DISSERTATION

submitted to the

Combined Faculties for the
Natural Sciences and for Mathematics

of the Ruperto-Carola University of Heidelberg
Germany

for the degree of
Doctor of Natural Sciences

put forward by

Dipl.-Phys. Jiying Jia

born in: Shandong

Date of oral examination: 13.11.2019

Modeling and Study of the Organization of Chromosomes

Referees: Prof. Dr. Dieter W. Heermann
Prof. Dr. Ulrich Schwarz

Abstract

Various modeling and simulations have been done to study the organization of chromosomes under different circumstance and for different purposes. In this thesis, we would first study the influence of bending rigidity and spatial confinement on the organization of the chromatin. More concretely, the effect of heterogeneity and the definition of contacts will be addressed. We find that the definition of a contact does not change the asymptotic behavior of the contact probability. The heterogeneity of bending rigidity is shown to render the chain more flexible by comparing the persistent length and the contact probability of homogeneous and heterogeneous chains. In addition, we simulate semiflexible chains in rectangular confinements with different aspect ratios. An oscillation in the contact probability and the orientational correlation function is found due to the spiraling of polymer when the box size is small enough.

The entanglement of chains is another important aspect when studying chromatins. The processes of disentanglement of two flexible chains are studied using the Monte Carlo simulation. Specifically, several measurements such as the inter-contact of chains, dynamic structure factor are analyzed in the process. When only the excluded volume interaction exists in the system, the average time required for segregation is barely influenced by the initial configurations of the two chains according to our results. However, the intertwinement of chains indeed could impede the segregation at a small time scale. The number of contacts inside a self-avoiding chain is also analyzed. It is found that the total number N_c grows linearly with the length of a free chain, while in cubic confinement it grows quadratically. The distribution function of contacts number between two halves $N_c(AB)$ shows a power-law decay behavior and then an exponential decay for a free chain. In confinement, the function has a maximum. As the chain becomes longer, the percentage of inter-half contacts among the total contacts has a power-law decay behavior with an exponent close to -1, which supports that the number of contacts between two halves is finite even when the chain is infinitely long.

Finally, we studied the fractality and the topology in the self-avoiding walks. Specifically, we calculate the fractal dimension and growth rates of the Betti numbers of the system. These growth rates can be viewed as a topological signature for different systems. The intra-contacts of the self-avoiding walk is a subset of the original walk, and we find that this subset may have a slight multifractal property. In addition, the topological exponents are also different from the self-avoiding walk. Further, each contact gives rise to the formation of a loop. To elucidate how these loops influence the structure of the self-avoiding walk, we delete the loops in a similar way to the loop-erased random walk, thus producing a new walk: loop-deleted self-avoiding walk (LDSAW). The critical exponent of LDSAW is approximated by

studying the scaling behavior of mean end-to-end distance, and the dependence of the mean length of LDSAW on the length of the original self-avoiding walk. Afterward, the fractal dimension and growth rates of Betti numbers of this LDSAW are calculated. The same calculations are also performed on the projection and random subsets of self-avoiding walks.

Zusammenfassung

Verschiedene Modellierung und Simulationen wurden durchgeführt, um die Organisation von Chromosomen unter verschiedenen Umständen zu untersuchen und z verschiedene Zwecke. Im In dieser Arbeit untersuchen wir zunächst den Einfluss der Biegesteifigkeit und der räumlichen Begrenzung auf die Organisation von das Chromatin. Konkreter wird auf den Effekt der Heterogenität und die Definition von Kontakten eingegangen. Wir stellen fest, dass die Definition eines Kontakts das asymptotische Verhalten der Kontaktwahrscheinlichkeit nicht verändert. Die Heterogenität der Biegesteifigkeit macht die Kette flexibler, indem die Dauerlänge und die Kontaktwahrscheinlichkeit von homogenen und heterogenen Ketten verglichen werden. Darüber hinaus simulieren wir semiflexible Ketten in rechteckigen Begrenzungen mit unterschiedlichen Aspektverhältnissen. Eine Schwingung in die Kontaktwahrscheinlichkeit und die Orientierungskorrelationsfunktion ist gefunden aufgrund der Spiralbildung des Polymers, wenn die Schachtelgröße klein genug ist.

Die Verflechtung von Ketten ist ein weiterer wichtiger Aspekt bei der Untersuchung von Chromatinen. Die Entflechtungsprozesse zweier flexibler Ketten werden mit der Monte-Carlo-Simulation untersucht. Insbesondere mehrere Messungen wie die Dabei werden die Wechselwirkungen von Ketten und dynamischer Strukturfaktor analysiert. Wenn nur die ausgeschlossene Volumenwechselwirkung im System vorhanden ist, wird die durchschnittliche Zeit, die für die Entmischung benötigt wird, nach unseren Ergebnissen kaum von den anfänglichen Konfigurationen der beiden Ketten beeinflusst. Die Verflechtung von Ketten könnte jedoch tatsächlich die Trennung in einem kleinen Zeitmaßstab behindern. Die Anzahl der Kontakte innerhalb einer selbstvermeidenden Kette wird ebenfalls analysiert. Es zeigt sich, dass die Gesamtzahl N_c linear mit der Länge von wächst eine freie Kette, während sie in kubischem Einschluss quadratisch wächst. Die Verteilungsfunktion der Kontaktnummer zwischen zwei Hälften zeigt $N_c(AB)$ ein Potenzgesetz-Zerfallsverhalten und dann einen exponentiellen Zerfall für eine freie Kette. In Haft, Die Funktion hat ein Maximum. Wenn die Kette länger wird, wird der Prozentsatz der Kontakte zwischen den Hälften unter den Gesamtkontakten angegeben hat ein Potenzgesetz-Abklingverhalten mit einem Exponenten nahe -1, was unterstützt, dass die Anzahl der Kontakte zwischen zwei Hälften ist endlich, auch wenn die Kette unendlich lang ist.

Schließlich haben wir die Fraktalität und die Topologie in den selbstvermeidenden Wanderungen untersucht. Insbesondere berechnen wir die fraktale Dimension und die Wachstumsraten der Betti-Zahlen des Systems. Diese Wachstumsraten können als topologische Signatur für angesehen werden verschiedene Systeme. Die Intra-Kontakte des selbstvermeidenden Gehens sind eine Untergruppe des ursprünglichen Gehens, und wir stellen fest, dass diese Untergruppe eine leichte multifraktale Eigen-

schaft haben kann. Darüber hinaus unterscheiden sich auch die topologischen Exponenten von der selbstvermeidenden Wanderung. Ferner führt jeder Kontakt zur Bildung von a Schleife. Um herauszufinden, wie diese Schleifen die Struktur des selbstvermeidenden Gehens beeinflussen, löschen wir die Schleifen in ähnlicher Weise Weg zum gelöschten Random Walk, wodurch ein neuer Walk entsteht: Loop-Deleted Self-Avoiding Walk (LDSAW). Das kritische Der Exponent von LDSAW wird durch Untersuchung des Skalierungsverhaltens des mittleren End-to-End-Abstands und der Abhängigkeit von approximiert Die mittlere Länge von LDSAW auf der Länge des ursprünglichen, sich selbst vermeidenden Spaziergangs. Danach die fraktale Dimension und das Wachstum Raten von Betti-Nummern dieser LDSAW werden berechnet. Die gleichen Berechnungen werden auch für die Projektion und durchgeführt zufällige Untergruppen von selbstvermeidenden Spaziergängen.

Publications Related to this Thesis

Parts of this thesis have already been published. Papers in preparation are also listed.

- **J. Jia**, K. Li, A. Hofmann, D. W. Heermann The Effect of Bending Rigidity on Polymers. *Macromolecular Theory and Simulations* **(2019)**, *28*, 1800071

  DOI: 10.1002/mats.201970005

- **J. Jia**, D. W. Heermann The Disentanglement of Chains. *in preparation* **(2019)**

- **J. Jia**, D. W. Heermann Fractality and Topology in the Self-Avoiding Walks. *in preparation* **(2019)**

Contents

Acknowledgments

I am greatly thankful to my advisor Prof. Dieter W. Heermann for his enduring support throughout my doctoral study. His invaluable advices and instructions are indispensable for me to finish my work. I would like to thank my co-advisor Prof. Remus T. Dame for his great patience and suggestions.

I also appreciate the support of my group members, Wei Xiong, Min Chu, Andreas Hofmann, and Kunhe Li. They helped me a lot in both my work and my life in Heidelberg.

I am most grateful for the unconditional support from my family.

I am willing to acknowledge the funding from the China Scholarship Council, the DFG program, the Institute for Theoretical Physics, and the HGS MathComp Graduate School.

Chapter 1

Aims and Scope of this Thesis

1.1 Introduction

DNA is one of the most important molecules in living systems in that it carries the genetic information which controls the processes of functioning, growth, and reproduction in these systems. During these processes, DNA undergoes many transitions between different forms of organization, where many other molecules such as the RNAs and proteins may be involved. Although biologists have gained plentiful knowledge about their organization through experiments, the study of the underlying physical mechanism is still the focus of many research. The development of biophysics has integrated numerous physics concepts, theories, and techniques, especially those from statistical physics, into biology, and has seen great success in the illustration of many biological phenomena.

In eukaryotic cells, the native state of DNA is double-stranded and supercoiled. Moreover, it is usually combined with special proteins, forming a complex called chromatin. These special proteins are histones, which are usually positively charged, and have an electrostatic attraction to certain sites of DNA that has positive charges. The interactions between DNA and histones are also influenced by many other factors such as the salinity of the solvent [1], some remodeling enzymes [2]. The pattern of the DNA-histone binding is depicted by nucleosome, which is the repeating units of chromatin, and composed of a DNA segment about 146 base pairs and eight histone proteins [3, 4]. These nucleosomes and the linker DNA connecting them

are generally described by the "beads on a string" model. These complex reorganizations are of important biological significance. For example, the supercoiling of DNA reduces the spatial distance of operators that are bound on DNA strands and are separated by a large contour distance, which would assist the gene regulation process. Also, the formation of chromatin has many advantages: it organizes DNA into a more compact state, facilitates the mitosis and meiosis processes, prevents possible DNA damages, controls the expression and replication of genes.

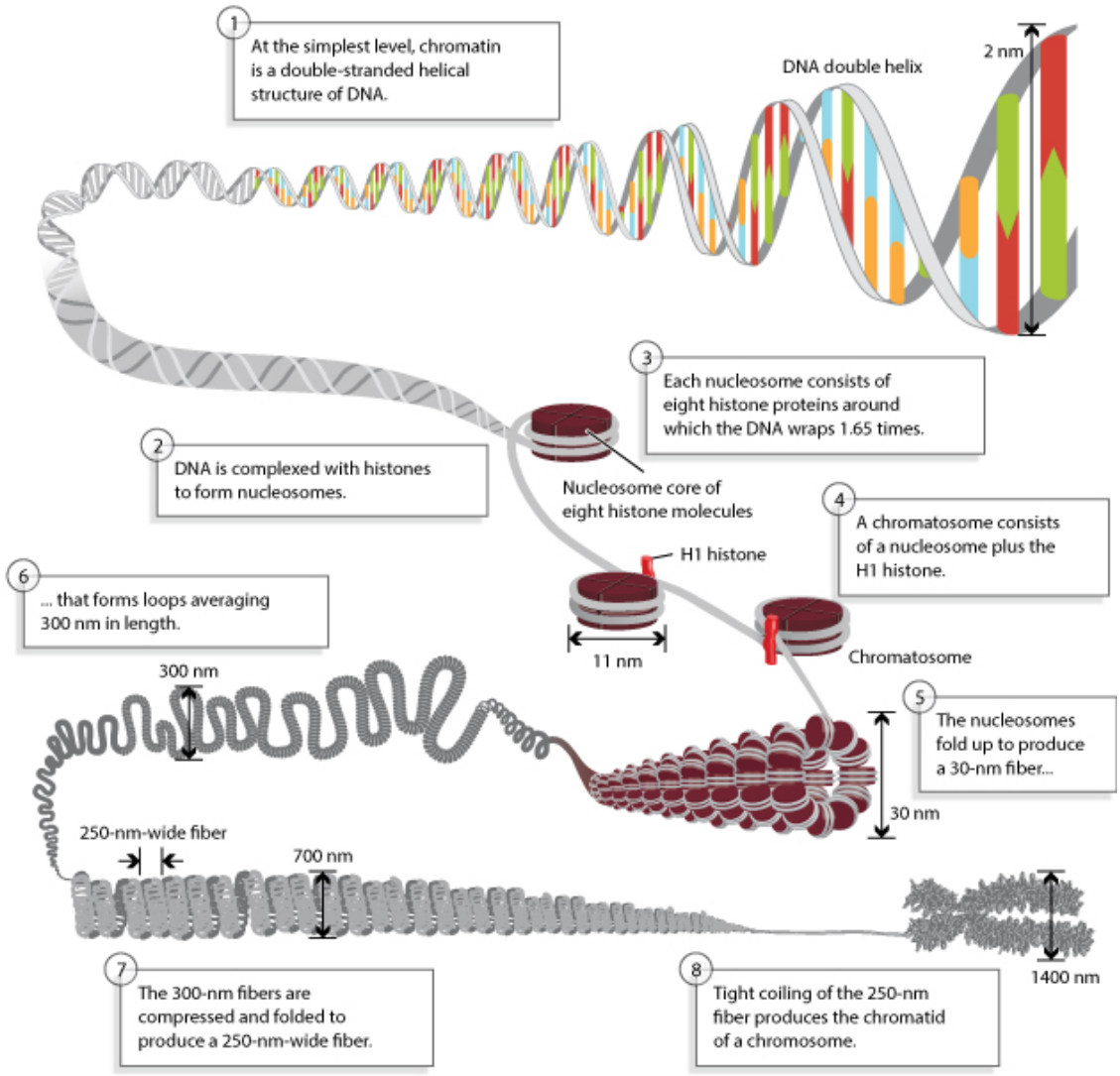


Figure 1.1: The organization of chromatin in different length scales. The figure is from [4].

The organization of chromatin in the nucleus is an important subject in biology. In a human cell, the total length of DNA is about 1 meter. The supercoiling of DNA and combining with histones result in chromatin of extended length about 14 centimeters. Since the diameter of the nucleus is only around 10 microns, the chromatin must adopt highly compacted structures. In molecular biology various techniques are used to analyze the organization of chromatin, such as the chromatin immunoprecipitation (ChIP) [5] and the chromosome conformation capture (3C) techniques [6]. It is found that the formation of loops is one of the most important features in the organization of chromatin. As an effective way to pack the chromatin in a narrow space, loops exist in many length scales [7]. What's more, they are closely related to the regulation of genes in that loops can lead to the physical proximity of genes with large genomic distance [8,9], which facilitates the expression or repression of certain genes. Many experimental techniques including the ChIP and 3C methods can detect these loops by examining the spatially proximal chromatin segments [10,11]. Several coarse-grained models have also been developed to describe the formation of loops, such as the random loop model [12], the dynamic loop model [13] and the fractal globular model [14,15]. The validities of these models are usually verified by comparison of some important measurements with experimental results. These measurements include the scaling behavior of the physical distance with relation to the genomic distance of chromatin sites, the distribution of loop size and so on. Successful models provide an insight into the mechanism of the loop formation. For example, the dynamic loop model attributes the loop formation to the diffusion motion of chromatin fibers. Another important feature of chromatin organization is the existence of the topologically associating domains (TADs) [16]. A topologically associating domain is defined as a compartment in which the chromatin segments have more interaction than those segments outside the compartment. An important difference of TAD with other kinds of self-interacting regions such as sub-TADs, loops is that TADs have similar domain boundaries in different cell types in certain organism [17]. It has been reported that TADs are important in many genetic activities. One example is that genes within the same TAD have correlated expression profiles [18, 19]. On the other side, disrupting the TADs could lead to unexpected contact between genes and thus may cause misregulation and some disease in human and mice [20]. The mechanism of TAD formation is not fully understood yet, but many proteins are believed to be involved in the TAD formation process, two of which are the protein CTCF [21] and the cohesin complex [22].

1.2 Scope of This Thesis

In this thesis, we address three aspects in the modeling of chromatin organization. Firstly we focus on the effect of bending rigidity as well as its heterogeneity on the polymers. Further, its competing effect with the confinement is also discussed. We found that heterogeneity makes the polymer more flexible on average. Next, we study the self-entanglement of a single chain and entanglement of two chains, in which the behavior of the number of intra-and inter-contact is analyzed. Finally, we want to find out the geometrical and topological features of the spatial distribution of contacts by estimating the fractal dimension and one kind of topological index: growth rates of Betti numbers. In addition, a new walk, the loop-deleted self-avoiding walk, is introduced and analyzed.

Many factors can influence the organization of chromatin, one of the internal factors are the intrinsic stiffness of the chromatin. The rigidity of chromatin has several origins. One origin is the stiffness of DNA. It is reported that the base pair stacking effects contributes to the rigidity of DNA strands [23]. Moreover, the interaction between nucleosomes via histone tails is also important in providing the rigidty [24]. There are some other factors that may influence the rigidity of chromatin, such as the methylation of histones [25]. On the other side, chromatin is usually heterogeneous as a result of the DNA sequence and the positioning of nucleosomes [26], which implies that chromatin could have different stiffness along the backbone. The modeling of the rigidity of chromatins has been intensively studied [27–30], where various kinds of rigidity such as the bending rigidity, torsional rigidity, and stretching rigidity are incorporated. Nonetheless, few studies have discussed the heterogeneity of the stiffness. In light of this, we will investigate the influence of this heterogeneity by studying the properties of a semiflexible chain with a distribution of rigidity along the backbone.

Another factor that is of vital importance to the organization of chromatin is the geometrical constraint. The shape of these constraints varies, depending on not only the species but also the scales of inspection. For eukaryotic cells, the nucleus where chromatins reside in is spherical or oval, while bacterial chromosomes are often assumed to be confined in a cylindrical volume. When examining specific chromatin, the constraints imposed by other molecules may be irregular. The effect of spatial confinement on the chromatin organization is an important topic and has also gained persistent attention [31–33]. Furthermore, the competing effect of stiffness, confinement, and entropy is also a crucial issue. Fritsche [34] and Ostermeir [35] have studied the spatial organization of homogeneous stiff ring polymers in rectangular and weak spherical confinement separately. Liu [36] investigated 2D linear semi-flexible polymers in confined space. A semiflexible chain in nanochannel was discussed in [37]. As a part of this thesis, we will study how the heterogeneity

of stiffness and the change of confinement size would influence the organization of a single chain.

The contacts of chromatins apparently play a key role in the accomplishment of many biological functions. For instance, the gene expression is achieved after the promoters and enhancers come into contact [22]. These contacts are also closely related to the organization of chromatins and can be detected by previously mentioned techniques [38]. Usually by analyzing the Hi-C data, one can obtain some interesting properties such as the contact probability [39–41] and chromosomal contact networks [42, 43]. Note that not all contacts correspond to certain physical or biological functions since they can be simply driven by stochastic process [44]. In the modeling of chromatins, the contact probability is also an important quantity to be compared with experimental results [45]. Some studies discussed the number of contacts. In [15,46] it shows that in the fractal globule model the number of contacts of a certain region with the rest is linear with its volume, rather than the surface area. In addition, the contacts number of two regions is proportional to the product of their volumes. This behavior is related to the scaling of contact probability. Besides, the number of contacts for a chain without constraints is also studied in the context of the self-avoiding walk. It shows that the average number of contacts has an asymptotic behavior as the length of the walk N goes to infinity [47]. Baiesi et. al showed that the contact number of two halves of a self-avoiding walk is finite even if the walk is infinitely long [48]. We would further study the dependence of the contact number on the length of segments and the size of geometrical constraints.

The segregation of chromosome is a significant phenomenon in living systems. In eukaryotic cells, two sister chromatids will separate from each other in order to realize the DNA replication or to create haploid cells during the mitotic and meiotic processes. These kinds of segregation are facilitated by cohesin [49] and many other factors [50, 51]. In prokaryotes, the chromosome segregation is also controlled by several mechanisms. For instance, it is found that several genes are involved in the segregation [52], while [53] states that active mitotic-like machinery contributes to the segregation of bacterial chromosomes. Aside from all these external factors, entropy and the structure of chromosomes are also important inducements to the chromosomal segregation under confinement [31, 54, 55]. The dynamics of the segregation process has also been studied by molecular dynamics simulation and Monte Carlo simulation [56, 57]. We further want to study how other factors may influence the segregation, such as the stiffness and the entanglement of chains. Further, we are also interested in how a linear chain is organized if we view it as two or more connected segments.

While most studies concentrate on the probability and number of contact of polymers, few foci are placed on the distribution of the contacts in physical space. It is obvious that the set of contact points a subset of the polymers, but whether this

subset has different distribution properties is an open question. Since the contacts cannot be viewed as connected like a chain, we turn to other quantities that may describe the structure of the contact points. One is the fractal dimension, and another one called growth rate of Betti number comes from a topological theory, i.e. the persistent homology. This theory is attracting considerable interest in many fields [58, 59] due to its powerful ability to characterize the topological features of an object [60]. In particular, it can also be applied in the analysis of biomolecule structures including proteins [61] and chromatins [62]. In addition, following the scheme of loop-erased random walk, we define a new walk: loop-deleted self-avoiding walk. We seek to find for this new walk what the values of critical exponent, fractal dimension and the growth rate of Betti number are.

1.3 Structure of this Thesis

In chapter 2 we introduce the modeling and simulation method in polymer studies. First, the principles of Monte Carlo simulation are introduced, where the Metropolis-Hastings sampling method is stressed. The estimation of errors when applying the MC simulation is also presented. Then some basic polymer models, i.e. ideal chain, self-avoiding chain, and worm-like chain are discussed. The main established results of these models are also exhibited. Typically, the simulation techniques of the self-avoiding walks are given since this is an important subject in simulations. In the last of this chapter, we review three intensively studied topics of the chain systems: chains in geometrical constraints, the topology in polymers and the contacts of polymers. These three topics are also closely related to the above-mentioned questions that will be discussed this thesis.

In chapter 3 we explore two factors that could influence the organization of chromatin: the bending rigidity and confinement. For the first factor, we additionally address the effect of heterogeneity of the polymer. The results show that heterogeneous polymers are more flexible than homogeneous polymers assuming that they have the same mean value of bending parameter. Whether the asymptotic behavior of contact probability depends on the definition of contact is also analyzed. Single semiflexible chain systems in rectangular boxes with various sizes and aspect ratios are simulated. The conformations are analyzed by some properties such as the contact probability, orientational correlation function, and an order parameter.

We in chapter 4 concentrate on the two topics: the self-contact and self-entanglement of a linear chain in free space and confinement, the disentanglement of two chains in confinement. The dependence of the contact of two halves of a chain on the length scale under inspection is investigated. It is reasonable that geometrical constraints would impose more contact inside a chain. We want to analyze whether there are

different kind of dependence when the constraints are imposed. Particularly, we study the limit when the total space is occupied by the chain, i.e. the Hamiltonian path. In addition, the packaging of the two halves in confinement is also studied. It has been revealed that two flexible chains tend to be segregated under confinement. The dynamic process of this segregation is inspected by tracking some measurement of the segregation, such as the dynamic structure factor, the distance of mass centers and the crossing number of the chain. Further, we also study possible factors that may influence the segregation process.

The chapter 5 deals with some geometrical and topological features of the distribution of contacts inside a chain. Specifically, we calculate the fractal dimension and growth rates of the Betti numbers of the system. In the simplest case, we consider the chain as a self-avoiding walk without any other interaction. It is obvious that the set of contacts is a subset of the original walk, and the question arises as to whether this subset has different geometrical or topological features with the self-avoiding walk. Further, each contact gives rise to the formation of a loop. To elucidate how these loops influence the structure of the self-avoiding walk, we delete the loops in a similar way to the loop-erased random walk, thus producing a new walk: loop-deleted self-avoiding walk (LDSAW). The critical exponent of LDSAW is approximated by studying the scaling behavior of mean end-to-end distance, and the dependence of the mean length of LDSAW on the length of the original self-avoiding walk. Afterward, the fractal dimension and growth rates of Betti numbers of this LDSAW are calculated. The same calculations are also performed on the projection and random subsets of self-avoiding walks.

In chapter 6 we give a summary of our work. Also, we discuss the possible directions for further studies.

Chapter 2

The Monte Carlo Method and Polymer Models

Chapter Summary

This chapter discusses the modeling and simulation method in polymer studies. First, the principles of Monte Carlo simulation are introduced, where the Metropolis-Hastings sampling method is stressed. The estimation of errors when applying the MC simulation is also presented. Next, we discuss some basic polymer models, i.e. ideal chain, self-avoiding chain, and worm-like chain. The main established results of these models are also exhibited. Typically, the simulation techniques of the self-avoiding walks are given. Finally, we review three intensively studied topics of the chain systems: chains in geometrical constraints, the topology in polymers and the contacts of polymers. These three topics are also closely related to the results in the following chapters.

2.1 Simulation Method

The molecular dynamics (MD) simulation and Monte Carlo (MC) simulation are two widely used methods in the numerical study of biological systems. The basic principle of MD simulation is to capture the evolution of the system state by solving the dynamic equations of the system numerically. In classical mechanics, the system state means the coordinates, velocities of every particle in the system and the dynamic equations are the Newton's equations of motion, where the forces include many kinds of interactions such as the electrostatic and van der Waals forces. To improve the accuracy of the simulation, in the hybrid QM/MM [63] approach some parts of the system are handled in a quantum mechanics way by introducing the QM energy. The MD simulation provides a way to understand the dynamics of a complex many-body system and thus is prevailing in many areas. Typically, it is a powerful tool to investigate the structure of proteins [64,65] and other biomolecules [66]. The MC simulation, on the other hand, emphasizes on the states of the system in thermodynamic equilibrium and studies its statistical properties. Most of the results in this thesis are obtained by the MC simulation. In this section, the principles of the MC simulation are introduced.

2.1.1 The Monte Carlo Method

Suppose the system under study has a phase space X which is a collection of the system state x: $X = \{x|x \in X\}$. Let $f(x)$ be the probability that the system is in state x. Then the expectation of a measurement $M(x)$ defined on the phase space X is:

$$\langle M \rangle = \int_X M(x) f(x) dx \tag{2.1}$$

The Monte Carlo method to calculate $\langle M \rangle$ numerically is to average it over sufficient samples in X: x_i. If x_i is uniformly distributed in X, then

$$\langle M \rangle = \sum_{i=1}^{N} M(x_i) f(x_i), \tag{2.2}$$

where N is the number of samples. This kind of sampling is simple random sampling. In most cases the simple sampling is quite inefficient, especially for high-dimensional systems, it is possible that most samples contribute to a small part of the summation, therefore one needs quite large amount of samples to estimate $\langle M \rangle$. An alternative way is to generate the samples x_i according to the probability distribution function $f(x)$, thus,

$$\langle M \rangle = \frac{1}{N} \sum_{i=1}^{N} M(x_i). \tag{2.3}$$

If $f(x)$ is simple enough, one can generate such samples using the inverse transform sampling. Unfortunately, $f(x)$ is too complex or implicit in many practical problems. The importance sampling suggests generating samples from a different distribution $g(x)$ when sampling from $f(x)$ is unfeasible. For each sample x_i from $g(x)$, the importance weight is $\omega_i = f(x_i)/g(x_i)$, the $\langle M \rangle$ is estimated by:

$$\langle M \rangle = \frac{\sum_{i=0}^{n} \omega_i M(x_i)}{\sum_{i=0}^{n} \omega_i}. \tag{2.4}$$

Apparently, the choice of $g(x)$ can influence the accuracy of $\langle M \rangle$, given a finite samples. To minimize the variance of the estimation, the best choice of $g(x)$ is proportional to $M(x)f(x)$ [67].

When dealing with a high-dimensional system in statistical physics, the Metropolis algorithm is often adopted to obtain samples from the probability distribution function $f(x)$ of the system. This algorithm uses the Markov chain to generate lots of samples with the strategy of accepting and rejecting until the distribution of these samples approximates $f(x)$, which is called the equilibrium state. The steps of the Metropolis algorithm are as follows:

1. Give an arbitrary initial sample x_0. The choice of this initial sample can influence the speed to reach the equilibrium state, but usually this computational cost is negligible compared to that of generating sufficient independent samples after reaching the equilibrium.

2. Generate a new sample x' from the current sample $x_t = x$ and a proposal distribution $g(x'|x)$.

3. Accept this new sample $x_{t+1} = x'$ with probability

$$A(x', x) = \min\left(1, \frac{f(x')g(x|x')}{f(x)g(x'|x)}\right), \tag{2.5}$$

   Otherwise $x_{t+1} = x$.

4. Repeat step 2 and 3.

Finally we get n samples $\{x_0, x_1, \cdots, x_{n-1}\}$. The acceptance probability $A(x', x)$ is derived so that the process observes the principle of detailed balance, which means that the process should be reversible at equilibrium. Let $\omega(x'|x)$ be the transition probability from state x to x', the detailed balance is:

$$\omega(x'|x)f(x) = \omega(x|x')f(x'). \tag{2.6}$$

Since $\omega(x'|x) = g(x'|x)A(x', x)$, we have

$$\frac{A(x', x)}{A(x, x')} = \frac{f(x')}{f(x)} \frac{g(x|x')}{g(x'|x)}. \tag{2.7}$$

It can be easily proved that the form of $A(x', x)$ in Equation 2.5 satisfies the above condition.

In statistical mechanics, for a canonical ensemble, the probability that the system is under a state x is given by:

$$f(x) = \frac{1}{Z} e^{-E(x)/k_B T}, \tag{2.8}$$

where $E(x)$ is the energy, T is the temperature of the system, k_B is the Boltzmann constant, $Z = \sum_x e^{-E(x)/k_B T}$ is the canonical partition function. According to the Metropolis algorithm described above, if the proposal distribution $g(x'|x)$ is symmetric: $g(x'|x) = g(x|x')$, then the probability of accepting a new state x' is determined by the energy difference ΔE:

$$A(x', x) = \min(1, e^{-\Delta E}),\ \Delta E = E(x') - E(x). \tag{2.9}$$

If the energy of new state x' is lower than the state x ($\Delta E < 0$), x' is always accepted, otherwise it is accepted with probability $e^{-\Delta E}$.

The equation 2.6 states an important principle of this Markov chain Monte Carlo method: the detailed balance. Another important principle is the ergodicity, which means the scheme of generating new sample should be guaranteed so that as the number of samples goes to infinite, the distribution of samples converges to $f(x)$, regardless of the initial sample. Only when these two principles are met can one get a trustworthy estimation of the system if we put aside the statistical errors.

2.1.2 The Statistical Errors of MC samples

Recall that the Metropolis algorithm starts from an arbitrary initial state, and the Markov process converges to the distribution $f(x)$ only after long time of iterations. This means that the samples before the system reaches the equilibrium state has a different distribution from $f(x)$, therefore they should be discarded when calculating the expectation of certain measurement $\langle M \rangle$. Actually the bias of $\langle M \rangle$ resulting from these initial conformation has the order $1/N$, where N is the number of samples to calculate $\langle M \rangle$, which is much smaller than the statistical error with the order $1/\sqrt{N}$.

Another important source of the bias of $\langle M \rangle$ is the autocorrelation of the samples generated by the Metropolis algorithm. If $x_1, x_2, ..., x_N$ are successive samples of the process after discarding the initial samples, and $M_t = M(x_t),\ t = 1, 2, ..., N$ are

the values of the measurement on these samples, the autocorrelation function of the measurement M is

$$C(t) = \langle (M_i - \langle M_i \rangle)(M_j - \langle M_j \rangle) |_{|i-j|=t} = \langle M_i M_j \rangle |_{|i-j|=t} - \langle M \rangle^2. \tag{2.10}$$

For large t, the asymptotic behavior of $C(t)$ can be described by:

$$C(t) \sim \exp\left(-\frac{t}{\tau_{\text{exp}}}\right), \ t \to \infty. \tag{2.11}$$

τ_{exp} in the above equation is called the exponential autocorrelation time. The statistical error of the expectation $\langle M \rangle$ is

$$\sigma^2(\langle M \rangle) = \left\langle \left[\frac{1}{N} \sum_{t=1}^{N} (M_i - \langle M \rangle) \right]^2 \right\rangle. \tag{2.12}$$

After some simplification [68], the error can be rewritten by:

$$\sigma^2(\langle M \rangle) = \frac{1}{N} \sigma^2(M) \left(1 + 2 \sum_{t=1}^{N-1} \left(1 - \frac{t}{N} \right) \frac{C(t)}{C(0)} \right). \tag{2.13}$$

Define the integrated autocorrelation time τ_{int} as:

$$\tau_{\text{int}} \equiv \sum_{t=1}^{N-1} \left(1 - \frac{t}{N} \right) \frac{C(t)}{C(0)}. \tag{2.14}$$

Thus, the error is

$$\sigma^2(\langle M \rangle) = (1 + 2\tau_{\text{int}}) \cdot \frac{\sigma^2(M)}{N}. \tag{2.15}$$

The last term $\frac{\sigma^2(M)}{N}$ is the statistical error when the samples $x_1, x_2, ..., x_N$ are independent, therefore the autocorrelation between the samples contributes to a factor of $2\tau_{\text{int}}$. This indicates one should use the samples at an interval of at least $2\tau_{\text{int}}$ to calculate the expectation $\langle M \rangle$. Suppose the chosen interval is denoted as Δt, then the error of the estimation is $(1 + 2\tau_{\text{int}}/\Delta t)\frac{\sigma^2(M)}{N}$.

Generally, the two autocorrelation times τ_{exp} and τ_{int} mentioned above are not equal. They characterize two processes of the MC simulation. The relaxation process from the initial state to equilibrium states is characterized by τ_{exp}, while τ_{int} influences the statistical errors of the measurement by sampling in the equilibrium states. To reduce the bias caused by the relaxation process, it is safe to discard the first $20\tau_{\text{exp}}$ samples before estimating $\langle M \rangle$ [69]. Nevertheless, one does not need to calculate both autocorrelation times since they are of the same order. The computation of τ_{int} is not exactly guided by the Equation 2.14 considering that for long time $C(t)$ contains more noise than the correlation information. A truncation of the summation needs to be implemented. Detailed discussion can be found in [70].

2.2 Polymer Models

Before performing any simulation, one should establish a well-defined model of the molecules under study. Suppose one wants to study the organization of chromatin or the folding of proteins, the most direct model is the one including all of the atoms, called all-atom model, and the simulation is all-atom simulation. Usually for small proteins, all-atom simulation is practical and reliable. However, most macromolecules in living systems are too large to perform all-atom simulations due to the restriction of computer resources. In thus cases, one can instead use the coarse-grained representation of the original system. In a coarse-grained model, a collection of atoms in the molecule are replaced by a monomer, resulting in a simplified molecule. The key in this coarse-graining process is to keep the most important characters of the system. The ideal chain and self-avoiding walk (SAW) are two simplest coarse-grained models that can describe molecules.

2.2.1 The Ideal Chain

Consider a chain with $N+1$ monomers connected by N segments with fixed length l, if each segment can point in any direction in space with equal probability, without any dependence on the direction of other segments and the position of monomers, then this chain is an ideal chain or freely-jointed chain. Because of the arbitrariness of each segment, two monomers have the chance to occupy the same position in space. If we denote the N segments by $\vec{r_i}(i = 1, 2, ..., N)$ (Figure 2.1), then the end-to-end vector is $\vec{R_e} = \sum_{i=1}^{N} \vec{r_i}$.

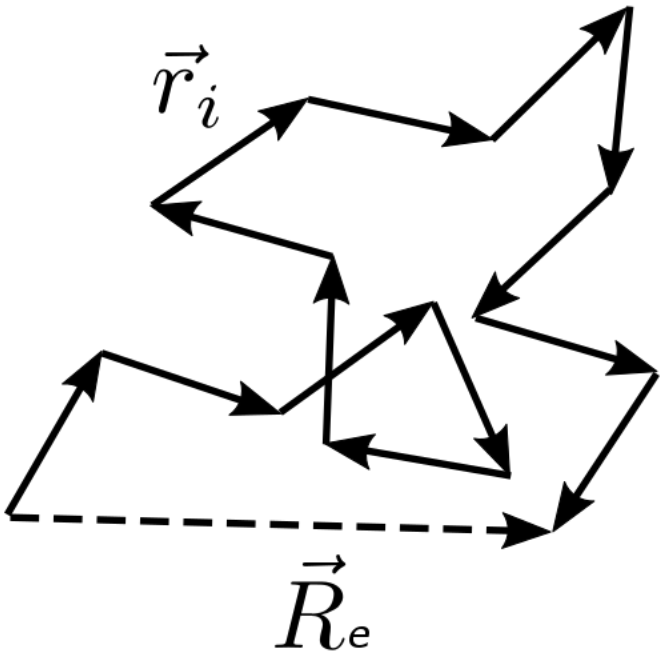


Figure 2.1: The ideal chain. The image is from wikipedia.

Since $\vec{r_i}$ is randomly oriented, $\langle \vec{R_e} \rangle = \sum_{i=1}^{N} \langle \vec{r_i} \rangle = \vec{0}$. On the other side, the mean square end-to-end distance $\langle R_e^2 \rangle$ is not zero:

$$\langle R_e^2 \rangle = \left\langle \left(\sum_{i=1}^{N} \vec{r_i} \right)^2 \right\rangle = \sum_{i=1}^{N} \langle \vec{r_i}^2 \rangle + \sum_{i \neq j} \langle \vec{r_i} \vec{r_j} \rangle \tag{2.16}$$

Because the directions of segments are totally independent from each other, the last term is zero. Therefore,

$$\langle R_e^2 \rangle = \sum_{i=1}^{N} \langle \vec{r_i}^2 \rangle = Nl^2 = Ll, \tag{2.17}$$

where $L = Nl$ is the total length of the walk. The average end-to-end distance $\langle R_e \rangle \sim \langle R_e^2 \rangle^{1/2} = \sqrt{N} l$ is much shorter than the total contour length $L = Nl$. The distribution of R_e is a Gaussian function. This distance is an important measure of the size of the chain, but it is not sufficient to describe the distribution of the monomers and the volume they occupy. One quantity for this is the gyration tensor S:

$$S_{mn} = \frac{1}{2(N+1)^2} \sum_{i=0}^{N} \sum_{j=0}^{N} (R_i^{(m)} - R_j^{(m)})(R_i^{(n)} - R_j^{(n)}) \tag{2.18}$$

where m, n are the indexes of the coordinate of each monomer $\vec{R}_i$. If the chain resides in 3D space, the gyration tensor $S = (S_{mn})$ is a 3×3 matrix. By diagonalizing the gyration tensor, one gets three eigenvalues $\lambda_x, \lambda_y, \lambda_z$, called the principle moments, where are nonnegative, and represent the distribution spans of these monomers along three principle axes. The radius of gyration R_g is defined by $R_g^2 = \lambda_x + \lambda_y + \lambda_z$, and can be rewritten as:

$$R_g^2 = \frac{1}{2(N+1)^2} \sum_{i=0}^{N} \sum_{j=0}^{N} (\vec{R}_i - \vec{R}_j)^2, \tag{2.19}$$

which describes the overall size of the distribution of the monomers. Other quantities that can depict the distribution are the asphericity $b = \lambda_z - (\lambda_x + \lambda_y)/2$, acylindricity $c = \lambda_y - \lambda_x$, if the eigenvalues are ordered so that $\lambda_x \leqslant \lambda_y \leqslant \lambda_z$.

For the ideal chain, the average square radius of gyration is $\langle R_g^2 \rangle = \frac{1}{6} N l^2 = \frac{1}{6} \langle R_e^2 \rangle$. The distribution of monomers for ideal chain is not symmetric in space, the asymptotic ratio of its three average eigenvalues when $N \to \infty$ is $\lambda_x : \lambda_y : \lambda_z = 1 : 2.7 : 12$ [71].

Since the ideal chain model does not consider any interaction, it is not applicable to most real polymer systems, where two monomers cannot occupy the same position according to the excluded volume effect. However, under certain conditions, the polymer systems can have some features of the ideal chain. One example is the polymer in a theta solvent, where the temperature of the solvent is T_θ so that the interaction between polymer and solvent cancels the interaction between segments for real polymer. Another example is the polymer melt, when many chains in the system are heavily overlapped. In this case, at large length scale the excluded volume effect is screened, therefore the chains behave like ideal chains [72].

The computational realization of ideal chain model is usually simplified into lattice random walk. The arbitrary direction of each segment is instead actualized by keeping moving in four directions for square lattice or eight directions for cubic lattice with equal probability. One interesting topic of the random walk is its recurrence. If the position of step i is denoted as $\vec{R}_i$, define p as:

$$p \equiv P\{\vec{R}_i = \vec{R}_0 \ \text{ for some } i > 0\}, \tag{2.20}$$

then a random walk is called recurrent if $p = 1$. For one dimensional random walk,

$$P\{\vec{R}_k = \vec{R}_0\} = p^k(1-p). \tag{2.21}$$

Thus the average number of steps $\langle k \rangle$ when the random walk is recurrent is:

$$\langle k \rangle = \sum_{k=1}^{\infty} k p^k (p-1) = \frac{p}{1-p}. \tag{2.22}$$

This means that one needs infinite steps to make sure that a random walk is recurrent.

2.2.2 The Self-Avoiding Chain

The Excluded Volume Interaction

There are many kinds of interactions in real polymer system, one of which is the excluded volume interaction. This kind of interaction corresponds to the hard sphere potential, defined by:

$$U(r) = \begin{cases} \infty & r < d_0, \\ 0 & r > d_0, \end{cases} \tag{2.23}$$

where r is the distance between two monomers, and it cannot be smaller than d_0, and when $r > d_0$ there is no interaction between monomers. Thus, in this hard sphere model, each monomer can be represented by a hard sphere with diameter d_0. In 3D, the excluded volume of each monomer is eight times its own volume. This model is a very rough simplification of the real system because it does not consider the attraction of monomers. Other forms of inter-monomer potential include the square-well potential, the Lennard-Jones potential and the Sutherland potential which take into account the attraction part. Nonetheless, considering the computational efficiency, the hard sphere potential is acceptable to explore the properties of many systems in MD and MC simulations.

The chain with this excluded volume effect is self-avoiding chain, which is expanded compared with the ideal chain because of the non-overlapping of monomers. Although the implementation of excluded volume seems simple, it leads to much more complex properties of the chain. The mean square end-to-end distance thus has a different asymptotic behavior. In the long chain limit $N \to \infty$, the self-avoiding chain has:

$$\langle R_e^2 \rangle \sim N^{2\nu} l^2, \ N \to \infty. \tag{2.24}$$

Various theories have been developed to study the asymptotic behavior of $\langle R_e^2 \rangle$. The most well-known is the Flory theory where the critical exponent is approximated as $\nu = 3/5$ for 3D self-avoiding chain. The renormalization group method gives $\nu = 0.588$ [73]. Other theories and methods include the perturbation calculation and mean field theory. Lots of Monte Carlo simulations of the self-avoiding chain also provide similar results of ν [74–76]. The same exponent holds for the mean radius of gyration:

$$\langle R_g^2 \rangle \sim N^{2\nu} l^2, \ N \to \infty. \tag{2.25}$$

For ideal chain, the asymptotic value of $\langle R_e^2 \rangle / \langle R_g^2 \rangle$ is 6, while for self-avoiding chain, $\langle R_e^2 \rangle / \langle R_g^2 \rangle \approx 6.45$ [77]. Also, the three principal moments of the gyration tensor have another ratio when $N \to \infty$, $\lambda_x : \lambda_y : \lambda_z = 1 : 2.98 : 14$ [71].

The distribution of the end vector $p(\vec{R}_e, N)$ is not Gaussian function any more, and has been studied both theoretically and numerically [78–84]. For R_e much larger

than the mean value, the probability distribution function is:

$$p(R_e, N) \sim \exp\left[-\alpha\left(\frac{R_e}{\widetilde{R_e}}\right)^{2.43}\right], \quad \text{for } R_e > \widetilde{R_e}. \tag{2.26}$$

For R_e much smaller than the mean value, due to the excluded volume effect, the function is approximated by:

$$p(R_e, N) \sim \left(\frac{R_e}{\widetilde{R_e}}\right)^{0.28}, \text{ for } R_e \ll \widetilde{R_e}. \tag{2.27}$$

These two equations can be combined together, hence

$$p(R_e, N) \sim \left(\frac{R_e}{\widetilde{R_e}}\right)^{0.28} \exp\left[-\alpha\left(\frac{R_e}{\widetilde{R_e}}\right)^{2.43}\right], \tag{2.28}$$

where $\widetilde{R_e}$ is the mean end-to-end distance and $\widetilde{R_e} \sim N^{\nu}$.

Now we consider the average number of monomers $n(r)$ of the chain within sphere of radius r. From Equation 2.25 we should have $\langle R_g \rangle \sim N^{\nu}$, therefore

$$n(r) \sim r^{1/\nu}. \tag{2.29}$$

This implies that the fractal dimension of the self-avoiding chain is $d = 1/\nu$. The fractality of self-avoiding chain is deduced in detail in [85, 86].

$n(r)$ can also be expressed by the pair distribution function $g(r)$ which describes the distribution of distances between monomers contained within sphere of radius r:

$$n(r) = \int_0^r g(r')4\pi r'^2 dr'. \tag{2.30}$$

From this equation and the relation $n(r) \sim r^{1/\nu}$, we have $g(r) \sim r^{(1-3\nu)/\nu}$.

