# Computing the spatial organization and dynamics of chromatin domains

*Submitted in partial fulfillment of the requirements of
the degree of
Doctor of Philosophy
of the
Indian Institute of Technology Bombay, India
and
Monash University, Australia*

*submitted by*

**Kiran Kumari**

The course of study for this award was developed jointly by
Monash University, Australia and the Indian Institute of Technology Bombay, India
and was given academic recognition by each of them.
The programme was administrated by The IITB-Monash Research Academy in
Indian Institute of Technology Bombay
Mumbai 400076 (India)

2021

# Copyright notice

## Abstract

The *state* of chromatin is not only determined by the protein arrangement along the chromatin contour but also by its 3D organization inside the nucleus. Packaging of DNA into chromatin requires high compaction, while reading the genetic code requires accessibility. How cell tackles these seemingly contradictory requirements is a puzzle. Recent advances in chromosome conformation capture experiments (3C, 4C, Hi-C etc.) provide partial information on the chromatin organization in a cell population, namely the contact count between any segment pairs. However, these experiments do not provide the interaction strengths that lead to these contact counts. Hence, given the contact matrix, determining the complete 3D organization of the whole chromatin polymer is an inverse problem. In this thesis, a novel Inverse Brownian Dynamics (IBD) method based on a coarse-grained bead-spring chain model has been proposed to compute the optimal interaction strengths between different segments of chromatin such that the experimentally measured contact probability constraints are satisfied.

We model the chromatin polymer as a bead-spring chain. The crosslinking between the beads (chromatin segments) is often represented by soft non-bonded potential such as Lennard-Jones potentials. We argue that this is a better strategy to study biopolymers like chromatin having crosslinks with a diverse range of interaction strengths and propensity to break and reform. We study the crosslinked single polymer using a Soddemann-Duenweg-Kremer (SDK) potential, which has a parameter that can vary the attractive interaction strength without affecting the repulsive nature, and a parameter that will introduce a natural cut-off in the attractive interaction strength. Through extensive Brownian Dynamics simulations with hydrodynamic interactions, we investigate parameters that satisfy the scaling behaviour in all solvent regimes. Simulating crosslinks as bead-pairs bonded by the SDK potential and applying the IBD method to the $\alpha$-globin gene locus in two different cell types, we predict the 3D organizations corresponding to active and repressed states of chromatin at the locus. We show that the average distance between any two segments of the region has a broad distribution and cannot be computed as a simple inverse relation based on the contact probability alone. The results presented for multiple normalization methods suggest that all measurable quantities may crucially depend on the

nature of normalization. We argue that by experimentally measuring predicted quantities, one may infer the appropriate form of normalization.

Chromatin is observed to be organized into multiple domains of varying sizes and compaction. While these domains are often imagined as static structures, they are highly dynamic and show cell-to-cell variability. Since processes such as gene regulation and DNA replication occur in the context of these domains, it is important to understand their organization, fluctuation and dynamics. To quantify how epigenetic changes can alter the spatio-temporal nature of the domains, we perturbed the interaction strengths systematically. Computing distance-distributions and relaxation times for different chromatin states, we show that weak and strong interactions cooperatively determine the organization of the domains and how the solid-like and liquid-like nature of chromatin gets altered as we vary epigenetic states. Quantifying the dynamics of chromatin segments within a domain, we show how the competition between polymer entropy and interaction energy influence the timescales of loop formation and maintenance of stable loops. Our results suggest that chromatin cannot be completely described by its average properties; the statistical distributions of configurations are crucial. Analysis of chromatin fluctuations implies that, as chromatin folds, the domains get more stiff as well as experience more viscous drag. Our study of chromatin dynamics finds power-law distribution of loop formation times indicating multiple timescales.

## Declaration

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

The core theme of the thesis is understanding the spatial organization and dynamics of chromatin domains. The ideas, development and writing up of all the papers in the thesis were the principal responsibility of myself, working within the Department of Biosciences and Bioengineering (IIT Bombay) under the supervision of Prof. Ranjith Padinhateeri and within the Department of Chemical Engineering (Monash University) under the supervision of Prof. J. Ravi Prakash. This thesis includes the following two original paper published in peer-reviewed journals and one submitted manuscripts.

1. K. Kumari, J. R. Prakash, R. Padinhateeri, *Spatiotemporal organization of chromatin domains: role of interaction energy and polymer entropy* (under review), bioRxiv (2021), doi: 10.1101/2021.02.17.431616.

2. K. Kumari, B. Duenweg, R. Padinhateeri, J. R. Prakash, *Computing 3D chromatin configurations from contact probability maps by Inverse Brownian Dynamics*, Biophys J. 118: 2193-2208 (2020), doi: 10.1016/j.bpj.2020.02.017.

3. A. Santra, K. Kumari, R. Padinhateeri, B. Duenweg, J. R. Prakash, *Universality of the collapse transition of sticky polymers*, Soft Matter, 15, 7876-7887 (2019), doi: 10.1039/C9SM01361J.

_____                                                          Wednesday 30th July, 2021

Kiran Kumari

IIT Bombay Roll no. 164304001

Monash ID no. 28285271

# Acknowledgement

My journey till here is not mine alone. Apart from my choices, other people, their actions and circumstances brought me to the point where I am today. I want to express my gratitude towards everyone and everything in the universe for making me achieve this.

Firstly, let me express my gratitude towards my thesis advisors, Prof. Ranjith Padinhateeri, and Prof. J. Ravi Prakash who supervised and inspired me throughout the research. They have always supported me during the highs and lows of my Ph.D. journey. I have learnt a lot from them, not only about research but also how to be a good human being and respect others. I am extremely thankful to Prof. Burkhard Duenweg and Prof. P. Sunthar for guiding me and being available whenever I needed their advice. I would also like to thank my research progress committee (RPC) members, Prof. Dibyendu Das, Prof. Rajarshi Chakrabarti and Prof. Chen Davidovich. They provided constructive suggestions during the annual progress seminars (APS) and the pre-synopsis seminar.

I gratefully acknowledge the computational resources provided by the MonARCH and SpaceTime of Monash University and IIT Bombay, respectively. I also acknowledge the funding and general support received from the IITB-Monash Research Academy, DST, SERB and DBT India.

I would like to take this opportunity to thank the members of Molecular Rheology group and the Physical Biology group for patiently listening to my mock presentations and providing me valuable feedback. Specifically, I would specially like to thank Kailasham, Aritra, Gaurav and Rakesh for being collegues with whom I could discuss nitty-gritties of the research.

# Contents

# Chapter 1

# Introduction

All living organisms are made of cells, and every cell contains a long semi-flexible polymer known as deoxyribonucleic acid (DNA). It contains the recipe—the genetic information—to produce the entire material necessary to make and maintain a cell. DNA is made up of four types of nucleotide subunits (also, "bases"), namely adenine (A), cytosine (C), guanine (G) and thymine (T), and exists in cells as a double-stranded molecule having A-T and G-C paired bases as shown in Fig. 1.1. The sequence of these four nucleotides acts like a code (the genetic code) for all cellular processes. This code is inherited from each cell to its daughter cell as the cells divide. The total length of DNA in a typical human cell is $\approx$ two meters and is packed into a small volume ($\approx 10\mu\text{m}^3$) of the cell nucleus [2, 41]. All the DNA in a human cell does not exist as one single long piece of polymer; rather, they are 46 pieces of polymers, each having a length of several centimetres, which collectively span over 2m ($6 \times 10^9$ base pairs (bp)).

Even though all the cells in a multi-cellular organism have the same DNA sequence, they function differently based on the cell type. For example, the phenotype of a skin cell is significantly different from that of a neuronal cell. In a given cell type, only certain parts of the DNA will be read to produce proteins, while other parts will be highly folded and inaccessible to read. In another cell type, some other parts will be accessible, and only those genes will be read. That is, each cell type effectively reads different parts of the DNA (hence, effectively, different genetic code) due to its packaging and 3D organization [75, 56, 48, 17, 97, 137]. Therefore, reading of the genes (and producing genetic material) should be done in a regulated manner to maintain cell fate and homeostasis.

How is the regulated reading of genes achieved? On the one hand, the DNA is required
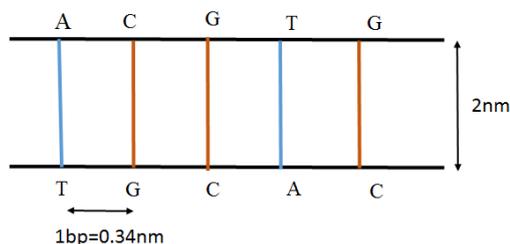
**Figure 1.1:** Schematic representation of the DNA with its nucleotide sequence. Distance between consecutive nucleotides is 0.34nm. The width of the DNA is 2nm.

to fold and condense to accommodate it in the tiny space of the nucleus; on the other hand, the reading of the genetic code for cell functionality requires regulated access to the DNA. How cell tackles these seemingly contradictory requirements is a puzzle. DNA achieves this challenging task of regulated packing of the genome with the help of a number of specialised proteins present inside the cell, and it results in a higher-order structure known as chromatin. That is, in cells, the accessibility of the genetic code is controlled by covering it with a large number of proteins, folding it further and packaging it into a 3-dimensional structure known as chromatin. It must be noted that every cellular process occurs in the context of the chromatin polymer.

In the first level of packaging, the DNA molecule is wrapped around histone proteins forming a structure reminiscent of a string of beads where each bead is known as a nucleosome having a width of 11nm as shown in Fig. 1.2 [70, 71]. Earlier, based on in vitro experiments and theoretical predictions arising from minimal models, it was argued that this string of nucleosomes does further fold in a highly regular solenoid or zig-zag manner to form a structure having 30nm width. However, current evidence suggests that, in vivo, chromatin may not exist in its 30nm fibre form; recent studies rather observe an irregular structure in the length scale of a few nucleosomes. [82, 96, 7, 63, 58, 117, 5, 36, 26, 118, 95]. However, how chromatin gets further organised in the interphase and finally to form mitotic chromosomes is still being investigated.

In early Eukaryotic organisms such as yeast, chromatin is organized into a *Rabl* configuration with centromeres of all the chromosomes forming a cluster tethered to the nuclear periphery and telomeres on the opposite direction [14]. Fig. 1.3(a) and (b) pictorially represent the Rabl configuration where the centromere cluster is marked in black, and each chromosome is shown in a different colour. It appears that the chromatin organiza-
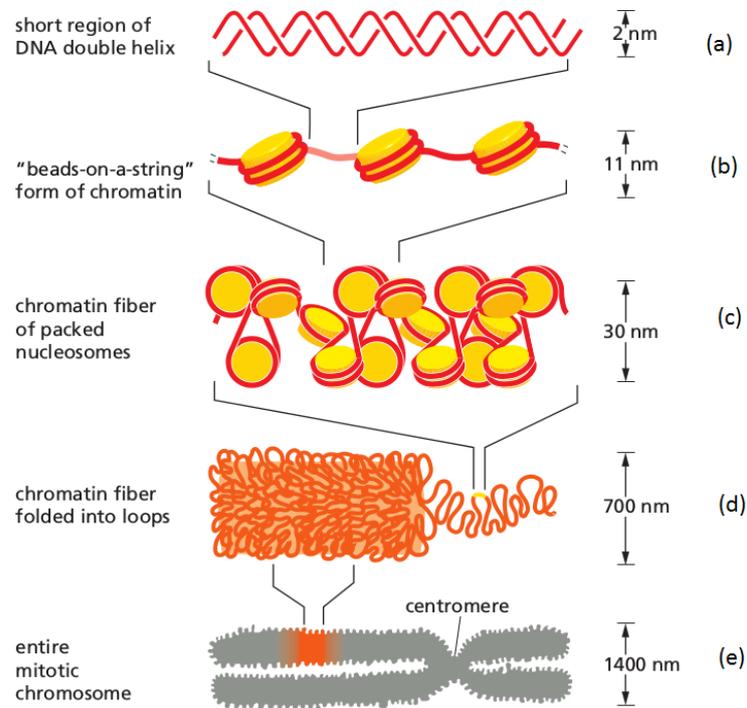
**Figure 1.2:** A typical model given in biology text books showing many levels of chromatin packing giving rise to the highly condensed mitotic chromosome. This figure is reproduced from Alberts [2] . However, new ongoing research is questioning many of the prevalent notions of chromatin organisation and a new picture is emerging.



