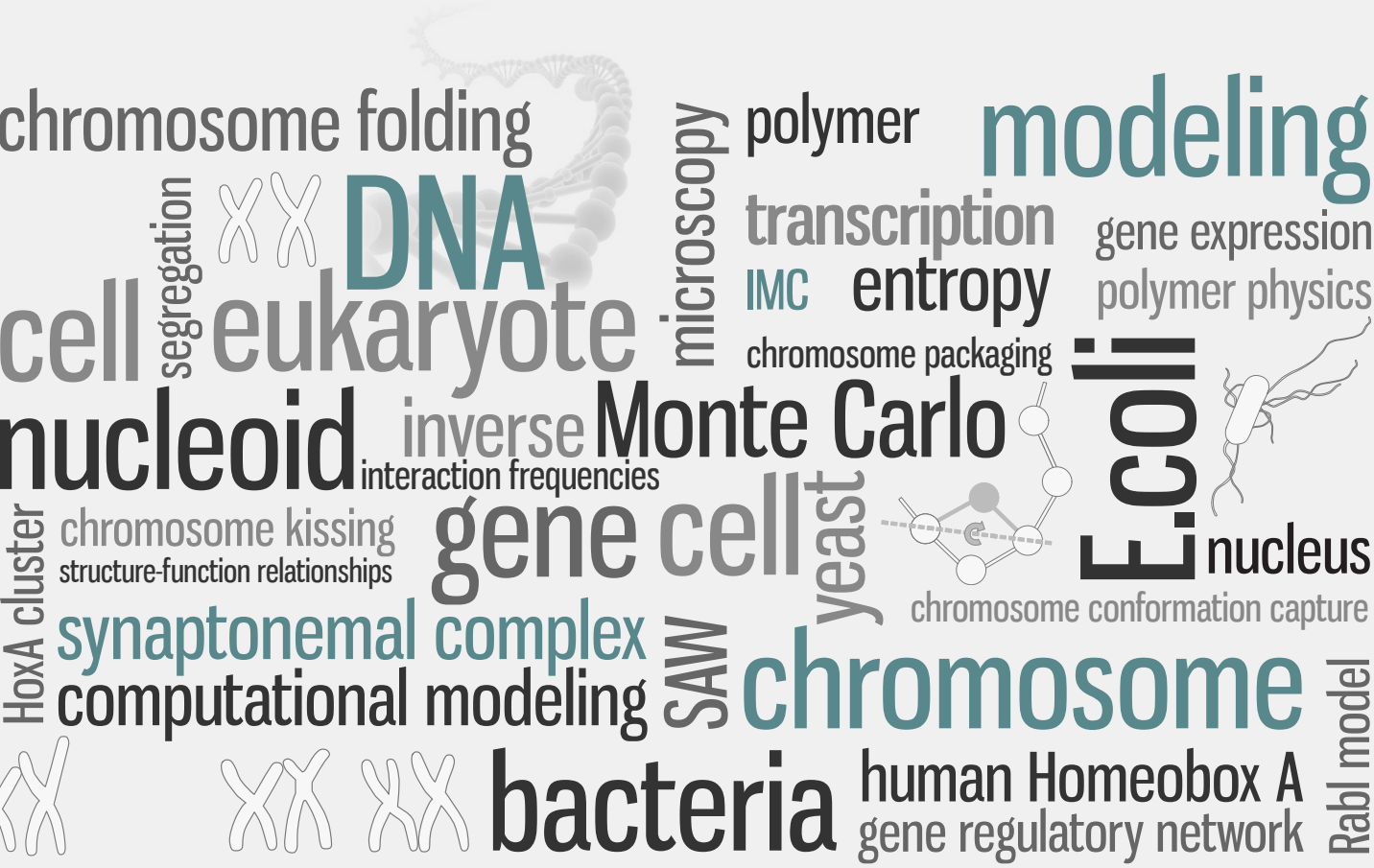
Chromosome Kissing and Chromosome Folding in Eukaryotic and Bacterial Cells



Miriam Fritsche

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

M.Sc. Miriam Fritsche

born in: Groß-Gerau

Date of oral examination: 26.10.2011

Chromosome Kissing and Chromosome Folding in Eukaryotic and Bacterial Cells

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

Meiner Familie

Abstract

The genomes of bacteria and eukaryotes are intricately organized by the mutual interplay between three-dimensional (3D) genome folding and functional cell activities. The aim of this thesis is to investigate such structure-function relationships.

I have identified four key features that greatly affect biopolymer architecture: polymer topology, confinement, semiflexibility and tethering. One project of my thesis shows that the 3D organization of synaptonemal complexes during meiosis is strongly influenced by the entropic interplay between confinement and double-tethering to the confinement walls. In another collaboration, we have demonstrated that the spatial packaging and dynamics of yeast chromosomes can effectively be described by a Rabl model of organization based on confinement, tethering and the appropriate chromatin packaging density.

A main part of my thesis contributes to a better understanding of spatial E. coli chromosome organization and segregation. We could show that looped star polymers, whose structural makeup reflects E. coli chromosome topology, show a conformational transition in free space from spherical to flat toroidal shapes, which might facilitate proper genome packaging within a rod-shaped, bacterial envelope. To this end, we provide evidence that the coupling of chromosome topology and confinement is relevant for the bacterial cell to overcome the chromosome's propensity to mix. We propose and test one possible mechanism for chromosomal domain formation: Colocalization of transcription factors and target genes, mimicking the regulatory control defined by the E. coli transcriptional regulatory network, can explain the experimentally found high precision of subnuclear positioning of genetic sites. The developed framework of chromosome packaging is also relevant for a better understanding of E. coli chromosome segregation, suggesting that excluded volume effects, specific polymer topologies and geometrical confinement compete with entropy to drive dynamic processes such as the segregation of highly compacted chromosomes during cell division.

Last but not least, a computational tool to unravel chromosome architecture was developed which models the chromatin fiber as a worm-like chain and solves for a (unique) maximum-entropy ensemble of conformations consistent with structural constraints based on chromosome conformation capture data (contact interaction maps) or fluorescence-based/microscopy data (radial distribution functions).

Zusammenfassung

Das bakterielle und eukaryotische Genom wird durch ein differenziertes Zusammenspiel von dreidimensionaler (3D) Kompaktifizierung und Zellfunkionalität, beispielsweise bezüglich Genexpression, organisiert. Das Ziel dieser Arbeit ist es, solche Struktur- und Funktionszusammenhänge zu untersuchen.

Ich habe vier Schlüsselmerkmale identifiziert, die einen grossen Einfluss auf die Architektur von Biopolymeren haben: Polymertopologie, (räumlicher) Einschluss, Semiflexibilität und Fixierung. In einem Projekt meiner Doktorarbeit habe ich gezeigt, dass die 3D-Organisation der Synaptonemalen Komplexe (SK) während der Meiose stark durch das Wechselspiel von räumlichem Einschluss und doppelter Fixierung der SK-Enden beeinflusst wird. In einer anderen Kollaboration konnte gezeigt werden, dass die räumlich Anordnung und Dynamik von Hefe-Chromosomen im Rahmen eines Rabl Models, basierend auf Einschluss, Fixierung und der entsprechenden Packungsdichte, beschrieben werden kann.

Ein grosser Teil der Dissertation trägt zu einem besseren Verständnis der räumlichen E. coli Chromosomorganisation und -segregation bei. Wir konnten zeigen, dass Sternpolymere mit zirkulären Armen, deren struktureller Aufbau dem des E. coli Chromosomes entspricht, einen Übergang von sphärischen zu flachen, gestreckten Strukturen durchlaufen, welcher es erleichtern könnte, das bakterielle Genom in einer gestreckten Huelle unterzubringen. Wir konnten den Nachweis liefern, dass die Kopplung von Chromosomentopologie und Einschluss relevant ist, um den Hang des bakterielle Chromosomes zu überwinden, sich zu mischen. Wir schlagen einen Mechanismus zur Bildung von Domänen im E. coli Chromosom vor und testen diesen: Der Prozess der Kolokalisation von Transkriptionsfaktoren und deren Zielgenen, der die durch das regulative Transkriptionsnetzwerk definierte Kontrolle imitiert, kann die experimentell beobachtete, präzise Positionierung der genetischen Segmente erklären. Der hier definierte Rahmen zur Chromosomenanordnung erlaubt es ausserdem, die Segregation von E. coli Chromosomenpaaren während der Zellteilung im Kontext von Volumenausschlusswechselwirkung, spezifischen Chromosomentopologien und geometrischen Einschränkungen zu verstehen.

Im Schulterschluss mit einer weiteren Kollaboration wurde eine Methode entwickelt, die die räumliche Chromosomenorganisation auflösen kann, indem sie den Chromatinstrang als “Worm-like Chain” modelliert und ein Ensemble von Chromatinstrukturen generiert, die Strukturrestriktionen basierend auf “Chromosome Conformation Capture”- oder Mikroskopiedaten konsistent abbilden.

Publications Related to this Thesis

Large parts of this thesis have already been published or are currently under peer-review. Papers in preparation are also listed. (Information as of August 15th, 2011)

- **M. Fritsche**, D.W. Heermann, M. Dutra and C.E. Cordeiro, Conformational and Dynamical Properties of the Isolated, Three-Dimensional Single- and Double-Tethered Polymer Chain on an Infinite Surface. *Macromolecular Theory and Simulations* **(2010)**, *19*, 440-448.

  DOI: 10.1002/mats.201000025

- **M. Fritsche** and D.W. Heermann, Temperature-dependent structural behavior of self-avoiding walks on three-dimensional Sierpinski sponges. *Physical Review E* **(2010)**, *81*, 051119.

  DOI: 10.1103/PhysRevE.81.051119

- **M. Fritsche** and D.W. Heermann, Confinement driven spatial organization of semiflexible ring polymers: Implications for biopolymer packaging . *Soft Matter* **(2011)**, *7*, 6906-6913.

  DOI: 10.1039/C1SM05445G

- **M. Fritsche**, S. Li, D.W. Heermann and P.A. Wiggins, Transcription factor induced DNA domain formation provides a natural framework for understanding the physical structure of the E. coli chromosome. **(2011)**, under peer-review.

- P. Reiß*, **M. Fritsche***, and D.W. Heermann, Looped star polymers show conformational transition from spherical to flat toroidal shapes. **(2011)**, under peer-review.

- **M. Fritsche**, R.B. Pandey, B.L. Farmer and D.W. Heermann, Multi-scale Structural Dynamics of a Histone H2AX: A Coarse-Grained Monte Carlo Approach Via Knowledge-based Interaction Potentials. **(2011)**, under peer-review.

- **M. Fritsche***, P.M. Diesinger*, D.W. Heermann and M. Bathe, A Maximum Entropy Approach for 3D Genome Conformation Reconstruction. **(2011)**, in preparation.

- **M. Fritsche**, L. Reinholdt, M. Lessard, M.A. Handel, J. Bewersdorf and D.W. Heermann, Entropy-Driven Spatial Organization of Synaptonemal Complexes within the Cell Nucleus. **(2011)**, in preparation.

- S. Gordon-Messer, B. Avşaroğlu, **M. Fritsche**, J. Martin, P.A. Wiggins, J. Sedat, D.W. Heermann, J. Haber and J. Kondev, Rabl model of yeast interphase chromosomes quantitatively accounts for their spatial organization and dynamics. **(2011)**, in preparation.

*Shared first authorship due to equal contribution

- Y. Liu, **M. Fritsche**, D.W. Heermann and B. Chakraborty, Segregation of Polymers in Confined Spaces. **(2011)**, in preparation.

Additional Publications

- R. Mendez*, **M. Fritsche***, M. Porto and U. Bastolla Mutation bias favors protein folding stability in the evolution of small populations. *PLoS Computational Biology* **(2010)**, *6*, e1000767.

  DOI: 10.1371/journal.pcbi.1000767

- **M. Fritsche**, E. Roman and M. Porto Temperature-dependent structural behavior of self-avoiding walks on Sierpinski carpets. *Physical Review E* **(2007)**, *76*, 061101.

  DOI: 10.1103/PhysRevE.76.061101

*Shared first authorship due to equal contribution

Contents

Acknowledgments

I would like to thank my advisor, Prof. Dieter W. Heermann for his enduring support throughout my doctoral studies. His advice on physics and computational modeling as well as on the art of effectively presenting research results have been invaluable in my scientific development. I am also very grateful to him for giving me the opportunity of spending two fruitful research stays at the Whitehead Institute of Biomedical Research and the Department of Biological Engineering, Massachusetts Institute of Technology (MIT), in Cambridge, USA.

I owe many thanks to my co-advisor, Prof. Christoph Cremer, for his knowledgeable and yet playful way of communicating science.

I would like to thank Prof. Paul A. Wiggins for his great hospitality during my two research stays in his group at the Whitehead Institute and for our very helpful and inspiring discussions on E. coli chromosome organization.

My special thanks go to Prof. Mark Bathe and Dr. Philipp M. Diesinger for generously hosting me at the Department of Biological Engineering at MIT. I am especially grateful to Dr. Philipp M. Diesinger's for his ongoing professional assistance and friendship.

I owe many thanks to Prof. Ras B. Pandey for inviting me to the University of Southern Mississippi, where we have shared many insightful discussions on protein modeling. His and his wife's amazing hospitality is much appreciated.

I would also like to thank Prof. Jané Kondev and Baris Avşaroğlu for the fascinating collaboration on yeast chromosome packaging, where I have learned a lot about polymer modeling at each one of our meetings.

Prof. Jörg Bewersdorf, Prof. Lindsay Shopland, Dr. Laura Rheinholdt and Mark Lessard are very much appreciated for their hospitality, for sharing their data with us as well as for their patience in explaining biology to me.

Many thanks go to Prof. Bulbul Chakraborty and Dr. Ya Liu for our ongoing collaboration on polymer segregation.

This work was funded by the Heidelberg Graduate School of Mathematical and Computational Methods for the Science (HGS MathComp) as well as the German Academic Exchange Service (DAAD).

Many thanks go to Cornelia Merkel, Melanie Steiert and Sonja Bartsch for being the heart and soul of Philosophenweg 19.

I would like to thank Gabriell Maté, Songling Li, Hans-Jörg Jerabek, Yang Zhang, Pascal

Reiß, Marcel Hellmann and Mischa Gerstenlauer for the excellent working atmosphere, their support and entertaining distraction. I am thankful to Dr. Manfred Bohn for his helpfulness when I joined the group three years ago as well as for sharing the latex template of his thesis with me.

I am grateful for the support of my dear friends Maî Zahran and Benoît Knecht and for the loyalty of my most trusted friend, Alexander Mergler. I am thankful to Dr. Gregor Obernosterer for his constant encouragement, affection and wisdom. I could not have done this without them.

Throughout my life, my family's love, encouragement, and understanding have supported and comforted me. I am deeply grateful for their ongoing devotion.

Chapter 1

Intention and Structure of this Thesis

A Short Overview

1.1 Introduction

The major goal of the human genome project completed in 2003 was to decipher the sequence of the 3×10^{12} base pairs that human DNA consists of [1]. However, eight years after the complete sequencing of the human genome it is becoming more and more evident that the pure knowledge about the linear ordering of its base pairs is not enough to understand its functioning.

"We finished the genome map, now we can't figure out how to fold it."

Figure 1.1: Illustration of the genome folding problem. Image adapted from cartoonist John Chase (`http://www.chasetoons.com/`).

Instead, many of the functional aspects of the genome are governed by its three-dimensional packaging, which involves the organizational challenge of ensuring both compactification and accessibility of the hereditary information: The genetic material, which is meters long, has to be tightly folded in order to fit into the limited space of a micrometer sized cell, while dynamic access to the genetic information stored in the DNA molecule is required for nuclear processes like transcription, replication, DNA repair and recombination.

Bacterial genomes face similar constraints as their eukaryotic counterparts. The Escherichia coli (E. coli) chromosome is about 1000 times longer than the confinement of the bacterial cell [2–4], and replication, segregation, and transcription/translation of the genetic material constitute an intricate organizational challenge [2]. Recent advances in cell-imaging technology have revealed that the E. coli genome exists in an extremely precise spatial ordering within the nucleoid [5] as opposed to the "bowl of spaghetti" configuration assumed in the past [2].

How does the three-dimensional (3D) architecture of bacterial and eukaryotic genomes influence vital cell functions and, in turn, what impact do nuclear processes have on genome structure [6]?

In order to address these questions a better understanding of 3D genome packaging and the underlying physical principles is imperative [6]. Increasingly, computational modeling coupled to high-resolution microscopy and/or chromosome conformation capture (CCC) techniques are used for this purpose. CCC techniques probe spatial genome folding by determining the relative contact frequency with which pairs of genomic loci are in direct physical contact, thus generating interaction frequency maps [7]. While such biochemical methods measure interaction frequencies between DNA segments for a population of cells, they are not able to provide information about the functional relevance of these complex (epi)genetic interactions [7]. Thus, high-resolution (live) cell imaging is pivotal for validating and quantifying many findings about spatial genome organization [7, 8].

Regarding eukaryotic cells, various experiments have revealed that there is a mutual interplay between 3D chromatin organization and transcriptional activity [9–12]. In fact, the 3D genome architecture, and in particular the formation of chromosomal loops, is supposed to facilitate the spatial colocalization of regulatory elements with their genomically distant target genes [9, 13].

Much like their eukaryotic counterpart, the organization of transcription within the prokaryotic nucleoid is expected to both depend on and determine the spatial packaging of the bacterial chromosome [5, 14–20]. Indeed, transcriptional regulation is carried out through a complex network of interactions among regulatory elements and their target sites [19–21]. In a recent approach, Llopis and colleagues discuss the use of FISH in Caulobacter crescentus and E. coli to demonstrate that mRNAs display limited dispersion from their site of transcription [22]. The high localization of mRNA implies that bacterial chromosome architecture might act as a spatial organizer which compartmentalizes the cell interior such that dedicated (regulatory) proteins are produced within those subcellular regions, where their regulatory intervention is needed [22]. However, in spite of considerable advances in elucidating the nature of the different players that determine nucleoid structure and its segregation, our knowledge of the mechanisms linking gene expression with a given nucleoid organization or driving nucleoid segregation is still rather limited.

This is a field in which computational/analytical modeling approaches may guide our intuition [2, 5, 23, 24]. Disregarding molecular details, coarse-grained models are particularly suited to provide an overall picture of the major driving forces and underlying organizational principles of biological systems. For this purpose, one might chose a "multi-level" approach, where each level of detail emerges from the previous, more coarse-grained one [6]. In a first step, the focus is on answering the important question that Erwin Schrödinger has already raised in 1944 [25]: "How can the events in space and time which take place within the spatial boundary of a living organism be accounted for by physics and chemistry?" For this purpose, the impact of key features which are present in a wide range of biophysical systems needs to be investigated and theoretically understood. Those key features are (i) polymer topology, (ii) confinement, (iii) semiflexibility as well as (iv) tethering.

(i) Ring polymers are increasingly stirring theoretical and experimental studies since they play an important role in many biological contexts where DNA is constrained [26, 27]: the segregation of the compacted circular E. coli genome or the storage of knotted viral DNA in a capsid [28, 29]. Additionally, the structure of the E. coli chromosome is more sophisticated than the simple ring form due to its organization into multiple chromosomal domains [2, 15, 30].

(ii) Confinement is omnipresent in biological systems where it is imposed by cell walls or membranes, the cell nucleus as well as bacterial or viral envelopes [9,31,32]. Indeed, spatial

constraints of different geometries affect polymer conformation, induce conformational transitions and, in combination with ring closure, have a strong effect on the statistics of biopolymers [31–33].

(iii) A key feature determining the conformation of biopolymers is their resistance against bending on specific length scales, i.e. their semiflexibility. A popular example of a semiflexible biopolymer is double-stranded DNA with a persistence length of about 50 nm [9]. In fact, it is the mutual interplay between bending energy and conformational entropy that drives the equilibrium form and the dynamics of semiflexible biopolymers and consequently influences their function [26, 27, 31–33].

(iv) Last but not least, protein-mediated anchors or tethering to the nuclear membrane, nuclear lamina, or to other cellular structures constrain the chromatin fiber in various biological systems such as in bacteria [34–36], yeast [37–39], Drosophila melanogaster [40, 41], and mammalian cells [10, 42, 43].

While the investigation of the impact of such basic features is important from a theoretical point of view as well as a way of qualitatively understanding biological systems of larger complexity, a subsequent level of detail regards coarse-grained models which incorporate more specific aspects of experimental evidence to provide a quantitative approach.

For this purpose, CCC studies coupled to modeling approaches have become a thriving field in recent years. Computer modeling has allowed for the generation of chromatin conformations consistent with contact frequency constraints measured by CCC methods. Such a combined effort was applied to study chromosome conformation signatures of cellular differentiation of the human Homeobox (Hox) A cluster [44]. Regarding bacterial chromosome packaging, a huge amount of experimental evidence on various aspects of spatial chromosome organization or segregation has inspired several computational models of nucleoid organization, which focus either on the folding of the circular chromosome [5, 45] or on its segregation during cell division [23, 24]. However, there is so far no integrative model which is able to connect bacterial chromosome packaging with chromosome segregation.

A truly ingenious, integrative approach combining state-of-the-art technology, such as high-resolution microscopy and CCC techniques, with computational/analytical modeling and polymer theory is needed to deepen our current understanding on the coupling between spatial genome organization and genome functioning in eukaryotic and bacterial cells.

1.2 Structure of this Thesis

The work presented in this thesis is inspired by a wide range of biological systems. Even though they display different dynamics, length scales, organizational complexity as well as functional purpose, all of the complex systems studied here “can be accounted for by physics and chemistry” [25] within a proper explanatory framework. The aim of this thesis is to develop precise models to confirm and understand current experimental evidence. The predictive power of the developed frameworks is expected to stimulate future experiments to challenge our present knowledge on structure-function relationships ranging from bacterial to eukaryotic cells.

- Since the work presented in this thesis is of an interdisciplinary nature, the next two chapters provide an introduction to selected aspects in biology and physics. In **chapter 2**, the basic biology of eukaryotic and prokaryotic cells and, in particular, the biology of the spatial packaging of the genetic information is discussed. Addi-

tionally, the two major experimental methods probing the 3D genome are presented and confronted to each other.

- In **chapter 3**, basic concepts on polymer physics known for the last decades are discussed. The *freely-jointed*, the *Gaussian* and the *worm-like chain* model are presented. Measures describing the properties of polymers are introduced and will come in handy in the other chapters. Eventually, the physics of polymer chains with excluded volume interactions is developed and a short overview of the principles underlying *universality* and coarse-graining is given. The second major part of this chapter introduces the mathematical principles of *Markov chain Monte Carlo* and presents the two algorithms for generating polymer conformations used in the following chapters.

- In **chapter 4**, the conformational and dynamical properties of a polymer chain, which is single- or double-tethered to an impenetrable surface, are investigated as a basic model for tethering in various biological or physical systems. Prominent examples are polymer chains grafted to colloidal particles or protein-mediated anchors or tethering of the chromatin fiber to the nuclear membrane, nuclear lamina, or to other cellular structures [2, 46, 47].

- The other key features mentioned above, which are present in a wide range of biophysical contexts, namely ring closure, confinement and semiflexibility, are addressed in **chapter 5**. The conformational properties of a semiflexible ring polymer in confined spaces are investigated. Taking into account the competing interplay between configurational entropy, bending energy and excluded volume, the role that different geometrical constraints can play in shaping the spatial organization of biopolymers is elucidated.

- After having studied the impact of these basic underlying features by applying simplified toy models, we proceed with more detailed modeling approaches. In **chapter 6**, we study the spatial response dynamics of the histone H2AX by a coarse-grained bond fluctuating model for a broad range of normalized temperatures. This histone protein is not only important due to its vital role in cellular processes such as DNA compaction, replication and DNA repair but also shows intriguing structural properties that might be exploited for bioengineering purposes such as the development of nano-materials.

- Sexually reproducing organisms employ a specialized cell division cycle called *meiosis* to produce *haploid gametes* from *diploid* nuclei [48]. This process is accomplished by first pairing *homologous chromosomes*, recombining and subsequently segregating them from each other [49, 50]. Prior to their segregation away from each other, homologous chromosomes pair along their entire length and are held together by a proteinaceous structure known as the *synaptonemal complex* (SC) [48–52]. In **chapter 7**, 4Pi-microscopy of synaptonemal complexes (SCs) in mouse spermatocyte nuclei coupled to computational modeling of SCs in the cell nucleus is presented with the purpose of better understanding the mechanisms driving SC organization at the *pachytene stage* during meiosis.

- The *Rabl model* of chromosome organization is tested quantitatively in yeast interphase cells in **chapter 8**. Key features of this model are the tethering of the *centromeres* at the *spindle pole body* and the localization of *telomeres* to the nuclear

periphery. Using a polymer model to account for the flexibility of yeast chromatin and taking into account the constraints assumed by the Rabl model we compute a number of different quantities associated with the spatial organization and dynamics of yeast chromosomes and compare them to experiments on fluorescently labeled chromosomes.

- A rather large part of this thesis is dedicated to the spatial packaging of the E. coli chromosome within the nucleoid as well as to E. coli chromosome segregation during cell division. In three chapters, the fundamental question of the physical mechanism which leads to organization and segregation of a highly-condensed, and confined circular chromosome is addressed. In **chapter 9**, we are inspired by the topological organization of the circular E. coli chromosome, which is compacted into separate chromosomal domains. Thus, we study a similar unconfined polymer architecture in order to elucidate the role of topology in shaping E. coli chromosome packaging motifs.

- However, the E. coli chromosome is not located in free space but it is strongly confined within the bacterial nucleoid. Thus, in the subsequent **chapter 10**, we investigate the interplay between the specific topology underlying the E. coli chromosome and the geometrical constraints imposed by the envelope of the confining cavity. In particular, a mechanism responsible for the formation of the chromosomal domains is introduced and and its implications discussed.

- After having established a physical framework for E. coli chromosome packaging, in **chapter 11**, we study how excluded volume effects, specific polymer topologies and geometrical confinement compete with entropy to drive dynamical processes such as the segregation of highly compacted chromosomes during cell division.

- **Chapter 12** introduces a Monte Carlo-based computational tool that inverts CCC data to provide the (unique) maximum entropy ensemble of conformations consistent with given contact frequency constraints. The underlying idea is to probe genome architecture by accounting for the complex (epi)genetic interactions that contribute to its folding pattern in a modeling approach.

- Last but not least, **chapter 13** presents a concise summary of all research projects and the obtained results, closing with a discussion of future challenges.

Chapter 2

Introduction to the Biology of Eukaryotic and Bacterial Cells

In this chapter, the basic biology of eukaryotic and prokaryotic cells and, in particular, the biology of the spatial packaging of the genetic information is discussed. Additionally, the two major experimental methods probing the three-dimensional (3D) genome are presented and confronted with each other. For a detailed review of the molecular biology of the cell, the reader is referred to Refs. [9,53,54], whereas a reader already familiar with these topics might want to skip this chapter altogether.

2.1 Cells as Building Blocks of Life

The simplest definition of a living organism names three general features: metabolism, growth and reproduction [54]. By this definition, the *cell* is the functional basic unit of life which is why it can rightly be called the "building block of life" [9].

There are two general classes of cells: *eukaryotic* and *prokaryotic* ones. While eukaryotic cells are usually part of multicellular organisms, prokaryotic cells are independent. Prokaryotes can, in turn, be divided into two main *phyla*, *bacteria* and *archaea*. We will focus on bacteria since they become of importance in the next chapters. In the following section, generic features are described which, however, need not be present in all bacterial or all eukaryotic cells. Fig. 2.1 illustrates such a general setup of eukaryotic and prokaryotic cells.

A *cell* or *plasma membrane* is a selective barrier found in all cells. It allows certain chemicals (water, proteins, nutrients, waste material etc.) to pass or to be retained, thus maintaining the specific chemical equilibrium inside the cell [9, 53, 54]. The *cell wall* is a structure surrounding the plasma membrane and gives rigidity to a cell. Some bacteria have an additional covering layer called a *capsule*. On the outside of the bacterial cell, *flagella* and *pili* protrude from its surface with the scope to facilitate movement and communication between cells. Eukaryotic cells are further stabilized by a scaffold called *cytoskeleton*, which maintains cellular shape, anchors organelles and supports cell growth and separation of daughter chromosomes during cell division [54].

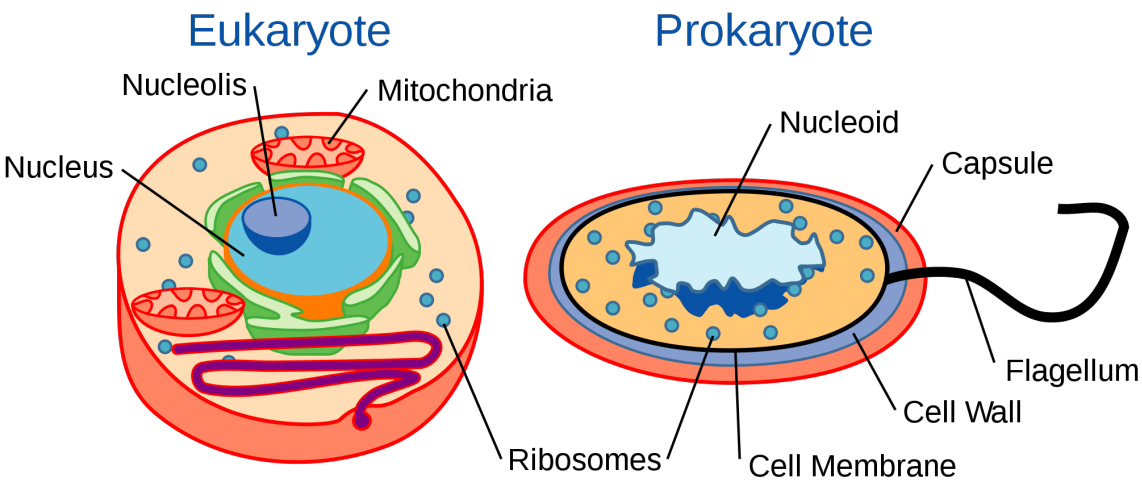


Figure 2.1: The general setup of eukaryotic and prokaryotic cells. Figure adapted from [55].

Inside the cell is the *cytoplasm* within which many metabolic cycles and protein synthesis take place. A significant difference between eukaryotes and prokaryotes is that eukaryotes, besides having a nucleus, contain membrane-bound organelles, while prokaryotes do not. Organelles perform dedicated cell functions such as energy generation (*mitchondria*). For a detailed discussion of their functioning the reader is referred to Refs. [9, 53, 54]. *Ribosomes* occur in both eukaryotic and bacterial cells, do not have a membrane and are responsible for protein synthesis which will be discussed later on.

Hereditary material, in the form of a double-stranded *DNA* or *RNA* polymer is needed for all cells to store information about all vital cell functions from basic tasks, such as nutrient and energy uptake, to complex coordinated ones, such as cell division. Prokaryotic DNA is organized as a *circular chromosome* which is packaged in the cytoplasm into a distinct structure known as the *bacterial nucleoid*. Additionally, bacteria can carry extrachromosomal DNA elements called *plasmids*, which are usually circular, too. Eukaryotic DNA is organized in linear structures, the *eukaryotic chromosomes*, within the nuclear envelope, a double layer membrane. The whole compartment is referred to as the *cell nucleus*.

The most important function of DNA is that it carries *genes*, which encode the instructions for producing *proteins*. Proteins are, in turn, known to be the "chief actors within the cell" carrying out the duties specified by the information encoded in genes [53]. The DNA message is spelled out by the four letter alphabet of the *bases* adenine (A), cytosine (C), guanine (G) and thymine (T). It is the genetic code which bridges the gap between the linear sequence of *nucleotides* (triplets of bases) in a gene and the linear sequence of *amino acids* that a protein consists of. The process of reading the information stored in a particular stretch of DNA is called *transcription* and results in the production of a complementary RNA copy, *messenger RNA* (mRNA). In a subsequent step, the mRNA in eukaryotes diffuses through the nuclear membrane, and is used by ribosomes in the cytoplasm as a template to assemble the specific sequence of amino acids that the protein is made of. Unlike eukaryotic cells, bacteria do not have a nucleus that separates DNA from ribosomes, which means that translation and transcription occur in spatial proximity.

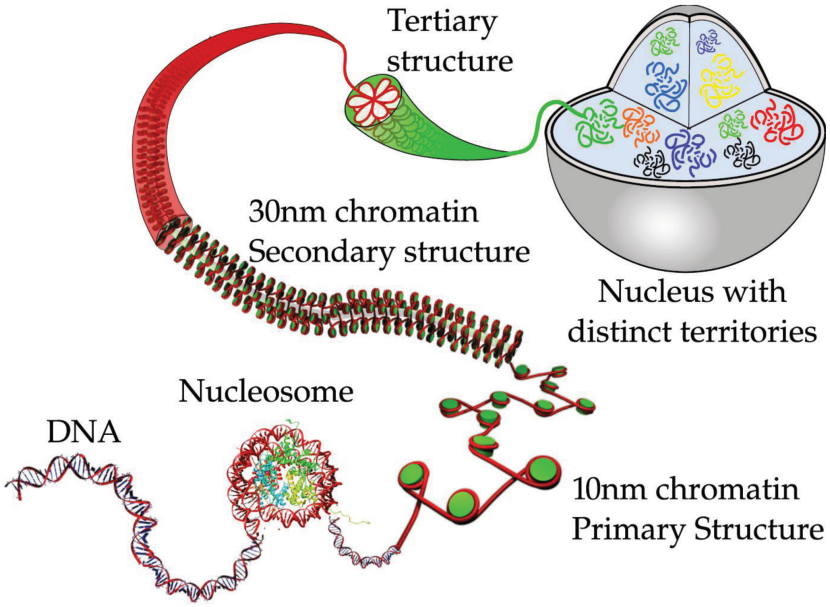


Figure 2.2: Compaction stages of the genetic information in the eukaryotic nucleus. The DNA molecule folds around histones to from a "beads-on-a-string" fiber of approximately 11 nm. An additional level of compaction consists in the formation of the much discussed 30 nm fiber. Higher-order chromatin folding motifs are driven by loop formation, which are assumed to drive the formation of chromosome territories [56, 57]. Image adapted from [6].

2.2 Organization of Eukaryotic Chromosomes in the Cell Nucleus

The genomes of eukaryotic organisms are packaged into chromosomes, which are highly dynamic entities. Nevertheless, one can still identify a hierarchical process by which the double-stranded DNA is compacted within the confined space of the cell as illustrated in Fig. 2.2.

2.2.1 Cell Cycle Dependent Chromosome Packaging

The cell spends most of its time in *interphase*, which is marked by an increased size of the nucleus as well as gene expression, protein synthesis and DNA replication. Consequently, interphase chromosomes are less tightly packed (10 times less) than mitotic chromosomes in the subsequent *M phase* [9]. During M phase, *mitosis* takes place and the nucleus is divided into two daughter nuclei [9]. For this purpose, the mitotic spindle captures the condensed, replicated chromosomes and then pulls each set of chromosomes to opposite ends of the cell, where a nuclear envelope reforms around them [9]. Finally, cell division takes place and results in two separate two daughter cells [9]. The most condensed higher-order DNA structure is observed during mitosis, forming the well-known four-arm structure revealed by light microscopy studies [9, 53, 54, 58], where the centromere joins the two duplicated chromosomes. Such a compact organization is needed in order to facilitate faithful segregation of daughter chromosomes [6, 9, 58].

Hierarchy of Chromosome Packaging

The first level of chromosomal DNA compaction is DNA wrapped around a protein core formed by *histone octamers* as shown in Fig. 2.2. This combined histone octamer-DNA complex is called the *nucleosome* [6] and is interconnected by linker DNAs forming a rather loose 11 nm "beads-on-a-string" *chromatin fiber* illustrated in Fig. 2.2.

The next step of compaction, i.e. the structural organization of the notorious *30 nm chromatin fiber* is a subject of ongoing research. Electron microscopy of *in vitro* isolated nucleosomal arrays show the transition from 11 nm "beads on a string" structures to 30 nm condensed fibers with increasing salt concentrations [6, 59, 60]. The packaging of nucleosomal chromatin into a transcriptionally silent 30 nm fiber is achieved by the incorporation of *linker histone* H1 during interphase [58]. However, the details of the way the nucleosomes are arranged on top of each other to from the 30 nm fiber are under debate. Indirect ways of deducing the 3D chromatin architecture, using single-molecule pulling techniques [61] have inspired several different structural models : zigzag ribbon models [62–66] and helical solenoid models [60,67,68]. Moreover, other authors claim that there simply exists no regular 30 nm fiber at all [69, 70], while a recent work proposes that missing linker histones and nucleosomes lead to an "irregular", less dense 30 nm structure [71, 72]. In general, such a compact chromatin fiber is constantly subjected to remodeling activities (*histone modifications* [73, 74], *sliding*, *depletion of nucleosomes* [75, 76] and incorporation of histone variants [77]) which modify the accessibility of the DNA sequence for gene expression, DNA replication and repair.

Even if it were present and despite its structural uncertainties, the 30 nm fiber does not sufficiently compact the DNA polymer to fit into the confined space of the cell [9,53,54,58]. Thus, more levels of chromatin folding above 30 nm have to exist within interphase chromosomes [6,9,58]. While the very condensed higher-order structure of chromatin observed for mitotic chromosomes disappears during interphase, there is strong evidence from light microscopy studies that the interphase chromatin fiber is organized into *loops* [12,78]. Looping constitutes the organizational principle for colocalizing distant enhancer and promoter regions necessary for gene activation, regulation and recombination [9,10,56,57,79–83]. Additionally, anchoring the chromatin fiber to the nuclear periphery via chromatin-associated proteins such as nuclear lamins might be an additional mechanism for generating larger looped chromatin domains (300-700 nm) [10, 58].

At an even higher-order organizational level labeling techniques like *fluorescent in situ hybridization* (FISH) [11,84] show that nuclear architecture is compartmentalized by interchromatin compartments, while *whole chromosome painting* reveals distinct chromosome territories [9, 11, 85]. Fig. 2.2 schematically illustrates these territories, while Fig. 2.3 displays FISH labeling of all 24 human chromosomes. Computational modeling of confined coarse-grained polymers coupled to light-microscopy and chromosome conformation capture experiments identify looped structures [12,82,86] and/or confined fractal organization [10,87,88] as basic organizational principles driving the observed nuclear architecture.

2.2.2 The Interplay between Gene Regulation and Genome Folding

Due to its mutual coupling, genome folding cannot be addressed without discussing its interplay with *gene expression*. Genes have to be expressed differentially since their products serve dedicated cellular purposes at specific times during the cell cycle. While only approximately 1.5% of the human genome encodes for proteins and structural/catalytic RNAs [9], non-coding regions contain lots of regulatory elements [9]. Notably, gene ex-

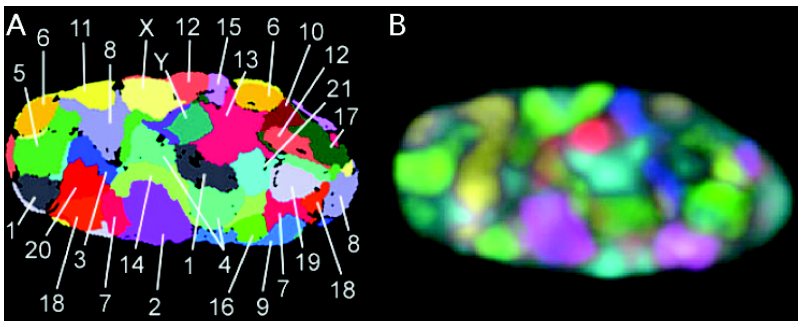


Figure 2.3: (A) FISH labeling of all 24 human chromosomes which are organized into distinct territories in a fibroblast nucleus. (B) False color representation of chromosome territories after computer classification. Image adapted from [8].

pression can be regulated at many steps in the pathway from DNA to RNA to protein. However, since there is increasing evidence that genome folding constitutes one mechanism that regulates the expression of genes we focus on the initiation of transcription as the most important point of control [9]. Fig. 2.4 illustrates the interplay between the major players of transcriptional regulation: *promoters, RNA polymerase, enhancers, silencers, insulators* and *general transcription factors*.

Transcription of a gene is initiated when an enzyme called *RNA polymerase* binds to the DNA upstream of that gene at a specialized sequence called a *promoter* [9, 53]. In contrast to bacteria, three different types of RNA polymerase exist in eukaryotic cells, which can be distinguished by whether they translate protein molecules or RNAs [9, 53]. It is RNA polymerase II that transcribes all protein-coding genes, which have a *TATA box* (consensus sequence TATTAA) that indicates the start point of transcription [9, 53].

In addition to RNA polymerase, a set of sequence-specific DNA-binding proteins called *general transcription factors* need to be assembled at the promoter for transcription to begin [9]. To this end, gene regulatory proteins, *activators* or *repressors*, come into play. Activator proteins bind to DNA sites, called *enhancers*, which can be located up to several mega base pais (Mbps) away from the genes they regulate. The DNA between the enhancer and the promoter loops out such that the activator proteins bound to the enhancer colocalize (form a contact) with proteins bound to the promoter [9, 53]. By the same looping mechanism, repressor proteins silence or downregulate genes by binding to silencer regions.

In fact, the main function of regulatory proteins is to position and modify the RNA polymerase and the general transcription factors at the promoter as well as to facilitate or block DNA accessibility by changing the local chromatin structure (histone modifications and nucleosome remodeling) around the promoter [9, 53].

Since gene regulatory proteins can act across vast genomic regions, an "insulating" mechanism is needed that prevents different control regions from interfering with each other. *Insulator* elements are DNA sequences that protect domains of gene expression or block the activity of enhancers and silencers depending on the insulators' location with respect to the other players of transcriptional control [9, 53, 89, 90]. Evidence suggests that insulators cluster together to form chromosomal loops, which might establish and segregate (functionally) different chromosomal domains [89–91]. These insulator clusters are thought to colocalize through the interaction of proteins such as the insulator protein CCCTC-binding factor CTCF [90,91]. Interestingly, a recent work suggests that chromatin

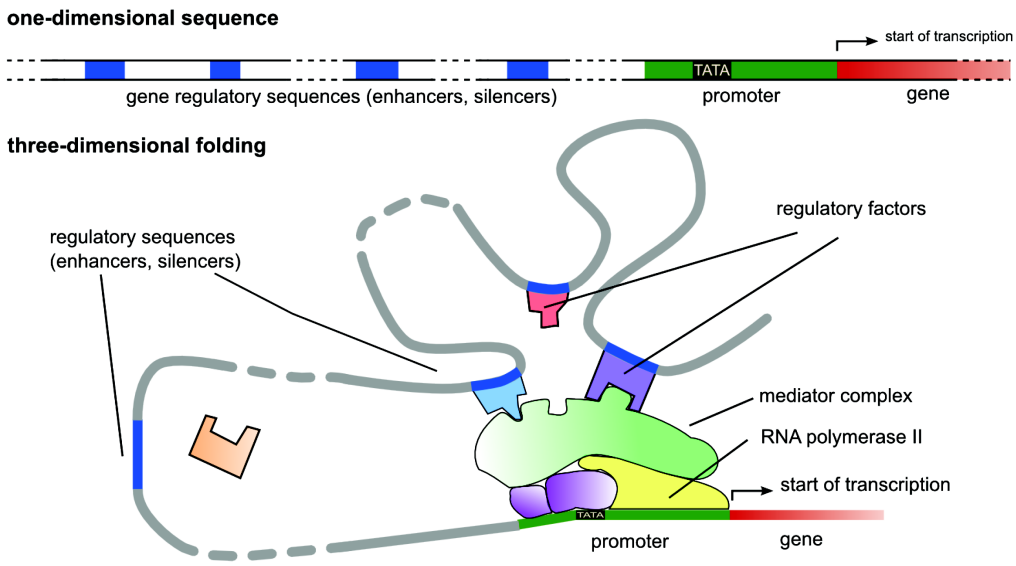


Figure 2.4: Schematical illustration of transcriptional regulation. Gene transcription factors and polymerase assemble at the promoter sequence. The regulatory sequences serve as binding sites for the gene regulatory proteins. These proteins interaction with the transcription machinery at the promoter by DNA looping. Image adapted from [9] and [92].

barrier activity of CTCF is instrumental to the coordinated establishment of chromatin structure, higher-order architecture, and developmental expression of the homeobox gene A (HoxA) locus [91] which is investigated in chapter 12.

A further control of transcriptional activity might emerge due to the formation of *"transcription factories"*, which are hubs of spatially proximate co-regulated genes, RNA polymerases and transcription factors [7, 93, 94]. The idea is that transcription of genes relying on the same transcription factors and cofactors could be performed more efficiently if all players of the transcription process were preassembled at high local concentrations in dedicated compartments [7]. This assumption is supported by the observation that RNA polymerase II accumulates in certain foci of high concentration in the nucleoplasm [7, 83, 95].

Regarding the organization of nuclear architecture both on the level of subchromosomal compartments as well as on the level of chromosome territories some interesting observations can be made [93]: (i) Territories of transcriptionally active chromosomes are usually located in the body of the nucleus [82,93,96,97]. (ii) Transcriptionally active genes are positioned at the periphery of the chromosomal territories they belong to [93,98–100]. (iii) Transcriptionally silent regions interact more frequently among themselves than with transcriptionally active regions [87, 93, 101].

Last but not least, the *nuclear lamina* has been shown to interact with chromatin such that loci at this interface show lower expression levels [7, 10, 102]. In contrast, chromatin interacting with components of the *nuclear pore complex* tends to be more actively transcribed [7, 10, 103].

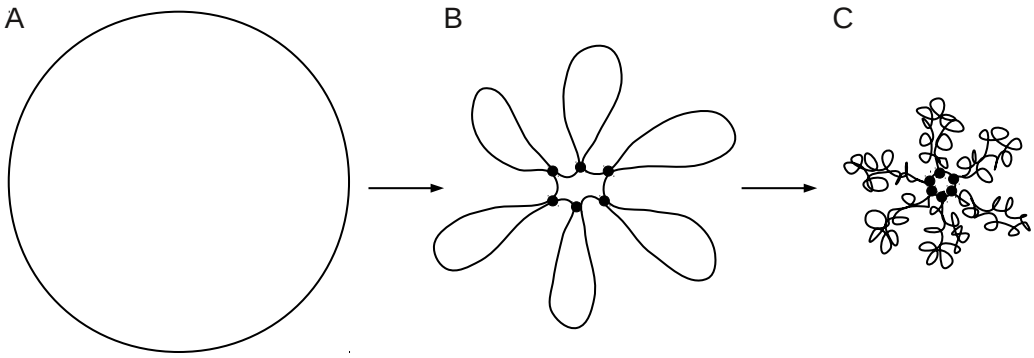


Figure 2.5: (A) The unfolded, circular chromosome of E. coli depicted as a single line for simplicity. (B) The DNA is folded into chromosomal domains by protein-DNA associations. The proteins are depicted as the black circles, interacting with both the DNA and with each other. (C) Supercoiling and other interactions cause the chromosome to further compact. Image adapted from [108].

2.3 Organization of the Bacterial Nucleoid

In contrast to eukaryotes, bacteria do not contain a separate membrane compartment, in which their DNA is stored. Instead, the genetic material is packaged into a pseudo-compartment, called nucleoid, which floats in the cytoplasm [2, 9, 104].

2.3.1 Hierarchy of Bacterial Chromosome Packaging

While eukaryotes wrap their DNA around histone proteins for DNA compaction, bacteria do not utilize histones to package their genomes [9]. Instead, the compaction of the bacterial chromosome involves *DNA binding proteins* that help to form initial loops followed by coiling and subsequent *supercoiling* of DNA as is illustrated in Fig. 2.5. The DNA of most bacteria is twisted in the opposite direction of the double helix, leading to negative supercoiling [2], whose loss is lethal [105].

The enzymes *DNA gyrase* and *DNA topoisomerase I* act together with the so-called *nucleoid-associated proteins* (NAPs) to maintain the vital amount of DNA supercoiling [106]. In particular, the HU protein (histone-like protein) works together with topoisomerase I to bend the DNA polymer [9], which is then wrapped around tetramers of HU [107]. Additionally, NAPs are supposed to be in charge of chromosomal remodeling tasks and hence have a functional association with the regulation of gene expression [15, 16, 106]. These regulatory NAPs are: Fis (factor for inversion stimulation), HU (histone-like protein), H-NS (histone-like nucleoid structuring protein), and IHF (integration host factor) [106].

The stabilizing interplay of multiple DNA binding proteins helps to divide the chromosome into topologically isolated loops or *chromosomal domains*. According to [30], a domain is defined as a region which is relaxed when DNA is cut [30]. Even though NAPs are expected to contribute to the formation of structures at larger scales, such as topologically isolated supercoiled domains, they are abundant throughout the nucleoid and often bind DNA with little sequence specificity [15]. Thus, a recent work speculates that their direct impact on DNA topology might be limited to the nano meter scale [15].

Bacterial chromosomes are suggested to be compartmentalized into four large organi-

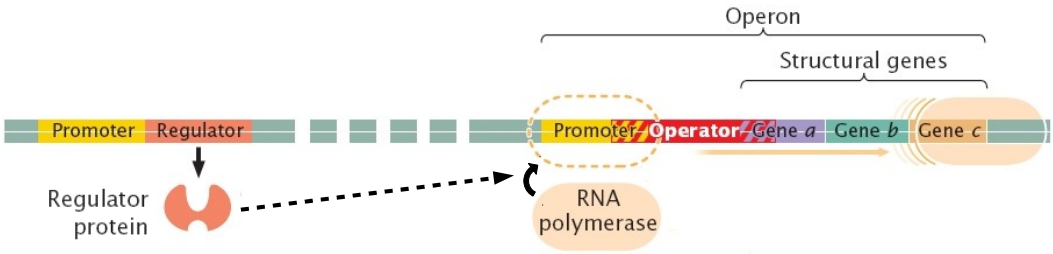


Figure 2.6: An operon is a single transcriptional unit that includes a series of structural genes, a promoter, and an operator. The operon is regulated by regulatory proteins that can be encoded dispersedly on the genome. Image adapted from [117].

zational units at the μm-scale, so-called macrodomains [109]. On the other hand, different average chromosomal domain sizes in the range between 10 kbp [30] and 100 kbp [110,111] have been reported, too.

Several drivers responsible for the observed reliable orientation and high level of organization of the E. coli chromosome have been suggested, including intranucleoid interactions such as (i) macromolecular crowding [112], (ii) DNA supercoiling [30] or (iii) protein-DNA interactions [15, 16, 113] as well as explicit mechanisms of external positioning such as (iv) cellular confinement [5, 33] or (v) tethering of the chromosome [114, 115]. However, consensus mechanisms that give rise to chromosomal domains are yet to be unveiled and will be discussed in more detail in chapter 10.

2.3.2 The Interplay between Gene Regulation and Genome Folding

Much like the coupling between spatial genome organization and gene regulation in eukaryotes, bacterial chromosome packaging affects gene expression [15, 30, 106, 109, 116].

Bacterial genes are organized into *operons* which are clusters of co-regulated genes [9]. Fig. 2.6 shows the topological setup of an operon. Within the organizational unit of an operon, the genes are not only physically grouped together, but they are also regulated such that all of them are either silenced or turn on at the same time [9]. Notably, these genes are transcribed into a single mRNA molecule and only subsequently translated into individual proteins [9].

As in case of eukaryotic transcriptional control, an operon contains a promoter region, which is recognized by the transcriptional machinery [9]. Within the promoter lies an additional regulatory sequence called *operator* [9]. An operator sequence can be recognized by either a gene repressor or activator protein (transcription factor, TF). In case of *negative control*, the repressor binds to the operator in order to block access to the promoter by RNA polymerase and thus prevent gene expression [9]. The reverse case, i.e. *positive control*, activates gene expression by the binding of a gene activator protein to the operator [9]. *Genetic switches* combining both positive and negative control exist, too [9]. A prominent example is the Lac operon in E. coli [9].

However, while operons can control a small amount of co-regulated genes, it is an insufficient organizational unit for the coordinated expression of a large number of genes. To this end, a second level of transcriptional control is the *regulon* [9]. It is a set of operons/genes that are co-regulated by the same specific regulatory protein (TF) which can be dispersed over the chromosome [9].

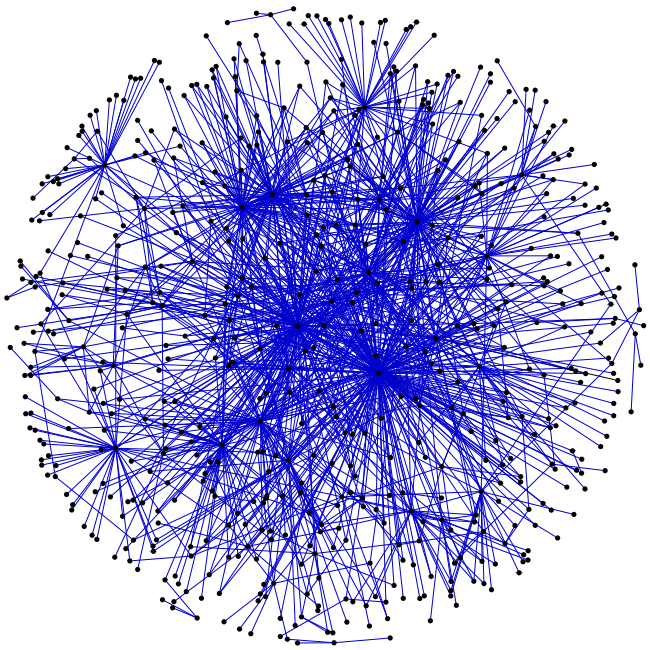


Figure 2.7: Graphical representation of the E. coli gene regulatory network. Black nodes represent coarse-grained genes and lines represent regulatory interactions between them. Global regulators are those TF genes that regulate lots of target genes, while other regulatory proteins are local, dedicated regulators. The network was generated using data from RegulonDB (`http://regulondb.ccg.unam.mx`) and is applied in chapter 10. Image adapted from [120].

2.3.3 Transcriptional Regulatory Networks

The decision about gene expression or repression is controlled by TFs which use metabolic or environmental signals to trigger a transcriptional response [3, 9, 106]. A functional network of regulatory interactions between TFs and target genes, which can themselves be TFs, is formed in each organism [3,21,118,119]. The transcriptional regulatory network of E. coli is illustrated in Fig. 2.7 and will come in handy in chapter 10.

A critical step of transcriptional control is the concise targeting of the operator region by regulatory proteins [18, 21, 106]. While it was assumed that DNA-binding site localization can be driven by Brownian diffusion, this explanation was dismissed when it was found that the LacI repressor in E. coli finds its DNA-targets 90-100 times faster than predicted by a purely diffusive mechanism [21, 121]. This observation has inspired the hypothesis of a "facilitated TF diffusion" mechanism consisting of alternating 3D jumps between DNA-strands and 1D sliding along the DNA [21, 122–124]. Janga and colleagues suggest [21] that local TFs governing small regulons are located genomically close to their regulated genes on the chromosome, which helps the newly synthesized TF to target its recognition site through the so-called sliding and hopping mechanism [21, 121, 123]. In agreement with this assumption, local regulators are normally found to be expressed in low cellular concentrations [21]. In contrast to local regulators, global regulators regulate a large number of genes and hence are distantly located on the genome with respect to their target sites. In this case, Janga et al. propose that targeting DNA could be achieved by maintaining large cellular concentrations of TFs coupled to the sliding and hopping mechanism mentioned before [21].

However, this picture neglects a further level for gene regulation in bacteria, namely the way the DNA molecule is packaged within the nucleoid [16–18]. The idea is that the underlying topology of the bacterial chromosome (chromosomal domains) which is fine-tuned by the dedicated action of NAPs (and probably other proteins such as MatP) facilitates access to the bacterial DNA for the activity of regulatory proteins [18, 19, 120]. In fact, the advantage of spatial gene clusters in bacterial chromosomes with respect to gene co-transcription and functional coupling has often been stressed [22, 125]. Indeed, immunofluorescence localization of transcriptional regulators has revealed foci of regulatory proteins and their DNA binding sites in actively transcribing E. coli cells [14]. Furthermore, structural and biochemical approaches suggest that the E. coli genome is organized into a nucleoid-filament type of structure compacted by the formation of (regulatory) chromosomal domains [5, 14]. Thus, regulation mediated by topological modifications of the bacterial chromosome can be thought of as an "analog control" and complements the "digital control" mediated by dedicated transcription factors [18, 19].

In a recent approach, Llopis and colleagues discuss the use of FISH in Caulobacter crescentus and E. coli to demonstrate that mRNAs display limited dispersion from their site of transcription [22]. The high localization of mRNA implies that chromosome architecture might act as a spatial organizer which compartmentalizes the cell interior such that dedicated (regulatory) proteins are produced within those subcellular regions, where their regulatory intervention is needed [22]. By a similar argument, we investigate the effect of TF-mediated domain formation on the packaging of the E. coli chromosome in chapter 10.

2.4 Experimental Methods for Unraveling the 3D Genome

Only the basic principles of chromosome conformation capture and cell imaging techniques for the investigation of genome architecture are discussed. At the end of the chapter, the case is made for an integrative approach combining both technologies in order to shed light onto the relationship between genome folding and function. The reader is referred to Refs. [7, 9, 126] for a more detailed review on both approaches.

2.4.1 Chromosome Conformation Capture Technologies

Chromosome conformation capture (CCC) technologies are able to resolve chromatin folding motifs in the native cellular state at a resolution beyond the current standard of microscopy techniques [10, 87, 87, 127, 128]. The spatial folding of chromatin is probed by determining the relative frequency with which pairs of genomic loci come in spatial proximity, thus generating interaction frequency maps. While the 3C technology probes only small genomic regions (6-600 kbp) [127], several detection methods including chromosome conformation capture on-chip [129] or circular chromosome conformation capture [130] (4C), chromosome conformation capture carbon copy (5C) [127] and Hi-C technology [87] have been developed in an effort to map chromatin interactions at a genome-wide scale. Application of the Hi-C method to the human genome has enabled the construction of interaction frequency maps at a resolution of 0.1-1 Mbp [87].

In principle, all CCC technologies consist of four consecutive steps, which are then followed by method-specific processing, as illustrated in Fig. 2.8. First, the specimen is fixed with formaldehyde resulting in fixed DNA-protein and protein-protein complexes (covalently bound). After cross-linking, a restriction digest (enzyme) separates non-cross-linked DNA fragments from cross-linked ones. Subsequent isolation and dilution of sample

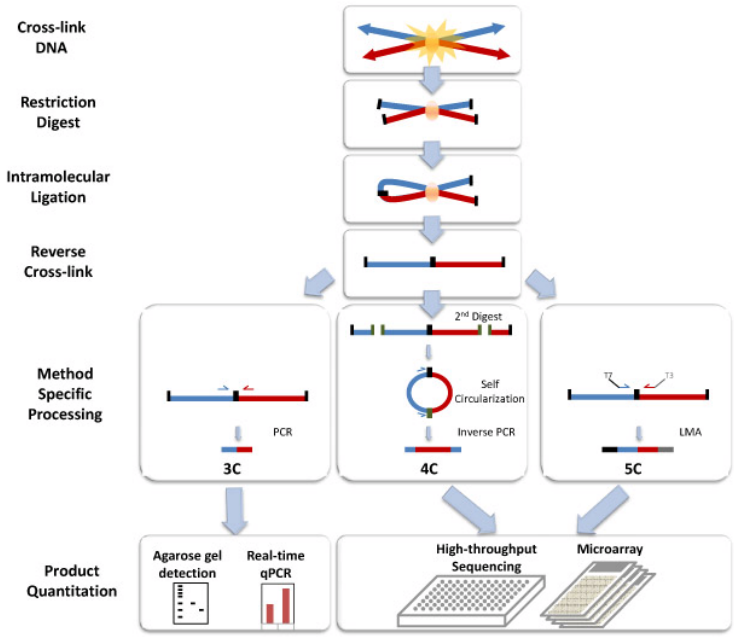


Figure 2.8: Cross-linking of DNA segments, restriction digest, ligation of cross-linked DNA ends and reverse cross-linking are common steps to all CCC technologies. Quantification of specific ligation events is performed by quantitative polymerase chain reaction (qPCR) for 3C [131], microarray analysis for 4C/5C [127, 129] or massive parallel sequencing for Hi-C [87]. Image adapted from [132].

cells is followed by ligation of cross-linked DNA ends. Eventually, cross-linking is reversed, which results in a set, called library, of linear DNA fragments with specific restriction ends and a central restriction site corresponding to the site of ligation. Quantification of specific ligation events is performed by quantitative polymerase chain reaction (qPCR) for 3C [131], microarray analysis for 4C/5C [127, 129] or massive parallel sequencing for Hi-C [87].

2.4.2 Cell Imaging Techniques

While current biochemical chromatin folding methods measure the interaction frequencies between DNA segments for a population of cells, these techniques do not provide information about the functional relevance of these (epi)genetic interactions [7]. Labeling and fixation techniques that are as little invasive as possible coupled to advances in microscopy development have provided insights into the functional nuclear and cellular architecture

While conventional light microscopy is limited to a resolution of 200 nm, making it impossible to track the path of the chromatin fiber inside the nucleus, localization light microscopy offers a resolution on the scale of a single histone [11] and, hence, is able to spatially and temporally resolve the specific positioning of nuclear players with respect to each other [7].

The use of DNA and RNA FISH reveals interphase chromosomes as a whole ("whole-chromosome painting" [8]), the genomic and spatial distance relationship between smaller DNA segments [12] or specific transcription events [7]. FISH uses fluorescent sequence-specific DNA-binding probes which fuse to DNA segments with high sequence similarity

and which can subsequently be tracked by fluorescence microscopy. FISH has successfully been applied to living cells to establish a relationship between the genomic distance of FISH markers and their physical distance in various eukaryotes [12]. Bolzer and colleagues were the first to use 3D-FISH (allowing for quantitative 3D positional mapping) to simultaneously detect all chromosomes in single diploid human fibroblasts in interphase and in prometaphase rosettes [8]. Other labeling technologies such as Fluorescent Repressor Operator System (FROS) and the ParB-parS system have been applied to investigate nucleoid structure in bacterial live cells [5] allowing genomic sites to be visualized by the aggregation of spectrally distinct fluorescently labeled proteins at the loci of interest.

Compared to light microscopy, the most immediate advantage of electron microscopy is its higher resolution in the range of 0.5 nm [126], while its most obvious disadvantage is the restriction to fixed, i.e. dead cells. However, the careful preparation of biological specimen for electron microscopy has helped to resolve the fine structures of cells such as the spindle pole body [9] or the nuclear pore complex [9]. Additionally, electron microscopy of isolated polynucleosomes has revealed an open "beads on a string" conformation in low salt concentration [59] in agreement with the first level of DNA compaction discussed above. However, determination of higher-order chromatin structure using light and electron microscopy has so far remained elusive since nucleosomes and linker DNA/the 30 nm fiber cannot be adequately resolved in the compact state that occurs in the nucleus [59].

2.4.3 A Comparison of CCC and Cell Imaging Technologies: Why Not Use Both?

The 3D architecture of a whole genome or a genomic region such as the HoxA cluster which will be discussed in chapter 12 facilitates the colocalization in space of sequentially distant loci, which share a regulatory dependency. Thus, an integrative approach for elucidating the relationship between genome folding and function is needed. However, current biochemical and imaging techniques alone suffer from limitations.

CCC-based approaches aim at identifying spatial genome organization by determining contact interactions between genetic loci [10,87,87,127,128]. The so generated interaction datasets do not contain absolute interaction frequencies and lack proper calibration [10]. Thus, CCC technologies mainly provide a qualitative picture. The reported interaction frequencies values are "arbitrary" in the sense that they are greatly affected by the experimental setup, i.e. the composition of the primer library as well as of the amount that is hybridized to an array. Both aspects affect the raw signals and consequently the interaction frequency values, which are, in turn, rarely consistent from experiment to experiment. Nevertheless, the basic idea is that, within the same experiment and with the appropriate controls, such as a gene desert region, it is possible to normalize between cellular samples for comparison purposes.

Currently, it is possible to approximate absolute interaction frequencies by FISH measurements [7, 10]. In fact, data from FISH/high precision fluorescence microscopy is able to measure spatial distance distributions between sites of genomic regions under investigation [13], where the precise knowledge about Euclidian distances would also strengthen chromatin modeling approaches.

A further drawback of CCC technologies is that CCC datasets are (so far) generated from cell populations at various cell cycle states [128]. Since spatial organization of the chromatin fiber is dynamically linked to nuclear events during the cell cycle such as transcription, replication, DNA repair and recombination, interactions emerging from dedicated folding motifs at a specific cell cycle state (which we would like to study) might

be averaged out by this approach. Here, high-resolution imaging is expected to facilitate the comparison of chromatin arrangements at the single cell level and the quantification of cell-to-cell variability [7].

Last but not least, CCC techniques require a huge amount of sample cells to capture a particular interaction and do not allow for the distinction between simultaneously or disconnectedly occurring interactions [7]. In contrast, with the advent of a "rainbow of fluorescent proteins" [133], FISH allows to temporally and spatially visualize several DNA loci and proteins. It has to be taken into account, though, that, as with most fluorescence techniques, a significant problem is photo damage of cells due to excessive exposure or photobleaching resulting in a worsening of the signal-to-noise ratio [9].

In conclusion, a truly ingenious, integrative approach combining high-resolution imaging and CCC technologies with computational modeling for identifying spatial genome organization is needed. To this end, CCC measurements should be seen as providing an overall qualitative picture, while high-resolution (live) cell imaging is able to fine-tune our knowledge by spatially and temporally tracking the nuclear players at the single-cell level. In a subsequent step, computational modeling (applying such a tool as presented in chapter 12), can leverage the wealth of experimental CCC or microscopy data on chromatin folding to hint to specific folding motifs of interest, where high-resolution microscopy allowing for the precise measurement of distances might bridge the gap from a qualitative overview to a truly quantitative model of genome organization.

Chapter 3

Introduction to Polymer Physics and Modeling

In this chapter, basic concepts known for the last decades are discussed. The *freely-jointed*, the *Gaussian* and the *worm-like chain model* are presented. Measures describing the properties of polymers are introduced and will come handy in the following chapters. Eventually, the physics of polymer chains with excluded volume interactions is developed and a short overview over the principles underlying *universality* and coarse-graining is given.

The second major part of this chapter introduces the basic idea of Markov Chain Monte Carlo and presents the two algorithms used in the following chapters.

The reader is referred to Refs. [134–145] for a detailed review on polymer physics and to Refs. [144, 146, 147] for an introduction to computer simulation techniques and their application to polymer physics. A reader who is already familiar with these topics might want to skip this chapter altogether.

3.1 Polymer Models

The word polymer originates from the Greek words *poly* meaning "many" and *mer* meaning "part" [148], thus hinting to its composition of small chemical units joined together by chemical interactions. The small chemical units are called *monomers* and the number of monomers constituting a polymer is called the *chain length* or *degree of polymerization* N [140]. Since the degree of polymerization of macromolecules can be huge, concepts of statistical physics have to be applied in order to study their properties [134–145]. Thus, characteristic measures are given as mean values over the huge conformational space that one single polymer can explore [134–145].

3.1.1 The Freely-Jointed Chain Model

The simplest, yet insightful model, to describe polymers is the ideal or freely-jointed chain (FJC) model [135]. The freely-jointed chain consists of $N+1$ monomers at positions $\mathbf{r}_0, \ldots, \mathbf{r}_N$. Neighboring monomers, $\mathbf{r}_i$ and $\mathbf{r}_{i-1}$, are connected by rigid linkers $\mathbf{b}_i = \mathbf{r}_i - \mathbf{r}_{i-1}$

of length b. The orientation of these linkers is completely independent of the orientation and position of other linking segments allowing two monomers to overlap. Thus, the freely-jointed chain model describes a polymer as a *random walker* [134, 138, 144], meaning as a trajectory consisting of successive random steps neglecting any kind of monomer-monomer interactions [134, 141].

A polymer's *end-to-end distance* $\mathbf{R}_{\text{end}} = \mathbf{r}_N - \mathbf{r}_0 = \sum_{i=1}^{N-1} \mathbf{b}_i$ is a measure of its spatial extension. In case of the freely-jointed chain model, correlations between different bond vectors vanish, such that the *mean squared end-to-end distance* $\langle \mathbf{R}_{\text{end}}^2 \rangle$ can be calculated according to

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

and hence $\langle \mathbf{R}_{\text{end}}^2 \rangle \sim N^{2\nu}$ with $\nu = 0.5$.

A further shape descriptor is the *radius of gyration* of a polymer chain defined as

$$R_{\text{gyr}}^2 = \frac{1}{N+1} \sum_{i=0}^{N} (\mathbf{r}_i - \mathbf{r}_{\text{cm}})^2, \tag{3.2}$$

where $\mathbf{r}_{\text{cm}}$ denotes the polymer's *center of mass*. For the freely-jointed chain model one gets

$$\begin{aligned} R_{\text{gyr}}^2 &= \frac{1}{2N^2} \sum_{i=0}^{N} \sum_{j=0}^{N} \langle (\mathbf{r}_i - \mathbf{r}_j)^2 \rangle && (3.3) \\ &= \frac{1}{2N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} |i-j| b^2 && (3.4) \\ &= \frac{1}{6} N b^2, && (3.5) \end{aligned}$$

which shows that the mean squared radius of gyration is smaller than the mean squared end-to-end distance by a factor of 6.

The radius of gyration is closely related to the *gyration tensor* [149] which measures the distribution of monomers in space and is given by

$$S_{mn} = \frac{1}{N+1} \sum_{i=0}^{N} r_m^{(i)} r_n^{(i)}. \tag{3.6}$$

Here, $r_m^{(i)}$ is the m-th Cartesian coordinate of the position vector $\mathbf{r}^{(i)}$ of the i-th monomer with respect to the polymer's center of mass. The eigenvalues $\lambda_1 \leq \lambda_2 \leq \lambda_3$ of the gyration tensor correspond to the squared lengths of the *principal axes* of gyration and the squared radius of gyration can be expressed as the sum

$$R_{\text{gyr}}^2 = \lambda_1 + \lambda_2 + \lambda_3 \,.$$

The ratios of the eigenvalues indicate the deviation from a sphere-like shape having a value of unity for a perfect sphere. Notably, the gyration tensor for polymers in good solvent is not sphere-like but has a pronounced asphericity, given by $\langle \lambda_3 \rangle : \langle \lambda_2 \rangle : \langle \lambda_1 \rangle \to 12 : 2.7 : 1$ [150, 151].

The applicability of the freely-jointed chain model for describing "real" polymers is limited since it does not take into account excluded-volume interactions or monomer-solvent/monomer-monomer interactions. However, in a dense melt of many polymers,

where the three-dimensional space is filled uniformly and each polymer interacts with many other polymers besides its own, polymers really behave ideally on large length scales [135, 141, 144]. Other fluctuating polymer models, that consider neither monomer-monomer interactions/monomer-solvent interactions nor excluded volume, such as the Gaussian or the worm-like chain model, converge to the freely-jointed chain model in the thermodynamic limit [137, 139–141].

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, which can be taken into account by introducing a fluctuating bond vector $\mathbf{b}$. In the *Gaussian chain* model [140, 141], a polymer is made up by such freely-jointed bond vectors b, which are not constant but follow a Gaussian distribution

$$G(\mathbf{b}) = \left(\frac{3}{2\pi b^2}\right)^{3/2} \exp\left(-\frac{3\mathbf{b}^2}{2b^2}\right), \tag{3.7}$$

where $\langle \mathbf{b}^2 \rangle = b^2$. Hence, the probability distribution for the end-to-end vector $\mathbf{R}_{\text{end}}$ is

$$P(\mathbf{R}_{\text{end}}) = \prod_{i=1}^{N} \left(\frac{3}{2\pi b^2}\right)^{3/2} \exp\left(-\frac{3\mathbf{b}_i^2}{2b^2}\right) \tag{3.8}$$

$$= \left(\frac{3}{2\pi b^2}\right)^{3/2} \exp\left(-\sum_{i=1}^{N} \frac{3(\mathbf{r}_i - \mathbf{r}_{i-1})^2}{2b^2}\right). \tag{3.9}$$

Notably, the Gaussian chain model has a physical analogy with a system of beads connected by harmonic strings [138–141]. This can be understood by noting that the Gaussian chain is equivalent to a many-body system with interaction energy

$$U(\mathbf{r}_1, \ldots, \mathbf{r}_N) = \frac{1}{2}\frac{3kT}{b^2} \sum_{i=1}^{N} (\mathbf{r}_i - \mathbf{r}_{i-1})^2, \tag{3.10}$$

where the equilibrium probability distribution of the monomer positions is given by the Boltzmann factor $\exp(-U/kT)$ and $\kappa = \frac{3}{b^2}kT$ [138–141].

For the mean squared end-to-end distance we find

$$\langle \mathbf{R}_{\text{end}}^2 \rangle = Nb^2, \tag{3.11}$$

which is the same as for the freely-jointed chain model since the short-range interaction potential represented by the bond vector distribution $G(\mathbf{b})$ does not influence large-scale quantities. In fact, in case of the freely jointed chain, the central limit theorem in statistics [152] states that the distribution of end-to-end distances $\mathbf{R}_{\text{end}} = \sum \mathbf{b}_i$ converges to a Gaussian distribution in the limit of large chain length N [139–141].

3.1.3 The Worm-like Chain Model

One of the most basic characteristics of macromolecules is chain flexibility or the lack thereof, namely chain stiffness [134, 136, 137, 139–141]. In particular, important biopolymers such as DNA, proteins, rod-like viruses or actin filaments belong to the class of semi-flexible polymers [143, 145, 153–156]. The Kratky-Porod *worm-like chain* (WLC) model

describes a semiflexible polymer of contour length L as a differential space curve $\mathbf{r}(s)$ of length L parameterized by the unit tangent vector $\mathbf{u}(s)$ at $s \in [0, L]$,

$$\mathbf{u}(s) \equiv \frac{\partial \mathbf{r}(s)}{\partial s}. \tag{3.12}$$

The worm-like chain model can be described by a Hamiltonian (in the continuum limit)

$$\mathcal{H} = \frac{\kappa}{2} \int_0^L \mathrm{d}s \left(\frac{\partial^2 \mathbf{r}(s)}{\partial s^2} \right)^2, \tag{3.13}$$

where the parameter κ describing the bending stiffness is related to the *persistence length* by $\kappa = l_\mathrm{p} kT$ [134, 137, 139, 157].

The internal structure of the worm-like polymer chain, i.e. stretching as opposed to coiling, can be investigated by computing the correlation of two unit tangent vectors $\mathbf{u}(i)$ and $\mathbf{u}(j)$ separated by a contour distance $|i - j| \in [0, L]$ along the polymer backbone

$$C(|i - j|) = \langle \mathbf{u}(i)\mathbf{u}(j) \rangle. \tag{3.14}$$

The average is taken over the whole ensemble of possible chain conformations. For chain molecules in the absence of excluded volume, where distances between monomers that are far apart satisfy Gaussian statistics, the orientational correlation function decays exponentially [134, 137, 139, 141]

$$\langle \mathbf{u}(i)\mathbf{u}(0) \rangle = \exp(-|i - j|/l_\mathrm{p}). \tag{3.15}$$

Thus, for sufficiently large distances $|i - j| \in [0, L]$ the orientational correlation function tends to zero, $C(|i - j| \to \infty) \to 0$ [139–141].

The Kratky-Porod model yields for the mean squared end-to-end distance

$$\begin{aligned}
\langle \mathbf{R}_\mathrm{end}^2 \rangle &= \langle \mathbf{R}_\mathrm{end} \mathbf{R}_\mathrm{end} \rangle && (3.16)\\
&= \Big\langle \int_0^L \mathbf{u}(s)\,\mathrm{d}s \int_0^L \mathbf{u}(s')\,\mathrm{d}s' \Big\rangle && (3.17)\\
&= \int_0^L \mathrm{d}s \int_0^L \langle \mathbf{u}(s)\mathbf{u}(s')\,\mathrm{d}s' \rangle && (3.18)\\
&= \int_0^L \mathrm{d}s \int_0^L \exp(-|s - s'|/l_\mathrm{p})\,\mathrm{d}s' && (3.19)\\
&= 2 l_\mathrm{p} L \Big\{ 1 - \frac{l_p}{L} [1 - \exp(-L/l_\mathrm{p})] \Big\}, && (3.20)
\end{aligned}$$

which will come in handy for the validation of the worm-like chain algorithm used in chapter 12. Eqn. 3.20 shows the standard Gaussian behavior for $L \to \infty$, $\langle R_\mathrm{end}^2 \rangle = 2 l_\mathrm{p} L = 2 l_\mathrm{p} b N$ since $L = Nb$, while rod-like behavior $\langle \mathbf{R}_\mathrm{end}^2 \rangle = L^2$ is recovered for $L \ll l_\mathrm{p}$.

3.1.4 Polymers with Self-Avoidance

In the polymer models presented so far the polymer chain is able to cross itself or trace back the same path, allowing monomers to occupy the same position in space. In physical systems such as flexible polymers under good solvent conditions [134, 136, 137, 139, 144, 145], this is impossible since steric hindrance induces monomer-monomer excluded volume restriction. Thus, a more "realistic" model of a polymer chain is a self-avoiding random

walk (SAW), which is a path on a lattice that does not visit the same site more than once [134–136, 144, 145].

The size of a polymer chain with excluded volume is larger than that of a corresponding ideal polymer. This can be understood by noting that, since an excluded volume chain is forbidden to overlap, its configurational space is biased towards more extended conformations. However, intrinsic difficulties arising from the emergence of long range monomer-monomer correlations along self-avoiding chains, hamper a rigorous mathematical treatment of self-avoiding random walks [144, 145].

A simple argument put forward by the chemist Paul J. Flory can be used to estimate the size of a swollen chain by considering two counteracting forces, namely a free energy contribution penalizing a strong elongation of the polymer chain due to decreasing conformational entropy and an opposed free energy contribution acting against strong compression due to excluded volume interactions(for more details see [135, 138–142, 144]). Minimizing the sum of both free energy contributions with respect to the end-to-end distance vector $\mathbf{R}_{\text{end}}$ yields a general scaling law

$$\langle \mathbf{R}_{\text{end}}^2 \rangle \sim N^{2\nu}, \tag{3.21}$$

with $\nu \approx 3/5$. While the exponent for a random walker, $\nu = 0.5$, is exact, the estimate of $\nu \approx 3/5$ for a self-avoiding walker is still remarkably close to more sophisticated computations based on perturbation and renormalization group theories, which find $\nu \approx 0.588$ [144, 145, 158].

3.1.5 Universality and Coarse-graining

The reason why polymers can be modeled by random or self-avoiding random walks (depending on the physical circumstances) is their belonging to the same *universality class* as the random or self-avoiding random walk [144]. In particular, this means that they have the same critical exponents [135,141], which can be understood by considering a polymer's end-to-end distance. For a polymer of length N, the end-to-end distance is expected to scale as DN^{ν} for $N \to \infty$, where the exponent ν is believed to be universal, while D is not [144]. In fact, ν is expected to be the same for all polymer systems falling into the same universality class such as linear polymers in good solvents [144, 146, 147]. As opposed to the exponent ν, the factor D is influenced by the microscopic details of the system such as the specific monomeric units and solvent molecules [144]. To this end, the scale invariant features exhibited by systems belonging to the same universality class can be exploited for comparison with experimental data.