In experiments one often uses the neutron scattering to investigate the structure of polymers [87, 88]. The structure factor $S(q)$ obtained from these experiments can measure the spatial correlation of monomers. It is defined as follows:

$$S(q) = \frac{1}{N+1}\langle \sum_{i=0}^{N}\sum_{j=0}^{N} \exp[i\vec{q}\cdot(\vec{R}_i - \vec{R}_j)]\rangle \tag{2.31}$$

$S(q)$ is the Fourier transformation of the pair distribution function:

$$S(q) = \frac{N+1}{V}\int g(\vec{r})e^{-i\vec{q}\cdot\vec{r}}d\vec{r}, \tag{2.32}$$

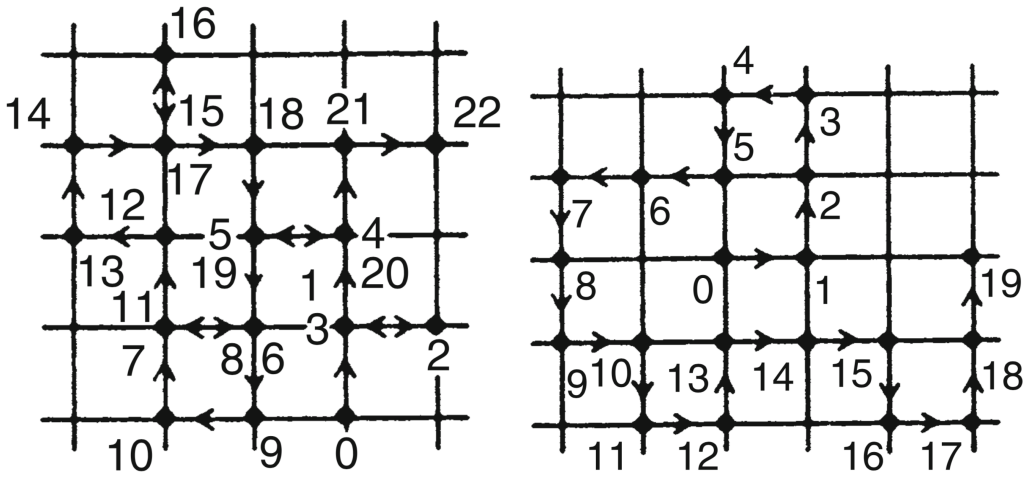
where V is the average volume the chain occupies. Since $g(r) \sim r^{(1-3\nu)/\nu}$, the second term in $S(q)$ has the power law relation:

$$S(q) \sim q^{-1/\nu} \tag{2.33}$$

For the ideal chain, $\nu = 0.5$, $S(q) \sim q^{-2}$, and for the self-avoiding chain, $\nu \approx 0.588$, $S(q) \sim q^{-1.7}$.

MC Simulation of the Self-Avoiding Chain

Compared with the ideal chain, the computer simulation of the self-avoiding chain is much more complex. Similarly, the self-avoiding chain can be remodeled as self-avoiding walk on lattice. The difference from lattice random walk is that each lattice site should be occupied by only one monomer 2.2.



(a) (b)

Figure 2.2: The random walk (a) and self-avoiding walk (b) on square lattice. The figures are from [68].

By generating sufficient and independent conformations of the self-avoiding walk, the asymptotic behaviors of the mean square end-to-end distance and many other quantities could be obtained. Lots of algorithms based on the Monte Carlo simulation have been developed to get samples of the self-avoiding walk [89,90]. Generally, they are classified into two groups: static Monte Carlo methods and dynamic Monte Carlo methods.

Static Monte Carlo Sampling Methods

(1) *The Simple Sampling Method.* The most straightforward method is simple sampling method, the idea of which is to generate lots of random walks, and discard those that are not self-avoiding. This simple sampling method is not practical since when N is large, most samples of the random walk are not self-avoiding. Specifically, the probability that a N-step walk is self-avoiding is:

$$\frac{N^{\nu-1} z_{\text{eff}}^N}{z^N} = \frac{N^{\nu-1}}{(z/z_{\text{eff}})^N}, \ N \to \infty, \tag{2.34}$$

where $z = 4$ for square lattice and $z = 6$ for cubic lattice, $z_{\text{eff}} < z - 1$ is the effective coordination number. As N gets larger, the probability decreases rapidly to zero, meaning that it is very hard to find a self-avoiding walk. Actually the simple sampling method can only produce samples of self-avoiding walk up to length 100.

(2) *The Rosenbluth Sampling and Pruned-Enriched Rosenbluth Method.* One reason for the inefficiency of simple sampling is that the whole walk is discarded once a new step leads to overlapping. The Rosenbluth sampling method suggests that one can continue the walk by discarding the overlapped step and reproducing a new step. This operation would impose a bias towards the distribution of the self-avoiding walks since each walk is not sampled with equal probability any more. This bias is eliminated by assigning a weight w_{N_i} to each sample of walk x_i generated by this Rosenbluth sampling.

$$w_{N_i} = \prod_{k=1}^{N} a_k, \tag{2.35}$$

where a_k is the number of available choices for the k-th step. The corresponding unnormalized probability for this sample x_i is $p_i = 1/w_{N_i}$. Any measurement A on the self-avoiding walk can be estimated by:

$$\langle A \rangle = \frac{\sum_i p_i A(x_i)}{\sum_i p_i} = \frac{\sum_i A(x_i)/w_{N_i}}{\sum_i 1/w_{N_i}}. \tag{2.36}$$

The performance of the Rosenbluth sampling is obviously better than the simple sampling. Still, it can only sample the self-avoiding walk up to a few hundred steps because there is high possibility that no choice is available for the new step when k is large.

To simulate longer walks, the Pruned-Enriched Rosenbluth method (PERM) was proposed by Grassberger [91]. The essence of the PERM algorithm is to increase the samples of walk with high weight and to reduce the samples with low weight. To avoid the bias of sampling, the related weights have to be adjusted while these increasing and reducing are carried out. First one needs to define what are high weight and low weight by two thresholds w_m^h and w_m^l at step m. Suppose that when

the walk is grown to length m, the weight $w_m = \prod_{k=1}^{m} a_k$ is larger than the predefined high threshold w_m^h, this m step walk is enriched by adding s copies of it. The weight of each copy is adjusted to $\frac{w_m}{s+1}$ to eliminate the concomitant bias. Afterwards each copy of the walk continues to grow till length n following the same procedure of Rosenbluth sampling. On the other side, if the weight w_m is smaller than the low threshold w_m^l, this walk is discarded with probability q, and kept with probability $1-q$. The weight is adjusted to $\frac{w_m}{q}$ before the m-step walk continues to grow if it is kept.

There are several parameters to be determined in the PERM algorithm, namely, w_m^h, w_m^l, s and q. Typically, s is set to 1 and q is set to 1/2. The values of w_m^h, w_m^l are usually chosen to be roughly proportional to the partition function at step m:

$$w_m^h = C^h Z_m,\ w_m^l = C^l Z_m. \tag{2.37}$$

Z_m is an approximation of the partition function. It is tested that a ratio of $C^h/C^l \approx 10$ can give good results. The first sample is generated with $w_m^l = 0$ and a very large w_m^h, corresponding to the Rosenbluth sampling method.

The Pruned-Enriched Rosenbluth method is high-efficient, it can simulate θ polymers up to length 1 000 000. Further, various variants of PERM have been developed to study different kinds of systems, such as the Multicanonical PERM [92], the flat-histogram PERM [93] and the dynamic PERM [94].

Dynamic Monte Carlo Sampling Methods

In section 2.1.1 the Metropolis algorithm is introduced to obtain samples from the probability distribution function of the system. This algorithm is especially applicable to high-dimensional systems, and can also be used to generate samples of the self-avoiding walk. The process starts from an arbitrary initial sample. From a N-step walk, the most simple one is a walk with all segments having the same direction. Next, a new sample x_{i+1} needs to be obtained from the old one x_i, which is the kernel of the Monte Carlo simulation of the self-avoiding walk. This step is normally achieved by an operation on the old sample, i.e. elementary moves. Many kinds of moves are designed to achieve this sampling, generally classified as local moves and non-local moves. Since the excluded volume interaction is described by the hard sphere potential (Equation 2.23), according to Equation 2.9, if the elementary move leads to overlapping of monomers, the energy difference $\Delta E = \infty$, thus the accepting probability $A(x_{i+1}, x_i) = 0$, which means this elementary move should be rejected and the new walk x_{i+1} is same as the old one x_i. Otherwise it is accepted as a new walk.

(1) *Single-monomer move.* The simplest kind of move is single-monomer move, which means generating a new walk by moving a random monomer of the old walk

(Figure 2.3). First a random monomer is selected from the $N+1$ monomers, then the number of possible moves m of this monomer that will not lead to overlapping is checked. If $m = 0$, then $x_{i+1} = x_i$. If $m \geqslant 1$, randomly choose one of these m moves and create the new walk x_{i+1}.

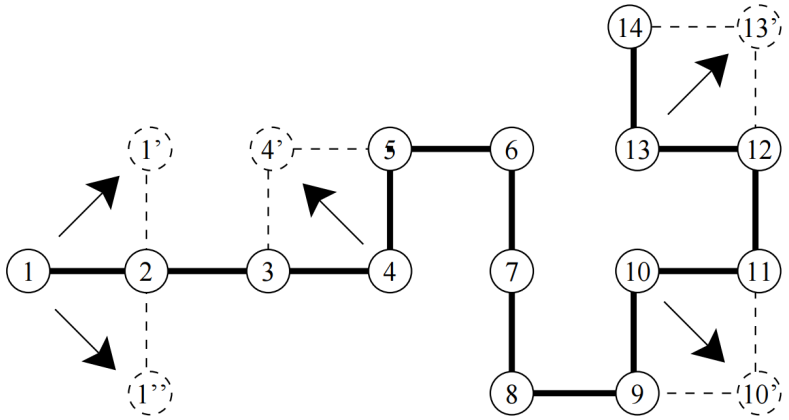


Figure 2.3: The single-monomer move of self-avoiding walk on square lattice. The figure is from [95].

Although simple, the single-monomer move has two problems: the samples are nonergodic and strongly correlated. The nonergodicity can be illustrated by some "double cul-de-sac" conformations (Figure 2.4), in which any single-monomer move is impossible. In fact, other kinds of local moves such as the two-monomer moves face the same nonergodicity problem [96]. However, the local moves are still very popular in many studies because the samples like these "double cul-de-sac" conformations only occupy a small part of the entire phase space [97].

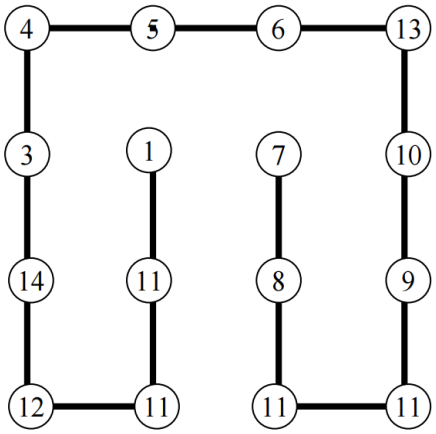


Figure 2.4: One possible conformation that cannot be changed by local moves. The figure is from [95].

The strong correlation between samples leads to a quite large autocorrelation time τ_{int} especially when the length of walk N is large, which means that one needs huge amount of Monte Carlo steps to reduce the bias caused by initialization and correlation. According to [98], the autorcorrelation time scales as $\tau_{\text{int}} \sim N^3$. Therefore, other kinds of moves are needed to simulate long chains.

(2) *The reptation algorithm.* This algorithm suggests that in each Monte Carlo step the monomer at one random end of the walk is removed, then at the other end, an additional monomer is placed in a random available position that could avoid overlapping. If there is no such available position, the move is discarded and the new sample is same as the old one $x_{i+1} = x_i$. This procedure is repeated until sufficient samples are generated. The reptation is not a local move, and thus has a smaller autocorrelation time compared with the local moves, $\tau_{\text{int}} \sim N^{1.02}$ for cubic lattice and $\tau_{\text{int}} \sim N^{1.15}$ for square lattie [99]. This algorithm also encounters the nonergodicity problem, for example it cannot change the conformation in Figure 2.4. Due to the same reason as the local moves, the statistical error caused by the nonergodicity is inconsiderable.

(3) *The pivot algorithm.* The pivot algorithm was widely used in simulating long chains due to its high efficiency and facility to implement [100,101]. The first step is to randomly select a monomer of the self-avoiding walk as the pivot point; secondly, a random symmetry operation is selected and applied to the latter part of the walk starting from the pivot monomer; thirdly, the new walk after the symmetry operation is accepted if it is self-avoiding, otherwise the new walk is discarded. Rotation and reflection are two main types of symmetry operation. The number of

possible operation depends on the type of lattice. For example there are 8 kinds of symmetry operation on a square lattice and 48 kinds on a cubic lattice [100]. A schematic illustration of an operation on square lattice is shown in Figure 2.5.

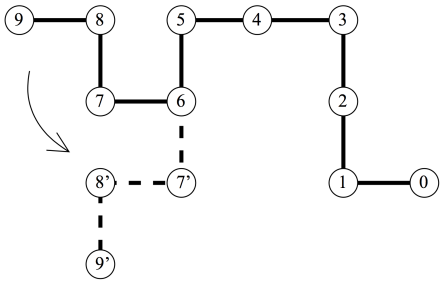


Figure 2.5: One symmetry operation on a square lattice: rotate the latter part (from monomer 6) by 90° Counterclockwise. The figure is from [95].

An off-lattice pivot algorithm was implemented by my college K. Li on the basis of the algorithm by Kennedy [102]. In the off-lattice algorithm, the pivot move is achieved by two transformation matrices: the rotation matrix and the reflection matrix. The rotation of an random angle $\theta \in (0, 2\pi)$ about an axis with unit vector (l, m, n) in 3D space is described by the rotation matrix (Equation 2.38) [103]. The reflection matrix (Equation 2.39) reflects a point through a plane defined by $ax + by + cz = 0$.

$$\mathbf{T_1} = \begin{bmatrix} ll(1-\cos\theta)+\cos\theta & ml(1-\cos\theta)-n\sin\theta & nl(1-\cos\theta)+m\sin\theta \\ lm(1-\cos\theta)+n\sin\theta & mm(1-\cos\theta)+\cos\theta & nm(1-\cos\theta)-l\sin\theta \\ ln(1-\cos\theta)-m\sin\theta & mn(1-\cos\theta)+l\sin\theta & nn(1-\cos\theta)+\cos\theta \end{bmatrix} \tag{2.38}$$

$$\mathbf{T_2} = \begin{bmatrix} 1-2a^2 & -2ab & -2ac \\ -2ab & 1-2b^2 & -2bc \\ -2ac & -2bc & 1-2c^2 \end{bmatrix} \tag{2.39}$$

The two unit vectors (l, m, n) and (a, b, c) are set by four random angles as in Equation 2.40, where $\phi_1, \phi_2 \in (0, \pi), \varphi_1, \varphi_2 \in (0, 2\pi)$ are random variables.

$$\begin{aligned} l = \sin\phi_1\cos\varphi_1, m = \sin\phi_1\sin\varphi_1, n = \cos\phi_1 \\ a = \sin\phi_2\cos\varphi_2, b = \sin\phi_2\sin\varphi_2, c = \cos\phi_2 \end{aligned} \tag{2.40}$$

A monomer $\vec{r}$ to be pivoted is transformed to a new position $\mathbf{T_2}\mathbf{T_1}\vec{r}$. In this off-lattice chain, each monomer is represented by a sphere of diameter d_0, therefore the distance between any two monomers should be larger than d_0, otherwise this pivot move is rejected.

The two principles of the Monte Carlo simulation, detailed balance and ergodicity, are fulfilled by the pivot algorithm with symmetry operations. Besides this, the main advantage over the previous two algorithm is its high efficiency in generating independent samples. Since the symmetry operation changes the conformation of walk globally, it is apparent that one need only far less accepted steps to decorrelate the samples compared with local moves. The problem is that the fraction of accepted pivots decreases as N^{-p}, where p is a positive number and depends on the dimension of space [100]. Also, the check for self-avoiding after the pivot requires much time than the local moves. Kennedy [102, 104] argued that one should emphasize on the time required for accepting a pivot, and proposed an implementation that can reduce this time to $O(N^q)$, where q is about 0.7 for the 2D walk and 0.9 for the 3D walk.

In [105], a generalization of the pivot algorithm, the "cut-and-paste" algorithm, is combined with the local moves to study self-avoiding walks with variable length and fixed endpoints. The idea of the "cut-and-paste" algorithm is to cut the walk into several pieces, then apply a random symmetry operation to each piece individually, finally these pieces are connected in a random order.

The pivot algorithm has exceptional performance for chains in free space and polymers in dilute solvent. Nonetheless, it encounters an obstacle when simulating chains in dense melt and finite constraints [106, 107] since the global pivot move can easily result in the overlapping of monomers or violation of spatial constraints. Especially, [106] proved that the pivot algorithm is not ergodic a two-dimensional strip. For polymers grafted to a surface, Causo [108] used a hybrid of the pivot move and "cut-and-paste" move to avoid the quasi-ergodic problem of pure pivot algorithm. [106] introduced several bilocal algorithms to simulate self-avoiding walk in finite space. When studying long chains melt system, Auhl et al. [107] applied a double-bridging algorithm to change the conformations of two chains simultaneously by splitting two bonds of the original chains and creating two bonds to bridge two parts belonging to different old chains.

(4) *The bond fluctuation model.* Most of the Monte Carlo simulations of self-avoiding walk, implemented with the local moves or non-local moves, are carried out on a square lattice or a cubic lattice owing to its simplicity. However, this kind of lattice has many practical problems [109,110]. The first one is the mentioned nonergodicity of local moves and reptation algorithm. Although the consequent error is negligible for many systems, this is not ensured for some special systems, such as the dense system. While the pivot algorithm is ergodic, the pivot move is not related to any dynamics of the real chains. Secondly, the simulation of self-avoiding walk on a square lattice is quite problematic since the excluded volume effect is more pronounced than in 3D space [111]. This leads to small fraction of acceptance of

local moves and thus large relaxation time. In addition, the simulation of branched polymers are hard to implement in cubic lattice because the branching point cannot move according to the scheme of local moves. The bond fluctuation model (BFM) was therefore proposed to clear up these problems.

In the 2D BFM, each monomer is represented by a unit cell of a square lattice, and thus occupies 4 lattice sites. To avoid the overlapping of monomers and bond crossing, the bond length l has 6 possible values: $2, \sqrt{5}, \sqrt{8}, 3, \sqrt{10}, \sqrt{13}$. There are in total 36 bond vectors and 41 possible bond angles. The local moves of monomers in this model have to satisfy with this constrict of bond length and the self-avoiding condition. The extention to 3D BFM is straightforward, in which each monomer occupies 8 lattice sites, i.e., a cubic cell (Figure 2.6). The possible values of bond length l are $2, \sqrt{5}, \sqrt{6}, 3, \sqrt{10}$. The possible bond vectors are obtained by all permutation and sign combinations of the basic vectors (2,0,0), (2,1,0), (2,1,1), (2,2,1), (3,0,0) and (3,1,0), thus resulting in 87 possible values of bond angle.

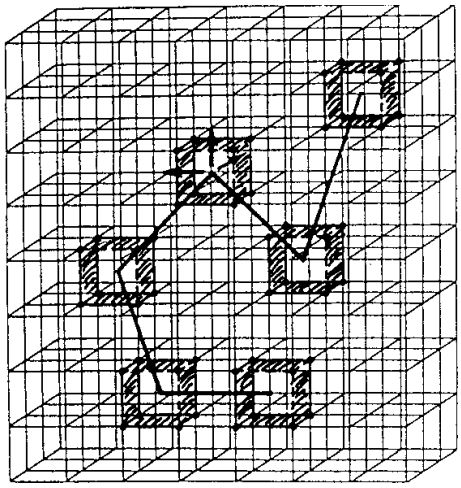


Figure 2.6: The 3D bond fluctuation model. Each monomer are represented by a cubic cell that occupies 8 lattice sites. This figure is from [112].

In the Monte Carlo simulation of polymers using the 3D BFM, a single-monomer move typically has 6 possible direction, leading to a local conformation change of polymers. But for long chains this "L6" move usually leads to long relaxiation and strong autocorrelation between samples. In [113] Wittmer et al. used a hybrid of the "L26" moves, the reptation moves and the double-bridging algorithm to simulate long-range correlations in polymer melts. To study the behavior of bottle-brush polymers a combination of the "L26" moves and pivot moves is applied in [114]. In chapter 3 and 4, we use the bond fluctuation model and the "L6" move to simulate chains up to length $N = 320$ in confinement.

The problems mentioned above does not exist in the BFM. The nonergodicity is avoided by the allowance of various bond vectors, thus with local moves the walk can escape from the "double cul-de-sac" . It is also possible to reproduce the realistic dynamics of polymer system in the BFM with local moves [115]. The simulations of 2D self-avoiding walk and branched polymers are feasible since the branching monomer can move under this framework.

2.2.3 The Worm-Like Chain

Since the ideal chain in section 2.2.1 considers no interaction among monomers and segments, and the self-avoiding chain in section 2.2.2 only considers the excluded volume interaction of monomers, they can only be applied to model limited systems. The real polymers such as chromatins and proteins have many kinds of interactions, among which the bending stiffness is one kind of simple interaction in DNA [116, 117]. The bending stiffness describes how hard it is to bend a polymer. This interaction acts locally, and on large length scales the polymer is still flexible due to the entropy. Clearly, the bending stiffness can influence the conformation of polymers. Further, the combined influence of this stiffness, entropy and spatial constraints on the organization of polymers is an important subject in polymer studies [36, 118–120].

This section starts with the worm-like chain, where only the bending interaction is considered. The worm-like chain is a continuous linear chain of length L, described by the position vector $\vec{r}(s)$ where $s \in (0, L)$ (Figure 2.7).

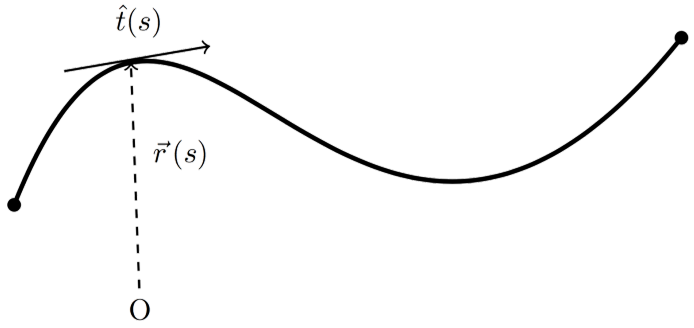


Figure 2.7: The worm-like chain. $\vec{r}(s)$ is the position vector and $\vec{t}(s)$ is the normalized direction vector. This figure is from Wikipedia.

The direction vector $\vec{t}(s)$ is the derivative of the position vector: $\vec{t}(s) = \frac{\partial \vec{r}(s)}{\partial s}$. Thus the bending energy E_b can be represented by:

$$E_b = \frac{1}{2} k_B T \int_0^L l_p \left(\frac{\partial \vec{t}(s)}{\partial s} \right)^2 ds, \tag{2.41}$$

where k_B is the Boltzmann constant, T is the temperature, l_p is a stiffness parameter which is defined as the persistent length.

In order to obtain the orientation correlation function $\langle \vec{t}(s) \cdot \vec{t}(0) \rangle = \langle \cos\theta(s) \rangle$ of the worm-like chain, we first consider the bending energy for a sufficient small fragment of the chain Δs:

$$\Delta E_b = \frac{k_B T}{2} \int_{s_0}^{s_0+\Delta s} l_p \left(\frac{\partial \vec{t}(s)}{\partial s} \right)^2 ds = \frac{k_B T l_p \Delta s}{2} \left(\frac{\theta}{\Delta s} \right)^2, \tag{2.42}$$

where θ is the bending angle of this fragment. This bending angle is a random variable due to the thermal disturbance and its mean square value is [121]

$$\left\langle \theta^2 \right\rangle = 2 \frac{\int \exp(-\Delta E_b / k_B T) \theta^2 d\theta}{\int \exp(-\Delta E_b / k_B T) d\theta} = 2 \frac{\Delta s}{l_p} \tag{2.43}$$

The factor 2 in the above equation means that the fragment can bend in two direction. Thus the mean cosine of θ is:

$$\langle \cos\theta \rangle \approx 1 - \frac{1}{2} \langle \theta^2 \rangle = 1 - \frac{\Delta s}{l_p}, \Delta s \ll 1 \tag{2.44}$$

On the other side, the orientation correlation function $\langle \cos\theta(s) \rangle$ has the property of multiplicativity [122], which means that if a worm-like chain is composed of two successive chains of length s_1 and s_2, then

$$\langle \cos\theta(s_1 + s_2) \rangle = \langle \cos\theta(s_1) \rangle \langle \cos\theta(s_2) \rangle. \tag{2.45}$$

This multiplicativity indicates that the bending behaviors of different parts of the worm-like chain are independent [72], i.e. the interaction is local. The orientation correlation function $\langle \cos\theta(s) \rangle$ can be derived by the Equation 2.44 and Equation 2.45:

$$\begin{aligned} \langle \vec{t}(s) \cdot \vec{t}(0) \rangle &= \langle \cos\theta(s) \rangle \\ &= \lim_{\Delta s \to 0} \prod_{i=1}^{s/\Delta s} \langle \cos\theta \rangle = \lim_{\Delta s \to 0} \left(1 - \frac{\Delta s}{l_p} \right)^{s/\Delta s} \\ &= \exp\left(-\frac{s}{l_p} \right) \end{aligned} \tag{2.46}$$

This equation shows that the persistence length l_p characterizes the length scale of the correlated orientations and the stiffness. On the length scale smaller than l_p, the chain behaves like a rod, and on length scale much larger than l_p, the chain behaves like a flexible ideal chain. The persistence length varies for different kinds of polymers. For double-strand DNA l_p is about 50 nm, and for the actin filament l_p is about 10 μm.

The end vector of the worm-like chain is $\vec{R}_e = \int_0^L \vec{t}(s)ds$. With Equation 2.46, the mean square end-to-end distance $\langle R_e^2 \rangle$ is:

$$\begin{aligned}
\langle R_e^2 \rangle &= \langle \overrightarrow{R_e} \cdot \overrightarrow{R_e} \rangle \\
&= \left\langle \int_0^L \int_0^L \vec{t}(s_1) \cdot \vec{t}(s_2) ds_1 ds_2 \right\rangle \\
&= \int_0^L \int_0^L \langle \vec{t}(s_1) \cdot \vec{t}(s_2) \rangle ds_1 ds_2 = \int_0^L \int_0^L \exp\left(-\frac{|s_1 - s_2|}{l_p}\right) ds_1 ds_2 \\
&= 2Ll_p \left[1 - \frac{l_p}{L}(1 - e^{-L/l_p})\right]
\end{aligned} \tag{2.47}$$

Two limits are usually considered in the above equation. One is when the persistence length is much larger than the total length $l_p \gg L$, $\langle R_e^2 \rangle \approx L^2$, which means that the chain behaves like a rod. Another limit is when $l_p \ll L$, $\langle R_e^2 \rangle \approx 2l_pL$. Recall that for random walk, $\langle R_e^2 \rangle = Nl^2 = Ll$ (Equation 2.17), this means that in this limit the worm like chain behaves like a random walk with bond length $2l_p$.

Similarly, the mean square radius of gyration $\langle R_g^2 \rangle$ can also be calculated. In the continuous case,

$$\begin{aligned}
\langle R_g^2 \rangle &= \frac{1}{2L^2} \int_0^L \int_0^L \langle [\vec{r}(s_1) - \vec{r}(s_2)]^2 \rangle ds_1 ds_2 \\
&= \frac{1}{2L^2} \int_0^L \int_0^L 2l_p |s_1 - s_2| \{1 - \frac{l_p}{|s_1 - s_2|}(1 - e^{-|s_1 - s_2|/l_p})\} ds_1 ds_2 \\
&= \frac{l_p}{3L^2} \left[L^3 - 3L^2 l_p + 6Ll_p^2 - 6l_p^3 \left(1 - \exp\left(-\frac{L}{l_p}\right)\right)\right].
\end{aligned} \tag{2.48}$$

In the two limits mentioned above, the mean square radius of gyration has much simpler forms:

$$\langle R_g^2 \rangle = \begin{cases} \frac{1}{12}L^2 & L \ll l_p, \\ \frac{1}{3}Ll_p & L \gg l_p, \end{cases} \tag{2.49}$$

corresponding to the rod-like and random walk like behaviors respectively. The values of the ratio $\langle R_e^2 \rangle / \langle R_g^2 \rangle$ are 12 and 6 in the two limits.

The structure factor $S(q)$ of the worm-like chain also has different behaviors over different length scales. The range $q \ll \frac{1}{l_p}$ corresponds to large length scale, where the chain behaves like random walk, therefore $S(q) \sim q^{-2}$. In the range $q > \frac{1}{l_p}$, the chain is like a rod, $S(q) \sim q^{-1}$ [123, 124].

The discrete forms of the above formulas in a discrete model where monomers are connected by bonds $\{\vec{r_i}, i = 1, 2, ..., N\}$ can be easily obtained. For example, the exponential decay of the orientation correlation function is $\langle \vec{r_i} \cdot \vec{r_j} \rangle = e^{-|i-j|l/l_p}$, where

l is the bond length. The persistent length defined in this way is a quite important characteristic of polymers that can quantify its flexibility. However, the exponential decay in Equation 2.46 is not valid for all polymers. In [86] and [125] Havlin and Schafer stressed that since the self-avoiding walk is self-similar, its orientation correlation function should also have a scaling behavior:

$$\langle \cos\theta(s) \rangle = \langle \vec{r}_i \cdot \vec{r}_{i+s} \rangle \sim s^{-2(1-\nu)}, \quad 1 \ll s \ll N \tag{2.50}$$

where $\nu \approx 0.588$ is the critical exponent of self-avoiding walk. Consequently, for a semiflexible real chain where both the bending energy and excluded volume potential are present, the orientation correlation function is:

$$\langle \cos\theta(s) \rangle \sim \begin{cases} \exp(-s/l_p) & 1 < s < \xi_T \\ s^{-2(1-\nu)} & \xi_T \ll s \ll N, \end{cases} \tag{2.51}$$

where $\xi_T (> l_p)$ is the length scale below which the excluded volume effect does not kick in. This length scale is also detectable in the structure factor of semiflexible real chain, where three crossovers exists [126]. Further, although real chains in melt condition and theta-solvent have the same scaling of the end-to-end distance $\langle R_e^2 \rangle \sim N$, $\langle \cos\theta(s) \rangle$ also shows a power law decay for certain range of s [127, 128]:

$$\langle \cos\theta(s) \rangle \sim s^{-3/2}, \quad 1 \ll s \ll N \tag{2.52}$$

For this reason, different alternative definitions of the persistent length were put forward for real chain systems [129–131], see [132] for a review of these definitions.

The simulation of semiflexible chains by the dynamics Monte Carlo method has the same procedure as that of the self-avoiding walk discussed in section 2.2.2. The difference is that their equilibrium states have different probability distribution function. For the N-step self-avoiding walk, the distribution function is $f(x) = 1/Z_N$, where Z_N is the total number of possible conformations of the N-step self-avoiding walk. For the semiflexible real chain, $f(x) = E(x)/\sum_x E(x)$. The energy $E(x)$ is:

$$E(x) = E_b(x) + E_{\text{excluded}}(x) = \frac{k_B T}{2} \sum_{i=1}^{N-1} l_p (\vec{r}_i - \vec{r}_{i+1})^2 + E_{\text{excluded}}(x), \tag{2.53}$$

where the first term is the discrete form of the bending energy ($\vec{r}_i$ is bond vector), the second term is the excluded volume potential, x denotes a conformation of the walk. To simulate semiflexible ideal chain, only the bending energy term is considered. In each Monte Carlo step, the elementary move such as the local move and pivot move could lead to the change of bond angles, and thus the energy difference

ΔE. In a cubic lattice, the bond angle has only two values: 0 and $\pi/2$, while in the 3D BFM lattice, there are 87 possible values of bond angle. This distinction in the available values of bond angle gives rise to different approximation of the persistent length [133, 134]. Theoretically the persistence length obtained by fitting the orientation correlation function should be equal to the l_p in the energy term in Equation 2.53 if ignoring the second term E_{excluded}. Nonetheless, the existence of the excluded volume effect could enhance the correlation of bond vectors on length scale larger than the persistence length. On the other side, the discreteness of the chain and limited choice of the bond angle could also affect the approximation of the persistence length. Specifically, the chains on the BFM lattice are more flexible than those on the cubic lattice [133]. We estimated the persistence length for semiflexible real chain in an off-lattice model using the continuous pivot algorithm and got a relatively reliable estimation of l_p [135].

2.2.4 Chains in Confinement

The chains represented in the above sections are not subject to any external influence. In most real systems, the chains are influenced by many other factors, such as the interaction with other chains, external forces, and geometrical constraints. How the constraints would influence the organization of chains is a long-standing topic due to its importance and complicatedness in the field of biology and chemistry, where different shapes of the constraints and various kinds of chains system arise. One classic type of geometrical constraints is a tube. The simplest cases, a single linear chain and two linear chains confined in a tube will be discussed in this section.

A single chain in a tube

For a chain in confinement, the scaling behaviors are still unaffected in length scale smaller than the dimension of the confinement, while beyond this dimension, the organization of the chain depends on the shape of the confinement. Consider a flexible chain consisting of N segments confined in a tube with diameter D (Figure 2.8), within length scale D, the chain segments behave like they are in free space and thus have the same scaling behaviors. The area these segments occupy is named as blob which is of size D. Suppose there are g segments inside each blob, then according to the scaling behavior, one can get $D \approx ag^{\nu}$, where a is a prefactor, ν is the critical exponent. The number of blobs is given by $n_{\text{bl}} = N/g \approx N(\frac{D}{a})^{-1/\nu}$. The length that the chain takes up along the tube is

$$L_{2c} \approx \frac{N}{g} D = ND \left(\frac{D}{a}\right)^{-1/\nu} = aN \left(\frac{D}{a}\right)^{(\nu-1)/\nu} . \tag{2.54}$$

And the volume fraction of the chain is

$$c_m \approx \frac{4N}{\pi D^2 L_{2c}} = \frac{4}{\pi a^3}\left(\frac{D}{a}\right)^{(1-3\nu)/\nu} \tag{2.55}$$

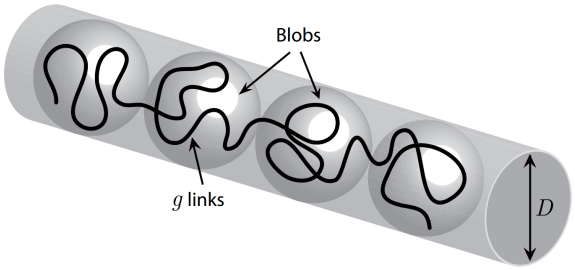


Figure 2.8: A single chain in a tube with diameter D is reorganized as many connected blobs. This figure is from [121].

The free energy of this confined chain is

$$F \approx k_B T n_{\mathrm{bl}} = k_B T \frac{N}{g} = k_B T N \left(\frac{D}{a}\right)^{-1/\nu}. \tag{2.56}$$

The tube is not the only kind of constraint where the chain inside can be viewed as a series of blobs. For example, for a chain confined in a spherical cavity, the picture of connected blobs also works if the volume fraction of the chain is smaller than 0.15 [136]. The difference is that the blobs are not arranged along a straight line.

For a semiflexible chain with length L and persistence length l_p, there is a length scale $\xi_{_T}$ below which the excluded volume effect is not present as mentioned in the previous section. If the diameter of the tube is much larger than this length scale $D > \xi_{_T}$, the above descriptions of blobs still hold. If $l_p \ll D \ll \xi_{_T}$, the blobs are stretched into ellipsoids. Further, if the diameter is much smaller than the persistence length $D \ll l_p$, then the chain has a wave-like organization in the narrow tube (the Odijk regime, Figure 2.9). In this case, the deflection length λ and deflection angle θ can be defined [137].

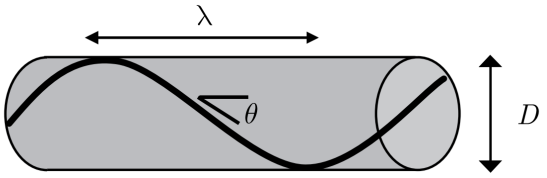


Figure 2.9: A semiflexible chain in a narrow tube. This figure is from [121].

For a very small deflection angle, $\theta \approx \tan\theta = \frac{D}{\lambda}$. With the equation 2.43, we get the following relation:

$$\langle\theta^2\rangle = \frac{2s}{l_p} \approx \frac{2\lambda}{l_p} = \left(\frac{D}{\lambda}\right)^2. \tag{2.57}$$

Therefore $\lambda \approx \left(\frac{D^2 l_p}{2}\right)^{1/3}$. The length that the chain takes up along the tube is

$$L_{2c} = L\langle\cos\theta\rangle \approx L\left[1 - 0.7\left(\frac{D}{l_p}\right)^{2/3}\right]. \tag{2.58}$$

The free energy of this stiff chain can be written as $F \approx k_B T \frac{L}{\lambda} \approx k_B T L l_p^{-1/3} D^{-2/3}$.

Two chains in a tube

In the two chains system, besides the geometrical constraints, the organization of each chain is impacted by the interchain interaction. The states of mixing and segregation are of special interest. First consider two chains of same length confined in a infinite long tube. A typical state of the two chains is shown in Figure 2.10. There are three regions, two of which are occupied by the two chains separately, with distance l_r along the tube. In the middle the chains are mixed or overlapped, and take up distance l_o along the tube. According to [138], in the mixed region, the single chain description remains valid, but each chain is confined in an effective tube with diameter $D_{\text{eff}} = D/\sqrt{2}$.

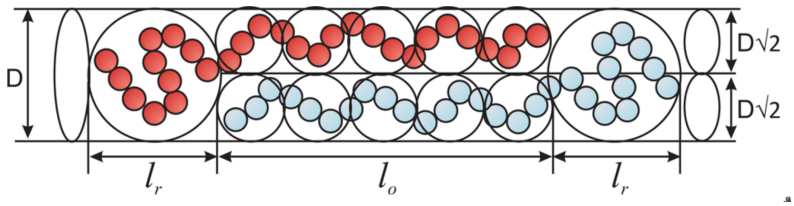


Figure 2.10: Two linear chains of same length in a infinite long tube. Shown is the state between totally mixed and segregated. In the overlapping region the chains take up distance l_o along the tube. This figure is from [139].

The free energy of the system is again purely contributed to the entropy term, and can be written as [139]:

$$F(n_0) = 2k_B T[n_0 + (n_{\text{bl}} - n_0)2^{1/2\nu}], \tag{2.59}$$

where n_0 is the number of blobs in the region at one end, $n_{\text{bl}} = N/g$ is the number of blobs when one chain is confined in the tube. Since $n_{\mathbf{bl}}$ does not change with

the state of the two chains, the free energy is only dependent on n_0. The totally mixed state corresponds to $n_0 = 0$, and has free energy $F_{\text{mix}} = 2^{1+1/2\nu} n_{\text{bl}} k_B T$, while the totally segregated states corresponds to $n_0 = n_{\text{bl}}$, thus $F_{\text{seg}} = 2 n_{\text{bl}} k_B T$. Clearly, $F_{\text{seg}} < F_{\text{mix}}$, therefore the two chains tend to segregate. The dynamics process from the mixed state to segregated state is discussed in [139] by studying the segregate rate, where they concluded that there is an acceleration of segregation during the process.

Things become more complicated in the case of a closed tube. The state of two chains depends on the size of the tube and the concentration of the chains. For a tube with length L_t equal to $2L_{2c}$ (see Equation 2.54), the free energy difference between mixed state and segregated state is same with the case of a infinite long tube. As the tube length becomes smaller, the free energy difference also decreases since the available conformations of each chain are less when segregated. When $L_t = D$, the chains tend to mix with each other. The phase diagram of the chains states is shown in Figure 2.11, where the x-axis denote the size of the tube, and the y-axis is the concentration of the chains. Detailed discussion about the regimes can be found in [54].

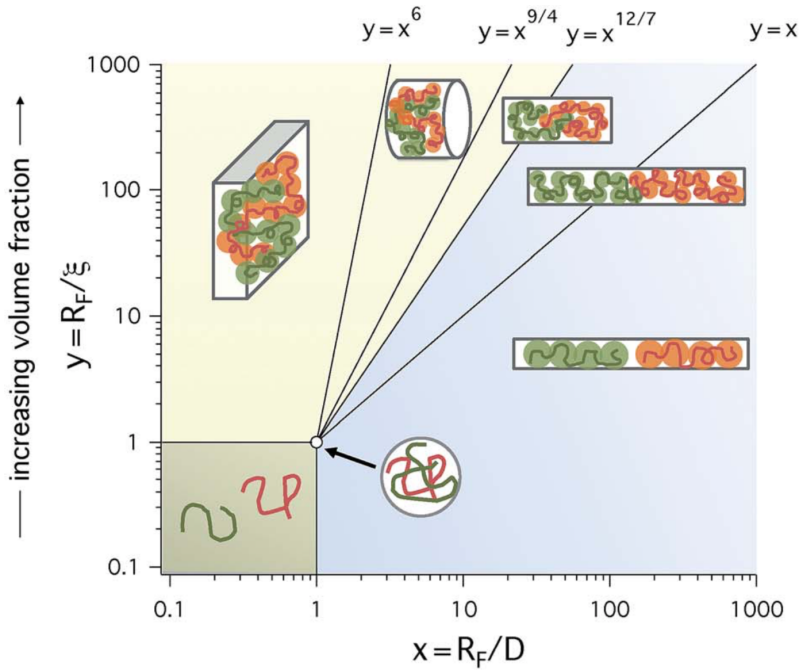


Figure 2.11: The phase diagram of the state of two chains confined in a tube. This figure is from [136].

2.2.5 Topology in Polymers

Besides many geometrical properties of polymers, their topological features also play an important role in their organization. One important kind of such features is the entanglement of polymers. In many biological processes, the entanglement or disentanglement of biomolecules is very common and important. For example, two strands of DNA can be transformed between the intertwined state and isolated state with the help of topoisomerase [140, 141]. In another case, the folding time of protein is correlated to the topology of its native state by $k_f \propto p^Q$, where k_f is the folding rate, Q is the crossing number of the protein [142, 143]. For closed polymers there are lots of studies on their topological complexity, such as the topology invariants [144, 145]. Some similar measurements have also been proposed to describe the entanglement complexity of open polymers [146, 147]. The writhe and average crossing number are two important quantities to describe the self-entanglement of a open or closed polymer [148, 149].

In knot theory, the writhe of diagram is the sum of all signed crossings. By signed crossing, it means that there are two kinds of crossings, which can be defined into positive crossing and negative crossing. The usual way of defining this is the right-hand rule. For a oriented curve, if the lower part goes from right to left, the it is a positive crossing, otherwise it is a negative crossing (Figure 2.12).

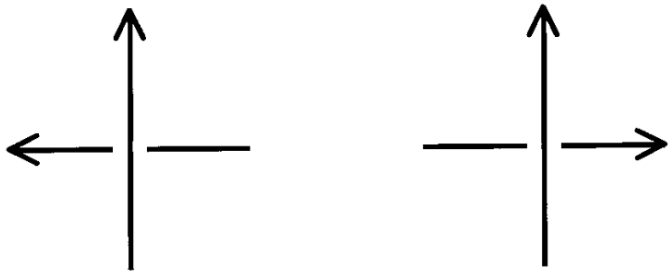


Figure 2.12: The signs of crossing: the left one is positive crossing and the right one is negative crossing according to the right-hand rule.

For a curve in 3D space, its writhe is the average writhe of the projections of the curve from all vantage points. If C is the curve with length L and $\vec{r}(s)$ is the position vector, $s \in (0, L)$, the writhe Wr is

$$Wr = \frac{1}{4\pi} \int_0^L \int_0^L \frac{(\dot{\vec{r}}(s) \times \dot{\vec{r}}(t)) \cdot (\vec{r}(s) - \vec{r}(t))}{|\vec{r}(s) - \vec{r}(t)|^3} ds\ dt \tag{2.60}$$

For a chain consisting of N segments, the following form is often used in the numer-

ical calculation of the writhe:

$$Wr = \sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\Omega_{ij}}{4\pi} = 2\sum_{i=2}^{N}\sum_{j<i}\frac{\Omega_{ij}}{4\pi}, \tag{2.61}$$

where $\Omega_{ij}/4\pi$ is the integral of the ith and jth segments of the chain. The second equivalence comes from the fact that $\Omega_{ij} = \Omega_{ji}$ and $\Omega_{i,i+1} = \Omega_{ii} = 0$. There are several ways to calculate the term $\Omega_{ij}/4\pi$, one of which is a geometrical approach [150]. First assign the monomers connected by the ith and jth segments to four numbers $1 \sim 4$, and let $\vec{r}_{mn}$ be the vector from monomer p to monomer q, and define four unit vectors as follows:

$$\vec{n}_1 = \frac{\vec{r}_{13}\times\vec{r}_{14}}{|\vec{r}_{13}\times\vec{r}_{14}|}, \vec{n}_2 = \frac{\vec{r}_{14}\times\vec{r}_{24}}{|\vec{r}_{14}\times\vec{r}_{24}|}, \vec{n}_3 = \frac{\vec{r}_{24}\times\vec{r}_{23}}{|\vec{r}_{24}\times\vec{r}_{23}|}, \vec{n}_4 = \frac{\vec{r}_{23}\times\vec{r}_{13}}{|\vec{r}_{23}\times\vec{r}_{13}|}. \tag{2.62}$$

Next step is to calculate Ω^*:

$$\Omega^* = \arcsin(\vec{n}_1\cdot\vec{n}_2) + \arcsin(\vec{n}_2\cdot\vec{n}_3) + \arcsin(\vec{n}_3\cdot\vec{n}_4) + \arcsin(\vec{n}_4\cdot\vec{n}_1) \tag{2.63}$$

The sign of Ω is determined by $(\vec{r}_{34}\times\vec{r}_{12})\cdot\vec{r}_{13}$, therefore:

$$\frac{\Omega}{4\pi} = \frac{\Omega^*}{4\pi}\,\mathrm{sign}\left((\vec{r}_{34}\times\vec{r}_{12})\cdot\vec{r}_{13}\right) \tag{2.64}$$

Following this algorithm, one could calculate the writhe of a closed or open chain. The calculation of the average crossing number $Cn$of a chain is similar, but the sign of the crossing is neglected.

$$\mathrm{ACN} = \frac{1}{4\pi}\int_0^L\int_0^L \frac{|(\dot{\vec{r}}(s)\times\dot{\vec{r}}(t))\cdot(\vec{r}(s)-\vec{r}(t))|}{|\vec{r}(s)-\vec{r}(t)|^3}ds\ dt \tag{2.65}$$

The increasing of writhe and average crossing number with the size of random walk or self-avoiding walk have been studied in lots of literatures. In [148] Orlandini et al. found that both the average writhe and mean average crossing number of self-avoiding walks has a power law increase with the length of the walks with exponents around 0.5 and 1.1 respectively. Another kind of increasing law of the mean average crossing number $aN\log N + bN$ for random walk and self-avoidng walk was discussed in [149, 151, 152]. For random walks $a = \frac{3}{16}$, and for self-avoiding walks $a \approx 0.039$.