(a) side view          (b) top view
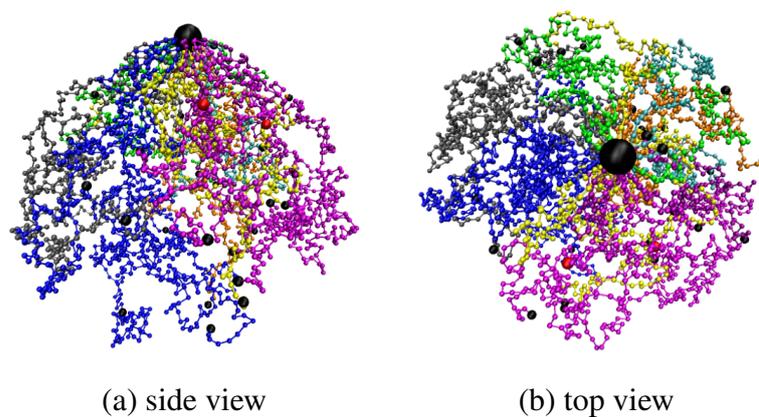
**Figure 1.3:** Pictorial representation of Rabl configuration of yeast genome where the centromeres are clustered together (big black bead). Different colors represent different chromosomes.
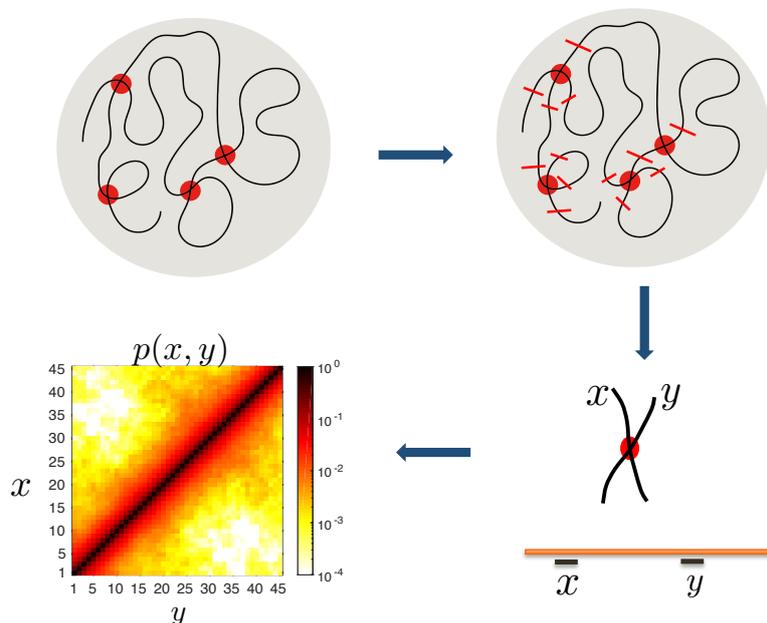
**Figure 1.4:** Schematic representation of chromosome conformation capture techniques. Chromatin polymers inside the cell nucleus are crosslinked with chemical agents like formaldehyde. Contacts between the chromatin segments are represented by the red circles. Crosslinked chromatin is then chopped off using restriction enzymes and the crosslinked fragments are separated. These fragments are then sequenced and locations are mapped back to the known genomic coordinates to identify the segments in the same crosslinked fragment. The results of these experiments can give us the information about the contact counts between all segment pairs of a chromatin.

tion in these fungal organisms is relatively simpler compared to the mammals like mouse or humans. In mammals, the chromatin is organized into many domains of different sizes and compaction. How these domains are formed and maintained is a matter of current research. To give this thesis a coherent prospective, a brief history of the discovery of chromatin is given below, followed by a summary of some of the relevant recent findings necessary to develop this thesis.

## 1.1 Experimental studies to understand chromatin

Experimentally, several attempts have been made to unravel the structure of chromatin, which can be broadly classified into two categories: biochemical methods and imaging

methods. Earlier studies using biochemical methods and electron microscopy techniques revealed a *beads-on-a-string* form of chromatin — a string of nucleosomes [71, 78] having a width of $\approx$ 10nm (also see Fig. 1.2). Later there have been intense investigations of how nucleosomes are organized in 3D, using imaging, X-ray scattering and other experimental techniques and led to the hypothesis that nucleosomes are organized in a zig-zag or solenoidal manner [53, 70, 80, 79, 11, 139, 45]. The focus then moved to understand the nucleosome crystal structure (using X-ray) [80], and positioning of nucleosomes along with the DNA (using sequencing methods) [119, 68]. Based on earlier studies of single mitotic chromosome structure, it was hypothesized that beads on string chromatin were further folded around a contiguous scaffold-like protein structure. However, single-molecule experiments showed that there is no such contiguous protein scaffold structure in mitotic chromosomes [104]. Detailed studies from the group of Maeshima and other groups showed that 10nm-wide chromatin is arranged in an irregular fashion with no well-defined structure [82, 96, 91, 7]

Around the same time, an imaging method known as fluorescence in situ hybridization (FISH) was used to visualize specific loci of each chromosome. These experiments, at the interphase, revealed the existence of chromosome territories — the finding that each chromosome has a preferred (relative) spatial location within the nucleus. It has been suggested that active genes within a chromosome prefer to be at the interior of the cell nucleus while the repressed genes are often found at the nuclear periphery [29, 123, 83, 61, 28]. Biochemical techniques have been successful in providing insights into our understanding of the structure of each chromosome and how chromatin is organized genomewide. One of the very useful methods that became popular in the last decade is the chromosome conformation capture method (3C) and its derivative techniques like 4C, 5C, HiC and microC. In these techniques, first, physically proximal chromatin segments are crosslinked by certain chemicals such as formaldehyde, as shown in Fig. 1.4. The crosslinked regions are then cut using restriction enzymes, separated out and sequenced. The sequenced segments are mapped back to the known genomic regions revealing the identity of the physically proximal regions. The result of these experiments are often represented in the form of a heatmap where the $x$ and $y$ axes indicate genomic segment locations along the contour; the elements of the heatmap (a matrix) reflect the contact frequency between the segments. One can easily observe high contacts (higher intensity) along the diagonal, indicating more contacts between the genomic segments which are nearby along the con-
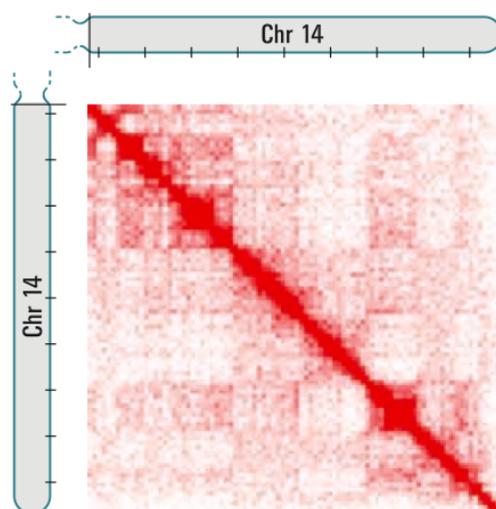
Hi-C contact map

**Figure 1.5:** Contact probability map from Hi-C experiment. This shows the intense area near diagonal indicating contact between the neighbouring chromatin regions. Many interesting structure can be seen throughout the matrix. This plot is reproduced from Lieberman-Aiden *et al.* [77].

tour. Far away regions come in contact less often. However, interestingly, there are many sub-regions that are very often in contact (see various structures in Fig. 1.5).

These experiments provided evidence for the existence of A and B compartments, with A being mostly composed of open and active chromatin while compartment B typically consists of repressed and condensed heterochromatin. It has been observed that these compartments are enriched with intra-compartmental interactions with minimal inter-compartmental interactions. The high throughput data on contacts between chromatin segments generated from these biochemical experiments have been extremely useful to study chromatin organization.

Detailed Hi-C studies revealed the importance of many high-intensity square structures, mostly near the diagonal axis of the Hi-C contact maps. These suggest that chromatin is organised in the form of domains with more contact within the domain and lesser contact between the different domains [92, 37]. It was found that, broadly, there are (at least) two kinds of domains [107]. One set of domains that are bounded by CCCTC-binding factor (CTCF) forms looped structures. Other set of domains are self-assembled without CTCF forming a loosely compact structure. Even though all these domains are

often termed as "topologically associated domains", there is a debate on the precise name and nature of these domains [140, 109].

It has been hypothesized that loop extrusion and/or phase separation could be the mechanism for the formation of these domains [3, 109, 86]. In the loop extrusion picture, chromatin regions are actively brought together with the help of proteins like cohesins/condensins and held together by CCCTC-binding factor (CTCF) [107, 109, 57, 49]. However, CTCF-dependent loops are found only in a fraction of the domains [107, 69]. Hence an alternative proposal is that chromatin domains may also be passively formed via phase separation [86, 88, 62]. Recently, it has been shown that in the absence of loop extruding factors, chromatin does still form domains and execute the necessary biological function[86, 18, 13, 69, 55] indicating that micro phase-separation might be an important mechanism. Phase separation is also known to bring together certain enhancers and promoters, segregating them from other regions [88, 62, 121]. In certain cases, as far as biological function is concerned, there is an ongoing debate whether the actual contact is crucial or proximity — closeness in 3D without being physically in contact — would suffice [86, 13, 144, 138].

The findings of these experiments have remarkably improved our understanding of chromatin. However, these Hi-C experiments have certain limitations. In all these experiments, cells are "fixed", and what we observe are frozen snapshots. Another limitation is that these experiments have limited resolution (often many kbs) and cannot reveal the details at the highest resolution (1bp - 10kb) even if required. The Hi-C method often gives us population-averaged data, and obtaining single-cell details are relatively more difficult with Hi-C. As a complementary method, new live high resolution microscopy experiments [93, 137, 54, 15, 131] are emerging; however, they are yet to provide us genomewide information.

## 1.2   Computational approaches to studying chromatin organization

Given the recent advances in experimental techniques, there is a large amount of imaging and biochemical experimental data on chromatin organization that are available. Theoretical and computational studies are indispensable and complementary tools to bridge the gap in our understanding of genome architecture and cell function. Several attempts have

been made to understand the 3D organization of the genome using a variety of techniques developed previously to understand the statics and dynamics of polymers [87, 116, 66, 133, 50, 8, 30, 52, 142, 33, 35, 34].

Similar to developments in experimental studies, computational/theoretical investigations too have been trying to study chromatin at different lengthscales and timescales. There have been many models that investigated the nucleosome organization and subsequent 3D organization of nucleosomes into 30nm fibres and other possible higher-order organization at the nucleosome resolution [11, 42]. There have also been models that investigate how chromatin is organized into territories. In a set of recent papers [50, 1], it has been argued that non-thermal activities arising from transcription and other cellular processes can be crucial for the understanding organization of chromatin into territories.

There have been a set of models that focused on understanding the 3D organization of single chromatin based on the Hi-C and imaging data [34, 35, 130]. Since the typical resolution of these experiments are not very high, and the nucleosome-resolution model for the whole chromatin is computationally very expensive, nearly all these models coarse-grained the chromatin into a polymer with the basic monomer unit representing 1mb or many kbs of the genome. Early Hi-C experiments revealed that the contact probability ($p$) decays as a function of genomic length ($s$) as $p \sim s^{-\gamma_s}$, where the exponent is found to be $\gamma_s \approx 1$. This exponent was different from the equilibrium globule ($p \sim s^{-1.5}$). This led to the proposal that chromatin is organized as a fractal globule ($p \sim s^{-1}$) [76]. The formation of loops of various sizes was evident from these experiments, but the mechanism of loop formation was still debated. This led to the hypothesis that such long loops need to be actively formed, and a loop extrusion mechanism was proposed [3]. The finding of CTCF dependent loop formation [107] and in vitro extrusion by loop extruding factors like cohesin lent credence to this hypothesis [51]. Since then, there have been many studies that investigated various aspects of loop extrusion and the role of loop extruding factors in chromatin organization [57, 112].

Several computational/theoretical efforts have been made to unravel chromatin organization features based on the contacts information provided by the HiC-like experiments. One of the approaches is the String and Binder Switch (SBS) model, where they consider chromatin as a polymer string with many diffusing binders promoting interaction. Studies based on this model have investigated various aspects of chromatin, including the coexistence of euchromatin and heterochromatin (polymer chain from good solvent to poor

solvent limit) [27, 31, 4, 16]. Chromatin has also been modeled as a copolymer chain where different kinds of beads representing different regions/domains of chromatin. This led to many interesting features such as observation of glassy behaviour [120]. Polymer models such as Chrom3D [101, 23, 98, 46, 24] consider HiC as well as lamin associated data and predict certain features of lamin associated domain regulation in cells. Polymer models accounting for histone modifications and other ChIP-Seq data investigate the role of "epigenomics-driven interactions in shaping the 3D genome" [34, 90] and investigate generic principles of chromatin folding. Polymer models at the nucleosome level explicitly modeling the histone tail modifications have shown interesting 3D organization of chromatin-based on 1D epigenetic data [67, 81, 113].

**Forward models and the need of an inverse model** Nearly all the computational studies do conventional "forward" polymer simulations. They start with a known set of potential energy functions and simulate the chromatin using Monte Carlo or Brownian dynamics methods. However, in the case of chromatin, we do not know the precise potential energy parameters for intra-chromatin interactions. HiC-like experiments tell us that there are many intra-chromatin crosslinking-like interactions; however, we do not know the strength of these interactions. Hence, to understand chromatin organization accurately based on HiC experimental data, one requires an "inverse" approach. That is, we need to compute the optimal interaction strength, given the contact probability from HiC experiments.

Many of the models that use Hi-C data convert contact counts obtained from Hi-C experiments into spatial distances, using a pre-decided formula [47, 38, 132, 108, 9, 100, 99]. That means, given a contact count matrix, such methods do not predict the distances between different chromatin segments; rather, they take the distance values as inputs based on certain assumptions. They then use conventional Monte Carlo (or equivalent) methods to find steady-state configurations of the chromatin, given a distance map between different DNA segment pairs. In other words, the existing models consider this as a "forward" problem of computing equilibrium configurations of chromatin as a consequence of assuming a certain spatial distance between bead-pairs. However, the problem of computing 3D configurations of chromatin polymer, given a contact map, is not a "forward" problem but rather an "inverse" problem [84]. The question is, given a contact map, what are the

optimal interactions between different segments of chromatin such that the experimentally seen contact map emerges.

