Multiple scale modeling of Bacterial Chromosome

Elham Ghobadpour

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

École doctorale : ISCE - Ingénierie pour la Santé la Cognition et l'Environnement
Spécialité : MBS - Modèles, méthodes et algorithmes en biologie, santé et environnement
Unité de recherche : Translational Innovation in Medicine and Complexity

Modélisation de chromosomes bactériens à plusieurs échelles

Multiple scale modeling of Bacterial Chromosome

Présentée par :

Elham GHOBADPOUR

Direction de thèse :

Ivan JUNIER Directeur de recherche, CNRS Délégation Alpes	Directeur de thèse
Ralf EVERAERS Professeur des Universités, ENS de Lyon	Co-directeur de thèse

Rapporteurs :

MARIA BARBI
Professeur des Universités, SORBONNE UNIVERSITE
ROMAIN KOSZUL
Directeur de recherche, INSTITUT PASTEUR

Thèse soutenue publiquement le **29 mars 2023**, devant le jury composé de :

IVAN JUNIER Directeur de recherche, CNRS DELEGATION ALPES	Directeur de thèse
RALF EVERAERS Professeur des Universités, ENS DE LYON	Co-directeur de thèse
MARIA BARBI Professeur des Universités, SORBONNE UNIVERSITE	Rapporteure
ROMAIN KOSZUL Directeur de recherche, INSTITUT PASTEUR	Rapporteur
IRINA MIHALCESCU Professeur des Universités, UNIVERSITE GRENOBLE ALPES	Présidente
ALEXANDER GROSBERG Professeur, New York University	Examinateur

"*The important thing in science is not so much to obtain new facts as to discover new ways of thinking about them.*"

Sir William Bragg

Abstract

Supercoiled DNA often adopts tree-like double-folded branching configurations. In this context, I propose a framework to generate expected bacterial chromosome structures at multiple scales.

Stage I: A lattice model for the dynamics of randomly branching double-folded ring polymers

First, I studied an elastic lattice model for tightly double-folded ring polymers, which allows for the spontaneous creation and deletion of side branches coupled to a diffusive mass transport, which is local both in space and on the connectivity graph of the tree. I performed Monte Carlo simulations and studies systems belonging to three different universality classes: ideal double-folded rings without excluded volume interactions, self-avoiding double-folded rings and double-folded rings in the melt state. The observed static properties are in good agreement with exact results, simulations, and predictions from Flory theory for randomly branching polymers. For example, in the melt state rings adopt compact configurations and exhibit territorial behaviors [1].

Stage II: Coarse-grained models of supercoiled DNA at multiple scales

Second, I built a coarse-grained model of bacterial DNA, which is known to adopt tree-like plectonemic structures due to negative DNA supercoiling. To that end, starting from the first model, I included the possibility to generate long branches, with the average length becoming the parameter of the model.Considering DNA concentration similar to the in vivo situation, we adjusted this average plectoneme length parameter in order to reproduce as well as possible contact properties between chromosomal loci as obtained from high-throughput chromosome conformation capture methods (Hi-C).

Finally, we obtained various coarse-grained models that are consistent with each other and that allow capturing contact properties of various bacteria, from 2 kb to 4 Mb scale. In other words, we are able to rationalize from first principles contact properties between bacterial chromosomal loci as measured from Hi-C methods.

References

1. E. Ghobadpour, M. Kolb, M. R. Ejtehadi, and R. Everaers, “Monte Carlo simulation of a lattice model for the dynamics of randomly branching double-folded ring polymers,” Phys. Rev. E, vol. 104, p. 014501, Jul 2021.

Acknowledgements

When I embarked on my PhD journey three years ago, little did I know the profound experiences that lay ahead. Fulfilling my long-held dream of pursuing a PhD, time seemed to pass swiftly as I immersed myself in the world of academia. Transitioning from several years in professional practice back to the university environment was both daunting and exhilarating.

Now, it's time to take a moment to express my deepest gratitude to the incredible individuals who have been with me every step of the way. First and foremost, my profound appreciation goes out to my incredible supervisors, Ralf and Ivan. Your mentorship, patience, and guidance have been instrumental in shaping this thesis. I highly valued the weekly meetings we held, which not only served as crucial checkpoints to keep me on track academically, but also provided me with plenty of encouragement. I am profoundly grateful for the immeasurable contributions you made to my development and for playing a pivotal role in helping me establish myself within the academic community.

In addition to my supervisors, I am indebted to my exceptional colleague, whose support has been a constant source of motivation, Max. Words cannot express how much your unwavering belief in me has meant. You have been my biggest cheerleader throughout this journey, and I cannot thank you enough for your love and support.

To my partner, Ali, your unwavering support and belief in my abilities have been my rock throughout this journey. Your sacrifices and understanding through the highs and lows of this academic journey did not go unnoticed, and I couldn't have done this without you.

Mom and Dad, your love and encouragement have shaped me into the person I am today. Your support and belief in me have been a source of strength and motivation. I am eternally grateful for everything you've done.

To my siblings, Elahe, Mohammad, and Mohsen, thank you for being my confidants and sources of inspiration. Your belief in me, even from afar, have meant the world to me.

To my dear friends, Béa and Fabrice, thanks for helping Ali and I when we first arrived in France. You truly made France feel like a second home. Your kindness and support were invaluable.

My dear friends and colleagues, Mandy, Janka, Marija, Renata, Tim, Pascal, Lorenza, Mehrnaz, and Navid, you provided the much-needed breaks, laughter, and intellectual conversations that kept me going. Your camaraderie has made this journey memorable.

Among the most memorable highlights of my PhD journey was the opportunity to visit Collège de France. This enriching experience would not have been possible without the dedication and expertise of Dr. Espeli and Céline. I am grateful for the knowledge and friendships that were forged during that time.

Thanks to generous CPU time at PSMN and CBP and the people working at CBP, Diane, Lois, Emmanuel.

Thanks to my great life coach, Fariba, who challenged me to reach new heights. Your coaching and belief in my potential played a significant role in my successful completion of this PhD and in my life. I am immensely thankful for your role in my accomplishments.

Thanks to my great homeopath, Saba, your unwavering faith in my dreams has been my driving force. You consistently provided motivation, wisdom, and a vital perspective that transcended the academic sphere, propelling me to overcome challenges and achieve success.

Finally, to anyone whose name I may have omitted but has contributed to my academic growth in any way, thank you from the bottom of my heart.

As I bring my three-year-long PhD journey to a close, I can genuinely say that I am immensely proud of my accomplishments. This transformative journey has not only shaped me as a researcher but also as an individual, imparting the invaluable lesson that perseverance yields fruitful rewards in the long run. The unwavering commitment and determination displayed by those around me have profoundly influenced my character.

Contents

List of Figures

List of Tables

Chapter 1

Intention and Organization of this Thesis

Bacterial genomes face similar constraints as their eukaryotic counterparts. The genetic material must be tightly folded to fit into the limited space of the cell. The (*E. coli*) chromosome is about 1000 times longer than the confinement of the bacterial cell. Cell-imaging technology has revealed that the bacterial genome exists in an extremely precise spatial ordering within the nucleoid (Robinow, 1956) as opposed to the "bowl of spaghetti" configuration assumed in the past. Moreover, this compaction should ensure that cell processes like replication, segregation, and transcription/translation of the genetic material are possible.

The sequencing of the bacterial DNA was deciphered years ago. However, it is more and more evident that pure knowledge about the linear ordering of DNA base pairs is not enough to understand its functioning. Instead, a fundamental understanding of the 3*D* organization of the genome and the physical principles governing its structure is the first vital step for fully understanding this biomolecule.

FIGURE 1.1: "Illustration of the Genome Folding Problem". From cartoonist John Chase. Copied from (Iyer, Kenward, and Arya, 2011).

How do bacteria fold their genetic materials in the three-dimensional (3D) architecture?

In order to address this question, computational modeling coupled with high-throughput chromosome conformation capture techniques (Hi-C) and high-resolution microscopy is used. Hi-C methods investigate spatial genome folding by determining the relative contact frequency between pairs of genomic loci and generating interaction frequency maps. Moreover, high-resolution (live) cell imaging is pivotal for quantifying and validating many findings about spatial genome organization. However, despite considerable advances in elucidating the nature of the nucleoid structure and the role of different players in this organization, our knowledge of it still needs to be improved. This is a field in which computational/analytical modeling approaches may guide our intuition. Disregarding molecular details, coarse-grained models are particularly suited to provide an overall picture of biological systems' major driving forces and underlying organizational principles. For this purpose, one might choose a "multi-scale" approach, where each coarse-grained level emerges from the previous, more detailed one.

1.1 Aim and Structure of the thesis

In this context, the work presented in this thesis is inspired by a range of experimental observations for bacterial chromosomes. Even though these organisms display different dynamics and organizational complexity, all bacterial chromosomes studied here follow a common pattern, suggesting a similar general mechanism behind it.

This thesis aims to develop a polymer physics model to study the $3D$ conformation of bacterial chromosomes (*P. aeruginosa*, *C. crescentus*, and *E. coli*) at different length scales. We do not aim to explain the specific patterns in the contact map like other theoretical studies but instead investigate the possibility of capturing the main features of the internal structure of bacterial chromosomes using a branched double-folded ring polymer model inspired by the topological organization of negative supercoiled DNA, which adopts tree-like structure. In the model, the branches represent plectonemes. Their length is the only free parameter in our model, which was adjusted to reproduce as well as possible contact probabilities between chromosomal loci as obtained from Hi-C experimental data. Our model can provide theoretical insights to guide future experimental approaches that try to obtain information about the shape of the prokaryotic genome.

In **chapter 2**, the basic biology of bacterial DNA is discussed in the first place. The second part of this chapter introduces the mathematical principles of calculating the free energy of DNA supercoiling. We examine how the radius and the pitch of the plectoneme scale with supercoiling density.

Two major experimental methods probing the $3D$ organization of genome are presented and confronted to each other in **chapter 3**. Additionally, previous models and computational methods used to study bacterial chromosome architecture are addressed. These methods differ in their approach, the assumptions they make, and the adopted implementations.

Since our work is of an interdisciplinary nature, **chapter 4** provides an introduction to basic concepts of polymer physics. Exponents describing the properties of polymers are introduced and will come in handy in the other chapters. Finally, the physics of ring polymer chains in

different systems is explained, and a short overview of the principles underlying universality and coarse-graining is given.

In **chapter 5**, our elastic lattice double-folded ring polymer model is presented in the first part. The second major part of this chapter introduces the mathematical principles of Monte Carlo simulations and presents the algorithm for generating double-folded ring polymer conformations. Eventually, the conformational properties of the double-folded ring polymers in various physical systems are investigated. The observed static properties are in good agreement with exact results, simulations, and predictions from Flory theory for different known polymer systems. For example, rings adopt compact configurations in the melt state and exhibit territorial behaviors. A part of the results of this chapter has been published in Phys. Rev. E (Ghobadpour et al., 2021).

After establishing a physical framework for bacterial chromosome conformation, in **chapter 6**, we proceed with extracting simulation parameters to model bacterial DNA. These parameters are the lattice constant, the amount of DNA contained in each tree node, the system's density to mimic the bacteria condition, and the average plectonemes length. Next, the contact probabilities derived from Hi-C matrices of different bacterial DNA are then investigated. Two scales of organization of chromosomes are observed in different bacteria. These features are observed for all bacterial chromosomes irrespective of the cell type, the details of the Hi-C interaction frequencies, and even the details of the Hi-C methods (since the data were obtained from different labs). Eventually, the role that branch lengths can play in shaping the spatial organization of bacterial DNA is elucidated, taking into account the competing interplay between secondary and tertiary structure contact.

Last but not least, **chapter 7** presents a concise summary of the results obtained, closing with a discussion of future challenges.

Chapter 2

Introduction to the Biology of Bacterial DNA

Deoxyribonucleic acid, or "*DNA*" for short, is one of the most important chemical compounds of living organisms responsible for storing genetic information. Although the DNA molecule was extracted from the cell for the first time in 1869 by Friedrich Miescher, its biological importance was only understood years later. Miescher thought that DNA's role in the cell is phosphorus storage. However, after near a century of scientific effort led to the discovery of the nature and role of the DNA molecule.

2.1 Molecular Structure of DNA

DNA is a heteropolymer of 4 elementary building blocks called nucleotides. The nucleotide is a compound of 20 atoms, grouped into three main structures: a phosphate group PO_4^-, a five-carbon sugar ring (ribose or deoxyribose), and a nitrogen base (Adenine(A), Guanine(G), Cytosine(C) and Thymine(T)), which form a single-stranded DNA (ssDNA) (Calladine and Drew, 1997) (Fig.2.1). The backbone of this strand is made of the succession of a covalent bond between the sugar of one nucleotide and the phosphate group of the adjacent one (Fig.2.1(A)) with a directionality quoted $5^{'} \rightarrow 3^{'}$ called conventional direction. Two ssDNA are coupled via hydrogen bonds between the bases, which always pair in a way that A with T via two hydrogen bonds and C with G via three hydrogen bonds forming a rather flat complex called base pair, *bp*. In order for the bases to be able to pair, the strands are run in opposite directions (Fig.2.1(B,C)). These two single strands are twisted together, and the backbones are separated by two helically-shaped grooves of different widths, the major groove and the minor groove (Fig. 2.1(C)). These grooves provide access to base pairs, which are essential in the DNA-proteins interactions and also DNA-DNA interactions.

2.2 Three Major Forms of DNA

Double-helix can adopt different shapes depending on pH and physicochemical conditions (e.g., changing the temperature and salt concentrations). B-form is the most common form of

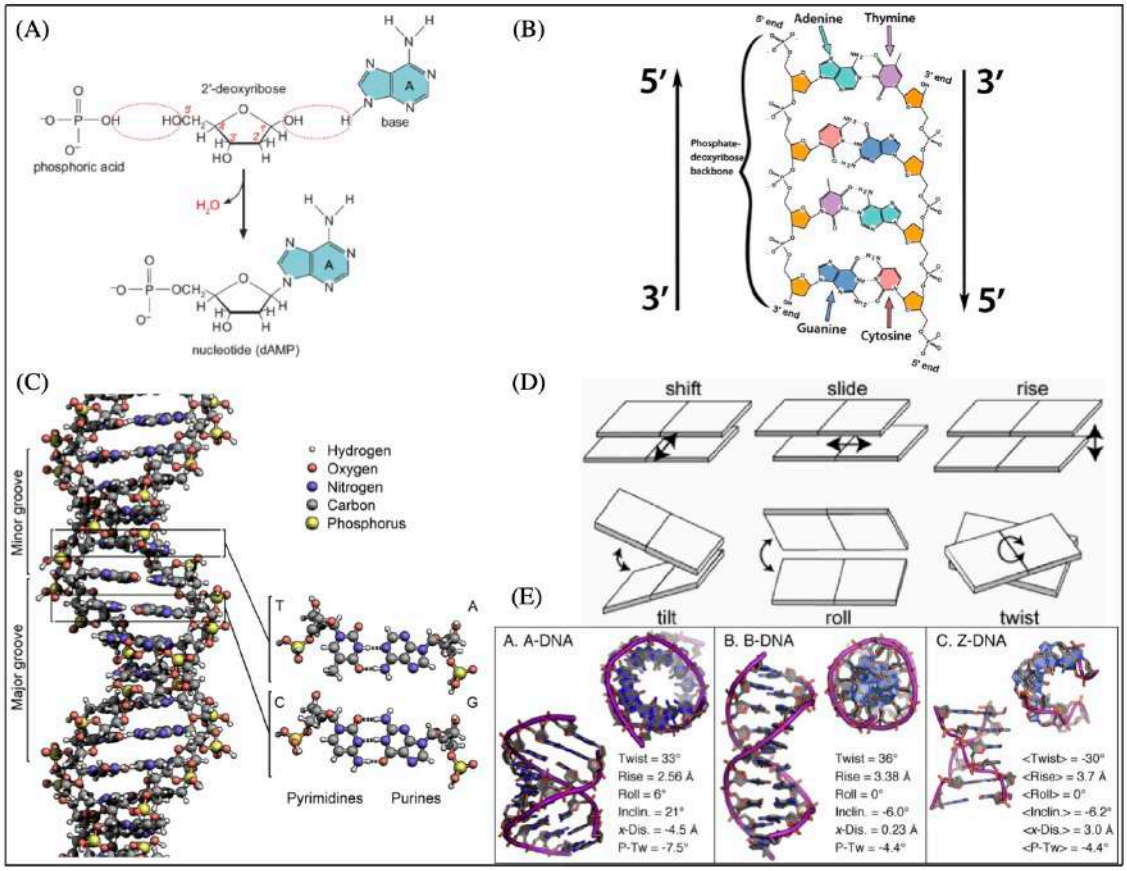


FIGURE 2.1: DNA Structure. (A) Formation of a nucleotide by removal of water. (B) Schematic view of the DNA backbone and the base pairing interactions. (C) 3D-Helical structure of B-DNA. (D) Schematic representation of the six degrees of freedom between successive base pairs. (E) Different forms of DNA. A, B, and Z forms. The image is copied from (http://en.wikipedia.org).

DNA under normal physiological conditions. It is a right-handed double helix whose parameters were discovered by the X-ray diffraction experiments at high humidity (92%) (Franklin and Gosling, 1953; Wilkins, Stokes, and Wilson, 1953). The phosphates in a base-pair are $20\,A^\circ$ apart (the diameter of the dsDNA), the thickness of one base is $3.4\,A^\circ$, two successive base-pair are rotated by a twist angle of around 36°, i.e., the helical pitch is about $10.5\,bp$. This molecule is stable and stiff in water at room temperature and neutral pH and has a persistence length of around 50 nm ($\sim$ 150 bp).

The A-form of DNA can be observed at lower relative humidity (75%) than the B-form, which means that it is rarely present in normal physiological conditions (Franklin and Gosling, 1953; Ussery, 2002). The A-DNA is a right-handed helix that is a bit wider than B-DNA (and also Z-DNA), with a diameter of $23\,A^\circ$, since the base-pairs stack a bit off-center in the A-form DNA compared to B-form which the base pairs stack nearly on top of each other. The rise between consecutive base-pairs $(2.56\,A^\circ)$ is shorter than the one in B-DNA, and they are not perpendicular to the helix-axis. Moreover, looking down at the helix, there is a hole in the A-form DNA (Fig. 2.1(E)). The minor groove is shallow, while the major groove is thin and

deep.

The other structure of DNA is known as Z-form, which is a left-handed double helix, with the local twist-angle approximately -30° opposite to the Watson–Crick helix. Its rise and pitch are the largest of all the forms of DNA ($3.7A^\circ$, and 12 bp per helical turn respectively) (Ussery, 2002). Its backbone is not a smooth helix but is irregular and zigzag in shape, therefore its name. Moreover, Z-DNA has no major groove. (Fig. 2.1(E)). The formation of Z-DNA *in vitro* was initially tested under very high concentrations of NaCl. The lowest energy-level ground state of DNA in a physiological solution is B-DNA, and the Z conformation is a higher energy state and is favored by DNA supercoiling.

One feature of the helices is how base pairs are stacked on top of each other. Due to the stiffness of the base pairs, we can describe them as rigid bodies like solid rectangular blocks and describe the state of the dsDNA by the relative orientation (twist, tilt, roll) and translation (shift, slide, rise) between two neighboring base pairs (Figure. 2.1(D)). The real DNA helices can be characterized by only three degrees of freedom, twist angle, roll, and slide at the first approximation. The local twist angle is a rotation about the axis perpendicular to the plane of the base-pair, that runs along (or nearly along) the line joining the center of two consecutive base pairs. The role refers to a rotation around the longest axis in a base pair. The slide describes a translation in the direction of the longest axis of the base pair. The values of the parameters that correspond to three forms of the DNA are summarized in Figure. 2.1(E). The schematic representation of the A, B, and Z forms of DNA are depicted in Fig. 2.1(E).

2.3 Bacterial Genome

The bacterial genome usually consists of one single DNA molecule, which is significantly shorter in length compared to eukaryotes, typically about a few million base pairs against billions of base pairs. Most bacterial genomes are circular, and their size varies over two orders of magnitudes, from almost 100 kb to 15 Mb, with the majority between 1 Mb and 6 Mb. The distribution of the chromosome length is displayed in Fig. 2.3. For a long time, it was thought that the bacterial chromosome, which tends to stain uniformly with basic dyes, was unstructured. It took until the 1930s that light microscopy using DNA dyes with acid-treated cells clearly showed that the bacterial chromosome was concentrated in distinct entities with soft, irregular edges (Fig.2.2 (A),(B)). Now we know that prokaryotic cells, unlike eukaryotes, lack a nucleus as a compartment enclosed by a membrane. They organize their genetic material into a specific region (Sherratt, 2003) of the cytoplasm called the nucleoid (composed of DNA, RNA, and proteins) (Fig.2.2 (C)). Early in the 1970s, Pettijohn and colleagues developed a technique for lysing *E. coli*gently and obtaining nucleoids for direct electron microscopy imaging, which could provide an image of the bacterial chromosome as a collection of plectonemic loops (Fig.2.2 (D)).

Protein-coding regions are, on average, 88% of the bacterial genome with a typical range of $85-90\%$, and can reach up to 97% (Land et al., 2015; Kirchberger, Schmidt, and Ochman,

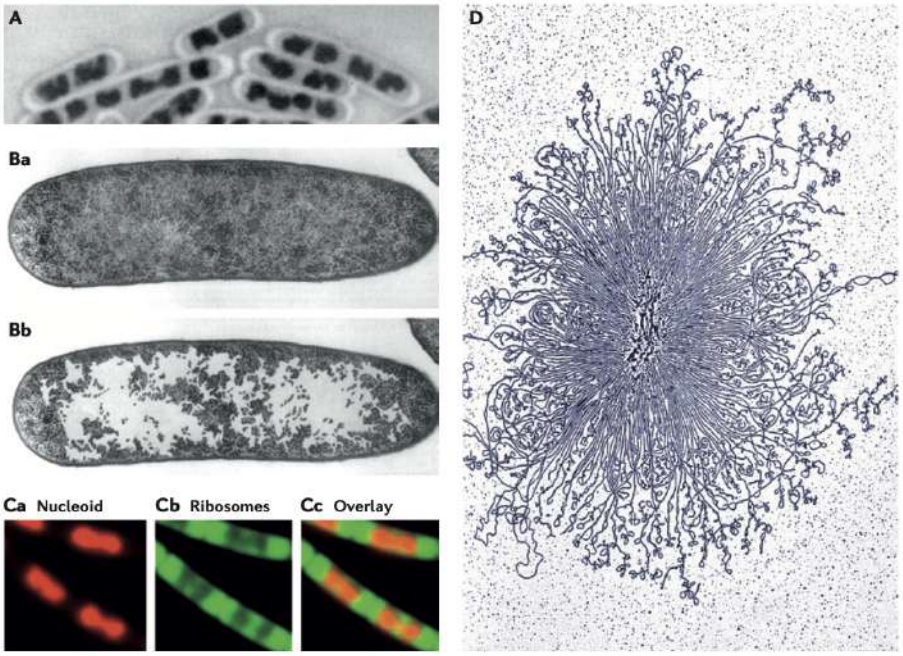


FIGURE 2.2: (A) Bacillus subtilis nucleoid. (B) The nucleoid of *E. coli* in the growing phase. Panel b shows the same section as in panel a, except the ribosome-free spaces, were colored. The image is copied from (Robinow and Kellenberger, 1994). (Ca) Nucleoid of *B.subtilis* cells growing in rich media colored red (stained with $4'$, 6-diamidino-2-phenylindole (DAPI); (Cb) ribosomes colored green; (Cc) overlay of the two images. (D) An isolated *E. coli* nucleoid spread on an electron microscope grid, stained with uranyl acetate, and visualized by transmission electron microscopy. A bacterial cell like *E. coli* is usually around $0.5-3\mu m$ in length, whereas the length of the nucleoid in this panel is approximately $20\mu m$. In memory of Dr Ruth Kavenoff $1944-1999$. The image is copied from (Wang, Montero Llopis, and Rudner, 2013).

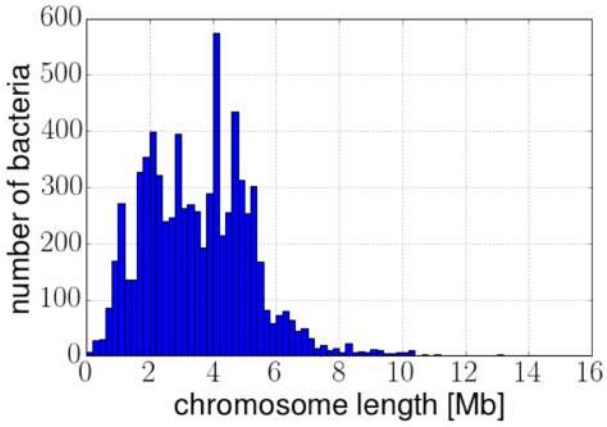


FIGURE 2.3: . Distribution of the chromosome length in bacterial kingdom.

2020). It means that almost all of the bacterial genome has a function, in contrast to most eukaryotic genomes, where the amount of coding DNA is much smaller (e.g., the human genome

contains somewhere between $1-2\%$ coding DNA). As a result, the bacterial genome displays a high gene density (i.e., the proportion of a genome composed of genes), and there is a strong link between the genome size and gene number (Kuo, Moran, and Ochman, 2009). Given the small volume of cells compared to the size of chromosomes, they must be packed inside the cells in a way that simultaneously meets the requirements of DNA replication, chromosome segregation, and gene transcription. Hence, the bacterial nucleoid is organized at different length scales with the help of DNA enzymes and many proteins. Over several decades, details of which have been investigated using various *in vivo*, *in vitro*, and *in silico* approaches. For instance, a multidisciplinary approach involving high-throughput techniques like chromosomal conformation capture (3C) (reviewed in detail in Sec.3.1) combined with powerful imaging approaches like super-resolution microscopy techniques (reviewed in detail in Sec.3.2) has helped us to understand the multi-scale hierarchical structural organization of bacterial chromatin similar to that of eukaryotic chromatin. A schematic of the hierarchical organization of bacterial genomes is given in Figure 2.4.

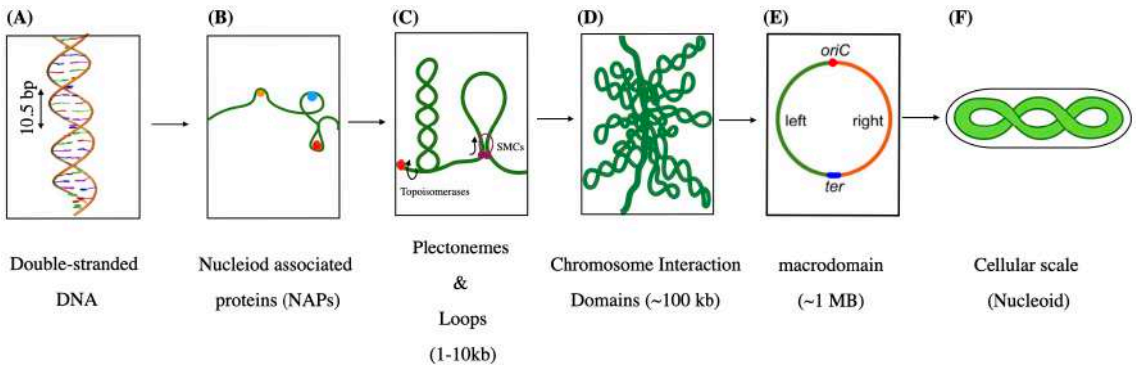


FIGURE 2.4: Schematic of hierarchical organization in bacteria. (A) dsDNA is the basic component of all chromosomes. (B) Nucleoid-associated proteins contribute to modulating the local structure of DNA structure. (C) Supercoiling is generated by topoisomerases' activities. Most DNA in bacteria forms plectonemes since it is negatively supercoiled. Moreover, bridging proteins and SMCs form loop-like structures. (D) Chromosomal Interaction Domains, CID, are the regions of the genome where loci interact more frequently than they do with other loci from different domains.(E) On a larger scale, the circular bacterial genome is organized in macro-domains. (F) The cytosol and cell border contain the bacterial chromosome, also known as the nucleoid. Modified from (Dame and Tark-Dame, 2016; Birnie and Dekker, 2021)

At the smallest scale, nucleoid-associated proteins contribute to modulating the local structure of DNA and in the regulation of specific genes (Fig. 2.4(B) and reviewed in Sec.2.4). At a larger scale, supercoiling is generated by transcription, replication, and topoisomerases' activities (Fig. 2.4(C) and reviewed in Sec.2.5). Also, at this scale, bridging proteins and structural maintenance of chromosomes (SMC) complexes form loop-like structures (Sec.2.4.4). One of the characteristics of the nucleoid is that supercoils are arranged into several topological domains. A topological domain forms because of a supercoiling-diffusion barrier. In other words, a single cut in one domain will only relax that domain and not the others (Postow et

al., 2004). Several independent investigations have shown that the topological domains have different sizes from $10-400$ *kb* (Lagomarsino, Espéli, and Junier, 2015) (Sec.2.5.3).

Moreover, the recent development of HiC techniques (Sec.3.1) has revealed that the bacterial chromosome is segmented into a succession of tens of kb-long highly self-interacting regions named "chromosome interaction domains" (CID) (Fig. 2.4(D)). They appear as squares along the main diagonal in Hi-C matrices. These are the regions of the genome where loci interact more frequently than they do with other loci from different domains. CIDs ranged in length from 30 to 400 kb (Le et al., 2013). CIDs are, at first sight, equivalent to topologically associating domains (TADs) observed in eukaryotic cells. Another characteristic of CIDs is the presence of a boundary between CIDs that prevents physical interactions between genomic regions of two neighboring CIDs. Boundaries of CIDs seem to correlate with the expression of highly expressed genes. Inhibition of transcription (like treating cells with rifampicin, which inhibits RNA polymerase) disrupts CID boundaries. In particular, moving a locus with highly expressed genes into a poorly expressed genome region generates a new CID boundary(Le et al., 2013). Further studies are needed to determine the nature of CID boundaries unambiguously.

On even larger scales, the DNA is compartmentalized into large spatially distinct domains called macrodomains, which are on the order of $\sim 1-$Mb (Fig. 2.4(E)). For instance, it has been shown that the *E. coli* genomes can be divided into four macrodomains (Ori, Ter, left, and Right) plus two non-structured domains (Valens et al., 2004). Even at this scale, loci within a given macrodomain recombined at higher frequencies with other loci in the same macrodomain.

2.4 Condensation and Organization of DNA by Nucleoid-Associated Proteins (NAPs)

2.4.1 Nucleoid-Associated Proteins

Histones modulate the local structure of the DNA in eukaryotic cells. However, in bacteria, small proteins cause them to fold into a more compact structure. At the kilo-base (kb) scale, these small essential proteins condense DNA by bending (Fis, HU), wrapping (IHF, Dps), and bridging (H-NS) relatively distant DNA strands to reduce the persistence length. These groups of proteins are called histone-like proteins or nucleoid-associated proteins (NAPs).

Although the presence of NAPs is a universal feature among bacteria, there is a significant variation in NAPs in different bacteria. Even their cellular abundance in one specific bacteria (like *E. coli*) depends on growth phase, growth condition, and environmental changes in general (Dillon and Dorman, 2010; Hołówka and Zakrzewska-Czerwińska, 2020). Figure. 2.5 shows the evolution of NAPs concentrations in *E. coli* bacteria across the different growth phases.

The DNA binding activities of NAPs induce topological and structural changes in the chromosomal DNA to ensure proper compaction inside the cell. In addition to their architectural

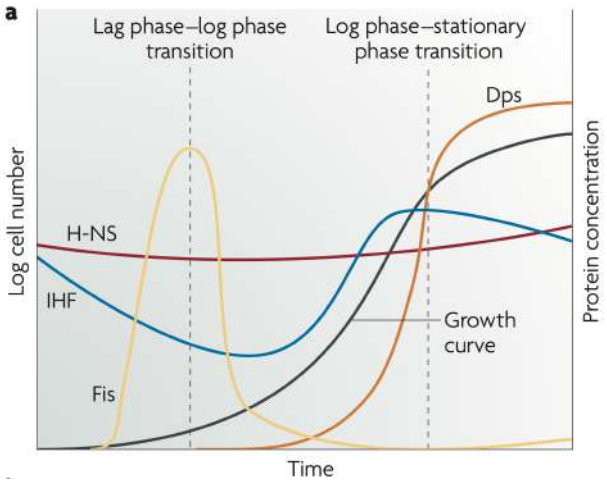


FIGURE 2.5: Evolution of four NAPs concentrations across the different growth phases in *E. coli*. A typical bacterial growth curve is shown in black. Dashed lines indicate phase–log phase and stationary phase transitions. The image is copied from (Dillon and Dorman, 2010).

rules, NAPs are also involved in cellular processes such as transcription, replication, and DNA recombinations (Swinger and Rice, 2004; Chodavarapu et al., 2008; Dillon and Dorman, 2010; Browning, Grainger, and Busby, 2010; Dorman, 2014).

There are at least 12 NAPs identified in *E. coli* (Verma, Qian, and Adhya, 2019). Here, a brief review of the most extensively studied ones is presented. Based on their topological effect on DNA, these proteins are categorized into two groups: DNA-bridging proteins and DNA-bending proteins. Their binding properties are summarized in Fig. 2.6.

2.4.2 DNA bridging proteins

The Heat-stable Nucleoid Structuring protein (H-NS)

The heat-stable nucleoid structuring protein is conserved among Gram-negative bacteria. Atomic force microscopic images of H-NS-DNA complexes show that H-NS binding makes bridges between adjacent DNA duplexes (Fig. 2.6) (Dame, Wyman, and Goosen, 2000; Luijsterburg et al., 2006). Bridging relatively distant DNA segments have a significant impact on the general organization and compactness of the nucleoid.

2.4.3 DNA bending proteins

The Heat-Unstable protein (HU)

HU is a non-sequence-specific DNA binding protein that regulates DNA flexibility. This protein prefers to bind into distorted regions, such as bends (Dillon and Dorman, 2010). HU plays its role in DNA flexibility by bending the duplex. At low protein concentrations, HU reduces the stiffness of DNA over short distances. However, at high concentrations, it stiffens DNA.

DNA flexibility caused by HU facilitates DNA loop formation, which affects chromosome architecture.

The Integration Host Factor (IHF)

IHF is structurally almost identical to HU but behaves differently from HU in many aspects (Rice et al., 1996; Verma, Qian, and Adhya, 2019). Unlike HU, it preferentially binds to a specific DNA sequence and condenses DNA by inducing by $> 160^{circ}$ bend angle (Rice et al., 1996; Verma, Qian, and Adhya, 2019).

The Factor of Inversion Stimulation (FIS)

FIS has a homodimer structure, and like IHF, it induces DNA bending at cognate sites. FIS has an ellipsoidal structure and is made of two helix-turn-helix motifs involved in DNA binding. The two helix-turn-helix motifs bind successive major grooves and have the distance $25A^{circ}$, which is shorter than the pitch of a B-DNA (about $8A^{circ}$), So FIS must force a bend into DNA to bind stably (Skoko et al., 2006; Verma, Qian, and Adhya, 2019). Although FIS is often considered as a sequence-specific DNA binding protein, it binds to a random sequence DNA as well, which this non-specific binding of FIS can contribute to DNA condensation and organization shoed by Magnetic tweezers experiments (Skoko et al., 2006)

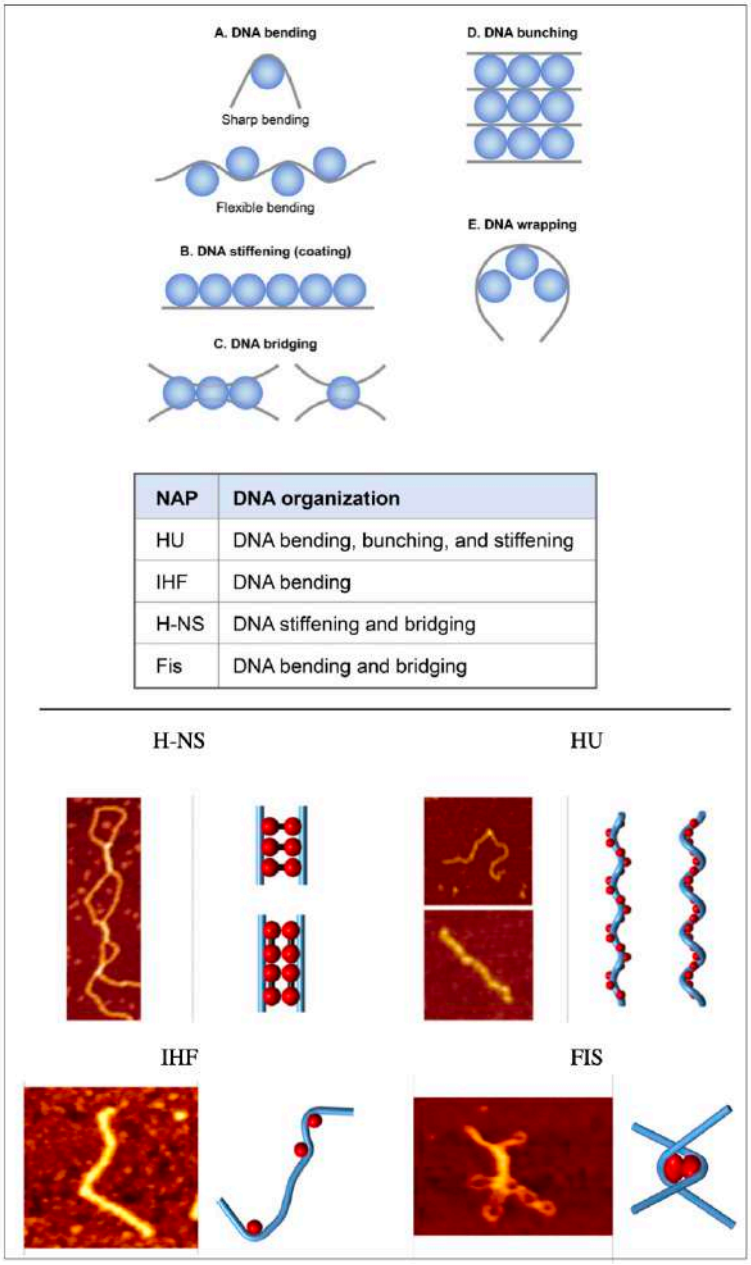
2.4.4 Structural Maintenance of Chromosomes (SMC) complexes

As in eukaryotes, bacterial cells partition their genetic material via chromosome condensation. Structural maintenance of chromosomes (SMC) acts as a condensin since they can bridge and extrude DNA loops. SMC complexes are large V-shaped homo-dimeric molecules. They have a higher molecular weight than other NAPs $(\approx 150-200kDa)$ (Luijsterburg et al., 2006). They are found in almost all organisms: mammals, nematodes, plants, and prokaryotes(Cobbe and Heck, 2004).

Many bacteria contain three SMC complexes: Smc-ScpAB, MukBEF, and MksBEF, defined as bacterial condensins which play a central role in chromosome organization and segregation (Lioy, Junier, and Boccard, 2021). Smc-ScpAB condensins are the most conserved complex and are found in the vast majority of bacteria. It works together with the ParABS system and is essential for chromosome segregation. Some $\gamma-$proteobacteria like *E. coli* have the second family, MukBEF. Mutational inactivation of this complex results in chromosome disorganization and decondensation. MksBEF is another condensin occasionally found in bacterial genomes.

2.5 Condensation and Organization of DNA by Supercoiling

The structure of the bacterial genome is partially achieved by significant supercoiling. DNA supercoiling is a fundamental, inescapable property of DNA due to DNA machinery that transcribes and replicates the DNA. Activities of topoisomerases, nucleoid-associated proteins



NAP	DNA organization
HU	DNA bending, bunching, and stiffening
IHF	DNA bending
H-NS	DNA stiffening and bridging
Fis	DNA bending and bridging

FIGURE 2.6: Upper row: Schematic representation of DNA folding properties of bacterial nucleoid proteins (NAPs). Gray lines depict DNA, and blue spheres show NAPs. (A) DNA bending. For instance, HU causes flexible bends (varying between $10-180°$, whereas IHF introduces sharp bending ($> 160^{circ}$). FIS can bend DNA between $60-75°$. (B) Straightening or stiffening of DNA (e.g., H-NS spreads along DNA, making it stiff. (C) Bridging relatively distant DNA strands, like H-NS. (D) DNA bunching or bundling (e.g., HU bindings brings several parallel DNA segments together). (E) NAPs bound together and wrap DNA. Lower row: SFM images of NAPs and a model of their roles in the compaction of DNA. The image is copied and modified from (Verma, Qian, and Adhya, 2019; Luijsterburg et al., 2006)

(NAPs), and transcription are three main factors in generating and maintaining supercoiling in bacterial DNA. In this section, this property of DNA is reviewed in detail.

2.5.1 DNA Supercoiling

The two strands of DNA are held together via hydrogen bonds between base pairs (two bonds between A and T and three bonds between C and G). For B-DNA, the two strands once intertwined each 10.5 base pair to form a right-handed helix. Any mechanical manipulation which leads to overwinds/underwinds of the helix generates torsional strain that can be relieved by distorting the shape of DNA, or in other words, DNA writhes upon itself. These overwinds/underwinds of the helix are referred to as DNA *supercoiling*. The closed circular form of DNA or linear form with fixed ends (whose ends are not allowed to rotate) is topologically constrained. In other words, the number of times the two strands wind around each other is topologically invariant, known as linking number Lk. This topological parameter is always an integer and related to two geometrical properties: the twist Tw and the writhe Wr. The twist is the number of helical turns in the DNA molecule, and the writhe is the number of self-crossing of the dsDNA centerline (Mirkin, 2001). Unlike Lk, which is an integer number, Wr can be any value. Upon White-Fuller-Calugarenau theorem (Fuller, 1978) the linking number of DNA at any time t is the sum of the twist and writhe:

$$Lk(t) = Tw(t) + Wr(t). \tag{2.1}$$

The characteristic feature of the topologically-closed DNA, Lk, cannot be altered by any deformation of the DNA strands as long as it does not involve cutting one or both strands. However, twist and writhe can be interconvertible due to changes in ambient conditions, temperature, and during DNA functioning as long as Eq.2.1 is satisfied. In an overwound DNA, the excess in twist can be converted into positive writhe, and for the underwound molecule, the deficit in turns can be converted into negative writhe (Fig. 2.8). The orientation of the crossing the center line of the helix distinguishes between positive and negative writhe (Figure. 2.7).

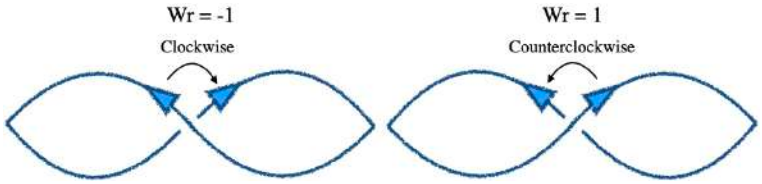


FIGURE 2.7: Positive and negative signs of writhe. In a $(-)$ crossing, the overlying direction arrow must be turned clockwise to align with the underlying arrow. In a $(+)$ crossing, this turn is counterclockwise.

The linking number of N base-pairs long circular DNA in the relaxed form is defined as Lk_0. It can be calculated by

$$Lk_0 = Tw_0 = \frac{N}{\gamma}, \tag{2.2}$$

where γ is the number of base pairs per helical turn (Fig. 2.8). If the linking number of a DNA molecule is different from the relaxed one, Lk_0, DNA is said to be supercoiled. In other words, linking difference $\Delta Lk = Lk - Lk_0$ measures DNA supercoiling. When $\Delta Lk < 0$, DNA is underwound compared to relaxed DNA and is called negatively supercoiled. While $\Delta Lk > 0$,

DNA is overwound compared to relaxed DNA and is called positive supercoiled (Fig.2.8). A valuable characteristic of supercoiled DNA that is independent of the size of DNA molecules is the superhelical density σ, which estimates the number of supercoils per helical turn,

$$\sigma = \frac{Lk - Lk_0}{Lk_0}. \tag{2.3}$$

When the DNA is supercoiled, the entire double helix is stressed. This stress can either cause a change in the number of base pairs per helical turn in closed circular DNA or leads to spatial deformation of the helix axes. In other words, the change in Lk can be decomposed into a twist (local) and a writhe (nonlocal) contribution:

$$\Delta Lk = \Delta Tw + Wr. \tag{2.4}$$

ΔTw represents the change in the twist, and Wr represents the number of writhes in the system (Fig. 2.8).

In general, bacterial DNA is negatively supercoiled (underwound), which can lead to the formation of plectonemes and solenoids through writhe changes. Fuller (Fuller, 1971), by theoretical analysis of the shape of the supercoiled DNA, illustrated that interwound superhelix is more favorable from the energetic point of view over a solenoid shape. Years Later, Vologodskii (Vologodskii et al., 1992) provided numerical evidence for it. The electron microscope images of isolated supercoiled DNA revealed the presence of plectonemes.(Fig. 2.9) (Delius and Worcel, 1974; Kavenoff and Bowen, 1976).

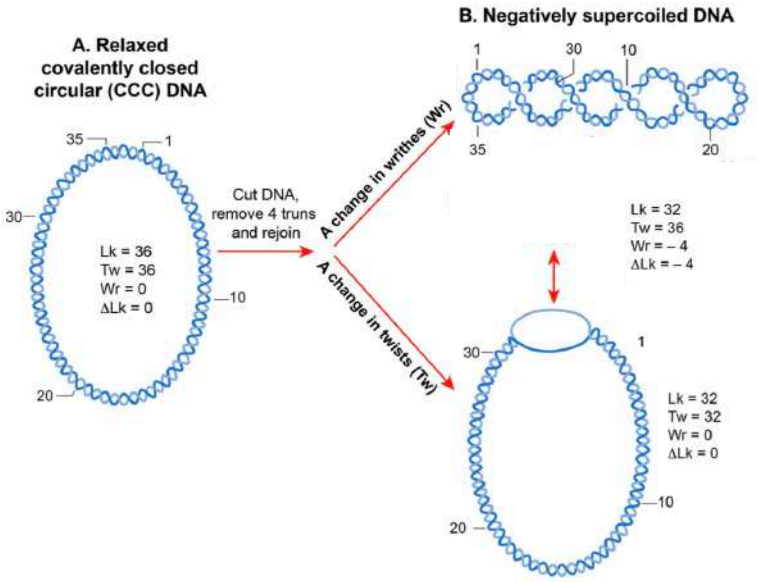


FIGURE 2.8: (A) A relaxed closed circular DNA. (B) A decrease in Lk induces negative supercoiling in DNA. Writhe can adopt a geometrical structure called plectoneme. The image is modified from (Verma, Qian, and Adhya, 2019).

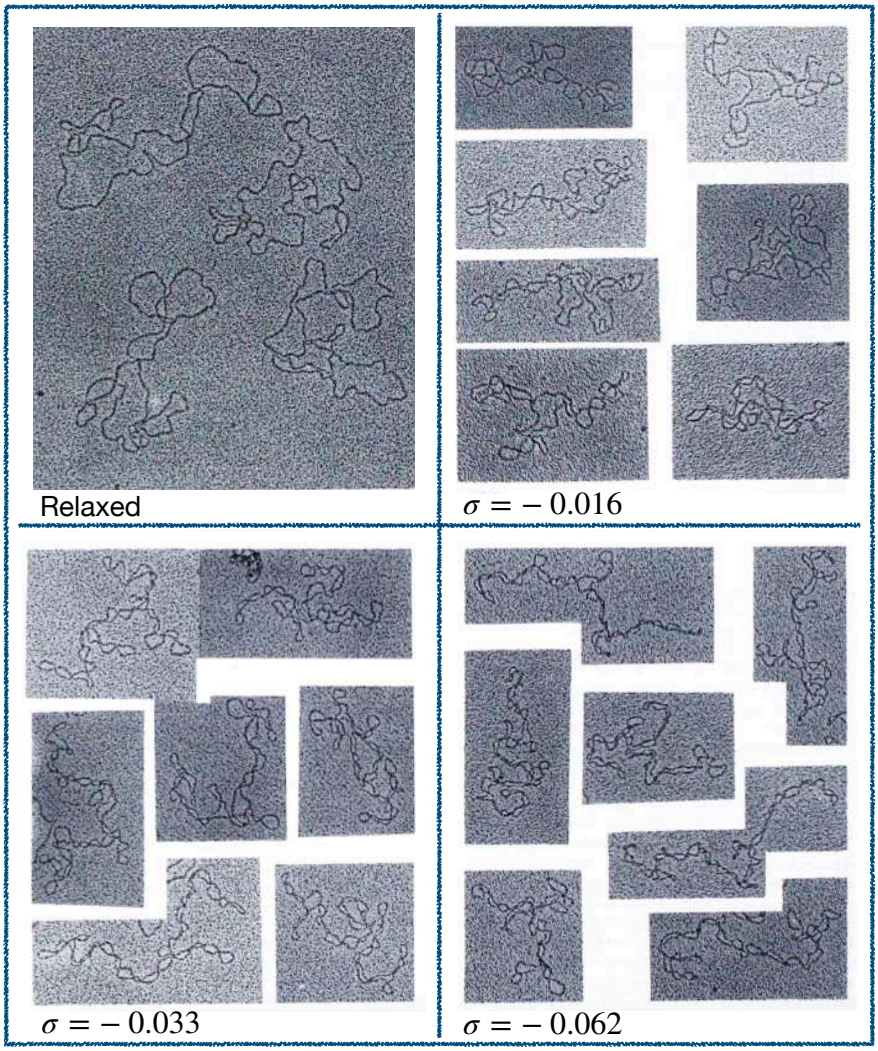


FIGURE 2.9: The electron microscope images of supercoiled DNA with length 7kb at different supercoiling densities, σ. The superhelices are very long, thin, and branched. Clearly, DNA adopts plectonemic form even at very low values of supercoiling. It was modified from (Boles, White, and Cozzarelli, 1990).

2.5.2 DNA Topoisomerases

DNA topoisomerases are essential proteins that play vital roles in regulating supercoiling. Topoisomerases are present in all domains of life and are highly conserved (Forterre et al., 2007). There are some functionality overlaps between DNA topoisomerases (e.g., interconverting the topological states of DNA). However, depending on whether they introduce transient breaks into ssDNA or dsDNA breaks, they are divided into two major groups, type I or type II respectively. Topoisomerase mechanisms and their effect on DNA topology are summarized in Figure 2.10.

The type I topoisomerases are further subdivided into type IA, IB, and IC according to

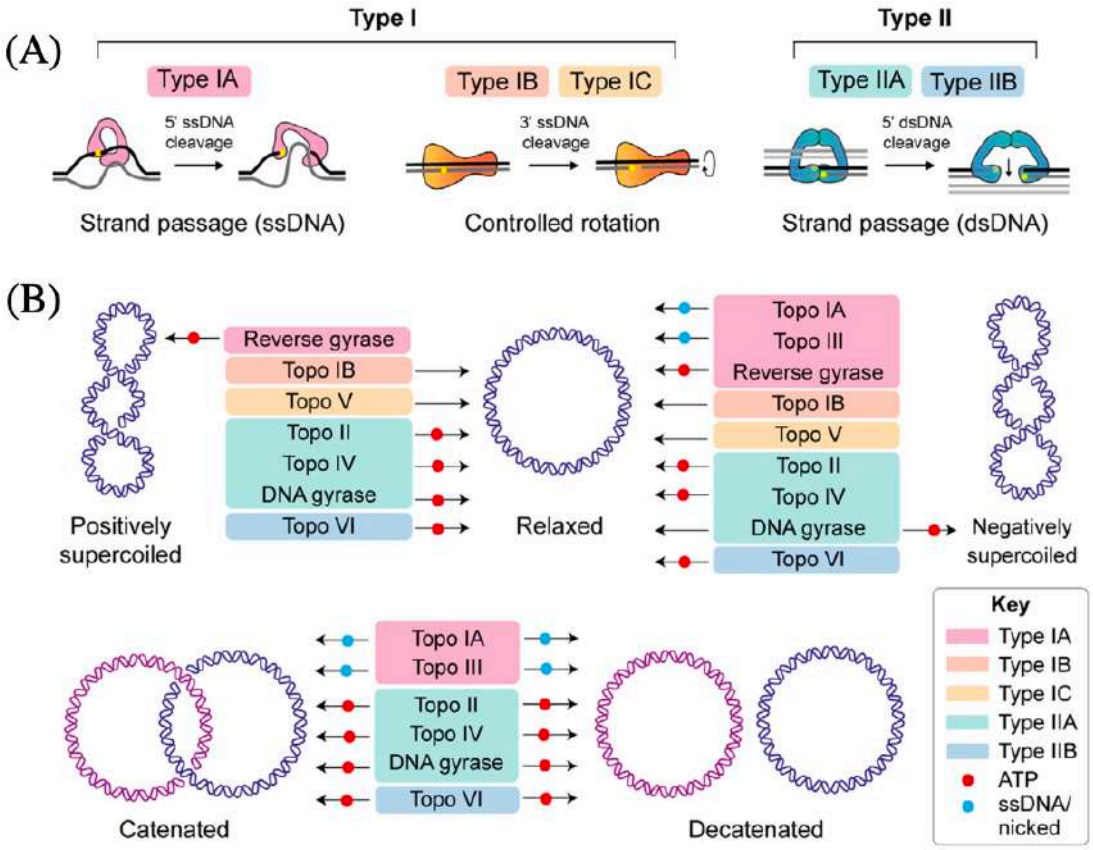


FIGURE 2.10: (A) Mechanisms of DNA topoisomerases. The topoisomerases bind to DNA and cut the sugar-phosphate backbone of either one (type I) or both (type II) of the DNA strands. The type I topoisomerases are further subdivided to type IA, IB and IC. Type IA opens a break (gate) through the formation of a transient covalent bond to the $5^{'}$ DNA phosphate and allows the other segment to pass through its cleavage (strand passage mechanism). Type IB and IC open a gate through the formation of a transient covalent bond to the $3^{'}$ DNA phosphate and rotate the free end around the intact strand (controlled-rotation mechanism). Type II topoisomerases are further subdivided into type IIA and IIB. They change the DNA topology by forming a transient covalent bond to the $5'$ DNA phosphate of both strands and function via a strand-passage mechanism. (B) Schematic of the topological manipulations in DNA by topoisomerases, relaxation of positively and negatively supercoiled DNA, and decatenation. The requirement of ATP or ssDNA for activity is indicated by a red or blue circle, respectively. The image is copied from (McKie, Neuman, and Maxwell, 2021).

mechanistic, sequence, and structural similarities. Type IA enzymes cleave the DNA backbone and generate a covalent linkage to the $5'$-phosphate in an Mg^{2+}- dependent and ATP-independent manner, while type IB and IC topoisomerases form a covalent linkage to the $3'$-phosphate, independently of both ATP and Mg^{2+}. Type IA topoisomerases change DNA topology using an enzyme-bridged strand passage mechanism (McKie, Neuman, and Maxwell, 2021). In this mechanism, ssDNA is transiently cleaved to allow passage of the other strand through the break. This mechanism removes one supercoil and strictly changes the linking

number (Lk) in steps of $+1$. As a result, Type IA enzymes can relax negatively supercoiled DNA (But not positively supercoiled). However, type IB enzymes leave the broken strand free to rotate around the intact strand (Champoux, 2001; Bates AD, 2005; Bush, Evans-Roberts, and Maxwell, 2015).

The type II topoisomerases are subcategorized as type IIA and IIB. They are structurally different, even though both catalyze dsDNA breaks through cleavage of the DNA backbone, generating a covalent linkage to the $5'$-phosphate on both duplex strands in an ATP/Mg^{2+} dependent manner.

The main bacterial topoisomerases in *E.coli* are Topo I and III (Type I), and DNA gyrase and TopoIV (Type II) (Champoux, 2001). Although there are some functionality overlaps between these DNA topoisomerases, each has its particular role in topological manipulations.

Topo I was the first topoisomerase discovered (Wang, 1971) and was initially named ω protein. It is found in eukaryotes and prokaryotes. Eukaryotic topo I topoisomerase are type IB enzymes and can relax both positive and negative supercoiled DNA. Bacterial topoI are type IA and can relax only negatively supercoiled. They are evolutionarily and mechanistically distinct from the Eukaryotic enzymes.

Topo III is highly conserved across evolutionary lineages. It is found in prokaryotes, eukaryotes, and archaea. Topo III is Mg^{2+}- dependent. This enzyme is a type IA topoisomerase that relaxes DNA; however, this does not appear to be its primary function in *E. coli*. It has been shown that it is involved in the segregation of chromosomes during replication (Bush, Evans-Roberts, and Maxwell, 2015).

Topo IV is a bacterial type IIA enzyme and is an ATP-dependent enzyme. It is capable of relaxing both positive and negative supercoils. Topo VI also has significant decatenate activity in a Mg^{2+}/ATP-dependent manner (McKie, Neuman, and Maxwell, 2021). *E.coli* topo IV is a heterotetramer consists of two subunits, encoded by the *parC* and *parE* genes.

DNA gyrase; Bacteria also have a unique type IIA topoisomerase called DNA gyrase. All bacteria contain DNA gyrase, but it is absent in higher eukaryotes. This enzyme is the only topoisomerase (among other types II topoisomerases) that can introduce negative supercoils into DNA at the expense of ATP, resulting in the global generation of negative supercoils in the bacterial chromosome at all times. DNA gyrase is a heterotetramer with two GyrA and two GyrB subunits. It is argued that the balance between Topo I and DNA gyrase's opposing activities maintains a steady-state level of average negative supercoiling in *E.coli*.

2.5.3 Dynamic properties of DNA supercoiling

DNA supercoiling not only can change in response to environmental changes (i.e., during the lifecycle of the bacterium (Krogh, Møller-Jensen, and Kaleta, 2018), environmental changes such as pH (Martis B et al., 2019)) but also different regions of the chromosomes experience different supercoiling levels (Guo et al., 2021; Visser et al., 2022). As reviewed in Sec. 2.5.2, topoisomerases play a crucial role in regulating supercoiling in bacteria. Moreover, NAPs such as H-NS are expected to prevent supercoiling (Junier and Rivoire, 2016). They prevent supercoiling from relaxing at their fixed points and hence play a topological role in shaping the

topological domains with different SC levels. In vivo and in vitro studies have shown that the *E. coli* chromosome is divided into approximately 400 supercoiled domains with an average size of 10 kbp (Postow et al., 2004; Deng, Stein, and Higgins, 2005). These domains appear topologically separated, meaning supercoils cannot move between adjacent domains (Deng, Stein, and Higgins, 2005). Another factor in local SC level variations is the transcription process of DNA to RNA (Fig.2.11). When an RNA polymerase (RNAP) transcribes a gene, strand separation leads to positive supercoiling through overwinding the DNA molecule ahead of the advancing protein machinery and negative supercoil formation behind (Liu and Wang, 1987). Moreover, during DNA replication, the strands are separated, which leads to the formation of positively supercoiled ahead of the replication fork, and the daughter strands form precatenanes behind the replication fork (Fig.2.11).

Genomic variations of the supercoiling of the *E. coli* chromosome is shown in Figure 2.12. Since supercoiling also regulates the transcription, there is a dynamic interaction between the transcription levels of neighboring genes (Meyer and Beslon, 2014; Grohens, Meyer, and Beslon, 2022).

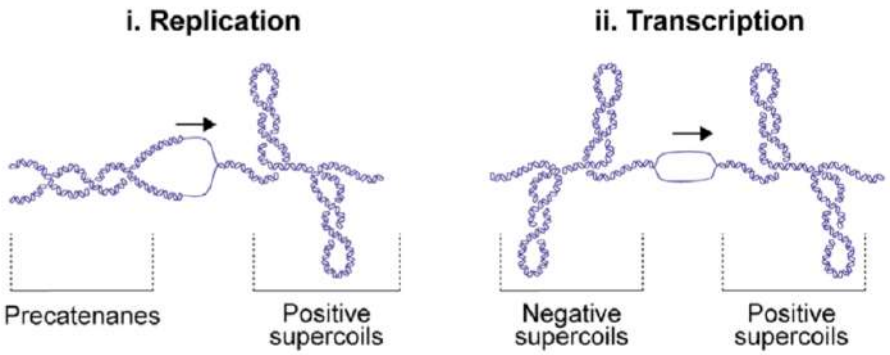


FIGURE 2.11: DNA metabolism has topological consequences. i. The replication process, separating the strands, leads to the formation of positively supercoiled ahead of the replication fork and the formation of precatenane behind. ii. Transcription process, separation of the strands leads to positive supercoiling through overwinding the DNA molecule ahead of the advancing protein machinery and negative supercoil formation behind. The image is copied from (McKie, Neuman, and Maxwell, 2021).

2.5.4 DNA supercoiling detection

Different approaches can be used to measure the level of supercoiling in DNA. Historically, one of the first methods to measure DNA supercoiling was centrifugation. This method is based on the differential migration of DNA molecules with different compaction through a sucrose density gradient (Mirkin, 2001; Corless and Gilbert, 2017). More compact molecules have a higher sedimentation rate than relaxed counterparts in a manner proportional to the level of supercoiling/compaction (Fig. 2.13 (A)).

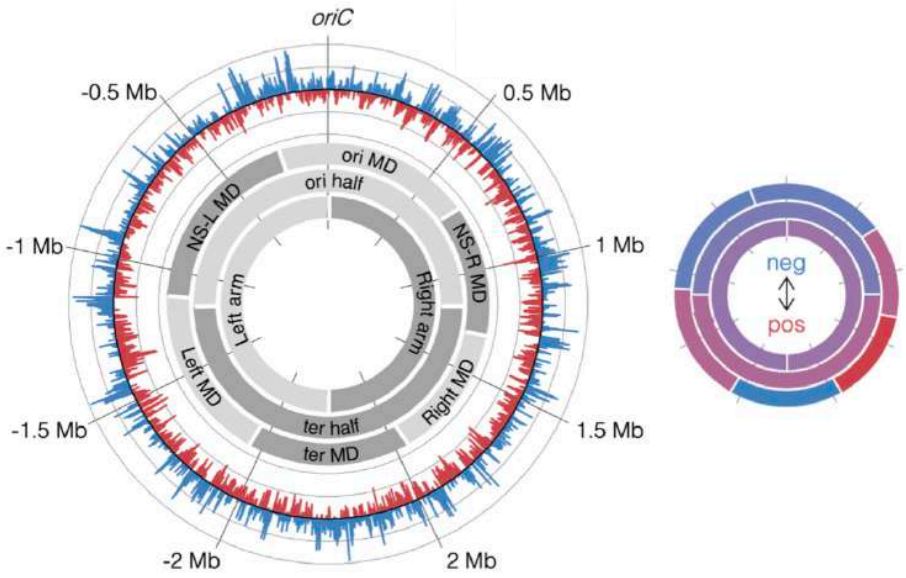


FIGURE 2.12: Large-scale supercoiling features of the E.coli chromosome. Left: Circular map at 1 *kb* resolution. Blue/red tracks indicate more negative/positive supercoiling than average ($\sigma = -0.06$). Shaded grey arcs indicate the positions of macro-domain regions. Right: corresponding heat map. The image is copied from (Visser et al., 2022).

This method was very laborious, so it has been largely superseded by the less complex agarose gel electrophoresis method. The electrophoresis approach is also based on the electrophoretic mobility of DNA molecules. Supercoiled DNA migrates faster than relaxed ones through an agarose gel in a way that molecules with different ΔLk form separate bands in the electrophoretic pattern. The band of higher mobility corresponds to compact supercoiled molecules, whereas the lower-mobility band comprises nicked-circular (open-circular) molecules (Fig. 2.13 (B1)).

The resolution of the agarose gel is not enough to differentiate topoisomers with different levels of DNA supercoiling, and they co-migrate as a single band. Therefore, agarose gel electrophoresis must be run in the presence of an intercalator agent, usually chloroquine Fig. 2.13 (B2)). An intercalator can insert itself between the stacked DNA base pairs and causes a local unwinding of the DNA helix of 26° per each molecule between adjacent base pairs (Fig. 2.13 (C)). As a result, chloroquine unwinds the double helix and releases negative supercoils, which leads to changing the DNA molecules' migration rates (Mirkin, 2001; Bates AD, 2005; Corless and Gilbert, 2017).

A $(2D)$ agarose gel electrophoresis technique is applied to study supercoiled templates that might simultaneously include positive and negative topoisomers. In this approach, the gel is run in the usual way, so separation and co-migration of both positively and negatively supercoiled DNA will occur. Electrophoresis is then performed in the presence of chloroquine in the perpendicular direction (90° rotation of gel). This method yields more detailed information about the supercoiling density of circular DNA (Fig. 2.13 (d));

In addition to using centrifugation and electrophoresis methods, there are other methods

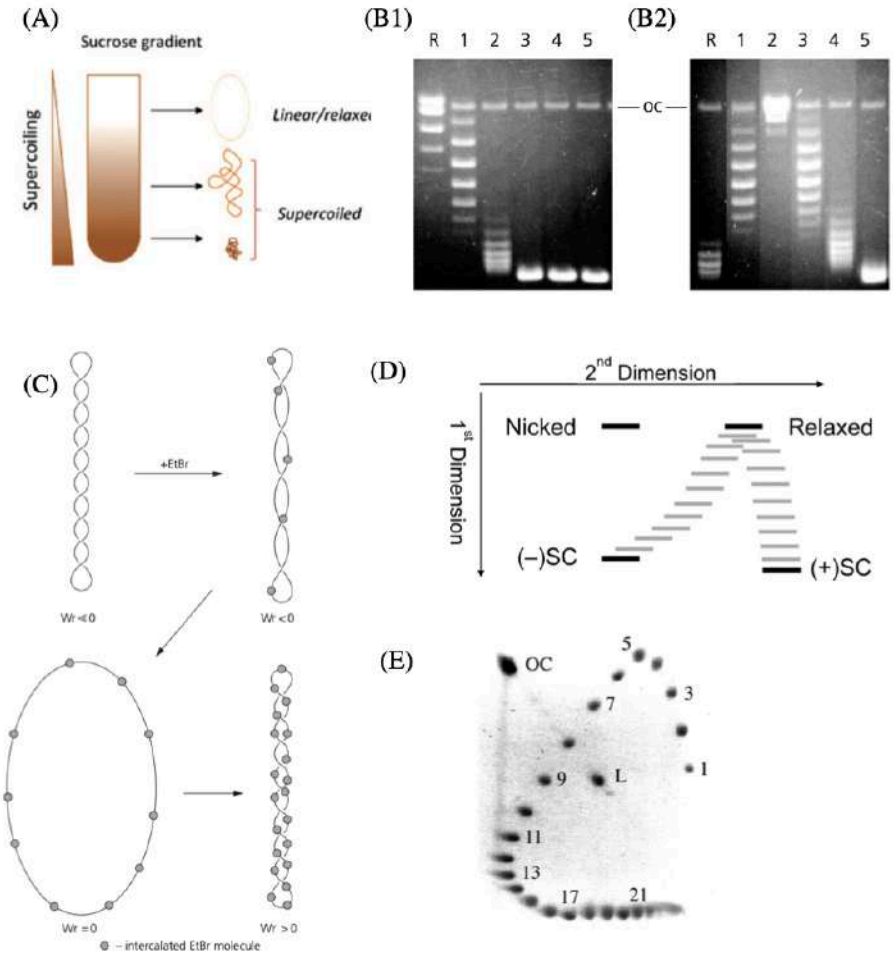


FIGURE 2.13: DNA Supercoiling Measurements. (A) The centrifugation method. Based on the migration rate of highly supercoiled/compact molecules, differentiate DNA supercoil level. The image is copied from (Corless and Gilbert, 2017). (B) Effect of an intercalator on electrophoresis. (B1) shows an agarose gel of increasing negative supercoiling, and (B2) shows the same samples electrophoresed in the presence of an intercalator. The relaxed sample (R) has positive writhe and higher mobility in (B2). Sample 3 is negatively supercoiled since in *B*2 has reduced negative writhe and migrates more slowly. Samples $3-5$ appear identical in (B1); however, they are resolved from each other in (B2). Sample 1 has the same mobility in both but is negatively writhed in (B1) and positively writhed in (B2). oc: open-circular DNA. The image is copied from (Bates AD, 2005). (C) The effect of intercalator on an initially negatively supercoiled DNA. It causes a reduction in the twist and a concomitant increase in writhe. The image is copied from (Bates AD, 2005) (D) Schematic of two-dimensional agarose gel electrophoresis. Positively supercoiled molecules migrate as the arc between $(+)SC$ and the relaxed band. Negatively supercoiled molecules move as the arc between relaxed and $(-)SC$. The image is copied from (Gibson, Oviatt, and Osheroff, 2020). (E) An example of separation of DNA topoisomers by two-dimensional gel electrophoresis. $1-4$ have positive supercoiling, and the rest are negatively supercoiled. The image is copied from (Higgins and Vologodskii, 2015).

like psoralen-based methods (Mirkin, 2001; Corless and Gilbert, 2017) to measure the plasmids' supercoiling density. Topology of DNA molecules can study with recently developed methods like fluorescent evaluation of DNA supercoiling (FEDS) (Duprey and Groisman, 2020), recombination assays (Rovinskiy et al., 2019). It can also be analyzed by single-molecule methods such as transmission electron microscopy (Viguera et al., 2000), atomic force microscopy (López et al., 2012) and magnetic tweezers (Charvin, Bensimon, and Croquette, 2003). However, detailed information about the supercoiled molecule's shape and the local supercoiled level is still missing. (Mirkin, 2001; Bates AD, 2005; Higgins and Vologodskii, 2015; Corless and Gilbert, 2017; Cebrián et al., 2021). Future works must go beyond image limitations and technical capacities to provide detailed insight into DNA supercoiling *in vivo*.

2.6 Free-Energy of DNA supercoiling

Getting precise knowledge of the plectonemes inside the cell is almost impossible. In this context, we use a thermodynamics approach to estimate the feature of the plectoneme. The plectonemic superhelix consists of two right helices that are interwound and the presence of small loops located at the tips of the plectonemes are neglected in the calculations. Moreover, sequence effects are also neglected at this level of modeling (Marko and Siggia, 1994; Marko and Siggia, 1995b; Marko, 2015; Barde, Destainville, and Manghi, 2018).

The equation for the right-handed helical curve can be written as (White and Bauer, 1986) (Fig. 2.14):

$$\vec{r}(s) = (R\cos\theta, R\sin\theta, p\theta), \tag{2.5}$$

where R is the radius of the helix and $2\pi p$ is the pitch.

The total supercoil free energy is obtained as the sum of bending energy, twist elastic energy, entropic and electrostatic contributions (Marko and Siggia, 1994; Marko and Siggia, 1995b; Marko, 2015; Barde, Destainville, and Manghi, 2018):

$$\beta F = \beta E_b + \beta E_t + \Delta s + w(R) + w(\pi p) \tag{2.6}$$

The bending energy is:

$$E_b = \frac{l_p}{2}\int_0^L \kappa^2 ds \tag{2.7}$$

where l_p is the bending persistence length and κ is the curvature of the curve. By applying the curvature definition (eq. 2.40), we get:

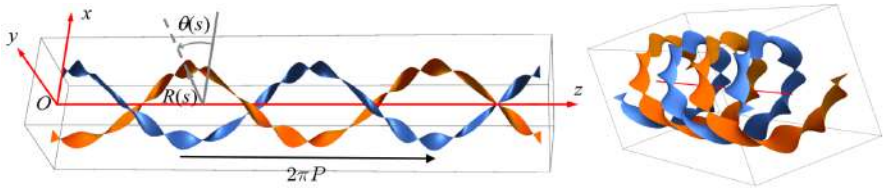


FIGURE 2.14: Left: Schematic of plectoneme, which is constructed of two ribbons, each of them representing dsDNA. The curvilinear abscissa along the molecular axes is denoted by s. The position $\vec{r}(s)$ is defined in the cylindrical coordinates. Right: a different point of view of the same plectoneme. The image is copied from (Barde, Destainville, and Manghi, 2018).

$$E_b = \frac{l_p L}{2}\left(\frac{R}{R^2+p^2}\right)^2 \tag{2.8}$$

The twist elastic energy is defined as (Marko, 2015):

$$E_t = \frac{C}{2L}\Theta^2 \tag{2.9}$$

where C is the torsional persistence length and $\Theta = 2\pi T w$ is the net twist angle along the double helix (DNA twisting). By substituting equations 2.3, 2.4, and writhe definition (eq. 2.45):

$$E_t = \frac{C}{2L}(2\pi(\Delta Lk - Wr))^2 \tag{2.10}$$

$$= \frac{2\pi^2 C}{L}\left(\sigma L k_0 + \frac{L}{2\pi}\frac{p}{R^2+p^2}\right)^2 \tag{2.11}$$

$$= \frac{2\pi^2 C}{L}\left(\sigma \frac{L}{n l_0} + \frac{L}{2\pi}\frac{p}{R^2+p^2}\right)^2 \tag{2.12}$$

$$= \frac{CL}{2}\left(\sigma\omega_0 + \frac{p}{R^2+p^2}\right)^2 \tag{2.13}$$

where $\omega_0 = \frac{2\pi}{n l_0}$, $n = 10.5$ bp per ds-DNA helix turn and $l_0 = 34$ nm the ds-DNA rise per base pair.

The entropic cost of bringing the filament close to itself is given by Eq. 2.14. This term prevents the collapsing of the plectoneme into a line.

$$\Delta s = \frac{L}{(l_p R^2)^{1/3}} + \frac{L}{(l_p p^2)^{1/3}} \tag{2.14}$$

The electrostatic contribution is (Marko, 2015):

$$w(R) + w(\pi p) \tag{2.15}$$

$$w(x) = l_B \nu^2 K_0(2x/\lambda_D) \tag{2.16}$$

$l_B = \frac{e^2}{\varepsilon k_B T} = 0.7$ nm is the Bjerrum length (the distance at which the electrostatic interaction is comparable in magnitude to the thermal energy scale, $k_B T$), $\lambda_D = 0.8nm$, $\nu = 8.4nm^{-1}$ and $K_0(x)$ is a modified Bessel function that decays exponentially for large x and diverges logarithmically for small x (Marko and Siggia, 1995b). As the electrostatic term is very small compared to other terms, we neglect it in our calculation.

Hence, the total free energy per length is obtained by:

$$f = \frac{\beta F}{L} = \frac{l_p}{2}(\frac{R}{R^2+p^2})^2 + \frac{C}{2}(\sigma\omega_0 + \frac{p}{(R^2+p^2)})^2 + \frac{1}{(l_p R^2)^{1/3}} + \frac{1}{(l_p p^2)^{1/3}}. \tag{2.17}$$

The free energy depends on two parameters R and p (correspondingly curvature κ and torsion τ of the curve, for more detail, see 2.8). By minimizing the free energy, we can determine the equilibrium values of R and p. Fig.2.15 shows this free energy as a function of supercoiling density $|\sigma|$.

By minimization of the free energy with respect to R and p, we get:

$$\partial_p f = -2l_p \frac{R^2 p}{(R^2+p^2)^3} + C(\sigma\omega_0 + \frac{p}{(R^2+p^2)})(\frac{R^2-p^2}{(R^2+p^2)^2}) = 0 \tag{2.18}$$

$$\tag{2.19}$$

$$\partial_R f = -l_p \frac{R(R^2-p^2)}{(R^2+p^2)^3} + C(\sigma\omega_0 + \frac{p}{(R^2+p^2)}) \times (\frac{-2Rp}{(R^2+p^2)^2}) - \frac{2}{3}\frac{1}{(l_p R^5)^{1/3}} = 0 \tag{2.20}$$

Figure 2.16 shows the result for $C = 110$ nm, $l_p = 50$ nm, $\omega_0 = 2\pi/nl_0 = 176$ rad/nm, ($n = 10.5$ bp per ds-DNA helix turn and $l_0 = 34$ nm the ds-DNA rise per base pair). The equations are solved numerically, and the same graph as reference (Barde, Destainville, and Manghi, 2018) was reproduced. The order of magnitude of R and p are quite realistic compared with experimental and numerical results (Barde, Destainville, and Manghi, 2018).

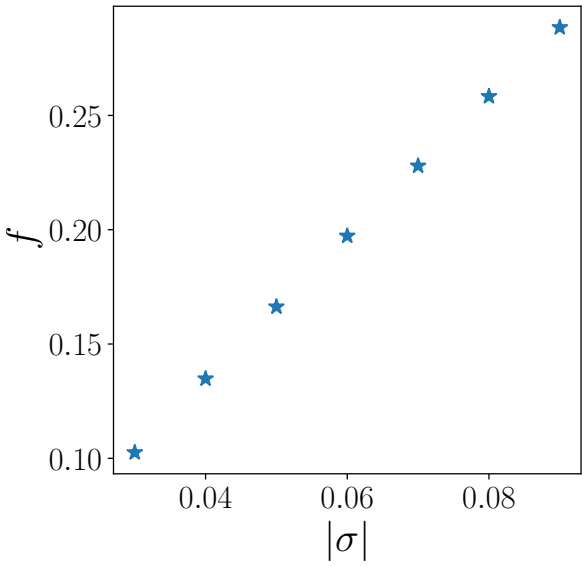


FIGURE 2.15: Free energy per length f, of plectoneme as a function of the supercoiling density $|\sigma|$.

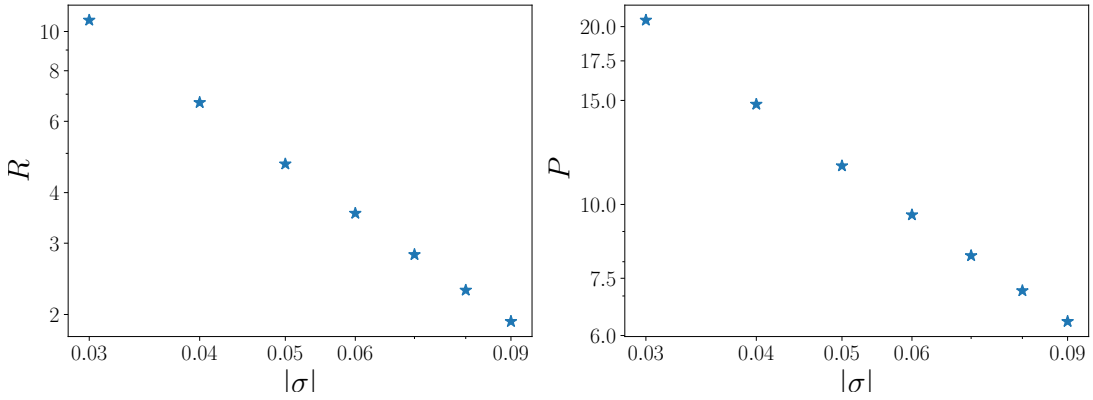


FIGURE 2.16: Plectoneme radius R (left) and its pitch p (right), both in nm, as a function of the supercoiling density $|\sigma|$ for $C = 110$ nm and $\omega_0 = 2\pi/nl_0 = 1.76$ rad/nm.

2.7 Appendix A: Frenet-Serret frame

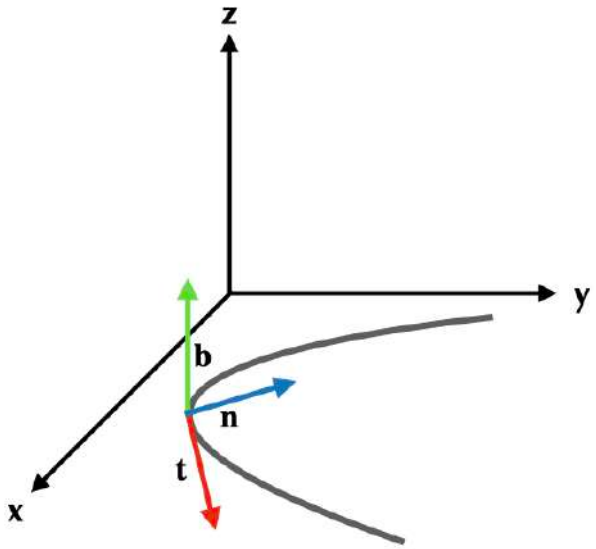


FIGURE 2.17: A curve in space. At any point $r(s)$, there is a tangent vector, $\vec{t}$, a normal vector $\vec{n}$, and the binormal vector $\vec{b}$. I need to indicate s and r in it.

The local Frenet-Serret frame can be used to study a curve $\vec{r}(s)$ parameterized by the arc length along the curve s, $0 \leq s \leq L$. This frame is unique, i.e., for every curve, the frame is uniquely defined (Kamien, 2002). It consists of three vectors that can describe the curve, the tangent vector $\vec{t}$, the normal vector $\vec{n}$, and the binormal vector $\vec{b}$.

Tangent vector, The unit tangent vector at a point s on the curve:

$$\vec{t} = \frac{\vec{dr}}{ds} / \| \frac{\vec{dr}}{ds} \|, \tag{2.21}$$

Normal vector,

$$\vec{n} = \frac{\vec{dt}}{ds} / \| \frac{\vec{dt}}{ds} \|, \tag{2.22}$$

The magnitude of the rate of variation of the unit tangent vector is the curvature κ of the curve.

$$\kappa = \| \frac{\vec{dt}}{ds} \|, \tag{2.23}$$

Binormal vector,

$$\vec{b} = \vec{t} \times \vec{n}, \tag{2.24}$$

The binormal vector is a unit vector perpendicular to both $\vec{t}$ and $\vec{n}$.

Moving on the curve $r(s)$ along s, the normal vector can change its direction if

1) The normal vector rotates together with the tangent vector to stay perpendicular to the curve. In this case, the curve stays in the same flat plane.

2) By rotating around the tangent vector. This case corresponds to the rotation of the plane of the curve.

$$\frac{\vec{dn}}{ds} = \alpha\vec{t} + \tau\vec{b}, \tag{2.25}$$

Here, τ is defined as the torsion of the curve, and it is measured as the rate of the variation of the binormal.

To calculate α, since the derivative of any unit vector is perpendicular to itself, differentiating $\vec{t}\cdot\vec{n} = 0$ ends up in:

$$\frac{\vec{dt}}{ds}\cdot\vec{n} + \vec{t}\cdot\frac{\vec{dn}}{ds} = 0, \tag{2.26}$$

by substituting equations 2.22 and 2.25 in equation 2.26, we get $\alpha = -\kappa$. In order to complete the analysis of the curve, we calculate:

$$\frac{\vec{db}}{ds} = \frac{d}{ds}(\vec{t}\times\vec{n}) = \frac{dt}{ds}\times\vec{n} + \vec{t}\times\frac{\vec{dn}}{ds} \tag{2.27}$$

$$= \kappa\vec{n}\times\vec{n} - \kappa\vec{t}\times\vec{t} + \tau\vec{t}\times\vec{b} \tag{2.28}$$

$$= \tau\vec{t}\times\vec{b}. \tag{2.29}$$

By applying the rule:

$$\vec{a}\times(\vec{b}\times\vec{c}) = \vec{b}(\vec{a}.\vec{c}) - \vec{c}(\vec{a}.\vec{b}). \tag{2.30}$$

equation 2.29, will be:

$$\frac{\vec{db}}{ds} = \tau\vec{t}\times\vec{b} = \tau\vec{t}\times(\vec{t}\times\vec{n}) \tag{2.31}$$

$$= \tau[\vec{t}(\vec{t}.\vec{n}) - \vec{n}(\vec{t}.\vec{t})] \tag{2.32}$$

$$= -\tau\vec{n} \tag{2.33}$$

We can summarize the Frenet-Serret equations for a curve in three dimensions as (Kamien, 2002):

$$\frac{d}{ds}\begin{bmatrix}\vec{t}\\ \vec{n}\\ \vec{b}\end{bmatrix} = \begin{bmatrix}0 & \kappa & 0\\ -\kappa & 0 & \tau\\ 0 & -\tau & 0\end{bmatrix}\begin{bmatrix}\vec{t}\\ \vec{n}\\ \vec{b}\end{bmatrix}. \tag{2.34}$$

2.8 Appendix B: Solenoidal Helix

The plectonemic superhelix consists of two right helices that are interwound. The equation for the right-handed helical curve can be written as (White and Bauer, 1986) (Fig. 2.14):

$$\vec{r}(s) = (R\cos\theta, R\sin\theta, p\theta), \tag{2.35}$$

where R is the radius of the helix and $2\pi p$ is the pitch. The arc length in 1 radian of the superhelix is $l = \sqrt{R^2 + p^2}$, therefore the arc length is

$$ds = ld\theta = \sqrt{R^2 + p^2}d\theta, \tag{2.36}$$

where $0 \leq s \leq L$.

The unit tangent vector is:

$$\vec{t} = \frac{\vec{dr}}{ds} = \frac{\vec{dr}}{d\theta}\frac{\vec{d\theta}}{ds} = \frac{1}{l}(-R\sin\theta, R\cos\theta, p) \tag{2.37}$$

By applying eq.2.22 and eq.2.23, the curvature of the curve will be:

$$\frac{\vec{dt}}{ds} = \frac{\vec{dt}}{d\theta}\frac{\vec{d\theta}}{ds} = \tag{2.38}$$

$$\frac{1}{l^2}(-R\cos\theta, -R\sin\theta, 0) \tag{2.39}$$

$$\kappa = \|\frac{\vec{dt}}{ds}\| = \frac{R}{R^2 + p^2} \tag{2.40}$$

Hence the normal vector $\vec{n} = (-\cos\theta, -\sin\theta, 0)$ has no component in the z direction and remains always horizontal.

The binormal vector is obtained from equation 2.24:

$$\vec{b} = \frac{1}{l}(p\sin\theta, -p\cos\theta, R), \tag{2.41}$$

So the Frenet equation:

$$\frac{\vec{db}}{ds} = \frac{1}{l^2}(p\cos\theta, p\sin\theta, 0) \tag{2.42}$$

Hence:

$$\tau = \frac{p}{R^2 + p^2}. \tag{2.43}$$

So torsion and curvature are constant along the superhelix.

The total torsion of the curve of length L (which is indeed the twist of the curve) is defined as (White and Bauer, 1986):

$$T_w = \frac{1}{2\pi}\int_0^L \tau(s)ds = \frac{L}{2\pi}\frac{p}{R^2+p^2}, \tag{2.44}$$

The factor $\frac{1}{2\pi}$ is used for normalization and helps to express the twist in the number of turns instead of radians.

Since the linking number is topologically invariant, the linking number does not change when the plectonemes are formed, this means (White and Bauer, 1986; Boles, White, and Cozzarelli, 1990; Marko, 2015; Barde, Destainville, and Manghi, 2018)[ref: Calugareanu-Fuller-White theorem] that $\Delta Lk = 0$, hence:

$$Wr = -Tw = -\frac{L}{2\pi}\frac{p}{R^2+p^2}. \tag{2.45}$$

Chapter 3

Investigating $3D$ Organization of Chromosome

To answer one of the fundamental questions in biological studies on how the spatial conformation of the DNA affects its functions like gene regulation, gene expression, DNA transcription, and replication (Oluwadare, Highsmith, and Cheng, 2019), we first need to understand chromosome conformation. Powerful experimental methods, including microscopy-based techniques such as DNA Fluorescence in situ Hybridization (DNA FISH) and chromosome conformation capture methods, such as Hi-C, have opened up new possibilities for understanding chromosome architecture. Even though experimental research has come a long way, there is still not enough information to fully comprehend the structure and function of the genome. Computational models are a crucial complement to interpret experimental data and provide quantitative comprehension of how chromosomes fold, move, and interact. The following sections present a brief review of selected experimental approaches and studies developed over the recent years to model the $3D$ chromosome architecture.

3.1 Chromosome conformation capture experiments

3.1.1 Captured conformations of chromosomes

Chromosome conformation capture $(3C)$-based techniques study the $3D$ architecture of the genome by connecting genomic position with the spatial organization. The $3C$ assay was first developed by Dekker et al. in 2002 (Dekker et al., 2002) and measured the contact frequency between two specific genomic sites (or loci) in a cell population (one vs. one). Several 3C derivative techniques have been developed to overcome the limitation of 3C. Later improvements consisted of counting a single loci's contact with several other loci on the chromosome, $4C$ (Lomvardas et al., 2006; Simonis et al., 2006; Würtele and Chartrand, 2006; Zhao et al., 2006). These one-to-all approaches help capture genome-wide interactions made by a single locus (Sati and Cavalli, 2017). Since $3C$ and $4C$ methods study interactions of a single preselected locus, they are insufficient to capture

general principles of chromosome organization. The 5*C* technique allows the acquirement of the contact information of multiple genomic loci in a large genomic region, many vs. many, (Dostie et al., 2006). Finally, Hi-C, an all-to-all assay (Lieberman-Aiden et al., 2009a), which combines the 3*C* technique with a high throughput sequencing method, can detect the contacts between all loci genome-wide. Today, Hi-C is the most widely used and well-known technique. During the last decade, an enormous amount of data were generated using the Hi-C methods. It has revolutionized chromosome studies by measuring the number of interactions between chromosome fragments. The Hi-C data is commonly summarized in the form of a genome-wide contact frequency matrix (Berkum et al., 2010). The principle steps in (3*C*) and methods derived from it are outlined in Figure 3.1.

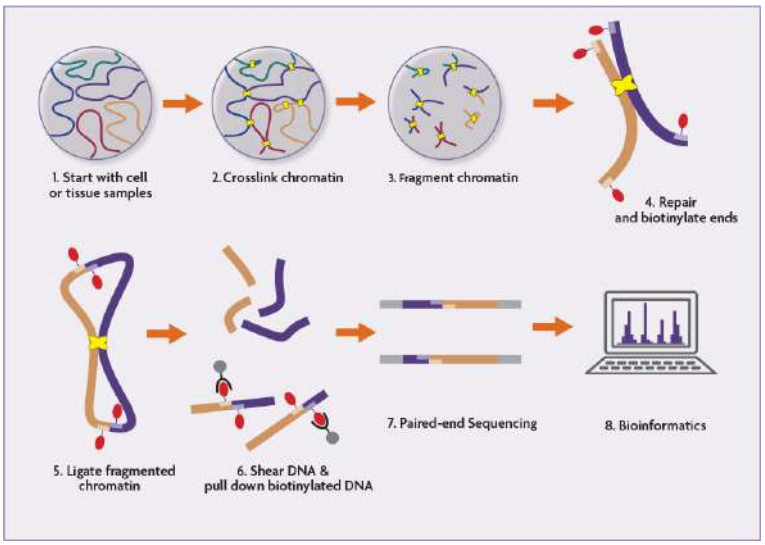


FIGURE 3.1: Schematic of the Hi-C Workflow. This figure was obtained from (https://www.activemotif.com/blog-hi-c)

3.1.2 Hi-C method

In a typical Hi-C experiment, the protocol (as depicted in Figure 3.1) involves the following steps (Lieberman-Aiden et al., 2009a). First, to freeze contacts in place, a population of cells is cross-linked with formaldehyde and lysed. Therefore, the covalent links between adjacent chromatin segments are formed. Second, DNA is digested with a restriction enzyme. The restriction enzyme is able to break whole chromosomes at specific restriction sites (nuclease) and produce restriction fragments with a $5'$ overhang. Then the $5'$ overhang in each restriction fragment pair is joined via the ligation procedure. As a result, small DNA rings are made of two restriction fragments. This step also removes the restriction sites and adds biotin tags in place for use in the next step. Fourth, the DNA solution is purified, and the resulting DNA samples contain ligated fragments marked with biotin at the junction. In the last step, the ligated fragments are amplified by

PCR and sequenced using paired-end high-throughput sequencing, yielding a collection of "reads". A complex bioinformatics treatment is required to map the reads onto a reference genome and identify interacting restriction fragments and their frequency count. This last step has many caveats and is known to be prone to error.

3.1.3 Contact Matrix

Once the Hi-C library has been obtained, raw data of read-pairs is mapped onto a location on the genome to obtain a count map with elements n_{ij} via bioinformatic methods. n_{ij} is the number of contact events between two genomic fragments i and j. Fragments, also known as genomic loci or bins, are the regions in which the chromosome has been partitioned. The size of the genomic loci is defined by the number of base pairs in it (range from 1kb to 1Mb). Then the interaction frequencies (also known as contact frequencies), c_{ij} between any pairs (i, j) of loci are assessed by normalizing the counts' map. The loci-to-loci interactions along the DNA are summarized in a two-dimensional matrix (also referred to as a contact matrix or contact map), where rows and columns represent the number of fragments into which the DNA was divided. Heatmaps are usually used to visually represent the contact matrix where the intensity displays the interaction frequency. The size of the fragments defines the resolution (e.g., $1kbp$) of the contact map. Normalizing count maps into genome-wide maps of relative contact probability matrices requires the elimination of systematic biases.

The resolution of a Hi-C dataset has increased in the last decade by roughly 1-fold every three years. The best resolution available to date is 1 to 5kb. However, the resolution is limited by multiple factors, and the repeatability and reliability of Hi-C experiments and their bioinformatics steps come under scrutiny. Artifacts can emerge from various sources, including protocol efficiency (e.g., ligation and fill-in efficiency), PCR amplification of the purified reads, the size of the cross-link between formaldehyde, and varying lengths of DNA.

3.1.4 Normalization

In Hi-C experiments, the contact frequencies between two genomic loci are roughly proportional to the reads counts obtained between two regions after sequencing. Raw contact frequencies contain several systematic biases, like uneven distribution of restriction fragment sizes, the distance between restriction sites, mappability of individual reads and sequencing platforms (e.g., differences in GC content) (Yaffe and Tanay, 2011).

In the Hi-C method, an enormous number of cells are used to analyze the contact frequencies between all loci (e.g., at least 25 million for eukaryotes cells (Lieberman-Aiden et al., 2009a; Berkum et al., 2010)) since a population of cells has significantly larger sequence reads compared to a single cell. Only read pairs in which both reads were uniquely aligned to the genome are considered and interpreted as the interaction count.

In other words, higher read counts mean higher interaction frequencies. However, the systematic biases affect the Hi-C read counts, which means some fragments will be involved in fewer interactions than expected. In contrast, others will be overrepresented in the final bank. Therefore, estimating and smoothing out all the biases is essential to provide a cleaner view of the interaction frequencies and ensure accurate analysis (An example of Hi-C data before and after normalization is shown in Fig.3.2). Theoretically, the normalization factor in transforming counts into contact probabilities should be the total number of cells in the experimental samples. As a result, the contact probability is defined as $c_{ij} = n_{ij}/N$, where n_{ij} is the number of cells that have contact between loci i and j. In practice, it is not the case since N is unknown.

In parallel with the development of experimental methods, several algorithms and software packages have been developed to normalize Hi-C data. Historically, normalization is conducted by dividing each n_{ij} by the average over all genomic loci pairs with the same genomic contour distance, $c_{ij} = n_{ij}/\langle n_{ij} \rangle$ (Lieberman-Aiden et al., 2009a). However, there is no stringent justification for this normalization method. All the proposed normalization methods can be categorized into two groups: explicit factor correction algorithms and matrix balancing methods. (Ay and Noble, 2015; Servant et al., 2018).

Explicit-factor normalization methods

These methods need a priori information about the sources and factors involved in Hi-C systematic biases. Yaffe and Tanay (Yaffe and Tanay, 2011; Ay and Noble, 2015; Servant et al., 2018) proposed a probabilistic model for the analysis of the contact maps to estimate the probability of observing a contact between two loci given these biases. The downside to this method is the relatively high computational cost. However, the HiCNorm. (Hu et al., 2012) method, based on regression models (Poisson or Negative Binomial regression), was later proposed that was significantly faster.

Matrix balancing methods

These methods collectively correct all factors that affect experimental visibility without taking any specific biases into consideration. Theoretically, this method should be able to correct all variances in the contact maps. The critical assumption in these types of methods is that each locus in the genome would be "equally visible" in the absence of any biases. The normalization is transformed to a matrix balancing where the goal is to find a decomposition of the raw contact matrix $C = \vec{b}^T N\vec{b}$. Where b is a (column) vector of the bias terms and N is a normalized contact map (all rows have equal sums, $\sum_i N_t ij = k$).

Imakaev et al. (Imakaev et al., 2012) proposed **ICE**, Iterative correction and eigenvector decomposition method. In this method, the systematic biases between two loci i and j are

represented as the product of their individual biases, $N_{ij} = b_i b_j C_{ij}$. The maximum likelihood solution for the individual biases is obtained by applying an iterative correction procedure.

Cournac et al. (Cournac et al., 2012; Lyu, Liu, and Wu, 2020), proposed **SCN**, Sequential component normalization method, which is very similar to **ICE**. In this method, all columns and rows are successively scaled to one by using the euclidian norm. The process is repeated sequentially until the matrix becomes symmetric again, with each row and each column normalized to one.

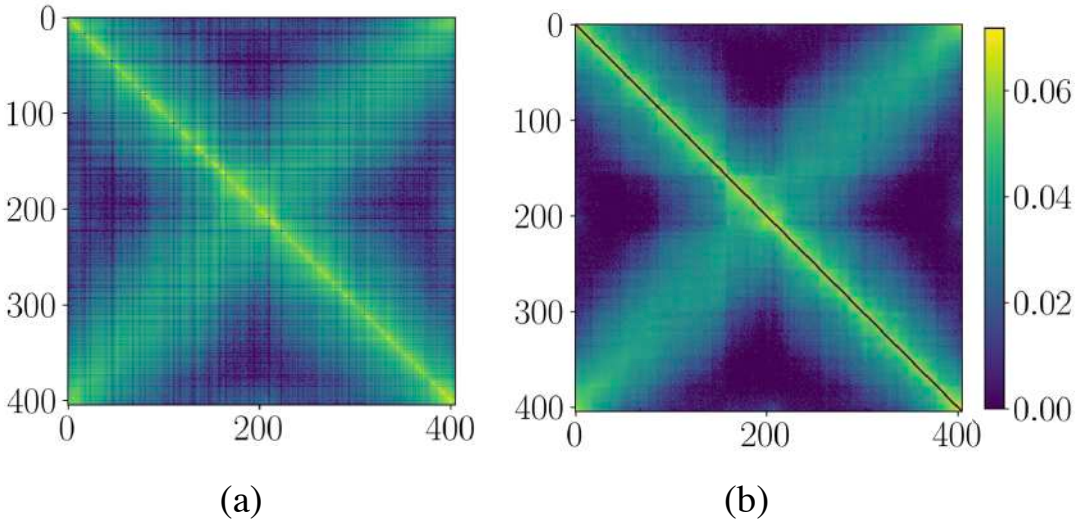


FIGURE 3.2: Hi-C matrixes for *Caulobacter crescentus* bacteria before (a) and after (b) normalization. The raw data is from (Le et al., 2013).

3.2 Microscopy-Based Techniques

Although our study uses and highlights Hi-C data, it is noteworthy to examine the complementary microscopy-based methods used to study the DNA configurations. An optical microscope (light microscope) uses visible light to detect the structural details in the cell. However, the diffraction barrier limits the spatial resolutions to approximately half the wavelength under observation. Super-resolution microscopy (SMR) approaches have been proposed to circumvent the diffraction limit of light microscopy, which can achieve spatial resolutions smaller than 10 nm (Miriklis et al., 2021).

Earlier techniques to study chromosome conformation are cytogenetic techniques like fluorescence in situ hybridization (FISH). Fish employs complementary fluorescently-labeled DNA probes to detect a specific region and labeled locus on the chromosome. FISH has been widely used to measure the spatial distance between two regions and provide direct visualization of the conformation of the chromosome. For instance, studying

E.coli by FISH revealed two so-called macro-domain near the origin and terminus of replication, called Ori and Ter, respectively (Le and Laub, 2014).

Unfortunately, such studies are challenging due to the need for photobleaching. Photobleaching causes the fading of the dye or a fluorophore molecule, which makes it unable to fluoresce. Moreover, fluorescent imaging requires the bacteria to be fixed and permeabilized, which may alter the chromosome conformation. Another limitation is that cell fixation hinders the study of the internal dynamics of the chromosome (Wu et al., 2019). These limitations were partially overcome by the developments of fluorescent repressor-operator systems (FROS). In one derivation of this method, using ParB/*parS* systems derived from plasmid and chromosome partitioning systems yields to visualization of specific loci. ParB proteins bind to *parS* site, initiate the polymerization, and spread additional ParB proteins into DNA. So by inserting a single *parS* site into specific loci, the expression of a fluorescently-tagged ParB makes the loci visible under the microscope. An image of *E.coli* is shown in Fig.3.3.

One restriction of this method is that only $2-3$ loci may be detected at once. One common method to visualize the global organization of bacterial chromosomes *in vivo* is imaging cells stained with dyes (e.g., DAPI) that non-specifically bind to DNA. An image of *E.coli* is shown in Fig.3.3

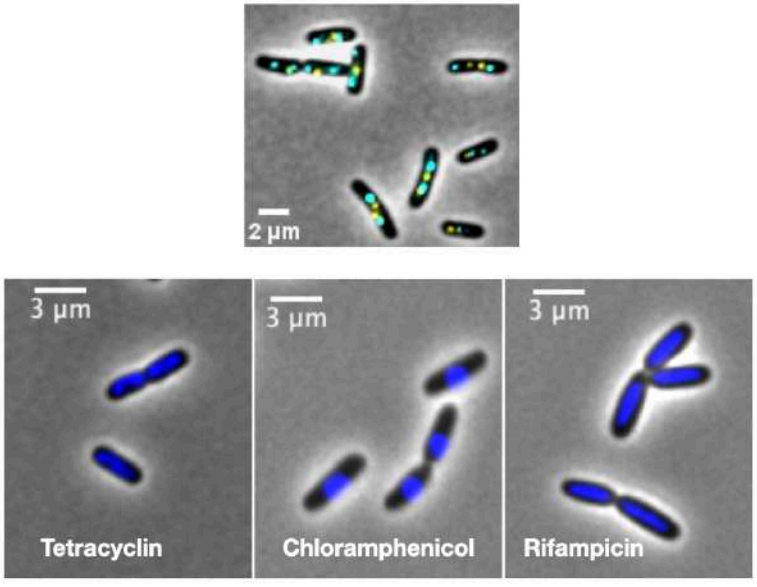


FIGURE 3.3: Fluorescent images of *E. coli*. Top: Fluorescently-tagged ParB reveals the presence of $2-3$ loci, making them visible under the microscope. Bottom: Images of cells stained with DNA-binding dyes (e.g., DAPI), which bind non-specifically to DNA. The lower images showcase *E. coli* subjected to different antibiotics, illustrating alterations in the global DNA structure in response to environmental conditions.

DAPI stainng also uses fixed cells; hence this technique also precludes analyses of living

cells. New methods involve fluorescently-tagged proteins that bind throughout the chromosome (like RNA polymerase subunits and nucleoid-associated proteins). The distribution of these proteins provides a reasonable chromosome proxy and leads to important new insights (Le and Laub, 2014).

Although microscopy-based methods have played a significant role in shedding light on how bacterial genomic DNA is packed, it is challenging to answer fundamental questions regarding the (sub)structure and dynamics, the interplay between structure and function (Brocken, Tark-Dame, and Dame, 2018) using this method.

In conclusion, a truly ingenious, integrative approach combining high-resolution imaging and Hi-C technologies with computational modeling for identifying $3D$ genome organization is needed. To this end, computational modeling can leverage the wealth of experimental Hi-C or microscopy data on chromatin folding to hint to specific folding motifs of interest.

3.3 From experimental data to chromosome folding

Understanding the organizational principles of chromosomes is a very attractive field for biologists, physicists, and computational researchers. Over the last couple of years, several models and computational methods have been developed to construct chromosome architecture using experimental data. These methods differ in their approach, the assumptions they make, and the adopted implementations. Here, some of these methods are reviewed based on how they model chromosomes.

3.3.1 Approaches

Since the chromosome is a polymer, preexisting theoretical and computational works in the field of statistical physics of polymers are uniquely suited to study chromosomes. These physical models can provide predictive mechanistic insights into the chromosome structures at a quantitive level (Bottom-up approaches) (Rosa and Zimmer, 2014). Bottom-up or direct models look at the entire system and try to distill insights from the larger view by testing whether certain known or hypothesized physical or biological principles can lead to ensembles that agree with experimental data. In other words, the highest priority is deriving universal behavior by applying physics laws and physical analysis. As a result, these models use a minimum set of assumptions and parameters and have strong predictive power.

Another approach, data-driven or inverse methods, uses large sets of experimental data directly to reconstruct chromosomal configurations *in silico* that are consistent with the input data. These models can then be used to study other observables of the genomic architecture. Both approaches have notable different assumptions, implementations, and

numerical or mathematical tools, while the distinction between bottom-up and data-driven models is not entirely black and white. According to the number of parameters used in either model, a useful distinction can be drawn between them. Bottom-up models typically use several parameters (a handful of parameters), and data-driven approaches typically use thousands or more values from large datasets, often genome-wide ones. Here we review a selection of these models and methods that have tried reproducing essential features of bacterial chromosomes.

3.3.2 Data-driven methods

As described in the previous section, experimental data are used directly in this approach to produce a detailed spatial configuration of the chromosome *in silico* that are as consistent as possible with those experimental data. The most common form of experimental data to reconstruct the 3*D* organization of DNA are chromosome contact maps resulting from 3*C* methods.

Most data-driven approaches directly or indirectly adopt polymer descriptions of DNA, a chain of interacting monomers. Therefore, the bonds between adjacent monomers describe the connectivity of the polymer, e.g., harmonic or fixed bonds. In addition, excluded volume interactions between each pair of monomers are needed to model self-avoiding polymers, e.g., hard or soft repulsion due to collisions. The chain's stiffness can also be manipulated by controlling the angle between linearly connected monomers. Finally, on top of these fundamental polymer topologies and interactions, a set of pairwise interactions between monomers (these restraints are assumed based on some assumptions) is added and then tuned to get results that are as consistent as possible with the experimental data. Once a customized polymer model is at hand, Monte Carlo or Molecular Dynamics simulations are used to sample different configurations.

Some of these methods attempt to generate a single-cell 3*D* structure by imposing a fixed relationship between the interaction frequency and the average spatial distance between pairs of genomic loci. Another group of these methods adopts a more flexible approach, assuming a range of constraints to reconstruct an ensemble of structures that match experimental Hi-C data. In both cases, a step-wise approach is employed, utilizing various assumptions and optimization methods to reconstruct the 3D chromosome structure. The following section briefly reviews the two most important categories, Distance-Based restraints, and Probability-Based methods.

I. Distance-Based restraints

This group of data-driven methods have two steps: first, translate the interaction frequencies into spatial distances, and second, apply non-linear optimization techniques to get the 3*D* chromosome structure from spatial distances. What distinguishes these methods from each other is the method used to convert experimental data to distances

and the optimization technique used to derive 3D structures from locus distances. The fundamental assumption is that the equilibrium distance between two loci is related to contact probabilities via $d_{ij} \propto c_{ij}^{-\alpha}$, where α is the best scale parameter, conversion factor (Oluwadare, Highsmith, and Cheng, 2019). The commonly used conversion factor is 1, which means that the distances are inversely proportional to the interaction frequencies, $d_{ij} = \frac{1}{c_{ij}}$ (Lieberman-Aiden et al., 2009a), however, there is no fundamental reason to support this assumption. The correlation between interaction frequencies and distances may vary from one Hi-C data to the other and also from one organism to the other. The other important caveat in interpreting Hi-C data is that the contact matrix does not represent the contact of chromosomes in one cell but represents the average contact frequency in the population of cells. The so generated interaction datasets do not contain absolute interaction frequencies. Thus, only some of the Hi-C matrix's contact probabilities may represent the true contact between the fragments in the 3D space. A short review of three distance-based restraint methods is presented in the following.

I.1. Reconstruction of *Caulobacter crescentus* from 5C data

One of the first computational methods to model the $3D$ architecture of the bacterial genome was proposed by Umbarger and colleagues (Umbarger et al., 2011). They applied the same method which has been used in their group for modelling the 3D conformation of the $\alpha-$globin genomic domain in the human chromosome 16 (Alber et al., 2007; Baù et al., 2011). Restriction fragments are represented by points and harmonic interactions are introduced between all pairs of loci (i, j), to reconstruct the equilibrium chromosome conformation from the 5C data, (Fig. 3.4 i, ii),

$$H(\vec{r}_{ij}) = k(\vec{r}_{ij} - d_{ij})^2, \tag{3.1}$$

Where $\vec{r}_{ij}$ is the distance between the two fragments i and j, the spring constant, k, was set at the square of the corresponding normalized contact frequency, and d_{ij} is the equilibrium (expected) distance derived from the contact frequency at which these fragments contacted.

To obtain spatial distances between fragments, they first converted the interaction frequencies into Z-scores (the number of standard deviations by which the value of a raw score is above or below the mean). High Z-scores correspond to high interacting pairs, while low Z-scores correspond to little or no interaction. The Z-scores were then combined with the information from fluorescence microscopy, the measured distances between loci at different genomic separations (Viollier et al., 2004). Eventually, Z-scores were converted to spatial distances using a calibration curve (fifth-order polynomial). Additional restraints were then introduced between pairs with very high or very low

contact frequencies to prevent them from moving too close to or too far away from each other.

Furthermore, they applied another constraint: all of the fragments lie within a box slightly larger than the *Caulobacter* cell size (Fig. 3.4(iii)). Optimization was performed by initializing the randomly 3*D* coordinates of all points, then altering these positions to derive the structures with the lowest objective function scores (sum of all the individual restraints) (Alber et al., 2007; Baù et al., 2011; Umbarger et al., 2011). This initiation and optimization process was repeated thousands of times to create an ensemble of structures. Then the structures were combined and grouped based on their coordinates, producing clusters of configurations with highest structural similarity. In the configurations, circular genome was elongated and twisted ~ 1.5 times around its longest axes (Fig. 3.4(iv)).

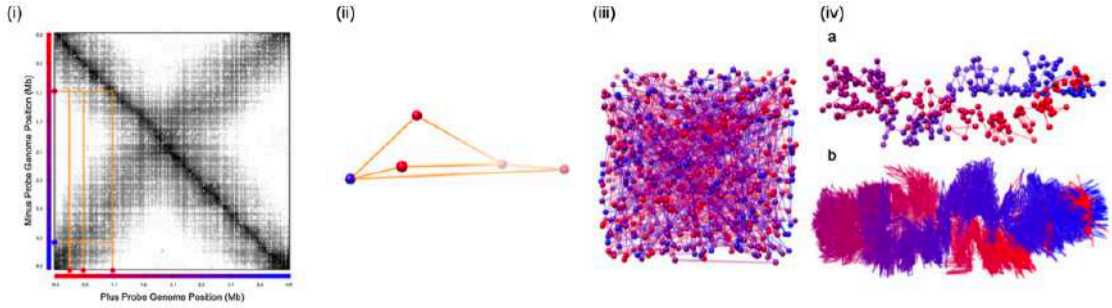


FIGURE 3.4: Distance-Based restraints methods, Harmonic restraints. Restriction fragments were represented as a point connected by springs. (i) Contact frequencies were converted into spatial distances between fragments. (ii) the position of all points were randomly initialized in space, (iii) a structure that minimally violates the equilibrium (averaged) lengths derived via optimization procedure. (iv.a) To generate an ensemble of structures, initialization and optimization procedure was repeated thousands of times. (iv, b) these structures were superimposed and grouped into clusters based on their structural similarity. Figure is copied from (Umbarger et al., 2011)

I.2. Reconstruction of *E.coli* at various replication stages

Another similar approach proposed by Wasim and colleagues, as described in the reference (Wasim, Gupta, and Mondal, 2021), aims to reconstruct the three-dimensional (3D) architecture of *E.coli* using ensemble average Hi-C interactions at different cell cycles. The authors considered three scenarios with different levels of DNA replication:

Non-replicated single chromosome ($G = 1$): In this case, the amount of DNA in the cell is equal to the length of a single non-replicated chromosome, which is approximately 4.6 Mbp (megabase pairs). The value of G represents the ratio of the amount of DNA in the cell to the length of the DNA in a non-replicated chromosome.

Partially replicated single chromosome ($G = 1.8$): Here, the amount of DNA in the cell is 1.8 times the length of a non-replicated chromosome.

Partially replicated pair of chromosomes ($G = 3.6$): In this scenario, the total amount of DNA in the cell is 3.6 times the length of a non-replicated chromosome.

They modeled DNA as bead-on-spring polymers with resolution 5 kbp (with diameter $\sigma \sim 50$ nm), at the same resolution of Hi-C map, subject to sphero-cylindrical confinement with the same size of *E.coli*. Adjacent beads were connected via a very strong harmonic potential with σ as the equilibrium length, and non-adjacent beads interacted via a purely repulsive potential, $\sim \frac{1}{r^{12}}$. Moreover, Hi-C contacts were modeled as harmonic springs via eq. 3.1, where the equilibrium distances were inversely proportional to the interaction frequencies, $d_{ij} = \frac{\sigma}{c_{ij}}$. To consider the fact that the very weak interaction frequencies will yield ∞ distances, they modeled the spring constant as a gaussian,

$$k_{ij} = k_0 e^{-\frac{(D_{ij}-\sigma)}{\omega}} \tag{3.2}$$

k_0 and ω are the parameters that give the best fit to experimental data. They made two assumptions as simplifications in the proposed approach. The first assumption states that there are no Hi-C interactions between two different replication forks and two independent replicated chromosomes. However, in actual biological systems, it is possible for interactions to occur between these regions. Disregarding such interactions may overlook important spatial relationships and potential influences on the 3D organization of the genome. The second assumption assumes that the interactions between two beads in a replication fork are equivalent to those between the backbone beads to which the fork beads have been mapped. This assumption simplifies the analysis by considering only the interactions of the backbone beads. However, it neglects the potential differences in interactions between the beads within the replication fork and the backbone beads. These differences could play a role in shaping the overall 3D structure of the genome.

The energy-minimized conformations were subjected first to long molecular dynamics followed by stochastic dynamics (SD). Then equally interval snapshots from the SD trajectories were extracted and subjected in another SD as the initial conformations to reach equilibrium and obtain an ensemble of conformations. Finally, the bead distances were used to calculate a simulated Hi-C contact matrix, and then it was filtered to compare with the experimental Hi-C matrix.

The scaling behavior of the contact probability in the reconstructed chromosome configurations proposed by Wasim and colleagues deviates from that expected for a fractal globule. They observed that the contact probability scales approximately as, $P(s) \sim s^{-0.55}$ in the range of 10 kb to 1 Mb, whereas a fractal globule is predicted to exhibit a scaling behavior of $\sim s^{-1.1}$. However, the authors did not provide explicit reasons for the deviation

in the scaling behavior of contact probabilities, but they argued that their configurations are knots-free, supporting the notion of a fractal nature for their structures.

I.3. Reconstruction high-resolution 3D models of *Caulobacter crescentus*

Yildirim and Feig proposed a multi-scale modeling protocol to reconstruct the three-dimensional (3D) architecture of *Caulobacter crescentus*, as depicted in Figure 3.5, in their study (Yildirim and Feig, 2018). Their approach involved modeling the bacterial DNA at the plectonemic level, representing it as a hyper-branched polymer chain.

In their model, the DNA was represented by central ring beads connected by branch beads, with each ring bead and its associated branch beads representing a micro-domain. The central ring segments represented double-stranded DNA, which connected each micro-domain, while the segments between the branch beads represented supercoiled DNA. The topology of the branches was fixed, but they allowed for movement and re-connection of the branches using a Hamiltonian-based approach.

To impose constraints on the system, the authors applied harmonic potentials to restrict the angles, bonds, and degree of branching of all the segments. These potentials were chosen based on reasonable discussions to ensure biophysical plausibility. Additionally, the polymer chains were confined within a small sphere to compact the conformations.

To guide their Monte Carlo simulations and energy minimization procedure, the authors utilized distance-based restraints derived from Hi-C interaction frequencies. These restraints were obtained using a procedure similar to the one introduced by Umbarger and colleagues (Umbarger et al., 2011). However, you mentioned that there are no clear reasons supporting Umbarger's idea.

After the initial modeling of the plectonemic segments, Yildirim and Feig proceeded to wrap higher resolution beads around each of these segments using Langevin dynamics simulations. This allowed them to model a higher resolution coarse-grained representation of the chromosomal DNA, with each bead representing a segment of 15 base pairs (bp). At this higher resolution level, the total energy of the chain included contributions from stretching, bending, torsional, and electrostatic interactions. These energy terms were taken into account to capture the relevant physical forces acting on the DNA chain.

Rather than applying rigid forces to enforce the distance restraints, the authors utilized weak forces. This approach was based on the assumption that while the overall topology of chromosome structures might be similar across different cells, the local structural organization can vary. By applying weak forces, they aimed to generate an ensemble of structures that captured the variability in local structural arrangements. To achieve this, the authors reweighted the configurations obtained from their simulations. They aimed to create an ensemble of structures that not only had an average contact distribution matching the Hi-C scores but also maximally matched the distribution of contacts across the configurations.

By combining the coarse-grained modeling, weak forces for distance restraints, and reweighting of configurations, Yildirim and Feig sought to create a model that represented the chromosomal DNA's 3D architecture in *Caulobacter crescentus* while incorporating the experimental information from Hi-C data.

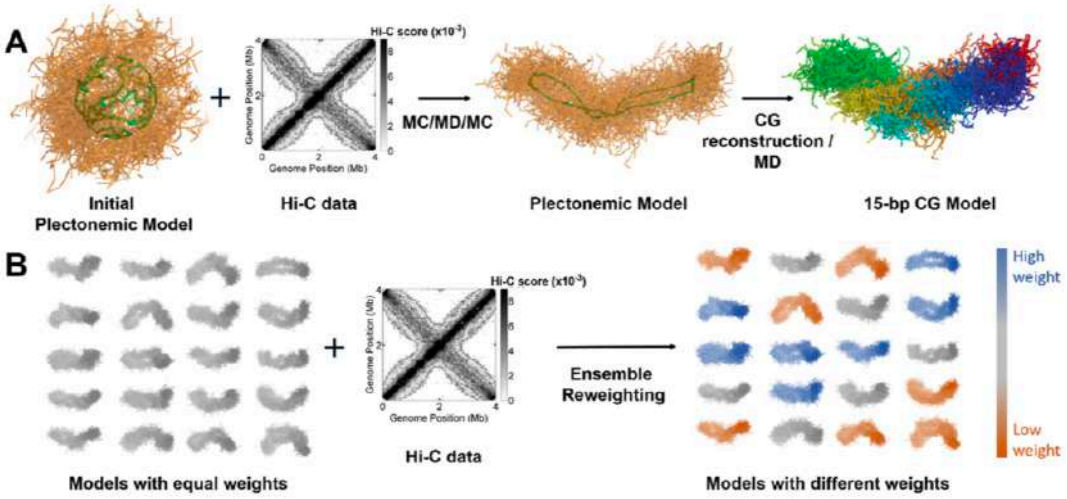


FIGURE 3.5: Multi-scale modeling procedure based on Hi-C data. (A) plectonemic and CG configurations. (B) reweighting the configurations. Figure is copied from (Yildirim and Feig, 2018)

II. Probability-Based Methods

These methods uses a probabilistic approach to model 3*D* chromosome structure. Interaction frequencies can be considered as average since the Hi-C data are derived from cell populations. Hence, assuming an ensemble of structures underlies a contact map is common. These methods do not assume Hi-C score-distance relation. One of this method is **Maximum Entropy Method (MaxEnt)**.

The central assumption in MaxEnt methods is that the experimental Hi-C maps have sufficient information to drive the distribution of chromosome conformations. These methods do not assume Hi-C score-distance relation. "This approach infers the least-structured distribution of chromosome conformations that fits Hi–C experiments, capturing population heterogeneity at the single-cell level". Specifically, the approach is based on seeking the statistical weights $P(s)$ with the largest Shannon entropy under experimental constraints,

$$S = -\sum_{\sigma} P(\sigma) ln\, P(\sigma), \tag{3.3}$$

where σ is the set of all chromosome configurations within the cellular confinement (Messelink et al., 2021).

In order to apply the MaxEnt method to experimental Hi–C data, they employ polymer representation of the chromosome under two constraints: First, the contact frequency between any two genomic regions i and j from the model and experiment should match. Secondly, the distribution should be normalized. The first one introduced Lagrange multiplier λ_{ij} for each experimental constraint and λ_0 to ensure normalization. Usually, for Hi-C data on a bacterial chromosome, there are 10^5 experimental constraints (parameters) to fit. Then apply the inverse Monte Carlo algorithm iteratively to find all the Lagrange multipliers. MaxEnt models can reproduce essential features of the experimental Hi–C map, including the fine structure of the CIDs and the secondary diagonal. However, the model can not truly capture the organizational features of the DNA and the physical laws behind the $3D$ architecture of DNA.

3.3.3 Challenges

One significant difficulty in developing spatial models of chromosomes is that Hi-C maps generally represent the population-average map of contact probabilities and do not represent information from a single cell in vivo conformations. Thus, Hi-C technologies mainly provide a qualitative picture. The so generated interaction datasets do not contain absolute interaction frequencies. A noticeable feature of Hi-C maps is that almost all genomic locus may be found in contact with each other; as a result, there are almost no regions of zero contact probability in the map (Imakaev, Fudenberg, and Mirny, 2015).

3.4 Bottom-Up approaches

As discussed in Section 3.3.1, bottom-up models employ a minimal set of physical assumptions and parameters to develop polymer models of DNA. These models can be categorized based on the level of complexity considered in chromosome folding and its interactions with proteins.

In the case where the complexity of chromosome folding and interactions with all proteins are taken into account, these models can study bacterial chromosomes at a subgenomic length scale with high resolution. By incorporating detailed information about protein binding and sequence effects, these models aim to capture the intricacies of chromosomal organization and its functional implications. On the other hand, if the microscopical details such as protein binding and sequence effects are coarse-grained, these models have the ability to simulate bacterial chromosomes at the full genomic length but with lower resolution. By simplifying these details, these studies can explore larger genomic scales and provide insights into the overall behavior of bacterial chromosomes. Some notable examples of the bottom-up models of bacterial chromosomes are briefly reviewed in the following.

3.4.1 Polymer models of chromosome organization

In the study of polymer chains, the radius of gyration and end-to-end distance are two important measures that characterize the chain's conformation. The scaling behavior of these measures ,$R \sim N^{\nu}$, with respect to the chain length (number of monomers) can provide insights into the overall architecture of the polymer chain. In the literature, two types of globules, equilibrium globules and fractal globules, have been proposed to explain the architecture of chromosomes.

Equilibrium globule

When a polymer chain experiences confinement within a small volume or exhibits stronger self-attraction compared to its surrounding solution, it undergoes a coil-globule transition. This transition leads to the formation of an equilibrium globule, which is a "disordered dense packing of coil" that minimizes contacts with its surroundings. In the equilibrium globule, the behavior of sub-chains can be described by a random walk-like scaling relationship for the end-to-end distance.

For sub-chains within the equilibrium globule, the scaling of the end-to-end distance ($R(s)$) can be approximated as:

$$R(s) \sim \begin{cases} s^{1/2} & s \leq N^{2/3} \\ constant & otherwise \end{cases} \tag{3.4}$$

This means that for sub-chains with a size (s) smaller than or equal to $N^{2/3}$, where N is the total chain length, the end-to-end distance scales as $s^{1/2}$, resembling a random walk behavior. However, for larger sub-chains, the end-to-end distance remains relatively constant. Similarly, the contact probability of sub-chains within the equilibrium globule scales as:

$$p_c(s) \sim \begin{cases} s^{-3/2} & s \leq N^{2/3} \\ constant & otherwise \end{cases} \tag{3.5}$$

Several theoretical and simulation studies (Mirny, 2011; Imakaev et al., 2015; Imakaev, Fudenberg, and Mirny, 2015), have raised questions about its relevance as a model for DNA packing. It has been shown that the equilibrium globule state is highly entangled and abundant with polymer knots, which may not accurately reflect the behavior of chromosomal DNA. Additionally, the process of forming an equilibrium globule is time-consuming, with an equilibrium time scaling as N^3, making it less likely to be a biologically relevant model for DNA packing in cells.

Fractal (or Crumpled) globule

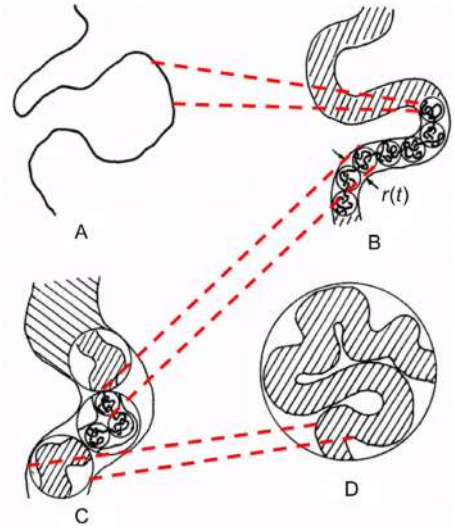


FIGURE 3.6: The crumpled globule model. The formation of crumples happens at all length scales. The construction of crumpled globules can occur by fast switching the monomers' interactions from repulsion to attraction (e.g., changing solvent quality). This picture can also be viewed as a chain of blobs made of smaller blobs and so on. Figure is copied from (Grosberg, Nechaev, and Shakhnovich, 1988; Rosa and Zimmer, 2014)

The fractal globule is a proposed globular form of polymer structure that was introduced by Grosberg in 1988. In their work, Grosberg and his colleagues (Grosberg, Nechaev, and Shakhnovich, 1988; Grosberg et al., 1993) argued that chromosomal DNA likely adopts a long-lived, unentangled, unknotted, and out-of-equilibrium state known as a crumpled globule, which arises due to topological constraints. These constraints play a crucial role in shaping the organization of the polymer.

The construction of crumpled globules can occur by fast switching of monomers' interactions from repulsion to attraction.. Furthermore, polymer crumples form in all length scales since this collapse is due to self-attraction and topological constraints. In other words, the polymer can be viewed as a recursive coiling of mass that appears like a collapsed globule at any length scale (larger than the entanglement length N_e) within the globule. (Fig. 3.6). Importantly, pair segments belonging to different crumples exhibit minimal interaction with each other, maintaining a relatively unentangled and unknotted configuration.

The Rouse relaxation time for this system is $\tau_R \sim L^2/D_R$, where L is the chromosome length, and D_R is the diffusion coefficient, based on the Stokes-Einstein relation:

$$D_R = \frac{k_B T}{(6\pi r \eta)}. \tag{3.6}$$

where k_B is Boltzmann's constant, T is the temperature (approximately 293 K for room temperature), r is the segment radius (approximately 1 nm), and η is the viscosity of water (approximately 10^{-3} Pa s). By substituting $L = 30$ mm, The estimated collapse time is approximately 100 days.This duration is much longer than the cell cycle, implying that within the timeframe of a cell cycle, the crumpled globule picture of DNA does not have sufficient time to develop knots.

Years later, Rosa and Everaers (Rosa and Everaers, 2008) provided computational evidence that supported the "topological origin" idea for the observed structure and dynamics of eukaryotic chromosomes. In their study, they modeled interphase chromosomes as tightly compacted and well-separated polymers. They estimated that achieving "proper mixing" within chromosomes would take hundreds of years due to the presence of topological constraints (Rosa and Everaers, 2008).

They argued that the physics underlying the organization of chromosome territories, where long chromosomes occupy distinct regions, is analogous to the segregation of unentangled ring polymers in a melted state. In this scenario, unknotted and non-catenated ring polymers form distinct territories and cannot interpenetrate each other due to their construction. Similarly, in very long polymers like chromosomes, topological constraints play the same role. Importantly, Rosa and Everaers' model was able to reproduce the $3D$ structure and dynamics observed in experimental chromosome data (Fig. 3.7). This computational validation further supports the notion that the topological constraints imposed on chromosomes contribute to their organization and behavior within the nucleus.

In a nutshell, the original theory of the fractal globule suggests the following: (i) A chain without ends, such as a ring, remains in the fractal globule state indefinitely. This implies that the chain does not exhibit entanglement or knotting.

(ii) The lifetime of a fractal globule is determined by the time it takes for the polymer ends to thread through the entire globule, enabling the formation of knots. The threading process takes approximately N^3 time, where N is the length of the polymer.

In support of the fractal globule model, Lieberman-Aiden et al. (Lieberman-Aiden et al., 2009a) argued that the scaling law observed in Hi-C experiments for human chromosomes in the genomic range of 0.5 to 7 Mbp are consistent with the predictions of the fractal globule model, with a scaling exponent of approximately s^{-1}.

In addition, Leonid Mirny (Mirny, 2011) performed simulations to confront the two different types of globules. The results are shown in Figure 3.8, highlighting how different the two states (equilibrium and fractal) look in terms of knotting and separation of regions of the polymer. The end-to-end vector of the fractal (crumpled) globule subchain scales with the length of the subchain as:

$$R(s) \sim s^{1/3} \tag{3.7}$$

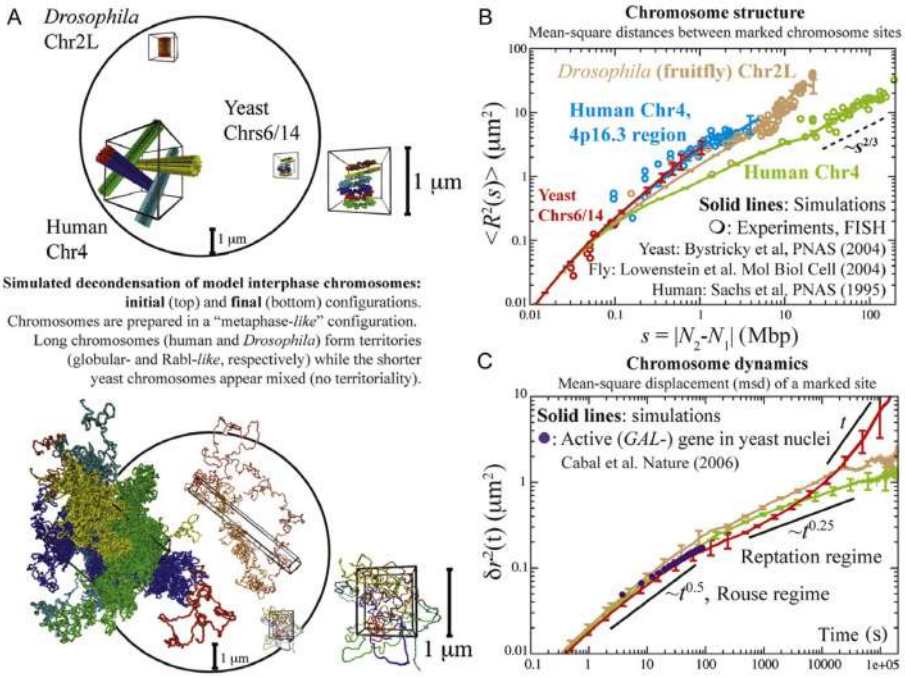


FIGURE 3.7: Topological model of the eukaryotic chromosomes. (A) Initial and final configurations of three different chromosomes: human Chr4, Drosophila Chr2L, and yeast Chr6. The bead-spring model is used to model each chromosome (each bead represents 3kb). (B)(Average square spatial distances) and (C) (mean square displacement) proves that their model could reproduce the 3*D* structure and dynamics of experimental chromosome data. Figure is copied from (Rosa and Everaers, 2008; Rosa and Zimmer, 2014)

and the contact probability of the subchain scales as:

$$P_c(s) \sim s^{-1} \tag{3.8}$$

Model bacterial DNA as a charged polymer

Another model which pays special attention to the role of physical interactions in the compaction of the genomic DNA of a prokaryotic cell is to consider DNA as a charged polymer based on the fact that DNA is highly charged with linear charge $-2e$ per base pairs. On the other hand, cytoplasm behaves like a polyelectrolyte since it contains K^+ or Na^+ cations as counter-ions, also a large number of divalent metal ions like Mg^{2+} or Ca^{+2}. In the long charged polymers, part of the counter-ions is trapped in the volume of the macromolecular coil. In contrast, the other part escapes to the remainder of the volume, so DNA retains a non-zero charge even in the salt solution. This DNA-DNA electrostatic repulsion tends to oppose DNA compaction.

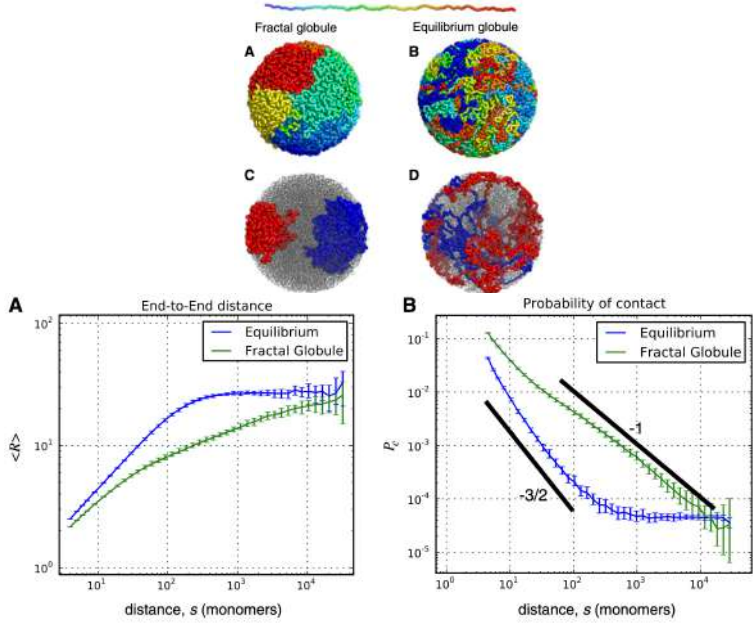


FIGURE 3.8: Top: simulated examples of conformation of fractal (A,C) and equilibrium globules (B,D). The fractal globule show territorial organization. Bottom: (A) end-to-end distance, and (B) the probability of contact as a function of genomic distance s, for equilibrium globule (blue) and fractal globule (green). Figure is copied from (Mirny, 2011)

To overcome this repulsion, the author (Joyeux, 2015) suggested that since the solution contains macromolecules with like-charges, two proteins or a protein and a DNA repulsion are stronger than two DNA sites. Therefore, compacting the DNA in a separate cell region is energetically favourable, leading to the segregative phase separation (Zimmerman and Murphy, 1996; Shin, Cherstvy, and Metzler, 2014; Joyeux, 2015).

To test this idea, the author (Joyeux, 2015) modelled DNA as a ring polymer with 1440 beads representing 21600 bp (each bead represents 15 DNA bp), with bending, stretching, and electrostatic energy enclosed in a sphere with a radius of 120 nm at a concentration close to the physiological ones. Then proteins are induced by 3000 additional beads that interact with other proteins and DNA through repulsive electrostatic interactions. The dynamics of the system composed of the DNA chain and the protein beads are investigated by integrating overdamped Langevin equations. Figure3.9 shows that the repulsion between all the beads leads to substantial compaction of DNA, despite the repulsion between neighbouring segments and the sterical hindrance of protein bead.

3.4.2 Other polymer models

Bottlebrush Model

One of the polymer models to reconstruct the chromosome architecture from Hi-C data was a semi-flexible polymer containing a central backbone and an array of plectonemes

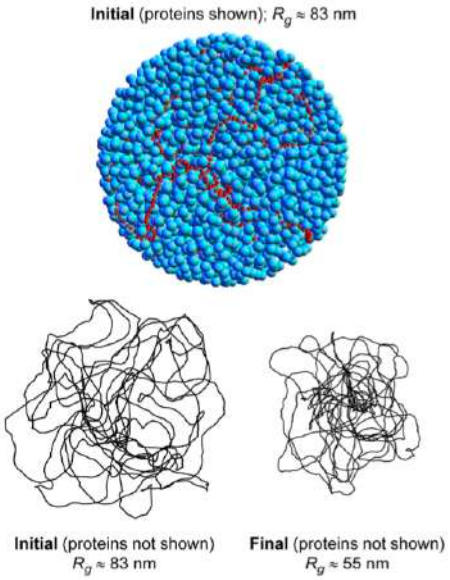


FIGURE 3.9: Top: Snapshot of a conformation of the equilibrated DNA chain enclosed in the confining sphere in the presence of the negatively charged protein beads. Bottom: the same snapshots with the proteins have been removed before and after equilibrium. Figure is copied from (Joyeux, 2015)

attached to it, resembling a bottlebrush (Fig.3.10) (Le et al., 2013). Plectoneme Free Regions (PFRs) were introduced at specific regions on the backbone, as shown in Fig.3.10(C). They are located at the location of the highly expressed genes to the emergence of CIDs based on the assumption that highly expressed genes unwind the DNA duplex and create the PFRs (i.e., PFRs prevent the diffusion of supercoils and separate CIDs). The polymer was confined to the cell volume. Then, an ensemble of equilibrated polymers was obtained by applying Brownian Dynamics simulations. In total, five fitting parameters were applied to characterize the structure of the chromosome: plectoneme length, gap length between adjacent plectonemes, the diameter of plectonemes, the flexibility of plectonemes, and contact radius (Fig.3.10). The goal was to select values of all five parameters to best fit the scaling of the contact probability $P(s)$ observed by Hi-C data and to reproduce other features of the Hi-C data, like interactions between chromosomal arms and the presence of CIDs. The configurations with almost ~ 300 plectonemes with length ~ 15 kb and separated by less than 300 bp gave the best agreement with wild-type experimental Hi-C data.

To examine the role of gene expression, They performed the simulations by removing PFRs, which produced domain-free contact maps similar to swarmer cells treated with rifampicin—the antibiotic effect as an inhibitor of the transcription elongation (Le et al., 2013).

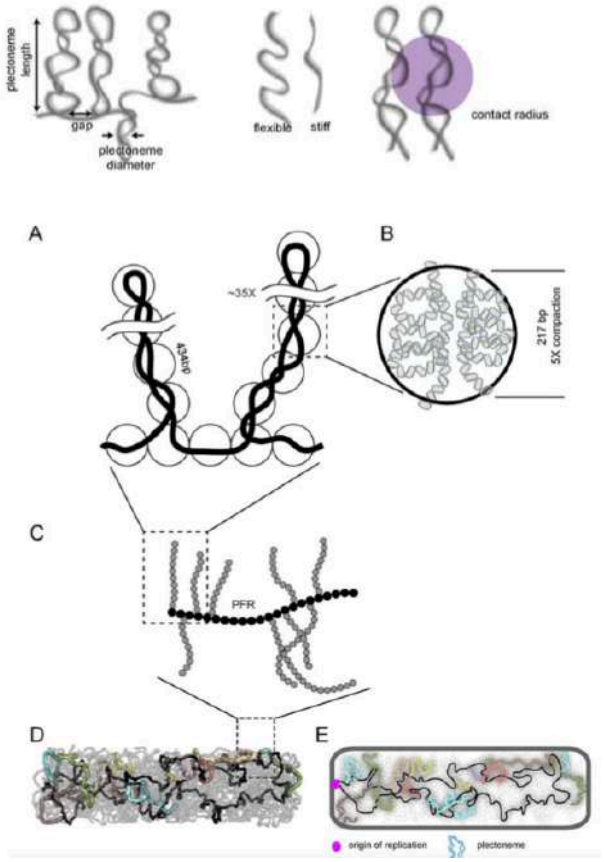


FIGURE 3.10: Schematic of the Bottlebrush model of the Bacterial DNA. Five parameters describing the polymer model. Figure is copied from (Le et al., 2013)

Lattice Models of Bacterial Nucleoids

This model is an idealized model of bacterial nucleoids developed by Goodsell and colleagues (Goodsell, Autin, and Olson, 2018). In their model, a single DNA is divided into equal-length unbranched hairpin loops representing plectonemes, connected by equal-length segments. The central connecting segments make a ring on the lattice, and the plectonemes are generated by adding a random walk rooted at specific locations on the central ring, representing the helix axis of the plectonemes (Fig.3.11). To form the local DNA double-helical strands, the plectoneme's lattice points are divided into two points and are rotated to generate the desired supercoiling density. The DNA chains are confined in an ideal spherical or capsular shape depending on the bacterial cell, i.e., sphere for *Mycoplasma*, and capsule-shaped volume for *E.coli* nucleoid. Finally, lattice configurations are relaxed to produce the non-lattice coarse-grain configurations with 10 bp bead resolution (Fig.3.11). The length of the plectonemes, plectoneme's separations, and DNA persistence length are chosen to give the best agreement with experimental data. The model is idealized by considering the plectonemes as equal lengths of unbranched

superhelical hairpins spaced equally around the circular genome. Secondly, the equilibrium of the configurations is under suspicion.

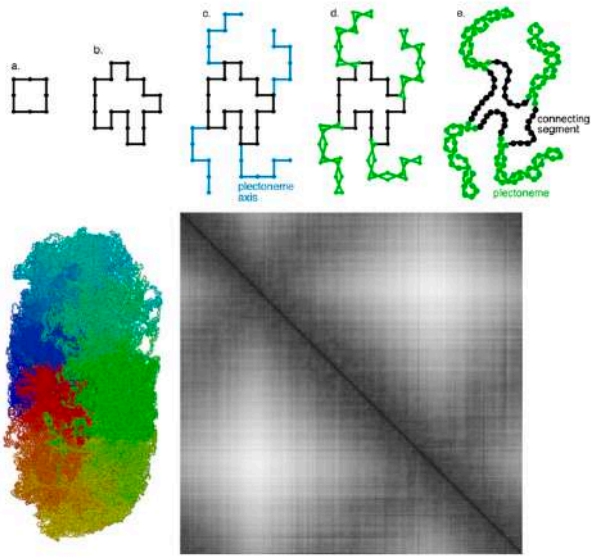


FIGURE 3.11: Schematic of lattice model of Bacterial Nucleoids.Top: summary of the method on how to grow the chain. Bottom, right: One configuration of *E.coli* nucleoid, left: Distance map at 10 kbp resolution. Figure is copied from (Goodsell, Autin, and Olson, 2018)

3.4.3 Role of topological constraint, confinement, molecular crowders in the organization of bacterial DNA

Another polymer approach based on confinement and topological constraints could provide insights into understanding bacterial chromosome folding and genome architecture. In the following, several models based on topological constraints of bacterial DNA are reviewed. These articles mainly did not provide any direct comparison with experimental data but investigated some DNA properties from a polymer physics point of view.

Crowding effects on chromosome organization

The simulation of ring polymers with uncharged molecular crowders in cylindrical confinement shows that the interaction between the polymer and the molecular crowds causes a decrease in the radius of gyration R_g of the ring polymer as the density of the mutually repulsive crowding molecules (modeled by repulsive Weeks-Chandler potential) is increased (Shin, Cherstvy, and Metzler, 2014). In particular, it was shown that the component of R_g along the longitudinal direction of the cylinder decreases $\approx 8\%$ when the volume fraction ϕ of the crowders is increased from $\phi = 0$ to $\phi = 0.3$, while the component of R_g along the radial axis of the cylinder remains largely unchanged.

These investigations and others involving various crowders proved that the crowding environment causes the DNA molecule to condense in the nucleoid region inside the cell (McGuffee and Elcock, 2010). These studies (Jung et al., 2012; Jeon, Jung, and Ha, 2017) give essential insights into the role of molecular crowders in the organization of the DNA polymer but completely neglect the role of either DNA-binding proteins or chain crossing.

Entropy-driven spatial organization of highly confined polymers

The authors investigated the role of entropy as the driving force for chromosome segregation (Jun and Mulder, 2006; Jun and Wright, 2010). They explored several aspects of the spatial organization and segregation of strongly confined entropic polymers using computer simulations and scaling arguments.

To study the role of chain topology in the tendency of demixing of chains, they put two chains with three different topologies (linear, ring, and randomly branched polymer) in the same rod-shaped confinement and the same volume fraction ($\approx 10\%$). By applying Monte Carlo simulations, they showed that increasing the topological complexity of the chains corresponds to stronger repulsion between them so that the branched polymers have a higher tendency to demix more than the other two cases (branched polymer $>$ ring polymer $>$ linear chain).

Then they proposed a simple coarse-grained polymer model of a bacterial chromosome during replications in rod-shaped confinement mimicking the nucleoid. By applying Monte Carlo simulations, they could capture the main features of the experimental observation of the reorganization of DNA during replication (Jun and Mulder, 2006). At the early stage, two replicated daughters are highly entangled and strongly confined within the cell. From a polymer physics point of view, overlapping chains have fewer conformational degrees of freedom, or less conformational entropy, than the ones that are completely separated. As a result, the combination of entanglement and confinement leads a strong entropic force to the chromosome, repels them, and ends in the segregation of the chromosomes.

To investigate the effect of the shape of the confinement on the tendency of demixing, they confined two chains in a three-dimensional box of a given size and shape. They kept their volume constant (same confinement free energy) (Jun and Wright, 2010). They summarised their results in a phase diagram to show the segregatability of two confined chains. They observed that the longer boxes have more tendency to demix and the stronger the linear ordering of the confined chain (Fig.3.12).

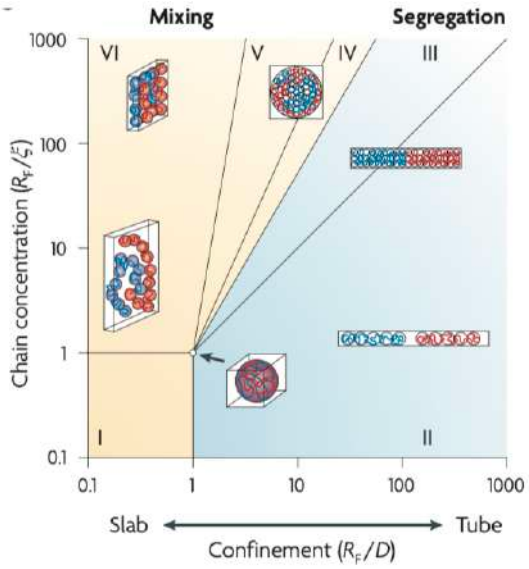


FIGURE 3.12: A phase diagram explaining 'segregatability' of two confined chains in a three dimensional box. It shows how two long chains will segregate or mix according on their degree of confinement and their concentration within a box. Regime I describes polymers in dilute solution without any confinement. Figure is copied from (Jun and Wright, 2010)

3.5 Summary

In this chapter, two major experimental methods probing the 3*D* organization of the genome are presented and confronted with each other.

The chapter has also far presented a short review of some quantitative models that investigated the 3*D* architecture of bacterial chromosomes. We also discussed the main motivations and limitations (optimization of multi-parameters in the bottom-up approaches or taking all the contact map data in data-driven models) of these models.

Data-driven methods can reproduce essential features of the experimental Hi–C maps by taking into account rich experimental data; however, biologically interpreting the parameters is not relatively easy. Moreover, tuning many different sets of parameters can give the best agreement with experimental contact maps; however, regularizing these approaches has been frequently called into question. On the other hand, bottom-up approaches despite their simplification can account for quantitive properties such as scaling laws of average contact frequencies with genomic distance. This group of models investigates whether certain known or hypothesized physical or biological principles can lead to understanding the experimental data and try to drive universal behavior by applying physics laws. As a result, these models use a minimum set of assumptions and parameters and have strong predictive power. Although the line between the bottom-up approach and data-driven methods is not as sharp as black and white, some data-driven models can be considered extensions of polymer models.

Chapter 4

Theoretical Background

4.1 Introduction

From the viewpoint of polymer physics, a plectonemic conformation of a circular DNA corresponds to the situation of the *double-folding* of a ring chain. In order to appreciate the large-scale behavior of supercoiled DNA, I thus first remind the differences in the equilibrium properties between ring and linear polymers, more specifically in the melt case. Namely, the equilibrium statistics and dynamics of ring polymers are fundamentally different from those of their linear counterparts (In our article, we have also investigated the dynamics of double-folded rings vs. linear chains (Ghobadpour et al., 2021)). To predict the equilibrium properties of a ring polymer in a melt situation (e.g., in the presence of many other rings), a convenient approach consists in viewing the ring as moving among an array of fixed obstacles (Khokhlov and Nechaev, 1985; Rubinstein, 1986; Grosberg, 2014; Smrek and Grosberg, 2015) (Fig.4.1(c)). In this picture, the ring adopts three strategies to maximize its entropy: double-folding, branching, and swelling. Firstly, and most importantly, the ring adopts a double-folded conformation to reduce the importance of the topological constraints imposes on each other since double-folding minimizes the threadable surface. Secondly, double-folded rings make branches so that to increase their entropy. Thirdly, they swell due to partially screened excluded volume interactions. This swelling leads to asymptotically compact conformations characterized by an exponent $\nu = 1/d$ in $d \leq 4$ dimensions. These strategies are exactly those at play in supercoiled DNA (Marko and Siggia, 1995a; Dorman, 2006; Mondal et al., 2011; Lepage, Képès, and Junier, 2015; Lepage and Junier, 2019) (Fig.4.1(b)) – note also that viral RNA behave like double-folded branched polymers (Kelly, Grosberg, and Bruinsma, 2016) (Fig.4.1(a)). Moreover, as explained in Sec.3.4.1, a long, unknotted, and non-concatenated ring polymer in the melt state, hence also adopting these strategies, appears to provide a natural explanation for the territorial behavior of eukaryotic chromosomes during interphase (Gutin, Grosberg, and Shakhnovich, 1993; Rosa and Everaers, 2008; Lieberman-Aiden et al., 2009b). As a result, understanding the behavior of self-avoiding randomly branched polymers, or trees should provide us with a general understanding of the principles governing the cellular organization of DNA, from bacteria to eukaryotes.

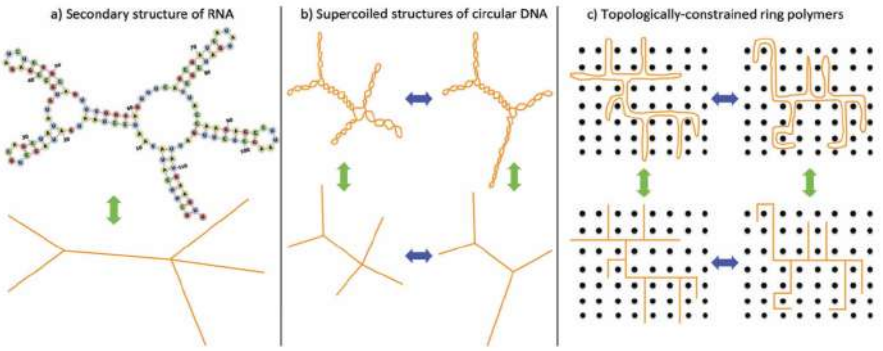


FIGURE 4.1: (a) A viral RNA has a branched secondary structure. (b) Bacterial negatively supercoiled DNA has a plectonemic structure that can be mapped on the branched trees. (c) A ring polymer in an array of obstacles resembles a branched polymer. This figure is copied from (Everaers et al., 2017).

The known results about these aspects are reviewed in this chapter according to the following structure. First, linear polymer chains are reviewed in Sec.4.2. Then in Sec.4.3, relevant observables and the related exponents for randomly branched polymers are introduced. Also, the theoretical background is briefly summarized at the end.

4.2 Linear polymer chains

4.2.1 Ideal chain

An ideal chain (i.e., a random phantom walk) is the simplest model to describe the properties of a flexible polymer chain. In this model, the motion of neighboring monomers along the chain is not correlated (except by the chain connectivity), and monomers do not physically interact so that the polymer chain behaves like a random walk in $3D$ space. Therefore, this model is reasonable for studying a fully-flexible chain where at a first approximation, interactions between monomers are screened. However, it is unrealistic in most cases since monomers allow to overlap in space. The configuration of such an ideal chain is described by the simplest polymer model, a freely joint chain (FJC). The polymer is approximated as a series of N monomers, with each bond vector independent of the rest.

In general, polymer size is described as the root-mean-square end-to-end distance $\langle R^2(N)\rangle^{1/2}$ where R is the distance from one end of the polymer chain to the other in $3D$ space. $\langle R^2(N)\rangle^{1/2}$ is the square root of the average of this distance squared. Squaring the distance eliminates any negative values so that $\langle R^2(N)\rangle^{1/2} \geq 0$. Averaging can be taken over a single chain over time as it explores conformational space or over many such chains simultaneously. In this manner, the average account for the fluctuations observed for

polymer chains with a single parameter. Therefore, the average size of the chain in the FJC model is given by

$$\langle R^2(N)\rangle^{1/2} = l_k {N_k}^{1/2} \tag{4.1}$$

N is the number of Kuhn segments, l_k is the Kuhn length, which is the bond length and measures the chain's bending rigidity. It is a measure of the segment length over which the chain's direction is relatively independent. Hence, the total length is $L = l_k\, N_k - l_k$ thus represents the minimum length between two points on a polymer that are uncorrelated. Expressed as a function of the chain's contour lengths L, will end to:

$$\langle R^2(L)\rangle^{1/2} = \sqrt{l_k L}. \tag{4.2}$$

It is also common to characterize the average chain size by the radius of gyration where the square radius of gyration of a polymer conformation is defined as the average square distance between monomers ($\vec{r}_i$) and the center of mass ($\vec{r}_c$):

$$R_g^2 = \frac{1}{N}\sum_{i=1}^{N}(\vec{r}_i - \vec{r}_c)^2, \tag{4.3}$$

By substituting the definition of the position vector of the center of mass, one can rewrite the following:

$$R_g^2 = \frac{1}{2N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\vec{r}_i - \vec{r}_j)^2. \tag{4.4}$$

Simple calculation for the ideal chain will get:

$$\langle R_g^2(L)\rangle = \frac{l_k\, L}{6} = \frac{1}{6}\langle R^2(L)\rangle. \tag{4.5}$$

The probability of contact between two monomers (loci) with (genomic) distance s for the ideal chain is (Appendix 4.5):

$$P(s) \sim s^{-3/2}. \tag{4.6}$$

For example, a naive representation of the *E. coli* genome as an ideal chain, using $l_k =$ 300 bp (the Kuhn length is equal to two times the persistence length (150 bp for *E. coli*)

of semi-flexible chains) would lead to the following size:

$$\langle R^2(N)\rangle^{1/2} = \sqrt{l_k\, L} = \sqrt{300 \cdot 4.6 \cdot 10^6} bp \approx 12\mu m, \tag{4.7}$$

This is much greater than the size of the bacterium cell ($\approx 3\mu m$ for the long axis).

4.2.2 Real chains

In ideal chains, monomers are assumed to be point particles, and the interactions between them are entirely ignored. However, these interactions can significantly affect the conformations and dynamics of the chain. In reality, monomers occupy non-zero volume; therefore, excluded volume interactions exist. The interaction between pairs of monomers depends on the favorability of the direct interaction between monomer–monomer or between the monomer and other surrounding solvent molecules. When the interactions between the monomer and solvent are highly favorable, the monomers will stay away from one another, and the chain will swell (good solvent). Monomers will prefer to be close to one other in the presence of highly unfavorable solvent-monomer interactions, and the chain will collapse (poor solvent). Interestingly, the polymer adopts ideal chain conformations if the interaction between the monomer and solvent is "just right" (De Gennes, 1979). This occurs when there is no effective energetic difference between monomer–monomer, and monomer–solvent interactions (θ solvent). The Flory theory of polymer conformations and the scaling behavior of the real chains in good solvents will be investigated in the following section.

4.2.3 Flory theory of real polymer chains

Paul J. Flory (Nobel Prize Chemistry, 1974) could successfully address the problem of the conformations of real polymer chains, now known as the Flory theory of polymer chains. Flory used a mean-field approach to treat the question of equilibrium conformation of real chains to avoid the need to count every pairwise energetic contribution.

For a polymer in a good solvent, the overall spatial size of a chain reflects a balance between excluded volume, which tends to expand the chain size, and a restoring force due to loss of conformational entropy due to swelling (Flory, 1953).

$$\mathscr{F} = \mathscr{F}_{el}(N,R) + \mathscr{F}_{inter}(N,R)\ , \tag{4.8}$$

In this view, the entropic contribution is the energy for stretching a polymer of linear contour length L at its ends:

$$\frac{\mathscr{F}_{el}}{\beta} \sim \frac{R^2}{l_K L}\ , \tag{4.9}$$

For linear chains in a good solvent, the excluded volume interaction represents the two-body repulsion between polymer segments, which dominates in good solvent:

$$\frac{\mathscr{F}_{inter}(N,R)}{\beta} \sim v_2 \frac{N^2}{R^3} \ , \tag{4.10}$$

Where v_2 is the two-body interaction coefficient, $v > 0$ corresponds to excluded-volume repulsion, $v < 0$ corresponds to attraction, and $v = 0$ corresponds to balancing the effect of the hardcore repulsion and the attraction between monomers, so the chain adopt an ideal conformation, (θ-solvent).

The optimum end-to-end distance can be determined from the minimum of free energy with respect to R:

$$\partial_R \mathscr{F}(N,R) = \beta(-3v_2 \frac{N^2}{R^4} + 2\frac{R}{Nl_k^2}) = 0 \tag{4.11}$$

whose solution reads:

$$R^5 = \frac{3v_2 l_k^2 N^3}{2} \tag{4.12}$$

so that:

$$R \sim N^{3/5} \tag{4.13}$$

According to Flory, a real polymer chain in a good solvent is expected to scale as $v = 0.6$. More sophisticated computations based on perturbation, renormalization group theory, and numerical simulations (de Gennes, 1972; Madras and Slade, 1996; Le Guillou and Zinn-Justin, 1980) led to a value of $v \cong 0.588$, which indicates that the Flory theory, although simple, provides a reasonable description of the real chain behavior.

Suppose the real polymer chain picture is applied to *E.coli* bacteria, $R \sim N^{3/5}$, the chain size will be even larger than the ideal chain picture, $R \sim N^{1/2}$ ($\approx 100\mu m$ vs. $\approx 12\mu m$). Hence it is unlikely that this is the polymer model to explain the configuration of bacterial DNA.

4.3 Randomly branching trees

A polymer chain with a branch structure free of loops is called a tree. An additional metric that characterized the embedded graphs is introduced by a ring polymer's (tight) wrapping of a tree (Rosa and Everaers, 2019). In analogy to protein or RNA structures, such conformations can be discussed in terms of primary, secondary, and tertiary structures (Rosa and Everaers, 2019). The primary structure is simply defined through the connectivity of the ring monomers. The secondary structure arises from the double folding and can be specified through the mapping of the ring onto a graph with the connectivity of the primitive tree. The tertiary structure describes embedding the rings and trees into the three-dimensional space. In the following, the corresponding observables (sections 4.3.1 – 4.3.3) that are used to study the statistics of randomly branching trees

are reviewed. Next, a brief summary of scaling arguments for interacting tree systems is presented in section 4.3.4 (A part of this review is copied from our published paper (Ghobadpour et al., 2021)). Finally, the probability distributions of the observables are discussed in Sec.4.7 and 4.6

4.3.1 Secondary (tree) structure

Two standard measures of the tree connectivity are the mean contour distance, $\langle L \rangle$, between tree nodes, and the average weight of branches, $\langle N_{br} \rangle$.

For one tree conformation, L is defined as:

$$L(N) = \frac{1}{N} \sum_{i=1}^{N-1} \sum_{j>i}^{N} L(i,j) \tag{4.14}$$

Where the tree contour distance $L(i,j)$ is the length of the linear path on the tree connecting i and j. $L(i,j)$ only depends on the tree connectivity and is completely independent of the spatial embedding of the tree. The mean behavior of the tree contour distance can be written as (Everaers et al., 2017)

$$\langle L(N) \rangle \equiv \sum_{l=0}^{N} l p_N(l), \tag{4.15}$$

where $p_N(l)$ is the probability of finding two nodes at a particular tree contour distance l, with

$$p_N(l=1) = \frac{N}{N(N+1)/2} = \frac{2}{N+1}. \tag{4.16}$$

The ensemble average of the length of paths on the tree depends on the weight, N, through the power law relation

$$\langle L(N) \rangle \quad \sim \quad N^{\rho}. \tag{4.17}$$

Alternatively, a tree can be viewed as an ensemble of sub-trees which generated by cutting the tree at a randomly chosen segment. Removing the segment divides the tree into two tree-like parts with lengths $n < N/2$ and $N-1-n$ (Everaers et al., 2017). By defining the sub-tree with a shorter length (or weight, as the weight is defined as the number of monomers in the sub-tree) as a "branch", the average branch weight is defined as

$$\langle N_{br}(N) \rangle \equiv 2 \sum_{n=0}^{(N-1)/2} n p_N(n), \tag{4.18}$$

where $p_N(n)$ is the probability of splitting the tree with branch size n (Everaers et al., 2017). The expectation value for the average branch weight is scaled as

$$\langle N_{br}(N)\rangle \quad \sim \quad N^{\varepsilon}, \tag{4.19}$$

where $\varepsilon = \rho$ is expected to hold in general (Janse van Rensburg and Madras, 1992). The central quantity is the length of the shortest path on the *tree* or tree contour distance, L, between two monomers, i and j, along the ring. For short *ring* contour distances, $n = |i-j|$, one simply expects $\langle L(n)\rangle \sim n$. However, beyond the typical distance between branch points, the ring does not follow a linear path on the tree but wraps side branches. For $n \ll N$, Eq. (4.17) suggests $\langle L(n)\rangle \sim n^{\rho}$. Due to the ring closure $\langle L(n)\rangle \equiv \langle L(N-n)\rangle$ reaches its maximum for $n = N/2$ before reducing to zero at the total ring size, $\langle L(N)\rangle \equiv 0$. The simplest functional form accounting for this constraint is (Rosa and Everaers, 2019)

$$\langle L(n)\rangle_N \sim \left(n\left(1-\frac{n}{N}\right)\right)^{\rho}. \tag{4.20}$$

4.3.2 Tertiary (spatial) structure

The simplest measure of the tertiary structure is given by the tree gyration radius:

$$\langle R_g^2(N)\rangle \quad \sim \quad N^{2\nu}, \tag{4.21}$$

as a function of the chain length. For a more detailed understanding, it is useful to consider the mean-square spatial distance between nodes,

$$\langle R^2(L)\rangle \quad \sim \quad L^{2\nu_{path}}, \tag{4.22}$$

as a function of their contour distance on the tree, where $\nu = \rho\,\nu_{\text{path}}$. Combining Eqs. (4.22) and (4.20) suggests (Rosa and Everaers, 2019)

$$\langle R^2(n)\rangle_N \sim \left(n\left(1-\frac{n}{N}\right)\right)^{2\nu} \tag{4.23}$$

for the mean-square spatial distance of monomers as a function of their distance, $n = |i-j|$, along the ring.

4.3.3 Secondary structure contacts, tertiary structure contacts, and total contacts

The total contact probability $\langle p_c(n)\rangle$ between pair of ring monomers at ring contour distance n is the sum of the secondary and the tertiary contact probabilities. The secondary structure contact is considered as a pair of monomers neighbors on the tree (i.e., $L_{ij} \leq l_k$) but not along the ring (i.e., $|j-i| > 1$) – l_k is the Kuhn length or lattice constant. The

probability of secondary structure contacts (described in detail in Sec4.7) as a function of the contour distance along the ring, n, scales as (Rosa and Everaers, 2019)

$$\langle p_{2c}(n)\rangle_N \sim n^{-(2-\varepsilon)}\,, \tag{4.24}$$

or

$$\langle p_{2c}(n)\rangle_N \sim \left(n\left(1-\frac{n}{N}\right)\right)^{-(2-\varepsilon)}\,. \tag{4.25}$$

A pair of monomers neighbors in space (i.e., $|\vec{r_{ij}}| \leq l_k$) but neither on the tree (i.e., $L_{ij} > l_k$) nor along the ring (i.e., $|j-i| > 1$) is considered as a tertiary structure contact.

By taking into account the ring closure constraint, the total contact probability scales as:

$$\langle p_c(n)\rangle_N \sim \left(n\left(1-\frac{n}{N}\right)\right)^{-\gamma_r}\,, \tag{4.26}$$

Where the novel exponent $\gamma_r = \nu(d+\theta_r))$ (described in detail in Sec4.6) introduced and discussed in detail in (Rosa and Everaers, 2019).

4.3.4 Flory theory of interacting tree systems

Exact values for the exponents are known only for a very small number of cases. For ideal non-interacting trees, the exponents $\rho^{ideal} = \varepsilon^{ideal} = \nu_{path}^{ideal} = 1/2$ and $\nu^{ideal} = 1/4$ (Zimm and Stockmayer, 1949; De Gennes, 1968). For interacting trees, the only known exact result (Parisi and Sourlas, 1981) is the value $\nu = 1/2$ for self-avoiding trees in $d = 3$.

Flory theories (Isaacson and Lubensky, 1980; Daoud and Joanny, 1981; Gutin, Grosberg, and Shakhnovich, 1993; Grosberg, 2014; Everaers et al., 2017) of interacting tree systems are formulated as a balance of an entropic elastic term and an interaction energy (Flory, 1953)

$$\mathscr{F} = \mathscr{F}_{el}(N,R) + \mathscr{F}_{inter}(N,R)\,, \tag{4.27}$$

In the present case, the elastic free energy takes the form (Gutin, Grosberg, and Shakhnovich, 1993)

$$\frac{\mathscr{F}_{el}}{\beta} \sim \frac{R^2}{l_K L} + \frac{L^2}{N l_K^2}\,, \tag{4.28}$$

The first term of Eq. (4.28) is the usual elastic energy contribution for stretching a polymer of linear contour length L at its ends. The second term penalizes deviations from the ideal branching statistics, which lead to longer paths and hence spatially more extended trees (Gutin, Grosberg, and Shakhnovich, 1993). In annealed trees, the interactions have the ability to alter both the branching statistics and the spatial conformations. As the contour distances increase, resulting in greater spatial separations between monomers that repel each other, it becomes necessary to minimize the Flory energy with respect to both variables, R and L, simultaneously. Optimizing L for a *given* size, $R \sim N^{\nu}$, yields (Gutin,

TABLE 4.1: Summary of scaling exponents for the double-folded ideal linear chain, double-folded self-avoiding linear chain, double-folded ideal tree, double-folded self-avoiding ring and double-folded ring in the melt state.

Exponents	**ideal linear**	**S.A linear**	**Ideal trees**	**S.A trees**	**3*D* melt of trees**
ν	$1/2$	$3/5$	$1/4$	$1/2$	$1/3$
ρ	1	1	$1/2$	$2/3$	$5/9$
ε	–	–	$1/2$	$2/3$	$5/9$
‘ $(2-\varepsilon)$	–	–	$3/2$	$4/3$	$13/9$
$\gamma_r = \nu(d+\theta_r)$	0	0	0.75	$1.41-1.43$	$1.09-1.10$

Grosberg, and Shakhnovich, 1993; Everaers et al., 2017):

$$\rho = \frac{1+2\nu}{3}, \tag{4.29}$$

$$\nu_{path} = \frac{3\nu}{1+2\nu}, \tag{4.30}$$

independently of the type of volume interactions causing the swelling in the first place. Plausibly, a fully extended system (i.e., $\nu = 1$) is predicted not to branch (i.e., $\rho = 1$) and to have a fully stretched stem (i.e., $\nu_{path} = \nu = 1$). For the radius of ideal randomly branched polymers for which $\nu = 1/4$, one recovers $\rho = 1/2$ and a Gaussian path statistics, i.e., $\nu_{path} = 1/2$.

The work in (Everaers et al., 2017) reviews the predictions of the Flory theory for randomly branching polymers for a wide range of conditions characterized by different expressions for the interaction energy in Eq. (4.27). For self-avoiding trees, the interaction part is described in Eq.4.10. In this case, Flory theory predicts (Gutin, Grosberg, and Shakhnovich, 1993)

$$\nu = \frac{7}{3d+4} \quad for 1 \leq d \leq 8\,, \tag{4.31}$$

in excellent agreement with the exact results (Rosa and Everaers, 2016b). In the melt case, all terms of the virial expansion of the partially screened excluded volume interactions become relevant, and the trees are expected to be compact (Everaers et al., 2017):

$$\nu = \frac{1}{d} \quad for 1 \leq d \leq 4\,. \tag{4.32}$$

A summary of scaling exponents is given in Table 4.1.

While Flory theory describes the average behavior of the tree observables mentioned above, the corresponding non-Gaussian distribution functions are usually of the Redner-des Cloizeaux (RdC) form of a power law multiplied with a stretched exponential which will be explained in detail in the following sections.

4.3.5 Melt of branching polymers and blobs

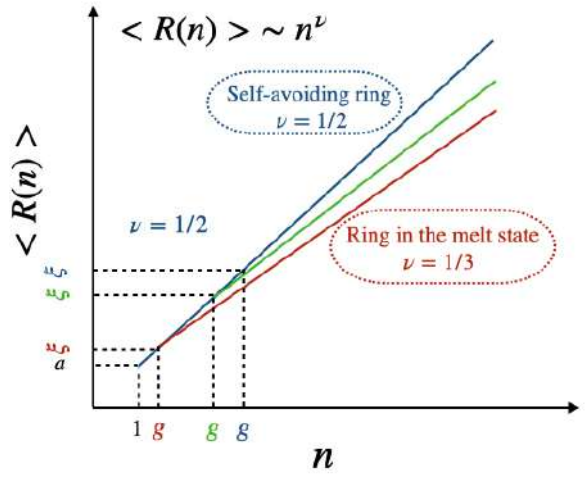


FIGURE 4.2: Schematic representation of mean-square internal distances $< R(n) >$ verses n in various concentration. Within the length-scale ξ the polymers are not aware of the existence of the external perturbations and show self-avoiding statistics.

De Gennes pioneered the blob-scaling approach in polymer physics (De Gennes, 1979). It is a very useful concept for understanding how external perturbations, like interaction with other chains and confinement, influence chain statistics and conformations. In this picture, the chain is viewed as a succession of blobs containing in average g monomers and having an average size ξ. The ring polymer, like any polymer chain, exhibits self-similar fractal behavior when it is not confined by boundaries (Rubinstein, 2003). This behavior can be described by the scaling relationship $R \sim an^{\nu}$, where R represents the size of the polymer chain and n is the number of monomers. Consequently, this scaling approach can be applied to any subsection of the chain as well.

For a subsection of the chain, the relationship between the size (ξ) and the number of monomers in the blob (g) can be approximated as follows:

$$\xi \simeq ag^{\nu} \Rightarrow g \simeq (\xi/a)^{1/\nu}, \tag{4.33}$$

Here, a represents the size of a monomer (assuming $a = 1$ for simplicity), and the Flory exponent $\nu = 1/2$ characterizes the self-avoiding rings in three dimensions. The chain density (concentration), ρ, is defined as $\rho = N/V$. Thus:

$$\rho \simeq \frac{g}{\xi^d}, \tag{4.34}$$

Hence,

$$g \simeq \rho^{\frac{1}{1-d\nu}}, \tag{4.35}$$

and

$$\xi \simeq \rho^{\frac{\nu}{1-d\nu}}. \tag{4.36}$$

As Eq. (4.35 and 4.36) show, the blobs are density-dependent, and the blob size g shrinks as the concentration increases. Figure 4.2 show the schematic of the behavior mean-square internal distances $< R(n) >$ verses n in various concentrations. Within the length-scale ξ, the polymer demonstrates self-avoiding statistics and remains unaffected by external perturbations. It is worth noting that the value of ξ is independent of the chain length. In Section 5.9.4, we conducted a thorough investigation of the blob and analyzed the influence of concentration on chain statistics within our model.

4.4 Conformation of branched polymer

Daoud and Joanny (Daoud and Joanny, 1981) proposed a calculation to determine the number of branches in the ideal trees as n_3. This calculation involves the use of two parameters, branching probability, Λ, tree length, N_t, as well as the partition function Z_{N_t} (Λ^2 referred to as the activity of the branch points in their original work). The expression for n_3 is given by:

$$\langle n_3 \rangle = \frac{\Lambda N_t}{Z_{N_t}} \frac{\partial Z_{N_t}}{\partial (2\lambda N_t)}, \tag{4.37}$$

The partition function Z_{N_t} is defined by

$$Z_{N_t} = \frac{I_1(2\lambda N_t)}{\lambda N_t} \tag{4.38}$$

and involves the modified Bessel function of the first kind, denoted as I_1. The top panel of Fig.4.3 illustrates the relationship between the number of branches and tree length for various branching probabilities. The figure demonstrates that as the branching probability increases, the number of branches also increases, holding the tree length constant.

The number of linear segments in the tree is given by $2\langle n_3 \rangle + 1$. Consequently, the length of the linear sections in the tree will automatically be:

$$\langle l_{lin} \rangle = \frac{N_t - 1}{2\langle n_3 \rangle + 1} \tag{4.39}$$

The results are depicted in the bottom panel of Fig.4.3. In Chapter 5, we will further explore these relationships in our model.

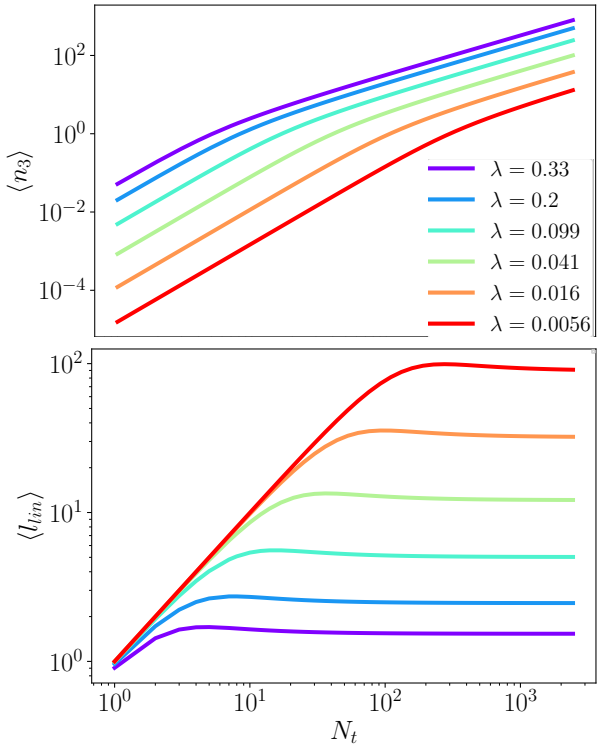


FIGURE 4.3: The relationship between the number of branches and branching probability in ideal trees. In the top panel, we illustrate how the number of branches, denoted as $\langle n_3 \rangle$, relates to tree length across different branching probabilities. The bottom panel displays the average length of linear segments within the tree, denoted as $\langle l_{lin} \rangle$, as a function of the tree length. This measurement include both branch lengths and the lengths between branch points. The graph represents ideal trees with varying branching probabilities.

4.5 Appendix A: Probability Distribution of end-to-end Distances

To develop an expression for the probability distribution of end-to-end distances of a polymer chain, let's start from a simple random walk in space and then generalize the results to $3D$ dimension.

In $1D$ random walk, the probability of arriving a distance x from the origin after N steps of unit size is given by:

$$p(N,x) = \frac{1}{2^N} \frac{N!}{n_-!\, n_+!} \tag{4.40}$$

where $n_{+/-}$ are the number of steps in positive/negative directions. By applying Stirling's approximation, $N! \approx \sqrt{2\pi N}(N/e)^N$and considering $x \ll N$, the probability $p(N,x)dx$ will be,

$$p(N,x) = \frac{1}{\sqrt{2\pi N}} \exp\left(\frac{-x^2}{2N}\right) \tag{4.41}$$

$$\langle x^2 \rangle = \int_{\infty}^{\infty} x^2 P(N,x)dx = N \tag{4.42}$$

$$p(N,x) = \frac{1}{\sqrt{2\pi\langle x^2\rangle}} \exp\left(\frac{-x^2}{2\langle x^2\rangle}\right) \tag{4.43}$$

Polymer chains in real space can explore all three dimensions. Since the three dimensions are completely uncorrelated or independent, the probability $P(N,R)$ in $3D$ space is just the product of $P(N,R_x)P(N,R_y)P(N,R_z)dx\,dy\,dz$, and $\langle R^2\rangle = l_k^2 N$. Hence,

$$p(N,R) = \left(\frac{3}{2\pi l_k^2 N}\right)^{3/2} \exp\left(\frac{-3R^2}{2l_k^2 N}\right). \tag{4.44}$$

4.6 Appendix B: End-to-end distance distributions for ring sections

The mean-square internal distance and the contact probability law can be derived from the end-to-end distance distribution, $p(\vec{r}|n)$, for ring sections of length n as:

$$\langle R^2(n)\rangle = \int_0^{\infty} |\vec{r}|^2\, p(\vec{r}|n)d\vec{r} \tag{4.45}$$

$$\langle p_c(n)\rangle = \int_0^{l_k} p(\vec{r}|n)d\vec{r}. \tag{4.46}$$

The end-to-end distance distributions is expected to have the form

$$p(\vec{r}|n) = \frac{1}{\langle R^2(n)\rangle^{d/2}} q(\vec{x}), \tag{4.47}$$

with $\vec{x} = \frac{\vec{r}}{\sqrt{\langle R^2(n)\rangle}}$. This equation indicates that $\sqrt{\langle R^2(n)\rangle}$ is the only relevant length scale. For the limit of $x \to 0$, the power law behavior of the form $q(\vec{x}) \sim q^{\theta}$ is expected to be held. However for the far apart end points ($x \gg 1$), the distribution is predicted to decay via $q(\vec{x}) \sim x^g \times exp\,((-Kx)^t)$.

This non-Gaussian distribution is well described by the one dimensional Redner-des Cloizeaux (RdC) distribution (Rosa and Everaers, 2017; Rosa and Everaers, 2019):

$$q(\vec{x}) = C(\theta,t)|x|^{\theta} exp(-(K(\theta,t)|x|)^{t}), \tag{4.48}$$

which is a power law multiplied with a stretched exponential and the characteristic exponents θ and t control the small and large behavior of the distribution.

The constants C and K are not arbitrary and are determined by applying the conditions that 1) the end-to-end distance distribution is normalized to one, ($\int q(\vec{x})dx = 1$), and 2) the second moment was chosen as the scaling length ($\int |x|^2 q(\vec{x})dx = 1$) (Rosa and Everaers, 2019). By substituting Eq.4.48 and Eq.4.47 in Eq.4.26 and considering $\langle R^2(n)\rangle \sim N^{2\nu}$, the scaling exponent of the total contact will be given by

$$\gamma_r = \nu(d+\theta_r). \tag{4.49}$$

Rosa and Everaers proposed a coherent framework to analyze the distribution functions for various quantities characterizing tree conformations and connectivities and relating the RdC exponents to the Flory exponents for interacting trees (Rosa and Everaers, 2017). A summary of scaling exponents γ_r for different systems is given in table 4.1.

4.7 Appendix C: Tree contour distance distributions

The mean tree contour distance and the secondary structure contact law can be derived from the tree contour distance distribution, $p(L|n)$, for ring sections of length n as:

$$\langle L(n)\rangle = \int_0^L L\, p(L|n)dL \tag{4.50}$$

$$\langle p_{2c}(n)\rangle = \int_0^{l_k} p(L|n)dL \tag{4.51}$$

where the tree contour distance distribution, $p(L|n)$, is non-Gaussian and is well described by the one dimensional Redner-des Cloizeaux (RdC) distribution (Rosa and Everaers, 2017; Rosa and Everaers, 2019):

$$p(L|n) = \frac{1}{\langle L(n)\rangle} q(\frac{L}{\langle L(n)\rangle}), \tag{4.52}$$

where

$$q(\vec{x}) = C(\theta_L,t_L)|x|^{\theta_L} exp(-(K(\theta_L,t_L)|x|)^{t_L}). \tag{4.53}$$

The characteristic exponents θ and t control the small and large behavior of the distribution.

The constants C and K, are determined by applying the conditions that 1) the tree contour distance distribution is normalized by one, ($\int P(L|n)dL = 1$), and 2) the first moment is the scaling length ($\int LP(L|n)dL = \langle L \rangle$) (Rosa and Everaers, 2019).

Chapter 5

Model and method

Since chromosomes are complex entities involving many molecular actors, simplification is the essential first step to developing a physical model. In this regard, polymer models have emerged as a minimalist but nontrivial physical model of chromosomes since they still retain essential features such as chain connectivity and excluded volume interactions between chain segments. This chapter describes and discusses the polymer model I have developed to model the bacterial chromosome. The first section describes the coarse-grained model and the mapping on a lattice. The second part moves on to describe in detail how to simulate the folding of this model using an elastic approach for branching polymer model (5.2). Then our Monte Carlo simulations are reviewed in detail in sections 5.3-5.5. In order to validate our model, we studied several known systems (sections 5.6 and 5.7) and showed that the static properties of our model are in very good agreement with theoretical and numerical works (section 5.9). Finally, we briefly conclude in section 5.10. Note that a part of this chapter has been published in Phys. Rev. E (Ghobadpour et al., 2021).

5.1 A coarse-grained branching lattice tree model of supercoild DNA

In general, bacterial DNA is negatively supercoiled (underwound), which leads to the formation of plectonemes through writhe changes. Experiments, simulations, and even general statistical mechanical arguments show that plectonemic supercoils behave as flexible branched polymers (Marko and Siggia, 1995b; Trun NJ, 1998) (Fig. 5.1(a)). Therefore, when studying supercoiled DNA, they are viewed as randomly branched polymers at the first step (Fig. 5.1(b)). Given the circular nature of the bacterial DNA, we consider rings tightly wrapped around the trees (double-folding chain) (Fig. 5.1(b)).

We then discretize the ring chain and use a lattice approach to simulate its dynamics (Fig.5.1(c,d)). The procedure is achieved by dividing the space into cells where the centers of these cells form an FCC lattice. The DNA conformation and its dynamics

are discretized by moving all the bp residing in a cell to the center. The lattice constant and the number of base pairs per site is the model parameters that depend on the coarse-graining level.

Fig. 5.1(c) shows the $2D$ representation of the FCC lattice. We chose a face-centered cubic (FCC) lattice to optimize the calculations as it is more spherical, and thus it converges faster to continuum results. It is well known that the scaling properties do not depend on the lattice details and only depend on the lattice dimension (De Gennes, 1979; Kardar, 2007) (the Ising model is an excellent example as numerous studies extensively prove that scaling behavior is independent of the lattice details (Kardar, 2007)). Similarly, our simulation results are not affected by choice of the lattice.

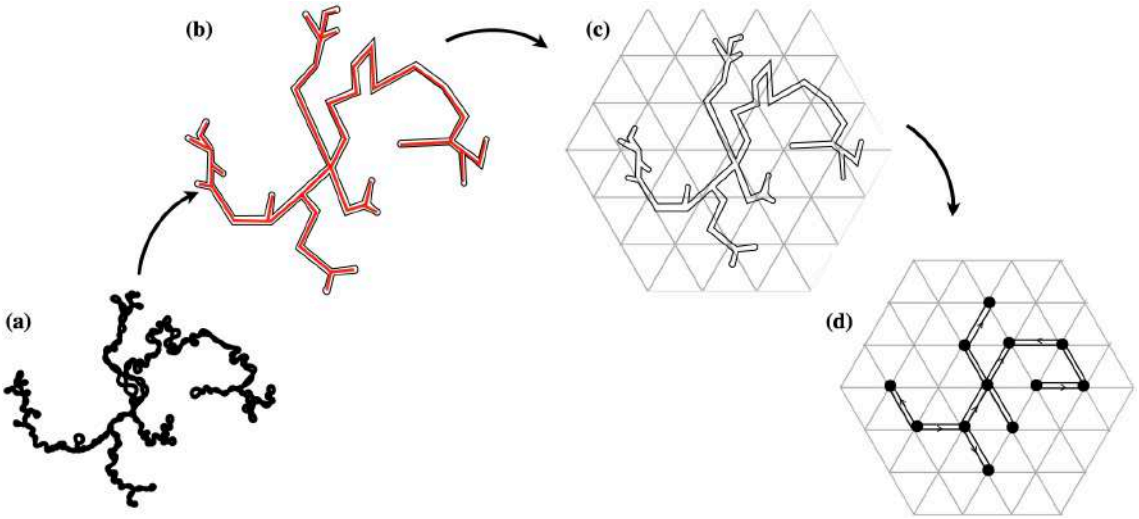


FIGURE 5.1: (a) A schematic of supercoiled DNA. (b) Plectonemic supercoils behave as flexible branched polymers. Given the circular nature of the bacterial DNA, we consider rings tightly wrapped around the trees. (c) and (d) We plug the ring onto the FCC lattice and discretize it. The lattice is a $2d$ representation of the FCC lattice.

5.2 Using an elastic approach to simulate the dynamics of the tree

We model the circular bacterial chromosome as an elastic lattice double-folded ring polymer. Our Model is based on Rubinstein repton model (Obukhov, Rubinstein, and Duke, 1994). As illustrated in the schematic graph in Fig.5.2 (A), tree nodes are located on lattice sites, and the ring is (tightly) wrapped around the tree (Fig.5.2 (B)). Lattice bonds represent tree segments that can only be occupied by two oppositely oriented bonds between underlying ring monomers. The ring polymer consists of n_m monomers, where consecutive monomers are either on adjacent or the same lattice tree nodes. The bond between two consecutive monomers on the same tree node is a unit of "stored length".

Stored lengths add elasticity to the system, and their local redistribution is responsible for the dynamics. (Schram, Barkema, and Schiessel, 2013). Hence the total number of tree nodes N_t can fluctuate around the equilibrium value N_0. I considered having on averaged 2.6 monomers on each tree node , which means $N_0 = n_m/2.6$, (For further details on the rationale behind choosing the value of 2.6, please refer to Appendix 5.12).

The functionality of the tree nodes is restricted to the values $f = 1$ (a leaf or branch tip), $f = 2$ (linear chain section), and $f \geq 3$ (branch point) (Fig.5.2(A)). The number of ring monomers belonging to the same tree node is equal to $f + \alpha$, where f is the functionality of the lattice tree node, and α is the number of "stored lengths". Eventually, the total number of monomers represents the number of bp of DNA we are studying, and the degrees of freedom are the positions of the ring monomers.

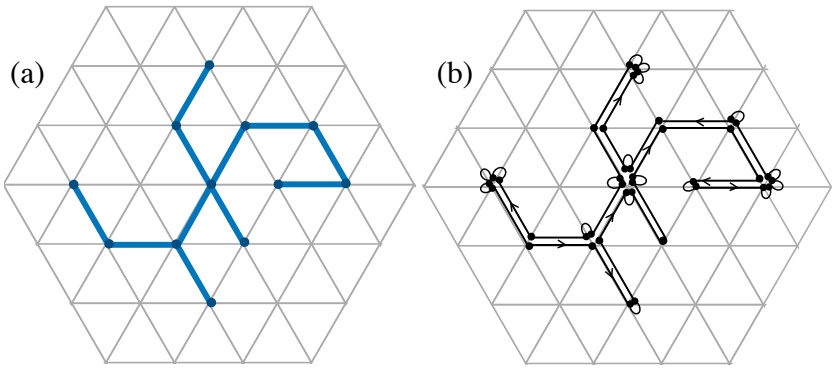


FIGURE 5.2: A schematic of a branched tree on a trigonal lattice. The functionality of the tree nodes is restricted to the values $f = 1$ (a leaf or branch tip), $f = 2$ (linear chain section), and $f \geq 3$ (branch point). (b) A corresponding (tightly wrapped) double-folded ring polymer. Small loops represent bonds of zero length, where adjacent monomers along the ring occupy identical lattice sites.

We can summarise the imposed rules as follows:

- **Lattice**: tree nodes are placed on the FCC lattice sites. We choose the FCC lattice because it is isotropic and it has the maximum number of nearest neighbors.
- **Ring monomers and tree nodes**: tree nodes are made of ring monomers. The number of ring monomers belonging to the same tree node equals $f + \alpha$, where f is the functionality of the lattice tree node and α is the number of "stored length" on that node. We chose the average number of monomers on each tree node to be 2.6.
- **Double-folding**: our model preserves the double folds of tree bonds. Each extended bond of the tree nodes is paired with exactly one extended bond pointing in the opposite direction.
- **Connectivity**: bonded monomers can occupy either the same site (a repton or loop of stored length) or nearest neighbor sites (an extended bond).

- **Order**: Ring monomers remain ordered even if several connected monomers occupy the same site

The double-folded branching tree conformation (C) is described by a set of tree node positions, $\{R_I\}$, on the FCC lattice sites. While satisfying the double-folding constraint, the conformations are sampled from the canonical ensemble with a Hamiltonian that specifies intrachromosomal interactions and the branching probability.

5.3 Energy functions

The hamiltonian for the double-folded branching tree model has four contributions: pairwise nearest-neighbor interactions, non-bonded tree node interactions, harmonic interaction acting as a restraint to control the tree contour length, and branching potential. For the sake of simplicity, our model does not include bending energy.

$$\mathscr{H}(C) = \mathscr{H}^{n.n}(C) + \mathscr{H}^{int}(C) + \mathscr{H}^{c.l}(C) + \mathscr{H}^{br}(C) \tag{5.1}$$

The contribution to the Hamiltonian due to pairwise nearest-neighbor interactions is

$$\mathscr{H}^{n.n}(C) = \beta^{-1} \sum_{I=1}^{N_t-1} \sum_{J=I+1}^{N_t} \delta(|R_I - R_J| - a), \tag{5.2}$$

where $\delta(x-a)$ is the Dirac delta function which is zero everywhere except at a, where it is infinite and we used to tune the excluded-volume effects of DNA.

The contribution to the Hamiltonian due to interactions between non-bonded tree nodes is

$$\mathscr{H}^{int}(C) = \nu_k \sum_l \frac{X_l(X_l-1)}{2}, \tag{5.3}$$

where X_l is the total number of tree nodes inside the elementary cell centered at the lattice site l, and ν_k tunes the free energy penalty for overlapping pairs of tree nodes in the coarse-grained resolution. For our model, we employ a large free energy penalty, $\nu_k = 30k_BT$, to avoid any overlapping between pairs of tree nodes.

The contribution to the Hamiltonian to control the contour length of DNA is a quadratic potential associated with the total number of tree nodes

$$\mathscr{H}^{c.l}(C) = \frac{k}{2}(N_t - N_0)^2, \tag{5.4}$$

where N_t is the total number of tree nodes (i.e. the total length of the tree) in the conformation. This term controls the tree contour length so that it fluctuates around the minimum length, N_0, which is the contour length of the tree that perfectly matches the

chromosome data given the resolution a. k controls the size of fluctuations, and we use $k = k_B T$.

Lastly, the branching probability is controlled via a chemical potential of the form:

$$\mathscr{H}^{br} = \mu_{br} N_{f=1}, \tag{5.5}$$

Where $N_{f=1}$ is the total number of branch tips, $f = 1$, in the tree, and μ is the chemical potential to control the branching probability. For $\mu = 0$, the tree is randomly branched.

5.4 Monte Carlo simulations

All simulations are carried out on a $3D$ FCC lattice with either periodic boundary conditions or fixed boundaries depending on the resolution of the model. In the dilute systems, we adapt a large box, so the average density of monomers per site is less than one percent; in the dense (melt) condition, a high lattice density, $\rho = 85\% - 90\%$ was used.

A Markov chain Monte Carlo algorithm was employed to sample the conformations from the canonical ensemble with the defined Hamiltonian. Trial move from the initial configuration for the final one, $C_i \to C_f$, is accepted via Metropolis criterion, and if the new conformation respects the conformational rules (introduced in Sec. 5.2),

$$acc_{i \to f} = min(1, \frac{P_{tri}(f \to i)}{P_{tri}(i \to f)} \frac{P(C_f)}{P(C_i)}). \tag{5.6}$$

Here, $P_{\text{tri}}(i \to f) = P_{\text{tri}}(f \to i)$ represents the proposal transition probability from state i to state j. In the context of the Metropolis method, this probability is equivalent to the reverse move probability from state j to i.

The tree configurations are observed with a probability,

$$P(C) = \frac{w(C)}{Z}, \tag{5.7}$$

Where $w(C)$ is the statistical (Boltzmann) weight of the tree conformation C,

$$w(C) = e^{-\beta \mathscr{H}(C)}. \tag{5.8}$$

and Z is the partition function of the system found by summing over all possible configurations of N_t tree nodes on the FCC lattice,

$$Z = \sum_C e^{-\beta \mathscr{H}(C)}. \tag{5.9}$$

Simply, by substituting equation 5.8 in 5.6, the acceptance probability is a function of the energy difference, $\Delta E = \mathscr{H}(C_f) - \mathscr{H}(C_i)$,

$$acc_{i\to f} = min(1, e^{-\beta\Delta E}) \tag{5.10}$$

5.5 Double-folding preserving moves

All the moves in our elastic lattice polymer model have the constraint of preserving the double-folding structure. They are divided into three groups: **tree node move**, **one monomer move**, and **pair monomer move**. In the following, these moves are explained in detail.

5.5.1 Tree node move

This move consists of trying to move a randomly selected tree node to a randomly chosen site out of twelve possible nearest neighbors. This move preserves the tree structure i.e., and it does not change the underlying monomer ring connectivity. The move is only accepted if the new conformation respects the double-folding constraint. The potential energy difference caused by the move is:

$$\Delta E = \mathscr{H}^{int}(C_f) - \mathscr{H}^{int}(C_i) \tag{5.11}$$

Examples of these MC moves for the present model are represented in Fig.5.3.

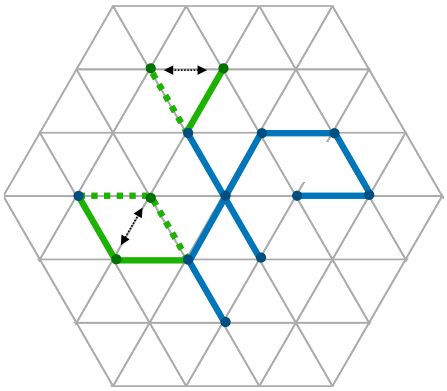


FIGURE 5.3: Two examples of tree node move. The move consists of trying to move a randomly selected tree node to a randomly chosen site out of twelve possible nearest neighbors under a double-folding preserving constraint.

5.5.2 One monomer move

This move involves making a monomer hop into a nearest neighbor lattice site. All allowed hopping moves for tightly folded rings can be classified in terms of two different move types:

The Repton move: A unit of stored length hops one unit along the tree without changing its structure or the system's energy, but the local redistribution of stored length is responsible for the dynamics. An example of a repton move is shown in Fig. 5.4(a).

The Hairpin move: There is at least one connected loop of stored length (in a site) on each side of a monomer; both loops can unfold and result in an extended bond (creation). Naturally, the inverse move removes an extended bond pair and thus shortens or removes a side branch from the tree structure (annihilation). The possible energy difference caused by the move is:

$$\Delta E = \Delta\mathscr{H}^{int} + \Delta\mathscr{H}^{br} + \Delta\mathscr{H}^{c.l} \tag{5.12}$$

$\Delta\mathscr{H}^{br}$ is the energy change due to branching. It can hence be equal to $\pm\mu$ (creation/deletion of a side branch) or 0 (corresponds to adding/deleting a segment to/from the end of the branch tip). $\Delta\mathscr{H}^{c.l}$ is responsible for controlling the total contour length of the tree. An example of the hairpin move is shown in Fig. 5.4(b).

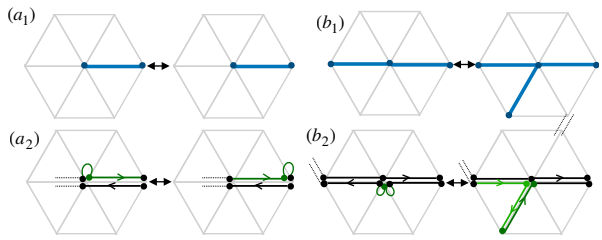


FIGURE 5.4: One monomer move is hopping a randomly selected monomer into the nearest neighbor lattice site. (a) The Repton move; This move is responsible for the local redistribution of stored lengths. (b) The Hairpin moves; This move is responsible for the creation or annihilation of side branches. The top row shows the tree structure, and the bottom row shows the underlying double-folded ring.

5.5.3 Pair monomer move

In the double-folded ring model, each extended bond is paired with another bond pointing in the opposite direction. In this move, one of these paired bonds is randomly selected; since two tree nodes belong to these paired bonds (two endpoints of the bonds), one of them is also randomly selected, resulting in the paired monomers. The move is accepted under the condition of preserving the double-folded structure. Paired moves are classified in terms of two different move types:

The transversal move: The bond vectors adjacent to the selected pair monomers are of the form $(\vec{w},0)$ and $(0,-\vec{w})$; this move results in a new position with the new vectors $(\vec{u},\vec{v})$, and $(-\vec{v},-\vec{u})$, with $\vec{w}=\vec{u}+\vec{v}$ (Fig. 5.5(a)).

This move introduces elasticity into the system by making the chain length fluctuate. Therefore, the possible energy difference caused by the move is:

$$\Delta E = \Delta\mathscr{H}^{int} + \Delta\mathscr{H}^{c.l} \tag{5.13}$$

The slip-link move: The result of this move is slipping a branch point along the backbone. An example of this move is shown in Fig. 5.5(b). The bond vectors of the paired monomers are in the form:

$$\begin{bmatrix} (0,\vec{v}) \\ (-\vec{v},\vec{u}) \end{bmatrix} \tag{5.14}$$

Then, this move will change the bond vector as:

$$\begin{bmatrix} (0,\vec{v}) & \leftrightarrow & (\vec{u},\vec{w}) \\ (-\vec{v},\vec{u}) & \leftrightarrow & (-\vec{w},0) \end{bmatrix} \tag{5.15}$$

where, $\vec{w} = \vec{v} - \vec{u}$.

This move mimics the plectoneme's motion by letting a side branch slip along the backbone. The possible energy difference caused by this move is:

$$\Delta E = \Delta\mathscr{H}^{br} \tag{5.16}$$

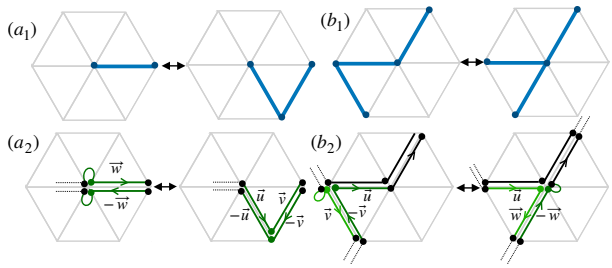


FIGURE 5.5: Examples of pair monomer moves. (a) Transversal move; This move makes the chains' length fluctuate, introducing elasticity into the system. (b) Slip-link move; Letting a side branch slip along the backbone mimics the plectoneme's motion. The paired monomers and their bonds are colored green.

5.6 Studied systems

The simulations were carried out for chain lengths varying between 216 to 1200 monomers for different systems like ideal double-folded rings, self-avoiding double-folded rings, and rings in the melt state. We studied the effect of the strength of excluded volume

interactions, the average length of branches, chain lengths, concentration, and confinement on the conformational statistics of the systems. The box size was set large enough to avoid self-interaction of the chain due to the periodic boundary condition (except for the simulation in which we studied the effect of confinement and concentration).

5.7 Initialization and equilibration

Our model's initialization and equilibration procedure are explained in detail in (Ghobadpour et al., 2021). Here, we give a brief review of it.

To monitor the state of equilibration of the polymer systems, the radius of gyration is the structural property that has been investigated. The systems have reached equilibrium when this quantity no longer changes. All reported quantities are taken from the simulation regime where the polymers are fully equilibrated.

Figure5.6 shows the radius of gyration vs MC time counted in sweeps for different trees – one sweep corresponds to a number of trials such that the probability for a monomer to be considered is fixed with respect to the number of monomers. To validate the equilibrium values of the radius of gyration, we ran the simulations from two totally different initial conditions. The left column shows the equilibration of compact, highly branched initial configurations, which swell as the simulation progresses with time. The right column shows the equilibration of more open, initially unbranched double-folded rings, where average ring size decreases as branches appear. The horizontal lines show the average values after equilibrium. In particular, this figure confirms that the simulation results do not depend on the initial conformations of the ring polymer chains.

5.8 Qualitative insights

Figure 5.7 illustrates configurations of our lattice model for double-folded ring polymers as they emerge from our simulations after the systems have reached equilibrium. Ring monomers are shown as small spheres which are displaced from their actual position by a small random offset. This representation reveals (i) multiple occupancies of lattice sites and (ii) double-folding.

Specifically, we show a sample configuration of (a) an ideal randomly branched double-folded ring, (b) a self-avoiding randomly branched double-folded ring, and (c) a randomly branched double-folded ring in the melt state with 216 monomers as well as (d) a view of a corresponding melt configuration (unfolded from the simulation box) for 12 double-folded rings. The gray tubes show the longest paths along the trees. For the ideal case, there is no restriction on the number of monomers on each site. Rings in the melt state and, in particular, isolated self-avoiding rings appear swollen relative to the ideal case. At least qualitatively, this is the expected (Everaers et al., 2017) consequence of

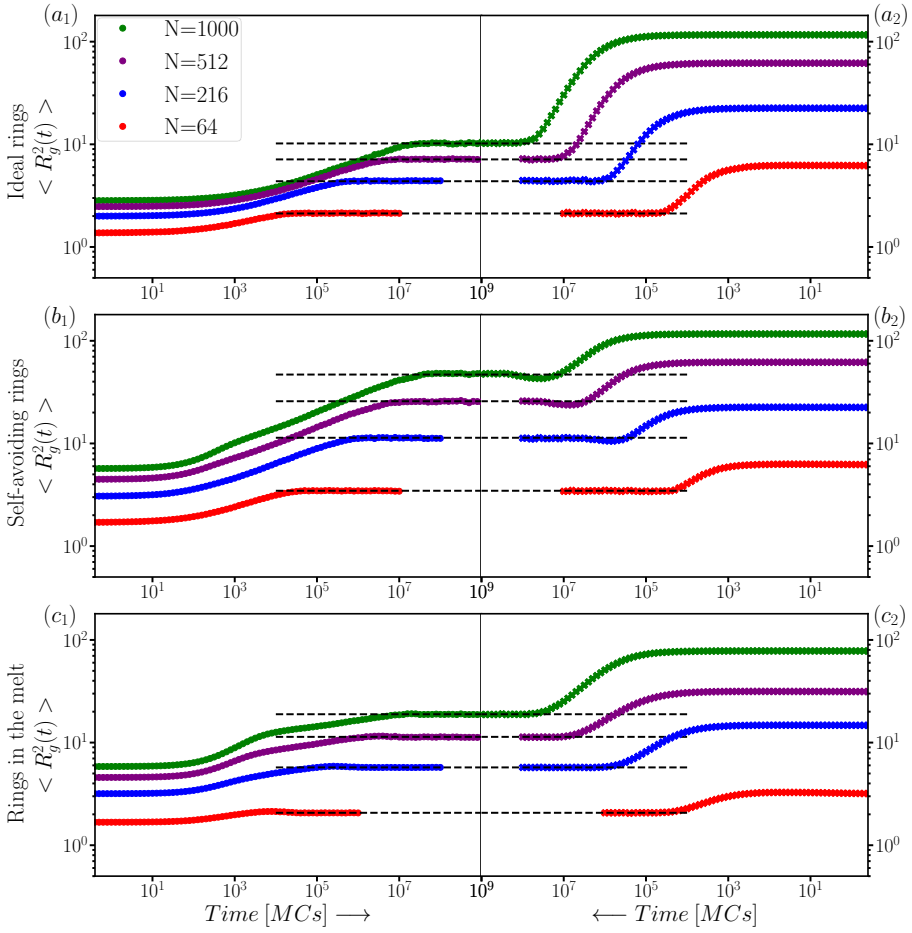


FIGURE 5.6: Equilibration of the polymer model for 4 different chain lengths for three different systems: (a) ideal double-folded rings, (b) self-avoiding double-folded rings, and (c) double-folded rings in the melt state. Comparison between initially branched, compact double-folded chains (left column), and initially double-folded chains with no branches (right column). The horizontal lines show the average values after the equilibrium. For better visualization, the time direction is reversed in the right column. In all three systems, at large times both initial states (from the left and right) reach the same equilibrium values. It proves that the polymer model results are independent of starting condition.

excluded volume interactions and the partial screening in melts. The structure is quite anisotropic in the self-avoiding case, and the longest path is aligned along the longest axis. In the melt case, the structure is more compact and spherically symmetric, and the mass is almost equally distributed between the branches.

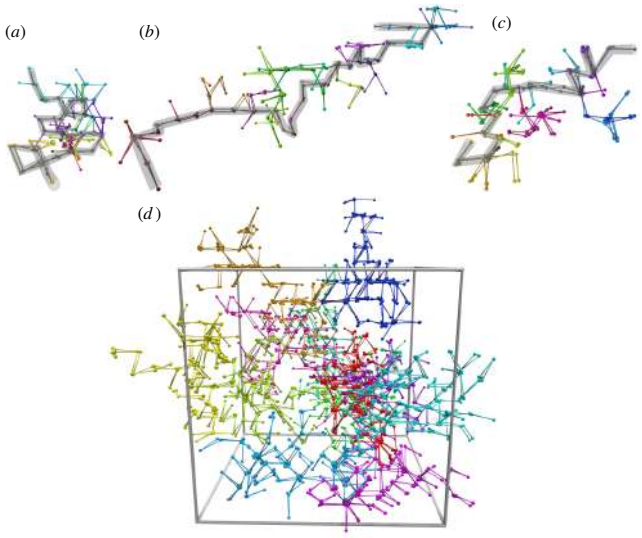


FIGURE 5.7: Equilibrated simulation snapshots of (randomly selected) configurations of the double-folded rings with $N = 216$. Successive segments are represented with an HSV cyclic color map. (a) a single double-folded ideal ring; (b) a single double-folded self-avoiding ring; (c) a single double-folded ring in the melt state. The gray tubes show the longest paths of the trees. All the trees have the same bond scale. The size of the ring in the melt is larger than the ideal tree and smaller than the self-avoiding tree. (d) Sample configuration of the melt with 12 double-folded rings. Each ring is represented by a different color. The snapshots were produced using Blender 2.8 Hess, 2010. 3D views of these configurations are available in our published article, supplementary materials, videos S1, S2, and S3 (Ghobadpour et al., 2021).

5.9 Conformational statistics

To analyze the secondary and tertiary structure of our double-folded ring polymers, we have calculated the tree contour distance $L(i, j)$ and square spatial distance $R^2(i, j)$ between all pairs of ring monomers i and j. Moreover we have calculated the secondary structure contacts, $p_{2c}(i, j)$, tertiary structure contacts $p_{3c}(i, j)$, and total contacts, $p_c(i, j)$ for all pairs of monomers.

The tree contour distance $L(i, j)$ is defined as the length of the shortest path on the tree connecting i and j. The secondary structure contact, $p_{2c}(i, j)$, is equal to 1 if i and j are neighbors along the tree and not along the ring ($|i - j| > 1$); otherwise, it is equal to 0. $L(i, j)$ and $p_{2c}(i, j)$ only depend on the tree connectivity and are completely independent of the spatial embedding of the double-folded ring polymer.

Conversely, the calculation of the spatial distance, $R^2(i, j)$, is straightforward given the monomer positions and completely independent of the secondary structure. The tertiary structure contacts $p_{3c}(i, j)$ is equal to 1 if i and j are neighbors in space but neither on the tree nor along the ring; otherwise, it is equal to 0. Total contact, $p_c(i, j)$, is the sum of $p_{2c}(i, j)$ and $p_{3c}(i, j)$. A schematic of the definition of the structure contacts in our lattice model is shown in Fig. 5.8.

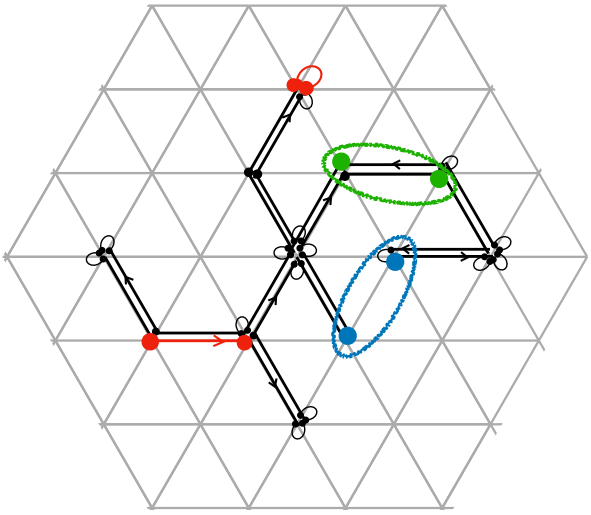


FIGURE 5.8: Schematic of the definition of the secondary and tertiary structure contacts. Red: Examples of pairs that are neighbors along the ring. Green: Secondary structure contact p_{2c}, two monomers are neighbors along the tree and not along the ring. Blue: Tertiary structure contact p_{3c}, pair of ring monomers are neighbors in space but neither on the tree nor along the ring. The total contact of a pair of monomers is the sum of secondary and tertiary structure contacts.

In the following subsections, results of averaged $L(n)$ and $R^2(n)$ over monomer pairs with identical ring contour distance, $n = |i - j|$, are shown in panels (a_1) and (b_1) of Figs. 5.9-5.17, and the finite ring size effects are effectively dealt with by expressing observables as a function of $n_{eff} = n(1 - n/N)$ in panels (a_2) and (b_2) of the same figures. This representation makes a meaningful comparison with the expected power law relations $\langle L(n)\rangle \sim n^{\rho}$ and $\langle R^2(n)\rangle \sim n^{2\nu}$ (Introduced in Sec.4.3.1).

In addition, we have extracted effective exponents by calculating the derivatives using the logarithm of neighboring data points, $(\frac{\Delta log\langle L(n_{\text{eff}})\rangle}{\Delta log n_{\text{eff}}})$ and $(\frac{\Delta log\langle R^2(n_{\text{eff}})\rangle}{\Delta log n_{\text{eff}}})$ (Ghobadpour et al., 2021). The results are shown in the inset of the corresponding figures' panels (a_2) and (b_2). The horizontal lines show the expected exponents corresponding to the known exact value or the predictions of Flory theory for the systems under investigation in the asymptotic limit of infinite ring size.

Moreover, values of the $p_{2c}(n)$, $p_{3c}(n)$, and $p_c(n)$ between all possible pairs of monomers as a function of $n_{eff} = n(1-n/N)$ are shown in panels (a), (b), and (c) in second Figures in the following sections. The scaling arguments allow refining the values of exponents, $\langle p_{2c}(n)\rangle \sim n^{-(2-\varepsilon)}$, and $\langle p_c(n)\rangle \sim n^{-\gamma_r}$ beyond the average regime (Rosa and Everaers, 2019). The value of these effective exponents has shown in the insets of the panels by dashed lines.

5.9.1 Maximum tree node functionality effect

The functionality of tree nodes depends on the number of emerging tree segments, $f = 1$ (a leaf of a branch tip), $f = 2$ (linear chain section), and $f \geq 2$ (branch point). The maximum tree node functionality for the trees on the FCC lattice is $f_{max} = 12$ since each lattice site has 12 nearest neighbors.

The number of branch points and branch tips are related via (Rosa and Everaers, 2016a):

$$N_{f=1} = 2 + \sum_{f=3}^{\infty} (f-2)N_f \tag{5.17}$$

For trees with maximum functionality 3, this equation simplifies as $N_{f=1} = 2 + N_{f=3}$.

Figures 5.9, and 5.10 show the results for the isolated self-avoiding chains with $N = 1200$ monomers with different maximum functionalities. For $f_{max} = 2$ which corresponds to the double-folded linear chains, $\langle R^2(n)\rangle \sim n^{1.2}$, and $\langle L(n)\rangle \sim n^1$ are expected to be held (Fig. 5.9). For all other systems $f_{max} \geq 3$, we expected to see the behavior of isolated self-avoiding trees, $\langle R^2(n)\rangle \sim n^1$ and $\langle L(n)\rangle \sim n^{2/3}$, independent of the maximum functionality of the system (Fig. 5.9).

All the contact probabilities are shown in Fig.5.10. The secondary structure contacts the pairs of monomers for a linear chain ($f_{max} = 2$) is constant (Fig.5.10 (a)). For all other cases, the data map on top of each other and show the expected behavior of isolated self-avoiding trees. Hence, for simplicity, all results in this thesis are derived from the maximum functionality $f_{max} = 3$.

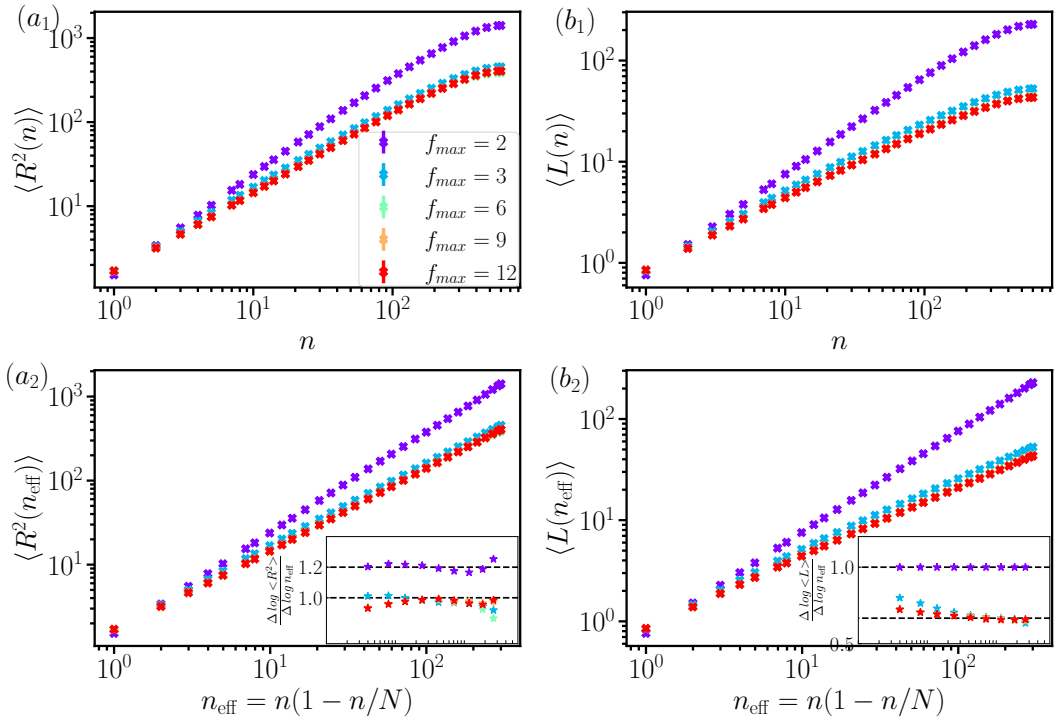


FIGURE 5.9: Conformational statistics of double-folded rings with length 1200 for five systems with different maximum functionalities (described in the legend). Data are shown for ring contour distances up to $N/2$. Column (a) plots the squared internal distances as a function of n, $\langle R^2(n) \sim n^{2\nu}$. Column (b) are the average values of the tree contour distances between all possible pairs of monomers, $\langle L(n) \sim n^{\rho}$. The exponents for the isolated self-avoiding trees are $\nu = 1/2$ and $\rho = 2/3$, and for the linear chains ($f_{max} = 2$), they are $\nu = 5/6$ and $\rho = 1$. In panels (a_2) and (b_2), data are plotted as a function of n_{eff}, which effectively reduces finite size effects. (a_2) and (b_2) insets show the local slopes of the data in panels (a_2) and (b_2), respectively. These effective exponents appear to converge to the theoretical exponents (dashed horizontal lines). The expected exponents reported in Table 4.1.Error bars are the same size or smaller than the symbols.

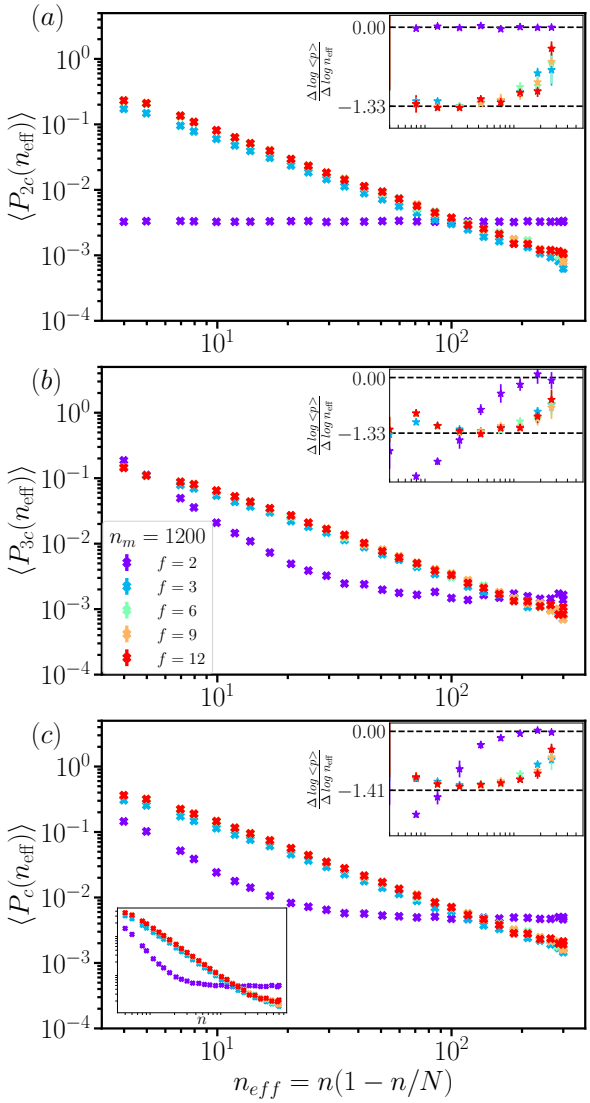


FIGURE 5.10: Contact probabilities of double-folded rings with length 1200 for five systems with different maximum functionalities (described in the legend). Data are shown for ring contour distances up to $N/2$. (a) $\langle p_{2c}(n) \sim n^{-(2-\varepsilon)}$, average probability of the secondary structure contacts. (b) the tertiary structure contact. (c) $\langle p_c(n) \sim n^{-\gamma_r}$, the total contact probabilities as a function of n_{eff} which effectively reduces finite size effects. The insets show the local slopes of the data in corresponding panels. The straight lines show the expected exponents reported in Table 4.1. Error bars are the same size or smaller than the symbols.

5.9.2 Effect of excluded-volume interaction

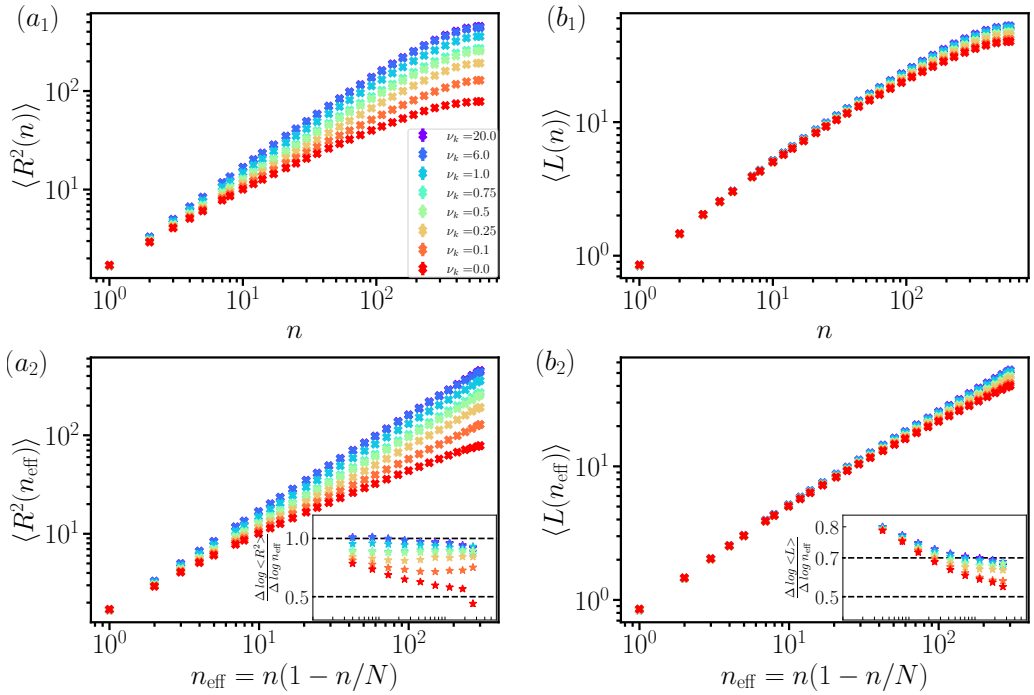


FIGURE 5.11: Conformational statistics of double-folded rings with length 1200 for eight systems with different excluded-volume interaction strengths, ν_k. All other terms of the energy functions are zero, $\mu = 0$, and $k = 0$. Column (a) plots the squared internal distance as a function of n. Column (b) plots the tree contour distances' average value between all possible monomers pairs. The exact exponents for the ideal case, $\nu_k = 0$, are $\nu = 1/4$ and $\rho = 1/2$. For the self-avoiding case, $\nu_k >> 0$, the exact scaling exponent is $\nu = 1/2$, and Flory theory predicts $\rho = 2/3$. By increasing the interaction strengths, the expected behavior that the data should migrate from an ideal case to an effectively self-avoiding case is well supported by our data. Notations and symbols are as in Fig. 5.9.

As mentioned in Sec.5.3, we can manipulate the free energy penalty for overlapping pairs of tree nodes via the parameter ν_k. Figure 5.11 and 5.12 show the results for the chains with $N = 1200$ monomers with different interaction strengths, ν. $\nu_k = 0$ corresponds to the ideal case where there are no interactions between tree nodes, and $\nu_k \gg 0$ corresponds to effectively self-avoiding systems. Mean-square internal distances, $\langle R^2(n)\rangle \sim n^{2\nu}$, and tree contour distances, $\langle L(n)\rangle \sim n^{\rho}$, have the exponents $\nu = 1/4$ and $\rho = 1/2$ for the isolated ideal trees, and $\nu = 1/2$ and $\rho = 2/3$ for the isolated self-avoiding trees, respectively. The systems with weak excluded-volume interactions have in-between exponents depending on how strong the interactions are. Moreover, contact probabilities, $\langle p_{2c}(n) \sim n^{-(2-\varepsilon)}$ and $\langle p_c(n) \sim n^{\gamma_r}$, have exponents $\varepsilon = 1/2$, $\gamma_r \sim 0.75$ and $\varepsilon = 2/3$, $\gamma_r \sim 1.43$ for isolated ideal trees and self-avoiding trees, respectively. Data

in Figure 5.11 and 5.12 show how the trees gradually change from the ideal case (no interaction) to the hard excluded volume case.

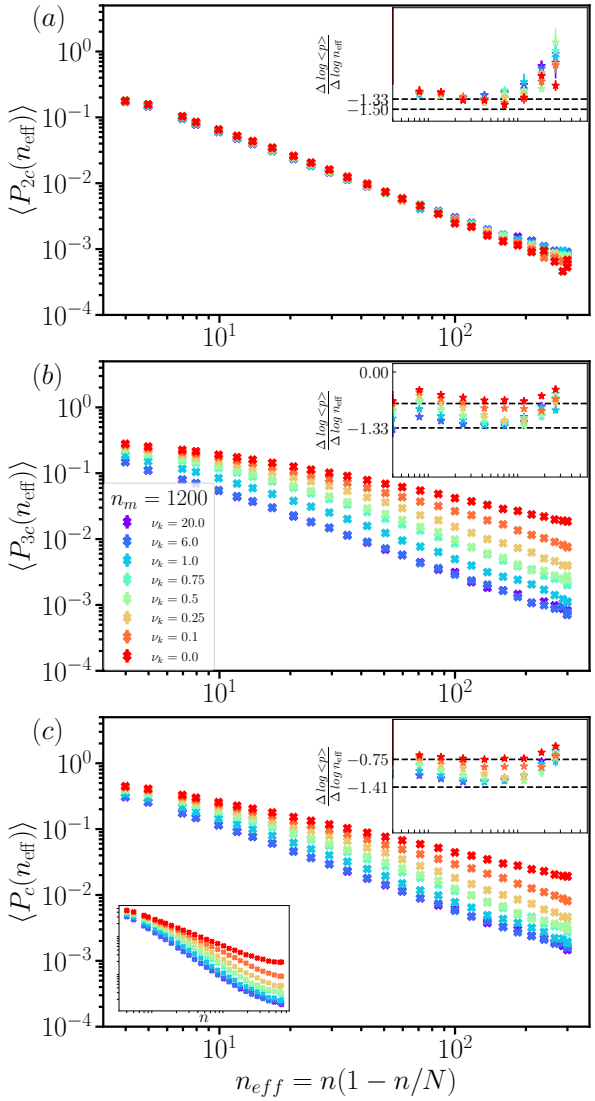


FIGURE 5.12: Contact probabilities of double-folded rings with length 1200 for eight systems with different excluded-volume interaction strengths, ν_k. All other terms of the energy functions are zero, $\mu = 0$, and $k = 0$. (a) the average value of secondary structure contacts, (b) tertiary structure contacts, and (c) total contacts between all possible monomers pairs. The expectation behavior $\langle p_{2c}(n) \sim n^{-1.5}$ and $\langle p_c(n) \sim n^{-0.75}$ for the ideal case, and $\langle p_{2c}(n) \sim n^{-1.33}$ and $\langle p_c(n) \sim n^{-1.41}$ for self-avoiding trees are supported by our numerical results. Moreover, the transition from ideal behavior to self-avoiding one by changing the interaction strengths is clear in our data. Notations and symbols are as in Fig. 5.10.

5.9.3 Branching probability effect

As mentioned in section 5.3, the asymptotic branching probability, λ, is controlled via the chemical potential for the branch points μ, in the energy term $\mathscr{H}^{br} = \mu N_{f=1}$. $N_{f=1}$ is the number of $1-$functional nodes in the tree (the number of branch points and the number of branch tips are related via $N_{f=3} = N_{f=1} - 2$). $\mu = 0$ corresponds to a randomly branching chain. By increasing μ, the branching probability decreases; hence the average length of branches increases. Therefore, the chain is effectively linear for $\mu >> 0$. Figs. 5.13 and 5.14 show that this is well supported by our data.

λ is defined as

$$\lambda = \lim_{N\to\infty} \frac{\langle N_{f=3}(N)\rangle}{N} \tag{5.18}$$

where, $\langle N_{f=3}(N)\rangle$ is the average number of branched points and N is the chain contour length. λ for different value of μ are listed in table 5.1.

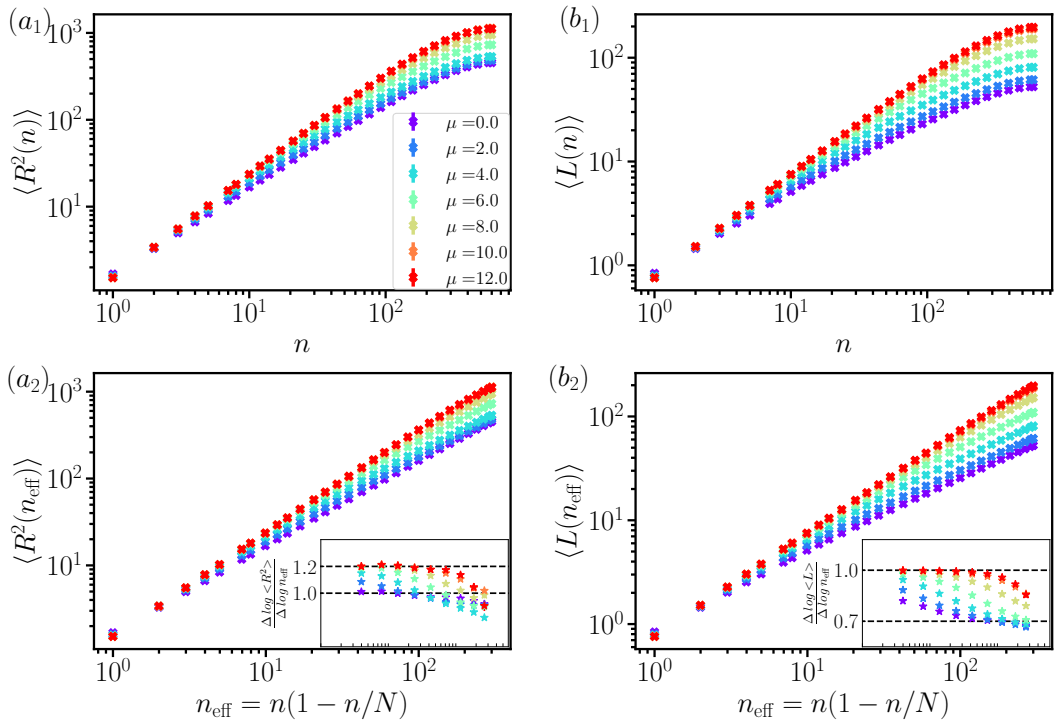


FIGURE 5.13: The effect of the branching probability on the conformational statistics of the double-folded self-avoiding trees with length 1200. The other terms of the energy functions are fixed, $\nu_k = 20$, and $k = 0$. By increasing the Hamiltonian chemical potential for the branch points, μ, the expected behavior that the data should migrate from randomly branching trees to effectively linear chains is well supported by our results. In other words, the scaling exponents gradually change from $\nu = 1$ and $\rho = 2/3$ to $\nu = 1.2$ and $\rho = 1$ by decreasing the chemical potential for the branch points. Notations and symbols are as in Fig. 5.9.

TABLE 5.1: Branch statistics for the double-folded self-avoiding trees with length 1200 with different chemical potential for branch points, μ. $\langle l_{lin}(N)\rangle$: The average length of linear segments in the tree, in the unit of lattice constant; $\langle n_{f=3}(N)\rangle$: the average number of branch points; This includes both the branch lengths and the lengths between branch points. λ: branching probability. All measurements are performed after reaching equilibrium.

μ $[k_BT]$	$\langle l_{lin}\rangle$	$\langle n_{f=3}(N)\rangle$	λ
0	1.50 ± 0.00	158.75 ± 0.39	0.319
2.0	2.40 ± 0.01	92.5 ± 0.33	0.195
4.0	5.05 ± 0.04	41.64 ± 0.26	0.089
6.0	13.29 ± 0.22	16.10 ± 0.18	0.035
8.0	38.03 ± 1.14	5.45 ± 0.10	0.011
10.0	128.03 ± 5.03	1.5 ± 0.06	0.003
12.0	198.17 ± 7.05	0.77 ± 0.04	0.001

The Relationship Between Branching Probability and Number of Branch Points

As discussed in Section 4.4 of Chapter 4, Daoud and Joanny established a relationship between the number of branch points in a tree and the branching probability. This relationship is illustrated in Figure 4.3. In order to gain a deeper understanding of our model, we compare the theoretical predictions of the Daoud theory (lines) with simulation results (star symbols) for ideal trees ($\nu_k = 0$), with different chain lengths and various chemical potentials (μ). The results, shown in Fig. 5.15, demonstrate a good agreement between the model and the simulation, except for a slight deviation in the branch lengths $\langle l_{lin}\rangle$ for very small chain lengths. This deviation is likely due to the fact that the chains are not in the branched regime.

Subsequently, we attempted to derive a connection between the chemical potential μ in our model and the branching probability λ in the theory. The relation is depicted in Figure 5.16, and the best fit of the data is given by the equation:

$$\lambda = \frac{1}{2 + e^{-\beta\mu/2}}. \tag{5.19}$$

By knowing this relation, we can tune the chemicasl potential in our model for the specific observables of average branch length.

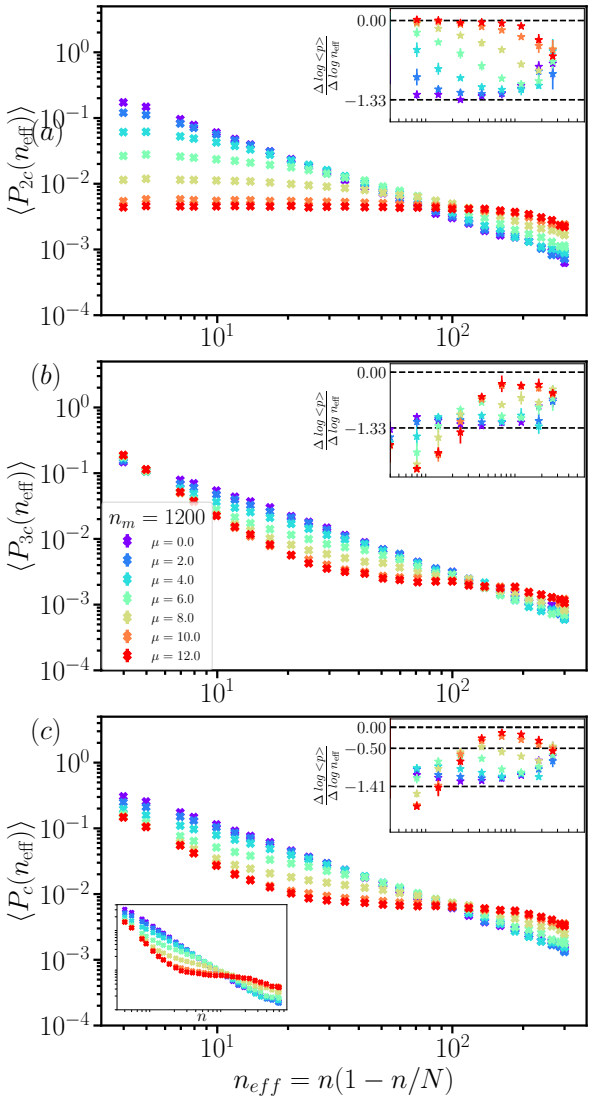


FIGURE 5.14: The average branch length effect on the contact probabilities of the double-folded self-avoiding trees with length 1200. The other terms of the energy functions are fixed, $\nu_k = 20$, and $k = 0$. By increasing the Hamiltonian chemical potential for the branch points, μ, the average length of branches increases, eventually for large enough μ, the chains are effectively linear without side branches. The expected behavior for migrating the chains from randomly branching trees to effectively linear chains is well supported by our data. In other words, the scaling exponents gradually change from $\langle p_{2c}(n) \sim n^{-1.33}$ and $\langle p_c(n) \sim n^{-1.41}$ to $\langle p_{2c}(n) \sim n^{0.0}$ and $\langle p_c(n) \sim n^{0.0}$ by decreasing the branching probability. Notations and symbols are as in Fig. 5.10.

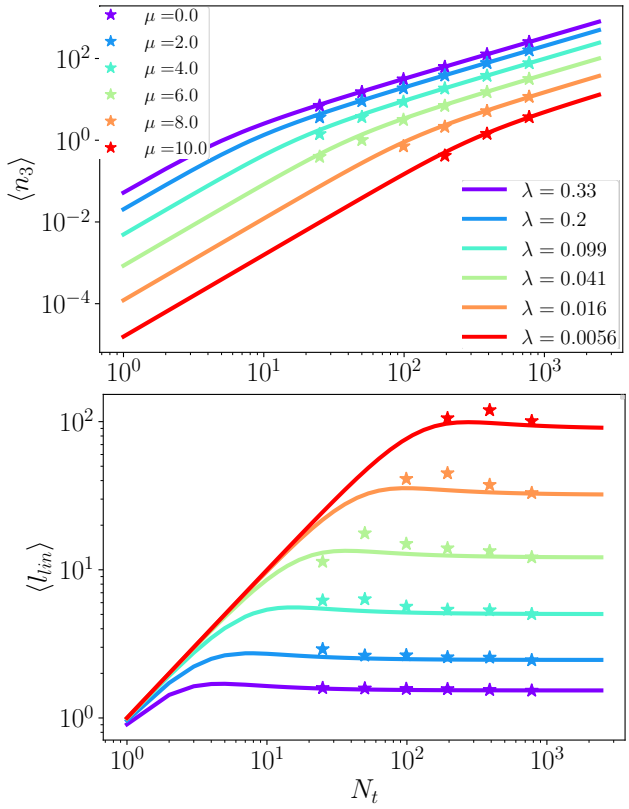


FIGURE 5.15: The figure compares the predictions of the Daoud theory (lines) and simulation results (star symbols) for ideal trees with varying branching probabilities. The model shows good agreement with the theory's predictions.

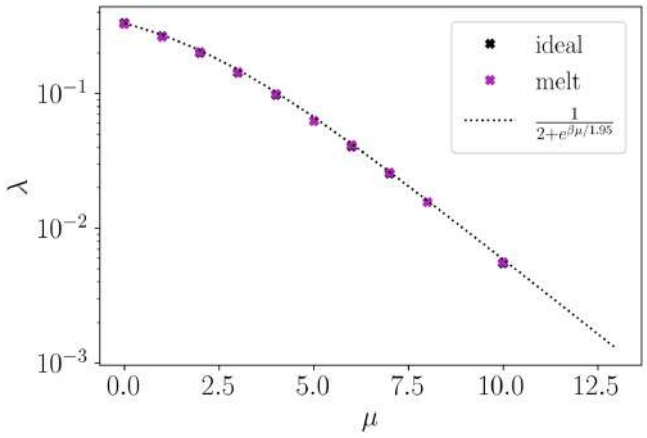


FIGURE 5.16: The relationship between branching probability λ and chemical potential μ.

TABLE 5.2: Blob sizes in different concentrations for the chains with 1200 monomers. ρ: Lattice density which is the ratio of the number of occupied sites to the total number of sites. Blobs containing in average g monomers and having an average size ξ. The blobs are density-dependent, and the blob size g shrinks as the concentration increases.

ρ	g	ξ
10%	93.53	9.67
30%	10.64	3.26
60%	2.89	1.70
80%	1.53	1.24
93%	1.16	1.08

5.9.4 Concentration effect

We can generalize our study to the systems of the polymer chains in the melt state, describing the behavior of the chromosomes. To this aim, we applied our model to study the systems with 1200 length at different concentrations. We have controlled lattice density in each system by changing the number of polymers in them ($N_p = 1,3,6,9,12$) while maintaining a fixed box size under periodic boundary conditions, and keeping all other parameters fixed. The chains in our study are randomly branched rings,($\mu = 0.0 k_b T$) and exhibit hard excluded volume interactions with $v_k = 20 k_b T$. The results are shown in Figs. 5.17-5.18.

Another point regarding the study of the concentration effect is that we did not control the total number of tree nodes in our systems (denoted as $k = 0$ in equation 5.4). Instead, we allowed the total number of tree nodes in each system to change according to the concentration variation. Consequently, the average number of monomers in each system is not constant due to the increase in concentration. This variation in the average number of monomers resulted in change in the average bond length between monomers. Specifically, in the absence of reptons with a bond length of zero ($\langle b \rangle = 1$), increasing the number of reptons caused a decrease in the average bond length ($\langle b \rangle$ decreased from 0.83 to 0.60, moving from a more dilute to a denser system). To facilitate a more accurate comparison between the systems, I multiplied the ring contour distance n by this average bond length $\langle b \rangle$ in Figs. 5.17-5.18.

The results illustrate by increasing the density in the system. We have a smooth crossover from a dilute condition to a melt case. The melt case shows the expected territorial behavior, $\nu = 1/3$.

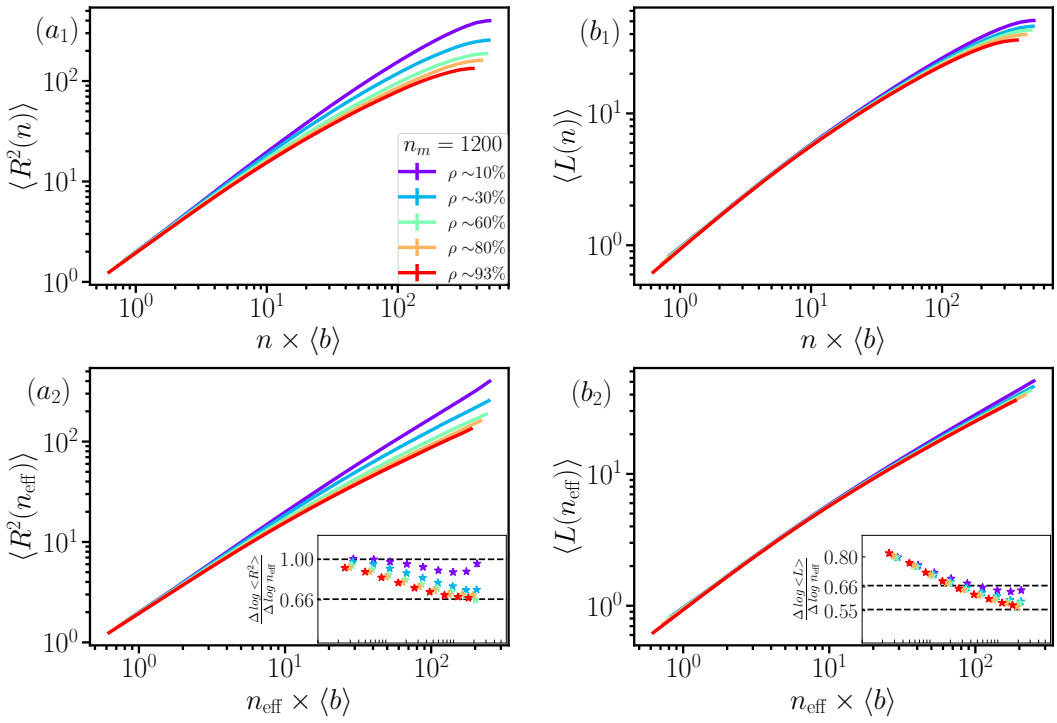


FIGURE 5.17: Concentration effect on the conformational statistics of the double-folded self-avoiding trees with length 1200. All the parameters of the energy functions are fixed, $\nu_k = 30\,k_BT$, $\mu = 0$, $k = 0$. By increasing the number of chains ($N_p = 1,3,6,9,12$) in the simulation box, the result illustrates the crossover from isolated self-avoiding behavior to the self-avoiding tree in the melt state. In other words, the scaling exponents gradually change from $\nu = 1.0$ and $\rho = 2/3$ to $\nu = 1/3$ and $\rho = 5/9$ by increasing the density in the system. Notations and symbols are as in Fig. 5.13.

Blobs and Flory exponent

To make the quantitive sense of the effect of concentrations on chain statistics (See Sec.4.3.5), we have considered the mean-square internal distances $< R^2(l) >$ as a function of an effective ring contour distance, $n_{eff} = n(1 - n/N)$ in different concentrations for the chains with 1200 monomers. The chains in our study are randomly branched rings,($\mu = 0.0 k_b T$) and exhibit hard excluded volume interactions with $\nu_k = 20 k_b T$. To modify the concentration while maintaining a fixed box size under periodic boundary conditions, we increased the number of polymer chains in the systems. The dilute systems consisted of a single isolated chain ($\rho = 10\%$), while the most densely packed systems contained 12 chains of equal lengths ($\rho = 93\%$).

The results are shown in Fig. 5.19. We rescaled the internal distances $< R^2(n) >$ by the length of each blob ξ^2 and ring contour distances, n, by the number of monomers in each blob, g, (See Sec.4.3.5). g and ξ for the chains in different concentrations are reported in Table 5.2. The results show the crossover from a dilute system to the melt case.

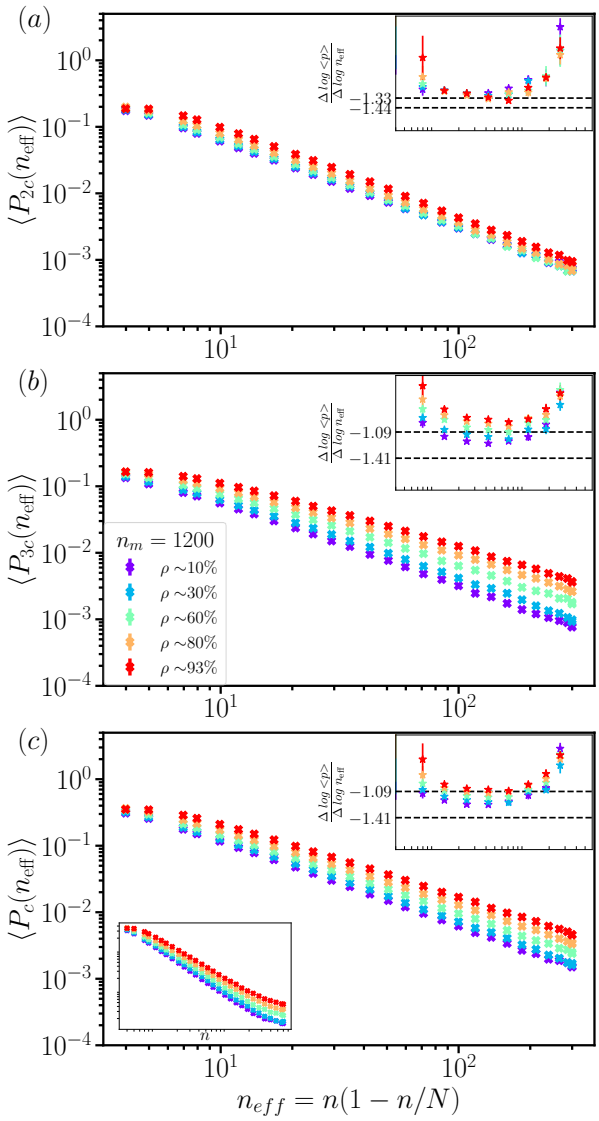


FIGURE 5.18: The concentration effect on the contact probabilities of the double-folded self-avoiding trees with length 1200. All the parameters of the energy functions are fixed, $\nu_k = 30\, k_B T$, $\mu = 0, k = 0$. By increasing the number of chains from 1 to 12, the expected behavior that the data should move from a self-avoiding isolated tree to an effectively self-avoiding tree in the melt state is well supported by our data. In other words, the scaling exponents gradually change from $\langle p_{2c}(n) \sim n^{-1.41}$ and $\langle p_c(n) \sim n^{-1.44}$ to $\langle p_{2c}(n) \sim n^{-1.33}$ and $\langle p_c(n) \sim n^{-1.09}$. Notations and symbols are as in Fig. 5.13..

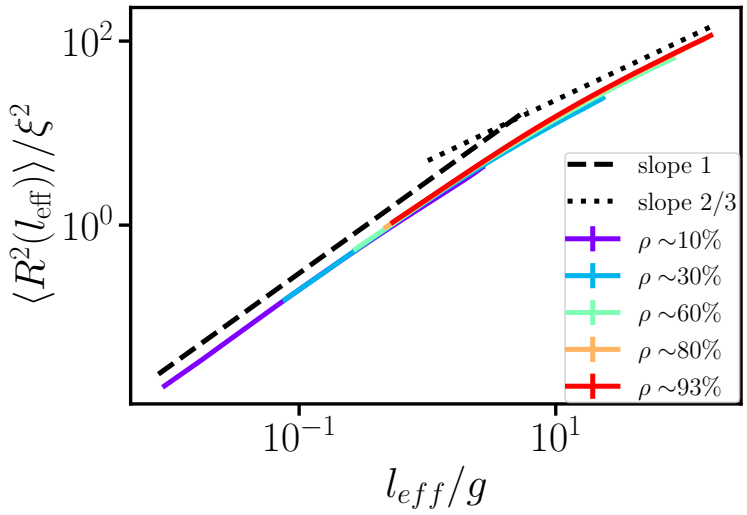


FIGURE 5.19: Mean-square internal distances, $< R^2(l) >$, verses the effective ring contour distance, $l_{eff} = l(1 - l/N)$. We rescaled the internal distances $< R^2(l) >$ by the length of each blob, ξ^2, and rescaled the ring contour distances, l_{eff}, by the number of monomers in each blob, g. This rescaling allows better exploration of the crossover from dilute to melt regimes. g and ξ for each chain lengths and different concentrations are reported in Table 5.2.

5.10 Summary and conclusion

In this chapter, we introduced our elastic lattice model for tightly double-folded ring polymers to explore the statistical properties of different systems (Sec.5.1 and 5.2). Then introduces the mathematical principles of our Monte Carlo simulations. The energy function has two parameters: 1) the chemical potential for the branch points, μ, in the energy term $\mathscr{H}^{br}$, which controls the asymptotic branching probability, λ; 2) parameter v_k manipulates the free energy penalty for overlapping pairs of tree nodes in the interaction part of the hamiltonian, $\mathscr{H}^{int}$. By changing these parameters, we can explore cross-over between different systems (Sec.5.3-5.5). In this chapter, we demonstrated how manipulating these two parameters allows us to investigate a wide range of systems, ranging from linear to highly branched structures, and from ideal chains to self-avoiding trees. By varying the branching probability and the chemical potential, we were able to explore and understand the behavior of different systems in a comprehensive manner.

Specifically, we performed Monte Carlo simulations and studied systems belonging to different universality classes: ideal double-folded rings without excluded volume interactions, self-avoiding double-folded rings, and double-folded rings in the melt state. The observed static properties are in good agreement with the exact results and predictions of Flory theory for polymer chains. For example, (double-folded) rings adopt compact configurations in the melt state and exhibit territorial behavior. In particular, we published our model, and in that article, we showed that the emergent dynamics are in excellent agreement with a recent scaling theory. We illustrated the qualitative differences with the familiar reptation dynamics of linear chains (Ghobadpour et al., 2021).

Moreover, we have illustrated how the chemical potential μ in our hamiltonian can be adjusted to control the branching probability and achieve chains with specific branch lengths. By tuning the value of μ, we have the ability to modulate the degree of branching in the system. This capability provides us with a powerful tool to design trees with specific branch lengths.

In the next chapter, we apply our coarse-grained model to study bacterial DNA, which is known to adopt tree-like plectonemic structures due to negative DNA supercoiling. To accurately capture the contact properties between chromosomal loci, as determined by high-throughput chromosome conformation capture methods (Hi-C), we fine-tune the average length parameter in our model. This allows us to closely reproduce the experimental observations and gain insights into the structural organization of bacterial DNA at the chromosomal level.

5.11 Appendix A: Chain length effect

The simulations were carried out for chain lengths varying between 64 to 1000 monomers for the double-folded self-avoiding rings to study the chain length effect. Results for averaging $L(n)$ and $R^2(n)$ for the system under investigation over monomer pairs with identical ring contour distance,$n = |i - j|$, are shown in panels (a_1), (b_1) of Fig. 5.20. As expected, the results are ring size independent at small scales and reach a plateau on approaching the maximal ring contour distance of $n = N/2$. Then as mentioned in the Sec.5.9, data were plotted as a function of an effective ring contour distance, $n_{eff} = n(1 - n/N)$ to reduce the finite ring size effects sufficiently for a meaningful comparison with the expected power law relations.

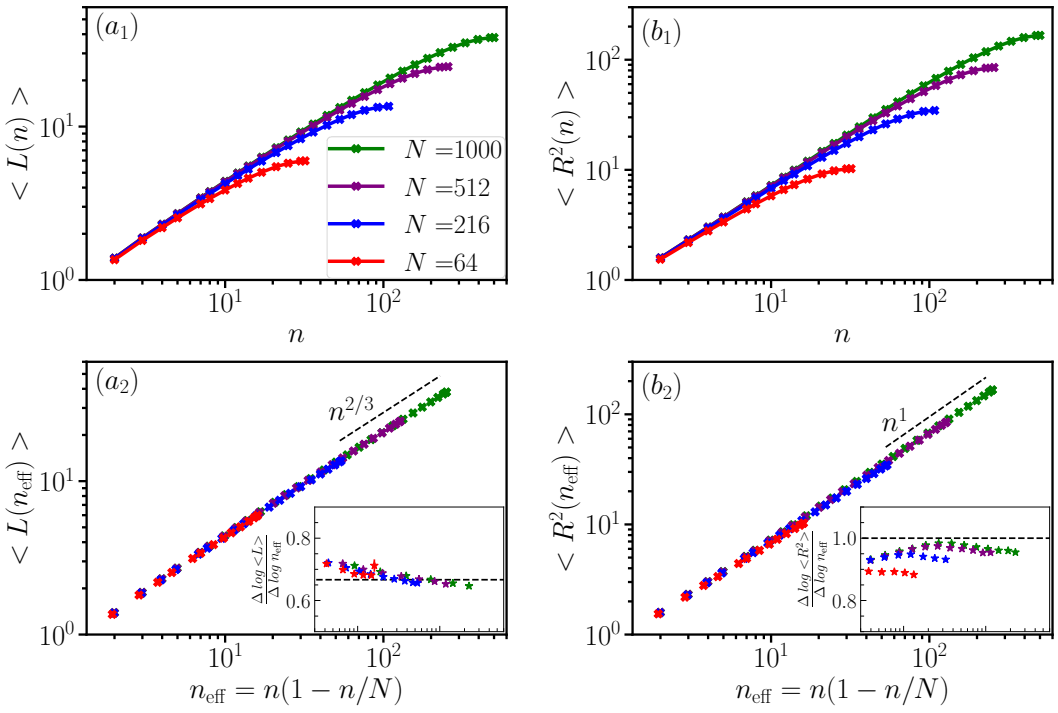


FIGURE 5.20: Conformational statistics of randomly branched self-avoiding double-folded rings for four different chain lengths (described in the legend). Data are shown for ring contour distances up to N/2. Column (a) are the average values of the tree contour distances between all possible pairs of monomers, $\langle L(n) \rangle \sim n^{\rho}$. Column (b) plots the squared internal distances as a function of n, $\langle R(n)^2 \rangle \sim n^{2\nu}$. The exact exponents for the ideal case are $\rho = 1/2$ and $\nu = 1/4$. In panels (a_2) and (b_2) data are plotted as a function of n_{eff}, which effectively reduces finite size effects. The straight dashed lines correspond to the expectation scaling exponents. (a_2) and (b_2) insets show the local slopes of the data in panels (a_2) and (b_2), respectively. These effective exponents appear to converge to the theoretical exponents (dashed horizontal lines). Error bars are the same size or smaller than the symbols.

5.12 Appendix B: Relation between ring contour length and tree contour length

In our elastic lattice polymer model, rings are tightly wrapped around the trees. The ring contour length, n_m, is related to tree contour length N_t via

$$n_m = N_t \times \langle \alpha \rangle, \tag{5.20}$$

where $\langle \alpha \rangle$ is the average number of monomers in each tree node. $\langle \alpha \rangle$ can be at least 2 to have double-folded bonds in each tree node. The term $\mathscr{H}^{c.l}$ in the energy function controls the tree contour length to fluctuate around the desired length, N_0. Fig.5.21 confirms that increasing the ring contour length at fixed N_0 does not affect the statistics . To compare these systems, we multiplied the ring contour distances by the average bonds, $\langle b \rangle$, in the system, (increasing the number of reptons caused a decrease in the average bond length).

The details of the systems are reported in Table.5.3. In order to study the supercoiled DNA, in the following chapter, we used $\langle \alpha \rangle = 2.6$.

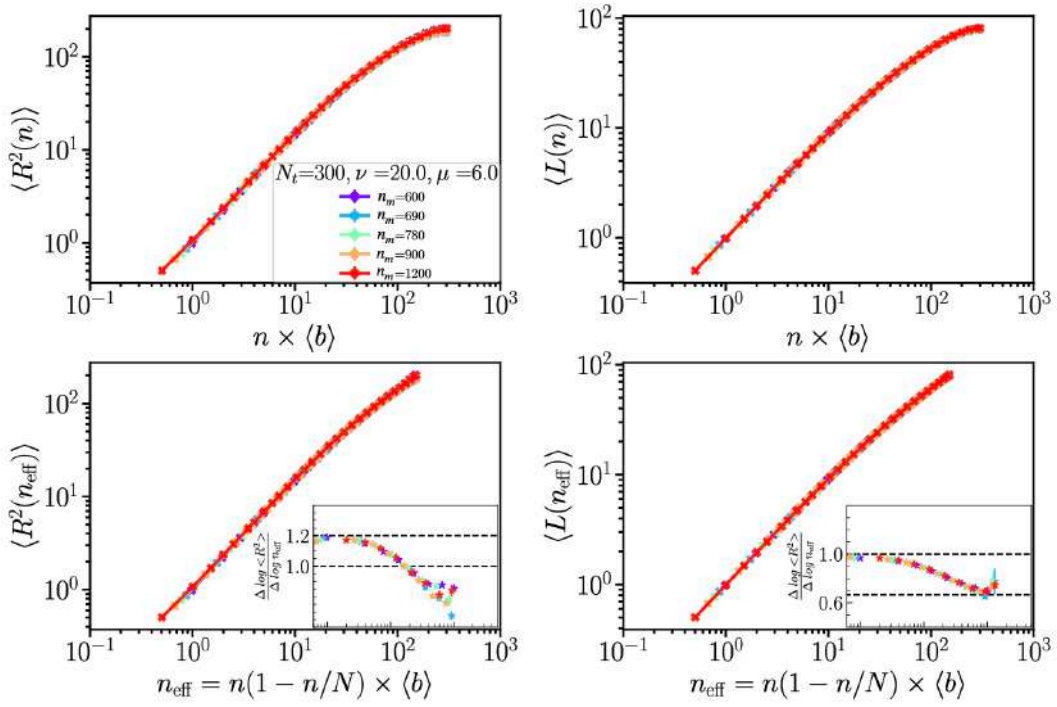


FIGURE 5.21: Statistical results for the systems with different ring contour lengths but the same tree contour length. It illustrates that increasing the ring contour length at fixed N_0 does not affect the statistics.

TABLE 5.3: A summary of the results for five studied systems in Fig.5.21 with different ring contour distances at the same tree contour length. $\langle\alpha\rangle$: the average number of monomers in each tree node; N_t: tree contour distance fluctuates around the desired contour length (in this case study, N_0=300); $\langle b\rangle$: average bond length, by increasing the ring contour length at fixed tree contour length, the number of "stored length" on the tree nodes increases, and the average bond length decreases.

$\langle\alpha\rangle$	N_t	$\langle b\rangle$
2.0	295.37 ± 0.06	0.98
2.3	299.75 ± 0.07	0.86
2.6	300.93 ± 0.06	0.77
3.0	301.87 ± 0.06	0.67
4.0	303.31 ± 0.07	0.50

Chapter 6

Coarse-grained models of supercoiled DNA at multiple scales

Coarse-graining approaches aim at reducing the number of degrees of freedom in a system for the sake of efficiency, so simulations of the coarse-grained systems require far less computing power than simulations of the fine-grained primary system. Consequently, both the simulation time and the size of the system can be substantially decreased. Thus coarse-grained models are helpful in their own right to study large-scale phenomena (Kremer and Müller-Plathe, 2002). However, sufficient details need to be preserved for the sake of accuracy.

As far as bacterial DNA is concerned, the *in vivo* internal structure of the bacterial chromosome remains unknown, with many potential molecular actors that may affect it. Inspired by the topological organization of negative supercoiled DNA, which adopts tree-like spatial conformations, I have investigated the possibility of capturing the main features of the internal structure of bacterial chromosomes using a double-folded ring polymer model presented in Sec.5. In this coarse-grained model of bacterial DNA, the branches represent plectonemes, while the actual size of these plectonemes *in vivo* is still uncertain. As a consequence, the mean size of the branches in the model is a parameter (the unique one, actually), and I investigate whether it is possible to capture the contact properties between loci as measured using Hi-C methods by testing different values of this parameter.

The chapter structure is as follows: In Sec.6.1, the values of the simulation parameters allowing the coarse-grained lattice double-folded ring chain to map onto supercoiled DNA are derived. A review of an additional coarse-graining procedure for tackling the problem of large-scale properties is presented in Sec.6.2. The conformational properties of bacterial DNA in three dimensions based on the observed contact probabilities extracted from Hi-C experimental data are analyzed in Sec.6.3. In Sec.6.3.2, the impact of the average branch lengths on the DNA conformational statistics is elucidated. In Sec.**??**, the possibility of tuning the average branch length parameter μ in order to reproduce as well as possible contact properties between chromosomal loci as obtained from Hi-C data

is investigated. Eventually, the results for different bacteria are summarized in Sec.6.4 before drawing some conclusions.

6.1 Simulation parameters

In order to build the double-folded ring polymer model to study chromosome conformation, the simulation parameters need to be determined. These parameters are the lattice constant, the amount of DNA contained in each tree node, and the system's density to mimic bacterial conditions. In the following, the procedures to get these parameters starting from a physical description of plectonemic DNA resulting from supercoiling constraints, are described. Note that to build the model, a constant and uniform value of the supercoiling density, $\sigma = -0.04$, is considered. The latter is a typical value of the "free" supercoiling (i.e., once the NAPs contribution is removed, see section 2.5.3) that has been reported *in vivo* (Bliska and Cozzarelli, 1987).

Lattice constant

Figure 6.1 shows a schematic mapping of a plectoneme with radius R and pitch p (i.e., the length of the superhelix is given by $2\pi p$) onto a lattice tree. The coarse-grained resolution is parametrized by the lattice constant, a, and the number of base pairs per tree node, m_t. The supercoiling density $\sigma = -0.04$ leads to $2R = 13$ nm, and $2\pi p = 92$ nm (detailed calculation can be found in Section 2.6). Since the radius is the smallest length scale, it defines the highest resolution of the double-folding coarse-grained representation of a plectoneme (later on, I will further use lower-resolution models by coarse-graining the resulting tree-like chain itself). In the following, I denote by a_0 this highest resolution so that $a_0 = 2R = 13$ nm. Importantly, *in vivo* DNA is bound by various protein complexes (e.g., transcription machinery) but also by NAPs (HU, IHF, H-NS,...). The effective diameter of the plectonemes bound by these proteins is thus expected to be more significant. In fact, previous experimental results have reported a higher-order chromatin-fiber-like structure whose diameter is on the order of 40 nm (Kim et al., 2004), in accord with the extra layer of proteins bound all over the plectonemes (Fig.6.2 (left)). Therefore, we eventually consider as the highest resolution this width of the fiber, so that the lattice constant is given by $a_0 = 40\,nm$ (Fig.6.2 (right)).

Number of base pairs per tree node

In the most general case, the number m_a of base pairs in each tree node given a mesh size a is given by:

$$m_a = \frac{2\int_0^{l_a} ds}{b}, \tag{6.1}$$

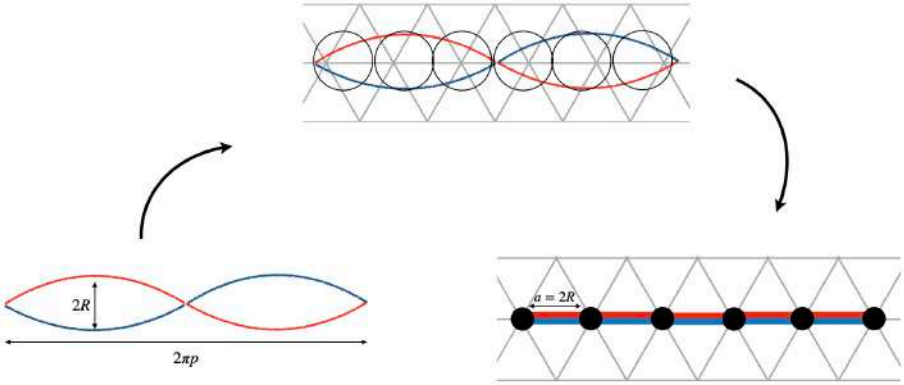


FIGURE 6.1: Schematic of mapping a plectoneme with radius R and pitch $2\pi p$ on a lattice tree. The plectoneme conformation is discretized by moving all the bp residing in a sphere to the center.

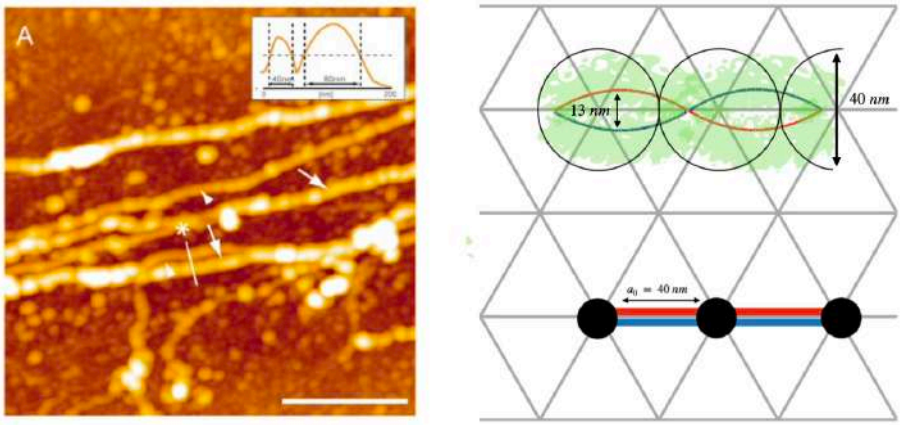


FIGURE 6.2: Left: The DNA and its binding proteins (HU, IHF, H-NS, ...) form a nucleosome-like structure, folded into a fiber with 40 nm width. Image is copied from (Kim et al., 2004). Right: Schematic mapping the 40 nm fiber on the lattice and moving all the bp residing in a sphere to the center, the plectoneme conformation is discretized.

where $b = 0.34$ nm is the distance between base pairs along the double-helix and l_a is the contour length of DNA along a demi super-helix in the resolution a. ds in the integral indicates the curvilinear abscissa along the DNA double-helix, and the factor 2 reflects the fact that there are two demi-super-helices in a super-helix. Noticing that $ds = \sqrt{R^2+p^2}d\theta$ where θ is the polar angle of the cylinder coordinate, such that $l_a = \sqrt{R^2+p^2}a/p$, Hence:

$$m_a = \frac{2a\sqrt{R^2+p^2}}{bp}. \tag{6.2}$$

Which is equal to $\simeq 257$ bp at the highest resolution.

The number of tree nodes N_t, to simulate a DNA with length L_{DNA}, (during slow growth, *E. coli* typically has one chromosome, $L_{DNA} \approx 4.6 \times 10^6$ bp) is

TABLE 6.1: Summary of simulation parameters for the model with the highest resolution in order to study a supercoiled bacterial DNA via a double-folded self-avoiding lattice tree. L_{DNA}: Length of the DNA in the unit of bp; a: FCC lattice unit length; m_t: number of base pairs on each tree node in the unit of bp; N_t: number of tree nodes required to simulate a DNA with length L_{DNA}; n_m: number of monomers correspond to N_t tree nodes; $\langle\alpha\rangle = 2.6$: average number of monomers in each tree node; τ_{max}: relaxation time to reach equilibrium in the unit of Monte Carlo sweep [MCs], which scales as $N_t^{2+\rho}$, where $\rho = 5/9$ for the tree in the melt state. For a DNA with supercoiled density, $\sigma = -0.04$, the radius is equal to 13 nm, and the pitch is equal to $2\pi p = 92$ nm. Considering the proteins that bind to DNA, we eventually have a plectoneme diameter equal to $a_0 = 2R = 40$ nm.

Parameters	**Formula**	E_{coli}	$200\,kbp$
L_{DNA} [bp]	L_{DNA}	4.6×10^6	$200\,kbp$
a [nm]	a_0	40	40
m_t [bp]	$\frac{2a_0}{0.34\times p}\sqrt{R^2+p^2}$	257	257
N_t	$\frac{L_{DNA}}{m_t}$	1.8×10^4	780
N_m	$\langle\alpha\rangle \times N_t$	5.4×10^4	1800
$\tau[MC_{sweep}]$	$(N_t)^{2.55}$	1.1×10^{12}	2×10^8

$$N_t = \frac{L_{DNA}}{m_a} = L_{DNA} \times \frac{bp}{2a\sqrt{R^2+p^2}}. \tag{6.3}$$

In this context, the number of monomers, N_m, to build up the ring, which is tightly wrapped around the tree, is expected to be given by:

$$N_m = \langle\alpha\rangle \times N_t \tag{6.4}$$

Where $\langle\alpha\rangle$, the average number of monomers per node, is chosen i) to be not too large to avoid time-consuming simulations and ii) to have typically more than 2 monomers per node in order to handle properly the Monte-Carlo moves, in particular the creation of new branches (see Sec.5.2 in chapter5). In this manuscript, I discuss the case where $\langle\alpha\rangle = 2.6$. Simulation parameters for the model with the highest resolution are summarized in Table 6.1.

Density

In order to carry out the simulations under the bacterial DNA conditions, the ring polymers are compacted within a rod-shaped geometry representing the nucleoid. Therefore, the parameters that govern the nucleoid's DNA concentration must be determined.

Considering a nucleoid as a cylinder of length L_n capped by two semi-spheres of radius $R_n = 400$ nm, the volume of the nucleoid is given by:

$$V_{Nucleoid} = \frac{4}{3}\pi R_n^3 + \pi R_n^2 (L_n - 2R_n) \tag{6.5}$$

Remarkably, the length L_n has been shown to follow the length of the cell L according to $L_n \simeq 0.51 \times L^{1.18}$ (Junier, Boccard, and Espéli, 2014) Fig.6.3 (B). The cell lengths range from $\sim 2\mu m$ at birth to $\sim 4.6\mu m$ at the division. In the following, we use $L \sim 2.4\mu m$ so that the nucleoid volume $V_{Nucleoid} \sim 7.8 \times 10^8\ nm^3$.

Next, from the perspective of the model, the volume of the primitive unit cell in an FCC lattice is defined by $v_{fcell} = \frac{\sqrt{2}a^3}{2}$ (Fig.6.4), where a is the lattice constant (nearest neighbor distances). Therefore the total number of primitive unit FCC cells to fill the whole volume of the nucleoid will be:

$$N_{fcells} = \frac{V_{Nucleoid}}{\frac{\sqrt{2}a^3}{2}}, \tag{6.6}$$

By substituting the above value of $_{Nucleoid}$ and using $a = 40\ nm$ (highest-resolution model), we obtain:

$$N_{fcells} \sim 17500. \tag{6.7}$$

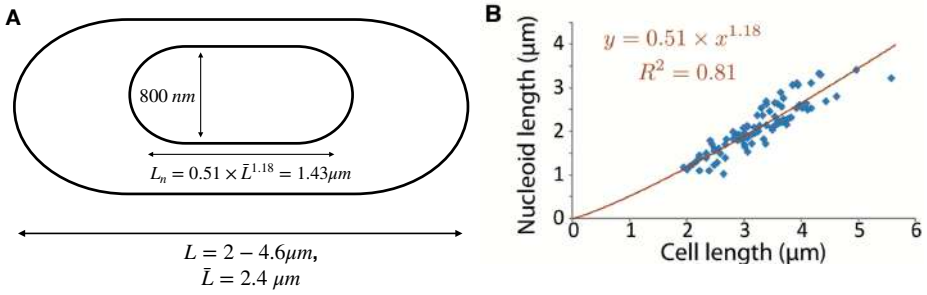


FIGURE 6.3: (A) Bacterial cell and nucleoid dimensions. (B) Nucleoid length as a function of the cell length of the *E.coli* chromosome. Graph copied from**??**.

During slow growth conditions, *E. coli* typically has one chromosome, $L_{DNA} \approx 4.6 \times 10^6$ bp. Since the number of base pairs per tree node at the highest resolution is around 257 bp, the number of tree nodes to model the whole DNA is equal to $N_t = \frac{L_{DNA}}{257} \simeq 18000$. Therefore, the occupation number, which is the ratio of the number of occupied lattice sites (tree nodes) to the total number of sites in the simulation, is $\frac{N_t}{N_{fcells}} \sim 1$. Based on this calculation, **trees must be simulated in the melt state to mimic the bacterial condition**.

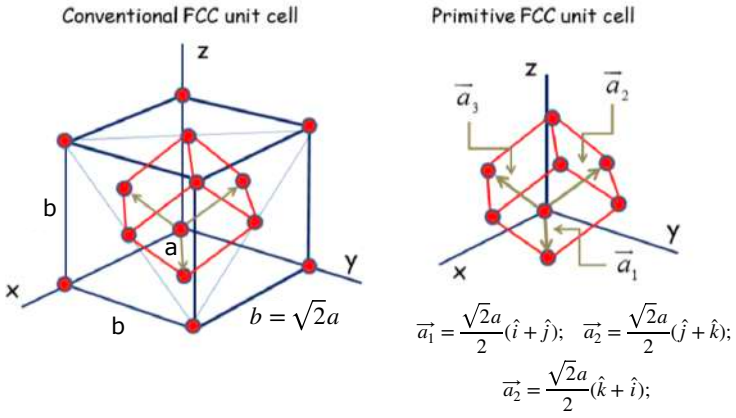


FIGURE 6.4: Left: conventional FCC unit cell. Right: primitive FCC unit cell. Volume of the primitive unit cell is $V_{cell} = \frac{b^3}{4} = \frac{\sqrt{2}a^3}{2}$.

6.2 Coarse-graining further the tree-like models: an effort to capture large-scale structuring properties of bacterial chromosomes

The simulation of large-size systems is computationally expensive. Given our simulation method and computational power, we can simulate systems consisting of 100 kb DNA in up to one day. The computational cost increases substantially with an increase in the DNA length in the system due to the required computational power for the equilibration step. Therefore, additional coarse-graining methods are required to study chromosomes at full genomic lengths (e.g., *E.coli* DNA with length ~ 4.6 Mbp).

Specifically, in our highest-resolution model, each tree node contains 257 bp DNA. The number of required tree nodes to model the whole *E. coli* is therefore $\sim N_t = 18000$. In the melt case, the relaxation time for a polymer chain is $\tau_{relax} = N_t^{2+\rho}$, where $\rho = 5/9$ is the exponent describing the scaling behavior of the mean tree contour distances (Sec.4.3.1 and Sec. *IV.C* in (Ghobadpour et al., 2021)) in the melt state. As a result, the relaxation time associated with the whole *E. coli* chromosome is given by:

$$\tau_{relax}[MC_{sweep}] = 18000^{2.55} \sim 7 \times 10^{10} \tag{6.8}$$

One MC sweep is equivalent to one MC trial for every tree node in the system (1 MC sweep $= 18000$ trials to study *E.coli* at full genomic length). The average CPU wall clock time for one MC sweep in this system is around ~ 0.018 second on a 3.3 GHz CPU. Achieving equilibrium for the whole *E. coli* DNA would occur after ~ 3.9 years. In other words, coarse-graining further the bacterial DNA is a crucial step in studying the large-scale structure of an entire bacterial chromosome.

6.2.1 Coarse graining method

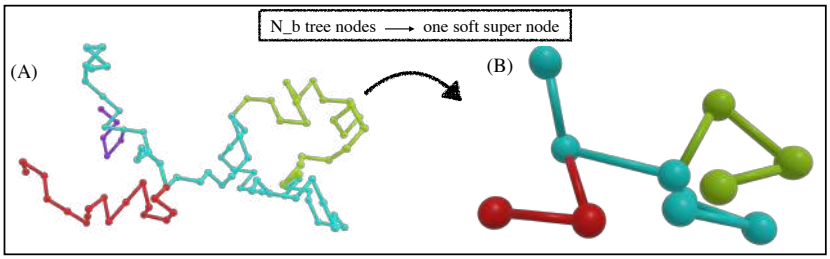


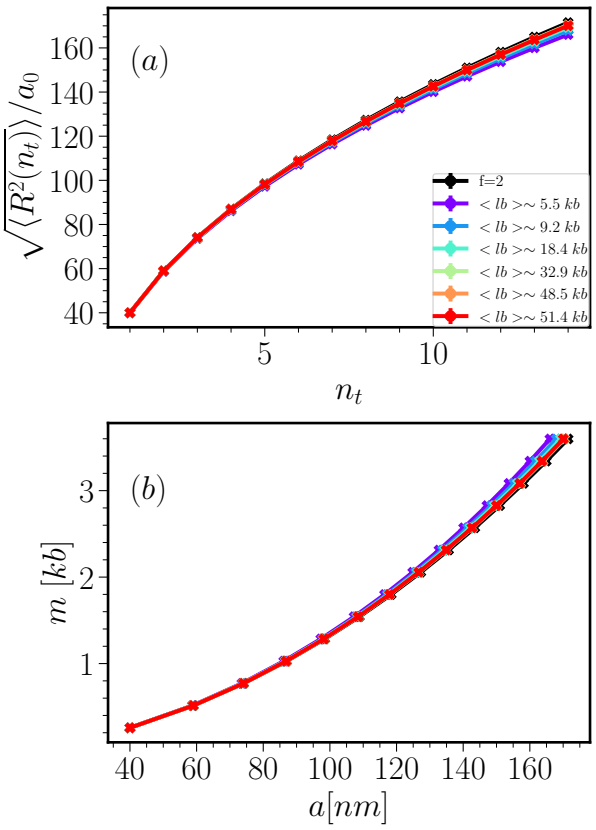
FIGURE 6.5: Schematic of the process of coarse-graining by replacing some tree nodes with one super node. (A) A configuration of the double-folded lattice tree with 90 tree nodes (highest-resolution model). (B) A coarse-grained tree with far fewer super nodes. Contrary to the highest-resolution model, the super nodes have weak excluded volume interaction as they can overlap with each other.

Coarse-graining methods vary substantially (Kremer and Müller-Plathe, 2002). No "standard" coarse-graining procedure is known to provide a zoomed-out description of a system. One way of achieving a coarse-grained mapping of our system into a less detailed level consists in grouping tree nodes into super nodes as illustrated in Fig.6.5. More specifically, the additional coarse-graining consists of the following steps:

1. Defining the level of coarse-graining. Let remind that at the highest resolution, the lattice size $a = 40$ nm. I then consider two additional levels corresponding to double ($a = 80$ nm) and quadruple ($a = 160$ nm) the lattice size, respectively.
2. Computing the amount of DNA per coarse-grained super-node, reminding that at the highest resolution, we have 257 bp per node. To this end, we multiply 257 by the average number of nodes found in a super-node.
3. Estimating the Hamiltonian parameters for the model to reproduce the statistical properties of the model at the immediately lower resolution.

Defining super nodes To compute the number of nodes per super-node, we use the curve of the root-mean-square of spatial distances as a function of the tree node distance n_t between any pair of nodes (Fig.6.6 (a))– the tree distance between a pair of tree nodes is defined as the shortest path along the tree. For a given new coarse-grained value a_{new} of the lattice site, the number of nodes in a super-node is then given by the value of n_t such as the root-mean-square of the spatial distances is equal to a_{new}, leading to the curve Figure6.6 (b). As a result, for $a = 80$ nm and $a = 160$ nm, one super node respectively contains 3.5 and 12.5 tree nodes of the highest-resolution model. Given these numbers, which were computed using a 200 kb long DNA, the chain size of the whole chromosome for $a = 160$ nm would be around ~ 1440, and the time to reach equilibrium would decrease from ~ 3.9years to ~ 28 days. Note that since the system is in the melt state and the trees adopt territories at all length scales, these results are

independent of the average branch lengths as soon as these are much larger than the coarse-grained unit (Figure6.6).



Model	a [nm]	n_t	m [bp]
First level Coarse-Graining (CG1)	40	1	257
Second level Coarse-Graining (CG2)	80	3.5	900
Third level Coarse-Graining (CG3)	160	12.5	3100

FIGURE 6.6: Defining super nodes for 200 *kb* long chains, with different average branch lengths as well as for a linear DNA with length 200 kb (i.e., for which $f = 2$). (a) The root-mean-square of spatial distances in the unit of lattice constant $a_0 = 40\,nm$ between all pairs of tree nodes (i, j), up to $n_t = |i - j| = 15$; (b) Average amount of DNA as a function of lattice size. The results for the three levels of coarse-graining are summarized in the bottom table.

Estimating Hamiltonian parameters. A good coarse-grained model preserves just enough information to reproduce certain aspects of the system under the corresponding conditions. In other words, the coarse-grained model should reproduce the statistical properties of the referred system.

The total number of tree nodes and the average branch lengths were reduced by $3.5/12.5$ for the second/third level of coarse-graining (see above), so the corresponding parameters, μ (will be derived in Sec.**??**), and N_0 can be directly implemented in the Hamiltonian definition. μ is the Hamiltonian chemical potential to control the branching probability, and N_0 is the equilibrium tree contour length of the DNA. Next, the ν_k (strength of the interaction potential) parameter should be adjusted until the statistical properties of the higher-resolution model are reproduced satisfactorily. However, since the DNA is in the melt state, and the density is homogenous, DNA can be modeled in the second and third level with hard excluded volume interaction.

6.3 From Hi-C data to contact probability

Measurements based on the Hi-C method provide data in the form of pairwise interaction frequencies between specific genetic loci. Since our coarse-grained model does not include specific structuring, such as the formation of CIDs, we confront our modeling data to experimental data by comparing the average interaction frequencies as a function of the genomic distance separating loci, also known as the probability of contact and usually referred to as $P(s)$ where s is the genomic distance. In Fig. 6.7, I show the results for wildtype *P. aeruginosa* bacteria with logarithmically spaced bins, starting at $2kb$ and increasing by a factor 1.12: (2 kb, int(2 kb$\times$1.12), int(2 kb$\times 1.12^2$),...); Then, for each bin, the number of observed Hi-C interaction frequencies at that distance is averaged. Finally, $P(s)$ is normalized such that the integral of $P(s)$ over the range of distances is one – note that Hi-C frequencies are not absolute frequencies but measured with respect to an unknown factor.

6.3.1 A ubiquitous two-regime scenario for the bacterial $P(s)$

$P(s)$ for *P. aeruginosa*, *E. coli*, and *C. crescentus* at different conditions and growth phases are plotted in Fig.6.8. *P. aeruginosa* and *E. coli* Hi-C data have 1-kb resolution and *C. crescentus* Hi-C data have 10-kb resolution.

The results reveal that all bacterial chromosomes have similar internal organization, independent of the bacterial type and length. They display a slow decrease in contact probability $P(s) \sim s^{-0.5}$ from 1kb to 100kb, followed by a rapid fall-off $P(s) \sim s^{-1.1}$ at $\sim$ 100kb. These features are observed for all these three bacterial chromosomes irrespective of the cell type, the details of the Hi-C interaction frequencies, and even the details of Hi-C methods since the data were obtained from different labs. This feature suggests that a 100-kb is a fundamental scale of chromosome structuring in these bacteria, as noticed in the first reported Hi-C matrix obtained in *C. crescentus* (Le et al., 2013).

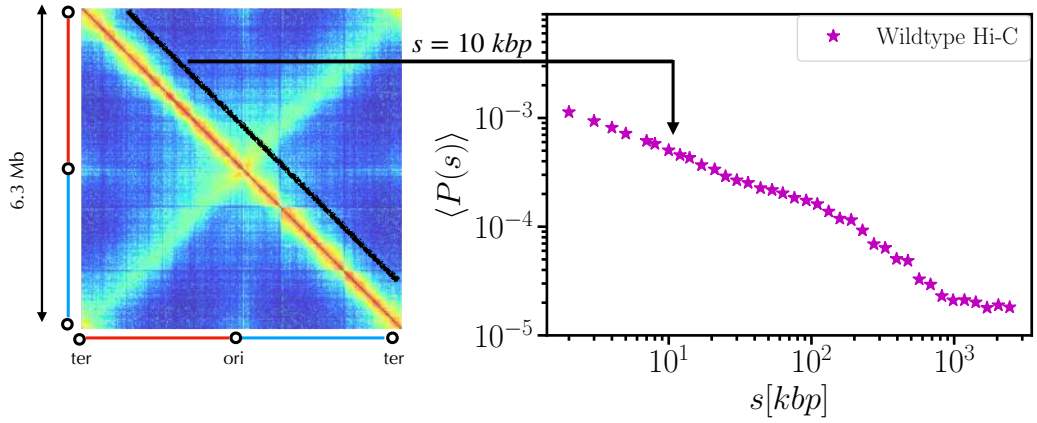


FIGURE 6.7: Contact probability as a function of genomic distances is obtained by averaging normalized Hi-C interaction frequency matrices, left: An example of a wild-type HiC matrix for *P. aeruginosa*. Right: Contact probability as a function of genomic distances. To compare experiments here and below, all $P(s)$ plots are normalized to integrate into one.

The two regimes observed in $P(s)$ suggest that the chromosome is organized differently above and below 100 kb. Above 100 kb, the decay with an exponent -1.1 is reminiscent of the crumpled globule picture where the melt state rings adopt compact configurations and exhibit territorial behavior (Sec.**??**) (Rosa and Everaers, 2014). Below 100 kb region, one needs to invoke a different mechanism since loci interact much more frequently with each other as indicated by the exponent -0.5 (Fig.6.8).

To understand the bacterial chromosome organization within a 100kb region, I compared the contact probability behavior observed with the randomly branched tree (where branches are of size one) and a linear chain without any branch. The linear chain displays a plateau, $P(s) = C^{te}$ (Fig.6.9). In contrast, as expected, the randomly branched tree exhibits territorial regions and behaves as $P(s) \sim s^{-1.1}$. The results for $P(s) \sim s^{-0.5}$ thus indicate a behavior in-between these two distinct polymer states (Fig.6.9). The idea, then, is to investigate whether there exists an intermediate average branch size (between 1 and the length of the chain) such that the model can capture the two regimes of the bacterial contact probability.

6.3.2 Effect of the average plectoneme length

To quantitatively test the assumption that the average plectoneme length is the length scale where the crossover between two regimes in $P(s)$ takes place, I performed simulations with $L = 50 - 100 - 200$ kb long DNA molecules with different plectoneme lengths, $\langle l_{lin} \rangle$. The average branch length in the model is adjusted by changing the value of the chemical potential, μ, in the Hamiltonian, which controls the branching probability. The results are displayed in Fig. 6.10. The gray curve represents $P(s)$ for *P. aeruginosa* as a reference, while the black curve corresponds to the simulation results for highly branched

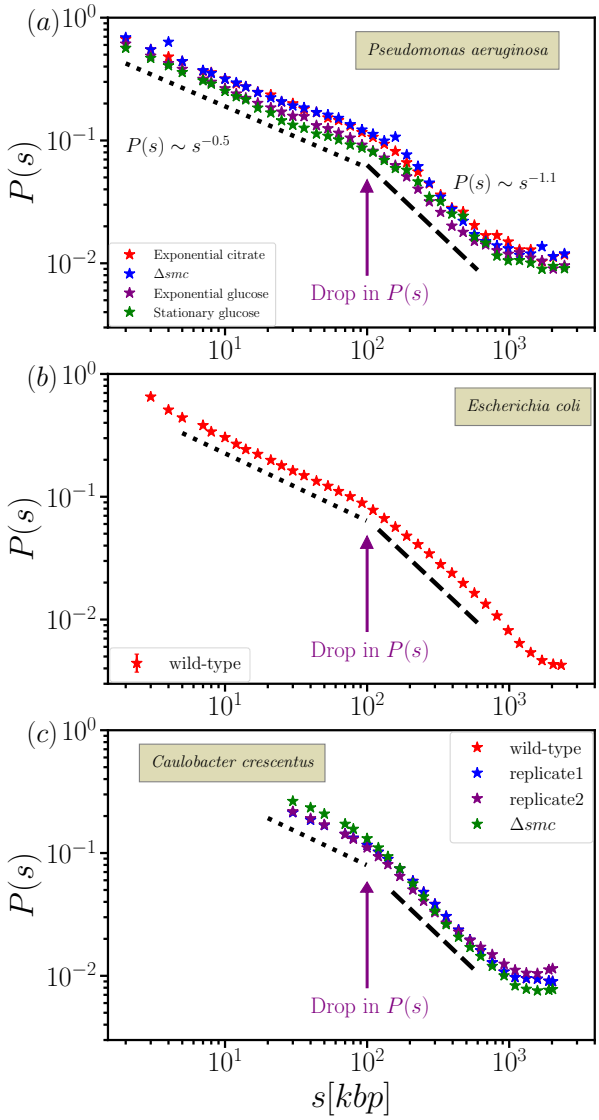


FIGURE 6.8: Contact probability as a function of genomic distance for three different bacterial chromosomes. The results show that the contact probability behavior is highly consistent among bacterial cell types. The three bacteria display a $\sim s^{-0.5}$-like decay for loci separated by fewer than $\sim$ 100kb and a $\sim s^{-1.1}$-like decay for loci separated by more than 100kb.

chains with an average branch length of $\langle l_{lin} \rangle \sim 650$ bp, also for reference. In addition, the data are plotted as a function of an effective genomic distance $s_{eff} = s(1 - \frac{s}{L})$, introduced in Sec.4.3.1 to reduce the finite ring size effects so that to have a meaningful comparison with experimental data. Fig.6.10 reveals that increasing the average branch length leads to a later drop in $P(s)$, which indeed occurs at the average branch length, i.e., the average plectoneme length.

I also extracted the effective exponents by calculating the derivatives using the logarithm of successive data points, $(\frac{\Delta log \langle P(s_{\text{eff}}) \rangle}{\Delta log\, s_{\text{eff}}})$. Results are shown in the insets of the panels in Fig. 6.10. The horizontal lines show the observed scaling exponents in the experimental

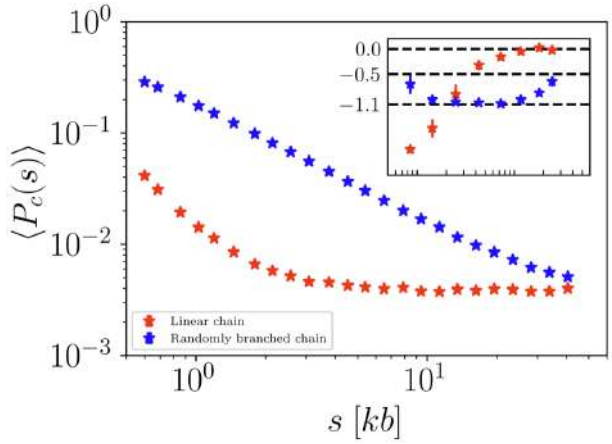


FIGURE 6.9: Left: DNA contact probability $P(s)$, below 100kb, plotted for randomly branched trees and linear chains. DNA's observed exponent falls between a linear state, where the contact probability is independent of the genomic distances, and the randomly branched tree, where regions are spatially segregated. Right: Equilibrated simulation snapshots of (randomly selected) configurations of the DNA with $L = 100$ kb on each state. Successive segments are represented with an HSV cyclic color map.

data. The effective exponents appear to converge to the experimental exponent -0.5 by increasing the branch length.

Next, we conducted rescaling of the results for the 200 kb long chains with different branch lengths. This was achieved by dividing the genomic distances by their respective branch lengths. Furthermore, we rescaled the experimental data for *P. aeruginosa* using 100 kb lengths, which is likely the length scale associated with plectonemes. The rescaled data was then presented in Figure 6.11. To enhance the clarity of the rescaled results, we normalized all the data points relative to the contact probabilities at the length scale equivalent to the branch length. Additionally, following the methodology employed in this study, we plotted the slopes of the data points in inset graphs. The slope was calculated by determining the derivatives using the logarithm of neighboring data points. This approach provides a visual representation of the rate of change within the data. As shown in Figure 6.11, a crossover in the behavior of the data is observed, with the exponent shifting from -0.5 to -1.1. This transition signifies the relationship between branch length and the observed drop in contact probabilities.

Based on these findings, we can confidently conclude that the decrease at 100 kb is likely correlated with the average branch lengths.

Based on the findings presented in Figure 6.11, it is strongly indicative that the average branch size in our model should ideally be set to 100;kb in order to accurately reproduce the observed transition in $P(s)$. Consequently, I conducted an extensive search to identify the optimal model parameters that would correspond to this specific average branch length, denoted by μ. Figure 6.12 effectively illustrates how an increase in the chemical potential μ (accompanied by decreasing in branching probability) results in the longer

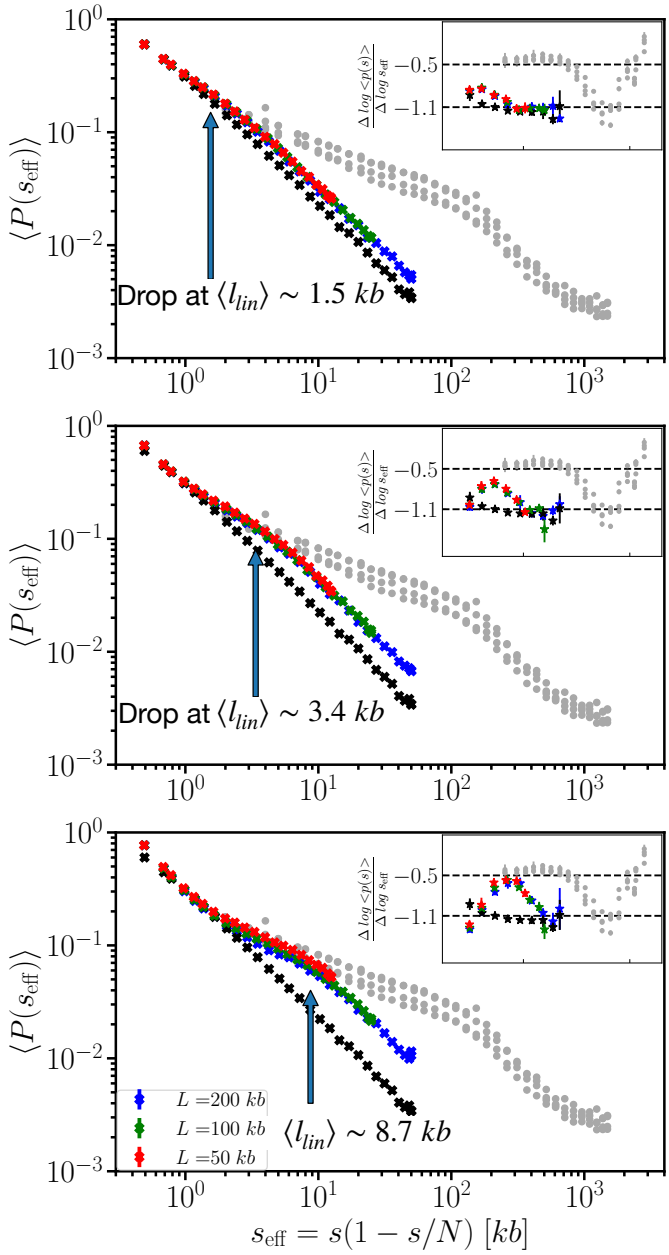


FIGURE 6.10: The effect of average branch lengths on $P(s)$ for DNA with $L_{DNA} = 50 - 100 - 200$ kb lengths. $P(s)$ for *Pseudomonas* bacteria are shown in gray as the reference data. Moreover, the data are plotted as a function of an effective genomic distance $s_{eff} = s(1 - \frac{s}{L})$, which effectively reduces finite-size effects. The insets show the local slopes of the data in corresponding panels. Increasing the average branch length migrates the position of the crossover between two regimes in $P(s)$. In other words, the average branch length is the scale where the drop in $P(s)$ takes place.

branch lengths. Notably, it discloses that setting μ to $10.5; k_BT$ and $8; k_BT$ at the second and third levels of coarse-graining aligns with an average branch length of 100;kb.

However, it is important to note that the first level of coarse-graining, which involves DNA lengths of $L = 50 - 100 - 200$;kb, fails to capture the two regimes evident in the

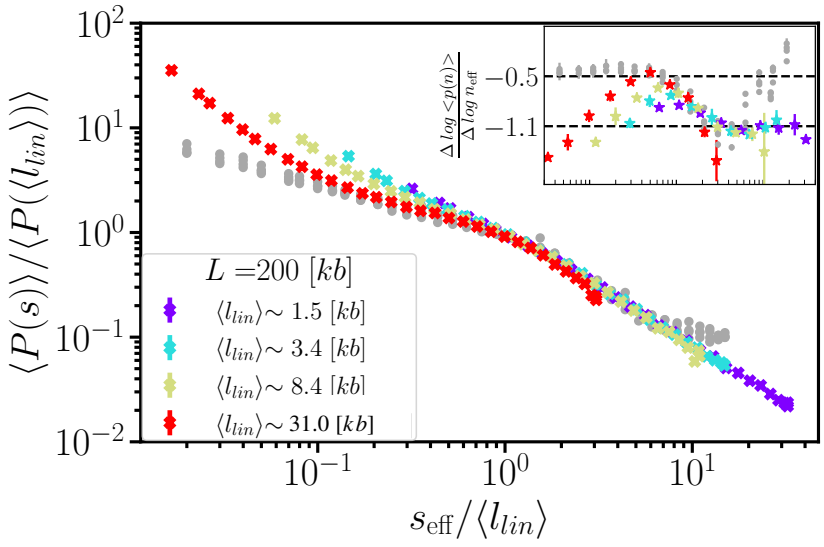


FIGURE 6.11: Rescaling of results for 200 kb long chains with varying branch lengths. Additionally, experimental data for *P. aeruginosa* is rescaled to 100 kb lengths, likely corresponding to the length scale associated with plectonemes. The insets depict local slopes of the data, which appear to converge with dashed horizontal lines representing experimental exponents. As evident in the graph, a notable decrease in $P(s)$ occurs at the average plectoneme length, strongly indicating that our model's average branch size must be set to 100;kb to faithfully reprduce the observed regime change in $P(s)$.

contact probabilities. Detailed simulation results and corresponding parameters for the first level of coarse-graining are concisely summarized in Table 6.3.

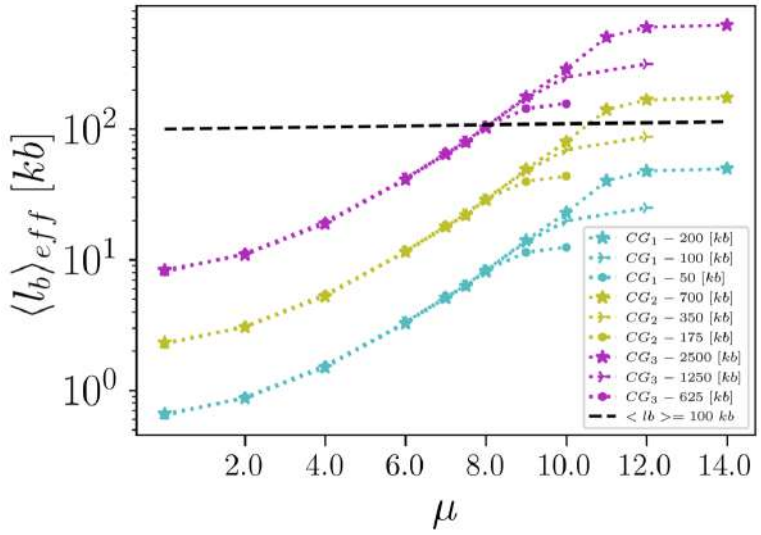


FIGURE 6.12: The quest for the optimal μ value, corresponding to the expected average branch length of 100kb, aimed at reproducing the observed scaling patterns in $P(s)$ as derived from the Hi-C matrix of bacterial genomes. For modeling the entire genome using the third level of the coarse-grained model, the parameter governing branching probability is set to $\mu = 8.0; k_B T$, while for the second level CG model, $\mu = 10.5; k_B T$ is selected. However, it's noteworthy that the first-level CG model falls short in capturing the crossover behavior in $P(s)$.

TABLE 6.2: Hi-C datasets.

C. crescentus from Laub group (Le et al., 2013)
Swarmer cells + $20\mu g/ml$ novobiocin for 30 min   At $0, 10, 30, 45, 60, 75$ min post-synchronization
P. aeruginosa lacking MukBESF from Boccard lab (Lioy et al., 2020)
E. coli from Koszul lab (Lioy et al., 2018)
B. subtilis from Runder lab (Wang et al., 2015)

6.4 Integrating the multiple scales: explaining Hi-C data of various bacteria in various conditions

In this section, I study other bacteria in various conditions and show that my modeling framework can systematically capture the probability of contact and predict the length of plectonemes. The list of Hi-C data investigated is summarized in Table 6.2.

P. aeruginosa As a first result, I summarize the results found above for *P. aeruginosa* in Fig. 6.13 where I chose the value of the parameters at each coarse-graining level to acquire an average plectoneme length equal to 100 kb. Again, note that to make a clear comparison with experimental data, the data are sketched as a function of s_{eff} to reduce finite-size effects effectively. Comparing the modeling results with the experimental ones (gray), one can see the excellent agreement in reproducing the experimental data's local (to some extent) and global behavior.

C. crescentus Hi-C data were recorded at 0, 10, 30, 45, 60, and 75 minutes post-synchronization (Le et al., 2013). The probabilities of contact remain identical, confirming again that a 100 kb length scale is fundamental in this bacterium (Le et al., 2013). Fig.6.14 (b) then shows an excellent agreement between experimental (in gray) and simulated (in purple) $P(s)$ for the third coarse-graining level, using an average branch length $\langle l_b \rangle \sim 100$ kb.