2.2.6 The Contacts of Polymes

The contacts of intra- and inter-polymers are another subject of vital importance in the field of biology because many biological functions are achieved by the contacts

of different parts of macromolecules. Moreover, the presence of contacts can also influence the dynamics of biomolecules. For example, proteins with more local contacts are found to fold more rapidly and experience less transition states than those with more non-local interactions [142]. For polymers in solution, the average and variance of the contact number are related to the internal energy and specific heat of system [153]. Usually many specific contacts of biomolecules are regulated by short-range interaction or some agents such as enzymes. However, the thermal fluctuation is also one importance source of contacts. For self-avoiding walks where only excluded volume potential is considered, the contacts between monomers are purely due to the randomness. The contact probability and contact number are two important focuses when studying the contacts of self-avoiding walks.

The contact probability is the probability that two monomers with certain contour length come into contact. The scaling behavior of this probability was found in [154, 155] by studying the correlation properties of a long polymer. The function they studied is the probability distribution of the vector $\vec{r}$ connecting two monomers inside a chain, denoted as $P_{s,N}(\vec{r})$, where N is the contour length of the two monomers, s is used to label three situations of the positioning of the two monomers (Figure 2.13). $s = 0$ is the case when two monomers are at the end of a chain with length N, $s = 1$ is the case when only one monomer is at one end of a chain, $s = 2$ is the case when the two monomers are both inside a chain. In the first case $P_{s,N}(\vec{r})$ is the distribution of the end vector of a self-avoiding walk discussed in section 2.2.2.

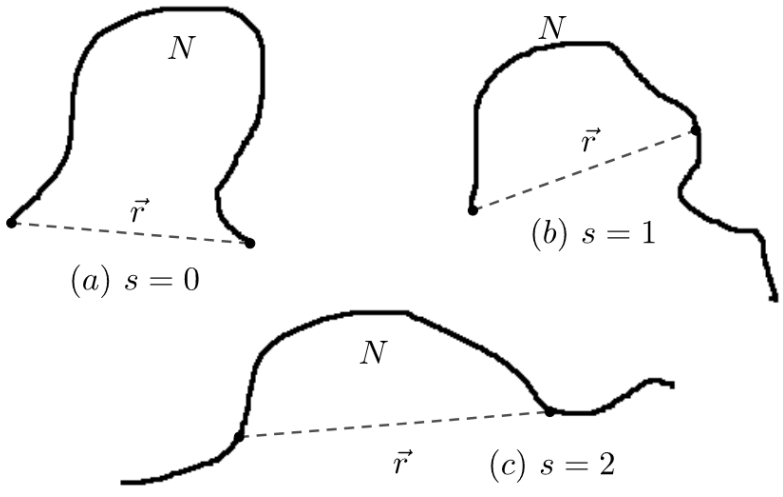


Figure 2.13: Three situations of the positioning of the two monomers.

Using the renormalization method, the probability distribution function has the

following form:

$$P_{s,N}(\vec{r}) = N^{-\nu d} f_s\left(r/N^{\nu}\right),\ f_s(x) \propto x^{\theta_s}, \tag{2.66}$$

where d is the dimension of the space, ν is the critical exponent of self-avoiding walk. For self-avoiding walks in 3D space, the values of θ_s is $\theta_0 = 0.273, \theta_1 = 0.46, \theta_2 = 0.71$. Redner [155] gave the estimation of $\theta_2 \approx 0.67$ and 1.93 for self-avoiding walks on the FCC and triangular lattices by exact enumeration. On a simple cubic lattice, if the contact is defined when two nonconsecutive monomers are nearest neighbors, i.e. $r = 1$, the contact probability of a long chain can be deduced from Equation 2.66:

$$P_{2,N} \propto N^{-\gamma}, \gamma = (3 + \theta_2)\nu \approx 2.18, \tag{2.67}$$

N is the contour length between the two monomers in contact. The results for 2D self-avoiding walks were given in [156]: $\theta_1 = \frac{5}{6}, \theta_2 = \frac{19}{12}$.

Another important topic, the number of contacts, was also studied both for random walks and for self-avoiding walks. In both cases the average number of nearest-neighbor contact $\langle m \rangle$ has an asymptotic behavior as the length of the walk N goes to infinity [47]:

$$\langle m \rangle \sim aN + bN^{\phi} + c, \quad N \to \infty, \tag{2.68}$$

where a, b and c are constants, the exponent ϕ is dependent on the dimension d of space, for random walks $\phi = 2 - d/2$. The values of the three constants and the exponent for various dimensions were estimated by Douglas using direct enumeration and $1/d$ expansion methods. For 3D self-avoiding walk, $a \approx 0.2, b \approx -0.55, c \approx 0.06, \phi \approx 0.15$. In addition, for rather short chains, an algebraic method to determine the number of conformations of self-avoiding walk with fixed number of contact was developed in [157].

Chapter 3

The Effect of Bending Rigidity on Polymers

References

The main part of this chapter is adapted from the following publication that was written by myself:

- J. Jia, K. Li, A. Hoffmann and D.W. Heermann, *The Effect of Bending Rigidity on Polymers.*

Chapter Summary

The conformations of chromatin can be manipulated by many factors. Specifically, in the regulation of gene expression the bending rigidity of the chromatin polymer and its heterogeneity play a critical role in the conformational change. To elucidate this we investigate the effect of bending rigidity as well as its heterogeneity on various polymer properties. In the study of chromatin organization, the contact probability is a significant measure. We analyze whether there is any ambiguity in the definition of a contact and show that results for the contact probability do not depend on the range of contact in the limit of a large contour length between monomers. We further compute the persistence length as a function of the bending rigidity in the homogeneous and heterogeneous cases and find the persistence length is systematically smaller in the heterogeneous case. Another important factor is confinement. Chromosomes are confined in the nucleus and by examining specific loci, the local chromatin part changes much faster than the entire environment. In conjunction with bending rigidity we analyze conformations in rectangular confinements with several aspect ratios. An oscillation in the contact probability and the orientational correlation function is found due to the spiraling of polymer when the box size is small enough. In addition, the influence of the bending rigidity and confinement on the crossing number are analyzed.

3.1 Introduction

Contact probabilities are at center stage in measuring the conformations of chromosome [158, 159]. One measures the number of self-contacts inside chromosomes as well as the inter-chromosomal contacts because these could give topological information on the spatial organization of chromosomes in these experiments. Since the chromosomes are confined in the nucleus, how they are packaged and how the packaging influences the intra-chromosomal contacts are questions to be studied. Further, how would the stiffness of chromatin influence all of this?

The packaging of chromosomes is of vital importance to the biological function [160, 161]. One example is the packing of DNA with the help of histone proteins in the formation the beads-on-string chain [4, 162]. A further packaging is the 30nm fiber (chromatin) and the packaging of the fiber into the nucleus [163]. Packing on the scale beyond 30nm is mainly accomplished by the dynamic formation of loops and higher order loop structures (loops of loops) [164]. These build up local compartments of varying densities which in turn are organized into chromosome territories [165, 166]. Therefore the fiber cannot be considered to be in free space and the contacts that the chain can have with itself are largely influenced by two factors: the kind of local confinement the fiber finds itself in and the bending rigidity of the fiber which for example is controlled by chromosome remodeling [167, 168].

The compartmentalization of the nucleus (such as in human cells) implies the existence of a rather symmetric confinement of the chain. For E. coli, on the other hand, the confinement is rectangularly shaped and this confinement influences the interaction [169]. Then one may ask: how does the shape influence the contact probabilities?

There are several factors that can influence the bending rigidity of a chromosome. For human chromosomes the existence of nucleosomes and their distribution along the backbone of the chain [170–172] indicate a distribution of the bending rigidity along the fiber. A further factor is the repulsion of the histone tails, i.e. methylation [173]. Furthermore histone H1 depletion has a great impact on the flexibility of the chain [174, 175].

In brief, chromatin is not totally flexible, that is to say, chromatin is a semiflexible polymer fiber with bending rigidity. Moreover, generally speaking, chromatin is heterogeneous, which means that different parts of the polymer could have different distributions of bending rigidity. This possible causation could be the genome sequence or the distribution of nucleosomes along the backbone [176]. Apparently this heterogeneity itself is important to the chromatin organization as well [177] and thus has an impact on the contact probability.

Chromatin undergoes structural transformation to perform biological functions properly. For a long chromatin chain, the change of the overall conformation is slow,

while at smaller scales the change is much faster and is confined in a narrower space in contrast to the overall volume the chromatin takes up. Therefore, in this chapter, when studying polymers in confinement, we concentrate on short chains, where the influence of bending rigidity, the size and aspect ratio of the confining volume is studied.

This chapter is organized as follows. In section 3.2 we introduce two polymer models that we use and the implementation of the heterogeneity of the bending rigidity with different distributions. Actually the heterogeneity of polymer has been modeled in several approaches [178, 179]. In our model it is implemented via the variance of the bending parameter κ along the chain. In section 3.3 the results on our main questions are presented. In section 3.3.1 we address the question: What is a contact? This question arises both in the context of lattice polymers as well as in continuum. Further, how the bending rigidity, especially its heterogeneity, affects the contact is illustrated. Next, we focus on properties like persistence length and the structure factor in section 3.3.2. In section 3.3.3 we address the question how the linear semiflexible chain is organized in confinements with different sizes and shapes. Further the influence of the heterogeneity of bending rigidity on this organization is investigated. In terms of the second question, Fritsche [34] and Ostermeir [35] have studied the spatial organization of homogeneous stiff ring polymers in rectangular and weak spherical confinement separately. Liu [36] investigated 2D linear semiflexible polymers in confined space. In this context we study how the stiffness and its heterogeneity affect the organization of linear polymers in 3D rectangular confinement with different sizes and aspect ratios. Our conclusions are presented in section 3.4.

3.2 The Model

The polymer is modeled as a self-avoiding walk. We introduce the bending rigidity as in the Kratky-Porod model, or the worm-like chain model in continuum. In the Kratky-Porod model the origin of stiffness of a polymer is the intrinsic bending energy H_b, which is the sum of energies of successive segments:

$$H_b = -\sum_{i=1}^{N-2} \kappa_i \mathbf{u_i} \cdot \mathbf{u_{i+1}}, \tag{3.1}$$

where κ_i is the stiffness parameter, $\mathbf{u_i}$ is the normalized bond vector, N is the number of monomers. To model a heterogeneous chromatin chain having a variable bending rigidity along the chain, κ_i can be set to obey a distribution of interest, while $\kappa_i = \kappa$ for a homogeneous chromatin chain. When studying heterogeneous chains, we assume that κ_i obeys the Gaussian distribution with mean value $\langle\kappa\rangle$ and

standard deviation σ.

The continuous version of the Kratky-Porod model is the worm-like chain model, where the persistence length l_p is defined through the exponential decay of the orientational correlation function:

$$\langle \mathbf{u}(s_1 + s) \cdot \mathbf{u}(s_1) \rangle = \langle \cos\theta(s) \rangle = e^{-s/l_p}. \tag{3.2}$$

Here $\mathbf{u}(s) = \frac{\partial \mathbf{r}(s)}{\partial s}$ is the unit tangent vector to the chain at contour distance s, and $\mathbf{r}(s)$ is the position vector along the chain. Although chains in a dense melt or at the Θ-point in solution behave like ideal chains without excluded volume effect, as the worm-like chain does, recently it was shown that the orientational correlation function for chains in these conditions shows a power law decay $s^{-3/2}$ instead of the above exponential decay for certain range of contour length $1 \ll s \ll N$ [180, 181]. For real chains Hsu et al. [132] have shown that the standard definition of persistence length does not describe the local "intrinsic" stiffness either, with $\langle \cos\theta(s) \rangle \sim s^{-\beta}$ for $1 \ll s \ll N$, β being a different power law exponent $\beta = 2(1-\nu) \approx 0.824$. However, the exponential decay fits well at short length scales s for simple linear chains without a complex architecture such as side chains, and it is capable of approximating the stiffness parameter κ fairly. In free space and for the homogeneous chain, the stiffness parameter $\kappa_i = \kappa$ is actually related to l_p defined in equation (3.2) via $l_p \approx \kappa \bar{l}_b$ (where energy is measured in the units of $k_B T$). The deviation of l_p results from the discretization of the continuous worm like chain which makes l_p slightly smaller than κ , and the self-avoiding effect, which makes l_p larger compared to random walk. But the latter is negligible when κ is large enough. $\bar{l}_b$ is the averaged bond length. For the heterogeneous chain, the average persistence length over the entire chain is determined by the distribution ($\langle \kappa \rangle$ and σ for Gaussian distribution), which will be discussed in Section 3.3.2.

In this paper we use two models to perform the Monte Carlo simulation and study the questions defined in section 3.1. First, when simulating very long chains in order to investigate the key question on the definition of a contact we employ a pivot algorithm based on the original idea of Sokal and Kennedy [70, 102] in continuous space. There have been several applications of the continuous pivot algorithm in different polymer models. Adamo and Pelissetto [182] have implemented the off-lattice pivot algorithm to study the impact of the thickness of monomers, i.e. the effectiveness of the excluded volume interaction, on the asymptotic behavior of polymer chains. Also, a continuous pivot algorithm with narrower choice of pivot angles is used to study the effects of macromolecular crowding on protein stability [183]. Horwath, Clisby and Virnau [184] use the standard implementation of the pivot algorithm to investigate knots in finite memory walks where the excluded volume effects are considered only at short length scales.

In this algorithm, a pivot with a random pivot point on the chain and a random

symmetry matrix is carried out at each Monte Carlo move, producing a global conformation change of the chain. This algorithm is highly efficient in that it reduces remarkably the relaxation time to reach the equilibrium state and de-correlates conformations much faster compared to algorithms based on local moves. Kennedy [102] proposed a faster implementation of the existing pivot algorithm for self-avoiding walks on a lattice, requiring a time $O(N_b^q)$ per accepted pivot with $q < 0.85$ for a 3D lattice instead of $O(N_b)$ for other pivot algorithms. N_b is the number of bonds ($N_b = N - 1$ for a linear chain). We extended this faster on-lattice pivot algorithm into a continuous one, each monomer being a hard sphere of radius $r = 0.4$. Furthermore, the bending energy is also implemented to simulate long semiflexible chains.

The second model is a lattice polymer model, specifically, we are using the Bond Fluctuation Model (BFM) [109] to simulate short linear chains of size up the $N = 160$ in cubic and rectangular confinement. The local "L6" move is used at each Monte Carlo move. These conformations are correlated due to the local moves. We calculate the autocorrelation time τ_{int} following the routine outlined in Sokal [70] based on the radius of gyration. We took conformations into account that are separated at least $2\tau_{int}$ Monte Carlo steps [185]. About 10 000-15 000 independent conformations were generated for each parameter set.

The autocorrelation time τ_{int} for longer and highly stiff chains can be extremely high. A combination of the local "L26" move and a pivot move are employed within the BFM to simulate longer and stiffer bottle-brush polymers owing to the reduction of relaxation and autocorrelation time [114]. For polymer chains in confinement, the local moves are kind of indispensable because of the high rejection rate of global moves in finite space. In our case, the BFM with "L6" move is adequate to simulate short chains in cubic and rectangular confinement.

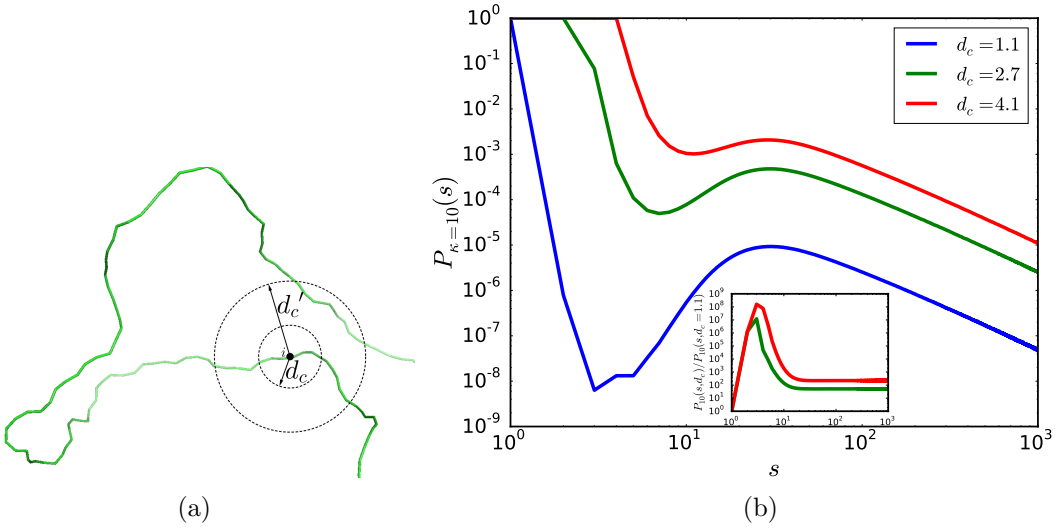
3.3 Results

3.3.1 Contact probability of the semi- and flexible chains

Does the contact probability as a function of the contour distances depend on the definition of what a contact is? On a lattice we can define a contact if two monomers occupy nearest neighbor sites. But then, we could also define a contact taking place at next nearest neighbor sites. In continuum we need to define a distance such that whenever two monomers are within the defined distance this would count as a contact. We refer to this defined distance as the cut-off distance d_c. **Figure 3.1(a)** shows the number of monomers which are in contact with monomer i (solid point). The number of contacts depends on the value of the cut-off distance d_c. Specifically, the contact probability with d_c fixed is calculated as follows: if the distance d_{ij}

between monomer i and monomer j is smaller than d_c, then the contribution to the contact probability is $p_\kappa(|i-j|) = 1$ (κ is the stiffness parameter of the chain), otherwise $p_\kappa(|i-j|) = 0$. The entire contact probability is the average over all pairs of i, j and sufficient independent conformations in equilibrium: $P_\kappa(s) = \langle\langle p_\kappa(|i-j|)\rangle_{|i-j|=s}\rangle_c$ where s is the contour distance between monomers.

To establish the asymptotic behavior of the contact probability $P_\kappa(s)$ with respect to contour length s, we simulated long chains (see Figure 3.1(b)). In Figure 3.1(b) the result for a semiflexible homogeneous chain $N_b = 10000, \kappa = 10$ using the continuous pivot algorithm is shown. The contact probability of this chain is calculated with three cut-off distances $d_c = 1.1, 2.7, 4.1$. Only the range $s < 1000$ is shown since $P_\kappa(s)$ for bigger s has rather large statistical fluctuations. The results prove that how we define the cut-off distance d_c for the contact of monomers does not change the asymptotic behavior of the contact probability $P_\kappa(s)$. As long as the value of d_c is not too large compared to the persistence length, $P_\kappa(s)$ has a similar structure over all length scales: a minimum (only for relatively large κ, discussed later) when s is small, and the same power law decay (roughly $s^{-2.2}$) when $s \gg l_p$ (c.f. inset). Shown in the inset are the ratios of contact probabilities for $d_c = 2.7$ and 4.1 over $P_\kappa(s)$ for $d_c = 1.1$. When s is large enough, the ratios level out, showing that the contact probabilities have the same asymptotic behavior only with different prefactors.



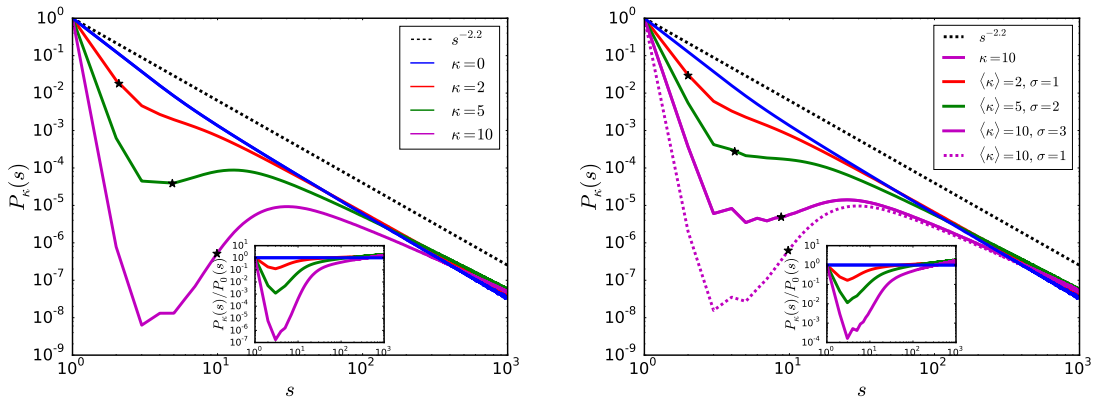
(a) (b)

Figure 3.1: Panel **A** shows the definition of a contact within the cut-off distance d_c. All the other monomers inside the dashed circle (sphere in 3D) with radius d_c are in contact with monomer i (solid point). Certainly if the cut-off distance d_c is larger, there are potentially more monomers contributing as contacts. Panel **B** shows the contact probabilities for different cut-off distances d_c. The results are for a chain of length $N_b = 10000$, $\kappa = 10$, bond length $l_b = 1$ and the radius of hard sphere representing one monomer of $r = 0.4$. These results were obtained using the continuous pivot algorithm. Only the range $s < 1000$ is shown. Different d_c only affect $P_\kappa(s)$ in the range of small s, while the asymptotic power law decay behavior of $P_\kappa(s)$ is recovered as is shown in the inset, where we plot the ratios of the probabilities. Other values of polymer length N_b and κ give similar results.

The influence of bending rigidity and its distribution on the contact probability $P_\kappa(s)$ (the cut-off distance d_c is set to be 1.1) is shown in **Figure 3.2**, where results for chains of homogeneous (panel **A**) and heterogeneous stiffness (panel **B**) are presented. In panel **A**, the contact probabilities of semiflexible chains exhibit a drop in the range of small s compared to the flexible chain (blue line). This is because of the fact that bending energy contributes to the parallel of successive chain segments, inducing larger separation between monomers than flexible chains. A local minimum exists if κ is large enough (roughly $\kappa > 3$). When $s \gg 1$, $P_{\kappa=0}(s)$ shows the asymptotic behavior of $P_{\kappa=0}(s) \sim s^{\gamma_0}$ for flexible chains. The exponent γ_0 is approximated by $3(1-3\nu) \approx -2.3$ if the monomers are considered as particles independently distributed in space, or more precisely by $(3+\theta)\nu \approx -2.2$ where θ is a parameter about 0.70 [154, 155] and $\nu \approx 0.588$ is the critical exponent of self-avoiding walk. For $\kappa > 0$, in the region $l_p \ll s < 1000$ the exponent γ_0 deviates from

this value as can be seen from the inset of Figure 3.2, instead it's a larger exponent for $\kappa > 0$. Nevertheless, this doesn't mean that for even larger s the semiflexible chains have a different exponent from the flexible chain, since $P_{\kappa>0}(s)$ is leaning down when s grows to 1000.

Figure 3.2(b) shows the contact probabilities of chains with Gaussian distributed stiffness parameter κ_i. The mean values of $\kappa_i = 2, 5, 10$ are same as for the homogeneous chains. The corresponding standard deviations are $\sigma = 1, 2, 3, 1$. Comparison of Figure 3.2(a) and 3.2(b) reveals that the heterogeneous chains have more contact in the small s region than the homogeneous chains with the same averaged stiffness parameter. What's more, for the Gaussian distribution of κ_i we studied, the contact probability increases with the standard deviation σ. This means that κ_i smaller than $\langle\kappa\rangle$ has more influence on the conformation than κ_i which is larger than $\langle\kappa\rangle$. In other words, the heterogeneity flexibilizes the chain. Nonetheless, in length scale much larger than the persistence length, they have the same contact probability despite the heterogeneity.



(a) (b)

Figure 3.2: Contact probability $P_\kappa(s)$ for flexible and semiflexible chains with homogeneous and heterogeneous stiffness. The homogeneous chains have the rigidity parameter $\kappa = 2, 5, 10$ (panel **A**). In the heterogeneous chains each κ_i is sampled from a Gaussian distribution with mean values $\langle\kappa\rangle = 2, 5, 10, 10$ and corresponding standard deviation $\sigma = 1, 2, 3, 1$ (panel **B**). Only the region $s < 1000$ is showed since $P_\kappa(s)$ for bigger s has large statistical fluctuations. The black dashed line in both figures is the power law $s^{-2.2}$ which is the predicted asymptotic behavior for the self-avoiding walk [154]. The star points indicate the values of persistence length l_p in these cases. The chains $N_b = 10000$ are simulated using the pivot algorithm. For the homogeneous chains (panel **A**), when $s < l_p$, the bending energy that tends to align neighboring bond vectors prevails over the entropy, therefore $P_\kappa(s)$ shows a drop compared to the flexible chain ($\kappa = 0$), and a minimum exists if κ is large enough. In the range $l_p \ll s < 1000$, $P_{\kappa\neq0}(s)$ shows a power law decay with an exponent slightly larger than the flexible chain. The persistence lengths extracted from the orientational correlation function are $l_p = 2.09, 4.88, 9.90$ for $\kappa = 2, 5, 10$. For the heterogeneous chains (panel **B**), the contact probabilities drop less in the small contour length range compared to the homogeneous chains due to the heterogeneity of stiffness along the chain even though the mean values are the same ($l_p = 2.00, 4.20, 8.77, 9.80$ for the four cases). Nevertheless, they have similar asymptotic behavior in the range $l_p \ll s < 1000$.

Hence, the bending rigidity and its heterogeneity mainly exert influence on contact probability in the region where s is smaller than the persistence length l_p for the polymers in free space, while the asymptotic behavior of $P_\kappa(s)$ are similar. For semiflexible chains in finite space, the bending rigidity not only leads to a drop of contact probability in region $s < l_p$, but also introduces an oscillation for $s > l_p$, as will be shown later.

3.3.2 Persistence Length and Structure Factor

For the worm-like chain without the excluded volume effect, the mean square end-to-end distance is:

$$\langle R_e^2 \rangle = 2l_p L \left[1 - \frac{l_p}{L}(1 - e^{-L/l_p}) \right] , \tag{3.3}$$

where the persistence length l_p is defined by $\langle \cos(\theta(s)) \rangle = \exp(-s/l_p)$. In the limit $L \gg l_p$, $\langle R_e^2 \rangle = 2l_p L \propto N_b$.

For most of the real chain systems except those in the melt condition or in the θ-solvent where polymers act like ideal chains, the excluded volume effect leads to chain swelling, resulting in different scaling exponent $\nu \approx 0.588$ for the end-to-end distance and radius of gyration according to the renormalization group method, or $\nu = 0.6$ by the Flory approximation:

$$\langle R_e^2 \rangle = C_e N_b^{2\nu}, \langle R_g^2 \rangle = C_g N_b^{2\nu}, N_b \to \infty, \tag{3.4}$$

where C_e, C_g are related to the persistence length.

However, despite the existence of excluded volume effects, equation (3.3) does validate itself in semiflexible real chain systems for limited length scale determined by the persistence length l_p [126]. In fact there are two regimes where equation (3.3) does apply: the first one is $s \leqslant l_p$ where the chain behaves like a rod; the second one is $l_p \ll s < s^*$ where the chain can be viewed as ideal since monomers can hardly "collide" and consequently the excluded volume effects are negligible. The value of s^* depends on the persistence length l_p as $s^* \propto l_p^3$ according to the Flory argument, or numerically $s^* \propto l_p^{2.5}$ [186].

There are several ways to determine the persistence length. The traditional one is defined through the exponential decay of the orientational correlation function $\langle \cos\theta(s) \rangle$ (Equation (3.2)). Although for both random walk and self-avoiding walk the orientational correlation function shows a power law decay behavior [132, 180, 181] at a large length scale $s > s^*$, this stays a good estimator considering that it can recover the stiffness parameter κ (**Figure 3.4**) and that l_p should not depend on the polymer length (**Figure 3.3** blue solid line). Another way is to calculate l_p from equation 3.3 or simply $\langle R_e^2 \rangle / 2N_b l_b$ when L is large enough. Clearly for real chains this is not reliable since $\langle R_e^2 \rangle \sim N_b^{2\nu}$ due to the excluded volume effect, thus $\langle R_e^2 \rangle / 2N_b l_b$ would increase with N_b (Figure 3.3 green solid line). On the other side, $\langle R_e^2 \rangle / 2N_b^{2\nu} l_b$ doesn't give reliable results either as shown in Figure 3.3 (red solid line) because when N is not very large the stiffness weakens the excluded volume effect. Shown in Figure 3.4 is the dependence of persistence length l_p on the bending rigidity parameter κ and its distribution, in which l_p is extracted by fitting the exponential decay to the orientational correlation function. The values of persistence length

for homogeneous $N_b = 1000$ chains are represented by blue open circles, the linear fitting of which has a slope equal to 1, suggesting the relation $l_p \approx \kappa\langle l_b\rangle$ ($l_b = 1$).

The heterogeneous chains have various stiffness parameter κ_i, hence persistence length l_{p_i}, along the backbone. We are interested in how the average persistence length over the chain would change due to the heterogeneity. As mentioned above, we assume that κ_i obeys the Gaussian distribution with mean value $\langle\kappa\rangle$ and standard deviation σ. Considering the exponential decay of orientational correlation function in equation (3.2), when $s = 1$ the persistence length is roughly approximated by $l_p \approx -1/\ln\langle\cos\theta\rangle$. For homogeneous chains, $\langle\cos\theta\rangle \approx \exp(-1/\kappa)$, while for heterogeneous chains,

$$\langle\cos\theta\rangle \approx \langle\exp(-1/\kappa)\rangle = \int_0^\infty \exp(-1/\kappa) f(\kappa) d\kappa, \tag{3.5}$$

where $f(\kappa)$ is the distribution function of the stiffness parameter. The integration starts from 0 because we don't make allowances for negative stiffness parameter κ. The consequential bias from Gaussian distribution is negligible when we take the standard deviation $\sigma \leqslant \langle\kappa\rangle/3$. Obviously on the right hand side of this equation the smaller values of κ contribute more to the integration, leading to a smaller $\langle\cos\theta\rangle$ and hence a smaller l_p compared to the homogeneous case. The dashed lines in Figure 3.3 and open circles in Figure 3.4 show the simulation results of the average persistence length for heterogeneous chains. These chains have smaller persistence lengths and end-to-end distances, which leads to the conclusion that the heterogeneity flexibilizes the chain.

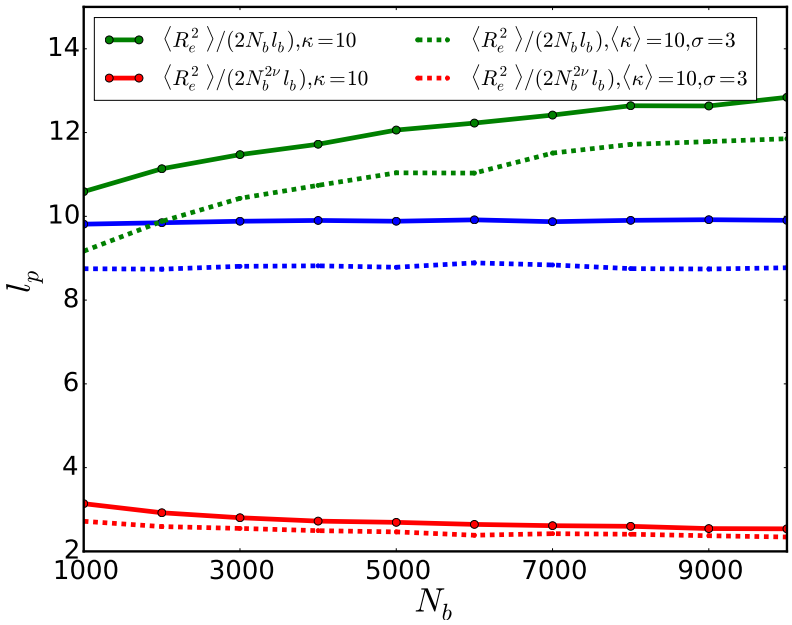


Figure 3.3: Shown is the persistence length l_p for homogeneous chains (solid lines) and heterogeneous chains (dashed lines) calculated from (1) blue lines: the exponential fit to the orientational correlation function (see equation 3.2); (2) green lines: $\langle R_e^2 \rangle / 2N_b l_b$; (3) red lines: $\langle R_e^2 \rangle / 2N_b^{2\nu} l_b$. The average bending rigidity parameter is $\langle \kappa \rangle = 10$, and $\sigma = 3$ for the heterogeneous case.

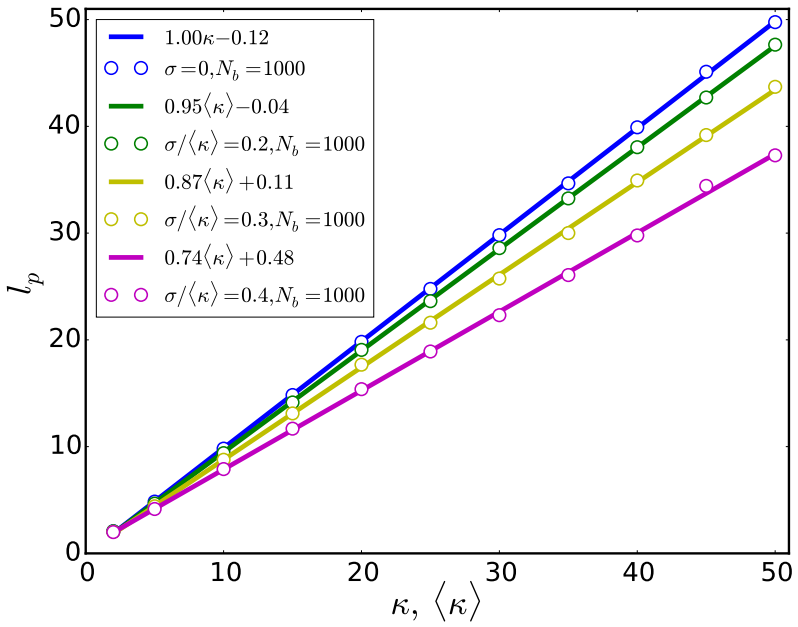


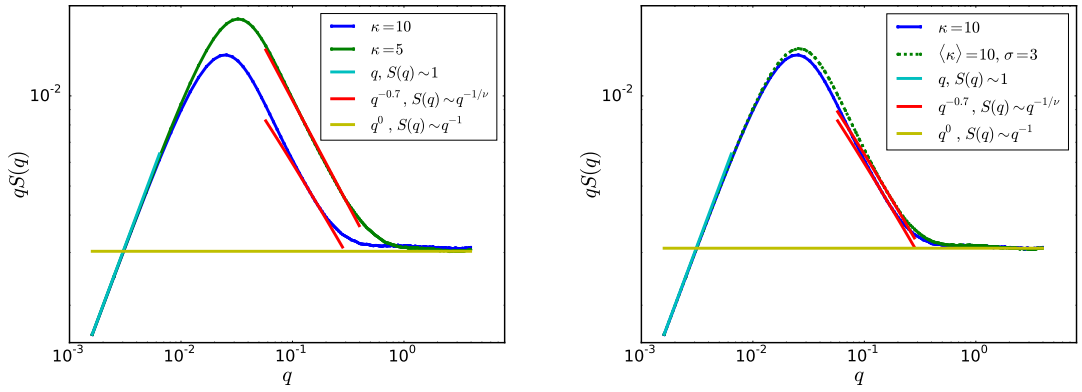
Figure 3.4: Dependence of average persistence length l_p on the bending rigidity parameter κ and its distribution with different standard deviation $\sigma = 0, 0.2, 0.3, 0.4$. The relation between l_p and $\langle\kappa\rangle$ can be considered linear, with slope dependent on σ. The open circles indicate values of l_p, while the solid lines are the linear fitting results. l_p is extracted by fitting an exponential decay to the orientational correlation function. The bond length is $l_b = 1$ in the continuous pivot algorithm. Results for $N_b = 100$ and 500 are not shown as they are almost on top of the data for $N_b = 1000$. The results show for homogeneous chains that the relation between persistence length and stiffness parameter is $l_p \approx \kappa\langle l_b\rangle$. In the heterogeneous case the average persistence length would be smaller with increasing σ.

Experimentally the persistence length l_p is usually calculated from the structure factor $S(q)$ which can be measured by the neutron scattering experiments, q is the wavenumber. The structure factor $S(q)$ is defined as:

$$S(q) = \frac{1}{N^2}\langle\sum_{i=1}^{N}\sum_{j=1}^{N}\exp[i\vec{q}\cdot(\vec{r}_i - \vec{r}_j)]\rangle \tag{3.6}$$

The semiflexible chain behaves rod-like at a small length scale $s < l_p$, and recovers the self-avoiding property at a much larger length scale $s^* \gg l_p$. This indicates the existence of two regimes in $S(q)$: the self-avoiding regime and the rod-like regime. In addition, as discussed above, in the region $l_p \ll s < s^*$, the semiflexible chain behaves more like a random walk since the excluded volume effect can be

ignored. Thus there are several corresponding features for these different regimes in the structure factor $S(q)$. When q is quite small, $S(q) \approx 1 - q^2\langle R_g^2\rangle/3$, which is the Guinier regime. In the region $q > 1/\langle R_g^2\rangle$, $S(q)$ shows the self-avoiding regime: $S(q) \propto q^{-1/\nu}$. Then the crossover from self-avoiding region to random walk region occurs at $qR^* = 1$ ($R^* \propto l_p^2$) [126], where $S(q)$ changes to $S(q) \propto q^{-2}$. When $ql_p > 1$, $S(q)$ exhibits the rod-like property $S(q) \propto q^{-1}$. In the Kratky plot $qS(q)$ (**Figure 3.5**), the rod-like region is the "Holtzer plateau". Therefore the persistence length l_p can be approximated from the onset of the horizontal region in the Kratky plot. Based on the above discussion, there should be three crossovers for $S(q)$ [126, 133], but not all the crossover can be seen clearly in the $S(q)$-plot. The Guinier, self-avoiding and rod-like regimes are present in Figure 3.5, while the random walk regime is hidden. The Gaussian random walk regime can be visible only when the persistence length l_p is large enough [126]. The structure factor of homogeneous $\kappa = 5, 10$ chains and heterogeneous chain with $\langle\kappa\rangle = 10, \sigma = 3$ are shown in Figure 3.5. The latter has a smaller persistence length, hence its structure factor is shifted compared to the homogeneous $\kappa = 10$ chain.



(a) (b)

Figure 3.5: Kratky log-log plot of $qS(q)$ versus q for homogeneous and heterogeneous chains. The chain length is $N_b = 1000$. In panel **A**, the green and blue lines are for homogeneous chains with $\kappa = 5$ and $\kappa = 10$. Three regimes can clearly be seen: the Guinier regime $S(q) \approx 1$ when $q \ll 1$ (cyan line), the self-avoiding regime $qS(q) \propto q^{1-1/\nu}$ (red lines), the rod-like regime $qS(q) \propto q^0$ (yellow line). The random walk regime is absent because the persistence length l_p or κ is not large enough. Panel **B** compares the structure factors of homogeneous chain and heterogeneous chain with the same $\langle\kappa\rangle = 10$. The latter has a smaller persistence length, hence its structure factor is shifted compared to the homogeneous chain.

3.3.3 The Chain in Confinement

There have been studies on semiflexible linear and ring polymers under different kinds of confinements, for example in a spherical capsule [187], in a channel and in a cavity [188], in a cylinder [189] and in rectangles [34, 36]. Here we will use the bond fluctuation model to explore different aspects of the structure of semiflexible chains in cubic and rectangular confinement, including the contact probability, the ordering of chain segments and the orientational correlation function.

Within finite space, the conformations of semiflexible chains depend on the persistence length l_p and the linear dimension a of the enveloping space, resulting in a "shape transition" [36, 187, 188]. When $l_p \ll a$, chain segments are randomly orientated (**Figure 3.6(a), 3.6(b)**), although at length scales smaller than l_p, they are more ordered due to the bending rigidity. However, when the persistence length l_p is comparable to or larger than the linear dimension a, the chain has to adopt an ordering (Figure 3.6(c), 3.6(d)) as a consequence of the competition between confinement, bending energy and entropy.

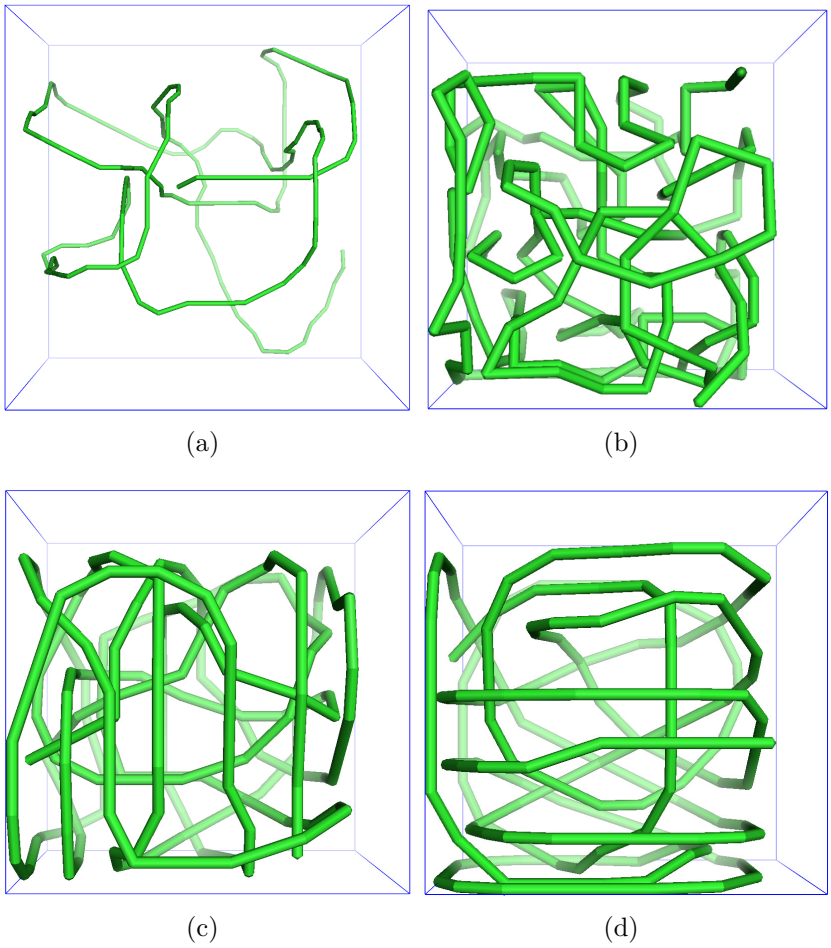


Figure 3.6: Typical conformations of $N = 160$ in a cubic box with side length a and bending rigidity parameter κ: **A** $a = 40, \kappa = 9, l_p \approx 24.3$; **B** $a = 16, \kappa = 2, l_p \approx 5.4$; **C** $a = 16, \kappa = 9, l_p \approx 24.3$; **D** $a = 16, \kappa = 20, l_p \approx 54$. The persistence length l_p here refers to the value when no confinement is imposed, and is roughly approximated by $l_p \approx \langle l_b \rangle \kappa \approx 2.7\kappa$. When l_p is smaller than the box, the chain forms spirals but they are randomly ordered (see **A, B**). When l_p is comparable to or larger than the box size, the spirally chain has to arrange itself in an orderly way (see **C, D**). Meanwhile, figures **C, D** show that the conformations do not differ significantly as l_p/a becomes even bigger.

The significant difference between the contact probability of polymers in symmetric confinement and in free space is that for the former it does not drop when $s > l_p$. Instead, considering for example a cubic box with side length $a = 40$, $P_\kappa(s)$ levels off after $s > s_c$ (**Figure 3.7(a), 3.7(b)**) where s_c depends on the box size and κ, even when κ is not zero, corresponding to the conformations before the "shape transition" occurs. In this case $s_c < N/2$, which means that monomers that are

separated by $N/2$ or more monomers actually have the same probability to contact each other, suggesting that the maximum distance of monomers has been reached in the finite box [190, 191]. Within this space, chains with small κ do not form spirals, while for larger κ, the semiflexible chains begin to spiral but the spirals are not regularly organized in size and direction (Figure 3.6(a)). As the space becomes smaller and κ is larger, the "shape transition" condition is satisfied, the semiflexible chain will organize into spirals to accommodate itself in the finite space (Figure 3.6(c)). The formation of these spirals leads to an oscillation of the contact probability for the length scale larger than the size of spirals (Figure 3.7(c), 3.7(d), 3.7(e), 3.7(f)).

In the left column of Figure 3.7 are the contact probabilities for homogeneous $N = 160$ chains with $\kappa = 0, 3, 7, 9$ confined in cubic boxes of side length $a = 40, 25, 16$. The right column shows contact probabilities of corresponding heterogeneous chains, with standard deviation $\sigma = 1$ or 3. In section 3.3.1 and 3.3.2 we have mentioned that the heterogeneity flexibilizes the chain and induces more contact at a length scale smaller than the persistence length. Here, when in confinement, the heterogeneity could also weaken the oscillation of contact probability in the large s regime because of the enhanced flexibility (Figure 3.7(d), 3.7(f)).