In the literature so far, minimal work is done to recover an ensemble of chromatin conformations consistent with the experimentally generated contact probabilities. Meluzzi and Arya [85] proposed a prototype model that does not require conversion of contact probabilities to mean spatial distance based on a predetermined relation, unlike most computational studies. Instead, they suggest an iterative technique to optimize the potential parameters, which would satisfy the experimental contact probability constraints. However, they did not apply this to study any real chromatin regions using experimentally measured contact maps. Also, a limitation of their model is that they have considered a harmonic spring potential responsible for the interaction between any pair of beads. This is not fully consistent with the short-range nature of intra-chromatin interactions that breaks beyond certain extension. An inverse model like this with short-range potential mimicking the protein-protein interaction is important to appropriately study chromatin organization and dynamics.

## Our approach

Our broad aim is to study 3D chromatin organization and dynamics consistent with currently known experimental findings. Given a set of experimental data, we want to use computational methods to investigate and predict chromatin organization in space and time. Given that Hi-C experiments do not give us any information about chromatin polymer dynamics, the high-intensity elements of the contact matrix is often imagined as a set of static contacts holding together different regions of chromatin [96]. In this work, we question this notion and study chromatin polymer fluctuations and dynamics by converting the contact probability matrix to optimal interaction energies through an inverse strategy. As a first step, note that experimentally inferred contact probability values for nearly all segments pairs are very small, mainly of the order of $10^{-1}$, $10^{-2}$ or smaller. This implies that the contacts will be often broken, and the chromatin polymer can be highly dynamic. Even in a steady state, there are likely to have large fluctuations, cell to cell variability and temporal variability within a single cell. While there have been many studies to understand the 3D shapes and contacts, there is very little understanding of polymer fluctuations and cell-to-cell variability. It is important to quantify the fluctuations and dynamics in a way that is consistent with what is observed experimentally.

## 1.3   Specific objectives of the thesis

To fulfil the research shortcomings and to improve our understanding of chromatin organisation and its relevance in function of the cell, the objective of the current work are listed below:

1. **Modeling intra-chromatin interactions**: Conventionally, studies have been using harmonic springs between interacting pairs of chromatin. This implies an attractive force between these pairs that do not decay with distance but instead increases. Even if one takes harmonic spring with a cut-off length, there is an artificially induced sudden change in the potential leading to a discontinuous derivative like force. To overcome this limitation, we intend to use a short-range potential suitable for chromatin polymer where the interaction energy decays with distance and smoothly reaches zero at the cut-off distance.

2. **Development of an algorithm to calculate intra-chromatin interaction strengths based on chromatin conformation capture experiments**: Since we do not know the precise interaction strengths of intra-chromatin interactions, we want to develop an algorithm that would compute the optimal interaction strengths between different chromatin segments based on the contact probability values from chromatin conformation capture experiments. In other words, our objective is to develop an inverse algorithm to compute interaction strengths, taking the contact map as the input. We will test the robustness of the algorithm on a prototype and also apply this for a real chromatin domain.

3. **Investigating the 3D organization of chromatin on the lengthscale of a few genes**: Once we compute the intra-chromatin interaction strength, we will simulate chromatin organization and study structural properties such as radius of gyration, shape functions and density profiles. To do this, we will consider a chromatin domain spanning a few gene loci and different epigenetic states such as active and repressed states.

4. **Prediction of fluctuations in chromatin domain and its dynamics**: Going beyond the average static properties, we aim to study quantities that capture the fluctuations of chromatin segments, such as the distance distribution of chromatin segments. We plan to examine distance distribution between different segments of a chromatin

domain under different epigenetic conditions. We will also study the dynamics of
chromatin segments.

5. **Study the role of hydrodynamic interactions (HI) in chromatin folding**: Crucial
   molecular scale physics such as hydrodynamic interactions play an important role in
   determining the dynamic properties of a polymer chain. Here in the present work,
   we aim to study the influence of hydrodynamic interactions on all the dynamic
   properties of chromatin polymer.

In this thesis, we use Brownian dynamics simulations to study chromatin configura-
tions, their fluctuations around the steady state configurations and dynamics within the
domains, taking $\alpha$-globin gene region as an example. In Chapter 2, we present the gov-
erning equation and simulation techniques used in this thesis. We used a novel potential to
represent the intra-chromatin interaction. The details of the potential and the estimation of
the cut-off parameters is provided in Chapter 3. We then develop an "inverse" algorithm
to obtain the optimal interaction strength parameters that recover the known measurable
properties such as contact probability. Chapter 4 provides the details on the development
of the "inverse" technique as well as its validation with a prototype. In Chapter 5, we
solve the inverse problem and obtain the optimal interaction parameters for the $\alpha$-globin
domain. In Chapter 6, we compute fluctuations in spatial and temporal quantities associ-
ated with a chromatin domain; we quantify this by computing the probability distribution
of distances between various segments of chromatin and argue that the chromatin state is
not dictated by an average 3D distance rather a wider distribution having multiple peaks.
We also compute the distributions in loop formation time and contact time. Finally, in
Chapter 7, we conclude the thesis by discussing the significance of these findings and
providing a summary & future outlook.

# Chapter 2

# Brownian dynamics simulations to compute the statics and dynamics of chromatin polymer

This chapter describes our model for chromatin, and the simulation method we implement. We will be using a bead-spring chain polymer model to study chromatin and simulate it using the Brownian dynamics simulation method. The aim is to investigate static and dynamic properties of chromatin polymer on the lengthscale of a few genes. The set of governing equations for the bead-spring chain model with a spring force, and with hydrodynamic and excluded volume interactions incorporated, are presented here.

## 2.1 Chromatin as a bead-spring chain

Since chromatin is a long polymer, we coarse grain the system and model it as a bead-spring chain made of $N$ beads each having radius $a$, with polymer connectivity introduced by $N-1$ elastic springs, as shown in Fig. 2.1. In general, the bonded potential between the adjacent beads can be modeled by the spring potential such as Hookean, Fraenkel, FENE or a potential derived from worm-like chain force-extension relation etc [19, 110]. In this thesis we have mostly used the Fraenkel and FENE springs. Below we describe these two spring potentials in detail.

**Fraenkel spring:** A Fraenkel spring is similar to a Hookean spring with a pre-determined natural length. The force in the Fraenkel spring is directly proportional to the extension or compression of the spring from its natural length, and the corresponding energy is
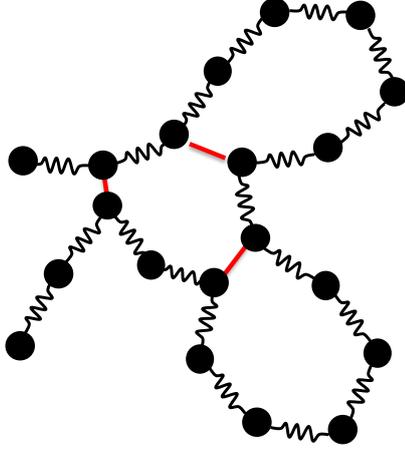
**Figure 2.1:** Schematic representation of bead-spring chain model, The adjacent beads are connected by elastic springs. The red bonds between beads represent protein-mediated intra-chromatin interactions.

described as

$$U^{\text{S}} = \sum_{\mu} \frac{H}{2}(|\boldsymbol{r}_\mu - \boldsymbol{r}_{\mu+1}| - r_0)^2 \tag{2.1}$$

where, $\boldsymbol{r}_\mu$ is the position vector of bead $\mu$, $r_0$ is the natural length and $H$ is the stiffness of the spring [21].

**FENE spring:** In the bead-spring simulation literature, a finitely extensible nonlinear elastic (FENE) spring is used to model a situation where a spring cannot be extended beyond a certain length, and where the tension force increases non-linearly as the spring length reaches its maximum permissible length. The FENE spring potential is given by

$$U_{\text{FENE}} = -\frac{1}{2}Q_0^2 \ln\left(1 - \frac{r_{\mu\nu}^2}{Q_0^2}\right) \tag{2.2}$$

where $Q_0$ is the maximum stretchable length of a single spring, $r_{\mu\nu}$ is the distance between the $\mu$th and $\nu$th bead.

**Potential energy representing excluded volume and protein-protein interactions:** Excluded volume refers to the idea that one part of a chain can not occupy space that is already occupied by another part of the chain. This is also known steric hindrance. Conventionally, the excluded volume or steric hindrance is modelled by the Weeks-Chandler-

Andersen (WCA) potential or the repulsive part ($r \leq 2^{1/6}\sigma$) of the Lennard-Jones (LJ) potential given as

$$
U_{\text{WCA}} =
\begin{cases}
4\epsilon_{\text{LJ}} \left[ \left( \dfrac{\sigma}{r} \right)^{12} - \left( \dfrac{\sigma}{r} \right)^{6} + \dfrac{1}{4} \right]; & r \leq 2^{1/6}\sigma \\[4mm]
0; & r > 2^{1/6}\sigma
\end{cases}
\tag{2.3}
$$

$$
U_{\text{LJ}} = 4\epsilon_{\text{LJ}} \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^{6} \right]
\tag{2.4}
$$

Here $r = (r_\mu - r_\nu)$ is the distance between beads $\mu$ and $\nu$, $\epsilon_{\text{LJ}}$ is the parameter to control the bead-bead interaction strength. See Fig. 2.2 where these potentials are plotted.

Apart from the excluded volume interaction, different parts of the chromatin can have short-ranged attractive interactions mediated by certain specific proteins as shown by the red bond in Fig. 2.1. This would appear as if different segments of chromatin are crosslinked. Such specific interactions between specific chromatin segments can be modelled by the attractive part of the LJ or by a harmonic spring. However, none of these potentials can accurately model the chromatin crosslinking. For example, in the case of the harmonic spring potential, the attractive force between segment pairs does not decay with distance, but rather increases and the attractive part in the LJ potential goes to zero only at infinite distance. Unlike these potentials, the intra-chromatin interaction is zero when the two segments are not in proximity.

To overcome the drawbacks of the above potentials, we have used a novel Soddemann-Dünweg-Kremer (SDK) potential [124] to model intra-chromatin interactions. The energy function is given by:

$$
U_{\text{SDK}} =
\begin{cases}
4 \left[ \left( \dfrac{\sigma}{r} \right)^{12} - \left( \dfrac{\sigma}{r} \right)^{6} + \dfrac{1}{4} \right] - \epsilon; & r \leq 2^{1/6}\sigma \\[4mm]
\dfrac{1}{2}\epsilon \left[ \cos \left( \alpha \left( \dfrac{r}{\sigma} \right)^{2} + \beta \right) - 1 \right]; & 2^{1/6}\sigma \leq r \leq r_c \\[4mm]
0; & r \geq r_c
\end{cases}
\tag{2.5}
$$

The parameters $\alpha$ and $\beta$ are determined by applying the two boundary conditions, namely, $U^{\text{SDK}} = 0$ at $r = r_c$ and $U^{\text{SDK}} = -\epsilon$ at $r = 2^{1/6}\sigma$. The appropriate choice of the cut-off radius $r_c$ has been investigated extensively and is present in the next chapter. It has been
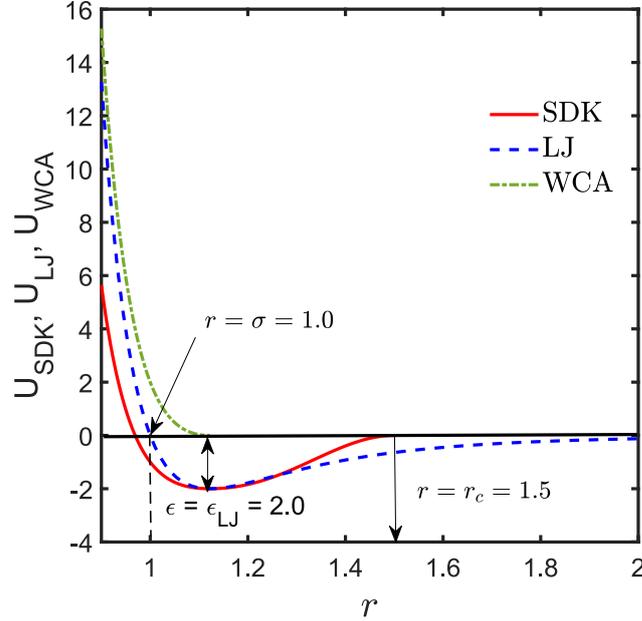
**Figure 2.2:** Three different interactions potentials are plotted as a function of the radial distance, $r$. The Lennard-Jones (LJ) potential, the Weeks-Chandler-Andersen (WCA) potential and the Soddemann-Dünweg-Kremer (SDK) potential. The parameters used here are $\epsilon = \epsilon_{LJ} = 2.0$, $\sigma = 1$, and $r_c = 1.5$.

shown that a value of $r_c = 1.82\sigma$ leads to an accurate prediction of the static properties of a polymer chain in poor, theta and good solvents. The same value is adopted here in the present study. A comparison is drawn in Fig. 2.2 between the SDK potential and the conventional LJ and WCA potentials, the expressions of which are given in Eq. (2.4) and Eq. (2.3), respectively.

The SDK potential has the following advantages compared to the LJ potential:

1. the complete range of solvent qualities, from poor to athermal, can be explored by varying the single parameter, $\epsilon$, which can alter the attractive component of the SDK potential without affecting the repulsive force (i.e. steric hindrace). On the contrary, for the LJ potential it is not possible to vary the attractive interaction without also affecting the repulsive part.

2. Protein-mediated interactions in chromatin are like effective "bonds" formed and broken with a finite range of interaction. Unlike the LJ potential, the SDK potential has a finite attractive range —the SDK potential energy smoothly reaches zero at

the cut off radius, $r_c$, whose value is set by the choice of two parameters $\alpha$ and $\beta$.