Next, *C. crescentus* bacteria were treated with $50\mu g/ml$ novobiocin for 30 min before drugs were washed away with plain media and then subjected to Hi-C analysis (Le et al., 2013) – novobiocin inhibits DNA gyrase and strongly reduces transcription elongation. The contact probability as a function of genomic distance for untreated wild-type swarmers versus novobiocin-treated samples is illustrated in Fig. 6.15 (a). As Fig. 6.15 (a) reveals, this drug modestly increases the interactions in the $200 - 800$ kb relative to the untreated wild type (Le et al., 2013). To model the effect of novobiocin, we increased

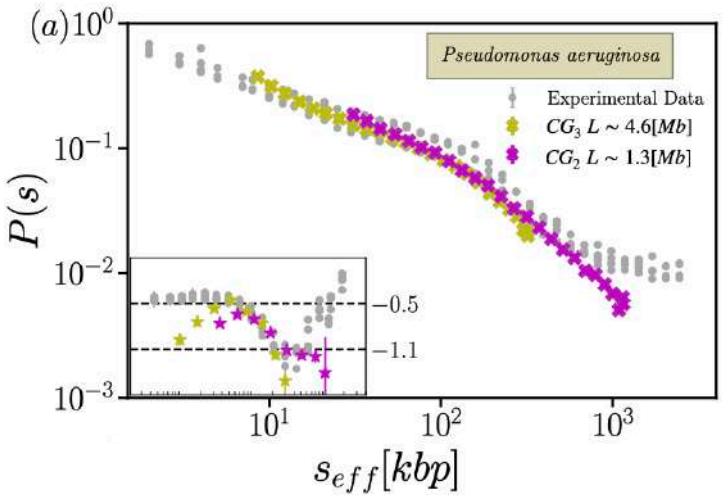


FIGURE 6.13: Contact probability for the two levels systematic coarse-grained models compared with observed $P(s)$ in *Pseudomonas* bacteria (gray, as in Fig. 6.8 (a)). Systematic coarse-graining procedures result in reproducing the whole experimental data behavior. The inset displays the local slopes of the data. The effective exponent appears to converge to the experimental exponents (dashed horizontal lines).

the branch length from 100 kb to 350 kb. As shown in Fig. 6.15 (b), the model could reproduce the observed experimental behavior very well.

E. coli Results for *E. coli* are shown in Fig.6.16 using the three coarse-graining levels with, again, an average branch length $\langle l_b \rangle \sim 100$ kb. I found the same behavior, where the systematic coarse-graining procedures reproduce the whole experimental data behavior, therefore confirming the ubiquitous behavior $P(s) \sim s^{-0.5}$ followed by $P(s) \sim s^{-1.1}$.

B. subtilis I finally applied the same procedure to investigate the situation in *B. subtilis*. The behavior of this bacteria is shown in Fig.6.17. In contrast to the observed property in the three other bacteria, the first regime extends to longer distances, with a crossover to the second regime occurring at a larger scale. Our coarse-grained double-folded model with $\langle l_b \rangle = 100$ kb branch length does not reproduce the experimental contact probabilities between pair loci in *B. subtilis* bacteria. A model with longer branches (approximately 500 kb in length) is required to accurately depict the behavior of this particular bacteria. The difference with the other bacteria suggests a different mechanism for the local folding of this bacterium.

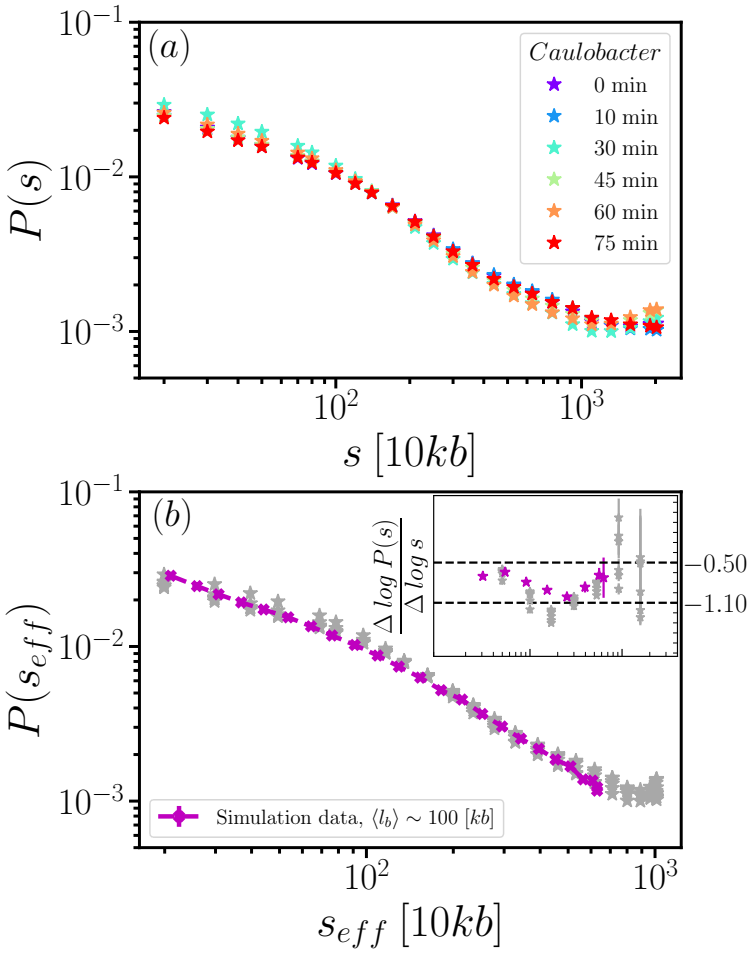


FIGURE 6.14: $P(s)$ of *Caulobacter* cells at various cell cycle stages. (a) Hi-C was performed at 0, 10, 30, 45, 60, and 75 minutes post-synchronization by Le et. al. (Le et al., 2013). Plots of chromosomal contact probability as a function of the genomic distance between restriction sites were reproduced. (b) results for cells at various cell cycle stages (grey) from panel (a) are compared to simulated contact probability data (purple). The simulated model is the third-level CG double-folded branch model with average branch length $\langle l_b \rangle \sim 100$ kb, reproducing the experimentally observed pair-wise contact probabilities. Panel (b) data are plotted as a function of s_{eff}, effectively reducing finite-size effects. Inset in panel (b) displays the local slopes of the data. The effective exponent appears to converge to the experimental exponents (dashed horizontal lines). Error bars are the same size or smaller than the symbols.

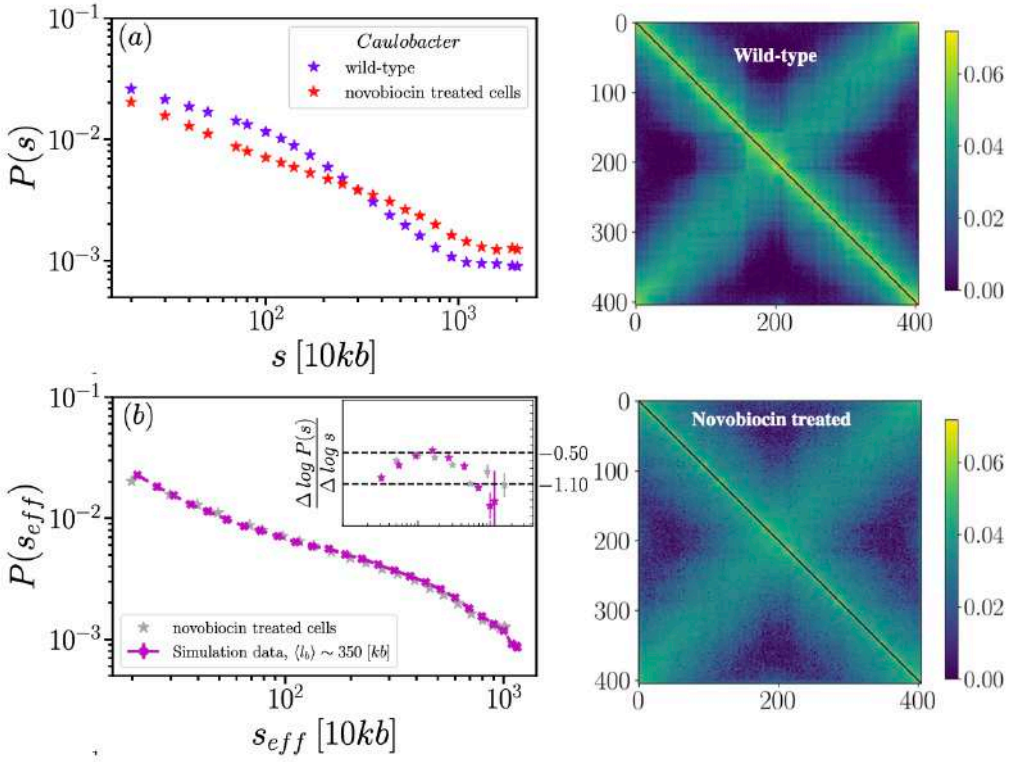


FIGURE 6.15: Polymer model of the chromosome upon gyrase inhibition by novobiocin. Top: Effect of novobiocin treatment on Hi-C data. (a) Comparison of contact probability between wild-type and gyrase inhibition data. (b) Comparison of contact probability between gyrase inhibition experimental data and the models. Gyrase inhibition models were modifications of a wild-type model with increased average branch lengths into $\langle l_b \rangle \sim 350$ kb. The model could correctly reproduce the behavior of experimental data. Notation and symbols are as in Fig.6.14.

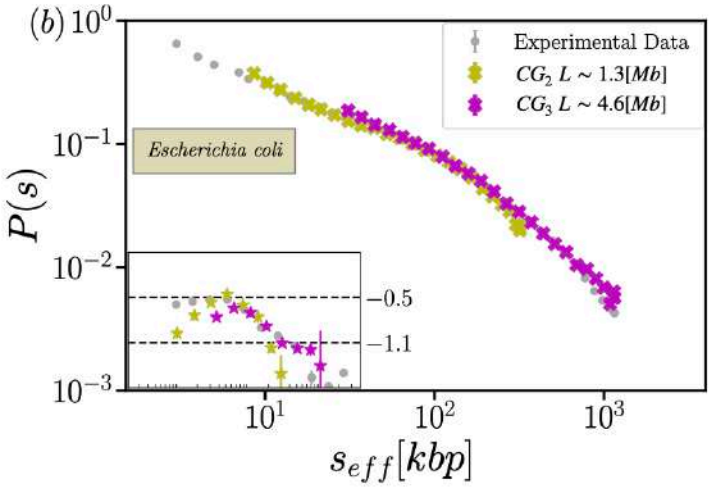


FIGURE 6.16: Contact probability for the two levels systematic coarse-grained models compared with observed $P(s)$ in *E.coli* bacteria (gray, the same data as in Fig. 6.8 (b)). Systematic coarse-graining procedures reproduce the whole experimental data behavior. The inset displays the local slopes of the data. The effective exponent appears to converge to the experimental exponents (dashed horizontal lines).

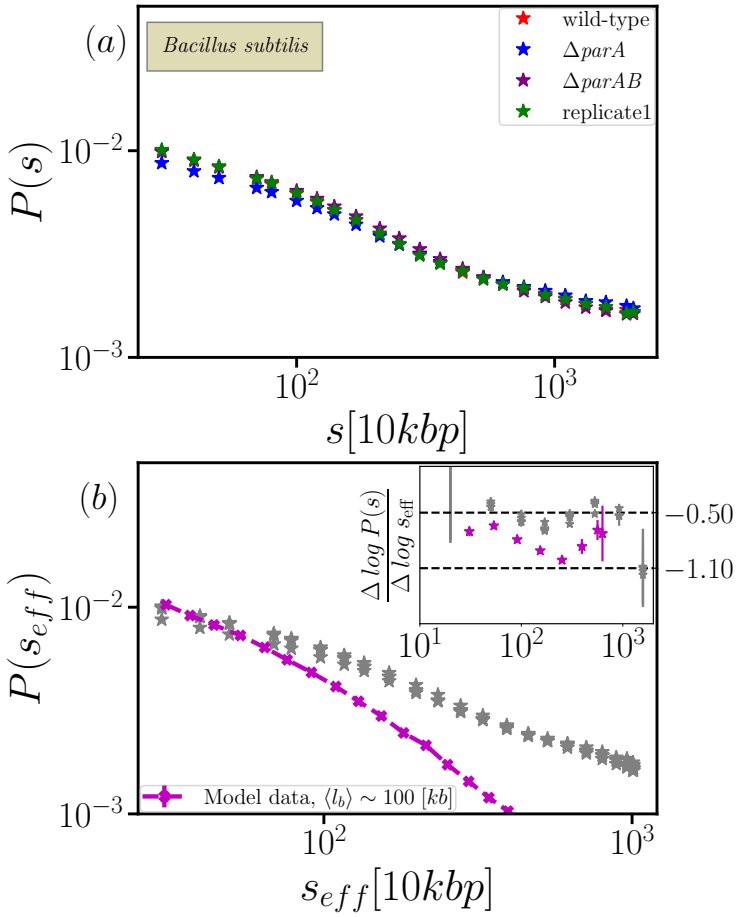


FIGURE 6.17: Top: Contact probability as a function of genomic distance for *Bacillus subtilis* bacterial chromosome. Bottom: The simulated CG double-folded branch model with average branch length $\langle l_b \rangle \sim 100$ kb can not reproduce the experimentally observed pair-wise contact probabilities, and the mechanism behind it is still unknown to us. Panel (b) data are plotted as a function of s_{eff}, effectively reducing finite-size effects. The inset displays the local slopes of the data.

6.5 Conclusion

In this chapter, we have elucidated the simulation parameters of a single coarse-grained lattice double-folded ring chain to model the DNA as a first step before addressing the bacterial DNA universal behavior. These simulation parameters are the lattice constant, the amount of DNA contained in each tree node, and the system's density to mimic the bacterial condition. By mapping a plectoneme and its binding proteins onto the FCC lattice, we found that for our model, the lattice constant is 40 nm, the amount of DNA that each tree node represent is 257bp and the actual density to mimic the bacteria condition *in vivo* is the melt state.

The wide range of time and length scales challenge our modeling approach. Therefore, we have investigated a recurring theme in the development of polymer models and simulations, which is the idea of coarse-graining, which, however, implies simplification: On the one hand, the underlying complexity of bacterial 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. To this aim, the number of tree nodes in the volume of the spheres with a radius equal to $2a_0 = 80$ nm and $4a_0 = 160$ nm, which corresponds to the second and third level of coarse-graining, respectively, were investigated. The result suggests the $3.5/12.5$-to-1 mapping. In other words, $3.5/12.5$ lattice tree nodes can be represented by one super node in the second/third coarse-grained models (Fig.6.6).

By investigating the contact probability as a function of genomic distances and examining the scaling behavior in $P(s)$, we gain further insights into the $3D$ organization of the DNA inside the nucleoid. Contact probability is obtained by averaging normalized Hi-C interaction frequency matrices and summarizing the results in a plot as $P(s)$ vs. genomic contour length. The results reveal that all bacterial chromosomes have similar internal organization. They display a slow decrease in contact probability $P(s) \sim s^{-0.5}$ from 1kb to 100kb, followed by a rapid fall-off $P(s) \sim s^{-1.1}$ at $\sim$ 100kb. These features are observed for all bacterial chromosomes, except that of *B. subtilis*, irrespective of the environmental conditions, the details of the Hi-C interaction frequencies, and even the details of Hi-C methods since the data were obtained from different labs. This feature suggests that 100-kb is a fundamental length scale of bacterial chromosomes (Fig.6.8).

Since branch length might be the key to the ubiquitous $P(s)$ behavior, I have investigated the branch length effect on the contact probability. Coarse-grained lattice double-folded ring polymer models were applied, which consist of $n_m = 507, 1012, 2024$ monomers with different average branch lengths to quantitively analyze the drop in $P(s)$ behavior. Upon increasing the average branch lengths, the drop in $P(s)$ migrates towards longer length scales. Our results confirm that regions separated by more than $\langle l_b \rangle$ occupy distinct spatial positions since they rarely interact, and they show the exponent $P(s) \sim s^{-1.1}$. However, loci within any $\langle l_b \rangle$ region interact with each other frequently and have the exponent $P(s) \sim s^{-0.5}$, (Fig.6.10).

By exhaustively testing the value of the chemical potential, μ, in the Hamiltonian, which controls the branching probability, I aimed to find the chemical potential corresponding to the observed branch length 100 kb. This massive search suggests that $\mu = 8\ k_BT$ and $\mu = 10.5\ k_BT$ can reproduce as well as the contact properties between chromosomal loci as obtained from Hi-C data by applying the second and third levels coarse-grained models, respectively (Fig.6.12).

As a result, I obtained various coarse-grained models that are consistent with each other and allow the capture of various bacteria's contact properties from 10 kb to 1 Mb scale. In other words, we are able to rationalize from first principles contact properties between bacterial chromosomal loci as measured from high-throughput conformation capture (Hi-C) methods (Fig.6.13-6.14-6.16).

Moreover, the effect of novobiocin is modeled by taking into account the average longer branches. Novobiocin inhibits DNA gyrase and strongly reduces transcription elongation. (Fig.6.15). Finally, the source of the different behavior from *Bacillus subtilis* raises two questions, 1) whether our approach is appropriate for this bacterium? and 2) is the nature of the local folding in this bacterium different?

In a nutshell, our double-folded ring polymer model is a very simple model based on a polymer physics background and without any unphysical assumption or extra restrictions, taking into account only the connectivity of the monomers on the tree and the ring topology in the dense systems. Our model stresses the impact of the branch lengths in the $3D$ conformations of the bacterial chromosomes, which might be deliberately exploited by nature (due to selective pressure) to drive proper biological functioning. Our model can provide theoretical insights to guide future experimental approaches that try to obtain information about the shape of the prokaryotic genome.

6.6 Appendix A: Simulation parameters for the DNA polymer models

TABLE 6.3: Simulation parameters for the polymer models of DNA with length 50 kb and 100 kb. Hamiltonian chemical potential; $\langle l_b(N)\rangle$: The average length of branches in the unit kb; λ: branching probability. L: DNA length in the unit kb, n_m: Number of monomers per chain; n_p= Total number of chains per simulated system; L_{box}: Size of the simulations box in the unit of nm. Note: An FCC lattice contains $4l_xl_yl_z$ sites; ρ: Lattice density which is the ratio of the number of occupied sites to the total number of sites; N_{sample}: Number of independent MC samples; T_{tot}: Simulation run time in Monte Carlo sweep $[MCs]$. All measurements are performed after reaching equilibrium.

$\mu\ [k_BT]$	$\langle l_b\rangle\ [kb]$	$\langle N_{f=3}(N_t)\rangle$	$\lambda\ \sim$
$L=50\,kb$, $\rho=86\%$,	$n_m=507$ $A=2.6$	$n_p=4$ $N_{sample}=800$	$L_{box}=(280,280,504)nm$ $T_{tot}=2\times10^8$
0	0.65 ± 0.01	62.3 ± 0.1	0.3190 ± 0.0007
2.0	0.89 ± 0.01	38.5 ± 0.1	0.1976 ± 0.0007
4.0	1.56 ± 0.02	18.48 ± 0.09	0.0951 ± 0.0005
6.0	3.56 ± 0.05	6.96 ± 0.06	0.0358 ± 0.0003
7.0	5.9 ± 0.1	4.03 ± 0.06	0.0207 ± 0.0003
8.0	10.6 ± 0.3	2.12 ± 0.04	0.0109 ± 0.0002
9.0	17.5 ± 0.5	0.98 ± 0.03	0.0051 ± 0.0002
10.0	24.0 ± 0.5	0.42 ± 0.02	0.0022 ± 0.0001
10.5	25.1 ± 0.5	0.27 ± 0.02	0.0014 ± 0.0001
11.0	26.7 ± 0.5	0.16 ± 0.01	0.0008 ± 0.0001
$N_t=100\,kb$, $\rho=90\%$,	$n_m=1012$ $A=2.6$	$n_p=4$ $N_{sample}=800$	$L_{box}=(336,336,672)nm$ $T_{tot}=2\times10^8$
0.0	0.64 ± 0.01	126.4 ± 0.2	0.3251 ± 0.0005
2.0	0.89 ± 0.01	77.7 ± 0.2	0.2000 ± 0.0005
4.0	1.56 ± 0.02	37.6 ± 0.1	0.0975 ± 0.0004
6.0	3.46 ± 0.04	15.1 ± 0.1	0.0390 ± 0.0003
7.0	5.41 ± 0.08	9.14 ± 0.09	0.0243 ± 0.0002
8.0	9.1 ± 0.2	5.14 ± 0.06	0.0133 ± 0.0002
9.0	16.3 ± 0.4	2.70 ± 0.05	0.0069 ± 0.0001
10.0	27.4 ± 0.8	1.45 ± 0.04	0.0037 ± 0.0001
10.5	44 ± 2	0.92 ± 0.07	0.0024 ± 0.0002
12.0	51 ± 1	0.22 ± 0.02	0.00056 ± 0.00005

TABLE 6.4: Free parameter search for the polymer model of DNA with 200 kb length. Notation and symbols are as in Tab.6.3.

$\mu\ [k_BT]$	$\langle l_b \rangle\ [kb]$	$\langle N_{f=3}(N_t) \rangle$	$\lambda\ \sim$
$N_t = 200\ kb$, $\rho = 87\%$,	$n_m = 2024$ $A = 2.6$	$n_p = 1$ $N_{sample} = 200$	$L_{box} = (280, 280, 504) nm$ $T_{tot} = 6 \times 10^8$
0.0	0.65 ± 0.07	253.6 ± 0.5	0.3258 ± 0.0006
2.0	0.87 ± 0.02	157.9 ± 0.5	0.2030 ± 0.0007
4.0	1.51 ± 0.03	77.0 ± 0.4	0.0991 ± 0.0006
6.0	3.32 ± 0.06	31.7 ± 0.3	0.0408 ± 0.0004
7.0	5.25 ± 0.09	19.4 ± 0.2	0.0250 ± 0.0003
8.0	8.6 ± 0.3	11.6 ± 0.2	0.0150 ± 0.0003
9.0	15.3 ± 0.5	6.5 ± 0.2	0.0084 ± 0.0002
10.0	26 ± 1	3.7 ± 0.1	0.0047 ± 0.0001
10.5	37 ± 2	2.8 ± 0.1	0.0036 ± 0.0001
11.0	56 ± 3	1.70 ± 0.08	0.0021 ± 0.0001
12.0	79 ± 4	1.04 ± 0.07	0.00134 ± 0.00009
14.0	122 ± 4	0.33 ± 0.04	0.00042 ± 0.00005

Chapter 7

Conclusions and Outlook

7.1 A short summary of the results

To answer one of the fundamental questions in biological studies on how the spatial conformation of the DNA affects its functions like gene expression, DNA replication, and segregation, we first need to understand chromosome conformation. Even though experimental research has come a long way, there is still not enough information to fully comprehend the structure and function of the genome. Computational models are a crucial complement to interpreting experimental data and providing quantitative comprehension of how chromosomes fold, move, and interact. At the end of the day, a genuinely integrative approach combining experimental data, such as high-resolution microscopy and Hi-C techniques, with computational/analytical modeling and polymer theory can deepen our current understanding of the spatial genome organization in bacterial cells.

The work presented in this thesis is inspired by a range of experimental observations for bacterial chromosomes. This thesis aims to develop a polymer physics model to study the $3D$ conformation of bacterial chromosomes (*P. aeruginosa*, *C. crescentus*, and *E. coli*) at different length scales. We do not aim to explain the specific patterns in the contact map like other theoretical studies but instead investigate the possibility of capturing the main features of the internal structure of bacterial chromosomes using our branched double-folded ring polymer model inspired by the topological organization of negative supercoiled DNA, which adopts tree-like structure.

Stage I: A lattice model for the dynamics of randomly branching double-folded ring polymers.

First, I introduced an elastic lattice model for tightly double-folded ring polymers, which allows for the spontaneous creation and deletion of side branches coupled to a diffusive mass transport, which is local both in space and on the connectivity graph of the tree. I performed Monte Carlo simulations and explored the statistical properties of different systems.

The energy function of the system has two parameters: 1) the chemical potential for the branch points, μ, in the energy term $\mathscr{H}^{br}$, which controls the asymptotic branching probability, λ; 2) parameter ν_k reflects the free energy penalty for overlapping pairs of tree nodes in the interaction part of the hamiltonian, $\mathscr{H}^{int}$. By changing these parameters, we can explore cross-over between different systems.

In particular, I studied systems belonging to three different universality classes: ideal double-folded rings without excluded volume interactions, self-avoiding double-folded rings, and double-folded rings in the melt state. The observed static properties are in very good agreement with exact results, simulations, and predictions from Flory theory for randomly branching polymers. For example, rings adopt compact configurations in the melt state and exhibit territorial behaviors.

Stage II: Coarse-grained models of supercoiled DNA at multiple scales

Second, I built a coarse-grained model of bacterial DNA, which is known to adopt tree-like plectonemic structures due to negative DNA supercoiling. To this aim, first, we have elucidated the simulation parameters of a single coarse-grained lattice double-folded ring chain to model the bacterial DNA. These simulation parameters are the lattice constant, the amount of DNA contained in each tree node, and the system's density to mimic the bacterial condition. By mapping a plectoneme and its binding proteins onto the FCC lattice, we found that for our model, the lattice constant is 40 nm, the amount of DNA that each tree node represent is 257 bp and the actual density to mimic the bacteria condition *in vivo* is the melt state.

The wide range of time and length scales challenge our capacity to simulate the model. Therefore, we have investigated a recurring theme in the development of polymer models and simulations, which is the idea of coarse-graining, which, however, implies simplification: On the one hand, the underlying complexity of bacterial 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. To this aim, the number of tree nodes in the volume of the spheres with a radius equal to $2a_0 = 80$ nm and $4a_0 = 160$ nm, which corresponds to the second and third level of coarse-graining, respectively, were investigated. The result suggests the $3.5/12.5$-to-1 mapping. In other words, $3.5/12.5$ lattice tree nodes (900 bp/3200 bp) can be represented by one super node in the second/third coarse-grained models.

By investigating the contact probability as a function of genomic distances and examining the scaling behavior in $P(s)$, we gained further insights into the $3D$ organization of the DNA inside the nucleoid. Contact probability was obtained by averaging normalized Hi-C interaction frequency matrices and summarizing the results in a plot as $P(s)$ vs. genomic contour length. The results reveal that all bacterial chromosomes have similar internal organization. They display a slow decrease in contact probability $P(s) \sim s^{-0.5}$ from 1kb to 100kb, followed by a rapid fall-off $P(s) \sim s^{-1.1}$ at $\sim$ 100kb. These features

are observed for all bacterial chromosomes, except that of *B. subtilis*, irrespective of the environmental conditions, the details of the Hi-C interaction frequencies, and even the details of Hi-C methods since the data were obtained from different labs. This feature suggests that 100-kb is a fundamental length scale of bacterial chromosomes.

In our model, the branches represent plectonemes. By exhaustively testing the value of the chemical potential, μ, in the Hamiltonian, which controls the branching probability, I aimed to find the chemical potential corresponding to the observed branch length 100 kb. An exhaustive search suggests that $\mu = 8\ k_BT$ and $\mu = 10.5\ k_BT$ can reproduce the contact properties between chromosomal loci as obtained from Hi-C data by applying the second and third levels coarse-grained models, respectively.

As a result, I obtained various coarse-grained models that are consistent with each other and allow the capture of various bacteria's contact properties from 10 kb to 1 Mb scale. In other words, we were able to rationalize from first principles contact properties between bacterial chromosomal loci as measured from high-throughput conformation capture (Hi-C) methods. Our model stresses the impact of the branch lengths in the $3D$ conformations of the bacterial chromosomes and suggests that 100-kb length scale regions are the fundamental unit of chromosome structure in many organisms.

In a nutshell, our approach offers a robust framework not to reflect the "true" in vivo structures but to understand the basic physical principles underlying bacterial chromosome organization. Its advantage is that it does not depend on the bacteria type, microscopic details of the chromosome and nucleoid condition, or even on specific DNA-protein interactions. In contrast to other modeling approaches, the model is not based on unphysical assumptions and restraints. It only takes into account the connectivity of the chromosome and thus accounts for the topological complexity (including topological constraints) of its three-dimensional fold. Thus, our model is based on the idea that nature experiences entropy, excluded volume, and specific chromosome topologies as a driving force to create the right physical conditions for chromosome packaging.

Last but not least, our coarse-grained model provides theoretical insights to guide future experimental approaches that try to obtain information about the shape of the prokaryotic genome.

7.2 Future Challenges

One may raise the question about reproducing bacterial chromosomes' specific internal structural properties, such as the biophysics of chromosome interaction domains (CIDs) in Hi-C data. While many models and mechanisms have been proposed to understand CIDs and their boundaries, no consensus mechanistic view has emerged so far. Consequently, we are interested in investigating the presence of CIDs in our model by fixing the position of the branch points on the DNA backbone.

Besides elucidating the statistical properties of the DNA, we are interested in investigating the possibility of rationalizing the folding dynamics at several Mbp for time scales equivalent to those of the bacterial cell cycle (i.e. the hour). To this aim, one of the major concerns is the implementation of replication in our model.

An additional step toward a better understanding of the processes that govern chromosome structure is the impact of sequence effects on the structural properties of bacterial chromosomes by using our model.

In the future, we will study how excluded volume effects, specific polymer topologies, and geometrical confinement in our model could drive the segregation of replicated chromosomes during cell division. It is worth the effort to take NAPs and SMCs into account. They are expected to influence both the level of supercoiling as well as the length of branches, which, in turn, modifies the complexity of chain topology and, thus, the strength of segregation due to entropic forces.

Appendix A

A Side-Study: Simulating DNA using the WLC model

Self-avoiding supercoiled worm-like chain (ssWLC) model can be used to study several kilobase pairs (kb) to tens kb long double-stranded DNA molecules. I have tested and benchmarked Molecular Dynamics, and Monte Carlo simulation techniques for supercoiled WLC developed in Prof. Everaers' and Dr. Junier's groups.

This model includes supercoiling constraints and self-avoiding properties. In this context, the bending and torsional energies are defined as:

$$\begin{aligned} \beta E_b(\theta) &= k_b(1-cos\theta) \cong \frac{k_b}{2}\theta^2 \\ \beta E_t(\phi) &= k_t\frac{\phi^2}{2} \end{aligned} \tag{A.1}$$

Where the bending angle, θ_i, and twist angle, ϕ_i, can be described explicitly as a function of the local frames attached to the beads (Carrivain, Barbi, and Victor, 2014) (Fig. A.2). Using $\vec{v}_i = \vec{t}_i \times \vec{u}_i$:

$$\begin{aligned} cos\phi_i &= \frac{\vec{u}_{i-1}.\vec{u}_i + \vec{v}_{i-1}.\vec{v}i}{1+\vec{t}_{i-1}.\vec{t}_i} \\ sin\phi_i &= \frac{\vec{v}_{i-1}.\vec{u}_i + \vec{u}_{i-1}.\vec{v}_i}{1+\vec{t}_{i-1}.\vec{t}_i} \\ cos\theta_i &= \vec{t}_{i-1}.\vec{t}_i \end{aligned} \tag{A.2}$$

k_b and k_t are dimensionless rigidity constants related to bending persistence length (ℓ_p) and twisting persistence length (C) respectively. Calculating k_t is straightforward as $k_t = \frac{\ell_p}{l_b}$,where l_b is the unit of distance, i.e bond length (Lepage and Junier, 2017). In order to calculate k_b we need to solve the equations below (Auhl et al., 2003):

$$\begin{aligned} <cos\theta> &= \frac{\int_0^\pi cos\theta sin\theta e^{-\beta E_b(\theta)}d\theta}{\int_0^\pi sin\theta e^{-\beta E_b(\theta)}d\theta} = coth(k_b) - \frac{1}{k_b}, \\ <cos\theta> &= \frac{2\ell_p - l_b}{2\ell_p + l_b}. \end{aligned} \tag{A.3}$$

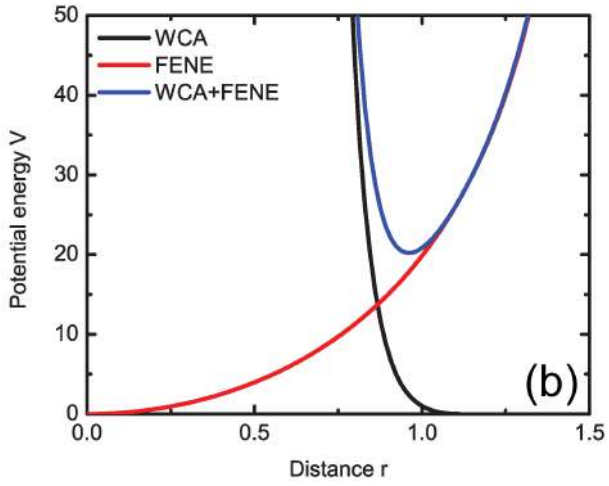


FIGURE A.1: Interaction potential between non-bonded and bonded beads in Kremer-Grest model. Nnn-bonded beads interact via WCA potential, and connected beads interact via both WCA and FENE potential.

In order to benchmark and compare the MD and MC method together, four properties of the DNA molecule were monitored while slowly decreasing $|\sigma|$: Twist, Write, the radius of gyration, and shape anisotropy (The parameters of the simulation of the DNA molecule are given in Table A.1).

Twist (Tw) and Writhe (Wr) can be written as:

$$Tw = \frac{1}{2\pi}\sum_{i=1}^{N}\phi_i \tag{A.4}$$

$$Wr = 2\sum_{i=2}^{N}\sum_{j<i}\frac{\Omega_{ij}}{4\pi} \tag{A.5}$$

where $\frac{\Omega_{ij}}{4\pi}$ is the Gauss integral along the segments i,j (Klenin and Langowski, 2000). The radius of gyration (R_g^2) and shape anisotropy (κ^2) are given by:

$$R_g^2 = \lambda_x^2 + \lambda_y^2 + \lambda_z^2 \tag{A.6}$$

$$\kappa^2 = \frac{3}{2}\frac{\lambda_x^4 + \lambda_y^4 + \lambda_z^4}{(\lambda_x^2 + \lambda_y^2 + \lambda_z^2)^2} - \frac{1}{2} \tag{A.7}$$

where $\lambda_x^2 \leq \lambda_y^2 \leq \lambda_z^2$ are the diagonal elements of the gyration tensor. Shape anisotropy is bounded between zero(all points are spherically symmetric) and one(all points lie on a line).

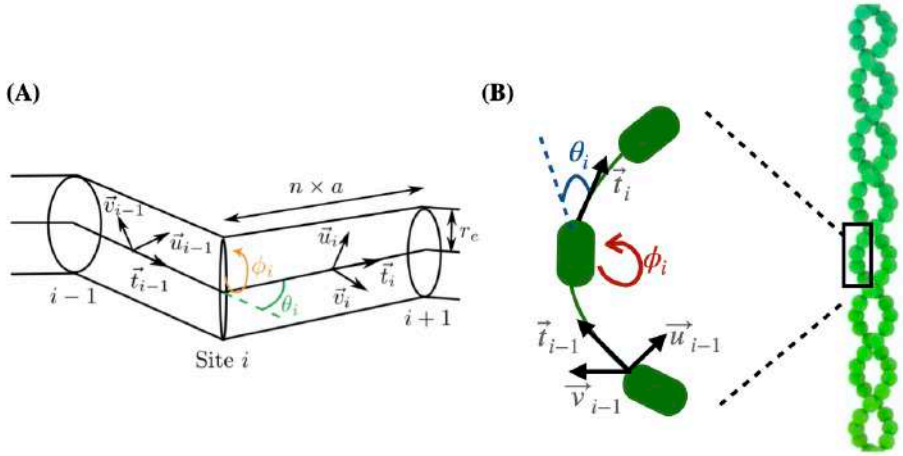


FIGURE A.2: Schematic of the ssWLC. The DNA double-helix is coarse-grained into a polymer of discrete beads. Each bead i is associated with a frame of u_i,v_i, and t_i unit vectors. (A) Cylinder model used in MC simulation (picture from (Lepage and Junier, 2017)), (B)Bead-spring model used in MD simulation.

A.0.1 Molecular dynamics simulation

The molecular dynamics (MD) simulation of a self-avoiding supercoiled Worm-Like Chain can be achieved using the Kremer-Grest polymer (bead-spring) chain (Fig. A.2 (B)) (Grest and Kremer, 1986). The method is develped in Excluded volume interactions between beads are set in place using the Weeks-Chandler-Anderson (WCA) potential, and the bond lengths are regulated with a finitely extensible nonlinear elastic (FENE) potential. FENE + WCA potential guarantees that the chain beads only oscillate within a specific distance (Fig. A.1).

$$V_{\mathrm{WCA}}(r_{ij}) = 4\varepsilon\left[\left(\frac{l_b}{r_{ij}}\right)^{12} - \left(\frac{l_b}{r_{ij}}\right)^{6} + \frac{1}{4}\right] \tag{A.8}$$

$$V_{\mathrm{FENE}}(r_{ij}) = -0.5kR^2 ln\left[1 - \left(\frac{r_{ij}}{R}\right)^2\right] \tag{A.9}$$

where ε is in units of energy, ł$_b$ in units of distance (bond length), $R = 1.5 \times l_b$, and $k = 30\varepsilon/l_b^2$.

A stochastic isothermal MD is implemented, which means integrating the particle's equation of motion using Langevin's equation (Grønbech-Jensen and Farago, 2013):

$$\dot{\vec{r}} = \vec{v} \tag{A.10}$$

$$m\dot{\vec{v}} = \vec{f} - \alpha\vec{v} + \vec{\beta}(t) \tag{A.11}$$

Entity	Parameter	Typical value	Definition
DNA parameters for $[NaCl] = 100\ mM$	r_e	$2\ nm$	radius of cylinders
	l_b	$3.57\ nm$	bond length $n \times a(10.5bp)$
	ℓ_p	$50\ nm$	bending persistence length
	C	$86\ nm$	twisting persistence length
	k_t	24.08	twisting rigidity constant
	k_b	14.50	bending rigidity constant
	N	$476 \approx 5\ kb$	number of DNA cylinders
MD simulation Parameters	τ	$19.12\ 10^{-12} s$	time units $\tau = l_b\sqrt{\frac{m_0}{k_B T}}$
	dt	0.01τ	simulation time step
	T_{tot}	$19.12\ 10^{-6}$ (μ s)	each simulation time
	n_{sample}	20	number of independent simulation samples
MC simulation Parameters	M_1	10^6	number of sweeps for quickest simulation ($4.76\ 10^6\ MC_{step}$)
	M_6	$3.2\ 10^7$	number of sweeps for slowest simulation ($1.5232\ 10^8\ MC_{step}$)
	n_{sample}	20	number of independent simulation samples

TABLE A.1: Set of parameters used to WLC model to study DNA.

where $\beta(t)$ is a thermal white noise and α is a friction coefficient. $\beta(t)$ has a Gaussian distribution and mean zero. α is connected to the thermal noise through the fluctuation-dissipation theorem,

$$\left\langle \vec{\beta}(t) \right\rangle = 0 \tag{A.12}$$

$$\left\langle \vec{\beta}(t)\vec{\beta}(t') \right\rangle = 2\alpha k_B T \delta(t-t'). \tag{A.13}$$

The results for simulating the negatively supercoiled DNA molecule composed of $N = 1547$ cylinders $\approx 60\ kb$ is presented in Fig. A.3. In order to validate the values, we ran the simulations from two strictly different initial conditions, circular chains, and double-folded linear chains. The data confirms that both initial conditions nicely map on top of each other. In other words, it confirms that the simulation results do not depend on the initial conformations of the ring polymer chains.

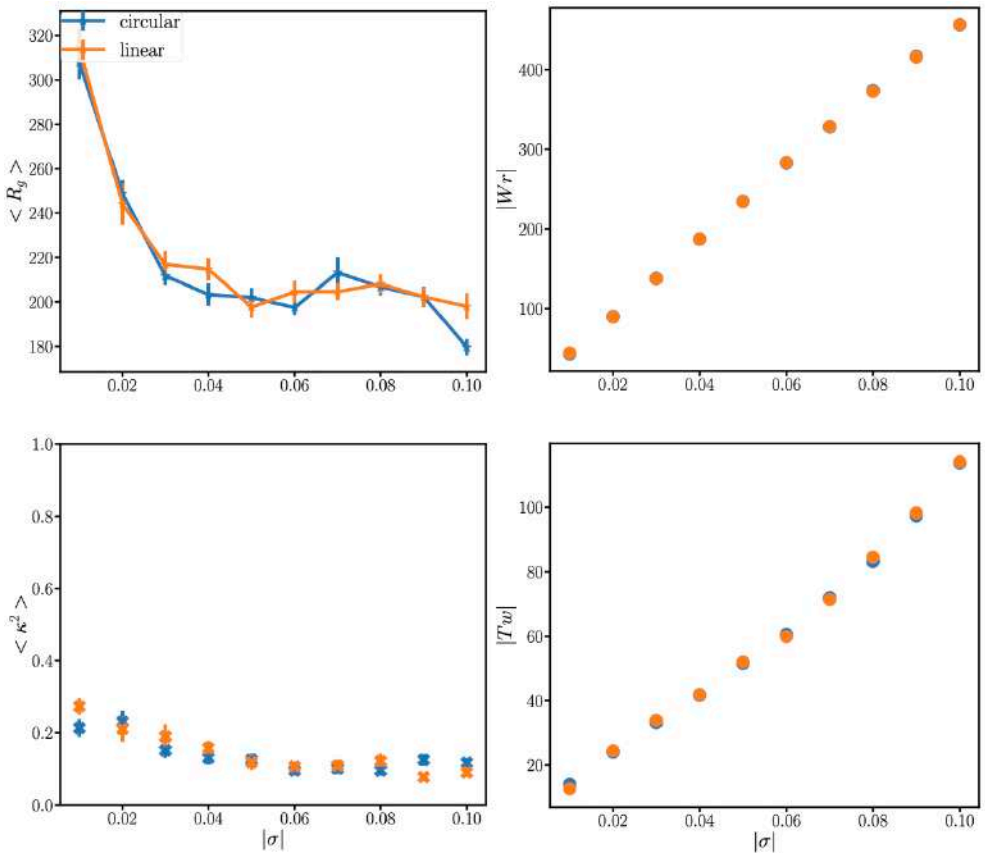


FIGURE A.3: Radius of gyration (R_g^2), shape anisotropy (κ^2), Twist (Tw), and Writhe (Wr) measured as a function of $|\sigma|$ for a ssWLC model with two different initial configurations. 20 independent simulations with $N = 1547$ cylinders ≈ 60 *kb* were initialized in circular chains and double-folded linear chains configurations.

A.0.2 Monte-Carlo Simulation

In the Monte-Carlo simulation of the ssWLC (Vologodskii and Cozzarelli, 1994), the chain is divided into cylinders (Fig. A.2 (A)). The length of each cylinder is fixed $(l_b = n \times a)$ and they are impenetrable. The energy of any conformation of the chain can be calculated by:

$$\beta E(C) = \beta E_b(\theta) + \beta E_t(\phi) \tag{A.14}$$

where $E_b(\theta)$ and $E_t(\phi)$ are defined by equations in (A.1). The Monte-Carlo simulation is evolved through the following MC move (Lepage and Junier, 2017; Lepage and Junier, 2019):

i) Two cylinders are selected for the crankshaft move(Fig. A.4),

ii) Check if the move does not violate collision and crossing rules,

iii) Accept the move using the Metropolis algorithm.

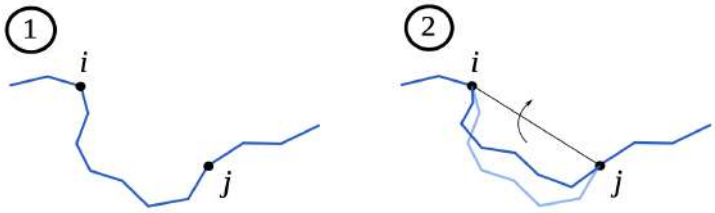


FIGURE A.4: Crankshaft move: Pick a random block of cylinders $[i, j]$. Rotate the block around the axis connecting the ends of the block by a random angle. (picture from (Lepage and Junier, 2017))

Now, in order to apply Monte-Carlo simulation to study negatively supercoiled DNA molecules, the protocol is as follows:
1) Begin with a circular chain with initial $\sigma = 0$,
2) Evolve the system with N MC moves (explained above),
3) Decrease σ by a $\delta\sigma$ (usually 0.005),
4) Repeat steps (2) and (3) iteratively until desired σ is obtained.

We ran 20 different simulations with $N = 476$ cylinders ≈ 5 *kb* with six different supercoiling rates ($\delta\sigma$ to N ratios). Fig. A.5 shows the data obtained from these different simulation runs. Interestingly, the rings' radii are sensitive to the supercoiling rates and equilibrate into two distinct values. The underlying cause for such behavior is under investigation. In this MC simulation, it is important to change σ slowly to achieve the desired density. By changing it rapidly, the configurations get trapped in local minima states and do not screen all conformations in the phase space.

A.0.3 Caparison of MD and MC simulations (IN PROGRESS)

Four properties of the negatively supercoiled DNA molecule are calculated for both the MC and MD simulations. As you can see in Fig. A.6 small deviations are visible between the results from the two simulation methods. We are investigating this deviation and searching for the answer to the following questions.

1- Are the simulations healthy? Assuming that the two simulation methods are equilibrated, we are trying to check separately if the simulation parameters for these two methods can produce good physics. To find the answer to this question, we used the MD method to simulate a linear WLC. There is an exact solution for the mean square internal distances of a linear WLC as a function of the contour length (Rubinstein and Colby, 2003). We can compare the simulation results to the theory.

2- Are the systems equilibrated? We are calculating and measuring the number of steps required by the MD and MC systems to reach equilibrium. We are repeating the simulation in different environments (different thermostats) for the MD method.

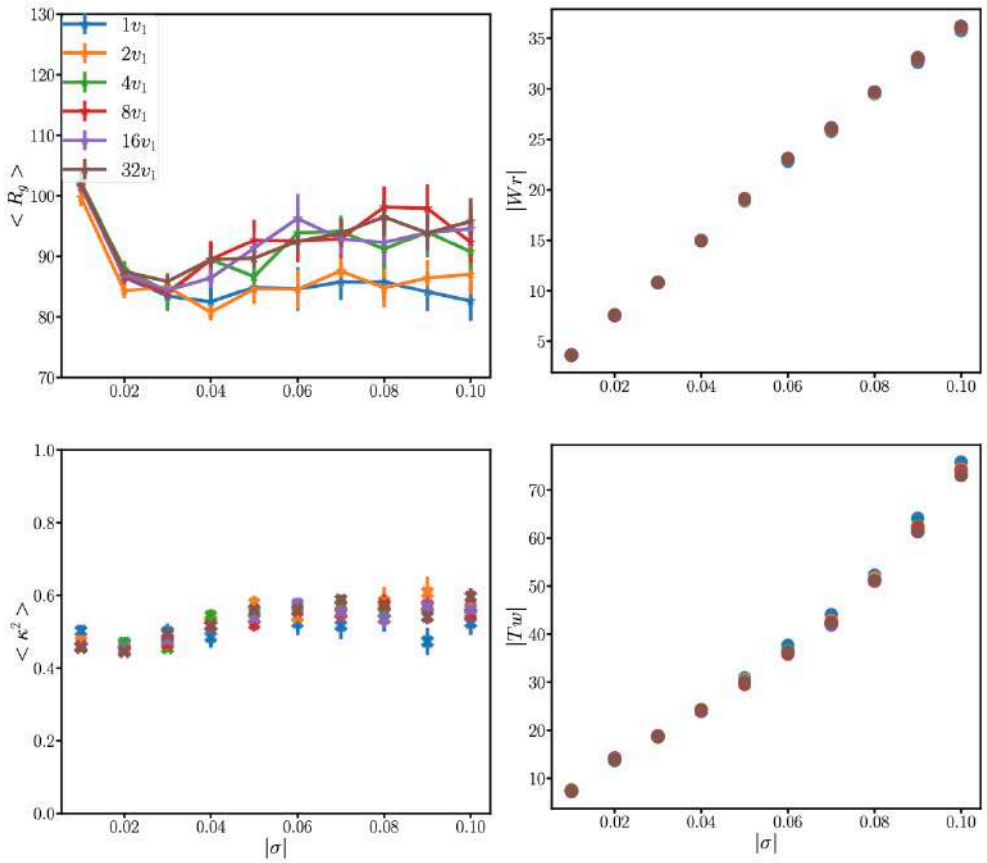


FIGURE A.5: radius of gyration (R_g^2), shape anisotropy (κ^2), Twist (Tw) and Writhe (Wr) as a function of $|\sigma|$ for different supercoiling rates from 20 different Monte-Carlo simulation runs with $N = 476$ cylinders ≈ 5 *kb*

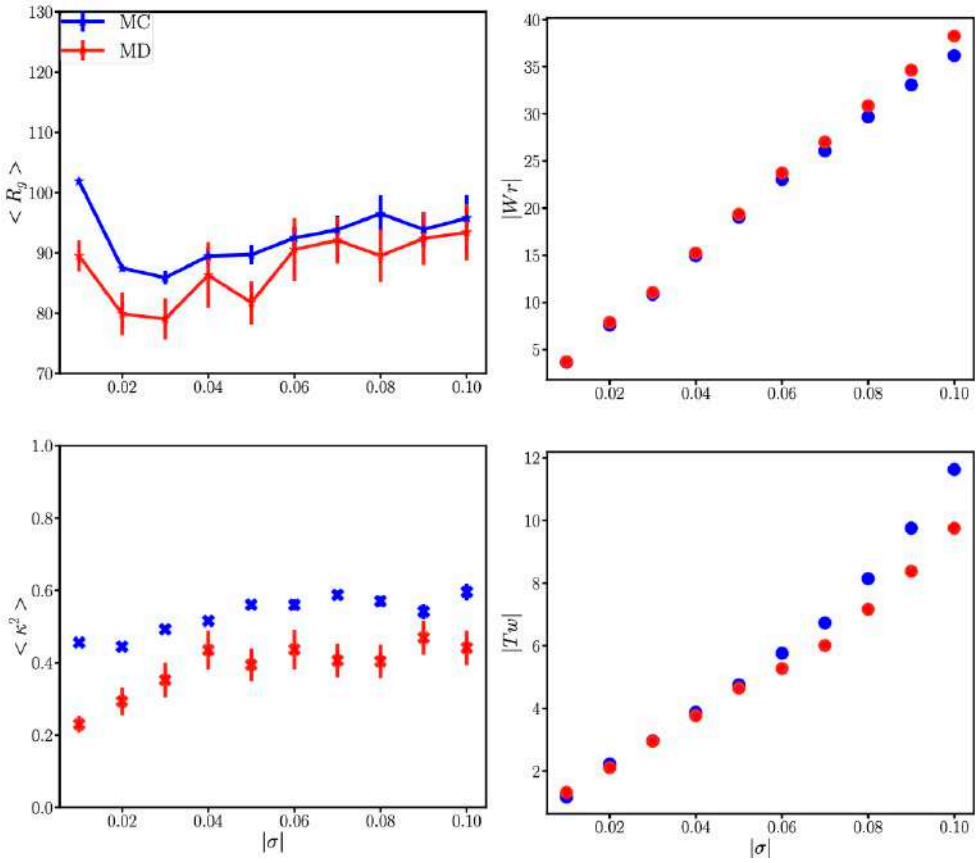


FIGURE A.6: Radius of gyration (R_g^2), shape anisotropy (κ^2), Twist (Tw) and Writhe (Wr) as a function of $|\sigma|$ for Molecular Dynamics and Monte-Carlo simulations with $N = 476$ cylinders ≈ 5 *kb*.

Monte Carlo simulation of a lattice model for the dynamics of randomly branching double-folded ring polymers

Elham Ghobadpour,[1,2] Max Kolb,[3] Mohammad Reza Ejtehadi,[4] and Ralf Everaers[3,*]

[1]*Université Grenoble Alpes, Centre National de la Recherche Scientifique (CNRS), TIMC, F-38000 Grenoble, France*

[2]*School of Nano Science, Institute for Research in Fundamental Sciences (IPM), 19395-5531, Tehran, Iran*

[3]*Université de Lyon, École Normale Supérieure (ENS) de Lyon, CNRS, Laboratoire de Physique and Centre Blaise Pascal de l'ENS de Lyon, F-69342 Lyon, France*

[4]*Department of Physics, Sharif University of Technology, 11155-9161, Tehran, Iran*

(Received 19 January 2020; revised 27 April 2021; accepted 7 June 2021; published 13 July 2021)

Supercoiled DNA, crumpled interphase chromosomes, and topologically constrained ring polymers often adopt treelike, double-folded, randomly branching configurations. Here we study an elastic lattice model for tightly double-folded ring polymers, which allows for the spontaneous creation and deletion of side branches coupled to a diffusive mass transport, which is local both in space and on the connectivity graph of the tree. We use Monte Carlo simulations to study systems falling into three different universality classes: ideal double-folded rings without excluded volume interactions, self-avoiding double-folded rings, and double-folded rings in the melt state. The observed static properties are in good agreement with exact results, simulations, and predictions of Flory theory for randomly branching polymers. For example, in the melt state rings adopt compact configurations and exhibit territorial behavior. In particular, we show that the emergent dynamics is in excellent agreement with a recent scaling theory and illustrate the qualitative differences with the familiar reptation dynamics of linear chains.

DOI: 10.1103/PhysRevE.104.014501

I. INTRODUCTION

The behavior of melts of nonconcatenated ring polymers has caught the interest of physicists over many years [1–12] and appears to provide a natural explanation for the territorial chromosomal arrangement in eukaryotic cells during interphase [13–15]. With their microscopic topological state permanently quenched, the equilibrium statistics and dynamics of nonconcatenated ring polymers is fundamentally different from the behavior of their linear counterparts. A powerful approximation is available through the analogy with ring polymers in an array of fixed obstacles [1,3,7,8]. In this view, crumpling can be understood by the successive application of three different strategies for entropy maximization: double folding, branching, and swelling. Firstly, and most importantly, the rings adopt double-folded configurations to minimize the threadable surface, as this reduces the importance of the topological constraints they impose on each other. Secondly, double-folded rings can increase their entropy by branching. Thirdly, there is a certain amount of swelling due to partially screened excluded volume interactions leading to asymptotically compact conformations characterized by the scaling exponent of the radius of gyration, $\nu = 1/d$ for $d \leqslant 4$ dimensions. Double-folding, branching, and swelling due to excluded volume interactions also occur in plectonemic configurations of supercoiled circular DNA [16–20]. Similarly, viral RNA may effectively behave like branched polymers [21–23]. From a more general perspective, the mapping of (double-folded) ring polymers to randomly branched polymers or trees suggests analogies to phenomena such as gelation [24], percolation [25], and the critical behavior of magnetic systems [26–29]. This may explain, why polymer *physics* [24,30–32] is often concerned with exponents, which characterize the essence of the behavior of all members of a universality class and are independent of microscopic details differentiating particular experimental polymers as well as lattice and off-lattice models from each other.

Recent numerical work on the *static* properties of self-avoiding trees and lattice tree melts [33–35] has shown that the behavior of randomly branching chains under different solvent conditions is in excellent qualitative agreement with a suitably generalized Flory theory [13,36]. A multiscale approach to the construction of ring melts based on this analogy faithfully captures many aspects of the conformational statistics of properly equilibrated systems [11,12]. However, Monte Carlo algorithms optimized for rapidly equilibrating the static structure of randomly branching chains [37,38] [Fig. 1(c)] generate an artificial dynamics.

To generate a physically more realistic dynamics, Monte Carlo simulations [39–41] need to obey the same conservation laws [42] as the modeled target systems. In the present case, this requires a scheme where the mass transport is local both in space and on the connectivity graph of the tree.

Below we present Monte Carlo simulations of a suitable elastic lattice polymer model which accounts for double folding [4,43], the local accumulation of contour length on the

*ralf.everaers@ens-lyon.fr

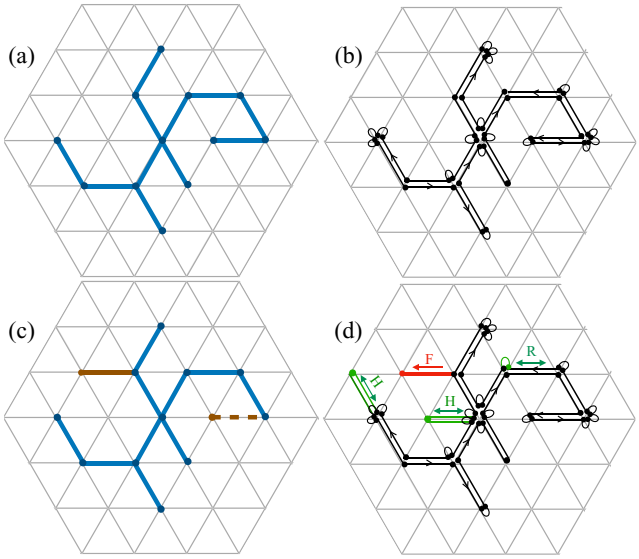


FIG. 1. (a) Branched tree on a trigonal lattice and (b) a corresponding (tightly wrapped) double-folded ring polymer. Small loops represent bonds of zero length, where adjacent monomers along the ring occupy identical lattice sites. (c) Example of a nonlocal "amoeba" Monte Carlo move [37,38] altering the tree structure. The dashed brown line shows the location of a branch prior to the MC move, and the solid brown line shows an arbitrary location where the branch could be reattached to the tree. (d) Examples of local MC moves for the present model of double-folded ring polymers. Dots represent monomers, and black lines represent an allowed conformation of the double-folded chain. The allowed (forbidden) moves are indicated by the green (red) color. R: the Repton move. H: the Hairpin move. F: Forbidden move that does not preserve the double-folded structure.

primitive tree [4,40,44–46], as well as excluded volume interactions [44–46].

The manuscript is structured as follows: In Sec. II we introduce relevant observables and the related exponents. Also, we briefly summarize the theoretical background. The model and the simulations are described in Sec. III. In Sec. IV we present and discuss our results. After some first qualitative insights in Sec. IV A, we validate that the static properties of our model are in good agreement with theoretical and numerical work on trees (Sec. IV B). The next step, Sec. IV C, focuses on comparison of the single ring dynamics with the predictions of a recent scaling theory [3,8]. Furthermore, we compare the dynamics of double-folded rings to the motion of linear chains in the tube model [47] (Sec. IV D). Finally, we briefly conclude in Sec. V.

II. THEORETICAL BACKGROUND

A double-folded ring polymer can be mapped on a randomly branched primitive tree [1,3,4,7,8,11,48,49]. In analogy to protein or RNA structures, such conformations can be discussed in terms of a primary, a secondary, and a tertiary structure [50]. The primary structure is simply defined through the connectivity of the ring monomers. The secondary structure arises from the double folding and can be specified through the mapping of the ring onto a graph with the connectivity of the primitive tree. The tertiary structure describes the embedding of the rings and trees into (three-dimensional) space. We define corresponding observables in Secs. II A and II B. Sections II C and II D briefly summarize scaling arguments for the effect of excluded volume interactions and the dynamics of randomly branched double-folded ring polymers.

A. Secondary structure

Two standard measures of the tree connectivity are the mean contour distance L between tree nodes and the average weight of branches $N_{\rm br}$, separated from the tree by severing a randomly chosen link. Both depend on the weight N of the rings through power law relations

$$\langle L(N)\rangle \sim N^{\rho}, \tag{1}$$

$$\langle N_{\rm br}(N)\rangle \sim N^{\epsilon}, \tag{2}$$

where $\epsilon = \rho$ is expected to hold in general [38]. The (tight) wrapping of a tree by a ring polymer introduces an additional metric on the embedded graph [50]. The central quantity is the length of the shortest path on the *tree* or tree contour distance, L, between two monomers i and j along the ring. For short *ring* contour distances, $n = |i - j|$, one simply expects $\langle L(n)\rangle \sim n$. However, beyond the typical distance between branch points, the ring does not follow a linear path on the tree but wraps side branches. For $n \ll N$, Eq. (1) suggests $\langle L(n)\rangle \sim n^{\rho}$. Due to the ring closure $\langle L(n)\rangle \equiv \langle L(N-n)\rangle$ reaches its maximum for $n = N/2$ before reducing to zero at the total ring size, $\langle L(N)\rangle \equiv 0$. The simplest functional form accounting for this constraint is [50]

$$\langle L(n)\rangle_N \sim \left[n\left(1 - \frac{n}{N}\right)\right]^{\rho}. \tag{3}$$

B. Tertiary structure

The simplest measures of the tertiary structure are the overall tree gyration radii,

$$\left\langle R_g^2(N)\right\rangle \sim N^{2\nu}, \tag{4}$$

as a function of the chain length. For a more detailed understanding, it is useful to consider the mean-square spatial distance between nodes,

$$\langle R^2(L)\rangle \sim L^{2\nu_{\rm path}}, \tag{5}$$

as a function of their contour distance on the tree, where $\nu = \rho\,\nu_{\rm path}$. Combining Eqs. (5) and (3) suggests [50]

$$\langle R^2(n)\rangle_N \sim \left[n\left(1 - \frac{n}{N}\right)\right]^{2\nu} \tag{6}$$

for the mean-square spatial distance of monomers as a function of their distance, $n = |i - j|$, along the ring.

C. Flory theory

Exact values for the exponents are known only for a very small number of cases. For ideal noninteracting trees, the exponents $\rho^{\rm ideal} = \epsilon^{\rm ideal} = \nu_{\rm path}^{\rm ideal} = 1/2$ and $\nu^{\rm ideal} = 1/4$ [51,52].

For interacting trees, the only known exact result [26] is the value $\nu = 1/2$ for self-avoiding trees in $d = 3$.

Flory theories [7,13,36,53,54] of interacting tree systems are formulated as a balance of an entropic elastic term and an interaction energy [55]

$$\mathcal{F} = \mathcal{F}_{el}(N, R) + \mathcal{F}_{\text{inter}}(N, R). \tag{7}$$

In the present case, the elastic free energy takes the form [13]

$$\frac{\mathcal{F}_{el}}{k_B T} \sim \frac{R^2}{l_K L} + \frac{L^2}{N l_K^2}. \tag{8}$$

The first term of Eq. (8) is the usual elastic energy contribution for stretching a polymer of linear contour length L at its ends [13]. The second term penalizes deviations from the ideal branching statistics, which lead to longer paths and hence spatially more extended trees. Optimizing L for annealed trees for a *given* asymptotic, $R \sim N^{\nu}$, yields [13,36]

$$\rho = \frac{1 + 2\nu}{3}, \tag{9}$$

$$\nu_{\text{path}} = \frac{3\nu}{1 + 2\nu}, \tag{10}$$

independently of the type of volume interactions causing the swelling in the first place. Plausibly, a fully extended system, $\nu = 1$, is predicted not to branch, $\rho = 1$, and to have a fully stretched stem, $\nu_{\text{path}} = \nu = 1$. For the radius of ideal randomly branched polymers, $\nu = 1/4$, one recovers $\rho = 1/2$ and Gaussian path statistics, $\nu_{\text{path}} = 1/2$.

Reference [36] reviews the predictions of the Flory theory for randomly branching polymers for a a wide range of conditions characterized by different expressions for the interaction energy in Eq. (7). For self-avoiding trees, $\frac{\mathcal{F}_{\text{inter}}(N,R)}{k_B T} \sim v_2 \frac{N^2}{R^d}$ represents the two-body repulsion between segments, which dominates in good solvent. In this case, Flory theory predicts [13]

$$\nu = \frac{7}{3d + 4} \quad 1 \leqslant d \leqslant 8, \tag{11}$$

in qualitatively excellent and almost quantitative agreement with the exact results [33]. In dense melts, all terms of the virial expansion of the partially screened excluded volume interactions become relevant and the trees are expected to be compact [36]:

$$\nu = \frac{1}{d} \quad 1 \leqslant d \leqslant 4. \tag{12}$$

While Flory theory describes the average behavior of the tree observables mentioned above, we note for completeness that the corresponding non-Gaussian distribution functions are typically of the Redner–des Cloizeaux (RdC) form of a power law multiplied with a stretched exponential. Most of the additional exponents characterizing the tails of the distributions can be related to each other and the standard tree exponents [35,50].

D. Dynamics

In the following we summarize the arguments for the dynamics of randomly branched double-folded ring polymers from Refs. [3,4,8]. Consider the division of a tree into its trunk (the longest path on the tree) and the branches hanging off this trunk. The trunk has a length of $L \sim aN^{\rho}$, where a is the lattice constant and the number of branches is proportional to L. The slowest relaxation process is the transport of mass along the trunk, while the intrabranch dynamics is irrelevant and may be neglected [3,4]. The elementary step of the dynamics is the reptonlike exchange of mass between neighboring branches along the trunk [56]. Each elementary event changes the average projected position of the monomers along the trunk by an amount $\delta s_{CM} \sim a/N$. The number of such events by an elementary unit of time τ_0 is proportional to the number of branches. As a consequence, $\langle \delta s_{CM}^2(\tau_0) \rangle \sim (a/N)^2 N^{\rho}$, corresponding to a diffusion constant for the curvilinear motion along the trunk of $D_{||}(N) \sim \langle \delta s_{CM}^2(\tau_0) \rangle / \tau_0 \sim (a^2/\tau_0) N^{\rho-2}$. To completely relax the internal tree structure, the tree CM has to diffuse over the entire trunk length. As a consequence,

$$\tau_{\max}(N) \sim \frac{(aN^{\rho})^2}{(a^2/\tau_0) N^{\rho-2}} \sim \tau_0 N^{\rho+2}, \tag{13}$$

or, using Eq. (9), $\tau_{\max}(N) \sim \tau_0 N^{(7+2\nu)/3}$. As this corresponds to a mean-square spatial displacement of $\langle \delta R_{CM}^2(\tau_{\max}) \rangle \sim \langle R_g^2(N) \rangle \sim a^2 N^{2\nu}$, the long-time CM and monomer diffusion are given by

$$g_{1,3}(t \gg \tau_{\max}(N)) \sim D_{CM}(N)\, t \tag{14}$$

with

$$D_{CM}(N) \sim \frac{a^2}{\tau_0} N^{2\nu-\rho-2} \tag{15}$$

or $D_{CM}(N) \sim \frac{a^2}{\tau_0} N^{(4\nu-7)/3}$.

Furthermore, one can invert Eq. (13) to obtain the mass,

$$n(t) \sim \left(\frac{t}{\tau_0}\right)^{\frac{1}{\rho+2}}, \tag{16}$$

of rings (or ring sections) which are equilibrated after a given time, $\tau_0 < t < \tau_{\max}$. During a corresponding time interval, monomers move over a spatial distance of the order of $a^2 n(t)^{2\nu}$:

$$g_1(\tau_0 < t < \tau_{\max}) \sim a^2 \left(\frac{t}{\tau_0}\right)^{\frac{2\nu}{\rho+2}}, \tag{17}$$

which is independent of the total ring length. Similarly, one can estimate the CM motion by noting that each of the $n(t)/N$ equilibrated ring sections independently moves over a distance of the order of $a^2 n(t)^{2\nu}$. As a consequence,

$$g_3(\tau_0 < t < \tau_{\max}) \sim g_1(t) \frac{n(t)}{N} \sim \frac{1}{N} a^2 \left(\frac{t}{\tau_0}\right)^{\frac{2\nu+1}{\rho+2}}. \tag{18}$$

III. MODEL AND METHOD

Monte Carlo simulations [39–41] can be expected to generate physically realistic results if they obey the same conservation laws [42] as the modeled target systems. As an illustration in the present context, consider first an algorithm on the tree level, which removes or adds segments with a probability governed by a chemical potential. While this allows control of the *average* tree weight, such an algorithm is clearly inappropriate to simulate (double-folded) ring polymers of *fixed* weight. This difficulty is partially overcome by the amoeba algorithm of Seitz and Klein [37], which attempts to move one-functional tree "leaves" to random locations on

the tree [Fig. 1(c)]. Since this operation conserves the tree weight, the algorithm can be meaningfully employed to study static aspects of the ring polymer and the chromosome folding problem [11,14,34,49]. The same holds true for a variant of the same idea by Janse van Rensburg and Madras [38], which achieves a much higher efficiency in dilute systems by cutting and relocating entire branches. Similarly, one could envision a (probably highly efficient) variant of the connectivity altering double-bridging scheme [57] for tree melts, where neighboring trees swap entire branches of equal weight without moving them in space. Such moves can be expected to have a much higher acceptance probability in dense systems, since they preserve the uniform monomer density. However, none of these algorithms can be used to simulate the configurational *dynamics* of double-folded ring polymers. Instead we require a scheme where *the mass transport is local in space and on the connectivity graph* of the tree.

In the following, we first review the elastic lattice polymer model [40,43–46], which is a simple and efficient Monte Carlo algorithm for studying the dynamics of entangled linear chains (Sec. III A). In Sec. III B we describe the generalization to randomly branching double-folded ring polymers [4]. Sections III C and III D provide more details on the systems studied and on how we initialized and equilibrated our runs.

A. Elastic lattice polymer models

The dynamics of topologically constrained *linear* polymers can be efficiently studied in Monte Carlo (MC) simulations of the coarse-grained elastic lattice polymer model [40,43–46]. In this model the continuum dynamics of a polymer melt is replaced by a lattice version. The mapping is achieved by dividing the space into cells where the centers of these cells form a regular lattice. By moving all the monomers residing in a cell to the center, the polymer conformation and its dynamics are discretized. The maximum number of monomers per site is a free parameter of the model that depends on the coarse graining, e.g., it can be calculated by considering the volume of the monomers and the cell volume. The projection of real space onto a lattice will also limit the dynamics to nearest-neighbor hops of the monomers. As a result of the coarse graining, the only allowed bond lengths between neighboring monomers can be 0 or 1, 0 for monomers that are in the same cell, and 1 for monomers residing in neighboring cells. This also means that any MC move should only be accepted if it preserves this constraint.

In order to guarantee the constraints of excluded volume and noncrossing of strands, without the loss of generality, a minor restraint is introduced to the occupancy of the cells [4]. Multiple occupancy on a lattice site is allowed only for monomers that are directly connected to each other by monomers on the same site, i.e., they form a polymer strand of variable length (a subchain of chemically bonded monomers), where the length of such a strand is limited by the cell size. This also guarantees that monomers belonging to different polymers can never occupy the same site. With this constraint the implementation of excluded volume interaction and noncrossability becomes operationally trivial while fully maintaining a repton [56]-like dynamics along the primitive chain [47].

When the elastic lattice polymer model is used to simulate ring polymers [46,49], the algorithm conserves the microscopic topological state of the starting configuration. In particular, melts of long, nonconcatenated rings exhibit compact and characteristically crumpled conformations.

B. Generalization to randomly branching double-folded ring polymers

When the algorithm is generalized to double-folded polymers [4], lattice bonds represent tree segments which can only be occupied by two oppositely oriented bonds between ring monomers [Fig. 1(b)]. Tree nodes are located on lattice sites. Their functionality depends on the number of emerging tree segments, $f = 1$ (a leaf or branch tip), $f = 2$ (linear chain section), and $f \geqslant 3$ (branch point). As in the elastic lattice polymer model and in contrast to standard tree models [33,37,38] (which can be wrapped *a posteriori* with ring polymers [50]), the degrees of freedom are the positions of the ring monomers. The local redistribution of stored length is responsible for the dynamics. In contrast to the elastic lattice polymer model, the connectivity graph for the connected lattice sites is a dynamically branching object. In particular, new side branches are created when a monomer from inside an accumulation of stored length moves to a neighboring lattice side. Conversely, side branches vanish when the last monomer reintegrates into the main branch and retracts the two remaining extended ring bonds representing the tree segment [Fig. 1(d)]. Specifically, we impose the following rules:

Lattice: Ring monomers are placed on the sites of a periodic FCC lattice. We choose the FCC lattice because it is isotropic and it has the maximum number of nearest neighbors.

Connectivity: Bonded monomers can occupy either the same site (a repton or loop of stored length) or nearest neighbor sites (an extended bond).

Order: Ring monomers remain ordered even if several connected monomers occupy the same site [56].

Double-folding: Each extended bond of the polymer is paired with exactly one extended bond pointing in the opposite direction [Fig. 1(b)].

Ring monomers and tree nodes: The number of ring monomers belonging to the same tree node is equal to $f + \alpha$, where f is the functionality of the lattice tree node and α is the number of loops of stored length on that node.

Excluded volume interaction: Different tree nodes are not allowed to occupy the same lattice site.

Dynamics: We employ a simple Metropolis scheme [58] which consists of trying to move a randomly selected monomer to a randomly chosen site out of the twelve possible nearest neighbors. The move is only accepted if the new conformation respects the conformational rules above.

A two-dimensional (hexagonal) representation of a double-folded polymer chain conformation produced using the model is shown in Fig. 1(d). The green (red) color indicates the allowed (forbidden) moves according to the conformational

TABLE I. System parameters for the double-folded ideal rings (Ideal), double-folded self-avoiding rings (S.A.T), and double-folded rings in the melt state (Melt). N: Number of monomers per chain; n_p: Total number of chains per simulated system; L_{box}: Size of the simulation box. Note: An FCC lattice of size L contains $\frac{L^3}{2}$ sites; Site occupation number: The average number of monomers in an occupied lattice site; ρ: Lattice density which is the ratio of the number of occupied sites to the total number of sites; N_{sample}: Number of independent MC samples; T_{tot}: simulation run time in Monte Carlo sweep [MCs]; CPU time: The CPU wall clock time for N_{sample} samples on a single core; N_{ind}: The number of independent configurations (see Sec. IV C); τ_{max}: Configuration relaxation time measured after reaching equilibrium calculated using, $\tau_{\text{max}} = \frac{T_{\text{tot}}}{2N_{\text{ind}}}$. Radius of gyration $\langle R_g^2 \rangle$, MSID $\langle R^2(N/2) \rangle$, and average value of contour distance $\langle L_N \rangle$ are defined in the text. All times are measured in Monte Carlo sweep [MCs], all distances are in units of lattice constants, and all measurements are performed after reaching equilibrium.

N	n_p	L_{box}	Site occupation	ρ	N_{sample} [$\times 10^2$]	T_{tot} [$\times 10^4$]	CPU time [days]	N_{ind} [$\times 10^4$]	τ_{max} [$\times 10^4$]	$\langle R_g^2 \rangle$	$\langle R^2(N/2) \rangle$	$\langle L(N) \rangle$
					Ideal							
64	1	8	2.7(7)	$9.1(1) \times 10^{-2}$	16	10^3	∼0.4	∼10^2	∼0.8	2.13 ± 0.01	5.74 ± 0.13	4.38 ± 0.03
216	1	12	3.4(7)	$7.2(6) \times 10^{-2}$	16	10^4	∼5	∼38	∼21	4.27 ± 0.03	11.13 ± 0.26	8.51 ± 0.06
512	1	16	3.9(8)	$6.3(4) \times 10^{-2}$	16	10^4	∼16.5	∼3.6	$\sim 2.2 \times 10^2$	7.13 ± 0.06	18.56 ± 0.44	13.59 ± 0.1
1000	1	20	4.3(2)	$5.8(6) \times 10^{-2}$	6.4	10^5	∼74	∼2.4	$\sim 1.3 \times 10^3$	10.52 ± 0.14	27.37 ± 1.02	19.18 ± 0.23
					S.A.T							
64	1	20	2.3(6)	$6.7(8) \times 10^{-3}$	16	10^3	∼0.2	∼81	∼0.9	3.53 ± 0.02	10.25 ± 0.21	4.84 ± 0.04
216	1	32	2.3(5)	$5.6(1) \times 10^{-3}$	16	10^4	∼4.5	∼27	∼29	11.76 ± 0.08	34.61 ± 0.72	10.96 ± 0.09
512	1	80	2.3(4)	$8.5(5) \times 10^{-4}$	16	10^4	∼17.5	∼2.6	$\sim 3 \times 10^2$	28.65 ± 0.20	85.01 ± 1.78	19.77 ± 0.17
1000	1	100	2.3(5)	$8.5(2) \times 10^{-4}$	6.4	10^5	∼30.7	∼1.8	$\sim 1.7 \times 10^3$	55.95 ± 0.62	166.42 ± 5.55	30.64 ± 0.35
					Melt							
64	12	8	3.1(5)	$9.5(3) \times 10^{-1}$	16	10^3	∼4.5	$\sim 2 \times 10^3$	∼0.4	2.08 ± 0.01	5.82 ± 0.03	3.63 ± 0.04
216	12	12	3.1(8)	$9.4(4) \times 10^{-1}$	8	10^4	∼76	$\sim 3.4 \times 10^2$	∼14.1	5.72 ± 0.01	15.99 ± 0.13	8.03 ± 0.06
512	12	16	3.1(8)	$9.4(4) \times 10^{-1}$	6.4	10^4	∼165	∼22	$\sim 1.6 \times 10^2$	11.33 ± 0.03	31.33 ± 0.30	13.69 ± 0.18
1000	12	20	3.1(8)	$9.4(3) \times 10^{-1}$	2.56	10^5	∼749	∼15	$\sim 9.7 \times 10^2$	18.89 ± 0.14	52.04 ± 1.15	20.46 ± 0.42

constraints. All allowed hopping moves for tightly folded rings can be classified in terms of two different move types:

The Repton move: A unit of stored length hops one unit along the tree without changing its structure [indicated with the green letter (R) in Fig. 1(d)].

The Hairpin move: If there is at least one connected loop of stored length (in a site) on each side of a monomer, both loops can unfold and result in an extended bond (creation). Naturally, the inverse move removes an extended bond pair and thus shortens or removes a side branch from the tree structure (annihilation). In fact, branched structures are introduced by the formation of hairpins [green letter (H) in Fig. 1(d)].

An example of a rejected move is also shown in Fig. 1(d). The red move (F) will lead to a forbidden conformation because it does not preserve the double-folded structure. For an impression of the dynamics of the appearance and disappearance of branches due to Hairpin moves, we refer the reader to the opening sequences of the Supplemental Material, video S4 [59].

C. Studied systems

The simulations were carried out for chain lengths varying between 64 to 1000 for three systems: ideal double-folded rings, self-avoiding double-folded rings, and rings in the melt state. The self-avoiding case and the rings in the melt state have excluded volume interactions. For the ideal case, there is no restriction on the number of tree nodes on any site of the lattice (no excluded volume interaction). In the melt state a high lattice density, $\rho = 0.95$, was used. Since polymer chains diffuse very slowly in compact systems, we have performed long simulation runs to have a large number of independent samples. For the self-avoiding double-folded ring simulations the size of the box was set large enough to avoid self-interaction of the chain as a result of the periodic boundary condition. A summary of the simulation parameters and data is given in Table I.

D. Initialization and equilibration

Initial configurations are produced through a growth process. First, double-folded rings are seeded as trimers which are located on a common lattice site. Then the process comprises two operations:

(i) The diffusion of the monomers, in agreement with the previously stated dynamic rules. If branching is allowed during the growth process, highly branched compact chain configurations appear on the lattice. However, if branching is not allowed during growth, the Hairpin move is restricted to the chain ends. As a result, a double-folded ring configuration will be built which resembles a self-avoiding random walk in space.

(ii) The occasional addition of new monomers. A new monomer is introduced to a chain by selecting a random monomer on the chain and inserting the new one between the selected monomer and the next neighbor along the chain, on

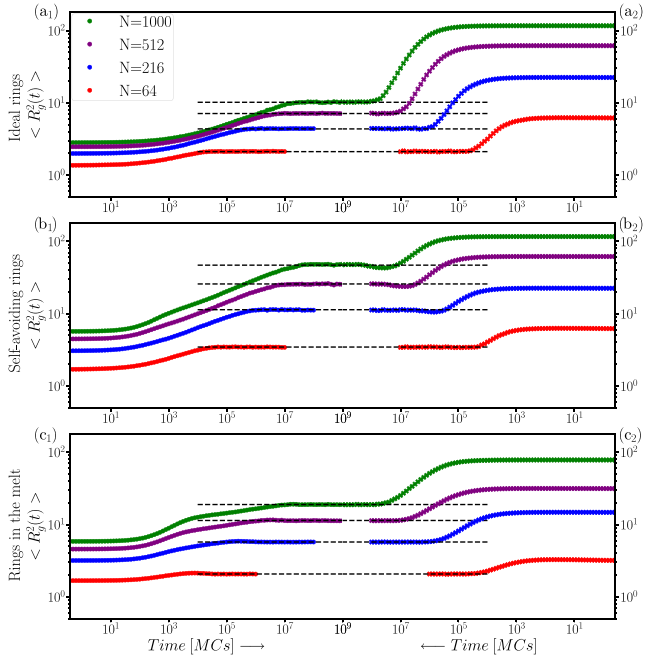


FIG. 2. Equilibration monitored using the mean-square radius of gyration as a function of time. Comparison between initially branched, compact double-folded chains (left column), and initially double-folded chains with no branches (right column) for three different systems: (a) ideal double-folded rings, (b) self-avoiding double-folded rings, and (c) double-folded rings in the melt state. The horizontal lines represent the average values after the chains have reached equilibrium. For better visualization, the time direction is reversed in the right column. In all three systems, at large times both initial states (from left and right) reach the same equilibrium values (reported in Table I).

the same site. In other words, we add a loop of stored length. The insertion of a loop is attempted with low probability, 0.01, to assure a good balance between growth and equilibration. The addition of new monomers is halted once all chains have the desired length. This process results in unknotted and non-concatenated rings.

The radius of gyration is the structural property that has been investigated to monitor the state of equilibration of the polymer systems. The mean-square gyration radius $\langle R_g^2 \rangle$ is the average squared distance of any point in the ring polymer chain from its center of mass. The systems have reached equilibrium when this quantity no longer changes. All reported quantities are taken from the simulation regime where the polymers are fully equilibrated.

Figure 2 shows the radius of gyration vs MC time (one MC sweep is equivalent to one MC trial for every monomer in the system) for different rings. In order to validate the equilibrium values of radius of gyration, we ran the simulations from two totally different initial conditions. The left column in Fig. 2 shows equilibration of compact initial configurations, which swell as the simulation progresses with time. The right column shows equilibration of more open, initially unbranched double-folded rings, where average ring size decreases as branches appear. The horizontal lines represent the average values after the chains have reached equilibrium (values are reported in Table I). In particular, Fig. 2 confirms that the simulation results do not depend on the initial conformations of the ring polymer chains.

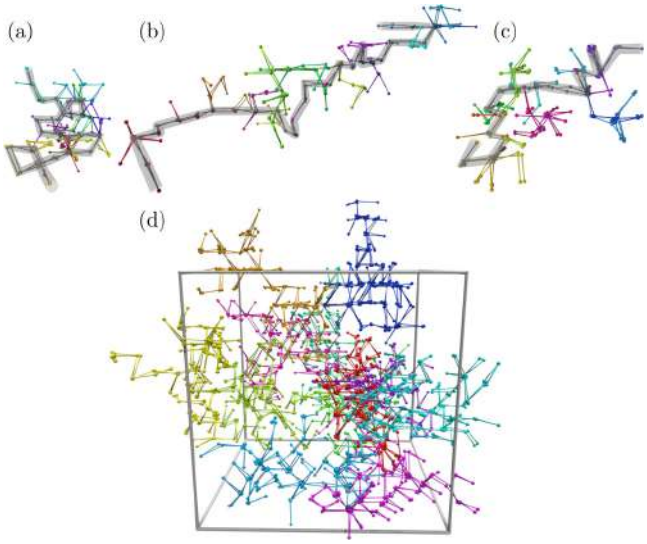


FIG. 3. Equilibrated simulation snapshots of (randomly selected) configurations of the double-folded rings with $N = 216$. Successive segments are represented with a HSV cyclic color map. (a) A single double-folded ideal ring; (b) a single double-folded self-avoiding ring; (c) a single double-folded ring in the melt state. The gray tubes show the longest paths of the trees. All the trees have the same bond scale. The size of the ring in the melt is larger than the ideal tree and smaller than the self-avoiding tree. (d) Sample configuration of the melt with 12 double-folded rings. Each ring is represented with a different color. The snapshots were produced using Blender 2.8 [60]. 3D views of these configurations are available in the Supplemental Materials, videos S1, S2, and S3 [59].

IV. RESULTS AND DISCUSSIONS

After some first qualitative insights in Sec. IV A, we present a quantitative analysis of the conformational statistics and dynamics of our randomly branching, double-folded ring polymers and compare our observations to available exact results and predictions of the scaling theories presented in Secs. IV B and IV C. We close by comparing the dynamics of double-folded rings to the motion of linear chains in the tube model [47] (Sec. IV D).

A. Qualitative insights

Figure 3 illustrates configurations of our lattice model for double-folded ring polymers as they emerge from our simulations after the systems have reached equilibrium. Ring monomers are shown as small spheres which are displaced from their actual position by a small random offset. This representation reveals (i) multiple occupancy of lattice sites and (ii) Double-folding. Specifically, we show a sample configuration of (a) an ideal double-folded ring, (b) a self-avoiding double-folded ring, and (c) a double-folded ring in the melt state with 216 monomers, as well as (d) a view of a corresponding melt configuration (unfolded from the simulation box) for 12 double-folded rings. The gray tubes show the

longest paths along the trees. Three-dimensional (3D) views of these configurations are available in the Supplemental Material, videos S1, S2, and S3 [59]. For the ideal case there is no restriction on the number of monomers on each site. Rings in the melt state and, in particular, isolated self-avoiding rings appear swollen relative to the ideal case. At least qualitatively, this is the expected [36] consequence of excluded volume interactions and the partial screening in melts. In the self-avoiding case, the structure is quite anisotropic and the longest path is aligned along the longest axis. In the melt case, the structure is more compact and spherically symmetric, and the mass is almost equally distributed between the branches.

A sequence of snapshots of the time evolution of a (randomly selected) self-avoiding double-folded ring with $N = 216$ monomers at logarithmic time intervals (indicated on the top left) is available in the Supplemental Material, video S4 [59]. The gray tube shows the longest path along the tree at $T = 0$, where T is set to zero after reaching equilibrium. As a result of the local mass transport (both in space and along the tree), three distinct dynamical regimes can be observed. (i) At short timescales ($T \lesssim 10^2$[MCs]), at the beginning of the video, the relaxation is dominated by the small intra-branch dynamics and the spontaneous formation and deletion of short side-branches in the tree structure. (ii) At intermediate timescales, longer side-branches appear and disappear but the core trunk remains stable. (iii) Near the end of the video ($T \gtrsim 10^6$[MCs]), the entire tree diffuses in space.

Finally, the Supplemental Material video S5 [59] follows the motion of the same ring over even longer times to illustrate that its internal structure completely relaxes on timescales over which the ring diffuses over a distance corresponding to its own size.

B. Conformational statistics

To analyze the secondary and tertiary structure of our double-folded ring polymers as discussed in Secs. II A and II B, we have calculated the tree contour distance $L(i, j)$ and square spatial distance $R^2(i, j)$ between all pairs of ring monomers i and j.

The tree contour distance $L(i, j)$ is defined as the length of the shortest path on the tree connecting i and j. $L(i, j)$ only depends on the tree connectivity and is completely independent of the spatial embedding of the double-folded ring polymer (details in the Appendix). Conversely, the calculation of the spatial distance, $R^2(i, j)$, is straightforward given the monomer positions and completely independent of the secondary structure.

Nevertheless, $L(i, j)$ and $R^2(i, j)$ are closely related, since the configurational statistics of the shortest path between two monomers on the tree is expected to follow a typical power-law relation, $\langle R^2(L)\rangle \sim L^{2\nu_{\rm path}}$, for linear chains [Eq. (5)]. For ideal chains, $\nu_{\rm path} = 1/2$ so that $\langle R^2(L_{ij})\rangle \sim L_{ij}$. Excluded volume interactions cause a characteristic swelling with $\nu_{\rm path} > 1/2$. To allow for a direct comparison, the various panels in Figs. 4–7 with our results for the secondary and tertiary structure always show corresponding data for these two quantities side by side.

Results for averaging $L(n)$ and $R^2(n)$ for the three systems under investigation over monomer pairs with identical ring

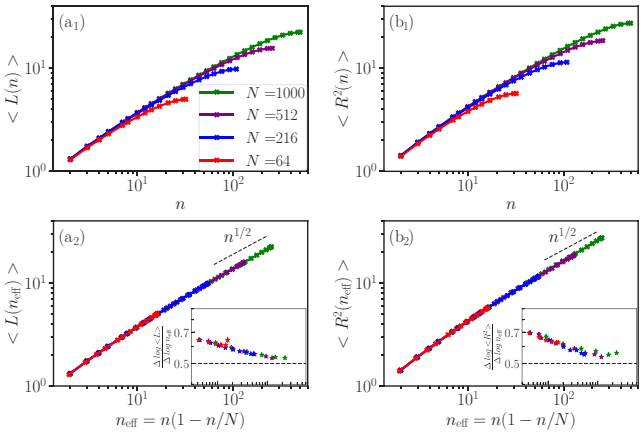


FIG. 4. Conformational statistics of ideal double-folded rings for four different chain lengths (described in the legend). Data are shown for ring contour distances up to $N/2$. Column (a) are the average values of the tree contour distances between all possible pairs of monomers, $\langle L(n)\rangle \sim n^{\rho}$. Column (b) plots the squared internal distances as a function of n, $\langle R(n)^2\rangle \sim n^{2\nu}$. The exact exponents for the ideal case are $\rho = 1/2$ and $\nu = 1/4$. In panels (a_2) and (b_2) data are plotted as a function of $n_{\rm eff}$, which effectively reduces finite size effects. The straight dashed lines correspond to the expectation scaling exponents. (a_2) and (b_2) insets show the local slopes of the data in panels (a_2) and (b_2), respectively. These effective exponents appear to converge to the theoretical exponents (dashed horizontal lines). Error bars are the same size or smaller than the symbols.

contour distance, $n = |i - j|$, are shown in panels (a_1) and (b_1) of Figs. 4–6. As expected, the results are ring size independent at small scales and reach a plateau on approaching the maximal ring contour distance of $n = N/2$.

Panels (a_2) and (b_2) in Figs. 4–6 show the same data plotted as a function of an effective ring contour distance, $n_{\rm eff} = n(1 - n/N)$, introduced in Sec. II B. This representation reduces the finite ring size effects sufficiently for a

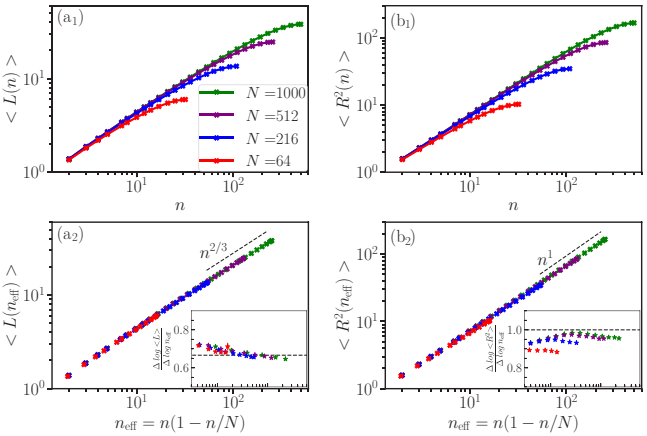


FIG. 5. Conformational statistics of self-avoiding double-folded rings. Column (a) plots the average value of the tree contour distances between all possible pairs of monomers. Flory theory predicts $\langle L(n)\rangle \sim n^{2/3}$. Column (b) plots the squared internal distance as a function of n. The exact scaling exponent is $\langle R(n)^2\rangle \sim n^1$. Notation and symbols are as in Fig. 4.

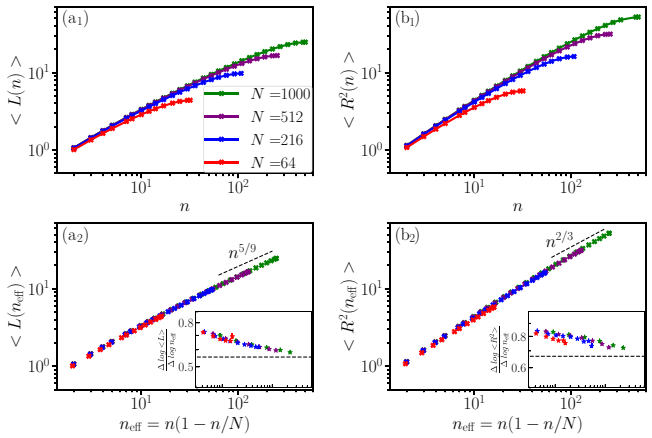


FIG. 6. Conformational statistics of double-folded annealed trees in the melt state. Column (a) plots the average value of the tree contour distances between all possible pairs of monomers. Flory theory predicts, $\langle L(n)\rangle \sim n^{5/9}$. Column (b) plots the squared internal distance as a function of n. Flory theory predicts, $\langle R(n)^2\rangle \sim n^{2/3}$. Notations and symbols are as in Fig. 4.

meaningful comparison with the expected power law relations $\langle L(n)\rangle \sim n^{\rho}$ and $\langle R^2(n)\rangle \sim n^{2\nu}$. The dashed lines have slopes corresponding to the exact value or the predictions of Flory theory for these exponents in the asymptotic limit of infinite ring size. In addition, we have extracted effective exponents by calculating the derivatives using the logarithm of neighboring data points, ($\frac{\Delta \log\langle L(n_{\rm eff})\rangle}{\Delta \log n_{\rm eff}}$) and ($\frac{\Delta \log\langle R^2(n_{\rm eff})\rangle}{\Delta \log n_{\rm eff}}$). Our results are shown in the inset of panels (a_2) and (b_2) of Figs. 4–6. The horizontal lines show again the expected exponents in the asymptotic limit of infinite ring size.

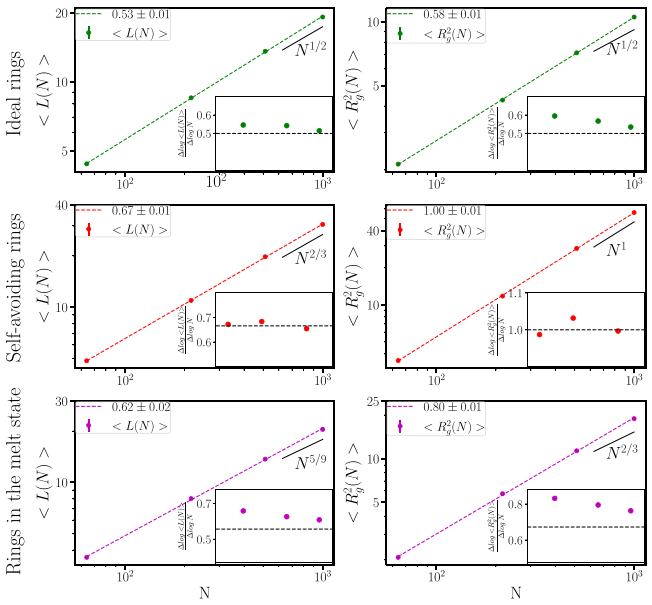


FIG. 7. Conformational statistics of double-folded rings. Left column: average tree contour distance $\langle L\rangle$ as a function of the chain length N. Straight lines correspond to the large-N behavior, $\langle L(N)\rangle \sim N^{\rho}$. Right column: ring mean-square gyration radius $\langle R_g^2\rangle$ as a function of the chain length. Straight lines correspond to the large-N behavior, $\langle R(N)^2\rangle \sim N^{2\nu}$. Insets show the local slopes of the data. These effective exponents appear to converge to the theoretical exponents (dashed horizontal lines). Error bars are the same size or smaller than the symbols.

As a complement, we have analyzed the average tree contour distance $\langle L(N)\rangle$ and the mean-square gyration radius $\langle R_g^2(N)\rangle$ as a function of the chain length (Fig. 7), where the averages of the tree contour and spatial distances is calculated over all monomer pairs irrespective of their distance along the ring. A summary of these values for the studied systems is provided in Table I. Again, we have calculated the local exponents based on the slopes of the data points. The results are shown in the inset of panels of Fig. 7.

Like in comparable simulations of lattice trees [33,34], none of our systems has truly reached the asymptotic regime. Nevertheless, the observed values and trends (which represent crossovers between numerous regimes for linear or branched structures without or with full or partially screened excluded volume interactions [36]) are in good agreement with the theoretical expectations.

C. Dynamics

Having obtained a brief characterization of conformational and structural properties of the double-folded rings, we can now turn our attention to their dynamics. Polymer dynamics is usually analyzed by monitoring the mean-square displacements (MSD) of individual monomers and of the chain centers of mass (CM) with time (as mentioned in Sec. II D). Figures 8–10 show our results for

(i) The total monomer mean-square displacement,

$$g_1(t) = \langle |\mathbf{r}_i(t) - \mathbf{r}_i(0)|^2\rangle,$$

(ii) The monomer mean-square displacement relative to the chain's center of mass,

$$g_2(t) = \langle |\mathbf{r}_i(t) - \mathbf{r}_i(0) - \mathbf{r}_{CM}(t) + \mathbf{r}_{CM}(0)|^2\rangle,$$

(iii) The mean-square displacement of the chain center of mass,

$$g_3(t) = \langle |\mathbf{r}_{CM}(t) - \mathbf{r}_{CM}(0)|^2\rangle,$$

for single ideal and self-avoiding double-folded ring polymers as well as for double-folded ring polymers in the melt state.

Panels (a) in Figs. 8–10 shows a comparison of $g_1(t)$, $g_2(t)$, and $g_3(t)$ for one chain length ($N = 216$). In all three systems we see that $g_1(t)$ is dominated by $g_2(t)$ at early MC times and by $g_3(t)$ at large times. Up to $\tau_0 \approx 10$ MC sweeps (MCs) the monomer and CM motion follow a trivial diffusive dynamics. The characteristic dynamics of double-folded rings discussed in Sec. II D sets in beyond τ_0 and extends up to a maximal time $\tau_{\rm max}$, where the internal monomer motion reaches a plateau, $g_2(t) = 2\langle R_g^2\rangle$, while $g_3(t)$ crosses over to free diffusion.

Panels (b) and (c) in Figs. 8–10 show data for the monomer motion $g_2(t)$ and $g_1(t)$ for different chain lengths, N. In all cases, the monomer MSD in the early and the tree regime are independent of N before crossing over to a ring-size-dependent free diffusion regime. In particular, the data in the tree regime follow an effective power law close to the prediction $t^{\frac{2\nu}{(\rho+2)}}$ [Eq. (17)], indicated by dashed lines. The

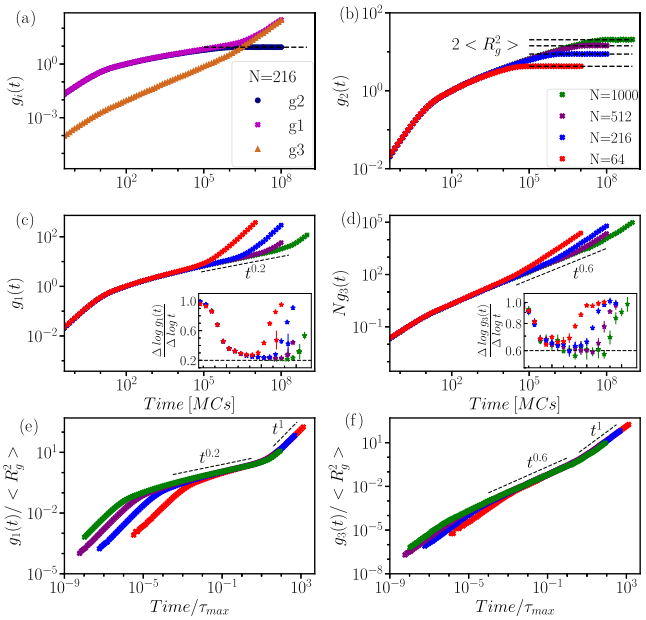


FIG. 8. MSDs for ideal double-folded rings. Panel (a) shows g_1, g_2, and g_3 for the ring with 216 monomers. The horizontal line corresponds to $2 \times \langle R_g^2 \rangle$. Panels (b), (c), and (d) plot $g_2(t)$, $g_1(t)$, and $N \times g_3(t)$ vs time in the unit of MCs, respectively. In panel (b) the horizontal lines correspond to $2 \times \langle R_g^2 \rangle$. In panels (c) and (d) the dashed lines have slopes corresponding to the prediction of the theory, $g_1(t) \sim t^{\frac{2\nu}{(\rho+2)}}$ and $g_3(t) \sim t^{\frac{2\nu+1}{(\rho+2)}}$. (c, d) Insets show the local slopes of the data. The effective exponents appear to converge to the theoretical exponents (dashed horizontal lines). Panels (e) and (f) show rescaled $g_1(t)$ and $g_3(t)$ with the mean-square gyration radii vs the rescaled time with the diffusion relaxation times.

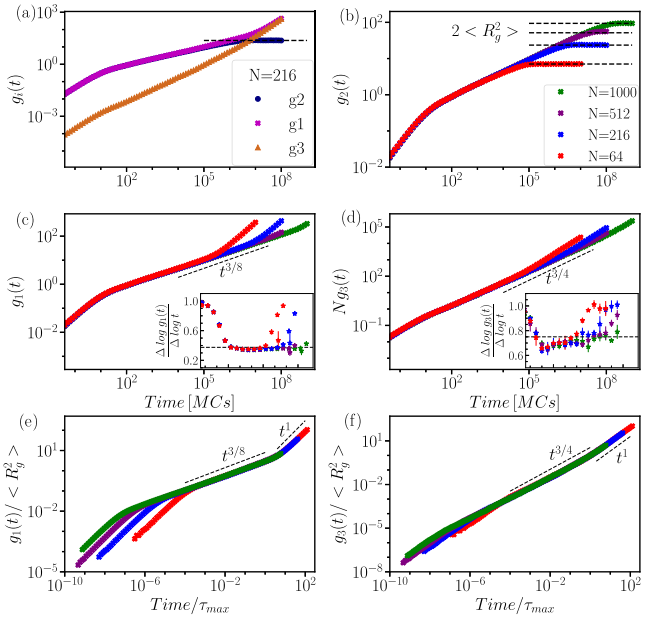


FIG. 9. MSDs for self-avoiding double-folded rings. Notation and symbols are as in Fig. 8.

crossovers between the three regimes are nicely revealed by the effective exponents, $(\frac{\Delta \log (g_1(t))}{\Delta(t)})$, shown in the insets of panels (c). While there are finite ring size effects, they essentially concern the width of the tree regime. The agreement with the expected exponents in the tree regime is excellent.

Panels (d) in Figs. 8–10 show the rescaled dynamics of the center of mass $g_3(t)$, multiplied by the ring size N, to compensate for the expected ring size dependence in the early and in the tree regime, Eq. (18). In both regimes data for different rings sizes collapse indeed on a universal scaling curve. In the tree regime, the data follow an effective power law close to the prediction $t^{\frac{2\nu+1}{(\rho+2)}}$ [Eq. (18)], indicated by dashed lines. Again, the various regimes can be clearly identified when considering the effective exponents, $(\frac{\Delta \log (g_3(t))}{\Delta(t)})$, shown in the insets of panels (c). Interestingly, the effective exponent for the dynamics of self-avoiding double-folded rings initially drops close to the value expected in the ideal case before increasing to a value in good agreement with the theoretical prediction.

Panels (e) and (f) in Figs. 8–10 explore the crossover of the monomer and the CM MSD from the tree to the free diffusion regime. For this purpose we rescale $g_1(t)$ and $g_3(t)$ with the mean-square gyration radii, $\langle R_g^2(N) \rangle$, of the corresponding rings. To rescale the time axis, we empirically defined a "diffusion relaxation time" as the time when the mean-square displacement of the center of mass has moved a distance equal to the radius of gyration, $g_3(\tau_{\max}) = \langle R_g^2 \rangle$. The operation leads to a perfect data collapse for all but the earliest times before the rings enter the tree regime.

The scaling theory predicts that the maximal relaxation time should vary as $\tau_{\max} \sim N^{2+\rho}$ with the ring size, Eq. (13). Besides the "diffusion relaxation time" defined above, we have also tested this relation for the correlation time of the mean-square gyration radius, $\langle R_g^2(N) \rangle$, which characterizes the tertiary structure of our double-folded rings. We have

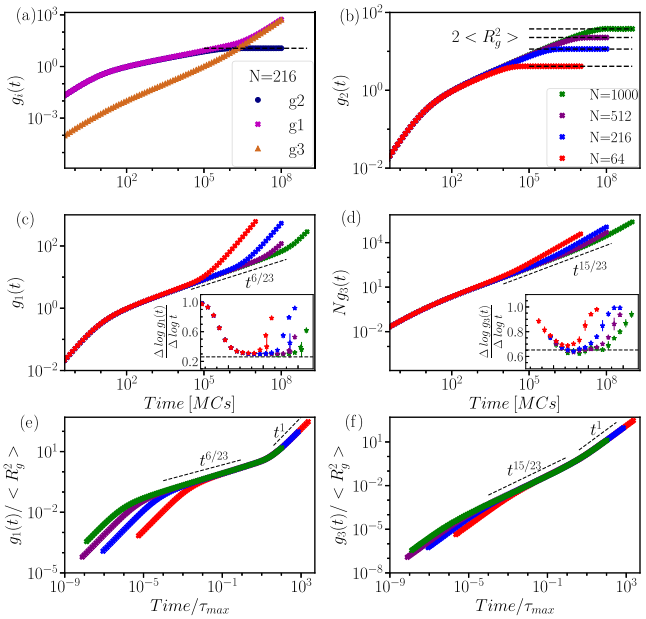


FIG. 10. MSDs for double-folded rings in the melt state. Notation and symbols are as in Fig. 8.

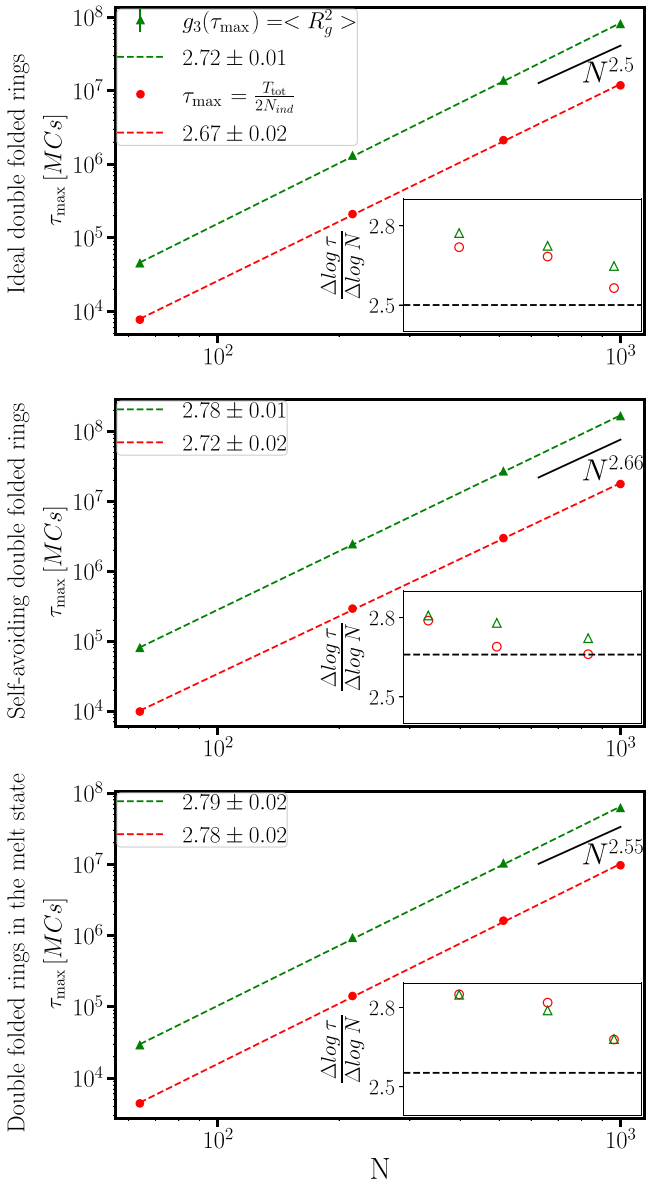


FIG. 11. Relaxation time in units of MCs vs N (chain length). Diffusion relaxation time (triangles) is calculated where $g_3(\tau_{max})$ and $\langle R_g^2 \rangle$ are equal. Configurational relaxation time (circles) is calculated using $\tau_{max} = \frac{T_{tot}}{2N_{ind}}$, where N_{ind} is number of independent samples. The black solid lines indicate the theoretically predicted slopes, $\tau_{max} \sim N^{2+\rho}$, while the dashed lines are the best fit. Insets show the local slopes of the data. These effective exponents appear to converge to the theoretical exponents (dashed horizontal lines).

inferred this configurational relaxation time via the equation $\tau_{max} = \frac{T_{tot}}{2N_{ind}}$ [61] from the number, N_{ind}, of independent samples we have obtained for the observable as estimated from a block averaging procedure [62]. Our results for the three investigated classes of double-folded ring polymers are plotted in Fig. 11. In all three cases, the configurational relaxation times are smaller than the diffusion relaxation times, but both estimates of τ_{max} scale in the same way. The corresponding effective exponents shown in the inset are somewhat larger than expected. While our values are compatible with an approach to the asymptotically expected exponent, a quantitative analysis probably requires data for larger systems.

D. Dynamics of double-folded rings vs linear chains

As summarized in Sec. IV C, the scaling theory [3,4,8] of the dynamics of randomly branching double-folded ring polymers focuses on the mass transport along the longest linear path on the tree. Given the similarities to the tube model for linear chains [e.g., Eq. (13) for the maximal relaxation time, $\tau_{max}(N) \sim N^{\rho+2}$, applies in both cases], one could be tempted to think of the ring motion as a generalized form of reptation along their longest paths.

To test this analogy, we have visualized the equilibrium dynamics of the longest path of a randomly selected self-avoiding double-folded ring (Supplemental Material, video S6 [59]) and of a randomly selected double-folded ring in the melt state (Supplemental Material, video S7 [59]). Note that in these movies time progresses exponentially to cover the large gap between the timescales relevant to motion on the monomer and on the ring scale, respectively. Furthermore, we show fading images of previous conformations to simplify comparisons with the current conformation.

At early times, the behavior is very similar to contour length fluctuations for linear chains in the tube model: the bulk of the longest path remains unchanged, while the path ends fluctuate. But after a while much more drastic changes appear, where the longest path appears to jump in space. For a linear chain such a jump of the primitive chain would necessarily require a corresponding transport of mass. For our rings the movie insets show that the jumps in the position of the longest path are not accompanied by major changes in the ring configurations. Instead, the jumps are due to the continuous redistribution of mass between the side branches, which at some point cause a substantially different path to outgrow the original longest path.

There are qualitative differences between the two types of ring systems. In the self-avoiding case, during relaxation, short side branches relax first. This manifests itself at the ends of the backbone whose center portion remains unchanged. At later times, longer side branches relax and randomly one or two side branches start to grow. Beyond the relaxation time the backbone changes as a whole as a result of modified branching. In the melt case, the structure is more compact and spherically symmetric. The mass is almost equally distributed between the branches. Hence, a small change in the side branches may immediately lead to a completely new path for the longest backbone. This can be observed by comparing videos S6 and S7 [59]. A quantitative study [35] pursued the statistics of the branch weight distributions in the melt state, but a study of the dynamics of the distribution of side branches is still required.

Since it is not possible to follow the temporal evolution of a particular path on an internally rearranging tree, it is easier to focus on pairs of monomers (i^*, j^*), which at some time, $t = 0$, are located on opposite ends of the longest path on the primitive tree. Typically these monomers are also relatively remote along the ring contour, their contour distances following a broad distribution of around $|i^* - j^*| = N/2$. In the movie insets we used bigger spheres to mark two such pairs of

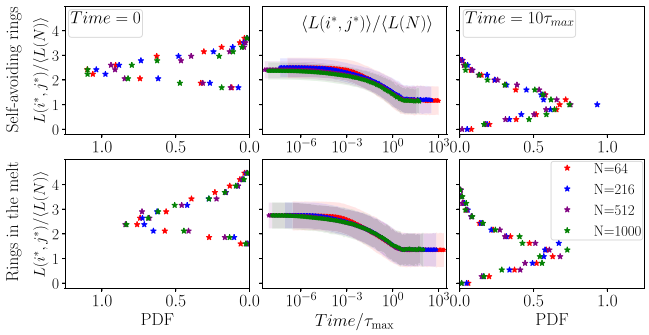


FIG. 12. Tree contour distances between monomers (i^*, j^*) flanking the longest path on the tree at an arbitrarily chosen time $t = 0$ after equilibration. Left-hand side column: Rescaled probability distributions of the longest path length, $L_{max} = L(i^*, j^*, t = 0)$. Middle column: Rescaled time evolution of $\langle L(i^*, j^*, t)\rangle$. Right-hand side column: Rescaled probability distributions of the tree contour distance, $L(i^*, j^*, t = 10\tau_{max})$, between (i^*, j^*) after all memory of the initial state at $t = 0$ is lost. Top row: Self-avoiding double-folded rings. Bottom row: Double-folded rings in the melt state.

monomers which flank the longest path at the beginning and the end of the visualized sequences.

Figure 12 presents a quantitative analysis of the tree contour distance $L(i^*, j^*, t)$ between these monomers rescaled by the average tree contour distance $\langle L(N)\rangle$ reported in Table I. The panels on the left-hand side shows the distribution of the tree contour distances at $t = 0$, i.e., for $L(i^*, j^*, t = 0) \equiv L_{max}(t = 0)$. Results for different ring sizes superpose, indicating that both the average and width of the distribution scale with $\langle L(N)\rangle \sim N^{\rho}$. In contrast, for linear chains contour length fluctuations, $\sqrt{\langle \delta L_{pp}^2\rangle} \sim \sqrt{N}$, become asymptotically negligible compared to the average length, $\langle L_{pp}\rangle \sim N$, of the primitive paths [31]. The central panels in Fig. 12 show the decay of $\langle L(i^*, j^*, t)\rangle$ from a value of $(2.44 \pm 0.03$ to $1.15 \pm 0.01) \times \langle L(N)\rangle$ for self-avoiding rings and $(2.77 \pm 0.01$ to $1.33 \pm 0.01) \times \langle L(N)\rangle$ for rings in the melt over a timescale of the order of the "diffusion relaxation time," τ_{max}. Again, the suitably rescaled data for different ring sizes exhibit a reasonable collapse. The true extent of the (tree) "contour length fluctuations" is better represented by the distribution functions of $L(i^*, j^*, t = 10\tau_{max})$ shown in the panels on the right-hand side of Fig. 12. While the monomers (i^*, j^*) located at opposite ends of the longest path on the tree at $t = 0$ have a finite chance to form secondary structure contacts [50] with $L(i^*, j^*, t) \equiv 0$ at later times, a corresponding deep retraction of one end of a linear chain to the opposite end of the tube is exponentially rare [31].

We conclude that randomly branching double-folded ring polymers move quite differently from reptating linear chains. In particular, the dynamics of rings of all sizes is dominated by the analog of contour length fluctuations occurring simultaneously between all pairs of monomers of the rings.

V. SUMMARY AND CONCLUSION

Supercoiled DNA, crumpled interphase chromosomes, and topologically constrained ring polymers often adopt treelike, double-folded, randomly branching configurations. To explore the statistical and dynamical properties of such objects, we have performed Monte Carlo simulations of a suitable elastic lattice polymer model which accounts for double folding [4,43], the local accumulation of contour length on the primitive tree [4,40,44–46], as well as excluded volume interactions [44–46]. In particular, we have studied single ideal double-folded rings, single self-avoiding double-folded rings, and double-folded rings in the melt state.

In our simulations, side branches of the primitive tree characterizing the double-folded rings are spontaneously created and deleted as a consequence of the local monomer motion. Since the diffusive mass transport is local both in space and on the connectivity graph of the tree, we expect our systems to fall into the same universality class as the experimental target systems.

The observed static properties are in good agreement with exact results and predictions of Flory theory for randomly branching polymers. For example, in the melt state rings adopt compact configurations and exhibit territorial behavior. In particular, the model reproduces results from previous simulations of double-folded ring polymers [50], which were wrapped *a posteriori* around randomly branching polymers generated in corresponding simulations on the tree level [33–35] in an attempt to devise a numerically efficient multiscale approach to the simulation of nonconcatenated ring polymer melts and interphase nuclei [11,12].

The present approach offers the advantage that the dynamics of the ring degrees of freedom can be followed together with the evolution of the tree structure. This is a key feature for the simulation of copolymer [63] models of crumpled [14,15] or supercoiled [17,19,20] interphase chromosomes. Here we have used this information to show that the diffusion of the monomers and the rings' center of mass are in excellent agreement with the predictions of a recent scaling theory [8]. Furthermore, we have explored a possible analogy between the motion of randomly branching double-folded ring polymers and reptating linear chains. While there exist formal similarities on the scaling level, we conclude that the dynamics of rings is rather dominated by the analog of contour length fluctuations, which constantly modify the distances between all monomers on the tree over a wide range.

ACKNOWLEDGMENTS

This work was only possible by generous grants of computer time by the PSMN computer center of the ENS-Lyon. We thank Angelo Rosa and Ivan Junier for valuable discussions and Ali Farnudi for his technical support. The authors would like to acknowledge the networking support by the "European Topology Interdisciplinary Action" (EUTOPIA) CA17139. R.E. acknowledges support by the National Science Foundation under Grant No. NSF PHY-1748958.

APPENDIX: CALCULATING TREE CONTOUR DISTANCES FOR TIGHTLY DOUBLE-FOLDED RINGS

1. Bridge bonds and tree contour distances

An example of a tree is illustrated in Fig. 13(a). The tree connectivity can be mapped on a circle as in Fig. 13(b), where consecutive monomers are represented with numbers. The

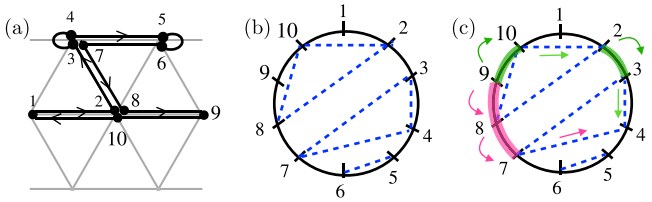


FIG. 13. (a) An example of a tree with side branches. (b) Ring connectivity of the example tree (a) mapped on a circle together with bridge bonds (dashed lines) that are formed during the "burning" process. (c) Illustration of the distance between monomers 9 and 4 following the bridge bonds along the way (green: clockwise, and pink: counterclockwise).

tree contour distance between two monomers of a double-folded ring can be calculated by following the ring contour while ignoring all the double-folded side branches along the way. Bridge bonds (represented with dashed lines) between monomers on each tree site can be defined to mark the location of possible side branches. The tree contour distance between any pair of monomers can be calculated by counting the number of steps taken on the circle and using the bridge bonds as shortcuts (shortcuts do not add to the number of steps). Obviously, the choice of direction (clockwise or counterclockwise) should result in the same contour length. For example, the tree contour distance between monomers 9 and 4 in Fig. 13(c) is 2.

2. Bridge bonds from spatial colocalization

In systems with excluded volume interactions, the identification of the bridge bonds is straightforward as they simply connect monomers which are colocalized in space (Fig. 13). However, as illustrated in Fig. 14, this method fails in the ideal case, where incorrectly identified extra bridges [red dashed line in Fig. 14(b) lead to lattice animal-like connectivity graphs containing falsely identified loops.

3. Bridge bonds from an analysis of the local connectivity

We used a "pinching" variant of the "burning" algorithm [25,33,64] that takes advantage of the local connectivity information. As the algorithm operates by iteratively removing (pinching off) branch tips, it avoids the false identification of loops. The protocol to find the bridges layer by layer is as follows:

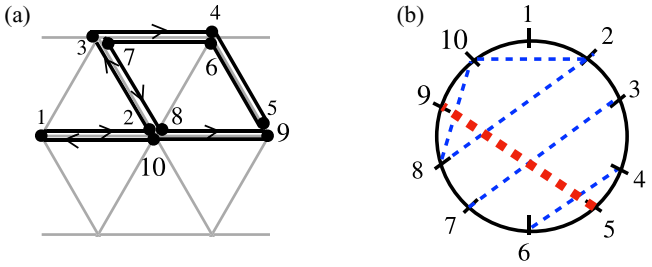


FIG. 14. (a) An example of the ideal double-folded ring. (b) Corresponding bridges. If all monomers on a site have bridges defined between them, it leads to an extra wrong bridge (red dashed line).

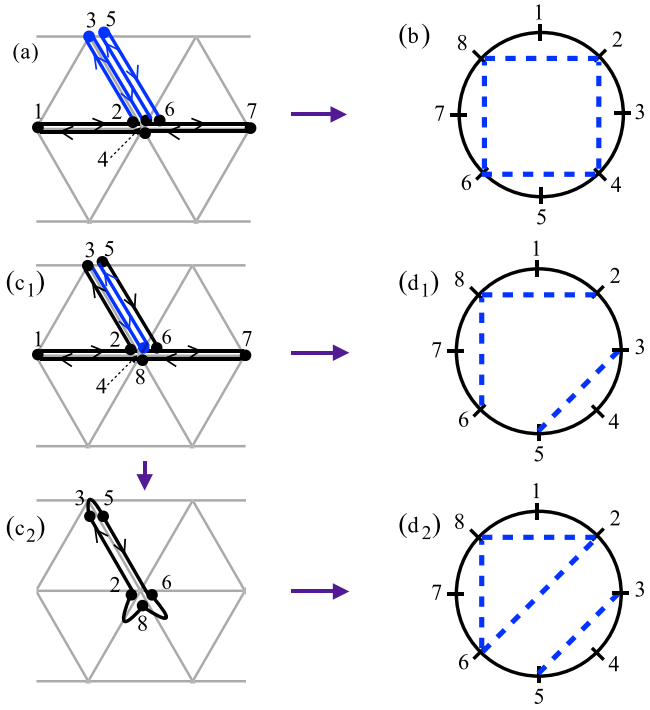


FIG. 15. An example of branch tip detection in an ideal double-folded ring. (a) and (c_1): Examples of side branches in the ideal case. Ambiguous branch tips and their corresponding bonds are shown in blue. The tree structure is the same, but different interpretations of branch tips are possible. (b) Bridge bonds corresponding to interpretation (a). Removal of the first layer of tips in (c_1) results in the formation of bridge bonds (d_1) and the tree structure (c_2). The removal of the second layer results in the completion of the bridge bond detection displayed in (d_2).

(A) Make a list of all the tree branch tips. A branch tip is defined as a monomer with attached bonds pointing in opposite directions, ignoring neighbor loops (with zero-length bonds). For example, in Fig. 13(a) monomers 1, 9, and 6 (5 is a loop) are branch tips.

(B) Randomly select a branch tip and remove it from the ring. For example, if the branch tip 9 is randomly selected, monomer 9 is removed and a bridge is defined between monomers 8 and 10.

(C) Repeat steps (A) and (B) until the branch tip list is exhausted.

(D) Steps (A)–(C) result in the removal of one "layer" of side branches. Repeat these steps (remove layer by layer) until all the branch tips are removed and all the bridge bonds are found.

We verified that the pinching algorithm reproduces the results from the spatial colocalization for double-folded rings with excluded volume interactions.

4. Ambiguities in the tree structure for ideal double-folded rings

The tree connectivity is not uniquely defined for the ideal double-folded rings. There is an intrinsic ambiguity in how side branches and the underlying tree structure is defined. An example is given in Figs. 15(a) and 15(c1), where branch tips

under consideration and their attached bonds are colored blue. They both have the same tree structure, but the number of branch tips is open for interpretation:

(1) Figure 15(a): Branch tips are 1, 3 (pointing up), 5 (pointing up), and 7. During the burning of the first layer, all the branch tips are removed, which results in the bridging bonds shown in Fig. 15(b).

(2) Figure 15(c1): Branch tips are 1, 7, and 4 (pointing down from the tree node {3, 5}). Figure 15(d1) shows the bridges corresponding to the burning of the first layer. In the next layer of burning, Fig. 15(c2), the bridge between monomers 2 and 6 is formed, Fig. 15(d2).

(1) and (2) are random outcomes of step (B) that result in different bridging bonds and therefore different tree contour lengths, as shown in Figs. 15(b) and 15(d2). In practice, we repeat the described procedure multiple times and consider the shortest tree contour distances among different interpretations.

[1] A. Khokhlov and S. Nechaev, Polymer chain in an array of obstacles, Phys. Lett. A **112**, 156 (1985).

[2] M. Cates and J. Deutsch, Conjectures on the statistics of ring polymers, J. Phys. **47**, 2121 (1986).

[3] M. Rubinstein, Dynamics of Ring Polymers in the Presence of Fixed Obstacles, Phys. Rev. Lett. **57**, 3023 (1986).

[4] S. P. Obukhov, M. Rubinstein, and T. Duke, Dynamics of a Ring Polymer in a Gel, Phys. Rev. Lett. **73**, 1263 (1994).

[5] J. D. Halverson, W. B. Lee, G. S. Grest, A. Y. Grosberg, and K. Kremer, Molecular dynamics simulation study of nonconcatenated ring polymers in a melt. I. Statics, J. Chem. Phys. **134**, 204904 (2011).

[6] M. Lang, Ring conformations in bidisperse blends of ring polymers, Macromolecules **46**, 1158 (2013).

[7] A. Y. Grosberg, Annealed lattice animal model and Flory theory for the melt of non-concatenated rings: Towards the physics of crumpling, Soft Matter **10**, 560 (2014).

[8] J. Smrek and A. Y. Grosberg, Understanding the dynamics of rings in the melt in terms of the annealed tree model, J. Phys.: Condens. Matter **27**, 064117 (2015).

[9] T. Ge, S. Panyukov, and M. Rubinstein, Self-similar conformations and dynamics in entangled melts and solutions of nonconcatenated ring polymers, Macromolecules **49**, 708 (2016).

[10] J. Smrek and A. Y. Grosberg, Minimal surfaces on unconcatenated polymer rings in melt, ACS Macro Lett. **5**, 750 (2016).

[11] A. Rosa and R. Everaers, Ring Polymers in the Melt State: The Physics of Crumpling, Phys. Rev. Lett. **112**, 118302 (2014).

[12] R. D. Schram, A. Rosa, and R. Everaers, Local loop opening in untangled ring polymer melts: A detailed Feynman test of models for the large scale structure, Soft Matter **15**, 2418 (2019).

[13] A. M. Gutin, A. Y. Grosberg, and E. I. Shakhnovich, Polymers with annealed and quenched branchings belong to different universality classes, Macromolecules **26**, 1293 (1993).

[14] A. Rosa and R. Everaers, Structure and dynamics of interphase chromosomes, PLoS Comput. Biol. **4**, e1000153 (2008).

[15] E. Lieberman-Aiden, N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander, and J. Dekker, Comprehensive mapping of long-range interactions reveals folding principles of the human genome, Science **326**, 289 (2009).

[16] J. F. Marko and E. D. Siggia, Statistical mechanics of supercoiled DNA, Phys. Rev. E **52**, 2912 (1995).

[17] C. J. Dorman, DNA supercoiling and bacterial gene expression, Sci. Prog. **89**, 151 (2006).

[18] 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 (2011).

[19] T. Lepage, F. Képès, and I. Junier, Thermodynamics of long supercoiled molecules: Insights from highly efficient Monte Carlo simulations, Biophys. J. **109**, 135 (2015).

[20] T. Lepage and I. Junier, A polymer model of bacterial supercoiled DNA including structural transitions of the double helix, Physica A **527**, 121196 (2019).

[21] L. Liu and C. Hyeon, Contact statistics highlight distinct organizing principles of proteins and RNA, Biophys. J. **110**, 2320 (2016).

[22] J. Kelly, A. Y. Grosberg, and R. Bruinsma, Sequence dependence of viral RNA encapsidation, J. Phys. Chem. B **120**, 6038 (2016).

[23] S. W. Singaram, A. Gopal, and A. Ben-Shaul, A Prüfer-sequence based algorithm for calculating the size of ideal randomly branched polymers, J. Phys. Chem. B **120**, 6231 (2016).

[24] M. Rubinstein and R. Colby, *Polymer Physics* (Oxford University Press, Oxford, 2003).

[25] A. Aharony and D. Stauffer, *Introduction to Percolation Theory* (Taylor & Francis, London, 2003).

[26] G. Parisi and N. Sourlas, Critical Behavior of Branched Polymers and the Lee-Yang Edge Singularity, Phys. Rev. Lett. **46**, 871 (1981).

[27] M. E. Fisher, Yang-Lee Edge Singularity and ϕ^3 Field Theory, Phys. Rev. Lett. **40**, 1610 (1978).

[28] D. A. Kurtze and M. E. Fisher, Yang-Lee edge singularities at high temperatures, Phys. Rev. B **20**, 2785 (1979).

[29] J. F. A. Bovier and U. Glaus, Branched polymers and dimensional reduction, in *Critical Phenomena, Random Systems, Gauge Theories*, edited by K. Osterwalder and R. Stora (North-Holland, Amsterdam, 1884).

[30] P.-G. De Gennes, *Scaling Concepts in Polymer Physics* (Cornell University Press, Ithaca, New York, 1979).

[31] M. Doi and S. F. Edwards, *The Theory of Polymer Dynamics* (Oxford University Press, New York, 1986).

[32] A. Y. Grosberg and A. R. Khokhlov, *Statistical Physics of Macromolecules* (AIP Press, New York, 1994).

[33] A. Rosa and R. Everaers, Computer simulations of randomly branching polymers: Annealed versus quenched branching structures, J. Phys. A: Math. Theor. **49**, 345001 (2016).

[34] A. Rosa and R. Everaers, Computer simulations of melts of randomly branching polymers, J. Chem. Phys. **145**, 164906 (2016).

[35] A. Rosa and R. Everaers, Beyond Flory theory: Distribution functions for interacting lattice trees, Phys. Rev. E **95**, 012117 (2017).

[36] R. Everaers, A. Y. Grosberg, M. Rubinstein, and A. Rosa, Flory theory of randomly branched polymers, Soft Matter **13**, 1223 (2017).

[37] W. A. Seitz and D. J. Klein, Excluded volume effects for branched polymers: Monte Carlo results, J. Chem. Phys. **75**, 5190 (1981).

[38] E. J. Janse van Rensburg and N. Madras, A nonlocal Monte Carlo algorithm for lattice trees, J. Phys. A: Math. Gen. **25**, 303 (1992).

[39] D. Landau and K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics* (Cambridge University Press, New York, 2005).

[40] M. Newman and G. Barkema, *Monte Carlo Methods in Statistical Physics* (Clarendon Press, Oxford, England, 1999).

[41] W. Krauth, *Statistical Mechanics: Algorithms and Computations*, ser. EBSCO ebook Academic Collection (Oxford University Press, Oxford, England, 2006).

[42] P. C. Hohenberg and B. I. Halperin, Theory of dynamic critical phenomena, Rev. Mod. Phys. **49**, 435 (1977).

[43] K. E. Evans and S. F. Edwards, Computer simulation of the dynamics of highly entangled polymers, Part 1.—Equilibrium dynamics, J. Chem. Soc., Faraday Trans. 2 **77**, 1891 (1981).

[44] A. van Heukelum and G. T. Barkema, Reaching large lengths and long times in polymer dynamics simulations, J. Chem. Phys. **119**, 8197 (2003).

[45] V. Hugouvieux, M. A. V. Axelos, and M. Kolb, Amphiphilic multiblock copolymers: From intramolecular pearl necklace to layered structures, Macromolecules **42**, 392 (2009).

[46] R. D. Schram, G. T. Barkema, and H. Schiessel, On the stability of fractal globules, J. Chem. Phys. **138**, 224901 (2013).

[47] P. G. de Gennes, Reptation of a polymer chain in the presence of fixed obstacles, J. Chem. Phys. **55**, 572 (1971).

[48] D. Michieletto, On the tree-like structure of rings in dense solutions, Soft Matter **12**, 9485 (2016).

[49] R. Schram and G. Barkema, Simulation of ring polymer melts with GPU acceleration, J. Comput. Phys. **363**, 128 (2018).

[50] A. Rosa and R. Everaers, Conformational statistics of randomly branching double-folded ring polymers, Eur. Phys. J. E **42**, 7 (2019).

[51] B. H. Zimm and W. H. Stockmayer, The dimensions of chain molecules containing branches and rings, J. Chem. Phys. **17**, 1301 (1949).

[52] P.-G. De Gennes, Statistics of branching and hairpin helices for the dAT copolymer, Biopolymers **6**, 715 (1968).

[53] J. Isaacson and T. C. Lubensky, Flory exponents for generalized polymer problems, J. Phys. Lett. **41**, 469 (1980).

[54] M. Daoud and J. F. Joanny, Conformation of branched polymers, J. Phys. **42**, 1359 (1981).

[55] P. J. Flory, *Principles of Polymer Chemistry* (Cornell University Press, Ithaca, New York, 1953).

[56] M. Rubinstein, Discretized Model of Entangled-Polymer Dynamics, Phys. Rev. Lett. **59**, 1946 (1987).

[57] N. C. Karayiannis, V. G. Mavrantzas, and D. N. Theodorou, A Novel Monte Carlo Scheme for the Rapid Equilibration of Atomistic Model Polymer Systems of Precisely Defined Molecular Architecture, Phys. Rev. Lett. **88**, 105503 (2002).

[58] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, Equation of state calculations by fast computing machines, J. Chem. Phys. **21**, 1087 (1953).

[59] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.104.014501 for generated videos.

[60] R. Hess, *Blender Foundations: The Essential Guide to Learning Blender 2.6* (Focal Press, Waltham, MA, 2010).

[61] K. Binder, Time-dependent Ginzburg-Landau theory of nonequilibrium relaxation, Phys. Rev. B **8**, 3423 (1973).

[62] D. Frenkel and B. Smit, *Understanding Molecular Simulation*, 2nd ed. (Academic Press, Inc., New York, 2001).

[63] D. Jost, P. Carrivain, G. Cavalli, and C. Vaillant, Modeling epigenome folding: Formation and dynamics of topologically associated chromatin domains, Nucleic Acids Res. **42**, 9553 (2014).

[64] H. J. Herrmann, D. C. Hong, and H. E. Stanley, Backbone and elastic backbone of percolation clusters obtained by the new method of "burning," J. Phys. A: Math. Gen. **17**, L261 (1984).

Bibliography

Alber, Frank et al. (2007). "Determining the architectures of macromolecular assemblies". In: *Nature* 450.7170, pp. 683–694. DOI: 10.1038/nature06404. URL: https://doi.org/10.1038/nature06404.

Auhl, Rolf et al. (2003). "Equilibration of long chain polymer melts in computer simulations". In: *The Journal of Chemical Physics* 119.24, pp. 12718–12728. DOI: 10.1063/1.1628670. eprint: https://doi.org/10.1063/1.1628670. URL: https://doi.org/10.1063/1.1628670.

Ay, Ferhat and William S. Noble (2015). "Analysis methods for studying the 3D architecture of the genome". In: *Genome Biology* 16.1, p. 183. DOI: 10.1186/s13059-015-0745-7. URL: https://doi.org/10.1186/s13059-015-0745-7.

Barde, Céline, Nicolas Destainville, and Manoel Manghi (2018). "Energy required to pinch a DNA plectoneme". In: *Phys. Rev. E* 97 (3), p. 032412.

Bates AD, Maxwell A (2005). *DNA Topology*. Oxford: Oxford University Press.

Baù, Davide et al. (2011). "The three-dimensional folding of the -globin gene domain reveals formation of chromatin globules". In: *Nature Structural & Molecular Biology* 18.1, pp. 107–114. DOI: 10.1038/nsmb.1936. URL: https://doi.org/10.1038/nsmb.1936.

Berkum, Nynke L van et al. (2010). "Hi-C: a method to study the three-dimensional architecture of genomes." eng. In: *J Vis Exp* 39. ISSN: 1940-087X (Electronic); 1940-087X (Linking). DOI: 10.3791/1869.

Birnie, Anthony and Cees Dekker (Jan. 2021). "Genome-in-a-Box: Building a Chromosome from the Bottom Up". In: *ACS Nano* 15.1, pp. 111–124. DOI: 10.1021/acsnano.0c07397. URL: https://doi.org/10.1021/acsnano.0c07397.

Bliska, James B. and Nicholas R. Cozzarelli (1987). "Use of site-specific recombination as a probe of DNA structure and metabolism in vivo". In: *Journal of Molecular Biology* 194.2, pp. 205–218. DOI: 10.1016/0022-2836(87)90369-X.

Boles, T C, J H White, and N R Cozzarelli (1990). "Structure of plectonemically supercoiled DNA." eng. In: *J Mol Biol* 213.4, pp. 931–951. ISSN: 0022-2836 (Print); 0022-2836 (Linking). DOI: 10.1016/S0022-2836(05)80272-4.

Brocken, Daan J.W., Mariliis Tark-Dame, and Remus T. Dame (2018). "The organization of bacterial genomes: Towards understanding the interplay between structure and function". In: *Current Opinion in Systems Biology* 8. * Regulatory and metabolic networks * Special Section: Single cell and noise, pp. 137 –143. ISSN: 2452-3100.

DOI: https://doi.org/10.1016/j.coisb.2018.02.007. URL: http://www.sciencedirect.com/science/article/pii/S2452310017302019.

Browning, Douglas F, David C Grainger, and Stephen Jw Busby (2010). "Effects of nucleoid-associated proteins on bacterial chromosome structure and gene expression." eng. In: *Curr Opin Microbiol* 13.6, pp. 773–780. ISSN: 1879-0364 (Electronic); 1369-5274 (Linking).

Bush, Natassja G., Katherine Evans-Roberts, and Anthony Maxwell (2015). "DNA Topoisomerases". In: *EcoSal Plus* 6.2. DOI: 10.1128/ecosalplus.ESP-0010-2014. eprint: https://journals.asm.org/doi/pdf/10.1128/ecosalplus.ESP-0010-2014. URL: https://journals.asm.org/doi/abs/10.1128/ecosalplus.ESP-0010-2014.

Calladine, C. R. and H. R. Drew (1997). *Understanding DNA : the molecule and how it works*. Academic Press, San Diego.

Carrivain, Pascal, Maria Barbi, and Jean-Marc Victor (Feb. 2014). "In Silico Single-Molecule Manipulation of DNA with Rigid Body Dynamics". In: *PLOS Computational Biology* 10.2, pp. 1–13. DOI: 10.1371/journal.pcbi.1003456. URL: https://doi.org/10.1371/journal.pcbi.1003456.

Cebrián, Jorge et al. (2021). "Two-Dimensional Gel Electrophoresis to Study the Activity of Type IIA Topoisomerases on Plasmid Replication Intermediates". In: *Biology* 10.11. ISSN: 2079-7737. DOI: 10.3390/biology10111195. URL: https://www.mdpi.com/2079-7737/10/11/1195.

Champoux, J J (2001). "DNA topoisomerases: structure, function, and mechanism." eng. In: *Annu Rev Biochem* 70, pp. 369–413. ISSN: 0066-4154 (Print); 0066-4154 (Linking). DOI: 10.1146/annurev.biochem.70.1.369.

Charvin, G., D. Bensimon, and V. Croquette (2003). "Single-molecule study of DNA unlinking by eukaryotic and prokaryotic type-II topoisomerases". In: *Proceedings of the National Academy of Sciences* 100.17, pp. 9820–9825. DOI: 10.1073/pnas.1631550100. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.1631550100. URL: https://www.pnas.org/doi/abs/10.1073/pnas.1631550100.

Chodavarapu, Sundari et al. (2008). "Escherichia coli DNA interacts with HU in initiation at the E. coli replication origin." eng. In: *Mol Microbiol* 67.4, pp. 781–792. ISSN: 0950-382X (Print); 0950-382X (Linking). DOI: 10.1111/j.1365-2958.2007.06094.x.

Cobbe, Neville and Margarete M S Heck (2004). "The evolution of SMC proteins: phylogenetic analysis and structural implications." eng. In: *Mol Biol Evol* 21.2, pp. 332–347. ISSN: 0737-4038 (Print); 0737-4038 (Linking). DOI: 10.1093/molbev/msh023.

Corless, Samuel and Nick Gilbert (Apr. 2017). "Investigating DNA supercoiling in eukaryotic genomes". In: *Briefings in Functional Genomics* 16.6, pp. 379–389. ISSN: 2041-2649. DOI: 10.1093/bfgp/elx007. eprint: https://academic.oup.

com/bfg/article-pdf/16/6/379/21735834/elx007.pdf. URL: https://doi.org/10.1093/bfgp/elx007.

Cournac, Axel et al. (2012). "Normalization of a chromosomal contact map." eng. In: *BMC Genomics* 13, p. 436. ISSN: 1471-2164 (Electronic); 1471-2164 (Linking). DOI: 10.1186/1471-2164-13-436.

Dame, R T, C Wyman, and N Goosen (2000). "H-NS mediated compaction of DNA visualised by atomic force microscopy." eng. In: *Nucleic Acids Res* 28.18, pp. 3504–3510. ISSN: 1362-4962 (Electronic); 0305-1048 (Print); 0305-1048 (Linking). DOI: 10.1093/nar/28.18.3504.

Dame, Remus T and Mariliis Tark-Dame (2016). "Bacterial chromatin: converging views at different scales." eng. In: *Curr Opin Cell Biol* 40, pp. 60–65. ISSN: 1879-0410 (Electronic); 0955-0674 (Linking). DOI: 10.1016/j.ceb.2016.02.015.

Daoud, M. and J. F. Joanny (1981). "Conformation of Branched Polymers". In: *J. Physique* 42, pp. 1359–1371.

De Gennes, P.-G. (1968). "Statistics of Branching and Hairpin Helices for the dAT Copolymer". In: *Biopolymers* 6, p. 715.

— (1979). *Scaling Concepts in Polymer Physics*. Ithaca: Cornell University Press.

de Gennes, P.G. (1972). "Exponents for the excluded volume problem as derived by the Wilson method". In: *Physics Letters A* 38.5, pp. 339–340. ISSN: 0375-9601. DOI: https://doi.org/10.1016/0375-9601(72)90149-1. URL: https://www.sciencedirect.com/science/article/pii/0375960172901491.

Dekker, Job et al. (2002). "Capturing Chromosome Conformation". In: *Science* 295.5558, pp. 1306–1311. DOI: 10.1126/science.1067799. eprint: https://www.science.org/doi/pdf/10.1126/science.1067799. URL: https://www.science.org/doi/abs/10.1126/science.1067799.

Delius, H. and A. Worcel (1974). "Electron microscopic visualization of the folded chromosome of Escherichia coli". In: *Journal of Molecular Biology* 82.1, pp. 107–109. ISSN: 0022-2836. DOI: https://doi.org/10.1016/0022-2836(74)90577-4. URL: https://www.sciencedirect.com/science/article/pii/0022283674905774.

Deng, Shuang, Richard A Stein, and N Patrick Higgins (2005). "Organization of supercoil domains and their reorganization by transcription." eng. In: *Mol Microbiol* 57.6, pp. 1511–1521. ISSN: 0950-382X (Print); 1365-2958 (Electronic); 0950-382X (Linking). DOI: 10.1111/j.1365-2958.2005.04796.x.

Dillon, Shane C. and Charles J. Dorman (2010). "Bacterial nucleoid-associated proteins, nucleoid structure and gene expression". In: *Nature Reviews Microbiology* 8.3, pp. 185–195. DOI: 10.1038/nrmicro2261. URL: https://doi.org/10.1038/nrmicro2261.

Dorman, Charles J. (2006). "DNA Supercoiling and Bacterial Gene Expression". In: *Science Progress* 89.3-4, pp. 151–166.

Dorman, Charles J (2014). "Function of nucleoid-associated proteins in chromosome structuring and transcriptional regulation." eng. In: *J Mol Microbiol Biotechnol* 24.5-6, pp. 316–331. ISSN: 1660-2412 (Electronic); 1464-1801 (Linking). DOI: 10.1159/000368850.

Dostie, Josée et al. (2006). "Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements." eng. In: *Genome Res* 16.10, pp. 1299–1309. ISSN: 1088-9051 (Print); 1549-5477 (Electronic); 1088-9051 (Linking). DOI: 10.1101/gr.5571506.

Duprey, Alexandre and Eduardo A. Groisman (2020). "FEDS: a Novel Fluorescence-Based High-Throughput Method for Measuring DNA Supercoiling In Vivo". In: *mBio* 11.4, e01053–20. DOI: 10.1128/mBio.01053-20. eprint: https://journals.asm.org/doi/pdf/10.1128/mBio.01053-20. URL: https://journals.asm.org/doi/abs/10.1128/mBio.01053-20.

Everaers, Ralf et al. (2017). "Flory theory of randomly branched polymers". In: *Soft Matter* 13 (6), pp. 1223–1234. DOI: 10.1039/C6SM02756C.

Flory, P. J. (1953). *Principles of Polymer Chemistry*. Ithaca (NY): Cornell University Press.

Forterre, Patrick et al. (2007). "Origin and evolution of DNA topoisomerases." eng. In: *Biochimie* 89.4, pp. 427–446. ISSN: 0300-9084 (Print); 0300-9084 (Linking). DOI: 10.1016/j.biochi.2006.12.009.

Franklin, Rosalind E. and R. G. Gosling (1953). "Molecular Configuration in Sodium Thymonucleate". In: *Nature* 171.4356, pp. 740–741. DOI: 10.1038/171740a0. URL: https://doi.org/10.1038/171740a0.

Fuller, F B (1978). "Decomposition of the linking number of a closed ribbon: A problem from molecular biology." eng. In: *Proc Natl Acad Sci U S A* 75.8, pp. 3557–3561. ISSN: 0027-8424 (Print); 1091-6490 (Electronic); 0027-8424 (Linking). DOI: 10.1073/pnas.75.8.3557.

Fuller, F. Brock (1971). "The Writhing Number of a Space Curve". In: *Proceedings of the National Academy of Sciences* 68.4, pp. 815–819.

Ghobadpour, Elham et al. (2021). "Monte Carlo simulation of a lattice model for the dynamics of randomly branching double-folded ring polymers". In: *Phys. Rev. E* 104 (1), p. 014501. DOI: 10.1103/PhysRevE.104.014501. URL: https://link.aps.org/doi/10.1103/PhysRevE.104.014501.

Gibson, Elizabeth G, Alexandria A Oviatt, and Neil Osheroff (2020). "Two-Dimensional Gel Electrophoresis to Resolve DNA Topoisomers." eng. In: *Methods Mol Biol* 2119, pp. 15–24. ISSN: 1940-6029 (Electronic); 1064-3745 (Print); 1064-3745 (Linking). DOI: 10.1007/978-1-0716-0323-9{_}2.

Goodsell, David S, Ludovic Autin, and Arthur J Olson (2018). "Lattice Models of Bacterial Nucleoids." eng. In: *J Phys Chem B* 122.21, pp. 5441–5447. ISSN: 1520-5207 (Electronic); 1520-6106 (Print); 1520-5207 (Linking). DOI: 10.1021/acs.jpcb.7b11770.

Grest, Gary S. and Kurt Kremer (1986). "Molecular dynamics simulation for polymers in the presence of a heat bath". In: *Phys. Rev. A* 33 (5), pp. 3628–3631. DOI: 10.1103/PhysRevA.33.3628. URL: https://link.aps.org/doi/10.1103/PhysRevA.33.3628.

Grohens, Théotime, Sam Meyer, and Guillaume Beslon (2022). "A Genome-Wide Evolutionary Simulation of the Transcription-Supercoiling Coupling: extended version". In: *Extended version of the ALIFE21 paper (https://hal.archives-ouvertes.fr/hal-03242696), to appear in the Artificial Life journal.*

Grønbech-Jensen, Niels and Oded Farago (2013). "A simple and effective Verlet-type algorithm for simulating Langevin dynamics". In: *Molecular Physics* 111.8, pp. 983–991. DOI: 10.1080/00268976.2012.760055. eprint: https://doi.org/10.1080/00268976.2012.760055. URL: https://doi.org/10.1080/00268976.2012.760055.

Grosberg, A. et al. (1993). "Crumpled Globule Model of the Three-Dimensional Structure of DNA". In: *Europhysics Letters* 23.5, p. 373.

Grosberg, Alexander Y. (2014). "Annealed lattice animal model and Flory theory for the melt of non-concatenated rings: Towards the physics of crumpling". In: *Soft Matter* 10 (4), pp. 560–565. DOI: 10.1039/C3SM52805G.

Grosberg, Alexander Y., Sergei Nechaev, and Eugene I. Shakhnovich (1988). "The role of topological constraints in the kinetics of collapse of macromolecules". In: *Journal De Physique* 49, pp. 2095–2100.

Guo, Monica S et al. (2021). "High-resolution, genome-wide mapping of positive supercoiling in chromosomes". In: *eLife* 10. Ed. by James M Berger and Naama Barkai, e67236. ISSN: 2050-084X. DOI: 10.7554/eLife.67236. URL: https://doi.org/10.7554/eLife.67236.

Gutin, A. M., A. Yu. Grosberg, and E. I. Shakhnovich (1993). "Polymers with annealed and quenched branchings belong to different universality classes". In: *Macromolecules* 26.6, pp. 1293–1295.

Hess, Roland (2010). *Blender Foundations: The Essential Guide to Learning Blender 2.6*. Focal Press. ISBN: 0240814304.

Higgins, N. Patrick and Alexander V. Vologodskii (2015). "Topological Behavior of Plasmid DNA". In: *Microbiology Spectrum* 3.2, p. 3.2.17. DOI: 10.1128/microbiolspec.PLAS-0036-2014. eprint: https://journals.asm.org/doi/pdf/10.1128/microbiolspec.PLAS-0036-2014. URL: https://journals.asm.org/doi/abs/10.1128/microbiolspec.PLAS-0036-2014.

Hołówka, Joanna and Jolanta Zakrzewska-Czerwińska (2020). "Nucleoid Associated Proteins: The Small Organizers That Help to Cope With Stress". In: *Frontiers in Microbiology* 11. ISSN: 1664-302X. DOI: 10.3389/fmicb.2020.00590. URL: https://www.frontiersin.org/articles/10.3389/fmicb.2020.00590.

Hu, Ming et al. (2012). "HiCNorm: removing biases in Hi-C data via Poisson regression." eng. In: *Bioinformatics* 28.23, pp. 3131–3133. ISSN: 1367-4811 (Electronic); 1367-4803 (Print); 1367-4803 (Linking). DOI: 10.1093/bioinformatics/bts570.

Imakaev, Maxim et al. (2012). "Iterative correction of Hi-C data reveals hallmarks of chromosome organization." eng. In: *Nat Methods* 9.10, pp. 999–1003. ISSN: 1548-7105 (Electronic); 1548-7091 (Print); 1548-7091 (Linking). DOI: 10.1038/nmeth.2148.

Imakaev, Maxim V, Geoffrey Fudenberg, and Leonid A Mirny (2015). "Modeling chromosomes: Beyond pretty pictures." eng. In: *FEBS Lett* 589.20 Pt A, pp. 3031–3036. ISSN: 1873-3468 (Electronic); 0014-5793 (Print); 0014-5793 (Linking). DOI: 10.1016/j.febslet.2015.09.004.

Imakaev, Maxim V. et al. (2015). "Effects of topological constraints on globular polymers". In: *Soft Matter* 11 (4), pp. 665–671. DOI: 10.1039/C4SM02099E. URL: http://dx.doi.org/10.1039/C4SM02099E.

Isaacson, J. and T. C. Lubensky (1980). "Flory exponents for generalized polymer problems". In: *J. Physique Lett.* 41, pp. L469–L471.

Iyer, Balaji VS, Martin Kenward, and Gaurav Arya (2011). "Hierarchies in eukaryotic genome organization: Insights from polymer theory and simulations". In: *BMC Biophysics* 4.1, p. 8. DOI: 10.1186/2046-1682-4-8. URL: https://doi.org/10.1186/2046-1682-4-8.

Janse van Rensburg, E. J. and N. Madras (1992). "A nonlocal Monte Carlo algorithm for lattice trees". In: *J. Phys. A: Math. Gen.* 25, pp. 303–333.

Jeon, Chanil, Youngkyun Jung, and Bae-Yeun Ha (2017). "A ring-polymer model shows how macromolecular crowding controls chromosome-arm organization in Escherichia coli". In: *Scientific Reports* 7.1, p. 11896. DOI: 10.1038/s41598-017-10421-y. URL: https://doi.org/10.1038/s41598-017-10421-y.

Joyeux, Marc (2015). "Compaction of bacterial genomic DNA: clarifying the concepts." eng. In: *J Phys Condens Matter* 27.38, p. 383001. ISSN: 1361-648X (Electronic); 0953-8984 (Linking). DOI: 10.1088/0953-8984/27/38/383001.

Jun, Suckjoon and Bela Mulder (2006). "Entropy-driven spatial organization of highly confined polymers: Lessons for the bacterial chromosome". In: *Proceedings of the National Academy of Sciences* 103.33, pp. 12388–12393. DOI: 10.1073/pnas.0605305103. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.0605305103. URL: https://www.pnas.org/doi/abs/10.1073/pnas.0605305103.

Jun, Suckjoon and Andrew Wright (2010). "Entropy as the driver of chromosome segregation". In: *Nature Reviews Microbiology* 8.8, pp. 600–607. DOI: 10.1038/nrmicro2391. URL: https://doi.org/10.1038/nrmicro2391.

Jung, Youngkyun et al. (2012). "Ring polymers as model bacterial chromosomes: confinement, chain topology, single chain statistics, and how they interact". In: *Soft Matter* 8 (7), pp. 2095–2102. DOI: 10.1039/C1SM05706E. URL: http://dx.doi.org/10.1039/C1SM05706E.

Junier, Ivan, Frédéric Boccard, and Olivier Espéli (2014). "Polymer modeling of the E. coli genome reveals the involvement of locus positioning and macrodomain structuring for the control of chromosome conformation and segregation." eng. In: *Nucleic Acids Res* 42.3, pp. 1461–1473. ISSN: 1362-4962 (Electronic); 0305-1048 (Print); 0305-1048 (Linking). DOI: 10.1093/nar/gkt1005.

Junier, Ivan and Olivier Rivoire (May 2016). "Conserved Units of Co-Expression in Bacterial Genomes: An Evolutionary Insight into Transcriptional Regulation". In: *PLOS ONE* 11.5, pp. 1–25. DOI: 10.1371/journal.pone.0155740. URL: https://doi.org/10.1371/journal.pone.0155740.

Kamien, Randall D. (2002). "The geometry of soft materials: a primer". In: *Rev. Mod. Phys.* 74 (4), pp. 953–971.

Kardar, Mehran (2007). *Statistical Physics of Fields*. Cambridge University Press. DOI: 10.1017/CBO9780511815881.

Kavenoff, R and B C Bowen (1976). "Electron microscopy of membrane-free folded chromosomes from Escherichia coli." eng. In: *Chromosoma* 59.2, pp. 89–101. ISSN: 0009-5915 (Print); 0009-5915 (Linking). DOI: 10.1007/BF00328479.

Kelly, Joshua, Alexander Y. Grosberg, and Robijn Bruinsma (July 2016). "Sequence Dependence of Viral RNA Encapsidation". In: *The Journal of Physical Chemistry B* 120.26, pp. 6038–6050. DOI: 10.1021/acs.jpcb.6b01964. URL: https://doi.org/10.1021/acs.jpcb.6b01964.

Khokhlov, A.R. and S.K. Nechaev (1985). "Polymer chain in an array of obstacles". In: *Physics Letters A* 112.3, pp. 156 –160. ISSN: 0375-9601. DOI: https://doi.org/10.1016/0375-9601(85)90678-4.

Kim, Joongbaek et al. (2004). "Fundamental structural units of the Escherichia coli nucleoid revealed by atomic force microscopy." eng. In: *Nucleic Acids Res* 32.6, pp. 1982–1992. ISSN: 1362-4962 (Electronic); 0305-1048 (Print); 0305-1048 (Linking). DOI: 10.1093/nar/gkh512.

Kirchberger, Paul C., Marian L. Schmidt, and Howard Ochman (2020). "The Ingenuity of Bacterial Genomes". In: *Annual Review of Microbiology* 74.1. PMID: 32692614, pp. 815–834. DOI: 10.1146/annurev-micro-020518-115822. eprint: https://doi.org/10.1146/annurev-micro-020518-115822. URL: https://doi.org/10.1146/annurev-micro-020518-115822.

Klenin, Konstantin and Jörg Langowski (2000). "Computation of writhe in modeling of supercoiled DNA". In: *Biopolymers* 54.5, pp. 307–31.

Kremer, Kurt and Florian Müller-Plathe (2002). "Multiscale simulation in polymer science". In: *Molecular Simulation* 28.8-9, pp. 729–750. DOI: 10.1080/0892702021000002458.

eprint: https://doi.org/10.1080/0892702021000002458. URL: https://doi.org/10.1080/0892702021000002458.

Krogh, Thøger J, Jakob Møller-Jensen, and Christoph Kaleta (2018). "Impact of Chromosomal Architecture on the Function and Evolution of Bacterial Genomes." eng. In: *Front Microbiol* 9, p. 2019. ISSN: 1664-302X (Print); 1664-302X (Electronic); 1664-302X (Linking). DOI: 10.3389/fmicb.2018.02019.

Kuo, Chih-Horng, Nancy A Moran, and Howard Ochman (2009). "The consequences of genetic drift for bacterial genome complexity." eng. In: *Genome Res* 19.8, pp. 1450–1454. ISSN: 1088-9051 (Print); 1549-5477 (Electronic); 1088-9051 (Linking). DOI: 10.1101/gr.091785.109.

Lagomarsino, Marco Cosentino, Olivier Espéli, and Ivan Junier (2015). "From structure to function of bacterial chromosomes: Evolutionary perspectives and ideas for new experiments". In: *FEBS Letters* 589.20, Part A. 3D Genome structure., pp. 2996–3004. ISSN: 0014-5793. DOI: https://doi.org/10.1016/j.febslet.2015.07.002. URL: https://www.sciencedirect.com/science/article/pii/S0014579315005761.

Land, Miriam et al. (2015). "Insights from 20 years of bacterial genome sequencing". In: *Functional & Integrative Genomics* 15.2, pp. 141–161. DOI: 10.1007/s10142-015-0433-4. URL: https://doi.org/10.1007/s10142-015-0433-4.

Le, Tung B. K. et al. (2013). "High-Resolution Mapping of the Spatial Organization of a Bacterial Chromosome". In: *Science* 342.6159, pp. 731–734. DOI: 10.1126/science.1242059. eprint: https://www.science.org/doi/pdf/10.1126/science.1242059. URL: https://www.science.org/doi/abs/10.1126/science.1242059.

Le, Tung Bk and Michael T Laub (2014). "New approaches to understanding the spatial organization of bacterial genomes." eng. In: *Curr Opin Microbiol* 22, pp. 15–21. ISSN: 1879-0364 (Electronic); 1369-5274 (Print); 1369-5274 (Linking). DOI: 10.1016/j.mib.2014.09.014.

Le Guillou, J. C. and J. Zinn-Justin (1980). "Critical exponents from field theory". In: *Phys. Rev. B* 21 (9), pp. 3976–3998. DOI: 10.1103/PhysRevB.21.3976. URL: https://link.aps.org/doi/10.1103/PhysRevB.21.3976.

Lepage, Thibaut and Ivan Junier (2017). "Modeling Bacterial DNA: Simulation of Self-Avoiding Supercoiled Worm-Like Chains Including Structural Transitions of the Helix". In: *The Bacterial Nucleoid: Methods and Protocols*. Ed. by Olivier Espéli. New York, NY: Springer New York, pp. 323–337. ISBN: 978-1-4939-7098-8. DOI: 10.1007/978-1-4939-7098-8_23. URL: https://doi.org/10.1007/978-1-4939-7098-8_23.

— (2019). "A polymer model of bacterial supercoiled DNA including structural transitions of the double helix". In: *Physica A: Statistical Mechanics and its Applications* 527, p. 121196. ISSN: 0378-4371.

Lepage, Thibaut, François Képès, and Ivan Junier (2015). "Thermodynamics of Long Supercoiled Molecules: Insights from Highly Efficient Monte Carlo Simulations". In: *Biophysical Journal* 109.1, pp. 135 –143. ISSN: 0006-3495.

Lieberman-Aiden, Erez et al. (2009a). "Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome". In: *Science* 326.5950, pp. 289–293. DOI: 10.1126/science.1181369. eprint: https://www.science.org/doi/pdf/10.1126/science.1181369. URL: https://www.science.org/doi/abs/10.1126/science.1181369.

— (2009b). "Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome". In: *Science* 326.5950, pp. 289–293. ISSN: 0036-8075. DOI: 10.1126/science.1181369.

Lioy, Virginia S, Ivan Junier, and Frédéric Boccard (2021). "Multiscale Dynamic Structuring of Bacterial Chromosomes." eng. In: *Annu Rev Microbiol* 75, pp. 541–561. ISSN: 1545-3251 (Electronic); 0066-4227 (Linking). DOI: 10.1146/annurev-micro-033021-113232.

Lioy, Virginia S et al. (2018). "Multiscale Structuring of the E. coli Chromosome by Nucleoid-Associated and Condensin Proteins." eng. In: *Cell* 172.4, pp. 771–783. ISSN: 1097-4172 (Electronic); 0092-8674 (Linking). DOI: 10.1016/j.cell.2017.12.027.

Lioy, Virginia S. et al. (2020). "Distinct Activities of Bacterial Condensins for Chromosome Management in Pseudomonas aeruginosa". In: *Cell Reports* 33.5, p. 108344. DOI: 10.1016/j.celrep.2020.108344.

Liu, L F and J C Wang (1987). "Supercoiling of the DNA template during transcription." eng. In: *Proc Natl Acad Sci U S A* 84.20, pp. 7024–7027. ISSN: 0027-8424 (Print); 1091-6490 (Electronic); 0027-8424 (Linking). DOI: 10.1073/pnas.84.20.7024.

Lomvardas, Stavros et al. (2006). "Interchromosomal Interactions and Olfactory Receptor Choice". In: *Cell* 126.2, pp. 403–413. ISSN: 0092-8674. DOI: https://doi.org/10.1016/j.cell.2006.06.035. URL: https://www.sciencedirect.com/science/article/pii/S0092867406008555.

López, Virginia et al. (2012). "Topo IV is the topoisomerase that knots and unknots sister duplexes during DNA replication." eng. In: *Nucleic Acids Res* 40.8, pp. 3563–3573. ISSN: 1362-4962 (Electronic); 0305-1048 (Print); 0305-1048 (Linking). DOI: 10.1093/nar/gkr1237.

Luijsterburg, Martijn S. et al. (2006). "The architectural role of nucleoid associated proteins in the organization of bacterial chromatin: A molecular perspective". In: *Journal of Structural Biology* 156.2, pp. 262–272. ISSN: 1047-8477. DOI: https://doi.org/10.1016/j.jsb.2006.05.006. URL: https://www.sciencedirect.com/science/article/pii/S1047847706001614.

Lyu, Hongqiang, Erhu Liu, and Zhifang Wu (2020). "Comparison of normalization methods for Hi-C data". In: *BioTechniques* 68.2. PMID: 31588782, pp. 56–64. DOI: 10.

2144/btn-2019-0105. eprint: https://doi.org/10.2144/btn-2019-0105. URL: https://doi.org/10.2144/btn-2019-0105.

Madras, Neal and Gordon Slade (1996). *The Self-Avoiding Walk*. Boston, MA: Birkhäuser Boston. ISBN: 978-1-4612-4132-4. DOI: 10.1007/978-1-4612-4132-4.

Marko, J. F. and E. D. Siggia (1995a). "Statistical mechanics of supercoiled DNA". In: *Phys. Rev. E* 52 (3), pp. 2912–2938.

Marko, J. F. and E.D. Siggia (1995b). "Statistical mechanics of supercoiled DNA". In: *Phys. Rev. E* 52 (3), pp. 2912–2938.

Marko, John F. (2015). "Biophysics of protein–DNA interactions and chromosome organization". In: *Physica A: Statistical Mechanics and its Applications* 418. Proceedings of the 13th International Summer School on Fundamental Problems in Statistical Physics, pp. 126–153. ISSN: 0378-4371.

Marko, John F. and Eric D. Siggia (1994). "Fluctuations and Supercoiling of DNA". In: *Science* 265.5171, pp. 506–508.

Martis B, Shiny et al. (2019). "DNA Supercoiling: an Ancestral Regulator of Gene Expression in Pathogenic Bacteria?" eng. In: *Comput Struct Biotechnol J* 17, pp. 1047–1055. ISSN: 2001-0370 (Print); 2001-0370 (Electronic); 2001-0370 (Linking). DOI: 10.1016/j.csbj.2019.07.013.

McGuffee, Sean R. and Adrian H. Elcock (Mar. 2010). "Diffusion, Crowding Protein Stability in a Dynamic Molecular Model of the Bacterial Cytoplasm". In: *PLOS Computational Biology* 6.3, pp. 1–18. DOI: 10.1371/journal.pcbi.1000694. URL: https://doi.org/10.1371/journal.pcbi.1000694.

McKie, Shannon J, Keir C Neuman, and Anthony Maxwell (2021). "DNA topoisomerases: Advances in understanding of cellular roles and multi-protein complexes via structure-function analysis." eng. In: *Bioessays* 43.4, e2000286. ISSN: 1521-1878 (Electronic); 0265-9247 (Linking). DOI: 10.1002/bies.202000286.

Messelink, Joris J. B. et al. (2021). "Learning the distribution of single-cell chromosome conformations in bacteria reveals emergent order across genomic scales". In: *Nature Communications* 12.1, p. 1963. DOI: 10.1038/s41467-021-22189-x. URL: https://doi.org/10.1038/s41467-021-22189-x.

Meyer, Sam and Guillaume Beslon (2014). "Torsion-mediated interaction between adjacent genes." eng. In: *PLoS Comput Biol* 10.9, e1003785. ISSN: 1553-7358 (Electronic); 1553-734X (Print); 1553-734X (Linking). DOI: 10.1371/journal.pcbi.1003785.

Miriklis, Esther L et al. (2021). "Understanding DNA organization, damage, and repair with super-resolution fluorescence microscopy." eng. In: *Methods Appl Fluoresc* 9.3. ISSN: 2050-6120 (Electronic); 2050-6120 (Linking). DOI: 10.1088/2050-6120/abf239.

Mirkin, Sergei M (2001). "DNA Topology: Fundamentals". In: *eLS*. John Wiley Sons, Ltd. ISBN: 9780470015902. DOI: https://doi.org/10.1038/npg.els.0001038. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1038/

npg.els.0001038. URL: https://onlinelibrary.wiley.com/doi/abs/10.1038/npg.els.0001038.

Mirny, Leonid A. (2011). "The fractal globule as a model of chromatin architecture in the cell". In: *Chromosome Research* 19.1, pp. 37–51. DOI: 10.1007/s10577-010-9177-0. URL: https://doi.org/10.1007/s10577-010-9177-0.

Mondal, Jagannath et al. (2011). "Entropy-Based Mechanism of Ribosome-Nucleoid Segregation in E.coli Cells". In: *Biophysical Journal* 100.11, pp. 2605–2613. DOI: 10.1016/j.bpj.2011.04.030.

Obukhov, Sergei P., Michael Rubinstein, and Thomas Duke (1994). "Dynamics of a ring polymer in a gel". In: *Phys. Rev. Lett.* 73 (9), pp. 1263–1266.

Oluwadare, Oluwatosin, Max Highsmith, and Jianlin Cheng (2019). "An Overview of Methods for Reconstructing 3-D Chromosome and Genome Structures from Hi-C Data". In: *Biological Procedures Online* 21.1, p. 7. DOI: 10.1186/s12575-019-0094-0. URL: https://doi.org/10.1186/s12575-019-0094-0.

Parisi, G. and N. Sourlas (1981). "Critical Behavior of Branched Polymers and the Lee-Yang Edge Singularity". In: *Phys. Rev. Lett.* 46, pp. 871–874.

Postow, Lisa et al. (2004). "Topological domain structure of the Escherichia coli chromosome." eng. In: *Genes Dev* 18.14, pp. 1766–1779. ISSN: 0890-9369 (Print); 0890-9369 (Linking). DOI: 10.1101/gad.1207504.

Rice, Phoebe A et al. (1996). "Crystal Structure of an IHF-DNA Complex: A Protein-Induced DNA U-Turn". In: *Cell* 87.7, pp. 1295–1306. ISSN: 0092-8674. DOI: https://doi.org/10.1016/S0092-8674(00)81824-3. URL: https://www.sciencedirect.com/science/article/pii/S0092867400818243.

Robinow, C and E Kellenberger (1994). "The bacterial nucleoid revisited". eng. In: *Microbiol Rev* 58.2, pp. 211–232. ISSN: 0146-0749 (Print); 0146-0749 (Linking). DOI: 10.1128/mr.58.2.211-232.1994.

Robinow, C. F. (1956). "THE CHROMATIN BODIES OF BACTERIA". In: *Bacteriological Reviews* 20.4, pp. 207–242. DOI: 10.1128/br.20.4.207-242.1956. eprint: https://journals.asm.org/doi/pdf/10.1128/br.20.4.207-242.1956. URL: https://journals.asm.org/doi/abs/10.1128/br.20.4.207-242.1956.

Rosa, Angelo and Ralf Everaers (2008). "Structure and Dynamics of Interphase Chromosomes". In: *PLOS Computational Biology* 4.8, pp. 1–10. DOI: 10.1371/journal.pcbi.1000153.

— (2014). "Ring Polymers in the Melt State: The Physics of Crumpling". In: *Phys. Rev. Lett.* 112 (11), p. 118302. DOI: 10.1103/PhysRevLett.112.118302.

— (2016a). "Computer simulations of melts of randomly branching polymers". In: *The Journal of Chemical Physics* 145.16, p. 164906. DOI: 10.1063/1.4965827.

— (2016b). "Computer simulations of randomly branching polymers: Annealed versus quenched branching structures". In: *Journal of Physics A: Mathematical and Theoretical* 49.34, p. 345001. DOI: 10.1088/1751-8113/49/34/345001.

Rosa, Angelo and Ralf Everaers (2017). "Beyond Flory theory: Distribution functions for interacting lattice trees". In: *Phys. Rev. E* 95 (1), p. 012117. DOI: 10.1103/PhysRevE.95.012117.

— (2019). "Conformational statistics of randomly branching double-folded ring polymers". In: *The European Physical Journal E* 42.1, p. 7. ISSN: 1292-895X. DOI: 10.1140/epje/i2019-11765-3.

Rosa, Angelo and Christophe Zimmer (2014). "Computational models of large-scale genome architecture." eng. In: *Int Rev Cell Mol Biol* 307, pp. 275–349. ISSN: 1937-6448 (Print); 1937-6448 (Linking).

Rovinskiy, Nikolay S. et al. (2019). "Supercoil Levels in E. coli and Salmonella Chromosomes Are Regulated by the C-Terminal 35–38 Amino Acids of GyrA". In: *Microorganisms* 7.3. ISSN: 2076-2607. DOI: 10.3390/microorganisms7030081. URL: https://www.mdpi.com/2076-2607/7/3/81.

Rubinstein, M. and R.H. Colby (2003). *Polymer Physics*. OUP Oxford. ISBN: 9780198520597.

Rubinstein, Michael (1986). "Dynamics of Ring Polymers in the Presence of Fixed Obstacles". In: *Phys. Rev. Lett.* 57 (24), pp. 3023–3026. DOI: 10.1103/PhysRevLett.57.3023.

Sati, Satish and Giacomo Cavalli (2017). "Chromosome conformation capture technologies and their impact in understanding genome function". In: *Chromosoma* 126.1, pp. 33–44. DOI: 10.1007/s00412-016-0593-6. URL: https://doi.org/10.1007/s00412-016-0593-6.

Schram, Raoul D., Gerard T. Barkema, and Helmut Schiessel (2013). "On the stability of fractal globules". In: *The Journal of Chemical Physics* 138.22, p. 224901. DOI: 10.1063/1.4807723.

Servant, Nicolas et al. (2018). "Effective normalization for copy number variation in Hi-C data". In: *BMC Bioinformatics* 19.1, p. 313. DOI: 10.1186/s12859-018-2256-5. URL: https://doi.org/10.1186/s12859-018-2256-5.

Sherratt, David J (2003). "Bacterial chromosome dynamics." eng. In: *Science* 301.5634, pp. 780–785. ISSN: 1095-9203 (Electronic); 0036-8075 (Linking). DOI: 10.1126/science.1084780.

Shin, Jaeoh, Andrey G Cherstvy, and Ralf Metzler (2014). "Mixing and segregation of ring polymers: spatial confinement and molecular crowding effects". In: *New Journal of Physics* 16.5, p. 053047. DOI: 10.1088/1367-2630/16/5/053047. URL: https://dx.doi.org/10.1088/1367-2630/16/5/053047.

Simonis, Marieke et al. (2006). "Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture–on-chip (4C)". In: *Nature Genetics* 38.11, pp. 1348–1354. DOI: 10.1038/ng1896. URL: https://doi.org/10.1038/ng1896.

Skoko, Dunja et al. (2006). "Mechanism of Chromosome Compaction and Looping by the Escherichia coli Nucleoid Protein Fis". In: *Journal of Molecular Biology* 364.4, pp. 777–798. ISSN: 0022-2836. DOI: https://doi.org/10.1016/j.jmb.2006.

09.043. URL: https://www.sciencedirect.com/science/article/pii/S0022283606012496.

Smrek, Jan and Alexander Y Grosberg (2015). "Understanding the dynamics of rings in the melt in terms of the annealed tree model". In: *Journal of Physics: Condensed Matter* 27.6, p. 064117. DOI: 10.1088/0953-8984/27/6/064117.

Swinger, Kerren K and Phoebe A Rice (2004). "IHF and HU: flexible architects of bent DNA." eng. In: *Curr Opin Struct Biol* 14.1, pp. 28–35. ISSN: 0959-440X (Print); 0959-440X (Linking).

Trun NJ, Marko JF (1998). "Architecture of a bacterial chromosome". In: *Asm News* 64.5, pp. 276–283.

Umbarger, Mark A et al. (2011). "The three-dimensional architecture of a bacterial genome and its alteration by genetic perturbation." eng. In: *Mol Cell* 44.2, pp. 252–264. ISSN: 1097-4164 (Electronic); 1097-2765 (Print); 1097-2765 (Linking). DOI: 10.1016/j.molcel.2011.09.010.

Ussery, David W (2002). "DNA Structure: A-, B- and Z-DNA Helix Families". In: *eLS*. John Wiley Sons, Ltd. ISBN: 9780470015902. DOI: https://doi.org/10.1038/npg.els.0003122. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1038/npg.els.0003122. URL: https://onlinelibrary.wiley.com/doi/abs/10.1038/npg.els.0003122.

Valens, Michèle et al. (2004). "Macrodomain organization of the Escherichia coli chromosome." eng. In: *EMBO J* 23.21, pp. 4330–4341. ISSN: 0261-4189 (Print); 1460-2075 (Electronic); 0261-4189 (Linking). DOI: 10.1038/sj.emboj.7600434.

Verma, Subhash C., Zhong Qian, and Sankar L. Adhya (Dec. 2019). "Architecture of the Escherichia coli nucleoid". In: *PLOS Genetics* 15.12, pp. 1–35. DOI: 10.1371/journal.pgen.1008456. URL: https://doi.org/10.1371/journal.pgen.1008456.

Viguera, E. et al. (Jan. 2000). "Visualisation of plasmid replication intermediates containing reversed forks". In: *Nucleic Acids Research* 28.2, pp. 498–503. ISSN: 0305-1048. DOI: 10.1093/nar/28.2.498. eprint: https://academic.oup.com/nar/article-pdf/28/2/498/9904733/280498.pdf. URL: https://doi.org/10.1093/nar/28.2.498.

Viollier, Patrick H et al. (2004). "Rapid and sequential movement of individual chromosomal loci to specific subcellular locations during bacterial DNA replication." eng. In: *Proc Natl Acad Sci U S A* 101.25, pp. 9257–9262. ISSN: 0027-8424 (Print); 1091-6490 (Electronic); 0027-8424 (Linking). DOI: 10.1073/pnas.0402606101.

Visser, Bryan J et al. (Apr. 2022). "Psoralen mapping reveals a bacterial genome supercoiling landscape dominated by transcription". In: *Nucleic Acids Research* 50.8, pp. 4436–4449. ISSN: 0305-1048. DOI: 10.1093/nar/gkac244. eprint: https://academic.oup.com/nar/article-pdf/50/8/4436/43547226/gkac244.pdf. URL: https://doi.org/10.1093/nar/gkac244.

Vologodskii, Alexander and N.R. Cozzarelli (Feb. 1994). "Conformational and Thermodynamic Properties of Supercoiled DNA". In: *Annual review of biophysics and biomolecular structure* 23, pp. 609–43. DOI: 10.1146/annurev.bb.23.060194.003141.

Vologodskii, Alexander V. et al. (1992). "Conformational and thermodynamic properties of supercoiled DNA". In: *Journal of Molecular Biology* 227.4, pp. 1224–1243. ISSN: 0022-2836. DOI: https://doi.org/10.1016/0022-2836(92)90533-P. URL: https://www.sciencedirect.com/science/article/pii/002228369290533P.

Wang, J C (1971). "Interaction between DNA and an Escherichia coli protein omega." eng. In: *J Mol Biol* 55.3, pp. 523–533. ISSN: 0022-2836 (Print); 0022-2836 (Linking). DOI: 10.1016/0022-2836(71)90334-2.

Wang, Xindan, Paula Montero Llopis, and David Z Rudner (2013). "Organization and segregation of bacterial chromosomes." eng. In: *Nat Rev Genet* 14.3, pp. 191–203. ISSN: 1471-0064 (Electronic); 1471-0056 (Print); 1471-0056 (Linking). DOI: 10.1038/nrg3375.

Wang, Xindan et al. (2015). "Condensin promotes the juxtaposition of DNA flanking its loading site in Bacillus subtilis." eng. In: *Genes Dev* 29.15, pp. 1661–1675. ISSN: 1549-5477 (Electronic); 0890-9369 (Print); 0890-9369 (Linking). DOI: 10.1101/gad.265876.115.

Wasim, Abdul, Ankit Gupta, and Jagannath Mondal (Feb. 2021). "A Hi–C data-integrated model elucidates E. coli chromosome's multiscale organization at various replication stages". In: *Nucleic Acids Research* 49.6, pp. 3077–3091. ISSN: 0305-1048. DOI: 10.1093/nar/gkab094. eprint: https://academic.oup.com/nar/article-pdf/49/6/3077/36885065/gkab094.pdf. URL: https://doi.org/10.1093/nar/gkab094.

White, J H and W R Bauer (1986). "Calculation of the twist and the writhe for representative models of DNA." eng. In: *J Mol Biol* 189.2, pp. 329–341. ISSN: 0022-2836 (Print); 0022-2836 (Linking). DOI: 10.1016/0022-2836(86)90513-9.

Wilkins, M H F, A R Stokes, and H R Wilson (1953). "Molecular structure of deoxypentose nucleic acids." eng. In: *Nature* 171.4356, pp. 738–740. ISSN: 0028-0836 (Print); 0028-0836 (Linking). DOI: 10.1038/171738a0.

Wu, Fabai et al. (2019). "Direct imaging of the circular chromosome in a live bacterium". In: *Nature Communications* 10.1, p. 2194. DOI: 10.1038/s41467-019-10221-0. URL: https://doi.org/10.1038/s41467-019-10221-0.

Würtele, Hugo and Pierre Chartrand (2006). "Genome-wide scanning of HoxB1-associated loci in mouse ES cells using an open-ended Chromosome Conformation Capture methodology". In: *Chromosome Research* 14.5, pp. 477–495. DOI: 10.1007/s10577-006-1075-0. URL: https://doi.org/10.1007/s10577-006-1075-0.

Yaffe, Eitan and Amos Tanay (2011). "Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture". In:

Nature Genetics 43.11, pp. 1059–1065. DOI: 10.1038/ng.947. URL: https://doi.org/10.1038/ng.947.

Yildirim, Asli and Michael Feig (2018). "High-resolution 3D models of Caulobacter crescentus chromosome reveal genome structural variability and organization." eng. In: *Nucleic Acids Res* 46.8, pp. 3937–3952. ISSN: 1362-4962 (Electronic); 0305-1048 (Print); 0305-1048 (Linking). DOI: 10.1093/nar/gky141.

Zhao, Zhihu et al. (2006). "Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions". In: *Nature Genetics* 38.11, pp. 1341–1347. DOI: 10.1038/ng1891. URL: https://doi.org/10.1038/ng1891.

Zimm, B. H. and W. H. Stockmayer (1949). "The Dimensions of Chain Molecules Containing Branches and Rings". In: *J. Chem. Phys.* 17, pp. 1301–1314.

Zimmerman, Steven B. and Lizabeth D. Murphy (1996). "Macromolecular crowding and the mandatory condensation of DNA in bacteria". In: *FEBS Letters* 390.3, pp. 245–248.