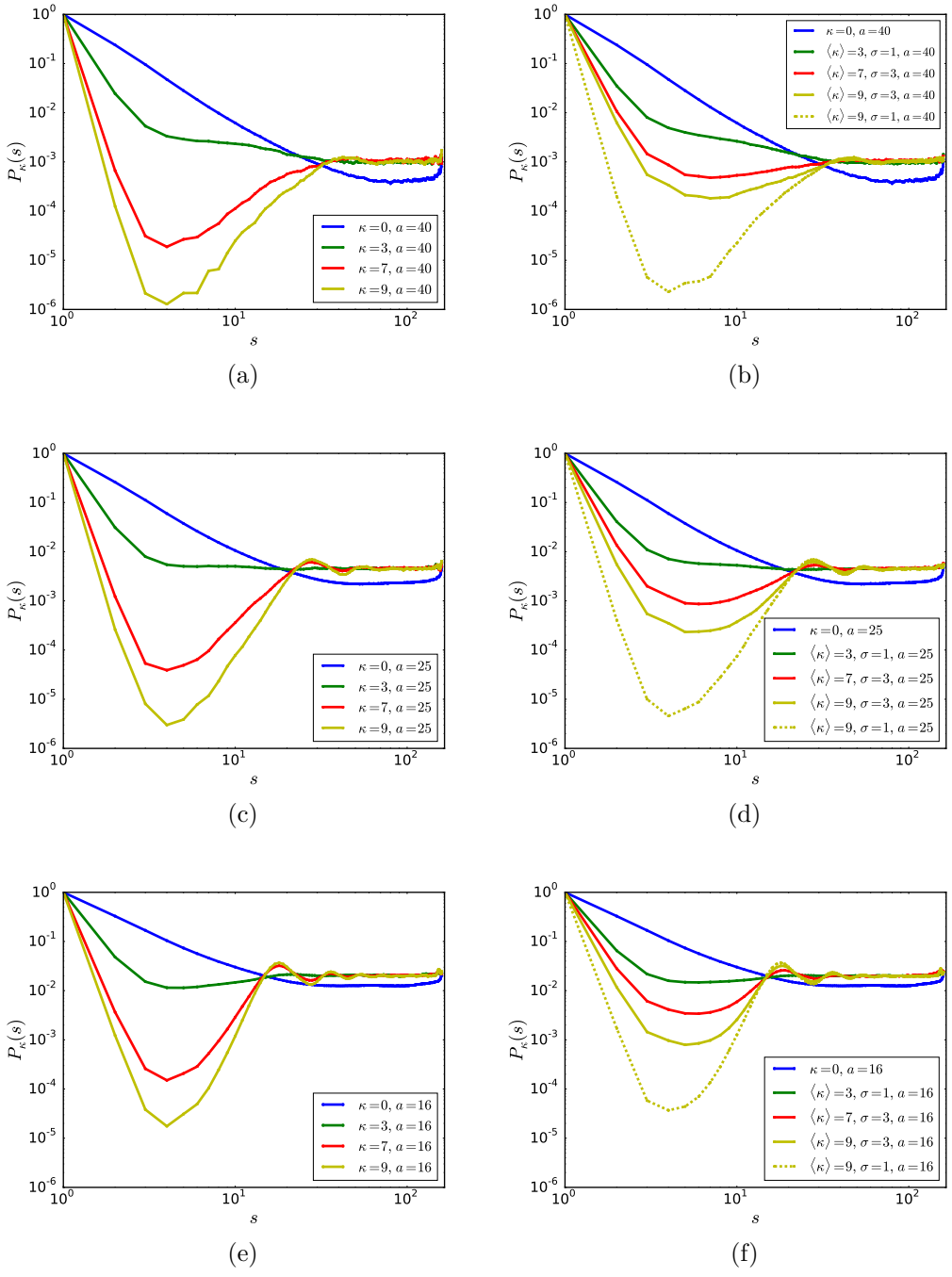


Figure 3.7: Contact probability $P_\kappa(s)$ for $N = 160$ homogeneous (left column) and heterogeneous (right column) chains with different bending rigidity $\langle\kappa\rangle = 0, 3, 7, 9$ in different sizes of cubic boxes: $a = 16, 25, 40$. When the space is finite but not too narrow, $P_\kappa(s)$ begins to level off after $s > s_c$. When κ is larger and the box size is smaller, the chain has to spiral, hence oscillations in $P_\kappa(s)$ appear in large s regime. The heterogeneity induces more contact and weakens oscillation in $P_\kappa(s)$.

For the chains in asymmetric space, the contact probability exhibits a slightly different behavior in the region $s > s_c$ compared to the symmetric case. Shown in **Figure 3.8** are the contact probabilities for flexible and semiflexible chains ($N = 160$) in rectangle boxes with different aspect ratios. As the box is elongated, the oscillation in $P_\kappa(s)$ is distorted for semiflexible chains. The volume of the boxes is about 4000. While in a symmetric box the spirals of the chain have the same radius in all directions on average, in rectangle boxes, the spirals are also elongated, like ellipsoids. The local minimum part is smaller than cubic box case since the space in this direction is narrower and the monomers have higher probability of contact.

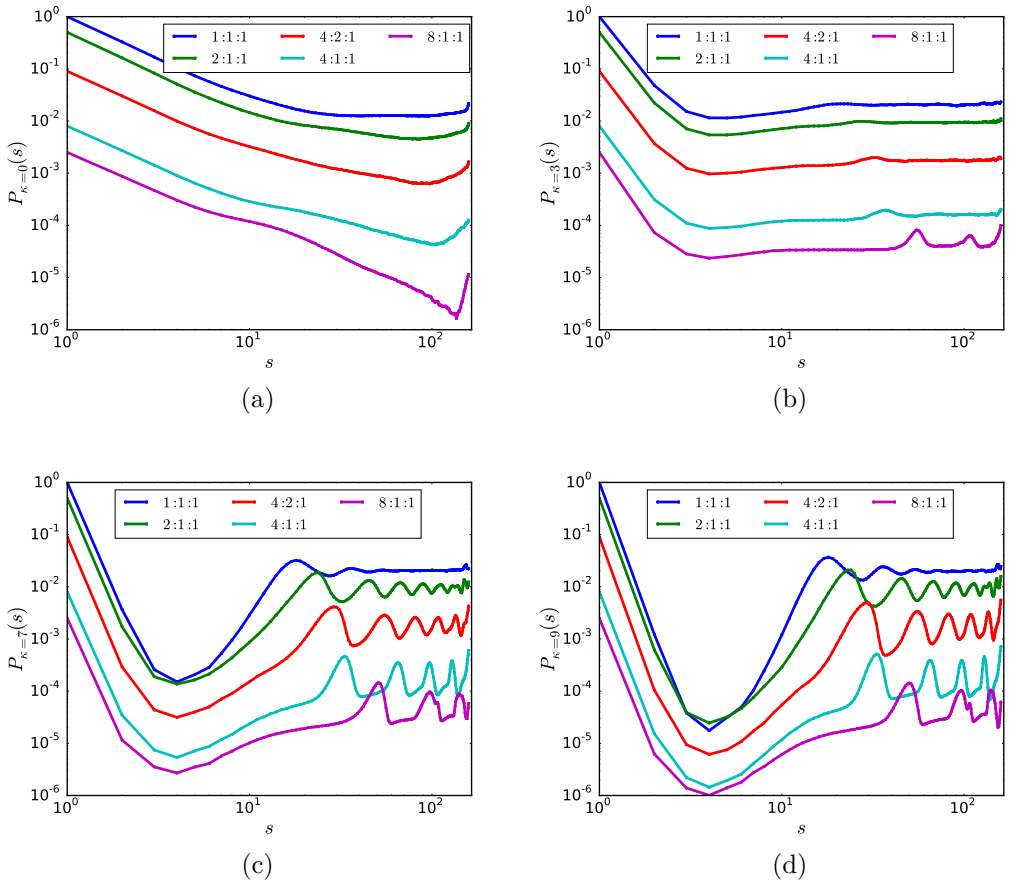


Figure 3.8: The contact probability of chains in rectangle boxes with different aspect ratios. In each figure the probability functions are shifted vertically in order to have a better view of them. (a)$\kappa = 0$; (b)$\kappa = 3$; (c)$\kappa = 7$; (d)$\kappa = 9$.

3.3.4 The effect of bending rigidity

Here we investigate how the shape of the confinement affects the packing of a semiflexible chain. We choose rectangular boxes of different aspect ratios but the same volume: $a : b : c$ = 1:1:1, 2:1:1, 4:1:1, 8:1:1, 4:2:1. Fritsche et al. [34] showed that a semiflexible ring polymer prefers the long axis of the surrounding envelope. This conclusion holds for semiflexible linear chains as well. **Figure 3.9** shows conformations for $N = 160$ and $\kappa = 9$ where the chain is confined in selected boxes.

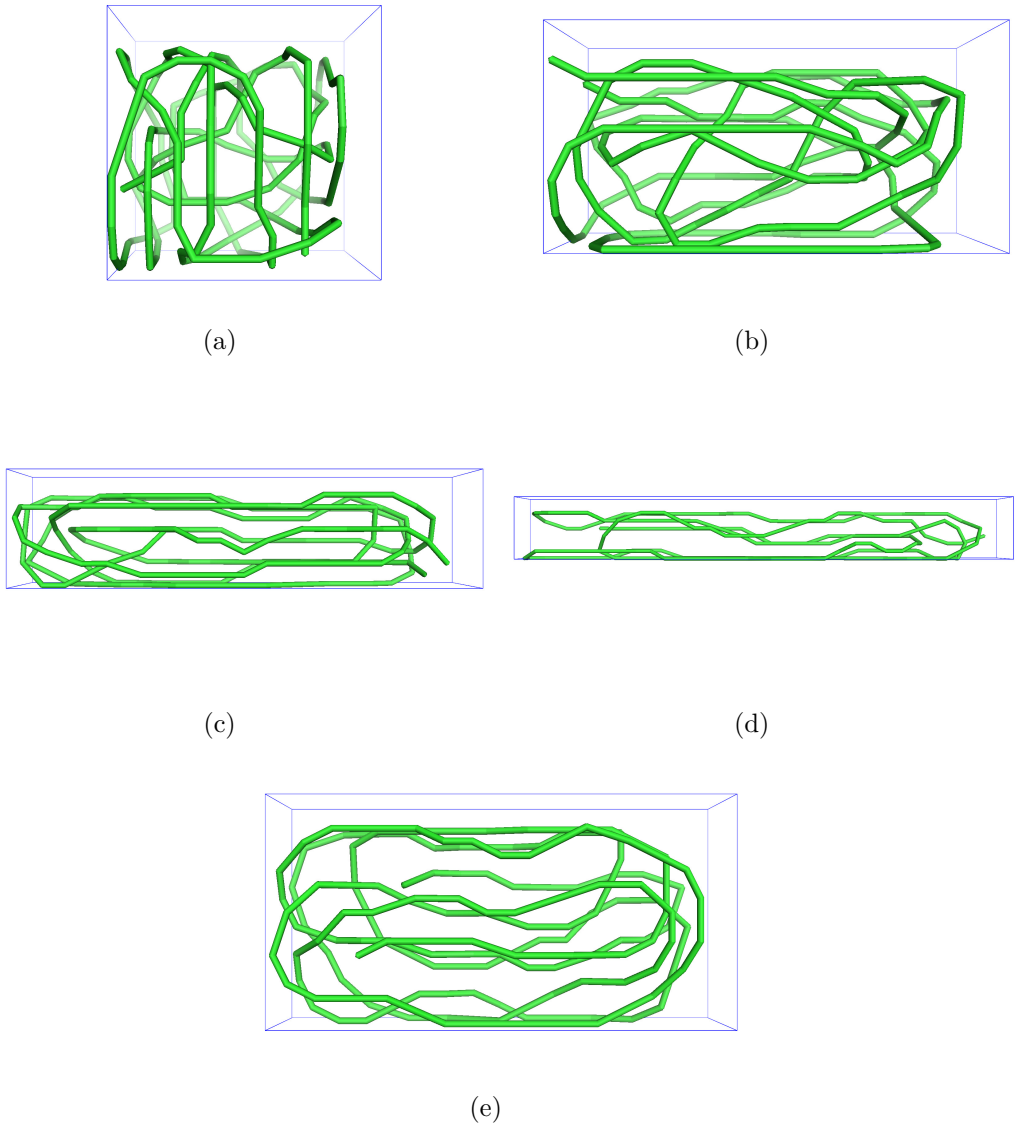


Figure 3.9: Chain ($N = 160, \kappa = 9$) conformations in rectangular boxes of different aspect ratios but the same volume $V \approx 4000$. The aspect ratios are: 1:1:1, 2:1:1, 4:1:1, 8:1:1, 4:2:1. To minimize the free energy in the narrow space, the semiflexible chain would stretch along the long axis, and spirals around the shortest axis.

To quantify the ordering of chain segments, we use the order parameter S following [34] which is defined as:

$$S = \frac{1}{N-1} \left\langle \sum_{i=1}^{N-1} \left(\frac{3}{2} \cos^2 \theta_i - \frac{1}{2} \right) \right\rangle, \tag{3.7}$$

where θ_i is the angle between chain segment $\mathbf{u}_i$ and the local direction $\mathbf{n}$ of the confined geometry of interest. In the rectangular confinement, $\mathbf{n}$ has three choices, which are parallel to the three sides, namely, $\mathbf{n}_x = (1, 0, 0), \mathbf{n}_y = (0, 1, 0)$ and $\mathbf{n}_z = (0, 0, 1)$. Thus, we have three order parameters S_x, S_y, S_z, each ranging from -0.5 to 1. If the chain segments $\mathbf{u}_i$ have no orientational preference along a given direction $\mathbf{n}$, the order parameter would be $S = 0$, whereas the chain with all $\mathbf{u}_i$ parallel to $\mathbf{n}$ gives $S = 1$, and chain with all $\mathbf{u}_i$ perpendicular to $\mathbf{n}$ has $S = -0.5$. Therefore, $S < 0$ means that the chain segments have a tendency to be organized perpendicularly to $\mathbf{n}$, $S > 0$ indicates the tendency of being parallel to $\mathbf{n}$.

Figure 3.10 shows the order parameters S_x, S_y, S_z for chains of different bending rigidity κ in rectangular boxes of different aspect ratios. In the cubic case, the chain segments have no orientational preference, S_x, S_y, S_z are almost zero, both for the flexible and semiflexible chains. When the box has a longer side (x direction), S_x is positive which means chain segments tend to be parallel to the x direction, while S_y, S_z are smaller than 0. Note that even for the flexible chain ($\kappa = 0$) in a rectangular box, Figure 3.10 shows a positive S_x, this is mainly due to the artificial lattice setting. When the chains are stiffer, we get a larger S_x and hence smaller S_y and S_z. This means that bending rigidity makes the chain order itself along the longer axis in a rectangle confinement.

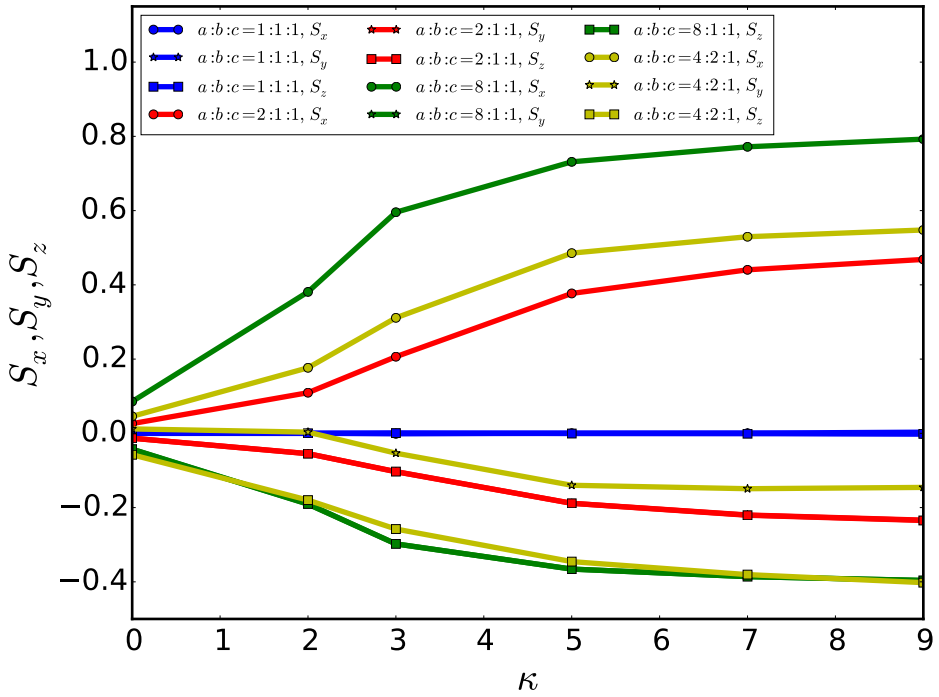


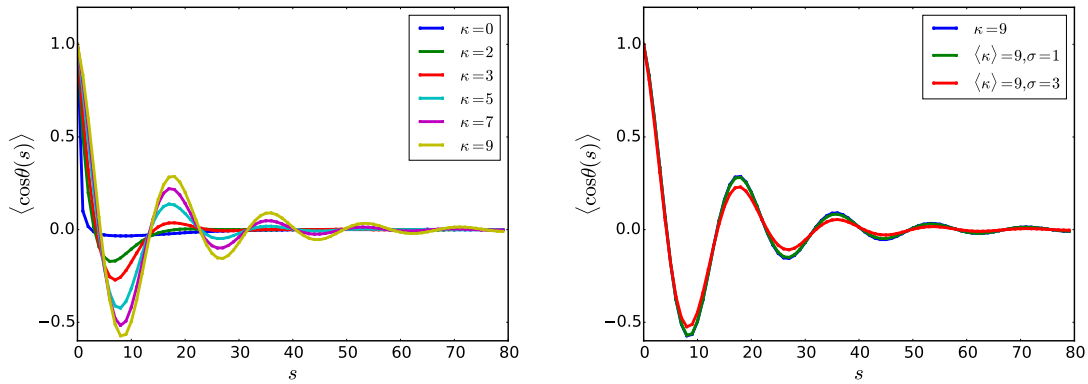
Figure 3.10: The order parameter S_x, S_y, S_z for chains of different bending rigidities κ in rectangular boxes of different aspect ratios: 1:1:1(cubic), 2:1:1, 8:1:1, 4:2:1. A positive S means that the chain segments are more parallel to the corresponding axis. Bending rigidity makes the semiflexible chain order itself along the longer axis, therefore S_x increases with κ for each aspect ratio except the cubic one.

3.3.5 The orientation of bond vectors

Since bending rigidity forces the chain to order itself along the long side of the confinement and to form spirals, the orientational correlation function $\langle\cos(\theta(s))\rangle$ also behaves differently from the exponential decay or power law decay in large contour length regime [132] in free space. Instead, the orientational correlation function shows an oscillation due to the existence of spirals. This correlation function for different confining geometries has been studied extensively both by simulations and experiments [192, 193].

Figure 3.11(a) shows the orientational correlation functions for $N = 160$ with different bending rigidity $\kappa = 0, 2, 3, 5, 7, 9$ confined in a cubic box with side length $a = 16$. When the chain becomes stiffer, the spirals get larger and better ordered. As a result, the oscillations in this function are more pronounced. Another fact from this figure is that the periodicity of the function doesn't change with the value of κ.

This may imply that it may be determined by the box size, which will be discussed later in this section. In Figure 3.11(b) we compare the orientational correlation function of the heterogeneous chains with the homogeneous one. The difference is more identifiable for the $\sigma = 3$ case (red line). This means that the heterogeneity weakens the oscillation by enhancing the flexibility.



(a) (b)

Figure 3.11: Orientational correlation function for different bending rigidity parameters and distributions. Panel **A** shows the orientational correlation function $\langle \cos\theta(s) \rangle$ for chains with $N = 160$ and bending rigidities $\kappa (= 0, 2, 3, 5, 7, 9)$ in boxes of size $a = 16$. Only the range $s \leqslant 80$ is shown. As κ becomes larger, the spirals in the finite space are more ordered, therefore $\langle \cos\theta(s) \rangle$ has larger oscillations. Panel **B** compares the correlation function of the heterogeneous chains with homogeneous one. The oscillation is weakened due to the heterogeneity.

Figure 3.12 shows the orientational correlation functions for $N = 160$ and $\kappa = 9$ in cubical boxes of different sizes. As the space becomes narrower, the chain has more spirals and they are more orderly, thus the amplitude and frequency of the oscillation in $\langle \cos(\theta(s)) \rangle$ get larger.

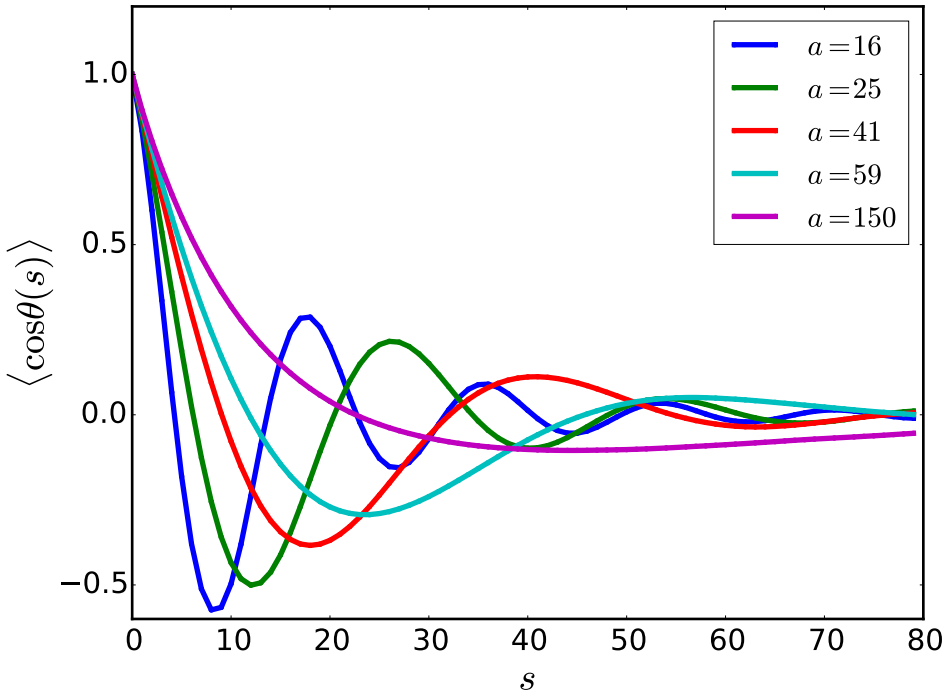


Figure 3.12: The orientational correlation function $\langle \cos\theta(s) \rangle$ for chains of size $N = 160$ and $\kappa = 9$ in different cubical boxes with side lengths $a = 16, 25, 41, 59, 150$. Only the range $s \leqslant 80$ is shown. As the space becomes narrower, the chain has to spiral more orderly, which contributes to the oscillation in $\langle \cos\theta(s) \rangle$.

Liu [36] studied the form of the orientational correlation function $\langle \cos\theta(s) \rangle$ in a 2D square confinement for a worm-like chain, concluding that the leading contribution to $\langle \cos\theta(s) \rangle$ is $e^{-\frac{s}{l_e}} \cos\frac{s}{d}$, where l_e is the effective persistence length, d is linearly related to the size of box a.

Figure 3.13 shows the fitting of the orientational correlation function to $e^{-\frac{s}{l_e}} \cos\frac{s}{d}$ for chain lengths of $N = 20, 40, 80, 160$ and bending rigidity parameter $\kappa = 9$. The fitting values of l_e and d are listed in the caption. The side lengths of the cubic boxes for these four chains are $a = 8, 10, 13, 16$. We have roughly the same ratios of $a/d : 8/1.526 = 5.24, 10/1.883 = 5.31, 13/2.375 = 5.47, 16/2.863 = 5.59$, which means that d is almost proportional to the box size.

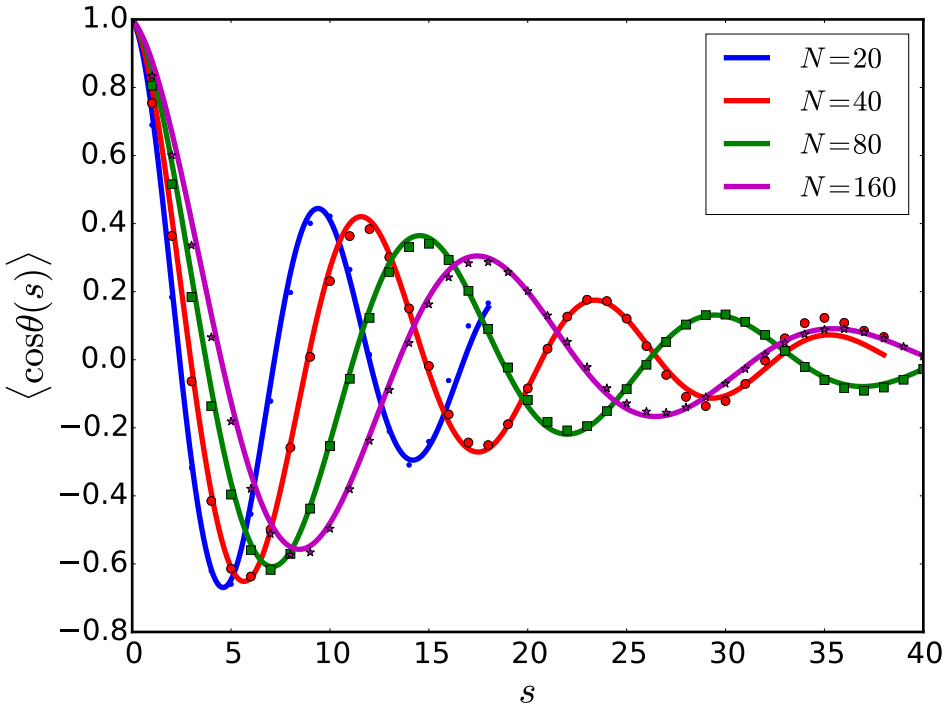


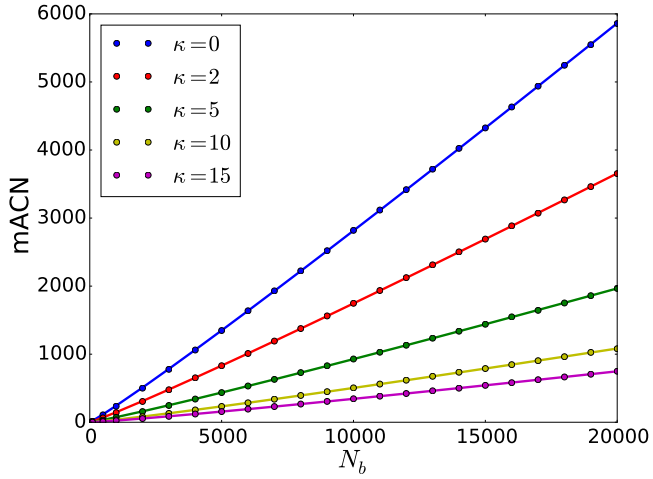
Figure 3.13: Fitting of the leading term $e^{-\frac{s}{l_e}} \cos\frac{s}{d}$ to the orientational correlation data for different chain lengths $N = 20, 40, 80, 160$ with bending rigidity parameter $\kappa = 9$. The points are data calculated from Monte Carlo simulations, and the solid lines are the curves fitted to corresponding points. Only the range $s \leqslant 40$ is shown. The fitting parameters are: $(1)N = 20, l_e = 11.71, d = 1.526; (2)N = 40, l_e = 13.51, d = 1.883; (3)N = 80, l_e = 14.61, d = 2.375; (4)N = 160, l_e = 14.92, d = 2.863$.

3.3.6 Average Crossing Number

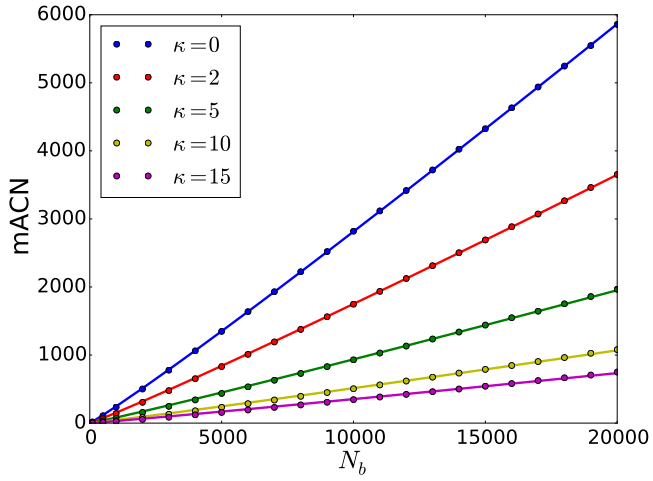
The average crossing number (ACN) is a parameter to quantify the self-entanglement of a chain, defined as the average of the crossing number of the chain when projected onto planes orthogonal to all the directions. Given a chain composed of N_b segments, we use the method in [150] to calculate the ACN. The mean average crossing number (mACN) is obtained by averaging the ACN over all the conformations in thermodynamic equilibrium.

In [151] Diao et al. showed that the relation between the mACN and the number of chain segments N_b can be approximated by $aN_b \ln N_b + bN_b$ for equilateral random walks and polygons. Diesinger and Heermann [152] investigated the influence of excluded volume interactions on the relation for Gaussian and equilateral random walks, concluding that the $N_b \ln N_b$-behavior and the power law aN_b^b fits well for the

mACN of self-avoiding chains.



(a) $N_b \ln N_b$ fitting



(b) power law fitting

Figure 3.14: The fit of the relations $a_1 N_b \ln N_b + b_1 N_b$ and $a_2 N_b^{b_2}$ to mACN versus the segments number N_b for chains with different bending rigidities in free space. N_b ranges from 100 to 20000, and $\kappa = 0, 2, 5, 10, 15$. The chains are simulated using the continuous pivot algorithm with fixed bond length. The $N_b \ln N_b$ and power law relations both fit well for flexible and semiflexible chains, with the coefficients having a dependence on the stiffness parameter κ. The approximated coefficients are: (1) $\kappa = 0, a_1 = 0.0168, b_1 = 0.1268, a_2 = 0.1629$; (2) $\kappa = 2, a_1 = 0.0119, b_1 = 0.0648, a_2 = 0.1014$; (3) $\kappa = 5, a_1 = 0.008, b_1 = 0.0193, a_2 = 0.0542$; (4) $\kappa = 10, a_1 = 0.0054, b_1 = 0.0004, a_2 = 0.0297$; (5) $\kappa = 15, a_1 = 0.0044, b_1 = -0.0059, a_2 = 0.0204$. The exponent b_2 in the power law relation is assumed to be independent of κ. Its value is estimated by fitting the power relation for flexible chain. The residual sum of squares (RSS) and the coefficient of determination (Rsquare) for the two kinds of fitting are shown in table 3.1.

In this section we study the mean average crossing number of semiflexible real chains. Fig 3.14 shows the fitting of mACN for semiflexible chains with excluded volume interactions. The chains are simulated by the continuous pivot algorithm, with chain length ranging from 100 to 20000, and bending rigidity parameter $\kappa =$ 0, 2, 5, 10, 15. By fitting the $N_b \ln N_b$ and power law aN_b^b relations to the data sets (see Figs 3.14(a),3.14(b)), we get the estimations of coefficients for flexible and semiflexible chains. In the fitting of power law, we assume that the exponent b does not change for different κ since on length scale much larger than the persistence length, the chain behaves like a flexible self-avoiding walk. This assumption is checked in figure 3.15 where the ratios between the mean average crossing number and the power law is presented. The leveling off of the ratios for large N_b indicates that the same scaling behavior holds for different κ. The corresponding residual sum of squares (RSS) and the coefficient of determination (Rsquare) which can quantify the goodness of fit are summarized in table 3.1.

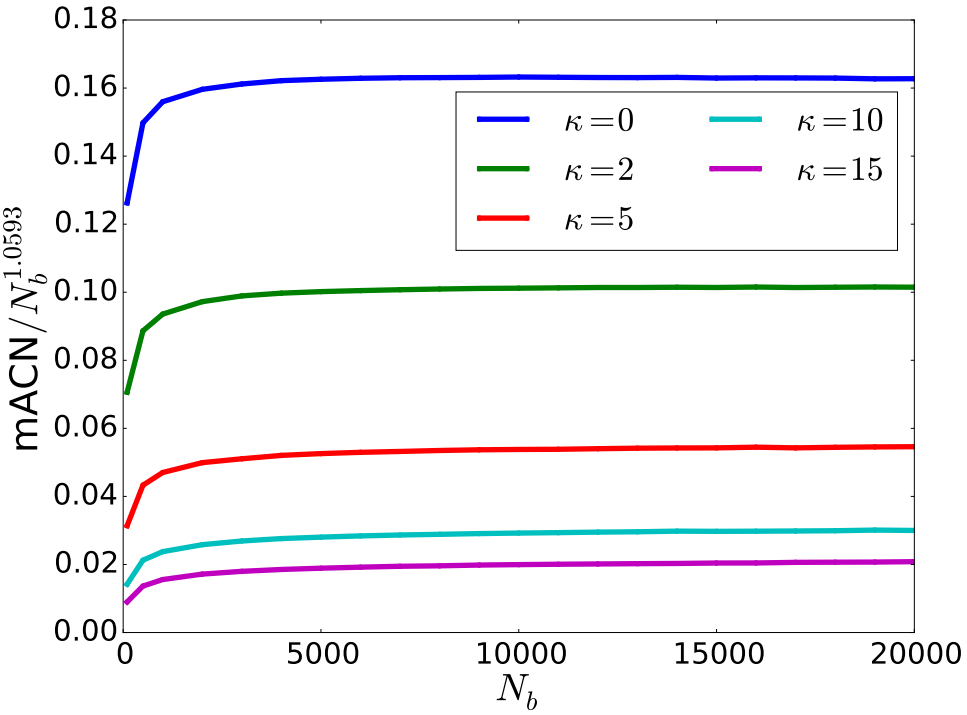
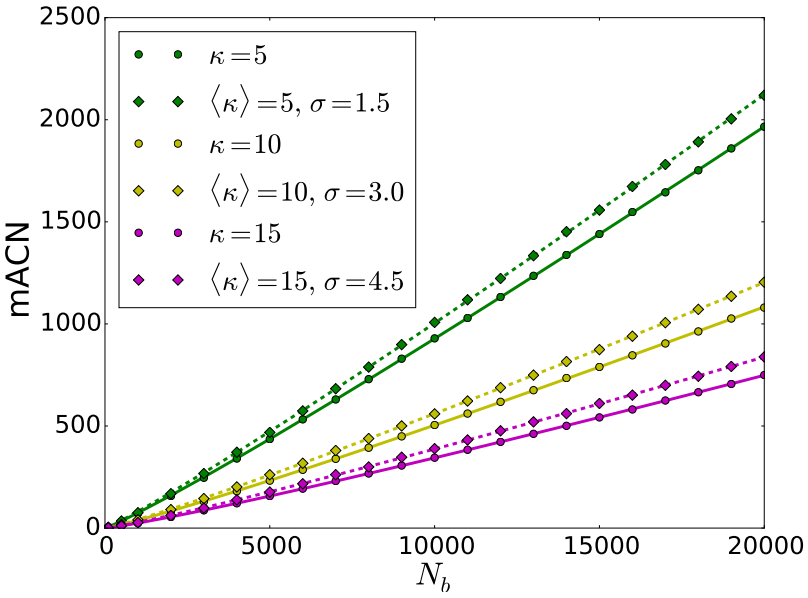


Figure 3.15: The ratio of mean average crossing number to the power law $N_b^{1.0593}$ for $\kappa = 0, 2, 5, 10, 15$. The leveling off for large N_b indicates that the same scaling behavior holds for different κ.

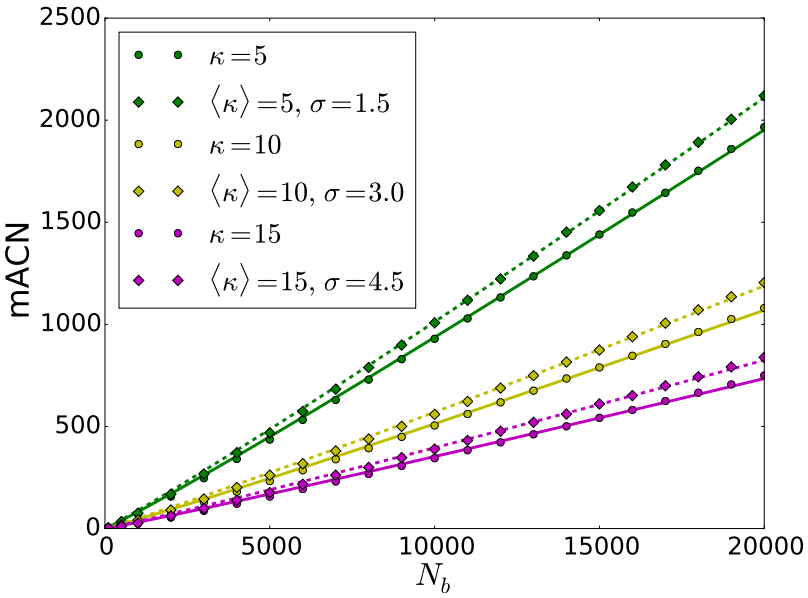
	$\kappa = 0$	$\kappa = 2$	$\kappa = 5$	$\kappa = 10$	$\kappa = 15$
RSS1	383.79	200.62	102.09	65.34	14.23
RSS2	653.89	1043.69	1989.48	1892.44	1684.77
Rsquare1	0.999995	0.999993	0.999988	0.999975	0.999988
Rsquare2	0.999991	0.999964	0.999764	0.999263	0.998628

Table 3.1: The residual sum of squares (RSS) and the coefficient of determination (Rsquare) of the $aN_b \ln N_b + bN_b$ relation (RSS1, Rsquare1) and the power law aN_b^b relation (RSS2, Rsquare2) of the mACN for flexible and semiflexible chains. The chain length N_b ranges from 100 to 20000. Both the two relations fit well with data sets of mACN according to the Rsquare, but the $N_b \ln N_b$ relation seems to be slightly better than the power law relation.

In figure 3.16 we calculate the mean average crossing number for heterogeneous chains (dashed lines) and compare it with homogeneous chains (solid lines). The values of mACN for heterogeneous chains are larger, which is consistent with the conclusion that the heterogeneity enhances flexibility and thus induces more crossing of chain segments.



(a) $N_b \ln N_b$ fitting



(b) power law fitting

Figure 3.16: The mean average crossing number of homogeneous chains (solid line) and heterogeneous chains (dashed line) with $\langle\kappa\rangle = 5, 10, 15$ and $\sigma = 1.5, 3.0, 4.5$. The heterogeneous chains have more crossing of chain segments due to the enhanced flexibility.

Geometrical constraints imposed on the chains could influence their crossing and packing within the finite space. It is reasonable that the size of the confinement would affect the mACN, i.e. the chains have more crossing in smaller space. Here we argue that the symmetry of the confinement will also influence the number of crossing. Fig 3.17 shows the mACN of linear chains in rectangular boxes with different aspect ratios and same volume. The aspect ratios are 1:1:1 (cubic), 2:1:1, 4:2:1, 4:1:1, 8:1:1. The more symmetric the boxes are, the larger mACN is, i.e., the chains are more intermingled.

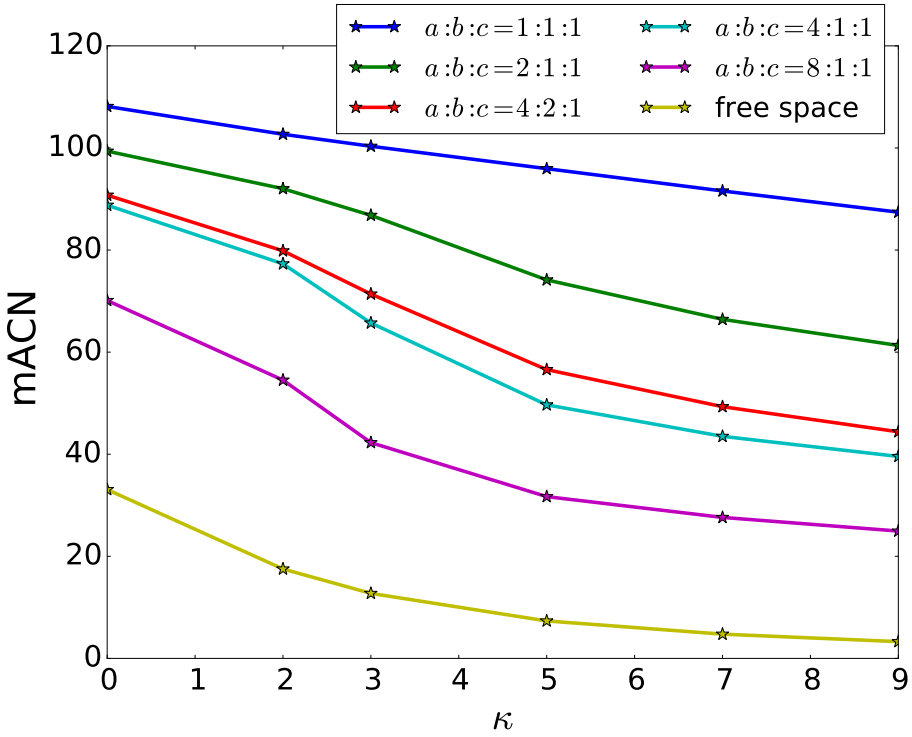


Figure 3.17: The mean average crossing number (mACN) of chains ($N = 160$) in cubic and rectangular boxes with different aspect ratios and same volume(≈ 4000). It is shown that the symmetry of confinement increases mACN, i.e. makes the chain more crossed. This is because the symmetry reduces the effect of bending energy that tends to minimize crossing by stretching along some preferred directions in confinement.

3.4 Conclusion

In this chapter we stressed the contact definition of monomers and the invariability of asymptotic behavior when different cut-off distances and bending rigidity come

into play. At very large length scale, the contact probability for linear chains in free space will exhibit the same power law decay $P(s) \sim s^{-2.2}$, with different coefficients when the cut-off distance for contact is involved.

Secondly, we investigated how the bending rigidity influences the conformations of a linear chain under geometric confinement, represented here by means of cubic and rectangle boxes. The bending potential reshapes the chain due to the competitive interplay of stiffness, entropy and confinement. Moreover, there exists a "shape transition" from overall randomness to orderliness when the persistence length is comparable to the size of confinement. One measure that can reflect the impact of bending rigidity and confinement is the contact probability. The contact probability of a flexible or semiflexible chain in sufficient small confinement has a plateau region in the large contour length region, as opposed to the power law decay in free space. Moreover, if the bending rigidity is big enough compared to the size of the confinement, this plateau region will turn into an oscillation (Figure 3.7), which indicates the existence of spirals formed by the chain.

The ordering of the chain according to the shape of geometric confinement can also be studied by constraining the chain into rectangle boxes of different aspect ratios. An order parameter S is defined to quantify the ordering of chain segments. It is shown that the semiflexible chain preferably chooses the long direction of the boxes to order the segments. The orientational correlation function $\langle \cos\theta(s) \rangle$ of bond vectors is also dramatically changed due to the bending rigidity and confinement, and the oscillation in it serves as a direct evidence of the formation of spirals. The leading term of the analytical expression of $\langle \cos\theta(s) \rangle$ consists of two parts: the first one is the exponential decay term that gives the effective persistence length of the semiflexible chain in confinement; the second part is a cosine function which determines the period of the oscillation mentioned above. This period is dependent on the box size.

The presence of bending rigidity leads to less crossing inside a chain. However, the asymptotic behavior of the mean average crossing number of semiflexible chains stays the same as the flexible case. Two kinds of relationships can describe the dependence of mACN on the chain length. On the other side, the geometrical constraints could induce more crossing, where both the size and symmetry of the constraints kick in. In particular, the more symmetric the confinement is, the more crossing the chain has. The reason is that the symmetry reduces the effect of bending energy that tends to minimize crossing by stretching along some preferred directions in confinement.

It has been pointed out that the three-dimensional organization of chromosomes is tightly coupled to the mechano-genomic code [194]. Our study shows that the modulation of the bending rigidity may be part of the mechano-genomic code regulating the contact probability and thus the three-dimensional organization. It re-

mains to decipher the mechano-genomic code. Here, one of the leading contender is the nucleosomal organization. Nucleosomes contribute due to their steric repulsion and their absence alone to the bending rigidity.

Chapter 4

The Contacts and Segregation in Chains System

Chapter Summary

The self-contacts inside a self-avoiding chain and the contacts between two halves of a chains in free space and in confinement are investigated in this chapter. The total number of self-contacts N_c grows linearly with the length of a free chain, while in cubic confinement it grows quadratically. The distribution function of contacts number between two halves $N_c(AB)$ shows a power law decay and then an exponential decay for a free chain. In confinement, the function has a maximum. As the chain becomes longer, the percentage of inter-half contacts among the total contacts has a power law decay with an exponent close to -1, which supports that the number of contacts between two halves is finite even when the chain is infinitely long. The segregation of two chains in a rectangular box starting from different initial conformations is studied. The results show that without monomer friction, the segregation is dominated by the slow diffusion, thus the required times are similar for these different cases. However, the intertwinement of chains indeed could impede the segregation at small time scale.

4.1 Introduction

The segregation of chromosome under certain circumstance is a significant phenomenon in living systems. In eukaryotic cells, two sister chromatids will separate from each other in order to accomplish the DNA replication or to create haploid cells during the mitotic and meiotic processes. These kinds of segregation are facilitated by cohesin [49] and several other factors [50, 51]. In prokaryotes, the chromosome segregation is also achieved under several mechanisms. For instance, it is found that several genes are involved in the segregation [52], while [53] states that active mitotic-like machinery contributes to the segregation of bacterial chromosomes. Aside from all these biological factors, it is found that the entropy and the structure of chromosomes are also important inducements to the chromosomal segregation under confinement [31, 54, 55].

To elucidate the effects of entropy, lots of analysis and simulations have been performed to study the segregation process and the dynamics of two polymer chains system. Specifically, [54] gave a detailed theoretical analysis on the free energy of mixed and segregated states while varying the size of the confinement and the length of chains. Arnold [56] studied the time scale τ of segregation of two flexible real chains confined in a infinitely long cylinder confinement due to the entropic effect using the molecular dynamics simulation, concluding that the time is proportional to $\tau \sim N^2$, wherer N is the length of each chain. Liu et al. [57] investigated the dependence of the segregation time τ on the length L of a 2D box in which two chains are confined by solving the first-passage time problem and by performing Monte Carlo simulations. Their results showed that there is a minimum of τ as a function of L. Polson et al. [195] measured the dependence of the free energy of the system on the distance between the centers of mass for various confinement aspect ratios and volume fractions of the chains, and found that the free energy at mixed state is reduced if the chains are semiflexible. In addition, the formation of loops inside chains can further enhance the segregation of chains according to [55, 138].

On the other side, the segregation time τ should not only depend on the the shape of the confinement, the size and structure of the chains, but also be related to how the two chains are initially organized. One may assume that it takes longer time for two highly entangled chains to be finally separated. In [196] the Baumgärtner and Muthukumar found that the disentangling of two intertwined chains in free space is characterized by two stages: the unwinding of double helix conformation to interpenetrating chains, and the segregation to isolated chains. The first stage has a time scale $\tau_1 \sim N^{3.0}$, while the second has a time scale $\tau_2 \sim N^{3.3}$. In another literature [197], Diddens et al. studied the positive correlation between segregation time and the number of contacts of the initial conformation of the two chains in a dilute solution. Moreover, they found that initial conformations with larger winding

number usually take longer time to arrive at the segregation state. Inspired by this, we in this chapter discuss how the number of contact and entanglement will influence the segregation process of two linear chains under confinement.

In [197], one method of creating the initial conformations of two chains is to cut a linear chain into two halves. Afterward, these two chains are segregated due to the entropy effect. However, one would ask how these two halves are organized if they are still bonded. We argue that they are more or less separated on average, and this is why the pivot algorithm works so efficiently. In order to unveil this, we analyze the number of contact between the two halves of a chain both in free space and confinement.

The intra-and inter-contact of polymers is also an important topic in biology since it is closely related to the function of these polymers [22]. These contacts are also closely related to the organization of chromatins and can be detected by previously mentioned techniques [38]. Usually by analyzing the Hi-C data, one can obtain some interesting properties such as the contact probability [39–41] and chromosomal contact networks [42, 43]. Some studies concentrated on the number of contacts. In [15, 46] it shows that in the fractal globule model the number of contacts of a certain region with the rest is linear with its volume, rather than the surface area. In addition, the contacts number of two regions is proportional to the product of their volumes. This behavior is related to the scaling of contact probability. Besides, the number of contacts for a chain without constraints is also studied in the context of the self-avoiding walk. It shows that the average number of contacts of a self-avoiding walk has an asymptotic behavior $aN + bN^{\phi} + c$ as the length of the walk N goes to infinity [47]. Baiesi et. al showed that the contact number of two halves of a self-avoiding walk is finite even if the walk is infinitely long [48]. In section 4.3.1 we discuss the contacts within one of the bisection pieces of the chain (c.f. Figure 4.1).

4.2 The Model and Simulation

In order to analyze the self-contact of a single chain, we use the Bond Fluctuation Model (BFM) to perform the Monte Carlo simulation of a self-avoiding walk with length $N = 160$ in free space and in cubic boxes with different side lengths a. In the limit case when the volume fraction of the chain is 1, the walk is a Hamiltonian path. The algorithm in [198,199] is used to generate different Hamiltonian paths. In addition, we also use the continuous pivot algorithm to simulate longer chains ($N = 1000$) in free space. In each case, about 10 000-15 000 independent conformations are generated. Two nonconsecutive monomers are in contact when their distance is smaller than a cut-off distance d_0. The maximum of bond length in the BFM is

$\sqrt{10}$, therefore we set $d_0 = \sqrt{10}$. In the pivot algorithm, the bond length is 1, and d_0 is set to 1.1. We shall be using the following measures of the self contacts: First is the number of contacts N_c that will be calculated as a function of the contour length s_m, where the subscript denoted *from the mid-point* (c.f. Figure 4.1). The second measures the number of contacts when distinguishing between the left and right hand side from the mid-point monomer $N_c(AB)$, and third and fourth $N_c(AA)$ and $N_c(BB)$ the contacts within left and right side from the mid-point, respectively.

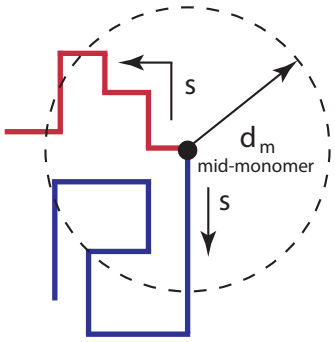
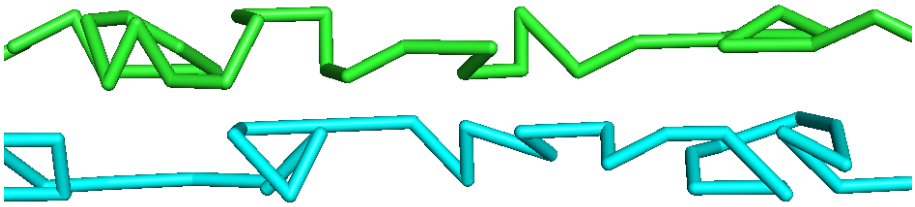
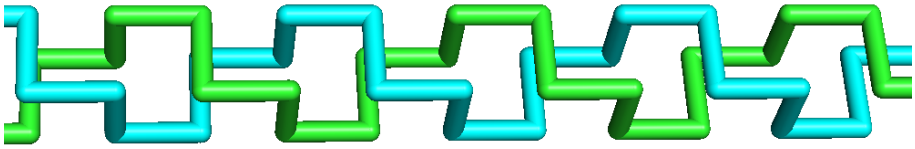


Figure 4.1: The sketch shows how we partition the self-avoiding walk into two equal arms from the mid-point monomer m. The contour length s for possible contacts is measured from the mid-point.

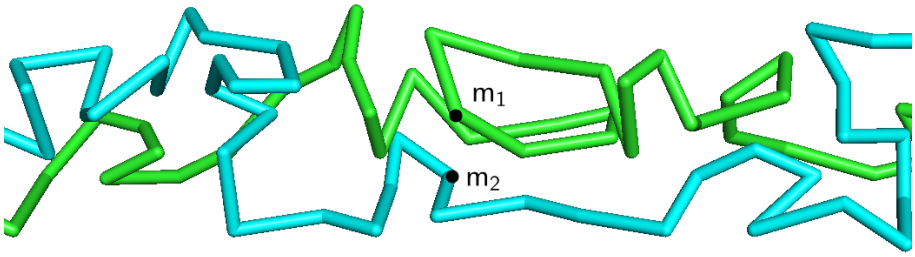
Two linear chains are confined in rectangular box when studying their segregation process. Since the influence of the sizes of the confinement and the chains on the segregation has already been intensively discussed in other literatures, here we set the length of the chain $N = 200$, and the width and height of the box $L_y = L_z = 8$. The average extension along the longitude of the box for one such chain is about 215. The length of the box is set to $L = 100$ and 200. The initial conformations of the two chains are prepared in three ways. Firstly, the two chains are placed so that they are fully overlapped along the longitude of the box but separated by the plane $z = 4$ (Figure 4.2(a)). Secondly, the two chains intertwine with each other [196], thus they have a large winding number (Figure 4.2(b)). Another way is that we tether the mid-monomer m_1, m_2 of the two chains and obtain by simulation the equilibrium states of them in the box (Figure 4.2(c)). Thus the two chains are always overlapped in the equilibrium states. Further, they would have different number of contacts. By studying the segregation of these initial conformations, we can explore the influence of contacts and intertwinement on the segregation time τ.



(a)



(b)



(c)

Figure 4.2: The initial conformations of two chains. (a) They are overlapped along the longitude but separated seeing from the transverse section. (b) The chains are intertwined with each other. (c) We get the equilibrium conformations of the two chains with the mid-monomer m_1, m_2 tethered, thus they are overlapped and have different number of contacts.

4.3 The Results

4.3.1 Two halves of a single polymer

In this section we will show the main results related to the self-contact of a single polymer chain. First we calculate the probability distribution function of the num-

ber of contacts $f(N_c(AB))$ between the two halves of a chain consisting of $N = 160$ monomers in free space and in confinement (Figure 4.3). We use the Monte Carlo method and the bond fluctuation model to simulate the chain in free space and in cubic boxes. The volume fractions of the chain are $\rho = 0.02, 0.04, 0.08, 0.16$. For the very dense system $\rho = 0.96$, the conformations of the chain are generated using the algorithm of Hamiltonian path. In dilute condition, the distribution function $f(N_c(AB))$ decreases monotonically with the $N_c(AB)$, which means that two halves of most conformations of the chain are nearly spatially separated. As the cubic confinement becomes narrower, $f(N_c(AB))$ has a maximum at certain value of $N_c(AB)$ as shown in Figure 4.3.

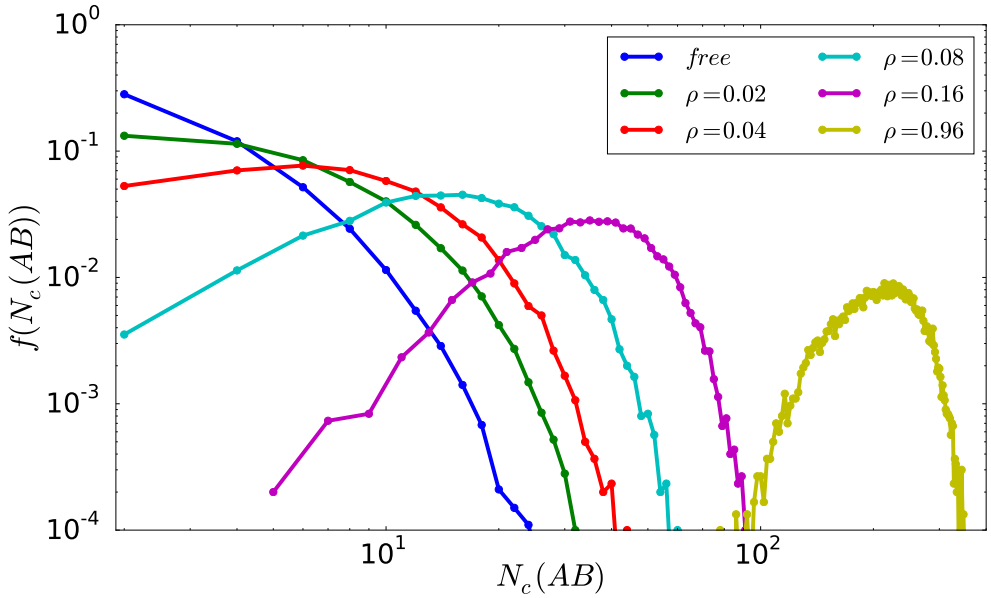


Figure 4.3: The log-log plot of the distribution function of the number of contact between two halves of a linear chain $f(N_c(AB))$ in free space and in cubic boxes. The length of the chain is $N = 160$. The volume fractions of the chain in cubic boxes are $\rho = 0.02, 0.04, 0.08, 0.16, 0.96$. In the case $\rho = 0.96$, the chain is simulated using the algorithm of Hamiltonian path, while the others are simulated using the Monte Carlo method and the BFM. The cut-off distance for contact is set to $\sqrt{10}$.

In [197] it is reported that there is a scaling law behavior in $f(N_c(AB))$ for free chains. This can be seen for longer chains (Figure 4.4). The pivot algorithm is used to simulate the chain with length $N = 10000$. Different definitions of d_c give similar behavior of $f(N_c(AB))$ but lead to different exponents. Here the cut-off distance for d_c is set to 2.0 in order to get a better approximation. For small $N_c(AB))$, the distribution function $f(N_c(AB))$ shows a power law decay $f(N_c(AB)) \propto N_c(AB)^{-1.02}$,

while in large $N_c(AB)$) region, the function has an exponential decay. The distribution of $N_c(AB)$) for different chain lengths also show that the number of contacts of the two halves is finite [48, 200].

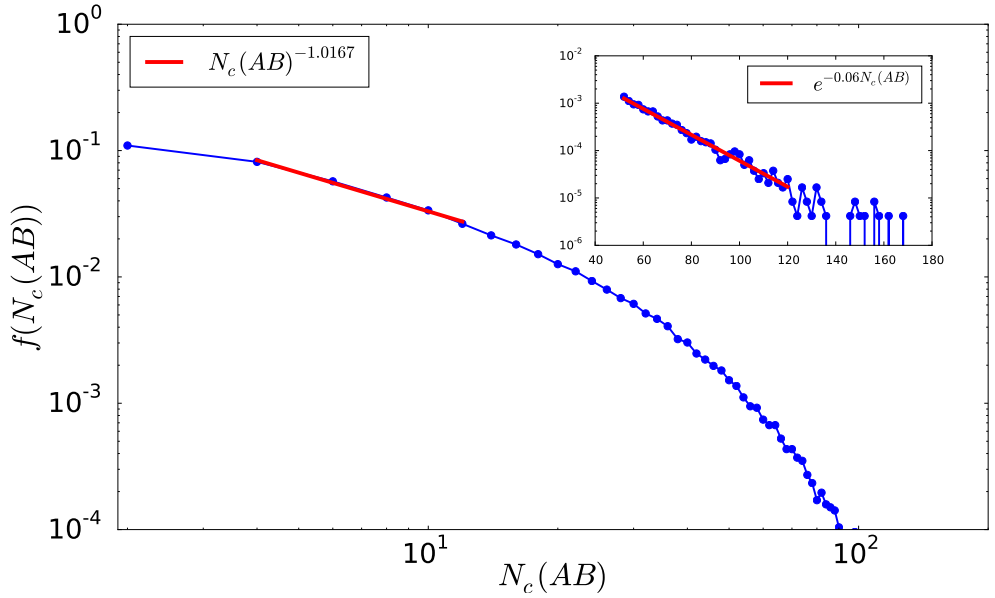


Figure 4.4: The log-log plot of the distribution function of the number of contact between two halves of a linear chain $f(N_c(AB))$. The length of the chain is $N = 10000$. The cut-off distance for contact is set to 2.0 in order to get a better approximation. For small $N_c(AB)$ the function has a power law decay with exponent around -1.02. At the tail it instead has an exponential decay. The change of cut-off distance for contact only affects the value of the exponent.

Secondly we analyze the dependence of the number of contacts on the segments we are looking at. The segment starts at the mid-monomer and extends to both sides by a contour length s_m (Figure 4.1). Therefore the segment is composed by two parts, denoted as A and B, and has the length $2s_m$. The total number of contacts within this segment N_c has the following relation: $N_c = N_c(AA)+N_c(BB)+N_c(AB)$. In figure 4.5 we plot the ratio of the number of contacts N_c to s_m versus s_m for a linear chain in free space and in cubic confinement. In the cubic boxes, the volume fractions of the chain are 0.16, 0.26, and 0.96 respectively. The length of the chain is $N = 160$. In free space, N_c grows linearly $N_c \sim s_m$ after the length of the segment exceeds some value $2s_m^*$. In contrast, when the cubic constraints is imposed, N_c grows quadratically $N_c \sim s_m^2$ when $s_m > s_m^*$. In other words, if we lengthen the

segment by $2\Delta s_m$, then for chains in free space

$$\Delta N_c \sim \Delta s_m. \tag{4.1}$$

The number of newly-formed contacts only depends on the new fragment. For chains in cubic confinement,

$$\Delta N_c \sim s_m \Delta s_m. \tag{4.2}$$

The number of newly-formed contacts also depends on the old segment, which means that the new fragment penetrates the region that the segment occupies under confinement.

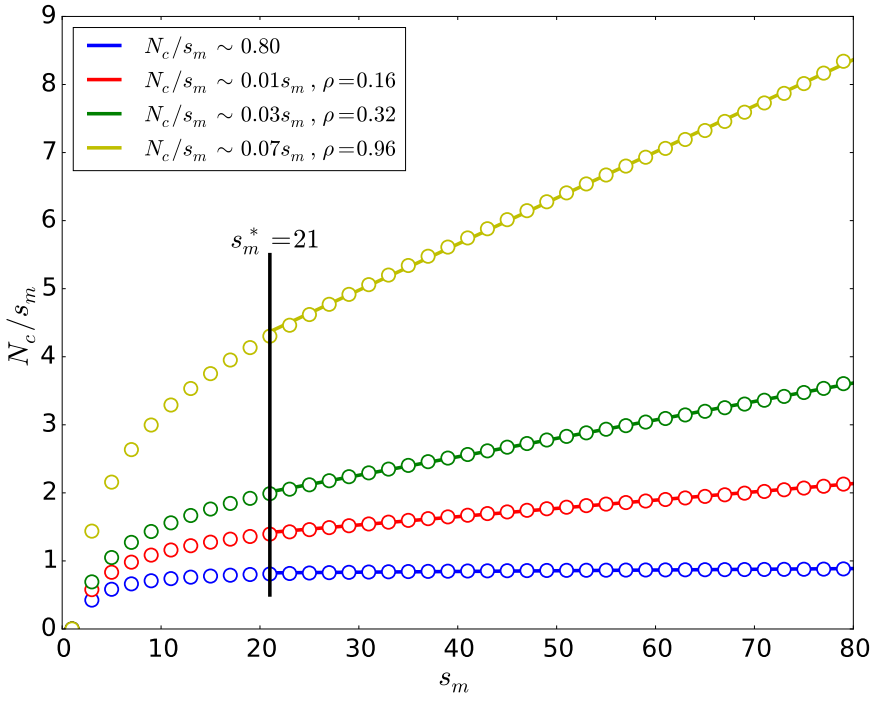


Figure 4.5: Shown is the ratio of number of contacts N_c to the contour length from the mid-monomer s_m versus s_m. The chain with $N = 160$ monomers is placed in free space (blue) and cubic boxes of different sizes. The solid lines are the results of a least square fit to the data from s_m^* onward. In free space the contact number grows linearly, while in constraint space it grows quadratically.

Next we consider the percentage of the inter-contacts of the two halves: $p_{AB}(s_m) = N_c(AB)/N_c$ (Figure 4.6). For a chain in free space, p_{AB} decreases monotonically with s_m, which means that as the chain becomes longer, the number of contacts within

a halve of the chain increases more rapidly. The reason is that the excluded volume effects are more pronounced, thus the two halves drift apart. In a cubic box, for rather small s_m, the percentage of inter-contacts is the same as the free chain case, i.e. the short segment does not feel the constraint. As the segment becomes longer, it is compressed by the remaining parts of the chain, inducing more contacts between two halves of the segment. In the limit when the segment occupies the whole space (Hamiltonian path), the percentage of $N_c(AB)$ is $p_{AB}^{HP} \approx 0.33$. This means that in this case we have $p_{AB}^{HP} = p_{AA}^{HP} = p_{BB}^{HP}$, which is reasonable since every site is occupied and each contact has the same probability of counting into $N_c(AA), N_c(BB)$, and $N_c(AB)$.

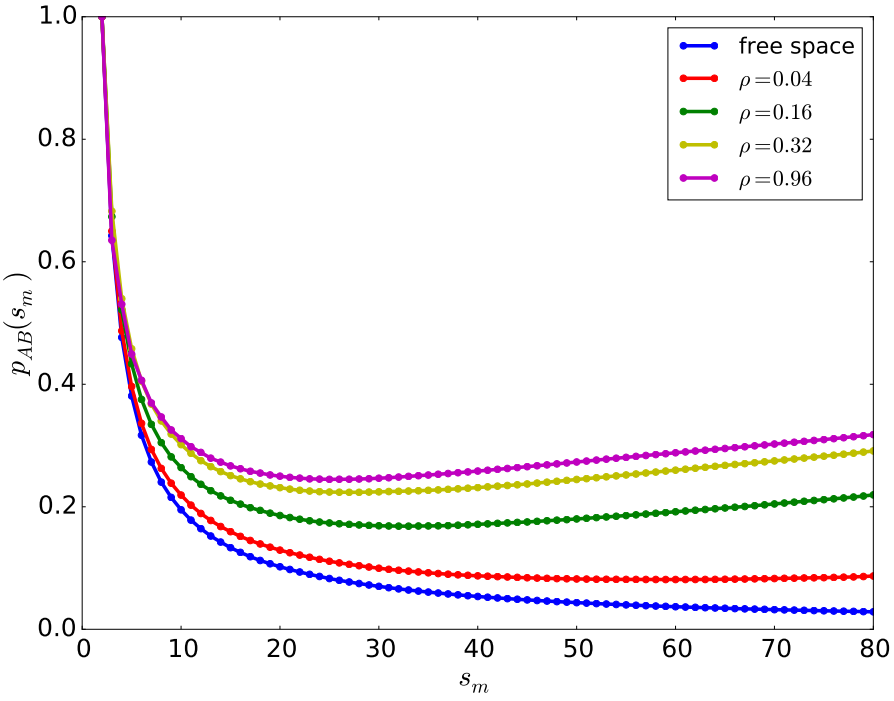


Figure 4.6: Shown is the proportion of inter-contacts of the two halves $N_c(AB)$ to the total contacts N_c versus s_m for chain in free space and in confinement. In free space, the proportion is always decreasing, while it sees an increase after some contour distance s_m if the box is sufficiently small. ($\rho > 0.04$).

For a linear chain in free space, the asymptotic behavior of $p_{AB}(s_m)$ for large s_m is plotted in Figure 4.7. The open circles are the data we calculate. By linearly fitting the region $s_m > 2000$ in the log-log plot, we get an exponent -0.97 of the power law decay: $p_{AB}(s_m) \sim s_m^{-0.97}$. Recall that from figure 4.5 we get for a free chain the total

number of contacts grows linearly with s_m. Actually this linear growth has already been discussed in [47], where they gave the asymptotic scaling $N_c \sim aN + bN^{\phi} + c$ (N is the length of the self-avoiding walk). The last two terms are rather small compared with the leading term aN. Approximately we have $N_c \sim N = 2s_m$. Thus, the number of inter-chain contacts is $N_c(AB) = p_{AB}N_c \sim s_m^{0.03}$. We surmise that for sufficiently large s_m or N, we should have $p_{AB}(s_m) \sim s_m^{-1}$ and $N_c(AB) \sim s_m^0$, which is consistent with the conclusion that the number of contacts between two halves is strictly finite [48]. Much longer chains should be analyzed to fill the gap.

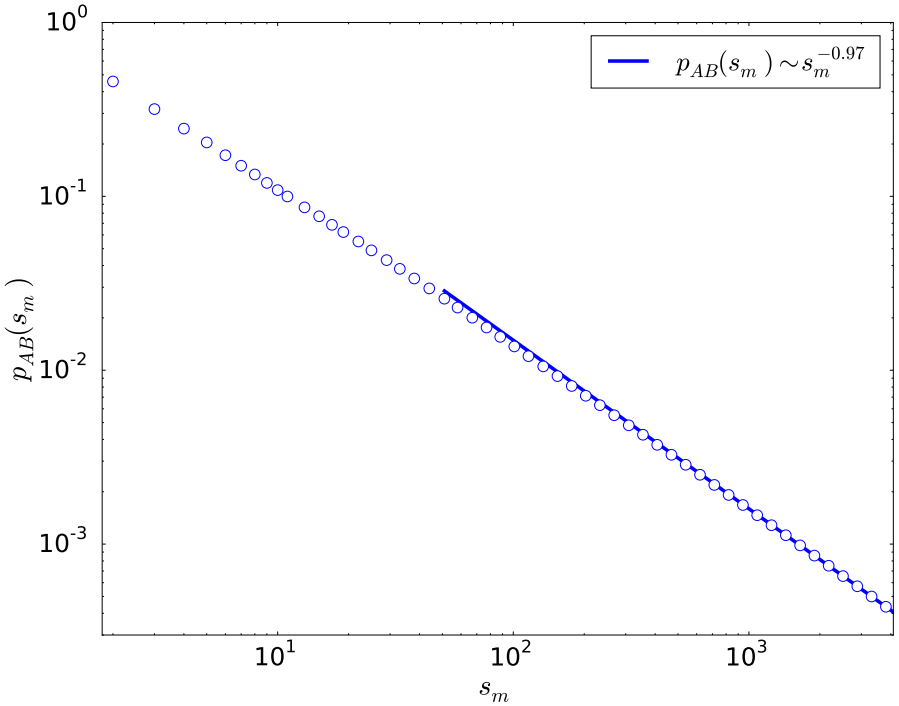


Figure 4.7: The power law decay of $p_{AB}(s_m)$ in the large s_m region. The open circles are the data we calculate. The blue line indicates the power law decay with exponent -0.97. We get this exponent by the linear fitting of the region $s_m > 2000$ in the log-log plot.

Finally we look at the overlap of the two halves by the projection of monomers on the line connecting the centers of mass of two halves. In the figure 4.8 we show the results of chains in free space (blue), cubic box (red) and rectangle box with aspect ration 4:1:1 (green) and 8:1:1 (yellow). The distance d is given in units of the center-of-mass distance. The gray line indicates the positions of the centers of mass. In free space and the longer rectangle box, the monomers of one halve that penetrates into the other halve have a very small proportion, which means that in

the two cases, the two halves of the chain are nearly separated. For chains in cubic box and shorter rectangle box, the two halves penetrate into each other, i.e. they are more mixed.

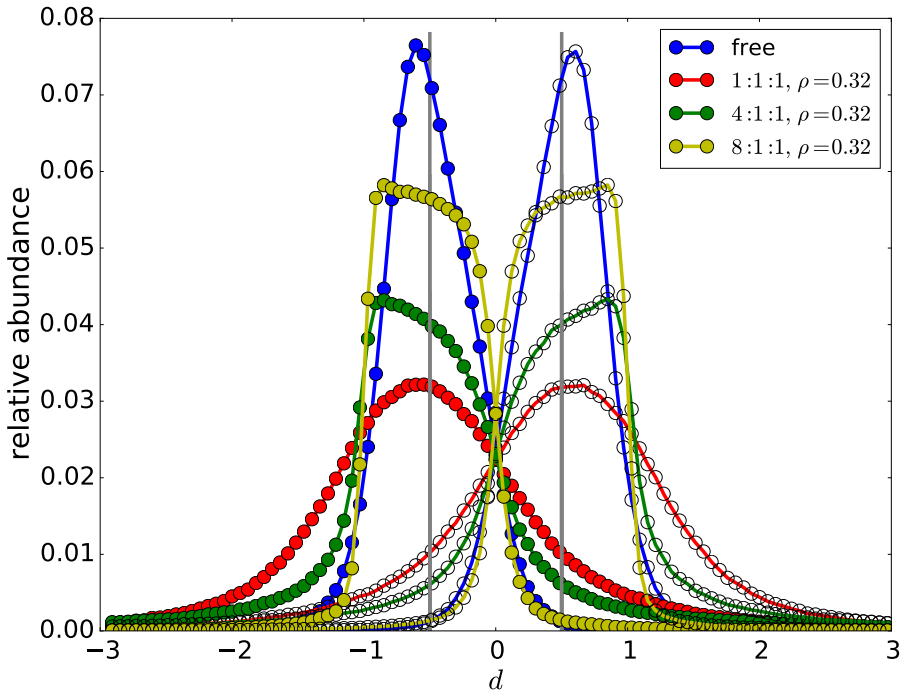


Figure 4.8: Distributions of monomers projected onto the line connecting the centers of mass of two halves of the linear chain. The distance d is given in units of the center-of-mass distance. The origin corresponds to the middle of the centers of mass, and the gray vertical lines represent the positions of the centers of mass. The distributions for chain in free space (blue), cubic box (red), rectangle box with aspect ration 4:1:1 (green) and 8:1:1 (yellow) are shown.

4.3.2 The Segregation of Two chains

The segregation process of two polymer chains could be influenced by many factors. One important and well-studied factor is the geometrical constraint, where both its shape and size matter. Also, the intra-and inter-actions of the chain systems play a crucial role in this process, for example, the bending rigidity and the attraction between the chains. The coaction of these factors is quantified by the free energy $F(\mathbf{x}) = E(\mathbf{x}) - TS(\mathbf{x})$, where $\mathbf{x}$ denotes the state of the system. By minimizing this free energy, one can obtain the equilibrium states. The relaxation to these equilib-

rium states from an initial state that have a high free energy is usually characterized by the relaxation time τ. On the free energy landscape, the relaxation process can be described by pathways. Although the pathways are influenced by thermal fluctuation, where this process starts also determines how the pathways go. The main results of the section show whether and how different initial states carry weight in the segregation of two chains.

We use two quantities to distinguish different initial conformations of the two chains: the number of their contacts N_c and the winding number W. The distance of centers of mass is often used to describe the states of the two chains, yet it only gives the overall information about their relative position. The number of contacts, on the other side, provides the local proximity of the two chains. In the presence of monomer friction, these contacts would impede the segregation process. We further want to study the effect of intertwining of the chains on the segregation, where the winding number W is used to measure this intertwining.

We only consider the excluded volume interaction, that is, there is no monomer friction in the system. The segregation is driven by the entropy of the system. Figure 4.9 shows the time dependence of the distance of mass centers $r_{cm}(t)$ for the case $L = 100$. The segregation starts from 7 different initial conformations: separated by the plane $z = 4$ (Figure 4.2(a)), intertwined (Figure 4.2(b)), and mid-monomers fixed but with different numbers of contacts $N_c(0)$ (Figure 4.2(c)). For each initial conformations, we perform 100 independent runs of Monte Carlo simulation. The total time required for segregation is almost same for these 7 initial conformations: $\tau \approx 3.5 \times 10^7$mcs. In the case $L = 200$ the results are similar, with $\tau \approx 8.0 \times 10^6$mcs. This means that with the entropy as the only driven force of segregation, the contacts and the intertwinement have little effects on the entire segregation process.

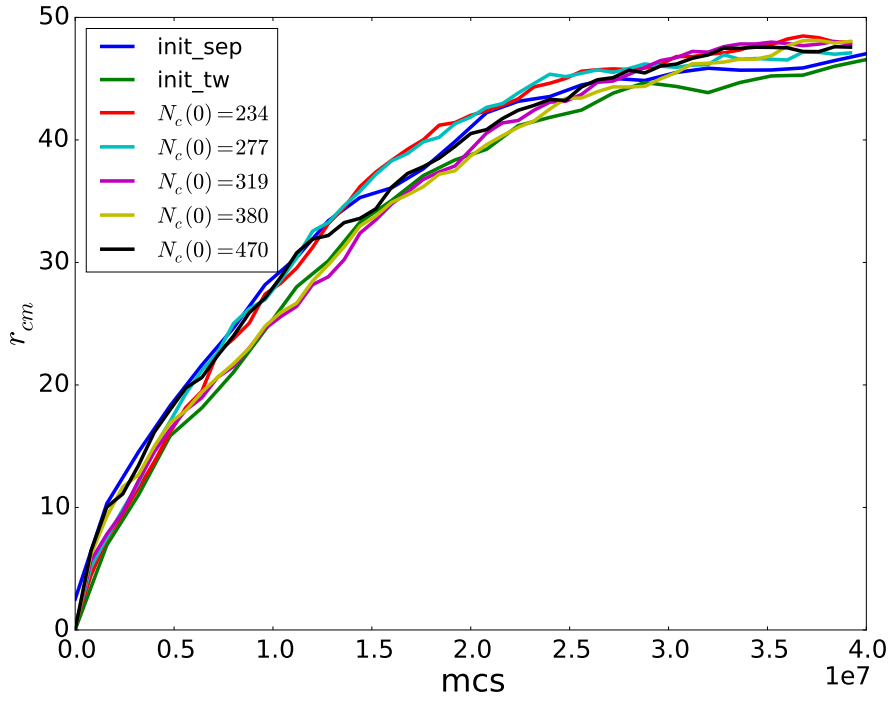


Figure 4.9: The time dependence of the distance of mass centers $r_{cm}(t)$ for the case $L = 100$. The segregation starts from 7 different initial conformations as indicated in the legend. The results are averaged over 100 independent runs.

The number of contacts of the two chains $N_c(t)$ is shown in Figure 4.10 ($L = 100$) and Figure 4.11 ($L = 200$). Although the contacts number for these initial conformation differs much at the beginning, they come closer shortly. The time scale of this τ_1 is much smaller than the scale of the overall segregation process, with $\tau_1 \approx 2 \times 10^6$mcs for $L = 100$ and $\tau_1 \approx 2 \times 10^5$mcs for $L = 200$. In particular, for two chains that are initially intertwined (the green line), it takes longer time for $N_c(t)$ to join the other curves. The time is $\tau_2 \approx 4.6 \times 10^6$mcs for $L = 100$, and $\tau_2 \approx 1 \times 10^6$mcs for $L = 200$. This implies that the intertwinement could indeed slow down the separation of two chains. The winding number of them $W(t)$ has similar behavior (Figure 4.12 and 4.13)

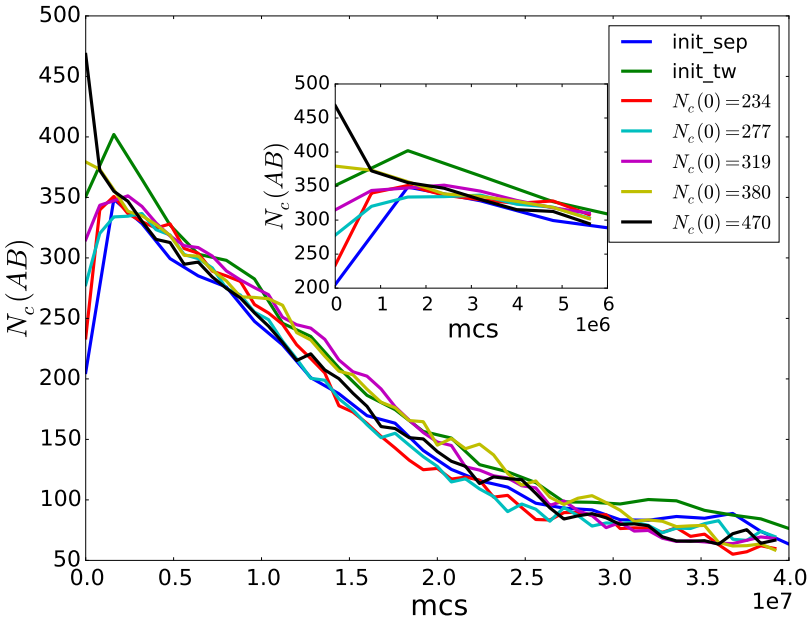


Figure 4.10: The number of contacts between the two chains $N_c(t)$ in the segregation process for different initial conformations in the case $L = 100$. The inset shows a zoom for the beginning phase of the segregation.

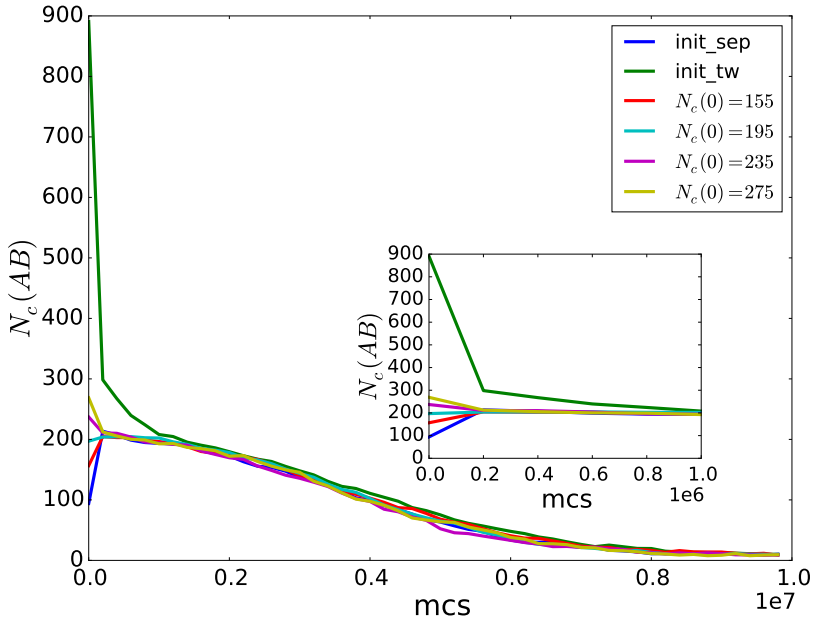


Figure 4.11: The number of contacts between the two chains $N_c(t)$ in the segregation process for different initial conformations in the case $L = 200$. The inset shows a zoom for the beginning phase of the segregation.

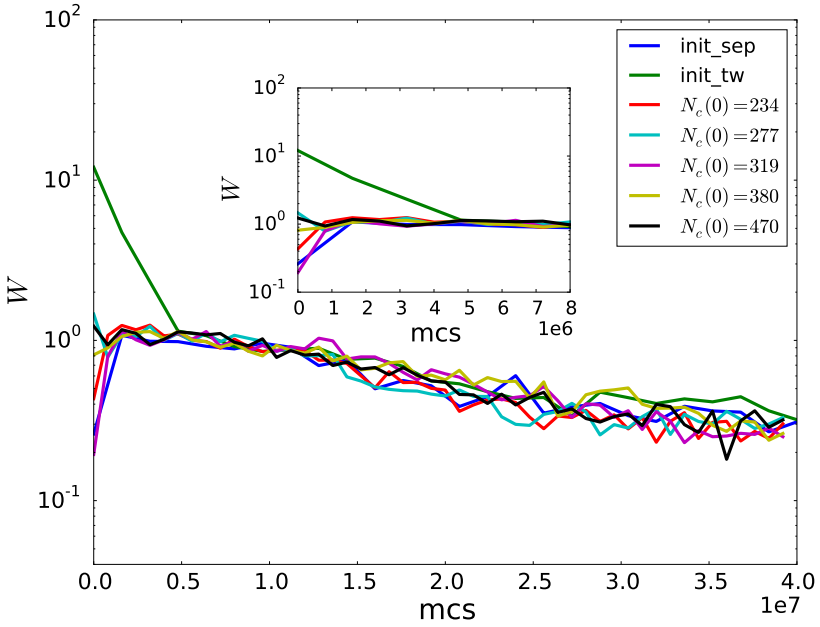


Figure 4.12: The semi-log plot of the winding number of the two chains $W(t)$ in the segregation process for different initial conformations the case $L = 100$. The inset shows a zoom for the beginning phase of the segregation.

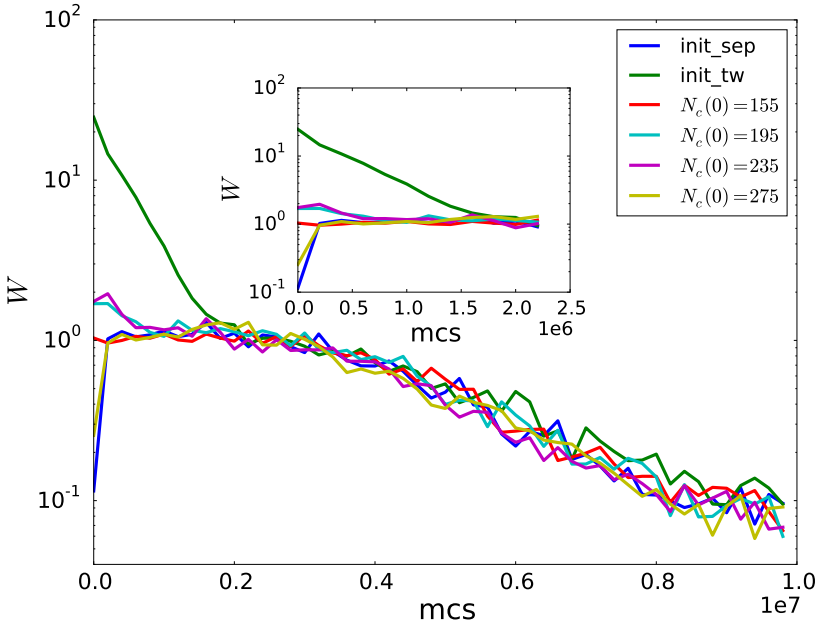


Figure 4.13: The semi-log plot of the winding number of the two chains $W(t)$ in the segregation process for different initial conformations the case $L = 200$. The inset shows a zoom for the beginning phase of the segregation.

Experimentally the neutron scattering is often used to analyze the structural disorder of alloys., where the structure factor is calculated to identify the organization of metals [201,202]. In a binary alloy model, the dynamics of unmixing process of two kinds of atoms A and B is associated with the following dynamic structure factor

$$S(\mathbf{q},t) = \frac{1}{L^2}\langle|\sum_i \exp[i\mathbf{q}\cdot\mathbf{r}_i]\{c_i^A(t) - c_i^B(t) - [\langle c_i^A\rangle - \langle c_i^B\rangle]\}|^2\rangle, \tag{4.3}$$

where $c_i^A(t), c_i^B(t)$ denote the occupancy of the lattice site i at time t, e.g., if it is occupied by an A-atom, $c_i^A = 1$. L is the size of the lattice. During the unmixing process of A-atom and B-atom, the peak of the $S(\mathbf{q},t)$ is growing [202]. Similarly, for two chains (A and B) with the same length N, the dynamic structure factor is rewritten as follows:

$$\begin{aligned} S(\mathbf{q},t) &= \frac{1}{N^2}\left\langle\left|\sum_m^N \exp[i\mathbf{q}\cdot\mathbf{r}_m^A(t)] - \sum_n^N \exp[i\mathbf{q}\cdot\mathbf{r}_n^B(t)]\right|^2\right\rangle \\ &= S^A(\mathbf{q},t) + S^B(\mathbf{q},t) - S^{AB}(\mathbf{q},t) \end{aligned} \tag{4.4}$$

where $\mathbf{r}_m^A(t), \mathbf{r}_n^B(t)$ are the positions of the m-th and n-th monomers of chain A and B respectively. The first two terms are the single-chain structure factors for chain

A and B, while the last term is the interchain structure factor:

$$S^{AB}(\mathbf{q},t) = \frac{1}{N^2}\sum_{m=1}^{N}\sum_{n=1}^{N}(\exp[i\mathbf{q}\cdot(\mathbf{r}_m^A(t) - \mathbf{r}_n^B(t))] + \exp[i\mathbf{q}\cdot(\mathbf{r}_n^B(t)) - \mathbf{r}_m^A(t)]) \quad (4.5)$$

Figure 4.14 shows the dynamic structure factor of the two chains at different Monte Carlo steps during the segregation process. At the beginning, the two chains are intertwined, and $S(\mathbf{q},0)$ has several small peaks. After that $S(\mathbf{q},t)$ only has a major peak and the peak is growing as the segregation goes on. Thus the growth of the peak can also reflect this segregation process (Figure 4.15).

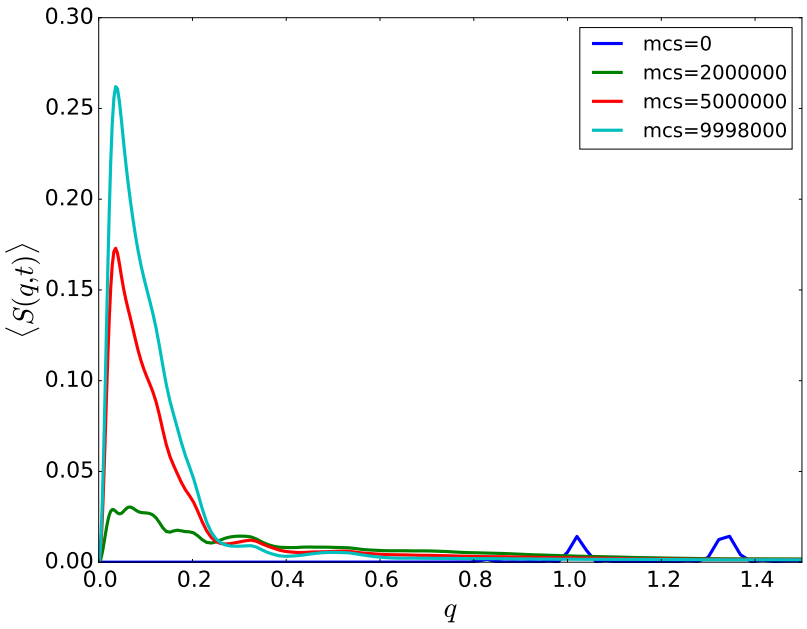


Figure 4.14: The dynamic structure factor $S(\mathbf{q},t)$ of the two chains at different Monte Carlo steps during the segregation process.

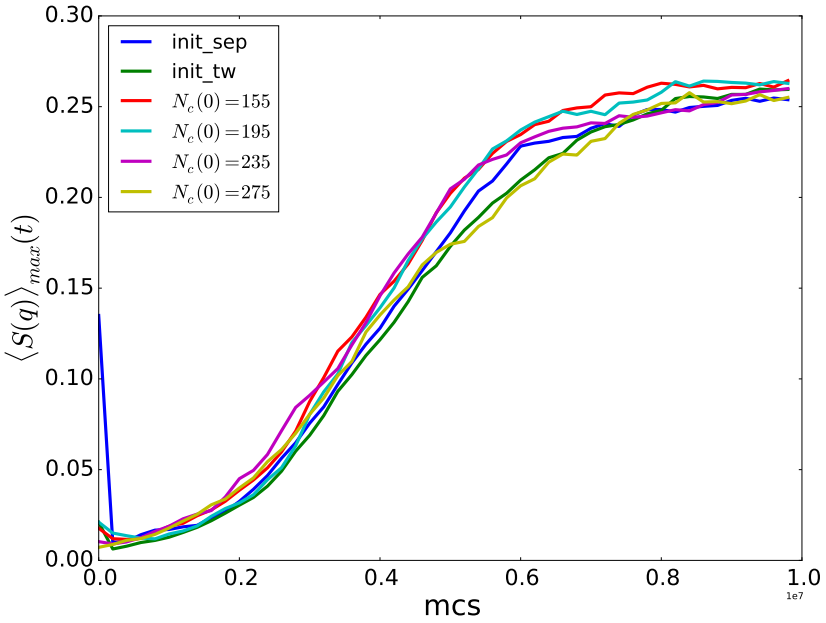


Figure 4.15: The growth of peak value of the dynamic structure factor $S(\mathbf{q}, t)$ during the segregation process for two chains confined in the $L = 200$ rectangular box.

4.4 Conclusion

In this chapter we focused on two main topics: the self-contact of two halves of a single polymer chain and the segregation of two chains in a rectangular box. First we studied the distribution function of the number of contacts $f(N_c(AB))$ of two halves for a single chain with and without confinement. For a chain in free space, the distribution function exhibits a power law decay in small $N_c(AB)$ region, and a exponential decay in large $N_c(AB)$ region. For a single chain confined in a cubic box, when the size of the box is small enough, the distribution function first increases with $N_c(AB)$, then decreases rapidly. Next we studied the dependence of total number of contacts N_c on the size of the segment we are looking at. For free chain, N_c grows linearly with the length of the segment s_m, while for confined chain, N_c grows quadratically. Further, we studied the percentage of the inter-contacts of the two halves of the segment $p_{AB}(s_m)$, finding that it has a power law decay for free chains, with an exponent slightly larger than -1. We assume that the discrepancy is caused by statistical error and finite size effects. This power law decay demonstrates that the number of contacts between two halves of a self-avoiding chain is finite as the length of the chain goes to infinity, which indicates that the two halves are

nearly separated. For confined chain, in small s_m region, $p_{AB}(s_m)$ overlaps with the free chain case, meaning that in small length scale, the segment does not feel the constraints. As s_m becomes larger, the confinement induces more contacts between two halves of the chain. Specifically, for a Hamiltonian path, the total number of inter-contacts is equal to the intra-contact, i.e. $p_{AB}^{HP} \approx 0.33$. The separation of two halves of a self-avoiding chain is also revealed by the distribution of monomers projected onto the line connecting the mass centers of the two halves.

The segregation of two chains in a rectangular box starting from different initial conformations is also investigated. We create three kinds of initial conformations: separated, intertwined, mid-monomers fixed but with different numbers of inter-contacts. By looking at the evolution of mass centers distance, number of inter-contacts, winding number, and the peak value of dynamic structure factor, we find that for the system only with excluded volume interaction, these different initial conformations have similar segregation time because this process is mainly governed by a much slower diffusion of chains. However, the intertwinement indeed would impede the segregation process at the beginning.

In the presence of other interaction such as the attraction between nearest neighbors, the dynamics of this segregation process will have more interesting features. It is found that in free space, the segregation time is proportional to the number of initial contacts between the two chains [197]. We will implement this kind of interaction and study this dynamics for two confined chains in future works.

Chapter 5

Fractality and Topology in the Self-Avoiding Walks

Chapter Summary

In this chapter we deal with some geometrical and topological features of the contacts inside a chain. Specifically, we calculate the fractal dimension and growth rates of the Betti numbers of the system. In the simplest case, we consider the chain as a self-avoiding walk without any other interaction. The set of contacts is a subset of the original walk, and we found that this subset may have a slight multifractal property. In addition, the topological exponents are also different from the self-avoiding walk. Further, each contact gives rise to the formation of a loop. To elucidate how these loops influence the structure of the self-avoiding walk, we delete the loops in a similar way to the loop-erased random walk, thus producing a new walk: loop-deleted self-avoiding walk (LDSAW). The critical exponent of LDSAW is approximated by studying the scaling behavior of mean end-to-end distance, and the dependence of the mean length of LDSAW on the length of the original self-avoiding walk. Afterward, the fractal dimension and growth rates of Betti numbers of this LDSAW are calculated. The same calculations are also performed on the projection and random subsets of self-avoiding walks.

5.1 Introduction

The self-avoiding walk (SAW) has been investigated both theoretically and numerically due to its applicability in the coarse-grained modeling of various polymer systems. The fractality of SAW is defined and measured by Havlin and Ben-Avraham [85, 86], with the fractal dimension $d_F = 1/\nu$ (ν is the critical exponent). The contacts of polymers, especially biomolecules, play an important role in corresponding systems. For example, certain biological functions are accomplished via the contact and interaction of different parts of molecules [22]. Experimentally the Hi-C contact maps are used to study the 3D organization of genome. The scaling behavior of contact probability of SAW as a function of contour length was studied in [135, 154, 155]. The average number of contacts $\langle m \rangle$ of SAW is found to have an asymptotic behavior $\langle m \rangle \sim a_\infty N$, where N is the length of the SAW [47]. Baiesi [48] studied the contacts between two halves of a SAW and found that they are strictly finite in number. The dependence of the number of contacts on the radius of gyration of a SAW also has a simple scaling law [203]. In this chapter, we would rather inspect the spatial distribution of these contacts from geometrical and topological aspects of view. Since every contact occurs when two monomers become the nearest neighbors, the set of these contact points is actually a subset of the original SAW. Then the question arises that what the differences between them are in terms of the fractality and topology.

Each contact of SAW indicates a formation of loop. Lawler [204, 205] introduced a different kind of self-avoiding walk by erasing the loops in a random walk, namely LERW (Loop-Erased Random Walk). Due to its close relation to the uniform spanning tree and the Laplacian random walk, the LERW has received lots of attention [206–208]. The LERW is self-avoiding but belongs to a different universality class from the normal SAW. It has a slightly larger critical exponent $\nu_{_{LERW}} \approx 0.616$ in 3D which was intensively estimated in [209–212]. Instead, one may consider what if we erase the loops of a self-avoiding walk. To distinguish from LERW, we denote this new kind of walk as loop-deleted self-avoiding walk (LDSAW). Clearly, the configuration space of the LDSAW is a subspace of that of the SAW, and a LDSAW is one subset of the original SAW. Whether this LDSAW has a different critical exponent is a question we will study.

Topologically, people have examined the knots [149, 184, 213], writhe [148, 150] in self-avoiding walks since these are very important features in the packing of DNA [214,215]. When calculating the writhe of a polymer, the 3D object is projected onto planes from all directions, then the writhe of each knot diagram is calculated and averaged. This means that the projections share some topological features of the 3D object. So how are the points projected on the plane distributed? What is the fractal dimension?

Recently, owing to the powerful ability of persistent homology to reveal the underlying topological features of data at different resolutions [216, 217], the application of persistent homology is quite prevailing in many areas such as machine learning [59, 218], disease examination [219, 220], brain network [221] and so forth. Specifically, the fractal property of data is linked to the persistent homology as they both measure the data at different scales. Robins proposed the growth rates in the number of connected components or holes as the resolution goes to zero, and explained the relation between these growth rates and fractal dimension in her PhD thesis [222]. Afterwards, Macpherson and Schweinhart [223] defined a persistent homology dimension based on the birth and death of holes to measure the complexity of the data. Another method to define the fractal dimension was introduced in [224]. Further, an estimator of fractal dimension in terms of the minimum spanning trees and higher dimensional persistent homology was proposed in [225–227]. Despite all these different definitions, the fractal dimension established on the persistent homology gives a description of the complexity of the system under investigation.

In this chapter we would first use geometrical methods to study the fractal properties of contacts of the SAW, loop-deleted self-avoiding walk (LDSAW) and projections of SAW. For comparison, a random subset of SAW would also be studied. Following are the analysis of these systems on the topological aspect. This chapter is organized as follows. In section 5.2 we introduce some concepts of fractal and several algorithms for calculating the fractal properties. Then the basic idea of persistent homology will be presented, together with the fractal dimensions defined on this. The results of our analysis for different systems are given in section 5.3. Finally in the last section we conclude our work.

5.2 Concepts and Methods

5.2.1 Definitions of Fractal Dimension

Fractals have an important property called self-similarity which means that they have similar patterns at different scales. This is the foundation for various definitions of fractal dimension. In general, self-similarity is only well founded for artificial fractals, or deterministic fractals, such as the Cantor set and Sierpinski triangle, while in nature, as a consequence of the ubiquitous randomness, objects of interest would rather repeat a pattern stochastically at different scales, i.e., they are statistical self-similar, therefore they are random fractals. Typical examples of random fractals include the coastline of England, Brownian motion and so on. The self-avoiding walk is also a random fractal.

The definition of box counting dimension is straightforward. If S is the fractal

under study,

$$D_{\text{box}}(S) = \lim_{L\to 0} \frac{\log N(L)}{\log(1/L)}, \tag{5.1}$$

where $N(L)$ is the number of boxes of side length L needed to cover S. An equivalent definition is given by

$$D_{\text{MB}}(S) = n - \lim_{L\to 0} \frac{\log \text{vol}(S_L)}{\log L}, \tag{5.2}$$

where $\text{vol}(S_L)$ is the influence volume of S by dilating S by a sphere of radius L [228, 229]. Another important definition of fractal dimension is the mass dimension, which is defined as the exponent of power law relation between the mass $M(L)$ within a ball of radius L centered at a point:

$$D_{\text{mass}} = \lim_{L\to 0} \frac{\log M(L)}{\log L}. \tag{5.3}$$

The mass dimension is quite intuitive if we consider that for a $n(= 1, 2, 3)$ dimensional object, $M(L)$ is proportional to L^n [230, 231]. The difference in the sequence of taking the limit and averaging $M(L)$ over every point gives the information dimension and the correlation dimension [232]. The box-counting dimension together with the information dimension and correlation dimension are special case of the generalized dimension which is:

$$D_q = \lim_{L\to 0} \frac{\frac{1}{1-q} \log \left(\sum_i p_i^q\right)}{\log(1/L)}, \tag{5.4}$$

where p_i is the percentage of points in box i, q is the moment. For a monofractal, the value of D_q will not change with q, while if an object is a combination of different fractals, then D_q exhibits a decreasing with q, which indicates the structure of this multifractal. Specifically, D_∞ and $D_{-\infty}$ correspond to the most dense and least dense areas.

Instead of counting the mass within a box of size L, Termonia and Alexandrowicz [233] defined the fractal dimension by finding the scaling behavior of the average radius $\langle R_m \rangle$ of m nearest-neighbor points with m:

$$\langle R_m \rangle \propto m^{1/D}. \tag{5.5}$$

Afterwards, it is proposed in [234] that the Equation (5.5) can be extended to the moment of order τ:

$$\langle R_m^\tau \rangle \propto m^{\tau/D(\tau)}, \tag{5.6}$$

where $\tau = (q-1)D_q, D(\tau) = D_q$. The algorithms based on Equation (5.1) and (5.4) are fixed-size algorithms, while algorithms based on Equation (5.5) and (5.6) are

fixed-mass algorithms [232]. The latter outperforms the fix-sized algorithms in some cases, especially when $q < 1$ [235, 236].

The critical exponent ν of self-avoiding walk or other kind of walk is always a key subject in the field of polymers, which can be estimated by the asymptotic behavior of the mean-square end to end distance and radius of gyration [74, 76, 101]. The corresponding fractal dimension of the walk is $D = 1/\nu$. In this chapter, since we intend to investigate the structure of contacts of a self-avoiding walk, which is not a typical walk, but rather a point-cloud, we want to use the typical algorithms for estimating the fractal dimension. Both the self-avoiding walk and its contacts are random fractals, therefore it's necessary to average over some point-clouds in order to get a better estimation of the fractal dimension. To check the performance of the algorithms, the self-avoiding walk is also analyzed. The algorithms we have assessed are: (1) a fast box-counting method implemented in [237]; (2) the Bouligand-Minkowski method [229, 238], which is an variant of the box-counting method [228]; (3) the Barycentric fixed-mass method [239]. The last two methods are capable of carrying out the multifractal analysis. The Bouligand-Minkowski method is based on the Equation (5.2) and is modified to quantify the multifractality using the derivative of the influence volume $\mathrm{vol}(S_L)$ in [238, 240]. The Barycentric fixed-mass method is based on the Equation (5.6) and incorporates the barycentric pivot point selection and nonoverlapping coverage criteria to reduce the edge effects.

5.2.2 Persistent Homology

The scheme of persistent homology is to discover how the topological features of a shape, such as the connected components, holes and voids, would change at different scales. In the first place, the point-cloud that describes the shape needs to be represented by a simplicial complex. Then a filtration is started by growing the balls centered at every point (Figure 5.1). The radius of the balls, or the resolution is denoted by ϵ. During this growing process, the number of the connected components is always nonincreasing, while the holes and voids could appear and disappear when increasing ϵ. The corresponding values of ϵ are usually denoted as ϵ_{birth} and ϵ_{death}. The intervals $(\epsilon_{birth}, \epsilon_{death})$ represent the underlying topological property of the given shape, and can be visualized as the persistent barcodes and the persistent diagram. The long intervals, which means that the holes or voids persist through a large range of resolution, indicate important topological features, and small intervals are considered as noise. For example, in [61] the barcodes are used to detect the existence of alpha helix and beta sheet in proteins. In addition, various definitions of distance between two persistent diagrams provide different measures for the similarity of two shapes, e.g. two γH2AX clusters [241]. Detailed mathematical description for the persistent homology can be found in [217, 242].

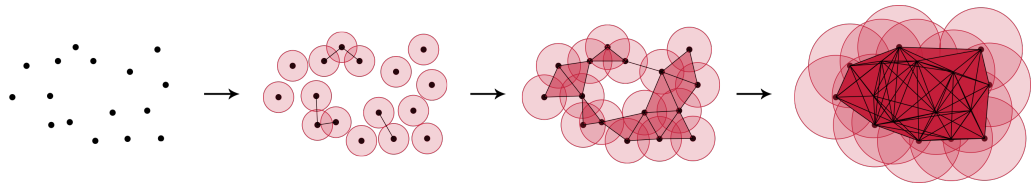


Figure 5.1: The filtration process for 2D points. The image is from Christian Bock [243].

The connection of persistent homology with fractality is proposed by Robins [222], where the growth rates of Betti numbers are studied. The numbers of connected components, holes and voids of a shape are the 0th, 1st, 2nd persistent Betti number $\beta_i (i = 0, 1, 2)$. As mentioned above, these numbers would change during the filtration process, that means β_i depends on ϵ. For a shape that is fractal, suppose as $\epsilon \to 0$, $\beta_i(\epsilon) \to \infty$, an exponent γ_i could be defined if the asymptotic behavior is a power law:

$$\gamma_i = \lim_{\epsilon \to 0} \frac{\log \beta_i(\epsilon)}{\log(1/\epsilon)}. \tag{5.7}$$

If the limit does not exist, alternatively limsup or liminf is used. These three topological growth rates could also serve as the characteristics of the fractal in discussion. By applying these to some well-defined fractals, Robins showed that the growth rates can distinguish sets with the same Hausdroff dimension but different homology. However, the relation between them and fractal dimension remains an open question. Several other definitions of persistent homology fractal dimension are studied in [223, 225–227].

Since a self-avoiding walk with equal bond length is already a minimal spanning tree, β_0 is N or 1. The loops in the self-avoiding walk are actually one kind of holes in the first dimension of persistent homology. In section 5.3, we will show the results of γ_1 and γ_2 for the self-avoiding walk, its contact and other related systems.

5.3 Results

5.3.1 The Fractal Dimension of the 3D Self-Avoiding Walk

The well known critical exponent for 3D SAW is $\nu \approx 0.588..$, which is the inverse of the fractal dimension $D = 1/\nu \approx 1.7$ [85]. The most common way to calculate the fractal property numerically is to use the asymptotic behavior of mean-square end to end distance or radius of gyration: $\langle R_e^2 \rangle \sim N^{2\nu}, \langle R_g^2 \rangle \sim N^{2\nu}$. Since our task is to get the fractal information from one or several point-clouds, such as the contacts of SAW, the performance of the algorithms mentioned in section 5.2.1 on self-avoiding walk could serve as a criterion for the accuracy of different methods in our systems.

The results of the fast box-counting algorithm on self-avoiding walks of length $N = 10^4, 10^5, 3 \times 10^5, 10^6, 2 \times 10^6$ are shown in Figure 5.2. Due to the intrinsic randomness of the system, single point-clould of the walk does not give a reliable estimation of the fractal dimension. Therefore, the box number $n(L)$ is averaged over a lot of conformations (point-clouds) for each N. The numbers of conformations are 15 000, 200, 200, 100, 50, respectively. By fitting the linear region of the log-log plot of $\langle n(L) \rangle$ versus L, the fractal dimension D is approximated. The values are 1.5155, 1.6319, 1.6442, 1.6673, 1.6782, respectively. Due to the finite resolution and edge effects, the box-counting method is quite sensitive to the system size and noise, therefore the estimated fractal dimension is not reliable for the SAW.

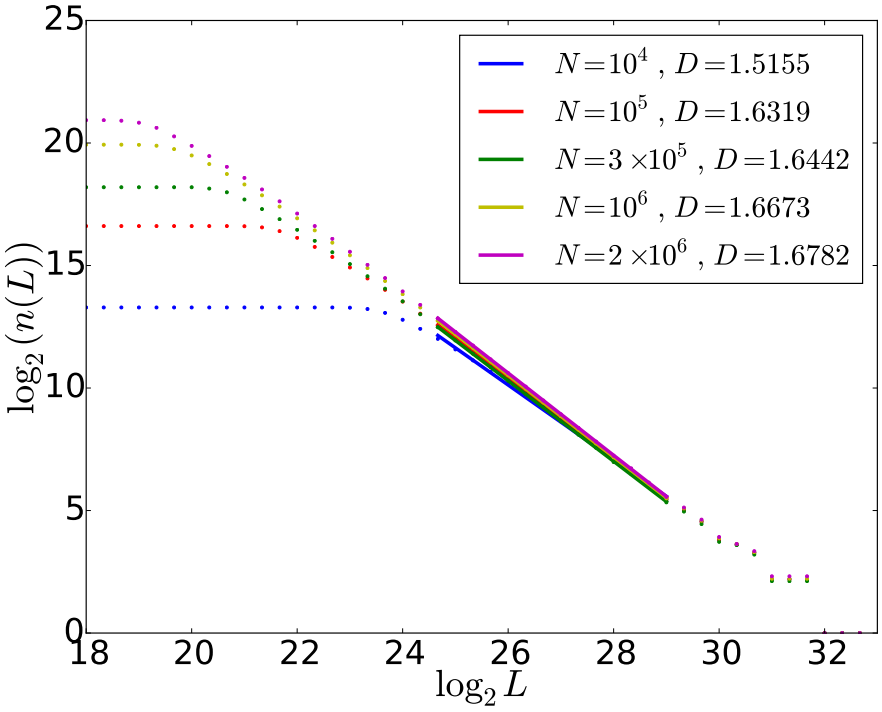


Figure 5.2: The log-log plot of the box number $n(L)$ versus the box size L. The fractal dimension D of SAW with different lengths are estimated by fitting the linear region. The fast box-counting algorithm is implemented by Kruger [237].

The Bouligand-Minkowski method is applied to the same systems of self-avoiding walks, with results shown in Figure 5.3. The estimated fractal dimensions for different lengths of SAW are 1.6187, 1.6769, 1.6754, 1.6901, 1.7014, which are better than the fast box-counting method, but still sensitive to the finite size effects. The multifractal analysis of this method is not robust in such a random system, and

therefore does not give the result that SAW is statistically monofractal.

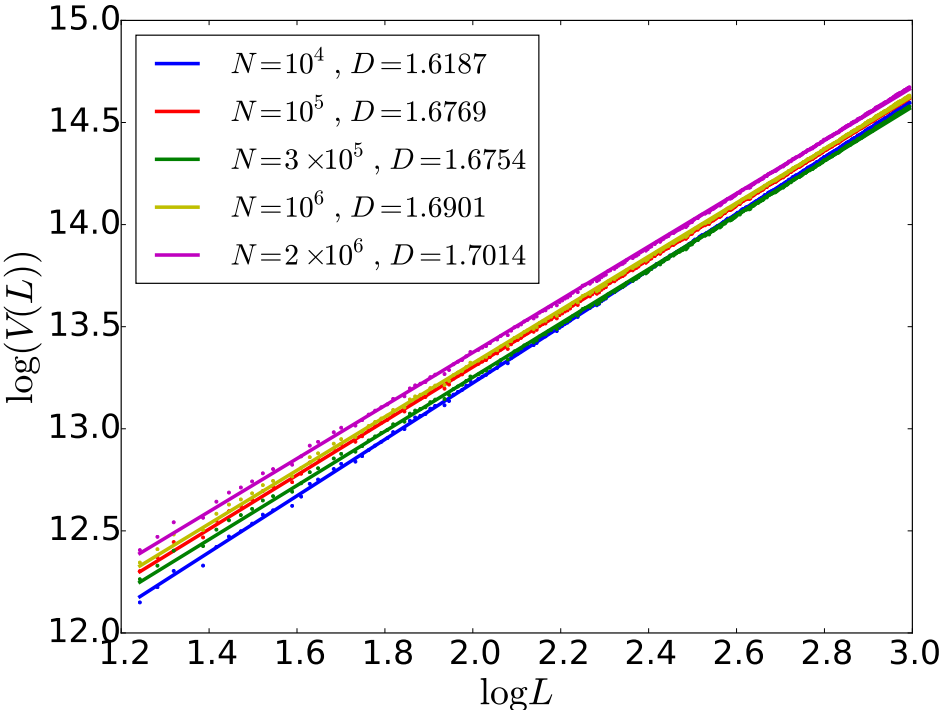


Figure 5.3: The log-log plot of the influence volume $V(L)$ versus the radius L (up to 20) of sphere at each point. The fractal dimension D is equal to $3-a$, where a is the coefficient by linear regression. Self-avoiding walks of different lengths are analyzed by this Bouligand-Minkowski method.

Because of the limitation of computer resources, the algorithm of Barycentric Fixed-Mass method provided by Kamer can only calculate the fractal information of systems with up to 3×10^5 points. The Figure 5.4 and 5.5 show respectively the log-log plot of Equation 5.6 for $q\approx0$ and D_q-q curve for the self-avoiding walks with $N=10^4, 10^5, 3\times10^5$. The numbers of conformations for averaging are 15000, 200, 100. When $q\approx0$, the estimated values of D_0 are 1.6962, 1.7012, 1.7008 for the three systems, which is much better than the results of the above two methods. Since the self-avoiding walk is statistically monofractal, D_q should be the same when varying q. The slight increase in Figure 5.5 is attributable to the finite size effect and statistical deviation. Considering these reliable results, this method will be employed in the analysis of the contacts of SAW and other systems.

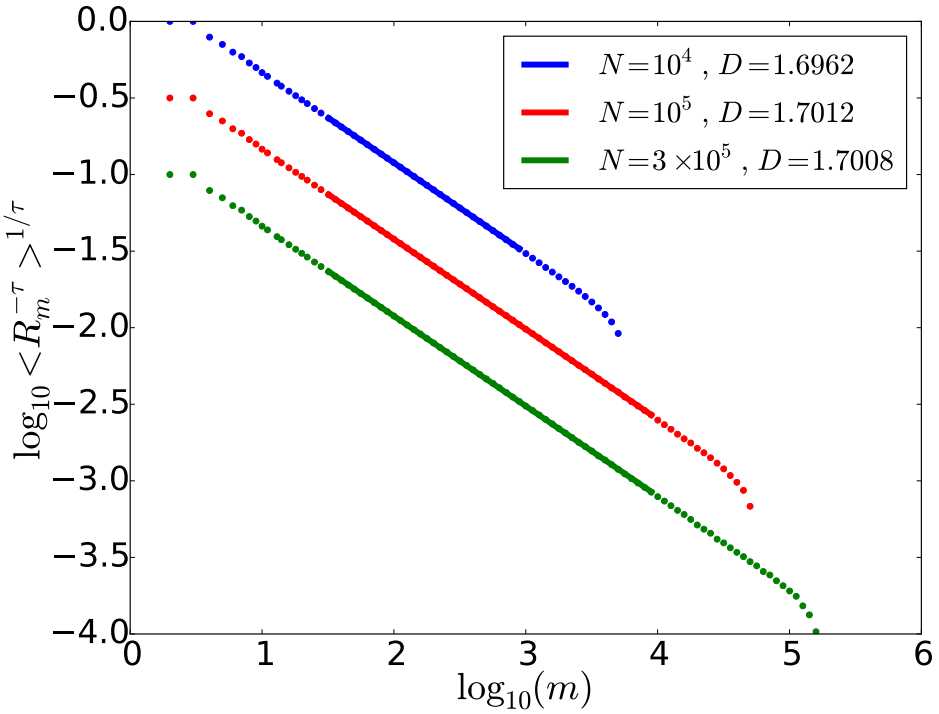


Figure 5.4: The log-log plot of one moment of R_m versus m. The moment order is selected so that $q \approx 0$. The system sizes are $N = 10^4, 10^5, 3 \times 10^5$. The fractal dimension D is related to the slope a of the linear parts of the points by $D = -1/a$. The data for $N = 10^5$ and $N = 3 \times 10^5$ are shifted downwards by 0.5 and 1 artificially to avoid the overlap of the points.

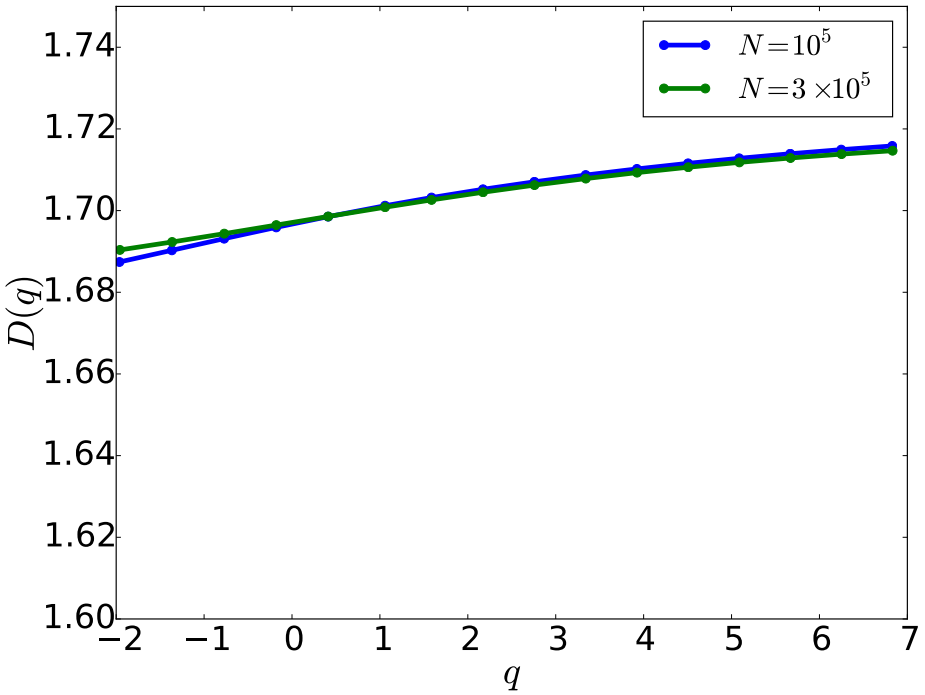


Figure 5.5: D_q versus q for self-avoiding walk with $N = 10^5$ and $N = 3 \times 10^5$. In principle D_q should not change with q for self-avoiding walk. The slight increase is due to the finite size effect and statistical error.

The routine of persistent homology is applied to the self avoiding walk using the package Dionysus. The alpha complexes are constructed for the filtration due to its high efficiency when dealing with large systems. We analyzed self-avoiding walks of length $N = 3 \times 10^5$ and $N = 10^6$, 500 and 300 independent conformations of which are generated to average. Figure 5.6 and 5.7 show the growth of the first and the second Betti number. The estimated values of γ_1 and γ_2 from linear fitting are about 1.75 and 1.72 respectively.

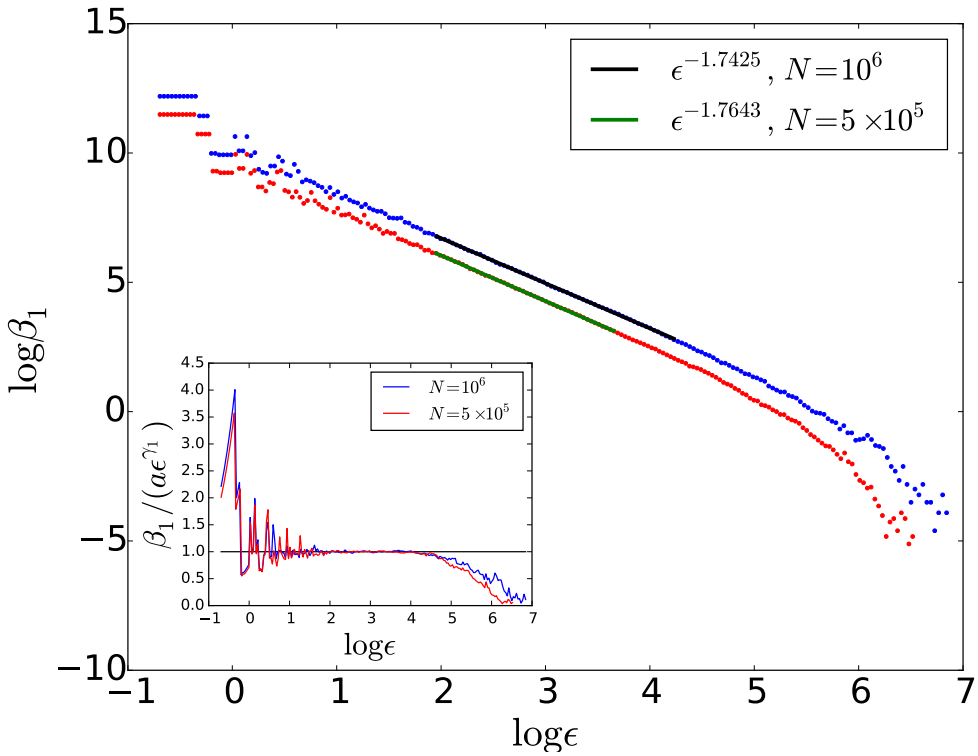


Figure 5.6: The log-log plot of the dependence of the first betti number β_1 on the resolution ϵ for self-avoiding walks. The lengths of the walks are $N = 10^6$ and $N = 5\times10^5$. The results are averaged by 300 and 500 independent conformations. By fitting the linear regions of both set of points, we get the growth rate $\gamma_1 \approx 1.76$. The inset shows the ratio between the original data and the power law.

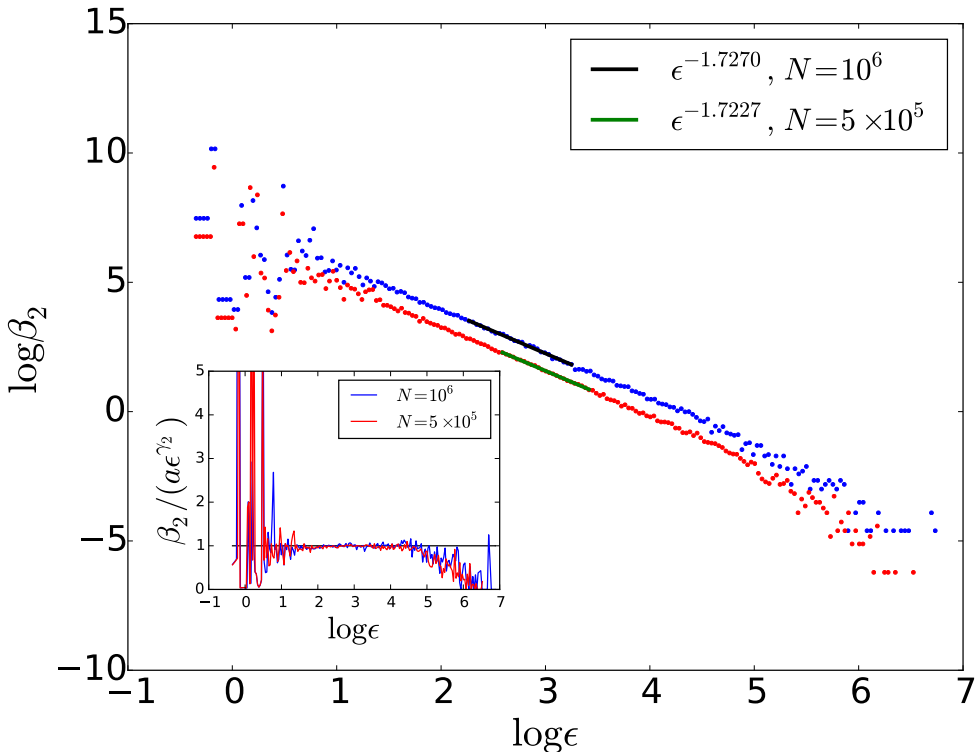


Figure 5.7: The log-log plot of the dependence of the second betti number β_2 on the resolution ϵ for self-avoiding walks. The lengths of the walks are $N = 10^6$ and $N = 5\times10^5$. The results are averaged by 300 and 500 independent conformations. By fitting the linear regions of both set of points, we get the growth rate $\gamma_2 \approx 1.72$. The inset shows the ratio between the original data and the power law.

5.3.2 The Geometric and Topological Properties of the Contact of Self-Avoiding Walk

The contacts of inter or intra biomolecules often indicate the realization of some biological functions. These contacts are driven by many kinds of interactions such as electrostatic forces and hydrogen bonds. For a totally flexible linear chain, the contacts are caused by randomness. The average number of contacts of self-avoiding walks is found to have an asymptotic behavior with respect to the length of the walks ($\langle N_c\rangle_{\mathrm{SAW}} \sim a_\infty N$) [47]. In this section we investigate the distribution of the contacts from geometric and topological aspects.

In the cubic lattice model of SAW, when two nonconsecutive monomers are nearest neighbors, they are counted as a contact, the position of which is the average

of these two monomers. The set of those contact points is clearly near the backbone of the original SAW, that means it is a subset of the SAW. However, the fractality is not necessarily the same. Due to the excluded volume effect, two densely contacted regions should be separated spatially (Figure 5.8).

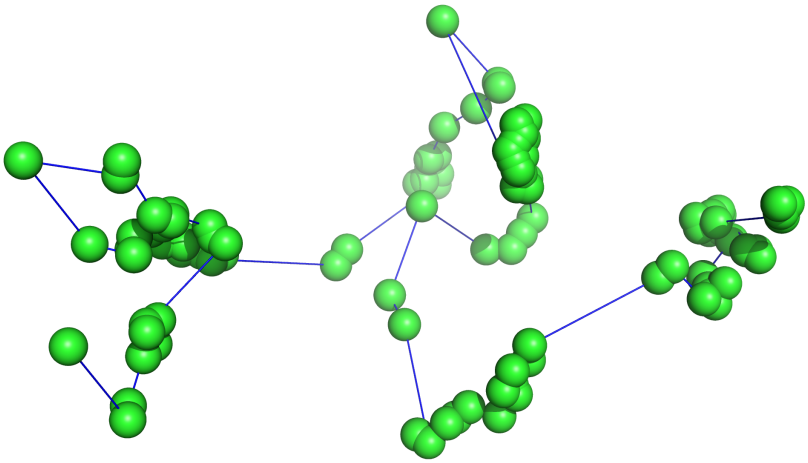


Figure 5.8: The contacts of a self-avoiding walk. Each contact is represented by a green sphere. The blue lines do not mean that the contacts are connected, just show the separation of different contacted regions.

Different from the self-avoiding walk, the point-cloud of contact seems to be disordered. But do they have the self-similar property? To explore this, the Barycentric Fixed-Mass method is applied to study its fractality. The contacts of $N = 10^6$ and $N = 2 \times 10^6$ self-avoiding walks are recorded, each with 250 and 200 conformations. By calculating the relation in Equation 5.6, we indeed find a power law dependence(Figure 5.9), which suggests that the set of contacts is fractal. However, D_q exhibits a decreasing with the moment order q (Figure 5.10), which implies that the set of contacts is multifractal.

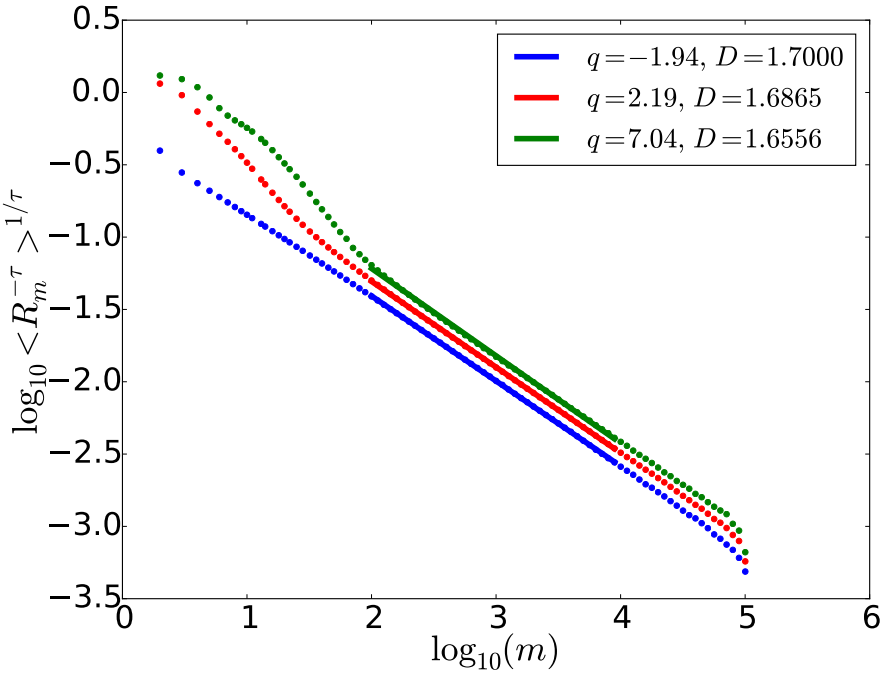


Figure 5.9: The log-log plot of three moments of R_m versus m for the contacts of $N = 10^6$ self-avoiding walk. The moments are $\tau = -5,\ 2,\ 10$.

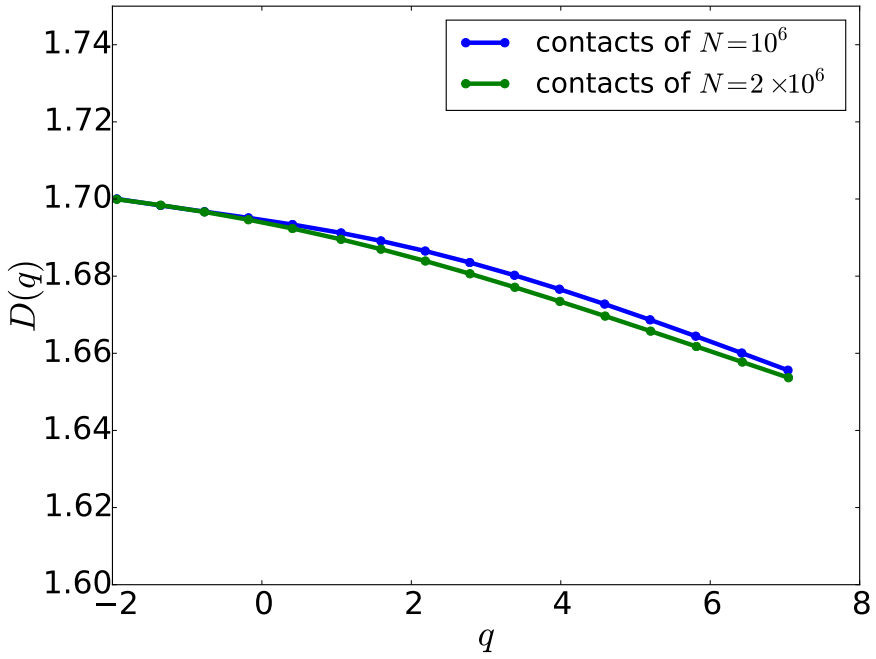


Figure 5.10: The $D_q - q$ relation for the contact points of the SAW $N = 10^6$ and $N = 2 \times 10^6$.

The growth of the first and second Betti number of the contacts is shown in Figure 5.11 and 5.12. The estimated values of γ_1 for the contacts of SAW $N = 2\times10^6$ and $N = 10^6$ are 1.7707 and 1.7735. The estimated values of γ_2 are 1.8108 and 1.8029.

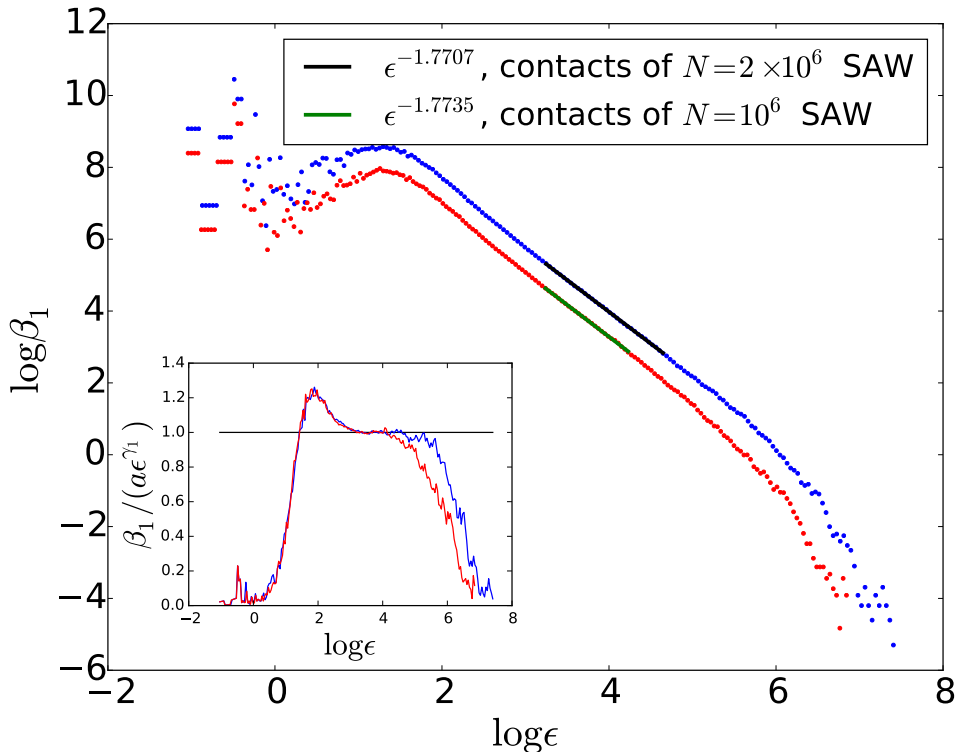


Figure 5.11: The dependence of first Betti number on resolution ε for contacts of $N = 2\times10^6$ and $N = 10^6$ SAW. The first Betti number is averaged over 200 and 250 independent conformations of SAW. Shown in the inset is the ratio between the Betti number and the power law.

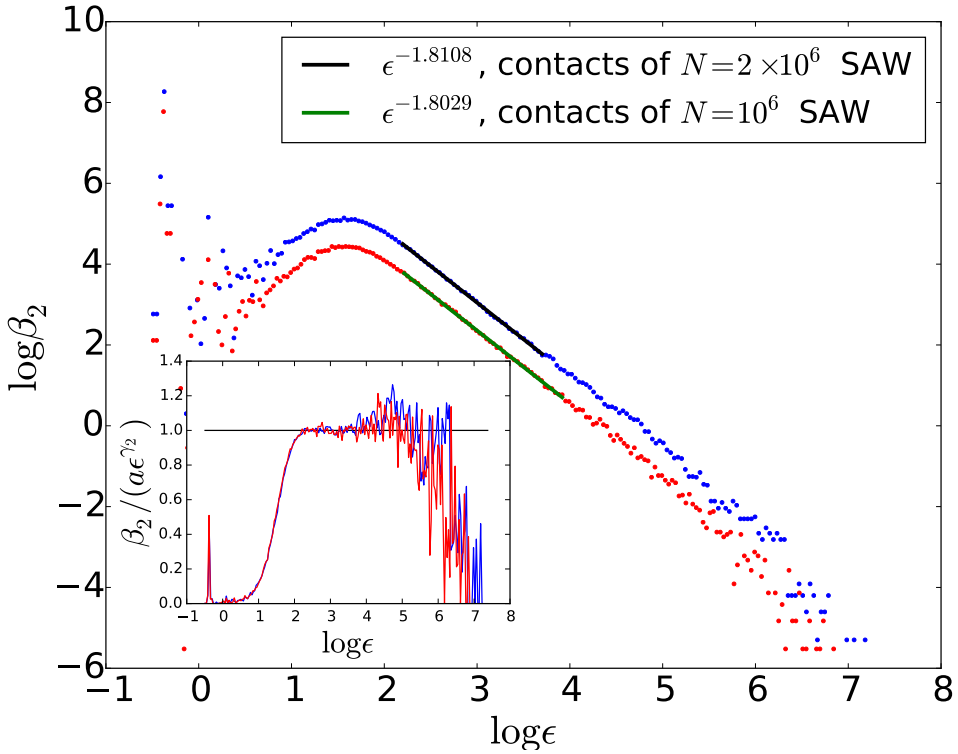


Figure 5.12: The dependence of second Betti number on resolution ε for contacts of $N = 2 \times 10^6$ and $N = 10^6$ SAW. The second Betti number is averaged over 200 and 250 independent conformations of SAW. Shown in the inset is the ratio between the Betti number and the power law.

5.3.3 The Loop-Deleted Self-Avoiding Walk

When two nonconsecutive monomers come into contact, a loop appears. What will the walk be like if we erase these loops? For random walk, the loop erasure produces a new kind of walk [204, 205]: the loop-erased random walk (LERW), or loop-erased self-avoiding walk (LESAW) in some literatures. To avoid any ambiguity, we denote the same procedure for self-avoiding walk as loop deletion, and the resulting new walk is loop-deleted self-avoiding walk (LDSAW) (Figure 5.13). Clearly, the LDSAW is also self-avoiding, but the minimal distance between nonconsecutive monomers is $\sqrt{2}$ if the SAW is on a square or cubic lattice. The LERW has a different critical exponent from the SAW: $\nu_{\text{LERW}} \approx 0.616$ [209–212]. Similarly, we expect another critical exponent λ for LDSAW.

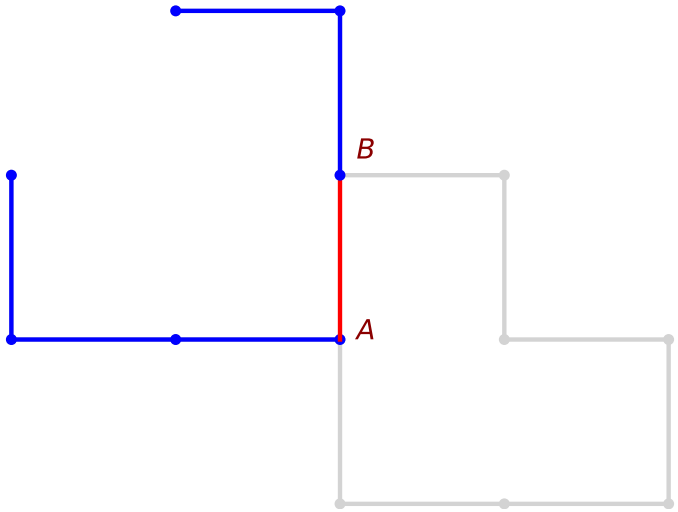


Figure 5.13: The loop deletion of a 2D self-avoiding walk on a square lattice. The nonconsecutive monomer A and B of SAW come into contact and a loop arises. By deleting all the monomers between A and B (the grey part) and connecting A, B, we get the LDSAW (blue lines).