## 2.2 Governing equation

The configurational state of the chain is completely specified by $\{r_\nu | \nu = 1, ...., N\}$, the set of position vectors of the beads. The configurational distribution function, $\psi(r_1, ..., r_N, t)$ for the chain is obtained using the Fokker-Planck equation:

$$\frac{\partial \psi}{\partial t} = -\sum_{\nu=1}^{N} \frac{\partial}{\partial r_\nu} \cdot \left[ \frac{1}{\zeta} \sum_{\mu=1}^{N} \boldsymbol{\Gamma}_{\mu\nu} \cdot \frac{\partial U}{\partial r_\mu} \right] \psi + \frac{k_B T}{\zeta} \sum_{\mu,\nu=1}^{N} \frac{\partial}{\partial r_\mu} \cdot \boldsymbol{\Gamma}_{\mu\nu} \cdot \frac{\partial \psi}{\partial r_\nu} \tag{2.6}$$

In this equation, the force $F_\nu$ ($F_\nu = -\partial U/\partial r_\nu$) is the net force on the $i$ th bead arising as a result of all the conservative intramolecular interactions. It is the sum of net force exerted on the $i$ th bead by the springs adjacent to it, and the force due to excluded volume interaction. $\boldsymbol{\Gamma}_{\mu\nu}$ is the diffusion tensor representing hydrodynamic interactions (HI). HI account for the modification of the drag force on a particle that arises due to the motion of other particles mediated by the solvent. A schematic diagram depicting hydrodynamic interactions is shown in figure 2.3.

We have non-dimensionalized the equations using the characteristic time-scale, $\lambda_H = \zeta/4H$ (where $\zeta$ is the Stokesian bead-friction coefficient, which is related to the bead radius a through $\zeta = 6\pi a \eta_s$ with $\eta_s$ being the viscosity of the solvent), and the characteristic length-scale, $l_H = \sqrt{k_B T/H}$ (where $k_B$ is the Boltzmann's constant, $T$ is the absolute temperature and $H$ is the spring constant). It is to be noted that non-dimensionalization can be performed in different ways. Appendix A.1 contains the non-dimensionalization in two ways and shows that both leads to same results, when converted to dimensional units. Dimensionless quantities are denoted with an asterisk as a superscript, such that

$$t^* = \frac{t}{\lambda_H}; \ \psi^* = \psi l_H^3; \ r^* = \frac{r}{l_H} \tag{2.7}$$

Scaling the variables as shown above, the dimensionless Fokker-Planck equation is obtained as follows.

$$\frac{\partial \psi^*}{\partial t^*} = -\sum_{\nu=1}^{N} \frac{\partial}{\partial r_\nu^*} \cdot \left[ \frac{1}{4} \sum_{\mu=1}^{N} \boldsymbol{\Gamma}_{\mu\nu} \cdot \frac{\partial U^*}{\partial r_\mu^*} \right] \psi^* + \frac{1}{4} \sum_{\nu,\mu=1}^{N} \frac{\partial}{\partial r_\nu^*} \cdot \boldsymbol{\Gamma}_{\mu\nu} \cdot \frac{\partial \psi^*}{\partial r_\mu^*} \tag{2.8}$$

The HI between any pair of beads is taken into account through the dimensionless diffusion tensor $\boldsymbol{\Gamma}_{\mu\nu} = \delta_{\mu\nu}\boldsymbol{\delta} - \zeta \boldsymbol{\Omega}_{\mu\nu}$, where $\delta_{\mu\nu}$ is the Kronecker-delta, $\boldsymbol{\delta}$ is the unit tensor and
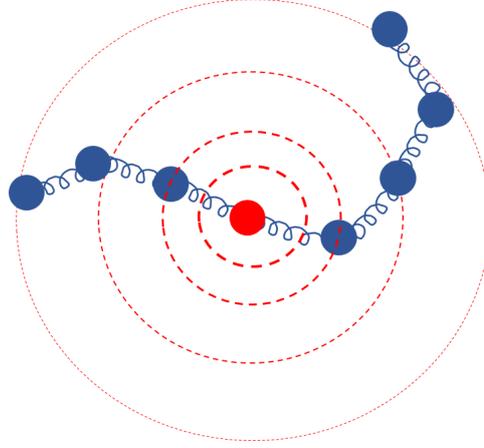
**Figure 2.3:** Pictorial representation of hydrodynamic interactions in a bead-spring chain. Disturbance in the motion of the red bead is propagated to all other beads through the solvent

$\mathbf{\Omega}_{\mu\nu}$ is the tensor representing the HI between the $i$th and $j$th beads.

$$\mathbf{\Omega}_{\mu\nu} = \mathbf{\Omega}(r_i^* - r_j^*) \tag{2.9}$$

We use the regularized Rotne-Prager-Yamakawa (RPY) tensor to represent HI; its form is

$$\Omega(\mathbf{r}^*) = \Omega_1 \boldsymbol{\delta} + \Omega_2 \frac{\mathbf{r}^* \mathbf{r}^*}{r^{*2}} \tag{2.10}$$

for $r^* \geq 2\sqrt{\pi}h^*$

$$\Omega_1 = \frac{3\sqrt{\pi}}{4} \frac{h^*}{r^*} \left(1 + \frac{2\pi}{3} \frac{h^{*2}}{r^{*2}}\right)$$

$$\Omega_2 = \frac{3\sqrt{\pi}}{4} \frac{h^*}{r^*} \left(1 - \frac{2\pi}{3} \frac{h^{*2}}{r^{*2}}\right) \tag{2.11}$$

and for $r^* \leq 2\sqrt{\pi}h^*$

$$\Omega_1 = 1 - \frac{9}{32} \frac{r^*}{h^* \sqrt{\pi}}$$

$$\Omega_2 = \frac{3}{32} \frac{r^*}{h^* \sqrt{\pi}} \tag{2.12}$$

where $h^*$ is the dimensionless bead radius in the bead-spring model, defined as $h^* = a/(l_H \sqrt{\pi})$. Typical values of $h^*$ lie between 0 and 0.5.

## 2.3 Brownian Dynamics simulation

The Itô stochastic differential equation for the chain configuration equivalent to the Fokker-Planck equation Eq-2.8 is given by

$$dr_\nu^* = D(r_\nu^*)dt^* + B(r_\nu^*) \cdot dW^* \tag{2.13}$$

where each component of $W$ is an independent Wiener process. The function $D(r_\nu^*)$ correspond to the drift term in the Fokker-Plank equation and the tensor $B(r_\nu^*)$ corresponds to the diffusion term, which is chosen in such a way that

$$B \cdot B^T = \frac{1}{2}\Gamma \tag{2.14}$$

Equation-2.13 can be solved through numerical integration. We will use the derivative free, second-order scheme as it is computationally better than other methods. A stochastic trajectory is generated by calculating the conformation of a single chain at the end of each successive time step. This can be performed in two steps, first predictor step which is followed by the corrector step [105].

**Predictor Step**

$$r_\nu^{'*}(t^* + \Delta t^*) = r_\nu^*(t^*) + D(r_\nu^*)\Delta t^* + B(r_\nu^*) \cdot \Delta W_\mu \tag{2.15}$$

where $r_\nu^{'*}(t^* + \Delta t^*)$ is the configuration at the end of predictor step of the $n + 1$th time step and $r_\nu^*$ is the configuration at the end of the correcter step of $n$th step.

**Corrector Step**

$$
\begin{aligned}
r_\nu^*(t^* + \Delta t^*) =& r_\nu^*(t^*) + \frac{1}{2}\left[D(r_\nu^{'*}) + D(r_\nu^*)\right]\Delta t^* - \frac{1}{2}B(r_\nu^*) \cdot \Delta W_\mu \\
&+ \frac{1}{4}\sum_n \left[B(\Gamma_\mu^n) + B(\Gamma_\mu^{'*})\right] \cdot \Delta W_\mu
\end{aligned} \tag{2.16}
$$

Using such an implicit formulation and solving for $r_i^*(t^* + \Delta t^*)$ leads to greater stability of the numerical algorithm which permits the use of larger time steps, which, in principle, should lead to a reduction in the CPU-time required for the simulation of a bead-spring chain.

**Table 2.1:** Definitions of shape functions in terms of eigenvalues of the gyration tensor, **G**. Note that, $I_1 = \lambda_1^2 + \lambda_2^2 + \lambda_3^2$, and $I_2 = \lambda_1^2\lambda_2^2 + \lambda_2^2\lambda_3^2 + \lambda_3^2\lambda_1^2$, are invariants of **G**.

| Shape function | Definition |
|---|---|
| Asphericity [134, 126] | $B = \langle \lambda_3^2 \rangle - \dfrac{1}{2}\left[\langle \lambda_1^2 \rangle + \langle \lambda_2^2 \rangle\right]$ |
| Acylindricity [134, 126] | $C = \langle \lambda_2^2 \rangle - \langle \lambda_1^2 \rangle$ |
| Degree of prolateness [22, 145, 126] | $S = \dfrac{\left\langle (3\lambda_1^2 - I_1)(3\lambda_2^2 - I_1)(3\lambda_3^2 - I_1) \right\rangle}{\left\langle (I_1)^3 \right\rangle}$ |
| Relative shape anisotropy [134, 145] | $\kappa^2 = 1 - 3\dfrac{\langle I_2 \rangle}{\langle I_1^2 \rangle}$ |

## 2.4  Quantifying chromatin macrostates

Since we are interested in the 3D organization of chromatin, we use a number of different static properties to describe the shape of the equilibrium chromatin chain. One such property defining a macrostate of chromatin polymer is the radius of gyration of the chain defined as, $R_g \equiv \sqrt{\langle R_g^2 \rangle}$, where $\langle R_g^2 \rangle$ can also be defined by

$$\langle R_g^2 \rangle = \langle \lambda_1^2 \rangle + \langle \lambda_2^2 \rangle + \langle \lambda_3^2 \rangle \tag{2.17}$$

with, $\lambda_1^2$, $\lambda_2^2$, and $\lambda_3^2$ being the eigenvalues of the gyration tensor **G** (arranged in ascending order), with

$$\mathbf{G} = \frac{1}{2N^2} \sum_{\nu=1}^{N} \sum_{\mu=1}^{N} \mathbf{r}_{\mu\nu} \mathbf{r}_{\mu\nu} \tag{2.18}$$

Note that, **G**, $\lambda_1^2$, $\lambda_2^2$, and $\lambda_3^2$ are calculated for each trajectory in the simulation before the ensemble averages are evaluated. The asymmetry in equilibrium chain shape has been studied previously in terms of various functions defined in terms of the eigenvalues of the gyration tensor [73, 125, 145, 59, 127, 134, 22]. Apart from $\lambda_1^2$, $\lambda_2^2$, and $\lambda_3^2$, themselves, we have examined the following *shape functions*: the asphericity ($B$), the acylindricity ($C$), the degree of prolateness ($S$), and the shape anisotropy ($\kappa^2$), as defined in Table 2.1.

Eigenvalues of the radius of gyration tensor for polymer chains are usually reported in terms of ratios, either between individual eigenvalues or with the mean-square radius of gyration. For a chain with a spherically symmetric shape about the centre of mass, we expect $\langle \lambda_i^2 \rangle / \langle R_g^2 \rangle = 1/3$, for $i = 1, 2, 3$, and $\langle \lambda_i^2 \rangle / \langle \lambda_j^2 \rangle = 1$ for all combinations $i$ and $j$. For chain shapes with tetrahedral or greater symmetry, the asphericity $B = 0$, otherwise $B > 0$. For chain shapes with cylindrical symmetry, the acylindricity $C = 0$, otherwise $C > 0$. With regard to the degree of prolateness, its sign determines whether chain shapes are preponderantly oblate ($S \in [-0.25, 0]$) or prolate ($S, \in [0, 2]$). The relative anisotropy ($\kappa^2$), on the other hand, lies between 0 (for spheres) and 1 (for rods). We have analysed the $\alpha$-globin gene locus using these macroscopic properties which is given in the following chapters.

## 2.5 Summary

In this section, we discuss the bead-spring model of polymer. We provided the potentials for different types of spring which represents the connectivity of the beads in a chain. We discussed various interbead potentials to mimic the steric hinderance and intra-chromatin interactions. A novel SDK potential to crosslink polymer segments, and its advantages over the conventional potential has been stated. We then provided the governing equation and the simulation techniques for the bead-spring chain. Some of the static macroscopic properties and its calculation from the ensemble of configurations has been given here. Finally, we point out certain limitation of our polymer model.

# Chapter 3

# Soddemann-Dünweg-Kremer (SDK) potential for intra-chromatin interactions

In this chapter, we discuss a novel potential to study intra-chromatin interactions, namely the Soddemann-Dünweg-Kremer (SDK) potential, that we briefly introduced in the previous chapter. The potential has many advantages as discussed earlier. However, we need to tune the parameters such that it reproduces the know scaling laws in all the solvent regimes ranging from good to poor.