The wide range of time and length scales challenge modeling approaches of polymer systems. A recurring theme in the development of polymer models and simulations is the idea of *coarse-graining* [135, 138, 141], which, however, implies frustration: On the one hand, the underlying complexity of biophysical systems has to be reduced for the sake of efficiency. On the other hand, sufficient details need to be preserved for the sake of accuracy. Fig. 3.1 illustrates a coarse-grained process zooming into a random walker down to the level of least coarse-graining, the DNA double helix.

3.2 Computational Modeling of Polymers

The two main simulation techniques for modeling macromolecules are molecular dynamics (MD) and Monte Carlo (MC). In a MD simulation Newton's equations of motion for a

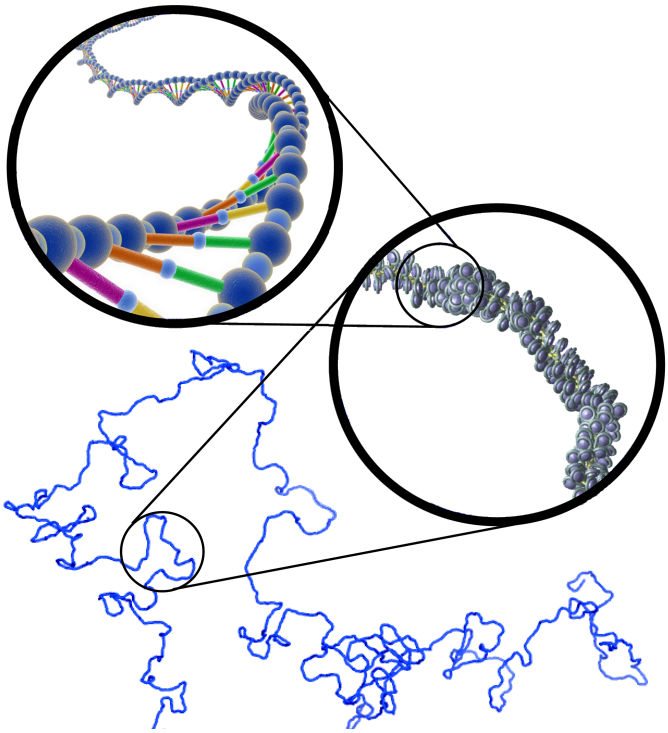


Figure 3.1: Zooming into a random walker and ending at the level of least coarse-graining, the DNA double helix. Parts of this image are adapted from [159] and [160].

system of interacting particles are numerically solved, where interaction potentials are applied to the system of particles [147, 161].

The general idea behind Monte Carlo techniques is to provide an efficient stochastic sampling of the configurational phase space (or of parts of it) in order to obtain approximations for statistical quantities such as expectation values or probabilities [146, 147, 162]. The obvious advantage of MD simulations over MC methods is that they give a route to dynamical properties of the system such as transport coefficients, time-dependent responses to perturbations etc. [161] On the other hand, MD simulations of complex systems on a microscopic scale often exceed the feasible time scale of the respective MD simulation [147, 162].

Luckily, answering many questions related to (biological) complex systems requires the study of the interplay between energy and entropy rather than the explicit consideration of intrinsic dynamics. The key to this is provided by MC simulations, which are applied in the following chapters to study the properties of (biological) polymer systems. Here, the basic idea of dynamic MC methods as well as two algorithms for the generation of polymer conformations are discussed. A brief introduction to MD simulations is given in chapter 10, where the simulation method is employed. For a detailed overview of computer simulation techniques and their application to polymer physics, the reader is referred to Refs. [144, 146, 147].

3.2.1 Markov Chain Monte Carlo

Monte Carlo methods can be classified as static, quasi-static and dynamic methods [144, 146, 147]. In this work, dynamic Monte Carlo methods are applied, which is why the

focus is on these methods. Dynamic methods generate new conformations by updating the immediate predecessor conformation instead of building up walks from scratch (static methods) [144].

In this context, the method of *Markov chains* in connection with Monte Carlo simulations was first introduced by Nicolas Metropolis [163]. The idea is the following:

Let π be a probability distribution on some set S, such that $\pi(i)$ is the probability for each $i \in S$ and $\sum_{i \in S} \pi_i = 1$. If one can find a *Markov chain* with state space S whose unique equilibrium distribution is π, then the fundamental theory of Markov chains [144, 146,147,162] states that running this chain for a long time will produce observations whose distribution approaches π.

In the following, $S = S_N$, where S_N is the set of all N-step polymer chain configurations and $\pi(\omega)$ is the probability for every polymer chain conformation ω in S_N. The state the system is in at discrete time steps $t = 0, 1, 2, \ldots$ is denoted by $\omega^{[t]}$. Now, one begins with a conformation $\omega^{[0]}$, which is locally or globally modified yielding a subsequent conformation $\omega^{[1]}$ [144, 146, 147, 162]. *Local update moves* for polymer models can be random changes of single monomers positions, bond angle modifications or rotations about covalent bond axis, while *global update moves* consists of combinations of local updates [146, 147]. In this way, one generates a sequence of conformations $\{\omega^{[t]} : t = 0, 1, \ldots\}$ such that for a sufficiently large t the distribution of $\omega^{[t]}$ is arbitrarily close to π. The crucial requirement is that the system evolves towards a unique equilibrium probability distribution, which is the desired one, and stays there. Thus, the basic problem is determining the conditions that satisfy this demand.

A Markov process is a process, that does not possess an explicit "memory", since the conditional probability is independent of all states but the immediate predecessor. Hence, the conditional probability that $\omega^{[t]} = i$ is $\Pr\{\omega^{[t]} = i | \omega^{[t-1]}, \ldots, \omega^{[1]}\} = \Pr\{i | \omega^{[t-1]}\}$ [144, 147, 162].

The advantage of this property is that one can introduce a set of transition rates between every two states. Let $P(i,j) = \Pr\{\omega^{[t+1]} = j | \omega^{[t]} = i\}$ be the one-step transition probability of the chain and $P^n(i,j) = \Pr\{\omega^{[t+n]} = j | \omega^{[t]} = i\}$ the n-step transition probability. The chain is said to be irreducible if for every i and j in S_N there exists an $0 < n < \infty$ such that $P^n(i,j) > 0$ [144, 147, 162]. This means that every state can be reached from every other state.

Additionally, an irreducible chain is said to be aperiodic if $P(i,i) > 0$ for some i [144]. A chain that is both irreducible and aperiodic is ergodic [144,147,162]. However, ergodicity is often hard to prove for a Monte Carlo algorithm and an insufficient choice of move sets might lead to systematic errors [144, 146, 147].

By demanding *detailed balance*, that is

$$\pi_i P(i,j) = \pi_j P(j,i) \tag{3.22}$$

for every i and j in S_N [144, 147, 162], one asserts that the probability of reaching a state $i \in S_N$ is equal to the probability of leaving it (no "absorbing" states), thus assuring that the system evolves towards equilibrium.

Based on the detailed balance condition, it is easy to construct the transition probabilities for the canonical ensemble, since the equilibrium distribution is known

$$\frac{P(i,j)}{P(j,i)} = \frac{\pi_j}{\pi_i} = \exp(-\beta \Delta E), \tag{3.23}$$

where $\Delta E = \mathcal{H}(\omega_j) - \mathcal{H}(\omega_i)$ is the energy difference between two conformations [144, 147, 162]. This transition probability comes into play in the Metropolis method [146, 147, 163],

which is the standard *importance sampling* variant. The underlying idea is that, due to the dominance of a certain restricted space of microstates, it is useful to concentrate in a simulation on a precise sampling of those microstates that determine the macrostate given a specific temperature [144, 146, 147, 162, 163]. Thus, the acceptance probability p is governed by

$$p = \min[1, \exp(-\beta \Delta E)]. \tag{3.24}$$

3.2.2 Autocorrelation Time and Initialization Bias

Two major difficulties arise when using dynamic Monte Carlo methods. The first one is the correlation between samples $\omega^{[t]}$ since each configuration is generated out of the previous one or is equal to it. Since the statistics is effectively reduced by the number of Monte Carlo steps until the correlations have decayed, the statistical error of an estimator $\langle A \rangle$ is given by the relation [144, 146, 147, 162]

$$\epsilon_{\langle A \rangle} = \frac{\sigma_{\langle A \rangle}}{\sqrt{n_{\text{eff}}}}. \tag{3.25}$$

Here, $n_{\text{eff}} = n/\tau_{\text{int}}$ represents the number of independent conformations, where n denotes the total number generated conformations $\omega^{[t]}$, and τ_{int} corresponds to the integrated autocorrelation time.

In the simulations performed in the coming chapters, the integrated autocorrelation time τ_{int} is computed by first determining the autocorrelation function $C(t)$ of the squared radius of gyration $A(t) = R^2_{\text{gyr}}(t)$ according to

$$R(t) = \frac{1}{K-t} \sum_{i=0}^{K-t} [A(i+t) - \langle A \rangle][A(i) - \langle A \rangle], \tag{3.26}$$

and thus,

$$C(t) = \frac{R(t)}{R(0)}. \tag{3.27}$$

Here, K denotes the number of performed MC steps and $C(t)$ gives the correlation of two conformations separated by t time steps along the Markov chain. Now, the integrated autocorrelation time τ_{int} can be calculated, which is defined as

$$\tau_{\text{int}} = \frac{1}{2} + \sum_{t=1}^{\infty} C(t). \tag{3.28}$$

Notably, the integrated autocorrelation time τ_{int} should be computed for the "slowest relation mode" of the system [147, 164], which is, however, not known a priori. Thus, the radius of gyration is used as a measure for the relaxation time [56, 57, 165].

Finally, an approximation of Eqn 3.28 is obtained by applying the "automatic windowing" procedure introduced by Sokal [164]. In a first step, the autocorrelation time is computed up to a cutoff T,

$$\hat{\tau}_{\text{int}}(T) = \frac{1}{2} + \sum_{t=1}^{T} C(t). \tag{3.29}$$

Now, T^* is the smallest integer, which satisfies

$$T^* \geq 10\, \hat{\tau}_{\text{int}}(T) \tag{3.30}$$

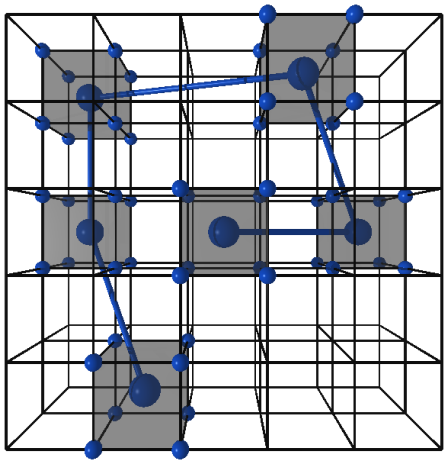


Figure 3.2: Schematic illustration of the bond fluctuation model adapted from [92, 167]. A monomer blocks the eight lattice sites, which then cannot be occupied by other monomers. The bond vectors connecting two monomers fluctuate within a set of "allowed" bond vectors. Chain configurations change by random monomer moves by one lattice unit.

and hence the integrated autocorrelation time for the observables $A(t)$ is defined as

$$\tau_{\text{int}} = \hat{\tau}_{\text{int}}(T^*). \tag{3.31}$$

The factor 10 is arbitrary but the idea is to make sure to include contributions from terms that are several τ_{int}'s apart [144]. Having in mind that the variance of the observable $A(t)$ is larger by $2\tau_{\text{int}}$ than it would be if the conformations $\omega^{[t]}$ we decorrelated, we consider two subsequent conformations as uncorrelated after $5\tau_{\text{int}}$ Monte Carlo steps.

The second difficulty arising from dynamic Monte Carlo techniques is initialization bias as a simulation typically starts from a state that is not chosen according to the equilibrium distribution. A simulation normally begins with an initial period that is "far from equilibrium" and it eventually "approaches equilibrium" [164]. This initial period has to be removed from the data since it introduces a bias to the estimator. To achieve this, one might watch some observables over time until they have stabilized [164].

3.2.3 Simulating Self-Avoiding Polymers

The bond fluctuation method (BFM) [166, 167] is a coarse-grained lattice Monte Carlo method and employed in all coming chapters with the exception of chapter 12. It has successfully been used to study the static and dynamic properties of polymer systems [33, 56, 57, 146, 165–167]. This algorithm constructs a chain as illustrated in Fig. 3.2.

The BFM satisfies the following criteria: (i) The set of allowed moves is local such that on small scales in space and time the Rouse model dynamics is reproduced [138, 168]. (ii) The monomers are not allowed to sit on top of each other on the simulation lattice in order to account for excluded volume interactions [166, 167]. (iii) No bond crossings can occur such that entanglement restrictions can be taken into account [138, 169]. (iv) The algorithm is ergodic [166, 167, 170]. Moreover, as a coarse-grained lattice model, its computational efficiency renders it more attractive than off-lattice models.

The polymer evolves on a cubic lattice, where each monomer blocks the eight lattice sites, which then cannot be occupied by other monomers as shown in Fig. 3.2. Monomers

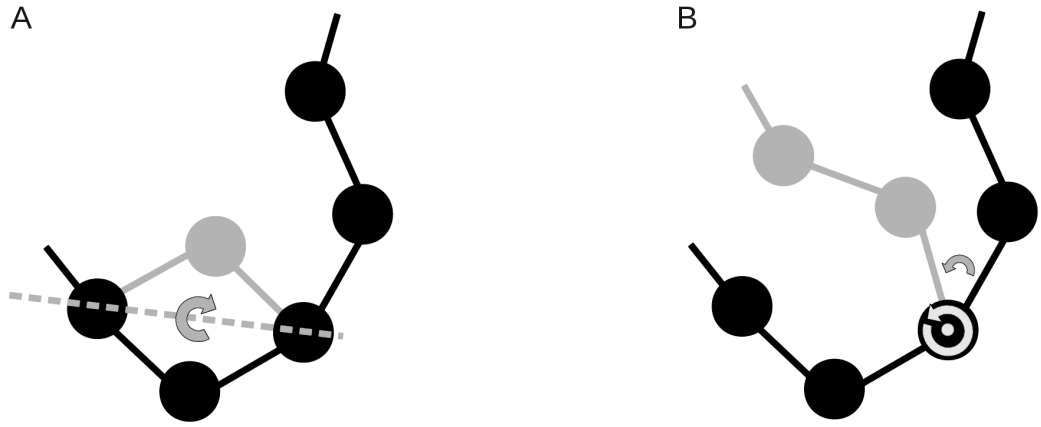


Figure 3.3: Schematic illustration of the Monte Carlo moves for the worm-like chain model applied in chapter 12.

are connected by fluctuating bond vectors of lengths 2, $\sqrt{5}$, $\sqrt{6}$, 3 and $\sqrt{10}$, leading to an average bond length of $\langle b \rangle = 2.7$. Allowed bond vectors are $\mathcal{B} = \{(2,0,0),(2,1,0),(2,1,1),(2,2,1),(3,0,0),(3,1,0) + \text{permutations} + \text{sign combinations}\}$.

A Monte Carlo move consists of randomly choosing one monomer to be moved to a randomly chosen lattice direction by one lattice unit. This trial motion is only accepted if neither the excluded volume constraint nor the restriction on the allowed range of bond lengths is violated. By additionally considering the Metropolis transition probability [146, 147, 163] for accepting or rejecting a move it is straightforward to include effects due to a finite interaction energy [167].

3.2.4 Simulating Worm-like Chains

In chapter 12 the worm-like chain model is applied to describe the linear chromatin fiber on a length scale between bare DNA and a highly condensed 30 nm fiber in agreement with [127, 171–174]. While the polymer is constraint to an underlying lattice in the bond fluctuation method, the worm-like chain is described by N segments of fixed length b freely evolving in space. Self-avoidance is not taken into account.

Two types of moves (A) and (B), schematically illustrated in Fig. 3.3, are applied at equal probability to update the polymer chain conformations.

(A) Two monomers i, j with position vectors $\mathbf{r}_i$ and $\mathbf{r}_j$ are chosen randomly. Let $i < j$. Then, all monomers m, for which $i < m < j$ holds, are rotated with respect to the rotation axis defined by $\mathbf{r}_\varphi = \mathbf{r}_i - \mathbf{r}_j$ by a randomly chosen angle $\varphi \in [-\pi, \pi]$. Local moves for the special case of $|i - j| = 2$ are performed with a probability of 70%.

(B) A monomer i with position vector $\mathbf{r}_i$ is randomly chosen. A rotation axis $\mathbf{r}_\theta$ orthogonal to the plane defined by the monomers $i - 1$, i and $i + i$ and intersecting this plane at $\mathbf{r}_i$ is determined. Then, a rotation of $\theta \in [-\pi, \pi]$ with respect to $\mathbf{r}_\theta$ is performed for all monomers m with either $0 \leq m < i$ or $i < m < N$ (randomly determined).

The rate of accepted Monte Carlo moves is determined for each individual move type (A) and (B). While the maximum range of rotation angles for φ and θ is chosen between $-\pi$ and $+\pi$, this range is automatically adjusted in the course of the simulation such that the rate of accepted moves for (A) and (B) is always between 40-60%.

In order to test whether the chosen set of moves correctly generates the expected

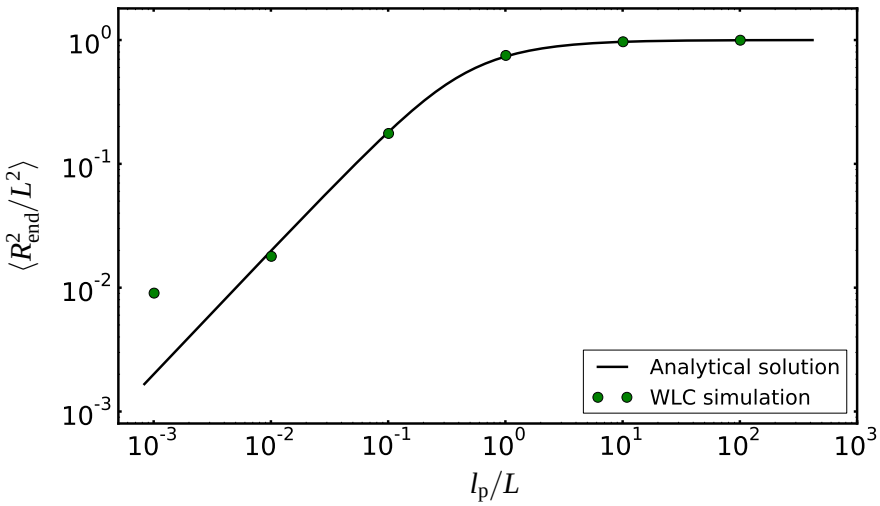


Figure 3.4: Analytical solution for the mean squared end-to-end distance $\langle \mathbf{R}^2_{\mathrm{end}} \rangle = 2Ll_{\mathrm{p}} - 2l_{\mathrm{p}}[1 - \exp(-L/l_{\mathrm{p}})]$ compared to the numerical solution of the WLC simulation. The deviation of the first data point (at $l_p/L = 10^{-3}$) from the analytical solution is due to finite size effects. At low persistence lengths the discretization of the chain prevents it from coiling into smaller volumes. The standard error is of the order of the line width. Image adapted from [175].

worm-like chain statistics, the numerical solution for the mean squared end-to-end distance $\langle \mathbf{R}^2_{\mathrm{end}} \rangle$ as a function of persistence length l_{p} is compared to the analytical one derived in Eqn. 3.20. Fig. 3.4 illustrates this comparison.

Chapter 4

Single- and Double-Tethered Polymers

Implications for Biological Tethering

References

The results presented in this chapter are published as and adapted from

- M. Fritsche, D.W. Heermann, M. Dutra, and C.E. Cordeiro (2010), *Conformational and Dynamical Properties of the Isolated, Three-Dimensional Single- and Double-Tethered Polymer Chain on an Infinite Surface.* Macromolecular Theory and Simulations, 19: 440-448.

We thank Susan M. Parkhurst for providing the chromosome image in Fig. 4.1 b as well as Manfred Bohn and Benoît Knecht for helpful discussions.

Chapter Summary

The conformational and dynamical properties of single polymer chains attached to an impenetrable surface are important both from a theoretical and a biological point of view. On the one hand, these simple models serve as a kind of preparation for tethered systems of larger complexity such as the genome organisation in eukaryotic and prokaryotic organisms, which is expected to be influenced by tethering interactions fixing parts of the chromosome(s) to the membrane of the nucleus or the cell membrane, respectively. On the other hand, in the context of the theory of polymer brushes, single-tethered (ST) chains grafted at one end to an interface have received considerable attention in recent years, while few studies have addressed the isolated, double-tethered (DT) chain in three dimensions.

Monte Carlo simulations have been performed to elucidate and compare the characteristics of two different models of three-dimensional ST and DT chains, respectively. In the first, either one or both ends (at a grafting distance d) are fixed on the surface, while, in the second model, the ST or DT chain is "annealed" by permitting the anchors to diffuse laterally along the surface. Besides recovering scaling relationships regarding the gyration radius and the end-to-end distance, we aim at providing theoretical insights to guide future experimental approaches that investigate the impact of tethering on the packaging of the genetic material by concurrent visualisation of multiple loci along the chromosome.

Thus, we study the crossover behavior of polymer chains of finite lengths rather than the "unbiological" limit of $N \to \infty$. In this context, we analyse the complexity of chain self-entanglement, correlations between intra-chain segments as well as the relationship between the mean square physical distance separating two monomers and the corresponding contour length. Finally, to our knowledge, we provide the first report on the diffusion behavior of the four model systems, finding subdiffusive dynamics at time scales that are small with respect to the chain relaxation times.

4.1 Introduction

Polymers tethered or grafted by one or both ends to a surface play an important role in various biological and physical systems. Prominent examples are polymer chains grafted to colloidal particles or chromosomes anchored to the cell membrane [2, 46, 47]. Layers of single-tethered chains grafted at one end to an interface, so called "polymer brushes", have been studied extensively for various grafting densities σ (number of chains per unit area) by theoretical and experimental approaches [176, 177]. At high σ, in the "brush" regime, chains interact with each other and the layer height h scales as $N\sigma^{1/3}$ [178] At low σ, inter chain interactions are essentially absent and chains assume an inverted mushroom shape. In this "mushroom" regime, they behave as random coils and the layer height h scales only with chain length, $h \sim N^{3/5}$ [134].

Few studies have, however, addressed the analogous case of double-grafted chain layers, where both chain ends are anchored to an interface. Jones and coworkers investigated the equilibrium conformations and dynamic relaxation of three-dimensional layers comprised of double-tethered polymer chains in the "brush" regime [179]. While they also elucidated the conformational properties of the isolated two-dimensional double-grafted chain [180],

a detailed study of the isolated case in three dimensions has not been performed yet, even though Shida and coworkers have already shed light on the total configuration number and the three-dimensional segment density profile [181]. Huang and coworkers have performed a study of the shape of the isolated, fixed three-dimensional single-tethered polymer chain [182].

In this chapter, the equilibrium conformational properties and dynamics of both the isolated single- and double-tethered polymer chain in three dimensions are analysed and compared. Two models of each single- and double-tethered chain are investigated. Fig. 4.1 schematically illustrates the four polymer systems. In the first, either one or both ends are grafted to the surface, where the last case allows the study of the impact of a finite grafting distance d between the two fixed chain ends. In the second model, the single- and double-tethered chains are "annealed" by permitting the anchors to diffuse laterally along the surface so that the effect of chain mobility can be studied.

The investigation of these model systems is important not only because of the elucidation of yet to be answered theoretical questions, but also as a way of better understanding tethered (biological) systems of larger complexity. Among these is the packaging of the genome in both eukaryotic and prokaryotic organisms, which is expected to be influenced by tethering interactions fixing parts of the chromosome(s) to the membrane of the nucleus or the cell membrane, respectively [2,47]. In this chapter, the aim is to provide theoretical insights to guide future experimental approaches that try to obtain information about the impact of tethering in shaping eukaryotic and prokaryotic genome organisation.

Multiple loci on the chromosome can be concurrently visualised with methods such as in situ hybridisation (FISH), the fluorescent repressor-operator system (FROS) or ParB-*par*S systems therewith determining the relationship between the physical distance of genomic sequence elements (chromosomal loci) and their genomic distance [183]. Consequently, the crossover behavior of polymer chains of finite lengths rather than the "unbiological" limit of $N \to \infty$ is of interest here.

4.2 Modeling

We apply a lattice polymer model, where either one or both chain ends are grafted to the infinite surface or permitted to diffuse laterally along it as illustrated in Fig. 4.1.

Thus, simulations are performed on a lattice of size $750 \times 750 \times 1700$ with periodic boundary conditions. In z-direction the dimension of the box is chosen to be large enough to ensure an unperturbed polymer coil in the box. The flat surface is assumed to be infinitely large and impenetrable.

We employ the bond-fluctuation method (BFM) [166] which has already been introduced in chapter 3. The simulation method produces unbiassed results, takes into account excluded volume interactions and ensures that no bond crossings can occur [166,167].

In order to generate thermodynamically equilibrated polymer conformations we use the Metropolis Monte Carlo method. Since subsequently created conformations are highly correlated, we determine, for each set of parameters (chain length N, grafting distance d), the autocorrelation function of the squared radius of gyration. Then, the intergrated autocorrelation time τ_{int} is computed by applying the windowing procedure introduced by Sokal [164]. For further details on the determination of the autocorrelation time please refer to chapter 3. We consider two subsequent conformations as uncorrelated after $5\tau_{\text{int}}$ Monte Carlo steps. By creating 10 000-100 000 independent configurations we are able to study polymer chains of lengths $N = 40, 80, 160, 320, 640$ and $N = 1280$. Regarding the

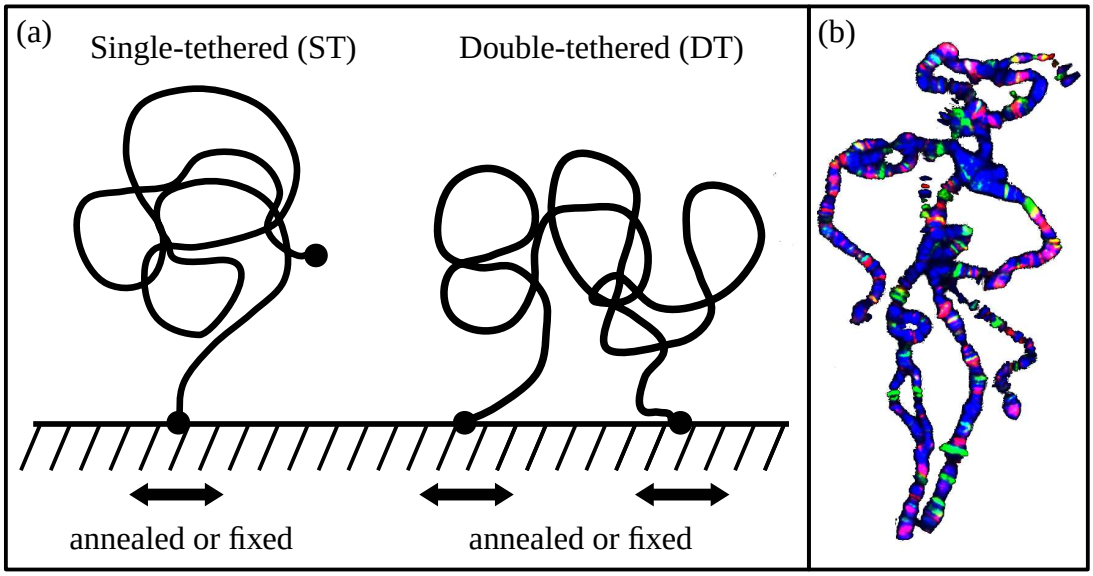


Figure 4.1: (a) Schematical illustration of the fixed and annealed single-tethered (ST) and double-tethered (DT) chain. Either one or both ends are grafted to the surface, where the last case allows the study of the impact of a finite grafting distance d between the two fixed chain ends. In case of the annealed ST and DT chain the anchors are permitted to diffuse laterally along the surface so that the effect of chain mobility can be studied. (b) A drosophila chromosome 2 is displayed, where genomic sites are labeled with fluorescence markers [184]. Polymer models can help to better understand the relationship between the mean square physical distance and the genomic distance between two such markers on the same chromosome, therewith providing insight into the folding and compactification of the genetic material in eukaryotic and prokaryotic cells. Image adapted from [185].

fixed DT chain, we performed simulations for the following ratios between the grafting distance d and the chain contour length L, $d/L = 1/2.7, 1/5.4, 1/10.6, 1/21.6$ and $1/54$. Furthermore, we investigate the situation, where the grafting distance d of the fixed DT chain corresponds to the mean end-to-end distance of the annealed DT chain of equal length N, $d = \langle R^a_{\text{end}} \rangle$.

4.3 Results and Discussion

4.3.1 Conformational Properties of Single- and Double-Tethered Chains

We investigate the mean-squared end-to-end distance $\langle R^2_{\text{end}} \rangle$ and the mean-squared radius of gyration $\langle R^2_{\text{gyr}} \rangle$. It is known that $\langle R^2_{\text{end}} \rangle$ and $\langle R^2_{\text{gyr}} \rangle$ scale with the chain length N according to $\langle R^2_{\text{end}} \rangle \sim N^{2\nu}$ and $\langle R^2_{\text{gyr}} \rangle \sim N^{2\nu}$ with $\nu = 0.589$ [169]. Fig. 4.2 shows the scaling plot for both quantities versus chain length, where fits are displayed as solid and dashed lines, respectively. In agreement with Huang et al. [182] The fixed ST chain has the same exponent as the free self-avoiding walk (SAW). For the annealed ST and DT chain, we find the same behavior, therewith confirming the universal exponent ν. Moreover, while the annealed and the fixed ST chain display exactly the same quantitative behavior, we find that the gyration radius of the fixed DT chain equals that of the annealed one if the fixed DT chains's grafting distance d equals the annealed DT chain's mean end-to-end

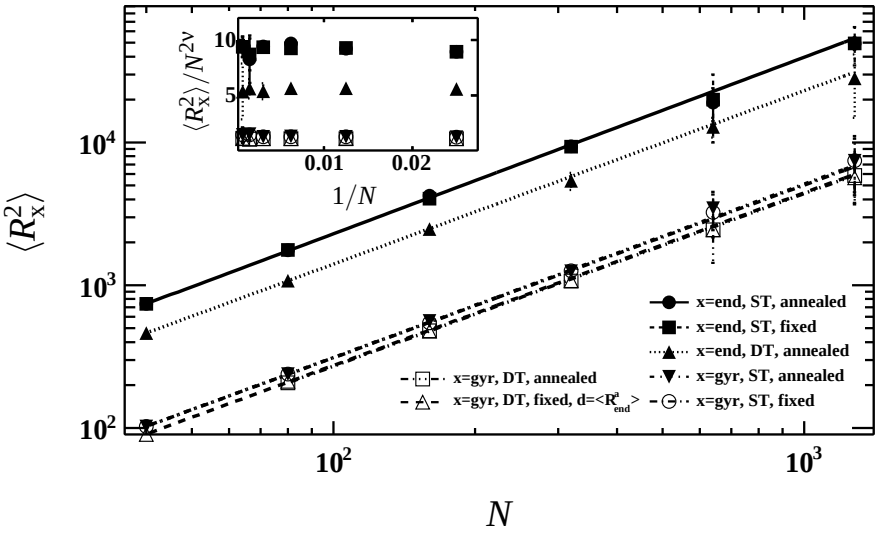


Figure 4.2: Log-log plots of the mean square end-to-end distance and the mean square radius of gyration vs chain length for the fixed and annealed ST chain as well as for the annealed DT chain and the fixed DT chain with $d = \langle R^a_{\text{end}} \rangle$. Fits are displayed as solid and dashed lines, respectively. The ratio $\langle R^2_{\text{end}} \rangle / \langle R^2_{\text{gyr}} \rangle$, being 6 for the SAW, is about 7.5 for both ST systems and about 5.2 for the annealed DT chain. This means that the repulsion of the flat surface forces both types of ST chains to stretch upward, whereas the annealed DT chain adopts a more localised conformation near the interface (due to the constraints placed upon the ends) than does a ST chain of equal length. Inset: Plot of $\langle R^2_{\text{end}} \rangle / N^{2\nu}$ and $\langle R^2_{\text{gyr}} \rangle / N^{2\nu}$ vs $1/N$ with $\nu = 0.588$. Image adapted from [185].

distance, $d = \langle R^a_{\text{end}} \rangle$. The ratio $\langle R^2_{\text{end}} \rangle / \langle R^2_{\text{gyr}} \rangle$, being 6 for the SAW, is about 7.5 for both ST systems and about 5.2 for the annealed DT chain. This means that the repulsion of the flat surface forces both types of ST chains to stretch upward, whereas the annealed DT chain adopts a more localised conformation near the interface (due to the constraints placed upon the ends) than does a ST chain of equal length.

A way to characterise the shape of a polymer is the gyration tensor, defined by

$$S_{mn} = \frac{1}{N} \sum_{i=1}^{N} r_m^{(i)} r_n^{(i)}. \tag{4.1}$$

Here, $\mathbf{r}^{(i)}$ is the coordinate vector of the ith monomer and the subindex denotes its Cartesian components. The eigenvalues $\lambda_1^2 \leq \lambda_2^2 \leq \lambda_3^2$ give the squared lengths of the principal axes of gyration, while their ratios indicate the deviation from a sphere-like shape, both having a value of unity for a sphere. It is well known that free SAW polymer chains show a pronounced asphericity, showing up in the ratio $\langle \lambda_1^2 \rangle : \langle \lambda_2^2 \rangle : \langle \lambda_3^2 \rangle \to 1 : 2.98 : 14.1$ [186]. In the present work, we determined the eigenvalue ratios as shown in Fig. 4.3. For the fixed ST chain, they approach asymptotically $1 : 3.0 : 15.0$ in good agreement with data published earlier [182] and the annealed ST chain is characterised by exactly the same triaxial ellipsoidal shape. For the annealed DT chain the ratio $1 : 3.2 : 12.0$ is found showing that it prefers a more spherical shape with respect to the ST case and the SAW. In case of the fixed DT chain, the ratio of the principal axis of inertia is strongly determined by the imposed grafting distance d, where the lateral stretching of the chain leads to its pronounced asphericity. In contrast to the ratio $\langle \lambda_2^2 \rangle / \langle \lambda_1^2 \rangle$, $\langle \lambda_3^2 \rangle / \langle \lambda_1^2 \rangle$ is mostly affected by

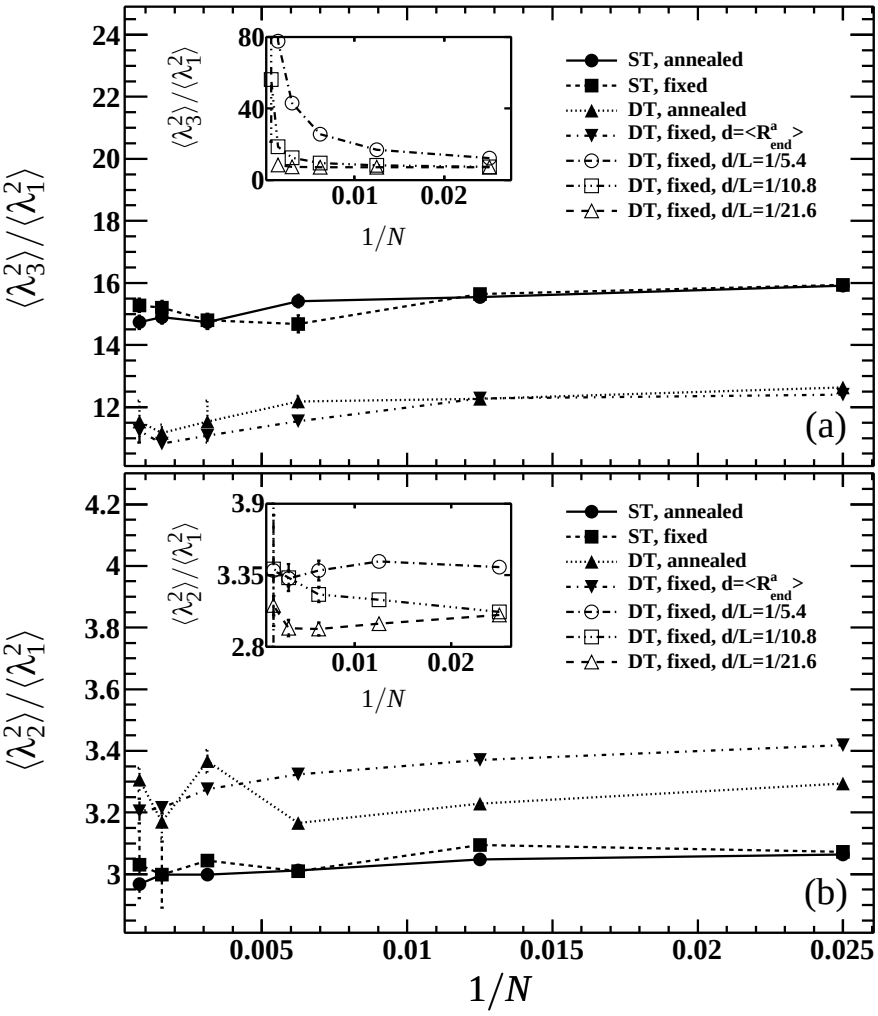


Figure 4.3: (a) $\langle\lambda_2^2\rangle/\langle\lambda_1^2\rangle$ vs $1/N$ for the fixed and annealed ST and DT chain with $d = \langle R^a_{\text{end}}\rangle$. Inset: $\langle\lambda_2^2\rangle/\langle\lambda_1^2\rangle$ vs $1/N$ for the fixed DT chain at various grafting distances. (b) $\langle\lambda_3^2\rangle/\langle\lambda_1^2\rangle$ vs $1/N$ for the fixed and annealed ST and DT chain with $d = \langle R^a_{\text{end}}\rangle$. Inset:$\langle\lambda_2^3\rangle/\langle\lambda_1^2\rangle$ vs $1/N$, for the fixed DT chain at various grafting distances d.
For the fixed ST chain, the ratios $\langle\lambda_1^2\rangle : \langle\lambda_2^2\rangle : \langle\lambda_3^2\rangle$ approach asymptotically $1 : 3.0 : 15.0$ and the annealed ST chain is characterised by exactly the same triaxial ellipsoidal shape. For the annealed DT chain the ratio $1 : 3.2 : 12.0$ is found showing that it prefers a more spherical shape with respect to the ST case and the SAW with $1 : 2.98 : 14.1$. [186] In case of the fixed DT chain, the ratio of the principal axis of inertia is strongly determined by the imposed grafting distance d, where the lateral stretching of the chain leads to its pronounced asphericity. Image adapted from [185].

an increase in d.

Other publications use different measures to describe the shape of the polymer chain. Two of these universal measures are the asphericity, $\langle A\rangle$, and the prolateness, $\langle P\rangle$. The asphericity value displays deviations from a spherelike shape, varying between 0 (sphere) and 1 (rod). It is defined as [187, 188]

$$\langle A\rangle = 1 - 3\left\langle \frac{\lambda_1^2\lambda_2^2 + \lambda_2^2\lambda_3^2 + \lambda_3^2\lambda_1^2}{(\lambda_1^2 + \lambda_2^2 + \lambda_3^2)^2} \right\rangle. \tag{4.2}$$

Prolate shapes may be distinguished from oblate ones by the ratio [187, 188]

$$\langle P \rangle = \left\langle \frac{(3\lambda_1^2 - R_{\mathrm{gyr}}^2)(3\lambda_2^2 - R_{\mathrm{gyr}}^2)(3\lambda_3^2 - R_{\mathrm{gyr}}^2)}{(R_{\mathrm{gyr}}^2)^3} \right\rangle, \tag{4.3}$$

which is bounded to the interval $[-0.25, 2]$. Negative values refer to predominantly oblate and positive values represent prolate shapes. For both the fixed and annealed ST chain, we find$\langle A_{\mathrm{ST}} \rangle = 0.445 \pm 0.003$ similar to the asphericity of a free SAW chain, $\langle A_{\mathrm{SAW}} \rangle = 0.447 \pm 0.011$ [189] and $\langle A_{\mathrm{SAW}} \rangle = 0.431 \pm 0.002$ [188]. The results for the prolateness of the two ST model systems, $\langle P_{\mathrm{ST}} \rangle = 0.57 \pm 0.004$, are again only slightly different from those obtained for a single polymer in free space, $\langle P_{\mathrm{SAW}} \rangle = 0.572 \pm 0.02$ [189] and $\langle P_{\mathrm{SAW}} \rangle = 0.541 \pm 0.004$ [188]. The asphericity and prolateness values of the annealed DT chain, $\langle A_{\mathrm{DT}}^{\mathrm{a}} \rangle = 0.36 \pm 0.012$ and $\langle P_{\mathrm{DT}}^{\mathrm{a}} \rangle = 0.37 \pm 0.015$, confirm its more sphere-like shape with respect to the ST and SAW chains as already indicated by the ratios of the eigenvalues of the gyration tensor. Regarding the impact of a fixed grafting distance d, we recover a rod-like, predominantly prolate organisation in the regime of high chain stretching with $\langle A_{\mathrm{DT}}^{\mathrm{f}} \rangle = 0.95 \pm 0.02$ and $\langle P_{\mathrm{DT}}^{\mathrm{f}} \rangle = 1.84 \pm 0.08$ for the largest investigated grafting distance, $d/L = 1/2.7$. A decrease in d allows the polymer chain to assume a more spherical shape resulting in a decrease of $\langle A_{\mathrm{DT}}^{\mathrm{f}} \rangle$ and $\langle P_{\mathrm{DT}}^{\mathrm{f}} \rangle$. In case of $d = \langle R_{\mathrm{end}}^{a} \rangle$, the asphericity and prolateness values of both the fixed and annealed DT chain are again equal. The regime of very small grafting distance, $d/L = 1/54$, is characterised by the most sphere-like conformations, $\langle A_{\mathrm{DT}}^{\mathrm{f}} \rangle = 0.25 \pm 0.015$ and $\langle P_{\mathrm{DT}}^{\mathrm{f}} \rangle = 0.19 \pm 0.02$. This can be understood by noting that the smaller the grafting distance the more the chain behaves as a ring polymer in the infinite chain limit $N \to \infty$, where the grafting distance acts as an immobile extra bond connecting the fixed chain ends to the interface. Thus, we find $\langle A_{\mathrm{DT}}^{\mathrm{f}} \rangle$ and $\langle P_{\mathrm{DT}}^{\mathrm{f}} \rangle$ to be similar to the asphericity and prolateness values determined for a free self-avoiding ring polymer, $\langle A_{\mathrm{RP}} \rangle = 0.262 \pm 0.001$ and $\langle P_{\mathrm{RP}} \rangle = 0.205 \pm 0.002$ [188].

In order to further quantify conformational differences associated with the polymer systems we compare monomer density profiles $\rho(z)$ as a function of the distance z (in lattice units) from the surface, as displayed in Fig. 4.4. The area under the curves has been normalised to unity. The higher localisation of both DT chains near the interfacial layer is again confirmed by the $\rho(z)$ curves. In this respect, we find that the distributions for DT chains are symmetric about a maximum position, whereas the ST chains exhibit a tail beyond this maximum in accordance with findings in two dimensions [180]. These features can be explained in terms of the loss of conformational entropy, which occurs only at one end for the ST chains but at two ends for the DT chains. In contrast to DT chains, ST chains are free at one end and therefore able to stretch out from the low-entropy interfacial layer. The end point mobility of the annealed cases, however, has no effect on the shape of the monomer density profiles. In addition, Fig. 4.4 reveals that an increasing grafting distance d leads to higher peak densities at lower peak positions, therewith indicating that the larger d the more concentrated are the fixed DT chains within the proximity of the surface.

A way to understand the impact of tethering sites on the local rearrangement of the polymer chains is to study correlations of intrachain segments. Consider two arbitrary bond vectors $\mathbf{b}_1 = (x_1, y_1, z_1)$ and $\mathbf{b}_2 = (x_2, y_2, z_2)$ connecting adjoined monomers. We evaluate the vector correlation coefficient c according to

$$c = \cos(\theta) = \frac{\mathbf{b}_1 \mathbf{b}_2}{|\mathbf{b}_1||\mathbf{b}_2|}, \tag{4.4}$$

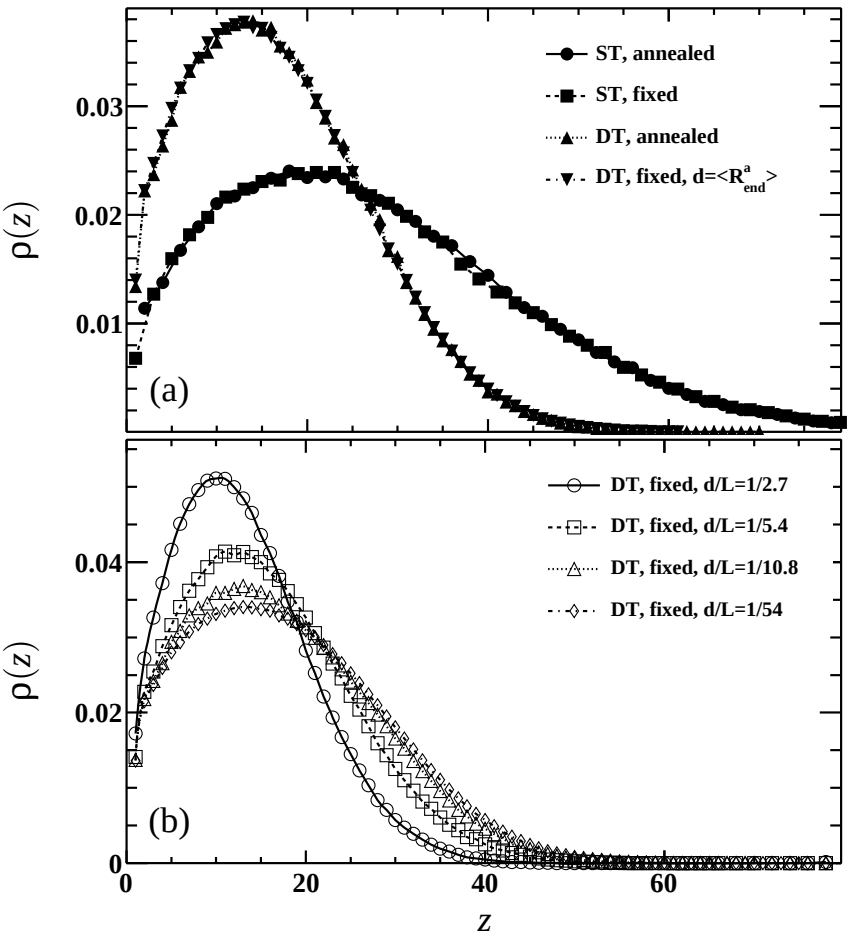


Figure 4.4: (a) Monomer density profiles $\rho(z)$ as a function of the distance z from the surface examplified for a chain of length $N = 160$ for the fixed and annealed ST and DT chain with $d = \langle R^a_{\text{end}} \rangle$. The distributions for DT chains appear symmetric about a peak position, whereas the ST chains exhibit a tail beyond this peak. In contrast to DT chains, ST chains are free at one end and therefore able to stretch out from the low-entropy interfacial layer. (b) $\rho(z)$ vs z is shown for the fixed DT chain at various grafting distances d. An increasing grafting distance d leads to higher peak densities at lower peak positions, therewith indicating that the larger d the more concentrated are the fixed DT chains within the proximity to the surface. Image adapted from [185].

where θ is the angle between $\mathbf{b}_1$ and $\mathbf{b}_2$. Fig. 4.5 displays the vector correlation coefficient in dependence of the relative contour length separating two bonds along the chain. For short contour length, there are high correlations because of the connectivity of the chain. However, while the vector correlations are becoming negative with contour length in case of both DT chains, they quickly decrease to zero for the ST chains. This behavior for intermediate and high contour lengths again confirms the ST chain end's stretching out of the low-entropy layer near the surface. In terms of vector correlations, there is no difference between the annealed and fixed ST chain as well as the annealed and fixed DT chain with $d = \langle R^a_{\text{end}} \rangle$. In case of the fixed DT chain, an increasing grafting distance prevents the bond correlations to drop below zero for intermediate contour length, as the lateral stretching due to the fixed tethering sites forces them to align with angles smaller

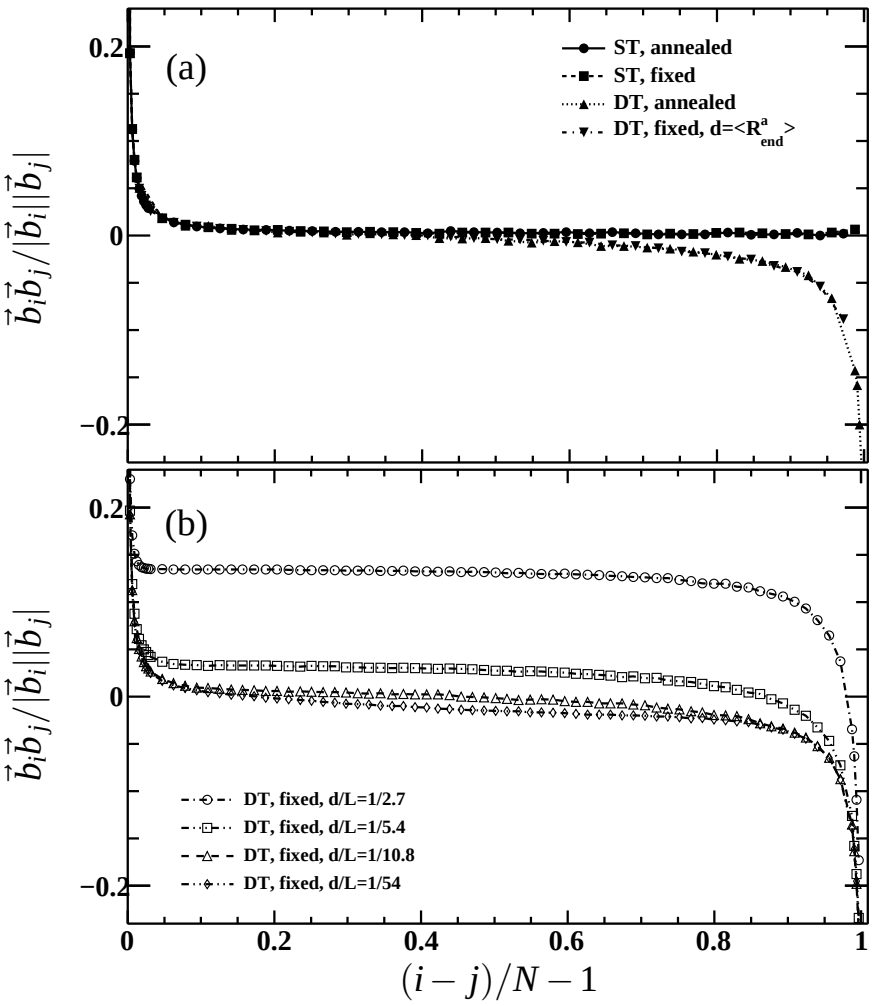


Figure 4.5: Vector correlation coefficient in dependence of the relative contour length separating two bonds along the chain for (a) the fixed and annealed ST and DT chain with $d = \langle R^a_{\text{end}} \rangle$ as well as (b) the fixed DT chain with various grafting distances d. The data shown is for $N = 320$. While the vector correlations are becoming negative with contour length in case of both DT chains, they quickly decrease to zero for the ST chains. Image adapted from [185].

than $\pi/2$. In case of low chain stretching (small grafting distance d), the correlations decay very fast to zero for intermediate contour lengths, whereas the chain's end segments are correlated again. This shows that, at low and intermediate stretching, the fixed DT chain behaves as a SAW, while the end segments are influenced by their mutual interaction.

The understanding of the impact of tethering on self-entanglement is of biological interest due to its occurrence in biopolymers such as viral or bacterial DNA [27, 31, 32, 190]. In order to quantify the complexity of chain overcrossings, i.e. to determine whether a polymer chain is highly or weakly self-entangled, we investigate the average crossing number [173, 191, 192] or mean number of "overcrossings" between bonds, when the polymer configurations are projected onto two dimensions [173, 193]. The mean average crossing number is then the average of the average crossing number over all possible configurations of a certain length N (and grafting distance d, respectively). In contrast to geometrical parameters such as for example the radius of gyration, this approach has the advantage

that it explicitly takes into account the connectivity of the polymer chain. However, while the type and complexity of entanglements in polymer chains are generally characterised by the analysis of all possible two-dimensional projections, we analyse projections on the xy-, xz- and yz-plane. This approach allows for the comparison between the so-determined mean average crossing number $\langle M \rangle$ and the mean bond-crossings in each of the three planes, $\langle M_{\mathrm{xy}} \rangle$, $\langle M_{\mathrm{xz}} \rangle$ and $\langle M_{\mathrm{yz}} \rangle$. Fig. 4.6 displays the mean average crossing number $\langle M \rangle$ for the annealed and fixed ST chain as well as for the annealed and fixed DT chain with $d = \langle R^a_{\mathrm{end}} \rangle$. In agreement with previous results for linear chains with excluded volume conditions [193], we find that the configurational averages of the average crossing number obey a power law in terms of the number of monomers, $\langle M \rangle \sim N^{\beta}$ with $\beta = 1.4 \pm 0.1$. Interestingly, as the more entangled structures will have a higher mean number of overcrossings [194], both ST chains as well as the annealed DT chain show the same degree of entanglement. In case of the fixed DT chain, the grafting distance introduces deviations from the above mentioned scaling law. In particular, an increase of the grafting distance d forces the polymer chain into more elongated configurations, therewith reducing the possibility of complex entanglements. While the mean bond crossings of fixed DT chains in the xy- and yz-planes, $\langle M_{\mathrm{xy}} \rangle$ and $\langle M_{\mathrm{yz}} \rangle$, are quantitatively the same, the mean number of overcrossings in the two-dimensional projection onto the xz-plane, $\langle M_{\mathrm{xz}} \rangle$, is different, which can be explained by noting that the two tethering sites are fixed at different y-positions on the xy-plane.

Polymer models can help predicting the relationship between the mean square physical distance and the genomic distance between two fluorescent markers on a genomic region of interest, therewith providing insight into specific folding patterns [5, 12].

In this work, we are interested in the effect that tethering, i.e. the fixation of one or more genetic loci to a membrane [2, 47], has on the relationship between mean square physical distance and the genomic distance. Classical polymer theory predicts that the mean square displacement between the end points of a polymer of length N scales as $\langle R^2_{\mathrm{end}} \rangle \sim N^{2\nu}$ with $\nu = 0.589$ in case of a SAW. Notably, this scaling law only holds for end-to-end distances, whereas we study intrachain distances. However, comparing the studied polymer systems to a SAW model, we divide out the leading order term $L^{2\nu}$ to analyse the ratio $\langle R^2 \rangle / L^{2\nu}$ as a function of the contour length in the inset of Fig. 4.7 for a polymer chain of length $N = 320$. The annealed and fixed ST chain show a power-law dependence of the mean square displacement in relation to the contour length on intermediate and large length scales. For the annealed and fixed DT chain with $d = \langle R^a_{\mathrm{end}} \rangle$, this scaling does not hold any longer, especially for large contour lengths.

Thus, in order to investigate double-tethered biological systems markers separated by large contour lengths should be taken into account. Regarding the crossover behavior of the fixed DT chain at various grafting distances d, we can identify two distinctive regimes, namely of small and of large contour lengths. In the regime of small contour lengths, the relationship between mean square physical distance and contour length is hardly influenced by tethering and we find that the above mentioned scaling form holds. For large contour lengths, the behavior of the mean square physical distance depends strongly on the imposed grafting distance between the tethers. The over-stretching of the polymer chain in case of large d leads to a crossover towards higher effective exponents in the already mentioned scaling form for large contour lengths. In contrast, small grafting distances force the chain's end regions to stay in proximity, ultimately resulting in a bending of the curves in Fig. 4.7.

Loop formation plays a central role for transcriptional control in eukaryotes [9], when

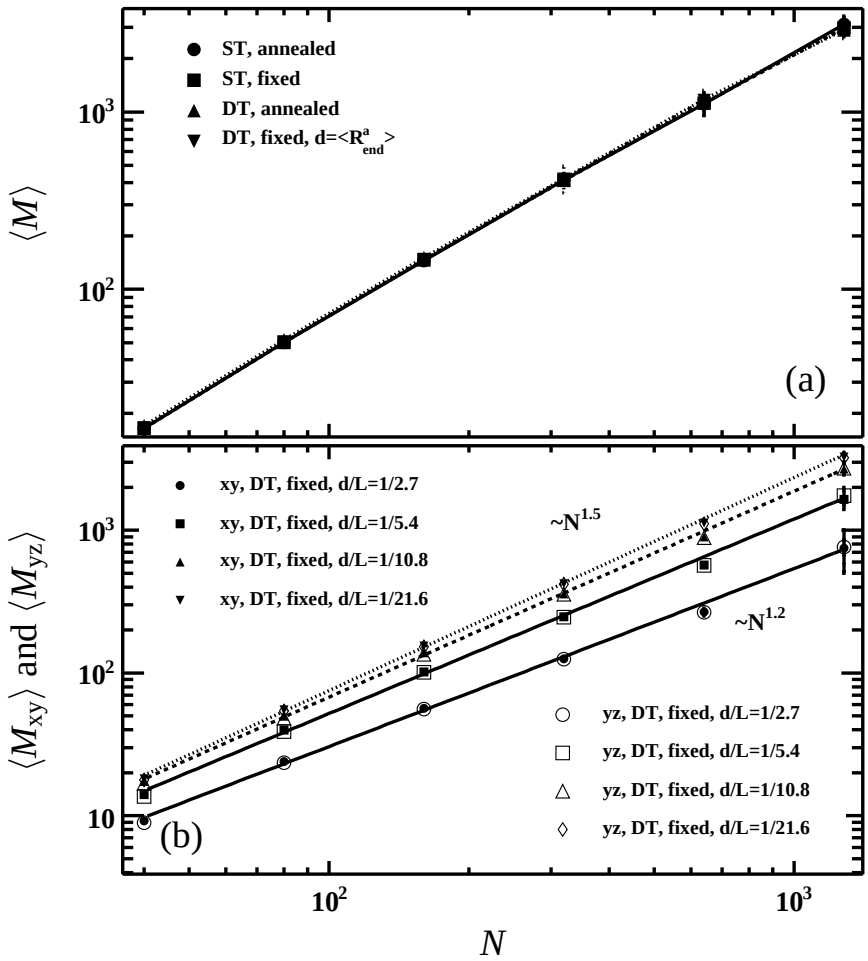


Figure 4.6: (a) Dependence of the mean average crossing number $\langle M \rangle$ on the chain length N presented on a double logarithmic plot for the annealed and fixed ST chain as well as the annealed and fixed DT chain with $d = \langle R^a_{\text{end}} \rangle$. The solid and dashed lines are power-law fits to the simulation data and indicate that $\langle M \rangle \sim N^\beta$ with $\beta = 1.4 \pm 0.1$. Thus, all four model systems show the same degree of entanglement as linear chains with excluded volume conditions. (b) Log-log plot of $\langle M_{\text{xy}} \rangle$ and $\langle M_{\text{yz}} \rangle$ vs N, respectively, for the fixed DT chain at various grafting distances, where fits are shown as solid and dashed lines. An increase in the grafting distance d forces the polymer chain into more elongated configurations, therewith reducing the possibility of complex entanglements and the effective exponent of the scaling behavior. Image adapted from [185].

colocalization of genetic sites results in gene expression or repression [82, 129, 195–201]. State-of-the-art technologies such as chromosome conformation capture techniques (3C/4C/5C/Hi-C [87, 127–129]) allow for the measurement of looping probabilities and distributions. In this work, we investigate how tethering sites can influence loop formation on the basis of diffusional collisions of the fiber.

Fig. 4.8 shows the contact probability for the four polymer models. Both ST chains as well as the annealed DT chain and its fixed counterpart with $d = \langle R^a_{\text{end}} \rangle$ display a power-law behavior of the colocalisation probability $p_{\text{c}}(L) \sim L^{-2.3 \pm 0.1}$. For tethered chains, the probability of specific contacts is only slightly decreased with respect to the free SAW chain. High chain stretching on the other hand has a strong impact on p_{c} since the formation of large loops is strongly surpressed. With respect to strongly tethered

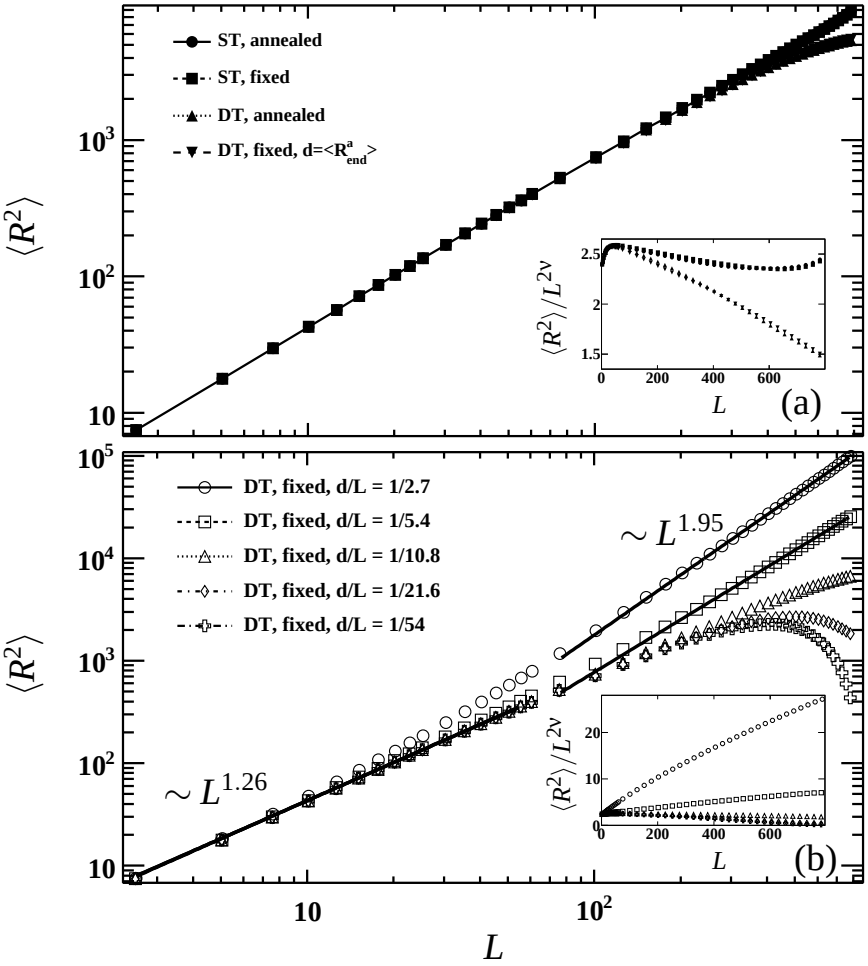


Figure 4.7: The relationship between the mean square physical distance separating two monomers and the corresponding contour length is exemplified for $N = 320$. (a) The solid line is a power-law fit to the simulation data with $\langle R^2 \rangle \sim L^{2\nu}$ and $\nu = 0.59 \pm 0.03$. Inset: Dividing out the leading order term $L^{2\nu}$, the impact of tethering is visible in the regime of large contour lengths, when there occur deviations from the scaling behavior. (b) The impact of the finite grafting distance d is shown. The power-law fits are displayed as solid lines and valid in the regime of small and large contour lengths, respectively. While the scaling $\langle R^2 \rangle \sim L^{2\nu_{\text{eff}}}$ with $\nu_{\text{eff}} = 0.6 \pm 0.03$ is recovered for small contour lengths, we find a crossover behavior for large contour lengths. Large grafting distances over-stretch the chain resulting in high effective exponents in the regime of large contour lengths, while small grafting distances lead to a bending of the curves in the same regime. Image adapted from [185].

biological systems, it can be stated that other mechanisms have to be taken into account to compensate the decreased contact probabiliy if colocalisation is needed to maintain biological functioning.

4.3.2 Dynamics of Single- and Double-tethered Polymer Chains

The dynamical properties of polymers arise from the stochastic monomer motions which are subject to connectivity constraints within the polymer chain. The first theoretical description of random motions of polymers within an isolated polymer coil was presented

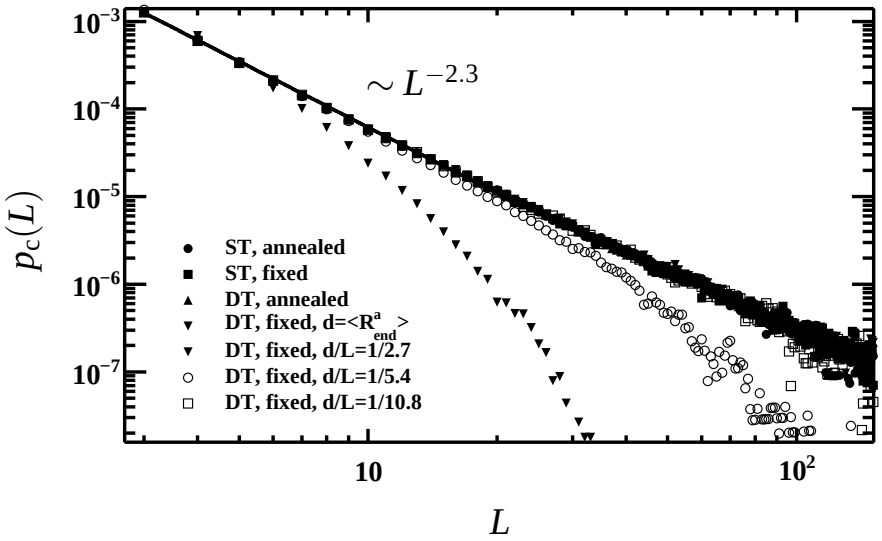


Figure 4.8: The relationship between the contact probability of two monomers forming a loop and the contour length separating them along the polymer chain is displayed for $N = 320$. For a SAW polymer chain the probability that two beads come into contact decreases with the separation according to $p_c(L) \simeq L^{-2.1}$. The solid line is a power-law fit to the simulation data with $p_c(L) \sim L^{-2.3\pm0.1}$. For tethered chains, the probability of specific contacts is only slightly decreased with respect to the free SAW chain. High chain stretching on the other hand has a strong impact on p_c since the formation of large loops is strongly surpressed. Image adapted from [185].

by Rouse as the classical bead-spring model [168] taking into account the entropic elasticity of polymer segments and the viscous friction of the solvent. In this work, we investigate the influence that excluded volume, tethering sites as well as an impenetrable flat surface exert on the dynamical properties of annealed and fixed ST and DT polymer chains.

We focus our attention on the mean square displacement of the monomers in the center of the chains, $g_1(t) = \langle[\mathrm{r}_{N/2}(t) - \mathrm{r}_{N/2}(0)]^2\rangle$. It is also instructive to have a look at the center-of-mass motion defined as $g_3(t) = \langle[\mathrm{r}_{\mathrm{CM}}(t) - \mathrm{r}_{\mathrm{CM}}(0)]^2\rangle$ as well as at the mean square displacement of monomers at the free ends of the chains $g_4(t) = \langle[\mathrm{r}_{\mathrm{end}}(t) - \mathrm{r}_{\mathrm{end}}(0)]^2\rangle$. Fig. 4.9 displays the mean square displacements for the four cases of the annealed and fixed ST and DT chain with $d/L = 1/2.7$.

In the case of all four studied model systems, the mean square displacement of the center monomer, $g_1(t)$, and of the free chain end(s), $g_4(t)$, display a distinct behavior for different time regimes, which are related to the relaxation time of the respective polymer chain. For $t \ll \tau_{\mathrm{int}}$, one finds subdiffusive behavior with $g_1(t) \sim t^{0.6\pm0.03}$ and $g_4(t) \sim t^{0.6\pm0.03}$. This result is similar to the dynamics of the free SAW chain, where scaling considerations in the short time Rouse model lead to $g_1(t) \sim t^{1/(1+(1/2\nu))}$ with $1/(1 + (1/2\nu)) \approx 0.54$. On large time scales, $t \gg \tau_{\mathrm{int}}$, the center node motion follows the motion of the center-of-mass, displaying normal Brownian motion with $g_3(t) \sim g_2(t) \sim g_1(t) \sim t$. In case of the fixed DT and ST chain, we find a leveling-off for $t \gg \tau_{\mathrm{int}}$ due to the motion-constraining tethering sites.

The dynamical properties of the four model systems are not only interesting from a theoretical point of view but they are also useful in better understanding the impact of nuclear achitecture on chromosome dynamics. In fact, studies of chromosome organisation in interphase nuclei revealed a high level or ordering where individual chromosomes occupy

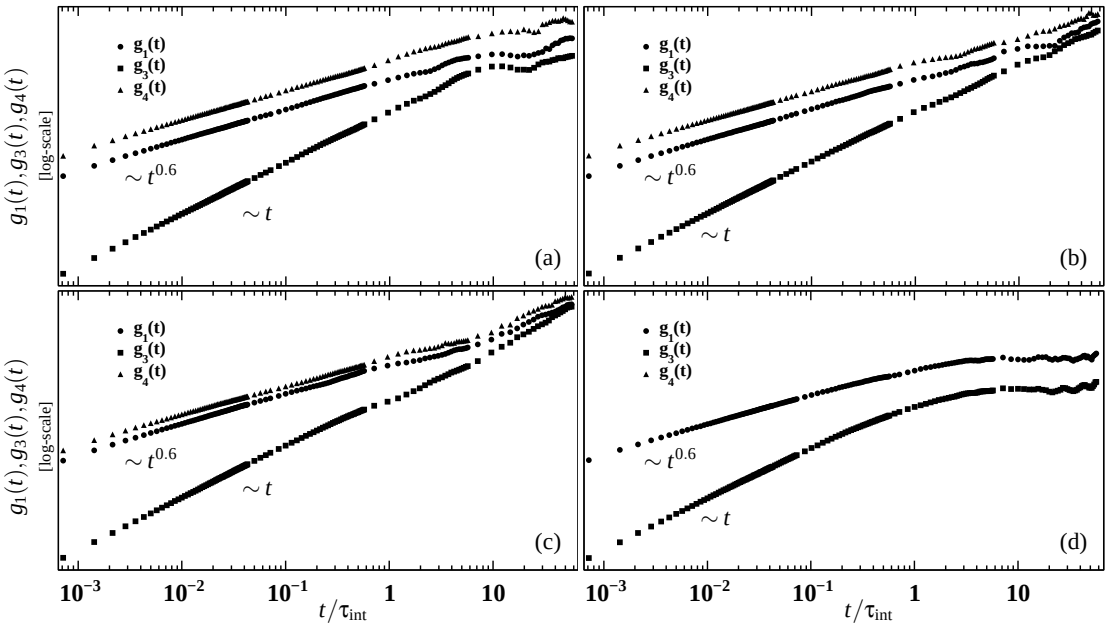


Figure 4.9: The motion of the central monomer $g_1(t)$, the center of mass $g_3(t)$ and the free chain end(s) $g_4(t)$ is shown for (a) the fixed ST, (b) the annealed ST, (c) the annealed DT and (d) the fixed DT chain with $d/L = 1/2.7$ and chain length $N = 160$, respectively. As for the SAW, the center-of-mass shows normal diffusion, $g_3(t) \sim t$, for all four model systems, while on intermediate time scales $g_1(t) \sim t^{0.6}$ and $g_4(t) \sim t^{0.6}$ holds. On large time scales the monomer motions follows the motion of the center of mass, displaying normal Brownian motion. In case of the fixed DT and ST chain, we find a leveling-off for $t \gg \tau_{\text{int}}$ due to the motion-constraining tethering sites. Image adapted from [185].

discrete territories and genomic regions are precisely positioned with respect to specific subnuclear positions [8, 10, 38, 202–204]. The study of the diffusional mobility of several different regions in the human genome as well as of individual genomic loci in drosophila and yeast has revealed subdiffusion in volumes that are much smaller than the nuclear size [38, 202–204]. Taking into account the diffusion results gained from our simulations, where tethering imposes subdiffusive behavior on short and a leveling-off on larger time scales, one might propose that physical attachment of chromatin to the nuclear envelop or to other substructures of the nucleus is a way of maintaining the organisation of chromatin in the nucleus. In fact, microscopy studies have indicated that specific regions of chromosomes are located in close proximity to the nuclear lamina, which in turn has led to the idea that certain genomic elements may be attached to it [10, 205]. Thus, future experiments tracking the diffusion behavior of more then one individual locus along the same chromosome can help to shed light into the nature of the attachment site(s) in the eukaryotic nucleus.

4.4 Conclusions

Experiments probing the physical organization of chromatin have found that both eukaryotic and prokaryotic cells have intricately structured chromosomes [5, 110, 206–212]. Their physical structure is expected to emerge from protein-protein or protein-DNA interactions which create chromosomal tethers and anchor the chromosome to the nuclear

membrane. [37–39, 172, 213–216].

Studies in bacteria [34–36], yeast [37–39], Drosophila melanogaster [40, 41], and mammalian cells [42, 43] confirm that dedicated chromatin regions, such as the centromere and telomeres, are restricted to specific subnuclear locations. In fact, the impact that the Rabl organization of yeast chromosomes has on their spatial organization and dynamics is studied in chapter 8. The coupled nature of tethering to gene expression was demonstrated by showing how tethering and removal of the tether from the nuclear membrane regulates gene expression [10, 217–220].

In this chapter, we have developed a tethered-polymer model to determine the impact of tethering on conformational and dynamical properties of single- and double-tethered chains. The simple tethered-polymer model presented in this chapter provides a framework for understanding the physical structure of a tethered chromatin fiber and might constitute a building block for more complex models. The emphasis of this chapter is on the interpretation of the experimentally accessible quantities, such as distance distributions between fluorescently labeled loci. We believe that the study of the four polymer systems might facilitate the interpretation of future experiments (FISH, FROS, ParB-*par*S system) with multiple marked loci which try to resolve chromosome organization in more detail.

Chapter 5

Semiflexible Ring Polymers in Confinement

Implications for Biopolymer Packaging

References

The results presented in this chapter are published as and adapted from

- M. Fritsche and D.W. Heermann (2011), *Confinement Driven Spatial Organization of Semiflexible Ring Polymers: Implications for Biopolymer Packaging.* Soft Matter, 7: 6906-13.

We thank Bulbul Chakraborty and Songling Li for fruitful discussion.

Chapter Summary

The conformational properties of a semiflexible ring polymer in confined spaces are investigated. Taking into account the competing interplay between configurational entropy, bending energy and excluded volume, we elucidate the role that different geometrical constraints can play in shaping the spatial organization of biopolymers.

While elongated, rod-like geometries reduce the amount of chain overcrossings and induce a pronounced ordering with respect to the long axis of the surrounding envelope, there exists no preferred orientational axis in case of spherical confinement. Upon increasing the system density and rigidity of the chain, the polymer migrates from the center of the accessible space towards the surrounding surface forming a spool-like structure know from DNA condensation within viral capsids. The existence of distinct loop sizes for different confining geometries might influence the colocalization of genomic sites in biopolymers necessary for the genome-wide coordination of gene expression. Thus, the advantages of specific geometric constraints such as spherical confinement of viral DNA in a capsid or the rod-shaped envelope of the circular chromosome in Escherichia coli could be one driving force for controlling proper biological functioning.

5.1 Introduction

Biological macromolecules are organized into confining geometries that have a crucial impact on their overall shape. A prominent and extensively studied example of strong confinement is viral DNA, which is tightly packaged into a capsid whose size is comparable to the DNA persistence length [221–226]. In fact, the shape of biopolymers has important implications for their functionality [31, 32]. With respect to transcriptional regulation in eukaryotes and prokaryotes, the shape of the DNA polymer modulates its accessibility to proteins/enzymes or to genes on genomically distant DNA segments and thus triggers gene expression or repression [9, 15, 223, 227–229].

In confinement, the equilibrium form and dynamics of semiflexible biopolymers is governed by the competing interplay between configurational entropy, excluded volume and bending energy [31, 32, 230–232]. The persistence length l_p is a measure for the competition between elastic against entropic contribution and it is the biopolymer's semiflexibility, that turns it into a fascinating material [232–235].

Linear semiflexible polymers have previously been the subject of investigations in free space as well as under conditions of strong confinement where the confining cavity is much smaller than the equilibrium coil of the biopolymer which is the case in viral capsids and bacterial envelopes [230]. Furthermore, theoretical investigations of linear semiflexible polymers in channels [236–238], on spherical surfaces [239–241] and on two-dimensional planes [230, 235] have been undertaken.

However, nature not only imposes geometrical constraints on biopolymers by confinement through cell membranes, the cell nucleus, the bacterial nucleoid or viral capsids but also exploits the advantages of certain underlying topologies such as the ring structure. In fact, most of the short genomes as well as plasmids are circular [9, 27] and actin or actin bundles can also form rings [242–245]. Alongside confinement the ring topology alone represents an important constraint for the polymer's shape which was studied in free space both by theoretical [27, 56, 165, 233, 234, 246] and experimental approaches [235, 247].

Yet, the physics of confined circular biopolymers is a field in its infancy and few studies have addressed semiflexible ring polymers in confinement so far. Only recently, the internal structure of semiflexible polymer rings in weak spherical confinement established by an impenetrable shell of about the average size of the polymer ring has theoretically been investigated [31, 32]. It was found that weak confinement induces buckling and a conformational transition to a figure eight form.

In this chapter, we investigate the shape and ordering of a semiflexible ring polymer strongly confined to two different geometries, namely a rectangular and a square box. Therefore, we are able to study the impact of different geometrical constraints on the competition between entropy and bending energy when the polymer's flexibility is varied.

We believe that the investigation of these model systems is important for understanding biological systems of larger complexity such as the behavior of DNA inside phages or the spatial organization of the bacterial nucleoid in E. coli [5, 112].

5.2 Modeling

Approximating biological "storage" such as the bacterial nucleoid or viral capsids [23, 24], we consider a semiflexible ring polymer in cubic confinement of dimensions $D_\mathrm{x} = D_\mathrm{y} = D_\mathrm{z}$ or within a rectangular cuboid of $D_\mathrm{x} = D_\mathrm{y} = D_\mathrm{z}/8$ in agreement with the aspect ratio of the E. coli nucleoid [23, 24].

To generate polymer conformations we employ the bond-fluctuation method [166], which has been applied successfully to model the static and dynamical properties of polymer systems in several investigations [146]. The reader is referred to chapter 3 for more details on the algorithm.

Semiflexible polymers may be characterized by their persistence length l_p, which is the typical length scale over which the chain backbone loses information about its direction due to thermal fluctuations [139]. In a recent study, alternative definitions were investigated for polymers with bottle-brush architecture [248] showing that standard definitions of persistence length do not describe the local "intrinsic" stiffness of real polymer chains. Thus, in this work, we interpret the decay of the orientational correlation function in terms of an effective "quasi" persistence length reflecting global conformational properties rather than local intrinsic stiffnesses.