Different from the LERW, to obtain the LDSAW, it is impractical to delete the loops while growing the SAW due to its low efficiency. Instead, we choose the second way: first generate sufficient samples of the SAW, then the loops are detected and deleted. The pivot algorithm [102] is used to simulate 3D SAW on a cubic lattice. Given a self-avoiding walk of length N, deleting all the loops in chronological order results in a walk of length M_N. Clearly M_N is a random number ranging from 2 to N. Following the definition for LERW in [209], if we consider the asymptotic behavior of the average length of the LDSAW:

$$\langle M_N \rangle \sim N^{1/2\mu}, N \to \infty, \tag{5.8}$$

Then we have:

$$\langle R_e^2 \rangle_{\text{LDSAW}} = \langle R_e^2 \rangle_{\text{SAW}} \sim N^{2\nu} \sim \langle M_N \rangle^{4\nu\mu}. \tag{5.9}$$

This gives a definition of the critical exponent of LDSAW: $\lambda = 2\nu\mu$, where $\nu \approx 0.588$ is the exponent for the SAW.

Alternatively, the exponent could also be extracted from the relation $\langle R_e^2 \rangle \sim m^{2\lambda}$ by generating sufficient conformations of LDSAW of length M. To get an almost complete of samples of LDSAW with length M, we delete the loops of SAW of length N, and take the segment of length M from all resulting new walks as long as $M_N \geqslant M$. N should be much larger than M so that samples of M-step LDSAW from deleting loops of SAW longer than N account for a negligible part of the ensemble of M-step LDSAW.

Figure 5.14 shows the relation between $\langle M_N \rangle$ and N. By a linear fitting in the log-log plot of $\langle M_N \rangle$ versus N, we get $\mu \approx 0.511$. Thus the estimated critical exponent for LDSAW is $\lambda = 2\mu\nu \approx 0.600$. The asympotitic behavior of the mean squared end-to-end distance of LDSAW is displayed in Figure 5.15. By linear regression, we get the exponent $\lambda \approx 0.601$. These two methods give the same values for λ, therefore we conclude that the new walk has a critical exponent 0.60. This new exponent is larger than the exponent ν of SAW, which is reasonable since the loop deletion basically stretches the self-avoiding walk. On the other hand, we could also define a stretched self-avoiding walk (SSAW) with minimal distance $\sqrt{2}$ between nonconsecutive monomers. Nonetheless, this SSAW has the same critical exponent with SAW. Therefore as with LERW and SAW, SSAW and LDSAW are not in the same universality class. Another point worth mentioning is that although the LDSAW is more stretched, it has a smaller critical exponent than the LERW (0.612). The reason is that the stretching of LDSAW is more local compared with the LERW.

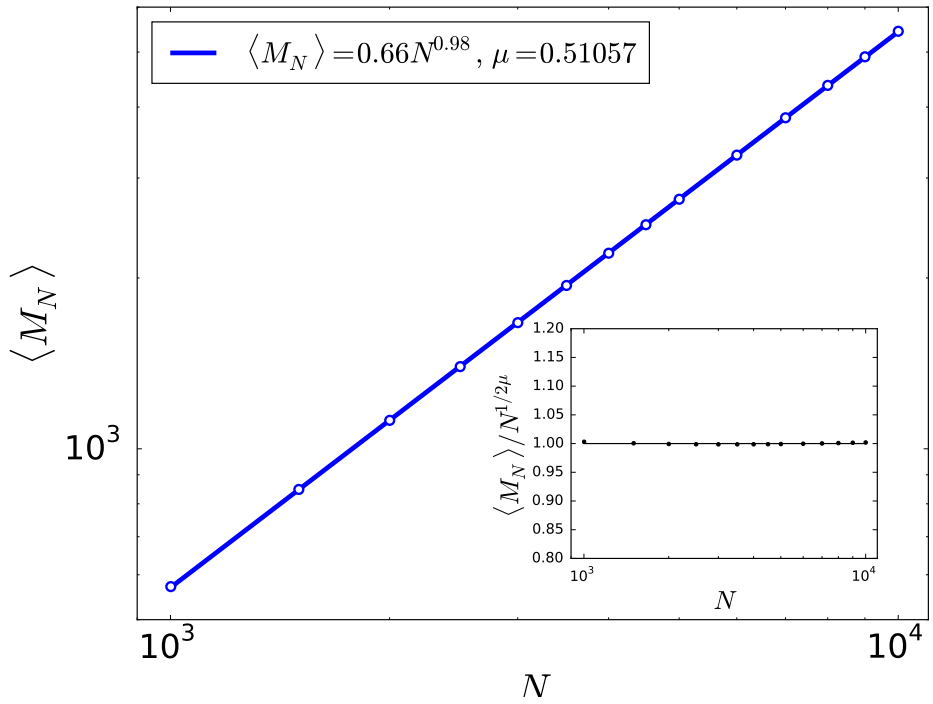


Figure 5.14: The log-log plot of average length of LDSAW $\langle M_N \rangle$ versus N. Shown in the inset is the ratio between $\langle M_N \rangle$ and the power law by fitting.

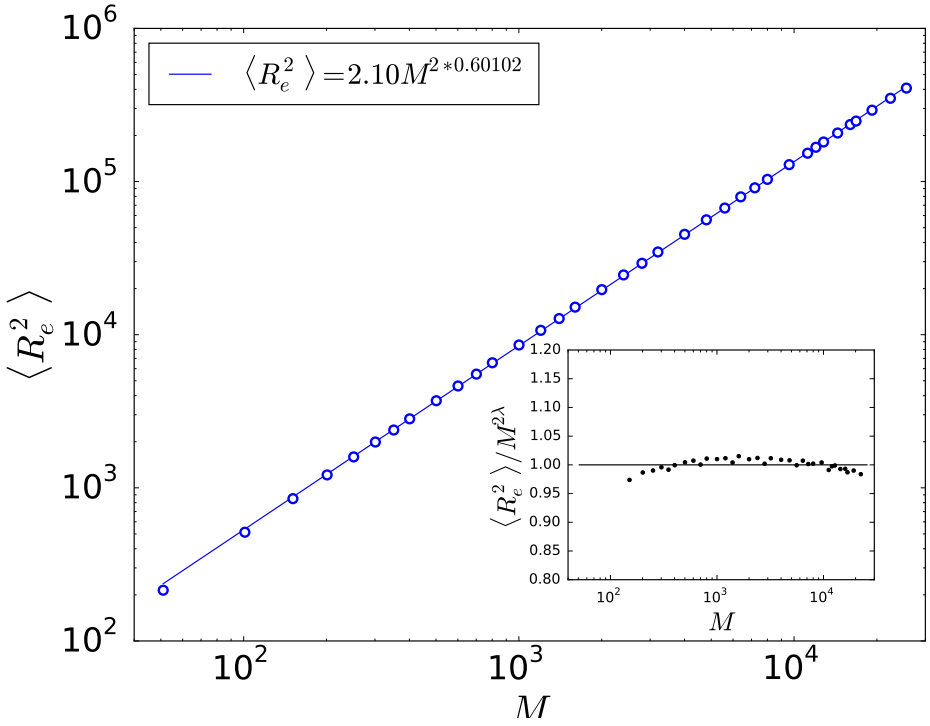


Figure 5.15: The dependence of the mean squared end-to-end distance $\langle R_e^2 \rangle$ of LDSAW on the length M. The inset is the ratio between $\langle R_e^2 \rangle$ and the power law.

Via the critical exponent, we can get the fractal dimension of the LDSAW $D^{\mathrm{ld}} = 1/\lambda \approx 1.667$. To verify this value, the Barycentric fixed-mass method is applied to the LDSAW with length $M = 10^5, 3 \times 10^5$. Figure 5.16 shows the log-log plot of one moment of R_m versus the mass m (Equation 5.6). τ is selected so that $q \approx 0$. The lengths of the LDSAW are $M = 10^5$ and $M = 3 \times 10^5$. The results are averaged over 200 conformations.

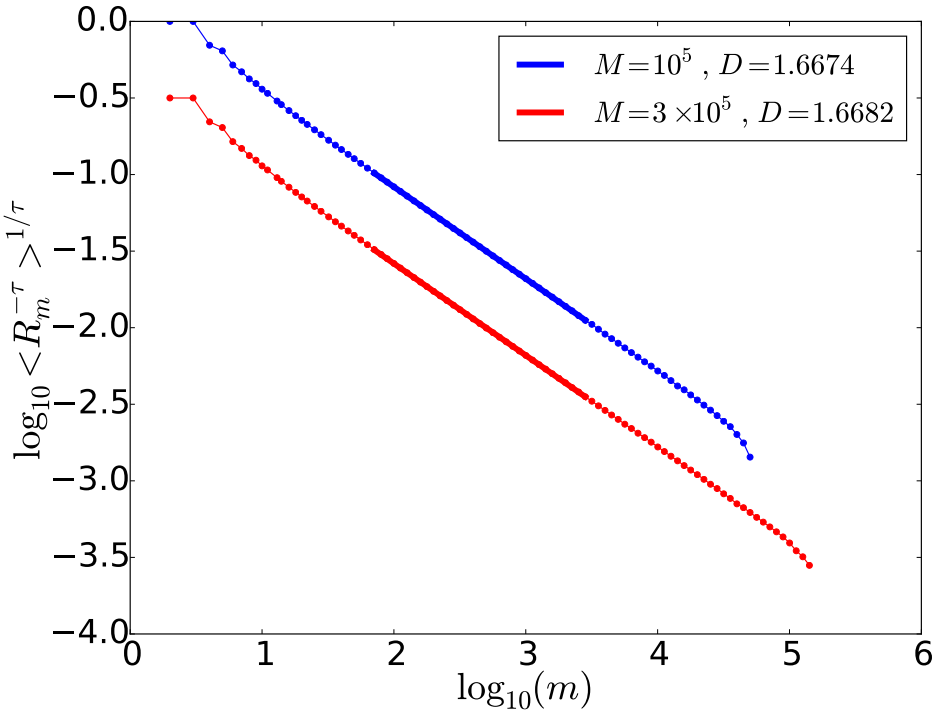


Figure 5.16: The log-log plot of one moment of R_m versus m. The moment order is selected so that $q \approx 0$. The lengths of the LDSAW are $M = 10^5$ and $M = 3 \times 10^5$. By determining the slope a of the linear parts, the fractal dimension of LDSAW is $D^{\text{ld}} \approx 1.6674$ and 1.6682.

The growth of the first and second Betti number of the LDSAW is shown in Figure 5.17 and Figure 5.18. The estimated values of γ_1 for deleting the loops of SAW $N = 5 \times 10^5$ and $N = 10^6$ are 1.6742 and 1.6689, which are very close to the fractal dimension 1.67. On the other side, the performance of estimating γ_2 is poor, with values 1.6980 and 1.6579. More conformations should be used to reduce the statistical error.

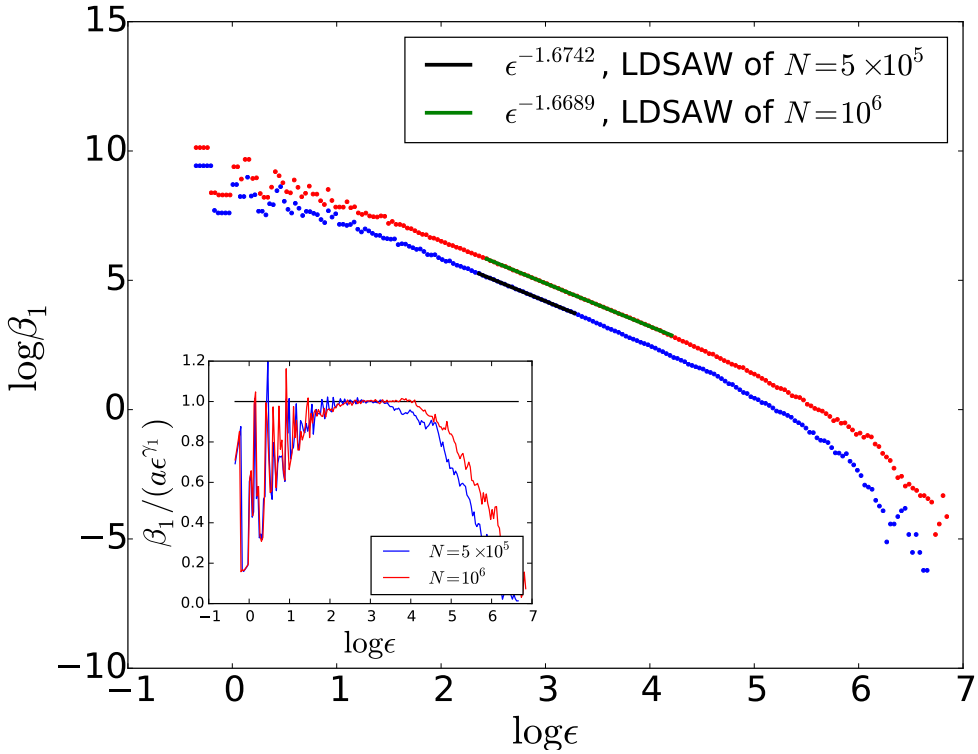


Figure 5.17: The dependence of first Betti number on resolution ε for LDSAW of length $M = 10^5$ and $M = 3 \times 10^5$. The first Betti number is averaged over 200 independent conformations of LDSAW. Shown in the inset is the ratio between the Betti number and the power law.

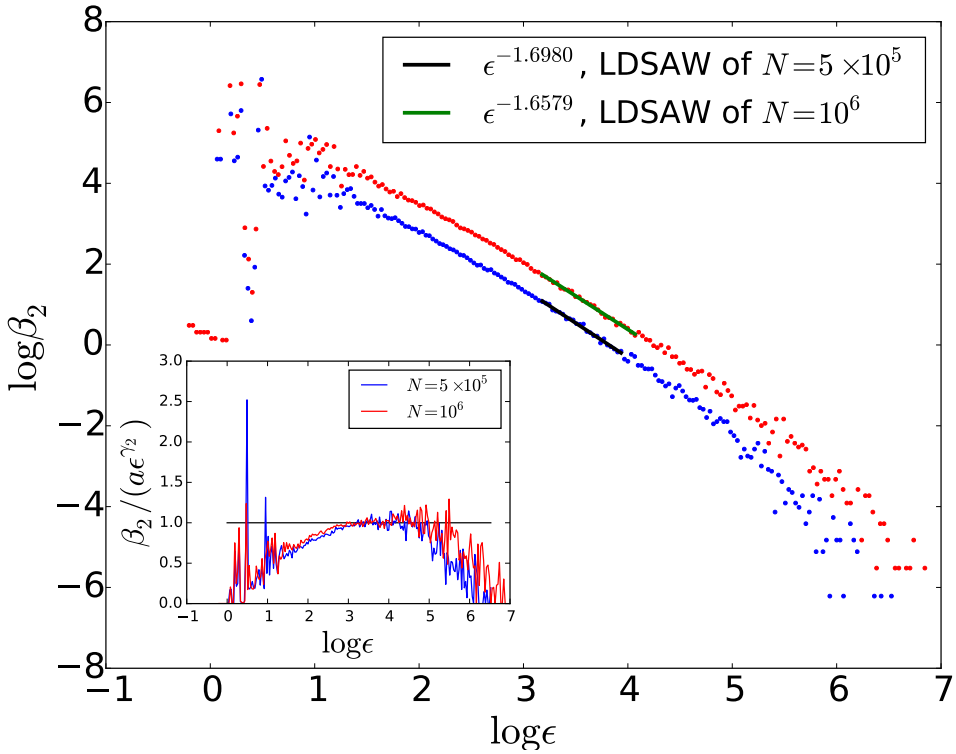


Figure 5.18: The dependence of second Betti number on resolution ε for LDSAW of length $M = 10^5$ and $M = 3 \times 10^5$. The second Betti number is averaged over 200 independent conformations of LDSAW. Shown in the inset is the ratio between the Betti number and the power law.

5.3.4 The Projection of Self-Avoiding Walk on A Plane

By projecting the 3D self-avoiding walk onto a plane, the knots information can be obtained, such as the writhe and generalized linking number [244, 245]. Shown in Figure 5.19 is the projection of a 3D SAW onto a plane determined by two vectors $\vec{e}_x = (0.804,\ 0.525,\ 0.280), \vec{e}_y = (0.238,\ 0.149,\ -0.960)$.

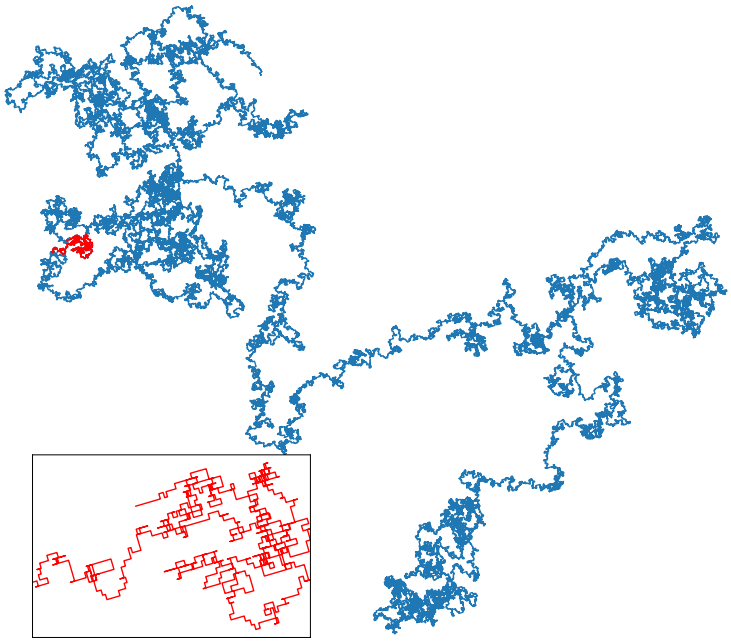


Figure 5.19: The projection of a 3D SAW onto a random plane. The red part is zoomed in the inset.

Clearly the walk on the plane is not self-avoiding any more, but it seems to be self-similar. We wonder what kinds of geometrical and topological information are hidden, and how they may be related to the properties of the original 3D SAW. By calculating the fractal dimension using the Barycentric Fixed-Mass method (Figure 5.20), the projection of SAW is monofractal and has a fractal dimension around 1.62.

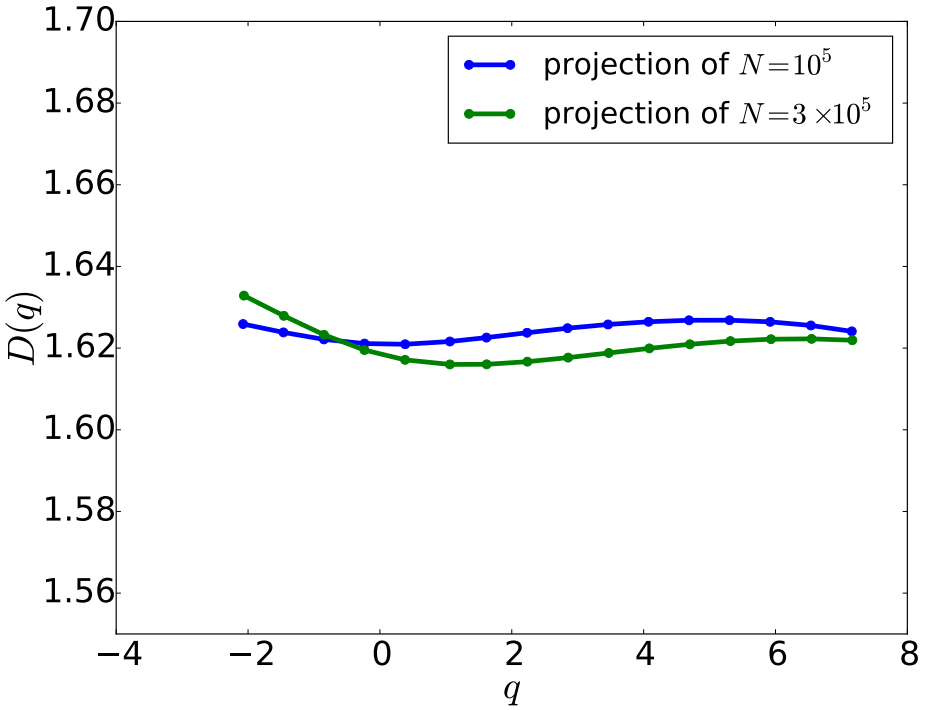


Figure 5.20: The $D(q) - q$ plot of the projection of a 3D SAW. The lengths of the SAW are $N = 10^5$ and $N = 3 \times 10^5$. 200 conformations are used for these two walks.

The growth of the first Betti number of the projection is shown in Figure 5.21. The estimated values of γ_1 for the projections of $N = 10^6$ and $N = 3 \times 10^5$ are 1.7834 and 1.8108.

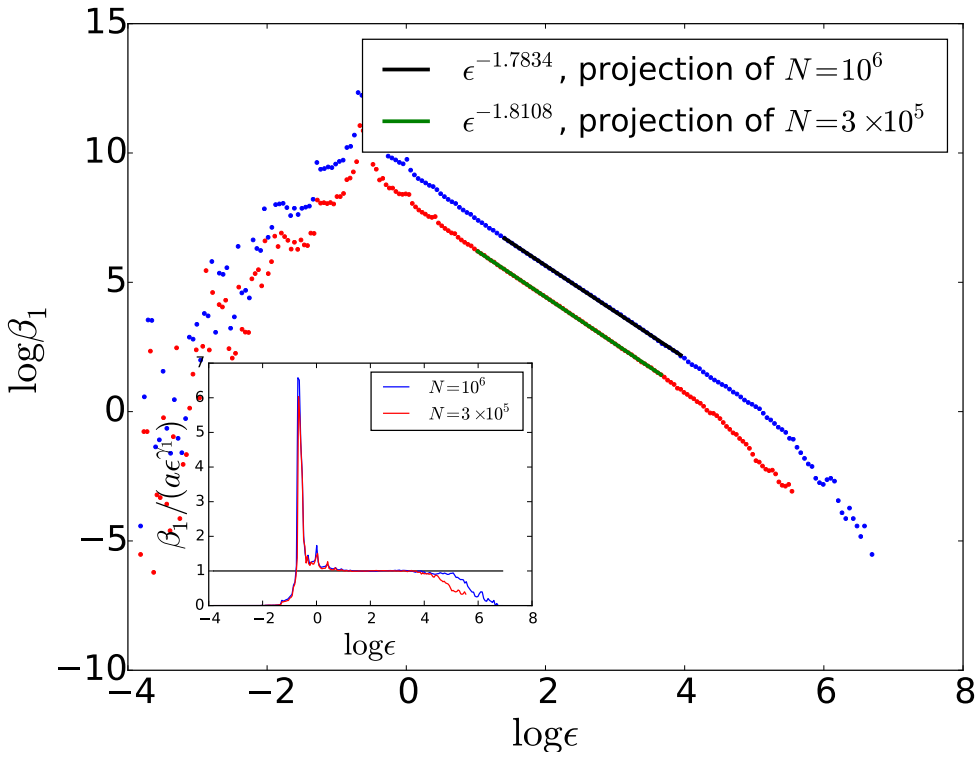


Figure 5.21: The dependence of first Betti number on resolution ε for projection of 3D SAW ($N = 10^5$ and $N = 3\times10^5$). The first Betti number is averaged over 200 independent conformations. Shown in the inset is the ratio between the Betti number and the power law.

5.3.5 Randomly Deleting Points of SAW

Either the contacts of SAW or the LDSAW in section 5.3.2 and 5.3.3 is one subset of the SAW, but they carry different geometrical and topological information. In this section, for comparison we study another subset of the SAW by deleting points in a SAW with equal probability. It is expected that this kind of subset shares the same information with the SAW. In Figure 5.22 is the multifractal analysis of three random subsets of the $N = 10^6$ SAW using the Barycentric Fixed-Mass method. The three subsets contain 10%, 20%, 30% points of the SAW respectively. The results are averaged over 300 conformations. It shows that this random subset has same fractality with the SAW: it is monofractal with dimension $D^{\mathrm{rd}} \approx 1.7$.

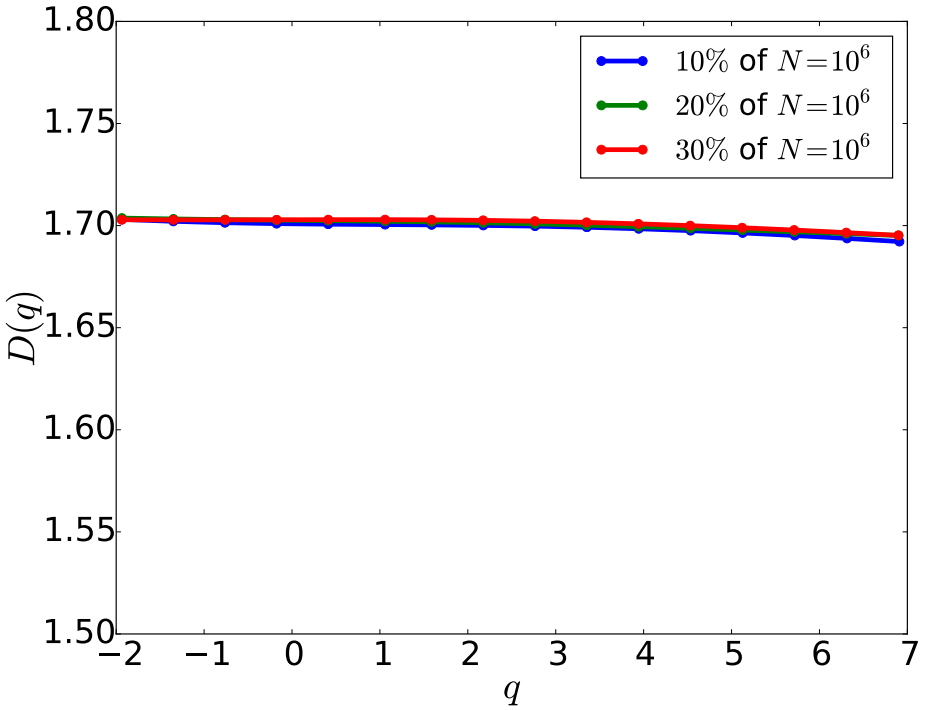


Figure 5.22: The $D(q) - q$ plot of three random subsets of the $N = 10^6$ SAW using the Barycentric Fixed-Mass method. The three subsets contain 10%, 20%, 30% points of the SAW respectively. The results are averaged over 300 conformations.

The growth of the first and second Betti number of the random subsets of SAW are shown in Figure 5.23 and Figure 5.24. The estimated values of γ_1 for 300 0000 points and 500 000 points are 1.7577 and 1.7682, while the values of γ_2 are 1.7559 and 1.7603. Therefore we have $\gamma_1 \approx \gamma_2 \approx 1.76$.

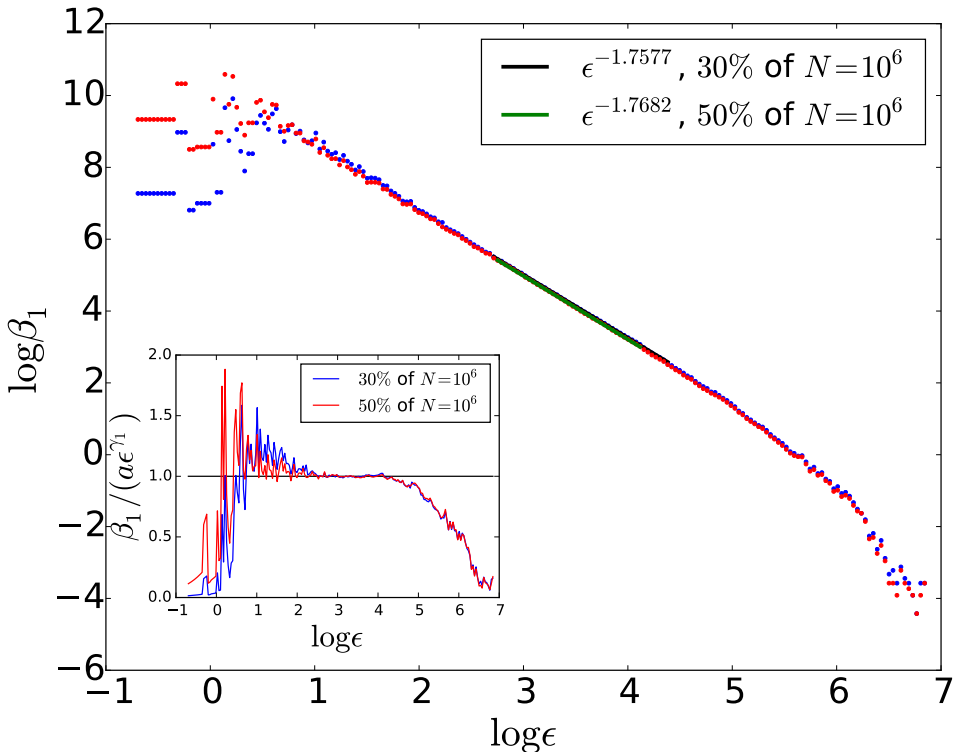


Figure 5.23: The dependence of first Betti number on resolution ε for random subsets of a SAW with $N = 10^6$. The first Betti number is averaged over 300 independent conformations. Shown in the inset is the ratio between the Betti number and the power law.

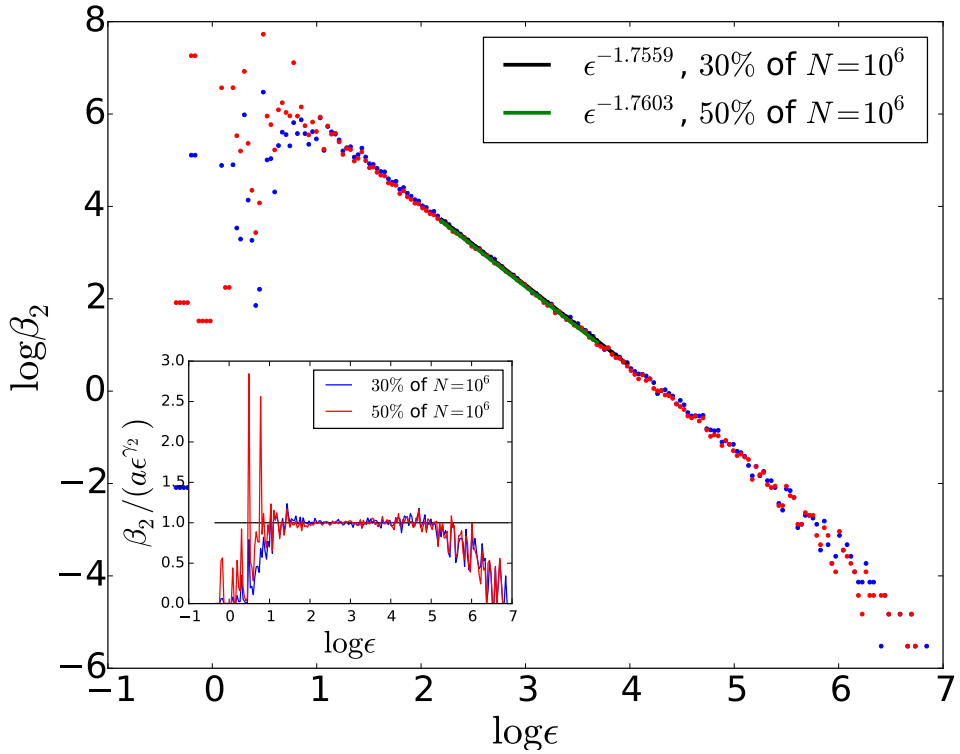


Figure 5.24: The dependence of second Betti number on resolution ε for random subsets of a SAW with $N = 10^6$. The second Betti number is averaged over 300 independent conformations. Shown in the inset is the ratio between the Betti number and the power law.

5.4 Conclusion

The fractal and topological features of the self-avoiding walk, its contacts, and the loop-deleted self-avoiding walk are studied in this chapter. First we select the Barycentric Fixed-Mass method to calculate the fractal dimension considering its accuracy in these systems. The fractal dimension of the self-avoiding walk is about 1.7, which is consistent with the existing result. We also calculate the fractal dimension of the contacts of a self-avoiding walk and find that the contacts exhibit a feature of multifractality. By deleting the loops inside a self-avoiding walk, we define the loop-deleted self-avoiding walk (LDSAW). Its critical exponent λ is estimated through the scaling behavior of average length and mean square end-to-end distance. The estimated value of λ is about 0.60. Then we use the same method to calculate its fractal dimension. The result is about 1.667, which is the reciprocal of

λ. The projection of a self-avoiding walk on a random plane is also analyzed, with a fractal dimension about 1.62. For comparison, we also study the random subset of a self-avoiding walk. As expected, its fractal dimension of the random set is also 1.7.

As for the topological features, we calculate the growth rates of the Betti numbers γ_1 and γ_2 within the framework of persistent homology, which provide the topological signature of different systems. Further, these two growth rates are related to the fractal dimension, but the relation has not been fully understood till now.

Chapter 6

Conclusion and Outlook

6.1 A Summary of the Results

We studied several important topics in the modeling of chromosomes organization in this thesis, including the bending rigidity and confinement, the contacts segregation in chains system, and the fractal and topological features in the self-avoiding walks. We believe that our findings could provide some new perspective into the study of chromosomes organization.

In chapter 3 we stressed the contact definition of monomers as well as the invariability of asymptotic behavior for different cut-off distances in the definition and when the bending rigidity comes into play. At very large length scale, the contact probability for linear chains in free space will exhibit the same power-law decay $P(s) \sim s^{-2.2}$, with the coefficient depending on the cut-off distance for contact. Secondly, we investigated the effects of bending rigidity and its heterogeneity on the conformations of a linear chain in free space and under geometric confinement. We found that the heterogeneity of bending rigidity makes the chain more flexible by measuring the contact probability and persistence length. Under confinement, the bending potential reshapes the chain due to the competitive interplay of stiffness, entropy, and the geometrical constraints. Moreover, there exists a "shape transition" from overall randomness to orderliness when the persistence length is comparable to the size of confinement. We analyzed the contact probability of a flexible or semiflexible chain and found that in sufficient small space it has a plateau region

in the large contour length region, as opposed to the power-law decay in free space. Moreover, if the bending rigidity is big enough compared to the size of the confinement, this plateau region will turn into an oscillation (Figure 3.7), which indicates the existence of spirals formed by the chain. The ordering of the chain according to the shape of geometric confinement can also be studied by constraining the chain into rectangle boxes of different aspect ratios. We used the order parameter S to quantify the ordering of chain segments. It is shown that the semiflexible chain preferably chooses the long direction of the boxes to order the segments. The orientational correlation function $\langle \cos\theta(s) \rangle$ of bond vectors is also dramatically changed due to the bending rigidity and confinement, and the oscillation in it serves as direct evidence of the formation of spirals. The presence of bending rigidity leads to less crossing inside a chain. However, the asymptotic behavior of the mean average crossing number of semiflexible chains stays the same as the flexible case. Two kinds of relationships can describe the dependence of mACN on the chain length. On the other side, the geometrical constraints could induce more crossing, where both the size and symmetry of the constraints kick in. In particular, the more symmetric the confinement is, the more crossing the chain has. The reason is that the symmetry reduces the effect of bending energy that tends to minimize crossing by stretching the chain along some preferred directions in confinement.

In chapter 4 we focused on two main topics: the self-contact of two halves of a single polymer chain and the segregation of two chains in a rectangular box. First we studied the distribution function of the number of contacts $f(N_c(AB))$ of two halves for a single chain with and without confinement. For a chain in free space, the distribution function exhibits a power law decay in small $N_c(AB)$ region, and a exponential decay in large $N_c(AB)$ region. For a single chain confined in a cubic box, when the size of the box is small enough, the distribution function first increases with $N_c(AB)$, then decreases rapidly. Next we studied the dependence of total number of contacts N_c on the size of the segment we are looking at. For free chain, N_c grows linearly with the length of the segment s_m, while for confined chain, N_c grows quadratically. Further, we studied the percentage of the inter-contacts of the two halves of the segment $p_{_{AB}}(s_m)$, finding that it has a power law decay for free chains, with an exponent slightly larger than -1. We assume that the discrepancy is caused by statistical error and finite size effects. This power law decay demonstrates that the number of contacts between two halves of a self-avoiding chain is finite as the length of the chain goes to infinity, which indicates that the two halves are nearly separated. For confined chain, in small s_m region, $p_{_{AB}}(s_m)$ overlaps with the free chain case, meaning that in small length scale, the segment does not feel the constraints. As s_m becomes larger, the confinement induces more contacts between two halves of the chain. Specifically, for a Hamiltonian path, the total number of inter-

contacts is equal to the intra-contact, i.e. $p_{AB}^{HP} \approx 0.33$. The separation of two halves of a self-avoiding chain is also revealed by the distribution of monomers projected onto the line connecting the mass centers of the two halves. The segregation of two chains in a rectangular box starting from different initial conformations is also investigated. We created three kinds of initial conformations: separated, intertwined, mid-monomers fixed but with different numbers of inter-contacts. By looking at the evolution of mass centers distance, number of inter-contacts, winding number, and the peak value of dynamic structure factor, we found that for the system only with excluded volume interaction, these different initial conformations have similar segregation time because this process is mainly governed by a much slower diffusion of chains. However, the intertwinement indeed would impede the segregation process at the beginning.

In chapter 5 the fractal and topological features of the self-avoiding walk, its contacts, and the loop-deleted self-avoiding walk were studied. First we used the Barycentric Fixed-Mass method to calculate the fractal dimension considering its accuracy in these systems. The fractal dimension of the self-avoiding walk is about 1.7, which is consistent with the existing result. We also calculated the fractal dimension of the contacts of a self-avoiding walk and found that the contacts exhibit a feature of multifractality. By deleting the loops inside a self-avoiding walk, we defined the loop-deleted self-avoiding walk (LDSAW). Its critical exponent λ was estimated through the scaling behavior of average length and mean square end-to-end distance. The estimated value of λ is about 0.60. Then we used the same method to calculate its fractal dimension. The result is about 1.667, which is the reciprocal of λ. The projection of a self-avoiding walk on a random plane was also analyzed, with a fractal dimension about 1.62. For comparison, we also studied the random subset of a self-avoiding walk. As expected, its fractal dimension of the random set is also 1.7. As for the topological features, we calculated the growth rates of the Betti numbers γ_1 and γ_2 within the framework of persistent homology, which provided the topological signature of different systems.

6.2 Outlook

We have presented some results on some topics in the modeling of chromosomes organization in our work. However, some questions remain open and are worth pondering. In chapter 3 we have studied the heterogeneity of the chromosome stiffness, where we assume that the bending parameter for consecutive bonds follows a Gaussian distribution. However, in real systems, the heterogeneity of chromosome is often manifested in a different manner. For example, in [246] the distribution of inter-

detection distances is used to measure the chromatin heterogeneity. One important source of the heterogeneity is the nonuniformity of the nucleosomes distribution on chromatin [247, 248], e.g., different intensity of interaction between alternate nucleosomes [249]. To illustrate the experimental results, the model should implement an appropriate form of interaction in the chromatin which can reflect the property of the heterogeneity.

The asymptotic behavior of the mean average crossing number (mACN) of the self-avoiding walk is now described by two forms: $aN \ln N + bN$ and aN^b, where N is the length of the walk. Both forms show an excellent compatibility with the simulation results. It is assumed that much longer chains are needed to determine which one is better.

The number of contacts is one main issue we dealt with in chapter 4. When studying the distribution function of contacts number between two halves of a chain, we found that the function shows a power law decay and then an exponential decay for a free chain. For a chain in a cubic box, the function has a maximum at certain contacts number. It is unclear whether the function has similar behaviors. To elucidate this, longer chains need to be simulated and more conformations should be used to reduce the statistical error. Besides, the dependence of this function on the shape and size of confinement is also interesting. In this chapter we also studied the segregation of two chains in a rectangular box, in which case only the excluded volume interaction is considered. We think that it would be more interesting if we implement other kinds of interaction, such as the attraction between nearest neighbors. Moreover, as shown in the results, two halves of a linear chain are more or less separated in free space and in confinement. The dynamics of this separation is also interesting.

In chapter 5 we studied the fractality and topology in the self-avoiding walks. In the organization of chromosomes, some fractal properties have also been detected by experiments and found to be able to indicate possible anomaly in living systems [250–252]. The multifractality of contacts of a self-avoiding walk is a new result in our study, but according to the $D(q) - q$ plot, the multifractality is weak. More studies should be done to corroborate this. In addition, we defined a new walk: loop-deleted self-avoiding walk, and found it has a critical exponent $\lambda \approx 0.60$. We are not very clear whether this walk has any counterpart in real systems. We further calculated the growth rates of the Betti number, which can serve as a topological signature of the system. However, the method is very sensitive to noise, therefore more robust methods should be developed. The relation of the growth rates with the fractal dimension is also an open question.

Bibliography

[1] Zeina Al-Natour and Ahmed H. Hassan. Effect of salt on the binding of the linker histone H1 to DNA and nucleosomes. *DNA and cell biology*, 26(6):445–452, June 2007.

[2] Patrick Cramer and Cynthia Wolberger. Proteins: Histones and Chromatin. *Current opinion in structural biology*, 21(6):695–697, December 2011.

[3] Karsten Rippe, Jacek Mazurkiewicz, and Nick Kepper. Interactions of Histones with DNA: Nucleosome Assembly, Stability, Dynamics, and Higher Order Structure. In *DNA Interactions with Polymers and Surfactants*, pages 135–172. John Wiley & Sons, Ltd, 2007.

[4] A Annunziato. Dna packaging: nucleosomes and chromatin. *Nature Education*, 1(1):26, 2008.

[5] Michael F Carey, Craig L Peterson, and Stephen T Smale. Chromatin immunoprecipitation (chip). *Cold Spring Harbor Protocols*, 2009(9):pdb–prot5279, 2009.

[6] Stefan Grob and Giacomo Cavalli. Technical review: a hitchhiker's guide to chromosome conformation capture. In *Plant Chromatin Dynamics*, pages 233–246. Springer, 2018.

[7] Mariliis Tark-Dame, Roel van Driel, and Dieter W Heermann. Chromatin folding–from biology to polymer models and back. *J Cell Sci*, 124(6):839–845, 2011.

[8] M. Bulger and M. Groudine. Looping versus linking: toward a model for long-distance gene activation. *Genes & development*, 13(19):2465–2477, 1999.

[9] Stephan Kadauke and Gerd A Blobel. Chromatin loops in gene regulation. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1789(1):17–25, 2009.

[10] Job Dekker and Tom Misteli. Long-Range Chromatin Interactions. *Cold Spring Harbor Perspectives in Biology*, 7(10):a019356, January 2015.

[11] Yousra Ben Zouari, Anne M. Molitor, Natalia Sikorska, Vera Pancaldi, and Tom Sexton. ChiCMaxima: A robust and simple pipeline for detection and visualization of chromatin looping in Capture Hi-C. *Genome Biology*, 20(1):102, May 2019.

[12] Manfred Bohn, Dieter W Heermann, and Roel van Driel. Random loop model for long polymers. *Physical Review E*, 76(5):051805, 2007.

[13] Manfred Bohn and Dieter W Heermann. Diffusion-driven looping provides a consistent framework for chromatin organization. *PloS one*, 5(8):e12218, 2010.

[14] Erez Lieberman-Aiden, Nynke L van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragoczy, Agnes Telling, Ido Amit, Bryan R Lajoie, Peter J Sabo, Michael O Dorschner, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *science*, 326(5950):289–293, 2009.

[15] Leonid A. Mirny. The fractal globule as a model of chromatin architecture in the cell. *Chromosome Research*, 19(1):37–51, January 2011.

[16] Benjamin D. Pope, Tyrone Ryba, Vishnu Dileep, Feng Yue, Weisheng Wu, Olgert Denas, Daniel L. Vera, Yanli Wang, R. Scott Hansen, Theresa K. Canfield, Robert E. Thurman, Yong Cheng, Günhan Gülsoy, Jonathan H. Dennis, Michael P. Snyder, John A. Stamatoyannopoulos, James Taylor, Ross C. Hardison, Tamer Kahveci, Bing Ren, and David M. Gilbert. Topologically associating domains are stable units of replication-timing regulation. *Nature*, 515(7527):402–405, November 2014.

[17] Jesse R. Dixon, Siddarth Selvaraj, Feng Yue, Audrey Kim, Yan Li, Yin Shen, Ming Hu, Jun S. Liu, and Bing Ren. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485(7398):376–380, 2012.

[18] Elphège P. Nora, Bryan R. Lajoie, Edda G. Schulz, Luca Giorgetti, Ikuhiro Okamoto, Nicolas Servant, Tristan Piolot, Nynke L. van Berkum, Johannes Meisig, John Sedat, Joost Gribnau, Emmanuel Barillot, Nils Blüthgen, Job Dekker, and Edith Heard. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, 485(7398):381–385, May 2012.

[19] Yin Shen, Feng Yue, David F. McCleary, Zhen Ye, Lee Edsall, Samantha Kuan, Ulrich Wagner, Jesse Dixon, Leonard Lee, Victor V. Lobanenkov, and Bing Ren. A map of the *cis*-regulatory sequences in the mouse genome. *Nature*, 488(7409):116–120, August 2012.

[20] Darío G. Lupiáñez, Katerina Kraft, Verena Heinrich, Peter Krawitz, Francesco Brancati, Eva Klopocki, Denise Horn, Hülya Kayserili, John M. Opitz, Renata Laxova, Fernando Santos-Simarro, Brigitte Gilbert-Dussardier, Lars Wittler, Marina Borschiwer, Stefan A. Haas, Marco Osterwalder, Martin Franke, Bernd Timmermann, Jochen Hecht, Malte Spielmann, Axel Visel, and Stefan Mundlos. Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell*, 161(5):1012–1025, May 2015.

[21] Rodolfo Ghirlando and Gary Felsenfeld. CTCF: Making the right connections. *Genes & Development*, 30(8):881–891, April 2016.

[22] Michael H. Kagey, Jamie J. Newman, Steve Bilodeau, Ye Zhan, David A. Orlando, Nynke L. van Berkum, Christopher C. Ebmeier, Jesse Goossens, Peter B. Rahl, Stuart S. Levine, Dylan J. Taatjes, Job Dekker, and Richard A. Young. Mediator and cohesin connect gene expression and chromatin architecture. *Nature*, 467(7314):430–435, September 2010.