The repulsive part of the SDK potential is similar to the WCA potential, but the attractive part is modelled by the cosine function as

$$
U_{\text{SDK}} = \begin{cases} 4\left[\left(\dfrac{\sigma}{r}\right)^{12} - \left(\dfrac{\sigma}{r}\right)^6 + \dfrac{1}{4}\right] - \epsilon; & r \leq 2^{1/6}\sigma \\[2em] \dfrac{1}{2}\epsilon\left[\cos\left(\alpha\left(\dfrac{r}{\sigma}\right)^2 + \beta\right) - 1\right]; & 2^{1/6}\sigma \leq r \leq r_c \\[2em] 0; & r \geq r_c \end{cases}
\tag{3.1}
$$

The constants $\alpha$ and $\beta$ are determined by applying the two boundary conditions, namely, $U_{\text{SDK}} = 0$ at $r = r_c$, and $U_{\text{SDK}} = -\epsilon$ at $r = 2^{1/6}\sigma$. Based on these two boundary conditions,

$\alpha$ and $\beta$ are calculated by solving the following set of equations,

$$2^{1/3}\alpha + \beta = \pi \tag{3.2}$$

$$\left(\frac{r_c}{\sigma}\right)^2 \alpha + \beta = 2\pi \tag{3.3}$$

In order to solve the above set of equations, it is required to choose a reasonable value of the cut-off radius, $r_c$. In the original study by Soddemann et al. [124], $r_c$ was chosen to be $1.5\,\sigma$ to include only the first neighboring shell of interactions, determined from the first minimum of the pair correlation function. For $r_c = 1.5\,\sigma$, the values of $\alpha$ and $\beta$ are calculated to be $3.1730728678$ and $-0.856228645$, respectively [124], and the resultant SDK potential has been used to investigate various equilibrium properties of the solutions of polymer chains using molecular dynamics (MD) and Monte Carlo (MC) simulations [124, 127].

In sec. 3.1, we determine the theta and onset of poor solvent for the SDK potential with $r_c = 1.5\sigma$ and investigate the scaling laws in all the solvent regimes. We observe the deviation from the know scaling law in poor solvents and consequently modify the SDK potential cutoff in sec. 3.2 for accurate reproduction of scaling behavior in all solvent regimes ranging from good to poor.

## 3.1 Theta point and poor solvent onset determination

For a linear polymer chain, the mean-squared radius of gyration follows the universal scaling law $R_g^2 \sim (N - 1)^{2\nu}$, where the value of the Flory exponent, $\nu$, depends on the solvent quality. At the $\theta$-temperature, linear polymer chains follow RW statistics, with $\nu = 1/2$, leading to the ratio $R_g^2/(N - 1)$ to be independent of the chain length, $N$. Whereas, in the case of good and poor solvents, $\nu$ takes the values $3/5$ and $1/3$, respectively [111]. For polymer chains with the SDK potential representing the excluded volume force, the temperature dependence can be captured with the potential well-depth, $\epsilon$. As mentioned earlier, $\epsilon = 0$ (which is equivalent to a WCA potential), represents the athermal limit, where the chain is fully swollen. With increasing values of $\epsilon$, a unique value is reached, where the repulsive and attractive interactions between pairs of beads are precisely balanced, leading to $\theta$-like conditions. The value of $\epsilon$ at the $\theta$-point can be estimated by plotting the ratio $R_g^2/(N - 1)$ versus $\epsilon$ for different chain lengths, $N$, and finding the point of intersection at which curves for different values of $N$ intersect, as shown in Fig. 3.1(a) [127].
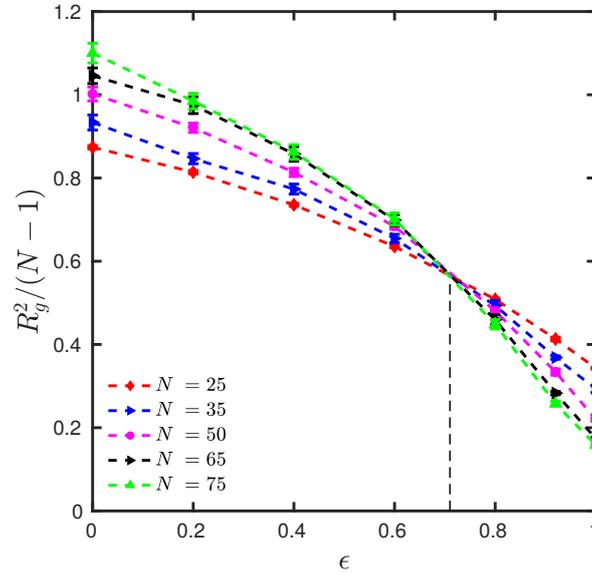
(a)
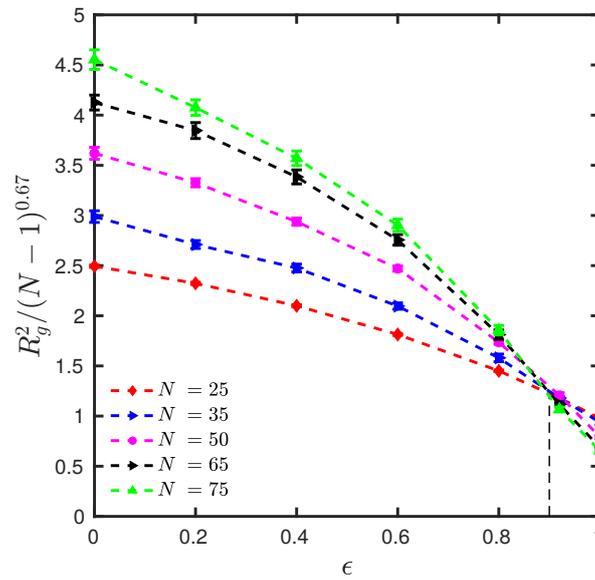


(b)

**Figure 3.1:** (a) The ratio $R_g^2/(N-1)$ as a function of the well depth of the SDK potential, $\epsilon$, used to estimate the $\theta$-point for the cutoff radius $r_c = 1.5\sigma$. (b) The ratio $R_g^2/(N-1)^{0.67}$ as a function of the well depth of the SDK potential, $\epsilon$, used to estimate the onset of poor-point for the cutoff radius $r_c = 1.5\sigma$. The symbols represent simulation data and the dotted lines are drawn to guide the eye.

Similarly, the onset of poor-point can be estimated by plotting the ratio $R_g^2/(N-1)^{0.67}$ versus $\epsilon$ for different chain lengths, $N$, and finding the point of intersection at which curves for different values of $N$ intersect, as shown in Fig. 3.1(b). Following this procedure, the $\theta$-point for a homopolymer chain with beads connected by FENE springs having a maximum stretchable length of $Q_0^2 = 50.0$, is found to be $\epsilon = 0.72$. All the simulations to study the cut-off radius of SDK potential use a value of $Q_0^2 = 50.0$. Note that the notation $Q_0^2$ used here is the same as the more commonly used FENE $b$-parameter. With increasing values of $\epsilon$ beyond $\epsilon = 0.72$, the chain begins to collapse due to decreasing solvent quality. As mentioned earlier, in the limit of a poor solvent, linear polymer chains obey the scaling law $R_g \sim (N-1)^{1/3}$, indicating that the chains are space filling. The simulations presented in this chapter are carried out for different chain lengths, $N$, ranging from 25 to 90 beads per chain, with an equilibration run of about 8 Rouse relaxation times [94] and a production run of 6 to 8 Rouse relaxation times with a non-dimensional time step size $\Delta t^* = 0.001$. Data from each independent trajectory in the simulations are collected at an interval of 1000 to 5000 non-dimensional time steps over the entire production run and time averages are calculated over each of the trajectories. Average equilibrium properties and error of mean estimates are evaluated over an ensemble of such independent time averages consisting of 1000 to 2000 independent trajectories.

Fig. 3.2 (a) studies the chain length dependence of $R_g^2$ for various well depths $\epsilon$. For $\epsilon = 0$ (the athermal limit) and $\epsilon = 0.72$ (the $\theta$-point), the expected power law exponents of 1.2 and 1.0, respectively, are observed. For intermediate values in the crossover regime, $0 < \epsilon < 0.72$, one expects, strictly speaking, a curve beginning with slope 1 at small values of $N$, and gradually increasing to 1.2 for asymptotically long chains. However, for the fairly short chains studied here, this curvature is very hard to observe; instead the data can be well described in terms of an effective exponent, whose variation with $\epsilon$ is shown in Fig. 3.2 (b). An analogous crossover from a slope of 1.0 to (2/3) is expected as the well depth is increased beyond the $\theta$ value of 0.72, with the effective exponent remaining at (2/3) for sufficiently large $\epsilon$. However, as can be seen from Figs. 3.2 (a) and (b), the "asymptotic" slope at $\epsilon = 1$ seems to be only 0.35, which is obviously unphysical, if interpreted as an asymptotic scaling law. We can only speculate here about the reasons for this behaviour — since we were able to "cure" the problem without a detailed investigation, we did not attempt to analyse it in depth. However, a few observations may be made.

Firstly, Fig. 3.2 (a) shows clearly that the data at $\epsilon = 1$ are hampered by equilibra-
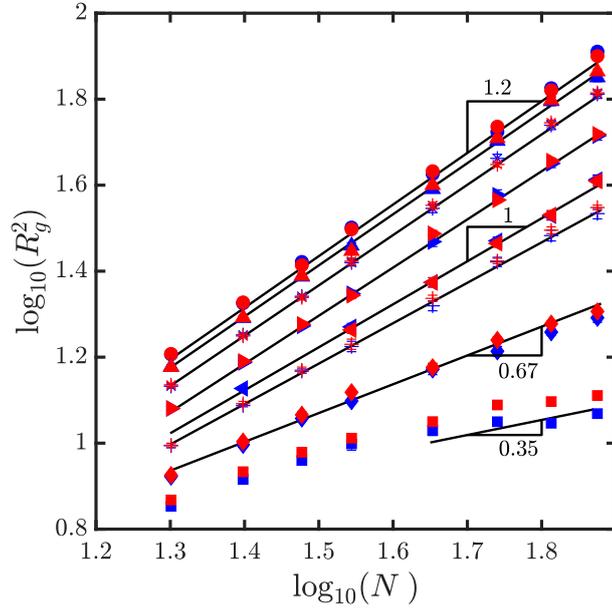
(a)
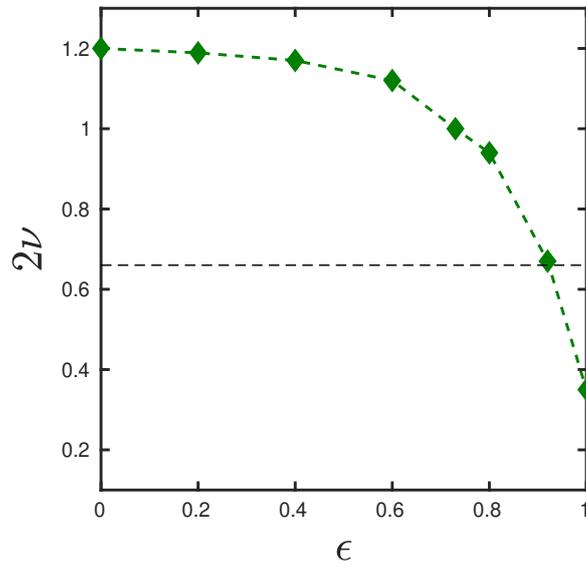


(b)

**Figure 3.2:** (a) The mean-squared radius of gyration as a function of the number of beads in a chain. The blue-coloured symbols are for different values of well-depth, $\epsilon$, in the absence of hydrodynamic interactions. • $\epsilon = 0$, ▲ $\epsilon = 0.2$, ∗ $\epsilon = 0.4$, ► $\epsilon = 0.6$, ◄ $\epsilon = 0.72$, + $\epsilon = 0.8$, ◆ $\epsilon = 0.92$ and ■ $\epsilon = 1$. The same symbols are used with a red colour for simulations with hydrodynamic interactions. The straight lines are of slope $2\nu$ at different values of $\epsilon$. (b) Effective exponent $2\nu$ versus the well-depth, $\epsilon$, for cutoff radius $r_c = 1.5\,\sigma$.
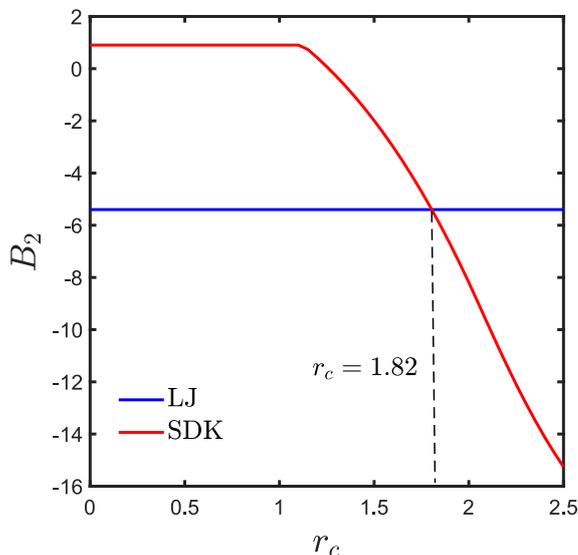
**Figure 3.3:** The second virial coefficient $B_2$ of the SDK potential, compared with the corresponding value for the LJ potential, as a function of the cutoff radius, $r_c$, for well-depths $\epsilon = \epsilon_{LJ} = 1.0$, and $\sigma = 1$.

tion problems. This becomes obvious via the comparison of data accumulated with and without hydrodynamic interactions, which, as static averages, must be identical if strict thermal equilibrium and sufficient sampling is achieved. Secondly, it has already been pointed out in works of Soddemann et al [124] that the SDK potential with $r_c = 1.5\,\sigma$ has a propensity to induce crystallisation, i.e., highly ordered structures. It is then quite conceivable that the growth of a highly collapsed globule with chain length occurs essentially in a layer-by-layer fashion, which would then give rise to a fairly abrupt increase of $R_g^2$ as soon as a new layer begins to be populated. The small slope of 0.35 that we observe in Fig. 3.2 (a) may then perhaps be part of a quasi-plateau that corresponds to oscillations that are added on top of the leading $N^{2/3}$ behaviour.