The bending energy H_b can be expressed as [249]

$$H_\mathrm{b} = \frac{l_\mathrm{p}}{2b} \sum_{i=1}^{N-1} (\mathbf{u}_{i+1} - \mathbf{u}_i)^2, \tag{5.1}$$

in a lattice representation such as the BFM. Here, N is the total number of monomers in the chain und $\mathbf{u}_i = \left(\frac{\mathbf{R}_i - \mathbf{R}_{i-1}}{|\mathbf{R}_i - \mathbf{R}_{i-1}|} + \frac{\mathbf{R}_{i+1} - \mathbf{R}_i}{|\mathbf{R}_{i+1} - \mathbf{R}_i|} \right) \Big/ \left| \frac{\mathbf{R}_i - \mathbf{R}_{i-1}}{|\mathbf{R}_i - \mathbf{R}_{i-1}|} + \frac{\mathbf{R}_{i+1} - \mathbf{R}_i}{|\mathbf{R}_{i+1} - \mathbf{R}_i|} \right|$ is a discrete realization of $\delta \mathrm{R}(s)/\delta s$, the unit tangent vector at arclength s, where $\mathrm{R}(s)$ is the position vector [230]. Periodic boundary conditions $\mathbf{u}_N = \mathbf{u}_1$ ensure the closure of the ring. In the following, all energies are measured in units of $k_\mathrm{B}T$.

In order to generate thermodynamically equilibrated polymer conformations we use the Metropolis Monte Carlo method [146]. Since subsequently created conformations are highly correlated, we determine, for each set of parameters (persistence length l_p, polymer length N, box geometry), the autocorrelation function [146] of the squared radius of gyration. Then, the integrated autocorrelation time τ_int is computed by applying the windowing procedure introduced by Sokal [164], which is explained in chapter 3. We

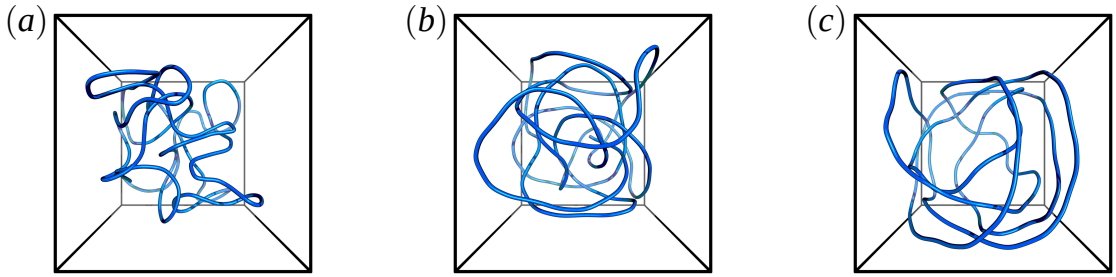


Figure 5.1: "Snapshot" of chain configurations. With increasing persistence length (a) $L/l_\mathrm{p} = 160$, (b) $L/l_\mathrm{p} = 8$ and (c) $L/l_\mathrm{p} = 2$ a spool-like structure emerges in cubic confinement $D_\mathrm{x} = D_\mathrm{y} = D_\mathrm{z}$. Image adapted from [33].

consider two subsequent conformations as uncorrelated after $5\tau_\mathrm{int}$ Monte Carlo steps and create 10 000-100 000 independent configurations.

Studying polymer rings of lengths $N = 80, 160$ and $N = 320$, we set up the linear dimensions D_x, D_y and D_z of the confining geometry such that the radius of gyration R_gyr of the unconfined semiflexible chain is larger than at least two of the linear box sizes. Cubic confinement is defined by $D_x = D_y = D_z = 20$, while $D_x = D_y = D_z/8 = 10$ holds for the rectangular envelope, retaining the same accessible volume for both geometries. The range of bending stiffnesses is varied from a totally flexible chain up to the stiff limit, representing a planar ring. Figure 5.1 shows three "snapshot" polymer configurations at low, intermediate and high chain flexibility.

5.3 Results and Discussion

5.3.1 Confinement Induces "Orientational Memory"

The internal structure of polymers can be investigated by computing the correlation of two unit tangent vectors $\mathbf{u}(i)$ and $\mathbf{u}(j)$ separated by a distance $|i - j| \in [0, L]$ along the polymer backbone

$$C(i, j) = \langle \mathbf{u}(i)\mathbf{u}(j) \rangle. \tag{5.2}$$

Tangent-tangent correlations provide details about the relative orientation of the polymer's whole contour line and can be measured in experiments [250–253]. Figure 5.2 shows the mean tangent-tangent correlation $C(i, j)$ for an unconfined semiflexible ring polymer, while Figure 5.3 displays the correlation function in cubic and rectangular confinement.

Due to the ring form, the correlation function is symmetric around the peak point of anticorrelation at $L/2$ along the polymer backbone [31, 32]. In case of a rigid ring, the tangent-tangent correlation yields $C(i, j) = \cos(2\pi|i - j|/L)$ [32]. Without confinement, a crossover takes place between the correlations of a rigid ring in the stiff regime and the exponential decay of the correlation function also observed for linear semiflexible polymers [32, 246]. This can be understood by noting that with increasing flexibility the topological constraint of the ring form becomes less and less important.

In confinement, the internal structure of a semiflexible polymer depends on the complex relationship between the linear size of the confining geometry, the polymer's contour length as well as its persistence length. A confined ring polymer is not able to adopt the elliptical configurations of free polymers if the major axis of the ellipse exceeds the available space in confinement. Instead buckling into a banana-like ellipse takes place [31]. In the regime,

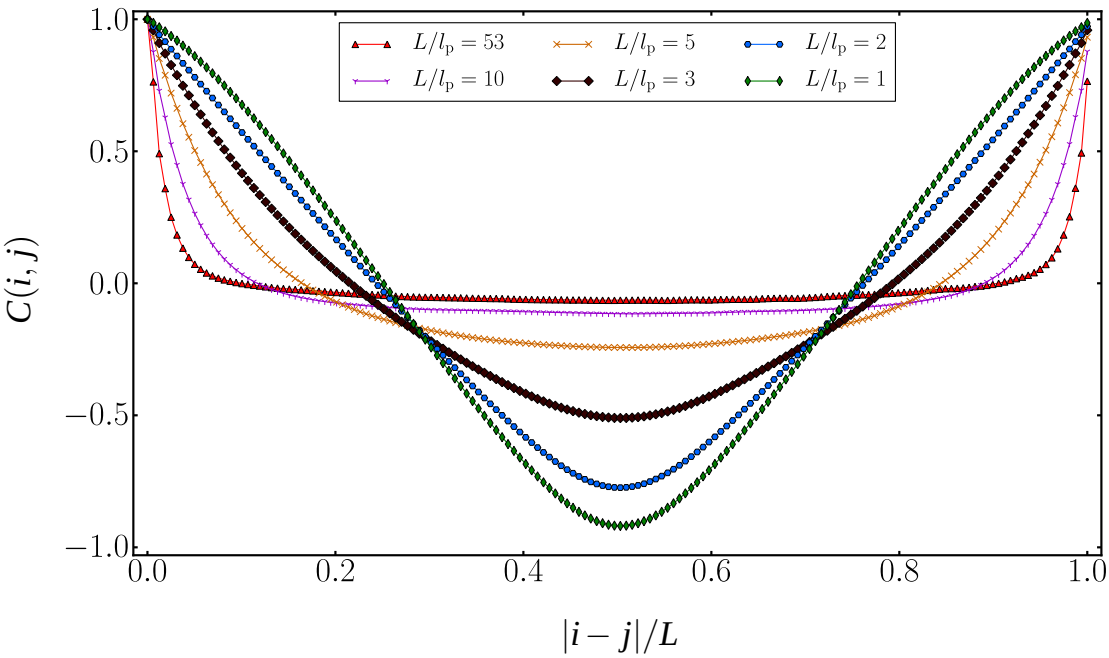


Figure 5.2: Mean tangent-tangent correlation function $C(i,j) = \langle \mathbf{u}(i)\mathbf{u}(j)\rangle$ along the polymer backbone $|i-j| \in [0, L]$ for a free semiflexible polymer ring over various persistence lengths L/l_p. The correlation of an unconfined polymer ring smoothly crosses over from the tangent-tangent correlation along an "ellipse" in the stiff regime to an exponential decay with periodic boundary conditions in the flexible limit in agreement with [32]. Image adapted from [33].

where l_p is much smaller than the dimensions of the surrounding box, the free energy of the polymer is dominated by entropy and the correlations resemble those of an unconfined semiflexible polymer as is displayed in Figure 5.2. Consequently, the correlation function gradually approaches a symmetric exponential decay with a length scale comparable to the bare persistence length [31, 32].

For l_p larger than or comparable to the size of the confining geometry, the correlation functions exhibits an oscillatory character, where a negative value of $C(i,j)$ indicates a reflection in the orientation of the tangent vector, as can be seen in Figure 5.3. Thus, the oscillations reflect the constraints that confinement imposes on the polymer's internal structure and depend strongly on the geometry of confinement as illustrated in Figure 5.3. Such oscillatory behavior of the correlation function has been observed in experiments: The tangent-tangent correlation function of actin filaments trapped in narrow channels display undulations and an effective persistence length deduced from it shows deviations from the bare persistence length [250–253]. A theoretical study of linear semiflexible polymers in 2D confinement used a mean-field approach to calculate the effective persistence length finding it to be strongly increased in the stiff regime [230]. Our results suggest that strong confinement and semiflexibility induce an enduring "orientational memory" where the correlation function doesn't decay to zero but shows a specific frequency of undulations depending on the geometry of the confining envelope. Figure 5.3 shows that the frequency of tangent-tangent correlations increases with increasing monomer to volume ratio since the polymer chain is more and more forced to adopt undulating crumbled configurations in order to fit in three-dimensional space ultimately leading to higher frequencies. Moreover,

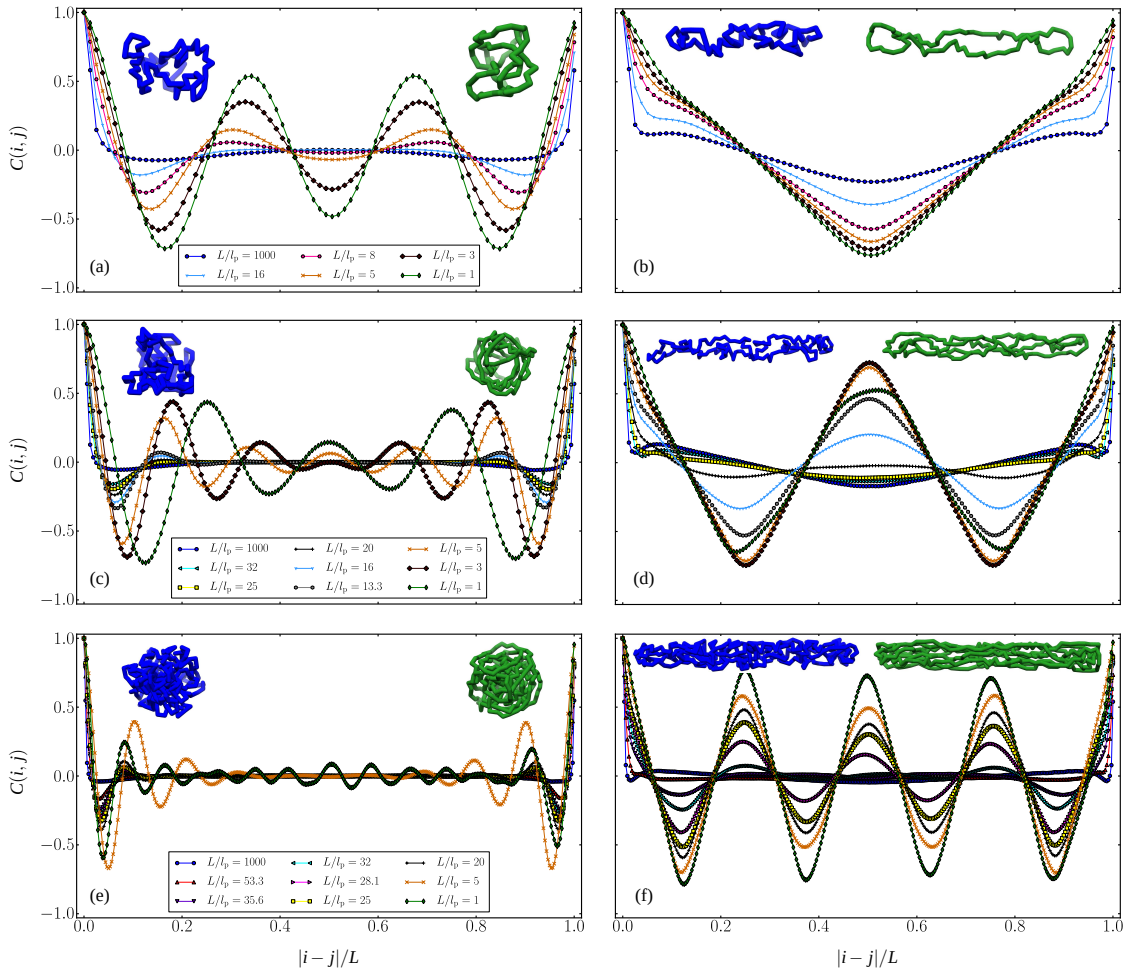


Figure 5.3: Mean tangent-tangent correlation function $C(i,j) = \langle \mathbf{u}(i)\mathbf{u}(j)\rangle$ along the polymer backbone $|i-j| \in [0, L]$ for a semiflexible polymer ring over various persistence lengths L/l_p in (a), (c), (e) cubic and (b), (d), (f) rectangular confinement and for (a), (b) $N = 80$ and (c), (d) $N = 160$ and (e),(f) $N = 320$ monomers. In confinement, the internal structure of a semiflexible polymer depends on the complex relationship between the linear size of the confining geometry, the polymer's contour length as well as its persistence length. For l_p larger than or comparable to the size of the confining geometry the correlation function exhibits an oscillatory character, where a negative value of $C(i,j)$ indicates a reflection in the orientation of the tangent vector. Image adapted from [33].

at fixed system density cubic confinement induces higher oscillation frequencies than rectangular confinement. This can be understood by considering the organization of the ring polymer in the confined space. Fig. 5.3 illustrates "snapshot" configurations in the flexible and the stiff regime. With increasing bending stiffness elongated geometries induce a polymer organization with respect to the long axis of the available space therewith requiring fewer windings to package the polymer. In contrast, the number of windings necessary to accomodate a ring polymer of equal size in squared confinement leads to a "spool" like arrangement which is reflects in the higher tangent-tangent correlation frequencies.

5.3.2 The Geometry of Confinement Mediates Loop Formation of Certain Lengths

The looping of polymers allows for reactions between chain segments that would otherwise be too distant to interact. On the one hand, loop formation is an elegant way for a polymer to fit into strongly confined spaces. On the other hand, polymer loops are particularly important in biology where their formation is key for the genome-wide coordination of gene expression, when colocalization of genetic sites results in gene expression or repression [82, 129,195–201]. To this end, chromosome conformation capture techniques (3C/4C/5C [127–129] and Hi-C [87]) allow for the measurement of looping probabilities and distributions.

In this chapter, we investigate how the competition between configurational entropy, excluded volume and bending energy influences the colocalization probability of sites on the basis of diffusional collisions of the circular fiber. Figure 5.4 shows the contact probability p_c for different system densities in both types of confinement.

Independent of the confining geometry, the probability of small loop formation (as opposed to intermediate and large loops) is strongly increased in the flexible regime. In contrast, with increasing bending stiffness the increasing energy cost of bending leads to a strong decrease in the probability of small loop formation. However, while loop formation in free space becomes more and more difficult with increasing chain stiffness due to the development of a rigid planar ring, the impact of the surrounding envelope changes this situation. Confinement induces an oscillatory behavior of the colocalization probability clearly favoring loops of certain lengths. The periodicity of the oscillations non-trivially depends on the system density and the geometry of confinement. Since the ring polymer assumes a more and more "spool"-like organization with increasing system density and bending stiffness, loops of all length scales can develop. Since the ring polymer arranges with respect to the long axis of the rectangular confinement, at low system density

The existence of distinct loops sizes for different confining geometries has important implications when considering the physical process of colocalization of genomic sites: Nature might deliberately exploit (by imposing selective pressure) the advantages of certain geometric constraints such as spherical confinement of viral DNA in a capsid or the rod-shaped envelope of the circular chromosome in E. coli to mediate long-range interactions between distant sites, thus facilitating the coordination of transcription and other processes acting on DNA.

5.3.3 Rod-Shaped Geometries Induce Less Chain Overcrossings Than Cubic Ones

Genome organization is not random [200, 254] and the highly packaged DNA polymer or chromatin fiber need to be locally unpacked for biological functioning [82, 197, 199, 200]. However, the tight spatial confinement produces a significant amount of entanglement hampering the accessibility and processing of the genetic information [112, 255]. While tangent-tangent correlations provide details about the relative orientation of the whole contour line of a polymer, the mean average crossing number mACN or mean number of "overcrossings" measures the complexity of self-entanglement of a polymer chain [173, 192, 193].

Thus, it is key to understand the impact of semiflexibility and confinement on the degree of self-entanglement. Projecting a three-dimensional polymer configuration into a plane defined by a normal vector $\mathbf{n}$ results in a two-dimensional curve which may exhibit crossings. Averaging the number of crossings over all angular perspectives given by all

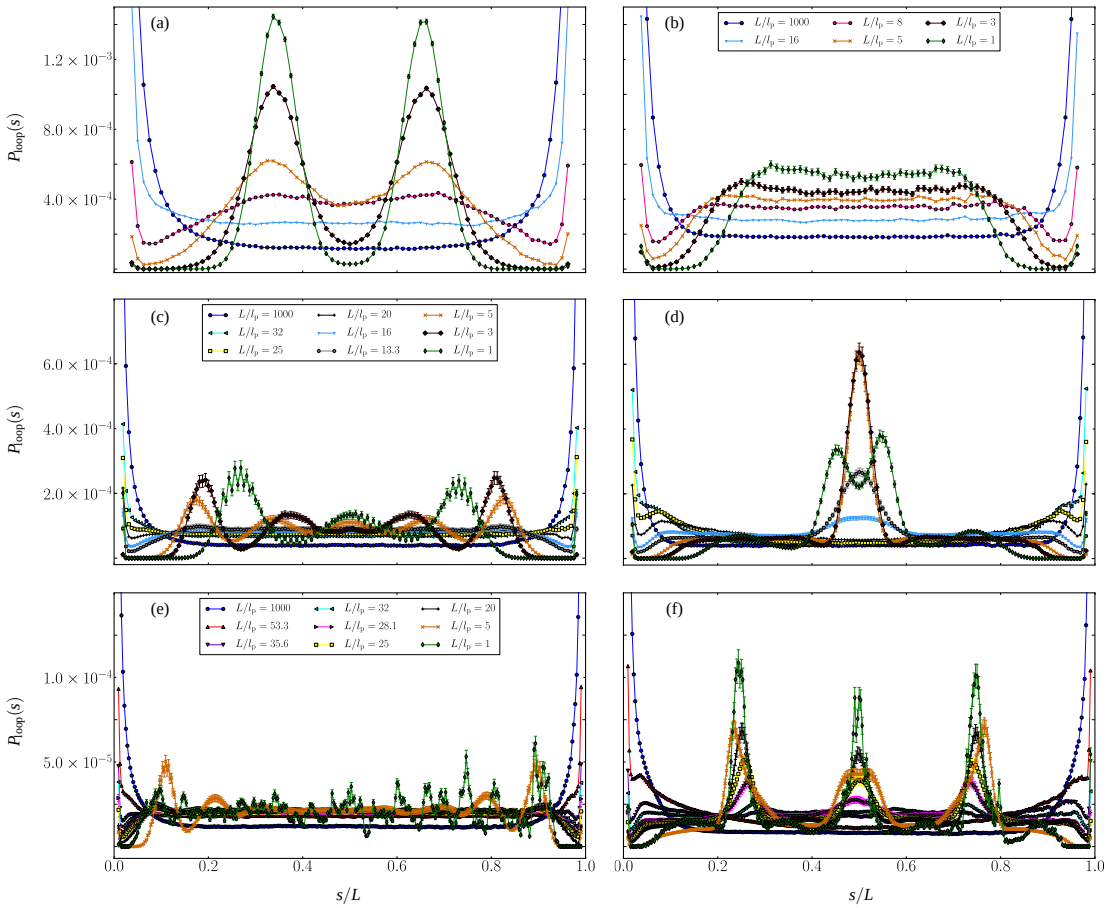


Figure 5.4: The relationship between the contact probability of two monomers forming a loop and the contour length $s \in [0, L]$ separating them along the polymer chain (a), (b) $N = 80$ and (c), (d) $N = 160$ and (e), (f) $N = 320$ in both cubic (a,c,e) and rectangular (b,d,f) confinement. Confinement induces an oscillatory behavior of the colocalization probability clearly favoring loops of certain sizes distributed over all lengths scales at high chain stiffness. The occurrence of distinguished loops sizes for different confining geometries and stiffnesses might be an important driving force for biological processes such as the colocalization of genomic sites to ensure error-free gene expression. Image adapted from [33].

possible normal vectors defines the average crossing number ACN. To calculate the average crossing number of polymer configurations generated by Monte Carlo simulations we follow [256]. The mean average crossing number mACN is then obtained by averaging over the average crossing number of all possible polymer configurations.

Figure 5.5 shows the mean average crossing number for an unconfined semiflexible ring polymer as well as in cubic and rectangular confinement. In free space, a perfectly rigid circular polymer forms a planar ring with no chain overcrossings. With increasing flexibility, small deviations out of the planar elliptical shape occur and induces an increasing amount of chain entanglement.

The mean average crossing number of confined ring polymers grows with chain flexibility, too. However, strongly confined ring polymers cannot form the desired elliptical configurations of free polymer rings due to the limited amount of available space. Consequently, increasing flexibility induces more and more undulations which lead to a larger

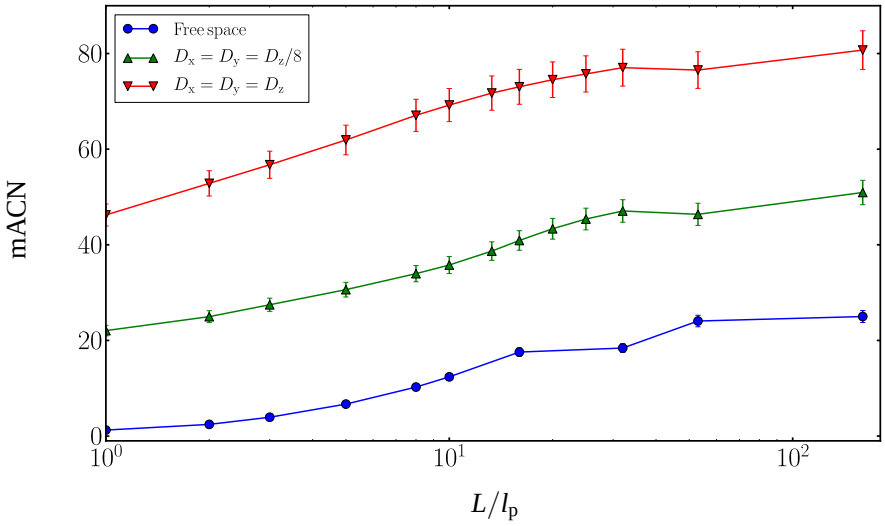


Figure 5.5: The relationship between the mean average crossing number mACN and the persistence length L/l_p for a free and a confined ring polymer of $N = 160$ monomers. Cubic geometries induce a larger amount of chain overcrossings than rectangular ones. An elongated geometry with a distinguished axis imposes chain stretching with respect to the long axis reducing the amount chain self-crossings necessary for compaction. Image adapted from [33].

mean number of overcrossings in confinement than in free space. Moreover, the geometry of confinement has a strong impact on the mean average crossing number as has already been pointed out by the frequency behavior of the tangent-tangent correlation function.

Figure 5.5 shows that cubic geometries induce a larger amount of chain overcrossings than rectangular ones for the same chain flexibility. This can be understood by noting that a polymer ring is forced to intertwine with itself to fit inside a strong confinement. An elongated geometry imposes stretched chain conformations with respect to the long axis reducing the necessary number of compacting chain overcrossings.

5.3.4 Rod-Shaped Geometries Induces Strong Segmental Ordering with Respect to the Confining Envelop

A way to further investigate overall shape changes is to consider the polymer arrangement inside the confining geometry as a liquid crystal. Orientational ordering with respect to the hard walls of the confining geometry can be assessed by defining an order parameter S according to

$$S = \frac{1}{N}\left\langle \sum_{i=1}^{N} \left(\frac{3}{2}\cos^2\theta_i - \frac{1}{2} \right) \right\rangle, \tag{5.3}$$

where θ_i is the angle between between the normalized chain segment $\mathbf{s}_i$ and a unit vector $\mathbf{n}$ along the confining envelope [257]. In this work, the relevant orientational variables are S_x, S_y and S_z, where the unit vector is taken to be $\mathbf{n}_x$, $\mathbf{n}_y$ and $\mathbf{n}_z$ along the x-, y- and z-axis, respectively. The order parameter varies from -0.5 to 1. The case $S = 1$ corresponds to completely planar packaging of the segments with respect to the confining envelope. When $S < 0$, the chain segments exhibit predominantly orthogonal alignment, while the limit $S = -0.5$ indicates an ideal orthogonal packing near the hard walls.

Figure 5.6 shows the segmental order parameters S_x and S_z for both the cube and

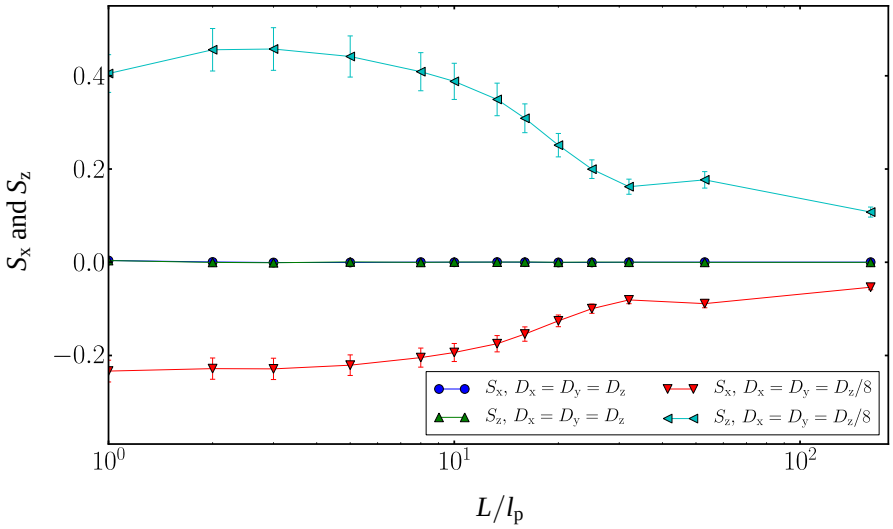


Figure 5.6: The segmental order parameters S_x and S_z for a polymer of $N = 160$ monomers. It provides a measure for the orientational ordering of chain segments with respect to the walls of the confining cavity. In case of rectangular confinement, the deviation from the cubic geometry induces strong segmental ordering parallel to the long axis of the confining envelope with increasing ring stiffness, while cubic confinement gives rise to equal alignment with regard to all three constraining axes. Image adapted from [33].

rectangular cuboid at various persistence lengths. In case of the rod-shaped confinement, the deviation from the cubic geometry induces strong segmental ordering. The polymer segments tend to orient predominantly parallel to the confining envelope. In fact, most chain segments align parallel to the long axis leading to $S_z > 0$ and $S_x < 0$.

In case of cubic confinement, there exists no preferred directional axis. Even though chain segments tend to occupy near-wall regions and align parallely with them, they equally show this behavior with respect to all three directional axis. Consequently, Figure 5.6 illustrates the segmental ordering parameter S fluctuating around zero as is the case for semiflexible SAW ring polymers in free space.

Figure 5.7 demonstrates the importance of chain rigidity and system density for orientational ordering and packaging at the surface of the confining geometry. In agreement with [258], a short and flexible polymer chain concentrates at the center of the confining space being less orientationally ordered, whereas upon increasing the length and rigidity of the chain, it migrates towards the surrounding surface. Regarding DNA condensation within viral capsids, our results suggest that the conformations of biopolymers obtained for the case of long and stiff rings in strong cubic confinement resemble the well-known spool-like, torus structure [112].

5.4 Conclusions

The ring form, semiflexiblity and confinement are key features in a wide range of biophysical systems [26], among which we can name the segregation of the compacted circular genome of E. coli discussed in chapter 11 or the storage of viral DNA [28, 29].

In this chapter, we have contributed to a deeper understanding of the basic properties of such biological systems. We have investigated the conformational properties of

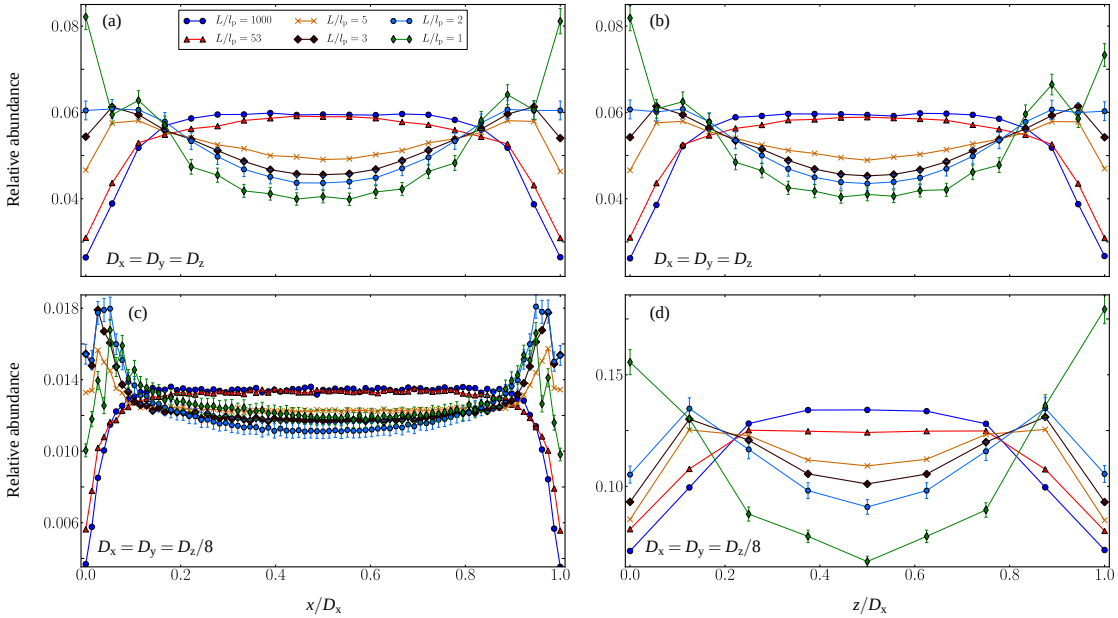


Figure 5.7: The relative abundance of chain segments with respect to two axes of the confining geometries for a polymer of $N = 160$ monomers. Instead of occupying the central region between the surrounding hard walls to maximize conformational entropy, the polymer is forced to parallely align towards the near near-wall regions. Image adapted from [33].

a semiflexible ring polymer confined to a cube and rectangular cuboid as a model for biopolymers packaged into different geometries. Taking into account the competing interplay between configurational entropy, bending energy and excluded volume, we elucidate the role that different geometrical constraints can play in shaping the spatial organization of biopolymers.

The nature of the tangent-tangent correlation is found to be highly influenced by non-trivial conformations that a ring polymer adopts due to strong confinement. The oscillatory behavior of the tangent-tangent correlation function becomes more pronounced with increasing system density and chain rigidity. These consequences of ring closure and strong confinement are underpinned by a recent work [26]. Witz and coworkers have probed the conformation of circular DNA molecules in two dimensions by means of atomic force microscopy. In agreement with our theoretical result, they find the DNA chains to adopt specific conformations illustrated in Fig. 5.8 in order to limit the bending-energy penalty. Notably, the authors point out that, under space constraint, ring and linear molecules behave similarly and, to a certain extent, share a set of conformations.

By investigating the mean number of chain overcrossings and introducing a segmental order parameter, we gain further insights into the position of polymer segments with respect to each other and with respect to the geometry of the confining envelope. Our results suggest that the geometry of confinement has a major influence on the packaging of biopolymers. While rod-shaped constraints lead to less chain self-entanglement and a higher ordering with respect to the long axis of the surrounding envelope, there exists no preferred orientational axis in case of spherical confinement. Upon increasing the system density and chain rigidity, the polymer chain migrates towards the surrounding surface. We find that the conformations of biopolymers obtained for the case of long and stiff chains

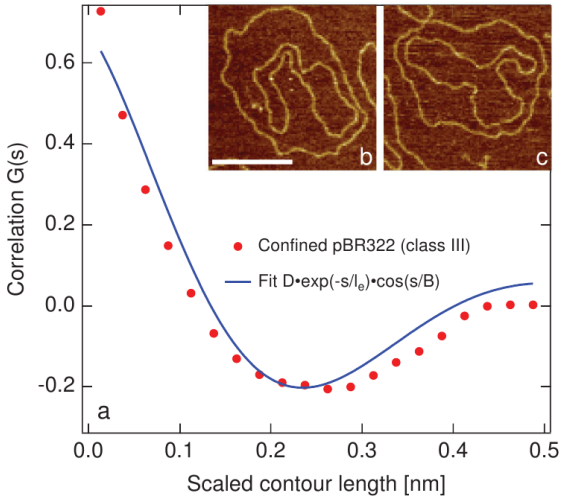


Figure 5.8: (a) The bond correlation function $G(s)$ for confined ring DNA molecules is compared to the analytical approximation proposed by Liu and Chakraborty [230]. Representative images of confined conformations are shown in (b) and (c), where the white scale bar is 250 nm. Image adopted from [26].

in strong cubic confinement resemble the well-known spool-like, torus structure observed for DNA condensation within viral capsids [112].

Since loop formation might be key for the genome-wide coordination of gene expression, we investigate the colocalization probability of strongly confined semiflexible ring polymers. Most importantly, depending on the geometry of confinement we find an oscillatory behavior of the colocalization probability clearly favoring loops of certain lengths. Our results suggest that the coordination of transcription and other processes that require genomically distant sites to be in spatial proximity might be supported by the underlying geometric constraints which could facilitate specific long-range interactions.

Thus, our model stresses the impact of conformational transitions due to administered changes in the confining geometry which might be deliberately exploited by nature (due to selective pressure) to drive proper biological functioning and may open up new perspectives for the study of biological processes and technological applications.

Chapter 6

Multi-Scale Structural Dynamics of the Histone H2AX

A Coarse-Grained Approach

References

The results presented in this chapter are published as and adapted from

- M. Fritsche, R.B. Pandey, B.L. Farmer and D.W. Heermann (2011), *Multi-scale Structural Dynamics of a Histone H2AX: A Coarse-Grained Monte Carlo Approach Via Knowledge-based Interaction Potentials*, under peer-review.

Chapter Summary

Histone proteins are not only important due to their vital role in cellular processes such as DNA compaction, replication and repair but also show intriguing structural properties that might be exploited for bioengineering purposes such as the development of nano-materials. Based on their biological and technological applications, it is interesting to investigate the structural properties of proteins as a function of temperature. In this chapter, we study the spatial response dynamics of the histone H2AX, consisting of 143 residues, by a coarse-grained bond fluctuating model for a broad range of normalized temperatures. A knowledge-based interaction matrix is used as input for the residue-residue Lennard-Jones potential.

We find a variety of equilibrium structures including global globular configurations at low normalized temperature ($T^* = 0.014$), combination of segmental globules and elongated chains ($T^* = 0.016, 0.017$), predominantly elongated chains ($T^* = 0.019, 0.020$), as well as universal SAW conformations at high normalized temperature ($T^* \geq 0.023$). Notably, the radius of gyration of the protein exhibits a non-monotonic temperature dependence with a maximum at a characteristic temperature ($T_c^* = 0.019$) where a crossover occurs from a positive (stretching at $T^* < T_c^*$) to negative (contraction at $T^* > T_c^*$) thermal response on increasing T^*.

6.1 Introduction

In the last decades, our perspective of proteins has changed from static to dynamical entities [259]. Increasing efforts are directed to understanding the functioning of biological macromolecules in order to integrate them with engineering for technological applications [259, 260]. In fact, proteins are powerful therapeutic agents [261, 262] and recent developments have stressed their impact on the bioengineering of nanomaterials, especially in biomedical imaging, drug delivery, biosensing and the design of functional nanocomposites [263, 264].

Protein self-assembly in suitable media offers unique advantages in the fabrication of protein-based nanodevices and avoids cost-intensive manufacturing processes [259]. However, bionanotechnological applications of proteins require thermally stable proteins [265–267].

Thus, based on the biological [268–271] and technological applications [259, 272, 273], it is interesting to investigate the structural properties of proteins as a function of temperature [274, 275].

But it is not only the technological aspects but also their biological impact that make the study of their temperature-dependent structural changes worthwhile. In fact, early work about temperature-protein interactions are concerned with temperature effects on catalytic rates or interspecific differences in thermal stability [274]. Notably, since the discovery of thermophilic organisms adaptive variations in structural and kinetic properties which may be pivotal in establishing an organisms' thermal optima have been a focus of intense research [276, 277].

In this work, we explore how the protein H2AX, a variant of the histone H2A, conforms and responds to temperature changes. Besides its possible technological applicability mentioned above, the protein H2AX plays a vital role in cellular functioning, in particular with

respect to DNA repair: Higher organisms have developed sophisticated mechanisms for the detection and repair of chromosome breaks in order to maintain genome integrity [270]. If left unrepaired, double-strand breaks can induce genetic alterations which, in turn, lead to a variety of illnesses including cancer [278,279]. There exists a relationship between a cell's response to DNA damage and histone postranslational modifications: Upon formation of double-strand breaks a subset of the core histone H2A is phosphorylated [279–281]. In most species, including humans, this modification occurs specifically on a variant of H2A, namely H2AX [279]. Formation of γ-H2AX occurs rapidly in the vicinity of double-strand breaks and might act as a trigger for the accumulation of many components needed in the DNA repair process [279–281]. In fact, it was found that γ-H2AX increases in a variety of conditions in relation to double-strand break generation processes, including radiation and high temperature [282].

Inspired by these findings that stress the importance of understanding the temperature-dependent behavior of H2AX, we consider a coarse-grained model [283,284] of this protein chain where knowledge-based residue-residue interactions [285, 286] are employed and effects of temperature are explored.

6.2 Modeling

The high dimensional space of protein conformations as well as the complexity of the energy surface make coarse-graining almost unavoidable in modeling the global structure and dynamics of proteins no matter whether one choses an all-atom approach, a minimalist description or a combination thereof [287]. In this work, we present a computer simulation study of the protein H2AX applying the bond-fluctuation method [146], which has been applied successfully to model the static and dynamical properties of polymer systems in several investigations. The reader is referred to chapter 3 for additional details on the simulation method.

The H2AX protein consists of 143 residues shown in Tab. 6.1. Each residue is described by a monomer of the bond-fluctuating protein chain [283,284]. This is a simplified representation of a residue without the all-atom structural details but the specificity of each residue is captured via the applied residue-residue interactions [288]. Moreover, our approach has the advantage of computational efficiency allowing for the covering large (biological) scales. In fact, the bond-fluctuation method has recently been used to study the conformational relaxation into native structure of a general HP protein chain [289] and even a specific protein, sensory rhodopsin, without severe constraints [290].

Apart from excluded volume interactions, each residue interacts with the neighboring residues within the range r_c using a generalized Lennard-Jones (LJ) potential

$$U_{i,j} = \left[|\epsilon_{ij}| \left(\frac{\sigma}{r_{ij}} \right)^{12} - \epsilon_{ij} \left(\frac{\sigma}{r_{ij}} \right)^{6} \right] \quad \text{for } r_{ij} < r_\mathrm{c}, \tag{6.1}$$

where r_{ij} is the distance between the residues i and j and $r_\mathrm{c} = \sqrt{8}$ and $\sigma = 1$ in units of the lattice constant. The strength of the pair potential ϵ_{ij} is unique for each pair of residues with appropriate positive (repulsive) or negative (attractive) values (for more details on the force field see [286, 291]). In contrast to our recent study of a HIV protease with a coarse-grained approach involving the relative hydropathy index of each amino acid as well as results from all-atom simulations [283], in this work, we use a knowledge-based interaction matrix for the residue-residue pair interactions. The knowledge-based interaction potential matrix is derived from an ensemble of a large number of protein structures in

the protein data bank (PDB). A number of such interaction tables are frequently used to investigate a range of questions related to protein structure including protein folding which has been studied extensively with a variety of models and methods involving all-atom details to minimalist coarse-grained descriptions [287, 292–297]. We resort here to the classic residue-residue contact interaction table [286] which is employed in studying scaffolding of short peptides [284]. Even though the knowledge-based matrix elements ϵ_{ij} are simplified estimates derived from residue-residue contacts we are confident that the phenomenological interaction matrix implicitly takes into account the secondary and tertiary structure of the proteins. In fact, X-ray crystallographic images of several thousands of proteins from the PDB are used to derive the residue-residue interaction matrix which is is applied as an input for our potential. In such a huge ensemble of proteins (from PDB), residues in secondary and tertiary structures are well represented by effective residue-residue interactions when applying a coarse-grained protein model.

Each randomly selected residue performs its stochastic movement according to the Metropolis algorithm subject to excluded volume constraints and the limits on changes in the covalent (i.e. peptide) bond length as in previous studies [283, 284]. A randomly selected residue at a site i is moved to one of its randomly selected neighboring lattice sites j with the Boltzmann probability $\exp(-\Delta E_{ij}/T^*)$, where $\Delta E_{ij} = E_j - E_i$ is the change in energy between the attempted E_j and current E_i configuration [146, 147].

T^* is the normalized temperature in units of the Boltzmann constant k_B and the energy ϵ_{ij}. Due to the lack of calibration with experimental data it is not possible to quantitatively relate the temperature T^* to physical temperature values. However, since the interresidue contact energies ϵ_{ij} allow for the calculation of realistic conformational energies of amino acids sequences in a number of different folds [291] we are able to relate temperature changes qualitatively to changes in the structural properties of proteins. Thus, our coarse-grained protein model provides a (qualitative) framework for understanding the temperature-dependent response of proteins which are so far inaccessible to experimental testing.

Initially, the protein chain is placed in a random conformation with excluded volume constraints. Simulations are then performed for a sufficiently long time (typically 10^7 time steps) with 150 independent samples. While one can monitor (thermodynamic) quantities (such as the radius of gyration or the energy) in a simulation in order to make sure that the system has reached asymptotic steady state, one has to take into account that the protein may or may not be in equilibrium due to the possibility of metastability (caused by frustration). Gerstman and Chapagain [298] provide an estimate for the time which is required for a protein to undergo the transition from a random coil to its native state. Using a simplified coarse-grained model and introducing a propensity energy to constrain appropriate segmental structures they suggest that 10^7 time steps (corresponding to about 0.01 sec) is large enough for a protein to reach its native structure. In our simulation, the protein chain is initially in a random coil configuration (with excluded volume constraints) and it takes about 10^7 time steps to reach an equilibrium conformation. Thus, while one has to take into account that this approximation might fail when considering additional details such as an effective medium etc., the time scale by Gerstman and Chapagain could be a rough estimate of the order of magnitude for our simulation. Different lattice sizes are used to test for finite size effects. Most of the data presented here are generated on a lattice of size $64 \times 64 \times 64$ since the qualitative results for different lattice sizes do not show significant differences.

Table 6.1: Sequence of residues of the histone H2AX. Hydrophobic residues are pink, polar residues are gold and electrostatic ones are blue. Image adapted from [299].

^{1}M	^{2}S	^{3}G	^{4}R	^{5}G	^{6}K	^{7}T	^{8}G	^{9}G	^{10}K	^{11}A	^{12}R	^{13}A	^{14}K	^{15}A	^{16}K	
^{17}S	^{18}R	^{19}S	^{20}S	^{21}R	^{22}A	^{23}G	^{24}l	^{25}Q	^{26}F	^{27}P	^{28}V	^{29}G	^{30}R	^{31}V	^{32}H	
^{33}R	^{34}L	^{35}L	^{36}R	^{37}K	^{38}G	^{39}H	^{40}Y	^{41}A	^{42}E	^{43}R	^{44}V	^{45}G	^{46}A	^{47}G	^{48}A	
^{49}P	^{50}V	^{51}Y	^{52}L	^{53}A	^{54}A	^{55}V	^{56}L	^{57}E	^{58}Y	^{59}L	^{60}T	^{61}A	^{62}E	^{63}I	^{64}L	
^{65}E	^{66}L	^{67}A	^{68}G	^{69}N	^{70}A	^{71}A	^{72}R	^{73}D	^{74}N	^{75}K	^{76}K	^{77}T	^{78}R	^{79}I	^{80}I	
^{81}P	^{82}R	^{83}H	^{84}L	^{85}Q	^{86}L	^{87}A	^{88}I	^{89}R	^{90}N	^{91}D	^{92}E	^{93}E	^{94}L	^{95}N	^{96}K	
^{97}L	^{98}L	^{99}G	^{100}G	^{101}V	^{102}T	^{103}I	^{104}A	^{105}Q	^{106}G	^{107}G	^{108}V	^{109}L	^{110}P	^{111}N	^{112}I	
^{113}Q	^{114}A	^{115}V	^{116}L	^{117}L	^{118}P	^{119}K	^{120}K	^{121}T	^{122}S	^{123}A	^{124}T	^{125}V	^{126}G	^{127}P	^{128}K	
^{129}A	^{130}P	^{131}S	^{132}G	^{133}G	^{134}K	^{135}K	^{136}A	^{137}T	^{138}Q	^{139}A	^{140}S	^{141}Q	^{142}E	^{143}Y		

6.3 Results and Discussion

6.3.1 H2AX Shows a Variety of Equilibration Structures

Fig. 6.1 illustrates snapshots of the histone H2AX for different representative temperatures (in reduced units) in the range of $T^* = 0.014 - 0.025$. Some of the general conformational characteristics such as globular structure formation (global aggregation of the intra-chain residues, $T^* = 0.014$), local segregation of selective residues ($T^* = 0.016$), large-scale stretching ($T^* = 0.020$), onset of randomization ($T^* = 0.022$) and thermal mixing ($T^* = 0.025$) are already apparent in the "snapshot" configurations. The interplay between the cooperative and competing interactions among the residues and the temperature constrained by the peptide bonds leads to a rich ensemble of protein structures. While a detailed analysis of such a structural ensemble in an in vivo system still remains an open challenge our approach offers some insight into the overall structural pattern changes that are so far inaccessible to experiment. Moreover, simulational studies (such as ours) preceding experimental tests may help e.g. in assessing the applicability of proteins for the design and fabrication of biomolecular devices in the bionanotechnology.

Fig. 6.2 and Fig. 6.3 show the energy and mobility profile of each residue. The energy of a residue is its interaction energy with neighboring residues within the range of interaction. The mobility of a residue is defined by the number of successful moves per unit MC time step. Note the contrasts in profiles at relatively low ($T^* = 0.014, 0.015$) and high ($T^* = 0.025$) temperature. Residues along the histone backbone appear to possess an isotropic distribution of (almost in equal number) attractive (cohesive, negative) and repulsive (positive) energy. The magnitude of the repulsive and the attractive energy and their differences in consecutive segments increases with temperature which is manifested in the segmental configuration as well as in the global (coil-to-globule) structure of the protein.

The mobility profile of the residues follows the energy profile relatively closely where the lower segmental energy differences translate into lower segmental mobility. At low temperatures ($T^* = 0.014, 0.015$), residues with the lowest mobility consist of 36R, 37K, 57E, 62E, 65E, 72R, 73D, 75K, 76K, 78R, 89R, 90N, 91D, 92E, 93E, 95N, 96K, 119K, 120K, 134K, 135K, 141Q and 142E. Nearly all the electrostatic residues (D, E, K, R) along with a few polar groups (Q,N) act as anchor/seed for segmental aggregation. Note that the pair interaction potentials of these residues have the largest well depth [286].

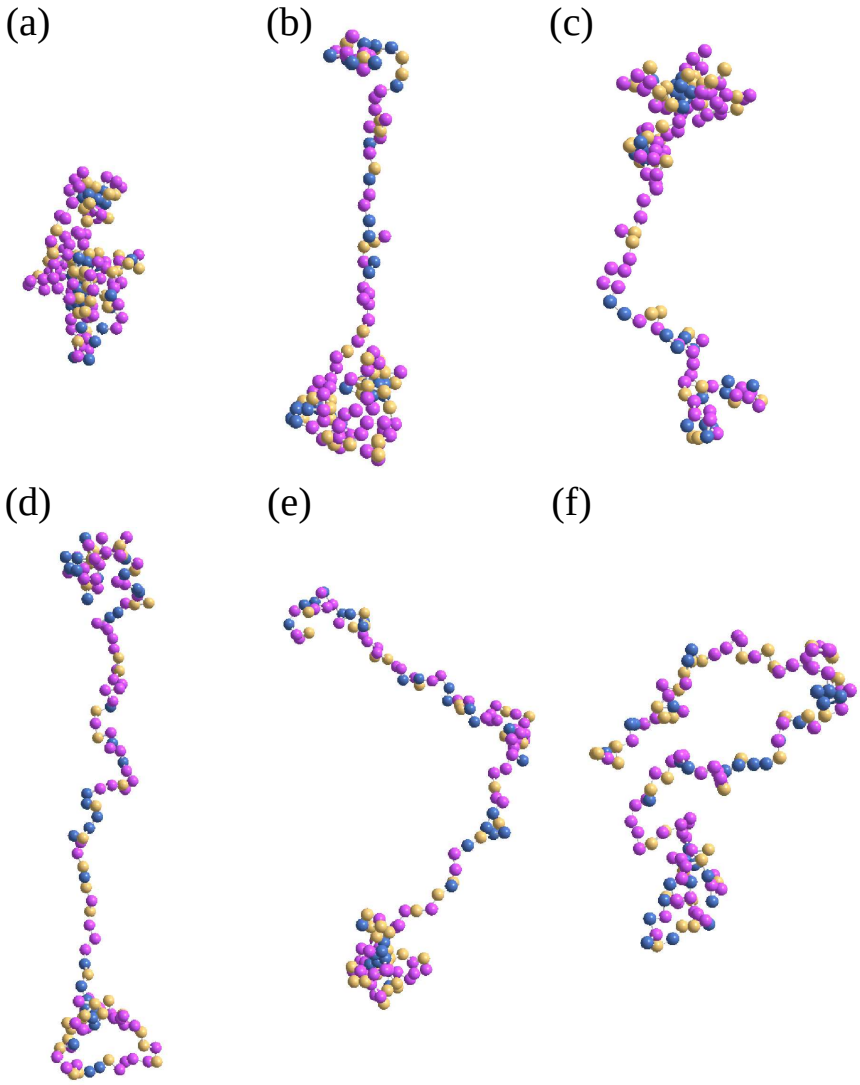


Figure 6.1: Snap shots of the histone H2AX at (a) $T^* = 0.014$, (b) $T^* = 0.015$, (c) $T^* = 0.016$, (d) $T^* = 0.020$, (e) $T^* = 0.022$ and (f) $T^* = 0.025$. Hydrophobic residues are shown in pink, polar residues in gold and electrostatic ones in blue. We find globular configurations at low temperature ($T^* = 0.014$), combination of segmental globules and elongated chains ($T^* = 0.015, 0.016$), predominantly elongated chains ($T^* = 0.020$), as well as universal SAW conformations at high temperature ($T^* \geq 0.022$). Image adapted from [299].

Most of these residues become more mobile when raising the temperature (i.e. see the segments 91D, 92E, 93E).

It should be pointed out that some residues (e.g. 42E and 57E) have surprisingly low mobility despite their positive energy while others with low energy have a high mobility index. In fact, the conformational energy of the amino acids sequence (the interaction energy) does not determine the local structure and mobility alone. Physical (covalent bonding) or topological (trapping) constraints also play an important role in the cooperative response.

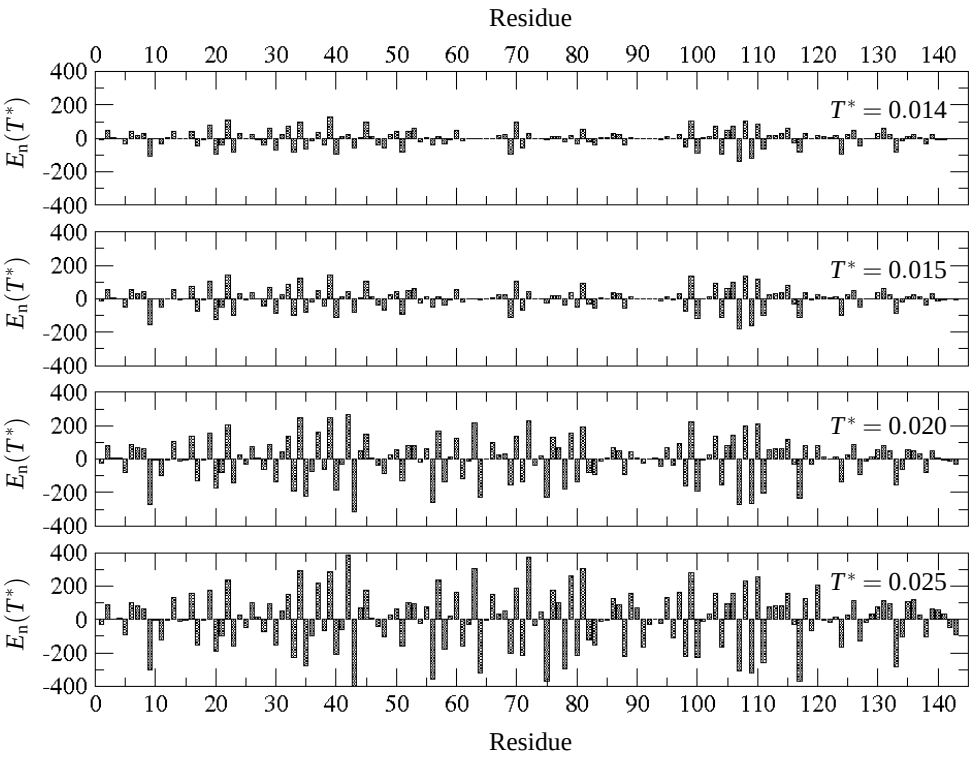


Figure 6.2: Energy E_n of each residue of histone H2AX at normalized temperatures $T^* = 0.014, 0.015, 0.020$, and 0.025. The energy of a residue is its interaction energy with neighboring residues within the range of interaction. The magnitude of the repulsive and the attractive energy and their differences in consecutive segments increases with temperature which is manifested in the global (coil-to-globule) structure of the protein. Image adapted from [299].

6.3.2 Non-Monotonic Temperature Dependence of the Gyration Radius

With respect to biotechnological applications it is interesting to study how the entire range of temperatures affects a protein's size and shape. The temporal variation of the radius of gyration R_{gyr} shows that it has reached its equilibrium at all temperatures except for the lowest one ($T^* = 0.014$) where the relaxation is too slow. Equilibration implies that the protein chain has explored a sufficient amount of conformations in structural phase space. The average value of the equilibrium radius of gyration can be evaluated from the asymptotic data sets at each temperature. R_{gyr} shows a non-monotonic dependence on temperature as can be seen in Fig. 6.4 with a maximum at a characteristic temperature T_{c}^* which is a specific property of the studied biomaterial. The radius of gyration increases on increasing the temperature ($T^* \approx 0.014 - 0.019$) from the low end until around $T_{\mathrm{c}}^* \approx 0.019$ followed by a linear decay ($T^* \approx 0.019 - 0.023$) before reaching its saturation at high temperature $T^* \geq 0.023$.

We examine how the competition between residue-residue interactions and thermal fluctuations leads to the observed non-monotonic temperature dependence of the radius of gyration. The attractive inter-residue interaction induces self-assembly of the protein segments towards a global globular structure ($T^* \to 0.014$) as the residues undergo their stochastic motion. As discussed above the highly interacting electrostatic residues act as

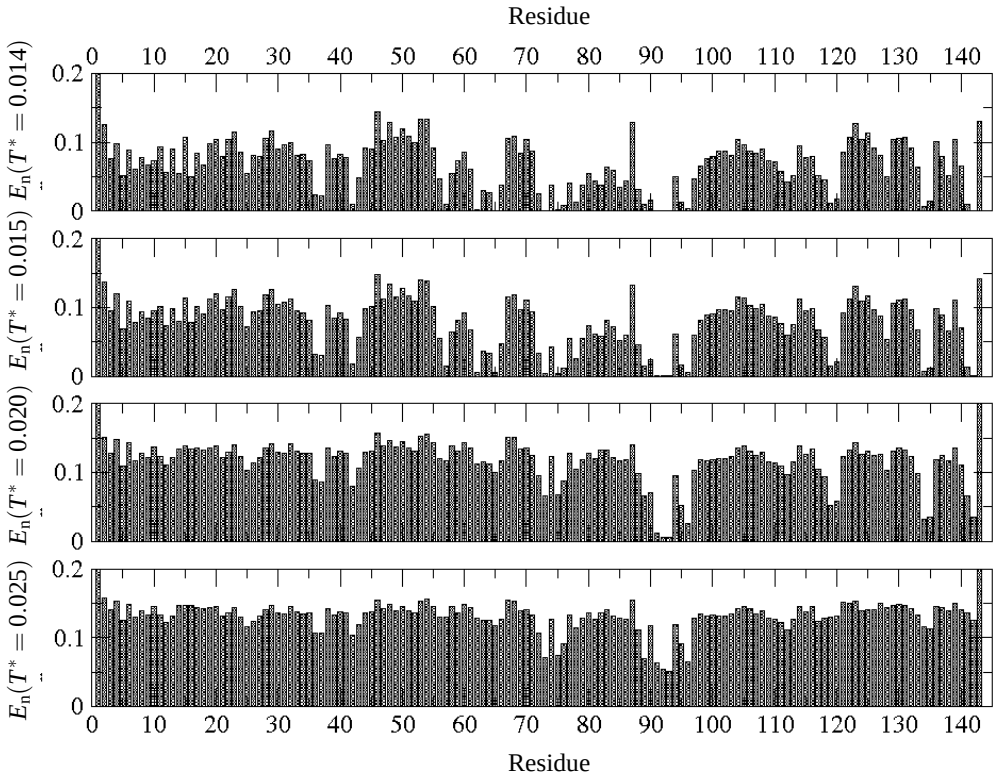


Figure 6.3: Mobility M_n of each residue (number of successful moves per unit time step) of the histone H2AX at temperatures $T^* = 0.014, 0.015, 0.020$, and 0.025. The mobility of a residue is defined by the number of successful moves per unit MC time step. With exceptions (e.g. 42E and 57E) the mobility profile of the residues follows the energy profile relatively closely where the lower segmental energy differences translate into lower segmental mobility. Image adapted from [299].

an anchor collecting even those tethered residues that are repulsive. Thus, cooling down the protein leads to a conformational collapse into its globular conformation. The self-organized protein structure in its globular conformation begins to break on raising the temperature even by a small amount ($T^* \approx 0.016$) as the constitutive residues dissociate while some local assembly (held together by non-covalent interactions) still persists. The local assembly de-segregates on further increasing the temperature ($T^* = 0.017$) which stretches the corresponding segments resulting in a larger radius of gyration. Stretching of the protein continues until the characteristic temperature ($T_c^* \approx 0.019$) is reached beyond which the protein chain begins to contract. The chain segments fluctuate introducing randomness into a relatively stable elongated structure as the onset of thermal fluctuations sets in. The process can be described as a "thermal-driven contraction" emerging due to the cooperative effect of segmental interaction (looping) and conformational entropy. Note that the protein H2AX expands (positive) on heating in the low temperature regime ($T^* \approx 0.014 - 0.019$) and contracts (negative) at the higher temperatures ($T^* \approx 0.019 - 0.023$). The crossover from a positive thermal response in low temperature to a negative thermal response in high temperature regimes with a well-defined transition temperature ($T_c^* \approx 0.019$) appears to be a specific characteristic of the H2AX protein structure. In a recently performed work, we have examined the thermal response of his-

tone H3.1 (of comparable size with 136 residues) with exactly the same method [300]. In contrast to the non-monotonic thermal response of H2AX, histone H3.1 exhibits a continuous transition from coil-to-globule on reducing the temperature. The difference in thermal response of H2AX and H3.1 however leads us to believe that this is due to the specific sequence of amino acids which might be exploited in technological applications requiring a material with such a distinct temperature response. Eventually, the radius of gyration saturates on further increasing the temperature beyond $T^* \geq 0.023$, where the protein conforms to a thermal-driven random (coil) structure.

The size of the protein as measured by its radius of gyration R_{gyr} can be compared at a low temperature ($T^* < T_{\mathrm{c}}^*$) in the positive thermal response regime and at a higher temperature ($T^* > T_{\mathrm{c}}^*$) in the negative thermal response regime. One has to point out that, despite having the same magnitude of R_{gyr}, the structure of the protein at these temperatures is very different. In particular, as shown in Fig. 6.1 we observe a local segmental segregation at low temperature while random configuration at high temperature dominate.

Although the radius of gyration can provide insight into the spatial extension of the protein, the specific dynamics of local structures are difficult to quantify with this measure. Thus, we have analyzed the root mean squared displacement of the center of mass of the protein with as a function of time for the entire temperature range. Fig. 6.4 shows these results for representative temperatures. The protein continues to diffuse at high temperatures while its motion slows down on reducing the temperature showing sub-diffusive asymptotic dynamics. At very low temperatures, $T^* = 0.014 - 0.015$, the dynamics are too slow since the protein is localized into its globular conformation.

The question has to be raised whether there is another property such as the specific heat C_{V} for which T_{c}^* has a special significance. C_{V} is evaluated from the fluctuation in the energy $\langle \Delta E^2 \rangle / T^2$ and Fig. 6.5 shows that the specific heat C_{V} does not show a peak characteristic per se. This can be understood by noting that the characteristic temperature T_{c}^* is related to the maximum thermal response in the spatial extension and not to the identification of a phase transition. However, we do observe a minimum in C_{V} around T_{c}^*. In fact, the relaxation of the protein is not only controlled by the competition between residue-residue interactions and temperature but also by the steric constraints imposed by the peptide bonds. Thus, the thermal response in C_{V} is expected to be different from that of the radius of gyration.

The structure factor $S(q)$ provides the spatial scaling of the distribution of constitutive elements

$$S(q) = \left\langle \frac{1}{N} \left| \sum_{j=1}^{N} \exp(-i\mathbf{q} \cdot \mathbf{r}_j) \right|^2 \right\rangle_{|\mathbf{q}|} \tag{6.2}$$

where $\mathbf{r}_j$ is the position vector of each residue and $|\mathbf{q}| = q = 2\pi\lambda$ is the average spherical wave vector of wave length λ. For the protein, the structure factor is useful in understanding the structural details over a range of length scales. Fig. 6.6 shows the variation of $S(q)$ with the wave vector on a log-log scale. Since the radius of gyration R_{gyr} is a measure of the residue spread, the distribution of residues in the range of $q = 0.17 - 0.50$ provides information about the global conformation of the protein. From the power-law scaling of the structure factor with the wave vector, $S(q) \sim q^{-1/\nu}$, we can estimate the distribution of the protein segments $R_{\mathrm{gyr}} \sim N^{\nu}$. We have estimated the slope of the power-laws in the appropriate range of the wave vector q in Fig. 6.6: $-1/\nu \approx 2.84$ at $T^{*\mathrm{low}} = 0.014$ and $-1/\nu \approx 1.70$ at $T^{*\mathrm{high}} = 0.025$ which provides $\nu^{\mathrm{low}} = 0.35$ and $\nu^{\mathrm{high}} = 0.59$, respectively. Thus, the protein has an effective dimension $D_{\mathrm{eff}}^{\mathrm{low}} \approx 1/\nu^{\mathrm{low}} \approx 2.84$ (almost solid,

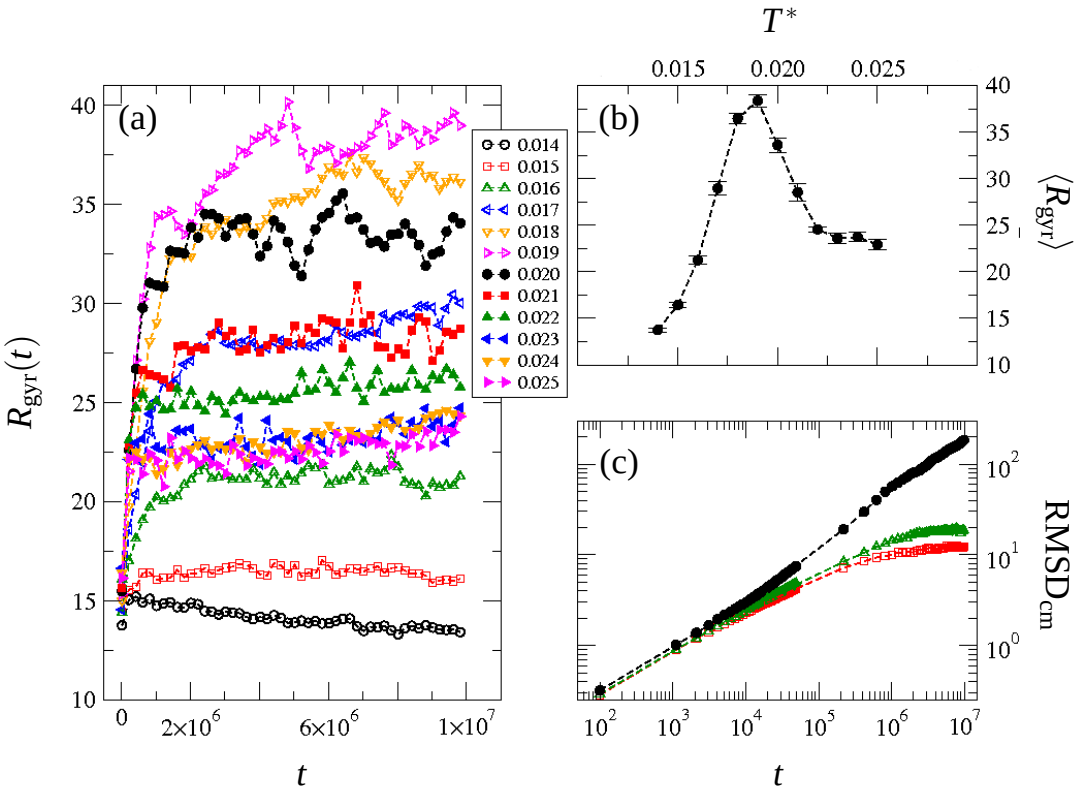


Figure 6.4: (a) Dependence of the radius of gyration $R_{\mathrm{gyr}}(t)$ on MC time steps t for various temperatures $T^* = 0.014 - 0.025$; (b) temperature-dependent behavior of the mean radius of gyration $\langle R_{\mathrm{gyr}} \rangle$; (c) root mean squared displacement of the protein's center of mass $\mathrm{RMSD}_{\mathrm{cm}}$ as a function of MC time steps t. The crossover from a positive thermal response of $\langle R_{\mathrm{gyr}} \rangle$ at low temperature to a negative thermal response at high temperature with a well-defined transition temperature $T_c^* \approx 0.019$ appears to be a specific characteristic of the H2AX protein structure. Image adapted from [299].

a globular structure) at $T^* = 0.014$ and $D_{\mathrm{eff}}^{\mathrm{high}} \approx 1.70$ (a ramified, tenuous SAW structure) at $T^* = 0.025$. As shown in Fig. 6.6 a systematic change in the mass distribution of the protein is clearly seen on increasing the temperature ($T^* = 0.014 - 0.025$). On the lower spatial scale (higher q) there are minor modifications in the mass distribution at $T^* = 0.014$ while the protein segments appear like an ideal chain with $\nu \approx 1/2$ at $T^* = 0.025$.

6.4 Conclusions

γ-H2AX [301], the serine 139 phosphorylated form of H2AX, is one of the earliest repair responses to DNA double-strand breaks which can lead to mutations that in turn are a cause of cancer and hereditary diseases [278].

There are increasing lines of evidence that damage-dependent changes in chromatin structure by chromatin remodeling complexes are required for the formation of γ-H2AX, which is expected to contribute to the preparation of repair [279]. In fact, reorganization of higher-order chromatin structures may be related to the rather large chromatin domains observed to contain γ-H2AX and cohesins [279]. Jörg Bewersdorf and coworkers have

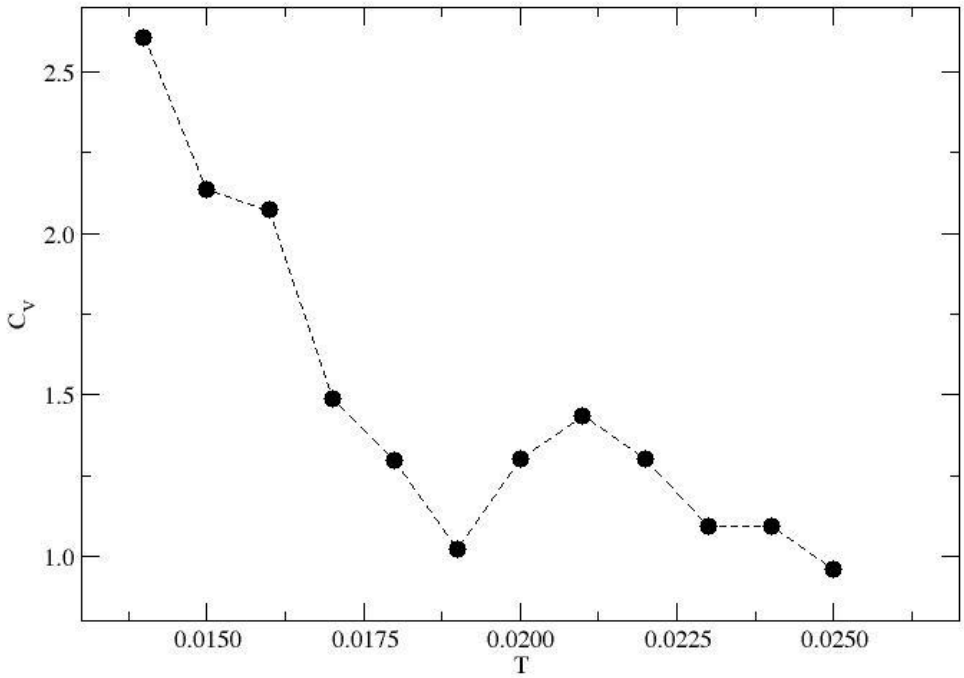


Figure 6.5: C_V is evaluated from the fluctuation in the energy $\langle \Delta E^2 \rangle / T^2$ and shows a minimum in C_V around T_c^*. In fact, the relaxation of the protein is not only controlled by the competition between residue-residue interactions and temperature but also by the steric constraints imposed by the peptide bonds. Image adapted from [299].

investigated the distribution of H2AX throughout the chromatin fiber during the time course of DNA damage and repair by means of 4Pi-microscopy [270]. The authors propose that the formation of H2AX clusters illustrated in Fig. 6.7 supports the immediate and robust repair response observed after DNA damage.

In this chapter, we have studied the structural dynamics of a single H2AX histone as a first step before addressing the histone's aggregation behavior (as a function of temperature and system density). A coarse-grained protein model is applied which consists of 143 residues tethered in a bond fluctuating chain on a cubic lattice. Although the atomistic details of residues are ignored, their specificity is captured via a knowledge-based interaction matrix as well as a LJ pair potential for residue-residue interactions. Each residue executes its stochastic motion according to the Metropolis criterion.

We have analyzed a number of local and global physical quantities such as the energy and mobility of each residue as well as the root mean squared displacement of the protein's center of mass, its radius of gyration, and its structure factor. The impact of temperature on these quantities is investigated and might be exploited for the design of biomaterials.

Our approach allows for the identification of segmental characteristics such as active regions and anchoring sites of the protein. We find that the electrostatic residues (e.g. 72R, 73D, 91D, 92E, 93E, 134K, 135K, etc.) are critical in orchestrating the segmental

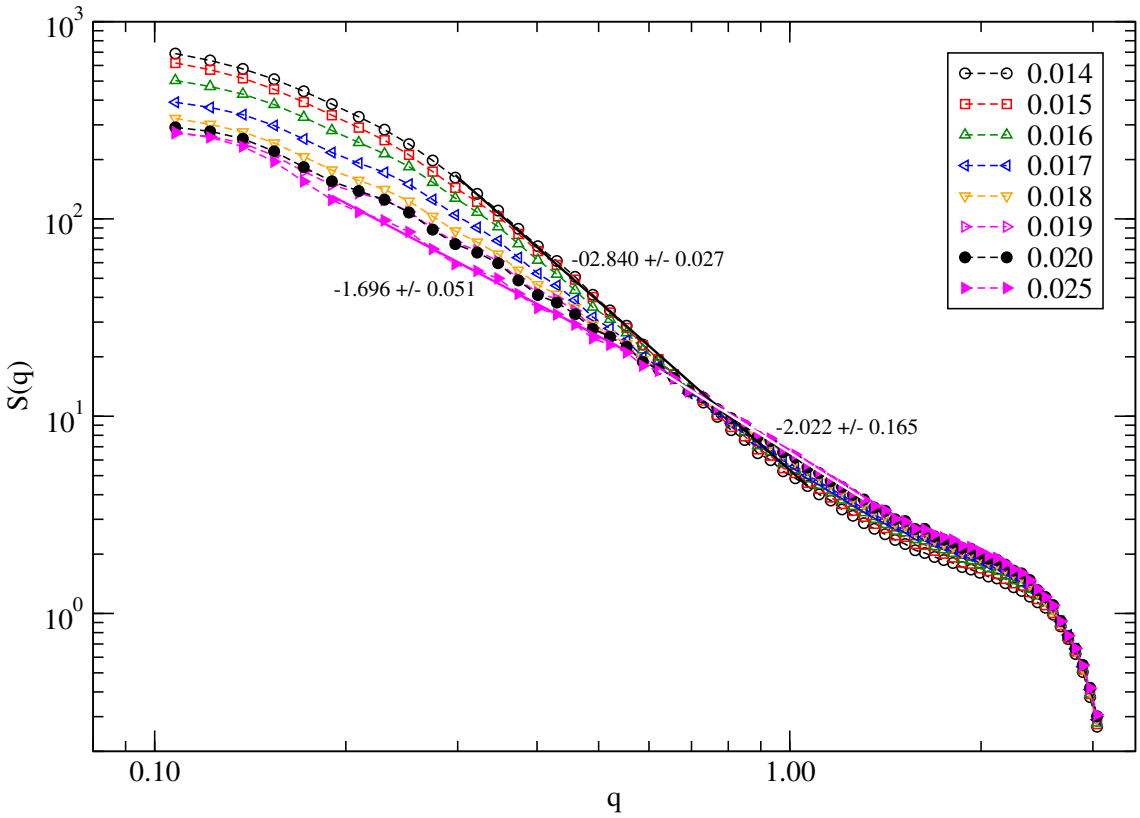


Figure 6.6: Structure factor $S(q)$ versus wave vector q. The effective dimension of the protein is $D_{\text{eff}} \approx 2.84$ (almost solid, a globular structure) at $T^* = 0.014$ and $D_{\text{eff}} \approx 1.70$ (a ramified, tenuous SAW structure) at $T^* = 0.025$. Image adapted from [299].

conformation, their self-assembly and de-segregation from the low to the moderately high temperature regime ($T^* = 0.014-0.025$). These highly interacting residues at their specific positions in the protein sequence appear to determine specificity and multi-scale structures. Accordingly, we observe global globular configurations at low ($T^* = 0.014$), a combination of chains segments and smaller segmental globules at intermediate ($T^* = 0.016, 0.017$), and elongated structures at moderately high temperatures ($T^* = 0.019, 0.020$). As expected, the specificity of residues vanishes at high temperatures ($T^* \geq 0.023$) where the mobility of most residues becomes considerably high and comparable. In this thermal-driven structural regime, the residues become indistinguishable leading to a SAW chain conformation.

The radius of gyration of the protein shows a non-monotonic dependence on the temperature with a maximum at a characteristic temperature which is determined by the competition between inter-residue interactions and temperature. The protein H2AX expands (positive thermal response) on heating in the low temperature regime ($T^* = 0.014-0.019$) and contracts (negative thermal response) at higher temperatures ($T^* = 0.019 - 0.023$). The crossover from a positive to negative thermal response occurs at a well-defined transition temperature ($T_c^* \approx 0.019$) which may be a specific characteristic of the histone H2AX and particularly interesting for bioengineering purposes. The variation in the global conformation of the protein is explained in the framework of self-assembly at the local scale.

Based on the analysis of the structure factor $S(\mathbf{q})$, we find that the radius of gyration scales with its molecular weight N as $R_{\text{gyr}} \sim N^{\nu}$, where $\nu = 0.35$ and $\nu = 0.59$ at

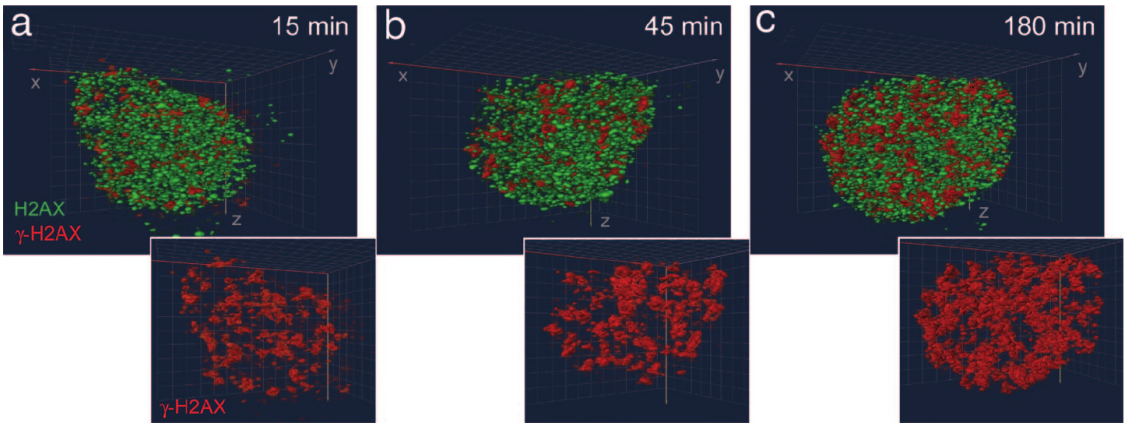


Figure 6.7: 4Pi images of H2AX (green) and γ-H2AX (red) clusters during a time course of DNA damage and repair. Image adapted from [270].

$T^* = 0.014$ and 0.025, respectively. The effective dimension of the protein is therefore $D_{\text{eff}} \approx 2.84$ (almost solid, a globular structure) at $T^* = 0.014$ and $D_{\text{eff}} \approx 1.70$ (a ramified, tenuous SAW structure) at $T^* = 0.025$. A systematic change in the mass distribution is clearly seen with an increase in temperature ($T^* = 0.014 - 0.025$).

Our coarse-grained protein model allows for a deeper understanding of local and global properties, which can so far not be gained by experimental testing. Besides the biological importance of proteins such as H2AX, we are able to provide a framework for analyzing potential candidates for the bioengineering of nano-materials. To this end, future experiments measuring physical quantities such as the spatial extension (radius of gyration) as a function of temperature would allow for the calibration of the temperature scale.

Chapter 7

Entropy-Driven Synaptonemal Complexes Organization within the Nucleus

A Combined 4Pi-Microscopy and Modeling Approach

References

The results presented in this chapter are published as and adapted from

- M. Fritsche, L. Reinholdt, M. Lessard, M.A. Handel, J. Bewersdorf and D.W. Heermann (2011), *Entropy-Driven Spatial Organization of Synaptonemal Complexes within the Cell Nucleus*, in preparation.

The authors would like to thank Lindsay Shopland for her comments and suggestions.

Chapter Summary

We employ 4Pi-microscopy to study the spatial organization of synaptonemal complexes (SCs) in mouse spermatocyte nuclei allowing for the three-dimensional reconstruction of their backbone arrangement. Additionally, we model the SCs in the cell nucleus by confined, self-avoiding polymers, whose chain ends are attached to the envelope of the confining cavity and diffuse along it. We aim at elucidating the role of entropy in shaping pachytene SC organization. The framework provided by the complex interplay between SC polymer rigidity, tethering and confinement is able to qualitatively explain features of SC organization, such as mean squared end-to-end distances, mean squared center of mass distances or SC densities distributions. However, it fails in correctly assessing SC entanglement within the nucleus. In fact, our analysis of the 4Pi-microscopy images reveals a higher ordering of SCs within the nuclear volume than what is expected by our numerical model. Thus, while entropic contributions constitute an essential organizational driving force, the dedicated action of proteins or actin cables might be needed in order to fine-tune the three-dimensional SC organization. Future experiments determining the bending rigidity of SCs within the cell nucleus might help to qualitatively test our assumptions.

7.1 Introduction

7.1.1 Meiosis

Sexually reproducing organisms use a dedicated cell division cycle called meiosis to produce *haploid gametes* (containing a single copy of each chromosome) from *diploid nuclei* (containing two *homologous*s copies of each chromosome) [48]. One member of a chromosome pair, where each member of the pair is derived from one parent, is referred to as a *homologue* [302]. A *reduction* and a *equational division* [9,53,54], *meiosi I* and *meiosis II*, are required to produce gametes as illustrated in Fig. 7.1.

Pre-meiotic S phase proceeds the beginning of meiosis, where cells duplicate their chromosomes in order to produce two identical *sister chromatids* [302]. *Prophase I* initiates meiosis and can again be divided into multiple time steps visualized in Fig. 7.1: *leptotene*, *zyogtene*, *pachytene* and *diplotene* (not shown in Fig. 7.1).

During leptotene individual chromosomes condense into visible strands within the nucleus [9], while alignment (pairing) of homologous chromosomes takes place during zygotene. Due to the way the telomeres cluster within the nucleus, this is called the bouquet stage. A proteinaceous structure, the synaptonemal complex (SC), develops around the paired homologues and holds them in close proximity, or synapsis [48–52]. The pachytene stage is the stage when chromosomal crossover occurs. The new combinations of DNA created by crossing over significantly contribute to genetic variation and the points, where non-sister chromatids crossover, are known as chiasmata. During diplotene the SCs dissolve. Eventually, the homologous chromosomes move to opposite cell poles, where they gather into separate nuclei and the original cell divides.

Following meiosis I, the daughter cells directly enter meiosis II. Meiosis II can be compared to a mitotic division, except that it results in four haploid cells that contain only a single copy of each chromosome. For more details on the whole meiotic process, the reader is referred to Refs. [9, 49, 50, 302, 303].

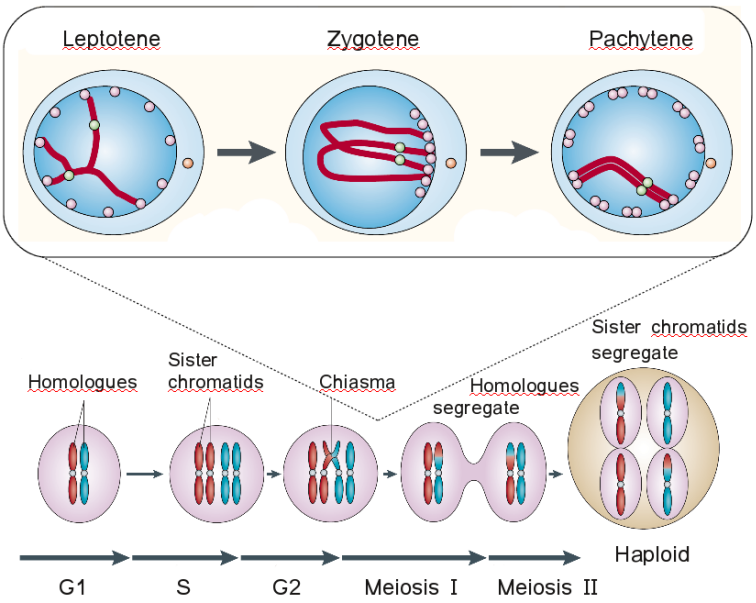


Figure 7.1: Meiosis I and meiosis II follow a single round of DNA replication during pre-meiotic S phase. Due to two successive nuclear divisions during meiosis, two sets of chromosomes are divided among four nuclei, each of which then has half as many chromosomes (haploid) as the original cell (diploid). Prophase I during meiosis I can be divided into leptotene, zygotene, pachytene and diplotene (not shown) [303]. Telomeres (pink) of elongated chromosomes attach randomly over the nuclear envelope during leptotene. Bouquet formation occurs during zygotene, followed by the initiation and progression of synapsis. At pachytene stage, homologues are paired by the synaptonemal complex. Telomeres of the paired homologues are distributed over the nuclear envelope. Images adapted from [302] and [303].

7.1.2 The Synaptonemal Complex

The SC's gross structure has been studied by a variety of imaging methods revealing its specific features within the nucleus [49,50,52,304,305]. Optical sectioning and fluorescence deconvolution light microscopy have shown that SCs undergo substantial rearrangements during meiotic prophase leading to the resolution of interlocks, where SCs appear well separated and uniformly distributed throughout the nucleus [304,306,307].

The main features of the SC are two lateral elements to which loops of the paired homologous chromosomes are attached, as well as a central element with linking transverse filaments, giving the SC a ladder-like appearance [9]. The lateral elements of SCs are twisted suggesting that SCs are not simply rigid rods but substantially (semi)flexible [50, 308].

Moreover, the SC ends are attached, through the telomeres, to the nuclear envelope, which enables them to diffuse along it [310,311]. Tethering of the ends is critical for proper SC organization. The absence of proteins such as Ndj1 and Sun1, which are required for telomere attachment, leads to the dissociation of telomeres from the nuclear envelope and SC organization and function (recombination) are disrupted [312].

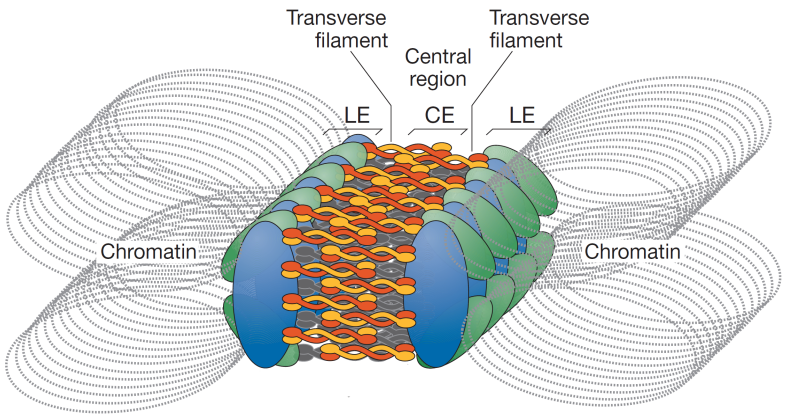


Figure 7.2: Model of the synaptonemal complex structure, which connects paternally and maternally derived chromosomes. It is a proteinaceous structure formed by two lateral elements (LE), a central element (CE) and transverse filaments. The LE comprises cohesins and structural proteins such as synaptonemal complex protein 2 (SCP2) and SCP3 [309]. Among other proteins, the transverse filaments are formed by the protein SC1 [309]. Image adapted from [309].

7.1.3 4Pi-Microscopy

Since the lateral elements of the SCs are 100-200 nm apart [49,50], conventional microscopy is unable to resolve them as this distance lies just below the diffraction limit of about 250 nm for green light [313]. In this work, we employ 4Pi-microscopy, a laser scanning fluorescence microscopy with an improved axial resolution [198, 314, 315], to study SC organization in mouse spermatocyte nuclei. This technology does not only permit the identification of the SC's spatial arrangement, including the differentiation between the two lateral elements, but also allows for the characterization of twists in three dimensions.

The resolution limit is described by the minimal fluorescence spot size [316], which is represented by the point-spread function [317]. A standard highest aperture confocal microscope allows for a minimal fluorescence spot size of about 180 nm in the transverse and about 500 nm in the axial direction [316]. Thus, improving the spatial resolution implies reducing the extent of the point-spread function in real space [316].

4Pi-microscopy renders a main maximum of the point-spread function three to seven times axially smaller than confocal microscopy by using two opposing objective lenses that are illuminated coherently and focus the same spot onto a sample [198, 314, 317]. The basic setup of the 4Pi-microscope is shown in Fig. 7.3. A beam splitter divides the laser light into two beams. Both beams are then directed by mirrors towards the two opposing objective lenses. At the common focal spot superposition of both focused light beams can occur. Excited sample molecules subsequently emit fluorescence light which is collected by the two objective lenses, combined by the same beam splitter and deflected onto a detector. At that point, constructive interference of both emitted light pathways can occur again [318].

Three different operation modes of the 4Pi-microscope can be used to increase resolution. Either the coherent superposition of excitation light (type A), or the constructive interference of the emission light (type B) [314] or the interference of both light pathways (type C) can be exploited to achieve a resolution increase, which is strongest for a 4Pi-microscope of type C [317].

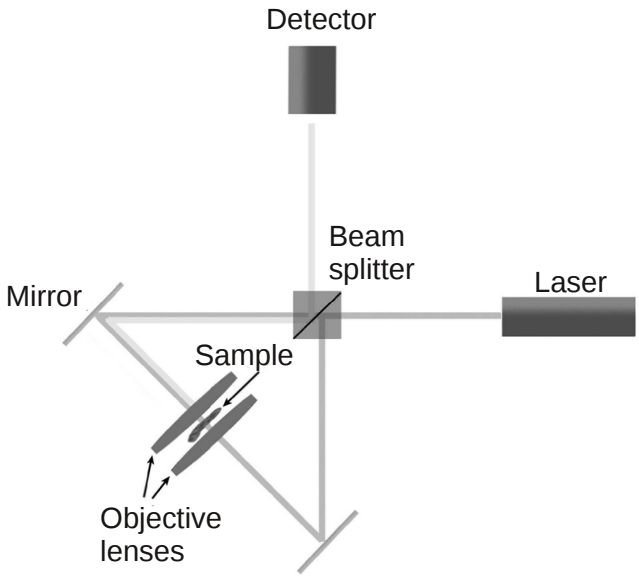


Figure 7.3: General setup of a 4Pi-microscope. Since there are no lenses that are able to produce a wavefront of a solid angle of 4π, the idea is to use two opposing objective lenses coherently, such that the illumination wavefronts and/or fluorescence wavefronts can "join forces" [317]. Image adapted from [318].

7.1.4 Synaptonemal Complexes within the Cell Nucleus

The SCs are assumed to facilitate the formation of chromosomal crossovers, i.e. the exchange of genetic material between homologous chromosomes, [49] and to dissolve afterwards [319]. But how can the SCs fulfill this role?

This is a place, in which physical modeling of polymers in confined space can help to derive an understanding of the basic principles underlying spatial SC organization. Notably, we focus on the SC organization at the pachytene stage, after which much of the dynamic activity related to telomere clustering, bouquet formation, "zippering" of the SCs and interlock resolution has occurred. To this end, the early recognition and colocalization of homologous DNA sequences has already been investigated by a coarse-grained polymer approach [310]. Nicodemi and colleagues have shown that entropy can drive the recognition/pairing mechanism whereby homologous sequences spontaneously recognize and become tethered to each other [310].

In this work, we aim at investigating the physical basis and principles of SC organization at pachytene. Based on the already known features of SC organization, such as confinement in the nuclear volume, mean SC length and number, size of the nucleus, telomere tethering to the nuclear envelope, as well as semiflexibility, we construct a coarse-grained polymer model. Besides imposing self-avoidance on the SC polymer chains, we do not take into account any additional interactions, therewith being able to study the role of (configurational) entropy in shaping SC organization.

Additionally, such a basic polymer model allows the investigation of the impact of key SC features: In particular, we study the role of (i) tethering the SCs to the nuclear

envelope as well as (ii) their semiflexibility on their spatial organization (and ordering).

The impact of (i) tethering is studied by comparing two models: In the SC model, the ends of the semiflexible polymers are tethered to the borders of the confining geometry and are only allowed to diffuse along it. A comparative "null model" consists of SC polymers which are untethered and allowed to freely explore the accessible (nuclear) space.

The combination of topological constraints, like tethering (i) and/or confinement, and (ii) semiflexibility plays a central role in a wide range of biophysical contexts, such as chromosome packaging [5, 33]. In fact, semiflexibility induces a competing interplay between configurational entropy, bending energy and excluded volume. While, there is no consensus on the range of bending stiffness of SCs so far, our modeling approach allows for the investigation of a broad range of bending rigidities.

The analysis of 4Pi-microscopy images allows for the determination of the SCs' three-dimensional coordinates within the cell nucleus. Thus, we are able to compute all quantities of interest, such as the mean squared end-to-end distance, the mean squared center of mass distance, the amount of chain overcrossings as well as densities distribution functions for both the computational model and the experimental dataset.

7.2 Experimental Approach and Modeling

7.2.1 Synaptonemal Complexes as Semiflexibility Polymers

Our model includes 19 autosomal mouse SCs described as 19 semiflexible self-avoiding polymers. Based on a mean SC length of 12 μm in male mice [319, 320] each polymer consists of $N_{\mathrm{Kuhn}} = 12$ Kuhn segments of length $b_{\mathrm{Kuhn}} = 1.029$ μm.

Approximating biological "storage" such as the cell nucleus of diameter $r_{\mathrm{D}} = 8$ μm, we consider cubic confinement of $D_{\mathrm{x}} = D_{\mathrm{y}} = D_{\mathrm{z}} = 8$ Kuhn segments.

To generate polymer conformations we employ the bond-fluctuation method [166] of which a detailed description can be found in chapter 3.

Semiflexible polymers may be characterized by their persistence length l_{p}, which is the typical length scale over which the chain backbone loses information about its direction due to thermal fluctuations [139]. Notably, a recent study [157, 248] has shown that standard definitions of persistence length might fail for chains with excluded-volume restrictions, stressing the importance of carefully checking in which regime experimental data belong. In this work, the contour length exceeds the persistence length only a few times so that excluded volume effects are not yet very important [248]. Thus, we interpret the decay of the orientational correlation function in terms of an effective "quasi" persistence length reflecting global conformational flexibility rather than local intrinsic stiffness. The range of chain rigidity is varied from a relatively flexible chain of $l_{\mathrm{p}}/L \to 0.086$ up to the stiff regime of $l_{\mathrm{p}}/L \to 0.77$.

In a lattice representation such as the BFM, the bending energy H_{b} can be expressed as [249]

$$H_{\mathrm{b}} = \frac{l_{\mathrm{p}}}{2b} \sum_{i=1}^{N-1} (\mathbf{u}_{i+1} - \mathbf{u}_i)^2. \tag{7.1}$$

Here, $\mathbf{u}_i = \left(\frac{\mathbf{R}_i - \mathbf{R}_{i-1}}{|\mathbf{R}_i - \mathbf{R}_{i-1}|} + \frac{\mathbf{R}_{i+1} - \mathbf{R}_i}{|\mathbf{R}_{i+1} - \mathbf{R}_i|} \right) \Big/ \left| \frac{\mathbf{R}_i - \mathbf{R}_{i-1}}{|\mathbf{R}_i - \mathbf{R}_{i-1}|} + \frac{\mathbf{R}_{i+1} - \mathbf{R}_i}{|\mathbf{R}_{i+1} - \mathbf{R}_i|} \right|$ is a discrete realization of $\delta \mathrm{R}(s)/\delta s$, the unit tangent vector at arclength s, where $\mathrm{R}(s)$ is the position vector [230], and N is the total number of monomers in the chain. All energies are measured in units of $k_{\mathrm{B}}T$.

In order to generate thermodynamically equilibrated polymer conformations we use the Metropolis Monte Carlo method [146, 147]. Since subsequently created conformations are highly correlated, we determine, for each set of parameters (persistence length l_p, tethered or untethered SC polymers), the autocorrelation function [146, 147] of the squared end-to-end distance R^2_end. The integrated autocorrelation time τ_int is computed by applying the windowing procedure introduced by Sokal [164], which is described in chapter 3. We consider two subsequent conformations as uncorrelated after $5\tau_\mathrm{int}$ Monte Carlo steps, thus creating 16 000 independent configurations.

7.2.2 "3D"-Preps for 4Pi-Microscopy Measurement

The subsection 7.2.2 is kindly provided by Laura Rheinholdt, Mark Lessard and Jörg Bewersdorf, who also performed the experiment described below.

Mice

Male B6SJLF1/J mice (JAX stock #100012) were obtained from The Jackson Laboratory and euthanized by cervical dislocation at 17 dpp. The testes were removed and placed in 1 ml of 1x PBS with protease inhibitors added (Roche, Complete Mini #11 836 153 001).

Spermatocytes

After removing the tunica, each testis was macerated in 1 ml of 1xPBS with protease inhibitors and triturated gently using a 1 ml pipette to create a suspension. This suspension was then layered over 6ml of 1xPBS with protease inhibitors, in a 15ml conical tube, and allowed to sit 10 minutes. Five, 1ml fractions were aspirated from the top of the layered suspension and placed in 1.5ml eppendorf tubes. These aliquots were centrifuged at 9000 RPMs for 10 minutes. The supernatant was aspirated and discarded, while each of the cell pellets were resuspended in 150 micro liters of 1x PBS with protease inhibitors. The above procedure was repeated for all testes. Each of the final 150 micro liter cell suspensions were placed on a Poly-L-lysine coated, coverslip and allowed to sit for 15 minutes at room temperature. The coated coverslips with the cell suspension were then fixed by immersion in 4% paraformaldehyde for 10 minutes at room temperature. After fixation the samples were washed 3 x 15 minutes in 1x PBS and stored at 4 deg. C until immunolabeling.

Immunofluorescence 4P-Microscopy

Samples for use on the 4Pi-microscope were permeabilized in 0.25% T-x-100 for 10 minutes at room temperature and then washed twice in 1x PBS with 0.025% T-x-100(PBST) for 5 minutes. The samples were then incubated in "MAXBlock" (Active Motif) for 1 hour at 37 degrees C in a humid chamber followed by washes in PBST. 200 micro liters of MLH1 antibody (BD Biosciences, 1:50), was placed on the sample and allowed to incubate overnight at 4 degrees C in a humid chamber. After washing in PBST, 200 micro liters of Alexa Fluor 594 (Molecular Probes) secondary antibody was added to the specimen at a 1:800 dilution and allowed to incubate at 37 degrees C for 30 minutes. The samples were again washed in PBST, followed by an incubation with 200 microliters of SYCP3 antibody (BD Biosciences, 1:800 dilution) for 1 hour at 37 degrees C in a humid chamber. The samples were washed again in PBST and then incubated with Alexa Fluor 488 secondary antibody for 30 minutes at 37 degrees C. After washing in PBST, the

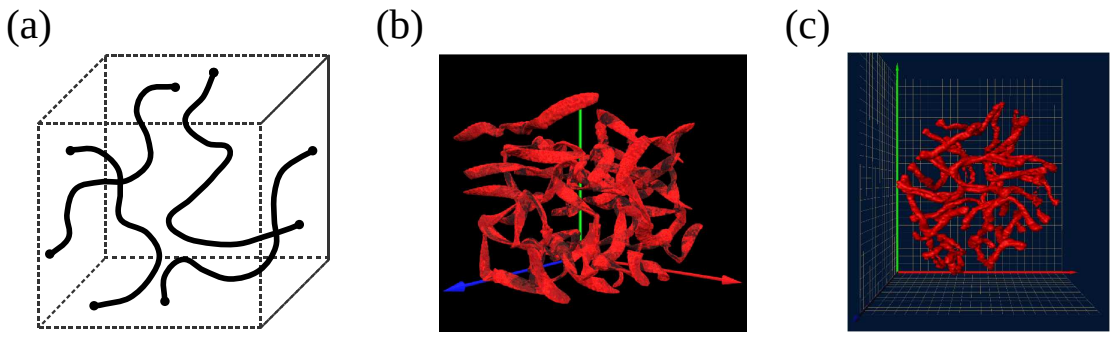


Figure 7.4: (a) Sketch of the applied SC polymer model with the tethered polymer ends being able to diffuse along the envelope of the confining geometry. (b) "Snapshot" of SC polymers in confinement based on the double-tethered SC polymer model. (c) "Snapshot" of synaptonemal complexes in spermatocyte nuclei based on 4Pi-microscopy data. Visual inspection of structural characteristics between the 4Pi-microscopy images and the SC model results such as their end-to-end distance as well as their orientation with respect to each other and the confining cavity indicate that entropy might be one driving force for structural SC organization complementing the dedicated action of specific proteins or actin cables [49, 308, 321]. Image adapted from [322].

samples were mounted in glycerol (n=1.460) and imaged using a Leica TCS 4Pi-microscope equipped with 100x glycerol objectives with an N.A. of 1.35.

Analysis of 4Pi-Microscopy Images

The SC's were reconstructed in Imaris software (Bitplane AG, Zürich, Switzerland). Using the Filament Tracer and Measurement Pro modules, the three-dimensional backbone of each SC was then determined and associated statistics (length, position coordinates) exported.

7.3 Results and Discussion

7.3.1 Shapes of SC Polymers

Fig. 7.4 shows a sketch of the applied SC polymer model as well as a "snapshot" of a model conformation and an image of the mouse SCs in spermatocyte nuclei based on 4Pi-microscopy data. Notably, visual inspection of structural characteristics between the experimental images and the SC model such as their end-to-end distance as well as their orientation with respect to each other and the confining cavity indicate that entropy might be one driving force for structural SC organization complementing the dedicated action of specific proteins or actin cables [49, 308, 321].

End-to-End Distance of SC Polymers A way to characterize the extend of a polymer is its mean squared end-to-end distance $\langle R^2_{\text{end}} \rangle$ as well as the probability distribution PDF thereof. Fig. 7.5 shows $\langle R^2_{\text{end}} \rangle$ as a function of the polymer chain's bending rigidity and the PDF for high and low chain rigidity in the inset. For all bending stiffnesses, tethering of the polymer's ends to the confining cavity forces the polymer into more stretched conformations illustrated by the larger mean squared end-to-end distances and the broader

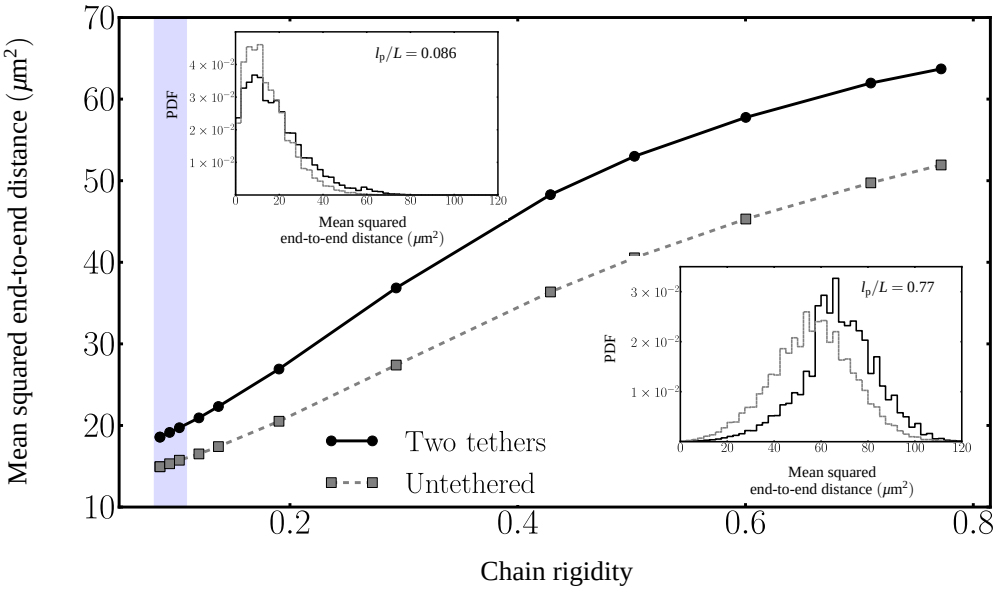


Figure 7.5: Mean squared end-to-end distance $\langle R^2_{\text{end}} \rangle$ as a function of chain rigidity l_p/L for the double-tethered SC polymer model ("Two tethers") as well as for the "null" model of free polymers ("Untethered"). The insets show the probability distribution function (PDF) of the end-to-end distance for the flexible regime $l_\text{p}/L = 0$ as well as for the stiff case $l_\text{p}/L = 0.833$. Semiflexible end-tethered polymers are forces to stretch out between the (moving) attachment sites, leading to larger mean end-to-end distances in agreement with visual inspection of the SC's end-to-end distances in 4Pi-microscopy images. The shaded region indicates the range of bending stiffness that generates the experimentally observed mean squared end-to-end distance of $\langle R^2_{\text{end,exp}} \rangle = 20.31 \pm 2$. Image adapted from [322].

distributions. This can be understood by noting that tethered chains as opposed to free ones can form fewer coil-like conformations which reduce the mean end-to-end distance.

Fig. 7.5 also shows the experimentally determined mean squared end-to-end distance based on the 4Pi-microscopy images, finding $\langle R^2_{\text{end,exp}} \rangle = 20.31 \pm 2$. Comparing the modeling results with the experimental ones, one can locate the SCs in the range of bending stiffness between $l_\text{p}/L = 0.08$ and $l_\text{p}/L = 0.11$, as indicated by the shaded region in Fig. 7.5.

Notably, tethering the polymer ends to the envelope of the confining cavity effectively creates a layer of double-grafted polymers, forming a so called "polymer brush" inside the confined space. A "polymer brush" consists of polymers attached by one or two ends to an interface at relatively high coverage (grafting density) [176]. The physics of polymer brushes has been studied extensively for various grafting densities in the past [176, 177, 185]. The underlying principle is that double-tethered polymers do not intermingle but entropically repel each other, which can be understood by noting that two ring polymers repel each other when brought together within a distance smaller than their gyration radius due to the loss of conformational entropy [23, 56, 165].

Segregational Tendency of SC Polymers This tendency to segregate can be assessed by computing the mean squared distance d^2_cm between the centers of mass, $\mathbf{r}^{(i)}_\text{cm}$ and $\mathbf{r}^{(j)}_\text{cm}$,

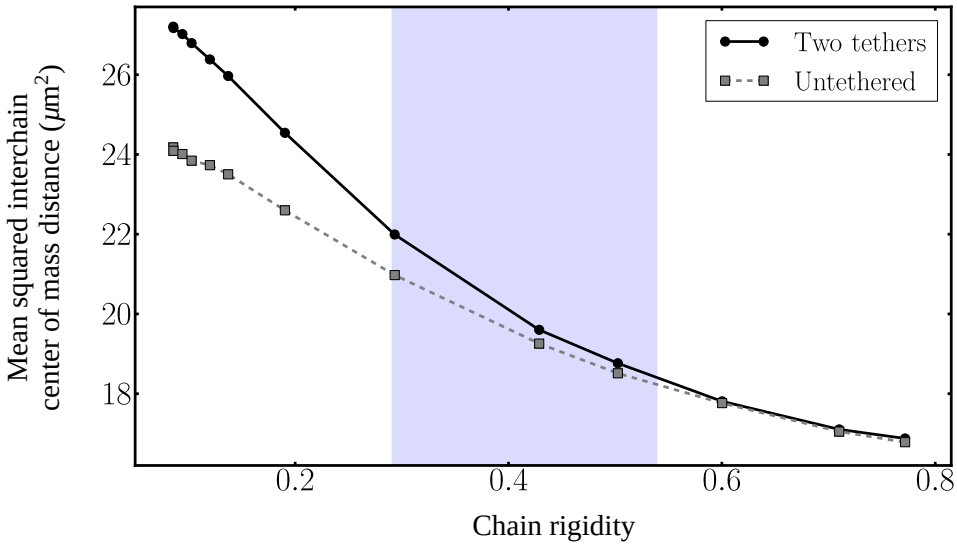


Figure 7.6: Mean squared interchain center of mass distance $d^2_{\mathrm{cm}} = \langle |\mathbf{r}^{(j)}_{\mathrm{cm}} - \mathbf{r}^{(i)}_{\mathrm{cm}}|^2 \rangle$ as a function of chain rigidity l_{p}/L for the double-tethered SC polymer model ("Two tethers") as well as for the "null" model of free polymers ("Untethered"). At low bending rigidity, double tethered SC polymers effectively form a "polymer brush", leading to stronger segregation between them in contrast to the intermingling of free polymers. The shaded region indicates the range of bending stiffness that generates the experimentally observed mean squared center of mass distance of $d^2_{\mathrm{cm,exp}} = 19.0 \pm 2$. Image adapted from [322].

of two polymers i and j, respectively

$$d^2_{\mathrm{cm}} = \langle |\mathbf{r}^{(j)}_{\mathrm{cm}} - \mathbf{r}^{(i)}_{\mathrm{cm}}|^2 \rangle. \tag{7.2}$$

In fact, Fig. 7.6 shows that d^2_{cm} is larger for tethered SC polymers confirming their stronger segregation in contrast to the intermingling of the free polymers.

An increase in bending rigidity reduces this effect which is accompanied by an increase in the end-to-end distance for both the free and the tethered SC model as shown in Fig. 7.5 and 7.6. In the stiff regime and in free space, rod-like, elongated conformations dominate, while in confinement the internal structure of a semiflexible polymer depends on the complex relationship between the accessible volume, the polymer's contour length as well as its persistence length [33]. Both the tethered and the free SC polymer chain are forced to adopt undulating crumbled configurations assuming spool-like structures (which explains $d^{2(\mathrm{tethered})}_{\mathrm{cm}} \approx d^{2(\mathrm{free})}_{\mathrm{cm}}$ in the stiff regime) in order to fit in three-dimensional space.

Comparing again the modeling results with the experimental ones, we find the best agreement for an experimentally determined mean squared center of mass distance of $d^2_{\mathrm{cm,exp}} = 19.0 \pm 2$ with a SC polymer of bending rigidity between $l_{\mathrm{p}}/L = 0.3$ and $l_{\mathrm{p}}/L = 0.53$.

7.3.2 Semiflexibility Induces Frustration: Implications for Meiotic Chromosome Entanglement

A polymer's "crumpledness" can be assessed by the average crossing number ACN, which is a measure for the mean number of chain "overcrossings" [33, 173, 192, 323]. In this

work, we study the impact that tethering and semiflexibility can have on the intrachain-entanglement (self-entanglement) as well as on interchain-entanglement between SC polymers in confinement. Additionally, we determine the amount of intrachain- and interchain-entanglement based on the 4Pi-microscopy images of SCs in male mouse spermatocytes.

Projecting a three-dimensional polymer configuration into a plane defined by a normal vector $\mathbf{n}$ results in a two-dimensional curve which may exhibit crossings. Averaging the number of crossings over all angular perspectives given by all possible normal vectors defines the average crossing number ACN. To calculate the average crossing number of polymer configurations generated by Monte Carlo simulations we follow [256]. The mean average crossing number mACN is then obtained by averaging over the average crossing number of all possible polymer configurations.

The resolution of chain entanglement plays an important role during meiosis, when the formation and the presence of SCs define the two phases, zygotene and pachytene [308,311]. Before and during zygotene, the chromosomes undergo substantial rearrangement, where their ends associate with the nuclear envelope and cluster into a restricted area forming the "bouquet" [49,50]. At pachytene, the chromosome ends are again redistributed throughout the nuclear membrane [49, 50, 304]. A recent study has analyzed key features of SC axial element behavior during zygotene and pachytene [304]. During zygotene, chromosomes are observed to form interlocks with other synapsing pairs of homologous chromosomes. [304]. Interestingly, by late pachytene, these interchain entanglements are removed, possibly by the coordinated breakage and rejoining of chromosomes [324–326] or by chromosome movement and SC disassembly during pachytene [308, 327].

Fig. 7.7 shows both the intra- and interchain mean average crossing number as a function of bending rigidity for the free and the double-tethered SC polymer model. In contrast to the free polymer model, tethering of the SC polymer ends to the envelope of the confining cavity induces less intra- and interchain entanglement.

Notably, with respect to the SC's experimentally observed semiflexibility, we find that an increase in bending stiffness leads to a decrease in interchain overcrossings for both polymer systems, while it induces the reverse trend for interchain-entanglement. In fact, we find a trade-off resulting from the impossibility to minimize both kinds of entanglement at the same time, a concept that is referred to as frustration.

Analysing intra- and interchain entanglement based on the 4Pi dataset, we find a surprisingly low amount of both types of chain overcrossings, $\mathrm{mACN}_{\mathrm{intra,exp}} = 0.1 \pm 0.75$ and $\mathrm{mACN}_{\mathrm{inter,exp}} = 0.13 \pm 0.8$, which cannot be explained within our SC polymer model. While tethering the SC polymer ends reduces the amount of (inter- and intra-) chain entanglement compared to the untethered case, the applied SC model still shows higher chain entanglement than what is experimentally observed.

7.3.3 Assessing the SC Density Distribution with the Cell Nucleus

The analysis of 4Pi-microscopy images allows for the computation of the SCs' three-dimensional configurations within the cell nucleus. In order to assess their distribution within the available volume of the cell nucleus and to compare it to our modeling results, we estimate a probability density function (PDF) from each observation using this PDF to represent the respective three-dimensional "snapshot" images. In this work, we use a classical, parameter-free density estimation technique, Kernel Density Estimation

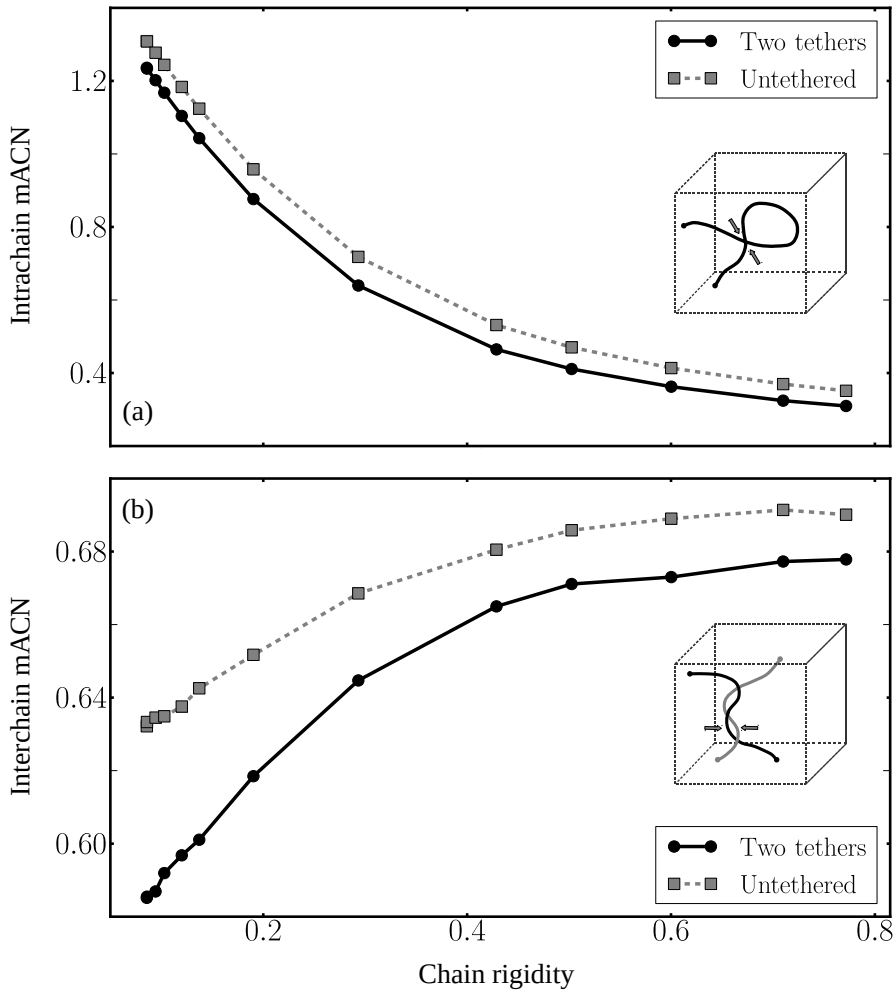


Figure 7.7: (a) Intrachain entanglement ("self-entanglement") and (b) interchain entanglement measured by the mean average crossing number mACN as a function of chain rigidity l_p/L. Tethering the SC polymer's ends to the borders of the confining cavity induces fewer chain overcrossings than the "null model" consisting of free semiflexible polymers in confinement, which suggests that the interplay between tethering and confinement might help to prevent an excess of chain overcrossings. However, semiflexibility induces a trade-off in both polymer systems between the amount of intrachain- and interchain-entanglement which has to be balanced with respect to interlock resolution. Notably, we find a surprisingly low amount of both types of chain overcrossings for the 4Pi-microscopy dataset, $\mathrm{mACN}_{\mathrm{intra,exp}} = 0.1 \pm 0.75$ and $\mathrm{mACN}_{\mathrm{inter,exp}} = 0.13 \pm 0.8$, which cannot be explained within our SC polymer model. Image adapted from [322].

(KDE) [328, 329]. According to this method, the density estimates are given by

$$\hat{f}_h(\vec{x}) = \frac{1}{Nh} \sum_{i=0}^{N-1} K\Big(\frac{\vec{x} - \vec{x}_i}{h}\Big), \tag{7.3}$$

where $K(\cdot)$ is a PDF known as the kernel function, $(\vec{x}_0, \vec{x}_1, \ldots, \vec{x}_{N-1})$ is a sample drawn from an unknown density function f and h a smoothing parameter. The role of the kernel is to "spread" the mass of the observations around its original position. Here, we use a Gaussian kernel.

The representation of an ensemble of conformations as a PDF has been applied successfully in the field of protein research ranking the space of conformations in agreement with NMR observations [330,331]. We are interested in the quantitative comparison of two ensembles of conformations, namely the experimentally determined one by 4Pi-microscopy and the ensemble generated from our SC models.

Fig. 7.8 shows the PDF calculated for sites along the backbone of each SCs within the nuclear volume for various 4Pi-microscopy samples as well as for our SC model sytems at high and low bending rigidity. Notably, the behavior of the experimental PDF fits qualitatively with our (double-tethered) SC polymer model, where the probability density is high along middle-chain regions and drops quickly towards the polymer end regions. This is in contrast to the untethered (free) SC model, where random coil formation induces a less steep decrease of the probability density towards both polymer end regions.

7.4 Discussion

Based on the already known basic features of SC organization, we model the SCs in the nuclear volume by semiflexible polymers whose ends are, in one case, double-tethered to the envelope of the confining cavity and only allowed to diffuse along it, while, in a "null" model, the polymer ends are untethered. Besides imposing self-avoidance on the SC polymer chains, we do not take into account any additional interactions, therewith being able to study the role of (configurational) entropy in shaping SC organization. Additionally, we investigate the impact that (i) tethering the SCs to the nuclear envelope as well as (ii) their semiflexibility have on their spatial organization (and ordering).

Assessing the extention of individual SCs as well their spatial organization with respect to each other we have computed their mean squared end-to-end distance as well as the mean squared distance between the centers of mass of two polymers as a function of their bending rigidity. At low bending rigidity, double-tethered SC polymers effectively form a polymer brush, leading to stronger segregation between them in contrast to the intermingling of free polymers. Semiflexible end-tethered polymers are forced to stretch out between the (moving) attachment sites, leading to larger mean end-to-end distances. Comparing these modeling results with the experimental ones (based on the 4Pi dataset), we find an order-of-magnitude agreement for the investigated range of bending stiffness.

A measure for intra- and interchain entanglement is the mean average crossing number, mACN. Notably, tethering the polymer ends to the borders of the confining cavity induces fewer chain overcrossings than the "null model" consisting of rather flexible polymers in confinement, suggesting that the interplay between tethering and confinement might help to prevent an excess of chain overcrossings. However, semiflexibility induces a trade-off in both polymer systems between the amount of intrachain- and interchain-entanglement which has to be balanced with respect to interlock resolution. Notably, analysing intra- and interchain entanglement based on the 4Pi dataset, we find a surprisingly low amount of both types of chain overcrossings, that cannot be explained by our polymer model.

In order to assess the experimentally determined distribution of the SCs within the available volume of the nucleus and to compare it to our modeling results, we estimate a probability density function (PDF) from each observation using this PDF to represent the respective three-dimensional "snapshot" images. We find a qualitative agreement between the PDF based on 4Pi-microscopy images and the distribution function based on the numerical results for the double-tethered SC model.

Summarizing, this work elucidates the role of entropy in shaping pachytene SC orga-

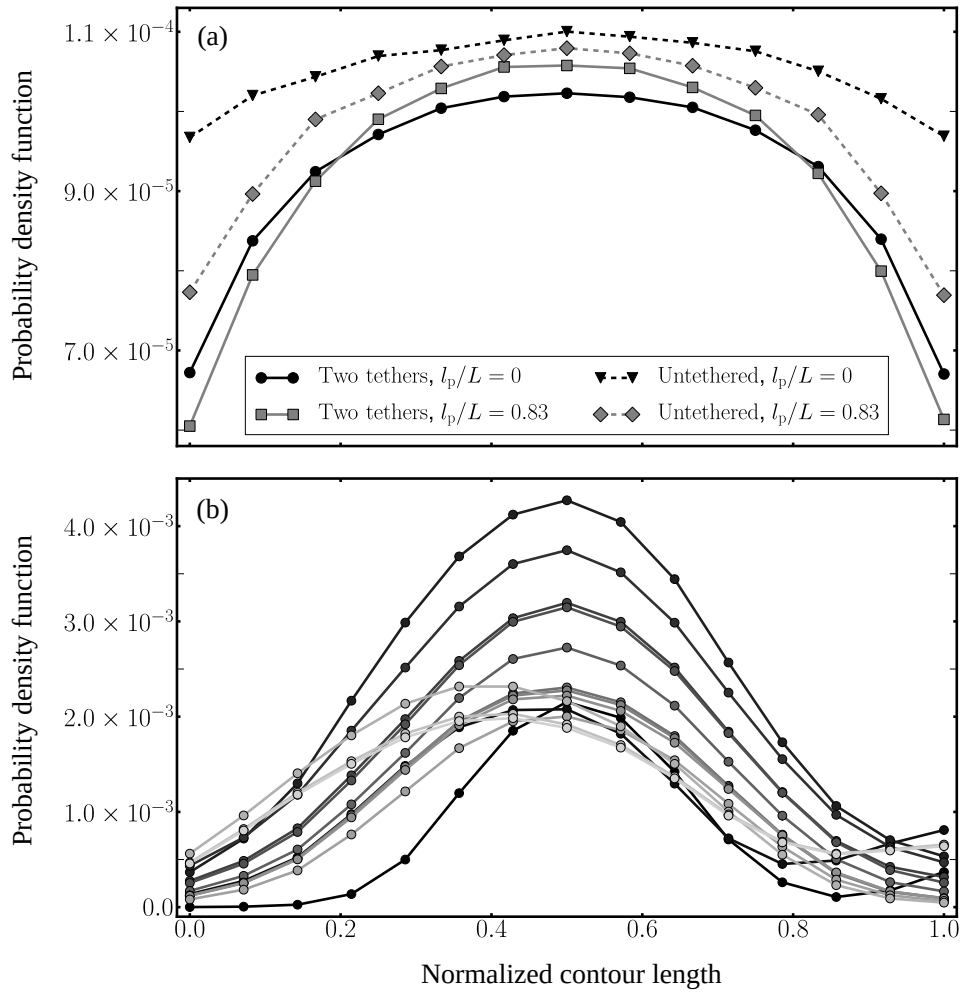


Figure 7.8: Probability distribution function (PDF) calculated for (b) sites along the backbone of each SCs within the nuclear volume for various 4Pi-microscopy samples as well as (a) for our SC model sytems at high and low bending rigidity, $l_p/L = 0.83$ and $l_p/L = 0$, respectively. The behavior of the experimental PDF fits qualitatively to our (double-tethered) SC polymer model, where the probability density is high for middle-chain regions and drops quickly towards the polymer end regions. This is in contrast to the untethered model, where random coil formation induces a less steep decrease of the probability density function towards both polymer end regions. Image adapted from [322].

nization. The framework provided by the complex interplay between SC polymer rigidity, tethering and confinement is able to qualitatively explain features of SC organization, such as mean squared end-to-end distances, mean squared center of mass distances or SC densities distributions. However, it fails in correctly assessing SC entanglement within the nucleus. Our analysis of the 4Pi-microscopy images reveals a higher ordering of SCs within the nuclear volume than what is expected by our numerical model. While entropic contributions are a driving organizational driving force, the dedicated action of proteins or actin cables [49, 308, 321]) might be needed in order to fine-tune the three-dimensional SC organization. Future experiments determining the bending rigidity of SCs within the cell nucleus might help to qualitatively test our assumptions.

Chapter 8

Rabl Model of Yeast Interphase Chromosomes

Spatial Organization and Dynamics

References

The results presented in this chapter are published as and adapted from

- S. Gordon-Messer, B. Avşaroğlu, M. Fritsche, J. Martin, P. Wiggins, J. Sedat, D.W. Heermann, J. Haber and J. Kondev (2011), *Rabl model of yeast interphase chromosomes quantitatively accounts for their spatial organization and dynamics*, in preparation.

Chapter Summary

Different lines of evidence lead to the Rabl model of organization of interphase chromosomes in yeast. Key features of this model are the tethering of the centromeres at the spindle pole body and the localization of telomeres to the nuclear periphery. Using a polymer model to account for the flexibility of yeast chromatin and taking into account the constraints assumed by the Rabl model we compute a number of different quantities associated with the spatial organization and dynamics of yeast chromosomes and compare them to experiments on fluorescently labeled chromosomes. We find good quantitative agreement between theory and experiment for wild type cells. When we remove a protein that is known to tether telomeres to the nuclear periphery the polymer model is no longer able to account for the observations, suggesting that in these mutants the chromosomes are not simply untethered as was previously argued. Our combined analytical, computational, and experimental studies reveal the importance of telomere tethering and nuclear confinement on the spatial organization of yeast chromosomes.

8.1 Introduction

The spatial organization of chromosomes within the nucleus is nonrandom [40,332] and has a strong influence on cellular functioning with respect to gene expression, DNA-damage repair, recombination and replication [217, 333–336]. In fact, different lines of evidence suggest a structure-function relationship for chromosomes where the structure refers to the physical location of genetic loci within the cell [37, 172, 213, 215–217, 337, 338]. However, the question about the mechanism driving the organization of chromosomes in cell nuclei has so far remained unanswered.

In this chapter, we examine the impact of tethering, chromosome flexibility and confinement on the organization of yeast interphase chromosomes. A number of recent studies have investigated the organization of yeast chromosomes in quantitative detail pointing towards a Rabl organization of chromosomes in the interphase nucleus of budding yeast [339–341]. The Rabl model of chromosome configuration is one in which the centromere is located on one side of the nucleus and the telomeres reside on the opposite side [339–341]. In this chapter, we attempt a direct comparison between theoretical calculations and experimental results [37, 172, 213, 215–217, 337, 338].

The tethering interactions considered are those leading to the localization of the centromere to the spindle pole body located at the nuclear periphery, and the localization of the ends of chromosomes, the telomeres, also to the nuclear periphery [339, 342]. Confinement refers to the requirement that the chromosomes are located within the nucleus. Here, we try to answer the question whether this mechanical description of the yeast chromosomes taking into account the constraints imposed by tethering the centromeres and telomeres, and confinement in the nucleus, can quantitatively describe the observed organization and dynamics. For this purpose, we compute observable quantities that describe the organization and dynamics of yeast chromosomes and then directly test the results of these calculation against published data as well as data from our own experiments. The goal is to determine to what extent a simple polymer model can account for the observed organization and dynamics in a quantitative way. This model can then serve as a departure

point for quantitative discussions about the role of chromosome organization in directing important chromosome functions such as gene expression, recombination and replication.

We are particularly interested in the conformational properties of the yeast chromosome III due to its role in mating type switching [216, 343]. Budding yeast haploid cells have the ability to change their mating type as often as every generation. During this process, the MAT locus, located on the right arm of chromosome III, is cut by the HO endonuclease and repaired via gene conversion using one of two homologous donor sequences, HML (Hidden MAT Left) or HMR (Hidden MAT Right), located on the left and right arm of chromosome III, respectively [343]. Depending on the mating type of the mother cell, there is a strong preference for the right or left arm donor in the repair process, leading to mating type switching of 90% of the time. However, the mechanism underlying donor preference has remained elusive so far. Recent studies suggest that it may be linked to the folding of chromosome III therewith promoting the pairing of either MAT and HML or MAT and HMR [172, 343].

Information regarding static chromosome positioning can come from chromosome painting, chromosome conformation capture technologies (3C/4C/5C/High-C) [87, 87, 131, 344–346] as well as from imaging fluorescently labeled loci along the chromosome [35, 37, 172, 198, 216, 217, 337, 338, 347]. We have experimentally and theoretically studied 3D end-to-end distances between HML and the spindle pole body (SPB) in populations of fixed cells, where the removal of YKU80 and ESC1, two proteins involved in anchoring the telomere to the nuclear membrane [38, 338, 348], yields a decrease in the variance of the mutant distance distribution with respect to the wild type distribution. This suggests that telomere anchoring has an effect on the position of HML and its distance from the wild type.

Since chromosomes are dynamic within the nucleus and the majority of loci are engaged in constant, constrained motion [349], this is a place, in which ingenious labeling experiments and physical modeling of polymers in confined space can guide our intuition. In fact, chromosomal diffusion has been measured in bacterial [35, 347], yeast [40], Drosophila [349, 350] and mammalian cells [351, 352]. We measure the mean square displacement between HML and the SPB on chromosome III by applying the OMX system [353] which allows us to take live cell images for long periods of time. Removing the telomere anchor leds to a higher plateau value suggesting that motion of HML is increased when the anchor is untethered. This finding seems to be in contrast to the fixed cell measurements where the variance of the distance distribution between HML and the SPB decreases in an untethered mutant.

8.2 Experimental Approach and Modeling

8.2.1 Measurement of Distance Distributions in Fixed Cells

Susannah Gordon-Messer has performed the fixed and live cell measurements [354], where individual chromosomal loci are labeled such that specific positional information can be extracted [37, 172, 198, 217, 336–338]. Within the yeast nucleus, telomeres are arranged in clusters and tethered to the nuclear envelope [340, 341]. The spindle pole body (SPD) is known to be embedded in the nuclear envelope [339], too, and on average the microtubules connecting it to the centromere are 50 nm in length [354]. These elements are expected to position the chromosomes such that they orient themselves according to the Rabl model of organization [339–341].

The spatial organization of yeast chromosome III is of particularly interest due to its role in mating type switching, which has been explained above. While there is evidence

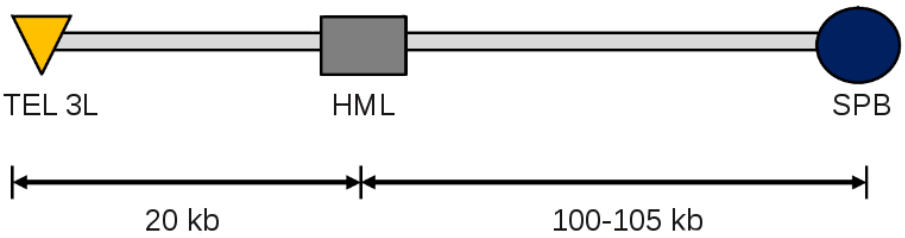


Figure 8.1: Schematic view of the yeast chromosome III arm. The genetic locus HML and the SPB are separated by 100 kbp. Image adapted from [355].

that yeast chromosome III is spatially organized according to tethering from the SPB and from the telomeres, there is limited data that come from the direct measurement between two labeled loci along the same chromosome arm.

Thus, we use fluorescent markers in yeast near the HML locus on chromosome III and at the SPB, which serves as a marker of centromere attachment, to determine the distribution of distances between these two loci. Fig. 8.1 schematically illustrates the left arm of yeast chromosome III.

If we assume a toy model of the left arm of chromosome III in which the SPB and the left telomere are the only anchoring points, we expect a change in the distance distribution between HML and the SPB if the telomere is untethered. In the Ku80/Esc1 deletion mutant cells, the telomere tethers are removed by deleting the proteins Ku80 and Esc1, which are known to play a role in the anchoring mechanism [38, 338, 348].

Moreover, it was found that the chromosome arm length influences the positioning of the telomeres [342], which is why we study the Δ12kbp mutant comprising of a 12 kbp shortened chromosome III left arm. Further details on the performed fixed cell measurements can be found in section 8.5.

8.2.2 Live Cell Imaging Reveals the Dynamics of a Chromosome Arm

Fixed cell measurements provide information about the organization of chromosomes *in vivo*, but do not give insight into the dynamics of the chromosomes. In order to look at chromosomal movement, it is necessary to take live cells images over the course of time. However, such an approach is often hindered by the fact that cells are sensitive to excitation light, which can result in both cell damage and photobleaching as was discussed in chapter 3. In order to minimize these effects, the OMX system is applied [353]. It takes images by employing a low excitation light intensity and increases the signal-to-noise ratio of the resulting images through the use of a denoising algorithm [353].

By employing this system, we are able to follow the diffusive movement of the left arm of chromosome III. We measure the diffusive motion of the vector connecting the position of HML and the SPB for wild type and Ku80/Esc1 deletion mutant cells. Further details on the performed live cell measurements can be found in section 8.5.

8.2.3 Random Walk Model of Yeast Chromosomes

Baris Avşaroğlu and Jane Kondev have performed the distance distribution calculations for the random walk model. We apply such a simple model of yeast chromosomes in order to determine whether or not the Rabl model quantitatively describes the experimental data on the left arm of chromosome III in the yeast nucleus. Random walks are zeroth

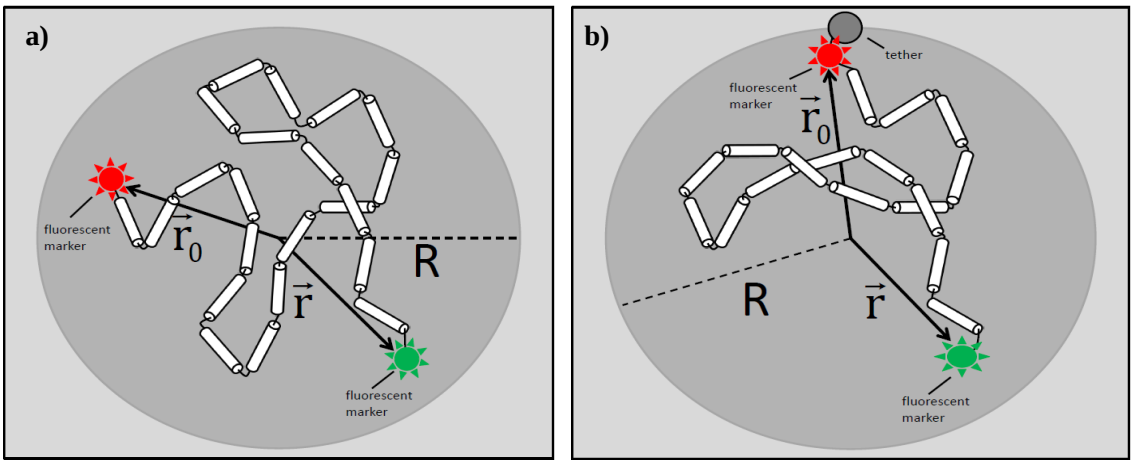


Figure 8.2: Schematic view of a random walk polymer confined to a sphere. The number of conformations that the random walk polymer can explore is reduced due to a) the compressing effect of the sphere with radius R and/or b) being double- or single-tethered to the surface of the sphere. Here, the magnitude of the vector connecting green and red fluorescent markers, $\rho = |\vec{r} - \vec{r_0}|$, denotes the end-to-end distance. Image adapted from [355].

order models to study the chain statistics of biopolymers and consists of N randomly oriented, rigid rods with Kuhn length a, connected to each other to form the chain of contour length $L = Na$.

Here, we assume screening of the self-avoiding interaction between monomers of a single chain (chromosome) which is brought about by the presence of multiple chains (chromosomes) that interpenetrate in the nucleus. This screening is effective above a particular length scale, the so-called mesh size, which depends on the overall monomer density, and the polymer properties of the chromosomes, in particular the packing density and the Kuhn length.

For the yeast chromosomes the mesh size is estimated to correspond to about 14 kbp of DNA which means that for genetic loci that are separated by larger genomic distances the chromatin fiber that connects them can be treated as an ideal polymer chain. 3C data on yeast chromosomes which finding the frequency of contacts scales as $G^{-1.5}$, where G is the genomic distance between chromosome loci, confirms this predictions [131, 344, 346].

To this end, yeast interphase chromosomes have been examined using a number of different experimental techniques and the emerging consensus is that on the scale of a few hundred nanometers their mechanical properties are well described by the model of a semiflexible polymer with a Kuhn length of 200 nm and a packing density of 25 bp/nm [172, 346]. The reported packing density is smaller than the number usually quoted for the highly debated 30 nm-fiber model of chromatin structure observed in higher eukaryotes, indicating a looser structure, more akin to the 10 nm-fiber. Note that the packing density of double-stranded DNA is 3 bp/nm.

In addition to the total number of subunits and the size of an individual subunit, the spatial organization of a random walk polymer is effected by extrinsic constraints. In this work, the random walk polymer chain is packed in a volume and/or is anchored to the surrounding cavity as illustrated in Fig. 8.2, where the chain's limited spatial positioning and motion of its subunits leads to a loss of conformational entropy.

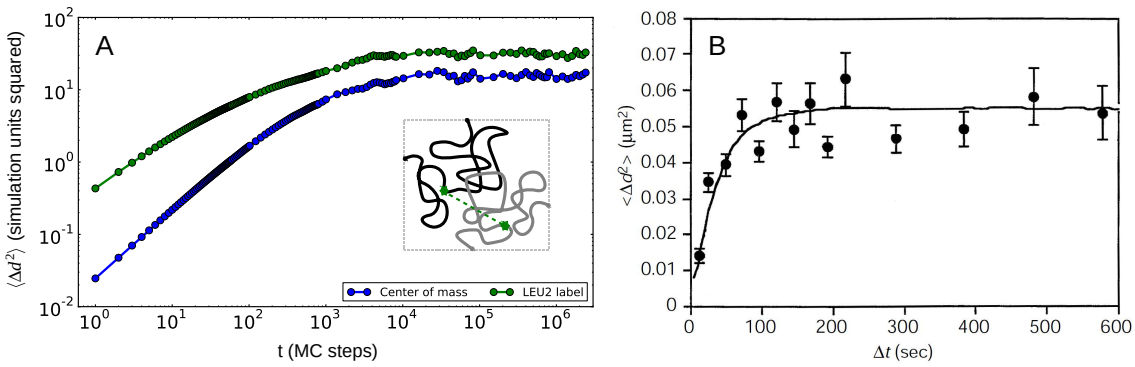


Figure 8.3: (A) Based on our yeast chromosome model, we perform a simulation with two 120 kbp double-tethered chromosomes and determine the mean squared distance displacement between both LEU2 sites $\langle \Delta d^2 \rangle$ in order to compare it to (B) the experimentally determined data by Sedat et al. [40]. (B) Mean squared displacement between fluorescent spots (LEU2) on two yeast chromosomes in live yeast cells plotted against the elapsed time interval. The solid line is the best-fit curve derived by computer simulation using the parameters $D = 5 \times 10^{-12} \mathrm{cm}^2/\mathrm{sec}$ and confinement radius $R = 0.3\,\mu\mathrm{m}$. For additional details on (B), the reader is referred to Ref. [40]. We identify the 1000 MC steps needed to arrive at the plateau of 30 simulation units squared (SU^2) in our polymer model with the plateau at $0.055\,\mu\mathrm{m}^2$ after 100 sec in the Sedat et al. data. This leads to a convergence factor for the distance of $C_{\mathrm{distance}} = 0.055\,\mu\mathrm{m}^2/30\,\mathrm{SU} = 0.0018\,\mu\mathrm{m}^2/\mathrm{SU}$ and for the time of $C_{\mathrm{time}} = 100\,\mathrm{sec}/1000\,\mathrm{MC\,steps} = 0.1\,\mathrm{sec}/\mathrm{MC\,steps}$. Images adapted from [355] and [40].

8.2.4 Simulating the Dynamics of Yeast Chromosomes

The dynamics of the 120 kbp chromosome III arm are modeled by a coarse-grained polymer of $N = 24$ Kuhn segments assuming a packing density of 25 bp/nm and a Kuhn length of 200 nm as explained above [131,172]. The polymer is confined to a cube whose dimensions are $D_{\mathrm{x}} = D_{\mathrm{y}} = D_{\mathrm{z}} = 8$ Kuhn segments corresponding to a nucleus of diameter $r_{\mathrm{D}} = 2\,\mu\mathrm{m}$.

We employ the bond-fluctuation method (BFM) [166] without self-avoidance to generate chain conformations. The reader is referred to chapter 3 for a detailed description of the algorithm.

For the model that described chromosomes in wild-type cells, both polymer ends are tethered to the envelope of the confining box and only allowed to diffuse along it. The mutant case is modeled by a polymer with one untethered end which can move freely within the confining volume and one tethered end forced to move along the boundaries of the simulation box representing the nucleus. The box geometry was chosen such as to simplify the calculations and we checked that it reproduces the same equilibrium distribution as was obtained above for spherical confinement.

In order to compare the simulational to the experimental results we have to relate Monte Carlo (MC) steps to real time: Sedat and coworkers have measured the diffusional motion of chromatin in yeast by sub-micrometer single-particle tracking of the LEU2 locus (located 23 kbp away from the centromere) finding a diffusion constant of $D_{\mathrm{exp}} = 5 \times 10^{-12}\,\mathrm{cm}^2/\mathrm{sec}$ [40].

Based on our yeast chromosome model, we perform a simulation with two 120 kbp double-tethered chromosomes and determine the mean squared displacement, $\langle \Delta d^2 \rangle$, between the two segments representing the LEU2 sites as illustrated in Fig. 8.3 A. We compare the so determined mean squared displacement with the experimentally deter-

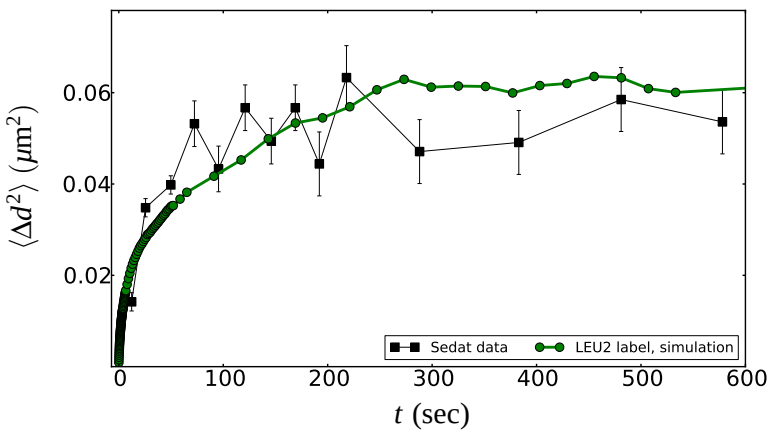


Figure 8.4: Fine-tuning of the conversion factors allows for an even better agreement between theory (green) and the experimental data from Sedat et al. (black). In the following, $C_{\text{distance}} = 0.0021$ and $C_{\text{time}} = 0.065$ are used to "calibrate" the dynamics of our yeast chromosome model. Image adapted from [355].

mined results of Sedat and coworkers in Fig. 8.3 B by applying the following approach: The Sedat et al. data shows a plateau height of $0.055\,\mu\text{m}^2$, which is reached after 100 sec. Thus, we identify the 1000 MC steps needed to arrive at the plateau of 30 simulation units squared (SU^2) in our model with the plateau at $0.055\,\mu\text{m}^2$ after 100 sec in the Sedat et al. data. This leads to a convergence factor for the distance of

$$C_{\text{distance}} = \frac{0.055\,\mu\text{m}^2}{30\,\text{SU}} = 0.0018\,\frac{\mu\text{m}^2}{\text{SU}}$$

and for the time of

$$C_{\text{time}} = \frac{100\,\text{sec}}{1000\,\text{MC steps}} = 0.1\,\frac{\text{sec}}{\text{MC steps}}.$$

Notably, we already have a conversion factor for the distances based on the Kuhn segment length of $l_{\text{K}} = 200$ nm in our model. This conversion factor is $C_{\text{distance}}^{\text{Kuhn}} = 0.0049\,\frac{\mu\text{m}^2}{\text{SU}}$ and is of the same order of magnitude as $C_{\text{distance}} = 0.0018\,\frac{\mu\text{m}^2}{\text{SU}}$. In fact, fine-tuning of the conversion factors as illustrated in Fig. 8.4 shows that an even better agreement with experiment can be achieved by using conversion factors of $C_{\text{distance}} = 0.0021$ and $C_{\text{time}} = 0.065$ which are applied in the following to "calibrate" the dynamics in our yeast chromosome model.

8.3 Results and Discussion

8.3.1 Confined and Tethered Random Walk Polymers

A good measure of the spatial organization of a polymer is the end-to-end distance ρ, which has already been introduced in chapter 3. The end-to-end distance distribution of a random walk polymer is equivalent to the statistical distribution of walks for a particle undergoing Brownian motion, which is given by the solution of the diffusion equation in the continuum limit, where $N >> 1$ [134, 135, 137–141].

To investigate the effect of confinement and tethering, we computed the distributions of end-to-end distances of the random walk polymer under these conditions. We use

the Green's function method to calculate the end-to-end distance distributions of which a detailed derivation is given in section 8.5. For a random walk polymer of contour length $L = Na$ in a sphere of radius R, whose initial and final segments are located at $\vec{r_0} = (r_0, \theta_0, \phi_0)$ and $\vec{r} = (r, \theta, \phi)$, respectively, the end-to-end distance is given by $\rho = |\vec{\rho}| = |\vec{r} - \vec{r_0}|$. Thus, the end-to-end distance distribution $P(\rho; Na)$ can be computed as

$$P(\rho; Na) = \int P(\vec{\rho}; Na)\, d\Omega = \int P(\vec{r}, \vec{r_0}; Na) |\frac{d^3\vec{r}}{d^3\vec{\rho}}|\, d\Omega, \tag{8.1}$$

where

$$P(\vec{r}, \vec{r_0}; Na) = \left[\sum_{l=0}^{\infty} \sum_{s=1}^{\infty} \sum_{m=-l}^{l} \frac{2}{\sqrt{r_0} R^2} \frac{J_{l+\frac{1}{2}}(\mu_{ls} \frac{r_0}{R})}{[J_{l+\frac{3}{2}}(\mu_{ls})]^2} Y_l^{m*}(\theta_0, \phi_0) \times \right.$$
$$\left. e^{(-\frac{a^2}{6}(\frac{\mu_{ls}}{R})^2 N)} \frac{J_{l+\frac{1}{2}}(\mu_{ls} \frac{r}{R})}{\sqrt{r}} Y_l^m(\theta, \phi) \right] \times$$
$$\left[\sum_{s=1}^{\infty} \sqrt{\frac{8R}{r_0 \pi}} \frac{-\mu_{0s} \cos(\mu_{0s}) + \sin(\mu_{0s})}{\mu_{0s}^{5/2}} \frac{J_{1/2}(\mu_{0s} \frac{r_0}{R})}{[J_{3/2}(\mu_{0s}]^2} e^{(-\frac{a^2}{6}(\frac{\mu_{0s}}{R})^2 N)} \right]^{-1}. \tag{8.2}$$

Here, $|\frac{d^3\vec{r}}{d^3\vec{\rho}}|$ is the Jacobian matrix for the coordinate change between $\vec{r}$ and $\vec{\rho}$ and $d\Omega$ is the infinitesimal solid angle. $J_{l+1/2}(\mu_{ls} \frac{r}{R})$ are the Bessel functions of first kind of order $l + 1/2$, μ_{ls} is the sth zero of the first kind Bessel function of order $l + 1/2$ and $Y_l^m(\theta, \phi)$ are the spherical harmonics.

Figure 8.5 shows the results of end-to-end distance distribution calculations for a chain of contour length $5\,\mu$m and segment length $0.2\,\mu$m under three different conditions: (i) freely diffusing in infinite medium, (ii) diffusing in a sphere of radius $1\,\mu$m and (iii) diffusing with one end tethered at the surface of a sphere of radius $1\,\mu$m.

When compared to the freely diffusing chain that obeys the Gaussian distribution discussed in chapter 3, the confined random walk polymer shows a narrower distribution with smaller end-to-end distances. For the chain with one end tethered to the surface of the sphere, the distributions are wider compared to the untethered, but confined, case with an increased mean end-to-end distance. This can be understood by noting that the repulsion between the free chain end and the walls limits the exploration of the chain in conformational space.

So far, we have investigated the effect of confinement and tethering on the spatial organization of random walk polymers, where either one or both of the ends are free. However, a polymer model with both ends tethered is required to address the spatial organization of chromosomes showing Rabl type orientation [339–341].

For this purpose, we exploit the convolution of two distance distributions for chains with one tethered and one free end. The positions where the free ends of both chains are in contact are taken under consideration. Thus, the distance distributions between two segments can be computed as

$$P_{\text{convolution}}(\rho; N_1 a) = \int P_{\text{convolution}}(\vec{\rho}; N_1 a)\, d\Omega, \tag{8.3}$$

where

$$P_{\text{convolution}}(\vec{\rho}; N_1 a) = \frac{P(\vec{r}, \vec{r}_{01}; N_1 a) P(\vec{r}, \vec{r}_{02}; N_2 a)}{\int P(\vec{r}, \vec{r}_{01}; N_1 a) P(\vec{r}, \vec{r}_{02}; N_2 a) d^3\vec{r}} P(\theta_{12}) |\frac{d^3\vec{r}}{d^3\vec{\rho}}|. \tag{8.4}$$

Here, $P(\theta_{12})$ denotes the probability to find the polar angle θ between the two chain ends. The individual end-to-end distance distribution functions are exactly the same as the ones

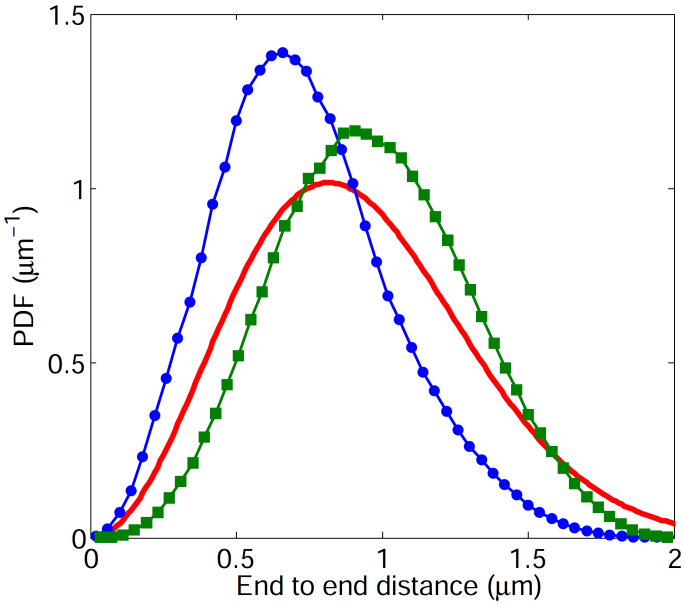


Figure 8.5: End-to-end distance distribution calculations for a chain of contour length 5 μm and segment length 0.2 μm under three different conditions: (i) freely diffusing in infinite medium (red line), (ii) diffusing in a sphere of radius 1 μm (blue line with spheres) and (iii) diffusing with one end tethered at the surface of a sphere of radius 1 μm (green line with squares). Image adapted from [355].

that we have used in our previous calculations, including the term Na^2. This is the average squared end-to-end distance term for a random walk polymer. For chromosomes it can be expressed in terms of genomic distance, G, and a parameter that combines chromosome flexibility and DNA packing density, γ, according to $Na^2 = G\gamma$ [131, 138, 356].

Experimental observations addressing the location of chromosome ends suggest that telomeres are mobile on the nuclear periphery rather than being tethered to specific locations [342, 356]. To test this hypothesis, we compute the end-to-end distance distributions for chains that have fixed tethered ends with angular separation θ. We confront the fixed tethered with the mobile case, where we consider unfixed, weighted polar angle separations between the tethered chain ends. The statistical weights of the tethering angles were computed as a function of chain length, nuclear size and chromosome flexibility by using the Green's function method for random walk polymers diffusing only on the surface of a sphere. These calculations have been performed by assuming a long chromosome arm of $G_{\text{long}} = 100$ kbp (represented by a chain of length 17 segments between red and green markers in Fig. 8.2) connected to a shorter chromosome part of length $G_{\text{short}} = 20$ kbp (represented by a chain of 4 segments in Fig. 8.2). A chromosome flexibility of $\gamma = 10\,\mu\text{m}^2/\text{Mbp}$ and a nucleus of radius $R = 1\,\mu$m is assumed.

Fig. 8.6 A shows the effect of tethering location and mobility on the distance distributions between two distant segments, representing the SPB and the HML locus. While the non-moving tether distributions showed peaked characteristics, the distance distributions computed for mobile tethers are broader. Thus, mobile chain ends wandering over the surface are entropically more favorable, maximizing the distribution variance, than their fixed counterparts. This finding supports the experimental observations of distributed angular positioning of yeast subtelomeric regions [342], suggesting that telomeres wander

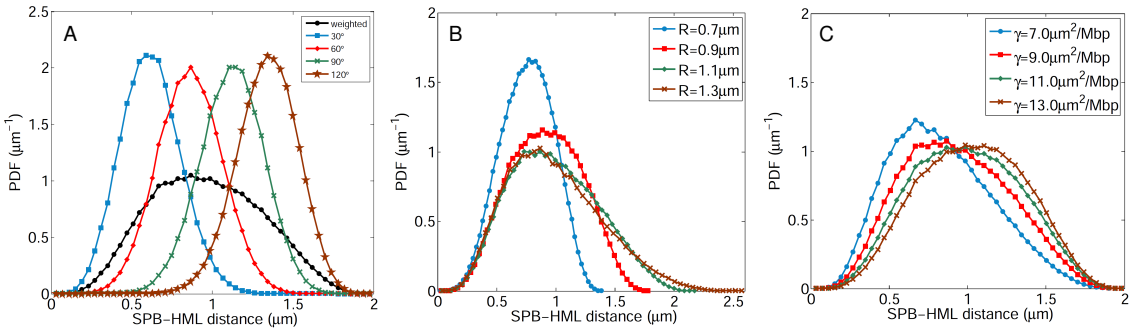


Figure 8.6: (A) The effect of tether location on the distance distributions between two segments. The distributions of distances between segments separated by genomic distance 100 kbp, for a double-tethered polymer with total genomic length of 120 kbp were computed with fixed 30° (blue line with squares), 60° (red line with diamonds), 90° (green line with crosses), 120° (brown line with stars) and unfixed (black line with circles) polar angle separations between tethers. While the non-moving tether distributions show peaked characteristics, the distance distributions computed with a weighted angular separation between tethers are broader. These findings support the experimental observations of distributed angular positioning of yeast subtelomeric regions [342], suggesting the hypothesis that telomeres wander on the nuclear periphery during (early)-interphase. (B) The effect of confinement size on the distance distributions. Distance distributions are shown between loci with genomic distance 100 kbp in spheres of radii 0.7 μm (blue line with circles), 0.9 μm (red line with squares), 1.1 μm (green line with diamonds) and 1.3 μ m (brown line with crosses), respectively. Chains that are confined to volumes of the order of their radius of gyration experience the effective repulsion of the confining walls, while large enough confinement allows random walk polymers to diffuse freely, converging to the ideal behavior obeying Gaussian statistics. The chromosome flexibility parameter, γ, is 10 μm^2/Mbpp for all cases. (C) The effect of chromosome flexibility on the distributions of distances. The distance distributions for varying γ values of 7 μm^2/Mbp (blue line with circles), 9 μm^2/Mbp (red line with squares), 11 μm^2/Mbp (green line with diamonds), 13 μm^2/Mbp (brown line with crosses) are effected by chain flexibility. The box radius is kept constant at 1 μm. Images adapted from [355].

on the nuclear periphery during (early)-interphase.

Within the Rabl polymer model of yeast chromosomes, two additional parameters have an impact on chromosome organization: confinement size as well as chromosome flexibility. Since yeast nuclei can vary greatly in size [357], we analyze the wild type model in different nuclear radii in the range of 0.7 μm to 1.3 μm. The third parameter that was altered is the gyration coefficient of the yeast chromatin fiber, γ, where an increase in the gyration coefficient corresponds to a decrease in the compaction of the chromatin fiber.

Fig. 8.6 B reports the effect of confinement size on the distribution of distances between segments, representing the genetic loci SPB and HML, for a double-tethered random walk polymer. Chains which are compressed in confined spaces of the order of their radius of gyration experience the effective repulsion of the surrounding walls as has already been pointed out in chapter 4. On the other hand, large enough radii of the spherical volume allow random walk polymers to diffuse freely, such that they converge to the ideal chain behavior obeying Gaussian statistics.

Fig. 8.6 C displays the impact of chromosome flexibility on the distribution of distances between two genetic loci of genomic separation 100 kbp for a double-tethered random walk polymer of contour length 120 kbp. As expected, the mean segment-segment distances increase with decreasing chain flexibility and the distance distributions become broader.

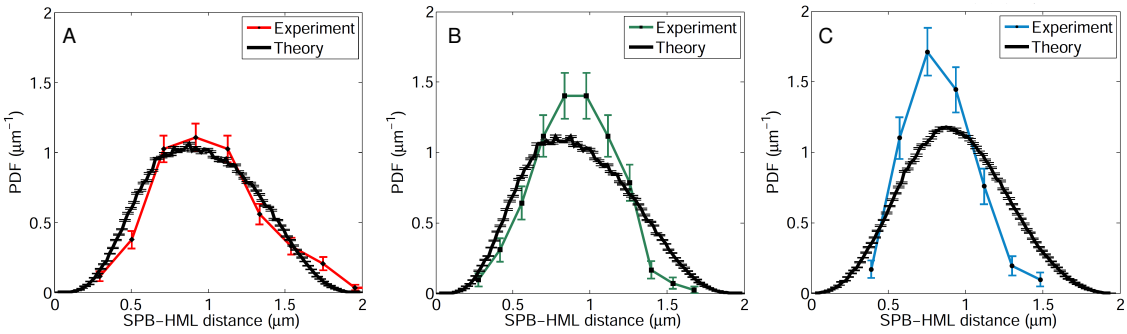


Figure 8.7: The theoretical predictions (black curves) are compared to the experimental results for the distribution of distances between SPB and HML in (A) fixed wild type cells (red lines with black diamonds), (B) fixed Δ12kbp mutant cells (green lines with black squares) and (C) ΔYKu80/Esc1 mutant cells (blue lines with black circles). Both the wild type and the Δ12kbp mutant distributions are in good agreement with the predictions from theory, while the polymer model is no longer able to account for the experimental observations when the telomere tether is removed. The chromosome flexibility parameter, γ, is $10\,\mu\text{m}^2/\text{Mbp}$ and the nuclear radius is $R = 1\,\mu\text{m}$ in all cases (A)-(C). Images adapted from [355].