[23] Janine B. Mills and Paul J. Hagerman. Origin of the intrinsic rigidity of DNA. *Nucleic Acids Research*, 32(13):4055–4059, 2004.

[24] Yuta Shimamoto, Sachiko Tamura, Hiroshi Masumoto, and Kazuhiro Maeshima. Nucleosome-nucleosome interactions via histone tails and linker DNA regulate nuclear rigidity. *Molecular Biology of the Cell*, 28(11):1580–1589, June 2017.

[25] Andrew D. Stephens, Patrick Z. Liu, Edward J. Banigan, Luay M. Almassalha, Vadim Backman, Stephen A. Adam, Robert D. Goldman, and John F. Marko. Chromatin histone modifications and rigidity affect nuclear morphology independent of lamins. *Molecular Biology of the Cell*, 29(2):220–233, January 2018.

[26] Maria Aurelia Ricci, Carlo Manzo, María Filomena García-Parajo, Melike Lakadamyali, and Maria Pia Cosma. Chromatin Fibers Are Formed by Heterogeneous Groups of Nucleosomes In Vivo. *Cell*, 160(6):1145–1158, March 2015.

[27] Helmut Schiessel. The physics of chromatin. *Journal of Physics: Condensed Matter*, 15(19):R699–R774, May 2003.

[28] Frank Aumann, Filip Lankas, Maïwen Caudron, and Jörg Langowski. Monte Carlo simulation of chromatin stretching. *Physical Review E*, 73(4):041927, April 2006.

[29] J. Langowski. Polymer chain models of DNA and chromatin. *The European Physical Journal E*, 19(3):241–249, March 2006.

[30] Jörg Langowski and Dieter W. Heermann. Computational modeling of the chromatin fiber. *Seminars in Cell & Developmental Biology*, 18(5):659–667, October 2007.

[31] Suckjoon Jun and Bela Mulder. Entropy-driven spatial organization of highly confined polymers: Lessons for the bacterial chromosome. *Proceedings of the National Academy of Sciences*, 103(33):12388–12393, August 2006.

[32] Julio Mateos-Langerak, Manfred Bohn, Wim de Leeuw, Osdilly Giromus, Erik M. M. Manders, Pernette J. Verschure, Mireille H. G. Indemans, Hinco J. Gierman, Dieter W. Heermann, Roel van Driel, and Sandra Goetze. Spatially confined folding of chromatin in the interphase nucleus. *Proceedings of the National Academy of Sciences*, 106(10):3812–3817, March 2009.

[33] Gadiel Saper, Stanislav Kler, Roi Asor, Ariella Oppenheim, Uri Raviv, and Daniel Harries. Effect of capsid confinement on the chromatin organization of the SV40 minichromosome. *Nucleic Acids Research*, 41(3):1569–1580, February 2013.

[34] Miriam Fritsche and Dieter W. Heermann. Confinement driven spatial organization of semiflexible ring polymers: Implications for biopolymer packaging. *Soft Matter*, 7:6906, 2011.

[35] Katja Ostermeir, Karen Alim, and Erwin Frey. Buckling of stiff polymer rings in weak spherical confinement. *Physical Review E*, 81(6):061802, 2010.

[36] Ya Liu and Bulbul Chakraborty. Shapes of semiflexible polymers in confined spaces. *Physical biology*, 5(2):026004, 2008.

[37] Douglas R. Tree, Yanwei Wang, and Kevin D. Dorfman. Mobility of a Semiflexible Chain Confined in a Nanochannel. *Physical Review Letters*, 108(22):228105, June 2012.

[38] Ferhat Ay, Thanh H. Vu, Michael J. Zeitz, Nelle Varoquaux, Jan E. Carette, Jean-Philippe Vert, Andrew R. Hoffman, and William S. Noble. Identifying multi-locus chromatin contacts in human cells using tethered multiple 3C. *BMC Genomics*, 16(1):121, February 2015.

[39] Geoffrey Fudenberg and Leonid A Mirny. Higher-order chromatin structure: Bridging physics and biology. *Current Opinion in Genetics & Development*, 22(2):115–124, April 2012.

[40] Maxim Imakaev, Geoffrey Fudenberg, Rachel Patton McCord, Natalia Naumova, Anton Goloborodko, Bryan R. Lajoie, Job Dekker, and Leonid A. Mirny. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nature Methods*, 9(10):999–1003, October 2012.

[41] Ferhat Ay, Timothy L. Bailey, and William Stafford Noble. Statistical confidence estimation for Hi-C data reveals regulatory chromatin contacts. *Genome Research*, 24(6):999–1011, January 2014.

[42] Stefanie Kaufmann, Christiane Fuchs, Mariya Gonik, Ekaterina E. Khrameeva, Andrey A. Mironov, and Dmitrij Frishman. Inter-Chromosomal Contact Networks Provide Insights into Mammalian Chromatin Organization. *PLOS ONE*, 10(5):e0126125, May 2015.

[43] Sepideh Babaei, Ahmed Mahfouz, Marc Hulsman, Boudewijn P. F. Lelieveldt, Jeroen de Ridder, and Marcel Reinders. Hi-C Chromatin Interaction Networks Predict Co-expression in the Mouse Cortex. *PLOS Computational Biology*, 11(5):e1004221, May 2015.

[44] Geoffrey Fudenberg and Leonid A Mirny. Higher-order chromatin structure: bridging physics and biology. *Current opinion in genetics & development*, 22(2):115–124, 2012.

[45] Mariano Barbieri, Mita Chotalia, James Fraser, Liron-Mark Lavitas, Josée Dostie, Ana Pombo, and Mario Nicodemi. Complexity of chromatin folding is captured by the strings and binders switch model. *Proceedings of the National Academy of Sciences*, 109(40):16173–16178, October 2012.

[46] Erez Lieberman-Aiden, Nynke L. van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragoczy, Agnes Telling, Ido Amit, Bryan R. Lajoie, Peter J. Sabo, Michael O. Dorschner, Richard Sandstrom, Bradley Bernstein, M. A. Bender, Mark Groudine, Andreas Gnirke, John Stamatoyannopoulos, Leonid A. Mirny, Eric S. Lander, and Job Dekker. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science (New York, N.Y.)*, 326(5950):289–293, October 2009.

[47] Jack F. Douglas and Takao Ishinabe. Self-avoiding-walk contacts and random-walk self-intersections in variable dimensionality. *Phys. Rev. E*, 51(3):1791–1817, 1995.

[48] Marco Baiesi, Enzo Orlandini, and Attilio L Stella. Peculiar scaling of self-avoiding walk contacts. *Physical review letters*, 87(7):070602, 2001.

[49] Yoshinori Watanabe and Paul Nurse. Cohesin Rec8 is required for reductional chromosome segregation at meiosis. *Nature*, 400(6743):461, July 1999.

[50] Antonio Tedeschi, Marilena Ciciarello, Rosamaria Mangiacasale, Emanuele Roscioli, Wilhelmina M. Rensen, and Patrizia Lavia. RANBP1 localizes a subset of mitotic regulatory factors on spindle microtubules and regulates chromosome segregation in human cells. *Journal of Cell Science*, 120(21):3748–3761, November 2007.

[51] Stephen M. Hinshaw and Stephen C. Harrison. An Iml3-Chl4 Heterodimer Links the Core Centromere to Factors Required for Accurate Chromosome Segregation. *Cell Reports*, 5(1):29–36, October 2013.

[52] Rasmus B Jensen and Lucy Shapiro. Chromosome segregation during the prokaryotic cell division cycle. *Current Opinion in Cell Biology*, 11(6):726–731, December 1999.

[53] Jakob Møller-Jensen, Rasmus B Jensen, and Kenn Gerdes. Plasmid and chromosome segregation in prokaryotes. *Trends in Microbiology*, 8(7):313–320, July 2000.

[54] Suckjoon Jun and Andrew Wright. Entropy as the driver of chromosome segregation. *Nat. Rev. Microbiol.*, 8(8):600–607, 2010.

[55] Manfred Bohn and Dieter W. Heermann. Repulsive Forces Between Looping Chromosomes Induce Entropy-Driven Segregation. *PLOS ONE*, 6(1):e14428, January 2011.

[56] Axel Arnold and Suckjoon Jun. Time scale of entropic segregation of flexible polymers in confinement: Implications for chromosome segregation in filamentous bacteria. *Physical Review E*, 76(3):031901, September 2007.

[57] Ya Liu and Bulbul Chakraborty. Segregation of polymers in confined spaces. *Physical Biology*, 9(6):066005, November 2012.

[58] Danijela Horak, Slobodan Maletić, and Milan Rajković. Persistent homology of complex networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2009(03):P03034, March 2009.

[59] Noah Giansiracusa, Robert Giansiracusa, and Chul Moon. Persistent homology machine learning for fingerprint classification. pages 1–15, 2017.

[60] Herbert Edelsbrunner and Dmitriy Morozov. Persistent homology: theory and practice. Technical report, Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States), 2012.

[61] Kelin Xia and Guo-Wei Wei. Persistent homology analysis of protein structure, flexibility, and folding: PERSISTENT HOMOLOGY FOR PROTEIN. *International Journal for Numerical Methods in Biomedical Engineering*, 30(8):814–844, August 2014.

[62] Kevin Emmett, Benjamin Schweinhart, and Raul Rabadan. Multiscale Topology of Chromatin Folding. In *Proceedings of the 9th EAI International Conference on Bio-Inspired Information and Communications Technologies (Formerly BIONETICS)*, BICT'15, pages 177–180, ICST, Brussels, Belgium, Belgium, 2016. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

[63] A. Warshel and M. Levitt. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of Molecular Biology*, 103(2):227–249, May 1976.

[64] M. Levitt. Molecular dynamics of native protein. I. Computer simulation of trajectories. *Journal of molecular biology*, 168(3):595–617, August 1983.

[65] M. Karplus and J. Kuriyan. Molecular dynamics and protein function. *Proceedings of the National Academy of Sciences*, 102(19):6679–6685, May 2005.

[66] Scott A. Hollingsworth and Ron O. Dror. Molecular Dynamics Simulation for All. *Neuron*, 99(6):1129–1143, September 2018.

[67] Reuven Y Rubinstein and Dirk P Kroese. *Simulation and the Monte Carlo method*, volume 10. John Wiley & Sons, 2016.

[68] Kurt Binder and Dieter W Heermann. Theoretical foundations of the monte carlo method and its applications in statistical physics. In *Monte Carlo Simulation in Statistical Physics*, pages 5–67. Springer, 2010.

[69] Kurt Binder. *Monte Carlo and Molecular Dynamics Simulations in Polymer Science*. Oxford University Press, 1995.

[70] A Sokal. Monte carlo methods in statistical mechanics: foundations and new algorithms. In *Functional integration*, pages 131–192. Springer, 1997.

[71] Wolfgang Bruns. Angular correlations within coiled polymer molecules. *Macromolecular Theory and Simulations*, 1(5):287–293, 1992.

[72] Gert R. Strobl. *The Physics of Polymers: Concepts for Understanding Their Structures and Behavior*. Springer-Verlag, Berlin Heidelberg, 3 edition, 2007.

[73] P. G. de Gennes. Exponents for the excluded volume problem as derived by the Wilson method. *Physics Letters A*, 38(5):339–340, February 1972.

[74] A. J. Guttmann. On the critical behaviour of self-avoiding walks. *Journal of Physics A: Mathematical and General*, 20(7):1839–1854, May 1987.

[75] Bin Li, Neal Madras, and Alan D Sokal. Critical exponents, hyperscaling, and universal amplitude ratios for two-and three-dimensional self-avoiding walks. *Journal of Statistical Physics*, 80(3-4):661–754, 1995.

[76] Nathan Clisby. Accurate Estimate of the Critical Exponent $\ensuremath{\nu}$ for Self-Avoiding Walks via a Fast Implementation of the Pivot Algorithm. *Physical Review Letters*, 104(5):055702, February 2010.

[77] Cyril Domb. Self-Avoiding Walks on Lattices. In *Advances in Chemical Physics*, pages 229–259. John Wiley & Sons, Ltd, 2007.

[78] Jacob Mazur. Distribution function of the end-to-end distances of linear polymers with excluded volume effects. *J. Res. Natl Bur. Stand. A*, 69:355, 1965.

[79] C. Domb, J. Gillis, and G. Wilmers. On the shape and configuration of polymer molecules. *Proceedings of the Physical Society*, 85(4):625–645, April 1965.

[80] Michael E. Fisher. Shape of a Self-Avoiding Walk or Polymer Chain. *The Journal of Chemical Physics*, 44(2):616–622, January 1966.

[81] D. S. McKenzie and M. A. Moore. Shape of self-avoiding walk or polymer chain. *Journal of Physics A: General Physics*, 4(5):L82–L85, September 1971.

[82] Sam F. Edwards and Pooran Singh. Size of a polymer molecule in solution. Part 1.—Excluded volume problem. *Journal of the Chemical Society, Faraday Transactions 2: Molecular and Chemical Physics*, 75(0):1001–1019, 1979.

[83] R. Everaers, I. S. Graham, and M. J. Zuckermann. End-to-end distance distributions and asymptotic behaviour of self-avoiding walks in two and three dimensions. *Journal of Physics A: Mathematical and General*, 28(5):1271–1288, March 1995.

[84] Michael Rubinstein, Ralph H Colby, et al. *Polymer physics*, volume 23. Oxford university press New York, 2003.

[85] S. Havlin and D. Ben-Avraham. Theoretical and numerical study of fractal dimensionality in self-avoiding walks. *Physical Review A*, 26(3):1728–1734, September 1982.

[86] S. Havlin and D. Ben-Avraham. New approach to self-avoiding walks as a critical phenomenon. *Journal of Physics A: Mathematical and General*, 15(6):L321–L328, June 1982.

[87] M. Adam and M. Delsanti. Dynamical Properties of Polymer Solutions in Good Solvent by Rayleigh Scattering Experiments. *Macromolecules*, 10(6):1229–1237, November 1977.

[88] T. C. B. McLeish, J. Allgaier, D. K. Bick, G. Bishko, P. Biswas, R. Blackwell, B. Blottière, N. Clarke, B. Gibbs, D. J. Groves, A. Hakiki, R. K. Heenan, J. M. Johnson, R. Kant, D. J. Read, and R. N. Young. Dynamics of Entangled H-Polymers: Theory, Rheology, and Neutron-Scattering. *Macromolecules*, 32(20):6734–6758, October 1999.

[89] Alan D. Sokal. Monte Carlo methods for the self-avoiding walk. *Nuclear Physics B - Proceedings Supplements*, 47(1):172–179, March 1996.

[90] E. J. Janse van Rensburg. Monte Carlo methods for the self-avoiding walk. *Journal of Physics A: Mathematical and Theoretical*, 42(32):323001, July 2009.

[91] Peter Grassberger. Pruned-enriched rosenbluth method: Simulations of θ polymers of chain length up to 1 000 000. *Physical Review E*, 56(3):3682–3693, September 1997.

[92] Michael Bachmann and Wolfhard Janke. Multicanonical Chain-Growth Algorithm. *Physical Review Letters*, 91(20):208105, November 2003.

[93] Thomas Prellberg and Jarosław Krawczyk. Flat Histogram Version of the Pruned and Enriched Rosenbluth Method. *Physical Review Letters*, 92(12):120602, March 2004.

[94] NICOLAS COMBE, THIJS J. H. VLUGT, PIETER REIN TEN WOLDE, and DAAN FRENKEL. Dynamic pruned-enriched Rosenbluth method. *Molecular Physics*, 101(11):1675–1682, June 2003.

[95] Gerard T Barkema. Monte Carlo Simulation of Lattice Polymer Models.

[96] Neal Madras and Alan D. Sokal. Nonergodicity of local, length-conserving Monte Carlo algorithms for the self-avoiding walk. *Journal of Statistical Physics*, 47(3):573–595, May 1987.

[97] O. J. Heilmann and J. Rotne. Exact and Monte Carlo computations on a lattice model for change of conformation of a polymer. *Journal of Statistical Physics*, 27(1):19–35, January 1982.

[98] Peter H. Verdier. Monte Carlo Studies of Lattice-Model Polymer Chains. II. End-to-End Length. *The Journal of Chemical Physics*, 45(6):2122–2128, September 1966.

[99] M. Muthukumar. Monte Carlo renormalization group Calculations for polymers. *Journal of Statistical Physics*, 30(2):457–465, February 1983.

[100] Neal Madras and Alan D Sokal. The pivot algorithm: a highly efficient monte carlo method for the self-avoiding walk. *Journal of Statistical Physics*, 50(1-2):109–186, 1988.

[101] Bin Li, Neal Madras, and Alan D. Sokal. Critical exponents, hyperscaling, and universal amplitude ratios for two- and three-dimensional self-avoiding walks. *Journal of Statistical Physics*, 80(3):661–754, August 1995.

[102] Tom Kennedy. A faster implementation of the pivot algorithm for self-avoiding walks. *Journal of Statistical Physics*, 106(3-4):407–429, 2002.

[103] John E Szymanski. *Basic mathematics for electronic engineers: models and applications*, volume 16. Taylor & Francis, 1989.

[104] Nathan Clisby. Efficient Implementation of the Pivot Algorithm for Self-avoiding Walks. *Journal of Statistical Physics*, 140(2):349–392, July 2010.

[105] Sergio Caracciolo, Andrea Pelissetto, and Alan D. Sokal. Nonlocal Monte Carlo algorithm for self-avoiding walks with fixed endpoints. *Journal of Statistical Physics*, 60(1):1–53, July 1990.

[106] Sergio Caracciolo, Maria Serena Causo, Giovanni Ferraro, Mauro Papinutto, and Andrea Pelissetto. Bilocal Dynamics for Self-Avoiding Walks. *Journal of Statistical Physics*, 100(5):1111–1145, September 2000.

[107] Rolf Auhl, Ralf Everaers, Gary S. Grest, Kurt Kremer, and Steven J. Plimpton. Equilibration of long chain polymer melts in computer simulations. *The Journal of Chemical Physics*, 119(24):12718–12728, December 2003.

[108] Maria Serena Causo. Cut-and-Permute Algorithm for Self-Avoiding Walks in the Presence of Surfaces. *Journal of Statistical Physics*, 108(1):247–281, July 2002.

[109] I. Carmesin and Kurt Kremer. The bond fluctuation method: a new effective algorithm for the dynamics of polymers in all spatial dimensions. *Macromolecules*, 21(9):2819–2823, 09 1988.

[110] Kurt Kremer and Kurt Binder. Monte Carlo simulation of lattice models for macromolecules. *Computer Physics Reports*, 7(6):259–310, June 1988.

[111] H. J. Hilhorst and J. M. Deutch. Analysis of Monte Carlo results on the kinetics of lattice polymer chains with excluded volume. *The Journal of Chemical Physics*, 63(12):5153–5161, December 1975.

[112] Kurt Kremer and Gary S. Grest. Simulations for structural and dynamic properties of dense polymer systems. *Journal of the Chemical Society, Faraday Transactions*, 88(13):1707–1717, 1992.

[113] J. P. Wittmer, P. Beckrich, H. Meyer, A. Cavallo, A. Johner, and J. Baschnagel. Intramolecular long-range correlations in polymer melts: The segmental size distribution and its moments. *Physical Review E*, 76(1):011803, July 2007.

[114] Hsiao-Ping Hsu and Wolfgang Paul. A fast monte carlo algorithm for studying bottle-brush polymers. *Computer Physics Communications*, 182(10):2115–2121, 2011.

[115] J. Baschnagel, J. P. Wittmer, and H. Meyer. Monte Carlo SImulation of Polymers: Coarse-Grained Models. *arXiv:cond-mat/0407717*, July 2004.

[116] Paul J Hagerman. Flexibility of dna. *Annual review of biophysics and biophysical chemistry*, 17(1):265–286, 1988.

[117] Chongli Yuan, Huimin Chen, Xiong Wen Lou, and Lynden A Archer. Dna bending stiffness on small length scales. *Physical review letters*, 100(1):018102, 2008.

[118] Arun Yethiraj. Monte Carlo simulation of confined semiflexible polymer melts. *The Journal of Chemical Physics*, 101(3):2489–2497, August 1994.

[119] Iliya Kusner and Simcha Srebnik. Conformational behavior of semi-flexible polymers confined to a cylindrical surface. *Chemical Physics Letters*, 430(1):84–88, October 2006.

[120] Avinoam Ben-Shaul. Entropy, Energy, and Bending of DNA in Viral Capsids. *Biophysical Journal*, 104(10):L15–L17, May 2013.

[121] Johan RC Van der Maarel. Polymer conformation. In *Introduction to Biopolymer Physics*, pages 23–53. WORLD SCIENTIFIC, December 2007.

[122] Alexei R. Khokhlov, Alexander Yu Grosberg, and Vijay S. Pande. *Statistical Physics of Macromolecules*. Polymers and Complex Materials. AIP-Press, 1994.

[123] Jan Skov Pedersen and Peter Schurtenberger. Scattering Functions of Semiflexible Polymers with and without Excluded Volume Effects. *Macromolecules*, 29(23):7602–7612, January 1996.

[124] Xinghua Zhang, Ying Jiang, Bing Miao, Yunlin Chen, Dadong Yan, and Jeff Z. Y. Chen. The structure factor of a wormlike chain and the random-phase-approximation solution for the spinodal line of a diblock copolymer melt. *Soft Matter*, 10(29):5405–5416, July 2014.

[125] Lothar Schäfer, Andrea Ostendorf, and Johannes Hager. Scaling of the correlations among segment directions of a self-repelling polymer chain. *Journal of Physics A: Mathematical and General*, 32(45):7875–7899, October 1999.

[126] Hsiao-Ping Hsu, Wolfgang Paul, and Kurt Binder. Estimation of persistence lengths of semiflexible polymers: Insight from simulations. *Polymer Science Series C*, 55(1):39–59, 2013.

[127] J. P. Wittmer, H. Meyer, J. Baschnagel, A. Johner, S. Obukhov, L. Mattioni, M. Müller, and A. N. Semenov. Long Range Bond-Bond Correlations in Dense Polymer Solutions. *Physical Review Letters*, 93(14):147801, September 2004.

[128] David Shirvanyants, Sergey Panyukov, Qi Liao, and Michael Rubinstein. Long-Range Correlations in a Polymer Chain Due to Its Connectivity. *Macromolecules*, 41(4):1475–1485, February 2008.

[129] Hiromi Yamakawa. *Modern Theory of Polymer Solutions*. Harper & Row, 1971.

[130] L. Schäfer and K. Elsner. Calculation of the persistence length of a flexible polymer chain with short-range self-repulsion. *The European Physical Journal E*, 13(3):225–237, March 2004.

[131] Hsiao-Ping Hsu, Kurt Binder, and Wolfgang Paul. How to Define Variation of Physical Properties Normal to an Undulating One-Dimensional Object. *Physical Review Letters*, 103(19):198301, November 2009.

[132] Hsiao-Ping Hsu, Wolfgang Paul, and Kurt Binder. Standard definitions of persistence length do not describe the local "intrinsic" stiffness of real polymer chains. *Macromolecules*, 43(6):3094–3102, 2010.

[133] Hsiao-Ping Hsu. Monte carlo simulations of lattice models for single polymer systems. *The Journal of chemical physics*, 141(16):164903, 2014.

[134] Jing-Zi Zhang, Xiang-Yao Peng, Shan Liu, Bang-Ping Jiang, Shi-Chen Ji, and Xing-Can Shen. The Persistence Length of Semiflexible Polymers in Lattice Monte Carlo Simulations. *Polymers*, 11(2), February 2019.

[135] Jiying Jia, Kunhe Li, Andreas Hofmann, and Dieter W Heermann. The effect of bending rigidity on polymers. *Macromolecular Theory and Simulations*, page 1800071, 2019.

[136] Bae-Yeun Ha and Youngkyun Jung. Polymers under confinement: Single polymers, how they interact, and as model chromosomes. *Soft Matter*, 11(12):2333–2352, March 2015.

[137] Theo Odijk. The statistics and dynamics of confined or entangled stiff polymers. *Macromolecules*, 16(8):1340–1344, August 1983.

[138] Youngkyun Jung, Chanil Jeon, Juin Kim, Hawoong Jeong, Suckjoon Jun, and Bae-Yeun Ha. Ring polymers as model bacterial chromosomes: Confinement, chain topology, single chain statistics, and how they interact. *Soft Matter*, 8(7):2095–2102, 2012.

[139] Elena Minina and Axel Arnold. Entropic Segregation of Ring Polymers in Cylindrical Confinement. *Macromolecules*, 48(14):4998–5005, 2015.

[140] James C Wang. Dna topoisomerases. *Annual review of biochemistry*, 54(1):665–697, 1985.

[141] Allyn J. Schoeffler and James M. Berger. DNA topoisomerases: Harnessing and constraining energy to govern chromosome topology. *Quarterly Reviews of Biophysics*, 41(1):41–101, February 2008.

[142] Kevin W Plaxco, Kim T Simons, and David Baker. Contact order, transition state placement and the refolding rates of single domain proteins. *Journal of molecular biology*, 277(4):985–994, 1998.

[143] Kim Sneppen and Giovanni Zocchi. *Physics in Molecular Biology*. Cambridge University Press, August 2005.

[144] Peter Freyd, David Yetter, Jim Hoste, WB Raymond Lickorish, Kenneth Millett, and Adrian Ocneanu. A new polynomial invariant of knots and links. *Bulletin of the American Mathematical Society*, 12(2):239–246, 1985.

[145] JPJ Michels and FW Wiegel. On the topology of a polymer ring. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 403(1825):269–284, 1986.

[146] E. J. Janse van Rensburg, D. A. W. Sumners, E. Wasserman, and S. G. Whittington. Entanglement complexity of self-avoiding walks. *Journal of Physics A: Mathematical and General*, 25(24):6557–6566, December 1992.

[147] Dimos Goundaroulis, Neslihan Gügümcü, Sofia Lambropoulou, Julien Dorier, Andrzej Stasiak, and Louis Kauffman. Topological models for open-knotted protein chains using the concepts of knotoids and bonded knotoids. *Polymers*, 9(9):444, 2017.

[148] E. Orlandini, M. C. Tesi, S. G. Whittington, D. W. Sumners, and E. J. Janse van Rensburg. The writhe of a self-avoiding walk. *Journal of Physics A: Mathematical and General*, 27(10):L333–L338, May 1994.

[149] Michael Brill, Philipp M. Diesinger, and Dieter W. Heermann. Knots in Macromolecules in Constraint Space. *arXiv:cond-mat/0507020*, July 2005.

[150] Konstantin Klenin and Jörg Langowski. Computation of writhe in modeling of supercoiled dna. *Biopolymers*, 54(5):307–317, 2000.

[151] Yuanan Diao, Akos Dobay, Robert B Kusner, K Millett, and Andrzej Stasiak. The average crossing number of equilateral random polygons. *Journal of Physics A: Mathematical and General*, 36(46):11561, 2003.

[152] PM Diesinger and DW Heermann. Average crossing number of gaussian and equilateral chains with and without excluded volume. *The European Physical Journal B-Condensed Matter and Complex Systems*, 62(2):209–214, 2008.

[153] DC Rapaport. On the polymer phase transition. *Physics Letters A*, 48(5):339–340, 1974.

[154] J Des Cloizeaux. Short range correlation between elements of a long polymer in a good solvent. *Journal de Physique*, 41(3):223–238, 1980.

[155] S Redner. Distribution functions in the interior of polymer chains. *Journal of Physics A: Mathematical and General*, 13(11):3525, 1980.

[156] Bertrand Duplantier. Exact contact critical exponents of a self-avoiding polymer chain in two dimensions. *Physical Review B*, 35(10):5290–5293, April 1987.

[157] Marquerite Lautout-Magat. Number of contacts between distant segments of a polymer chain. a solution of the problem of self-avoiding random walk. *Macromolecules*, 10(6):1375–1381, 1977.

[158] Job Dekker. A closer look at long-range chromosomal interactions. *Trends Biochem Sci*, 28(6):277–80, Jun 2003.

[159] Erez Lieberman-Aiden, Nynke L van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragoczy, Agnes Telling, Ido Amit, Bryan R Lajoie, Peter J Sabo, Michael O Dorschner, Richard Sandstrom, Bradley Bernstein, M A Bender, Mark Groudine, Andreas Gnirke, John Stamatoyannopoulos, Leonid A Mirny, Eric S Lander, and Job Dekker. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326(5950):289–93, Oct 2009.

[160] G.M. Cooper. *The Cell: A Molecular Approach.*, volume Available from: https://www.ncbi.nlm.nih.gov/books/NBK9863/. 2nd edition. Sunderland (MA): Sinauer Associates, 2000.

[161] Dieter W Heermann. Physical nuclear organization: loops and entropy. *Curr Opin Cell Biol*, 23(3):332–7, Jun 2011.

[162] Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter. Chromosomal dna and its packaging in the chromatin fiber. 2002.

[163] H Schiessel, W M Gelbart, and R Bruinsma. Dna folding: structural and mechanical properties of the two-angle model for chromatin. *Biophys J*, 80(4):1940–56, Apr 2001.

[164] Manfred Bohn, Dieter W Heermann, and Roel van Driel. Random loop model for long polymers. *Phys Rev E Stat Nonlin Soft Matter Phys*, 76(5 Pt 1):051805, 2007 Nov.

[165] Sandra Goetze, Julio Mateos-Langerak, Hinco J Gierman, Wim de Leeuw, Osdilly Giromus, Mireille H G Indemans, Jan Koster, Vladan Ondrej, Rogier Versteeg, and Roel van Driel. The three-dimensional structure of human interphase chromosomes is related to the transcriptome map. *Mol Cell Biol*, 27(12):4475–87, Jun 2007.

[166] T Cremer and C Cremer. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet*, 2(4):292–301, Apr 2001.

[167] Anthony H B de Vries, Bea E Krenn, Roel van Driel, Vinod Subramaniam, and Johannes S Kanger. Direct observation of nanomechanical properties of chromatin in living cells. *Nano Lett*, 7(5):1424–1427, 2007 May.

[168] Vladimir B. Teif and Karsten Rippe. Predicting nucleosome positions on the dna: combining intrinsic sequence preferences and remodeler activities. *Nucleic Acids Research*, 37(17):5641–5655, 09 2009.

[169] Suckjoon Jun, Axel Arnold, and Bae-Yeun Ha. Confined space and effective interactions of multiple self-avoiding chains. *Phys Rev Lett*, 98(12):128303, Mar 2007.

[170] Eran Segal, Yvonne Fondufe-Mittendorf, Lingyi Chen, AnnChristine Thåström, Yair Field, Irene K Moore, Ji-Ping Z Wang, and Jonathan Widom. A genomic code for nucleosome positioning. *Nature*, 442(7104):772–778, 2006.

[171] Gernot Längst, Vladimir B. Teif, and Karsten Rippe. *Chromatin Remodeling and Nucleosome Positioning*, pages 111–138. Wiley-VCH Verlag GmbH & Co. KGaA, 2011.

[172] Kevin Struhl and Eran Segal. Determinants of nucleosome positioning. *Nature structural & molecular biology*, 20(3):267–273, 2013.

[173] Ju Yeon Lee and Tae-Hee Lee. Effects of histone acetylation and cpg methylation on the structure of nucleosomes. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 1824(8):974–982, 2012.

[174] Jörg Langowski and Dieter W. Heermann. Computational modeling of the chromatin fiber. *Seminars in Cell & Developmental Biology*, 18(5):659–667, 10 2007.

[175] Philipp M Diesinger, Susanne Kunkel, Jörg Langowski, and Dieter W Heermann. Histone depletion facilitates chromatin loops on the kilobasepair scale. *Biophysical Journal*, 99(9):2995–3001, 11 2010.

[176] Cédric Vaillant, Benjamin Audit, C Thermes, and Alain Arnéodo. Formation and positioning of nucleosomes: Effect of sequence-dependent long-range correlated structural disorder. *The European Physical Journal E: Soft Matter and Biological Physics*, 19(3):263–277, 2006.

[177] A Ja Varshavsky, VV Bakayev, and GP Georgiev. Heterogeneity of chromatin subunits in vitro and location of histone h1. *Nucleic acids research*, 3(2):477–492, 1976.

[178] D Bensimon, D Dohmi, and M Mezard. Stretching a heteropolymer. *EPL (Europhysics Letters)*, 42(1):97, 1998.

[179] D Bratko, AK Chakraborty, and EI Shakhnovich. The structure of a random heteropolymer in a disordered medium: Ensemble growth simulation. *The Journal of chemical physics*, 106(3):1264–1279, 1997.

[180] AN Semenov. Bond- vector correlation functions in dense polymer systems. *Macromolecules*, 43(21):9139–9154, 2010.

[181] David Shirvanyants, Sergey Panyukov, Qi Liao, and Michael Rubinstein. Long-range correlations in a polymer chain due to its connectivity. *Macromolecules*, 41(4):1475–1485, 2008.

[182] Giuseppe D'Adamo and Andrea Pelissetto. Polymer models with optimal good-solvent behavior. *Journal of Physics: Condensed Matter*, 29(43):435104, 2017.

[183] Guanghui Ping, Guoliang Yang, and Jian-Min Yuan. Depletion force from macromolecular crowding enhances mechanical stability of protein molecules. *Polymer*, 47(7):2564–2570, 2006.

[184] Eric Horwath, Nathan Clisby, and Peter Virnau. Knots in finite memory walks. In *Journal of Physics: Conference Series*, volume 750, page 012010. IOP Publishing, 2016.

[185] Wolfhard Janke. Statistical analysis of simulations: Data correlations and error estimation. *Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms*, 10:423–445, 2002.

[186] Hsiao-Ping Hsu and Kurt Binder. Stretching semiflexible polymer chains: Evidence for the importance of excluded volume effects from monte carlo simulation. *The Journal of chemical physics*, 136(2):024901, 2012.

[187] Peter Cifra and Tomas Bleha. Simulation of chain organization in encapsulated polymers. In *Macromolecular symposia*, volume 296, pages 336–341. Wiley Online Library, 2010.

[188] P Cifra and T Bleha. Shape transition of semi-flexible macromolecules confined in channel and cavity. *The European Physical Journal E*, 32(3):273–279, 2010.

[189] Pablo Vázquez-Montejo, Zachary McDargh, Markus Deserno, and Jemal Guven. Cylindrical confinement of semiflexible polymers. *Physical Review E*, 91(6):063203, 2015.

[190] Andrea M Chiariello, Carlo Annunziatella, Simona Bianco, Andrea Esposito, and Mario Nicodemi. Polymer physics of chromosome large-scale 3d organisation. *Scientific reports*, 6:29775, 2016.

[191] Marian Walhout, Marc Vidal, and Job Dekker. *Handbook of systems biology: concepts and insights*. Academic Press, 2012.

[192] Erik Werner, Fredrik Persson, Fredrik Westerlund, Jonas O Tegenfeldt, and Bernhard Mehlig. Orientational correlations in confined dna. *Physical Review E*, 86(4):041802, 2012.

[193] Peter Cifra, Zuzana Benková, and Tomáš Bleha. Effect of confinement on properties of stiff biological macromolecules. *Faraday discussions*, 139:377–392, 2008.

[194] Caroline Uhler and GV Shivashankar. Geometric control and modeling of genome reprogramming. *Bioarchitecture*, 6(4):76–84, 2016.

[195] James M Polson and Logan G Montgomery. Polymer segregation under confinement: Free energy calculations and segregation dynamics simulations. *The Journal of chemical physics*, 141(16):164902, 2014.

[196] A Baumgärtner and M Muthukumar. Disentangling of two intertwined chains. *The Journal of chemical physics*, 84(1):440–443, 1986.

[197] Diddo Diddens, Nam-Kyung Lee, Sergei Obukhov, Jorg Baschnagel, and Albert Johner. Disentanglement of two single polymer chains: Contacts and knots. *ACS Macro Letters*, 5(6):740–744, 2016.

[198] Manfred Bohn and Dieter W Heermann. Conformational properties of compact polymers. *The Journal of chemical physics*, 130(17):174901, 2009.

[199] Marc L Mansfield. Unbiased sampling of lattice hamilton path ensembles. *The Journal of chemical physics*, 125(15):154103, 2006.

[200] S Müller and L Schäfer. On the number of intersections of self-repelling polymer chains. *The European Physical Journal B-Condensed Matter and Complex Systems*, 2(3):351–369, 1998.

[201] Werner Schweika. *Disordered Alloys: Diffuse Scattering and Monte Carlo Simulations*. Springer Science & Business Media, 1998.

[202] Khwaja Yaldram and Kurt Binder. Unmixing of binary alloys by a vacancy mechanism of diffusion: a computer simulation. *Zeitschrift für Physik B Condensed Matter*, 82(3):405–418, 1991.

[203] J-M. Victor, J-B. Imbert, and D. Lhuillier. The number of contacts in a self-avoiding walk of variable radius of gyration in two and three dimensions. *The Journal of Chemical Physics*, 100(7):5372–5377, April 1994.

[204] Gregory F. Lawler. A self-avoiding random walk. *Duke Mathematical Journal*, 47(3):655–693, September 1980.

[205] Gregory F. Lawler. Loop-erased self-avoiding random walk in two and three dimensions. *Journal of Statistical Physics*, 50(1-2):91–108, January 1988.

[206] Philippe Marchal. Loop-Erased Random Walks, Spanning Trees and Hamiltonian Cycles. *Electronic Communications in Probability*, 5:39–50, 2000.

[207] Gregory F. Lawler. The Laplacian-b random walk and the Schramm-Loewner evolution. *Illinois Journal of Mathematics*, 50(1-4):701–746, 2006.

[208] Scaling Limits of Loop-Erased Random Walks and Uniform Spanning Trees. Selected Works in Probability and Statistics. New York, NY.

[209] A. J. Guttmann and R. J. Bursill. Critical exponent for the loop erased self-avoiding walk by Monte Carlo methods. *Journal of Statistical Physics*, 59(1-2):1–9, April 1990.

[210] R. E. Bradley and S. Windwer. Dynamic algorithm for loop-erased self-avoiding random walks in two and three dimensions. *Physical Review E*, 51(1):241–244, January 1995.

[211] David B. Wilson. The dimension of loop-erased random walk in 3D. *Physical Review E*, 82(6):062102, December 2010.

[212] Himanshu Agrawal and Deepak Dhar. Distribution of sizes of erased loops of loop-erased random walks in two and three dimensions. *Physical Review E*, 63(5):056115, April 2001.

[213] D. W. Sumners and S. G. Whittington. Knots in self-avoiding walks. *Journal of Physics A: Mathematical and General*, 21(7):1689–1694, April 1988.

[214] Javier Arsuaga, Mariel Vazquez, Paul McGuirk, Sonia Trigueros, De Witt Sumners, and Joaquim Roca. DNA knots reveal a chiral organization of DNA in phage capsids. *Proceedings of the National Academy of Sciences*, 102(26):9165–9169, June 2005.

[215] Xiaoyan R. Bao, Heun Jin Lee, and Stephen R. Quake. Behavior of Complex Knots in Single DNA Molecules. *Physical Review Letters*, 91(26):265506, December 2003.

[216] Peter Bubenik and Peter T. Kim. A statistical approach to persistent homology. pages 1–34, 2006.

[217] Ulderico Fugacci, Sara Scaramuccia, Federico Iuricich, and Leila De Floriani. Persistent homology: a step-by-step introduction for newcomers. *STAG: Smart Tools and Apps in Computer Graphics*, (October):1—-10, 2016.

[218] Christoph Hofer, Roland Kwitt, Marc Niethammer, and Andreas Uhl. Deep Learning with Topological Signatures. pages 1–14, 2017.

[219] Vidit Nanda and Radmila Sazdanovi. *Discrete and Topological Models in Molecular Biology*. 2014.

[220] Hyung-Jun Im, Jarang Hahm, Hyejin Kang, Hongyoon Choi, Hyekyoung Lee, Do Won Hwang, E. Edmund Kim, June-Key Chung, and Dong Soo Lee. Disrupted brain metabolic connectivity in a 6-OHDA-induced mouse model of Parkinson's disease examined using persistent homology-based analysis. *Scientific Reports*, 6:33875, September 2016.

[221] H. Lee, H. Kang, M. K. Chung, B. Kim, and D. S. Lee. Persistent Brain Network Homology From the Perspective of Dendrogram. *IEEE Transactions on Medical Imaging*, 31(12):2267–2277, December 2012.

[222] Vanessa Robins. *Computational Topology at Multiple Resolutions: Foundations and Applications to Fractals and Dynamics*. PhD thesis, Boulder, CO, USA, 2000. AAI9979393.

[223] Robert MacPherson and Benjamin Schweinhart. Measuring shape with topology. *Journal of Mathematical Physics*, 53(7):1–19, 2012.

[224] Gabriell Máté and Dieter W. Heermann. Persistence intervals of fractals. *Physica A: Statistical Mechanics and its Applications*, 405:252–259, 2014.

[225] Henry Adams, Manuchehr Aminian, Elin Farnell, Michael Kirby, Chris Peterson, Joshua Mirth, Rachel Neville, Patrick Shipman, and Clayton Shonkwiler. A fractal dimension for measures via persistent homology. pages 1–25, 2018.

[226] Benjamin Schweinhart. The Persistent Homology of Random Geometric Complexes on Fractals. *arXiv Computational Geometry*, (January 2019):1–37.

[227] Benjamin Schweinhart. Persistent Homology and the Upper Box Dimension. (March):1–25, 2018.

[228] B. Dubuc, J. F. Quiniou, C. Roques-Carmes, C. Tricot, and S. W. Zucker. Evaluating the fractal dimension of profiles. *Physical Review A*, 39(3):1500–1512, February 1989.

[229] Jessica Reichert, André R. Backes, Patrick Schubert, and Thomas Wilke. The power of 3D fractal dimensions for comparative shape and structural complexity analyses of irregularly shaped organisms. *Methods in Ecology and Evolution*, 8(12):1650–1658, 2017.

[230] Richard F Voss. Characterization and Measurement of Random Fractals. *Physica Scripta*, T13:27–32, January 1986.

[231] Heinz-Otto Peitgen and Dietmar Saupe, editors. *The Science of Fractal Images*. Springer-Verlag, New York, 1988.

[232] James Theiler. Estimating fractal dimension. *Journal of the Optical Society of America A*, 7(6):1055, 2008.

[233] Yves Termonia and Zeev Alexandrowicz. Fractal Dimension of Strange Attractors from Radius versus Size of Arbitrary Clusters. *Physical Review Letters*, 51(14):1265–1268, October 1983.

[234] Remo Radii and Antonio Politi. Statistical description of chaotic attractors: The dimension function. *Journal of Statistical Physics*, 40(5):725–750, September 1985.

[235] P. Grassberger, R. Badii, and A. Politi. Scaling laws for invariant measures on hyperbolic and nonhyperbolic atractors. *Journal of Statistical Physics*, 51(1):135–178, April 1988.

[236] R. Badii and G. Broggi. Measurement of the dimension spectrum $f(\alpha)$: Fixed-mass approach. *Physics Letters A*, 131(6):339–343, August 1988.

[237] A. Kruger. Implementation of a fast box-counting algorithm. *Computer Physics Communications*, 98(1-2):224–234, October 1996.

[238] André Ricardo Backes, Danilo Medeiros Eler, Rosane Minghim, and Odemir Martinez Bruno. Characterizing 3D shapes using fractal dimension. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6419 LNCS:14–21, 2010.

[239] Y. Kamer, G. Ouillon, and D. Sornette. Barycentric fixed-mass method for multifractal analysis. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 88(2):1–9, 2013.

[240] André Ricardo Backes and Odemir Martinez Bruno. Fractal and Multi-Scale Fractal Dimension analysis: a comparative study of Bouligand-Minkowski method. (January), 2012.

[241] Andreas Hofmann, Matthias Krufczik, Dieter W. Heermann, and Michael Hausmann. Using Persistent Homology as a New Approach for Super-Resolution Localization Microscopy Data Analysis and Classification of γH2AX Foci/Clusters. *International Journal of Molecular Sciences*, 19(8), August 2018.

[242] Nina Otter, Mason A. Porter, Ulrike Tillmann, Peter Grindrod, and Heather A. Harrington. A roadmap for the computation of persistent homology. *EPJ Data Science*, 6(1):17, December 2017.

[243] A gentle introduction to persistent homology · Christian Bock. https://christian.bock.ml/posts/persistent_homology/.

[244] The writhe of open and closed curves. *Journal of Physics A: Mathematical and General*, 39(26):8321–8348, 2006.

[245] E. Panagiotou, K. C. Millett, and S. Lambropoulou. The linking number and the writhe of uniform random walks and polygons in confined spaces. *Journal of Physics A: Mathematical and Theoretical*, 43(4):045208, January 2010.

[246] Vincent Récamier, Ignacio Izeddin, Lana Bosanac, Maxime Dahan, Florence Proux, and Xavier Darzacq. Single cell correlation fractal dimension of chromatin. *Nucleus*, 5(1):75–84, January 2014.

[247] Christophe Lavelle, Jean-Marc Victor, and Jordanka Zlatanova. Chromatin fiber dynamics under tension and torsion. *International Journal of Molecular Sciences*, 11(4):1557–1579, April 2010.

[248] Rosana Collepardo-Guevara and Tamar Schlick. The effect of linker histone's nucleosome binding affinity on chromatin unfolding mechanisms. *Biophysical Journal*, 101(7):1670–1680, October 2011.

[249] Sergei A. Grigoryev, Gaurav Arya, Sarah Correll, Christopher L. Woodcock, and Tamar Schlick. Evidence for heteromorphic chromatin fibers from analysis of nucleosome interactions. *Proceedings of the National Academy of Sciences*, 106(32):13317–13322, August 2009.

[250] D. V. Lebedev, M. V. Filatov, A. I. Kuklin, A. Kh Islamov, E. Kentzinger, R. Pantina, B. P. Toperverg, and V. V. Isaev-Ivanov. Fractal nature of chromatin organization in interphase chicken erythrocyte nuclei: DNA structure exhibits biphasic fractal properties. *FEBS Letters*, 579(6):1465–1468, 2005.

[251] Konradin Metze. Fractal dimension of chromatin and cancer prognosis. *Epigenomics*, 2(5):601–604, October 2010.

[252] Konradin Metze. Fractal dimension of chromatin: Potential molecular diagnostic applications for cancer prognosis. *Expert Review of Molecular Diagnostics*, 13(7):719–735, September 2013.