## 3.2    The cut-off radius for the SDK potential

Prompted by the experience with using a simple Lennard-Jones potential in analogous studies of collapsing polymer chains [115], which did not exhibit this problem, we attempted to solve it by modifying the SDK potential such that it would mimic more closely the attributes of the Lennard-Jones potential. In practice, we adjusted the range of the

SDK potential $r_c$ by requiring that, for $\epsilon_{LJ} = \epsilon = 1$ and $\sigma = 1$, both potentials give rise to the same value of the second virial coefficient, which is easily calculated by evaluating the Mayer $f$-function [111]. It is defined as the difference between the Boltzmann factor for two monomers at a distance $r$ and at infinite distance. The Mayer $f$-function is given by

$$f(r) = \exp\left[\frac{-U(r)}{k_B T}\right] - 1 \tag{3.4}$$

At short distance, the Mayer function is negative because of the large energy $U(r)$ at a short distance due to hard-core repulsion. The probability of finding monomers at this distance is highly reduced relative to the non-interacting case i.e. monomers at infinite distance. The excluded volume is defined as the negative integral of the Mayer function over the whole space.

$$
\begin{aligned}
B_2 &= -\int_0^\infty f(r)\, d^3\mathbf{r} \\
&= \int_0^\infty (1 - \exp[-U(r)/k_B T])\, d^3\mathbf{r} \\
&= \int_0^\infty 4\pi r^2 (1 - \exp[-U(r)/k_B T])\, dr
\end{aligned}
\tag{3.5}
$$

Matching this value with the corresponding LJ value results in $r_c = 1.82\,\sigma$ (see Fig. 3.3), for which $\alpha = 1.5306333121$ and $\beta = 1.213115524$. In view of the remarks made earlier, it is well conceivable that such a smoother potential will exhibit less pronounced oscillations or perhaps none at all.

As seen in Fig. 3.4 (a), the effective exponent shows a gradual decrease from 1.2 at $\epsilon = 0$ to 0.67 at $\epsilon = 0.55$, and it remains constant at 0.67 well beyond $\epsilon = 0.55$, as shown in Fig. 3.4 (b). The values of the mean-squared radius of gyration, $R_g^2$, are reproduced with HI for $\epsilon = 0$, 0.45 and 1.0, for different chain lengths and found to be consistent with the results without HI (as seen in Fig. 3.4 (a)). All subsequent results reported here with the SDK potential are for $r_c = 1.82\,\sigma$. Finally, we determine the $\theta$-point and onset of poor solvent point for this value of $r_c$ as before. As shown in Fig. 3.5, we find it occurs at $\epsilon = \epsilon_\theta = 0.45$. By varying the depth of the SDK potential, the entire range of solvent quality (from good to poor) can be achieved.

As we have seen previously that the system was not in equilibrium at certain $\epsilon$ values with SDK potential at cutoff $1.5\sigma$. We further need to confirm that the system is in its equilibrium state at all values of $\epsilon$ with SDK potential at cutoff $1.82\sigma$. To validate the
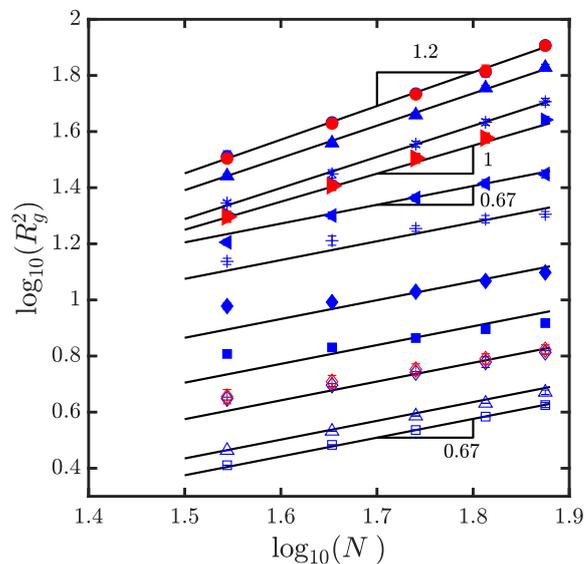
(a)
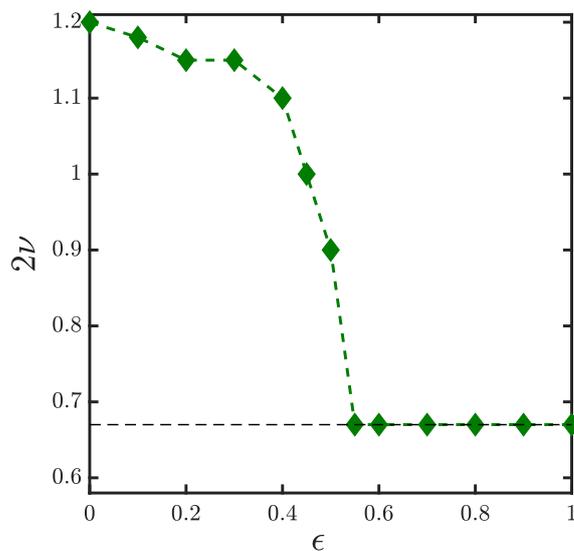


(b)

**Figure 3.4:** (a) The mean-squared radius gyration versus number of beads in a chain. The blue-coloured symbols are for different values of well-depth, $\epsilon$, in the absence of hydrodynamic interactions. ● $\epsilon = 0$, ▲ $\epsilon = 0.2$, * $\epsilon = 0.4$, ▶ $\epsilon = 0.45$, ◀ $\epsilon = 0.55$, + $\epsilon = 0.6$, ◆ $\epsilon = 0.7$ and ■ $\epsilon = 0.8$, ◇ $\epsilon = 1$, △ $\epsilon = 2$ and □ $\epsilon = 3$. The same symbols are used with a red colour for simulations with hydrodynamic interactions. The straight lines are of slope $2\nu$ at different values of $\epsilon$. (b) Exponent $2\nu$ versus the well-depth, $\epsilon$, at cutoff radius $r_c = 1.82\,\sigma$.
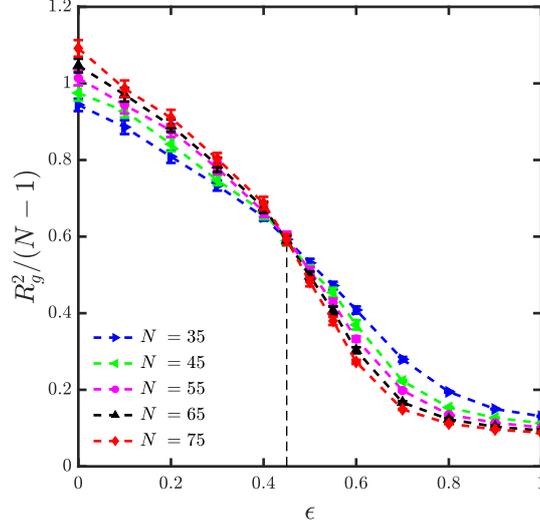
(a)
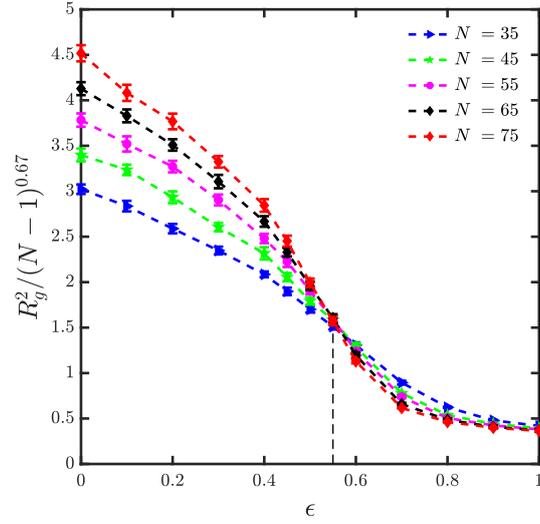


(b)

**Figure 3.5:** (a) The ratio $R_g^2/(N-1)$ versus the well depth of the SDK potential, $\epsilon$, to estimate the $\theta$-point for cutoff radius $r_c = 1.82\,\sigma$. (b) The ratio $R_g^2/(N-1)^{0.67}$ versus the well depth of the SDK potential, $\epsilon$, to estimate the onset of poor-point for cutoff radius $r_c = 1.82\,\sigma$. The symbols represent simulation data and the dotted lines are drawn to guide the eye. The $\theta$-point and the onset of poor-point is estimated as the intersection of all the curves and leads to $\epsilon = 0.45$ and $\epsilon = 0.55$, respectively.

equilibrium states, we studied the same system with SDK potential (cutoff=$1.82\sigma$) and HI at three different values of $\epsilon$ i.e. $0.0, 0.45$ (theta point) and at $1.0$, as shown in Fig. 3.4

## 3.3   Summary

A smooth and short range pair-wise potential has been introduced in the present chapter. Theta point and the onset of poor solvent has been investigated using the SDK potential at a cut-off of $1.5\sigma$ and the scaling at various solvent qualities has been studied. The scaling of the radius of gyration has been observed to be in disagreement with the universal swelling of homopolymers in the poor solvent regime. Also, the simulations with hydrodynamic interactions indicated that the polymer chains are not equilibrated after sufficiently long trajectory. In order to reproduce the excluded volume exerted by the LJ potential at $\epsilon = 1$ and $\sigma = 1$, the cut-off of the SDK potential has been modified to $1.82\sigma$. The swelling of the radius of gyration using SDK potential at this cutoff of $1.82\sigma$ has been observed to be identical with the universal swelling of homopolymers in the thermal crossover regime. Hence, the cut-off $1.82\sigma$ for SDK potential is realistic and may be utilised for further studies.

# Chapter 4

# Inverse Brownian Dynamics method to compute optimal intra-chromatin interaction strengths

The main challenge to simulate chromatin configuration and dynamics is that we do not know the interaction strength parameters among different segments. Chromosome conformation capture experiments such as 5C and Hi-C provide information on the contact counts between different segments of chromatin, but not the interaction strengths that lead to these counts. Here, in this chapter, we propose a novel method to obtain optimal interaction strengths between all chromatin segments, given the contact probability between different segments. This is an "inverse" method and we implement it here along with Brownian dynamics simulations and hence it is termed as the Inverse Brownian Dynamics (IBD) method. Firstly, in sec. 4.1, we present the IBD method in general terms, followed by its implementation in the context of a chromatin in sec. 4.2. We test the robustness of the algorithm by validating it with a prototype as discussed in sec. 4.3.

## 4.1 Inverse Brownian Dynamics method

The method is best explained in general terms. It is assumed that the system is described by a phase space variable $\Gamma$ and a model Hamiltonian $\mathcal{H}(\Gamma)$. Another assumption is that the simulation produces the canonical average of some observable, given by a phase-space
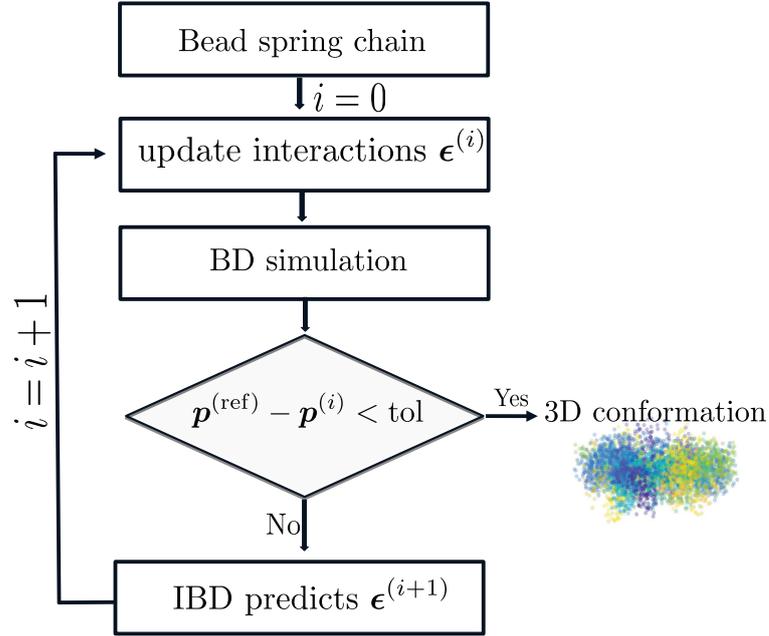
**Figure 4.1:** Flowchart for the Inverse Brownian Dynamics (IBD) method. Here $p^{(\text{ref})}$ represents the reference contact probability matrix, and $p^{(i)}$ represents the contact probability matrix from simulations at iteration $i$. The interaction strength between beads $i$ and $j$ is given by $\epsilon_{\mu\nu}$.

function $A(\Gamma)$:

$$\langle A \rangle = \frac{\int d\Gamma \, A(\Gamma) \exp(-\beta \mathcal{H}(\Gamma))}{\int d\Gamma \exp(-\beta(H)(\Gamma))}. \tag{4.1}$$

Here $\beta = 1/(k_B T)$. On the other hand, we have a given "target" value $A_t$ (e.g. from experiment), which will typically differ from our simulation result. We are now interested in the dependence of the Hamiltonian on some coupling parameter $J$, and we wish to adjust $J$ in order to bring $\langle A \rangle$ as closely to $A_t$ as possible, within the limitations of the Hamiltonian as such in general, and its dependence on $J$ in particular. In order to do this, it is desirable to obtain information on (i) in which direction $J$ should modified, and (ii) by what amount (at least by order of magnitude). If the change of the coupling constant, $\Delta J$, is small, we can write down a Taylor expansion around the value $J = J_0$ where we performed the simulation:

$$\langle A \rangle (J_0 + \Delta J) = \langle A \rangle (J_0) + \chi \Delta J + O(\Delta J^2), \tag{4.2}$$

where the "generalized susceptibility" $\chi$ is an abbreviation for the thermodynamic derivative

$$\chi = \left. \frac{\partial \langle A \rangle}{\partial J} \right|_{J=J_0}. \tag{4.3}$$

The crucial point is now that $\chi$ can be directly sampled in the simulation, by making use of a standard fluctuation relation. Indeed, taking the derivative of Eq. 4.1 with respect to $J$, one finds directly

$$\chi = \beta \left[ \langle AB \rangle - \langle A \rangle \langle B \rangle \right], \tag{4.4}$$

where $B$ denotes another phase-space function, which is just the observable conjugate to $J$:

$$B(\Gamma) = -\frac{\partial \mathcal{H}(\Gamma)}{\partial J}. \tag{4.5}$$

In deriving Eq. 4.4, it is assumed that the phase-space function $A(\Gamma)$ does not depend on $J$, i.e. $\partial A(\Gamma)/\partial J = 0$. This is the case for most typical applications, and certainly for the present investigation.

The simplest way to do IBD, therefore, consists of (i) neglecting all nonlinear terms in Eq. 4.2, (ii) setting its left hand side equal to $A_t$, (iii) solving for $\Delta J$, and (iv) taking $J_0 + \Delta J$ as a new and improved coupling parameter, for which a new simulation is done, and for which the whole process is done again. To avoid overshoots, it is often advisable to not update $J$ by the full increment $\Delta J$ that results from solving the linear equation, but rather only by $\Delta J = \lambda \Delta J$, where $\lambda$ is a damping factor with $0 < \lambda < 1$. The iteration is terminated as soon as $|\langle A \rangle| - A_t$ does not decrease any more, within some tolerance. One also has to stop as soon as $\chi$ becomes zero, within the statistical resolution of the simulation (this is, however, not a typical situation).

The method may be straightforwardly generalized to the case of several observables $A_m$ and several coupling parameters $J_n$, where the number of observables and the number of couplings may be different. The Taylor expansion then reads

$$\langle A_m \rangle (\boldsymbol{J}_0 + \Delta \boldsymbol{J}) = \langle A_m \rangle (\boldsymbol{J}_0) + \sum_n \chi_{mn} \Delta J_n + O(\Delta \boldsymbol{J}^2), \tag{4.6}$$

where the matrix of susceptibilities is evaluated as a cross–correlation matrix:

$$\chi_{mn} = \beta \left[ \langle A_m B_n \rangle - \langle A_m \rangle \langle B_n \rangle \right], \tag{4.7}$$

with

$$B_n(\Gamma) = -\frac{\partial \mathcal{H}(\Gamma)}{\partial J_n}. \tag{4.8}$$

Typically, the matrix $\chi_{mn}$ will not be invertible (in general, it is not even quadratic!). Therefore, one should treat the linear system of equations via a *singular-value decomposition* (SVD) and find $\Delta J$ via the *pseudo-inverse* (PI). In practice, this means that one updates the couplings only in those directions and by those amounts where one has a clear indication from the data that one should do so, while all other components remain untouched. For details on the concepts of SVD and PI, the reader may refer to [106] and [44].

In the chromatin context, the averages $\langle A_m \rangle$ are the contact probabilities as produced by the simulations, while the target values are the corresponding experimental values (discussed in greater detail below). The corresponding phase–space functions can be written as indicator functions, which are one in case of a contact and zero otherwise. The coupling parameters that we wish to adjust are the well depths of the SDK attractive interactions, which we allow to be different for each monomer pair. The schematic representation of the IBD algorithm is provided as a Flowchart in Fig. 4.1. The IBD algorithm discussed here in general terms is described in more detail in the section below.

## 4.2   IBD implementation

This section discusses the implementation of the IBD method in the context of a chromatin-like bead-spring chain as discussed in the Chapter 2. The phase-space variable ($\Gamma$) for the coarse-grained bead-spring chain used to represent chromatin is the set of bead position vectors $\boldsymbol{r}_\nu (\nu = 1, 2, ..., N)$ and the Hamiltonian $\mathcal{H}(\Gamma, \{\epsilon_{\mu\nu}\}) = U^{\text{S}*} + U^{\text{SDK}*}$, where $U^{\text{S}*} = \sum\limits_{\nu=1}^{N-1} U_\nu^{\text{S}*}$ and $U^{\text{SDK}*} = \sum\limits_{\nu,\mu=1}^{N} U_{\mu\nu}^{\text{SDK}*}$. Since we are interested in bead-pairs, we construct a single index to represent any particular bead-pair. For instance, the expression

$$m = \frac{1}{2}[\nu(\nu - 1)] + [\mu - (\nu - 1)] \tag{4.9}$$

converts any bead pair $(\nu, \mu)$ to a single index $m$. Here $\nu$ varies from 2 to $N$, and $\mu$ varies from 1 to $(\nu - 1)$ for a matrix of size $N$. In terms of the single index, the average contact probability $p_m$ of the bead-pair $m$ is given by

$$p_m = \langle \hat{p}_m \rangle = \frac{1}{Z} \int d\Gamma \, \hat{p}_m \exp(-\beta\mathcal{H}) \tag{4.10}$$

Here, the partition function $Z = \int d\Gamma \exp(-\beta\mathcal{H})$, and $\hat{p}_m$ is an indicator function which indicates when contact occurs between the bead pair represented by index $m$. $\hat{p}_m$ is 1 if the distance between the beads is less than the cut-off distance of the indicator function, $r_p^*$, and 0 otherwise. For this work $r_p^* = r_c^*$, the cut-off distance of the SDK potential. We intend to target the experimentally obtained contact probability $p_m^{\text{ref}}$ by adjusting the well-depth of SDK attractive interactions $\epsilon_m$. The Taylor series expansion of $\langle \hat{p}_m \rangle$ about the interaction strength $\epsilon_m$ after neglecting higher order terms is

$$\langle \hat{p}_m \rangle (\epsilon_m + \Delta\epsilon_m) = \langle \hat{p}_m \rangle (\epsilon_m) + \sum_n \chi_{mn} \Delta\epsilon_n \tag{4.11}$$

where $\Delta\epsilon_m$ is the change in the interaction strength, and the susceptibility matrix

$$\chi_{mn} = \frac{\partial \langle \hat{p}_m \rangle}{\partial \epsilon_n} = \frac{\partial}{\partial \epsilon_n} \left[ \frac{1}{Z} \int d\Gamma \, \hat{p}_m \exp(-\beta\mathcal{H}) \right] \tag{4.12}$$

Simplifying further

$$\chi_{mn} = \frac{1}{Z} \left[ \int \hat{p}_m \frac{\partial}{\partial \epsilon_n} \left( \exp(-\beta\mathcal{H}) d\Gamma \right) \right] - \left[ \frac{1}{Z} \left( \frac{1}{Z} \int \hat{p}_m \exp(-\beta\mathcal{H}) d\Gamma \right) \frac{\partial Z}{\partial \epsilon_m} \right] \tag{4.13}$$

or

$$\chi_{mn} = \frac{\beta}{Z} \left[ \int \hat{p}_m \exp(-\beta\mathcal{H}) \frac{\partial(-\mathcal{H})}{\partial \epsilon_n} d\Gamma \right] - \beta \left[ \langle \hat{p}_m \rangle \frac{1}{Z} \int \exp(-\beta\mathcal{H}) \frac{\partial(-\mathcal{H})}{\partial \epsilon_n} d\Gamma \right], \tag{4.14}$$

Defining the quantity $b_n$ by

$$b_n = -\frac{\partial\mathcal{H}}{\partial \epsilon_n} \tag{4.15}$$

and using the expression for SDK potential, since only $U_n^{\text{SDK}*}$ depends on $\epsilon_n$, one can show

$$b_n = \begin{cases} 1 & r_n^* \leq 2^{\frac{1}{6}}\sigma^* \\ \frac{1}{2}\left[1 - \cos(\alpha\, r_n^{*2} + \beta)\right] & 2^{\frac{1}{6}}\sigma^* \leq r_n^* \leq r_c^* \\ 0 & r_n^* \geq r_c^* \end{cases} \tag{4.16}$$

which leads to

$$\chi_{mn} = \beta \left[ \frac{1}{Z} \int \hat{p}_m b_n \exp(-\beta\mathcal{H}) d\Gamma - \langle \hat{p}_m \rangle \frac{1}{Z} \int b_n \exp(-\beta\mathcal{H}) d\Gamma \right] = \beta \left[ \langle \hat{p}_m b_n \rangle - \langle \hat{p}_m \rangle \langle b_n \rangle \right]$$

$$\tag{4.17}$$

Replacing the left hand side of Eq. 4.11 with the target contact probability $p_m^{\text{ref}}$ obtained from experiment, we get

$$p_m^{\text{ref}} - \langle \hat{p}_m \rangle = \sum_n \chi_{mn} \Delta\epsilon_n \tag{4.18}$$

Equation 4.18 can be solved for any particular iteration step as

$$\epsilon_n^{(i+1)} = \epsilon_n^{(i)} + \lambda \sum_m C_{nm}^{(i)} \left( p^{\text{ref}} - \langle \hat{p}_m \rangle^{(i)} \right) \tag{4.19}$$

where the matrix $C$ is the *pseudo-inverse* of the matrix $\chi$, superscript *i* represents the iteration number, $\lambda$ denotes the damping factor with $0 < \lambda < 1$, and $\epsilon_n^{(i+1)}$ is the well-depth of the SDK attractive interaction for the next iteration step. Since the susceptibility matrix $\chi$ is often singular (with the rank of the matrix being smaller than its size), it is necessary to define the matrix $C$ as a *pseudo-inverse* and determine it by using *singular-value decomposition* (SVD). By carrying out SVD, the susceptibility matrix is decomposed into three square matrices $U, S$ and $V$ such that $\chi = USV^T$ where $U$ and $V$ are orthogonal and $S$ is a diagonal matrix. The *pseudo-inverse* of the susceptibility matrix via SVD can then be written as

$$C = \chi^{-1} = V \cdot [\text{diag}(1/S_n)] \cdot U^T \tag{4.20}$$

Note that in all the cases considered here, $\chi, S, U$ and $V$ are square matrices since there is a one-to-one correspondence between contact probability and interaction strength.

As per the well-established procedure for finding the *pseudo-inverse*, when $S_n < 10^{-4}$, $1/S_n$ is replaced by 0. In other words, some values in the $S$ matrix whose inverse would lead to problems, are ignored.

An example of the $S$ matrix obtained in the case of the prototype scenario where we have considered a bead-spring chain with 45 beads is given here as illustration. The matrix is of size $N(N-1)/2 \times N(N-1)/2 = 990 \times 990$, which represent the number of interacting bead-pairs in the chain. Fig. 4.2 represents the matrix at an intermediate stage in the iteration process. It can been seen that $S$ has diagonal elements whose magnitude is of order $10^{-4}$ or less (top right in Fig 4.2).
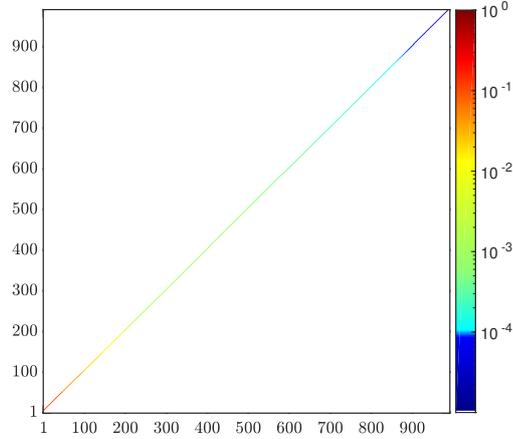
**Figure 4.2:** Representation of the matrix $\mathsf{S}$. The top right corner along the diagonal represents diagonal elements of the matrix, $S_n$, which are of order $10^{-4}$ or less.

## 4.3    Validation of the inverse Brownian dynamics method with a prototype

To validate the IBD method, a prototype of a chromatin-like polymer chain with artificially set interaction strengths ($\epsilon_{\mu\nu}$) was constructed. The data from this simulated chain was used to test the IBD algorithm, as described below. The IBD algorithm was validated for chains of length 10, 25 and 45 beads. Here we discuss the 45 bead chain case as a prototype. A few bead-pairs ($\mu\nu$) were connected arbitrarily with a prescribed value of the well-depth $\epsilon_{\mu\nu}^{(\text{ref})}$ of the SDK potential. The non-zero reference interaction strengths for the connected bead-pairs $\epsilon_{\mu\nu}^{(\text{ref})}$ are shown in Table 4.1; the remaining pairs were considered to have no attractive interaction ($\epsilon_{\mu\nu}^{(\text{ref})} = 0$). The beads-spring chain was simulated until it reached equilibrium, which was quantified by computing $R_g$ as a function of time. A stationary state was observed to be reached after eight Rouse relaxation times [20]. However, equilibration was continued for fifteen Rouse relaxation times. After equilibration, an ensemble of $10^5$ polymer configurations was collected from 100 independent trajectories, from each of which $10^3$ samples were taken at intervals of $10^3$ dimensionless time steps, which correspond to roughly 2 to 3 Rouse relaxation times. From this ensemble, the contact probability $p_{\mu\nu}^{(\text{ref})} = \langle \hat{p}_{\mu\nu} \rangle$ for each bead pair in the chain was computed. Here $\hat{p}_{\mu\nu}$ is an indicator function which is equal to 1 or 0 depending upon whether the $\mu^{\text{th}}$ and $\nu^{\text{th}}$

beads are within the cut-off distance of SDK potential ($r^*_{\mu\nu} \leq r^*_c$) or not ($r^*_{\mu\nu} > r^*_c$). The reference contact probabilities $p^{(\text{ref})}_{\mu\nu}$, determined in this manner, are shown in Fig. 4.4(b). In the present instance while $p^{(\text{ref})}_{\mu\nu}$ has been constructed by simulating the bead-spring chain for the given values of $\epsilon^{(\text{ref})}_{\mu\nu}$, in general it refers to the experimental contact probabilities.

The IBD method was then applied to recover the reference contact probabilities $p^{(\text{ref})}_{\mu\nu}$ starting with an initial guess of a self-avoiding walk where $\epsilon^{(0)}_{\mu\nu} = 0$, i.e., all the interaction strengths are set equal to zero. The contact probability for the initial state of self-avoiding walk is shown in Fig. 4.4(c). As illustrated in Fig. 4.1, at each iteration step $i$, Brownian dynamics was performed for the given $\epsilon^{(i)}_{\mu\nu}$ and an ensemble of $10^5$ conformations were collected. To quantify the difference between contact probabilities computed from simulation at iteration $i$ ($p^{(i)}_{\mu\nu}$) and reference contact probabilities ($p^{(\text{ref})}_{\mu\nu}$), the root mean-squared deviation $E^{(i)}_{\text{rmsd}}$ was calculated

$$E^{(i)}_{\text{rmsd}} = \sqrt{\frac{2}{N(N-1)} \sum_{1 \leq \mu < \nu \leq N} \left( p^{(i)}_{\mu\nu} - p^{(\text{ref})}_{\mu\nu} \right)^2} \tag{4.21}$$

at each iteration. The error criteria $E^{(i)}_{\text{rmsd}}$ has been used previously in Meluzzi and Arya [84], and is adopted here. At each iteration $i$, if the $E^{(i)}_{\text{rmsd}}$ value is greater than the preset tolerance limit (tol), the interaction strength parameters $\epsilon^{(i+1)}_{\mu\nu}$ for the next iteration were calculated. To avoid the overshoot in interaction strength $\epsilon^{(i+1)}_{\mu\nu}$, the range of $\epsilon^{(i+1)}_{\mu\nu}$ was constrained to [0, 10]. For the investigated polymer chain with 45 beads, the IBD algorithm converges ($E^{(i)}_{\text{rmsd}} < \text{tol}$) in approximately 50 iterations and $p^{\text{ref}}_{\mu\nu}$ was recovered. The error $E_{\text{rmsd}}$ for each iteration is shown in Fig. 4.4(a) while the recovered contact probability matrix is shown in Fig. 4.4(d). The recovered contact probability values along with the optimized interaction strengths $\epsilon_{\mu\nu}$ are shown in Table 4.1. The error in the recovered contact probabilities and interaction strengths is less than 5%, proving the reliability of the IBD method. The largest contact probabilities are for those bead-pairs for which values of the interaction strength were chosen a priori, as given in Table 4.1. However, the existence of these interactions leads to the existence of contact probabilities $p_{\mu\nu}$ between all bead-pairs $\mu$ and $\nu$. The IBD algorithm was applied to not just the specified bead-pairs but to recover all contact probabilities $p_{\mu\nu}$, for all possible pairs. The errors are given in Table 4.1 only for the specified values since they are the largest. To check the robustness of the IBD algorithm, the same reference contact probability of the prototype was recovered from an entirely different initial configuration of a collapsed chain where $\epsilon^{(0)}_{\mu\nu} = 1$ for all the bead pairs $\mu$ and $\nu$. The initial contact probability matrix of the collapsed chain is

(a)



(b)



(c)



(d)



(e)



(f)
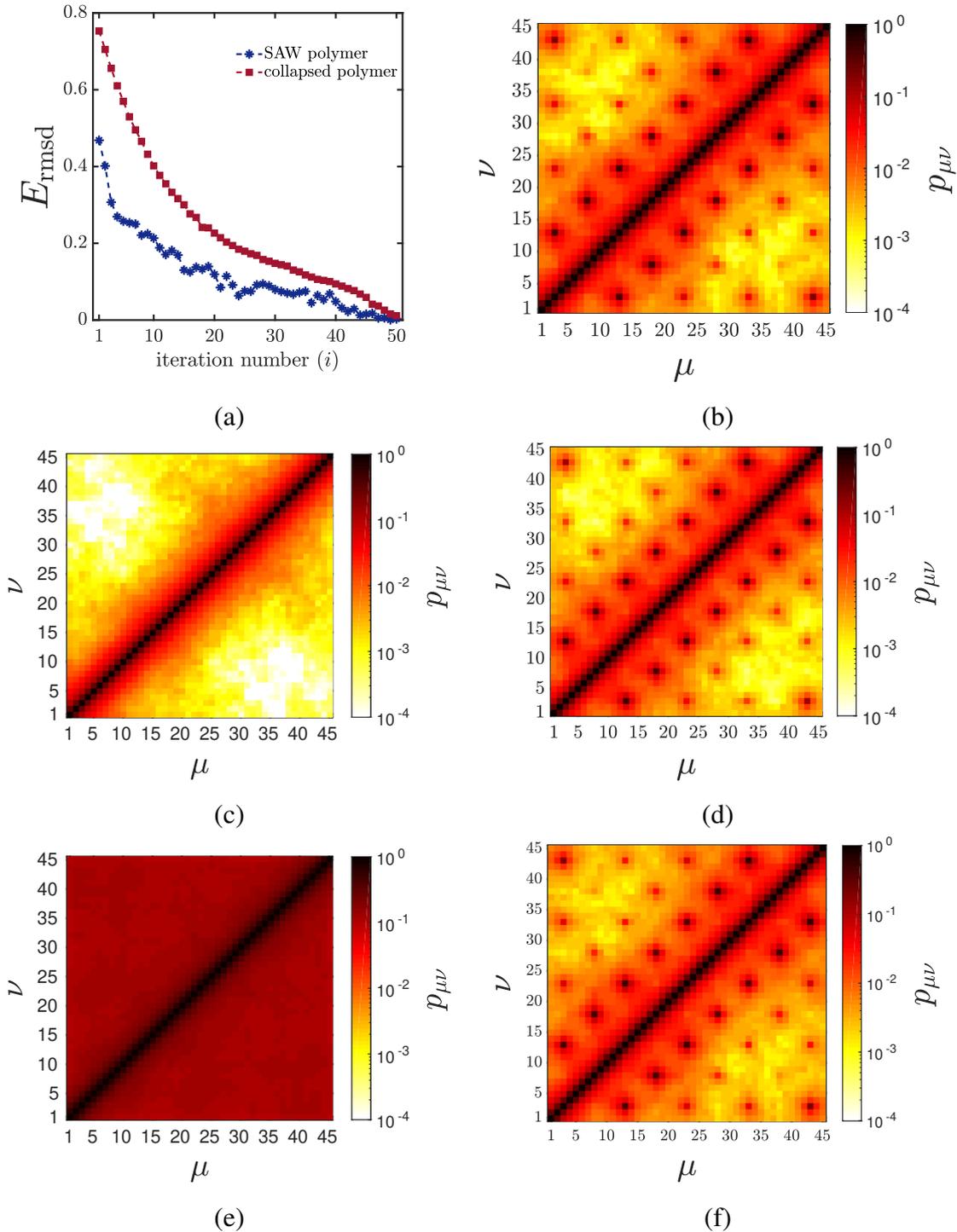
**Figure 4.3:** (a) Root-mean-square deviation $E_{\mathrm{rmsd}}$ (Eq. 4.21) as a function of iteration number showing convergence of the IBD method. (b) Reference contact probability matrix. Two different initial states have been considered for testing IBD convergence: (c) initial contact probability for the self-avoiding walk (SAW) where no bead-pairs have attractive interaction and (d) recovered contact probability matrix through IBD starting from the SAW state. Similarly (e) initial contact probability for the collapsed state where all the bead-pairs have attractive interaction, $\epsilon = 1$ and (f) recovered contact probability matrix through IBD starting with the collapsed state. .

**Table 4.1:** Interaction strengths $\epsilon_{\mu\nu}$ and contact probabilities $p_{\mu\nu}$ for selected bead pairs ($\mu\nu$) in a bead-spring chain with 45 beads. Values of these variables recovered using IBD are compared with those of the reference polymer chain, along with the percentage error between the reference and recovered values. Initial $\epsilon_{\mu\nu}$ values for all the bead-pairs were chosen to be 0 for the self-avoiding walk polymer while $\epsilon_{\mu\nu} = 1$ for all the bead-pairs in the collapsed polymer.

| | Initial state: self-avoiding walk polymer | | | | | | |
|---|---|---|---|---|---|---|---|
| bead-pair | interaction strength, $\epsilon_{\mu\nu}$ | | | contact probability, $p_{\mu\nu}$ | | | |
| | reference | recovered | % error | initial | reference | recovered | % error |
| 3-13 | 7.00 | 6.70 | 4.29 | 0.0033 | 0.44 | 0.46 | 4.55 |
| 13-23 | 7.00 | 7.28 | 4.00 | 0.0036 | 0.51 | 0.49 | 3.92 |
| 23-33 | 7.00 | 7.08 | 1.14 | 0.0057 | 0.39 | 0.37 | 5.13 |
| 33-43 | 7.00 | 7.35 | 5.00 | 0.0041 | 0.62 | 0.59 | 4.84 |
| 8-18 | 7.00 | 6.94 | 0.86 | 0.0056 | 0.47 | 0.47 | 0.00 |
| 18-28 | 7.00 | 6.89 | 1.57 | 0.0052 | 0.31 | 0.32 | 3.23 |
| 28-38 | 7.00 | 7.16 | 2.29 | 0.0071 | 0.55 | 0.53 | 3.64 |
| 3-43 | 7.00 | 7.18 | 2.57 | 0.0002 | 0.22 | 0.22 | 0.00 |
| | Initial state: collapsed polymer | | | | | | |
| bead-pair | interaction strength, $\epsilon_{\mu\nu}$ | | | contact probability, $p_{\mu\nu}$ | | | |
| | reference | recovered | % error | initial | reference | recovered | % error |
| 3-13 | 7.00 | 6.67 | 4.71 | 0.139 | 0.44 | 0.44 | 0.00 |
| 13-23 | 7.00 | 6.99 | 0.14 | 0.141 | 0.51 | 0.52 | 1.96 |
| 23-33 | 7.00 | 6.75 | 3.57 | 0.133 | 0.39 | 0.38 | 2.56 |
| 33-43 | 7.00 | 7.19 | 2.71 | 0.136 | 0.62 | 0.59 | 4.84 |
| 8-18 | 7.00 | 7.22 | 3.14 | 0.132 | 0.47 | 0.45 | 4.26 |
| 18-28 | 7.00 | 6.77 | 3.29 | 0.135 | 0.31 | 0.3 | 3.23 |
| 28-38 | 7.00 | 6.89 | 1.57 | 0.133 | 0.55 | 0.55 | 0.00 |
| 3-43 | 7.00 | 7.11 | 1.57 | 0.067 | 0.22 | 0.22 | 0.00 |
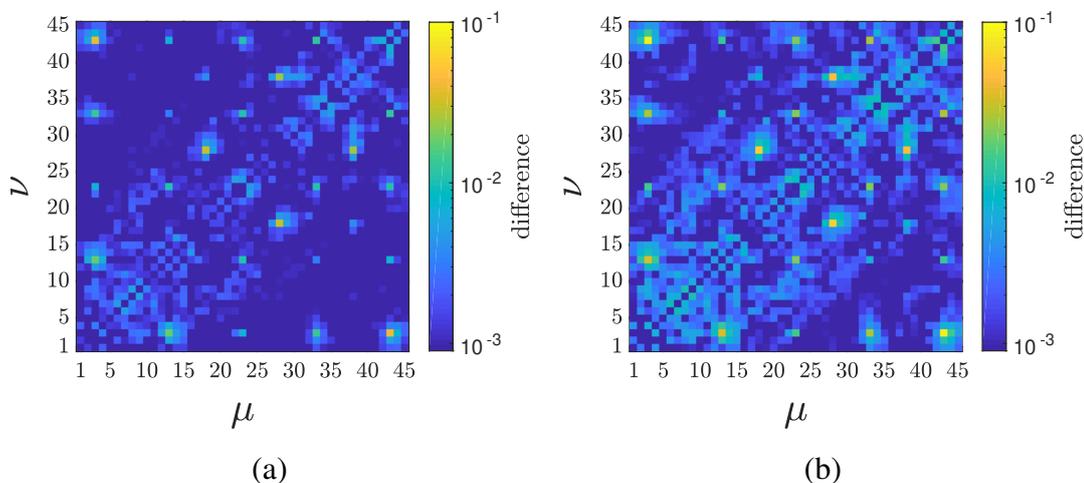
**Figure 4.4:** The difference between the reference and recovered contact probability for each bead-pairs are shown for simulations starting from (d) self-avoiding walk and (e) collapsed state.

shown in Fig. 4.4(e) and the recovered contact probability matrix starting from the collapsed chain is shown in Fig. 4.4(f). The recovered contact probability values along with the optimized interaction strengths $\epsilon_{\mu\nu}$ for a few bead-pairs are shown in Table 4.1. