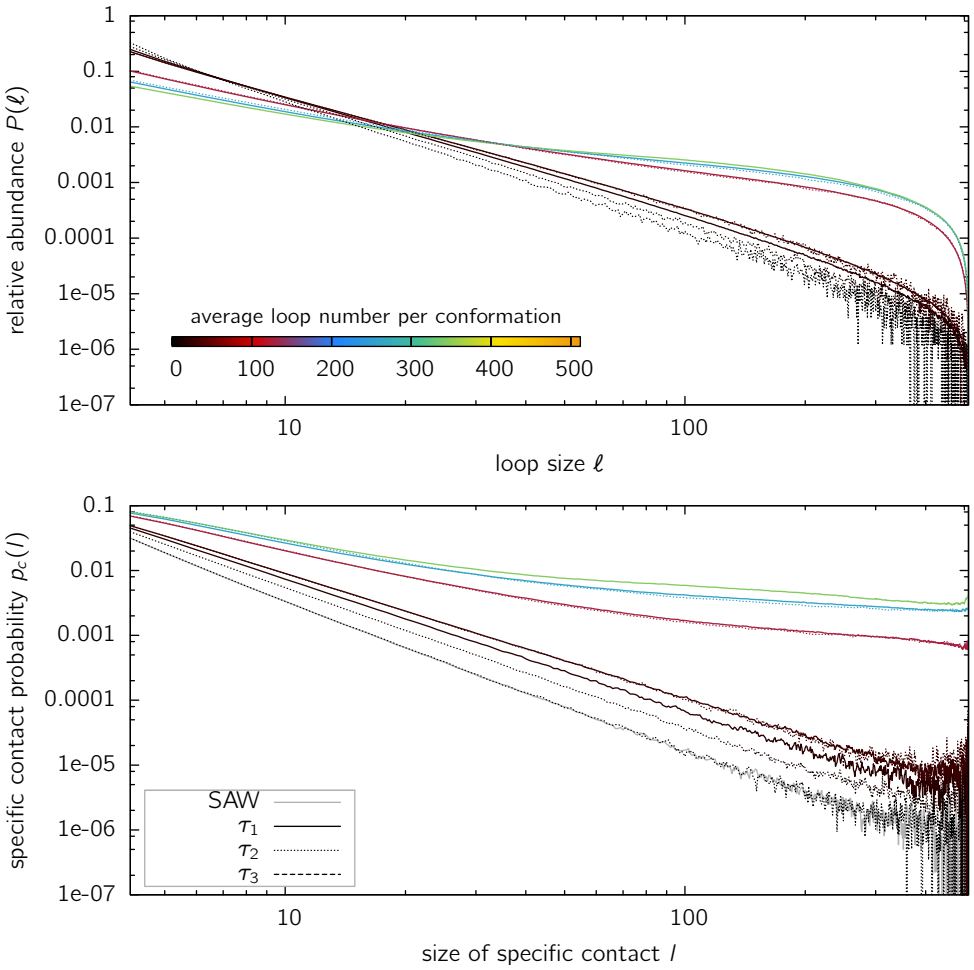
10.5.2 Supplementary Figure 2

Loop size distribution and specific contact probabilities for $N = 128$. Upper-most figure: Shown is the size distribution $P(\ell)$ of functional chromatin loops of model polymers with $N = 128$ beads. Model polymers were fully equilibrated and the loop size distribution was determined for various looping probabilities p (for reasons of comparison the average number of loops per conformation is displayed by a color code) and lifetimes τ of the functional loops. Looping lifetimes are chosen relative to the relaxation time (cf. Materials & Methods in the manuscript). Increasing the loop number results in a markedly smaller exponent, leading to a high probability for large loops. **Bottom-most figure:** The contact probability $p_c(l)$ for two specific sites with genomic separation (contour length) l to be co-localized. Shown are the results for equilibrated model polymers with $N = 128$ beads and various looping probabilities p. The contact probability decreases as a power-law $l^{-\beta}$ with a biphasic behaviour, the exponent changing at about $l \approx 15\%$ of the chromosome length. The grey line represents the self-avoiding walk. Again, the co-localization probability is strongly increasing due to diffusion-based looping.



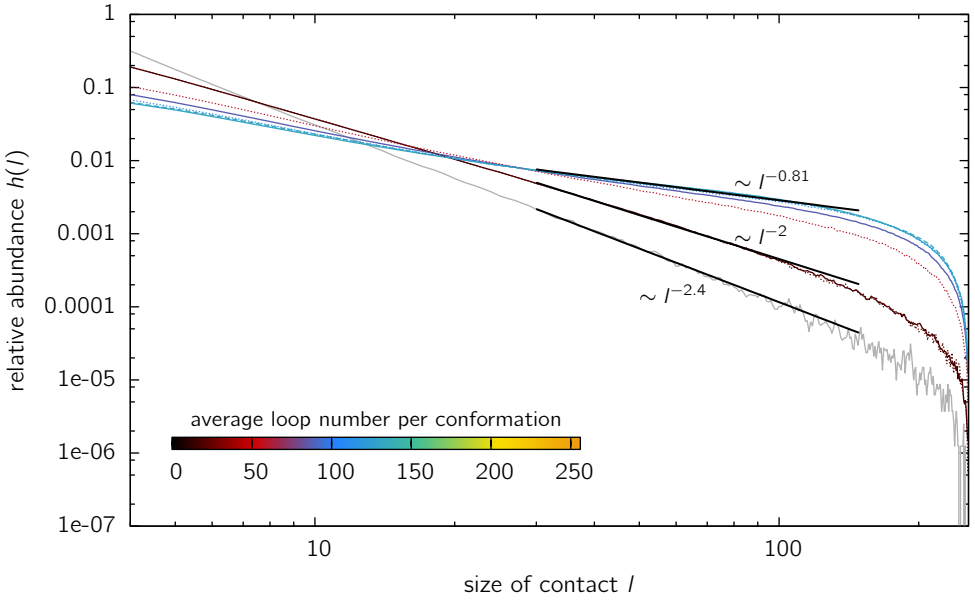
10.5.3 Supplementary Figure 3

Loop size distribution and specific contact probabilities for $N = 512$. Upper-most figure: Shown is the size distribution $P(\ell)$ of functional chromatin loops of model polymers with $N = 512$ beads. Model polymers were fully equilibrated and the loop size distribution was determined for various looping probabilities p (for reasons of comparison the average number of loops per conformation is displayed by a color code) and lifetimes τ of the functional loops. Looping lifetimes are chosen relative to the relaxation time (cf. Materials & Methods in the manuscript). Increasing the loop number results in a markedly smaller exponent, leading to a high probability for large loops. **Bottom-most figure:** The contact probability $p_c(l)$ for two specific sites with genomic separation (contour length) l to be co-localized. Shown are the results for equilibrated model polymers with $N = 512$ beads and various looping probabilities p.The contact probability decreases as a power-law $l^{-\beta}$ with a biphasic behaviour, the exponent changing at about $l \approx 15\%$ of the chromosome length. The grey line represents the self-avoiding walk. Again, the co-localization probability is strongly increasing due to diffusion-based looping.



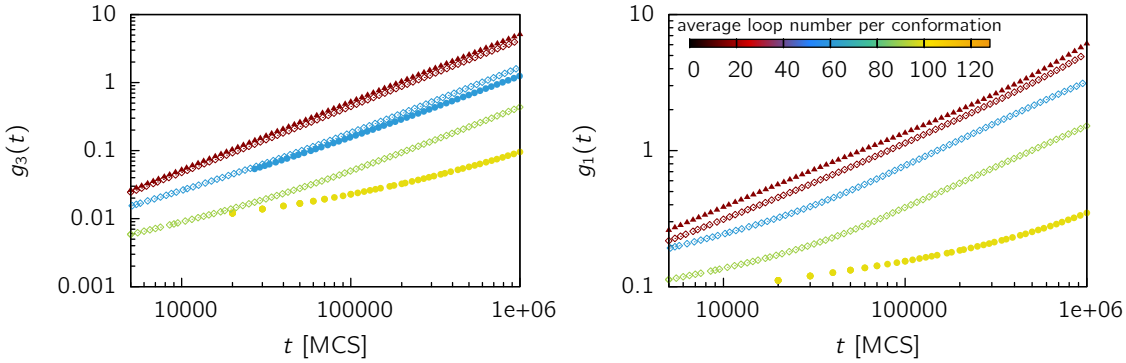
10.5.4 Supplementary Figure 4

Relative abundance of contacts $h(l)$ in relation to genomic separation l. The size distribution of random contacts $h(l)$ is shown for chains of length $N = 256$. Simulations for different looping probabilities are indicated by a color code. A crude power-law fit $h(l) \sim l^{-\alpha}$ has been performed to the data, showing that the exponent decreases with increasing looping probability.



10.5.5 Supplementary Figure 5

Dynamics of the center of mass and motion of the central monomers (unshifted plots). The left-hand shows the motion of the center of mass $g_3(t)$ using a chain length of $N = 128$. The data is shown for different values of the looping probability p, indicated by a color-code for the average number of loops. The lifetime of loops is indicated by the symbol (triangles [▲] for $\tau = \tau_1$, open diamonds [◇] for $\tau = \tau_2$ and filled circles [●] for $\tau = \tau_3$). The movement of the polymers' central monomer $g_1(t)$ is displayed in the right-hand figure. The unshifted data shows that introducing loops in the system slows down the motion of the polymer as well as of single monomers.



Chapter 11

Repulsive forces between chromosomes

References

The results presented in this chapter exist as a prerint and are going to be submitted as soon as the paper on the Dynamic Loop Model (Ref. [93]) is accepted.

- M. Bohn & D. W. Heermann, Repulsive forces between chromosomes. **2010**. *preprint, to be submitted.*

Chapter Summary

One striking feature of chromatin organization is the distinct separation of chromosomes into territories during interphase, showing only little intermingling. This behaviour has been explained by a non-equilibrium state of linear chains. Here, we show that entropic forces of loop formation alone without assuming other long-range interactions lead to a strong repulsion between chromosomes. We determine the effective potential between two chromosomes modelled by the Dynamic Loop model (↷ chapter 10). We find that introducing loops into the structure of chromatin results in a many-fold higher repulsion between chromosomes. Strong effects are observed for the tendency of a non-random alignment, the overlap volume between chromosomes decaying fast with increasing loop number.

11.1 Introduction

Chromosomes, in comparison to other polymeric systems, display a vast amount of unexpected types of behaviour. Most importantly, chromosomes are highly compartmentalized objects being well separated during interphase [7]. Such a separation is not only observed regarding complete chromosomes, rather Mb-sized stretches of chromatin, when labelled fluorescently with different colors, also display little intermingling [6]. A lot of speculation has been going on about the mechanisms driving such kind of segregation. While active ATP-consuming mechanisms might be possible, it has been proposed that compartmentalization results from non-equilibrium effects: During metaphase chromosomes are condensed and well separated. The entanglement time disregarding topoisomerase-II activity is supposed to be much larger than the lifetime of the cell [30]. Segregation of intrachromosomal region, however, is not explained in such a model. Recently, there is an ever growing body of evidence that the entropic effect of looping alone might be responsible for maintaining the compartmentalized state of the chromosomes [35]. A lot of work on entropic effects of loops has already been conducted in this thesis. We have shown that a polymer model allowing for dynamic loop formation induces a segregation of chromosomes (↷ chapter 10). Ring polymers were found to display a much stronger entropic repulsion than linear polymers (↷ chapter 9), the overlap volume between bonded rings being significantly smaller than for linear polymers (↷ chapter 8). However, a crucial question remains: How does the existence of multiple loops in the Dynamic Loop model affect the inter-chromosomal forces?

The scope of this chapter is to extend the study of effective interactions between ring polymers to the more complex system of looping chromosomes. Amongst others, we want to investigate how the strength of the repulsive interaction changes when introducing more and more loops into the system. While a coarse-grained model of the chromosome is employed here, a mapping to physical units can be conducted using results from fluorescence in situ hybridization (FISH) experiments [12]. Experimentally, the forces between two chromosomes have not been determined, but in principle such experiments are possible.

11.2 Methods

11.2.1 Computer simulations of chromosomes

Model chromosomes are simulated using the Dynamic Loop model introduced in chapter 10. Simulations are conducted on a lattice using the bond fluctuation method [155] to speed up the computation of excluded volume interactions. We perform simulations of isolated chromosomes with lengths ranging from $N = 128$ to $N = 1024$. After initial equilibration steps, loop formation is accomplished by diffusional motion: Whenever two polymer segments come into contact, a bond is created with probability p for a certain lifetime τ. Details are explained in section 10.2. Looping lifetimes are chosen relative to the relaxation time of the corresponding isolated polymers without loops. As it has been shown in chapter 10 that the chosen lifetime τ does not strongly influence equilibrium properties, parts of the analysis presented here are restricted to the choice $\tau = \tau_1 = 0.01\tau_{int}$ [see eq. (10.2)].

11.2.2 Calculation of the effective potential

To analyze the strength of the repulsive interactions, the potential acting between the chromosomes' centers of mass is determined using the method introduced by Dautenhahn and Hall [167] and described in detail in section 9.2.3. In short, two equilibrated isolated chromosome conformations are selected and shifted such that the distance between their centers of mass equals r. If the excluded volume condition is satisfied, i.e. no lattice site is occupied by more than one bond, the conformation is accepted, otherwise it is rejected. The fraction of accepted conformations N_{accepted} to the total number K of trial conformations defines the effective potential at distance r,

$$U_{\text{eff}}(r) = -k_B T \ln \frac{N_{\text{accepted}}}{K}$$

From the set of accepted two-chain conformations, the conformational properties can be calculated.

11.3 Results

11.3.1 The effective repulsion between chromosomes increases strongly with loop number

What happens when two polymeric coils are brought closely together? Clearly, in the absence of other interactions than excluded volume forces, polymers repell each other due to the constrained conformational space. Such a behaviour has been found both for linear self-avoiding walks [167] and ring polymers (Ref. [180] and chapter 9). Here, we investigate the potential of mean force between the centers of mass of two chromosomes modelled by the Dynamic Loop model (↷ chapter 10). Results are shown for chain lengths $N = 128, 256$ and $N = 512$ in Figure 11.1. In principle, depending on the coarse-graining used, such chains could represent small chromosomal regions up to whole chromosomes. To allow comparison for different sets of parameters (chain length N, looping probability p, lifetime of loops τ), the center-of-mass distance is scaled with the mean radius of gyration R_g of the corresponding isolated chains. The radius of gyration is a measure of the typical size of a chromosome, i.e. the chromosome territory.

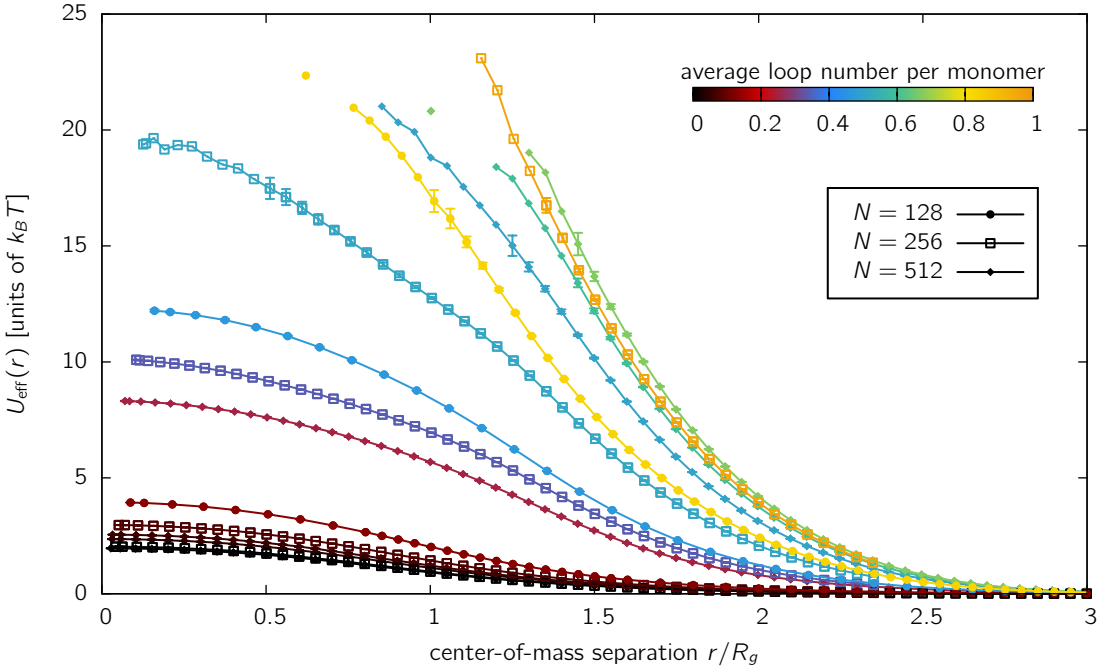


Figure 11.1: The effective potential $U_{\text{eff}}(r)$ between the centers of mass of looping polymers. Simulations have been conducted using the DL model (↷ chapter 10) for different chain lengths N and looping probabilities p. The average number of loops per monomer resulting from the parameter p is indicated by the color bar, different chain lengths by their point symbol. Data is scaled with the radius of gyration of isolated polymer chains to allow comparison between different parameters sets. The effective potential increases strongly with looping probability p, the order of magnitude being independent of chain length N.

Evidently, the effective potential increases when approaching the two chromosomes, i.e. lowering the center-of-mass distance r. This result is expected, as the accessible conformational space becomes smaller the more the monomer clouds are in proximity. More importantly, we find that the effective potential $U_{\text{eff}}(r)$ is pronouncedly stronger for chromosomes with a large average number of loops compared to linear chains. Contrarily, the dependence of the effective potential on chain length N is rather subtle, indicating that the level of coarse-graining does not effect the results. This is well-known to be true for self-avoiding walks and ring polymers, where the effective potential at full overlap adopts a constant value on the order of 1 $k_B T$ in the limit of infinite chain length. For the Dynamic Loop model, a comparison is more difficult, since the effective potential is also a non-trivial function of the looping probability p. Importantly, the repulsive potential increases most strongly in the range where $r \approx 1 - 2R_g$, i.e. around the size of the chromosome territory, indicating a huge energy cost for a high degree of CT intermingling.

To demonstrate how the data can be mapped onto physical units, we use model polymers of chain length $N = 256$. We set one coarse-grained bead to 400 kb in order to model a sufficiently long stretch in the size range of a typical chromosome. To determine the spatial extend of this bunch of chromatin, we employ long distance experimental data from chromosome 11 (Fig. 2.3C). In principal, such a mapping is always connected with lots of uncertainy: The detailed Kuhn length is not known, not allowing for a precise mapping on the short scale; As chromatin is organized in a much more complex manner than a linear chain, other parameters (looping, binding, heterogeneity) enter the calcula-

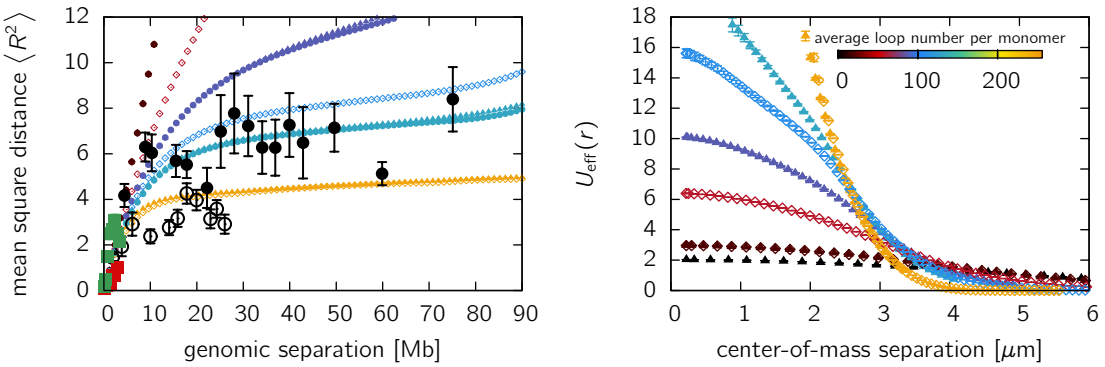


Figure 11.2: Mapping of coarse-grained polymers to physical parameters. Shown are results for a chain length of $N = 256$ using different looping probabilities. The chain is mapped to chromosome 11 by assuming one bead to comprise a 400kb-stretch of chromatin. Consistent with experimental data [12], this is set equal to 480 nm. Different symbols indicate different looping lifetimes (▲: $\tau = 0.01\tau_{int}$; ◇: $\tau = \tau_{int}$; •: $\tau = 100\tau_{int}$). **A.** This panel shows the mean square distance in relation to genomic separation of model and experimental data to assess the quality of the mapping. **B.** The potential of mean force between two model chromosomes in relation to physical distance r between the centers of mass. The effective potential strongly increases with increasing looping number at a separation of about 2-3 μm, i.e. the size range of the assumed chromosome territories.

tions. To obtain a simple mapping, we adjust the plateau level of the model polymers to that of experimental data. Figure 11.2A shows the results of the mapping using 177 nm for one lattice unit, the model displaying well the leveling-off observed in experiments for intermediate looping probabilities (the symbol ▲ corresponds to 131 loops on average). The effective potential $U_{\text{eff}}(r)$ in units of k_BT is displayed in Figure 11.2B. While the effective potential profile is rather flat for self-avoiding walks, the existence of loops leads to a strong increase in the potential at distances of about 2-3 μm, the region where the experimental data displays a leveling-off.

The quantitative increase in the effective potential of looping polymers over linear chains (self-avoiding walks) is shown in Figure 11.3. The factor $U_{\text{eff}}(r/R_g)/U_{\text{SAW}}(r/R_g)$ by which the effective potential of the model chromosomes is larger than that of the linear chain is plotted on the y-axis. Standard errors are in the size range of the symbols and therefore not displayed. Likewise, the abscissa shows the center-of-mass distance scaled by the radius of gyration R_g. We find that the repulsive potential is stronger by more

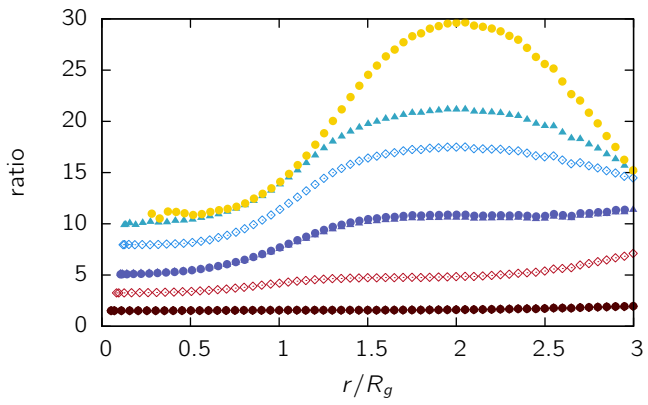


Figure 11.3: Ratio between the effective potential of looping polymers and linear chains. The data shows the ratio $U_{\text{eff}}(r/R_g)/U_{\text{SAW}}(r/R_g)$ for chains of length $N = 256$ and different looping probabilities p. The data is plotted against the center-of-mass separation scaled by the radius of gyration R_g of isolated polymers. The figure symbols and color codes are the same as in Figure 11.2.

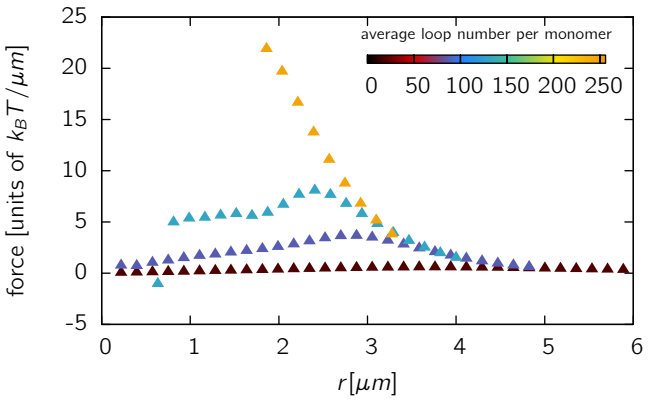


Figure 11.4: The force $F(r) = -\nabla U_{\text{eff}}(r)$ between a pair of chromosomes dependent on their center-of-mass distance r. Shown are simulational results for chains of length $N = 256$ using the mapping from Fig. 11.2. Results for different looping probabilities reveal, that the repulsive force becomes markedly stronger for chromosomes with loops compared to linear chains.

than one order of magnitude for chains in the parameter range where leveling off occurs.

11.3.2 The force between chromosomes can be calculated

From the effective potential $U_{\text{eff}}(r)$ the force acting between chromosomes can be derived using the relation $F = -\nabla U_{\text{eff}}(r)$. We determined the derivative of the effective potential $U_{\text{eff}}(r)$ at position r by locally approximating the potential with a cubic polynomial $p_3(r) = a_0 + a_1 x + a_2 x^2 + a_3 x^3$ using the four data points lying closest to r and calculating its derivative $F(r) = -p_3'(r)$ at position r. We have calculated the force curves for a chain length of $N = 256$ using the mapping indicated in Fig. 11.2, i.e. one coarse-grained monomer consists of a 400-kb-stretch of chromatin and has a diameter of 480 nm.

The force curve is shown in Figure 11.4. Clerly, the observed forces are markedly stronger for chromosomes with loops than for linear chains, having a maximum at approximatley 2-3 μm.

11.3.3 Looping polymers become aspherically elongated

Chromosomes, when brought into close proximity, not only reveal a strong repulsion between their centers of mass; besides this, their structural properties also undergo significant changes. Here, we investigate how size and shape of a model chromosome changes in the presence of a second one. Such effects play an important role inside the cell nucleus, as chromosomes are located in a complex environment being typically separated by only a few Mb; comparison to linear chains allows us to learn something about the effect of looping. The change in dimensionality is measured by the swelling factor s, given by

$$s = R_g(r)/R_g(r = \infty) \tag{11.1}$$

Here, $R_g(r)$ denotes the root mean squared radius of gyration [(3.8)] for a chromosome being in a distance r to a second one. $R_g \equiv R_g(r = \infty)$ denotes the corresponding quantity for isolated chromosomes. In our study on topological effects between ring polymers (↷ chapter 9) we found that both linear chains as well as ring polymers swell when being brought together, the swelling factor being about 10% for rings and slightly smaller for linear chains.

While linear chains and ring polymers only show a mild swelling in presence of a second chain, we find that model chromosomes swell enourmously. Figure 11.5A displays s for different looping probabilities p, i.e. different values of the average number of loops, in relation to the scaled center-of-mass distance r/R_g. To demonstrate a similar behaviour

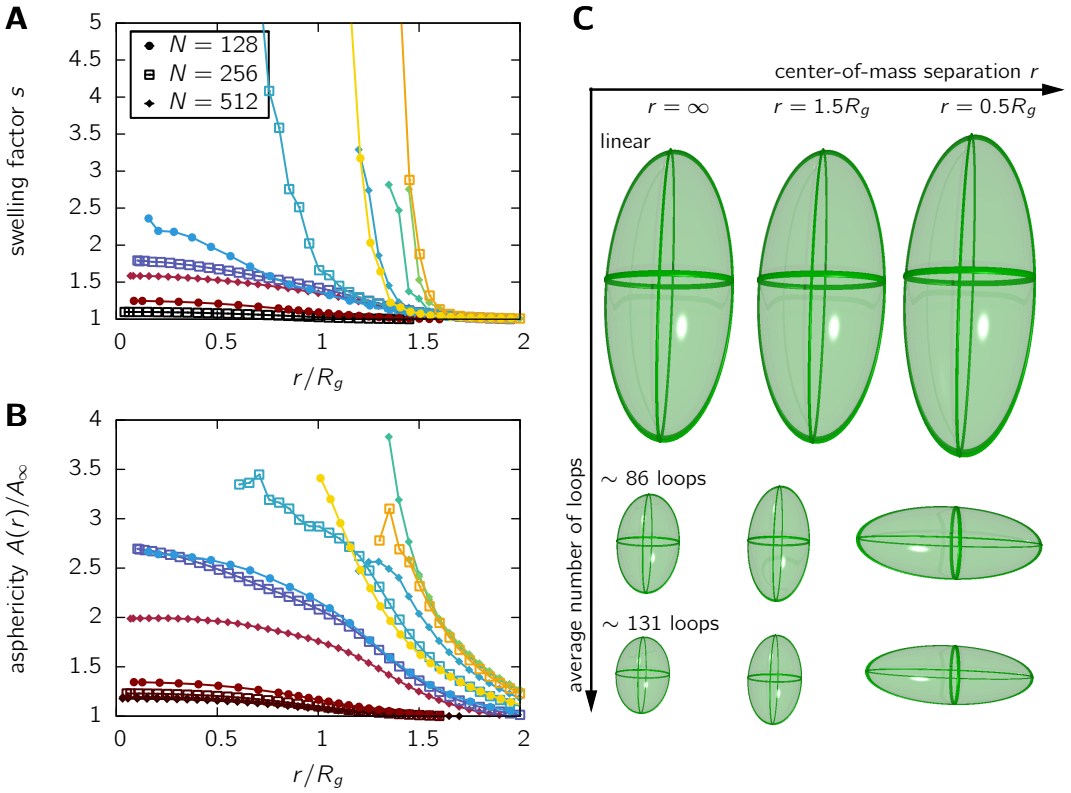


Figure 11.5: Structure of chromosomes being in proximity of a second one. **A.** The swelling factor $s = R_g(r)/R_g$ of the chromosome when being at a center-of-mass separation r compared to the isolated case. Data is shown for different chain lengths N and looping probabilities p. The symbols and colors used are the same as in Fig. 11.1. **B.** The scaled asphericity A/A_∞ for model chromosomes in proximity. The same data is used as in panel A. **C.** Illustrations of the average gyration ellipsoids of the chromosomes. Shown is the change in shape and size of the gyration ellipsoids for linear chains (no loops), chromosomes with an average number of 86 loops (symbol ▲ in Fig. 11.2B) and chromosomes with an average number of 131 loops (symbol ▲ in Fig. 11.2B).

independent of the level of coarse-graining, results are shown for different chain length N. Swelling factors s are strongly dependent on the average number of loops, increasing by a factor in the order of 2-10 for the range of looping probabilities where a leveling-off is observed (cf. also Fig. 11.2A). In fact, s diverges for large loop numbers, indicating that the chains can not be approached closer than approx. $1 - 2R_g$.

The swelling of the chromosomes might suggest that they open up to create space for the monomers of the other chromosome, i.e. allow for intermingling. In the following, we will show that this is not the case, rather the contrary is observed. To achieve this, we investigate how the shape of the chromosomes changes when being close together. The asphericity A of the gyration ellipsoid has been established (↷ 8.3.2) as a measure of shape, being zero for a spherically shaped polymer and unity for a rod-like polymer,

$$A(\lambda_1, \lambda_2, \lambda_3) = \frac{(\lambda_1 - \lambda_2)^2 + (\lambda_1 - \lambda_3)^2 + (\lambda_2 - \lambda_3)^2}{2(\lambda_1 + \lambda_2 + \lambda_3)^2}. \tag{11.2}$$

To highlight the changes in asphericity when approaching two chromosomes, we show the ratio $A(r)/A_\infty$ in Figure 11.5B, A_∞ being the asphericity of an isolated model chromosome with the same parameters. While the change in asphericity for linear chains (self-avoiding

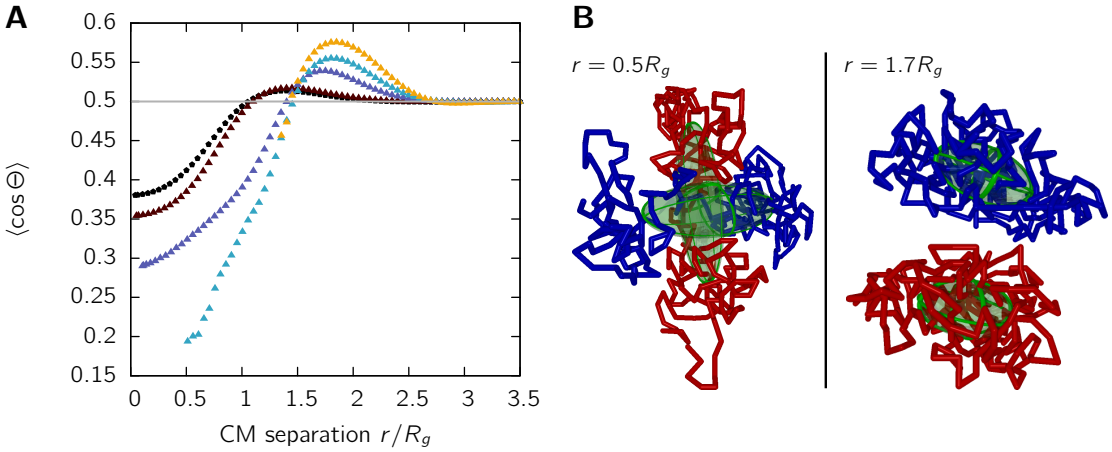


Figure 11.6: Mutual alignment of the gyration ellipsoids. **A.** The average angle $\langle \cos \Theta \rangle$ is shown in dependency of the center-of-mass separation r for chains of length $N = 256$. Black symbols correspond to a linear chain (SAW), colored symbols to chromosomes with loops, the color coding being the same as in Fig. 11.2. Error bars are smaller than the symbol size and therefore not shown. The grey line corresponds to a random orientation of the gyration ellipsoids, showing that chromosomes with loops induce a non-random mutual alignment. **B.** Two chromosomes with a fixed center-of-mass separation r. The right-hand images shows chromosomes in the regime where perpendicular alignment is observed ($r = 0.5R_g$), the left-hand image displays chromosomes with a center-of-mass separation of $r = 1.7R_g$.

walks, no loops) is rather small even at full overlap (about 20%), we find a pronounced aspherical deformation on our model chromosomes in the regime of looping probabilities that force a leveling-off in the mean square distance. Asphericity values increase by about 200-400% at genomic separations of $1R_g$, i.e. the typical size of the chromosome.

The changes in shape and dimension are visualized in Figure 11.5C. Shown are the average gyration ellispoids of three different model polymers: (i) linear chains (self-avoiding walks, 0 loops), (ii) chromosomes with 86 loops on average (symbol ▲ in Fig. 11.2B) and chromosomes with 131 loops on average (symbol ▲ in Fig. 11.2B). For each set of model parameters, the ellipsoids are displayed for three different center-of-mass distances: isolated chains (infinite CM distance), $r = 1.5R_g$ and $r = 0.5R_g$. We find that isolated linear chains require a huge amount of space, while looping polymers are pronouncedly smaller and more spherical. When being in contact with a second chromosome, the shape of self-avoiding walks changes only slightly, while chromosomes with loops become markedly aspherical compared to their isolated shape.

11.3.4 Looping polymers avoid intermingling

To answer the question whether chromosomes swell to create space for each other or rather try to avoid each other, the mutual alignment of the polymers is studied. An established measure for the mutual alignment is given by the average angle $\langle \cos \Theta \rangle$ between the gyration tensors largest principal axes (chapters 8 and 9). In case of the chromosomes being adjusted independently of each other, the average angle would adopt the value of $\langle \cos \Theta \rangle = 0.5$. Deviations from this value indicate a tendency of the polymers to align in a certain non-random way with respect to each other. Figure 11.6 displays results for linear chains (black symbols) and chromosomes with loops in relation to the center-of-mass distance for chains of length $N = 256$. The symbol and color codes used are the same as in

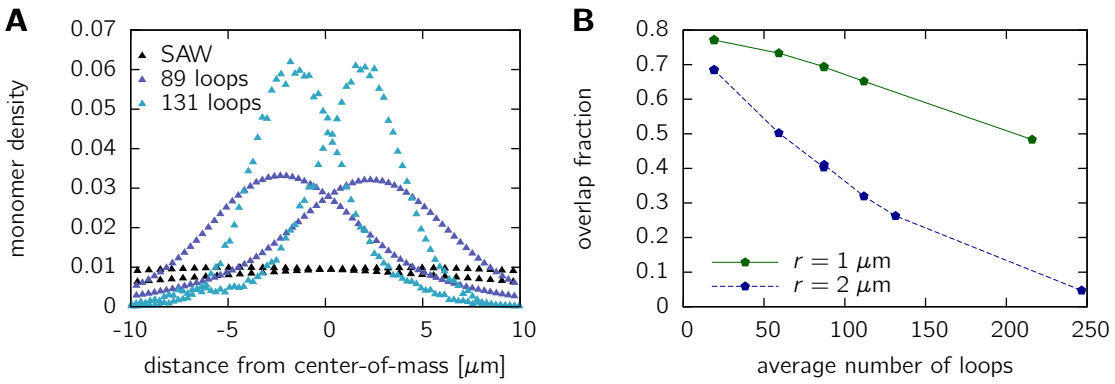


Figure 11.7: Segregation of chromosomes with loops. **A.** This panel shows the monomer density distribution projected onto the line connecting the centers of mass of both chromosomes. A distance of 0 indicates the point halfway between the centers of mass, which are in this example separated by 1 μm. Results are shown for chain lengths of $N = 256$ using the mapping of Figure 11.2. **B.** The degree of intermingling is measured by the overlap area of the monomer distributions from both chromosomes in panel A. The overlap fraction is given for two different center-of-mass separations: (i) 1 μm and (ii) 2 μm. Results show that the overlap fraction, i.e. the degree of intermingling decreases strongly with the average number of loops in the system.

Figure 11.2. We find that chromosomes display a pronouncedly stronger tendency to align perpendicularly at short center-of-mass separations than linear chains or ring polymers (cf. 9.10A). Similarly, a slightly parallel alignment can be found at intermediate distances.

These findings might be explained by a tendency of the chromosomes to minimize the overlap area. When the distance between chromosomes is lowered to values in the order of the size of the chromosome, they start to feel the presence of the second chromosome. Thus, the space of accessible conformations is reduced and the chromosome stretches in a direction perpendicular to the center-of-mass axis. However, when the chromosomes are forced even closer together, nearly overlapping completely, the observed perpendicular alignment together with the strong elongation minimizes the volume shared by both chains (Fig. 11.6B).

To quantitatively assess the amount of intermingling between the model chromosomes in dependency of the average number of loops, we project the monomer positions to the line connecting the centers of mass of both chromosomes. Thus, a density distribution can be obtained as shown in Figure 11.7A. Here, chromosomes with a coarse-grained length of $N = 256$ have been simulated. Mapping is done according to the procedure described above. Results are shown for different looping probabilities and a fixed center-of-mass distance $r = 1$ μm. The average number of loops being 0 (linear case, black symbol), 86 (symbol ▲) and 131 (symbol ▲). Clearly, introducing loops in the chromatin structure results in more compact polymers, the monomers being distributed closer around the centers of mass. We determine the overlap fraction by integrating the overlap area between the distributions of both chromosomes. The results are shown in Fig. 11.7B using two different center-of-mass separations $r = 1$ μm and $r = 2$ μm. For a center-of-mass distance of 2 μm, being comparable to the size of chromosomal regions [12], we find that the overlap fraction decreases strongly from about 0.7 down to less than 0.1 for large loop numbers. Interestingly, in the range where leveling-off occurs, overlap fractions are in the range of 20-30%. These values are, however, an overestimate, as the projection procedure does not capture segregation in the direction perpendicular to the line connecting the

centers of mass. Although not directly comparable to experimental results, these values are in the size range of experimental data from FISH cryo sections, where an overlap volume of 20% has been observed [60].

11.4 Conclusions

In this chapter, we have analyzed the effect of loops on the repulsive interactions between polymers. As a measure for these interactions we applied the theory of effective potentials, where monomeric degrees of freedom in the partition sum are traced out. The resulting effective potential $U_{\text{eff}}(r)$ gives the interaction between both polymers in dependence of their center-of-mass distance r. $U_{\text{eff}}(r)$ has been determined for self-avoiding walks in recent decades [167] and for ring polymers (↷ chapter 9) in this thesis. Both linear polymers and rings display a repulsive interaction at full overlap ($r = 0$), asymptotically converging to a finite value in the order of 1 k_BT in the limit of large chain lengths. Here, we applied the concept of effective interactions to the Dynamic Loop model, which has been proposed as a model for chromatin organization in chapter 10.

The major finding of this study is that introducing dynamic loops in the structure of chromatin results in a strong increase of the repulsive interactions by about one order of magnitude (Fig. 11.1). Using a mapping to physical units based on recent FISH experiments, we found that the repulsive forces are strongest at center-of-mass separations of $2 - 3$ μm, i.e. the size of the chromosome territory (Fig. 11.2). These observations indicate that chromatin looping plays a dominant role in the entropy-driven segregation of chromosomes.

Moreover, we found that the existence of loops introduces strong changes in the size and shape properties of the chromosomes. Indeed, when being brought close together, looping polymers swell and become pronouncedly aspherical, the observed effect being multi-fold larger than for linear chains or ring polymers. Chromosomes in proximity display a highly non-random orientation of their gyration ellipsoids. Increasing the number of loops leads to a significant decrease in the overlap fraction.

These findings indicate that chromatin loops not only play an important role in transcriptional regulation. Rather, they help to impose a certain state of order and segregation. Thus, loops seem to constitute a highly efficient regulatory mechanism concerning gene regulation as well as chromatin compartmentalization. The Dynamic Loop model used in this study refrains from assuming active driving mechanisms for loop formation, rather loops form by diffusional motion, minimizing the energy cost.

Chapter 12

Nuclear nanostructure detected with localization microscopy

References

The results presented in this chapter are published in:

- M. Bohn, P. Diesinger *et al.*, **2009**. Localization microscopy reveals expression-dependent parameters of chromatin nanostructure. *submitted to Biophysical Journal*.

The work described in this chapter is the result of a collaboration with the Applied Optics and Information Processing group at Kirchhoff Institute for Physics of Prof. Christoph Cremer, the group of Prof. Michael Hausmann and the Statistical Physics and Theoretical Biophysics group of Prof. Dieter W. Heermann at the Institute for Theoretical Physics.

The experimental work from cell cultivation and preparation to localization microscopy and raw data processing was performed in the group of Christoph Cremer and Michael Hausmann. The image segmentation and data analysis has been carried out by Philipp Diesinger and myself. Concerning the data analysis I will only present my work here, Philipp Diesinger has published an own analysis already in his PhD thesis [181].

Chapter Summary

A combined approach of 2D high-resolution localization light microscopy and statistical methods is presented to infer structural features and density fluctuations at the nuclear nanoscale. Hallmarks of nuclear nanostructure are found on the scale below 100 nm for both fibroblast and HeLa cells. Mechanical properties were extracted from the histone density fluctuations inside the cell. Results show that different mechanisms of expression of the same protein lead to significantly different patterns on the nanoscale and to pronounced differences in the detected compressibility of chromatin. While a linear self-avoiding walk polymer model is in disagreement with the experimental findings, the formation of dynamic loops explains well the observed large values of the compressibility.

12.1 Introduction

Chromatin nanostructure of eukaryotic cells has been hard to analyze by light optical techniques. Conventional light microscopy is limited physically to a resolution of about 200 nm, the Abbe limit. Structures below this length scale cannot be resolved by these microscopes.

Chromatin structures above the level of single nucleosomes, however, are typically in the size range between 10 nm and some μm [7]. The diameter of the presumed chromatin fiber between 10 nm and 45 nm renders it impossible to follow the path of the chromatin fiber by conventional light optical techniques. Thus, chromatin structure can only be inferred by using indirect approaches. Fluorescence in situ hybridization (FISH) has been applied to mark specific sites along the chromosome, measuring the physical distance between the FISH probes [27, 8, 9, 12]. These labels can then be localized with a conventional light microscope (↷ 2.3.4). Thus, a relationship between mean square displacement vs. genomic separation of two markers can be established. Large-scale organization on the level of the entire cell nucleus can be visualized by distributing FISH labels along the fiber [6] or by using fluorescent stains with a non-sequence specific DNA binding affinity, such as DAPI [182] which can be used to stain both fixed and live cells. However, structural information from conventional confocal light microscopy is limited by about 200 nm in the focal plane and about 600 nm in the optical plane. To overcome the resolution limits of conventional confocal light microscopy, confocal laser scanning fluorescence 4Pi microscopy [183] may be used where laser light is focussed from different sides, allowing for an axial resolution down to the 100 nm range [184].

The investigation of chromatin nanostructure, i.e. structures with dimensions below 100 nm, still faces severe experimental problems. Electron microscopy (EM) has been applied to study the both isolated chromatin segments in vitro or thin sections of chromatin in situ [185, 45]. Generally, transmission EM requires a high vacuum and thin samples to allow the beam penetrate the probe. One way of achieving this is to dehydrate the specimen, embed them in a plastic medium, finally cut thin sections out of it. Before staining the probe with heavy metals, they have to be chemically fixated due to the invasiveness of the staining procedure. A less invasive approach is cryo-EM, where whole unfixed nuclei are used to create frozen hydrated cryosections [186, 187]. Although actual optical resolution of nuclear structures in the 10 nm regime have been achieved using

advanced EM methods [188], some general restrictions of EM still remain which make it highly desirable to develop methods for light optical analyses of nuclear structure at an enhanced resolution beyond the Abbe Limit.

In the last years, there have been advances in light optical techniques allowing for a resolution of single fluorescence labelled molecules with a localization accuracy in the range down to 20 nm [189, 190, 191]. Basis of these techniques is to use different spectral signatures of light from a point-like source to localize positions of the fluorescent molecules, even if they are close together (< 200 nm). Generally speaking, when passing the optical microscopy setup, each point-like fluorophore will be blurred on the screen, the intensity distribution given by a Bessel function. Only if the distance between two fluorophores is larger than the half-width of the first maximum of this airy disk, the points can be separated. This, however, is not true, if the fluorescent spots have different colors. Then, two points can be *optically isolated* by inspecting the color-dependent maxima, allowing the separation of points much closer than given by the Abbe limit. Optical isolation can also be achieved by utilizing any kind of distinct optical signature, for example different blinking frequencies or consecutive emission times [192]. Using different emission times requires fluorescent molecules which can be switched on and off by an external source. First proof-of-principle experiments were done using spectral precision distance / spectral position determination microscopy (SPDM) [193, 194]. Since then an advanced method has been developed using reversibly bleached states of conventional fluorophores like GFP for localization microscopy [195]. The stochastic recovery of fluorophores from the dark state yields to an optical isolation, which allows an accurate position determination far below the Abbe limit. Acquiring a huge amount of images over time results in a good reconstruction of the original distribution of fluorescent markers.

Although these technical microscopy methods have been well-studied, a quantitative analysis of nuclear nanostructure is still lacking. In this study, we present a method combining high precision light microscopy with statistical methods to obtain information on both structure and mechanical properties. We apply and verify the method by studying the distribution of histones H2B inside the cell nucleus. The questions we ask are

1. Is there a recurrent nuclear (chromatin) nanostructure for different cell types?
2. Does this nanostructure show differences from cell type to cell type?
3. Does the nanostructure dependent critically on the type of expression of the fluorescent dye?
4. On which length scale do fluctuations vanish inside the nucleus? What is the corresponding compressiblity of localized fluorophores?
5. How do the results relate to polymer models?

To answer these questions, we apply localization microscopy to both fibroblast and HeLa cells. To examine the expression-dependency of the structure, we use different expression methods for HeLa cells. For the present experiments, a setup for SPDM with physically modifiable fluorophores (SPDM_{PHYMOD}) was used [195].

12.2 Experimental Methods and Image Segmenation

The sections “12.2.1 Specimen preparation”, “12.2.2 SPDM Setup” and “12.2.3 Data acquisition and evaluation” as well as the figures therein were kindly provided by Rainer

Kaufmann. The work described in these sections was done by the group of Prof. Christoph Cremer and Prof. Michael Hausmann, the persons involved are Rainer Kaufmann, Yanina Weiland, Patrick Müller, Manuel Gunkel, Alexa v. Ketteler and Paul Lemmer.

12.2.1 Specimen preparation

a) Fibroblast Cells
The VH7 diploid human fibroblast cells (kindly provided by Prof. Dr. Beauchamp from DKFZ, Heidelberg), used for the measurements, were cultivated in Dulbecco's Modified Eagle medium supplemented with 10% FCS, 1% L-glutamine, 1% penicillin/streptomycin in a standard CO_2 incubator. After seeding the cells onto cover slips, they were allowed to attach and grow over night. Using Organelle Lights (Invitrogen, Carlsbad, USA) according to the manufacturer's protocol, EmGFP conjugated histone proteins H2B were expressed. The cells were fixed with 4% formaldehyde in PBS and embedded with ProLong Gold antifade reagent (Invitrogen) 24 hours after the transfection.
b) HeLa Cells (strains I, II)
HeLa cells were cultivated in RPMI medium supplemented with 10% FCS, 1% L-glutamine, 1% penicillin/streptomycin in a standard CO_2 incubator.

1. HeLa cells of strain I were first seeded onto coverslips and allowed to attach and grow over night. After using Organelle Lights (Invitrogen, Carlsbad, USA) according to the manufactures protocol, emGFP conjugated histone proteins H2B were expressed. The cells were then fixed with 4% formaldehyde in PBS and embedded with ProLong Gold antifade reagent (Invitrogen) 20 hours after transfection.

2. HeLa cells (strain II, kindly provided by Dr. Tobias Knoch from Univerity Rotterdam and Bioquant Center, University of Heidelberg) stably expressing histone H2B-GFP proteins, were seeded onto coverslips, and were allowed to attach and grow over night. Afterwards the cells were fixed using 4% formaldehyde in PBS and embedded with ProLong Gold antifade reagent (Invitrogen).

12.2.2 SPDM setup

The microscopy was based on the principle of spectral precision distance/position determination microscopy (SPDM; [196, 193]; for review see [197]). For the present experiments, a setup for SPDM with physically modifiable fluorophores (SPDM_{PHYMOD}) was used [198]. In this case 2D localization of single fluorescent molecules was achieved by stochastic switching. This method is based on a light induced reversibly bleached state of conventional fluorophores [199, 200, 201]. By starting illumination of the sample with an excitation intensity in the 10 kW/cm^2 to several 100 kW/cm^2 range, some molecules are bleached irreversibly ($M_{fl} \rightarrow M_{ibl}$), another amount is transferred into a reversibly bleached state ($M_{fl} \rightarrow M_{rbl}$). The statistical recovery of fluorophores from this state ($M_{fl} \leftarrow M_{rbl}$) can be used for optical isolation of the single fluorescent molecules. This allows the SPDM setup single molecule detection and localization of conventional fluorophores specified as SPDM_{PHYMOD}.

For the experiments only one laser source was needed at an excitation wavelength of $\lambda = 488$ nm [Ar488] (Lexel 95-4, Lexel Laser, USA). It is used for fluorescence excitation as well as for the reversible bleaching of fluorophores. After deflection at mirror [M1] the laser beam is expanded by a factor of 10 using a collimator built up of two achromates

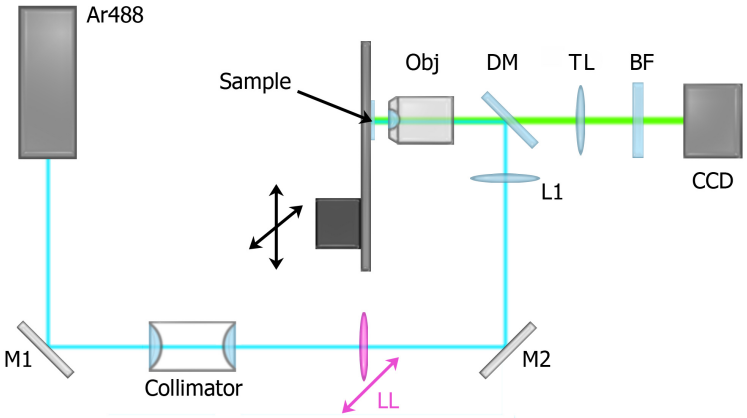


Figure 12.1: Schematic overview of the microscope setup used for the SPDM measurements.

with focal length of 10 mm and 100 mm (Linos Photonics, Göttingen, Germany). Then an additional lens [LL] leads to a more focused spot in the object region so one can achieve the high laser intensity, which is necessary for the localization mode. Via a mirror [M2], a dichroic mirror [DM] (AHF Analysetechnik AG, Tübingen, Germany) and a lens [L1] the beam is focused into the back focal plane of an oil immersion objective lens (x100, NA = 1.4, Leica, Bensheim, Germany).

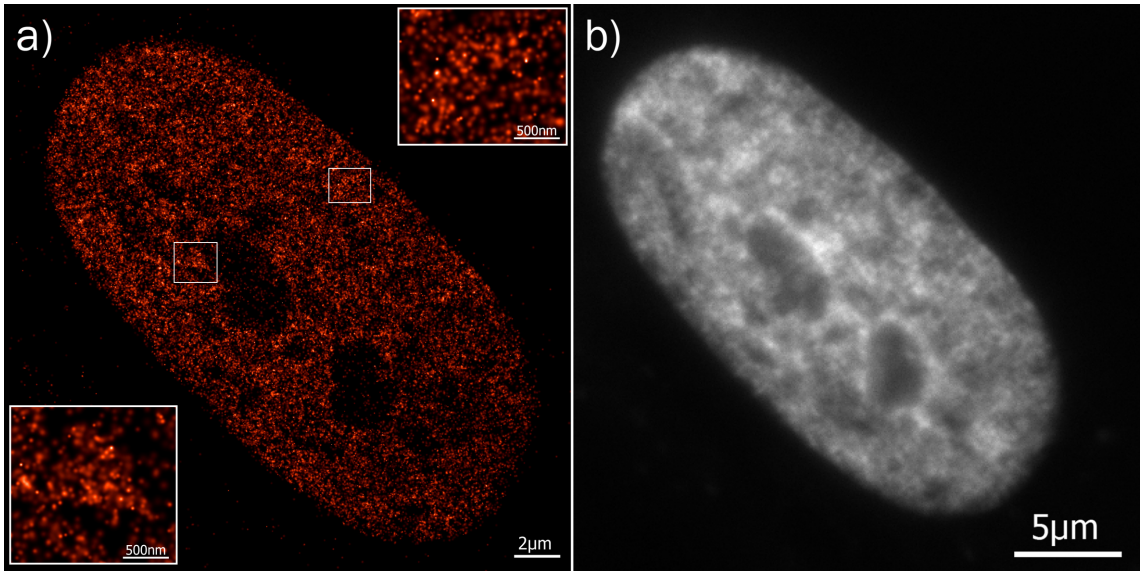


Figure 12.2: a): Localization image of histones H2B in the nucleus of a human fibroblast cell labeled with EmGFP. The localized fluorophores are blurred with a Gaussian corresponding to their individual localization accuracy. The inserts show the marked areas four times magnified. b): A conventional wide-field image of the same nucleus as in a). The image was acquired with a conventional epi-fluorescence microscope setup.

The detection pathway is a conventional epi-fluorescence setup. Emitted fluorescence light passes the dichroic mirror [DM] and is focused by a tube lens [TL] (x1.0, Leica, Bensheim, Germany) onto the CCD chip of a highly sensitive 12 bit black and white camera [CCD] (SensiCam QE, PCO Imaging, Kehlheim, Germany). A blocking filter [BF], which is mounted in front of the CCD chip, reduces the background signal of the excitation light.

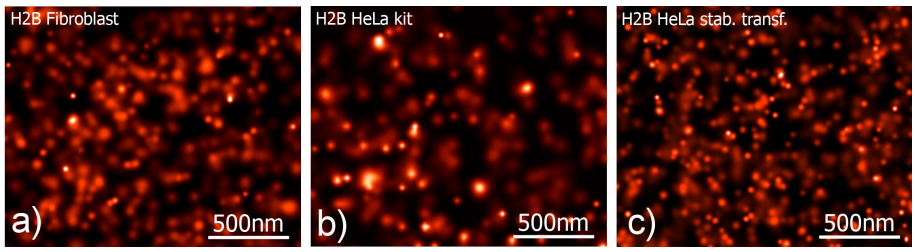


Figure 12.3: Sections of localization images of histones H2B in the nucleus of (a) human fibroblasts, (b) HeLa cells of strain I and (c) HeLa cells of strain II.

12.2.3 Data Acquisition and Evaluation

For data acquisition a time stack consisting of 2000 frames was recorded with a frame rate of about 18 fps for each image. The evaluation was done by algorithms implemented in MATLAB (7.4.0, The MathWorks, Natick, USA).

In a first step, the number of photons for each pixel of the CCD chip is estimated. The count number is translated into the number of incident photons by the multiplication of a factor dependent on the conversion rate of the CCD sensor:

$$N_{\text{Photons}}(x, y, t) = c_{conv} \cdot N_{Counts}(x, y, t),$$

where $N_{\text{Photons}}(x, y, t)$ is the number of incident photons as a function of the pixel position (x, y) and the time t denoting the individual frames of the image stack obtained by the CCD camera at times $t = t_1, t_2, ..., t_k, ...t_n$. c_{conv} is the conversion factor between the number of counts and the number of photoelectrons. $N_{Counts}(x, y, t)$ gives the number of counts (raw data) as a function of pixel position (x, y) and the time t. For example, using $c_{conv} = 2/$ count (low light mode) a photon number of $N_{\text{Photons}}(x, y, t) = 2N_{Counts}(x, y, t)$ is estimated.

In a second step, for signals with a low signal-to-noise ratio (i.e. high background and photo bleaching effects active during several succeeding frames) an additional computing step is required to segment signals originating from single molecules only. A differential photon stack $D_{\text{Photons}}(x, y, t')$ between the succeeding $(t = t_{k+1})$ and the preceding frame (t_k) is calculated to eliminate the background noise:

$$D_{\text{Photons}}(x, y, t') = N_{\text{Photons}}(x, y, t_{k+1}) - N_{\text{Photons}}(x, y, t_k).$$

The error σ in the photon number produced by the Poisson statistics of the incident photons and the noise σ_{CCD} of the detection at the CCD chip (approx. 4 counts per pixel) was estimated by the Gaussian law of error propagation:

$$\sigma[D_{\text{Photons}}(x, y, t')] = [N_{\text{Photons}}(x, y, t_{k+1})^2 + N_{\text{Photons}}(x, y, t_k)^2 + 2\sigma_{CCD}^2]^{1/2}.$$

In the last step, the data stack $N_{\text{Photons}}(x, y, t)$ (in case of low background) or in case of high background $D_{\text{Photons}}(x, y, t')$ is used for a lateral high precision localization of single molecules. To reduce computing efforts, Regions of Interest (ROIs) of typically 8x8 pixels were used. A model function is fitted to the acquired signals in the object plane using the Levenberg-Marquardt algorithm with a Gaussian distribution plus background:

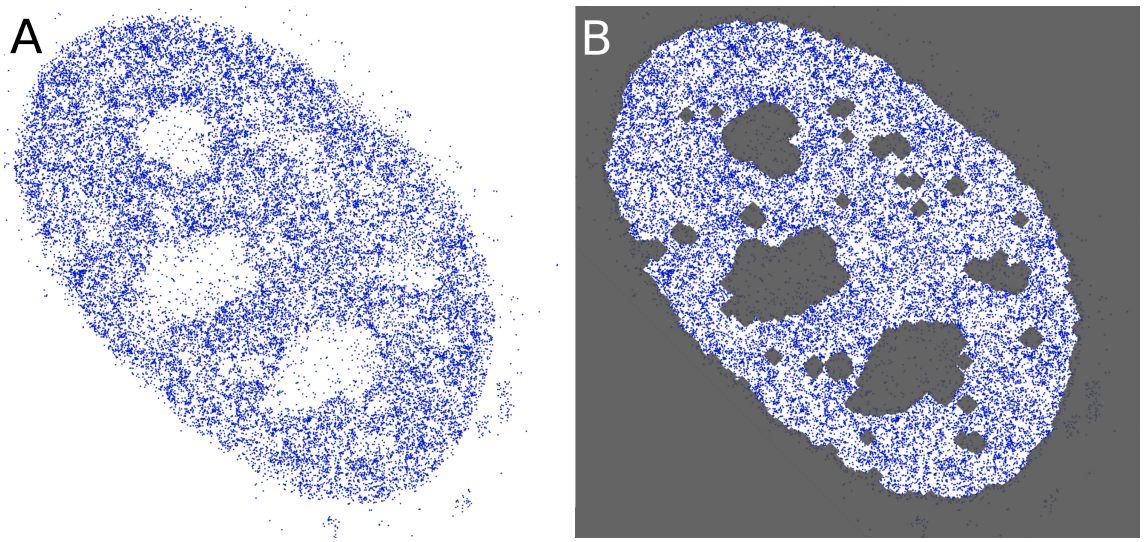


Figure 12.4: **A.** Localized histones of a different fibroblast cell after data processing. Each localized fluorophore is represented by a blue point. The average number of detected points is 124 per μm^2. **B.** Result of the segmentation procedure. Regions marked grey are not included in the data analysis. The image is segmented based on an density threshold procedure in order to exclude regions outside the nuclear membrane as well as chromatin-depleted regions inside the nucleus, e.g. nucleoli.

$$f(x,y) = Ae^{-\frac{(x-x_0)^2+(y-y_0)^2}{2\sigma^2}} + B_0 + B_1(x_0 - x) + B_2(y_0 - y).$$

x_0 and y_0 are start parameters for the position corresponding to the determined center of the segmented signals. A is the amplitude and σ the standard deviation of the Gaussian distribution. B_0, B_1 and B_2 represent a linear background.

In this manner the lateral positions of the single molecules can be determined with a localization accuracy (σ) in the 20 nm regime [195].

Fig. 12.2a) shows a localization image for histones H2B labeled with EmGFP in human fibroblast cells. The localized fluorophores give the positions of the single histone proteins with their individual localization accuracy displayed by a corresponding Gaussian blur. The wide-field image in Fig. 12.2b) was recorded with a conventional epi-fluorescence microscope setup. A comparison between localization images of fibroblasts and HeLa strain I and II is displayed in Fig. 12.3.

12.2.4 Image segmentation

After data acquisition and evaluation, the data is provided for each cell as a number of two-dimensional coordinates (x, y) together with a localization accuracy $(\Delta x, \Delta y)$, each point representing a localized fluorophore. An example of a Fibroblast cell is depicted in Fig. 12.4A.

To obtain information about the structure, a threshold-based density segmentation procedure was applied, which defines the region of pixels in which chromatin is found. The segmentation algorithm was parameterized such that regions outside the nuclear envelope as well as regions within the nucleus containing no or little chromatin (e.g. nucleoli) as defined by a low detected histone molecule density were excluded from the analysis.

Table 12.1: Number of cells and average number of detected fluorophores per square micrometer for human fibroblasts and HeLa strain I and II.

cell type	number of cells	av. # points / μm^2
fibroblasts	28	124
HeLa, strain I	9	206
HeLa, strain II	37	98

In order to achieve this, a quadratic box of linear dimension b was positioned at each possible position (with a resolution of 30 nm) and the density ρ_b of points inside this box was determined. If the density was below a certain percentage p_{thre} of the maximum density, the region of the box was marked as outside of the analysis region. The box dimension was always set to $b = 500$ nm. The threshold percentage was set to $p = 15\%$ for the HeLa cells of strain II and $p = 10\%$ for strain I and the fibroblasts. These values were choosen by visual inspection to ensure that no region with little fluorophore content is used for analysis. The segmentation algorithm is defined such that in tendency more area was removed from the analysis region than necessary. By using a rather large box dimension b we autmatically excluded heterochromatin regions near the nuclear envelope. Figure 12.4B shows the result of the segmentation procedure using the same cell as in Fiure 12.4A.

Table 12.1 lists the number of images of cell nuclei for each cell type and strain as well as the average number of detected fluorophores per square micrometer after the segmentation procedure.

12.3 Radial distribution function reveals nanoscale structure

In order to examine nanostructure we used the concept of the radial pair distribution function (RDF) $g(r)$ from solid state and soft matter physics [202], where it is used to analyze structural properties of systems and transferred it to the context of analyzing cell contents.

12.3.1 Definition

Structure information is coded in the radial distribution function (RDF) $g(r)$ as well as its Fourier transform, the structure factor $S(q)$. This radial distribution function describes the variation of the surrounding matter's density as a function of distance. Suppose, for example, a histone detected at the origin $\mathbf{0}$ of the coordinate system. What then is the density of histones in a certain spherical shell with distances $[r, r + dr]$ apart? Indeed, the RDF $g(r)$ is the factor by which this density differs from the average particle density ρ of the system. Consider a system of N particles in a volume V. For a given distance r let the number of particles being in the spherical shell $[r, r + dr]$ around the histone at the origin be $N(r)$. Then the RDF is defined as the ratio between the density $\rho(r) = N(r)/(\frac{4}{3}\pi r^3)$ and the average density $\rho = N/V$ of the system averaged over the ensemble. Therefore $g(r)$ is a measure for the local order of a system and deviations from a structureless random particle distribution of the histones inside the cell nucleus show off clearly.

In more statistical terms, the radial distribution function is defined by the normalized integral over the configurational part of the Hamiltonian of the system [202]

$$g(\mathbf{r}) = \frac{V^2}{N^2} N(N-1) \frac{\int \cdots \int \exp\left[-U_N(\mathbf{r}_1, \ldots, \mathbf{r}_N)\right] d\mathbf{r}_3 \ldots d\mathbf{r}_N}{\mathcal{Z}_N} .$$

Here, U_N is the system's potential energy function and $\mathcal{Z}_N$ is the configurational part of the partition sum.

The relation between the particle pair potential $U(r)$ and the RDF is given by [202]

$$g(r) = \exp\left(-U(r)/kT\right) .$$

The radial distribution function is normalized such that

$$\int_0^\infty \rho g(r) 4\pi r^2 dr = N .$$

12.3.2 The influence of projections

The images from SPDM microscopy are two-dimensional projections of three-dimensional cells. Projections comprise about 600 nm perpendicular to the focal plane, leading to a loss of structure information.

We studied a simple fluid as a model system to verify that projections do not conceal structural information. Simple fluids are often modelled as a Lennard-Jones fluid [202], i.e. the pairwise interaction potential is given by

$$U_{\mathrm{LJ}} = 4\epsilon \left\{ \left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right\} . \tag{12.1}$$

As the cell also is in a fluid state, such a system allows comparison to the experimental data. The Lennard-Jones system studied was composed of $N = 2048$ particles and a density of $\rho = 0.6\sigma^{-3}$ in the NVT-ensemble. Molecular Dynamics simulations are performed using the software package Espresso [121]. A time-step of $\Delta t = 0.005\tau$ was chosen. The Langevin thermostat was initialized with a temperature parameter of $T = 1\epsilon$ and a friction of $\Gamma = 0.5\tau^{-1}$, consistent with the values used in Ref. [121].

From a sample of 2000 independent conformations, we calculated the three-dimensional radial distribution function $g_{3D}(r)$ as well as projected two-dimensional RDFs. For the latter, all particles within a slice of thickness fraction f relative to the simulated system size were projected onto a plane, yielding the two-dimensional radial distribution function $g(r)$. As the cellular diameter is about 10 μm, the slice thickness of interest is $f \approx 600\,\mathrm{nm}/10\,\mu\mathrm{m} = 0.06$. While the three dimensional RDF displays the well-known shells, in which particles arrange, the information is blurred out slightly for projected particles (Figure 12.5), the effect becoming less the smaller the slice thickness. Although structures are blurred in the radial distribution function $g(r)$ for a value of $f = 0.06$, they are still markedly visible.

Labeling efficiency varies dependent on cell type and labeling method (table 12.1). How does this effect the radial distribution function? In fact, the RDF is independent of labeling efficiency as long as the localized fluorophores are drawn randomly from the ensemble of available histones. The reason for this is that $g(r)$ is a density-normalized average, thus density effects are entirely divided out. However, less labeling efficiency results in less statistics and therefore in larger errorbars.

12.3.3 Calculations of $g(r)$ for the data

For the calculation of $g(r)$ the valid analysis region has to be taken into account. It is to be emphasized that $g(r)$ is not determined using points, where the connecting line between them or parts of it lie in a region, which was marked invalid by the segmentation procedure.

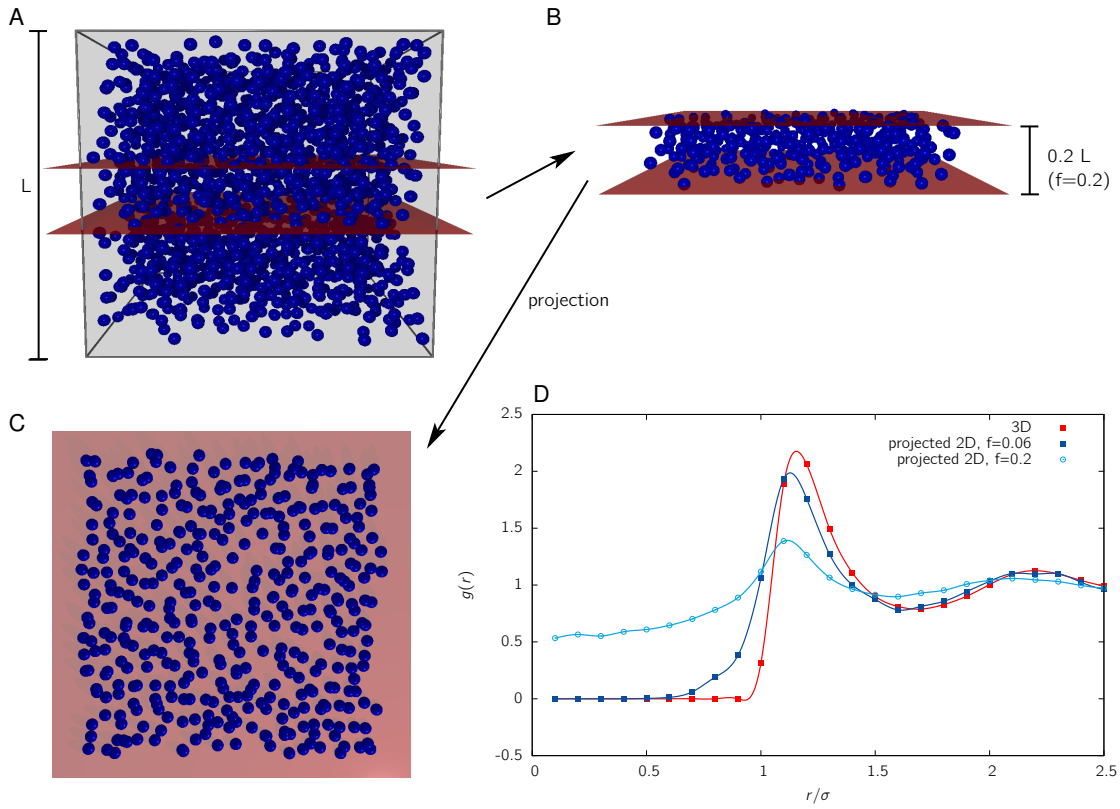


Figure 12.5: Calculation of the projected radial distribution function $g(r)$ for a Lennard-Jones fluid. To analyze the effect of 2D projections on the radial distribution function $g(r)$, a three-dimensional Lennard-Jones system is simulated. Panel A displays an example conformation. As microscopy data provides a 2D projection of points comprising about 600 nm along the optical axis, a section is cut out of the simulation box (marked by the red plane). Panel B displays this section which is in this example 20% ($f = 0.2$) of the linear system size L. This data is then projected onto a 2D plane (panel C) and the radial distribution function is calculated. Panel D shows the results for the radial distribution function $g(r)$. The red curve displays the results for the 3D radial distribution function. It displays the typical behaviour: Below the hard-sphere radius of the fluid particles σ, no other particles can be found. The first peak corresponds to the first shell of molecules arranging close to each other due to the attractive interactions. Calculations are shown for a slice thickness of the slice of 6% and 20% of the linear system size. Assuming typical diameter of a human cell of about 10 μm, the optical section used in microscopy of 600 nm corresponds to the 6%-curve (fraction $f = 0.06$). Obviously, the projected radial distribution function still displays the structure-defining maxima and minima, although they are less pronounced. This analysis indicates that the projected rdf does not conceal structural information.

Therefore, it leads to inconsistent results to just create a histogram of distances between localized points inside the analysis region and obtain $g(r)$ from it. Rather, we proceed as follows

- For each localized point $\mathbf{r}_i$ inside the valid region of the nucleus we
  - determine the maximum sphere around the localized fluorophore at $\mathbf{r}_i$ with radius R, for which the area is still completely inside the analysis region
  - and calculate the density $\rho_i(r)$ of points lying in a spherical shell with inner and

outer radii r and $r + dr$ with a resolution of $dr = 10$ nm up to the maximum value of R.

- We average over the radial densities $\rho_i(r)$ and calculate $g(r)$ using the averaged densities.

12.3.4 Results

The radial distribution function $g(r)$ is determined for the fibroblast cells and both HeLa strains. Results are shown in Figure 12.6. The RDF displays significant differences from an equidistribution of histones (corresponding to the black line) inside the cell nucleus for both fibroblasts and HeLa cells on the scale below 100 nm, indicating the existence of distinct structures. More interestingly, the RDF shows pronounced differences for HeLa cells stably expressing histone H2B-GFP proteins (strain II) and emGFP-labelled cells (strain I). The emGFP-labelled fibroblasts show the same structural features as strain I of HeLa. Thus, different mechanisms of expression of the same protein lead to a significantly different distribution of localized histones on the nanoscale. Errorbars in Fig. 12.6A represent the standard deviation in the ensemble of cells, showing that there is considerable structural variation from cell to cell (cf. also Figure 12.6B).

12.4 Density distribution analysis reveals large-scale fluctuations

12.4.1 Background and Method

Statistical Mechanics provides a useful tool to study both structural and mechanical properties of a system of particles [202]. The localized fluorophores at positions $\mathbf{r}_i$ can be considered as particles in a volume V interacting via a certain potential $U(\{\mathbf{r}_i\})$. The potential summarizes all interactions between localized particles as well as interactions with other constituents of the cell in the sense of an effective interaction. Thus, U might comprise pairwise or three-body or even higher-order interactions. What, then, does statistical mechanics tell us about the system?

A many-body system of volume V can be divided into blocks q_i of linear dimension b. As long as the block-size is larger than the correlation length ξ of the order parameters of the system, but smaller than the total size of the system $V^{1/3}$, all blocks represent one realization of the ensemble. Following the procedure described by Rovere *et al.* [203] we determine from the block densities ρ_i the density distribution $P_b(\rho)$ of the system. The distribution still carries an index b, as finite-size effects render the distribution dependent on the block size. In contrast to Ref. [203] we determine the distribution and its kth moments $\left\langle \rho_b^k \right\rangle$ for each cell separately and average the final measures. The rationale behind this procedure is that labeling and detection efficiency may vary from cell to cell, not displaying "real" biological and physical fluctuations.

In the canonical or NVT ensemble, where particle number, temperature and volume are fixed — as is the case inside the cell nucleus — the density fluctuations are linked to the radial distribution function $g(r)$ [202],

$$\frac{\left\langle \rho^2 \right\rangle - \left\langle \rho \right\rangle^2}{\left\langle \rho \right\rangle} = \frac{N}{V^2} \int 4\pi r^2 \left(g(r) - 1 \right) dr + 1 \,. \tag{12.2}$$

In case of a random distribution of points (corresponding to an ideal gas), the density

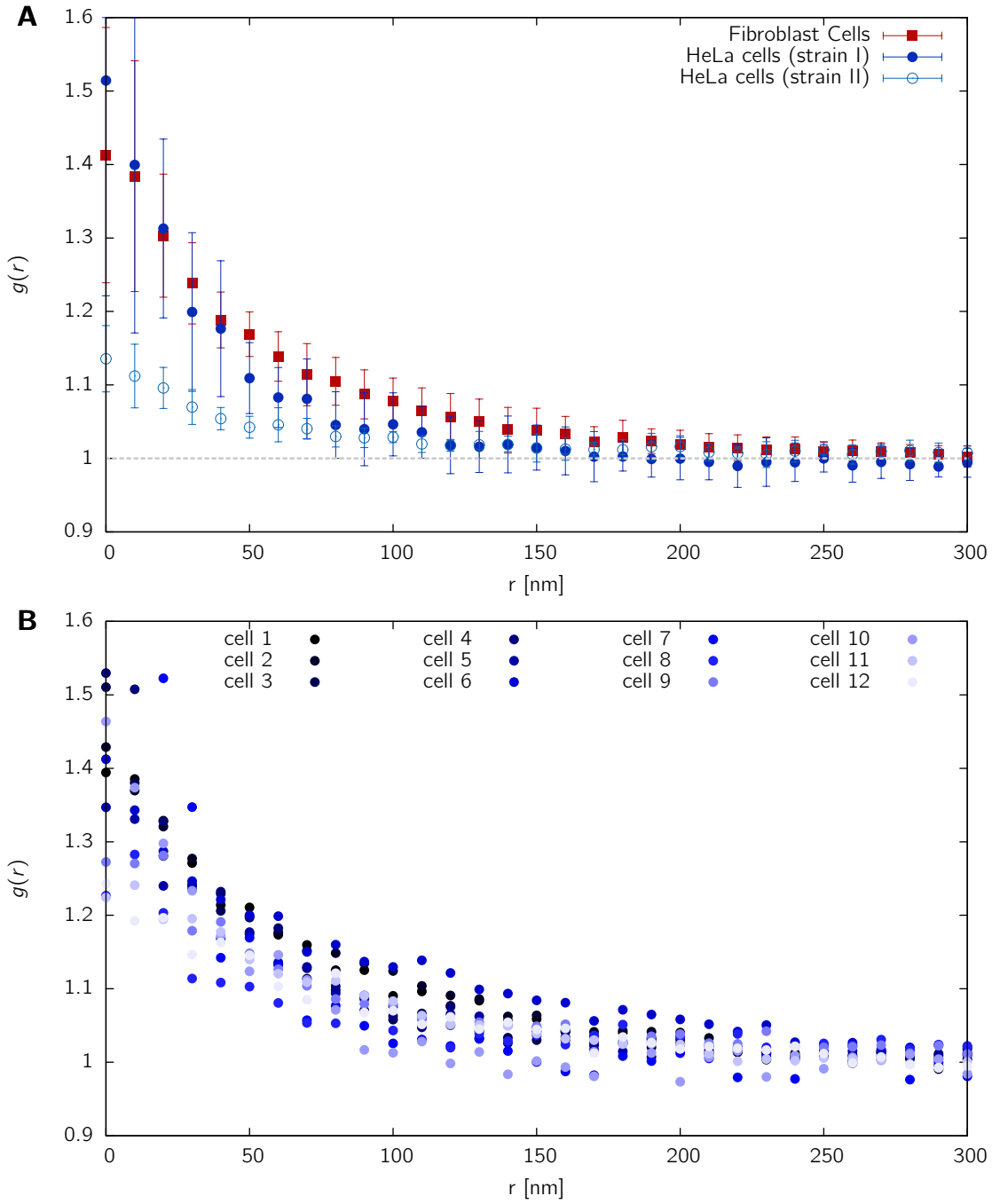


Figure 12.6: **A.** Nanostructure of nuclear histone distribution observed by high-resolution light microscopy. The radial distribution function $g(r)$ reveals distinct differences from a random distribution (i.e. $g(r) = 1$), highlighting the existence of distinct structures on the scale below 100 nm. Results are shown for human fibroblasts and two HeLa strains prepared by different labeling methods. Error bars represent the cell-to-cell variation (standard deviation). Structural hallmarks are found for HeLa as well as fibroblasts independent of labeling methods, however quantitative differences exist. The grey horizontal line corresponds to the (normalized) average density, i.e. a random spatial distribution of histones. **B.** The radial distribution function for single human fibroblast cells. Differences between different cells highlight the cell-to-cell variations.

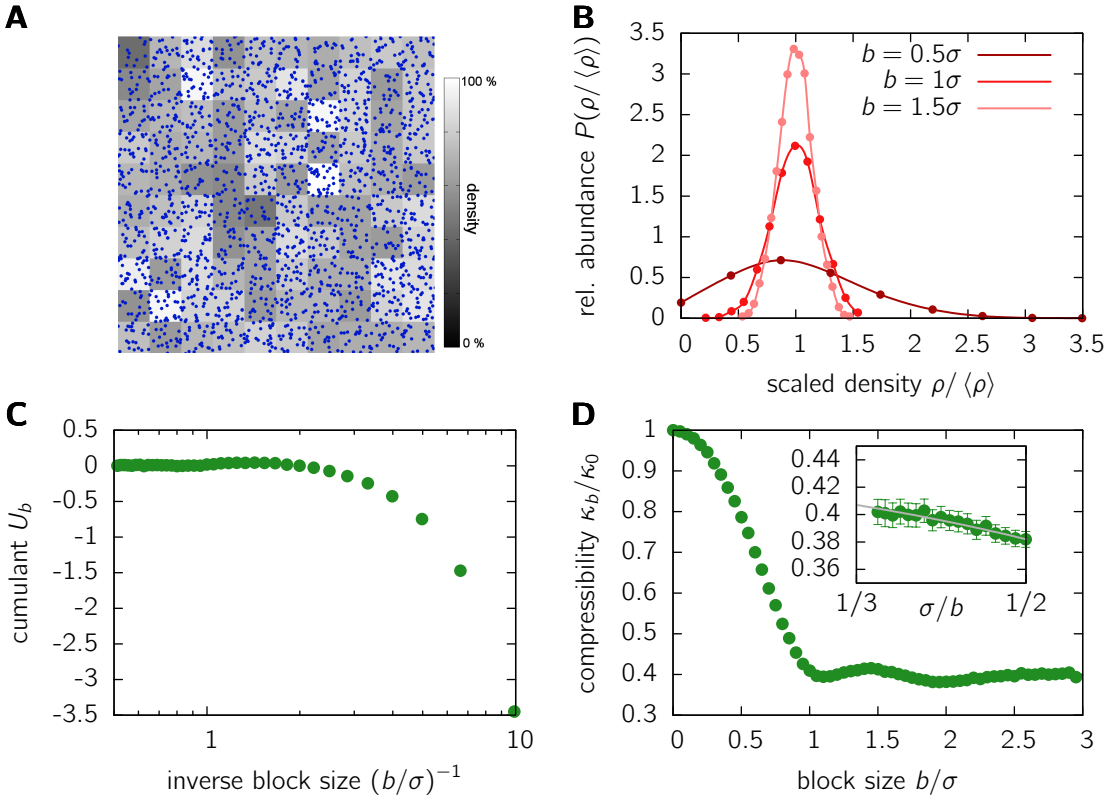


Figure 12.7: Block density analysis of a Lennard-Jones fluid. **A.** An image of the local densities of a projected Lennard-Jones fluid with a block size $b = \sigma$. Projected molecules are indicated by a point. Although σ is the hard sphere radius of the particles, more than one particle can be found per block as we deal with projections. **B.** Density distributions for three different values of the block size b. Below the correlation length $\xi = \sigma$ of the system, the distribution is not Gaussian. However, for $b > \xi$, the density is Gaussian distributed. **C.** The reduced fourth order cumulant U_b, eq. 12.3, approaches the value zero for large block sizes b, indicating a Gaussian density distribution above the typical interaction length scale $\xi = \sigma$ of the system. **D.** The compressibility of the system. The inset shows that for length scales larger than σ, the block-size dependent compressibility κ_b can be extrapolated to the limit $b \to \infty$, yielding the isothermal compressibility $\kappa_T/\kappa_0 = 0.46$ of the system. Compressibilities are scaled with the isothermal compressibility of an ideal gas $\kappa_0 = 1/p = (\rho k_B T)^{-1}$ to obtain independency of the systems average density.

distribution is binomial. The probability that box q_i contains N_i points is given by

$$\binom{N}{N_i}\left(\frac{V_i}{V}\right)^{N_i}\left(1-\frac{V_i}{V}\right)^{N-N_i}.$$

Apart from finite-size effects, this leads to a Gaussian distribution. The reduced fourth-order cumulant [203, 204]

$$U_b = 1 - \frac{\mu_4}{3\mu_2^2} \tag{12.3}$$

is zero in this case. $\mu_k = \left\langle (\rho_b - \langle\rho_b\rangle)^k \right\rangle$ denote the centered k^{th} moments. In case of pairwise or higher-order interactions, deviations from the Gaussian behaviour become visible here. However, above the length scale where interactions take place, i.e. $b \gg \xi$, the fluctuations are again Gaussian as every box can be viewed as an independent realization picked from the ensemble.

Besides structural information, the density fluctuations contain information about mechanical properties of the system. The isothermal compressibility κ_T is related to the density fluctuations by [203]

$$b^2 \left(\left\langle \rho^2 \right\rangle_b - \langle \rho \rangle_b^2 \right) = \langle \rho \rangle^2 \, k_B T \kappa_T^{(b)} \, . \tag{12.4}$$

Again, the actual isothermal compressibility κ_T is obtained only in the limit of $V \to \infty$ and $b \to \infty$. From the block density distributions, it has to be calculated by extrapolation

$$\kappa^{(b)} \to \kappa \qquad \text{for } b \to \infty, V \to \infty \tag{12.5}$$

Thus, the block distribution is in a sense complementary to the radial distribution function $g(r)$. While the latter gives information on the short scale, the compressibility is related to large scale behaviour of the system. Indeed, we have $\kappa \sim \lim_{q \to 0} S(q)$, thus representing the large scale $r \to \infty$.

12.4.2 Results

The test system: A Lennard-Jones fluid

In section 12.3.2 the Lennard-Jones fluid was introduced as a well-known model system. Here, we use simulations of the Lennard-Jones fluid as a reference for comparison to the experimental histone distributions. Simulations are performed in three dimensions, however, for the analysis, the particle positions are projected on the xy-plane to allow comparison to the experimental data. In Fig. 12.7A the block densities are visualized for a block size of $b = \sigma$, corresponding to the typical length scale ξ of the system. The density distributions $P(\rho_b)$ shown in Fig. 12.7B display the expected behaviour. Below the length scale $\xi = \sigma$ where interactions take place, the density distribution has pronounced deviations from a Gaussian shape. Above this length scale, $P(\sigma_b)$ becomes more and more Gaussian. The reduced fourth order cumulant U_b asymptotically adopts the value 0 (Fig. 12.7C). As expected and in agreement with simulations by Rovere *et al.*, the isothermal compressibility converges to a constant value in the limit of $b \to \infty$ (Fig. 12.7D). Strong deviations are visible on the scale below $\xi = \sigma$. Note that the compressibility is always plotted relative to the compressibility $\kappa_0 = 1/p$ of an ideal gas to get rid of the dependency on the average density.

Compressibility analysis reveals large-scale fluctuations

Fig. 12.8 shows the compressibility κ_b for the three cell types studied. Data is scaled with the compressibility of a random distribution of points (ideal gas) $\kappa_0 = 1/p = (\rho k_B T)^{-1}$ to remove effects of different labeling efficiencies, i.e. densities. Interestingly, the data does not converge for stably transfected HeLas (strain II), indicating that fluctuations exist on a length scale far above 1 μm. Fibroblast cells display less fluctuations, resulting in a smaller compressibility, however, pronounced differences are found to HeLa cells of strain I. The latter show a compressibility converging at about 1 μm, indicating that typical chromatin fluctuations are of this size range. Thus, differences in the expression of the protein leads to different structural patterns, not only on the short scale as indicated by the RDF $g(r)$, but also on the scale of the nucleus. The compressibility strongly differs both for different cell types, suggesting that large-scale properties depend on the cell's differentiation state, as well as different labeling methods. To exclude the possibility that insufficient statistics inside the cell is responsible for the observed behavior, we have simulated cells of the same

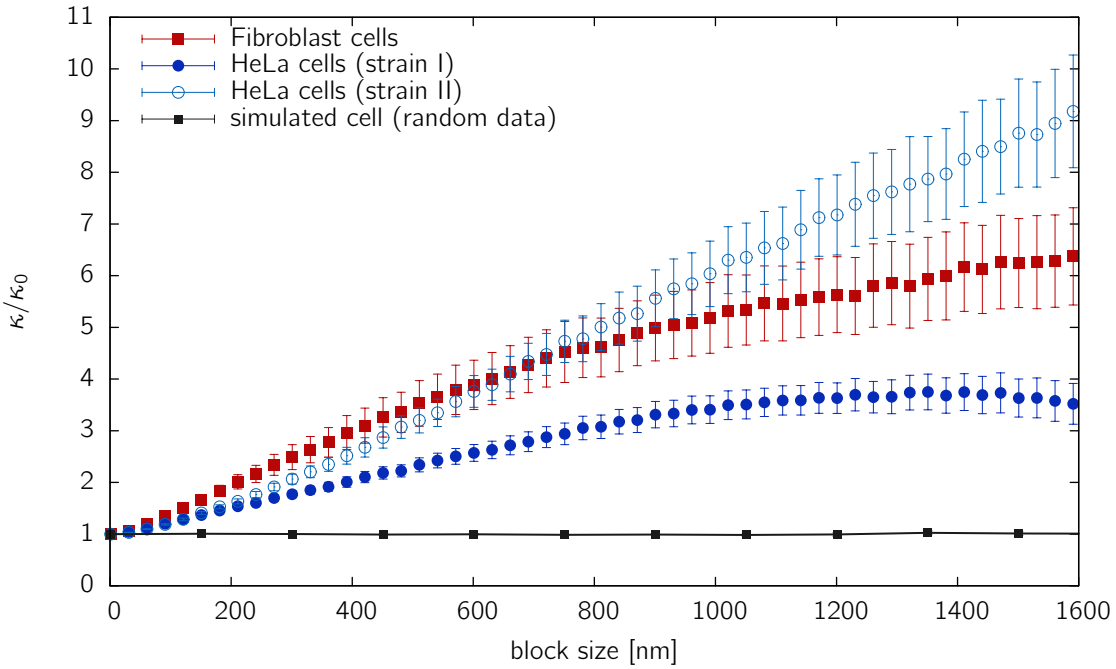


Figure 12.8: The compressibility ratio κ_b/κ_0 versus block-size. The compressibility κ_b is scaled with the compressibility of an ideal gas $\kappa_0 = 1/p$ obtain density-independent values. The non-convergence of the compressibility indicates large scale fluctuations of the histone distribution in the cell nucleus.

shape and density using a random distribution of points. For all block sizes studied, the simulation shows no deviation from the expected value $\kappa_b/\kappa_0 = 1$, rendering statistics issues unlikely.

The observed fluctuations on the large scale might reflect regions with locally different chromatin density. Experimental studies indicate that regions of increased gene density (ridges) as well as transcriptionally inactive regions (anti-ridges) mapped according to the human transcriptome map, spread over an average distance of up to 1-2 μm [12]. This value corresponds to estimates of the average size of chromosome territories [19]. Ridges and anti-ridges show pronounced differences in the compaction of chromatin, leading to histone density fluctuations. Furthermore, fluctuations might arise due to heterochromatic and euchromatic regions.

Interestingly, the compressibility is markedly larger for all cell types than for an ideal gas, i.e. a random distribution of points. This is somewhat unexpected as interactions typically lower the strength of the fluctuations. The Lennard-Jones system, for example, shows a compressibility κ_T which is lower by a factor of about 2 (Fig. 12.7D) compared to the ideal gas system.

Block density distribution reveals non-Gaussian distribution patterns

Further information on the fluctuations can be gained directly from the block density distributions $P(\rho)$, which should adopt a Gaussian distribution above the correlation length ξ. An illustration of the density distribution for a block size of $b = 500$ nm is illustrated in Fig. 12.9A and B for a fibroblast cell and a HeLa cell of strain I respectively. From visual inspection we find already that non-random distribution patterns prevail in the cell over

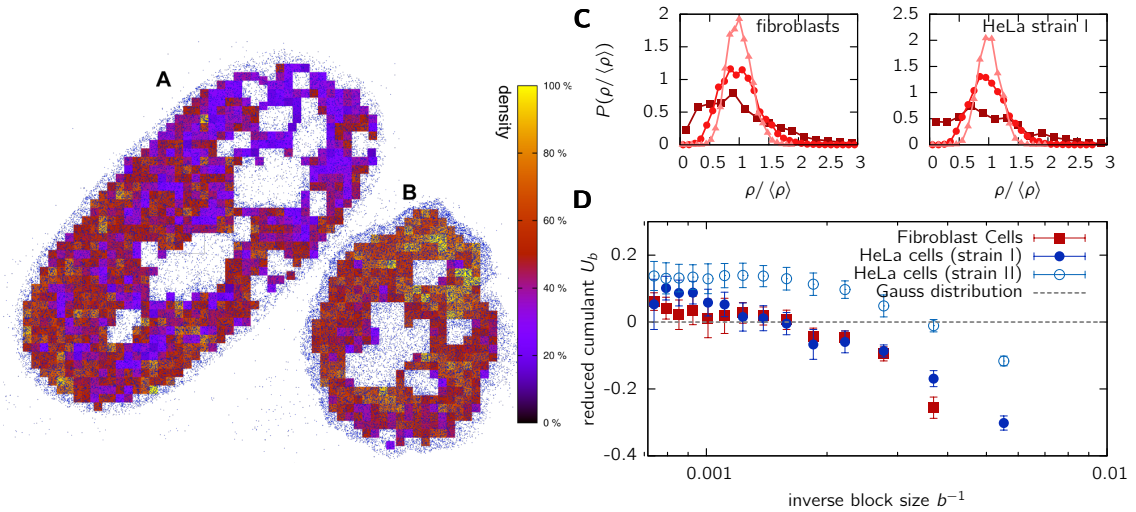


Figure 12.9: Visualized local densities for a fibroblast cell (**A**) and a HeLa cell of strain I (**B**). Shown are the densities of blocks of linear dimension $b = 500$ nm. They display strong deviation from a random distribution, indicating that structure in the histone distribution exists on that scale.Panel **C** displays the block density distributions $P(\rho/\langle\rho\rangle)$ for different values of block size: $b = 200$ nm (■), $b = 500$ nm (●), $b = 1000$ nm (▲). Data is scaled with the average density $\langle\rho\rangle$ to make the distribution independent of labeling efficiency. **D.** The cumulants U_b [eq. (12.3)] of the density distribution. For a thermodynamically equilibrated system they should display a value of zero (grey line) above a block size b larger than the scale of density fluctuations.

large length scales. For example, the upper part of the fibroblast cell in Fig. 12.9A seems to be depleted of fluorophores. This might be due to a heterochromatic region or to non-accessibility of the fluorescent markers in that region, which in turn might be related to chromatin density. Distinct connected regions with only a small density of localized points can be found throughout the cell spanning length scales of several μm. These regions again might be connected to differences in local chromatin compaction. Fig. 12.9C displays the density distributions $P(\rho)$ for both fibroblast cells as well as HeLa cells (strain I) for block sizes of $b = 200$ nm, $b = 500$ nm and $b = 1000$ nm. The distributions are scaled with the average density to obtain independency from labeling efficiency. Expectedly, clear deviations from a Gaussian shape are visible on the scale below 500 nm. For $b = 1000$ nm deviations from a Gaussian cannot be distinguished by visual inspection, therefore we apply a measure, the reduced fourth order cumulant U_b [eq. (12.3)], which should become zero for a Gaussian distribution. Interestingly, the fourth order cumulant does not vanish even for large block sizes up to $b \approx 1.5\ \mu$m (Fig. 12.9D), indicating non-Gaussian fluctuations on this scale.

12.5 Relating to polymer models

In this thesis, the Random Loop model (↷ chapter 5) and the Dynamic Loop model (↷ chapter 10) were introduced. Both models, basing on the same idea, make testable predictions on a huge variety of observable quantities. In fact, the Dynamic Loop model offered a unified framework of chromatin organization bridging the gap between genome folding and function. While measures of shape and organization can be easily evaluated in such a coarse-grained polymer model, simulations are not feasible on the scale of single histones.

Nevertheless, in the following sections, an approach is developed that allows comparison of the model with experimental data from localization microscopy.

We ask the question whether

- the observed pair distribution function (Fig. 12.6) can be explained in terms of the Dynamic Loop model.
- the observed high compressibility values compared to a random distribution of points (Fig. 12.8) are due to the disorder induced by looping.

12.5.1 Simulations of the Dynamic Loop model in a melt

Localization microscopy reveals the distribution of histone H2B molecules in a projected slice of the cell. Clearly, the structure of chromosomes can not be evaluated by this method, as it is not known at which chromosome one localized histone is attached. Mapping histone distribution onto simulated polymers requires taking into account realistic time scales and densities. While large-scale measures like size and shape have been shown to be independent on the level-of coarse-graining (↷ chapter 10), mimicking histone distributions requires a coarse-graining on the size scale of the persistence length (↷ 3.2). To obtain a mapping from lattice units to physical scales, we assume a persistence length of $l_p = 150$ nm. Although this value has been obtained by a fit to a linear worm-like chain model [8], which does not describe chromatin folding, it is the best estimate currently available. One lattice unit is set to 45 nm based on the following rationale: The minimum distance in lattice units (l.u.) between two bonds which can occur during the simulation run in the bond fluctuation method is $d_{\min} = 2/3$ l.u.. Assuming an average diameter and minimum contact distance of the chromatin fiber of 30 nm we have to set $d_{\min} = 30\,\text{nm} = 2/3$ l.u. As the average bond length of a simulated polymer is $\langle b \rangle \approx 2.71$ l.u. $= 122\,\text{nm} < l_p$, bending rigidity has to be introduced into the model. This is accomplished by adding a weight factor for the bond angles given by

$$U_{\text{bond}} = \kappa_\theta (1 - \cos\theta)$$

Similar potentials have been used in Refs. [30, 205]. To establish a relationship between persistence length l_p and the potential parameter κ_θ, we conduct simulations of self-avoiding walks in a semi-dilute solution of occupancy fraction 10%. Chromatin occupies approximately 10% of the nuclear volume, a value has also been used in other studies [35]. Simulations are run with different values of κ_θ and the angular correlations $\langle \cos\theta \rangle$ of neighbouring bonds are calculated. The relationship bewteen the potential parameter κ_θ and the angular correlation is shown in Fig. 12.10. A realistic chromatin model requires that $\langle \cos\theta \rangle = \exp(-l/l_p) = \exp(-2.71 \cdot 45\text{nm}/150\text{nm}) \approx 0.44$, yielding a interaction strength of $\kappa_\theta = 1.34$.

As simulations are conducted on a very low level of coarse-graining, it becomes intractable to simulate whole chromosomes. This, however, probably does not have a major impact on the small scale structure at the level of single histones. Simulations are performed in a system of size $L^3 = 64 \times 64 \times 64$ and a density of $\rho = 10\%$. 27 chains of length $N = 256$ are put inside the simulation box. The relaxation time τ_{int} of the system was determined using the integrated autocorrelation time of the radius of gyration according to the method described in section 10.2.2. We obtain a value of $\tau_{int} = 4\,500\,000$ for a system of linear chains (looping probability $p = 0$). Simulations of chains with loops are performed using an average lifetime of $\tau = 100\,000$, however, it has been shown in chapter 10 that the results are not critically dependent on the lifetime. Looping probabilities

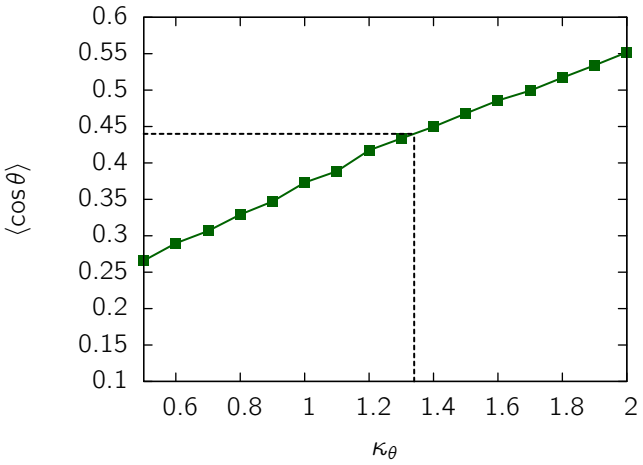


Figure 12.10: Relation between the bond-angle potential $U_{\text{bond}} = \kappa_\theta(1 - \cos\theta)$ and the persistence length for linear chains using the bond-fluctuation algorithm. Shown are simulational results of a system with density $\rho = 0.1$ and linear system size $L = 64$ lattice units. The parameter κ_θ was varied and the resulting average bond angle between neighbouring bonds $\langle\cos\theta\rangle$ was calculated. At $\kappa_\theta = 1.34$ the potential resembles the biological case where the persistence length is 150 nm.

chosen are $p = 0.03, p = 0.06$ and $p = 0.1$, corresponding to an average number of loops of 33, 82 and 144 per chain, respectively.

To test whether this modified Dynamic Loop model with bending rigidity still displays a confined folding, i.e. the formation of chromosome territories, the mean square distance $\langle R_n^2\rangle$ between two beads is calculated for the values of p above. Figure 12.11 displays the mean square distance for linear chains as well as looping polymers. A leveling-off in the spatial extent of the chain is recovered. The plateau level is below the experimental values of 1-2 μm due to the small chain lengths simulated.

12.5.2 Results

To compare simulational results with localization microscopy data, we simulated the distribution of GFP-labelled histones along the chromosomal fibers. Lacking knowledge on the detailed binding properties of GFP-labelled histone H2B and its distribution along the linear genome, a random (uniform) distribution of fluorophores along the fibers is assumed. The linear dimension of $L = 64$ l.u. corresponds to 2.88 μm according to the coarse-graining procedure described above (↷ 12.5.1). Experimental data evaluation is done in a 2D projection of a slice of thickness $d = 600$ nm. The average density of localized histones has been determined to about $150/\mu\text{m}^2$ (table 12.1). In order to obtain a

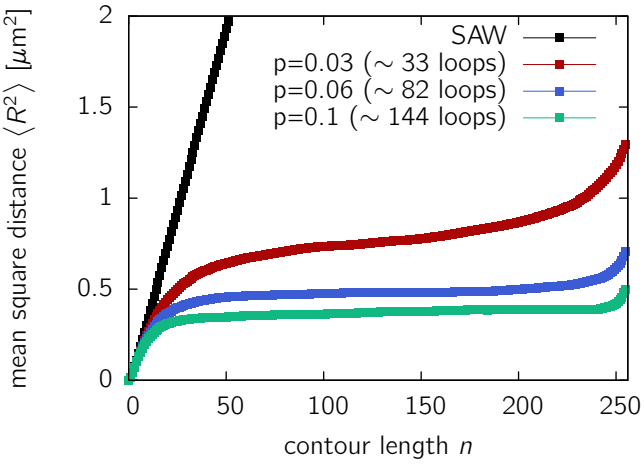


Figure 12.11: Mean square distance $\langle R_n^2\rangle$ bewteen two beads separated by n monomers. Simulations are performed using the Dynamic Loop model with a bending rigidity. The system consists of chains of length $N = 256$ at a density $\rho = 10\%$. Results are calculated for linear chains (black) and chains with different strengths of loop formation.

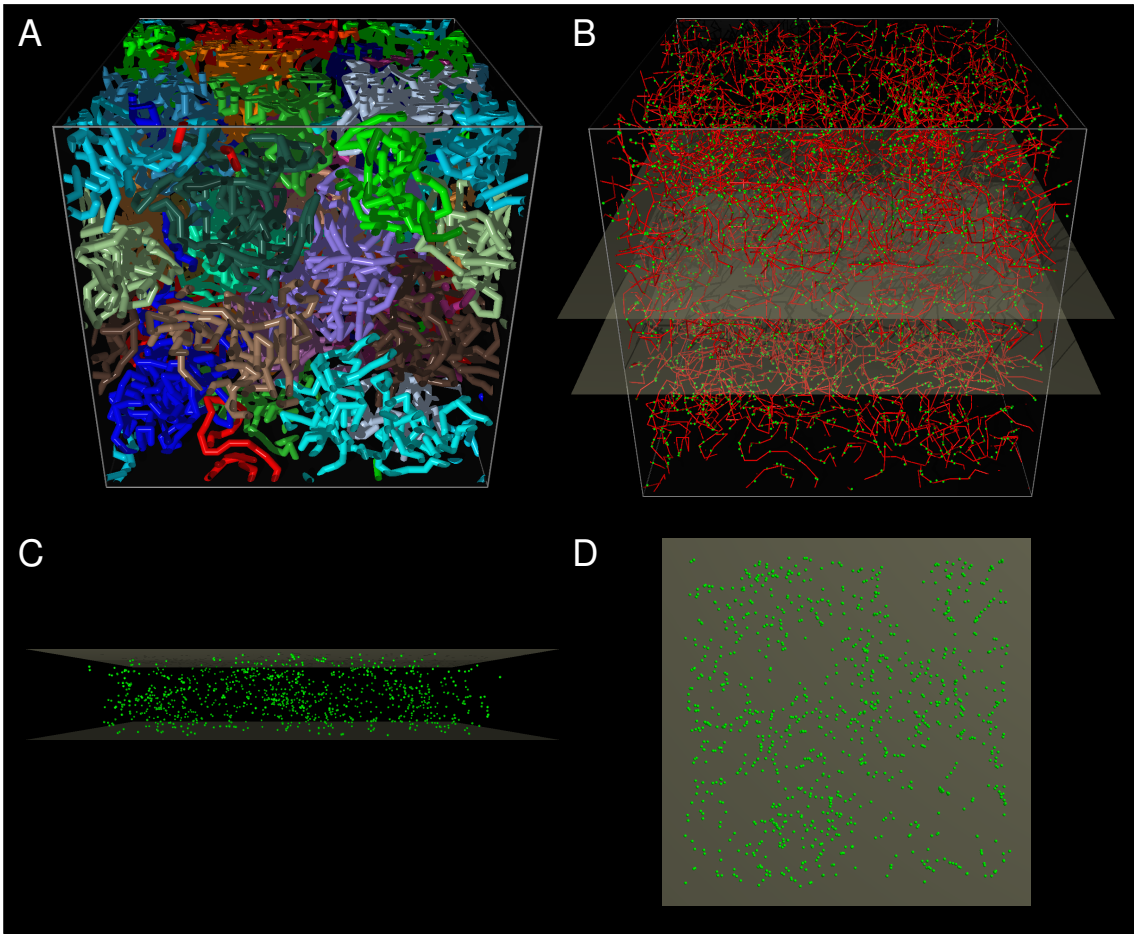


Figure 12.12: Histone distribution and projection for simulated chromosomes (Dynamic Loop Model). **A.** Example conformation with several chromosomes in the simulation box. Simulations are set up with an average number of loops per monomer of 0.32. **B.** Distribution of localized histones along the fibers (green points). The red threads mark the path of the chromosomes, the two planes have a distance of 600 nm. **C.** A slice of thickness 600 nm is cut out of the simulation box. **D.** The localized histones inside this slice are then projected onto the plane for conducting the block analysis.

similar density for simulated fibers, $K = 5971$ virtual histones are randomly distributed along the contour of the simulated chromosomes. To mimic the effect of the thickness of the chromatin fiber, each point is shifted randomly within a radius of 15 nm in the plane perpendicular to the chromatin fiber. Furthermore, excluded volume effects are taken into account by not allowing two virtual fluorophores be closer than 5 nm, the estimated radius of a nucleosome [1]. An example conformation of the system is shown in Fig. 12.12A. The distribution of histones is visualized in Fig. 12.12B. The chromosomes are shown as small red threads, the simulated localized fluorophores are marked by green points. After distributing histones along the simulated chromosomes, a z-slice of thickness 600 nm is cut out of the simulation box (Fig. 12.12C) and all histones inside this slice are projected onto the xy-plane (Fig. 12.12D). This projected data is then analyzed the same way as the experimental data, i.e. the block distribution analysis is conducted.

Obviously, the analysis is based on the assumptions that

1. the model does correctly describe chromatin folding

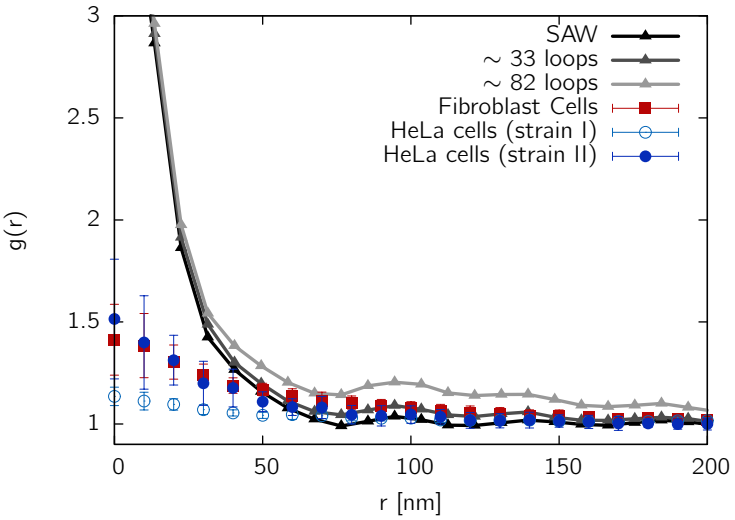


Figure 12.13: The radial distribution function (RDF) $g(r)$ for the DRL model compared to experimental results. A system of chromosomes with chromatin density $\rho = 0.1$ has been simulated using a persistence length of $l_p = 150$ nm, in agreement with the current knowledge on human chromatin. Both model and experiments display pronounced deviations from a random distribution of histones on the short scale, however, the model can not quantitatively explain the data.

2. the fluorophores are randomly distributed along the histones of the chromatin fiber

3. all fluorophores are attached to histones bound to chromatin

Surely, neither of the assumptions is fully correct: The model does not take into account the heterogeneity of the chromatin fiber; the positioning of fluorophores will be determined by accessibility to certain chromatin regions; finally, there are histones H2B which are not bound to chromatin. In view of these restrictions it is clear that a model is not able to quantitatively describe all the features observed, rather qualitative predictions can be derived.

The radial distribution function $g(r)$ is shown in Fig. 12.13 for the simulated chromosomes. Interestingly, similar to experimental data, the model chromosomes display strong deviations from a random distribution of points at short scales below 100 nm. As such a behaviour is observed both for linear as well as looping chains, the reason for this is the constrained positioning along the contour of the fibers. As the RDF $g(r)$ displays short scale behaviour, differences in the curve for different looping probabilities p are subtle. The values of $g(r)$ are much larger for the model polymers than for the experimental data (Fig. 12.6). One possible reason is that not all fluorophores are attached to histones bound to chromatin, so that the position constraint is relaxed to some extend in the biological system, resulting in a more "ideal gas"-like structure. Another important point is that the model is a coarse-grained model not taking into account the detailed small-scale structure of the chromatin fiber. Thus, as the length scale is below the scale of coarse-graining, one cannot expected the model to reproduce the data.

The block compressibility κ_b/κ_0 for the chromatin model is shown in Fig. 12.14. Compared to the case of linear chains (self-avoiding walks), polymers with looping interactions display stronger fluctuations increasing with the looping probability p. The block-size dependent compressibilities κ_b adopt a constant value in the limit of large b. However, the compressibility displays a saturation at block sizes of about $b \approx 700 - 800$ nm. This length scale corresponds to the typical size of the chromosome territories in our model 12.11, where the chain lengths used are much smaller than typical chromosomes. In agreement with the typical size of chromosome territories of 1-2 μm observed in experiments [12, 7], the results from our data analysis do not show a saturation below 1-1.5 μm.

Importantly, the obtained compressibility values for looping polymers are in the size

range of the experimental observations, which show pronounced differences from the linear polymer behaviour. Extrapolated compressibilities for our model range from $\kappa_b/\kappa_0 \approx 2$ (linear chains) up to $\kappa_b/\kappa_0 \approx 6.9$ (0.56 loops per segment on average). Thus, dynamic loop formation induces strong fluctuations in the overall density distribution of the system comparable to experimental findings.

Similarly, large fluctuations in the ensemble of cells exceeding the value of simple polymer models have already been discovered for the distance distribution between two fluorescent markers (Figures 2.3 and 10.4). Thus, a huge variety of densities and conformations seems to constitute a major feature of chromatin organization, which can be consistently explained by the dynamic formation of loops on the scale above the coarse-graining length.

12.6 Conclusions

We have applied 2D localization light microscopy and statistical techniques to infer structural features and mechanical properties at the nuclear nanoscale. Localization microscopy allowed us to take a look at the chromatin nanostructure with a non-invasive method. Such light optical methods to resolve structures by optical isolation are well-established and known to be able to resolve objects on the scale of single histones, allowing for the analysis of structures on the nanoscale.

To study the spatial distribution of histones and its dependency on cellular or expression parameters, we applied localization microscopy to two different cell types, human

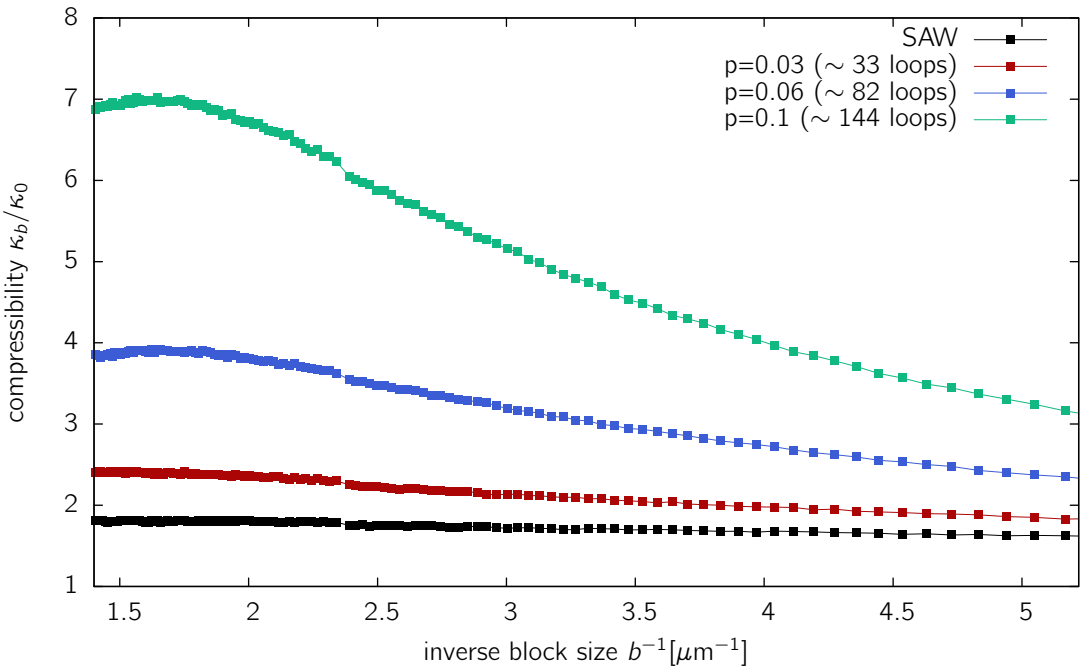


Figure 12.14: The block compressibility κ_b for the Dynamic Random Loop model. Shown are the results for a system of density $\rho = 0.1$ and a persistence length of $l_p = 150$ nm, in agreement with the current knowledge on human chromatin parameters [30]. The values of κ_b are normalized with the value expected for an ideal gas $\kappa_0 = 1/p$. Interestingly, compressibility and therefore fluctuations increase strongly with looping probability $\mathcal{P}$.

fibroblasts and HeLa cells. Two different HeLa strains were used. One HeLa strain is labeled with emGFP, the other strain is a stably transfected cell line expressing H2B-GFP fusionprotein.

Statistical analysis of the histone distribution revealed distinct differences from an average density distribution (Fig. 12.6). It could be evidenced that below the Abbe resolution limit of about 200 nm nucleosomal structures are existent in cell nuclei, a result which is shown to be robust with respect to cell type, strain and labeling method. Our results furthermore indicate that different mechanisms of expression (plasmid, stable integration) of the same protein may lead to significantly different localization patterns on the nuclear nanoscale.

Structural and mechanical properties of the histone distribution are obtained by evaluating local density fluctuations inside the cell nucleus. We find that the histone distribution shows pronounced differences from an ideal gas or a simple fluid organization. Interestingly, density fluctuations exist on scales larger than 1–2 μm, a result which is probably related to differences in large scale chromatin compaction as found in ridges and anti-ridges. One mechanical measure directly derived from the histone distribution is the compressibility of chromatin (Fig 12.8). We find that measured compressibilities deviate with respect to cell type and expression mechanism. A common feature of chromatin organization seems a higher compressibility compared to a fluid of equal density. This increase in compressibility which is even larger than the non-interacting ideal gas is in agreement with the recently proposed Dynamic Loop model [93], suggesting that loop formation induces the observed large fluctuations.

Although fluctuations are well explained by the Dynamic Loop model, it does not yet take into account local differences in compaction, i.e. gene-rich and gene-poor regions. Fluctuations arising thereof are not covered by the model. Incorporating such regional differences might even yield a better agreement with experimental observations. Furthermore, the distribution of localized fluorophores along the chromosomes is assumed to be uniform. This is an important simplification as it is well-known that genes are distributed non-randomly [13] along the contour of the chromosomes and it is hypothesized that histones itself code information (histone code) [54]. Notwithstanding these local deviations from the model's assumptions, interestingly, loop formation leads to fluctuations, a result which already could explain the finding of large cell-to-cell variation in distance measurements ($\curvearrowright$ 10.3.3).

A critical experimental issue when comparing to polymer models is the question, how many fluorophores do actually bind to histones which are attached to chromatin. In fact, some of the histones, thus some fluorophores, might be floating freely in solution. Also, there is evidence that certain GFP variants tend to cluster [206], therefore not directly representing chromatin structure. The qualitative differences found in this study dependent on cell type and expression method remain unaffected by these issues.

This method at hand can be used for a variety of applications. An expression sensitive analysis method could be used, for example, to study the nanostructure of gene-rich and highly transcribed regions in comparison to gene-poor regions with little transcriptional activity. These show different folding patterns on a scale of about 0.5–2 μm [12, 6], yet it its not clear whether differences exist on the nanoscale. Other interesting analysis regions comprise organization of telomers and heterochromatin foci. In general, the method might be useful to distinguish between cells in different states characterized by a variation in nanostructural chromatin organization.

Chapter 13

Pattern analysis of confocal images

References

The work described in this chapter is a collaboration with the group of Prof. Lindsay Shopland (Institute for Molecular Biology, The Jackson Laboratory, Bar Harbor, Maine) and Dr. Jörg Bewersdorf. Specimen preparation and image acquisition has been performed at the Jackson Laboratory. Data was kindly provided by Lindsay Shopland and Jörg Bewersdorf. The results are not yet published as our analyses raised a few questions. Currently, further experiments are undertaken and a publication is planned including both the analysis of the present chapter as well as that of future data.

Chapter Summary

Images from conventional confocal light microscopy are limited to a resolution of about 600 nm in axial direction and 200 nm in the focal plane, rendering it impossible to extract the detailed folding pathways of the chromatin fiber. Nevertheless, various analysis methods from pattern analysis to the comparison with simple models like the Ising model can be applied to obtain structural information. Mainly two questions are of interest, which can be investigated with conventional light microscopy: The distribution of active and inactive genes within the nucleus as well as its dependency on cell differentiation. In this chapter, we present measures to quantitatively analyse the nuclear distribution of active and inactive genes in undifferentiated mouse embryonic stem cells (ESCs) and provide a comparison to differentiated cell types. We find that the distribution of active genes is similar in both stem and differentiated cells, while repressed genes show different correlation patterns in stem cells. Furthermore, repressed genes are less clustered throughout the nucleus than active genes. We demonstrate that the observed patterns cannot be described by a Potts- or Ising-type of model. Polymer models, on the other hand, require more elaborate experimental input to provide a reasonable basis for comparison.

13.1 Introduction

The structure of chromosomes inside the nucleus of interphase cells can be assessed using different methods. Generally, the detailed folding path of the chromatin fiber cannot be visualized in living cells due to the limited resolution of conventional confocal light microscopy of about 200 nm in the focal plane. Indirect approaches, some of which have been mentioned in this thesis, have been applied to obtain information about the structure and folding of the chromatin fiber inside the nucleus. One prominent technique is to apply fluorescence in situ hybridization (FISH) to obtain information about the spatial distribution of certain genomic regions. This can be achieved either by labeling two chromatin segments and establishing a relationship between the physical distance and the genomic separation (↷ 2.3.4) or by determining the shape and positioning of larger chromosomal regions [6, 129].

High-resolution localization light microscopy allows for an alternative approach of studying chromatin structure offering a resolution down to the single histone level. In chapter 12 we have applied this novel technique to investigate chromatin nanostructure and its differences between fibroblasts and HeLa cells. Currently, this method is rather invasive, requiring powerful laser sources and specific fluorescent markers.

Despite its limited resolution, confocal light microscopy up to now allows for a higher throughput than localization microscopy as well as a larger variety on usable fluorescent dyes. The scope of this chapter is to develop quantitative and differential methods to investigate structure based on images from conventional (confocal) light microscopy.

Although it is well-known that the pluripotent state of embryonic stem cells (ESCs) requires specific transcription factors to perform tasks of transcriptional regulation [207, 55], the detailed structure of chromatin in ESCs and its changes towards more differentiated cells are not understood. Many studies indicate that the positioning of a gene greatly influences its activity: In general, gene-rich and transcriptional active regions tend to be

located towards the nuclear interior, while inactive regions are predominantly found at the periphery [6, 14, 15]. Tethering experiments, where active genes are moved towards the nuclear periphery, have revealed that indeed active genes become repressed close to the periphery [208]. The other way round, the spatial positioning was found to change on transcriptional activation [18]. However, the converse behavior has been observed in some experiments [16] indicating that the nuclear periphery is far from being depleted of active genes.

Embryonic stem cells are distinctly different from differentiated cell types. They are pluripotent, meaning that they can differentiate into a vast amount of specialized cells. Chromatin organization in stem cells also differs from that of more differentiated cells. First, ES cells display a rapid histone exchange, the structural chromatin proteins only binding loosly to the chromatin fiber. It has been suggested [209] that this hyperdynamic binding is an important hallmark of pluripotent ESCs being responsible for maintaining the plasticity of undifferentiated cells. Secondly, ES cells display a vast number of transcriptionally active foci, which are silenced upon differentiation [210].

A recent study by Luo *et al.* [55] focussing on the nuclear periphery has revealed that it contains substantial fractions of both active as well as inactive genes. Interestingly, the amount of transcriptional active genes is the same for ESCs as well as differentiated cells. However, the peripheral fraction of repressed genes is significantly higher in stem cells.

The aim of this chapter is to quantitatively analyze the structure of chromatin based on images from conventional confocal light microscopy. We establish measures and investigate differences between embryonic stem cells and more differentiated cells based on the distribution of markers for active and inactive genes. We focus on the questions of

- whether there are structural differences of chromatin in stem cells compared to differentiated cells,
- whether the distribution of active and inactive genes inside the nucleus differs.

After having quantitatively analyzed the data, we aim at developing models for the organization. First, we investigate how the results from the Dynamic Loop model (↷ chapter 10) can be compared to conventional microscopy images and discuss which information is needed in order to perform this task. Then, in a more abstract manner, we ask whether the observed intensity patterns can be explained by a Potts model or Ising model, which are often used in image segmentation and reconstruction.

13.2 Experimental Data and Image Processing

To investigate the distribution of active and inactive genes in differentiated as well as stem cells, three different types of mouse cells were used: (i) Embryonic stem (ES) cells, (ii) adult primary neuronal stem/progenitor cells (NPCs) and (iii) embryo-derived NIH-3T3 fibroblasts. The images used in this chapter have been kindly provided by Prof. Lindsay Shopland (Institute of Molecular Biology, The Jackson Laboratory, Maine).

Each of the cells was immunolabeled with antibodies raised against covalent histone modifications typically found on transcriptionally active and repressed genes. The histone H3 trimethylated on lysine 4 (H3K4-Me_3) is predominantly found on promoters of active genes, whereas histone H3 trimethylated on lysine 36 (H3K36-Me_3) accumulates on the transcribed regions of genes. Finally the histone modification (H3K27-Me_3) accumulates on the promoters of silenced genes as well as bivalent genes, which are repressed in ESCs and thus the coding regions do not accumulate H3K36.

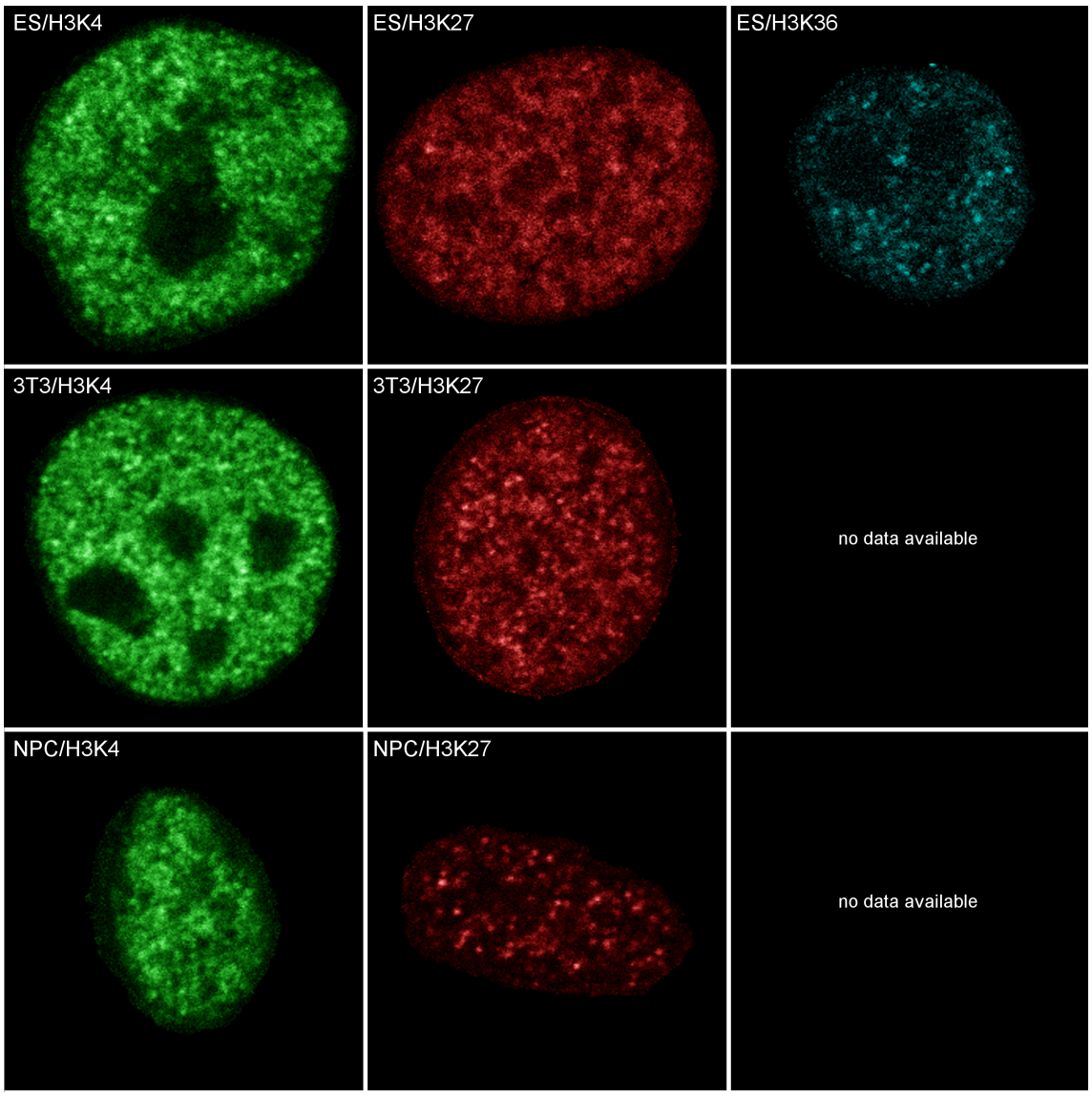


Figure 13.1: Samples of confocal microscopy images. For each cell type [embryonic stem cells (ESCs), 3T3 cells and neuronal progenitor cells (NPCs)] a z-stack is shown for various labelled histone types. Histone modification H3K4, indicating promoters of active genes, is artificially labelled green; histone modification H3K27, accumulating on promoters of silenced genes, is labelled red. Immunostaining for H3K36 has only been applied to stem cells.

3D images were acquired using a confocal microscope (see Ref. [55] for details). The "Erosion" program written by Joerg Bewersdorf has been applied to segment the images [184]. It produces a series of nested 3D shells defining the distance from the nuclear edge, the thickness of one shell being 200 nm. The volume outside the nuclear edge is given by a shell number of -7, making it easy for us to determine the region of interest (ROI), i.e. the cell nucleus, which is used for analysis. Typical two-dimensional slices of the cells are shown in Figure 13.1 for all three cell types as well as the different histone modifications. H3K4, representing active genes, is shown in green color, histone modification H3K27 accumulating on repressed gene promoters is shown in red color. The modification H3K36 has only been labelled for ESCs. Importantly, subsequent analysis methods are

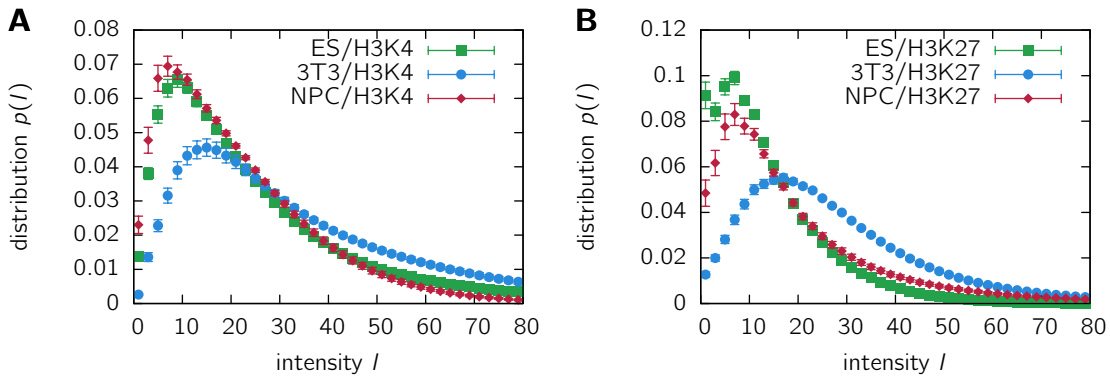


Figure 13.2: Intensity distribution of stained histone modifications. Shown is the distribution of intensity values inside the cell nucleus for embryonic stem cells (ESCs) as well as differentiated NPCs and 3T3 cells. Panel **A.** represents the distribution of intensities where immunostaining of H3K4, i.e. active genes, was applied. Panel **B.** displays the results for immunostaining of H3K27, i.e. repressed gene promoters. Errorbars represent the standard error of the cell-to-cell variation.

only applied inside the ROI defined by the Erosion mask.

13.3 Quantitative Analysis Methods

13.3.1 Intensity Distribution and Correlations reveal differences between cell types

To quantitatively assess differences in the observed chromatin distribution patterns, we first evaluate the intensity distribution $p(I)$. This intensity distribution is obtained by creating a histogram of all the voxels' intensity values inside the cell nucleus. The results are shown in Figure 13.2 for histone modification H3K4 and H3K27. As absolute intensity values are dependent on the specific microscope settings and the antibodies used, we restrict the comparison to the same histone modification evaluating its differences depending on the cell type.

Figure 13.2A shows the intensity distribution $p(I)$ for the histone H3 trimethylated on lysine 4, i.e. the intensity distribution of labelled active gene promoters. Interestingly, the intensity distribution in embryonic stem cells equals that of NPCs, the intensities being generally larger for 3T3 cells. Similar results are found for repressed gene promoters, i.e. histone modification H3K27, in Figure 13.2B. These results already indicate differences between ES cells and differentiated cells, more detailed measures are necessary to distinguish between ES and NP cells. Clearly, from visual inspection of Figure 13.1 we find that the distribution of repressed genes is markedly different between ES and NP cells, the latter showing concentrated spots.

To distinguish such concentrated spots from a more random intensity distribution, we study the spatial intensity correlations, given by the correlation function

$$g_I(\mathbf{r}) = \frac{\langle I(\mathbf{r}^\star + \mathbf{r})\rangle - \langle I(\mathbf{r}^\star)\rangle^2}{\langle I^2\rangle - \langle I\rangle^2}\,. \tag{13.1}$$

Here, $\langle I\rangle$ is the mean intensity of the cell. The average $\langle\cdot\rangle$ is over all points such that both $\mathbf{r}^\star$ and $\mathbf{r}$ are inside the cell nucleus (i.e. the ROI). The spherical-averaged correlation

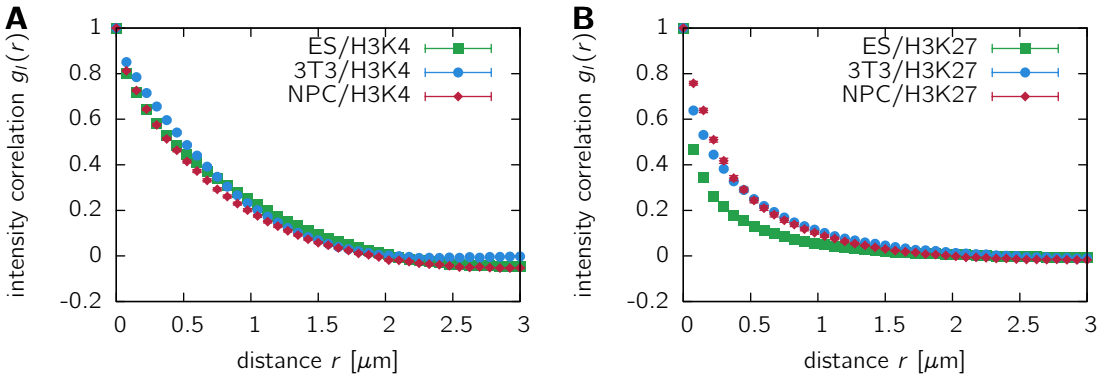


Figure 13.3: Intensity correlations for active and repressed genes. Shown is the correlation function eq. (13.1) for active genes stained for H3K4 in panel A and for inactive genes stained for H3K27 in panel B. Importantly, the distribution patterns for H3K4 are similar in each cells, while repressed genes have different distribution patterns in embryonic stem cells.

function $g_I(r)$ contains information about the distribution of intensity inside the nucleus as well as the long-range order. For example, if the intensities were randomly distributed, $g_I(r)$ would be zero unless $r = 0$. It is to be stressed that the spots marking promoters of active or repressed genes are blurred by the point spread function of the microscope. This point spread function can be approximated within an error of maximally 10% by a Gaussian with a resolution of 220 nm in the focal plane and 600 nm in axial direction [personal correspondence with Joerg Bewersdorf]. Surely, this induces correlations between neighboring voxels (the voxel size is 75 nm in each direction), thus quantitative conclusions on chromatin are prohibitive. However, as the point spread function is the same for every cell type and histone modification, qualitative comparisons are possible.

We find that the distribution of active genes (H3K4) is not strongly dependent on cell type (Figure 13.3A). This is in agreement with earlier findings [55], that the distribution of H3K4 is relatively similar in stem cells as well as differentiated cells. Differences between ESCs and differentiated cells, however, show up clearly in the correlation function of repressed gene promoters (Figure 13.3B). While 3T3 cells and NPCs display a quite similar decay of the correlations, the typical correlation length of ESCs is significantly smaller. Possibly, this finding is related to the more dynamic chromatin organization and the more rapid exchange of histones in stem cells [209, 210]. Interestingly, the typical decay length of the correlation function $g_I(r)$ is smaller for repressed genes than for active genes.

13.3.2 Active genes are more clustered than inactive genes

To measure the extend of clustering of active and inactive genes we apply a thresholding procedure to the microscopy image, where the $p = 20\%$ brightest pixels in the ROI are marked. This is visualized in Figure 13.4 where an embryonic stem (ES) cell is depicted (panel A) with the intensities coded according to the color bar. The voxels belonging to the fraction p of brightest voxels are marked red in panel B, the other voxels inside the ROI are marked black. The grey region indicates voxels lying outside of the ROI as determined by the erosion mask. The value of 20% has been used as a trade-off between making sure that background noise is not incorporated in the analysis and maintaining most of the structural features given by the markers.

We have evaluated both the average size of clusters as well as the cluster size distri-

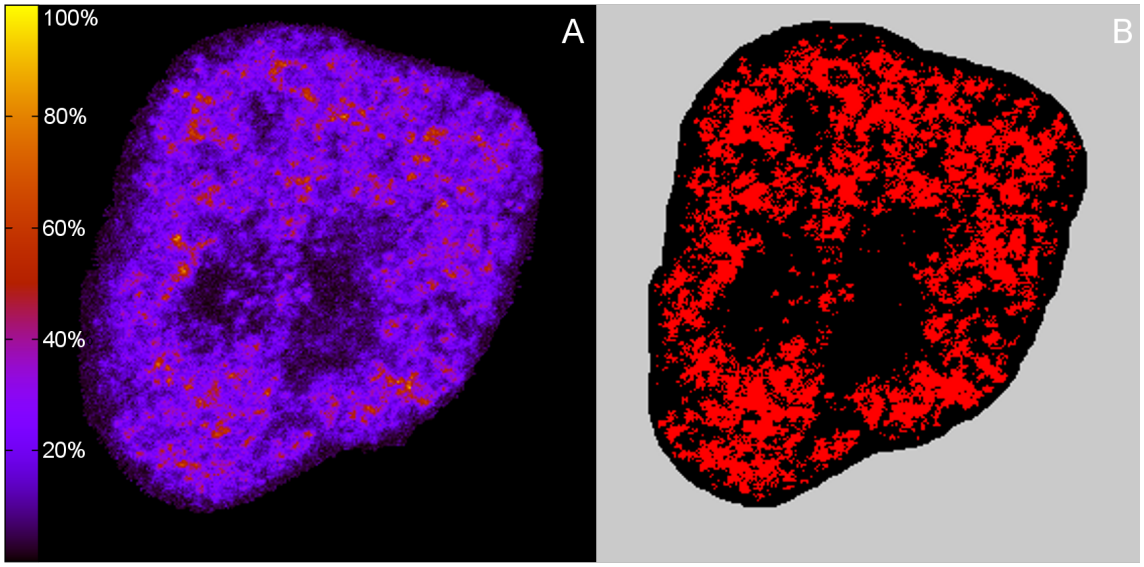


Figure 13.4: Determination of cluster statistics. To determine the amount of clustering of active and inactive genes, the original image (panel A) is segmented based on a thresholding procedure where $p = 20\%$ of the brightest pixels inside the cell nucleus are labelled red (panel B). The grey region corresponds to the outside of the nucleus. The images show a slice through an ES cell immunostained for H3K4.

bution. To achieve this, for each cell all the clusters were enumerated with a self-written algorithm and the average cluster size as well as a histogram of cluster sizes was calculated. The average cluster size is given in table 13.1 for each cell type and histone modification. Strikingly, promoters of repressed genes marked by H3K27 are forming smaller clusters than promoters of active genes marked by H3K4. This is in agreement with the intensity correlations (Figure 13.3) decaying faster for H3K27, indicating that repressed genes are more randomly distributed throughout the nucleus. We find that the average cluster size of repressed genes is significantly smaller in ES cells compared to differentiated cells (by a factor of 0.42 compared to NPCs and a factor of 0.58 compared to 3T3 cells), emphasizing the epigenetical role of spatial chromatin organization. In agreement with the results of the intensity correlations (Figure 13.3A), differences in the distribution of active genes are less pronounced.

Table 13.1: Average cluster sizes of active and inactive genes. This table shows the average size of clusters based on a segmentation procedure where the brightest pixels ($p = 20\%$) are selected. Cluster size averages are calculated for each cell. Uncertainties are standard errors of the cell-to-cell variation.

	ESC	NPC	3T3
H3K4	56.6 ± 1.5	60.0 ± 1.7	71.7 ± 3.8
H3K27	14.5 ± 0.2	34.2 ± 2.1	24.7 ± 0.8
H3K36	61.5 ± 5.3	—	—

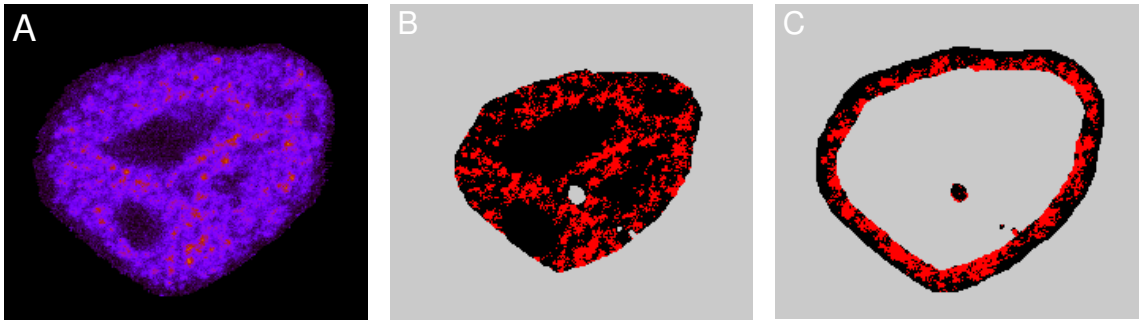


Figure 13.5: Cluster analysis for interior and peripheral chromatin distribution. The panels show an embryonic stem cell immunostained for H3K4, i.e. active gene promoters. The original image (A) is processed such that 20% of the brightest pixels are labeled red. The Erosion shells are used to distinguish between the interior part (B) and the periphery (C).

13.3.3 Repressed genes are more abundant in the nuclear periphery than active genes

The investigation of the influence of spatial position within the nucleus on gene activity has a long history. Regions of the genome with little activity have been found to be located more at the periphery than active genes [6, 129]. To differentiate between interior and peripheral regions we use the mask shells determined by the Erosion program. To distinguish between interior and periphery we assign each voxel having a shell number larger than or equal to zero to the interior region and each voxel with shell number < 0 to the periphery. Results of this segmentation are visualized in Figure 13.5 for an ES cell stained for H3K4.

To study the differences in the chromatin distribution concerning the interior and periphery regions we label the 20% brightest voxels inside the complete nucleus as in section 13.3.2. Then both for the interior region and the periphery region, the fraction of bright voxels is determined. The results are shown in table 13.2. The percentage of intense voxels is generally larger in the interior, which might be due to the erosion masks containing significant parts of the nuclear envelope which are not accessible for chromatin. However, it also indicates that both active and repressed genes are abundantly found in the nuclear interior.

One major finding is that the nuclear interior of ESCs contains significantly larger fractions of active genes than inactive ones, while the converse behavior is found for the periphery, a result being in agreement with Ref. [55]. In NPCs a similar tendency can be observed, which is however much less pronounced. In 3T3 cells, finally, both active and inactive genes are found in the interior as well as the periphery in similar volume fractions.

13.4 Modeling Approaches

While the quantitative analysis methods conducted above show clear differences between ES cells and differentiated cells concerning the distribution of inactive genes, a model explaining these differences is still lacking. In this section, we compare the microscopy data to several modeling approaches.

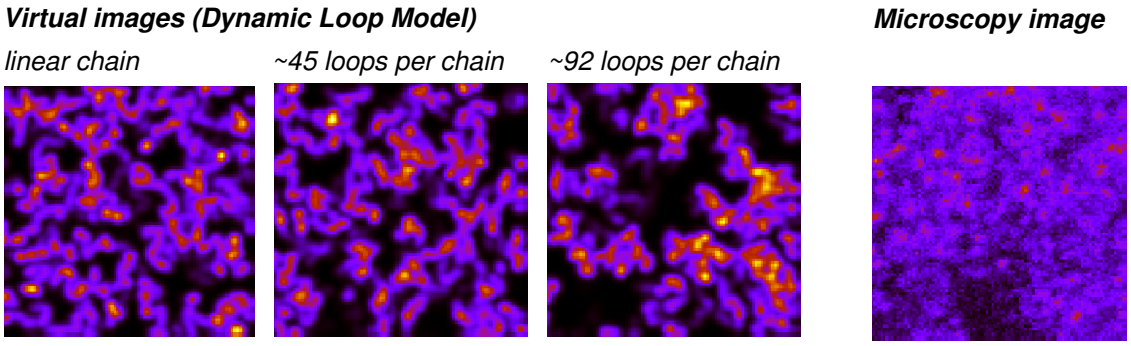


Figure 13.6: Virtual microscopy images of the Dynamic Loop Model. The images have been calculated using simulations of the DL model in a dense system of density $\rho = 0.125$ with a coarse-grained chromosomes of $N = 128$ segments. Fluorophores have been randomly distributed along the contour of the chromosomes. The right-handed image is a raw data microscopy image section provided for comparison.

13.4.1 The Dynamic Loop model

In chapter 10 we have introduced a model which was able to explain various experimental observations on chromatin folding. As this model has been implemented homogeneously, assuming each two points to interact with the same probability when being in close proximity, differences between active and inactive genes cannot be described by this model. Nevertheless we present a method here, how results from simulation can in principle be compared to confocal microscopy images.

In order to achieve this, we use simulational results of the Dynamic Loop model in a dense system as performed in section 10.3.6. Chromosomes of coarse-grained length $N = 128$ are thus simulated in a system of density $\rho = 0.125$ similar to the density inside the nucleus. The Kuhn segment length is chosen to be $l_K = 300$ nm, in agreement with the values used in Refs. [30, 211]. Then, individual labelled histones are distributed randomly along the simulated chromatin fibers. Each labelled histone is assumed to produce a three-dimensional Gaussian-shaped intensity distribution with a peak height of I_{peak}. The width of the distribution is chosen to be 220 nm in the focal plane and 600 nm in axial direction according to the point spread function of the microscope. Unfortunately, the actual peak intensity I_{peak} of a single fluorophore cannot be determined, thus we have to assume some value here. However, we found that the choice of the value does not significantly influence the statistics as long as it is chosen to be one order of magnitude smaller than the average voxel's intensity. The number of fluorophores positioned along

Table 13.2: Average chromatin content in the nuclear interior and periphery. This table shows the average fraction of brightest pixels for the interior region as well as the periphery region. Data is based on a segmentation procedure where the 20% brightest pixels inside the cell nucleus are selected. Then the voxels are classified being either at the periphery or nuclear interior according to the Erosion masks defining the distance from the nuclear envelope.

interior	ESC	NPC	3T3
H3K4	0.36	0.37	0.33
H3K27	0.24	0.31	0.30
H3K36	0.33	—	—

periphery	ESC	NPC	3T3
H3K4	0.10	0.12	0.14
H3K27	0.17	0.14	0.13
H3K36	0.12	—	—

Table 13.3: Average cluster sizes in the Dynamic Loop Model. This table shows the average cluster sizes of virtual images obtained from the Dynamic Loop Model (Figure 13.6). Clusters are determined by labeling 20% of the brightest voxels as described in the text.

	average cluster sizes
linear chain	2100 ± 25
∼ 45 l.p.c.	2278 ± 28
∼ 92 l.p.c.	2894 ± 38

the simulated fiber is chosen such that the average intensity equals that of the microscopy images. After the distribution of intensities, the intensity is summarized in voxels of linear size 75 nm, according to the voxel size of the microscopy data. The same analysis methods are then applied to these virtual microscopy images.

Samples of the resulting virtual images are shown in Figure 13.6 for different looping probabilities. For comparison, an equally sized section of microscopy data of ESCs stained for H3K4 is shown. Evidently, already from visual inspection it becomes clear that the simulation data does not fit the microscopy images. This is quantitatively confirmed by looking at the averages cluster sizes, which are enormously larger than for the microscopy data.

Why can the data not be adequately described by our computational model? The reasons for this are manifold. First, the Dynamic Loop model in its current implementation is homogeneous, i.e. it assumes the same local average looping density along the whole chromatin fiber. This is not the case in reality, as we have worked out in chapter 7. Second, the experiments used in this chapter do not label the chromatin fiber homogeneously, but select distinct histone modifications belonging to active or inactive genes. Clearly, this distinction is neglected in the approach of creating virtual images as there is no method known how to correctly assume a distribution of active and inactive genes used as a template for the distribution of fluorophores. This results in the whole fiber being labeled similarly, resulting in the large clusters observed. As a third reason for the deviations we have to mention that fluorophores not only bind to histones bound to chromatin but also to molecules being freely in solution. Thus, the microscopy image is much more noisy, a behavior which can be observed in Figure 13.6.

13.4.2 The Potts model

We have found that the Dynamic Loop model cannot be used in a simple manner to explain experimentally observed images. This is not due to the model being inconsistent, rather additional information on the distribution of certain histone modifications would have to be included. Here, we want to follow another approach borrowed from the field of image analysis and segmentation [212]. Consider the images as patterns containing a certain regularity but also a certain randomness. Which physical model can explain the given distribution of intensities?

Parameter estimation in the Potts model

The Potts model and especially the Ising model have been used to segment or reconstruct images in the framework of Markov Random Fields (MRF) in the past [213]. While the general aim of image reconstruction is to estimate the unknown "real" image from a disturbed (e.g. noisy) image, here we want to model basic image patterns. The Potts model has been often applied for noisy images as it smoothes the image by favoring equal nearest neighbor colors [212]. Given the image $\mathbf{x}$ as an array of intensity values, the

Hamiltonian reads

$$\mathcal{H}(\mathbf{x}) = -J \sum_{i \sim j} \delta(x_i - x_j) \tag{13.2}$$

Here, J is the interaction parameter and the sum is over all neighboring pairs of pixels or voxels. The interaction parameter determines the degree of smoothing: Small values of J produce more random image patterns while high values of J favor uniform images.

We want to answer the question of whether the Potts-model Hamiltonian can be used to generate image patterns similar to those observed in experiments. Let us in the following assume that the data $\mathbf{x}$ is given in q discrete intensity values. The probability for the image $\mathbf{x}$ using the parameter J is given in the Potts-model framework by

$$p(\mathbf{x}|J) = \mathcal{Z}(J)^{-1} \exp\left(-\mathcal{H}(\mathbf{x})\right) = \mathcal{Z}(J)^{-1} \exp\left(J \sum_{i \sim j} \delta(x_i - x_j)\right) , \tag{13.3}$$

$\mathcal{Z}$ being the partition function of the system. Let $N(\mathbf{x}) = \sum_{i \sim j} \delta(x_i - x_j)$ be the number of equal nearest neighbors in the image. Then, the probability for image $\mathbf{x}$ can be written as

$$p(\mathbf{x}|J) = Z^{-1} \exp\left(JN(\mathbf{x})\right) \tag{13.4}$$

Importantly, given the specific experimental data $\hat{\mathbf{x}}$, a likelihood estimate for the parameter J can be calculated (cf. Ref. [214]): The likelihood function is given by

$$\mathcal{L}(J|\mathbf{x} = \hat{\mathbf{x}}) = Z^{-1}(J) \exp\left(JN(\hat{\mathbf{x}})\right) \tag{13.5}$$

The maximum likelihood estimate is then calculated as the maximum of the logarithmic derivative,

$$\frac{\partial \log \mathcal{L}(J|\hat{\mathbf{x}})}{\partial J} = 0 \tag{13.6}$$

Thus, we find

$$\begin{aligned} 0 = \frac{\partial \log \mathcal{L}(J|\hat{\mathbf{x}})}{\partial J} &= N(\hat{\mathbf{x}}) - \frac{\partial_J \mathcal{Z}}{\mathcal{Z}} \\ &= N(\hat{\mathbf{x}}) - \frac{1}{\mathcal{Z}} \frac{\partial}{\partial J} \sum_{\{\mathbf{x}\}} \exp\left(-JN(\mathbf{x})\right) \\ &= N(\hat{\mathbf{x}}) - \frac{1}{\mathcal{Z}} \sum_{\{\mathbf{x}\}} N(\mathbf{x}) \exp\left(-JN(\mathbf{x})\right) \\ &= N(\hat{\mathbf{x}}) - \langle N(\mathbf{x}) \rangle \end{aligned} \tag{13.7}$$

We conclude, that the maximum likelihood estimate of J is given by the value of J for which

$$N(\hat{\mathbf{x}}) = \langle N(\mathbf{x}) \rangle \tag{13.8}$$

This quantity can be easily determined both for simulational results as well as for the experimental data.

Determination of Potts model quantities for the data

To determine the interaction parameters in the Potts model, we first discretized the image data. This was done by first determining the maximum intensity $I_{\max}$ for each cell image. Then, $q = 25$ intensity classes were created covering the interval $[0, I_{\max}]$ such that class

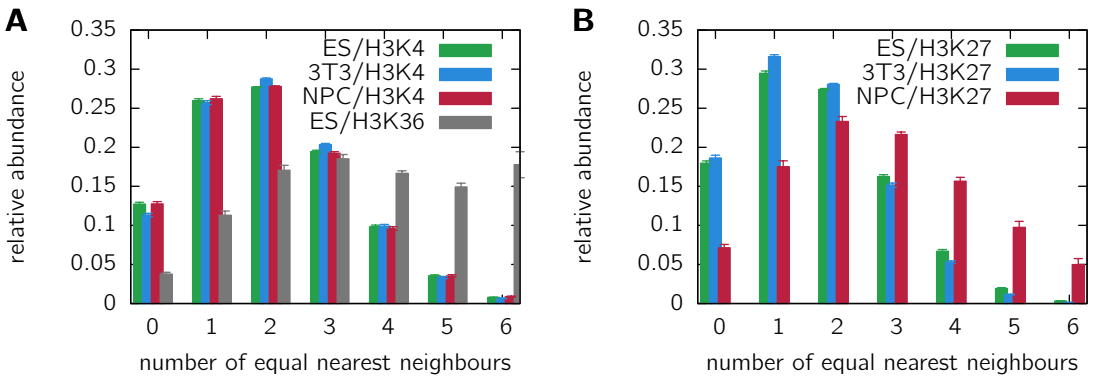


Figure 13.7: Distribution of equal nearest neighbors. Shown is the probability distribution of one voxel in the experimental data having a certain number of nearest neighbors of equal intensity. The intensity values have been discretized using $q = 25$ intensity levels of equal spacing. Panel A shows the results for histone modification H3K4 and H3K36, whereas in panel B the probability distribution for histone modification H3K27 is plotted.

i is the subinterval $[i/25\, I_{\max}, (i+1)/25\, I_{\max}]$ with $i = 0, \ldots, 24$. Then, each voxel was assigned one of these $q = 25$ intensity classes. From these discretized images, the average number of equal nearest neighbors per voxel $\langle \hat{N} \rangle$ were determined. The normalization per image was used to minimize effects of different cell sizes.

We find that the distribution of the number of equal nearest neighbors is similar for all cell types studied in case of histone modification H3K4 (Figure 13.7A), confirming our previous results that active genes are similarly distributed throughout the nucleus independent of cell type. However, differences are found for the distribution of inactive genes. While ESCs and 3T3 cells show a similar distribution of nearest neighbors, NPCs are significantly different and biased towards a higher number of equal nearest neighbors. This result is in agreement with the observed concentrated spots for NPC/H3K27 in Fig. 13.1. The average number of equal nearest neighbors are listed in table 13.4.

Estimation of Potts interaction parameter

To calculate the Potts model interaction parameter estimate J, we have to solve the estimator equation (13.8). This is only possible via computer simulations [214]. Therefore we have set up simulations of the q=25-state Potts model on a three-dimensional lattice of size $L = 64$. The Monte Carlo algorithm used randomly selects one lattice site (with a state q_{old}) and then randomly selects a new Potts state q_{new}. The change in Potts level $q_{\text{old}} \rightarrow q_{\text{new}}$ is accepted dependent on the energy difference ΔE with probability $\min\{1, \exp(-\Delta E)\}$.

The dependence of $\langle \hat{N} \rangle$ on the interaction parameters in the 25-state Potts model is

Table 13.4: Potts model statistics for the experimental data. The table shows the average number of equal nearest neighbors. For that purpose, the experimental data has been discretized to $q = 25$ discrete intensity levels of equal spacing. Uncertainties are given as standard error of the mean.

	ESC	NPC	3T3
H3K4	2.01 ± 0.10	2.00 ± 0.18	2.04 ± 0.14
H3K27	1.71 ± 0.10	2.70 ± 0.21	1.61 ± 0.12
H3K36	3.49 ± 0.17	—	—

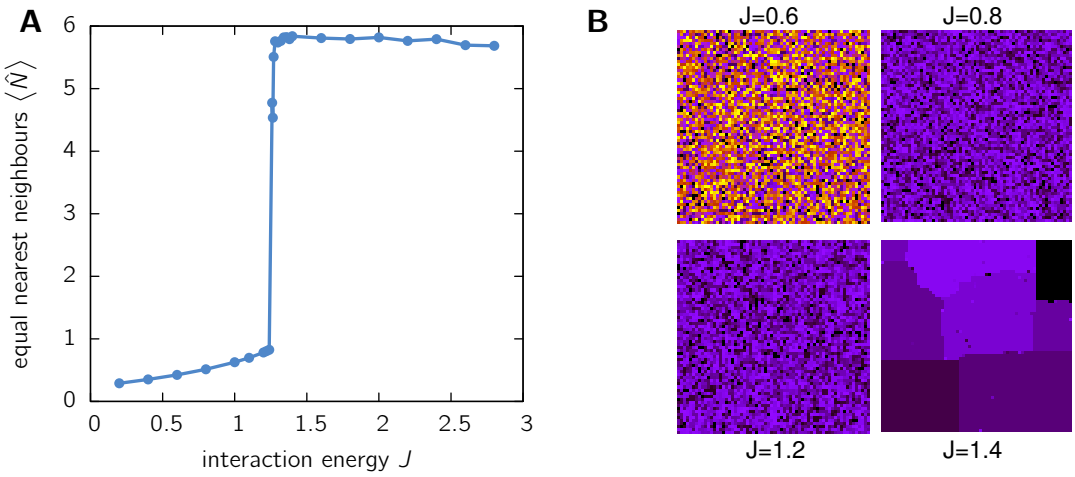


Figure 13.8: Simulation of the q=25-state Potts model. A. This panel displays the dependency of the average number of equal nearest neighbors per voxel $\langle \hat{N} \rangle$ on the interaction energy parameter J. A first-order phase transition can be observed. **B.** Two-dimensional sections of simulated Potts model configurations for different interaction parameters. For low J-values, a random distribution of levels is observed; approaching the phase transition point, clusters of voxels with the same intensity emerge.

shown in Figure 13.8A. It has the typical behavior of a phase transition, changing from a disordered state with small $\langle \hat{N} \rangle$ to a highly ordered state with large values of $\langle \hat{N} \rangle$. Although the simulational results close to the phase transition region are not very reliable due to the divergence of the autocorrelation time using the algorithm described above, it becomes clear that the maximum likelihood estimator for the interaction parameter is close to J-values at the phase transition. Due to the steepness of the change in the order parameter, the resulting J-value is basically the same for all cell types. This however, cannot give a correct description of experimental data, which show clear differences in the distribution of Potts energies (Fig. 13.7) as well as the correlations (Fig. 13.3). Notably, it has already been pointed out in Ref. [214], that the Potts model cannot be satisfactorily used for image segmentation.

13.4.3 The Ising model

In the last section we have seen that it is not possible to create images similar to the experimental data using a Potts model. The reasons for this is the steep first-order transition when having a large number of Potts levels q. What happens if we reduce the number of levels dramatically, say to $q = 2$, i.e. the Ising model? In section 13.3.2 we have evaluated the cluster properties of the experimental images by selecting the 20% brightest pixels. Here, we investigate whether it is possible to simulate the resulting black-white images (Figure 13.4B) based on an Ising Hamiltonian.

The Hamiltonian with which we want to generate patterns similar to the data is given by

$$\mathcal{H}(\mathbf{x}) = -J \sum_{x_i \sim x_j} x_i \, x_j \tag{13.9}$$

Importantly, the image data here is restricted to $x_i \in \{-1, 1\}$ and, given a lattice of size $L \times L \times L$, the number of spins with a value of 1 is restricted to $N_{\text{up}} = 0.2L^3$. To obtain samples of the system, we perform Monte-Carlo simulations. The restriction to a fixed

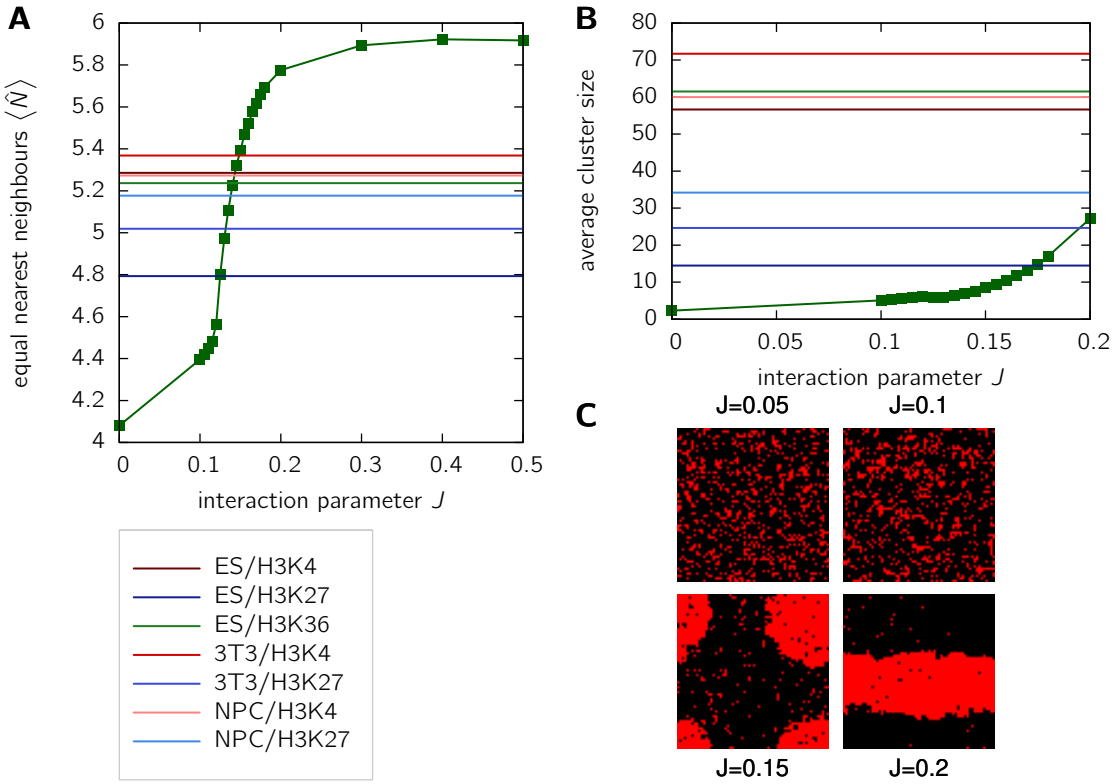


Figure 13.9: Ising model simulations compared to experimental data. A. The average number of equal nearest neighbors $\langle \hat{N} \rangle$ per lattice site in the Ising model on a lattice of size $64 \times 64 \times 64$ dependent on the interaction energy. Results are obtained by Monte Carlo simulations using a spin-flip algorithm. The number of spins, i.e. the total magnetization is kept constant such that 20% of the spins is $+1$. The experimental data average $\langle \hat{N} \rangle$ is indicated by the solid lines. **B.** The average cluster size of the Ising model is compared to experimental data, showing that the cluster sizes is the Ising model do not agree with the data based on the maximum likelihood estimate. **C.** Samples from the Ising model simulations at different interaction energies J.

number of up and down spins can be easily implemented using a spin-flip algorithm: We have chosen the start configuration such that $0.2L^3$ randomly chosen lattice sites have $x_i = 1$, the others $x_i = -1$. Then, in each Monte Carlo move, we randomly choose two lattice sites and propose the new configuration arising from the old one with the spins of these two lattice sites flipped. Again, the change in energy is calculated and the new configuration is accepted with probability $\min\{1, \exp(-\Delta E)\}$.

Simulations have been performed on a lattice of linear dimension $L = 64$ with periodic boundary conditions. Sample conformations are visualized in Figure 13.9C for different interaction energies J. Similar to the case of the Potts model (↷ 13.4.2) we first determine the maximum likelihood estimator for the parameter J. For that purpose, the average number of equal nearest neighbors $\langle \hat{N} \rangle$ is plotted in Figure 13.9A in dependence of J. Again, we find some kind of phase transition from a disordered system to an ordered system. We have not carefully analyzed this transition, nor can we ensure that the system is fully equilibrated at the transition point due to the diverging relaxation time. However, the exact transition point is not our so much of interest for us. More important is that the values of $\hat{N}$ determined from experimental data are lying in the region of the phase transition; they are shown as a solid straight line in Fig. 13.9A.

If one was able to derive different estimates J for different cell types or active/inactive genes, what would be the quality of this estimate, meaning that the obtained simulated patterns resemble the experimental ones? To investigate this issue, we compare the average cluster size given in table 13.1 to the average size of clusters with spin up in the Ising model. We find that the average cluster size is much smaller than predicted by the experimental data up to $J = 1.7$ (Figure 13.9B). Thus, in the range of J-values given by the likelihood estimator, the Ising model has a completely different cluster characteristics than the microscopy images.

These findings clearly show that neither the Ising nor the Potts model with $q > 2$ can be used to adequately describe experimental data on chromatin distribution inside the nucleus.

13.5 Conclusions and future work

In this chapter we have analyzed three-dimensional images from conventional confocal light microscopy to obtain information about differences between stem cells and differentiated cells in chromatin structure of active and inactive regions. One of our major findings is that repressed gene promoters, marked by histone modification H3K27, show different structural patterns in stem cells than in more differentiated NPCs and 3T3 cells. The spatial intensity correlations decay faster for stem cells (Figure 13.3B) than for differentiated cells. Furthermore, the average size of clusters of repressed gene promoters is significantly smaller in stem cells (table 13.1). Contrarily, histone H3K4 marking promoters of transcribed genes, shows similar correlations and less varying cluster sizes in all cell types observed. Importantly, repressed genes are less clustered than active genes (table 13.1). Also, the distribution of active and inactive genes is found to be dependend on cellular position, the degree of position dependency varying from stem cells to more differentiated cells. These findings indicate that the distribution of chromatin plays a dominant role in epigenetic regulation.

In general, our results offer an ambiguous picture. Concerning the intensity distribution (Figure 13.2), 3T3 cells display the largest deviations from the other cell types. However, this measure is highly dependent both on the microscope settings (which might have changed) as well as on the size of the cell and the brightness of the fluorophores. The distribution of equal nearest neighbors for repressed genes (Figure 13.7B) is different for NPCs, while stem cells and 3T3 cells show a similar behavior. These findings show that it is absolutely necessary to look at different measures for chromatin structure before drawing conclusions. It seems that each cell type indeed has its own characteristic behavior concerning chromatin organization, the detailed investigation of this requires more experimental input. Currently, experiments are conducted at the Jackson Laboratory relating the structure of active and inactive genes to each other using microscopes with a higher resolution.

From the modeling point of view, we could not present a satisfactory model which reconstructs the observed image patterns. The Ising and the Potts model cannot describe differences between cell types due to the maximum likelihood estimate of the interaction constant J being in the region of the phase transition (Figure 13.8). Future work will show whether more complicated models might explain the observed patterns.

Even more ambitious is to model the data using a polymer model. Although the Dynamic Loop model offers a versatile basis for explaining chromatin folding, various experimental information will have to be integrated. First of all, the Dynamic Loop

model has to be implemented heterogeneously, assuming different looping probabilities for ridges and anti-ridges. This can be easily done and has already been successful applied to the Random Loop model (↷ chapter 7). Secondly, The distribution of virtual fluorophores along the fibers has to follow the gene activity patterns found in the Human Transcriptome Map [13]. A recently published database could allow for a more direct approach: The Methylome data [215], mapping the occurrence of certain histone modifications along the whole genome, could be used to distribute the fluorophores. Most importantly, the polymer models presented in this thesis are flexible enough to incorporate such information.

Chapter 14

Summary & Outlook

14.1 A short summary of the results

The genomic content, coding the building plans of life, is stored in a highly complex hierarchical manner inside the cell nucleus of higher eukaryotes. The packing of about 2 meters of double-stranded DNA into a nucleus with a diameter of about 10 μm is accomplished on several scales. At the first stage of compaction, the DNA double helix is wound around histone octamers with whom it forms the chromatin fiber, a beads-on-a-string type of filament having a diameter of about 10 nm. A lot of discussion has been going on about the question whether this fiber in turn condenses to a filament with a diameter of 30 nm, the nucleosomes being arranged either in a zig-zag or a solenoidal manner. Even less is known about the higher-order folding of chromatin inside the interphase nucleus, i.e. the folding of complete chromosomes as well as the mutual arrangement of different chromosomes.

Limitations of experimental techniques render it hard to study the detailed folding pathways of chromatin inside living cells. Conventional light microscopy is limited to a resolution of about 200 nm, thus it is not possible to track the path of the fiber inside the nucleus. Localization light microscopy offers a promising approach to study nuclear nanostructure with a resolution on the scale of a single histone; However, up to now, the folding path of the fiber cannot be reconstructed from the data. Indirect approaches, on the other hand, from fluorescence in situ hybridization (FISH) to chromosome conformation capture (3C) techniques have allowed biologists to gather important information about folding mechanisms. There is now an ever growing body of evidence that chromatin folding is tightly related to genome function: Amongst others, the three-dimensional folding of chromosomal regions has been shown to depend on local transcriptional activity. In contrast to expectations from the physics of linear polymers, chromosomes do not intermingle, rather they are separated into distinct territories; the same behaviour is found for Mb-sized regions of single chromosomes. Importantly, intra- and interchromosomal contacts or chromatin loops on a broad range of scales have been shown to play an important role in transcriptional regulation.

The aim of this thesis has been to shed light onto the detailed folding principles and to develop a predictive framework for the connection between genome folding and function. Polymer models are useful to uncover the folding mechansims of chromatin, making

testable predictions on structural properties. We started by comparing experimental data from FISH measurements to classical polymer models: the random walk (RW) model, the self-avoiding walk (SAW) model and the globular state (GS) model (↷ chapter 4). All of them have been proposed to explain the folding of chromatin in a certain range of genomic separations (↷ 2.2) in earlier times. Using both 3D and 2D FISH measurements we have shown using information about the cell-to-cell variation of distance measurements that none of these models is able to explain the folding of chromatin on the short (100 kb – 10 Mb) and on the large scale (10 Mb – 80 Mb). Although chromatin models with loops have already been proposed, amongst others the Random-Walk/Giant-Loop model assuming fixed-sized loops of about 3 Mb and the Multi-Loop-Subcompartment model assuming rosette-like structures of 120kb-sized loops, these models are inconsistent with experimental data from FISH experiments clearly showing that chromosomes fold into a confined sub-space of the nucleus with a huge cell-to-cell variation.

Based on a variety of experimental results we have proposed a new polymer model, the Random Loop (RL) model. At the heart of this model is the assumption that chromatin looping is the major connection between genome folding and function. A lot of experimental evidence exists that this conjecture is indeed valid (↷ 2.3), the results and predictions of this thesis being in support of our claim. This RL model is founded on two important assumptions which have been neglected in other polymer models: (i) loops exist on all scales ranging from a few kb to several Mb (based on evidence from 4C and 5C experiments) and (ii) loops are formed dynamically and change during the lifetime of the cell.

In a first approach, these assumptions were implemented considering the absence of excluded volume interactions (↷ chapter 5). The loops were incorporated into the model by establishing harmonic interactions between the loop attachment points. Importantly, the pairs of monomers forming loop attachment points were selected randomly with a certain looping probability $\mathcal{P}$. The fiber was considered to be homogeneous, each pair of monomers forming a loop with the same probability. In fact, this model allowed for an analytic calculation of the thermal average, while the disorder average could be approximated by numerical methods. This simple model displays the observed leveling-off in the mean square distance between two genomic segments at separations larger than 10 Mb as well as the huge cell-to-cell variation. The cell-to-cell variation is assessed by evaluating the higher-order moment ratios of the distance distribution [eq. (4.3)], being a constant for the RW, SAW and GS model. Importantly, fluctuations exceeding the ones of the random walk model are observed due to the average over the disorder of loop-attachment points.

As the negligence of excluded volume might have a major impact on the observed confined folding, we have lifted this limitation in chapter 6. As analytical solutions of the Random Loop model become intractable when taking excluded volume into account, the model was implemented using Molecular Dynamics simulations with a combined FENE + WCA potential for the bonded interactions. Our results suggest that these excluded volume interactions do not qualitatively change the observed folding into a confined space.

Clearly, the chromatin fiber is not homogeneous as assumed so far. The Human Transcriptome Map reveals gene-rich regions spanning several Mb being transcriptional active (ridges) as well as gene-poor and transcriptional inactive regions (anti-ridges). While anti-ridges are rather compact and spherical, ridges have a more open structure. This heterogeneity has been incorporated into the RL model by assuming different local looping probabilities for ridge and anti-ridge regions (↷ chapter 7). Although still highly simplifying, a direct connection between genome folding (formation of loops) and gene function

(transcriptional activity) is established. The results of this heterogeneous RL model are in good agreement with experimental data on human chromosome 11 both on the short scale below 10 Mb as well as on the scale of the entire chromosome, thus offering a unified description of chromatin folding comprising several length scales.

The tight connection between chromatin loops and transcriptional regulation raised the question of which benefits exist for nature to force the formation of loops between genes and regulatory elements separated by several Mb. To investigate the effect of loop formation on the conformational properties of polymers we studied a system of two ring polymers in chapter 8. The non-catenated topology where two rings are bonded to each other resembles the behaviour of chromatin loops assembling in transcription factories. Indeed, we found that two rings in proximity display a much stronger tendency to segregate than corresponding linear-linear systems.

While a lot of studies indicated a strong effect of the non-catenation topology in a melt, resulting for example in a compact state, the forces acting between two ring polymer or loops have not been evaluated. To further deepen the understanding of the topological and entropic interactions exerted by rings on each other, we have quantitatively assessed the strength of these potentials (↷ chapter 9). We have evaluated the potential of mean force between the centers of mass of two ring polymers and found that the repulsion between them is threefold stronger than between linear chains. Forcing the rings to be non-catenated results in a further decrease of the accessible conformations at short separations of about 50%. These findings indicate that loop formation in chromatin not only acts as a tool for transcriptional regulation, but also represents a key mechanism to maintain a segregated and compartmentalized state of chomatin organization.

What mechanisms are necessary to make two chromatin segments become co-located? In fact, it has long been an open question how the cell nucleus manages to bring distant elements together. Using Monte Carlo simulations we have extended the Random Loop model to incorporate the dynamics of loop formation. In the framework of this Dynamic Loop model, loops are formed solely on the basis of diffusional motion without requiring active processes. Whenever two chromatin segments are in proximity, there is a certain probability for them to form a loop, mimicking e.g. the presence of binding proteins and the solvent. We have derived testable quantitative predictions for a variety of observable parameters. Amongst others, the Dynamic Loop model correctly predicts the folding of chromosomes into a confined sub-space of the nucleus as well as a huge cell-to-cell variation in the distance distribution. The analysis of the loop-size distribution reveals that loops in the size range of several Mb easily form on the basis of diffusional motion, the probability for their formation being more than two orders of magnitude higher than the corresponding contact probabilities for linear polymers. Consistent with experimental measurements in yeast, such a model predicts sub-diffusional motion of genes, the detailed exponent being dependent on the looping probability. Most importantly, Monte Carlo simulations of several chromosomes under biological conditions reveal that the chromosomes segregate, the overlap between chromosome territories decreasing strongly with increasing looping probability. Thus, this Dynamic Loop model provides a consistent explanatory framework for the relationship bewteen genome folding and function being in agreement with a vast amount of experimental data.

We have shown that two ring polymers exert strong repulsive forces when being brought close together. In the framework of the Dynamic Loop model, we have evaluated the forces acting between chromosomes (↷ chapter 11). This forces turned out to be much stronger than between an isolated system of two rings, confirming the strong influence of looping

and the induced compaction on the segregated state of chromosomes.

A model needs experimental data to compare with. Therefore we have analyzed images from high-resolution localization light microscopy labeling histone H2B with green fluorescent proteins (↷ chapter 12). The images at a resolution of single histones have allowed us to apply analysis methods borrowed from soft condensed matter theory. We find hallmarks of distinct structures in the images, the density fluctuations indicating more disorder to be present than expected in a thermodynamic equilibrium system. The large fluctuations have shown to be in good agreement with results of the Dynamic Loop model. Using more specialized markers labeling active and inactive genes in combination with conventional light microscopy, we found distinct structures in the distribution of repressed genes in comparison to differentiated cells (↷ chapter 13). For comparison to polymer models, however, more experimental input is needed.

14.2 Synthesis

Eukaryotic organisms are highly compartmentalized. Each cell acts as a functional unit on its own, carrying both the information and ability for reproduction by converting nutrients from the environment to energy. Such a hierarchical compartmentalization is continued by the cell nucleus, separating the chromosomes and the chemical equilibrium inside from the cellular environment by a lipid bilayer. In fact, compartmentalization is vital to maintain certain chemical processes by providing a high concentration of certain molecules. Chromosomes itself are highly compacted objects carrying the script of life. In order to be able to transcribe, replicate and store a vast amount of information, a concerted three-dimensional folding seems necessary, the folding of the chromatin fiber being strongly connected to genome function. Our results suggest that the connection between folding and function is imposed by the formation of chromatin loops between genes and regulatory sequences. These loops, possibly but not necessary assembled in transcription factories, are dynamic, meaning that loops form and unravel dependent on the needs of the current stage of development. Such loops are able to form by diffusional motion together with a DNA-DNA or DNA-protein binding affinity, requiring no additional energy from the nucleus.

However, besides its role in transcriptional regulation, chromation loops serve another important purpose. They allow for a purely entropic compartmentalization of chromosomes, lowering the intermingling between them. This compartmentalization might be important to allow intrachromosomal contacts to find each other, not being hindered by entanglement effects between chromosomes. Furthermore, segregation of chromosomes is vital for achieving a fast separation into the daughter cells during cell division. Thus, chromatin looping provides an energetically highly effective way of managing transcripitonal regulation as well as maintaining a certain order and segregation inside the cell nucleus.

14.3 Outlook

In this thesis, a versatile model for the relationship between genome folding and function has been established, being able to quantitatively explain important experimental results such as the folding into a confined sub-space of the nucleus as well as observed decay exponents in the intrachromatin contact probabilities. Due to important experimental knowledge being lacking, the model so far does not take into account any locus-specific interactions, assuming that loops can form between any two parts of the chromosome. Surely,

this approach is highly simplifying as chromatin-chromatin contacts are often established between specific genes and their regulatory elements. Among others, the influence of the solvent on the chromatin properties in terms of an effective interaction will have to be incorporated and the question of whether an equilibrium approach is sufficient to describe chromatin folding has to be answered.

Future experiments are necessary in order to be able to further specify the looping interactions of the Random Loop or Dynamic Loop model. Importantly, such information can be easily incorporated into the model. Several promising experimental techniques exist which can shed light on the detailed relationship between genome folding and function, amongst others: (i) The Human Transcriptome Map specifying transcriptional activity and gene density for each region of the genome, (ii) 4C/5C experiments revealing the abundance of contacts between two specific loci and (iii) the distribution of binding sites for loop-related proteins like CTCF.

(i) The Human Transcriptome Map displays gene expression profiles for any chromosomal region. It has been pointed out in this thesis that genes are not distributed randomly along the one-dimensional genome, rather gene-dense clusters with a high transcriptional activity are found as well as pronounced gene-poor regions. In chapter 7 we have related transcriptional activity to looping by assuming different looping probabilities for gene-rich and gene-poor regions using a very rough classification of the genome into ridges and anti-ridges. The detailed connection between local transcriptional activity and looping, however, remains an open issue of research, requiring elaborate experimental input.

(ii) Although 4C experiments have established an interaction map between one specific bait region and any other chromosomal locus (↷ 2.4.2), these data can hardly be used to model chromatin. First of all, these data only provide information between one specific locus and the rest of the genome, while modeling a chromosome requires information on interactions between any two loci. This short-coming is overcome in the 5C technique, measuring a two-dimensional map of interactions between any two loci on the genome. If it were possible to distinguish between functional and random contacts in these experiments, the data could be incorporated into our model. The functional contact probabilities from 5C could be incorporated as local looping probabilities $\mathcal{P}(i,j)$ into the heterogeneous RL model (↷ chapter 7). However, nC techniques do not provide information about topological constraints in chromatin folding. The observed interaction frequencies are an average over an ensemble of approximately one million cells, thus, correlations between different loops, i.e. the question which two loops can exist at the same time, cannot be answered with this method.

(iii) Several proteins have been shown to be involved in the formation of loops. For modeling purposes, the most promising is CTCF: About 15 000 binding sites for the evolutionary highly conserved protein CTCF along the genome indicate its vital importance. It has been shown that CTCF is required for looping in the β-globin locus. However, the detailed connection bewteen CTCF binding sites and looping remains to be established.

How can these experimental findings be incorporated into a more detailed locus-specific model of chromatin folding? To me, the most promision approach seems to be to establish a connection between 5C and the Dynamic Loop model. While in chapter 10 it has been shown that probabilistic unspecific looping leads to the correct contact probabilities averaged over all loci, a more specific implementation of loops might be possible. In principle, the interaction probabilities given by 5C experiments can be incorporated into the model by weighting the binding affinity between two loci according to the experimental data, always taking into account that 5C data contains random contacts as well as functional

contacts. Such a model could make very locus-specific predictions. Importantly, as 5C experiments do not yield correlations between loops related to topological constraints, these could be predicted by the model, establishing a map of which interaction sites might come together at the same time to initiate or repress transcription of a specific gene.

The abundant existence of CTCF binding sites reveals another promising approach for modeling specific loops. Looping could be restricted in the model by only allowing loops between these binding sites according to the distribution found in the cell nucleus. However, lots of experimental data is lacking for such an approach: Which binding sites are occupied by CTCF in which cell type? Which binding sites are actually involved in loop formation? Is CTCF the only mediator for loops? These questions need to be answered before building a model based on binding sites.

A general question emerging from the problem of large loop formation is how the cell can rely on the existence of a specific loop spanning several Mb? Clearly, the diffusional formation of loops greatly enhances the probability of such an event (↷ 10.3.2), however, one might argue that the probability is still to small for this loop to be required for a vital regulatory task. Although a definite answer to this question is far from reach, two speculative solutions have come to our mind while discussion this topic: First, loop formation might be less specific than previously thought. Perhaps, regulatory tasks can be accomplished in case of the regulatory element being in proximity rather than directly bound to the transcription complex. The interaction between regulatory element and gene might be mediated by certain proteins being able to diffuse the short remaining distance between the target sites. Second, cells in higher eukaryotes are assembled in a huge bunch of equally differentiated cells. Why should every single cell execute the same program? Possibly, it suffices if a vital loop is formed in a small number of cells for the whole network of cells to function. This picture would provide an easy explanation for the robustness of organisms being quite insensible to errors in single cells.

A lot of questions will have to be answered to further deepen our understanding on the folding of interphase chromosomes in higher eukaryotes. The dynamic formation of probabilistic loops as presented in this thesis offers a general framework, which can be subsequently filled with information from several experimental techniques by adjusting the local looping probabilities. Such a detailed model might offer deeper insights into the functioning of specific genes and regulatory complexes, especially with respect to differences between healthy and malignant cells. From such a model, specific predictions could be derived about topological constraints and steric hindrances for the formation of loops required for gene regulation, e.g. by allowing or disallowing certain loops to build up. However, it will be a long way to make such detailed predictions due to the amazing complexity of living organisms.

Publications

Large parts of this thesis have already been published or are currently submitted. The following list contains the journal references, or the journal to which the paper has been submitted. Papers in preparation are also listed. (Information as of January 29th, 2010)

- **Bohn, M.**, Heermann, D. W. & van Driel, R. Random loop model for long polymers. *Physical Review E* **(2007)**, *76*, 051805

  DOI: 10.1103/PhysRevE.76.051805 preprint: arXiv:0705.1470v3 [cond-mat.soft]

- Langerak, J. M.[1], **Bohn, M.**[1], de Leeuw, W., Giromus, O., Manders, E. M. M., Verschure, P. J., Indemans, M. H. G., Gierman, H. J., Heermann, D. W., van Driel, R. & Goetze, S. Spatially confined folding of chromatin in the interphase nucleus. *Proc. Natl. Acad. Sci. U.S.A.* **(2009)**, *106*, 3812-3817

  DOI: 10.1073/pnas.0809501106

- **Bohn, M.** & Heermann, D. W. Conformational properties of compact polymers. *Journal of Chemical Physics* **(2009)**, *130*, 174901+

  DOI: 10.1063/1.3126651 preprint: arXiv:0905.0798v1 [cond-mat.soft]

- **Bohn, M.** & Heermann, D.W. Topological interactions between ring polymers: Implications for chromatin loops. *Journal of Chemical Physics* **(2010)**, *132 (4)*, 044904

  DOI: 10.1063/1.3302812 preprint: arXiv:1001.4246 [cond-mat.soft]

- **Bohn, M.** & Heermann, D.W. Diffusion-driven looping provides a consistent framework for chromatin organization. *PLoS Comp. Biol. (submitted, under review)*

- **Bohn, M.**, Heermann, D. W., Lourenço, O. & Cordeiro, C. E. On the influence of topological catenation and bonding constraints on ring polymers. *Macromolecules (submitted)*

- **Bohn, M.**[1], Diesinger, P.[1] *et al.* Localization microscopy reveals expression-dependent parameters of chromatin nanostructure. *Biophys. J. (submitted)*

- Heermann, D. W., **Bohn, M.**, Diesinger, P. The Relation between the Gene Network and the Physical Structure of Chromosomes. *Proceedings of the 4th International Conference on High Performance Scientific Computing: Modeling, Simulation and Optimization of Complex Processes. (submitted)* **(2010)**

- **Bohn, M.** & Heermann, D.W. Repulsive Forces between Chromosomes. *(in preparation)*

- Tolhuis B., Kerkhoven R., Pagie L., **Bohn M.**, Teunissen H., Nieuwland M., Simonis M., de Laat W., Heermann D.W., v. Lohuizen M.& v. Steensel B. Interactions among Polycomb Domains are Guided by Chromosome Architecture. *Cell (submitted)*

[1] Shared first authorship due to equal contribution

Ruprecht-Karls
Universität Heidelberg
Institute of Theoretical Physics

Diffusion-driven looping provides a consistent framework for chromatin organization

M. Bohn & D. W. Heermann

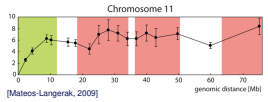
Abstract

Experimental results suggest that higher-order chromatin folding is tightly connected to genome function. The building blocks for a consistent overall description of chromatin folding have to be gathered from several experimental techniques. FISH measurements reveal a confined folding of chromosomes into a sub-space of the nucleus, the mean square distance between two markers becoming independent of genomic separation but at the same time revealing a huge cell-to-cell variation. 4C data provides evidence that loops exist on all scales and it has been shown that these take part in transcriptional regulation. On a large scale, chromosome form distinct territories. Our Dynamic Loop model presented here provides a unified description of chromatin folding on the basis of dynamic loop formation without invoking active transport mechanisms. In fact, a huge amount of experimental evidence from FISH distance measurements, 4C/5C data, diffusion measurements up to the formation of chromosome territories and its shape can be explained consistently in the framework of one model.

The Dynamic Loop model

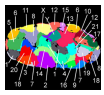
Experimental evidence

- Chromatin folds into a confined sub-space of the nucleus



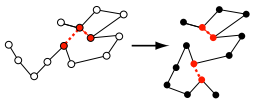
[Mateos-Langerak, 2009]

- Chromatin forms loops on all scales up to several mega basepairs (Mb) [Simonis, 2006]
- Loops play a role in gene regulation
- A huge cell-to-cell variation is observed
- Chromosomes do not intermingle (formation of territories) [Bolzer, 2005]



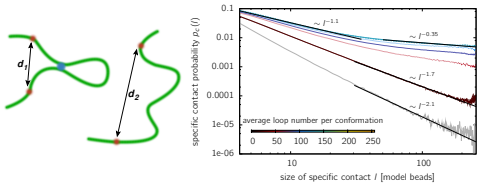
The model

- Chromosomes are modelled as coarse-grained polymers
- Loop formation is achieved on a purely diffusional basis
  - Whenever two segments co-localize by diffusional motion, they form a loop with a certain probability p
  - Loops only persist over a certain lifetime τ, afterwards they dissolve again
- Important: No long-range interactions



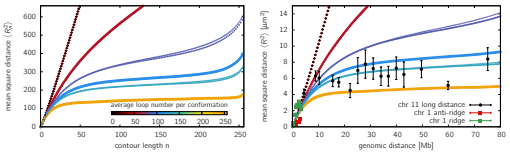
Hierarchical formation of large loops

- How do regulatory elements find their target?
- Co-localization probability for linear polymers $p_c(l) \sim l^{-1.5} \Rightarrow p_c(10\,Mb) \sim 10^{-4}$
- Our model predicts: $p_c(l) \sim l^{-1.1}$ on intermediate scales and $p_c(l) \sim l^{-0.35}$ on large scales
- Experimental finding: $p_c(l) \sim l^{-1.08}$ on intermediate scales (500 kb - 7 Mb) [Lieberman-Aiden, 2009]



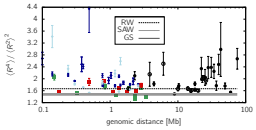
Folding into a confined space

- Chromosomal regions fold into a confined sub-space of the nucleus [Mateos-Langerak, 2009]
- This shows up in the mean square distance $\langle R^2 \rangle$ between two markers in relation to genomic separation g: $\langle R^2 \rangle \sim \mathcal{O}(1)$ above approx. 10 Mb.
- The Dynamic Loop model predicts such a confined folding dependent on the average number of loops

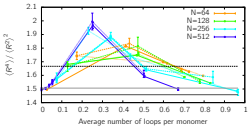


Cell-to-cell variation

- FISH distance measurements between two markers display a large cell-to-cell variation
- These fluctuations are much larger than expected for a random walk polymer model
- In our model, such large fluctuations arise from averaging over different loop configurations
- Fluctuations are evaluated by the dimensionless ratio $\langle R^4 \rangle / \langle R^2 \rangle^2$



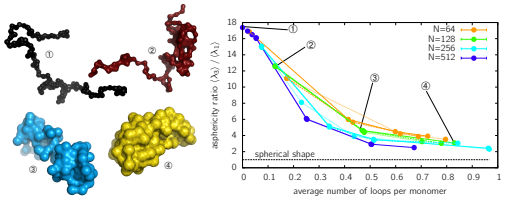
The following experimental data is shown: *Human fibroblasts Chr1*: ■ anti-ridge region, ■ ridge region, ◦ long distance measurements; *Human fibroblast Chr 11*: • long distance measurements; *Murine Igh locus*: ♦ pre-pro-B cells, ♦ pro-B cells. [Mateos-Langerak 2009, Jhunjhunwala 2008]



Simulational results for different chain lengths and looping probabilities. The dimensionless ratio $\langle R^4 \rangle / \langle R^2 \rangle^2$ is plotted against the average number of loops per monomer. Fluctuations are larger than the RW value due to looping disorder.

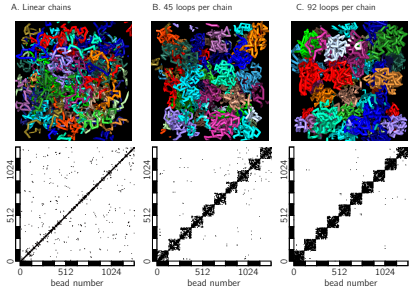
Aspherical shape of chromosomes

- Chromosome territories reveal an aspherical shape [Khalil et al., Chrom. Res. 2007]
- We use the ratio $\langle \lambda_3 \rangle / \langle \lambda_1 \rangle$ of the gyration tensors eigenvalues ($\lambda_3 \geq \lambda_2 \geq \lambda_1$) as a measure of shape
- The model polymers reveal an aspherical shape, being stronger elongated by a factor of $\sqrt{2}$ to $\sqrt{5}$ in one direction in the parameter range where a confined folding is observed.



The formation of chromosome territories

- Chromosomes have been shown to segregate into distinct territories [Cremer & Cremer, 2001]
- Linear polymers in a dense system, however, strongly intermingle
- Dynamic Loop formation induces an entropy-driven formation of chromosome territories, the number of inter-chromosomal contacts strongly decreasing



The figures show example conformations for linear chains (strong intermingling) and looping polymers (less intermingling). In the contact maps at the bottom, any contact between two beads is marked. Results show that the number of interchromosomal contacts strongly decreases upon loop formation.

Conclusions

- The Dynamic Random Loop provides a consistent description of chromatin folding on the basis of dynamic loop formation
- No active transport mechanisms are necessary, large loops are formed on the basis of diffusional motion
- The model correctly predicts a variety of experimental observations (distance measurements, cell-to-cell variation, formation of large loops, shape of territories, sub-diffusive motion of genes)
- Loop formation entropically drives chromosomes into a segregated state (chromosome territories)

Publications

1. M. Bohn & D. W. Heermann, Diffusion-driven looping provides a consistent framework for chromatin organization. **2009**, *submitted to PLoS Comp. Biol.*
2. J. M. Langerak, M. Bohn *et al.*. Spatially confined folding of chromatin in the interphase nucleus. *Proc. Natl. Acad. Sci. USA*, **2009**, *106 (10)*, 3812-3817.
3. M. Bohn & D. W. Heermann, Conformational properties of compact polymers. *Journal of Chemical Physics*, **2009**, *130 (17)*, 174901+.
4. M. Bohn, D. W. Heermann & R. van Driel, Random loop model for long polymers. *Physical Review E*, **2007**, *76*, 051805

Manfred Bohn • Institute for Theoretical Physics • University of Heidelberg
Philosophenweg 19 • D-69120 Heidelberg • Germany
e-mail: bohn@tphys.uni-heidelberg.de

A poster presented at the conference
"Jülich Soft Matter Days 2009" in Bonn.

Conference Contributions

The work described in this thesis has been presented at the following conferences and workshops:

- "Workshop on computer simulations of soft matter and biosystems", March 14 – 16, 2007, Heidelberg [poster]
- *StatPhys23*, International Conference on Statistical Physics, July 9 – 13, 2007, Genua [poster]
- Δ-Meeting, December 14 – 16, 2007, Heidelberg [talk]
- Joint Meeting of the Biophysical Society 52nd Annual Meeting & 16th International Biophysics Congress (IUPAB), February 2 – 6, 2008, Long Beach, California [poster]
- DPG-Frühjahrstagung, Februrary 25 – 29, 2008, Berlin [talk]
- HICCup – Heidelberg Initiative for Chromatin Computing, September 26th, 2008, Heidelberg [talk]
- Workshop "The physics of genome folding and function", Lorentz Center Leiden, October 20 – 23, 2008
- Jülich Soft Matter Days, November 11 – 14, 2008, Gustav-Stresemann-Institute, Bonn [poster]
- Workshop "Biophysics of Chromatin", February 4 – 6, 2009, Villa Bosch, Heidelberg [poster]
- Spring School on Multiscale Methods and Modelling in Biophysics and Systems Biology, May 18 – 29, 2009, Shanghai, China [talk + poster]
- Annual Colloquium of the Heidelberg Graduate School of Mathematical and Computational Methods for the Sciences, October 10th, 2009, Heidelberg [talk]
- Jülich Soft Matter Days, November 10 – 13, 2009, Gustav-Stresemann-Institute, Bonn [poster]

Ruprecht-Karls
Universität Heidelberg
Institute of Theoretical Physics

Higher-order folding of chromatin and the Random Loop Model

M. Bohn, D. W. Heermann & R. van Driel

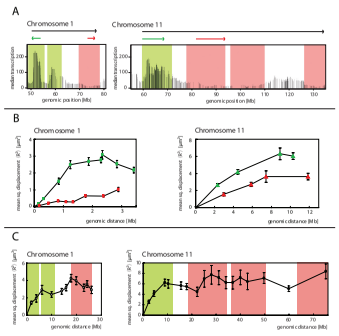
Abstract

Genome function in higher eukaryotes involves major changes in the folding of the chromatin fiber. Nevertheless, our understanding of chromatin folding is still remarkably limited. Polymer models have been used to describe chromatin folding; however none of the proposed models gives a satisfactory description of experimental data. In particularly, they ignore that each chromosome occupies a confined space, i.e. the chromosome territory. Here, we present a polymer model that is able to describe key properties of chromatin over length scales ranging from 0.5 to 75 Mb. This random loop (RL) model assumes a self avoiding random walk folding of the polymer backbone and defines a probability P for two monomers to interact, creating loops of a wide size range. It correctly predicts experimental observations on the large scale, where $\langle R^2 \rangle \sim \mathcal{O}(1)$ as well as on the short scale, where $\langle R^2 \rangle \sim N^{2\nu}$.

Key Question: What determines the large-scale structure of chromatin folding?

- About two meters of double-stranded DNA have to be densly packed into a nucleus of about 10 μm diameter.
- Simple polymer models predict a scaling like $\langle R^2 \rangle \sim N^{2\nu}$. This cannot apply to chromatin due to the confined space.

Recent experimental results



- Short genomic separations (up to $\sim$ 5 Mb)

$$\langle R^2 \rangle \sim N^{2\nu} \quad \text{with some } \nu$$

- Large genomic separations (above $\sim$ 10 Mb)

$$\langle R^2 \rangle \sim \mathcal{O}(1)$$

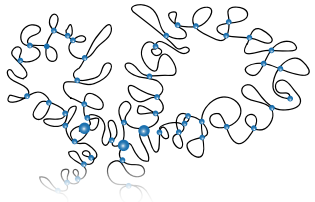
Idea: Large-scale structure determined by chromatin loops

- There is lots of experimental evidence that chromatin forms loops, e.g. promoter-enhancer complexes, transcription factories, β-globin locus...)
- These loops are functionally connected to transcriptional activity.
- Are loops the major driving force for the large-scale chromatin structure?

The Random Loop Model

Basic assumptions

- The backbone of the model is simulated by a FENE+WCA potential.
- Two parts of the chromatin fiber form loops with a certain probability $\mathcal{P}$
- Loops are not static but can change in the course of time; their size and position are chosen from a broad range
- Averaging is done over the thermal ensemble as well as over the ensemble of loops



The mathematical model (without excluded volume)

The backbone of our chain is Gaussian, while loops are modelled by a harmonic potential between non-adjacent beads. The potential for the Random Loop model looks as follows

$$U = \frac{1}{2} \sum_{i<j}^{N} \kappa_{ij} \, || \, \mathbf{x}_i - \mathbf{x}_j \, ||^2$$

Two different kinds of interactions

- The Gaussian chain is maintained by setting $\kappa_{ij} = \kappa$ for $|i-j| = 1$
- The κ_{ij} with $|i-j| > 1$ are Bernoulli distributed random variables with probability $\mathcal{P}$

$$\kappa_{ij} = \begin{cases} \kappa & \text{with probability } \mathcal{P} \\ 0 & \text{with probability } 1-\mathcal{P} \end{cases}$$

This corresponds to loops which are randomly distributed along the chain.

The mean square displacement between two beads

We introduce the connectivity matrix

$$\sigma^{-1} = K = \begin{pmatrix} \sum_{\substack{j=0 \\ j\neq 1}}^{N} \kappa_{1j} & -\kappa_{12} & \dots & -\kappa_{1N} \\ -\kappa_{21} & \sum_{\substack{j=0 \\ j\neq 2}}^{N} \kappa_{2j} & \dots & -\kappa_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ -\kappa_{N1} & -\kappa_{N2} & \dots & \sum_{\substack{j=0 \\ j\neq N}}^{N} \kappa_{Nj} \end{pmatrix}$$

and we can write the *thermal average* between bead I and J for one fixed set of interactions κ_{ij} as

$$\langle r_{IJ}^2 \rangle_{\text{thermal}} = 3(\sigma_{JJ} + \sigma_{II} - 2\sigma_{IJ})$$

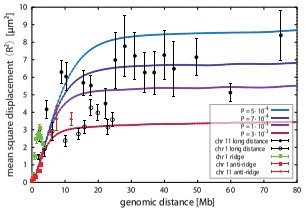
We now have to perform the average over the disorder of loops (i.e. our random variables κ_{ij}). This can be done numerically as an average over the random matrices K.

$$\langle r_{IJ}^2 \rangle = \langle \langle r_{IJ}^2 \rangle_{\text{thermal}} \rangle_{\text{loops}} = 3 \, \langle (\sigma_{JJ} + \sigma_{II} - 2\sigma_{IJ}) \rangle_{\text{loops}}$$

Results

1 The Random Loop (RL) Model leads to an independency of the mean square displacement of genomic distance without assuming a confined geometry

This holds for the model without excluded volume [1] as well as for simulations of the model with excluded volume [3].



2 Already a very small number of large loops leads to leveling-off

The chain length used in simulations is $N = 300$. About 40 loops on all scales are sufficient for observing a leveling-off (corresponding to $P = 5 \times 10^{-4}$).

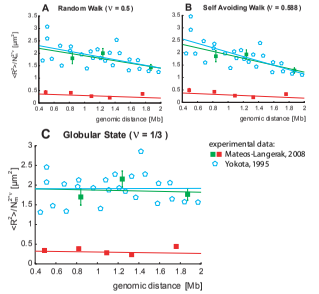
3 Loops on both small and large scales are necessary to obtain confinement

If we restrict the loop-sizes to loops that are smaller than a certain size or to loops that are larger than a certain size, we do not find a leveling-off. Allowing only small loops leads to RW behaviour with an effective bond length.

Folding motifs below the 5 Mb domain

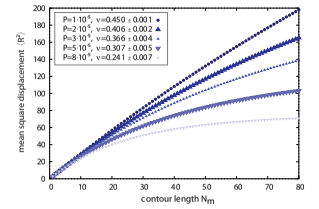
- We compare the data to classical polymer models where

$$\langle R^2 \rangle \sim N^{2\nu}$$



- The GS model fits the data best. *But:* dependent on the locus studied, one finds other exponents (e.g. $\nu \sim 0.1 - 0.2$ [Jhunjhunwala, 2008]).

4 RL model: Different exponents ν represent local differences in the looping probability



Contact & Publications

Manfred Bohn
Institute of Theoretical Physics
University of Heidelberg
Philosophenweg 19
D-69120 Heidelberg
bohn@tphys.uni-heidelberg.de

[1] Manfred Bohn, Dieter W. Heermann, and Roel van Driel. Random loop model for long polymers. *Physical Review E*, 76(5):051805, 2007.
[2] Dieter W. Heermann and Manfred Bohn. Percolation in a class of band structured random matrices. *arXiv:0705.1241v1 [cond-mat.stat-mech]*, 2007.
[3] Julio Mateos-Langerak, Manfred Bohn, Wim de Leeuw, Osdilly Giromus, Erik M. M. Manders, Pernette J. Verschure, Mireille H.G. Indemans, Hinco J. Gierman, Dieter W. Heermann, Roel van Driel and Sandra Goetze. Spatially confined folding of chromatin in the interphase nucleus. 2008. *(submitted to PNAS)*

This poster has been presented at the conference "Jülich Soft Matter Days 2008" in Bonn.

Bibliography

[1] Schiessel H, Gelbart WM, Bruinsma R **(2001)** DNA folding: structural and mechanical properties of the two-angle model for chromatin. *Biophys. J.* 80: 1940–1956.

[2] Woodcock C, Frado L, Rattner J **(1984)** The higher-order structure of chromatin: evidence for a helical ribbon arrangement. *J. Cell Biol.* 99: 42-52.

[3] Horowitz RA, Agard DA, Sedat JW, Woodcock CL **(1994)** The three-dimensional architecture of chromatin in situ: electron tomography reveals fibers composed of a continuously variable zig-zag nucleosomal ribbon. *J. Cell Biol.* 125: 1–10.

[4] van Holde K, Zlatanova J **(1996)** What determines the folding of the chromatin fiber? *Proc. Natl. Acad. Sci. U. S. A.* 93: 10548–10555.

[5] Diesinger PM, Heermann DW **(2009)** Depletion effects massively change chromatin properties and influence genome folding. *Biophys. J.* 97: 2146–2153.

[6] Goetze S, Mateos-Langerak J, Gierman HJ, de Leeuw W, Giromus O, et al. **(2007)** The three-dimensional structure of human interphase chromosomes is related to the transcriptome map. *Mol. Cell. Biol.* 27: 4475–4487.

[7] Cremer T, Cremer C **(2001)** Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat. Rev. Genet.* 2: 292–301.

[8] Bystricky K, Heun P, Gehlen L, Langowski J, Gasser SM **(2004)** Long-range compaction and flexibility of interphase chromatin in budding yeast analyzed by high-resolution imaging techniques. *Proc. Natl. Acad. Sci. U. S. A.* 101: 16495–16500.

[9] Lowenstein MG, Goddard TD, Sedat JW **(2004)** Long-range interphase chromosome organization in drosophila: A study using color barcoded fluorescence in situ hybridization and structural clustering analysis. *Mol. Biol. Cell* 15: 5678–5692.

[10] Bantignies F, Grimaud C, Lavrov S, Gabut M, Cavalli G **(2003)** Inheritance of Polycomb-dependent chromosomal interactions in Drosophila. *Genes & Development* 17: 2406-2420.

[11] Sachs R, Engh G, Trask B, Yokota H, Hearst J **(1995)** A Random-Walk/Giant-Loop Model for Interphase Chromosomes *Proc. Natl. Acad. Sci. U. S. A.* 92: 2710-2714.

[12] Mateos-Langerak J, Bohn M, de Leeuw W, Giromus O, Manders EMM, et al. **(2009)** Spatially confined folding of chromatin in the interphase nucleus. *Proceedings of the National Academy of Sciences* 106: 3812-3817.

[13] Versteeg R, van Schaik BDC, van Batenburg MF, Roos M, Monajemi R, et al. **(2003)** 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: 1998–2004.

[14] Scheuermann MO, Tajbakhsh J, Kurz A, Saracoglu K, Eils R, et al. **(2004)** Topology of genes and nontranscribed sequences in human interphase nuclei. *Exp. Cell Res.* 301: 266–279.

[15] Dietzel S, Zolghadr K, Hepperger C, Belmont AS **(2004)** Differential large-scale chromatin compaction and intranuclear positioning of transcribed versus non-transcribed transgene arrays containing beta-globin regulatory sequences. *J. Cell Sci.* 117: 4603-4614.

[16] Janicki SM, Tsukamoto T, Salghetti SE, Tansey WP, Sachidanandam R, et al. **(2004)** From silencing to gene expression: Real-time analysis in single cells. *Cell* 116: 683 - 698.

[17] Cabal GG, Genovesio A, Rodriguez-Navarro S, Zimmer C, Gadal O, et al. **(2006)** SAGA interacting factors confine sub-diffusion of transcribed genes to the nuclear envelope. *Nature* 441: 770–773.

[18] Zink D, Amaral MD, Englmann A, Lang S, Clarke LA, et al. **(2004)** Transcription-dependent spatial arrangements of CFTR and adjacent genes in human cell nuclei. *J. Cell Biol.* 166: 815–825.

[19] Meaburn KJ, Misteli T **(2007)** Cell biology: Chromosome territories. *Nature* 445: 379–381.

[20] Fraser P **(2006)** Transcriptional control thrown for a loop. *Current Opinion in Genetics & Development* 16: 490–495.

[21] Tolhuis B, Palstra RJ, Splinter E, Grosveld F, de Laat W **(2002)** Looping and interaction between hypersensitive sites in the active β-globin locus. *Mol. Cell* 10: 1453–1465.

[22] Dekker J **(2003)** A closer look at long-range chromosomal interactions. *Trends Biochem. Sci.* 28: 277–280.

[23] Cook PR **(2002)** Predicting three-dimensional genome structure from transcriptional activity. *Nat. Genet.* 32: 347–352.

[24] Marenduzzo D, Faro-Trindade I, Cook PR **(2007)** What are the molecular ties that maintain genomic loops? *Trends Genet.* 23: 126 - 133.

[25] Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, et al. **(2006)** Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat. Genet.* 38: 1348–1354.

[26] Hahnfeldt P, Hearst JE, Brenner DJ, Sachs RK, Hlatky LR **(1993)** Polymer models for interphase chromosomes. *Proc. Natl. Acad. Sci. U. S. A.* 90: 7854-7858.

[27] Yokota H, van den Engh G, Hearst J, Sachs R, Trask B **(1995)** Evidence for the organization of chromatin in megabase pair-sized loops arranged along a random walk path in the human G0/G1 interphase nucleus. *J. Cell Biol.* 130: 1239-1249.

[28] Münkel C, Langowski J **(1998)** Chromosome structure predicted by a polymer model. *Phys. Rev. E* 57: 5888–5896.

[29] Münkel C, Eils R, Dietzel S, Zink D, Mehring C, et al. **(1999)** Compartmentalization of interphase chromosomes observed in simulation and experiment *J. Mol. Biol.* 285: 1053-1065.

[30] Rosa A, Everaers R **(2008)** Structure and dynamics of interphase chromosomes. *PLoS Comput. Biol.* 4: e1000153.

[31] Emanuel M, Radja NH, Henriksson A, Schiessel H **(2009)** The physics behind the larger scale organization of dna in eukaryotes. *Physical Biology* 6: 025008 (11pp).

[32] Belmont AS **(2001)** Visualizing chromosome dynamics with GFP. *Trends Cell Biol* 11: 250–257.

[33] Lichter P, Cremer T, Borden J, Manuelidis L, Ward DC **(1988)** Delineation of individual human chromosomes in metaphase and interphase cells by in situ suppression hybridization using recombinant DNA libraries. *Human Genetics* V80: 224–234.

[34] Bolzer A, Kreth G, Solovei I, Koehler D, Saracoglu K, et al. **(2005)** Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol.* 5: e157.

[35] Cook PR, Marenduzzo D **(2009)** Entropic organization of interphase chromosomes. *J. Cell Biol.* 186: 825–834.

[36] Khalil A, Grant J, Caddle L, Atzema E, Mills K, et al. **(2007)** Chromosome territories have a highly nonspherical morphology and nonrandom positioning. *Chromosome Res.* 15: 899–916.

[37] Eils R, Dietzel S, Bertin E, Schröck E, Speicher MR, et al. **(1996)** Three-dimensional reconstruction of painted human interphase chromosomes: active and inactive X chromosome territories have similar volumes but differ in shape and surface structure. *J. Cell Biol.* 135: 1427–1440.

[38] Alberts B, Johnson A, Lewis J, Raff M, Roberts K, et al. **(2002)** Molecular Biology of the Cell. Garland Science, New York.

[39] Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, et al. **(2009)** Comprehensive mapping of long-range interactions reveals folding principles of the human genome *Science* 326: 289–293.

[40] Images are adapted from Wikipedia or Wikimedia Commons and have been published without copyright or under the CC-BY-SA licence. In the following the detailed references with the corresponding URLs are listed:
Fig. 2.1A: Cell (biology) — Wikipedia, The Free Encyclopedia **(2009)**, [Online; accessed 31-October-2009], URL http://en.wikipedia.org/w/index.php?title=Cell_(biology)&oldid=322985573 • Fig. 2.1B: Cell biology — Wikipedia, The Free Encyclopedia **(2009)**, [Online; accessed 31-October-2009], URL http://en.wikipedia.org/w/index.php?title=Cell_biology&oldid=322401936 • Fig. 2.2A/Fig. 3.3 and B: DNA — Wikipedia, The Free Encyclopedia **(2009)**, [Online; accessed 8-December-2009], URL http://en.wikipedia.org/w/index.php?title=DNA&oldid=328823261 • Fig. 2.2C: Histone — Wikipedia, The Free Encyclopedia **(2009)**, [Online; accessed 8-December-2009], URL http://en.wikipedia.org/w/index.php?title=Histone&oldid=326350120 • Fig. 2.2D and E/Fig. 3.3: Chromatin — Wikipedia, The Free Encyclopedia, **(2009)**, [Online, accessed 8-December-2009], URL http://en.wikipedia.org/w/index.php?title=Chromatin&oldid=329229798 • Fig. 2.6: Chromosome conformation capture — Wikipedia, The Free Encyclopedia, **(2009)**, [Online; accessed 14-December-2009] URL http://en.wikipedia.org/w/index.php?title=Chromosome_conformation_capture&oldid=321985874.

[41] Crick FHC, Watson JD **(1954)** The Complementary Structure of Deoxyribonucleic Acid. *Royal Society of London Proceedings Series A* 223: 80-96.

[42] ConsortiumInternational HGS **(2004)** Finishing the euchromatic sequence of the human genome. *Nature* 431: 931–945.

[43] Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ **(1997)** Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389: 251–260.

[44] Kornberg RD, Lorch Y **(1999)** Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell* 98: 285–294.

[45] Thoma F, Koller T, Klug A **(1979)** Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. *J. Cell Biol.* 83: 403–427.

[46] Widom J **(1986)** Physicochemical studies of the folding of the 100 A nucleosome filament into the 300 A filament. Cation dependence. *J. Mol. Biol.* 190: 411–424.

[47] Finch JT, Klug A **(1976)** Solenoidal model for superstructure in chromatin. *Proc. Natl. Acad. Sci. U. S. A.* 73: 1897–1901.

[48] Leuba SH, Yang G, Robert C, Samori B, van Holde K, et al. **(1994)** Three-dimensional structure of extended chromatin fibers as revealed by tapping-mode scanning force microscopy. *Proc. Natl. Acad. Sci. U. S. A.* 91: 11621–11625.

[49] Diesinger PM, Heermann DW **(2006)** Two-angle model and phase diagram for chromatin. *Phys. Rev. E* 74: 031904.

[50] Bode J, Goetze S, Heng H, Krawetz S, Benham C **(2003)** From DNA structure to gene expression: mediators of nuclear compartmentalization and dynamics. *Chromosome Res.* 11: 435–445.

[51] Cai S, Lee CC, Kohwi-Shigematsu T **(2006)** SATB1 packages densely looped, transcriptionally active chromatin for coordinated expression of cytokine genes. *Nat. Genet.* 38: 1278–1288.

[52] Splinter E, Heath H, Kooren J, Palstra RJ, Klous P, et al. **(2006)** CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes & Development* 20: 2349–2354.

[53] Caron H, van Schaik B, van der Mee M, Baas F, Riggins G, et al. **(2001)** The human transcriptome map: clustering of highly expressed genes in chromosomal domains. *Science* 291: 1289–1292.

[54] Strahl BD, Allis CD **(2000)** The language of covalent histone modifications. *Nature* 403: 41–45.

[55] Luo L, Gassman KL, Petell LM, Wilson CL, Bewersdorf J, et al. **(2009)** The nuclear periphery of embryonic stem cells is a transcriptionally permissive and repressive compartment. *J. Cell Sci.* 122: 3729-3737.

[56] Jenuwein T, Allis CD **(2001)** Translating the histone code. *Science* 293: 1074–1080.

[57] Gilbert N, Boyle S, Fiegler H, Woodfine K, Carter NP, et al. **(2004)** Chromatin architecture of the human genome: Gene-rich domains are enriched in open chromatin fibers. *Cell* 118: 555 - 566.

[58] Cremer T, Cremer C, Schneider T, Baumann H, Hens L, et al. **(1982)** Analysis of chromosome positions in the interphase nucleus of chinese hamster cells by laser-UV-microirradiation experiments. *Hum Genet* 62: 201–209.

[59] Cremer T, Lichter P, Borden J, Ward DC, Manuelidis L **(1988)** Detection of chromosome aberrations in metaphase and interphase tumor cells by in situ hybridization using chromosome-specific library probes. *Hum Genet* 80: 235–246.

[60] Branco M, Pombo A **(2006)** Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS Biol.* 4: 97-104.

[61] Nagele R, Freeman T, McMorrow L, Lee HY **(1995)** Precise spatial positioning of chromosomes during prometaphase: evidence for chromosomal order. *Science* 270: 1831–1835.

[62] Allison DC, Nestor AL **(1999)** Evidence for a relatively random array of human chromosomes on the mitotic ring. *J. Cell Biol.* 145: 1–14.

[63] E N C O D E Project Consortium, Birney E, Stamatoyannopoulos JA, Dutta A, Guigó R, et al. **(2007)** Identification and analysis of functional elements in 1*Nature* 447: 799–816.

[64] Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, et al. **(2007)** Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* 39: 311–318.

[65] de Laat W, Grosveld F **(2003)** Spatial organization of gene expression: the active chromatin hub. *Chromosome Res.* 11: 447–459.

[66] Dekker J, Rippe K, Dekker M, Kleckner N **(2002)** Capturing Chromosome Conformation. *Science* 295: 1306-1311.

[67] Simonis M, Kooren J, de Laat W **(2007)** An evaluation of 3C-based methods to capture dna interactions. *Nat. Methods* 4: 895–901.

[68] Dostie J, Richmond TA, Arnaout RA, Selzer RR, Lee WL, et al. **(2006)** Chromosome Conformation Capture Carbon Copy (5C): A massively parallel solution for mapping interactions between genomic elements. *Genome Res.* 16: 1299-1309.

[69] Jackson DA, Hassan AB, Errington RJ, Cook PR **(1993)** Visualization of focal sites of transcription within human nuclei. *EMBO J.* 12: 1059–1065.

[70] Fraser P, Bickmore W **(2007)** Nuclear organization of the genome and the potential for gene regulation. *Nature* 447: 413–417.

[71] Osborne CS, Chakalova L, Brown KE, Carter D, Horton A, et al. **(2004)** Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat. Genet.* 36: 1065–1071.

[72] Chubb JR, Trcek T, Shenoy SM, Singer RH **(2006)** Transcriptional pulsing of a developmental gene. *Curr. Biol.* 16: 1018–1025.

[73] Paulson JR, Laemmli UK **(1977)** The structure of histone-depleted metaphase chromosomes. *Cell* 12: 817–828.

[74] Pederson T **(2000)** Half a century of "the nuclear matrix". *Mol. Biol. Cell* 11: 799–805.

[75] Filippova GN, Fagerlie S, Klenova EM, Myers C, Dehner Y, et al. **(1996)** An exceptionally conserved transcriptional repressor, ctcf, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. *Mol. Cell. Biol.* 16: 2802–2813.

[76] Phillips JE, Corces VG **(2009)** CTCF: Master Weaver of the Genome. *Cell* 137: 1194–1211.

[77] Bell AC, West AG, Felsenfeld G **(1999)** The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* 98: 387–396.

[78] de Gennes PG **(1979)** Scaling concepts in polymer physics. Ithaca, N.Y., Cornell University Press.

[79] Grosberg AY, Khokhlov AR **(1994)** Statistical Physics of Macromolecules. AIP Press.

[80] Doi M, See HÜ **(1997)** Introduction to polymer physics. Oxford [u.a.]: Clarendon Press, IX, 120 S. pp.

[81] Doi M, Edwards SF **(1986)** The theory of polymer dynamics. International series of monographs on physics ; 73. Clarendon Pr., XIII, 391 S. pp.

[82] Feller W **(1968)** An introduction to probability theory and its applications/1. Wiley series in probability and mathematical statistics. New York, NY: Wiley, XVIII, 509 S. pp.

[83] Bruns W **(1992)** Angular-correlations within coiled polymer-molecules. *Makromolekulare Chemie-Theory And Simulations* 1: 287–293.

[84] Šolc K, Stockmayer WH **(1971)** Shape of a random-flight chain. *J. Chem. Phys.* 54: 2756-2757.

[85] Verdier PH, Stockmayer WH **(1962)** Monte Carlo calculations on the dynamics of polymers in dilute solution *J. Chem. Phys.* 36: 227-235.

[86] Madras N, Sokal A **(1988)** The pivot algorithm: a highly efficient Monte Carlo method for the self-avoiding walk. *J. Stat. Phys.* 50: 109-168.

[87] Kennedy T **(2002)** A Faster Implementation of the Pivot Algorithm for Self-Avoiding Walks *J. Stat. Phys.* 106: 407-429.

[88] Flory PJ **(1949)** The configuration of real polymer chains. *J. Chem. Phys.* 17: 303-310.

[89] Le Guillou JC, Zinn-Justin J **(1980)** Critical exponents from field theory. *Phys. Rev. B* 21: 3976–3998.

[90] Rippe K **(2001)** Making contacts on a nucleic acid polymer. *Trends Biochem. Sci.* 26: 773-740.

[91] Langowski J **(2006)** Polymer chain models of DNA and chromatin. *Eur Phys J E Soft Matter* 19: 241–249.

[92] van den Engh G, Sachs R, Trask BJ **(1992)** Estimating genomic distance from DNA sequence location in cell nuclei by a random walk model. *Science* 257: 1410–1412.

[93] Bohn M, Heermann DW **(2009)** Diffusion-driven looping provides a consistent framework for chromatin organization. *submitted to PLoS Computational Biology* -: -.

[94] Liu B, Sachs R **(1997)** A two-backbone polymer model for interphase chromosome geometry. *Bulletin of Mathematical Biology* 59: 325–337.

[95] Heng HH, Chamberlain JW, Shi XM, Spyropoulos B, Tsui LC, et al. **(1996)** Regulation of meiotic chromatin loop size by chromosomal position. *Proc. Natl. Acad. Sci. U. S. A.* 93: 2795–2800.

[96] Odenheimer J, Kreth G, Heermann DW **(2005)** Dynamic simulation of active/inactive chromatin domains. *Journal of Biological Physics* 31: 351 - 363.

[97] Sikorav JL, Jannink G **(1994)** Kinetics of chromosome condensation in the presence of topoisomerases - a phantom chain model. *Biophys. J.* 66: 827–837.

[98] Grosberg A, Rabin Y, Havlin S, Neer A **(1993)** Crumpled globule model of the three-dimensional structure of DNA. *EPL (Europhysics Letters)* 23: 373-378.

[99] Dill KA, Bromberg S, Yue K, Fiebig KM, Yee DP, et al. **(1995)** Principles of protein folding–a perspective from simple exact models. *Protein Sci.* 4: 561–602.

[100] Lau KF, Dill KA **(1989)** A lattice statistical-mechanics model of the conformational and sequence-spaces of proteins. *Macromolecules* 22: 3986–3997.

[101] Shakhnovich E, Gutin A **(1993)** Engineering of stable and fast-folding sequences of model proteins. *Proc. Natl. Acad. Sci. U. S. A.* 90: 7195-7199.

[102] Rapaport DC **(1985)** On three-dimensional self-avoiding walks. *Journal of Physics A: Mathematical and General* 18: 113-126.

[103] Domb C, Hioe FT **(1969)** Mean-square intrachain distances in a self-avoiding walk. *J. Chem. Phys.* 51: 1915-1919.

[104] Ramakrishnan R, Pekny JF, Caruthers JM **(1995)** A combinatorial algorithm for effective generation of long maximally compact lattice chains. *J. Chem. Phys.* 103: 7592-7604.

[105] Mansfield ML **(2006)** Unbiased sampling of lattice hamilton path ensembles. *J. Chem. Phys.* 125: 154103.

[106] Lua R, Borovinskiy AL, Grosberg AY **(2004)** Fractal and statistical properties of large compact polymers: a computational study. *Polymer* 45: 717–731.

[107] Pande VS, Joerg C, Grosberg AY, Tanaka T **(1996)** Enumeration of the Hamiltonian walks on a cubic sublattice. *Journal of Physics A Mathematical General* 29: 4753-4753.

[108] Fisher M **(1966)** Shape of a self-avoiding walk or polymer chain. *J. Chem. Phys.* 44: 616-622.

[109] Jhunjhunwala S, van Zelm MC, Peak MM, Cutchin S, Riblet R, et al. **(2008)** The 3D structure of the immunoglobulin heavy-chain locus: implications for long-range genomic interactions. *Cell* 133: 265–279.

[110] Binder K, editor **(1995)** Monte Carlo and Molecular Dynamics Simulations in Polymer Science. Oxford University Press.

[111] Binder K, Heermann DW **(2002)** Monte Carlo Simulations in Statistical Physics. An Introduction. Springer, 4 edition.

[112] Jacobsen JL **(2008)** Unbiased sampling of globular lattice proteins in three dimensions. *Phys. Rev. Lett.* 100: 118102.

[113] Paul W, Binder K, Heermann DW, Kremer K **(1991)** Crossover scaling in semidilute polymer-solutions - a Monte-Carlo test. *Journal De Physique Ii* 1: 37–60.

[114] Stephen MJ **(1978)** Mean-field theory and critical exponents for a random resistor network. *Phys. Rev. B* 17: 4444–4453.

[115] Heermann DW, Bohn M **(2007)** Percolation in a class of band structured random matrices. *arXiv:0705.1241v1 [cond-mat.stat-mech]* .

[116] Anderson E, Bai Z, Bischof C, Blackford S, Demmel J, et al. **(1999)** LAPACK Users' Guide. Philadelphia, PA: Society for Industrial and Applied Mathematics, third edition.

[117] Bohn M, Heermann DW, van Driel R **(2007)** Random loop model for long polymers. *Phys. Rev. E* 76: 051805.

[118] Case DA, III TEC, Darden T, Gohlke H, Luo R, et al. **(2005)** The Amber biomolecular simulation programs. *J. Comput. Chem.* 26: 1668–1688.

[119] Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, et al. **(1983)** CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* 4: 187–217.

[120] Spoel DVD, Lindahl E, Hess B, Groenhof G, Mark AE, et al. **(2005)** Gromacs: Fast, flexible, and free. *J. Comput. Chem.* 26: 1701–1718.

[121] Limbach H, Arnold A, Mann B, Holm C **(2006)** ESPResSo–an extensible simulation package for research on soft matter systems. *Comput. Phys. Commun.* 174: 704–727.

[122] Rapaport DC **(1995)** The Art of Molecular Dynamics Simulation. Cambridge: Cambridge University Press, XIV, 400 S. pp.

[123] Allen MP **(2004)** Introduction to Molecular Dynamics Simulation. In: Attig N, Binder K, Grubmuller H, (Eds) KK, editors, Computational Soft Matter: From Synthetic Polymers to Proteins,Lecture Notes., John von Neumann Institute for Computing, Juelich. ÂÍ, volume 23 of *NIC Series*. pp. 1-28.

[124] Andersen HC **(1980)** Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.* 72: 2384-2393.

[125] Weeks JD, Chandler D, Andersen HC **(1971)** Role of repulsive forces in determining the equilibrium structure of simple liquids. *J. Chem. Phys.* 54: 5237-5247.

[126] Kremer K, Grest GS **(1990)** Dynamics of entangled linear polymer melts: A molecular-dynamics simulation. *J. Chem. Phys.* 92: 5057-5086.

[127] Mavrich TN, Jiang C, Ioshikhes IP, Li X, Venters BJ, et al. **(2008)** Nucleosome organization in the drosophila genome. *Nature* 453: 358–362.

[128] Taylor J **(2005)** Clues to function in gene deserts. *Trends Biotechnol.* 23: 269–271.

[129] Shopland LS, Lynch CR, Peterson KA, Thornton K, Kepper N, et al. **(2006)** Folding and organization of a contiguous chromosome region according to the gene distribution pattern in primary genomic sequence. *J. Cell Biol.* 174: 27–38.

[130] Palstra RJ, Tolhuis B, Splinter E, Nijmeijer R, Grosveld F, et al. **(2003)** The beta-globin nuclear compartment in development and erythroid differentiation. *Nat. Genet.* 35: 190–194.

[131] Petrascheck M, Escher D, Mahmoudi T, Verrijzer CP, Schaffner W, et al. **(2005)** DNA looping induced by a transcriptional enhancer in vivo. *Nucleic Acids Res.* 33: 3743-3750.

[132] Fiers W, Sinsheimer RL **(1962)** The structure of the DNA of bacteriophage [phi]X174: III. Ultracentrifugal evidence for a ring structure. *J. Mol. Biol.* 5: 424–434.

[133] Dulbecco R, Vogt M **(1963)** Evidence for a ring structure of polyoma virus DNA. *Proc. Natl. Acad. Sci. U. S. A.* 50: 236–243.

[134] Weil R, Vinograd J **(1963)** The cyclic helix and cyclic coil forms of polyoma viral DNA. *Proc. Natl. Acad. Sci. U. S. A.* 50: 730–738.

[135] Roovers J, Toporowski PM **(1983)** Synthesis of high molecular weight ring polystyrenes *Macromolecules* 16: 843–849.

[136] McKenna GB, Hadziioannou G, Lutz P, Hild G, Strazielle C, et al. **(1987)** Dilute solution characterization of cyclic polystyrene molecules and their zero-shear viscosity in the melt *Macromolecules* 20: 498–512.

[137] Tead SF, Kramer EJ, Hadziioannou G, Antonietti M, Sillescu H, et al. **(1992)** Polymer topology and diffusion: a comparison of diffusion in linear and cyclic macromolecules *Macromolecules* 25: 3942–3947.

[138] Lodge TP, Rotstein NA, Prager S **(2007)** Dynamics of entangled polymer liquids: Do linear chains reptate? In: I Prigogine SAR, editor, Advances in Chemical Physics, John Wiley & Sons. pp. 1–132.

[139] Bishop M, Michels JPJ **(1985)** The shape of ring polymers. *J. Chem. Phys.* 82: 1059-1061.

[140] Bishop M, Saltiel CJ **(1988)** Polymer shapes in two, four, and five dimensions. *J. Chem. Phys.* 88: 3976–3980.

[141] Deutsch JM **(1999)** Equilibrium size of large ring molecules. *Phys. Rev. E* 59: R2539–R2541.

[142] Baumgartner A **(1982)** Statistics of self-avoiding ring polymers. *J. Chem. Phys.* 76: 4275-4280.

[143] des Cloizeaux J **(1981)** Ring polymers in solution : topological effects. *J. Physique Lett.* 42: 433-436.

[144] Grosberg **(2000)** Critical exponents for random knots *Phys Rev Lett* 85: 3858–3861.

[145] Moore NT, Lua RC, Grosberg AY **(2004)** Topologically driven swelling of a polymer loop. *Proc Natl Acad Sci U S A* 101: 13431–13435.

[146] Brown S, Lenczycki T, Szamel G **(2001)** Influence of topological constraints on the statics and dynamics of ring polymers. *Phys. Rev. E* 63: 052801.

[147] Müller M, Wittmer JP, Cates ME **(1996)** Topological effects in ring polymers: A computer simulation study. *Phys. Rev. E* 53: 5063–5074.

[148] Cates M, Deutsch J **(1986)** Conjectures on the statistics of ring polymers. *Journal de Physique* 47: 2121-2128.

[149] Müller M, Wittmer JP, Cates ME **(2000)** Topological effects in ring polymers. II. Influence of persistence length. *Phys. Rev. E* 61: 4078–4089.

[150] Depew DE, Wang JC **(1975)** Conformational fluctuations of DNA helix. *Proc. Natl. Acad. Sci. U. S. A.* 72: 4275–4279.

[151] Vologodskii AV, Zhang W, Rybenkov VV, Podtelezhnikov AA, Subramanian D, et al. **(2001)** Mechanism of topology simplification by type II DNA topoisomerases. *Proc. Natl. Acad. Sci. U. S. A.* 98: 3045–3049.

[152] Zhou HX **(2003)** Effect of catenation on protein folding stability. *J. Am. Chem. Soc.* 125: 9280–9281.

[153] Otto M **(2004)** Topological interactions in systems of mutually interlinked polymer rings. *Journal of Physics A: Mathematical and General* 37: 2881-2893.

[154] Sikorski A **(1994)** Monte Carlo study of catenated ring polymers. *Polymer* 35: 3792–3794.

[155] Carmesin I, Kremer K **(1988)** The bond fluctuation method: a new effective algorithm for the dynamics of polymers in all spatial dimensions. *Macromolecules* 21: 2819-2823.

[156] Deutsch HP, Binder K **(1991)** Interdiffusion and self-diffusion in polymer mixtures: A Monte Carlo study. *J. Chem. Phys.* 94: 2294-2304.

[157] Sokal A **(1996)** Monte carlo methods in statistical mechanics: foundations and new algorithms. In: Functional integration.

[158] Rawdon EJ, Kern JC, Piatek M, Plunkett P, Stasiak A, et al. **(2008)** Effect of knotting on the shape of polymers. *Macromolecules* 41: 8281–8287.

[159] Bohn M, Heermann DW **(2009)** Conformational properties of compact polymers. *J. Chem. Phys.* 130: 174901.

[160] Bishop M, Michels JPJ **(1986)** Scaling in two-dimensional linear and ring polymers. *J. Chem. Phys.* 85: 1074–1076.

[161] Marenduzzo D, Orlandini E **(2009)** Topological and entropic repulsion in biopolymers *Journal of Statistical Mechanics: Theory and Experiment* 2009: L09002–.

[162] Dorier J, Stasiak A **(2009)** Topological origins of chromosomal territories. *Nucleic Acids Res* 37: 6316–6322.

[163] Vettorel T, Grosberg AY, Kremer K **(2009)** Statistics of polymer rings in the melt: a numerical simulation study. *Physical Biology* 6: 025013 (10pp).

[164] Bohn M, Heermann DW, Lourenco O, Cordeiro C **(2009)** Influence of the catenation constraint on elongation and segregation of ring polymers *submitted to Macromolecules* -: -.

[165] Frank-Kamenetskii MD, Lukashin AV, Vologodskii AV **(1975)** Statistical mechanics and topology of polymer chains. *Nature* 258: 398–402.

[166] Likos CN **(2001)** Effective interactions in soft condensed matter physics. *Physics Reports* 348: 267 - 439.

[167] Dautenhahn J, Hall CK **(1994)** Monte Carlo simulation of off-lattice polymer chains: Effective pair potentials in dilute solution. *Macromolecules* 27: 5399-5412.

[168] Hsu HP, Grassberger P **(2004)** Effective interactions between star polymers. *Europhys. Lett.* 66: 874-880.

[169] Otto M, Vilgis TA **(1998)** Topological interactions in multiply linked DNA rings. *Phys. Rev. Lett.* 80: 881–.

[170] Flory PJ, Krigbaum WR **(1950)** Statistical mechanics of dilute polymer solutions. II *J. Chem. Phys.* 18: 1086-1094.

[171] Daoud M, Cotton JP, Farnoux B, Jannink G, Sarma G, et al. **(1975)** Solutions of flexible polymers. neutron experiments and interpretation. *Macromolecules* 8: 804–818.

[172] Grosberg AY, Khalatur PG, Khokhlov AR **(1982)** Polymeric coils with excluded volume in dilute solution: The invalidity of the model of impenetrable spheres and the influence of excluded volume on the rates of diffusion-controlled intermacromolecular reactions *Die Makromolekulare Chemie, Rapid Communications* 3: 709–713.

[173] Bolhuis PG, Louis AA, Hansen JP, Meijer EJ **(2001)** Accurate effective pair potentials for polymer solutions. *J. Chem. Phys.* 114: 4296-4311.

[174] Galande S, Purbey PK, Notani D, Kumar PP **(2007)** The third dimension of gene regulation: organization of dynamic chromatin loopscape by SATB1. *Current Opinion in Genetics & Development* 17: 408–414.

[175] Chandley AC, Speed RM, Leitch AR **(1996)** Different distributions of homologous chromosomes in adult human sertoli cells and in lymphocytes signify nuclear differentiation. *J. Cell Sci.* 109 (Pt 4): 773–776.

[176] Comings DE **(1980)** Arrangement of chromatin in the nucleus. *Hum Genet* 53: 131–143.

[177] Sun HB, Shen J, Yokota H **(2000)** Size-dependent positioning of human chromosomes in interphase nuclei. *Biophys. J.* 79: 184–190.

[178] Dundr M, Misteli T **(2001)** Functional architecture in the cell nucleus. *Biochem. J.* 356: 297–310.

[179] Gilbert N, Gilchrist S, Bickmore WA **(2005)** Chromatin organization in the mammalian nucleus. *Int Rev Cytol* 242: 283–336.

[180] Bohn M, Heermann DW **(2010)** Topological interactions between ring polymers: Implications for chromatin loops. *Journal of Chemical Physics* in press.

[181] Diesinger P **(2009)** Genome Folding at the 30nm Scale. Ph.D. thesis, University of Heidelberg.

[182] Morikawa K, Yanagida M **(1981)** Visualization of individual DNA molecules in solution by light microscopy: DAPI staining method. *J Biochem* 89: 693–696.

[183] Hell SW, Lindek S, Cremer C, Stelzer EHK **(1994)** Measurement of the 4Pi-confocal point spread function proves 75 nm axial resolution. *Appl. Phys. Lett.* 64: 1335-1337.

[184] Bewersdorf J, Bennett BT, Knight KL **(2006)** H2AX chromatin structures and their response to DNA damage revealed by 4Pi microscopy. *Proc. Natl. Acad. Sci. U. S. A.* 103: 18137-18142.

[185] Olins AL, Olins DE **(1974)** Spheroid chromatin units (ngr bodies). *Science* 183: 330–332.

[186] Dubochet J, Adrian M, Chang JJ, Homo JC, Lepault J, et al. **(1988)** Cryo-electron microscopy of vitrified specimens. *Q Rev Biophys* 21: 129–228.

[187] Bednar J, Woodcock CL **(1999)** Cryoelectron microscopic analysis of nucleosomes and chromatin. *Methods Enzymol* 304: 191–213.

[188] Dellaire G, Kepkay R, Bazett-Jones D **(2009)** High resolution imaging of changes in the structure and spatial organization of chromatin, γ-H2AX and the MRN complex within etoposide-induced dna repair foci. *Cell Cycle* 8.

[189] Betzig E, Patterson GH, Sougrat R, Lindwasser OW, Olenych S, et al. **(2006)** Imaging intracellular fluorescent proteins at nanometer resolution. *Science* 313: 1642–1645.

[190] Hess ST, Girirajan TPK, Mason MD **(2006)** Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophys. J.* 91: 4258–4272.

[191] Rust MJ, Bates M, Zhuang X **(2006)** Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat. Methods* 3: 793–795.

[192] Cremer C **(2009)** Lichtoptische Nanoskopie. *BioSpektrum* 01/2009: 27.

[193] Cremer C, Edelmann P, Bornfleth H, Kreth G, Muench H, et al. **(1999)** Principles of spectral precision distance confocal microscopy for the analysis of molecular nuclear structure. *Handbook of Computer Vision and Applications* 3.

[194] Heilemann M, Herten DP, Heintzmann R, Cremer C, Muller C, et al. **(2002)** High-resolution colocalization of single dye molecules by fluorescence lifetime imaging microscopy. *Anal. Chem.* 74: 3511–3517.

[195] Lemmer P, Gunkel M, Baddeley D, Kaufmann R, Urich A, et al. **(2008)** SPDM: light microscopy with single-molecule resolution at the nanoscale. *Applied Physics B* 93: 1-12.

[196] Bornfleth H, Satzler K, Eils R, Cremer C **(1998)** High-precision distance measurements and volume-conserving segmentation of objects near and below the resolution limit in three-dimensional confocal fluorescence microscopy. *Journal of Microscopy* 189: 118-136.

[197] Cremer C, von Ketteler A, Lemmer P, Kaufmann R, Weiland Y, et al. **(2009)** Far field fuorescence microscopy of cellular structures @ molecular resolution, nanoscopy and multidimensional optical fluorescence microscopy. *Taylor & Francis* in press.

[198] Kaufmann R, Lemmer P, Gunkel M, Weiland Y, Mueller P, et al. **(2009)** SPDM: single molecule superresolution of cellular nanostructures. In: Proceedings of SPIE. volume 7185, p. 71850J.

[199] Hendrix J, Flors C, Dedecker P, Hofkens J, Engelborghs Y **(2008)** Dark states in monomeric red fluorescent proteins studied by fluorescence correlation and single molecule spectroscopy. *Biophysical Journal* 94: 4103–4113.

[200] Patterson GH, Lippincott-Schwartz J **(2002)** A photoactivatable GFP for selective photolabeling of proteins and cells. *Science* 297: 1873–1877.

[201] Sinnecker D, Voigt P, Hellwig N, Schaefer M **(2005)** Reversible photobleaching of enhanced green fluorescent proteins. *Biochemistry* 44: 7085–7094.

[202] Kubo R **(1971)** Statistical mechanics : an advanced course with problems and solutions. North-Holland Publ. Comp.

[203] Rovere M, Heermann DW, Binder K **(1988)** Block density distribution function analysis of two-dimensional Lennard-Jones fluids. *Europhys. Lett.* 6: 585-590.

[204] Binder K **(1981)** Finite size scaling analysis of ising model block distribution functions. *Zeitschrift für Physik B Condensed Matter* 43: 119–140.

[205] Auhl R, Everaers R, Grest GS, Kremer K, Plimpton SJ **(2003)** Equilibration of long chain polymer melts in computer simulations. *J. Chem. Phys.* 119: 12718-12728.

[206] Wang H, Monteiro MJ **(2007)** Ubiquilin overexpression reduces gfp-polyalanine-induced protein aggregates and toxicity. *Exp. Cell Res.* 313: 2810–2820.

[207] Lunyak VV, Rosenfeld MG **(2008)** Epigenetic regulation of stem cell fate. *Hum Mol Genet* 17: R28–R36.

[208] Reddy KL, Zullo JM, Bertolino E, Singh H **(2008)** Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature* 452: 243–247.

[209] Meshorer E, Yellajoshula D, George E, Scambler PJ, Brown DT, et al. **(2006)** Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Dev Cell* 10: 105–116.

[210] Efroni S, Duttagupta R, Cheng J, Dehghani H, Hoeppner DJ, et al. **(2008)** Global transcription in pluripotent embryonic stem cells. *Cell Stem Cell* 2: 437–447.

[211] Odenheimer J, Heermann DW, Kreth G **(2009)** Brownian dynamics simulations reveal regulatory properties of higher-order chromatin structures. *Eur Biophys J* 38: 749–756.

[212] Winkler G **(2003)** Image analysis, random fields and Markov Chain Monte Carlo Methods. Applications of mathematics ; 27. Springer, XVI, 387 S. pp.

[213] Geman S, Geman D **(1984)** Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. In: IEEE Transactions on Pattern Analysis and Machine Intelligence. volume 6, pp. 721 - 741.

[214] Morris RD, Descombes X, Zerubia J **(1996)** The Ising/Potts model is not well suited to segmentation tasks. In: In Proceedings of the IEEE Digital Signal Processing Workshop. pp. 263-266.

[215] Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, et al. **(2009)** Human dna methylomes at base resolution show widespread epigenomic differences *Nature* advance online publication: –.

Acknowledgments

The work described in this thesis could not have been conducted without the support and knowledge of many people.

First of all, I would like to thank my advisor, ***Prof. Dieter W. Heermann*** for his strong and permanent support throughout the time at the Institute of Theoretical Physics. Without his knowledge in physical and computational topics, his effort in representing the results at workshops and conferences, this work would not have been possible.

I owe many thanks to my co-advisor ***Prof. Christoph Cremer*** for many fruitful discussions as well as his engagement in our collaboration on chromatin nanostructure.

Our collaborators ***Prof. Roel van Driel***, ***Dr. Sandra Goetze*** and ***Dr. Julio Mateos-Langerak*** are thanked for ongoing collaborations and fruitful discussions.

Prof. Lindsay Shopland and ***Dr. Joerg Bewersdorf*** are appreciated for their hospitality, the fruitful discussions and for sharing their data with us.

Prof. Wouter de Laat is regarded for sharing 4C data with us.

Prof. Heinz Horner is gratefully appreciated for fruitful discussions on analytical calculations of the Random Loop model as well as on ring polymers.

Prof. Wolfang Paul is thanked for providing us with a FORTRAN implementation of the bond fluctuation model.

This work was mainly funded by the Landesgraduiertenförderung Baden-Württemberg, partial support, especially for the participation at conferences, was offered by the Heidelberg Graduate School of Mathematical and Computational Methods for the Sciences as well as the Institute of Theoretical Physics (ITP) of the University of Heidelberg.

I would like to thank the former and present members of the Computational Biophysics group at the Institute of Theoretical Physics (ITP) for the time at the institute, especially ***Marcel Hellmann***, ***Philipp Diesinger***, ***Miriam Fritsche***, ***Benoît Knecht***, ***Timo Fischer***, ***Songling Li***, ***Yang Zhang*** and ***Michael Bach***.

Special thanks goes to the secretaries ***Sonja Bartsch***, ***Melanie Steiert*** and ***Cornelia Merkel*** for their very friendly support.

I thank ***Gudrun Baumann*** and ***Miriam Fritsche*** for proofreading my thesis.

Last but not least I owe special thanks to my family, especially my parents ***Dr. Hans Bohn*** and ***Gabriele Bohn***, for their constant support over all these years.