8.3.2 Deletion of the Proteins Involved in Telomere Tethering Yields a Shift in Distance Distribution

In order to explore the applicability of the Rabl model to the case of the yeast chromosome III, we compare our theoretical predictions with the experimental results. We have determined the three dimensional distance between HML and the SPB by fluorescently labeling these sites on yeast chromosome III. Measurements were performed for wild type cells and for two mutants: Δ12kbp mutant and Ku80/Esc1 deletion mutant cells.

Fig. 8.7 A shows the experimental and theoretical distance distributions for the wild type, while the distributions for the Δ12kbp mutant and the Ku80/Esc1 deletion mutant are displayed in Fig. 8.7 B and C, respectively. In all three cases, the chromosome arm is modeled with a flexibility of $\gamma = 10\mu\text{m}^2/\text{Mbp}$ within the nuclear volume of radius $R = 1\mu\text{m}$.

If the two main tethers of the yeast chromosome III are the SPB and the telomere and if there are no additional tethers then changing the chromosome arm length should reduce the mean of the distance distribution with respect to the wild type. In fact, both the wild type and the Δ12kbp mutant distributions are in good agreement with the predictions from theory.

However, removing the telomere tethers the polymer model is no longer able to account for the experimental observations. It was expected that removing the constraints on the telomere would allow HML to occupy a larger variety of conformations within the nucleus and therefore would give a larger distribution width as compared to that of the wild type. Thus, we posit that in these mutants the chromosomes are not simply untethered as was previously argued, but there might exist additional mechanisms responsible for localizing HML within the nucleus.

8.3.3 Deletion of Proteins Involved in Telomere Tethering Affects Chromosome Dynamics

Budding yeast provides an ideal model system in which to study chromosome dynamics, since there is evidence of constrained chromosome motion but existence of chromosome territories is disputed [40, 358]. In fact, the two main motion-constraining factors are the anchoring points of the yeast chromosome: the centromere and the telomeres. Using the same cells employed in the distance distribution measurements, we measured the mean squared displacement between the SPB and HML in wild type and Ku80/Esc1 deletion mutant cells as a function of time. The theoretical predictions were computed by simulating confined double-tethered and single-tethered random walk polymers of 120 kbp contour length.

We focus our attention on the mean square displacement (MSD) between the monomers representing the HML locus and the 100 kbp distant SPB

$$g(t) = \langle [\vec{r}(t) - \vec{r}(0)]^2 \rangle, \tag{8.5}$$

where $\vec{r}(t) = \vec{R}_{\mathrm{HML}}(t) - \vec{R}_{\mathrm{SPB}}(t)$ is the distance vector between the spatial positions of HML, $\vec{R}_{\mathrm{HML}}(t)$, and the SPB, $\vec{R}_{\mathrm{SPB}}(t)$. In analyzing diffusive motion, the mean square displacement increases according to $g(t) \sim t^{\alpha}$, where $\alpha = 1$ holds for standard two particle diffusion, $\alpha < 1$ refers to sub- and $\alpha > 1$ to superdiffusion.

Fig. 8.8 illustrates the MSD between HML and SPB for a time range of $\Delta t = 110$ sec for both the wild type and the Ku80/Esc1 deletion mutant cells, reporting good agreement between theory and experiment. At intermediate timescales comparable to the relaxation time of the system, we find $\alpha = 0.5 \pm 0.03$. This slope value is consistent with the subdiffusive motion that is seen in polymer diffusion as well as with values that have been measured for prokaryotic chromosome diffusion [347, 359] and fast time regimes for human telomeres [351]. Eventually, a leveling-off is reached due to the surrounding motion-constraining confinement. While the limited imaging timescales, do not allow for the experimental determination of the final plateau heights, we numerically find a crossover between wild type and mutant MSD curves. Fig. 8.8 C shows that this crossover leads to an increased final plateau height for the wild type with respect to the mutant.

Our results suggest that both the wild type and mutant chromosome III arm can be understood in terms of a polymer model. The underlying dynamics are governed by tethering the telomere to the nuclear periphery as well as by the complex interplay between entropic expansion of the chain and the repulsive forces exerted by the confining envelope.

8.4 Conclusions

Interphase chromosomes in budding yeast exhibit a high degree of organization, which we have studied by position tracking of fluorescently labeled chromosomal loci. In particular, we have employed this technique to study the organization and dynamics of the left arm of chromosome III. In a first step, we have analyzed the distances between HML and the SPB in fixed wild-type, fixed Δ12kbp mutant and fixed Ku80/Esc1 deletion mutant cells.

In addition to fixed cells, we then study the dynamics of the left arm in live cells. Using the OMX platform at UCSF, we were able to follow individual loci over long periods of time without the problems of photobleaching and phototoxicity.

The combination of fluorescence microscopy as well as analytical and computational modeling is used to investigate to which extent Rabl organization is sufficient to predict

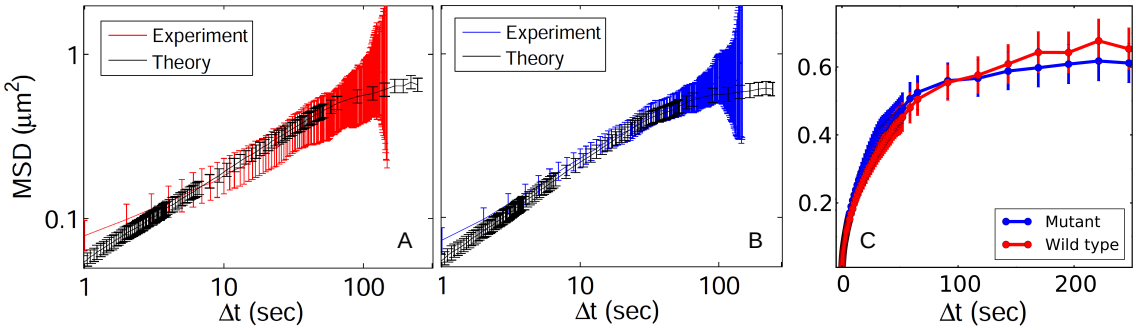


Figure 8.8: The comparison of theoretical predictions (black in A and B) and experimental results of the mean squared displacement (MSD) of the vector between SPB and HML in (A) live wild type yeast cells (red) and (B) ΔYKu80/Esc1 mutant cells (blue). Both the experimental and theoretical mean squared displacement lines display slopes around 0.5 on a log10-log10 scale, which supports the subdiffusive behavior of interphase chromosomes in wild type and mutant yeast cells. (C) Comparison of the theoretical MSD curves for wild type (red) and ΔYKu80/Esc1 mutant (blue) cells. The mutant shows an initial offset with respect to the wild type, while the wild type reaches a higher final plateau height. Images adapted from [355].

the experimental results. The applied polymer model accounts for the flexibility of yeast chromatin and takes into account the constraints assumed by the Rabl organization, i.e. telomere/centromere tethering and confinement in the nuclear volume.

Both the wild type and the Δ12kbp mutant distance distributions are in good agreement with the predictions from theory. However, the deletion of Yku80 and Esc1 leads to a decrease in the variance of the HML-SPB distance distribution as compared to the wild type. Notably, this suggests that the motion of HML is more constrained in the Ku80/Esc1 deletion mutant, which is opposite to what is expected if one polymer end was untethered. On the other hand, the experimental analysis of the dynamics of wild type and mutant cells shows that the mean squared displacement is offset in the mutant with respect to the wild type. This suggests that the motion of HML becomes less constrained in the mutant, which seems to be in contrast to the fixed cell measurements where the variance of the distance distribution between HML and the SPB decreases in an untethered mutant.

Here, one possible explanation is that, when the tether is released, HML is free to diffuse in three dimensions, which leads to faster equilibration, in contrast to the more constraint wild type case, where only "two-dimensional" movement of the telomeres on the envelope is possible. In fact, our dynamics simulations show a crossover between wild type and mutant MSD curves, which leads to an increased final plateau height for the wild type with respect to the mutant, at a crossover time of $t_{\text{crossover}}$ >110 sec. This period is, however, not covered by the experimentally studied life cell imaging time span of $t = 100$ sec.

8.5 Supplementary Information

8.5.1 Green's Function Method for Tethered and Confined Polymers

The subsection 8.5.1 on the Green's function method is kindly provided by Jane Kondev and Baris Avşaroğlu. Its use for the solution of inhomogeneous partial differential equations, subject to specific initial and boundary conditions, is well known in mathematical

physics. Here, it is applied to compute the end-to-end distance distribution of polymers that are confined into a sphere of radius R with one of the ends tethered to the inner surface of this sphere. The random walk model, i.e. the paths of particles under Brownian motion, is applied to resemble the polymer configurations.

Let us first define the Green's function, $G(\vec{r}, \vec{r}_0; N)$, that denotes the fraction, or the statistical weight, of random walk polymers with N segments, each with length a, whose ends are positioned at $\vec{r}$ and $\vec{r}_0$, respectively. Remembering the fact that the Brownian nature of the random walk polymers represents a Markovian process [137, 138, 144], the fraction of the chains of length $(N+1)a$, whose ends are at $\vec{r}$ and $\vec{r}_0$, can be given as the sum of fractions of chains of length Na, whose ends are located at $\vec{r\prime}_i$ and $\vec{r}_0$ times a multiplicity factor w

$$G(\vec{r}, \vec{r}_0; N+1) = w \sum_{i}^{NNN} G(\vec{r\prime}_i, \vec{r}_0; N). \tag{8.6}$$

Here, $\vec{r\prime}_i$ represents the position vector of the last segment at the ith adjacent point. For simplicity, we use the values 6 and 1/6 of a simple cubic lattice for the number of nearest neighbors, NNN, and the multiply constant, w, respectively.

This recursion relation can be approximated by a differential equation. Expanding the Green's function on the left hand side of Eqn. 8.6 into a Taylor series for $\vec{r}$ gives

$$G(\vec{r}, \vec{r}_0; N+1) = a^2 \nabla^2 G(\vec{r}, \vec{r}_0; N) \; + \; 6\, G(\vec{r}, \vec{r}_0; N) \tag{8.7}$$

for $N >> 1$. $G(\vec{r}, \vec{r}_0; N)$ varies slowly as a function of N. Now, Eqn. 8.7 can be rewritten as

$$\frac{\partial G(\vec{r}, \vec{r}_0; N)}{\partial N} = \frac{a^2}{6} \nabla^2 G(\vec{r}, \vec{r}_0; N) \tag{8.8}$$

and is idential to the diffusion equation. Its solution with the appropriate boundary and initial conditions gives the Green's function G. Since we are only interested in those random walk polymers that stay within the sphere, the absorbing boundary condition

$$G(\vec{r} = \vec{R}, \vec{r}_0; N) = 0 \tag{8.9}$$

is used. Moreover, the initial condition

$$G(\vec{r}, \vec{r}_0; 0) = \delta(\vec{r} - \vec{r}_0) \tag{8.10}$$

states where a polymer starts its random walk.

Normalization of G over the whole volume is required to express the probability density function, $P(\vec{r}, \vec{r}_0; N)$, for a chain of length Na that is confined in the sphere and whose ends are located at $\vec{r}$ and $\vec{r}_0$, respectively,

$$P(\vec{r}, \vec{r}_0; N) = \frac{G(\vec{r}, \vec{r}_0; N)}{\int G(\vec{r}, \vec{r}_0; N)\; d^3\vec{r}}. \tag{8.11}$$

After solving the diffusion equation with the corresponding boundary and initial conditions and normalizing the solution, the probability density function can be expressed as

$$P(\vec{r}, \vec{r}_0; Na) = \left[\sum_{l=0}^{\infty} \sum_{s=1}^{\infty} \frac{2l+1}{4\pi} \frac{J_{l+\frac{1}{2}}(\mu_{ls} \frac{r_0}{R})}{[J_{l+\frac{3}{2}}(\mu_{ls})]^2} \frac{J_{l+\frac{1}{2}}(\mu_{ls} \frac{r}{R})}{\sqrt{r}} P_l(\vec{r}.\vec{r}_0)\, e^{(-\frac{a^2}{6}(\frac{\mu_{ls}}{R})^2 N)} \right] \times$$
$$\left[\sum_{s=1}^{\infty} \sqrt{\frac{2}{\sqrt{R^5}\pi}} \frac{-\mu_{0s}\cos(\mu_{0s}) + \sin(\mu_{0s})}{\mu_{0s}^{5/2}} \frac{J_{1/2}\left(\mu_{0s} \frac{r_0}{R}\right)}{[J_{3/2}(\mu_{0s}]^2}\, e^{(-\frac{a^2}{6}(\frac{\mu_{0s}}{R})^2 N)} \right]^{-1}. \tag{8.12}$$

8.5.2 Strains

The subsections 8.5.2, 8.5.3 and 8.5.4 on the experimental details of the fixed and live cell measurements are kindly provided by Susannah Gordon-Messer, who also performed the experiments described in the following [354]. The yeast strains used in this study can be found in table X. All strains used were variants of YDB276 which has lacO(256)-LEU2 inserted 1 kb proximal to HML, LacI-GFP-(KAN), and SPC29-RFP-(kan::Ca-URA3-MX). YGM024 (yku80Δ esc1Δ) was created by deleting YKU80 using a BamHI/SalI restriction fragment from pJH1729 (yku80::URA3) and by delting ESC1 using transformation of a PCR-amplified fragment obtained from genomic DNA of the Research Genetics strain collection. For the 12kb deletion, a SalI/EcoR1 restriction fragment from pJH203 was used. This deleted the region between His4 and Leu2 (ChrIII: 67824-90765) using URA3 to obtain YGM014 (his4-leu2::URA3) [360].

Table 8.1: S. cerevisiae strains used in this study. Table adapted from [354].

Strain Number	Genotype
YDB276	Ho HMLα HMLprox::lacO(256)-LEU2 MATα HMRα-B ade1 ade3::GAL-HO leu2 trp1:hisG ura3-52 Spc29-RFP-(kan::Ca-URA3-MX) HIS3::URA3pro::lacI-GFP-(KAN)
YGM014	Same as YGM006 except his4-leu2::URA3
YGM024	Same as YGM006 except yku80::URA3 esc1::KAN

8.5.3 Fixed Cell Data Collection

The subsections 8.5.2, 8.5.3 and 8.5.4 on the experimental details of the fixed and live cell measurements are kindly provided by Susannah Gordon-Messer, who also performed the experiments described in the following [354].

Cell Preparation – Fixed Cells

Cells were grown to log phase in YEPD media that was filtered prior to use. Cells were fixed by addition of paraformaldehyde at a 2% final concentration for 10 minutes at room temperature. Following this, cells were pelleted and washed in 0.1M potassium phosphate, pH 6.6 for 10 minutes a room temperature. Cells were pelleted a second time and resuspended in 35-50 μl of 0.1M postassium phosphate, pH 6.6 and stored at 4°Cbefore imaging at room temperature.

Imaging – Deltavision

Fixed cell images were acquired on a DeltaVision Core deconvolution microscope (Applied Precision) equipped with a CoolSnap HQ2 CCD camera. 16 to 20 Z-sections were acquired at 0.2-μm steps using a 100x, 1.3 NA Olympus U-PlanApo objective with 1x1 binning. Each cell was imaged individually and cell size was kept consistent among strains by ensuring that each cell fit within a fixed window size of 128 x 128. Cells were imaged using a GFP exposure time of 0.2s and an RFP exposure time of 0.3s [216].

Image Processing

Images were processed using a series of customized MATLAB programs. The Deltavision (*.dv) files and Olympus (*.nd) files were converted in *.tiff files for each wavelength and slice. The full 3D stack was converted to a 2D projection in order to determine the x and y coordinates. Once these were determined, a Gaussian curve was fit to the intensity of the point throughout all z slice images. This allowed us to determine the z coordinate with sub-pixel resolution. The accuracy of the spot finding was manually checked by using a program in which an "x" was displayed on the original image in the location specified by the coordinates. Once coordinates were successfully found, 3D distances were calculated and histogrammed using a bin size of 200nm.

8.5.4 Live Cell Data Collection

The subsections 8.5.2, 8.5.3 and 8.5.4 on the experimental details of the fixed and live cell measurements are kindly provided by Susannah Gordon-Messer, who also performed the experiments described in the following [354].

Cell Preparation and Cell Imaging Set Up.

Cells were grown overnight in synthetic + dextrose (SD) medium (yeastminimal medium glucose) at 30℃in 5 mL cultures on a rotary shaker. SD medium with 2% agarose was liquified in a microwave and poured into a glass-bottomed Petri dish (Bioptechs Delta-T) to a depth of 5 mm. The bottom of the dish is coated with indium tin oxide, allowing heat to be generated by the application of current. Solidified agarose pads were removed from the dishes; 10 μL of cells were placed in the center of the dish, and covered with the pads. Dishes were placed on the OMX microscope via a custom-built adaptor connected to a power source (Bioptechs, Inc) that provided a current to keep the dish at 30℃. The objective used for imaging was kept constantly heated to 30℃ by a thermal ribbon and microcontroller (Minco, Inc) using a custom-built copper collar, 3 mm thick, coupling the thermal ribbon to the objective.

Image Acquisition and Processing

This subsubsection is adapted from [353]. The OMX microscope uses two simultaneous cameras to take images of two different wavelengths. In order to ensure that the wavelengths could be properly aligned in the final image, it was necessary to take images of fluorescent beads at the beginning of each day's experiments. For bead imaging, 100 nm red-emitting fluorescent latex beads (Molecular Probes, Inc) were diluted 1:10;000 in ethanol. A 1 μL drop of diluted beads was placed in the center of a plasma-cleaned coverslip, allowed to spread out and air-dry, then mounted on a slide with 5 μL of glycerol. Laser light at 488 nm and 532 nm was used for excitation of yeast cells and fluorescent latex beads, respectively.

Fluorescent Image Acquisition

Images were acquired on Andor iXon EMCCD cameras set at their highest gain level. Dark current was calculated by averaging 512 frames taken with no excitation light, and subtracted from images before further processing. Images were taken using a 10ms exposure for both 488nm and 532nm lasers. See SI Text of [353] for a complete description

of the OMX microscope. Images were taken using two simultaneous cameras, one of each wavelength. Each cell was imaged individually and 16 to 30 Z-sections were acquired at 0.2-μm steps. Cells were imaged at the following intervals: 1 second for 150 seconds, 5 seconds for 10 minutes, 150 seconds for 1 hour.

Image Processing

Resulting images were wavelength aligned, cropped and then run through a denoising program. For full details, the reader is referred to Refs. [353, 354]. After denoising was complete, images were deconvolved and processed using a series of customized MATLAB programs

Chapter 9

E. coli Chromosome Organization in Free Space

The Impact of A Specific Underlying Topology

References

The results presented in this chapter are published as and adapted from

- P. Reiss*, M. Fritsche* and D.W. Heermann (2011), *Looped star polymers show conformational transition from spherical to flat toroidal shapes*, under peer-review.

We thank Suckjoon Jun for fruitful discussions.

*These authors contributed equally

Chapter Summary

Inspired by the topological organization of the circular Escherichia coli chromosome, which is compacted by separate domains, we study a polymer architecture consisting of a central ring to which either looped or linear side chains are grafted. A shape change from a spherical to a toroidal organization takes place as soon as the inner ring becomes large enough for the attached arms to fit within its circumference. Building up a torus, the system flattens depending on the effective bending rigidity of the chain induced by entropic repulsion of the attached loops and, to a lesser extent, linear arms. Our results suggest that the natural formation of a toroidal structure with a decreased amount of writhe induced by a specific underlying topology could be one driving force, among others, that nature exploits to ensure proper packaging of the genetic material within a rod-shaped, bacterial envelope.

9.1 Introduction

The underlying topological architecture and the resulting shape of biopolymers have important implications for their functionality [31,32]. This can be seen in case of Escherichia coli (E. coli) where the coordination of gene expression is linked to the spatial organization of genes within the bacterial envelope called nucleoid [104,361]. It was recently shown that the E. coli chromosome cannot be modeled by an unstructured polymer, but measurements of the locus distributions reveal that it is precisely organized and compartmentalized [5]. Since it is usually the most likely polymer conformation that is relevant for biological processes [31, 32] physical modeling of polymers can guide our intuition by providing an understanding of the physics imposed by the underlying macromolecular architecture.

Thus, inspired by the topological organization of the E. coli chromosome, we study a polymer model consisting of an inner "core" ring to which either loops, representing the chromosomal domains, or linear arms of identical length are attached. Whereas the properties of star polymers, consisting of several linear polymers of identical length that are linked together by one of their ends to a common center, and of ring polymers have received considerable attention both theoretically [56, 165, 234, 362–364] and experimentally [365–367] the effects induced by the specific chromosome architecture underlying the E.coli chromosome have not been investigated yet.

Thus, we study the impact of a certain topology on the polymer's conformational state revealing insight into chromosome conformation in living cells with its potential role in the regulation of gene expression. Even though nature not only imposes intrinsic architectures on biopolymers, but also geometrical constraints induced by confinement, this work focuses on polymers in free space. In particular, we are able to elucidate the role that entropic repulsion of polymer rings plays in shaping the spatial chromosome arrangement. To this end, we investigate overall shape changes by studying the radius of gyration tensor and assess further global conformational properties such as chain rigidity and chain self-entanglement.

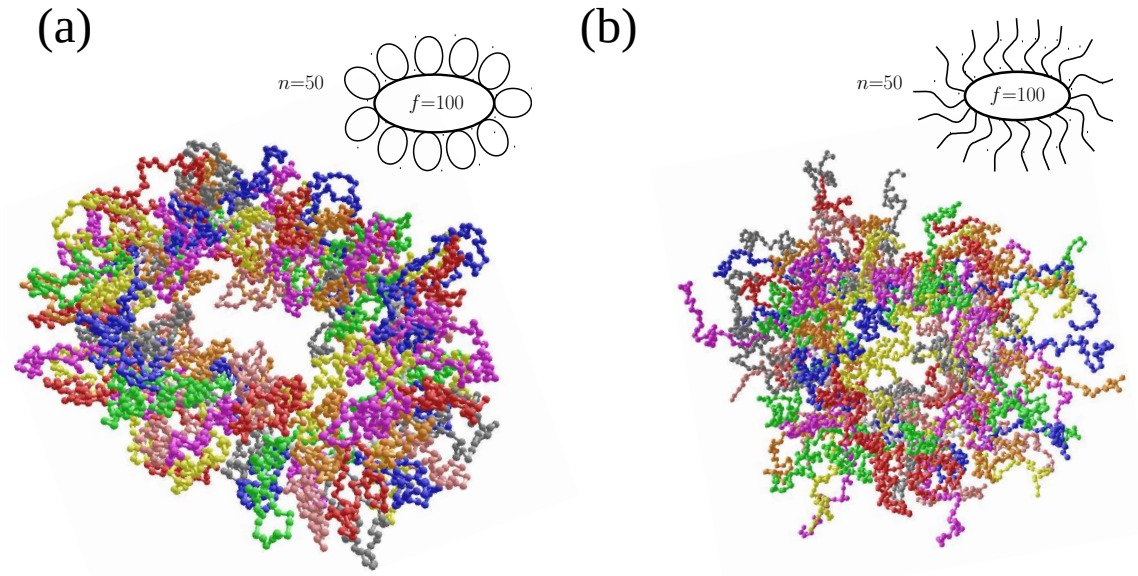


Figure 9.1: (a) A torus emerges for inner ring size $f = 100$ and looped arms of size $n = 50$, (b) while its emergence is less pronounced in the case of linear arms. This can be understood by noting that intra-side-chain excluded volume interactions are stronger between looped arms than between linear ones resulting in a higher effective bending rigidity. While a completely rigid core ring would be circular, thermal fluctuations induce a more planer, elongated elliptical shape in agreement with configurational changes found for semiflexible ring polymers in free space [234]. Image adapted from [368].

9.2 Modeling

The polymer models studied in this work are illustrated in Fig. 9.1 and consist of a central ring built up by f monomers to each of which is attached either a loop or a linear arm of size n. Since all loops or linear arms have the same size, the entire polymer is made of a total number of $N = (n + 1)f$ monomers.

To investigate the above defined models we employ the bond-fluctuation method (BFM) [166], which has been applied successfully to model the static and dynamical properties of polymer systems in several investigations [146]. Additional details on the BFM are given in chapter 3.

In order to generate thermodynamically equilibrated polymer conformations we use the Metropolis Monte Carlo method [146]. Since subsequent conformations in the Markov chain created by the Monte Carlo algorithm are highly correlated, one has to perform a certain number of Monte Carlo steps t_{MC} to obtain two independent conformations.

Thus, we determine, for each set of parameters (ring size f, loop size n), the autocorrelation function $C(t)$ of (i) the whole polymer's gyration radius as well as the autocorrelation function of the largest eigenvalue λ_3 of the gyration tensor for (ii) the entire polymer and for (iii) the central ring only. The integrated autocorrelation time τ_{int} is computed for the "slowest mode" of the three shape descriptors by applying the windowing method introduced by Sokal [164], which is explained in chapter 3. We consider two subsequent conformations as uncorrelated after $5\tau_{\mathrm{int}}$ Monte Carlo steps.

Fig. 9.2 illustrates the decay of the autocorrelation function $C(t_{\mathrm{MC}})$ as a function of the performed MC steps t_{MC} as well as the respective integrated autocorrelation times τ_{int}. The larger the system size the slower is the decay of the autocorrelation function. Looped

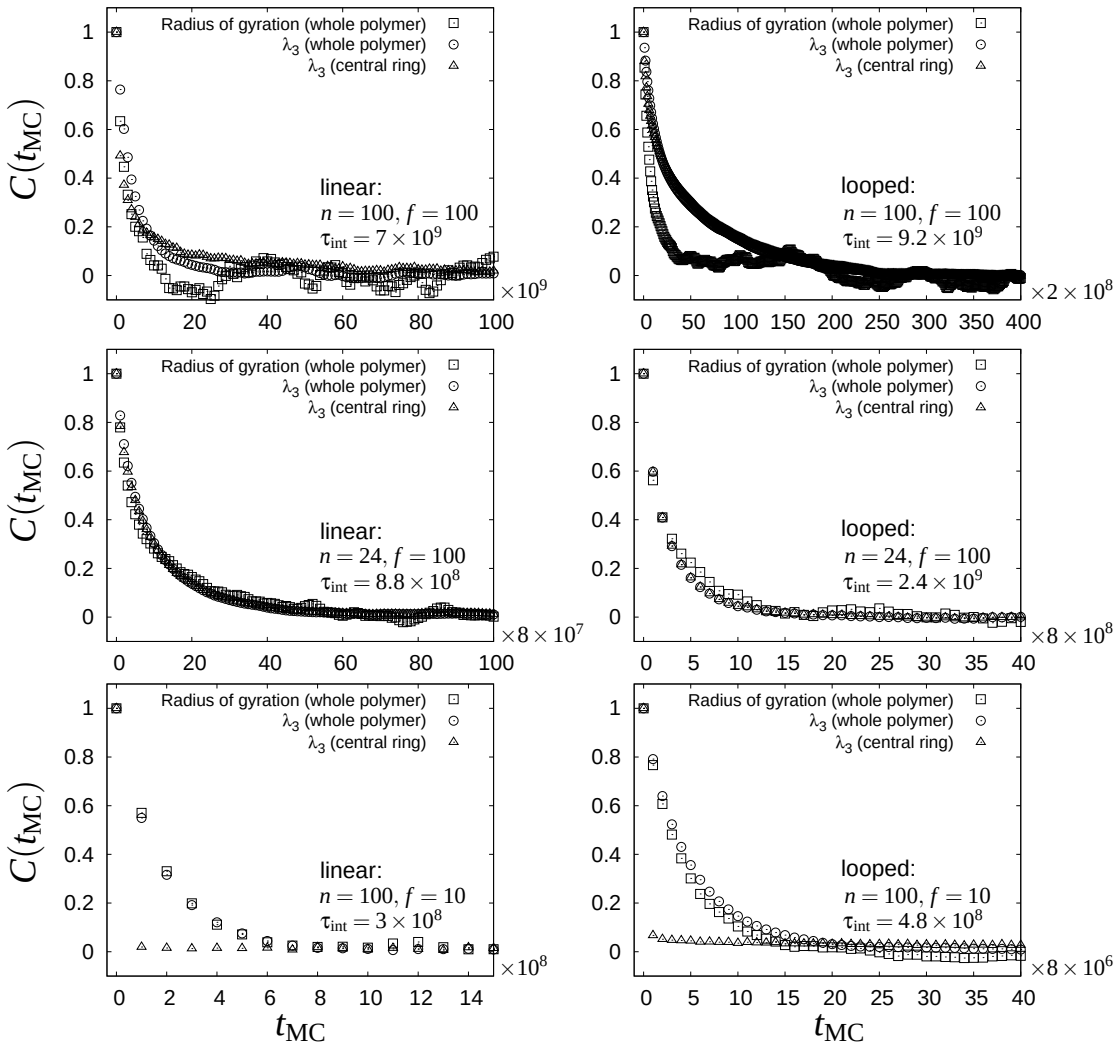


Figure 9.2: Since subsequently created conformations are highly correlated, we determine, for each set of parameters (ring size f, loop size n), the autocorrelation function of (i) the whole polymer's gyration radius as well as the autocorrelation function of the largest eigenvalue of the gyration tensor for (ii) the entire polymer and for (iii) the central ring only. The integrated autocorrelation time τ_{int} is computed for the "slowest mode" of the three shape descriptors by applying the windowing procedure introduced by Sokal [164]. Image adapted from [368].

side chains induce higher integrated autocorrelation times than linear ones for the same set of paramters (f and n), $\tau_{\text{int}}^{\text{linear}} > \tau_{\text{int}}^{\text{looped}}$. By creating 4000-40 000 independent configurations we can study polymers with parameters $f \in \mathcal{F} := \{2, 6, 8, 10, 16, 20, 30, 40, 60, 80, 100\}$ and $n \in \mathcal{N} := \{24, 50, 100, 200, 400\}$, resulting in polymer lengths in the range from $N = 50$ to $N = 40100$.

9.3 Results and Discussion

9.3.1 Shape Change from a Spherical to a Toroidal Organization

A measure characterizing the shape of a polymer is the gyration tensor [56, 151, 165], defined by

$$S_{mn} = \frac{1}{N} \sum_{i=1}^{N} r_m^{(i)} r_n^{(i)}. \tag{9.1}$$

N is the total number of monomers, $\mathbf{r}^{(i)}$ is the coordinate vector of the ith monomer and the subindex denotes its Cartesian components. The eigenvalues $\lambda_1^2 \leq \lambda_2^2 \leq \lambda_3^2$ give the squared lengths of the principal axes of gyration, while their ratios indicate deviations from a sphere-like shape, having a value of unity for a sphere.

Fig. 9.3 shows $\langle\lambda_3^2\rangle/\langle\lambda_2^2\rangle$ and $\langle\lambda_2^2\rangle/\langle\lambda_1^2\rangle$ for the entire polymer and for the central ring only (inset) as a function of the central ring size f in order to understand the contributions from the side chains. In the limit of small inner ring size f, we recover an almost spherical shape with the attached looped or linear arms isotropically stretching out, as reflected by $\langle\lambda_3^2\rangle/\langle\lambda_2^2\rangle = \langle\lambda_3^2\rangle/\langle\lambda_1^2\rangle \approx 1$.

The non-trivial interplay between the size f of the central loop and the size n of the attached looped or linear arms induces a shape change from a spherical and to a toroidal organization for increasing inner loop size f. While the two large axes, $\langle\lambda_3^2\rangle$ and $\langle\lambda_2^2\rangle$, assume equal lengths, the third axis, $\langle\lambda_1^2\rangle$, decreases indicating the emergence of a flat toroidal organization. Here, the impact of looped in contrast to linear arms becomes apparent, since the flattening of the overall polymer shape is more pronounced for looped arms. Comparing the shape changes of the entire polymer with those of the central ring only, we find the inner ring's structural crossover to be more pronounced, whereas the contribution of the sides chains smoothes out this effect with respect to the overall polymer shape.

Fig. 9.1 exemplifies such a spherical and toroidal polymer conformation. The overall shape change can be understood by noting that a torus can emerge only if the central ring is large enough to allow for the attached looped or linear arms to fit inside of its circumference. Thus, the limit of small central ring size leads to a high resemblance with star polymers, for which the known scaling law of the mean squared radius of gyration is recovered, $\langle R_{\text{gyr}}\rangle \sim n^{2\nu}$ with $\nu = 0.589$ [369, 370] (data not shown). With increasing ring size f, the spherical shape vanishes and toroidal conformations dominate.

The overall polymer shape is determined by the inner ring's effective bending rigidity induced by the interplay between the size of the inner ring and the attached arms. This can be understood by noting that ring polymers or polymer loops do not strongly intermingle but entropically repel each other both in free space and even stronger in confinement [24, 56, 165]. In fact, Figure 9.4 confirms the stronger entropic repulsion between adjacent side loops by illustrating the polymer's resemblance with a pearl necklace, where the ellipsoids spanned by the eigenvalues of the gyration tensor of each single attached loop appear well separated.

In order to further investigate the attached loop's preference of unmixing, we determine the rotation angle between two subsequent vectors $\mathbf{v}_i$ and $\mathbf{v}_{i+1}$ connecting the centers of mass of the i-th and the $(i+1)$-th side loop with their attachment site on the central ring, therewith gaining insight into their arrangement with respect to each other. Let $\mathbf{b}_i$ be the bond vector connecting subsequent attachment sites along the central ring corresponding to the i-th and the $(i+1)$-th loop. The vectors $\mathbf{v}_i$ and $\mathbf{v}_{i+1}$ are then projected to the

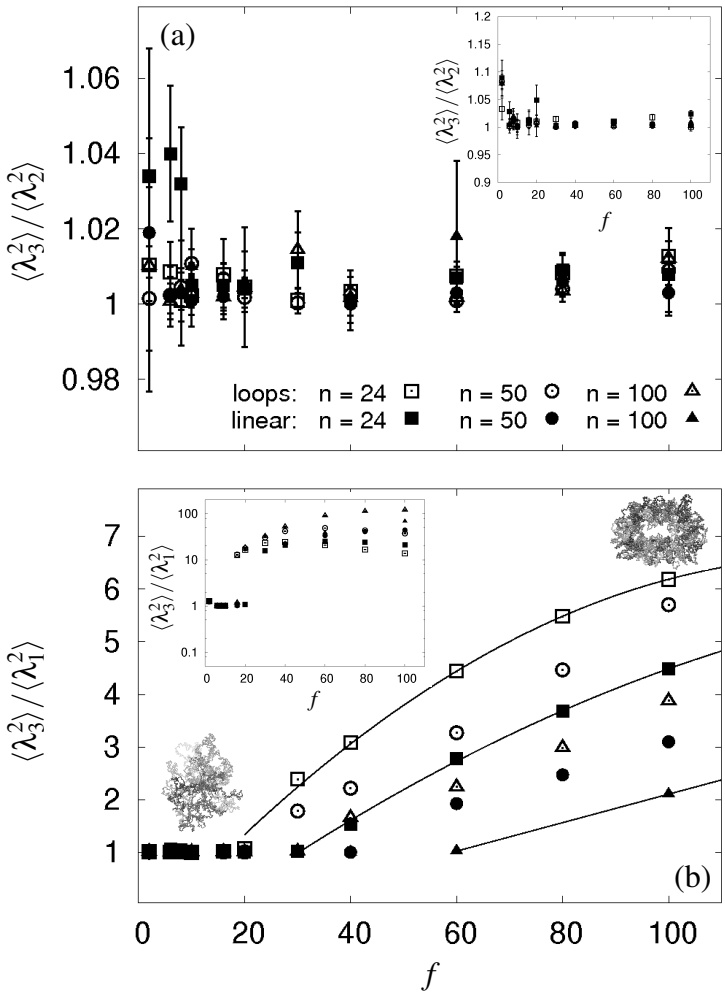


Figure 9.3: The ratios (a) $\langle\lambda_3^2\rangle/\langle\lambda_2^2\rangle$ and (b) $\langle\lambda_3^2\rangle/\langle\lambda_1^2\rangle$ are displayed for the entire polymer as well as for the inner ring only (inset). In (b) the errors are smaller than the symbol size. In the limit of small inner ring sizes, spherical polymer conformations dominate, while a toroidal shape emerges for increasing size of the central ring. Image adapted from [368].

plane perpendicular to $\mathbf{b}_i$ according to

$$\mathbf{V}_i = \mathbf{v}_i - \frac{(\mathbf{v}_i \cdot \mathbf{b})\mathbf{b}}{\mathbf{b}^2}. \tag{9.2}$$

Between $\mathbf{V}_i$ and $\mathbf{V}_{i+1}$ an angle φ_i can be computed describing the rotation of the loops with respect to the central ring allowing for the definition of a mean rotation angle defined as

$$\langle|\phi|\rangle = \frac{1}{f}\sum_{i=0}^{f-1}|\phi_i|. \tag{9.3}$$

Figure 9.5 shows the mean torsion angle for the polymer model with linear as well as looped side chains. Notably, for a given side chain length n, there exists a central ring size f for which the mean rotation angle is minimal. In the case of an overall spherical conformation ($n > f$) the sides chain stretch out isotropically and thus assume large angles with respect to each other. With increasing central ring size f the attached side chains start to fit

Figure 9.4: Illustration of the "pear necklace"-like alignement of the ellipsoids representing the gyration tensor of the loops. The looped side chains do not intermingle but entropically repel each other. Image adapted from [368].

more and more within the central ring making a rotated arrangement of adjacent side chains with a large rotation angle unnecessary. A further increase of the inner ring size eventually leads again to an increase in the rotation angles between adjacent side chains due to their tendency of demixing once the available space within the central ring is large enough. Since the entropic repulsion between looped side chains is stronger than between linear ones, the rotation angles between adjacent looped arms is larger.

In the limit of infinite core ring size $f \to \infty$, a polymer topology comparable to the bottle-brush structure emerges. Bottle-brush polymers are highly-branched macromolecules where linear side chains are bonded to a linear polymeric backbone in a dense manner [372]. Overcrowding of the side chains in the bottle-brush polymer leads to rather shape-persistent, stiff, cylindrical structures solely due to intra-molecular excluded volume interactions [373].

In this work, we are interested in relating the impact of a specific underlying topology (distinct loops around a central core) to the packaging of the genetic material in E. coli [5, 24]. Thus, we study the crossover behavior of circular polymer chains of finite lengths rather than the "unbiological" limit of $f \to \infty$ leading to either star or bottle-brush scaling behavior [157, 167, 248].

The overall polymer shape is determined by the inner ring's effective bending rigidity induced by the attached loops and linear arms, respectively. This can be understood by noting that ring polymers or polymer loops do not strongly intermingle but entropically repel each other both in free space and even stronger in confinement [24, 56, 165]. Consequently, the interplay between the size of the inner ring and the attached arms results in a change of the central ring's effective persistence length, $l_{\mathrm{p}}^{\mathrm{eff}}$, which is taken as the basic characteristics of polymer flexibility [137].

In order to determine the central ring's effective persistence length $l_{\mathrm{p}}^{\mathrm{eff}}$, we address the spatial correlations of two tangent vectors $\mathbf{u}(i)$ and $\mathbf{u}(j)$ separated by a distance $|i-j| \in [0, f/2]$ along the inner ring backbone

$$C(i,j) = \langle \mathbf{u}(i)\mathbf{u}(j) \rangle. \tag{9.4}$$

A negative value of $C(i,j)$ indicates a reflection in the orientation of the tangent vector due to the constraints imposed by the underlying ring architecture. We therefore represent $C(i,j)$ as $Ae^{-|i-j|/l_{\mathrm{p}}^{\mathrm{eff}}} \cos(|i-j|/B)$ with A, B and $l_{\mathrm{p}}^{\mathrm{eff}}$ as fitting parameters [230]. Fig. 9.2 exemplifies the mean tangent-tangent correlation $C(i,j)$ and shows that this form captures

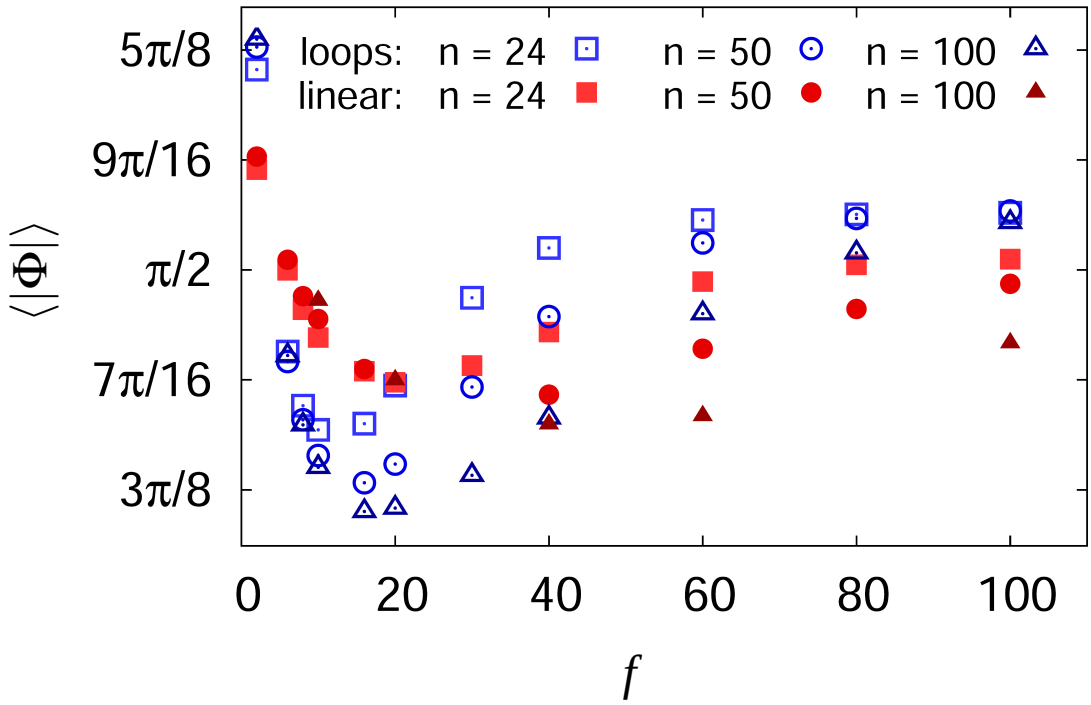


Figure 9.5: Mean rotation angle $\langle|\phi|\rangle$ between adjacent sides chains. In the case of an overall spherical conformation ($n > f$) the sides chain stretch out isotropically and thus assume large angles with respect to each other. With increasing central ring size f the attached side chains start to fit more and more within the central ring making a rotated arrangement of adjacent side chains with a large rotation angle unnecessary. A further increase of the inner ring size eventually leads again to an increase in the rotation angles between adjacent side chains due to their tendency of demixing once the available space within the central ring is large enough. Image adapted from [371].

the properties of $C(i,j)$. We can then extract the effective persistence length $l_{\mathrm{p}}^{\mathrm{eff}}$ from the fits, whose variation characterizes the rigidity of the inner ring for looped and linear arms, respectively.

Fig. 9.6 shows the dependence of the central ring's effective persistence length on its size f. At constant f, the longer the attached arms n are (either looped or linear) the larger is the effective inner ring rigidity due to the arms' increased entropic repulsion. However, comparing the impact of a domain organization by loops to an inner ring with attached linear arms, we find a higher bending stiffness of the inner core in case of looped arms. The reason for this is that polymer loops entropically repel each other stronger than linear ones and thus induce a higher bending rigidity alone the inner "core" ring [24, 56, 165].

However, since a recent study has shown that polymers with bottle-brush architecture need to be very large to reach the asymptotic limit where they satisfy self-avoiding walk statistics and where a well-defined persistence length can be extracted [157,248] we present another less controversial measure for polymer flexibility here, too. In fact, the inner ring's bending rigidity can be assessed by computing the standard deviation $\sigma_{\mathrm{bending}}$ of its normalized angular sum $\frac{1}{f}\sum_{i=0}^{f-1}\mathbf{b}_i\mathbf{b}_{i+1}$, where $\mathbf{b}_i$ is the normalized bond vector with periodic boundary conditions $\mathbf{b}_f = \mathbf{b}_0$. The larger the dispersion of the angular sum the more flexible is the inner ring.

Fig. 9.7 shows the inner ring's bending rigidity as a function of its size f for various side chain sizes n. At constant f, the longer the attached arms n are (either looped or

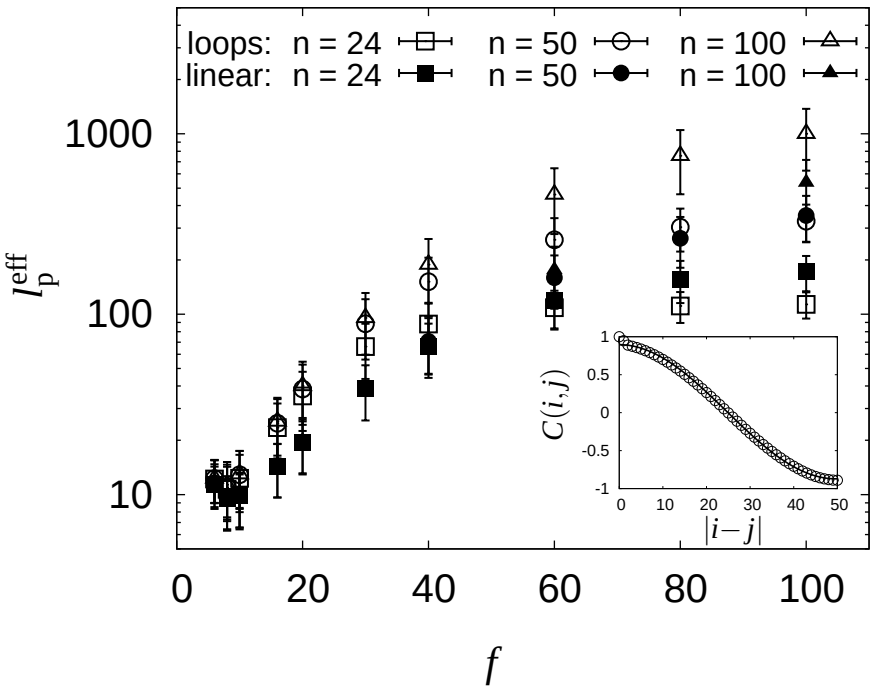


Figure 9.6: The effective persistence length $l_{\mathrm{p}}^{\mathrm{eff}}$ is shown as a function of f characterizing the rigidity of the inner ring. comparing the impact of a domain organization by loops to an inner ring with attached linear arms, we find a higher bending rigidity of the inner core in the case of looped arms. The reason for this is that polymer loops entropically repel each other stronger than linear ones and thus induce a higher bending rigidity along the inner "core" ring [24, 56, 165]. Inset: In order to determine $l_{\mathrm{p}}^{\mathrm{eff}}$, the spatial correlations $C(i,j)$ of two tangent vectors $\mathbf{u}(i)$ and $\mathbf{u}(j)$ separated by a distance $|i-j| \in [0, f/2]$ along the polymer backbone are fitted to $Ae^{-|i-j|/l_{\mathrm{p}}^{\mathrm{eff}}} \cos(|i-j|/B)$ with A, B and $l_{\mathrm{p}}^{\mathrm{eff}}$ as fit parameters. The fit is shown as solid line, exemplified for a polymer of "core" ring size $f = 100$ and attached loop size $n = 200$. Image adapted from [368].

linear) the larger is the effective inner ring rigidity due to the arms' increased entropic repulsion. However, comparing the impact of a domain organization by loops to an inner ring with attached linear arms, we find a higher bending stiffness of the inner core in case of looped arms. The reason for this is that polymer loops entropically repel each other stronger than linear ones and thus induce a higher bending rigidity along the inner "core" ring [24, 56, 165].

Noteably, $\sigma_{\mathrm{bending}}$ shows a non-monotonic dependence on f. Thus, for a given attached arm size n there exists an "optimal" inner ring size $f_{\mathrm{opt}}^{\mathrm{bending}}$ for which the inner ring rigidity is maximal leading to exclusively planar shapes ranging from total oblate to comparatively prolate in agreement with ring polymers in the stiff regime [234]. In contrast, towards the flexible regime, i.e. for large f and small n, the emerging conformations become three-dimensional and crumpled.

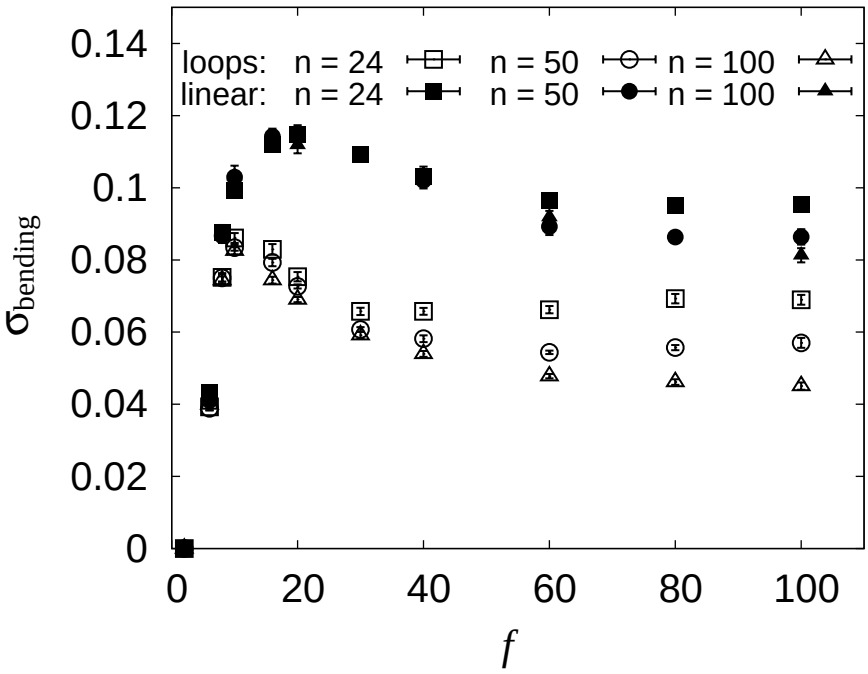


Figure 9.7: The inner ring's bending rigidity can be assessed by computing the standard deviation σ_{bending} of its normalized angular sum $\frac{1}{f}\sum_{i=0}^{f-1}\mathbf{b}_i\mathbf{b}_{i+1}$, where $\mathbf{b}_i$ is the normalized bond vector with periodic boundary conditions $\mathbf{b}_f = \mathbf{b}_0$. The larger the dispersion of the angular sum the more flexible is the inner ring. We find a higher bending rigidity of the inner core in the case of looped arms. The reason for this is that polymer loops entropically repel each other stronger than linear ones and thus induce a higher bending rigidity along the inner "core" ring [24, 56, 165]. Noteably, σ_{bending} shows a non-monotonic dependence on f. For a given attached arm size n there exists an "optimal" inner ring size $f_{\text{opt}}^{\text{bending}}$ for which the inner ring rigidity is maximal leading to exclusively planar shapes ranging from total oblate to comparatively prolate in agreement with ring polymers in the stiff regime [234]. Image adapted from [368].

9.3.2 Polymers with Looped Side Chains Display Lower Tendency for Chain Overcrossings

A measure for the "crumpledness" of the central ring, and ultimately the overall polymer shape, is the writhe Wr indicating the complexity of self-entanglement [374,375]. Projecting a three-dimensional ring trajectory into a plane defined by a normal vector $\mathbf{n}$ results in a two-dimensional curve which may exhibit crossings. Counting these crossings with ± 1 according to their handedness and averaging the number of crossings over all angular perspectives given by all possible normal vectors $\mathbf{n}$ defines the writhe Wr of the central ring. As only the orientation in which a trajectory is traced decides if the writhe is positive or negative, any writhe distribution is symmetric about the origin with the mean writhe being equal to zero. Insights are therefore gained when measuring the mean absolute writhe $|Wr|$ following [256].

Fig. 9.8 displays the absolute writhe $|Wr|$, which increases for increasing central ring size f at constant loop size n reflecting the fact that an increasing inner ring flexibility allows for a larger number of chain crossings. In contrast, in the limit of small central ring size and large attached loops, the loops' tendency of unmixing results in a high bending rigidity as shown above and thus low chain self-entanglement. In view of the non-

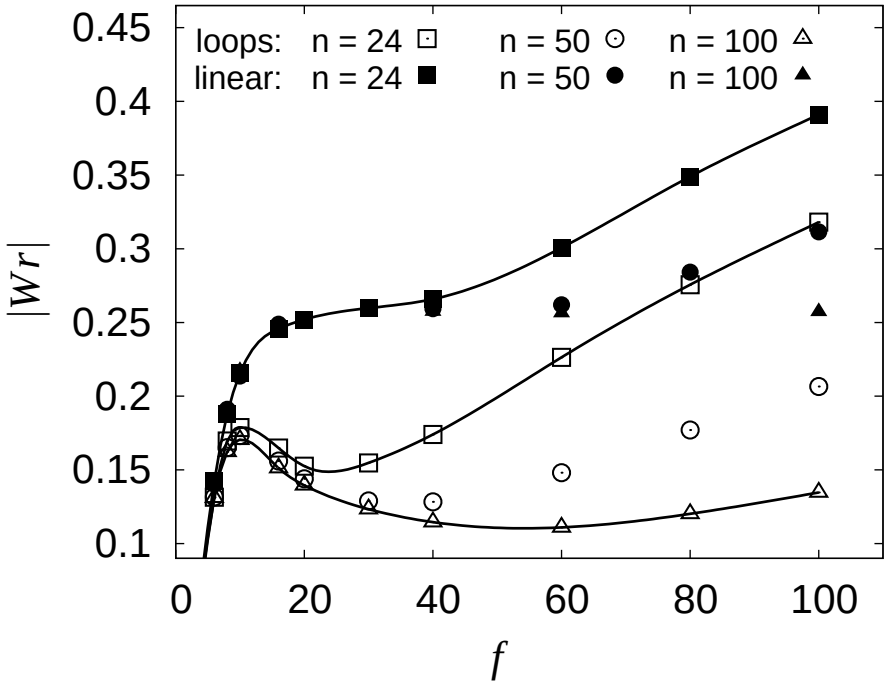


Figure 9.8: A measure for the "crumpledness" of the studied polymer model is the absolute writhe $|Wr|$ of the central ring indicating its complexity of self-entanglement. For a given attached arm size n there exists an "optimal" inner ring size f_{opt} for which the mean absolute writhe $|Wr|$ is minimal leading to conformations of minimal self-entanglement. According to the higher effective persistence length recovered in case of looped arms, we find a lower tendency for chain overcrossings compared to central rings with linear arms. The error bars are smaller than the symbol size. Image adapted from [368].

monotonic inner ring rigidity, there exists (for a given attached arm size n) an "optimal" inner ring size $f_{\text{opt}}^{\text{Wr}}$ for which the mean absolute writhe $|Wr|$ is minimal leading to conformations of minimal self-entanglement. In fact, nature might exploit the finite window of ratios between side-loop size and backbone radius of gyration (of order unity) resulting in optimal "structural" conditions to drive spatial chromosome organization [33, 57]. According to the higher bending rigidity recovered in the case of looped arms, we find a lower tendency for chain overcrossings compared to central rings with linear arms.

9.4 Conclusions

In this work, we have shown that a specific underlying polymer topology, similar to the structure found for the E. coli chromosome, leads to a flattened toroidal polymer organization. The stiffening of the central "core" ring as a consequence of the entropic repulsion between attached side chains ensures stable, less crumpled conformations. Recent experiments have found that loci along the circular E. coli chromosome show a precision of positioning of better than 10% of the cell length, while the precision of interlocus distance of genomically-proximate loci was found to be better than 4% [5].

Bridging the gap between the experimental findings and our results in free space, we suggest that the organization of the bacterial chromosome into distinct loops around a

central chromosome ”core ring“ might be one driving force, among others, to facilitate the proper packaging of the genetic material in rod-shaped confinement by inducing toroidal shape changes, a stiffening of the chromosomal fiber and reduced self-entanglement.

However, the E. coli chromosome is not located in free space but it is strongly confined within the bacterial nucleoid. Thus, in the next chapter, we investigate the interplay between the specific topology underlying the E. coli chromosome (namely its packaging due to chromosomal domains) and the geometrical constraints imposed by the envelope of the confining cavity. In particular, a mechanism responsible for the formation of the chromosomal domains is introduced and discussed.

Chapter 10

E. coli Chromosome Packaging in the Nucleoid

Transcription Factor Induced Domain Formation

References

The results presented in this chapter are published as and adapted from

- M. Fritsche, S. Li, D.W. Heermann and P.A. Wiggins (2011), *Transcription factor induced DNA domain formation provides a natural framework for understanding the physical structure of the E. coli chromosome*, under peer-review.

The authors would like to thank Jané Kondev, Suckjoon Jun, Jonathan Machta, Lindsay Shopland and Benoît Knecht for their comments and suggestions.

Chapter Summary

This chapter addresses a fundamental question of today's biophysical research: What physical mechanism leads to organization of a highly-condensed, and confined circular chromosome?

Measurement of position fluctuations of single loci by fluorescently labeling genetic sites along the circular chromosome reveal that loci in the body of the nucleoid show a precision of positioning within the cell of better than 10% of the cell length. Starting from the experimental results, we first of all investigate whether the interplay between confinement and the specific topology underlying the E. coli circular chromosome, namely its organization in chromosomal domains discussed in the previous chapter, is able to overcome the chromosome's propensity to mix and self-organizes into a nucleoid-filament type of structure (leading to the experimentally observed high precision of subcellular positioning).

Having shown that the entropic repulsion of chromosomal domains and the pressure exerted by the confining cavity is indeed able to create the right physical conditions for E. coli chromosome packaging, we establish one possible mechanism for actually forming chromosomal domains. To this end, more and more evidence is accumulating about the E. coli chromosome being organized into a complex 3D network constrained by long- and short-range interactions. We propose that it is the structure of the E. coli gene regulatory network that organizes the DNA chain into several domains (chromosomal loops). This process is driven by imposing colocalization of transcription factors and their target genes.

Thus, we investigate the consequences of this assumption and find the circular chromosome to indeed self-organize into an effective nucleoid filament-type of structure. We are able to explain the experimentally found high precision of subnuclear positioning by our framework. Moreover, to reproduce the observed precise ordering of the chromosome, we estimate that the domain sizes are distributed between 10 and 700 kb, in agreement with the size of topological domains identified in the context of DNA supercoiling.

10.1 Introduction

The E. coli chromosome consists of 4.6 million base pairs and has a contour length of 1.6 mm [3,4]. E. coli cells are typically rod-shaped and are about 2-4 μm long and 0.5 μm in diameter [3]. Since the chromosome is much longer than the typical prokaryotic cell it must be compacted to fit inside, where it forms a DNA-protein complex called the nucleoid [2, 104]. The spatial organization of the chromosome is shaped by the interplay between compaction of the genetic material and its accessibility in order to ensure vital cell functions such as replication, segregation, gene expression and DNA repair. In this work, we investigate an intriguing biological and physical problem: What physical mechanisms lead to organization of highly-condensed, and confined circular chromosomes?

Paul Wiggins and coworkers previously measured and analyzed the position fluctuations of 15 single genetic loci (in G and early S phase), among which are the origin of replication *oriC* as well as two loci named "lac" and "C4" for convenience (for more details see [5]). Loci in the body of the nucleoid were found to show a precision of positioning of better than 10% of the cell length and the precision of interlocus distance of genomically

proximate loci was found to be better than 4%. Moreover, the linear relationship between positions of the genes on the chromosome and their spatial location within the cell was confirmed. While most studies of the localization of chromosomal loci in bacteria have focused on their position along the length of the cell, a recent study reports data concerning loci positioning across the width of the *E. coli* cell finding *ter*-borne loci localized at the nucleoid periphery [376].

By physico-chemical approaches, it has long been shown that the circular *E. coli* DNA molecule is organized into separate chromosomal loops or superhelical domains that are relaxed independently when DNA is cut [30, 104, 377–379]. One purpose of these topological domains appears to be the prevention of chromosome unraveling as a result of DNA damage since the loss of chromosomal supercoiling leads to cell death [380]. However, cross-links between random positions on the chromosome would also protect against unwinding [104]. Instead of that genomic neighbors appear to be cross-linked suggesting that the so formed domains might serve structure-function relationships with respect to gene expression, too [15, 19, 21, 22, 381, 382].

Experimental determinations of the average sizes of these domains differ between 10 kb [30] and 100 kb [110], report intermediate values [111], and even indicate an organization into a ring compacted by four macrodomains and two less structured regions [15, 109]. Thus, there already is evidence of a non-trivial nucleoid structure where smaller domains are organized within higher-order "super" structures.

Several drivers responsible for the observed reliable orientation and high level of organization of the *E. coli* chromosome have been suggested, including intranucleoid interactions such (i) macromolecular crowding [112], (ii) DNA supercoiling [30] or (iii) protein-DNA interactions [15, 16, 113] as well as explicit mechanisms of external positioning such as (iv) cellular confinement [5, 33] or (v) tethering of the chromosome [114, 115].

Increasing lines of evidence link the three-dimensional packaging of genes to the proper coordination of gene expression [15, 19, 21, 22, 381, 382]. Transcription factors (TFs) are the key controlling elements for appropriate gene expression in bacteria, where a functional network of regulatory interactions between TFs and target genes, which can themselves be TFs, is formed [21, 106]. Assuming the spatial nucleoid structure to influence transcription and vice versa then specific correlations are expected to arise in gene expression patterns [104]. In fact, expression patterns correlate at short (< 16 kb), medium (~ 100 kb) and long (~ 600 - 700 kb) distances in E. coli [104, 383, 384]. The short-range correlations might result from small elementary domains of the nucleoid, while the long-range correlations could result from higher scales of organization, i.e. from long-range interactions between regions of the chromosome that are genomically distant, but spatial proximate when the chromosome is packaged within the nucleoid [104].

Underpinning the evidence for "regulatory domains", a recent experimental study on the spatial organization of mRNA in E. coli shows that mRNAs display limited dispersion from their site of transcription [22]. The high localization of mRNA implies that chromosome architecture might act as a spatial organizer which compartmentalizes the cell interior such that dedicated (regulatory) proteins are produced within those subcellular regions, where their regulatory intervention is needed [22, 113, 385, 386].

The assumption of functional domains is supported by another work on the role of transcriptional regulation in shaping the organization of genes on a chromosome [21]. It was demonstrated that the more target genes a TF regulates the higher is its need to be expressed in higher concentrations to regulate targets located dispersedly on the chromosome. In contrast, local or dedicated TFs were found to be expressed in much

lower concentrations explaining the reasons for their proximity on the chromosome to their target genes [21]. This aspect justifies an a posteriori conformational organization of DNA to produce colocalization phenomena since it is a natural way to make three-dimensional targeting and assembly of complexes more efficient and error-free.

Additionally, there is experimental evidence from other bacteria suggesting that there is a link between the final expression products and chromosome organization [387]. Visualization of replicated DNA within living cells of Bacillus subtilis show that genes from distant chromosomal regions colocalize within a similar subcellular location for the purpose of coregulation [387].

How do bacterial cells operate such coordinated movement of specific sites within the compacted genome rapidly and faithfully? This is a place, in which physical modeling of polymers in confined space can help to relate experimental observations on E. coli nucleoid structure and to quantitatively test our models for chromosome organization.

First, we demonstrate that confinement and condensation are not sufficient to spontaneously organize the chromosome. Without dividing the chromosome into topological domains, confinement-induced organization cannot overcome propensity of the chromosome to mix. The high precision of separate sub-cellular positioning of loci located on different chromosomal domains naturally emerges as a result of entropic demixing of these structural subunits. This concept might have important implications for chromosome segregation since the increasing topological complexity of the stacked sequence of chromosomal loops implies a stronger repulsion between replicated DNA chains and consequently more precise organization and faithful segregation [23, 24].

In a subsequent step, we propose and investigate one possible mechanism for organizing these domains: the gene regulatory network. It was demonstrated that in the gene regulatory network in E. coli, regulatory genes need to be expressed in different concentrations in dependence of the genomic distance from their target genes and of the number of target genes regulated [21]. Additionally, it was shown that expressed mRNAs largely display limited dispersion from their sites of transcription, which suggests that translation is spatially organized by using the chromosome layout as a template [22]. In light of these recent findings, we suggest that regulatory control requires the colocalization of TF genes and target genes. We do not propose a detailed mechanism generating these attractive interactions (which might result from protein-protein or from protein-RNA interactions [388, 389]) but rather explore its consequences in shaping the physical structure of the E. coli chromosome.

In fact, we find that the DNA chain self-organizes into several topologically distinguishable domains where the interplay between the entropic repulsion of chromosomal loops and their compression due to the confining geometry induces a formation into a stacked sequence of interlinked domains. These domains are sufficient to generate the observed precision of E. coli chromosome structure and we estimate the domain sizes to be distributed between 10 and 700 kb, in agreement with the size of topological domains identified in the context of DNA supercoiling.

10.2 Experimental Hints

Paul Wiggins and coworkers have investigated E. coli chromosome organization by fluorescently labeling genetic loci in live cells and tracking their position with respect to the long axis of the cell [5]. Jean-Christophe Meile and colleagues have assessed the mean positioning of chromosomal loci across the width of the E. coli cell using two-dimension

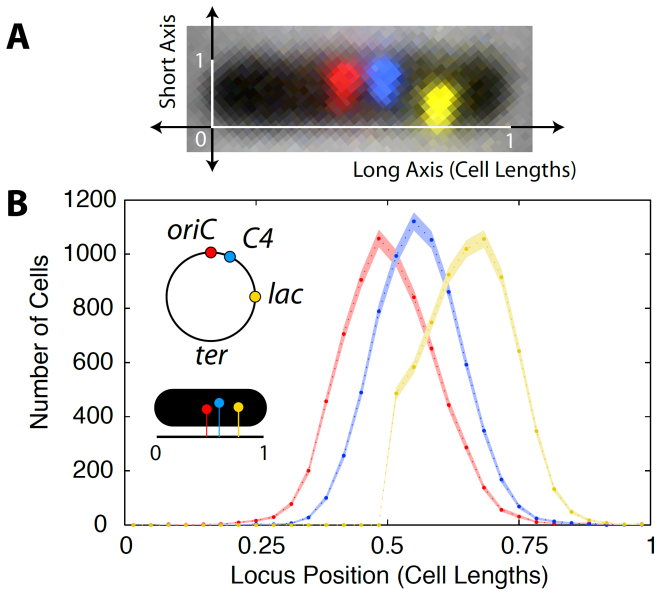


Figure 10.1: (A) Image of a typical E. coli cell with three spectrally distinct, fluorescently-labeled genetic loci. (B) Long-axis locus position distributions for oriC (red), C4 (cyan) and lac (yellow). The genomic positions of these sites along the E. coli chromosome are shown schematically in the inset. The labeled loci show a precision of subcellular positioning (variance of the distributions) of better than 10% of the cell length [5]. The images are adapted from [5].

images from wide-field fluorescence microscopy [376].

In the following, an overview over the set of performed experiments as well as of the experimental results which are important for our modeling approach is given. For further details on the experimental method (strains, plasmids, growth conditions, imaging protocols, etc.) and results the reader is referred to Ref. [5] and Ref. [376], respectively.

10.2.1 Loci Positioning with Respect to the Long Axis of the Nucleoid

To probe the nucleoid structure in live cells, Paul Wiggins and colleagues have applied both the Fluorescent Repressor Operator System (FROS) and the ParB-parS system to visualize three genetic loci concurrently [5]. Both labeling technologies allow genomic loci to be visualized by the aggregation of fluoresently labeled proteins at specific sequence introduced at the locus of interest [5]. Fig. 10.1 shows the image of a typical E. coli cell with three spectrally distinct, fluorescently labeled genetic loci.

Wiggins and coworkers have analyzed locus positioning in cells with only a single detectable oriC locus in a mix of cells in G and early S phase. To characterize the position dependence of the distributions, the experimental position distributions are fitted to Gaussian distributions in order to determine their mean and variance [5]. Loci along the terminus region are either positioned at one pole of the nucleoid, or unpositioned along the long axis of the E. coli cell [5]. These results, shown in Fig. 10.2, suggest that the two arms of the E. coli chromosome are interconnected by a low-packing-density fiber, the terminus region, that joins both nucleoid poles together [5].

The mean position of loci is shown as a function of genomic position in Fig. 10.3. The origin of replication is located at midcell and the left and right chromosome arms are positioned at separate cell halves [2, 5, 183]. Notably, the mean position is linear in the

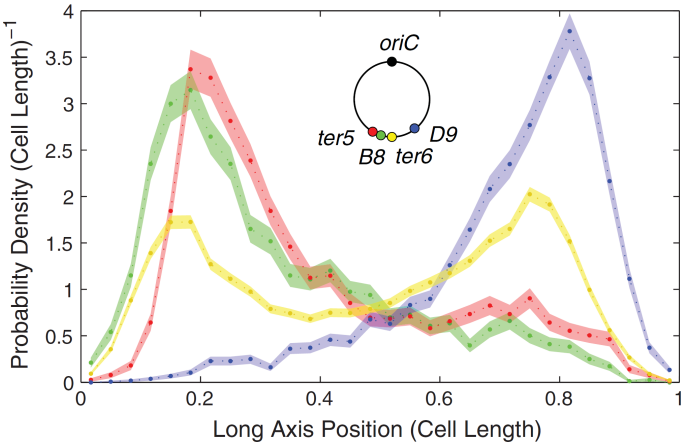


Figure 10.2: The position distributions of loci along the E. coli terminus region. The two arms of the E. coli chromosome are interconnected by a low-packing-density fiber, the terminus region, that joins both nucleoid poles together. This is why ter-proximate loci can either be well positioned (ter5, B8, D9) on one side of the nucleoid or unpositioned between the poles (ter6). The image is adapted from [5].

genomic position, indicating a linear organization at constant packing desity with respect to the long cell axis [5].

Regarding the precision of positioning indicated by the variance of the position distributions, the E. coli chromosome cannot be represented by an unstructured polymer, but measurements of the locus distributions reveal that the E. coli chromosome is precisely organized into a nucleoid filament with a linear order. In fact, Fig. 10.1 shows that loci in the body of the nucleoid show a precision of positioning within the cell of better than 10% of the cell length.

10.2.2 Loci Positioning with Respect to the Short Axis of the Nucleoid

Meile and coworkers have evaluated the position of chromosomal loci across the width of E. coli cells. Since bacteria are too thin for accurate 3D analysis by confocal microscopy, the position information of fluorescently tagged loci of the E. coli chromosome along the cell diameter has been extracted by statistical analysis of 2D images from wide-field fluorescence microscopy [376]. The distance along the cell diameter between foci and the membrane was measured as shown in Fig. 10.4. These datasets were then compared to simulated distributions based on cell width positioning models.

Meile and coworkers have detected different positioning patterns for different loci across the cell width. In particular, the terminus region of the chromosome is found to be excluded from the body of the nucleoid and preferentially located at its periphery.

10.3 Modeling

10.3.1 E. coli Chromosome Packaging in Confinement

The modeling results gained in the previous chapter suggest that the experimentally observed separate sub-cellular positioning of the three genetic loci in [5] could emerge due to their location on different chromosomal domains/loops. In fact, it is the mutual entropic

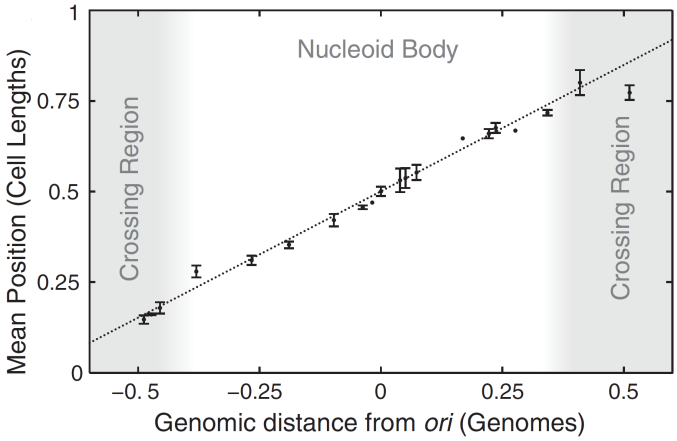


Figure 10.3: The mean long-axis position of fluorescently marked loci is plotted as a function of genomic distance from oriC showing the linear organization of the nucleoid. The image is adapted from [5].

repulsion between chromosomal domains that constitutes an elegant mechanism of ordering. In fact, polymer loops (or chromosomal domains) do not intermingle but entropically repel each other both in free space and even more strongly in confinement [23, 24, 56, 165]. This can be understood by noting that two ring polymers suffer a loss of conformational entropy when being brought together within a distance smaller than their gyration radius [23, 24, 165, 165]. This tendency to segregate, which holds to a lesser extend for linear domains, too, leads to compartmentalization.

The repulsive effect of loop formation is highlighted in Fig. 10.5 confirming that the density clouds of adjacent polymeric loops are indeed well separated. We project the segments of adjacent domains to the line connecting their centers of mass. We can thus determine the density distribution of segments of adjacent loops displaying their degree of intermingling or overlap. The data shows the relative abundance of projected segment positions of adjacent loops with respect to the axis between their 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 positions of the two centers of mass are located at $x_1 = 0.5$ and $x_2 = -0.5$, respectively. The density cloud of adjacent polymeric loops are indeed well separated. An general conclusion emerging from these results is that topological constraints imposed by looping play an important role in driving the segregation of other close-by loops or chromosomal domains.

In this chapter, we posit that it is the mutual entropic repulsion between chromosomal domains as well as the pressure exerted by the constraints imposed by the confining cavity (the nucleoid) that constitutes an elegant, self-organized mechanism of ordering in confinement due to compartmentalization. However, without dividing the chromosome into topological domains, confinement-induced organization cannot overcome the propensity of the chromosome to mix. Additionally, the experimental observation of equal variance of the position distributions displayed in Fig. 10.1 suggests the absence of tethering interactions.

In order to quantitatively test our assumptions, (i) we apply the Metropolis Monte Carlo method [146] to model the circular chromosome by a ring polymer being compacted due to fixed size loops within a rod-shaped geometry representing the nucleoid.

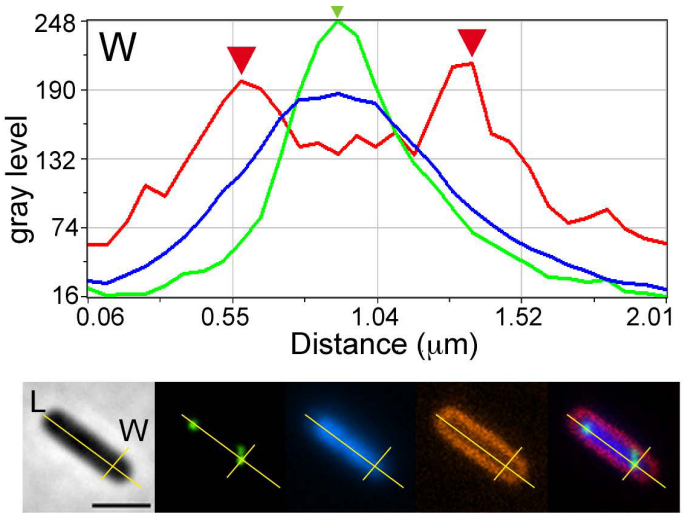


Figure 10.4: Analysis of fluorescence signals along the cell diameter (cell width). Linescans of fluorescence intensities (Y-axis, in Gray Level units) for the cell membrane (red), the DNA (blue) and YFP-ParB (green) are shown along the short cell axes (X-axis in μm). Red arrowheads indicate the cell boundaries and green arrowheads show the positions of YFP-ParB foci. The bottom panel shows micrographs of the cell scanned in the panel above with the two linescans used. (from left to right: phase contrast; YFP-ParB; DNA; membrane; overlay YFP-ParB/DNA/membrane). Scale bars indicate 2 μm. The terminus region of the chromosome is excluded from the body of the nucleoid and located at its periphery. The figure is adapted from [376].

We study various system densities ρ and various domain sizes l_{domain} being the "building blocks" of the chromosome model. In particular, the system density ρ is taken to be $\rho = \{10\%, 30\%, 40\%\}$, reflecting the typical volume fraction of the E. coli chromosome [23, 24]. Domain sizes of $l_{\text{domain}} = \{90 \text{ kb}, 120 \text{ kb}, 240 \text{ kb}\}$ are investigated, where sub-domains within the fixed size domains can dynamically form.

In agreement with the experimental results presented in Section 10.2, the *ter*-region is represented by a stretched linker connecting the two polymer arms. This approach is based on a recent work finding the E. coli chromosome organized with a linker that connects the outer edges of the nucleoid [390]. Consequently, it is modeled by a linear 80 kb polymer arm in agreement with a recent study finding the linking region to be composed of a segment of about 50 kb within the 400 kb *ter*-region [376]. The polymer chain is confined to a rectangular cuboid with an aspect ratio of the E. coli nucleoid of $r_{\text{aspect}} = 8$ [23, 24]. Fig. 10.6 illustrates the employed polymer model.

Comparing this approach to a "null model" which consists of a simple ring polymer confined to the same geometry allows for the investigation of the impact of domain formation on the packaging of highly confined chromosomes.

Chain conformations are generated by applying the bond-fluctuation method [166], which has been discussed in great detail in chapter 3. It is a coarse-grained lattice model with the advantage of avoiding non-ergodicity and its computational efficiency renders it more attractive than off-lattice models.

In order to generate thermodynamically equilibrated polymer conformations we use the Metropolis Monte Carlo method [146]. Since subsequently created conformations are highly correlated, we determine, for each set of parameters, the autocorrelation function of the squared radius of gyration. Then, the integrated autocorrelation time τ_{int} is com-

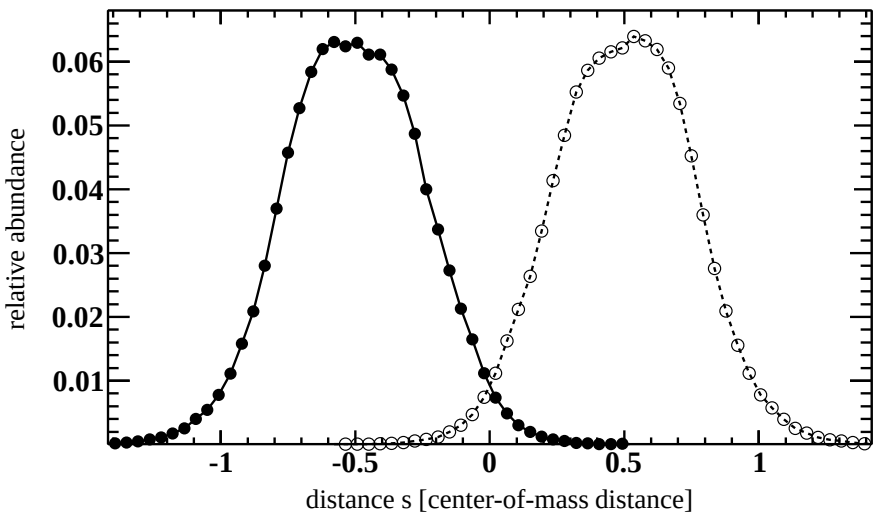


Figure 10.5: We project the segments of adjacent domains to the line connecting their centers of mass. We can thus determine the density distribution of segments of adjacent loops displaying their degree of intermingling or overlap. An general conclusion emerging from these results is that topological constraints imposed by looping or chromosomal domains play an important role in driving the segregation of other close-by loops or chromosomal domains. Image adapted from [120].

puted by applying the windowing procedure introduced by Sokal [164]. We consider two subsequent conformations as uncorrelated after $5\tau_{\text{int}}$ Monte Carlo steps therewith creating 5 000-10 000 independent configurations.

10.3.2 Gene Expression and Colocalization Phenomena

In a subsection step (ii), we propose that the structure of the E. coli gene regulatory network permits the genetic loci to identify and form domains with genetic neighbours. We investigate the consequences of this assumption by applying molecular dynamics (MD) simulations, which have been performed by Songling Li. The chromosome is modeled as a self-avoiding, but self-interacting, polymer in confinement. TFs along the chromosome are associated with the respective sites along the polymer chain. These sites interact with their target sites according to an effective attractive potential, mimicking a regulatory interplay and driving TFs and their target genes to colocalize in space. Figure 10.7 illustrates an excerpt of the flow of regulatory interactions between the TFs and their target genes as used in this work. We apply molecular dynamics simulations for two reasons: First, the (long-range) TF-gene interactions make Monte Carlo simulations ineffective. Second, already tested software packages, such as ESPResSo [393] allow for an efficient implementation of the particles and interactions.

The regulatory network of transcription factors and target genes (TF-gene interactions) was obtained from RegulonDB (`http://regulondb.ccg.unam.mx/data/network_tf_gene.txt`). Updated genome information about the E. coli K-12 strain can be found at EcoGene database (`http://maxd.cs.purdue.edu:9455/databasetable.php`). Based on the genome localization of TFs and target genes involved, we set up a coarse-grained polymer for simplicity. A monomer is set at every 1 kb genome distance (about 4640 kb for E. coli K-12 genomic size) leading to $N = 4641$ monomers building up the polymer. Then the TFs' and target gene's position is assigned. Each monomer is a hard sphere with

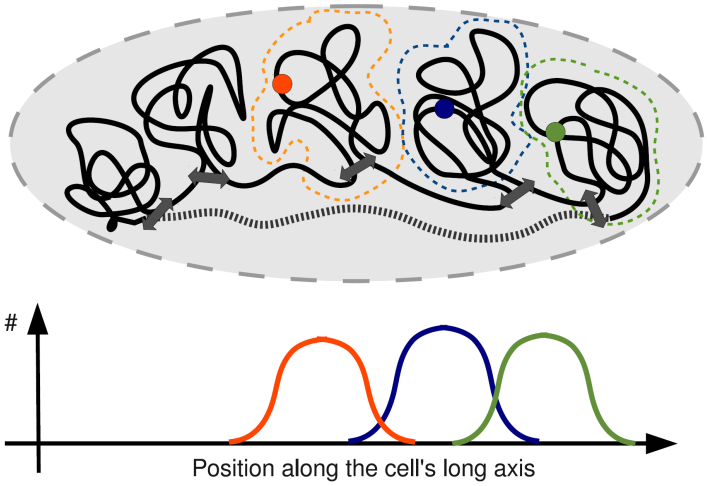


Figure 10.6: A cartoon of the chromosomal domains being the "building blocks" of the E. coli chromosome. The dark grey double arrows represent TF-gene interactions which "restrain" the domains. The domain "walls" indicated by the red, blue and green dashed lines are able to diffuse and thus subject to position fluctuations. The sub-cellular position distributions of the red, blue and green loci (fluorescently tagged in experiments [5, 390–392]) examplify the precise separate positioning of genetic sites on different domains due to their tendency of demixing. The grey dashed line represents the *ter*-proximate region acting as a linker that connects the two polymer arms. Image adapted from [120].

a bead diameter d_{bead}.

Initially, the polymer is created in the absence of confinement and eventually it is forced into the target geometry by carefully shrinking its accessible volume. In all simulations a single geometry of confinement is used: an elongated rectangular cuboid of aspect ratio 1 : 6 comparable to the aspect ratio of the nucleoid [23, 24]. The corresponding volume fraction of the chain is 10%. Our choice of parameters (aspect ratio of the cell and volume fraction of the chain) reflects the situation for E. coli and its chromosome [23, 24].

Molecular dynamics simulations are performed by applying the software ESPResSo [393] developed at the Max-Planck-Institute in Mainz (`http://espressowiki.mpip-mainz.mpg.de/wiki/index.php/Main_Page`). In the following, a short overview over molecular dynamics simulations is given, therein explaining the parameters used in the simulation, which are also summarized in Tab. 10.1. For a detailed introduction to molecular dynamics simulations the reader is referred to Ref. [394].

Molecular Dynamics

A molecular dynamics simulation consists of the numerical, step-by-step, solution of the classical equations of motion. For a system of N particles $i = 1, \ldots, N$ with coordinates $\mathbf{r}_i$ and momenta $\mathbf{p}_i$ evolving under the interaction potential $U(\mathbf{r}_i)$ these may be written as

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{p}_i = \mathbf{F}_i \qquad \text{and} \qquad \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{r}_i = \frac{\mathbf{p}_i}{m_i}, \tag{10.1}$$

where $\mathbf{F}_i = -\Delta_i U(\mathbf{r}_i)$ is the force acting on particle i. Thus, while the Monte Carlo method is generally suited for the investigation of equilibrium statistical properties of a

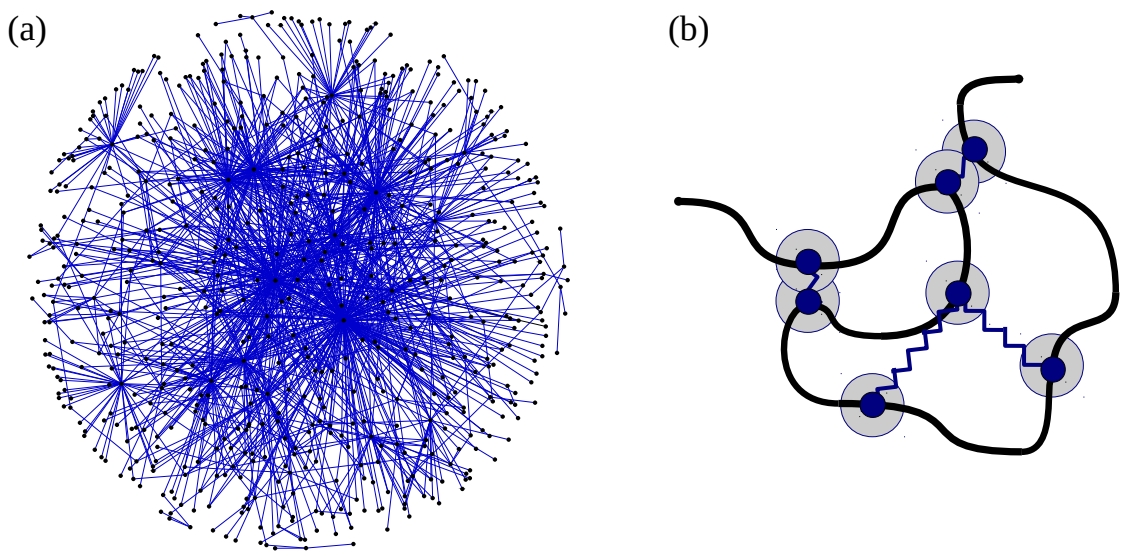


Figure 10.7: (a) Graphical illustration of the transcriptional regulatory network describing the regulatory interplay between the TFs and their target genes as applied in our polymer model. Black nodes represent genes and lines represent regulatory interactions between them. Global regulators are those TF genes that regulate lots of target genes, while other regulatory proteins are local, dedicated regulators. (b) 2D cartoon of the three-dimensional self-avoiding DNA chain (adapted from [196]) mimicking the regulatory control between TF genes and target genes by assuming a harmonic interaction between these sites. Sites that can interact with other sites are represented by small blue filled circles connected by blue springs. The outer blue circles define the strength of the harmonic interaction potential. Image adapted from [120].

system [146], molecular dynamics gives a route to dynamical properties such as transport coefficients, time-dependent responses to perturbations or rheological properties and spectra [161, 394].

There are various possibilities of solving Newton's equations of motion, among which the *velocity Verlet algorithm* [394] can be named (numerical Δt-wise integration). Depending on the system in question to be simulated there are various ways of implementing the microcanonical (NVE), the canonical (NVT), the isothermal-isobaric (NPT) or generalized ensembles (e.g. by parallel tempering [395]). A molecular dynamics simulation requires the definition of a potential function, or a description of the terms by which the particles in the simulation will interact [161]. In the following, the potentials used in this work are introduced and discussed

Excluded Volume Interactions Excluded volume interactions are simulated by a so-called Weeks-Chandler-Andersen (WCA) potential [396]. The Weeks-Chandler-Andersen (WCA) potential turns off the attractive part of the interaction between particles, while it is purely repulsive at short ranges when the distance is smaller than a cut-off distance r_{cut}

$$U_{\text{WCA}}(r) = \begin{cases} 4\epsilon[(\frac{\sigma}{r})^{12} - (\frac{\sigma}{r})^{6} + c_{\text{shift}}] & r < r_{\text{cut}}, \\ 0 & \text{otherwise.} \end{cases} \tag{10.2}$$

Table 10.1: Overview over the parameters used in the molecular dynamics simulation. Table adapted from [120].

$\rho_{\mathrm{i}} = 10^{-6}$	$K_{\mathrm{HARMONIC}} = 5 \times 10^{-5}$
$\rho_{\mathrm{f}} = 0.102$	$R_{\mathrm{HARMONIC}} = 1.5$
$T = 1.0$	nptiso_gamma0 = 1.0
$\Gamma = 1.0$	nptiso_gammaV = 1.0
$t = 0.01$	initial npt_p_ext = 1.0
$\sigma = 1.0$	initial piston = 1.0
$K_{\mathrm{FENE}} = 10.0$	npt_p_ext step = 0.08
$r_0 = 1.5$	piston step = 0.08
$\Delta r_{\max} = 1.5$	

Here, $r_{\mathrm{cut}} = \sqrt[6]{2}$ and $c_{\mathrm{shift}} = 1/4$, such that the potential is zero at the cutting distance $r_{mathrmcut}$. The WCA potential has two parameters σ, defining the radius of a monomer's hard core, and ϵ, controlling the energy penalty of two hard cores penetrating each other.

Backbone Potential The FENE (finite extension nonlinear expander) interaction is used to model the backbone interactions of the chain making sure that no bond crossings can happen. This bond type is a rubber band-like, symmetric interaction between two particles and is defined by two parameters, the prefactor K and the maximal stretching $\Delta r_{\max}$ according to

$$U_{\mathrm{FENE}}(r) = \begin{cases} -\frac{1}{2} K \Delta r_{\max}^2 \ln[1 - (\frac{r - r_0}{\Delta r_{\max}})^2] & r < \Delta r_{\max} \\ +\infty & r \geq \Delta r_{\max} \end{cases} \tag{10.3}$$

Thus, it is similar to the harmonic potential but grows to infinity at a predefined distance $\Delta r_{\max}$.

TFs Interaction Potential The classic harmonic potential is used for the regulatory interaction between TFs and target genes. This potential is determined by the prefactor K and a particle distance R, where it takes the minimal value at distance $r = R$

$$U_{\mathrm{HARMONIC}}(r) = \frac{1}{2} K (r - R)^2. \tag{10.4}$$

Thermostat for the Simulation The NPT ensemble and the npt_isotropic thermostat were applied in order to perform the isotropic changes of the box geometry, i.e. from cubic to rectangular during equilibration.

10.4 Results and Discussion

10.4.1 Division of the Chromosome into Domains is Key to Confinement-Induced Organization

The position distribution of four sites along the polymer backbone with respect to the long axis of the confining cavity both for the E. coli polymer model as well as for the "null model" are displayed in Fig. 10.8. In case of the bare ring polymer, the absence of distinguished domains leads to a loss of spatial ordering which is reflected in the polymer's mobility throughout the confining geometry. In contrast, the specific topology of the E. coli polymer model leads to an interplay between the domain's tendency of de-mixing and the pressure exerted by confinement. As a result the stacking of polymer loops inside the rod-shaped geometry induces a high level of spatial positioning illustrated by the well-defined and separated position distributions in Fig. 10.8.

The specific spatial organization is able to explain the general linear correlation between positioning of genes on the chromosome and their location in the cellular volume: adjacent loops are displaced along the long axis of the rod-shaped confining volume in such a way to form a sequence of stacked "neighbor" loops or an "effective nucleoid filament" [5]. The linear relationship between genomic and sub-cellular position is apparent in Fig. 10.8, too.

An analysis of how the SMC-like protein MukBEF condenses DNA revealed its likely involvement in organizing the chromosome in a series of loops orthogonal to the cell axis, which was argued to account for the orderly arrangement of the chromosome [397]. Here, we show that entropy might provide a purely physical driving force to support the dedicated action of specific proteins such as chromosome condensation proteins in order to create the right physical conditions for chromosome packaging. Notably, this result is confirmed by the experimental observation that cells with the MukBEF deletion are capable of almost wild type chromosome structure [398].

An established measure of compartmentalization, which is also open to experimental testing [87], are the contact probabilities between pairs of sites along the polymer as shown in Fig. 10.9. The sites are numbered consecutively and the contact map displays their probabilities of establishing a contact, i.e. whenever they are spatially closer than a threshold distance $d_{\text{threshold}}$. In fact, the very low contact probability of non-diagonal contacts in Fig. 10.9 shows that domain formation induces contacts among sites within the same and between sites of genomically neighboring domains. Thus, the experimentally observed high precision of sub-cellular locus positioning of better than 10% of the cell length [5] for genetic sites lying on different chromosomal domains has to be interpreted within the concept of structural units as the building blocks of the chromosome consisting of (supercoiled) DNA loops stabilized by DNA-binding proteins [399]. The size of the structural units simultaneously influences the organization of the chromosome (segregation of chromosomal domains versus mixing) and its conformation (ordered versus random) in confined space [23]. Consequently, the size of the structural unit is reflected in the decay width of the contact probability backbone shown in Fig. 10.9. While Valens et al. [109] provide one measure of contact probabilities Hi-C would provide another complementary approach with the opportunity to observe the interactions in a different biochemical context (cross-linking) and with a higher resolution in order to test our predictions [87].

Moreover, while other domain organization models conclude that domain barriers are not placed stably at fixed sites of the chromosome, but are effectively randomly distributed hard walls [30], we have shown that their position need not be stochastic, since the domain

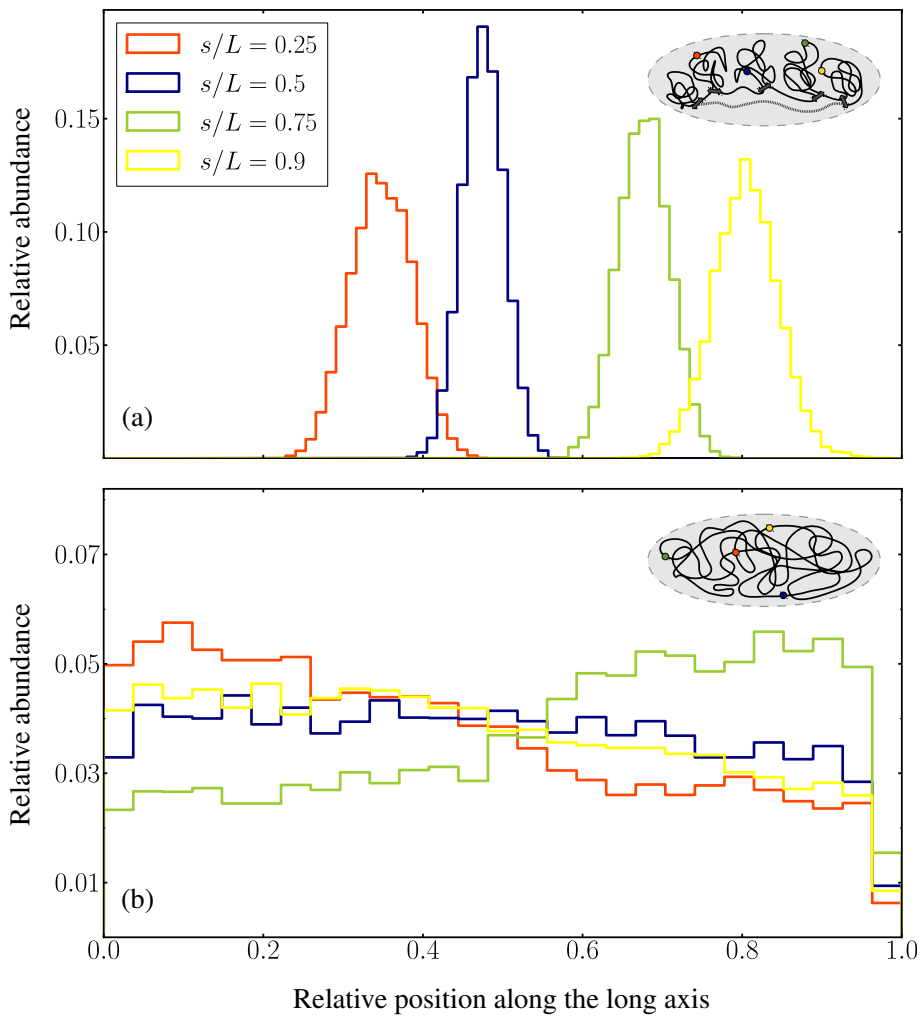


Figure 10.8: The position distribution of four sites located at the relative positions $s/L = 0.25$, 0.5, 0.75 and 0.9 along the polymer backbone L are shown with respect to the long axis of the confining cavity for (a) the domain model and (b) the "null model". In case of the bare ring polymer, the absence of distinguished domains leads to a loss of spatial ordering which is reflected in the polymer's mobility throughout the confining geometry. In contrast, the specific topology of the E. coli polymer model induces a high level of spatial positioning. The interplay between the domain's tendency of de-mixing and the pressure excerted by confinement leads to a stacking of polymer loops inside the rod-shaped geometry. Image adapted from [120].

walls are able to diffuse and thus subject to position fluctuations.

A recent study has evaluated the position of E. coli chromosomal loci across the width of the cell by tagging loci with fluorescent proteins and comparing the measured distributions with simulated ones from different cell width models [376]. The terminal region of the chromosome was found to be preferentially located at the periphery of the nucleoid consistent with its specific role in chromosome segregation [376, 400]. Our model of the E. coli chromosome topology displays the same characteristic feature. In our model, the polymer chain region, which represents the ter-borne chromosome part, is preferentially located at the periphery of the confining cavity with respect to its short axis as is shown in Fig. 10.10. Within our modeling approach, this result can be understood by noting that

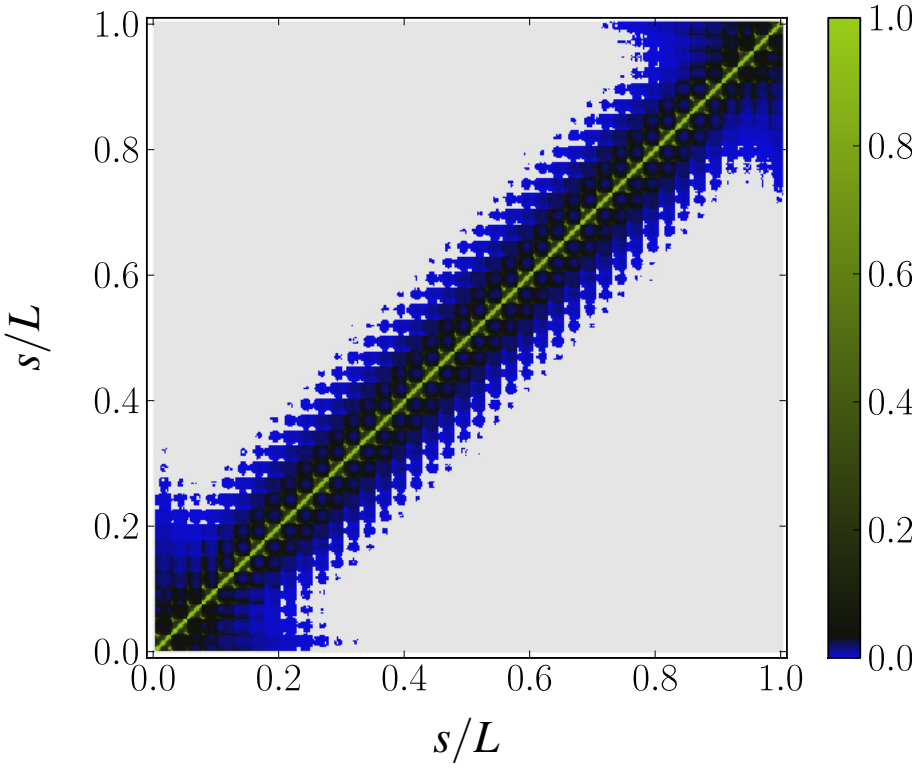


Figure 10.9: The contact probability map shows the probability for pairs of sites along the polymer backbone s/L to form a contact, i.e. whenever both sites are spatially closer than a threshold distance $d_{\text{threshold}}$. Domain formation induces contacts among sites within the same domain and between sites of neighboring domains. The experimentally observed high precision of sub-cellular locus positioning can be explained by noting that the mobility of sites belonging to different chromosomal domains is restricted to the radius of the so defined structural subunit. Here, high-throughput chromosome capture (Hi-C) technology can be a useful method to identify DNA elements that interact on a genome-wide scale and thus test our predictions [87]. Image adapted from [120].

the (mostly linearly stretched) *ter*-region has less topological complexity compared to the looped "filament"-like structure of the remaining nucleoid. The topological complexity of the chromosomal domains not only leads to an entropic repulsion between adjacent domains but also pushes the *ter*-region towards the envelope of the confining cavity. A donut like topology is observed if the topological complexity of chromosome is uniform [35].

10.4.2 Regulatory Control by Colocalization as a Mechanism for Domain Formation

We have shown that the interplay between entropic repulsion of chromosomal domains and pressure exerted by the envelope of the confining cavity can be one driving force for nucleoid organization (and segregation) supporting the action of dedicated cellular machinery [183, 391] or internal pushing forces [23, 24, 392]. However, while the existence of chromosomal domains is widely accepted, their size and the mechanism that gives rise to them is still under debate. In this work, we investigate one possible mechanism for organizing chromosomal domains: the gene regulatory network.

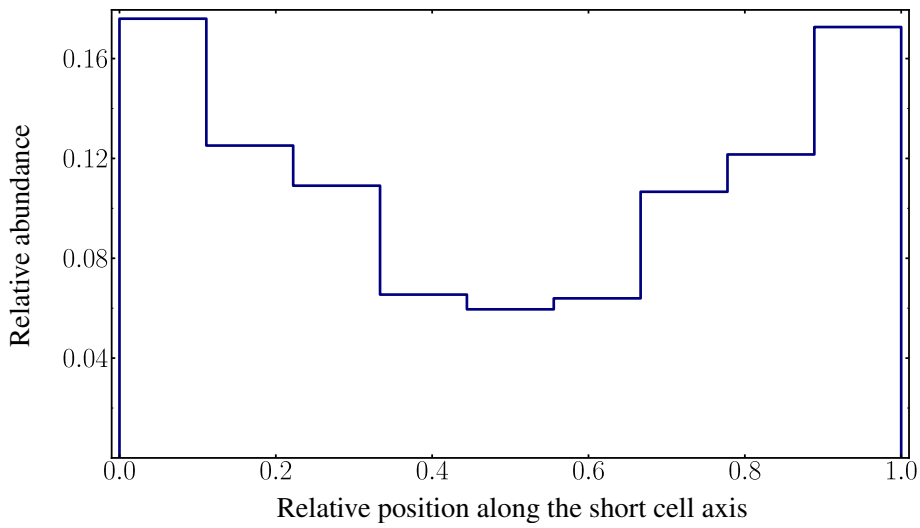


Figure 10.10: The relative abundance of *ter*-proximate sites as a function of their positioning with respect to the short axis of the confining geometry (cell's small axis). Meile and coworkers [376] have determined the position distributions of loci in the terminus region with respect to the short axis of the cell. They find these ter-borne loci localized at the nucleoid periphery. In our model, the polymer chain region representing the stretched *ter*-proximate region connecting the two nucleoid edges is preferentially located near the confining envelope, too, thus confirming the experimental observation. Within our modeling appraoch, this can be understood by noting that the (mostly linearly stretched) *ter*-region has less topological complexity compared to the looped "filament"-like structure of the remaining nucleoid. The topological complexity of the domains not only leads to an entropic repulsion between adjacent chromosomal domains but also pushes the *ter*-region towards the envelope of the confining cavity in agreement with experimental findings [376]. Image adapted from [120].

The decision about gene expression or repression is controlled by TFs which use metabolic or environmental signals to trigger a transcriptional response [3,9,106] within a functional network of regulatory interactions between TFs and target genes [19,20,118,119]. Additionally, the proper genome-wide coordination of gene expression has been shown to be linked to the spatial organization of the chromosome within the nucleoid [18,106,113,361]. In this respect, one can distinguish between "analog" control, i.e. regulatory action by chromosome topology, and "digital" control, i.e. regulation mediated by transcription factors [19,118,119].

Regarding the observed existence of chromosome domains in E. coli, we propose that one-dimensionally distant target genes, i.e. genes that are genomically far away from their regulative TF, colocalize with it in order to facilitate transcription. The formation of dedicated chromosomal domains, possibly stabilized by nucleoid-associated proteins who are in charge of most chromosomal remodeling tasks, could be seen as a feature of analog control.

Since a model for E. coli domain organization involves the recognition of a domain-specific pattern by a protein which would isolate it from other domains, various experimental groups have been looking for dedicated proteins that bind specifically to a single domain to organize the chromosome. A few examples of domain forming proteins have been identified. In B. Subtilis, the DNA-binding ParB-like protein Spo0J appears to ensure proper arrangement and partitioning of chromosomal DNA by recruiting the condensin structural

maintenance of chromosomes (SMC) complex to the replication origin region [401, 402]. In E. coli, SMC-like complex MukBEF appears to colocalize with the origin of replication, but the mechanism is unknown. A recent study has identified a protein MatP which structures the terminus macrodomain. ChIP experiments have revealed that proteins SeqA and SlmA both appear to be excluded from the terminus region [15, 403, 404]. However, the existence of these known examples is insufficient to explain the precision of structure observed on the chromosome [15]. Here, we propose that structure proteins have been hiding in plain sight. They are the transcription factors that are already known to target small regions of the chromosome specifically.

Here, we mimic the regulatory control between TF genes and target genes in the transciptional regulatory network by assuming a harmonic interaction between these sites as illustrated in Fig. 10.7. Within this rather general framework, we find that the DNA chain indeed self-organizes into several topologically distinguishable domains. Fig. 10.11 displays a snapshot of the genome organization as obtained after the equilibration of the self-interacting DNA chain, i.e. when the system fluctuates around its global energy minimum.

Local TFs regulating a small number of target genes tend to localize peripherally, while global TFs which regulate a large number of genes assume more central localizations within the confining geometry. Several models have been proposed that lead to such a macro-arrangement of the nucleoid [14,405,406]. In particular, the microarray experiments by Jeong et al [384] show a high degree of correlation between the transcriptional signal of genes close together on the chromosome, where the observed stability and range of correlations extend far beyond the expected size of the average operon. Such dependence offers an intriguing hypothesis about the physical basis of the short-range transcriptional correlations: the transcription of the genes within a chromosomal domain is more similar to each other than to genes in other domains [384].

The folding of the nucleoid in domains due to gene colocalization puts into spatial contact distant chromosomal regions. The inset of Fig. 10.11 shows the position distributions for the three genetic loci *oriC*, *C4* and *lac* concurrently visualized in [5] as obtained by our model of the E. coli nucleoid. We find the three genetic loci to be located on different domains emerging due to TF-gene colocalization. In agreement with this finding, the genomic distances between the loci *oriC* and *C4* as well as between the loci *C4* and *lac* being about 300 and 600 kb, respectively, formally confirm the sites' positioning on different domains even when an upper loop size limit of 700 kb (emerging from our numerical calculations) is assumed. Additionally, we find the terminus region to be localized at mid-cell connecting the two chromosome arms in agreement with [5, 376].

Moreover, the inset of Fig. 10.11 also shows that the assumption of TF-gene interactions as drivers of domain formation (among other possible mechanisms) are able to reproduce the linear correlation between the position of a gene on the chromosome (i.e. site along the polymer) and its sub-cellular position inside the nucleoid (i.e. the long axis of the confining cavity). In fact, the strong linear correlation is a direct consequence of the fact that DNA is compacted and confined. Notably, there is only a linear correlation for sufficiently large domain sizes since the absence of domains leads to a loss of precision of positioning and consequently linear ordering [45].

Thus, we study the distribution of loop sizes in Fig. 10.11 finding them to range between 10 and 700 kb in agreement with the size of topological domains identified in the context of DNA supercoiling [384, 405]. We find a mean domain size of $\langle l_{\text{domain}} \rangle = 86$ kb. The long tail of large loop sizes (> 150 kb) can be explained by noting that the larger

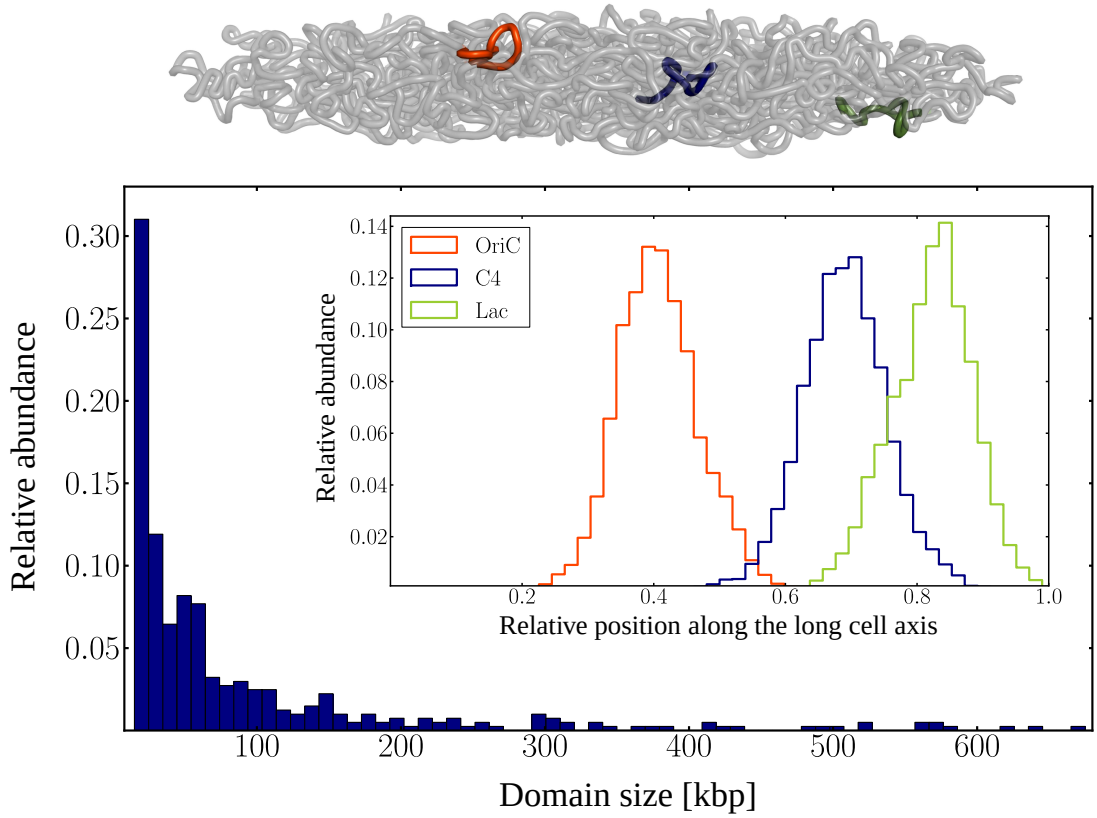


Figure 10.11: The three-dimensional chromosome organization is obtained after the equilibration of the interacting self-avoiding polymer chain in a rod-shaped geometry. Modelling the regulatory interplay between TFs and their targets based on the E. coli transcriptional regulatory network, we find that genes on a DNA chain self-organize into several topologically distinguishable domains of different sizes. The chromosomal loop to which *oriC* is associated with is displayed in red, while the blue and green marked chromosomal regions refer to the chromosome domains that contain the genetic loci *C4* and *lac*, respectively. It becomes clear that the three genetic loci are well separated with respect to projections on the long axis of the confining envelope. The formed domains are sufficient to generate the observed precision of E. coli chromosome structure and we estimate the domain sizes to be distributed between 10 and 700 kb. Image adapted from [120].

domains are themselves built up by smaller subdomains in agreement with the observed long-range correlations in gene expression patterns [104, 383, 384].

10.5 Conclusions

In this chapter, we have (i) proposed a mechanism by which chromosomal domains are formed and (ii) quantitatively investigated this model applying numerical simulations. We assume the structure of the E. coli transciptional regulatory network to give rise to colocalization of regulators and their target sites due to attractive interactions between genetic loci. Under these conditions, we find that the DNA chain self-organizes into several topologically distinguishable domains where the interplay between the entropic repulsion of chromosomal loops and their compression due to the confining geometry inducing a

formation into a stacked sequence of interlinked domains or "rosettes" [407]. Thus, the experimentally observed high precision of separate sub-cellular positioning of genetic loci located on different chromosomal domains naturally emerges as a result of entropic demixing of loops where the precision of localization is related to the position fluctuations of each chromosomal domain. To recover the precision of organization observed in E. coli, we estimate the domain sizes to be distributed between 10 and 700 kb, in agreement with the size of topological domains identified in the context of DNA supercoiling.

However, the question has to be raised whether the circular chromosome behaves as an equilibrated polymer [408]. Calculation of the rate of uncoiling of the DNA molecule can be computed by taking into account the increase of entropy on unwinding as well as the viscous resistance of the surrounding medium [409]. Thus, at approximately 300 K the bidirectional replication of a 4 Mbp chromosome would require about 20 minutes [409], which might leave a substantial amount of time for the non-replicating phase where chromosome domains could be rearranged. Notably, the study of E. coli chromosome organization is influenced by the experimental conditions, which might not be judicious due to different genetic background and growth conditions. Due to the high packing density of the nucleoid the bacterial chromosome might not be able to explore the whole configuration space of possible conformations but rather a restricted subspace starting from similar initial configurations emerging due to the progressive segregation and DNA compaction after each replication cycle. In fact, numerical calculations on time scales that are small compared to the relaxation time of our polymer system indicate even more precise locus positioning due to "frozen-in" configurations.

Our concept could have important implications for chromosome segregation, too. It was previously shown that compaction of the bacterial chromosome and conformational entropy alone could direct and facilitate the segregation of newly replicated daughter strands of DNA [23, 24]. In this work, the sequence of stacked domains has higher internal topological complexity compared to linear or circular chains leading to even stronger repulsive interactions. Thus, segregation by entropic forces induced by chain topology in strong confinement might constitute a reliable mechanism, where no additional drivers such as a mitotic-spindle-like machinery or dedicated proteins may be needed [23].

We have not taken into account the effect of the chromosomal domains being negatively supercoiled by either plectonemic or toroidal supercoils [405] since we have focused on the global physical properties of the chromosome on a more coarse-grained level. However, DNA gyrases, which cause branched supercoils and thus increase domains topological complexity, further promote entropic repulsion between the supercoiled chromosomal domains as pointed out above.

Summarizing, our approach offers a robust framework for understanding the basic physical principles underlying E. coli chromosome organization. Its advantage is that it does not depend on the microscopic details of the DNA chain or on specific DNA-protein interactions. In light of the difference in length scales between proteins and chromosomes the question has to be raised whether local actions of specific proteins alone are able to globally shaping chromosome organization [15, 23]. Thus, our model is based on the idea that nature exploits entropy, excluded volume, specific chromosome topologies and confinement as a driving force to create the right physical conditions for chromosome packaging eventually fine-tuned by the dedicated action of proteins such as NAPs that bridge and bend DNA [16] or chromosome condensation proteins (SMC [401] and MukBEF [410]).

In the future, various experimental techniques are available to further probe the precision of our E. coli chromosome model. Chromosome conformation capture (Hi-C) technol-

ogy [87] can be a useful method to identify DNA elements that interact on a genome-wide scale and thus test the contact probability map shown in Fig. 10.9. Moreover, fluorescently labeling more than three genetic loci separated by a genomic distance much smaller than the mean domains size of $\langle l_{\text{domain}} \rangle = 86$ kb might be able to resolve position distributions of genetic sites on the same chromosomal domains therewith revealing additional insights into the underlying chromosome topology. Last but not least, the investigation of E. coli chromosome dynamics, i.e. tracking the origin of replication oriC during the cell cycle might allow for a better understanding of mechanisms underlying chromosome segregation.

Chapter 11

Polymer Segregation in Confined Spaces

Implications for E. coli Chromosome Segregation

References

- Y. Liu, M. Fritsche, D.W. Heermann and B. Chakraborty (2011), *Segregation of Polymers in Confined Spaces*, in preparation.

Chapter Summary

In this chapter, we study how excluded volume effects, specific polymer topologies and geometrical confinement compete with entropy to drive dynamical processes such as the segregation of highly compacted chromosomes during cell division. Our work shows that the elongated geometry of the bacterial cell provides a natural axis along which the two copies of the chromosome separate. Notably, the velocity and completion of the segregation process in both cubic and rectangular confinement strongly depend on the flexibility of the biopolymer. A small increase in chain stiffness is already sufficient to induce a failure of the segregation progress. In the future, reliable assessment of the segregation time scale applying technologies such as microfluidics might provide a calibration of Monte Carlo time steps and thus help to underpin the concept of spontaneous, entropy-driven chromosome segregation.

11.1 Introduction

In the previous two chapters, we have investigated how the larger scale, bacterial DNA is organized within the nucleoid. In this chapter, the implications of this large scale topological organization with respect to the process of chromosome segregation are discussed. In particular, we would like to shed light onto the mechanism(s) that move the newly replicated DNA strands to opposite cell halves. For a concise and detailed overview over the biology of bacterial chromosome segregation the reader is referred to Ref. [411] and Ref. [412].

As in most bacteria, replication in E. coli is initiated at a single origin, oriC [411–413], and progresses bidirectionally towards the replication terminus region, ter [411–413], which has already been introduced in the previous chapter. Fig. 11.1 schematically illustrates the replication/segregation process. At replication initiation, the E. coli replication machinery assembles at the origin of replication (OriC), which is located close to midcell, and the two replisomes track independently around the chromosome [411–413]. After replication, the newly replicated sister genetic loci segregate to opposite cell halves leading to a spatial separation of sister loci sequentially 5 to 20 min after replication [411–413]. However, the processes that contribute to E. coli chromosome segregation are poorly understood [413].

There are several hypotheses on how proteins or dedicated mechanisms could facilitate chromosome segregation in bacteria [23,24,411,413,414]. An early model depicts bacterial chromosomes as attached to the membrane and their segregation as a result of the elongation of the membrane [415]. However, findings on the speed of chromosome movement and cell elongation prove that this model cannot explain the experimental observations, which show chromosomes to move much faster than the cell elongation rate [416].

Moreover, in contrast to other bacteria, such as Bacillus subtilis and C. crescentus, which apply tethering mechanisms attaching specific chromosome regions to cell poles and therewith facilitating chromosome segregation [115,417,418], no such tethering mechanism is know for the case of E. coli [413]. In fact, a dedicated “mitotic-like” mechanism, as is known for eukaryotes, has remained elusive until now.

Recently, it has been proposed that E. coli chromosomes, being modeled as linear DNA chains, could segregate spontaneously (without the help of dedicated actin-related

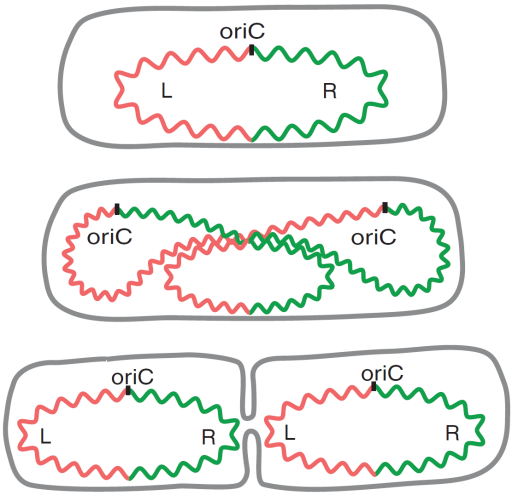


Figure 11.1: Chromosome organization in E. coli during the replication/segregation process. The origin of replication (OriC) is located at midcell, and each chromosome arm occupies separate cell half (left and right). Completion of replication and segregation results in two daughter chromosomes, which adopt the same configuration as their mother cell. Image adopted from [2].

proteins) due to their entropic repulsion in rectangular confinement, which could guide both polymer chains to less crowded positions within the confined space [23,24]. While this model successfully describes the replication-segregation process of a cell cycle it requires the introduction of an auxiliary construct in the form of an inner tube restraining the mother chromosome's movement and only allowing the daughter chromosome to occupy both the inner tube and the outer tube space. Yet, the existence of such a differential space restriction for the mother and daughter chromosomes awaits additional experimental evidence [23, 24, 414].

Two additional aspects, which are so far neglected in the entropy-driven segregation model [23, 24], are the specific E. coli chromosome topology as well as its bending stiffness. In fact, the chromosome is not linear but circular and further compacted due to chromosomal domain formation. Thus, additional entropic repulsion due to topological constraints (chain closure and domain organization) is expected, which, in turn, might minimize the segregation time.

The second aspect regarding the bending rigidity of the chromosome fiber has already been taken into account in chapter 5 and affects the (diffusive) movement of sites along the chromosome and consequently the segregation process [419]. While the effective bending stiffness of the bacterial E. coli DNA molecule is unknown [419], DNA molecules suspended in liquid have a thermal persistence length of 50 nm [420]. Thus, studying effect of bending rigidity is essential for (quantitative) understanding of cellular phenomena such as chromosome segregation.

To gain insights into the impact of chain topology as well as bending stiffness on chromosome segregation we apply Monte Carlo simulations: Two ring polymers, representing the circular sister chromosomes, at different chain stiffness segregate in different three-dimensional geometries (squared and rectangular). Additionally, this system is compared to a "null" model, consisting of two linear chains of the same length and in the same geometries as illustrated in Fig. 11.2.

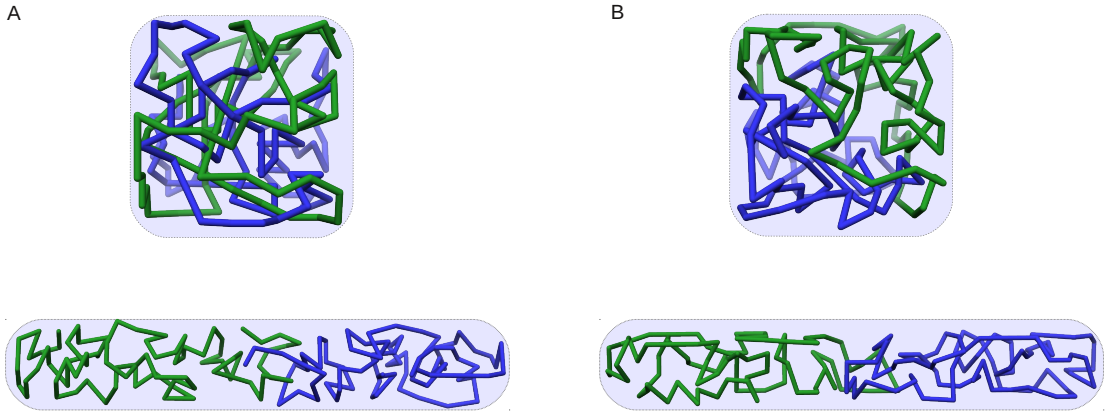


Figure 11.2: (A) "Snapshot" configurations of two linear polymers and (B) of two circular chains confined in a square and rectangular cuboid, respectively. In rectangular confinement (semiflexible) linear and circular chains segregate, while spontaneous segregation only happens for fully flexible ring polymers in cubic geometries. Linear chains as well as semiflexible ring polymers mix in squared confinement.

11.2 Modeling

Our model includes two E. coli sister chromosomes described as two self-avoiding ring polymers. Approximating biological "storage" such as the bacterial nucleoid an elongated rectangular cuboid of aspect ratio 1 : 6 comparable to the aspect ratio of the nucleoid [23, 24] confines both ring polymers. This geometry is compared to a cubic cavity retaining the same accessible volume and thus the same system density (monomer to volume ratio) for both geometries. Studying two polymer rings of lengths $N = 80$, the linear dimensions of the confining geometry are set up such that the radius of gyration R_{gyr} of the unconfined chain is larger than the linear square box sizes. Fig. 11.2 illustrates "snapshot" conformations of the two polymer chains in both confining geometries.

An overlapping configuration of two chains is created to initiate the segregation process as is illustrated in Fig. 11.3. Independent Monte Carlo trajectories (driven by different random number sequences) representing the dynamics of the segregation process are then sampled.

To generate polymer conformations we employ the bond-fluctuation method (BFM) [166], which has been applied successfully to model the static and dynamical properties of polymer systems in several investigations [146, 147]. As has been discussed in chapter 3, it is a coarse-grained lattice algorithm with the advantage of avoiding non-ergodicity and its computational efficiency, particularly with respect to dense systems, renders it more attractive than off-lattice models.

Semiflexible polymers may be characterized by their persistence length l_{p}, which is the typical length scale over which the chain backbone loses information about its direction due to thermal fluctuations [139]. In a lattice representation such as the BFM, the bending energy H_{b} can be expressed as [249]

$$H_{\mathrm{b}} = \frac{l_{\mathrm{p}}}{2b} \sum_{i=1}^{N-1} (\mathbf{u}_{i+1} - \mathbf{u}_i)^2. \tag{11.1}$$

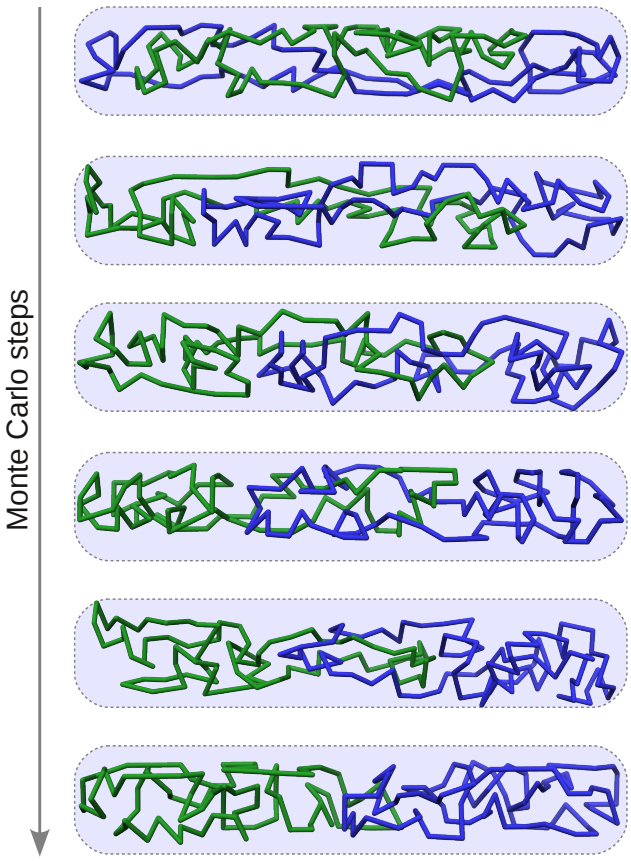


Figure 11.3: "Snapshots" of two segregating circular polymer chains in dependence on Monte Carlo time steps.

Here, $\mathbf{u}_i = \left(\frac{\mathbf{R}_i - \mathbf{R}_{i-1}}{|\mathbf{R}_i - \mathbf{R}_{i-1}|} + \frac{\mathbf{R}_{i+1} - \mathbf{R}_i}{|\mathbf{R}_{i+1} - \mathbf{R}_i|}\right) \Big/ \left|\frac{\mathbf{R}_i - \mathbf{R}_{i-1}}{|\mathbf{R}_i - \mathbf{R}_{i-1}|} + \frac{\mathbf{R}_{i+1} - \mathbf{R}_i}{|\mathbf{R}_{i+1} - \mathbf{R}_i|}\right|$ is a discrete realization of $\delta \mathrm{R}(s)/\delta s$, the unit tangent vector at arclength s, where $\mathrm{R}(s)$ is the position vector [230]. N is the total number of monomers in the chain. All energies are measured in units of $k_\mathrm{B}T$.

Notably, a recent study [157, 248] has shown that standard definitions of persistence length might fail for chains with excluded-volume restrictions, stressing the importance of carefully checking in which regime experimental data belong. In this work, we interpret the decay of the orientational correlation function (which is manipulated by the bending energy contribution H_b) in terms of an effective "quasi" persistence length reflecting global conformational flexibility rather than local intrinsic stiffness. The range of chain rigidity is varied from a totally flexible chain, $l_\mathrm{p}/L \to 0$ up to a lightly stiffer regime $l_\mathrm{p}/L \to= \{0.0125, 0.025, 0.05\}$ to investigate the effect of chain stiffness on the segregation properties.

Table 11.1: Summary of the mean normalized squared center of mass distances.

l_p/L	rectangular prism, circular chain	rectangular prism, linear chain	cube, circular chain	cube linear chain
0	0.76 ± 0.045	0.68 ± 0.051	0.5 ± 0.03	0.34 ± 0.02
0.0125	0.76 ± 0.05	0.11 ± 0.04	0.43 ± 0.03	0.21 ± 0.02
0.025	0.72 ± 0.049	0.02 ± 0.007	0.36 ± 0.029	0.14 ± 0.017
0.05	0.42 ± 0.046	0.03 ± 0.01	0.33 ± 0.024	0.12 ± 0.017

11.3 Results and Discussion

A measure which quantifies the degree of segregation between two polymer chains is the mean squared distance D'^2_{cm} between their centers of mass R_{cm} and r_{cm}, respectively,

$$D'^2_{\mathrm{cm}} = \langle (R_{\mathrm{cm}} - r_{\mathrm{cm}})^2 \rangle. \tag{11.2}$$

In order to be able to compare D'^2_{cm} for different confining geometries, it is normalized with respect to the maximum distance D^2_{max} found for all simulated trajectories such that $D^2_{\mathrm{cm}} = D'^2_{\mathrm{cm}}/D^2_{\mathrm{max}}$. As a first step, we are interested in D^2_{cm} as a function of Monte Carlo moves. However, the number of accepted Monte Carlo moves is influenced by the underlying circular or linear chain topology since the chain architecture itself imposes constraints on possible moves. In order to take this effect into account, the ratio of actually performed moves M_{accept} and overall suggested moves M_{total} is determined for each simulated segregation process. Then, the Monte Carlo segregation time $t^{\mathrm{norm}}_{\mathrm{MC}} = t_{\mathrm{MC}} \times \frac{M_{\mathrm{accepted}}}{M_{\mathrm{total}}}$ is normalized.

Fig. 11.4 shows D^2_{cm} as a function of $t^{\mathrm{norm}}_{\mathrm{MC}}$ for two segregating ring and linear polymers at different bending rigidities in rectangular confinement. Even a small raise in bending rigidity increases the time needed for demixing eventually resulting in the breakdown of spontaneous segregation. This effect is stronger for linear chains where failure of segregation occurs at even lower bending rigidity. While each individual segregation trajectory might display variations in the segregation velocity, the qualitative tendency when comparing chains of different bending stiffness holds as is illustrated in Fig. 11.5. Independent Monte Carlo trajectories (driven by different random number sequences) representing the dynamics of the segregation process are shown for the same parameters (geometry of confinement, chain topology, bending rigidity).

In contrast to rectangular confining, cubic geometries induces spontaneous segregation only in flexible ring polymers, while linear chains in general as well as semiflexible circular ones do not fully segregate. Tab. 11.1 summarizes the mean normalized squared center of mass distances.

11.4 Conclusions

In this chapter, we have studied how excluded volume effects, specific polymer topologies and geometrical confinement compete with entropy to drive dynamical processes such as the segregation of highly compacted chromosomes during cell division.

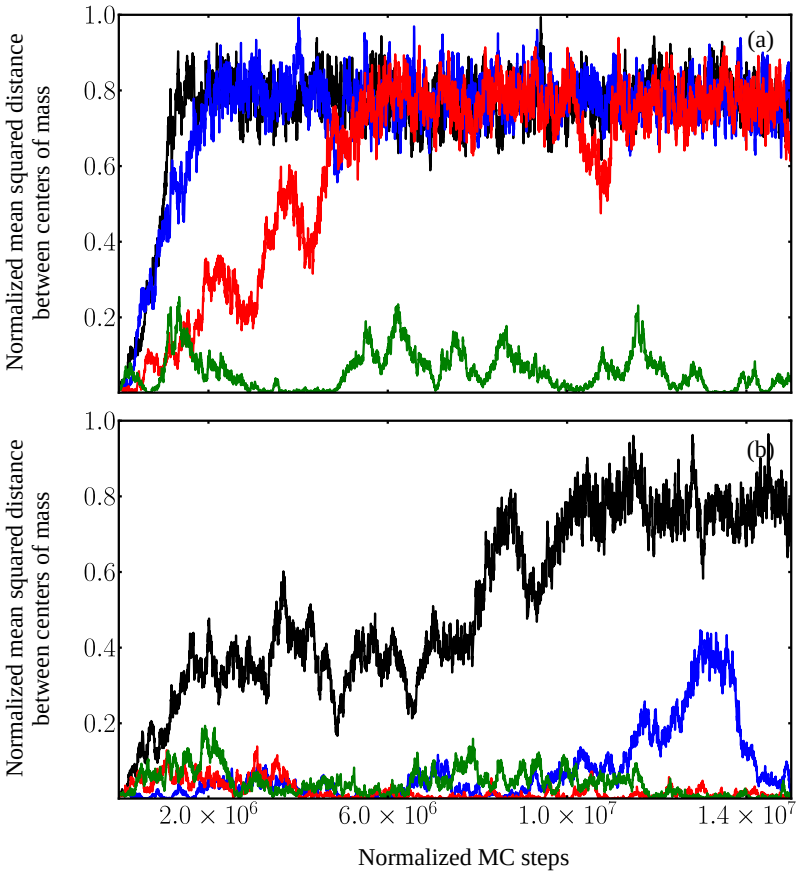


Figure 11.4: D^2_{cm} as a function of $t^{\mathrm{norm}}_{\mathrm{MC}}$ for (a) segregating ring and (b) linear polymers at different bending rigidities in rectangular confinement. Even a small raise in bending rigidity increases the time needed for demixing eventually resulting in the breakdown of spontaneous segregation. This effect is stronger for linear chains where failure of segregation occurs at even lower bending rigidity.

Our work shows that the elongated geometry of the bacterial cell provides a natural axis along which the two copies of the chromosome separate. Notably, the velocity and completion of the segregation process in both cubic and rectangular confinement strongly depend on the flexibility of the biopolymer. A small increase in chain stiffness is already sufficient to induce a failure of the segregation progress.

An important step towards adding more biological relevance to the analyses performed so far is the calibration of our simulational results with experimental data such that we can relate Monte Carlo time steps to "real" time. Reliable assessment of the segregation timescale might help to prove or disprove the concept of spontaneous chromosome segregation. To this end, cellular dynamics can be interrogated by microscopic tracking of fluorescently-tagged molecules or genetic sites in living cells. In particular, technologies such as microfluidics are increasingly able to capture cellular processes and interactions with high spatial and high temporal resolution as well as possibly high experimental throughput [421].

Additionally, the motion of two overlapping linear polymers confined to a 2D rectangle can be treated analytically. Our collaborators Bulbul Chakraborty and Ya Liu have proposed a statistical model based on blob free-energy arguments [422]. Based on such a

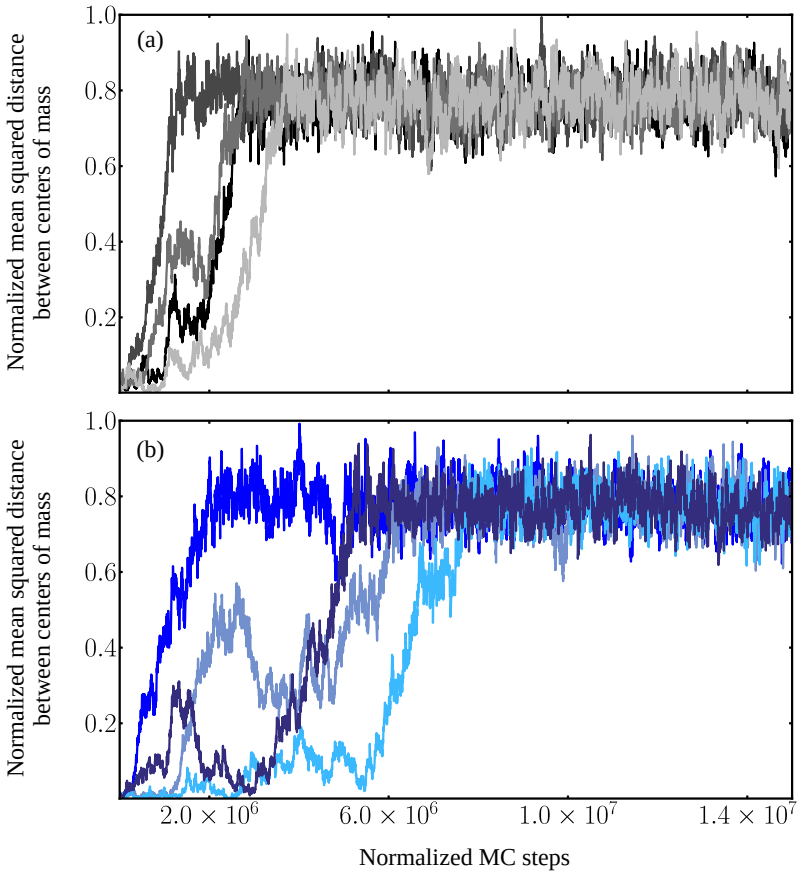


Figure 11.5: Independent Monte Carlo trajectories (driven by different random number sequences) representing the dynamics of the segregation process are shown for (a) fully flexible ring polymers ($l_\mathrm{p}/L \to 0$) and (b) semiflexible circular chains with $l_\mathrm{p}/L = 0.0125$ in rectangular confinement. While each individual segregation trajectory might display variations in the segregation velocity, the qualitative tendency when comparing chains of different bending stiffness holds.

model, spontaneous segregation is found under the condition that the (longitudinal) length of the rectangle is larger than the (longitudinal) polymer extension. Notably, in this case, the segregation time shows a minimum as a function of the length of the rectangle's long axis.

Applying the analytical approach to E. coli system parameters locates the bacterium in the segregation phase and close to the geometrical condition of minimum segregation time [422]. Although a chromosome strand is more complicated than a linear polymer and these results are restricted to two dimensions, one might speculate that genome segregation time could have applied evolutionary selection pressure to the relationship between genome lengths and the geometry of the surrounding envelope [422].

Chapter 12

Unraveling Chromosome Architecture

A Maximum Entropy Approach for 3D Genome Conformation Reconstruction

References

The results presented in this chapter are adapted from

- M. Fritsche*, P.M. Diesinger*, D.W. Heermann and M. Bathe (2011), *A Maximum Entropy Approach for 3D Genome Conformation Reconstruction*, in preparation.

We thank Job Dekker, Josée Dostie and James Fraser for access to their published data, as well as James Fraser for help in interpreting his CCC measurements. We also thank Philip Bransford and Hans-Jörg Jerabek for fruitful discussions.

*These authors contributed equally

Chapter Summary

The three-dimensional genomic architecture plays an important role in the regulation of gene expression through physical interactions of distant genomic sites. Chromosome conformation capture (CCC) techniques attempt to resolve the conformation of a genetic sequence or an entire genome in a population of fixed cells in order to understand (epi)genetic interactions that determine genomic architecture.

Here, we present a computational approach for genome conformation reconstruction, which generates the (unique) maximum entropy ensemble of conformations consistent with experimental contact probabilities measured by CCC techniques. Application of the procedure to the human HoxA cluster in differentiated and undifferentiated cells demonstrates that it is organized into multiple chromatin loops due to the formation of distinct contacts between HoXA genes. The present approach is equally applicable to fluorescence-based data as obtained, for example, from fluorescence in situ hybridization (FISH).

12.1 Introduction

Interphase chromosomes are intricately organized within the confinement of the nucleus. On the one hand, the chromatin fiber has to be highly compacted to fit into the confined space of the cell nucleus. On the other hand, chromosomes must be folded in a "accessible" way in order to allow for vital cell functions such as DNA transcription, replication and repair [93, 195, 423–425].

As discussed in chapter 2, the conformational state and flexibility of the chromatin fiber has an important impact on gene expression, which is modulated by colocalization of enhancers and their genomically distant target genes by the forming chromatin loops [44, 56,87,165,173,174,423,426]. In fact, chromosome organization is governed by the interplay between regulatory elements and their target genes, thus forming a highly complex 3D regulatory network [44, 80, 423, 427–431].

This is a field, in which new molecular and genome-wide experimental techniques can give insight into the folding principles of chromosomes, while three-dimensional modeling of the chromatin fiber can help to relate the experimental observations to build a model and to quantitatively test our predictions for chromosome organization.

The most widely used molecular method to probe the spatial folding of chromatin is chromosome conformation capture (3C) [432] which determines the relative frequency with which pairs of genomic loci are in direct physical contact (for details on the experimental methods see [432]). In an effort to map chromatin interactions at a genome-wide scale, several detection methods including chromosome conformation capture on-chip [129] or circular chromosome conformation capture [130] (4C), chromosome conformation capture carbon copy (5C) [131] and Hi-C technology [87] have been developed. However, while two-dimensional data analysis of CCC interaction maps is able to identify prominent changes in DNA contacts, this approach does not fully integrate the spatial aspect of 3D chromatin folding and information is "lost". For this purpose, we propose a computational approach that is founded on the principle of maximum entropy [163, 433, 434] and a wormlike chain model of the chromatin fiber.

Other modeling approaches have been applied to the immunoglobulin heavy chain [78], the Hox clusters [44, 427], the yeast genome [345] as well as the α-globin gene domain [13] and represent the chromatin fiber by a linear three-dimensional space curve defined by a certain number of points. These points are moved as to force them to conformations that minimally violate the imposed distance constraints which are determined by the assumption that measured chromatin interactions frequencies can be translated into spatial distances between interacting sites.

Our method offers the advantage that it treats the chromatin fiber as a worm-like chain (WLC), which is a concept that has been successfully applied to modeling of chromatin and DNA [173, 174, 435, 436]. Thus, by taking into account the connectivity of the chromatin fiber we are able to explicitly account for the topological complexity (including topological constraints) of its three-dimensional fold which is so far missing in other modeling approaches.

Moreover, the inverse Monte Carlo method applied in this work has been used successfully in structure determination of conformationally flexible biomacromolecules (based on NMR, X-ray or neutron scattering data) [437–439]. In light of the stochastic nature of the interactions and the huge fluctuations in sub-nuclear loci positioning [423] the advantage of the inverse procedure is the generation of a unique, thermodynamically consisted ensemble which is necessary for describing the correct distribution of conformations independent of the choice of adjustable parameters and simulation protocols employed [440, 441].

The goal of this work is to evaluate the utility of the inverse Monte Carlo (IMC) method for resolving chromosome architecture. For this purpose, we study several test systems for which the exact conformational ensemble of the chromatin fiber is known a priori [437,438]. We employ the Metropolis Monte Carlo method [146,147,163] to generate sets of "experimental data", i.e. chromatin fiber conformations from which we can then "measure" the interaction frequency maps. The utility of the IMC procedure is then evaluated by employing solely the previously generated interaction frequency map to solve for the ensemble of chromatin conformations. The accuracy of the resulting conformational ensemble is assessed by not only comparing the respective contact maps but also physical quantities such as the chromatin fiber's distance from its center of mass.

In a subsequent step, we apply the IMC approach to experimental CCC data for the human Homeobox A cluster [44] which encodes two oncogenes, HoxA9 and HoxA10, and plays an important role in promoting cellular proliferation of leukemia cells [442–445]. Moreover, HoX genes play a major role in the structural organization of organisms, especially during embryonic development [446, 447]. We used 5C-data [44] of the HoXA cluster in its undifferentiated as well as in its differentiated state as a first biological application for our computational conformation reconstruction approach and compare the results to the current biological interpretation of those data sets [44, 427].

12.2 Methods and Modeling

The inverse Monte Carlo (IMC) procedure computes a maximum entropy conformational ensemble of a given polymer class with some given constraints provided in the form of experimental (or simulated) distribution functions [440, 441]. It has already been applied to a single model protein [437], but it is a general method of structural modeling based on experimental data. For a formulation of such a macromolecular structure determination problem in a statistical mechanical framework, i.e. in the semigrand canonical ensemble, see Ref. [438].

Here, we apply the IMC procedure to compute a maximum entropy conformational ensemble of chromatin conformations. The given constraints are contact frequencies of pairs of genetic sites along the backbone of the chromatin fiber which have been determined by CCC measurements [44]. Fig. 12.1 illustrates the individual steps of the IMC scheme which will be explained in the following.

12.2.1 Chromatin Modeling

We used a coarse-grained approach to model chromatin on the length scale of hundreds of kbps up to Mbps. Coarse-grained models have the advantage of matching the features of systems on large length scales irrespective of their fine structure details [146, 147, 174].

On these length scales, chromatin can be modeled using a two parameter worm-like chain (WLC) model [448, 449] which has already been discusses in chapter 3. The WLC model is a standard model in polymer physics that represents polymers of contour length L which are continuously flexible with persistence length l_p.

The persistence length is a measure for the local stiffness of the chain. It might also be seen as the typical length scale over which the backbone loses information about its direction due to thermal fluctuations [139]. Notably, while the contour length remains constant during the inverse procedure, the persistence length is not a fixed parameter. Instead, it is iteratively adjusted by the potentials that drive the colocalization of genetic sites in space. However, we have to chose a start value for the persistence length at the beginning of the inverse simulation, which is then allowed to change locally in the course of the iterations.

In a discrete polymer representation, the bending energy H_b can be expressed as

$$H_\mathrm{b} = \frac{l_\mathrm{p}}{2b} \sum_{i=1}^{N-1} (\mathbf{u}_{i+1} - \mathbf{u}_i)^2. \tag{12.1}$$

Here, b is the bond length and $\mathbf{u}_i = \left(\frac{\mathbf{R}_i - \mathbf{R}_{i-1}}{|\mathbf{R}_i - \mathbf{R}_{i-1}|} + \frac{\mathbf{R}_{i+1} - \mathbf{R}_i}{|\mathbf{R}_{i+1} - \mathbf{R}_i|} \right) \Big/ \left| \frac{\mathbf{R}_i - \mathbf{R}_{i-1}}{|\mathbf{R}_i - \mathbf{R}_{i-1}|} + \frac{\mathbf{R}_{i+1} - \mathbf{R}_i}{|\mathbf{R}_{i+1} - \mathbf{R}_i|} \right|$ is a discrete realization of $\delta \mathrm{R}(s)/\delta s$, the unit tangent vector at arclength s, where $\mathrm{R}(s)$ is the position vector [249]. N is the total number of monomers in the chain. The WLC implementation is tested by comparing the analytical solution of the mean squared end-to-end distance $\langle R^2_\mathrm{end} \rangle = 2Ll_\mathrm{p} - 2l_\mathrm{p}[1 - \exp(-L/l_\mathrm{p})]$ to the numerical solution of the WLC simulation as shown in Fig. 3.4 in chapter 3.

The parameter choice of the chromatin model will in general depend on the length-scale on which conformational properties are of interest. Here, we aim at investigating conformational properties at the kbp-scale but the IMC approach can equally be applied to the nm-scale, provided CCC data are also available on this scale.

We assume a packing density of 30 bp/nm and an initial persistence length of 150 nm for the chromatin fiber [172, 346] similar to the parameters chosen to model the Rabl organization of yeast chromosomes in chapter 8. While the precise packing density and persistence length of chromatin is likely to vary across cells, cellular states, as well as locally within any given cell depending on its biological state, we choose values that are between the limiting cases of bare DNA (3 bp/nm) and a highly condensed 30 nm fiber in agreement with [131, 171, 172, 172–174].

The 150 kbp HoxA genomic region studied by 3C and 5C measurements in [44] is modeled by a 180 kbp chromatin fiber chain, where it is embedded between two 15 kbp

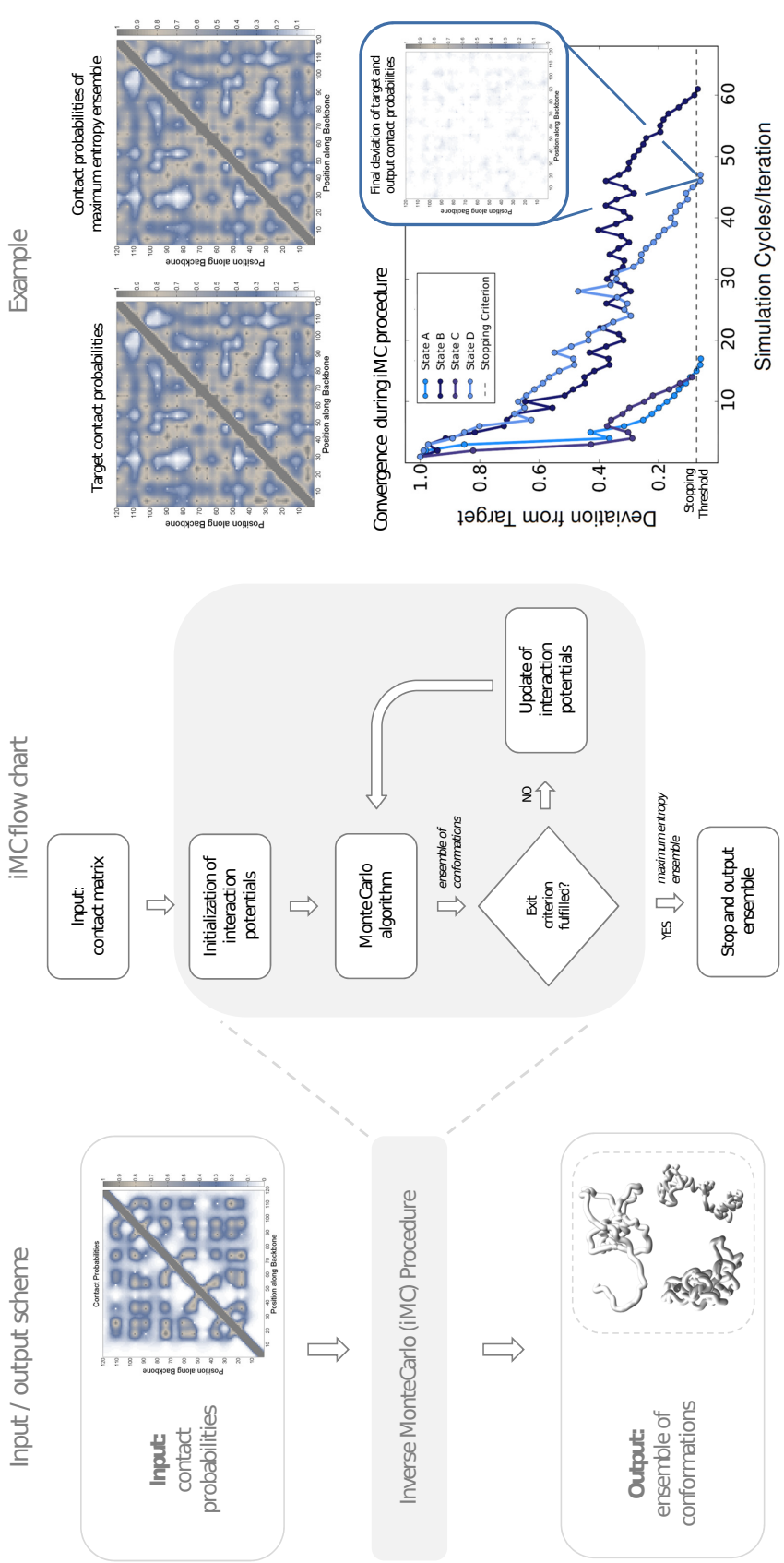


Figure 12.1: Flow chart of the IMC procedure. The IMC algorithm solves for a maximum entropy ensemble of conformations consistent with experimentally measured contact probability constraints between specific genetic sites i and j along the chromatin fiber. After a first initialization step, a Metropolis algorithm is used to generate worm-like chain conformations with interactions potentials along the chain to drive colocalization. After every simulation cycle the proximity of the contact matrix of the current (simulated) ensemble is checked with respect to the "target" contact matrix. If the convergence criterion is satisfied the algorithm stops. Otherwise the interaction potentials along the worm-like chains will be updated and the procedure and starts all over again until it converges. Image adapted from [175].

"fiber ends". Consequently, the whole 180 kbp long fiber is represented by a worm-like chain of $N = (180\text{kbp} * 1\text{nm} * 8)/(30\text{bp} * 150\text{nm}) = 320$ segments with segment length $b = (180\text{kbp} * 1\text{nm})/(30\text{bp} * 320) = 18.75$ nm.

12.2.2 Iterative Adjustment of Pair-Wise Contact Potentials Drives Colocalization of Genomically Distant Sites

Measurements based on CCC technologies provide data in the form of average pair-wise interaction frequencies between specific genetic sites. However, the developed inverse Monte Carlo method does not solve for interaction frequencies but for contact probabilities, which means that the experimentally determined interaction frequency map has to be converted to a target contact probability map. Notably, we are discussing contact probabilities in the following, while the conversion of interaction frequencies to contact probabilities is explained in subsection 12.2.3.

In order to account for the given contact probability constraints we employ pair-wise square-well potentials which drive colocalization of genomically distant, interacting sites. A contact is defined whenever two sites i and j come closer than a threshold distance $d_{\text{threshold}}$, which is taken to be $d_{\text{threshold}}^{\text{HoxA}} = 100$ nm for the HoxA cluster. Similar values have been used in other conformation reconstruction approaches [13]. The contact threshold distance equals the width of the square-well potentials.

The goal of the IMC procedure is to iteratively adjust the pair-wise square-well potentials along the self-interacting chromatin fiber [196] in such a way as to generate an ensemble of chromatin conformations whose contact probabilities match the experimentally measured ones. At the beginning of the inverse procedure, the pair-wise square-well potentials are initialized to zero. Notably, the pair potentials are not "real" physical potentials but artificial. They effectively account for all the constraints that are hidden in the given contact probabilities such confinement of the chromatin fiber by the cell nucleus, protein interactions, the state of the cell cycle, etc. Adjusting the square-well potentials changes chromatin properties such as the looping probability or the persistence length. The depths of the square-well potentials between two interacting sites i and j will be denoted by U_{ij} in the following. They can have positive or negative values, i.e. repulsive potentials when two sites need to be separated from each other or attractive potentials when two sites need to come into contact.

Each simulation cycle/iteration consists of a standard Metropolis Monte Carlo simulation [146, 147, 163] generating 10^4 independent conformations as illustrated in Fig. 12.1. After each simulation cycle/iteration a contact probability matrix E^I is calculated from the ensemble of chromatin fiber conformations and compared to the target contact probability matrix E^{target} in order to update the effective pair-wise square well interaction potentials between interacting chromatin sites.

12.2.3 Converting Interaction Frequencies to Contact Probabilites and Coarse-Graining of HoxA Target Matrix

The experimentally measured interaction frequency map, denoted by $\mathcal{E}^{\text{exp}}$, has to be converted to a contact probability map to form the "target" contact probability matrix E^{target} that the inverse approach can solve for. In the following E_{ij}^{target} and $\mathcal{E}_{ij}^{\text{exp}}$ denote entries for the interacting sites i and j and $\text{Max}(\mathcal{E}^{\text{exp}})$ denotes the maximum entry of the experimental interaction frequency map $\mathcal{E}^{\text{exp}}$.

The conversion between contact frequency maps and contact probability maps is based on the assumption that the largest experimental interaction frequency, $\mathrm{Max}(\mathcal{E}^{\mathrm{exp}})$, corresponds to an interaction between restriction fragments that are genomically so close to each other that they always form a contact, such that their contact probability is $p_{\mathrm{contact}} \approx 1$. Then, one can normalize the other interaction frequencies according to

$$E_{ij}^{\mathrm{target}} = \mathcal{E}_{ij}^{\mathrm{exp}}/\mathrm{Max}(\mathcal{E}^{\mathrm{exp}}). \tag{12.2}$$

This assumption is true for the diagonal elements of the experimentally determined interaction frequency matrices, i.e. for pairs of sites along the backbone of the chromatin fiber. In fact, these sites always show the highest interaction frequencies [13,44,131], thus confirming our conversion approach.

Additionally, computational feasability of the inverse procedure requires a coarse-graining of the experimentally determined HoxA interaction frequency map for differentiated and undifferentiated cells. The coarse-graining reduces the original interaction frequency map of 42×42 sites to 21×21 interaction sites by averaging over four adjacent contact interaction pairs therewith preserving connectivity constraints and contact information. Additionally, interacting sites are reduced from 42 to 21 by averaging over two (genomic) neighbor sites along the HoxA region.

12.2.4 Update of the Square-Well Potentials

Let E^I be the contact probability map calculated for the ensemble of chromatin fiber conformation at iteration I and E_{ij}^I the contact probability for the interacting sites i and j. Then, the update of the pair-wise square-well potential U_{ij}^{I+1} at the iteration step $I+1$ can be computed as

$$U_{ij}^{I+1} = \begin{cases} U_{ij}^I + f_{\mathrm{damping}} \times (E_{ij}^I - E_{ij}^{\mathrm{target}}) & \text{if } |E_{ij}^I - E_{ij}^{\mathrm{target}}| > 1.3 \times \mathrm{TOL}, \\ U_{ij}^I + f'_{\mathrm{damping}} \times (E_{ij}^I - E_{ij}^{\mathrm{target}}) & \text{else.} \end{cases} \tag{12.3}$$

Here, $f_{\mathrm{damping}} = 2$ and $f'_{\mathrm{damping}} = 0.5$ are damping factors to avoid excessive oscillations of the square-well potentials depths. After all potential updates have been performed, a new simulation cycle/iteration starts and this goes on until the simulated contact probability matrix E^I converges to the target contact probability matrix E^{target}.

12.2.5 Convergence Criterion

A measure for the proximity of E^I and E^{target}, based on the $p = 1$ matrix norm, is defined as

$$C = \sum_{ij} |E_{ij}^{\mathrm{target}} - E_{ij}^I|. \tag{12.4}$$

The inverse simulation is stopped if

$$C \leq \sum_{ij} \mathrm{TOL} \tag{12.5}$$

holds for four simulation cycles/iterations. TOL represents the tolerated deviation from the target contact matrix entries. CCC experiments have various sources of errors in the final data sets, which have been discussed in chapter 2. These errors are not easy to estimate but of the order of tens of percent of the experimental data [44]. Here, the

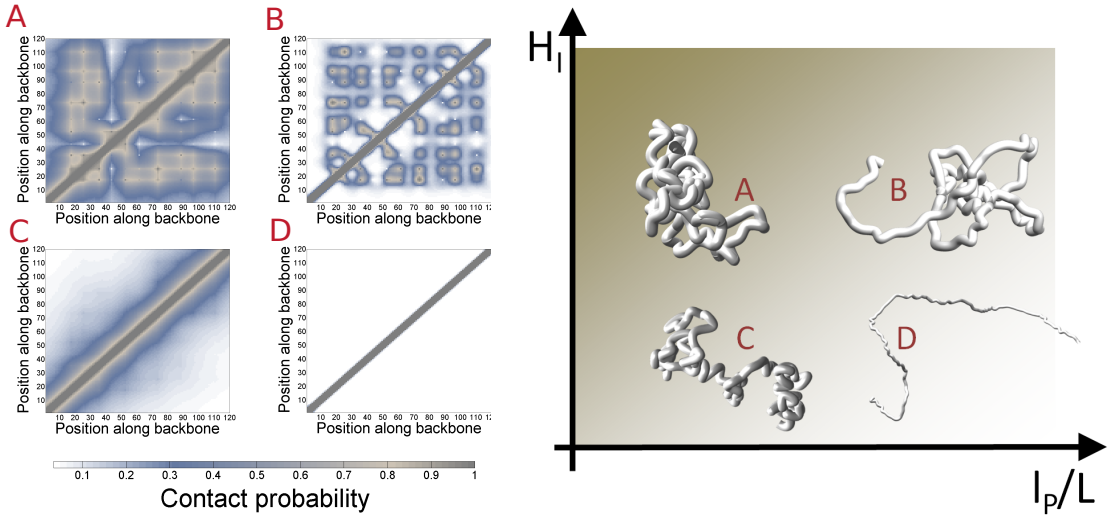


Figure 12.2: Conformational phase space of the four polymer test cases. The color gradient in the background illustrates the increasing compactness with higher interaction energy H_I and lower bending energy l_p/L. There is a competing interplay between bending energy (i.e. local chain stiffness) H_b and self-interaction energy H_I. The IMC procedure is able to provide a maximum entropy ensemble for different energy states in the conformational phase space. This demonstrates its robustness against variations of input data, e.g. euchromatin versus heterochromatin. Image adapted from [175].

precision TOL is taken to be in the range between $5\%\langle E^\mathrm{target}\rangle$ and $20\%\langle E^\mathrm{target}\rangle$, where $\langle E^\mathrm{target}\rangle = \sum_{ij} E_{ij}^\mathrm{target} / \sum_{ij} 1$. Notably, if $E_{ij}^\mathrm{target} < \mathrm{TOL}$, then the contact interaction between the sites i and j is not solved for by the inverse method. The "Example" in Fig. 12.1 illustrates the good agreement between the target probability matrix E^target and the iteratively found probability map E^I after the simulation has converged.

12.3 Validation of the Inverse Monte Carlo Method

12.3.1 Convergence for Different Conformational States in Phase Space

First of all, we validate the IMC procedure by showing its convergence for exemplary points in configurational phase space that is spanned by the chain's bending energy H_b and the self-interaction energy H_I determined by the square-well potentials along the polymer. Figure 12.2 shows a qualitative sketch illustrating the impact of the competing interplay between bending and interaction energy on the polymer organization for the four test case.

In the limiting case of high chain flexibility and low self-interaction energy (either due to few interacting sites or due to low interaction potentials) the polymer chain behaves as a self-avoiding random walk on length scales that are large compared to its persistence length, ultimately resulting in a "swollen" random coil. When the self-attraction of the polymer chain exceeds its bending rigidity, a compact globule emerges. Starting from the latter, a sufficiently high bending stiffness leads to a less compact globule, and eventually to the formation of a micro-structured polymer with co-localized sites and loops attached to it. In contrast, a rod-like polymer organization emerges if the bending energy is strongly increased and dominates the overall energy of the self-interacting polymer chain resulting in highly "stretched" conformations.

Table 12.1: Summary of the parameters used for the test cases in Fig. 12.2. It is $U = \{-4, -2, 0, 1\}$, $U' = \{-0.4, -0.2, 0, 0.1\}$ and $U'' = \{-40, -20, 0, 10\}$.

A	B	C	D
$l_\mathrm{p}/L \to 0$, U	$l_\mathrm{p}/L = 0.0625$, U''	$l_\mathrm{p}/L \to 0$, U'	$l_\mathrm{p}/L = 0.125$, U

In order to show the IMC procedure's convergence in phase space we employ a forward Metropolis Monte Carlo simulation to generate ensembles of self-interacting chain conformations for different ratios of bending energy H_b and self-interaction energy H_I. Ten interacting sites are distributed randomly along the chains of $N = 120$ monomers with the condition that neighboring sites i and j satisfy $|i-j| > 3$. The square-well potential depths for each pair of interacting sites are chosen randomly from the sets $U = \{-4, -2, 0, 1\}$, $U' = \{-0.4, -0.2, 0, 0.1\}$ or $U'' = \{-40, -20, 0, 10\}$. Tab. 12.1 summarizes the parameters used for the four test cases. Based on the ensemble of polymer conformations generated by the forward Metropolis Monte Carlo simulation contact probabilities are computed for all sites along the polymer chain. In a subsequent step, the contact probabilities of the interacting sites are assumed to be "target" contact probabilities used as input for the inverse approach. The utility of the IMC procedure is then evaluated by employing solely the previously generated contact probability map to solve for an ensemble of conformations.

The accuracy of the iteratively found conformational ensemble is assessed by comparing the converged contact probability map for all sites of the polymer (not just the interacting ones) with the "target" map (equally based on all sites) as is shown exemplarily for one test case in the "Example" of Fig. 12.1. Besides the visual inspection of both contact maps the convergence of the inverse procedure is illustrated by the %-deviation between target and iteratively determined contact probabilities in Fig. 12.1. The iteration process until the convergence criterion is satisfied is also exemplified in the same figure.

12.3.2 Reconstruction of "Non-Measured" Contact Probabilites

After having shown that the IMC procedure converges for different ratios of H_b and H_I in conformational phase space we investigate the method's robustness against missing contact information. This is usually the case for CCC data since CCC technologies are not able to determine interaction frequencies for every pair of interaction sites along the chromatin fiber but only allow for the investigation of a limited number of pair-wise interactions. Thus, it is necessary to understand the impact that a reduced target probability matrix has on our method's performance in "resolving" the fiber's structural properties. In particular, with respect to a 3D regulatory interaction network underlying chromosome organization [423] it is important to understand the impact that missing interaction information, e.g. for a major regulator ("transcription factory") in the network, might have on correctly modeling the spatial organization of the chromatin fiber.

We employ a forward Metropolis Monte Carlo simulation to generate an ensemble of self-interacting chain conformations. For this purpose, 20 interacting sites are distributed randomly along the chain of $N = 120$ monomers with the condition that neighboring interaction sites i and j satisfy $|i - j| > 3$. The pair-wise square-well potential depths under which the polymer chain evolves in the forward simulation, mimicking a regulatory interplay, are chosen randomly from $U = \{-4, -2, 0, 1, 2\}$ and assigned randomly to all the 200 pairs of interacting sites.

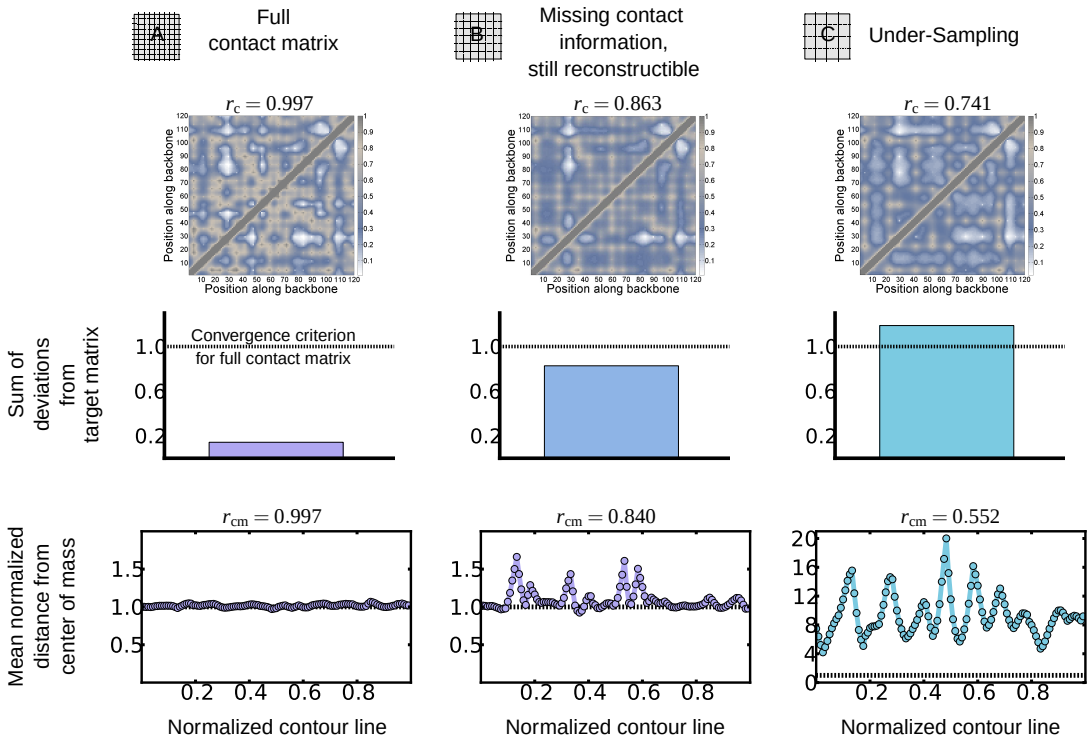


Figure 12.3: Evaluation of the IMC procedure's performance (robustness) when contact information is missing, which is usually the case for chromosome conformation capture data. In order to investigate the effect of missing contact information the IMC procedure solves for (A) the full target matrix $E^{\text{target}}_{20\times 20}$ as well as for "reduced" target matrices (B) $E^{\text{target}}_{17\times 17}$ and (C) $E^{\text{target}}_{14\times 14}$ with only 17×17 and 14×14 contact entries, respectively. The first row illustrates the contact probability matrices emerging from the inverse procedure as well as their Pearson correlation coefficient with respect to the full 20×20 "target" matrix. The second row shows the normalized sum of deviations from the overall "target" probability matrix based on the forward Monte Carlo ensemble, $\sum_{jk} |p^I_{jk} - p^{\text{target}}_{jk}| / \sum_{kj}$ TOL. Here, TOL is the maximally allowed deviation from the target probabilities as explained in subsection 12.2.5. Note that we do not just sum over the contact pairs i and j that the inverse method is solving for, but instead the sums contain all pairs $1 \leq k < N$ and $1 \leq k < N$ along the polymer chain. The third row shows each site's normalized mean distance from the polymer chain's center of mass thus comparing the target and IMC system's structural properties. The normalization is performed with respect to the forward Monte Carlo ensemble. It is possible to reconstruct contact and structural probabilities that have not been measured in the experiment when the density of the given contact data is high enough as in case (B). Image adapted from [175].

Based on the so generated ensemble of polymer conformations a full 20×20 "target" contact probability map $E^{\text{target}}_{20\times 20}$ is determined and used as input for the inverse approach. In order to investigate the effect of missing contact information we also apply "reduced" target matrices $E^{\text{target}}_{17\times 17}$ or $E^{\text{target}}_{14\times 14}$ with only 17×17 or 14×14 contact entries, respectively, to solve for an ensemble of conformations with the IMC procedure.

Fig. 12.3 illustrates the performance of the IMC procedure based on the full contact matrix $E^{\text{target}}_{20\times 20}$, the still reconstructible case of few missing contacts $E^{\text{target}}_{17\times 17}$ as well as the case where under-sampling due to too little contact information occurs.

The Pearson correlation coefficient r is a measure for the linear dependence of two variables. Let p^{target}_{kj} be the contact probability of the pair of sites $1 \leq k < N$ and

$1 \leq j < N$ computed for the forward Monte Carlo ensemble and p^I_{kj} be the contact probability of the same sites $1 \leq k < N$ and $1 \leq j < N$ determined for the inverse Monte Carlo ensemble after convergence. Here, we are interested in the correlation between the contact probabilities p^{target}_{kj} and p^I_{kj}, respectively. Fig. 12.3 shows that based on the contact probability matrix $E^{\text{target}}_{20\times 20}$ the IMC approach is able to solve for an ensemble of conformations whose contact probabilities have a Pearson correlation coefficient of $r_{\text{c}} = 0.997$ with the target probability map. Decreasing the number of contact probability constraints the inverse approach is solving for leads to a decreased resolution of the spatial chromatin organization. This is reflected in a decreasing Pearson correlation coefficient $r_{\text{c}} = 0.863$ and $r_{\text{c}} = 0.741$ for the "targets" $E^{\text{target}}_{17\times 17}$ and $E^{\text{target}}_{14\times 14}$, respectively.

Based on the forward and inverse Monte Carlo ensembles each site's mean distance from the polymer's center of mass can be computed in order to assess the IMC procedures ability of correctly reproducing physical properties given missing contact information. Fig. 12.3 shows that based on the full contact probability matrix $E^{\text{target}}_{20\times 20}$ the IMC approach is able to correctly reproduce the behavior each site's mean distance from the center of mass leading to a Pearson correlation coefficient between the "target" and the simulated system of $r_{\text{cm}} = 0.997$. The case of $E^{\text{target}}_{17\times 17}$ as input to the inverse method still leads to a good agreement between target and iteratively determined structure as can been seen from the correlation coefficient $r_{\text{cm}} = 0.840$. A further decrease of input contact information leads to a loss of convergence to "target" structural properties as can be seen by the low person correlation coefficient of $r_{\text{cm}} = 0.552$, while the respective contact map still shows a relatively high correlation of $r_{\text{cm}} = 0.741$ in this case.

12.3.3 Convergence to Unique Pair Potential Matrix

The question has to be raised whether the inverse Monte Carlo procedure generates a "unique" set of interaction potential depths that reproduce the experimentally observed pair-wise contact probabilities. R.L. Henderson has proved that for classical or quantum fluids with only pairwise interactions the pair potential which gives rise to a given radial distribution function is unique up to a constant. For a brief note on Henderson's short proof the reader is referred to Refs. [440] and [450].

However, the uniqueness theorem of Henderson is strictly valid with respect to the radial distribution function, which contains information at all values of $g(r)$ for all pairs considered, whereas, in this work, we do not solve for the radial distribution function but for contact probability constraints.

While we cannot show uniqueness by a rigorous mathematical proof, we can address this issue for the biological system studied in the next section: the 150 kbp undifferentiated human Homeobox (Hox) A cluster. To numerically tackle the question of uniqueness for the maximum entropy ensemble generated in section 12.4, we ran distinct inversion simulations using different random starting conformations and random number sequences. In each case the resulting potential matrix and consequently the contact probability map produced was identical to within statistical uncertainty in the stochastic Monte Carlo simulation as shown in Fig. 12.4.

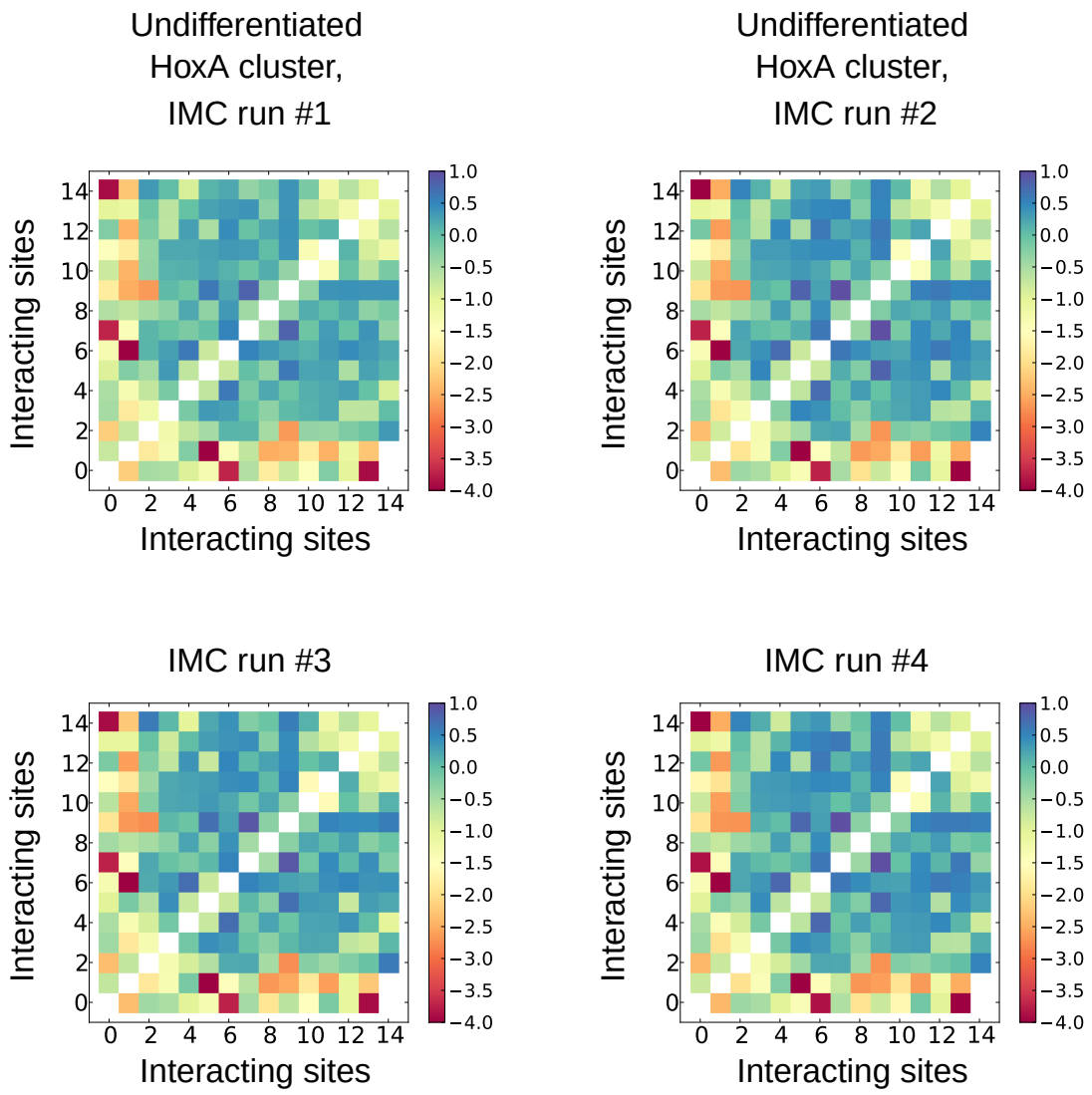


Figure 12.4: To test the uniqueness of the maximum entropy ensemble generated for the undifferentiated HoxA cluster, we ran distinct inversion simulations using random starting conformations and random number seeds. In each case the resulting potential matrix shown here and consequently the contact probability map were identical to within statistical uncertainty in the stochastic Monte Carlo simulation. Image adapted from [175].

12.4 Application of the IMC Method to the Human HoxA Cluster

12.4.1 Biological Relevance

Having evaluated the IMC method's performance for resolving the spatial organization and physical properties of test polymer systems we apply the procedure to experimental CCC data of the human Homeobox (Hox) A cluster schematically illustrated in Fig. 12.5. 3C and 5C techniques have been used to detect pair-wise interaction frequencies in the human HoxA cluster for differentiated and undifferentiated human acute monocytic leukemia

Figure 12.5: Schematic representation of the 150 kbp human HoxA gene cluster on chromosome 7 in a 3' (HoxA 1) to 5' (HoxA 13) orientation. Genes are indicated by arrows, where the left facing of the arrows indicates the direction of transcription. Image adapted from [44].

(THP1) cells [44, 427].

The HoxA cluster is located on human chromosome 7 and encodes transcription factors, which play a major role during development, when they regulate the formation of limbs and genitalia [44, 427]. Additionally, the HoxA cluster encodes two oncogenes, HoxA9 and HoxA10, which are overexpressed in THP1-cells [442–445].

It is known that undifferentiated THP1 cells express high levels of 5' end HoxA genes [44], which are partially repressed following differentiation [44, 451]. Dostie and coworkers have verified that HoxA genes are correctly regulated under their experimental conditions by determining steady-state mRNA levels with quantitative real-time PCR [44].

Since inappropriate gene expression is associated with a variety of illnesses, such as cancer, [452] and since spatial chromatin organization is an important mechanism for regulating gene expression [44, 87, 131, 172, 198, 423, 427] the modeling of chromatin architecture of the HoxA cluster might give insight into potential mechanisms underlying transcriptional regulation.

12.4.2 Convergence to HoxA Cluster Target Probabilites

Convergence of the inverse approach is checked by measuring the proximity between the target contact probability map E^{target} (based on [44]) and the contact probability matrix E^I that the IMC procedure has solved for according to subsection 12.2.5. Fig. 12.6 shows $(E^I_{ij} - E^{\text{target}}_{ij})/E^{\text{target}}_{ij}$ for all interacting pairs i and j for both the undifferentiated and differentiated HoxA cluster.

12.4.3 Interaction Profiles

3C and 5C measurements of the differentiated and undifferentiated human HoxA cluster provide pair-wise interaction frequencies between restriction fragments. While only a restricted number of interactions between genetic sites (restriction fragments) can be measured experimentally, application of the inverse Monte Carlo method allows for the tracking of every contact interaction along the chromatin fiber. Note, however, that the resolution of the numerical approach is limited due to the coarse-graining of the fiber.

Individual interaction profiles for specific sites (restriction fragments in CCC measurements) along the 150 kbp HoxA region illustrate the ability to reproduce quantitatively interaction profiles measured using CCC techniques [128] as is shown in Fig. 12.7. The offset between the measured interaction frequencies and the simulated ones emerges because of two reasons: The 150 kbp HoxA genomic regions is modeled by a 180 kbp chromatin fiber chain, where the HoxA cluster region is embedded between two 15 kbp "fiber ends". Both 15 kbp genomic regions are not constraint by interacting sites since there is no contact information available. As a result they loop out and thus affect the interacting sites at the beginning (RF 47) and end (RF 88) of the HoxA cluster. Additionally, the

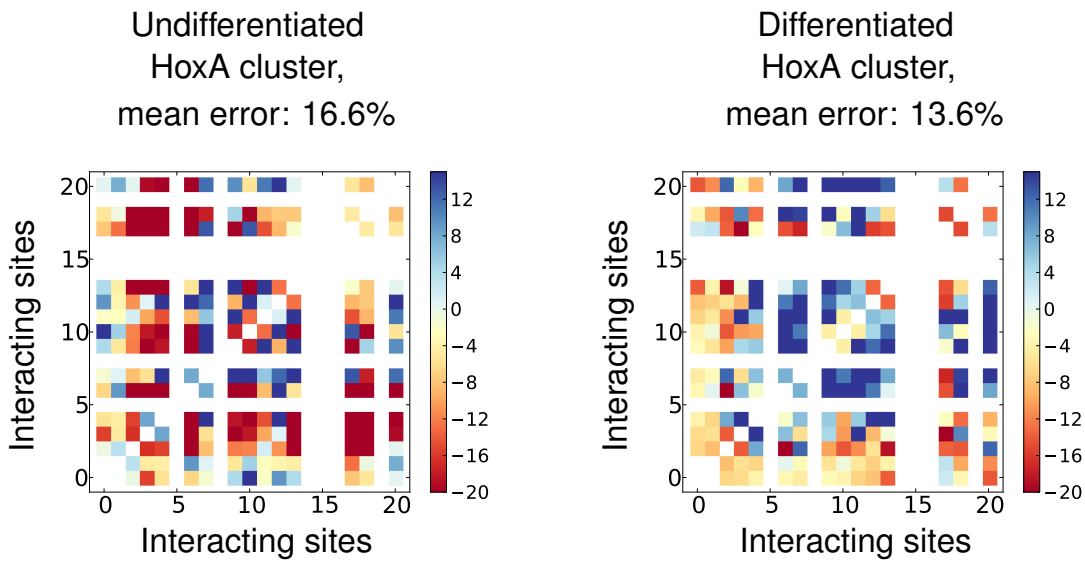


Figure 12.6: The deviation between target and IMC probabilities $(E^{I}_{ij} - E^{\text{target}}_{ij})/E^{\text{target}}_{ij}$ shown for all interacting pairs ij for the undifferentiated and differentiated HoxA cluster. The mean error (in %) is averaged over all deviations. Image adapted from [175].

coarse-graining of the experimental target matrix also reduces the precision with which interaction frequencies of single restriction fragments can be reproduced.

Based on the study of transcriptional activity of the undifferentiated and differentiated HoxA cluster by Dostie et al. as well as our modeling approach, we can try to relate 3D HoxA cluster architecture and gene expression. We find that the transcriptionally silent HoxA genes HoxA1- HoxA5 (fragments 47-50 in [44]) interact very strongly with the entire HoxA region, where the interaction frequency does not quickly decrease with increasing genomic distance in agreement with [44]. This leads to an effective spatial clustering which is also reflected by the radius of gyration. The gyration radius is a measure for the spatial extension of the chromatin fiber. In fact, the gyration radius of the genomic region containing the HoxA genes 1 to 5 is approximately 2.5 times smaller (both for the undifferentiated and differentiated case) than the extension of a non-interacting chromatin fiber (free polymer chain).

Moreover, in undifferentiated cells, the genes HoxA9, HoxA11 and HoxA13 are found to be highly expressed and in differentiated cells these genes are shown to still be partially expressed [44]. Based on the contact interaction constraints, we find the 5' end genomic region to loop away from the cluster (low interaction profile for fixed RF75 in Fig. 12.7) in both cellular states in agreement with [44].

12.4.4 Hierarchical Clustering

Visual inspection of the conformational ensemble generated by the IMC procedure shows that the undifferentiated and differentiated HoxA cluster is organized by the formation chromatin loops [427]. To this end, one might ask the question whether the chromatin fiber assumes specific folding motifs or whether its structural organization is determined

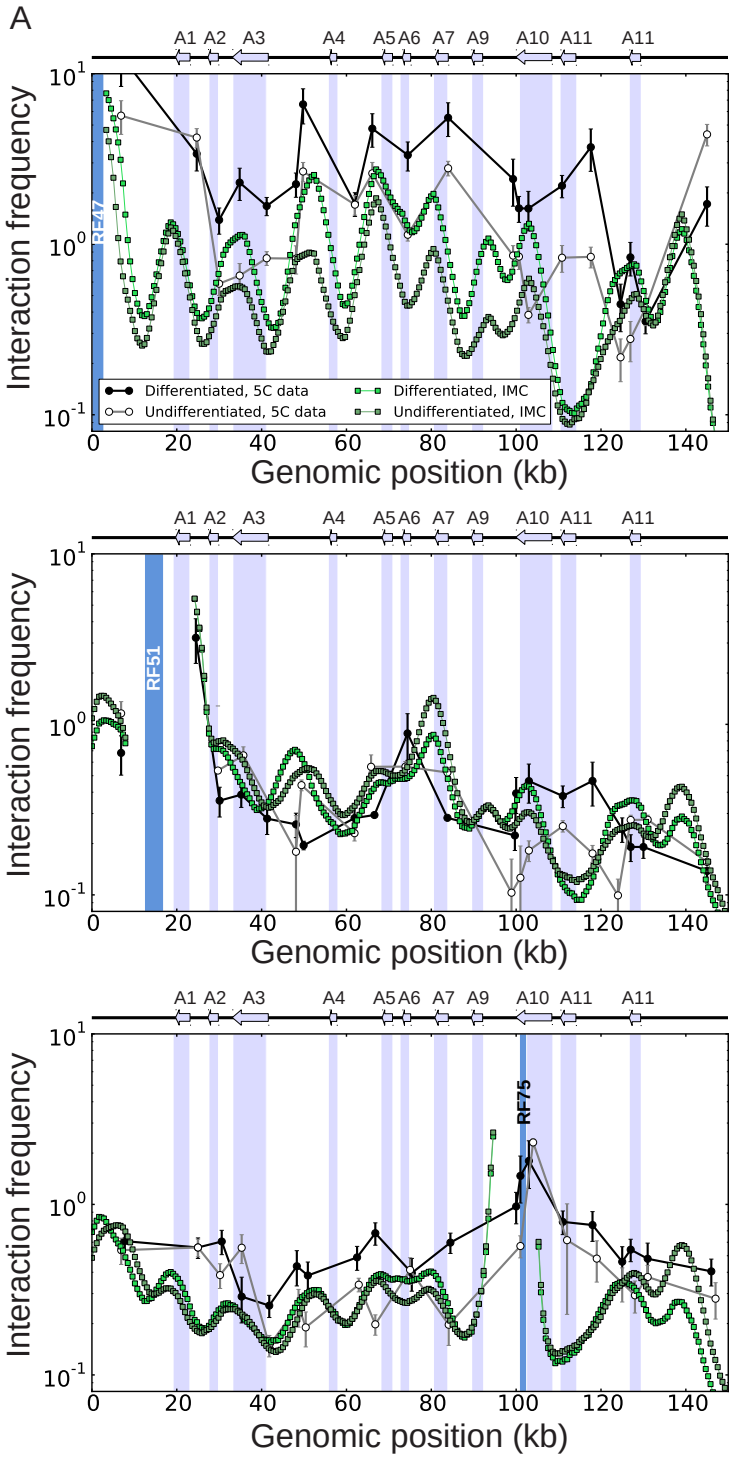


Figure 12.7: Interaction frequency profiles showing the experimentally measured data from Dostie and colleagues [44] (black and white circles) as well as the interaction frequencies based on our modeling approach (dark and light green squares). The dark blue vertical lines indicate the position of the fixed restriction fragments (A) RF47, (B) RF51 and (C) RF75, while the light blue vertical lines illustrate the positions of the HoxA genes. Individual interaction profiles for specific sites (restriction fragments in CCC measurements) along the 150 kbp HoxA region illustrate the ability of the inverse procedure to reproduce quantitatively interaction profiles measured using CCC techniques [128]. Image adapted from [175].

by large fluctuations. In order to gain further insight into its spatial organization we apply a hierarchical clustering approach to 4500 randomly picked conformations in the ensemble. The clustering analysis has been performed by Philipp M. Diesinger.

In order to decide which HoxA conformations should be combined to form a cluster a measure of dissimilarity between conformations is required. This is achieved by the use of an appropriate distance metric: To assign distances between pairs of conformations in the ensemble, we align the HoxA genes using *PyMol* (`http://www.pymol.org/`) and then take the sum of the squared deviations as the distance metric. Notably, all sites which are not HoxA genes are neglected. While other clustering metrics could be used, it is reasonable to cluster conformations by aligning those functional parts that are of interest: the HoxA genes. The so computed distances are the basis for the hierarchical clustering procedure. The conformations are assigned to classes according to the specific type of physical contact between the HoxA genes.interested in here

Structural clustering of the HoxA conformations reveals three different classes for the differentiated HoxA cluster and no distint classes for the undifferentiated HoxA cluster. In fact, the undifferentiated cluster forms a single dominant class of rather compact conformations as illustrated in SI Fig. 12.12. Regarding the differentiated HoxA cluster, Fig. 12.8 shows the overall contact matrices of each class together with the contact matrices for the HoxA genes only. The dendogram from the hierarchical clustering procedure is shown in SI Fig. 12.10. The first class is very small (1% of all conformations), but still significantly different from the other two, and consists of conformations which show a high number of contacts between the genes HoxA 5-HoxA 11. Typical conformations show a compact head region at the end of the HoxA cluster with a trail reqion sticking out of it. The second class is the largest class with 81% of all conformations and contains conformations with very high contact frequencies. The conformations are very compact. Class three is the second largest class (19% of the conformations) and shows a compact region at the beginning of the HoxA system (genes HoxA1 - HoxA4) with an unconstrained tail sticking out of it. "Snapshot" configurations examplify typical conformations for each class in Fig. 12.8.

The dominant structural motif for both the undifferentiated and differentiated HoxA genomic region is the "rosette" motif. This result is also supported by our analysis of the region's loop size distribution in SI Fig. 12.11 which shows that the abundance of intermediate ($\sim 20 - 80$ kb) and large loops ($\sim 80 - 120$ kb) does not rapidly decrease with increasing genomic size. However, this "rosette"-type structure is highly dynamic, since loops form and disolve in accordance to the measured contact probability constraints. In fact, the chromatin fiber does not show the structural stability of proteins, which assume specific folds, but chromatin undergoes strong position fluctuations. Thus, while structural clustering is a tool successfully applied in protein structure analysis, it has only limited applications in the fields of (bio)polymer physics.

12.4.5 Correlations Between Pairs of Interacting Genetic Sites

The maximum entropy conformational ensemble enables the calculation of correlations in structural quantities in addition to mean-field interaction frequencies. An interesting question relates to the correlations in distances between two genomic loci i and j that both interact with the third site k. Given the mean interaction frequencies for the pair ik and the pair jk, these pairs might be correlated or anti-correlated in their physical interactions.

To probe this we measured the Pearson correlation coefficient for distances between pairs of HoxA genes. Distance correlations are positive (correlation) if the distance between

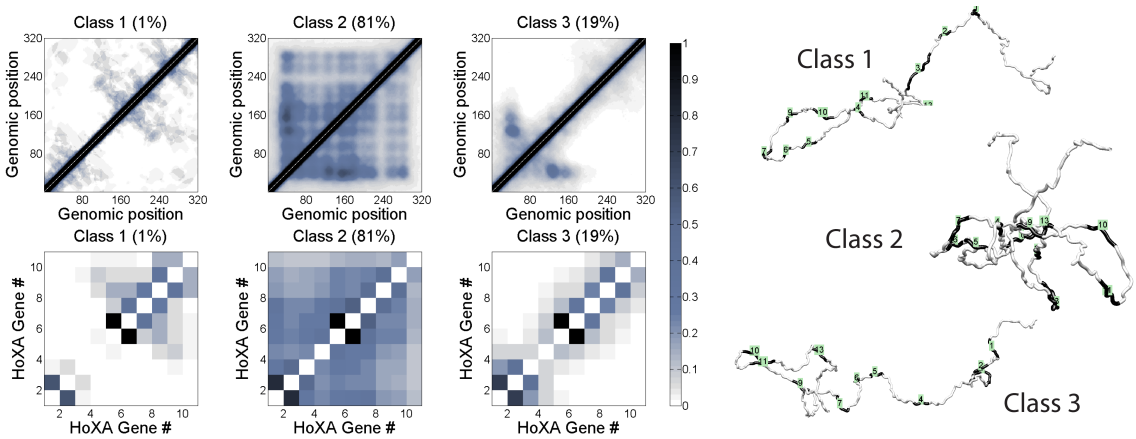


Figure 12.8: Structural clustering reveals three different classes for the differentiated HoxA cluster. The overall contact matrices of each class together with the contact matrices for the HoxA genes only are displayed. The dominant structural motif is the "rosette" motif. "Snapshot" configurations examplify typical conformations for each class. Image adapted from [175].

sites i and k is small when the distance between another pair j and k is small, too. Consequently, distance correlations are negative (anticorrelation) if the distance between sites i and k is small when the distance between another pair j and k is large.

Fig. 12.9 shows the pair distance correlations for the free polymer chain as well as for the undifferentiated and differentiated HoxA cluster. The Pearson correlation coefficient is computed for the distance between the fixed genetic site pair (A) A1-A5, (B) RF47-A13 and (C) A9-A13 with respect to all other distances between HoxA gene pairs. While the undifferentiated and differentiated HoxA cluster pair correlations are similar, which is consistent with our experimental and numerical results obtained so far, the pair correlations of the HoxA cluster region are different from those of a free polymer chain. In a free polymer fiber, the distance correlations emerge due to chain connectivity. In case of the HoxA cluster, the influence of clustering becomes apparent in the non-vanishing, positive correlation coefficients between pairs of genetic sites that are genomically distant, as illustrated in Fig. 12.9.

As opposed to the inverse Monte Carlo method, which takes into account the connectivity and the dynamic nature of the chromatin fiber, neither CCC measurements nor the existing computational modeling approaches [13, 44, 78, 345, 427] are able to obtain information about distance correlations so far, which, however, might play an important role in coexpression/corepression of genes [44, 423, 427, 436].

12.5 Conclusions

Genomes are are organized into dynamic three-dimensional networks of physical chromatin contacts controlling gene expression [44, 80, 423, 427–431]. Therefore, mapping the functional and the spatial dimension of genomes is essential to fully identify mechanisms involved in the regulation of genes.

Here, we present a computational tool to unravel chromosome architecture thus complementing experimental evidence based on CCC or FISH/high-resolution microscopy measurements. A detailed confrontation of biochemical (CCC techniques) and high-

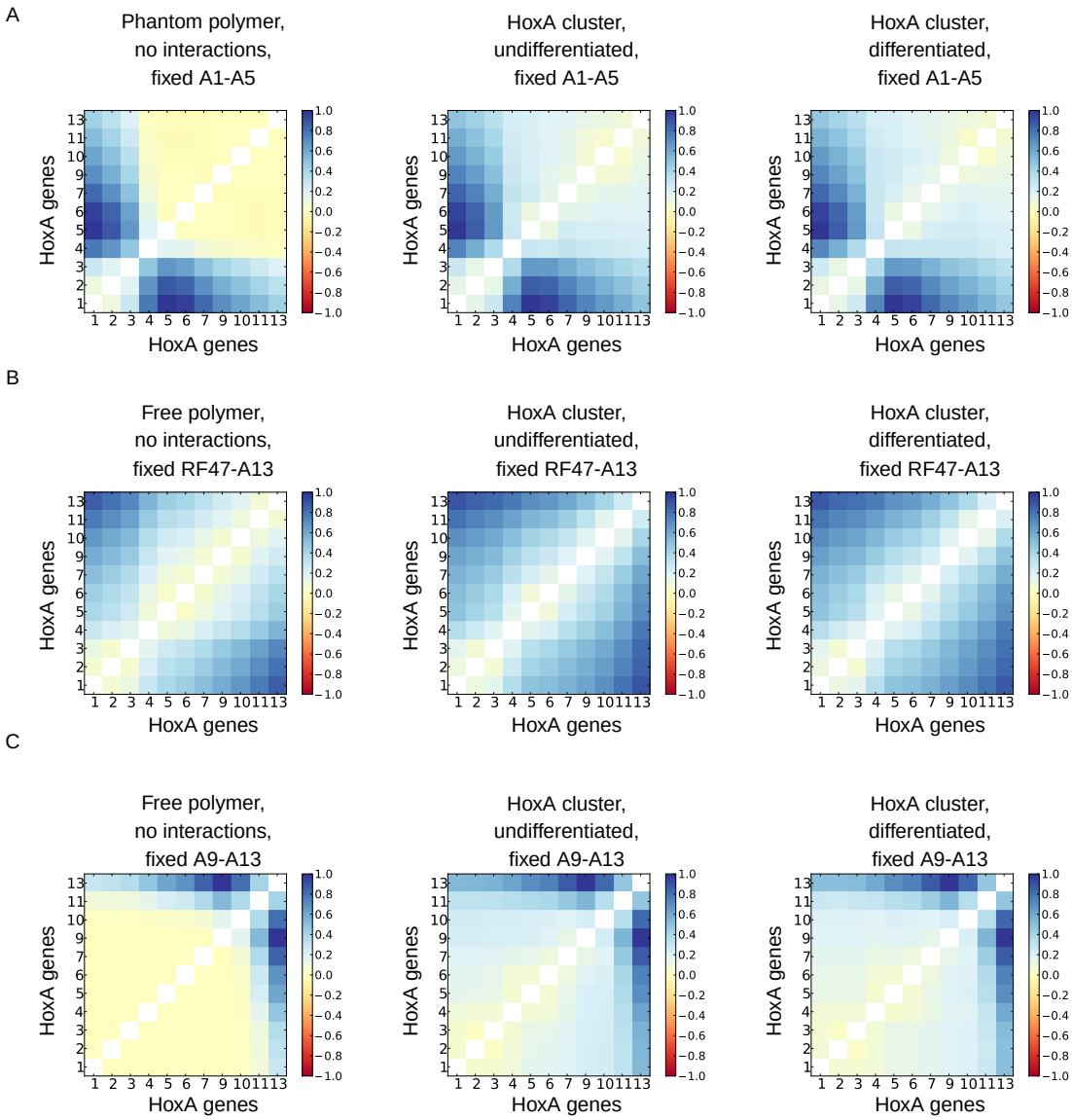


Figure 12.9: The Pearson correlation coefficient is calculated for (A) the distance between HoxA genes 1 and 5 and all other distances between HoxA genes, (B) the distance between restriction fragement (RF) 47 and HoxA gene 5 and all other distances between HoxA genes and (C) the distance between HoxA genes 9 and 13 and all other distances between HoxA genes. The comparison with a free polymer chain shows that the (positive) distance correlations for the differentiated and undifferentiated HoxA cluster emerge due to the connectivity of the fiber. In case of the HoxA cluster, the influence of clustering becomes apparent in the non-vanishing, positive correlation coefficients between pairs of genetic sites that are genomically distant. Image adapted from [175].

resolution imaging technologies can be found in chapter 2. The IMC method models the chromatin fiber as a worm-like chain and solves for an ensemble of conformations consistent with experimentally measured constraints (contact interaction maps, radial distribution functions, etc.). In contrast to other modeling approaches [13, 44, 78, 345, 427] the inverse procedure explicitly takes into account the connectivity of the chromatin fiber

and thus accounts for the topological complexity (including topological constraints) of its three-dimensional fold.

In light of the stochastic nature of the interactions and the huge fluctuations in subnuclear loci positioning [423] the advantage of the inverse procedure is the generation of a (unique) thermodynamically consistent maximum entropy ensemble. Notably, a single (individual) conformation alone does not reflect the "true" in vivo structures but only the full ensemble of chromatin structures allows for the investigation of structural properties [146].

We have applied the procedure to test systems to demonstrate its utility and performance in a controlled setting before investigating the human HoxA cluster. Our results suggests the it is organized into a "rosette"-like structures. Combining information on gene expression levels [44], our approach confirms looping and clustering as mechanisms related to gene repression [44, 87, 198, 427]. Moreover, the analysis of structural properties such as chromatin compaction by the radius of gyration or the distribution of loop sizes as well as hierarchical clustering shows that the IMC approach can leverage current CCC or FISH/microscopy data in order to reveal new insight into structure-function relationships of the chromatin fiber.

12.6 Supplementary Information

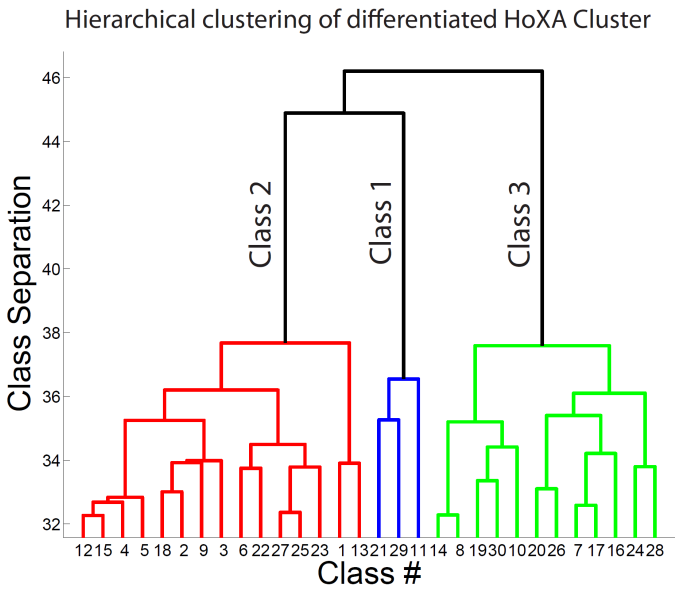


Figure 12.10: Dendrogram of the hierarchical clustering approach for the differentiated HoxA cluster. Clustering is performed in a hierarchical way, which allows to chose the optimal number of clusters. One way to judge the results of the clustering is the hierarchical tree (dendrogram), which is returned by the hierarchical clustering procedure. It illustrates where a cluster is split into (sub)classes according to the metric chosen. Image adapted from [175].

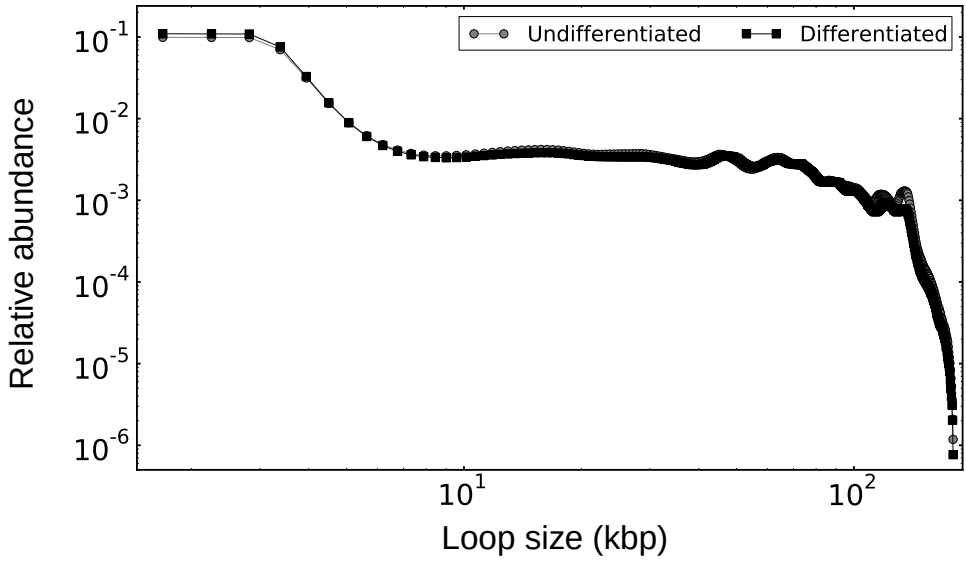


Figure 12.11: Relative abundance of loop sizes for the undifferentiated and differentiated HoxA cluster. The abundance of intermediate ($\sim 20 - 80$ kb) and large loops ($\sim 80 - 120$ kb) does not rapidly decrease with with increasing genomic size. This can be understood by noting that the 3' end of the HoxA cluster (fragments 47-50 in [128]) interacts very strongly with the entire HoxA region in agreement with [44]. Image adapted from [175].

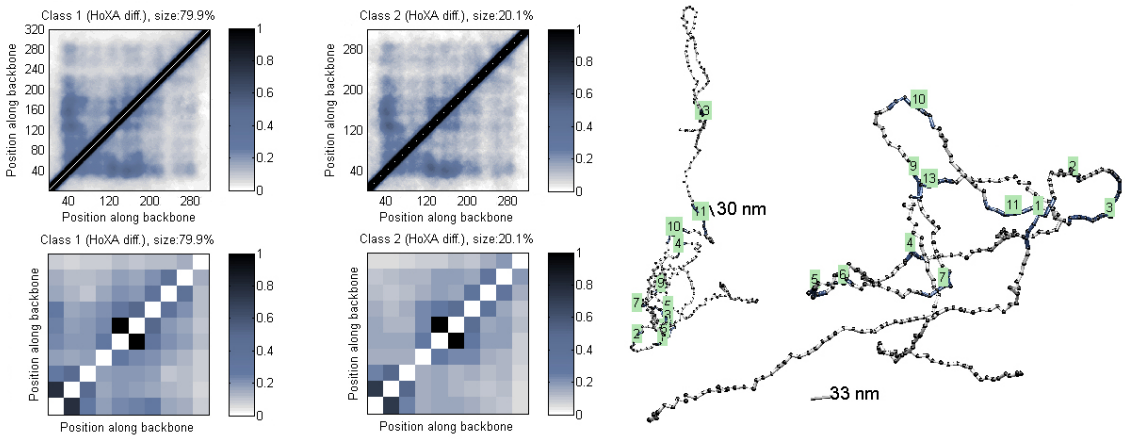


Figure 12.12: The undifferentiated HoxA cluster ensemble might be divided into two classes of conformations, which are, however, very similar as can be seen from the contact probability matrices representing the two classes. Thus, the undifferentiated HoxA region rather forms a single dominant class of compact fiber conformations, of which two 3D structures are exemplarily shown. Image adapted from [175].

Chapter 13

Conclusions and Outlook

13.1 A Short Summary of the Results

This thesis is inspired by a wide range of biophysical systems differing in their structural dynamics, length scales, organizational complexity as well as functional purpose. The aim of this thesis is to develop precise models to confirm and understand current experimental evidence. The predictive power of the developed frameworks is expected to stimulate future experiments to challenge our present knowledge on structure-function relationships ranging from bacterial to eukaryotic cells.

Biopolymer Tethering In the first two chapters of this thesis, the role of key features which are present in a wide range of biophysical contexts have been studied by applying basic polymer toy models. The investigated key features are (i) tethering, (ii) ring closure, (iii) confinement as well as (iv) semiflexibility.

We have developed a tethered-polymer model consisting of a single- or double-grafted chain on an infinite surface in order to determine the impact of (i) tethering on its structural and dynamical properties. The emphasis is on the interpretation of the experimentally accessible quantities, such as distance distributions between fluorescently labeled loci, which are significantly different in the presence of tethering interactions. We believe that the theoretical study of single- and double-tethered chains might facilitate the interpretation of future experiments (FISH, FROS, ParB-*par*S systems etc.) with multiple marked loci which try to resolve chromosome organization in more detail.

Semiflexible Ring Polymers in Confined Spaces In a subsequent step, the conformational properties of a semiflexible ring polymer in confined spaces are investigated therewith shedding light onto the role of the three key features (ii) ring closure, (iii) confinement and (iv) semiflexibility. In fact, taking into account the competing interplay between configurational entropy, bending energy and excluded volume, it is possible to elucidate the role that different geometrical constraints can play in shaping the spatial organization of semiflexible biopolymers. While elongated (rod-like) geometries reduce the amount of chain overcrossings and induce a pronounced ordering with respect to the long axis of the surrounding envelope, there exists no preferred orientational axis in the case of

spherical confinement. Upon increasing the system density and the rigidity of the chain, the polymer migrates from the center of the accessible space towards the surrounding surface forming a spool-like structure known for DNA condensation within viral capsids. The existence of distinct loop sizes for different confining geometries might influence colocalization in biopolymers necessary for the genome-wide coordination of gene expression. Thus, the advantages of certain geometric constraints such as spherical confinement of viral DNA in a capsid or the rod-shaped envelope of the circular chromosome in E. coli could be one driving force for controlling proper biological functioning.

Temperature-Dependent Structural Properties of Histone H2AX The toy models studied in this thesis allow for an intuitive picture on the static and dynamical properties of biopolymer systems. A more detailed and quantitative level of understanding can be gained from coarse-grained models which incorporate specific aspects of experimental evidence such as is the case for the study of the histone H2AX.

Histone proteins are not only important due to their vital role in cellular processes such as DNA compaction, replication and DNA repair but also show intriguing structural properties that might be exploited for bioengineering purposes such as the development of nano-materials. Given their biological and technological implications, it is interesting to investigate the structural properties of proteins as a function of temperature. In particular, we have studied the spatial response dynamics of the histone H2AX, consisting of 143 residues, with a coarse-grained bond fluctuating model for a broad range of normalized temperatures. A knowledge-based interaction matrix has been used as input for the residue-residue Lennard-Jones potential.

We find a variety of equilibrium structures including global globular configurations at low normalized temperature ($T^* = 0.014$), combination of segmental globules and elongated chains ($T^* = 0.016, 0.017$), predominantly elongated chains ($T^* = 0.019, 0.020$), as well as universal SAW conformations at high normalized temperature ($T^* \geq 0.023$). Notably, the radius of gyration of the protein exhibits a non-monotonic temperature dependence with a maximum at a characteristic temperature ($T_c^* = 0.019$) where a crossover occurs from a positive (stretching at $T^* < T_c^*$) to negative (contraction at $T^* > T_c^*$) thermal response on increasing T^*.

Entropy-Driven Organization of Synaptonemal Complexes Transitioning from the study of a single protein to the investigation of a proteinaceous structure, we have focused on the organization of synaptonemal complexes (SCs) to elucidate the role of entropy in shaping SC organization at the pachytene stage during meiosis.

Despite recent progress in visualization experiments, the mechanisms underlying the SC's spatial organization within the cell nucleus have so far remained elusive. 4Pi-microscopy is employed to study SC organization in mouse spermatocyte nuclei allowing for the three-dimensional reconstruction of the SC's backbone arrangement. Additionally, we model the SCs in the cell nucleus by confined, self-avoiding polymers, whose chain ends are attached to the envelope of the confining cavity and diffuse along it.

The framework provided by the complex interplay between SC polymer rigidity, tethering and confinement is able to qualitatively explain features of SC organization, such as mean squared end-to-end distances, mean squared center of mass distances or SC density distributions. However, it fails in correctly assessing SC entanglement within the nucleus. In fact, the analysis of the 4Pi-microscopy images reveals a higher ordering of SCs within the nuclear volume than what is expected in our numerical model. Thus, while entropic

contributions are an organizational driving force, the dedicated action of proteins or actin cables [49, 308, 321] might be needed in order to fine-tune the three-dimensional SC organization. To this end, future experiments determining the bending rigidity of SCs within the cell nucleus might help to qualitatively test our assumptions.

Rabl Model of Yeast Interphase Chromosomes A further transition of length scale, involves the quantitative testing of the Rabl model of chromosome organization in yeast interphase cells. Using a polymer model of yeast chromosomes that takes into account the Rabl organization and makes use of parameters such as packing density, elasticity and diffusion constants, which have all been reported previously, we compute the equilibrium distribution of distances between two genetic loci on the left arm of chromosome III, as well as their mutual diffusion, and compare the theoretical predictions to quantitative data obtained using fluorescently labeled chromosomes. Without the need for additional fitting parameters, our calculations are in line with experimental data. When the proteins responsible for telomere tethering are removed the measured distribution of distances changes but the observed change is markedly different than that calculated from the polymer model, suggesting that there is more to these mutant cells than simple untethering of telomeres from the nuclear periphery.

Impact of E. coli Domain Organization A rather large part of this thesis is dedicated to the spatial packaging of the E. coli chromosome within the nucleoid as well as to E. coli chromosome segregation during cell division. In a first step, we elucidate the role that the specific topological organization of the E. coli chromosome plays in shaping its structural properties in free space. Since the circular E. coli chromosome is compacted into separate chromosomal domains, we study a polymer architecture consisting of a central ring to which either looped or linear side chains are grafted. A shape change from a spherical to a toroidal organization takes place as soon as the inner ring becomes large enough for the attached arms to fit within its circumference. Building up a torus, the system flattens depending on the effective bending rigidity of the chain induced by entropic repulsion of the attached loops or, to a lesser extent, linear arms. Our results suggest that the specific underlying topology of the E. coli chromosome could be one driving force in the absence of confinement that nature exploits, to ensure proper packaging of the genetic material within a rod-shaped, bacterial envelope, by the formation of a toroidal structure with a decreased amount of writhe.

TF-Gene Driven E. coli Chromosome Packaging However, the E. coli chromosome is not located in free space but it is strongly confined within the bacterial nucleoid. In a subsequent step, we study the interplay between the specific topology of the E. coli chromosome and the geometrical constraints imposed by the bacterial envelope. To this end, measurement of position fluctuations of single loci by fluorescently labeling genetic sites along the circular chromosome reveal thats loci in the body of the nucleoid are positioning with a precision of better than 10% of the cell length.

Is the coupling between chromosome topology and confinement able to overcome the chromosome's propensity to mix and to self-organize into a nucleoid-filament type of structure (leading to the experimentally observed high precision of subcellular positioning)? Indeed, we can show that the entropic repulsion of chromosomal domains and the pressure exerted by the confining cavity creates the right physical conditions for E. coli chromosome packaging.

Additionally, we establish one possible mechanism for actually forming chromosomal domains which is supported by the observation that the E. coli chromosome is organized into a complex 3D network constrained by long- and short-range interactions. We propose that the E. coli gene regulatory interactions, which form a complex regulatory network, organize the DNA chain into several domains (chromosomal loops). This process is assumed to be driven by imposing colocalization of transcription factors and their target genes. Investigating the consequences of this assumption, we find the circular chromosome to indeed self-organize into an effective nucleoid filament-type of structure. Our framework is able to explain the experimentally found high precision of subnuclear positioning. Moreover, to reproduce the observed precise ordering of the chromosome, we estimate that the domain sizes are distributed between 10 and 700 kb, in agreement with the size of topological domains identified in the context of DNA supercoiling.

E. coli Chromosome Segregation Our developed framework of chromosome packaging allowed us to better understand the underlying principle of E. coli chromosome segregation. In particular, we studied how excluded volume effects, specific polymer topologies and geometrical confinement compete with entropy to drive dynamic processes such as the segregation of highly compacted chromosomes during cell division. Our work shows that the elongated geometry of the bacterial cell provides a natural axis along which the two copies of the chromosome can separate. Notably, the velocity and completion of the segregation process in both cubic and rectangular confinement strongly depend on the flexibility of the biopolymer. A small increase in chain stiffness is already sufficient to induce a failure of the segregation progress. In the future, a reliable assessment of the segregation time scale by applying technologies such as microfluidics might provide a calibration of numerical Monte Carlo time steps and thus help to underpin the concept of spontaneous, entropy-driven chromosome segregation.

Complementing CCC Techniques with an Inverse Monte Carlo Method Last but not least, we have developed a computational tool complementing chromosome conformation capture (CCC) techniques, such as 3C, 4C, 5C, ChIP-loop or Hi-C. These techniques probe the 3D architecture of a genetic sequence or an entire genome in order to understand complex (epi)genetic interactions that contribute to its folding pattern. Despite the significant amount of data that have been produced to date, limited insight towards detailed 3D chromosomal conformations has been gained because computational tools lack the capacity to invert CCC data in order to yield structural models of the genome.

Here, we have presented a Monte Carlo-based computational approach that inverts CCC data to provide the maximum entropy ensemble of conformations consistent with contact probabilities measured using CCC-based technologies. We have applied the procedure to test systems to demonstrate its utility and performance in a controlled setting.

Application of the procedure to experimental CCC data for the human HoxA cluster shows that it is organized into multiple chromatin loops due to the formation of distinct contacts between the HoxA genes. These regulatory loops form and disolve in coordination with the transcriptional activity of the HoxA genes.

13.2 Future Challenges

Polymer Toy Models Toy models which simplify a biophysical system by considering only its very basic features are a valuable first step towards gaining a general understanding about the system of interest and finding more specific models for it.

Semiflexible biopolymers, such as DNA, some proteins, rod-like viruses, or actin filaments, have increasingly stirred interest in the biophysical community [26, 31–33]. However, in light of a recent work by Binder and coworkers [157, 248], the applicability of the textbook definition of persistence length used in many experiments and modeling approaches as a measure for the local intrinsic stiffness of biopolymers might be limited. They have considered bottle-brush polymers as a model for semiflexible polymers, where the chain stiffness of the backbone can be changed by varying the length of the side chains. The authors have shown that the persistence length extracted from the exponential decay of the orientational correlation function is not a characteristic of the internal stiffness of the bottle-brush, but strongly depends on the backbone chain length. While Binder and colleagues have studied linear side chains, the question of the impact of looped side chains has to be raised since it is a common motif in biological systems such as E. coli [14, 30]. An investigation of bottle-brushes with looped side chains would clearly help the interpretation of future experiments, that find the persistence length to strongly depend on the method of data analysis and the experimental conditions such as temperature [453].

Another future toy model might be used to shed light onto bacterial chromosome segregation. During the segregation process topoisomerase IV (Topo IV) is responsible for unlinking the concatenated sister chromosomes to ensure that each daughter cell receives a complete copy of the genetic information [9, 53]. Concatenated ring polymers could be confined to different geometries, where the influence of Topo IV might be modeled by gradually disregarding excluded volume constraints when chain overcrossings take place. This approach would allow us to study the strength of the entropy-driven chromosome movements and might help to determine the (Monte Carlo) timescale for topoisomerases to disentangle the replicated sister chromosomes before cell division.

E. coli Chromosome Packaging and Segregation The folded bacterial genome is supposed to be organized by nucleoid-associated DNA-binding proteins (NAPs) [15, 116], DNA supercoiling [9, 30], and transcriptional regulatory interactions [5, 14–21]. In chapters 9 and 10, we have elucidated the impact of domain formation on nucleoid structure and have investigated one mechanism driving chromosomal domain formation: the gene regulatory network. Not surprisingly, global regulators in the E. coli transcriptional regulatory network are indeed NAPs such as FIS and H-NS [3]. One might want to combine state-of-the-art techniques such as CCC-based methods with high-resolution imaging approaches to track the position and function of NAPs within the nucleoid and relate them to chromosome folding motifs [454, 455].

Additionally, one may raise the question about the impact of NAPs such as MukBEF, H-NS and gyrase on bacterial chromosome segregation during cell division. While a broad range of proteins and mechanisms have been proposed to facilitate chromosome segregation, no consensus mechanistic view has emerged so far. In this thesis, we have studied how excluded volume effects, specific polymer topologies and geometrical confinement could drive segregation of highly compacted chromosomes during cell division. It is worthwhile to take NAPs into account. They are expected to influence both the level of supercoiling as well as the amount of domain loops [15, 116], which, in turn, modifies the complexity of chain topology and thus the strength of segregation due to entropic forces [23, 24]. Notably,

it was demonstrated that lowly expressed levels of MukBEF, which was expected to lead to failure of chromosome segregation, were completely balanced by alterations in gyrase activity [456, 457]. Consequently, these proteins alone cannot constitute the dedicated force behind segregation, but instead they might shape the physical conditions of chromosomes, i.e. with respect to topological complexity, in order to facilitate entropy-driven segregation [23, 24].

An additional step towards a better understanding of the processes that govern E. coli chromosome packaging and segregation could be achieved by fluorescently labeling a broad range of genetic loci separated by a genomic distance smaller than the mean domain size identified in chapter 10. Spatial and temporal tracking of these loci during the cell cycle is expected to show different dynamics of sites located on different chromosomal domains and might provide a "real" timescale allowing for the calibration Monte Carlo time steps as proposed in chapter 11.

Eukaryotic Genome Architecture The spatial genome organization within the cell nucleus and its functional implications are a thriving field of research. Coupling of the developed inverse Monte Carlo method to chromosome conformation capture experiments in chapter 12 reveals a pronounced structure-function relationship with respect to the regulation of gene expression, where the chromatin fiber is found to be highly dynamic. On the molecular biological level, the observed dynamics are driven by displacing and modifying the nucleosomes [6], while at a higher level of chromatin compaction, chromatin loops disolve and reform due to the (regulatory) activity of nuclear/architectural proteins [6, 9, 174].

These chromatin remodeling events influence the flexibility of the chromatin fiber [6]. In fact, computational modeling of the chromatin fiber in chapter 12 shows a wide range of fiber flexibilities, where specific genomic regions display distinct bending stiffnesses. Thus, experimental approaches which probe how the flexibility of chromatin correlates with intrinsic (histone modifications, presence of linker histones, gene expression/repression) [6] and extrinsic (salt concentration, temperature) conditions might reveal insightful details for future modeling purposes.

Chromatin flexibility can be characterized by its persistence length. However, measurements that try to assess the persistence length of the chromatin fiber report large variations between 30 nm to 200 nm [127, 172, 458]. To this end, one has to note that a recent work by Binder and coworkers [157, 248] has shown that the standard definitions of persistence length fail in describing the local intrinsic flexibility of the chromatin fiber. In this context, the need for toy models, as mentioned above, investigating particularly this issue is of great interest to both experimentalists and the computational/analytical modeling community.

The computational tool presented in chapter 12 is able to "translate" 2D interaction frequency data into an ensemble of 3D chromatin fiber conformations. This approach allows for the investigation of a vast amount of structural properties (persistence length, loop distributions, gyration radius, etc.), which cannot be assessed by experimental techniques so far. The most immediate consequence of the successfull application of the inverse procedure to the HoxA cluster is its application to other biological systems that have been studied by CCC techniques, such as the human alpha-globin locus [294].

However, the coupling of CCC technologies and computational modeling is not able to provide information about the functional relevance of the experimentally measured (epi)genetic interactions or the molecular drivers which fine-tune chromatin structure [7].

For example, it is known that the establishment and segregation of different chromatin domains (gene-rich and gene-poor) and maintenance of these structures are mediated by noncoding DNA elements known as insulators, which are bound by the CCCTC-binding factor (CTCF) [91]. Thus, future studies on the impact of epigenetic regulators, such as CTCF and cohesin, which stabilize DNA loops and chromosomal domains [459] are needed to complement our knowledge obtained from CCC studies and modeling approaches.

Last but not least, biochemical methods which map interaction frequencies between genomic sites suffer from a couple of drawbacks, which have been discussed in detail in chapter 2. First, it is difficult to calibrate the interaction datasets with respect to absolute contact frequencies [10]. Second, CCC datasets are (so far) generated from cell populations at various cell cycle states [128]. Third, CCC techniques require a huge amount of sample cells (of the order of millions) to improve the signal-to-noise ratio. Fourth, it is difficult to determine which interactions occur simultaneously in any given cell.

Here, FISH/high-resolution imaging techniques [454, 455] can provide information on the position and function of regulatory proteins as well as on the precise spatial distances between genomic sites of interest. In fact, the knowledge about Euclidian distances would strengthen chromatin modeling approaches [13]. Additionally, as opposed to CCC measurements, high-resolution imaging approaches facilitate the comparison of chromatin arrangements at the single cell level and the quantification of cell-to-cell variability [7].

In particular, high-resolution microscopy might allow for the determination of the pair distribution functions for genomic regions of interest, such as the HoxA cluster in chapter 12. With the advent of a "rainbow of fluorescent proteins" [133], this could be done by labeling a large amount of pairs of genetic sites in synchronized cells with different "colors" in order to be able to distinguish between them. Instead of applying contact frequency maps as "targets" for the inverse method, the so obtained pair distribution functions could be solved for by the inverse algorithm. In fact, pair distribution functions and structure factors have successfully been applied to study the structure of colloidal aggregates by reverse Monte Carlo modeling [460].

Besides the investigation of the human Hox clusters [44, 427] or the human alpha-globin locus [294], a biological process worth studying is the repair response to DNA double-strand breaks discussed in chapter 6. There are increasing lines of evidence that damage-dependent changes in chromatin structure (by chromatin remodeling complexes) are required for the formation of γ-H2AX, which is expected to contribute to the preparation of DNA repair [279]. Jörg Bewersdorf and coworkers have investigated the distribution of H2AX throughout the chromatin fiber during the time course of DNA damage and repair by means of 4Pi microscopy [270]. The authors propose that the observed formation of H2AX clusters supports the immediate and robust repair response. With the aim to determine the corresponding radial distribution function, distance distributions for several genetic sites on a genomic region subject to DNA damage could be measured. Using the so determined structural data to model this genomic region with our inverse approach might reveal repair-specific chromatin structures, which could then be confronted to the positioning of nuclear proteins that drive DNA repair mechanisms.

In conclusion, a truly ingenious, integrative approach combining high-resolution imaging and CCC technologies with computational modeling for identifying spatial genome organization is needed. To this end, CCC measurements should be seen as providing an overall qualitative picture, while high-resolution (live) cell imaging is able to fine-tune our knowledge by spatially and temporally tracking the nuclear players at the single-cell level. In a subsequent step, computational modeling (applying such a tool as presented in chap-

ter 12), can leverage the wealth of experimental CCC or microscopy data on chromatin folding to hint to specific folding motifs of interest, where high-resolution microscopy allowing for the precise measurement of distances might bridge the gap from a qualitative overview to a truly quantitative model of genome organization.

Conference/Workshop Participation

I have participated in the following conferences and workshops:

- Retreat of the Research Training Group "Simulational Methods in Physics", 2008, Würzburg [talk]
- HICCup – Heidelberg Initiative for Chromatin Computing, December 25th, 2008, Heidelberg [talk]
- 2nd Workshop on Monte Carlo Methods, January 22 – 24, 2009, Interdisciplinary Center for Scientific Computing (IWR), Heidelberg [talk]
- Workshop "Biophysics of Chromatin", February 4 – 6, 2009, Villa Bosch, Heidelberg
- Spring School on Multiscale Methods and Modelling in Biophysics and Systems Biology, May 18 – 29, 2009, Shanghai, China
- 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]
- "Darwin09 - 150 Years after Darwin: From Molecular Evolution to Language", November 23 – 27, 2009, Palma de Mallorca, Spain [poster]
- Joint Meeting of the Biophysical Society 54th Annual Meeting, February 20 – 24, 2010, San Francisco, California [poster]
- DPG-Frühjahrstagung, March 21 – 26, 2010, Regensburg [talk]
- Whitehead Institute Retreat, September, 2010, White Mountains, New Hampshire [poster]
- 5th Workshop on Monte Carlo Methods, January 14th, 2011, Interdisciplinary Center for Scientific Computing (IWR), Heidelberg [talk]
- Joint Meeting of the Biophysical Society 55th Annual Meeting, March 5 – 9, 2011, Baltimore, Maryland [poster]

Bibliography

[1] U.S. Department of Energy Genome Program. http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml. [Online; accessed July-2011].

[2] E. Toro and L. Shapiro. Bacterial Chromosome Organization and Segregation. *Cold Spring Harb Perspect Biol*, 2:a000349, 2010.

[3] A. Martínez-Antonio. Escherichia coli transcriptional regulatory network. *Network Biology*, 1:21–33, 2011.

[4] J. Mondal, B.P. Bratton, Y. Li, A. Yethiraj, and J.C. Weisshaar. Entropy-Based Mechanism of Ribosome-Nucleoid Segregation in E. coli Cells. *Biophys. J.*, 100:2605–2613, 2011.

[5] P.A. Wiggins, K.C. Cheveralls, J.S. Martin, R. Lintner, and J. Kondev. Strong intranucleoid interactions organize the Escherichia coli chromosome into a nucleoid filament. *Proc. Natl. Acad. Sci. U. S. A.*, 107:4991–95, 2010.

[6] B.V.S. Iyer, M. Kenward, and G. Arya. Hierarchies in eukaryotic genome organization: Insights from polymer theory and simulations. *BMC Biophysics*, 4:8, 2011.

[7] I. Rajapakse and M. Groudine. On emerging nuclear order. *J. Cell Biol.*, 192:711, 2011.

[8] A. Bolzer, G. Kreth, I. Solovei, D. Koehler, K. Saracoglu, C. Fauth, S. Müller, and R. Eils et al. Three-Dimensional Maps of All Chromosomes in Human Male Fibroblast Nuclei and Prometaphase Rosettes. *PLoS Biol.*, 3:e157, 2005.

[9] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of the Cell.* Garland Science, New York, 2007.

[10] B. van Steensel and J. Dekker. Genomics tools for unraveling chromosome architecture. *Nature Biotechnology*, 28:1089–1095, 2011.

[11] T. Cremer and C. Cremer. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat. Rev. Genet.*, 2:292–301, 2001.

[12] J. Mateos-Langerak, M. Bohn, W. de Leeuw, O. Giromus, E.M.M. Manders, P.J. Verschure, M.H.G. Indemans, H.J. Gierman, D.W. Heermann, R. van Driel, and S. Goetze. Spatially confined folding of chromatin in the interphase nucleus. *Proc. Natl. Acad. Sci. U. S. A.*, 106:3812–17, 2009.

[13] D. Baù and M.A. Marti-Renom. Structure determination of genomic domains by satisfaction of spatial restraints. *Chromosome Res.*, 19:25–35, 2011.

[14] F. Képès. Periodic transcriptional organization of the E. coli genome. *J. Mol. Biol.*, 340:957–964, 2004.

[15] R.T. Dame, O.J. Kalmykowa, and D.C. Grainger. Chromosomal Macrodomains and Associated Proteins: Implications for DNA Organization and Replication in Gram Negative Bacteria. *PLoS Genet.*, 7:e1002123, 2011.

[16] M.S. Luijsterburg, M.C. Noom, G.J.L. Wuite, and R.T. Dame. The architectural role of nucleoid-associated proteins in the organization of bacterial chromatin: a molecular perspective. *J. Struct. Biol.*, 156:262–272, 2006.

[17] S.B. Zimmerman. Shape and compaction of Escherichia coli nucleoids. *J. Struct. Biol.*, 156:255–261, 2006.

[18] N. Sonnenschein, M. Geertz, G. Muskhelishvili, and M.-T. Hütt. Analog regulation of metabolic demand. *BMC Microbiology*, 5:1–13, 2011.

[19] C. Marr, M. Geertz, M.-T. Hütt, and G. Muskhelishvili. Dissecting the logical types of network control in gene expression profiles. *BMC Syst. Biol.*, 2:18, 2008.

[20] E. Yeger-Lotem R. Hershberg and H. Margalit. Chromosomal organization is shaped by the transcription regulatory network. *Trends Genet.*, 21:138–42, 2005.

[21] S.C. Janga, H. Salgado, and A. Martínez-Antonio. Transcriptional regulation shapes the organization of genes on bacterial chromosomes. *Nuc. Acids Res.*, 37:3680–3688, 2009.

[22] P. Montero Llopis, A.F. Jackson, O. Sliusarenko, I. Surovtsev, J. Heinritz, T. Emonet, and C. Jacobs-Wagner. Spatial organization of the flow of genetic information in bacteria. *Nature*, 466:77–82, 2010.

[23] S. Jun and A. Wright. Entropy as the driver of chromosome segregation. *Nature*, 8:600–607, 2010.

[24] S. Jun and B. Mulder. Entropy-driven spatial organization of highly confined polymers: Lessons for the bacterial chromosome. *Proc. Natl. Acad. Sci. U. S. A.*, 103:12388, 2006.

[25] E. Schrödinger. *What is life?* Cambridge University Press, Cambridge, 1994.

[26] G. Witz, K. Rechendorff, J. Adamcik, and G. Dietler. Conformation of Ring Polymers in 2D Constrained Environments. *Phys. Rev. Lett.*, 106:248301, 2011.

[27] G. Witz, K. Rechendorff, J. Adamcik, and G. Dietler. Conformation of Circular DNA in Two Dimensions. *Phys. Rev. Lett.*, 101:148103, 2008.

[28] D. Marenduzzo and C. Micheletti. Thermodynamics of DNA packaging inside a viral capsid: The role of DNA intrinsic thickness. *J. Mol. Biol.*, 330:485–492, 2003.

[29] R. Matthews, A.A. Louis, and J.M. Yeomans. Knot-Controlled Ejection of a Polymer from a Virus Capsid. *Phys. Rev. Lett.*, 102:088101–5, 2009.

[30] L. Postow, C.D. Hardy, J. Arsuaga, and N.R. Cozzarelli. Topological domain structure of the Escherichia coli chromosome. *Genes. Dev.*, 18:1766–79, 2004.

[31] K. Ostermeir, K. Alim, and E. Frey. Buckling of stiff polymer rings in weak spherical confinement. *Phys. Rev. E*, 81:061802, 2010.

[32] K. Ostermeir, K. Alim, and E. Frey. Confinement induces conformational transition of semiflexible polymer rings to figure eight form. *Soft Matter*, 6:3467–3471, 2010.

[33] M. Fritsche and D.W. Heermann. Confinement Driven Spatial Organization of Semiflexible Ring Polymers: Implications for Biopolymer Packaging. *Soft Matter*, 7:6906–6913, 2011.

[34] P.H. Viollier, M. Thanbichler, P.T. McGrath, L. West, M. Meewan, H.H. McAdams, and L. Shapiro. Rapid and sequential movement of individual chromosomal loci to specific subcellular locations during bacterial DNA replication. *Proc. Natl. Acad. Sci. U. S. A.*, 101:9257–9262, 2004.

[35] H. Niki, Y. Yamaichi, and S. Hirage. Dynamic organization of chromosomal DNA in Escherichia coli. *Genes & Developement*, 14:212–223, 2000.

[36] K. Drlica, E. Burgi, and A. Worcel. Association of the folded chromosome with the cell envelope of Escherichia coli: Nature of the membrane-associated DNA. *Journal of Bacteriology*, 134:1108–1116, 1978.

[37] K. Bystricky, T. Laroche, G. van Houwe, M. Blaszczyk, and S.M. Gasser. Chromosome looping in yeast: Telomer pairing and coordinated movement reflect anchoring efficiency and territorial organization. *J.Cell Biol.*, 168:375–387, 2005.

[38] A. Taddei and S.M. Gasser. Multiple pathways for telomer tethering: Functional implications of subnuclear position for heterochromatin formation. *Biochimica et Biophysica Acta*, 1677:120–128, 2004.

[39] A. Rosa, J.H. Maddocks, F.R. Neumann, S.M. Gasser, and A. Stasiak. Measuring limits of telomere movement on nuclear envelope. *Biophys. J.*, 90:L24–26, 2006.

[40] W.F. Marshall, A. Straight, J.F. Marko, J. Swedlow, A. Dernburg, A. Belmont, A.W. Murray, D.A. Agard, and J.W. Sedat. Interphase chromosomes undergo constrained diffusional motion in living cells. *Curr. Biol.*, 7:930–9, 1997.

[41] W.F. Marshall, J.F. Marko, D.A. Agard, and J.W. Sedat. Chromosome elasticity and mitotic polar ejection force measured in living Drosophila embryos by four-dimensional microscopy-based motion analysis. *Current Biology*, 11:569–578, 2001.

[42] H. Yokota, G. van den Engh, J.E. Hearst, R.K. Sachs, and B.J. Trask. Evidence for the organization of chromatin in megabase pair-sized loops arranged along a random walk path in the human g0/g1 interphase nucleus. *J. Cell Biol.*, 130:1239–1249, 1995.

[43] R. Sachs, G. van den Engh, B. Trask, H. Yokota, and J. Hearst. A random-walk/giant-loop model for interphase chromosomes. *Proc. Natl. Acad. Sci. U. S. A.*, 92:2710–2714, 1995.

[44] J. Fraser, M. Rousseau, S. Shenker, M.A. Ferraiuolo, Y. Hayashizaki, M. Blanchette, and J. Dostie. Chromatin conformation signatures of cellular differentiation. *Genome Biology*, 10(4):R37, 2009.

[45] M. Buenemann and P. Lenz. A Geometrical Model for DNA Organization in Bacteria. *PLoS One*, 5:e13806, 2010.

[46] D. Napper. *Polymeric Stabilization of Colloidal Dispersions (Colloid Science).* Academic Press, London, 1983.

[47] L. Newnham, P. Jordan, B. Rockmill, G.S. Roeder, and E. Hoffmann. The synaptonemal complex protein, Zip1, promotes the segregation of nonexchange chromosomes at meiosis I. *Proc. Natl. Acad. Sci. U. S. A.*, 107:781–785, 2010.

[48] S.L. Page and R.S. Hawley. Chromosome Choreography. *Science*, 301:785–789, 2003.

[49] D. Zickler and N. Kleckner. Meiotic chromosomes: integrating structure and function. *Annu. Rev. Genet.*, 33:603, 1999.

[50] D. Zickler and N. Kleckner. Meiotic chromosomes: integrating structure and function. *Annu. Rev. Genet.*, 33:603–754, 1999.

[51] F.Z. Watts and E. Hoffmann. SUMO meets meiosis: An encounter at the synaptonemal complex. *BioEssays*, DOI: 10.1002/bies.201100002, 2011.

[52] M.J. Moses. Structure and function of the synaptonemal complex. *Genetics (Supplement)*, 61:41–51, 1969.

[53] H. Lodish, A. Berk, C.A. Kaiser, M. Krieger, M.P. Scott, A. Bretscher, and H.Ploegh. *Molecular Cell Biology.* W.H. Freeman, New York, 6th edition, 2007.

[54] N.S. Sharma. *Molecular Cell Biology.* International Science Publishing Academy, New Delhi, 1st edition, 2005.

[55] Wikipedia, The Free Encyclopedia. Celltypes.svg. `http://upload.wikimedia.org/wikipedia/commons/thumb/8/83/Celltypes.svg/2000px-Celltypes.svg.png`. [Online; accessed July-2011].

[56] M. Bohn and D. W. Heermann. Topological interactions between ring polymers: Implications for chromatin loops. *J. Chem. Phys.*, 132:044904–14, 2010.

[57] M. Bohn and D.W. Heermann. Repulsive Forces Between Looping Chromosomes Induce Entropy-Driven Segregation. *PLoS One*, 6:e14428, 2010.

[58] C.D. Allis, T. Jenuwein, and D. Rosenberg, editors. *Epigenetics.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 2007.

[59] C.L. Woodcock and R.P. Ghosh. Chromatin Higher-order Structure and Dynamics. *Cold Spring Harb. Perspect. Biol.*, 2:a000596, 2010.

[60] F. Thoma, T. Koller, and A. Klug. Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. *J. Cell Biol.*, 83:403–27, 1979.

[61] J. Zlatanova and S.H. Leuba. Chromatin structure and dynamics: lessons from single molecule approaches. *New Compr. Biochem.*, 39:369–396, 2004.

[62] T. Schalch, S. Duda, D.F. Sargent, and T.J. Richmond. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature*, 436:138–41., 2005.

[63] J. Bednar and S. Dimitrov. Chromatin under mechanical stress: from single 30 nm fibers to single nucleosomes. *FEBS Journal*, 278:2231–2243, 2011.

[64] Horowitz RA Whitaker N. Woodcock CL, Grigoryev SA. A chromatin folding model that incorporates linker variability generates fibers resembling the native structures. *Proc. Natl. Acad. Sci. U. S. A.*, 90:9021–5, 1993.

[65] H. Schiessel, W.M. Gelbart, and R. Bruinsma. DNA folding: structural and mechanical properties of the two-angle model for chromatin. *Biophys. J.*, 80:1940–56, 2001.

[66] B. Dorigo, T. Schalch, A. Kulangara, S. Duda, R.R. Schroeder, and T.J. Richmond TJ. Nucleosome arrays reveal the two-start organization of the chromatin fiber. *Science*, 306:1571–3, 2004.

[67] J.T. Finch and A. Klug. Solenoidal model for superstructure in chromatin. *Proc. Natl. Acad. Sci. U. S. A.*, 73:1897–901, 1976.

[68] J. Widom and A. Klug. Structure of the 300 A chromatin filament: X-ray diffraction from oriented samples. *Cell*, 43:207–13, 1985.

[69] K. van Holde and J. Zlatanova. Chromatin higher order structure: chasing a mirage? *J. Biol. Chem.*, 270:8373–6, 1995.

[70] K. Maeshima, S. Hihara, and H. Takata. New Insight into the Mitotic Chromosome Structure: Irregular Folding of Nucleosome Fibers Without 30-nm Chromatin Structure. *Cold Spring Harb. Symp. Quant. Biol.*, 75:439–44, 2010.

[71] P.M. Diesinger and D.W. Heermann. Monte Carlo Simulations indicate that Chromatin Nanostructure is accessible by Light Microscopy. *PMC Biophysics*, 3:11, 2010.

[72] P.M. Diesinger, S. Kunkel, J. Langowski, and D.W. Heermann. Histone Depletion Facilitates Chromatin Loops on the Kilobasepair Scale. *Biophys. J.*, 99:2995–3001, 2010.

[73] T. Kouzarides. Chromatin Modifications and Their Function. *Cell*, 128:693–705, 2007.

[74] X. Yang, R. Zaurin, M. Beato, and C.L. Peterson. Swi3p controls SWI/SNF assembly and ATP-dependent H2A-H2B displacement. *Nat. Struct. Mol. Biol.*, 14:540–547, 2007.

[75] I. Whitehouse, A. Flaus, B.R. Cairns, M.F. White, J.L. Workman, and T. Owen-Hughes. Nucleosome mobilization catalysed by the yeast SWI/SNF complex. *Nature*, 400:784–787, 1999.

[76] A. Hamiche, R. Sandaltzopoulos, D.A. Gdula, and C. Wu. ATP-Dependent Histone Octamer Sliding Mediated by the Chromatin Remodeling Complex NURF. *Cell*, 97:833–842, 1999.

[77] G. Mizuguchi, X. Shen, J. Landry, W.H. Wu, S. Sen, and C. Wu. ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. *Science*, 303:343–348, 2004.

[78] S. Jhunjhunwala, M.C. van Zelm, M.M. Peak, S. Cutchin, R. Roy, J.J.M. van Dongen, F.G. Grosveld, T.A. Knoch, and C. Murre. The 3D Structure of the Immunoglobulin Heavy-Chain Locus: Implications for Long-Range Genomic Interactions. *Cell*, 133:265–279, 2008.

[79] D. Carter, L. Chakalova, C.S. Osborne, Y. Dai, and P. Fraser. Long-range chromatin regulatory interactions in vivo. *Nat. Genet.*, 32:623–626, 2002.

[80] B. Tolhuis, R. Palstra, E. Splinter, F. Grosveld, and W. de Laat. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol. Cell*, 10:1453–65, 2002.

[81] C. Sayegh, S. Jhunjhunwala, R. Riblet, and C. Murre. Visualization of looping involving the immunoglobulin heavy-chain locus in developing B cells. *Genes & Dev.*, 19:322–327, 2005.

[82] P.R. Cook and D. Marenduzzo. Entropic organization of interphase chromosomes. *J. Cell Biol.*, 186:825–834, 2009.

[83] P.R. Cook. Predicting three-dimensional genome structure from transcriptional activity. *Nat. Genet.*, 32:347–52, 2002.

[84] L.A. Parada and T. Misteli. Chromosome positioning in the interphase nucleus. *Trends Cell Biol.*, 12:425–432, 2002.

[85] M. van der Ploeg. Cytochemical nucleic acid research during the twentieth century. *Eur. J. Histochem.*, 44:7–42, 2000.

[86] C. Münkel, R. Eils, S. Dietzel, D. Zink, C. Mehring, G. Wedemann, T. Cremer, and J. Langowski. Compartmentalization of interphase chromosomes observed in simulation and experiment. *J. Mol. Biol.*, 285:1053–1065, 1999.

[87] E. Lieberman-Aiden, N.L. van Berkum, L. Williams, M. Imakaev, and T. Ragoczy et al. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science*, 326:289–293, 2009.

[88] D.V. Lebedev, M.V. Filatov, A.I. Kuklin, A.K. Islamov, E. Kentzinger, R. Pantina, B.P. Toperverg, and V.V. Isaev-Ivanov. Fractal nature of chromatin organization in interphase chicken erythrocyte nuclei: DNA structure exhibits biphasic fractal properties. *FEBS Lett.*, 579:1465–8, 2005.

[89] J.A. Wallace and G. Felsenfeld. We gather together: Insulators and genome organization. *Curr. Opin. Genet. Dev.*, 17:400–407, 2007.

[90] C. Hou, R. Dale, and A. Dean. Cell type specificity of chromatin organization mediated by CTCF and cohesin. *Proc. Natl. Acad. Sci. U. S. A.*, 107:3651–3656, 2010.

[91] Y.J. Kim, K.R. Cecchini, and T.H. Kim. Conserved, developmentally regulated mechanism couples chromosomal looping and heterochromatin barrier activity at the homeobox gene A locus. *Proc. Natl. Acad. Sci. U. S. A.*, 108:7391–7396, 2011.

[92] M. Bohn. *Modelling of Interphase Chromosomes: From Genome Function to Spatial Organization.* PhD thesis, University of Heidelberg, 2010.

[93] J. Dorier and A. Stasiak. The role of transcription factories-mediated interchromosomal contacts in the organization of nuclear architecture. *Nuc. Acids Res.*, 38:7410–7421, 2010.

[94] J.A. Mitchell and P. Fraser. Transcription factories are nuclear subcompartments that remain in the absence of transcription. *Genes & Dev.*, 2006:20–25, 2008.

[95] T. Sexton, D. Umlauf, S. Kurukuti, and P. Fraser. The role of transcription factories in large-scale structure and dynamics of interphase chromatin. *Semin. Cell Dev. Biol.*, 18:691–697, 2007.

[96] S.T. Kosak and M. Groudine. Gene order and dynamic domains. *Science*, 306:644–647, 2004.

[97] T. Takizawa, K.J. Meaburn, and T. Misteli. The meaning of gene positioning. *Cell*, 135:9–13, 2008.

[98] T. Cremer, M. Cremer, S. Diezel, S. Muller, I. Solvei, and S. Fakan. Chromosome territories – a functional nuclear landscape. *Curr. Opin. Cell Biol.*, 18:307–316, 2006.

[99] M.R. Branco and A. Pombo. Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS Biol.*, 4:e138, 2006.

[100] M.O. Scheuermann, J. Tajbakhsh, A. Kurz, K. Saracoglu, R. Eils, and P. Lichter. Topology of genes and nontranscribed sequences in human interphase nuclei. *Exp. Cell Res.*, 301:266–279, 2004.

[101] J. Kang, B. Xu, Y. Yao, W. Lin, C. Hennessy, P. Fraser, and J. Feng. A Dynamical Model Reveals Gene Co-Localizations in Nucleus. *PLoS Comp. Biol.*, 7:e1002094, 2011.

[102] K.L. Reddy, J.M. Zullo, E. Bertolino, and H. Singh. Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature*, 452:243–247, 2008.

[103] S. Krull, J. Dörries, B. Boysen, S. Reidenbach, L. Magnius, H. Norder, J. Thyberg, and V.C. Cordes. Protein Tpr is required for establishing nuclear pore-associated zones of heterochromatin exclusion. *EMBO J.*, 29:1659–1673, 2010.

[104] E.P.C. Rocha. The Organization of the Bacterial Genome. *Annu. Rev. Genet.*, 42:211–233, 2008.

[105] E.L. Zechiedrich, A.B. Khodursky, and N.R. Cozzarelli. Topoisomerase IV, not gyrase, decatenates products of site-specific recombination in Escherichia coli. *Genes & Dev*, 11:2580–2592, 1997.

[106] A. Martínez-Antonio, A. Medina-Rivera, and J. Collado-Vides. Structural and functional map of a bacterial nucleoid. *Genome Biology*, 10:247–250, 2009.

[107] L.D. Murphy and S.B. Zimmerman. Isolation and characterization of spermidine nucleoids from Escherichia coli. *J. Struct. Biol.*, 119:321–335, 1997.

[108] D.E. Pettijohn. *The Nucleoid.* Escherichia coli and Salmonella: Cellular and Molecular Biology. ASM Press, Washington, DC, 2nd edition, 1996.

[109] M. Valens, S. Penaud, M. Rossignol, F. Cornet, and F. Boccard. Macrodomain organization of the Escherichia coli chromosome. *EMBO Journal*, 23:4330–4341, 2004.

[110] A. Worcel and E. Burgi. On the structure of the folded chromosome of Escherichia coli. *J. Mol. Biol.*, 71:127–147, 1972.

[111] T. Romantsov, I. Fishov, and O. Krichevsky. Internal structure and dynamics of isolated Escherichia coli nucleoids assessed by fluorescence correlation spectroscopy. *Biophys. J.*, 92:2875–84, 2007.

[112] D. Marenduzzo, I Faro-Trindade, and P.R. Cook. What are the molecular ties that maintain genomic loops? *Trends Genet.*, 23:126–133, 2007.

[113] T. Vora, A.K. Hottes, and S. Tavazoie. Protein Occupancy Landscape of a Bacterial Genome. *Molecular Cell*, 35:247–253, 2009.

[114] G.R. Bowman, L.R. Comolli, J. Zhu, M. Eckart, M. Koenig, K.H. Downing, W.E. Moerner, T. Earnest, and L. Shapiro. A self-associating protein critical for chromosome attachment, division, and polar organization in caulobacter. *Cell*, 134:945–55, 2008.

[115] G. Ebersbach, A. Briegel, G.J. Jensen, and C. Jacobs-Wagner. A self-associating protein critical for chromosome attachment, division, and polar organization in caulobacter. *Cell*, 134:956–968, 2008.

[116] S.C. Dillon and C.J. Dorman. Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat. Rev. Microbiol.*, 8:185–195, 2010.

[117] B. Pierce. *Genetics: A Conceptual Approach.* W.H. Freeman and Company, New York, 2nd edition, 2005.

[118] S.C. Janga and J. Collado-Vides. Structure and evolution of gene regulatory networks in microbial genomes. *Res. Microbiol.*, 158:787–94, 2007.

[119] M.M. Babu, N.M. Luscombe, L. Aravind, M. Gerstein, and S.A. Teichmann. Structure and evolution of transcriptional regulatory networks. *Curr. Opin. Struct. Biol.*, 14:283–91, 2004.

[120] M. Fritsche, S. Li, D.W. Heermann, and P.A. Wiggins. Transcription factor induced DNA domain formation provides a natural framework for understanding the physical structure of the E. coli chromosome. *Nuc. Acids Res.*, (in peer-review), 2011.

[121] Austin R.H. Wang, Y.M. and E.C. Cox. Single molecule measurements of repressor protein 1D diffusion on DNA. *Phys. Rev. Lett.*, 97:048302, 2006.

[122] R.B. Winter, O.G. Berg, and P.H. von Hippel. Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. the Escherichia coli lac repressor–operator interaction: kinetic measurements and conclusions. *Biochemistry*, 20:6961–77, 1981.

[123] J. Elf, G.W. Li, and X.S. Xie. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science*, 316:1191–1194, 2007.

[124] L. Mirny, M. Slutsky, Z. Wunderlich, A. Tafvizi, J. Leith, and A. Kosmrlj. How a protein searches for its site on DNA: the mechanism of facilitated diffusion. *J. Phys. A: Math. Theor.*, 42:434013, 2009.

[125] R. Overbeek, M. Fonstein, M. D'Souza, G.D. Pusch, and N. Maltsev. The use of gene clusters to infer functional coupling. *Proc. Natl. Acad. Sci. U. S. A.*, 96:2896–2901, 1999.

[126] J. Rouquette, C. Cremer, T. Cremer, and S. Fakan. Functional nuclear architecture studied by microscopy: present and future. *Int. Rev. Cell Mol. Biol.*, 282:1–90, 2010.

[127] J. Dostie, T.A. Richmond, R.A. Arnaout, R.R. Selzer, W.L. Lee, T.A. Honan, E.D. Rubio, A. Krumm, J. Lamb, C. Nusbaum, R.D. Green, and J. Dekker. Chromosome Conformation Capture Carbon Copy (5C): A massively parallel solution for mapping interactions between genomic elements. *Genome Research*, 16:1299–1309, 2006.

[128] J. Dostie and J. Dekker. Mapping networks of physical interactions between genomic elements using 5C technology. *Nature Protocols*, 2:988–1002, 2007.

[129] M. Simonis, P. Klous, E. Splinter, Y. Moshkin, R. Willemsen, E. de Witt, B. van Steensel, and W. de Laat. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat. Genet.*, 38:1348–1354, 2006.

[130] Z. Zhao, G. Tavoosidana, M. Sjölinder, A. Göndör, P. Mariano, S. Wang, C. Kanduri, and M. Lezcanoand K.S. Sandhu et al. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nature Genetics*, 38:1341–47, 2006.

[131] J. Dekker, K. Rippe, M. Dekker, and N. Kleckner. Capturing Chromosome Conformation. *Science*, 15:1306–11, 2002.

[132] Wikipedia, The Free Encyclopedia. Chromosome Conformation Capture Technology.jpg. `http://upload.wikimedia.org/wikipedia/commons/d/d2/Chromosome_Conformation_Capture_Technology.jpg`. [Online; accessed July-2011].

[133] D.W. Galbraith. The rainbow of fluorescent proteins. *Methods Cell Biol.*, 75:153–69, 2004.

[134] P.G. de Gennes. *Scaling Concepts in Polymer Physics.* Cornell University Press, Ithaca, London, 1979.

[135] P.J. Flory. *Principles of Polymer Chemistry.* Cornell University Press, Ithaca, NY, 1953.

[136] J. de Cloizeaux and G. Jannink. *Polymers in Solution: Their Modeling and Structure.* Clarendon Press, Oxford, 1990.

[137] A. Grosberg and A.R. Khokhlov. *Statistical Physics of Macromolecules.* AIP Press, Woodbury, NY, 1994.

[138] M. Doi and S.F. Edwards. *The Theory of Polymer Dynamics.* Oxford University Press, Oxford, 1999.

[139] M. Rubinstein and R.H. Colby. *Polymer Physics.* Oxford University Press, Oxford, 2003.

[140] U.W. Gedde. *Polymer Physics.* Springer, Heidelberg, 2007.

[141] G. Strobl. *The Physics of Polymers.* Springer, Berlin, 2007.

[142] P.J. Flory. *Statistical Mechanics of Chain Molecules.* Oxford University Press, Oxford, 1989.

[143] C.S. Tsai. *Biomacromolecules: Introduction to Structure, Function and Informatics.* Wiley, Canada, 2006.

[144] N. Madras and G. Slade. *The Self-Avoiding Walk.* Birkhäuser, Boston, 1993.

[145] L. Schaefer. *Excluded Volume Effects in Polymer Solutions as Explained by the Renormalization Group.* Springer, Berlin, 1999.

[146] K. Binder and D.W. Heermann. *Monte Carlo Simulation in Statistical Physics: An Introduction.* Springer, Berlin, 5th edition, 2010.

[147] K. Binder. *Monte Carlo and Molecular Dynamics Simulations in Polymer Science.* Oxford University Press, New York, 1995.

[148] http://de.wikipedia.org/wiki/Polymer. [Online; accessed June-2011].

[149] W.L. Mattice and U.W. Suter. *Conformational Theory of Large Molecules.* Wiley, New York, 1994.

[150] W. Bruns. Angular-correlations within coiled polymer-molecules. *Macromolecular Theory and Simulations*, 1:287–293, 1992.

[151] K. Šolc and W.H. Stockmayer. Shape of a random flight chain. *J. Chem. Phys.*, 54:2756–2757, 1971.

[152] W. Feller. *An introduction to probability theory and its applications.* Wiley, New York, NY, 1968.

[153] B.J. Kirby. *Micro- and Nanoscale Fluid Mechanics: Transport in Microfluidic Devices.* Cambridge University Press, Cambridge, 2010.

[154] C. Bustamante, J.F. Marko, E.D Siggia, and S. Smith. Entropic elasticity of lambda-phage DNA. *Science*, 265:1599–1600, 1994.

[155] J. Käs, H. Strey, J.X. Tang, D. Finger, R. Ezzell, E. Sackmann, and P.A. Janmey. F-actin, a model polymer for semiflexible chains in dilute, semidilute, and liquid crystalline solutions. *Biophys. J.*, 70:609–625, 1996.

[156] C.K. Ober. Shape Persistence of Synthetic Polymers. *Science*, 288:488–489, 2000.

[157] H.P. Hsu, W. Paul, and K. Binder. Polymer chain stiffness vs. excluded volume: A Monte Carlo study of the crossover towards the worm-like chain model. *EPL*, 92:28003–28009, 2010.

[158] J.C. Le Guillou and J. Zinn-Justin. Critical exponents from field theory. *Phys. Rev. B*, 21:3976–3998, 1980.

[159] P.M. Diesinger. *Genome Folding at the 30nm Scale.* PhD thesis, University of Heidelberg, 2009.

[160] shutterstock images. DNA 1. http://www.shutterstock.com/s/dna/search.html#id=17004592. [Online; accessed July-2011].

[161] M.P. Allen. Introduction to molecular dynamics simulation. In N. Attig, K. Binder, and H. Grubmuller, editors, *Computational Soft Matter: From Synthetic Polymers to Proteins*, volume 23 of NIC Series. John von Neumann Institute for Computing Jülich, 2004.

[162] M. Bachmann. Monte Carlo Simulations. In *42nd IFF Spring School: Macromolecular Systems in Soft and Living Matter*, Lecture Notes. Forschungszentrum Jülich, 2011.

[163] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equation of state calculation by fast computing machines. *Journal of Chemical Physics*, 21:1087–1092, 1953.

[164] A.D. Sokal. Monte Carlo Methods in Statistical Mechanics: Foundations and New Algorithms. In *Functional Integration: Basics and Applications*, Lecture Notes. Springer, 1997.

[165] M. Bohn. *private communication*, 2010.

[166] I. Carmesin and K. Kremer. The Bond Fluctuation Method: A New Effective Algorithm for the Dynamics of Polymers in All Spatial Dimensions. *Macromolecules*, 21:2819–2823, 1988.

[167] H.P. Deutsch and K. Binder. Interdiffusion and self-diffusion in polymer mixtures: A Monte Carlo study. *J. Chem. Phys.*, 94:2294–2304, 1990.

[168] P.E. Rouse. A Theory of the Linear Viscoelastic Properties of Dilute Solutions of Coiling Polymers. *J. Chem. Phys.*, 21:1272–1280, 1953.

[169] P.G. de Gennes. Reptation of a Polymer Chain in the Presence of Fixed Obstacles. *J. Chem. Phys.*, 55:572–579, 1971.

[170] N. Madras and A. Sokal. Nonergodicity of local, length-conserving Monte Carlo algorithms for the self-avoiding walk. *J. Stat. Phys.*, 47:573–595, 1987.

[171] P.M. Diesinger and D.W. Heermann. Depletion Effects massively change Chromatin Properties and influence Genome Folding. *Biophys. J.*, 97(8):2146–53, 2009.

[172] K. Bystricky, P. Heun, L. Gehlen, J. Langowski, and S.M. Gasser. Long-range compaction and flexibility of interphase chromatin in budding yeast analyzed by high-resolution imaging techniques. *Proc. Natl. Acad. Sci. U. S. A.*, 101:16495–500, 2004.

[173] P.M. Diesinger and D. W. Heermann. Average crossing number of Gaussian and equilateral chains with and without excluded volume. *EPJ*, 62:209–214, 2008.

[174] L.A. Mirny. The fractal globule as a model of chromatin architecture in the cell. *Chromosome Res.*, 19:37–51, 2011.

[175] M. Fritsche, P.M. Diesinger, D.W. Heermann, and M. Bathe. A Maximum Entropy Approach for 3D Genome Conformation Reconstruction. *Proc. Natl. Acad. Sci. U. S. A.*, (to be submitted), 2011.

[176] S. Milner. Polymer Brushes. *Science*, 251:905–914, 1991.

[177] A. Halperin, M. Tirrell, and T. Lodge. *Tethered chains in polymer microstructures*, volume 100 of *Advances in Polymer Science*. Springer, Berlin, 1992.

[178] S.T. Milner, T.A. Witten, and M.E. Cates. Theory of the grafted polymer brush. *Macromolecules*, 21:2610–2619, 1988.

[179] H.S. Gulati, C.K. Hall, R.L. Jones, and R.J. Spontak. Equilibrium conformations and dynamic relaxation of double-tethered chain molecules at an impenetrable interface. *J. Chem. Phys*, 105:7712, 1996.

[180] R.L. Jones and R.J. Spontak. Conformational analysis of double-tethered chain molecules at an impenetrable interface: A Monte Carlo study. *J. Chem. Phys*, 101:5179, 1994.

[181] K. Shida, A. Kasuya, and Y. Kawazoe. Entropic and conformational study of isolated double-tethered chains by three-dimensional lattice simulations. *Physica A*, 382:405–414, 2007.

[182] J. Huang, W. Jiang, and S. Han. Dynamic Monte Carlo Simulation on the Polymer Chain with One End Grafted on a Flat Surface. *Macromol. Theory Simul.*, 10:339–342, 2001.

[183] R. Reyes-Lamothe, X. Wang, and D. Sherratt. Escherichia coli and its chromosome. *Trends in Microbiology*, 16:238–245, 2008.

[184] S.M. Parkhurst. *private communication*, 2010.

[185] M. Fritsche, D.W. Heermann, M. Dutra, and C.E. Cordeiro. Conformational and Dynamical Properties of the Isolated, Three-Dimensional Single- and Double-Tethered Polymer Chain on an Infinite Surface. *Macromolecular Theory & Simulation*, 19:440–448, 2010.

[186] W. Bruns. Angular-correlations within coiled polymer-molecules. *Macromol. Chem. Theory Simul.*, 1:287–293, 1992.

[187] G. Zifferer. Shape distribution and correlation between size and shape of tetrahedral lattice chains in athermal and theta systems. *J. Chem. Phys.*, 109:3691, 1998.

[188] O. Jagodzinski, E. Eisenriegler, and K. Kremer. Universal shape properties of open and closed polymer chains: renormalization group analysis and Monte Carlo experiments. *J. Phys. I*, 2:2243, 1992.

[189] J.W. Cannon, J.A. Aronovitz, and P. Goldbart. Equilibrium distribution of shapes for linear and star macromolecules. *J. Phys. I*, 1:629–645, 1991.

[190] D. Marenduzzo, C. Micheletti, and E. Orlandini. Biopolymer organization upon confinement. *J. Phys.: Condens. Matter*, 23:283102, 2010.

[191] E. Orlandini, M.C Tesi, S.G. Whittington, D.W. Summers, and E.J. Janse van Rensburg. The writhe of a self-avoiding walk. *J. Phys. A.: Math. Gen.*, 27:L333, 1994.

[192] E.J. Janse van Rensburg, D.W. Summers, E. Wassermann, and S.G. Whittington. Entanglement complexity of self-avoiding walks. *J. Phys. A.: Math. Gen.*, 25:6557, 1992.

[193] G.A. Arteca. Small- and large-scale global shape features in macromolecular backbones. *J. Math. Chem.*, 12:37–51, 1993.

[194] G.A. Arteca. Overcrossing spectra of protein backbones: Characterization of three-dimensional molecular shape and global structural homologies. *Biopolymers*, 33:1829–1841, 1993.

[195] P. Fraser. Transcriptional control thrown for a loop. *Current Opinion in Genetics & Development*, 16:490–495, 2006.

[196] I. Junier, O. Martin, and F. Képès. Spatial and Topological Organization of DNA Chains Induced by Gene Co-localization. *PLoS Comp. Biol.*, 6:e1000678, 2010.

[197] H. Kimura and P.R. Cook. *Dynamic chromatin loops and the regulation of gene expression.* Springer, Japan, 2007.

[198] T. Cremer and C. Cremer. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat. Rev.*, 2:292–301, 2001.

[199] P. R. Cook. A model for all genomes; the role of transcription factories. *J. Mol. Biol.*, 395:1–10, 2010.

[200] A. Papantonisa and P.R. Cook. Genome architecture and the role of transcription. *Current Opinion in Cell Biology*, 22:271–276, 2010.

[201] C.S. Osborne, L. Chakalova, K.E. Brown, D. Carter, A. Horton, E. Debrand, and B. Goyenechea et al. Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat. Genet.*, 36:1065–1071, 2004.

[202] W.F. Marshall, A. Straight, J.F. Marko, J. Swedlow, A. Dernburg, and A. Belmont et al. Interphase chromosomes undergo constrained diffusional motion in living cells. *Current Biology*, 7:930–939, 1997.

[203] J.R. Chubb, S Boyle, P. Perry, and W.A. Bickmore. Chromatin Motion Is Constrained by Association with Nuclear Compartments in Human Cells. *Current Biology*, 12:439–445, 2002.

[204] C. Zimmer and E. Fabre. Principles of chromosomal organization: lessons from yeast. *JCB*, 192:723–33, 2011.

[205] L. Guelen, L Pagie, E. Brasset, W Meulemann, M.B. Faza, and W Talhout et al. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature*, 453:948–951, 2008.

[206] M. Thanbichler, S.C. Wang, and L. Shapiro. The bacterial nucleoid: A highly organized and dynamic structure. *Journal of Cellular Biochemistry*, 96:506–521, 2005.

[207] S. Cunha, C.L. Woldringh, and T. Odijk. Polymer-mediated compaction and internal dynamics of isolated Escherichia coli nucleoids. *J.Struct.Biol.*, 136:53–66, 2001.

[208] D. Zink, T. Cremer, R. Saffrich, R. Fischer, M.F. Trendelenburg, W. Ansorge, and E.H. Stelzer. Structure and dynamics of human interphase chromosome territories in vivo. *Human Genetics*, 102:241–251, 1998.

[209] S. Boyle, S. Gilchrist, J.M. Bridger, N.L. Mahy, J.A. Ellis, and W.A. Bickmore. The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Human Molecular Genetics*, 10:211–219, 2001.

[210] J.A. Croft, J.M. Bridge, S. Boyle, P. Perry, P. Teague, and W.A. Bickmore. Differences in the localization and morphology of chromosomes in the human nucleus. *J.Cell Biol.*, 145:1119–1131, 1999.

[211] C. Lanctôt, T. Cheutin, M. Cremer, G. Cavalli, and T. Cremer. Dynamic genome architecture in the nuclear space: Regulation of gene expression in three dimensions. *Nature*, 8:104–115, 2007.

[212] J. Dorier and A. Stasiak. Topological origins of chromosomal territories. *Nuc. Acids Res.*, 37:6316, 2009.

[213] G.G. Cabal, A. Genovesio, S. Rodriguez-Navarro, C. Zimmer, O. Gadal, A. Lesne, H. Buc, F. Feuerbach-Fournier, J.-C. Olivo-Marin, E.C. Hurt, and U. Nehrbass. Saga interacting factors confine sub-diffusion of transcribed genes to the nuclear envelope. *Nature*, 441:770–773, 2006.

[214] D.A. Drubin, A.M. Garakani, and P.A. Silver. Motion as a phenotype: the use of live-cell imaging and machine visual screening to characterize transcription-dependent chromosome dynamics. *Cell Biology*, 7:19, 2006.

[215] P. Heun, T. Laroche, K. Shimada, P. Furrer, and S.M. Gasser. Chromosome dynamics in the yeast interphase nucleus. *Science*, 294:2181–86, 2001.

[216] D.A. Bressan, J. Vazquez, and J.E. Haber. Mating type-dependent constraints on the mobility of the left arm of yeast chromosome III. *J. Cell Biol.*, 164:361–371, 2004.

[217] A. Taddei, G. van Houwe, F. Hediger, V. Kalck, F. Curbizolles, H. Schober, and S.M. Gasser. Nuclear pore association confers optimal expression levels for an inducible yeast gene. *Nature*, 411:774–778, 2006.

[218] J.M. Bridger, N. Foeger, I.R. Kill, and H. Herrmann. The Nuclear Lamina: both a structural framework and a platform for genome organization. *FEBS Journal*, 274:1354–1361, 2007.

[219] T. Misteli. Spatial positioning: A new dimension in genome function. *Cell*, 119:153–156, 2004.

[220] M.M. Cockell and S.M. Gasser. The nucleolus: nucleolar space for rent. *Currrent Biology*, 9:R575–R576, 1999.

[221] J. Kindt, S. Tzlil, S. Ben-Shaul, and W. Gelbart. DNA packaging and ejection forces in bacteriophage. *Proc. Natl. Acad. Sci. U. S. A.*, 98:13671–4, 2001.

[222] T. Odijk. Statistics and dynamics of condensed DNA within phages and globules. *Phil. Trans. R. Soc. Lond.*, 362:1497–1517, 2004.

[223] I. Ali, D. Marenduzzo, and J. Yeomans. Dynamics of polymer packaging. *J. Chem. Phys.*, 121:8635–41, 2004.

[224] W. Jiang, J. Chang, J. Jakana, J. K. P. Weigele, and W. Chiu. Structure of epsilon 15 bacteriophage reveals genome organization and DNA packaging/injection apparatus. *Nature*, 439:612–616, 2006.

[225] E. Katzav, M. Adda-Bedia, and A. Boudaoud. A statistical approach to close packing of elastic rods and to DNA packaging in viral capsids. *Proc. Natl. Acad. Sci. U. S. A.*, 103:18900–904, 2006.

[226] A. S. Petrov, M. B. Boz, and S. C. Harvey. The conformation of double-stranded DNA inside bacteriophages depends on capsid size and shape. *J. Struct. Biol.*, 160:241–248, 2007.

[227] T. Hu, A. Y. Grosberg, and B. I. Shklovskii. How Proteins Search for Their Specific Sites on DNA: The Role of DNA Conformation. *Biophys. J.*, 90:2731–2744, 2006.

[228] M. Muthukumar. Translocation of a confined polymer though a hole. *Phys. Rev. Lett.*, 86:3188, 2001.

[229] A. Arnold, B. Bozorgui, D. Frenkel, B. Y. Ha, and S. Jun. Unexpected relaxation dynamics of a self-avoiding polymer in cylindrical confinement. *J. Chem. Phys.*, 127:164903–12, 2007.

[230] A. Liu and B. Chakraborty. Shapes of semiflexible polymers in confined spaces. *Phys. Biol.*, 5:026004, 2008.

[231] J Kalb and B. Chakraborty. Single polymer confinement in a tube: Correlation between structure and dynamics. *J. Chem. Phys*, 130:025103–12, 2009.

[232] F. Thüroff, F. Wagner, and E. Frey. The effect of internal and global modes on the radial distribution function of confined semiflexible polymers. *EPL*, 91:38004, 2010.

[233] K. Alim and E. Frey. Fluctuating semiflexible polymer ribbon constrained to a ring. *Eur. Phys. J. E*, 24:185–191, 2007.

[234] K. Alim and E. Frey. Shapes of Semiflexible Polymer Rings. *Phys. Rev. Lett.*, 99:198102, 2007.

[235] F. Drube, K. Alim, G. Witz, G. Dietler, and E. Frey. Excluded volume effects on semiflexible ring polymers. *Nano Lett.*, 10:1445–9, 2010.

[236] F. Wagner, G. Lattanzi, and E. Frey. Conformations of confined biopolymers. *Phys. Rev. E*, 75:050902, 2007.

[237] Y. Yang, T. W. Burkhardt, and G. Gompper. Free energy and extension of a semiflexible polymer in cylindrical confining geometries. *Phys. Rev. E*, 76:011804, 2007.

[238] P. Cifra. Channel confinement of flexible and semiflexible macromolecules. *J. Chem. Phys.*, 131:224903–10, 2009.

[239] J. J. Cerdá, T. Sintes, and A. Chakrabarti. Excluded volume effects on polymer adsorption onto spherical surfaces. *Macromolecules*, 38:1469–1477, 2005.

[240] C.-H. Lin, Y.-C. Tsai, and C.-K. Hu. Wrapping conformations of a polymer on a curved surface. *Phys. Rev. E*, 75:031903, 2007.

[241] A.J. Spakowitz and Z.G. Wang. Semiflexible Polymer Confined to a Spherical Surface. *Phys. Rev. Lett.*, 91:166102, 2003.

[242] L. Limozin and E. Sackmann. Polymorphism of Cross-Linked Actin Networks in Giant Vesicles. *Phys. Rev. Lett.*, 89:168103, 2002.

[243] M.M.A.E. Claessens, M. Bathe, E. Frey, and A.R. Busch. Actin-binding proteins sensitively mediate F-actin bundle stiffness. *Nature Materials*, 5:748–753, 2006.

[244] T. Sanchez, I. M. Kulic, and Z. Dogic. Circularization, Photomechanical Switching, and a Supercoiling Transition of Actin Filaments. *Phys. Rev. Lett.*, 104:098103, 2010.

[245] I.F. Lau, S.R. Filipe, B. Søballe, O.-A. Økstad, B. Francois-Xavier, and D.J. Sherratt. Spatial and temporal organization of replicating Escherichia coli chromosomes. *Molecular Microbiology*, 49:731–741, 2003.

[246] Z. Benkova and P. Cifra. Stiffening Transition in Semiflexible Cyclic Macromolecules. *Macromolecular Theory and Simulation*, 20:65–74, 2010.

[247] S. Araki, T. Nakai, K. Hizume, K. Takeyasu, and K. Yoshikawa. Hydrodynamic radius of circular DNA is larger than that of linear DNA. *Chem. Phys. Lett.*, 418:255–259, 2006.

[248] H.P. Hsu, W. Paul, and K. Binder. Standard Definitions of Persistence Length Do Not Describe the Local "Intrinsic" Stiffness of Real Polymer Chains. *Macromolecules*, 43:3094–3102, 2010.

[249] J. Wilhelm and E. Frey. Radial distribution function of semiflexible polymers. *Phys. Rev. Lett.*, 77:2581, 1996.

[250] N. Mücke, K. Klenin, R. Kirmse, M. Bussiek, H. Herrmann, M. Hafner, and J. Langowski. Filamentous Biopolymers on Surfaces: Atomic Force Microscopy Images Compared with Brownian Dynamics Simulation of Filament Deposition. *PLoS One*, 4:e7756, 2009.

[251] S. Koester, D. Steinhauser, and T. Pfohl. Brownian motion of actin filaments in confining microchannels. *J. Phys.: Condens. Matter*, 17:S4091, 2007.

[252] S. Koester, J. Kierfeld, and T. Pfohl. Characterization of single semiflexible filaments under geometric constraints. *Eur. Phys. J. E*, 25:439–449, 2008.

[253] T. Odijk. The statistics and dynamics of confined or entangled stiff polymers. *Macromolecules*, 16:1340–1344, 1983.

[254] M. Emanuel, N.H. Radja, A. Henriksson, and H. Schiessel. The physics behind the larger scale organization of DNA in eukaryotes. *Phys. Biol.*, 6:025008, 2009.

[255] D. Reith and P. Virnau. Monte Carlo Simulation of a single polystyrene chain in spherical confinement. *Computer Physics Communications*, 182:1945–1948, 2011.

[256] K. Klenin and J. Langowski. Computation of Writhe in Modeling of Supercoiled DNA. *Biopolymers*, 54:307–317, 2000.

[257] S. Kumar. *Liquid Cristals: Experimental Study of Physical Properties and Phase Transitions*. Cambridge University Press, Cambridge, 2007.

[258] N.N. Oskolkov, P. Linse, I.I. Potemkin, and A.R. Khokhlov. Nematic Ordering of Polymers in Confined Geometry Applied to DNA Packaging in Viral Capsids. *J. Chem. Phys.*, 115:422–432, 2011.

[259] V. Renugopalakrishnan, R. Garuňo-Juárez, G. Narasimhan, C.S. Verma, X. Wei, and P. Li. Rational Design of Thermally Stable Proteins: Relevance to Bionanotechnology. *J. Nanosci. Nanotech.*, 5:1759–67, 2005.

[260] W. Huang, S. Krishnaji, X. Hu, D. Kaplan, and P. Cebe. Heat capacity of spider silk-like block copolymers. *Macromolecules*, 44:5299–309, 2011.

[261] N. Wang, D.M. Mattis, E.J. Sundberg, P.M. Schlievert, and D.M. Kranz. A Single, Engineered Protein Therapeutic Agent Neutralizes Exotoxins from Both Staphylococcus aureus and Streptococcus pyogenes. *Clinical and Vaccine Immunology*, 17:1781–1789, 2010.

[262] S. Kobsa and W.M. Saltzman. Bioengineering Approaches to Controlled Protein Delivery. *Pediatric Research*, 63:513–519, 2008.

[263] P. Asuri, S.S. Bale, S.S. Karajanagi, and R.S. Kan. The protein-nanomaterial interface. *Curr. Opinion Biotechnology*, 17:562–568, 2006.

[264] R.S. Kane and A.D. Stroock. Nanobiotechnology: Protein-Nanomaterial Interactions. *Biotechnology Progress*, 23:316–319, 2007.

[265] A. Korkegian, M.E. Black, D. Baker, and B.L. Stoddard. Computational Thermostabilization of an Enzyme. *Science*, 308:857, 2005.

[266] H.K. Liang, C.M. Huang, M.T. Ko, and J.K. Hwang. Amino Acid Coupling Patterns in Thermophilic Proteins. *Proteins*, 59:58, 2005.

[267] R. Das and M. Gerstein. The stability of thermophilic proteins: a study based on comprehensive genome comparison. *Funct. Integr. Genomics*, 1:76, 2000.

[268] T.T. Paull, E.P. Rogakou, Y. Yamazaki, C.U. Kirchgessner, M. Gellert, and W.M.A. Bonner. A critical role for histone H2AX in recruitment of repair factors to nuclear foci after DNA damage. *Curr. Biol.*, 10:886–895, 2000.

[269] O. Fernandez-Capitello, H.T. Chen, A. Celeste, I. Ward, and P.J. Romanienko et al. DNA damage-induced G2-M checkpoint activation by histone H2AX and 53BP1. *Nature Cell Biol.*, 4:993–997, 2002.

[270] J. Bewersdorf, B.T. Bennett, and K.L. Knight. H2AX chromatin structures and their response to DNA damage revealed by 4Pi microscopy. *Proc. Natl. Acad. Sci. U. S. A.*, 103:18137–18142, 2006.

[271] T. Ikura, S. Tashiro, A. Kakino, and H. Shima et al. DNA Damage-Dependent Acetylation and Ubiquitination of H2AX Enhances Chromatin Dynamics. *Mol. and Cell. Biol.*, 27:7028–40, 2007.

[272] B. van den Burg, A. de Kreij, P. van der Veek, J. Manfeld, and G. Venema. Characterization of a novel stable biocatalyst obtained by protein engineering. *Biotech. Appl. Biochem.*, 30:35, 1999.

[273] T.J. Graddis, R.L. Remmele, and J.T. McGrew. *Current Pharmaceutical Biotechnology*, 3:285, 2002.

[274] G.N. Somero. Proteins and Temperature. *Annu. Rev. Physiol.*, 57:43–68, 1995.

[275] M. Smerdon and I. Isenberg. The Effect of Temperature on Histone GRK Aggregation. *Biochemical and Biophysical Research Communications*, 55:1029–34, 1973.

[276] R. Sterner and W. Liebl. Thermophilic adaptation of proteins. *Crit. Rev. Biochem. Mol. Biol.*, 36:39, 2001.

[277] K.O. Stetter. Hyperthermophilic procaryotes. *FEMS Microbiol. Rev.*, 18:149–58, 1996.

[278] N. Ayoub, A.D. Jeyasekharan, J.A. Bernal, and A.R. Venkitaraman. HP1-β mobilization promotes chromatin changes that initiate the DNA damage response. *Nature*, 453:682–686, 2008.

[279] C. Thiriet and J.J. Hayes. Chromatin in Need of a Fix: Phosphorylation of H2AX Connects Chromatin to DNA Repair. *Mol. Cell*, 18:617–622, 2005.

[280] M. Stucki and S.P. Jackson. γH2AX and MDC1: Anchoring the DNA-damage-response machinery to broken chromosomes. *DNA Repair*, 5:534–543, 2006.

[281] C.L. Peterson and J. Côté. Cellular machineries for chromosomal DNA repair. *Genes Dev.*, 18:602–616, 2004.

[282] N. Kongruttanachok, C. Phuangphairoj, W. Ponyeam, and A. Mutirangura. Temperature dependent y-H2AX binding to DNA. *Science Asia*, 34:253–257, 2008.

[283] R.B. Pandey and B.L. Farmer. Globular structure of a human immunodeficiency virus-1 protease (1DIFA dimer) in an effective solvent medium by a Monte Carlo simulation. *J. Chem. Phys.*, 132:125101–05, 2010.

[284] D.M. Eby, G.R. Johnson, B.L. Farmer, and R.B. Pandey. Supramolecular assembly of a biomineralizing antimicrobial peptide in coarse-grained Monte Carlo simulations. *Phys. Chem. Chem. Phys.*, 13:1123–30, 2011.

[285] The Jernigan Laboratory. Knowledge-based potential server for proteins at coarse grained level. `http://gor.bb.iastate.edu/potential/`. [Online; accessed November-2010].

[286] S. Miyazawa and R.L. Jernigan. Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation. *Macromolecules*, 18:534–552, 1985.

[287] R.B. Pandey and B.L. Farmer. Conformation of a coarse-grained protein chain (an aspartic acid protease) model in effective solvent by a bond-fluctuating Monte Carlo simulation. *Phys. Rev. E*, 77:031902–10, 2008.

[288] J. Feng, R.B. Pandey, R.J. Berry, B.L. Farmer, R.R. Naik, and H. Heinz. Adsorption mechanism of single amino acid and surfactant molecules to Au 111 surfaces in aqueous solution: design rules for metal-binding molecules. *Soft Matter*, 7:2113–20, 2011.

[289] J. Bjursell and R.B. Pandey. Relaxation to native conformation of a bond-fluctuating protein chain with hydrophobic and polar nodes. *Phys. Rev. E*, 70:052904, 2004.

[290] C.-M. Chen and C.-C. Chen. Computer Simulations of Membrane Protein Folding: Structure and Dynamics. *Biophys. Journal*, 84:1902, 2003.

[291] S. Miyazawa and R.L. Jernigan. Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *J. Mol. Biol.*, 256:623–44, 1996.

[292] H. Behringer, A. Degenhard, and F. Schmid. Coarse-Grained Lattice Model for Molecular Recognition. *Phys. Rev. Lett.*, 97:128101–04, 2006.

[293] M. Skepo, P. Linse, and T. Arnebrant. Competition between a Macroion and a Polyelectrolyte in Complexation with an Oppositely Charged Polyelectrolyte. *J. Phys. Chem.*, 110:12141–48, 2006.

[294] Y. Zhou and M. Karplus. Interpreting the folding kinetics of helical proteins. *Nature*, 401:400–403, 1999.

[295] M. Shen and K.F. Freed. All-atom fast protein folding simulations: The villin headpiece. *Proteins*, 49:439–445, 2002.

[296] Y. Wei, W. Nadler, and U.H.E. Hansmann. Backbone and side-chain ordering in a small protein. *J. Chem. Phys.*, 128:025105, 2008.

[297] P.P. Chapagain and B. Gertsman. Removal of kinetic traps and enhanced protein folding by strategic substitution of amino acids in a model alpha-helical hairpin peptide. *Biopolym.*, 81:167–78, 2006.

[298] B.S. Gerstman and P.P. Chapagain. Self-organization in protein folding and the hydrophobic interaction. *J. Chem. Phys.*, 123:054901–6, 2005.

[299] M. Fritsche, R.B. Pandey, B.L. Farmer, and D.W. Heermann. Multi-Scale Structural Dynamics of a Histone H2AX: A Coarse-Grained Monte Carlo Approach Via Knowledge-based Interaction Potentials. *Biomacromolecules*, (in peer-review), 2011.

[300] R.B. Pandey and B.L. Farmer. 2011.

[301] E.P. Rogakou, D.R. Pilch, A.H. Orr, V.S. Ivanova, and W.M. Bonner. DNA Double-stranded Breaks Induce Histone H2AX Phosphorylation on Serine 139. *J. Biol. Chem.*, 273:5858–68, 1998.

[302] A.L. Marston and A. Amon. Meiosis: Cell-Cycle Controls Shuffle and Deal. *Nature Reviews Molecular Cell Biology*, 5:983–97, 2004.

[303] H. Scherthan. A bouquet makes ends meet. *Nature Reviews Molecular Cell Biology*, 2:621–27, 2001.

[304] C.-J.R. Wang, P.M. Carlton, and I.N. Golubovskaya. Interlock Formation and Coiling of Meiotic Chromosome Axis During Synapsis. *Genetics*, 183:905–15, 2009.

[305] P.M. Carlton. Three-dimensional structured illumination microscopy and its application to chromosome structure. *Chromosome Research*, 16:351–65, 2008.

[306] E.J. White, C. Cowan, W.Z. Cande, and D.B. Kaback. In vivo analysis of synaptonemal complex formation during yeast meiosis. *Genetics*, 167:51–63, 2004.

[307] M.E. Dresser. Chromosome Mechanics and Meiotic Engine Maintenance. *PLoS Genetics*, 4:e1000210, 2008.

[308] R. Koszul, K.P. Kim, M. Prentiss, N. Kleckner, and S. Kameoka. Actin-mediated motion of meiotic chromosomes. *Cell*, 133:1188–1201, 2008.

[309] S.L. Page and R.S. Hawley. The Genetics and Molecular Biology of the Synaptonemal Complex. *Annual Review of Cell and Developmental Biology*, 20:525–58, 2004.

[310] M. Nicodemi, B. Panning, and A. Prisco. A Thermodynamics Switch for Chromosome Colocalization. *Genetics*, 179:717–21, 2008.

[311] D. Jahn, S. Schramm, R. Benavente, and M. Alsheimer. Dynamic properties of meiosis-specific lamin C2 and its impact on nuclear envelope integrity. *Nucleus*, 1:273–283, 2010.

[312] H. Scherthan, A. Sfeir, and T. Lange. Rap1-independent telomere attachment and bouquet formation in mammalian meiosis. *Chromosoma*, 120:151–57, 2011.

[313] M. Born and E. Wolf. *Principles of Optics*. Cambridge University Press, Cambridge, 7th edition, 1999.

[314] S.W. Hell, E.H.K. Stelzer, S. Lindek, and C. Cremer. Confocal microscopy with an increased detection aperture: type-B 4Pi confocal microscopy. *Optics Letters*, 19:222–224, 1994.

[315] R. Schmidt, C.A. Wurm, S. Jakobs, J. Engelhardt, A. Egner, and S.W. Hell. Spherical nanosized focal spot unravels the interior of cells. *Nature Methods*, 5:539–544, 2008.

[316] A. Egner, S. Jakobs, and S.W. Hell. Fast 100-nm resolution three-dimensional microscope reveals structural plasticity of mitochondria in live yeast. *Proc. Natl. Acad. Sci. U. S. A.*, 99:3370–3375, 2002.

[317] S. Hell and E.H.K. Stelzer. Properties of a 4Pi confocal fluorescence microscope. *J. Opt. Soc. Am. A*, 9:2159–2166, 1992.

[318] D.B. Schmolze, C. Standley, K.E. Fogarty, and A.H. Fischer. Advances in Microscopy Techniques. *Arch. Pathol. Lab Med.*, 135:255–263, 2010.

[319] N.M. Vranis, G.W. van der Heijden, S. Malki, and A. Bortvin. Synaptonemal Complex Length Variation in Wild-Type Male Mice. *Genes*, 1:505–20, 2010.

[320] M. Codina-Pascual, M. Campillo, J. Kraus, M.R. Speicher, J. Egozcue, J. Navarro, and J. Benet. Crossover frequency and synaptonemal complex length: their variability and effects on human male meiosis. *Mol. Hum. Reprod.*, 12:123–133, 2006.

[321] A. Storlazzi, S. Gargano, G. Ruprich-Robert, M. Falque, M. David, N. Kleckner, and D. Zickler. Recombination Proteins Mediate Meiotic Spatial Chromosome Organization and Pairing. *Cell*, 141:94–106, 2010.

[322] M. Fritsche, L. Rheinholdt, M. Lessard, M.A. Handel, J. Bewersdorf, and D.W. Heermann. Entropy-Driven Spatial Organization of Synaptonemal Complexes within the Cell Nucleus. *PLoS One*, (to be submitted), 2011.

[323] G.A. Arteca. Scaling behavior of some molecular shape descriptors of polymer chains and protein backbones. *Phys. Rev. E*, 49:2417–2428, 1993.

[324] P.B. Holm, S.W. Rasmussen, and D. von Wettstein. Ultrastructural characterization of the meiotic prophase. A tool in the assessment of radiation damage in man. *Mutat. Res.*, 95:45–59, 1982.

[325] S.W. Rasmussen. Initiation of synapsis and interlocking of chromosomes during zygotene in Bombyx spermatocytes. *Carlsberg Res. Commun.*, 51:401–432, 1986.

[326] P.B. Moens. Unravelling meiotic chromosomes: topoisomerases II and other proteins. *J. Cell Sci.*, 97:1–3, 1990.

[327] M.N. Conrad, C.-Y. Lee, G.Chao, M. Shinohara, H. Kosaka, and A. Shinohara et al. Rapid Telomere Movement in Meiotic Prophase Is Promoted By NDJ1, MPS3 and CSM4 and Is Modulated by Recombination. *Cell*, 133:1175–87, 2008.

[328] M. Rosenblatt. Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, 27:832–837, 1956.

[329] E. Parzen. On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33:1065–1076, 1962.

[330] W. Rieping, M. Habeck, and M. Nilges. Inferential structure determination. *Science*, 309:303–6, 2005.

[331] D. Rother, G. Sapiro, and V. Pande. Statistical characterization of protein ensembles. *IEEE/ACM TRANSACTIONS ON COMPUTATIONAL BIOLOGY AND BIOINFORMATICS*, 5:1–14, 2008.

[332] W.F. Marshall, A.F. Dernbug, B. Harmon, D.A. Agard, and J.W. Sedat. Specific interactions of chromatin with the nuclear envelope: Positional determination within the nucleus in drosophila melanogaster. *Mol.Biol. Cell*, 7:825–842, 1996.

[333] A. Akhtar and S.M. Gasser. The nuclear envelop and transcriptional control. *Nat. Rev. Genet.*, 8:507–17, 2007.

[334] M. Cockell and S.M. Gasser. Nuclear compartments and gene regulation. *Curr. Opin. Genet. Dev.*, 9:199–205, 1999.

[335] A.G. Fisher and M. Merkenschlager. Gene silencing, cell fate and nuclear organization. *Curr. Opin. Genet. Dev.*, 12:193–197, 2002.

[336] T. Misteli. Beyond the sequence: cellular organization of genome function. *Cell*, 128:787–800, 2007.

[337] A. Taddei, G. Van Houwe, S. Nagai, I. Erb, E. van Nimwegen, and S.M. Gasser. The functional importance of telomere clustering: global changes in gene expression result from SIR factor dispersion. *Genome Res.*, 19:611–25, 2009.

[338] K. Bystricky, H. Van Attikum, M.D. Montiel, V. Dion, L. Gehlen, and S.M. Gasser. Regulation of nuclear positioning and dynamics of the silent mating type loci by the yeast Ku70/Ku80 complex. *Mol Cell Biol.*, 29:835–48, 2009.

[339] Q.W. Jin, J. Fuchs, and J. Loidl. Centromere clustering is a major determinant of yeast interphase nuclear organization. *J. Cell Sci.*, 113:1903–1912, 2000.

[340] J. Ostashevsky. A polymer model for large-scale chromatin organization in lower eukaryotes. *Mol. Biol. Cell*, 13:2157–69, 2002.

[341] D.L. Spector. The dynamics of chromosome organization and gene regulation. *Annu. Rev. Biochem.*, 72:573–608, 2003.

[342] P. Therizols, T. Duong, B. Dujon, C. Zimmer, and E. Fabre. Chromosome arm length and nuclear constraints determine the dynamic relationship of yeast subtelomeres. *Proc. Natl. Acad. Sci. U. S. A.*, 107:2025–30, 2010.

[343] J.E. Haber. Mating-type gene switching in Saccharomyces cerevisiae. *Annu. Rev. Genet.*, 32:561–599, 1998.

[344] A. Miele, K. Bystricky, and J. Dekker. Yeast silent mating type loci form heterochromatic clusters through silencer protein-dependent long-range interactions. *PLoS Genet.*, 5:e1000478, 2009.

[345] Z. Duan, M. Andronescu, K. Schutz, S. McIlwain, Y.J. Kim, C. Lee, J. Shendure, and S. Fields et al. A three-dimensional model of the yeast genome. *Nature*, 465:363–367, 2010.

[346] J. Dekker. Mapping in Vivo Chromatin Interactions in Yeast Suggests an Extended Chromatin Fiber with Regional Variation in Compaction. *J. Biol. Chem.*, 283(50):34532–40, 2008.

[347] S. Weber, A. Spakowitz, and J. Theriot. Bacterial chromosomal loci move subdiffusively through a viscoelastic cytoplasm. *Phys. Rev. Lett.*, 104:238102, 2010.

[348] F. Hediger, F.R. Neumann, G. Van Houwe, K. Dubrana, and S.M. Gasser. Live imaging of telomeres: yKu and Sir proteins define redundant telomere-anchoring pathways in yeast. *Curr. Biol.*, 12:2076–89, 2002.

[349] W.F. Marshall. Order and disorder in the nucleus. *Current Biology*, 12:R185, 2002.

[350] J. Vazquez, A.S. Belmont, and J.W. Sedat. Multiple regimes of constrained chromosome motion are regulated in the interphase Drosophila nucleus. *Curr. Biol.*, 11:1227–39, 2001.

[351] I Bronstein, Y. Isreal, E. Kepten, S. Mai, Y. Shav-Tal, E. Barkai, and Y. Garini. Transient anomalous diffusion of telomeres in the nucleus of mammalian cells. *Phys. Rev. Lett.*, 103:018102, 2009.

[352] N. Gilbert, S. Gilchrist, and W.A. Bickmore. Chromatin organization in the mammalian nucleus. *Int. Rev. Cytol.*, 242:283–336, 2005.

[353] P.M. Carlton, J. Boulanger, C. Kervrann, J.B. Sibarita, J. Salamero, S. Gordon-Messer, D. Bressan, J.E. Haber, S. Haase, and L. Shao et al. Fast live simultaneous multiwavelength four-dimensional optical microscopy. *Proc. Natl. Acad. Sci. U. S. A.*, 107:16005–6, 2010.

[354] S. Gordon-Messer. *The organization and dynamics of a yeast chromosome arm.* PhD thesis, Brandeis University, 2010.

[355] S. Gordon-Messer, B. Avşaroğlu, M. Fritsche, J. Martin, P. Wiggins, J. Sedat, D.W. Heermann, J. Haber, and J. Kondev. Rabl model of yeast interphase chromosomes quantitatively accounts for their spatial organization and dynamics. *PLoS Comp. Biol.*, (to be submitted), 2011.

[356] S.M. Gasser. Visualizing chromatin dynamics in interphase nuclei. *Science*, 296:1412–1416, 2002.

[357] M. Winey, D. Yarar, T.H. Giddings, and D.N. Mastronarde. Nuclear pore complex number and distribution throughout the Saccharomyces cerevisiae cell cycle by three-dimensional reconstruction from electron micrographs of nuclear envelopes. *Mol. Biol. Cell*, 8:2119–32, 1997.

[358] J.E. Haber and W.Y. Leung. Lack of chromosome territoriality in yeast: promiscuous rejoining of broken chromosome ends. *Proc. Natl. Acad. Sci. U. S. A.*, 93:13949–54, 1996.

[359] A. Fiebig, K. Keren, and J.A. Theriot. Fine-scale time-lapse analysis of the biphasic, dynamic behaviour of the two Vibrio cholerae chromosomes. *Mol. Microbiol.*, 2006:1164–78, 2006.

[360] M. Lichten, R.H. Borts, and J.E. Haber. Meiotic gene conversion and crossing over between dispersed homologous sequences occurs frequently in Saccharomyces cerevisiae. *Genetics*, 115:233–246, 1987.

[361] T. Sexton, H. Schober, P. Fraser, and S. Gasser. Gene regulation through nuclear organization. *Nat. Struct. Mol. Biol.*, 14:1049–1055, 2007.

[362] G.S. Grest, K. Kremer, and T.A. Witten. Structure of many arm star polymers: a molecular dynamics simulation. *Macromolecules*, 20:1376–1383, 1987.

[363] G.S. Grest, K. Kremer, S.T. Milner, and T.A. Witten. Relaxation of self-entangled many-arm star polymers. *Macromolecules*, 22:1904–1910, 1989.

[364] C.N. Likos. Effective interactions in soft condensed matter physics. *Phys. Rep.*, 348:267–439, 2001.

[365] D. Vlassopoulos, G. Fytas, T. Pakula, and J. Roovers. Multiarm star polymers dynamics. *J. Phys. Condens. Matter*, 13:R855, 2001.

[366] R.M. Robertson, S. Laib, and D.E. Smith. Diffusion of isolated DNA molecules: Dependence on length and topology. *Proc. Natl. Acad. Sci. U. S. A.*, 103:7310, 2006.

[367] A. Takano, Y. Kushida, K. Aoki, K. Masuoka, and K. Hayashida et al. HPLC Characterization of Cyclization Reaction Product Obtained by End-to-End Ring Closure Reaction of a Telechelic Polystyrene. *Macromolecules*, 40:679–681, 2007.

[368] P. Reiss, M. Fritsche, and D.W. Heermann. Looped star polymers show conformational transition from spherical to flat toroidal shapes. *Phys. Rev. Lett*, (in peer-review), 2011.

[369] H.P. Hsu, W. Nadler, and P. Grassberger. Scaling of Star Polymers with 1-80 Arms. *Macromolecules*, 37:4658–4663, 2004.

[370] H.R. Brown. Chain Pullout and Mobility Effects in Friction and Lubrication. *Science*, 263:1411–1413, 1994.

[371] P. Reiss. *Quasi star polymers as a new model for the simulations of rosette-like chromosome structures*, Bachelor Thesis, University of Heidelberg, 2010.

[372] M.F. Zhang and A.H.E. Müller. Cylindrical polymer brushes. *J. Polym. Sci. Part A*, 43:3461–81, 2005.

[373] S. Rathgeber, T. Pakula, A. Wilk, K. Matyjaszewski, H. Lee, and K.L. Beers. Bottlebrush macromolecules in solution: Comparison between results obtained from scattering experiments and computer simulations. *Polymer*, 47:7318–27, 2006.

[374] M. Baiesi, E. Orlandini, and S. Whittington. Interplay between writhe and knotting for swollen and compact polymers. *J. Chem. Phys.*, 131:154902, 2009.

[375] N.M. Toan, D. Marenduzzo, P.R. Cook, and C. Micheletti. Depletion Effects and Loop Formation in Self-Avoiding Polymers. *Phys. Rev. Lett.*, 97:178302, 2006.

[376] J.-C. Meile, R. Mercier, M. Stouf, C. Pages, J.-Y. Bouet, and F. Cornet. The terminal region of the E. coli chromosome localises at the periphery of the nucleoid. *BMC Microbiology*, 11:1–10, 2011.

[377] J.D. Griffith. Visualization of prokaryotic DNA in a regularly condensed chromatin-like fiber. *Proc. Nat. Acad. Sci. U. S. A.*, 73:563–67, 1976.

[378] M. Meyer, M.A. De Jong, C.L. Woldringh, and N. Nanninga. Factors affecting the release of folded chromosomes from Escherichia coli. *Eur. J. Biochem.*, 63:469–75, 1976.

[379] R. Kavenoff and O.A. Ryder. Electron microscopy of membrane-associated folded chromosomes of Escherichia coli. *Chromosoma*, 55:13–25, 1976.

[380] S. Deng, R.A. Stein, and N.P. Higgins. Organization of supercoil domains and their reorganization by transcription. *Mol. Microbiol.*, 57:1511–21, 2005.

[381] J.H. Frenster. Correlation of the binding to DNA loops or to DNA helices with the effect on RNA synthesis. *Nature*, 208:1093, 1965.

[382] A. Danchin, P. Guerdoux-Jamet, I. Moszer, and P. Nitschké. Mapping the bacterial cell architecture into the chromosome. *Philos. Trans. R Soc. Lond. B Biol. Sci.*, 355:179–90, 2000.

[383] A.S. Carpentier, B. Torresani, A. Grossmann, and A. Henaut. Decoding the nucleoid organisation of Bacillus subtilis and Escherichia coli through gene expression data. *BMC Genomics*, 6:84, 2005.

[384] K.S. Jeong, J. Ahn, and A.B. Khodursky. Spatial patterns of transcriptional activity in the chromosome of Escherichia coli. *Genome Biol*, 5:R86, 2004.

[385] M. Bailly-Bechet, A. Danchin, M. Iqbal, M. Marsili, and M. Vergassola. Codon Usage Domains over Bacterial Chromosomes. *PLoS Comput. Biol.*, 2:e37, 2006.

[386] B.-K. Cho and B.Ø. Palsson. Can the Protein Occupancy Landscape Show the Topologically Isolated Chromosomal Domains in the E. coli Genome?: An Exciting Prospect. *Mol. Cell*, 35:255–56, 2009.

[387] I.A. Berlatzky, A. Rouvinski, and S. Ben-Yehuda. Spatial organization of a replicating bacterial chromosome. *Proc. Natl. Acad. Sci. U. S. A.*, 105:14136–140, 2008.

[388] D.B. Wilsonb P.L. Foleya and M.L. Shuler. Macromolecular crowding can account for RNase-sensitive constraint of bacterial nucleoid structure. *Biochem. Biophys. Res. Commun.*, 395:42–47, 2010.

[389] R.L. Ohniwa, K. Morikawa, S.L. Takeshita, J. Kim, T. Ohta, C. Wada, and K. Takeyasu. Transcription-coupled nucleoid architecture in bacteria. *Genes Cells*, 12:1141–1152, 2007.

[390] X. Liu, X. Wang, R. Reyes-Lamothe, and D. Sherratt. Replication-directed sister chromosome alignment in Escherichia coli. *Molecular Microbiol*, 75:1090–1097, 2010.

[391] D. Bates and N. Kleckner. Chromosome and replisome dynamics in E. coli: Loss of sister cohesion triggers global chromosome movement and mediates chromosome segregation. *Cell*, 121:899–911, 2005.

[392] M.C. Joshi, A. Bourniquel, J. Fisher, T.H. Brian, N. Kleckner, and D. Bates. Escherichia coli sister chromosome separation includes an abrupt global transition with concomitant release of late-splitting intersister snaps. *Proc. Natl. Acad. Sci. U. S. A.*, 108:2765–2770, 2011.

[393] H. Limbach, A. Arnold, B. Mann, and C. Holm. ESPResSo - an extensible simulation package for research on soft matter systems. *Comput. Phys. Commun.*, 174:704–727, 2006.

[394] D.C. Rapaport. *The Art of Molecular Dynamics Simulation.* XIV. Cambridge University Press, Cambridge, 1995.

[395] Y. Sugita and O. Yuko. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Letters*, 314:141–151, 1999.

[396] J.D. Weeks, D. Chandler, and H.C. Andersen. Role of repulsive forces in determining the equilibrium structure of simple liquids. *J. Chem. Phys.*, 54:5237–5247, 1971.

[397] A.M. Breier and N.R. Cozzarelli. Linear ordering and dynamic segregation of the bacterial chromosome. *Proc. Natl. Acad. Sci. U. S. A.*, 101:9175–9176, 2004.

[398] P.A. Wiggins. *private communication*, 2011.

[399] J. Stavans and A. Oppenheim. Dna-protein interactions and bacterial chromosome architecture. *Phys. Biol.*, 3:R1–10, 2006.

[400] R. Mercier, M. Petit, S. Schbath, S. Robin, M. El Karoui, F. Boccard, and O. Espéli. The MatP/matS site-specific system organizes the terminus region of the E. coli chromosome into a macrodomain. *Cell*, 135:475–485, 2008.

[401] N.L. Sullivan, K.A. Marquis, and D.Z. Rudner. Recruitment of condensin to replication origin regions by ParB-parS organizes the origin and promotes efficient chromosome segregation. *Cell*, 137:697–707, 2009.

[402] S. Gruber and J. Errington. Recruitment of Condensin to Replication Origin Regions by ParB/SpoOJ Promotes Chromosome Segregation in B. subtilis. *Cell*, 137:685–96, 2009.

[403] M.A. Sánchez-Romero, S.J. Busby, N.P. Dyer, S. Ott, A.D. Millard, and et al. Dynamic distribution of SeqA protein across the chromosome of Escherichia coli K-12. *mbio*, 1:e00012, 2010.

[404] N.K. Tonthat, S.T. Arold, B.F. Pickering, M.W. Van Dyke, S. Liang, and et al. Molecular mechanism by which the nucleoid occlusion factor, SlmA, keeps cytokinesis in check. *EMBO J.*, 30:154–164, 2011.

[405] H. Willenbrock and D.W. Ussery. Chromatin architecture and gene expression in Escherichia coli. *Genome Biol.*, 5:252, 2004.

[406] M.A. Wright, P. Kharchenko, G.M. Chruch, and D. Segre. Chromosomal periodicity of evolutionary conserved gene pairs. *Proc. Natl. Acad. Sci. U. S. A.*, 104:10559–64, 2007.

[407] D. Bates. The bacterial replisome: back on track? *Mol. Macrobiol.*, 69:1341–1438, 2008.

[408] A. Rosa and R. Everaers. Structure and Dynamics of Interphase Chromosomes. *PLoS Comp. Biol.*, 4:e1000153, 2008.

[409] H.C. Longuet-Higgins and B.H. Zimm. Calculation of the Rate of Uncoiling of the DNA Molecule. *J. Mol. Biol.*, 2:1–4, 1960.

[410] O. Danilova, R. Reyes-Lamothe, M. Pinskaya, D. Sherratt, and C. Possoz. MukB colocalizes with the oric region and is required for organization of the two Escherichia coli chromosome arms into separate cell halves. *Mol. Microbiol.*, 65:1485–92, 2007.

[411] C.L. Woldringh and N. Nanninga. Structural and physical aspects of bacterial chromosome segregation. *Journal of Structural Biology*, 156:273–283, 2006.

[412] G.C. Draper and J.W. Gober. Bacterial chromosome segregation. *Annual Review of Microbiology*, 56:567–597, 2002.

[413] X. Wang, C. Lesterlin, R. Reyes-Lamothe, G. Ball, and D.J. Sherratt. Replication and segregation of an Escherichia coli chromosome with two replication origins. *Proc. Natl. Acad. Sci. U. S. A.*, 108:E243–E250, 2011.

[414] J. Fan, K. Tuncay, and P.J. Ortoleva. Chromosome segregation in Escherichia coli division: A free energy-driven string model. *Computational Biology and Chemistry*, 31:257–264, 2007.

[415] F. Jacob and S. Brenner. On the regulation of DNA synthesis in bacteria: the hypothesis of the replicon. *C. R. Hebd. Seances. Acad. Sci.*, 256:298–300, 1963.

[416] R.A. Daniel and J. Errington. Control of cell morphogenesis in bacteria: two distinct ways to make a rod-shaped cell. *Cell*, 113:767–776, 2003.

[417] S. Ben-Yehuda, D.Z. Rudner, and R. Losick. RacA, a bacterial protein that anchors chromosomes to the cell poles. *Science*, 299:532–536, 2003.

[418] R. Grant, LR. Bowman, J.Z. Comolli, M. Eckart4, M. Koenig, and K.H. Downing et al. A polymeric protein anchors the chromosomal origin/ParB complex at a bacterial cell pole. *Cell*, 134:945–955, 2008.

[419] T. Kalkbrenner, A. Arnold, and S.J. Tans. Internal Dynamics of Supercoiled DNA Molecules. *Biophys. J.*, 96:4951–4955, 2009.

[420] N.J. Trun and J.F. Marko. Architecture of a bacterial chromosome. *Am. Soc. Microbiol. News*, 64:276, 1998.

[421] D.R. Albrecht, G.H. Underhill, J. Resnikoff, A. Mendelson, S.N. Bhatiaacde, and J.V. Shah. Microfluidics-integrated time-lapse imaging for analysis of cellular dynamic. *Integr. Biol.*, 2:278–287, 2010.

[422] Y. Liu and B. Chakraborty. Segregation of Polymers in Confined Spaces. *arXiv:0906.4023v2*, 2009.

[423] B. van Steensel and J. Dekker. Genomics tools for unraveling chromosome architecture. *Nature Biotechnology*, 28:1089–95, 2010.

[424] W. de Laat and F. Grosveld. Spatial organization of gene expression: the active chromatin hub. *Chromosome Res.*, 11:447–459, 2003.

[425] J. Dekker. Gene regulation in the third dimension. *Science*, 319:1793–94, 2008.

[426] D.A. Kleinjan and V. van Heyningen. Long-range control of gene expression: emerging mechanisms and disruption in disease. *Am. J. Hum. Genet.*, 76:8–32, 2005.

[427] M.A. Ferraiuolo, M. Rousseau, C. Miyamoto, S. Shenker, X.Q.D. Wang, M. Nadler, M. Blanchette, and J. Dostie. The three-dimensional architecture of hox cluster silencing. *Nuc. Acids Res.*, 38:7472–84, 2010.

[428] Z. Liu and W.T. Garrard. Long-range interactions between three transcriptional enhancers, active v_k gene promoters, and a 3' boundary sequence spanning 46 kilobases. *Mol. Cell. Biol.*, 25:3220–31, 2005.

[429] S.L. Berger. The complex language of chromatin regulation during transcription. *Nature*, 447:407–412, 2007.

[430] E. Heard and W. Bickmore. The ins and outs of gene regulation and chromosome territory organization. *Curr. Opin. Cell Biol.*, 19:311–316, 2007.

[431] A.G. West and P. Fraser. Remote control of gene transcription. *Hum. Mol. Genet.*, 14:R101, 2005.

[432] J. Dekker. A closer look at long-range chromosomal interactions. *Trends Biochem. Sci.*, 28:227–280, 2003.

[433] E.T. Jaynes. Information Theory and Statistical Mechanics. *Phys. Rev.*, 106:620–30, 1957.

[434] E.T. Jaynes. Information Theory and Statistical Mechanics II. *Phys. Rev.*, 108:171–90, 1957.

[435] E.F. Koslover, C.J. Fuller, A.F. Straight, and A.J. Spakowitz. Local Geometry and Elasticity in Compact Chromatin Structure. *Biophys. J.*, 99:3941–50, 2010.

[436] M.A. Marti-Renom and L.A. Mirny. Bridging the Resolution Gap in Structural Modeling of 3D Genome Organization. *PLoS Comp. Biol.*, 7:e1002125, 2011.

[437] M. Bathe and G.C. Rutledge. Inverse Monte Carlo procedure for conformation determination of macromolecules. *Journal of Computational Chemistry*, 24:876–890, 2003.

[438] G.C. Rutledge. Modeling experimental data in a Monte Carlo simulation. *Phys. Rev. E*, 63:021111, 2001.

[439] D.A. Kofke. *Monte Carlo Methods Chemical Physics.* Advances in Chemical Physics. John Wiley and Sons, New York, 1st edition, 1999.

[440] R.L. Henderson. A uniqueness theorem for fluid pair correlation functions. *Phys. Lett.*, 49A:197–198, 1974.

[441] A.P. Lyubartsev and A. Laaksonen. Osmotic and activity coefficients from effective potentials for hydrated ions. *Phys. Rev. E*, 55:5689–5696, 1997.

[442] K. Orlovskya, A. Kalinkovichb, T. Rozovskaiaa, E. Shezenb, T. Itkinb, H. Alderc, H.G. Ozerd, and L. Carramusaa et al. Down-regulation of homeobox genes MEIS1 and HOXA in MLL-rearranged acute leukemia impairs engraftment and reduces proliferation. *Proc. Natl. Acad. Sci. U. S. A.*, 2011.

[443] P.M. Ayton and M.L. Cleary. Transformation of myeloid progenitors by mll oncoproteins is dependent on HoxA7 and HoxA9. *Genes Dev.*, 17:2298–2307, 2003.

[444] U. Thorsteinsdottir, G. Sauvageau, M.R. Hough, W. Dragowska, P.M. Lansdrop, H.J. Lawrence, C. Largman, and R.K. Humphries. Overexpression of HoxA10 in murine hematopoietic cells perturbs both myeloid and lymphoid differentiation and leads to acute myeloid leukemia. *Mol. Cell Biol.*, 17:495–505, 1997.

[445] A. Biondi, G. Cimino, R. Pieters, and C.H. Pui. Biological and therapeutic aspects of infant leukemia. *Blood*, 96:24–33, 2000.

[446] P. Fraser and W. Bickmore. Nuclear organization of the genome and the potential for gene regulation. *Nature*, 447:413–417, 2007.

[447] C. Morey, N.R. Da Silva, P. Perry, and W.A. Bickmore. Nuclear reorganisation and chromatin decondensation are conserved, but distinct, mechanisms linked to hox gene activation. *Development*, 134:909–919, 2007.

[448] O. Kratky and G. Porod. Röntgenuntersuchung gelöster Fadenmoleküle. *Rec. Trav. Chim. Pays-Bas.*, 68:1106–1123, 1949.

[449] N. Saito, J. Takahashi, and Y. Yunoki. The Statistical Mechanical Theory of Stiff Chains. *J. Phys. Soc. Jpn.*, 22:219–26, 1967.

[450] M. Deserno. `www.cmu.edu/biolphys/deserno/pdf/henderson.pdf`, 2003. [Online; accessed August-2011].

[451] The FANTOM Consortium, H. Suzuki, A. Forrest, E. van Nimwegen, C. Daub, P. Balwierz, K. Irvine, and T. Lassmann et al. The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. *Nat. Genet.*, 41:553–562, 2009.

[452] D.P. Harper and P.D. Aplan. Chromosomal Rearrangements leading to MLL gene fusions: clinical and biological aspects. *Cancer Res.*, 68:10024–27, 2008.

[453] S. Geggier, A. Kotlyar, and A.Vologodskii. Temperature dependence of DNA persistence length. *Nuc. Acids Res.*, 39:1419–26, 2011.

[454] Z. Gitai. New Fluorescence Microscopy Methods for Microbiology: Sharper, Faster, and Quantitative. *Curr. Opin. Microbiol.*, 12:341–346, 2009.

[455] X.S. Xie, P.J. Choi, G.-W. Li, N.K. Lee, and G. Lia. Single-Molecule Approach to Molecular Biology in Living Bacterial Cells. *Annu. Rev. Biophys.*, 37:417–44, 2008.

[456] V.F. Holmes and N.R. Cozzarelli. Closing the ring: links between SMC proteins and chromosome partitioning, condensation, and supercoiling. *Proc. Natl. Acad. Sci. U. S. A.*, 97:1322–1324, 2000.

[457] J.A. Sawitzke and S. Austin. Suppression of chromosome segregation defects of Escherichia coli muk mutants by mutations in topoisomerase I. *Proc. Natl. Acad. Sci. U. S. A.*, 97:1671–1676, 2000.

[458] L. Ringrose, S. Chabanis, P.O. Angrand, C. Woodroofe, and A.F. Stewart. Quantitative comparison of DNA looping in vitro and in vivo: chromatin increases effective DNA flexibility at short distances. *EMBO*, 18:6630–6641, 1999.

[459] A.J. Wood, A.F. Severson, and B.J. Meyer. Condensin and cohesin complexity: the expanding repertoire of functions. *Nat. Review Genet.*, 11:391–404, 2010.

[460] L. Pusztai, H. Dominguez, and O.A. Pizio. Reverse Monte Carlo modeling of the structure of colloidal aggregates. *J. Colloid Interface Sci.*, 277:327–34, 2004.

[461] Images shown on the cover page. `http://www.shutterstock.com/s/dna/search.html#id=17004592` and `http://www.glogster.com/media/3/14/24/60/14246098.jpg`. [Online; accessed August-2011].

2011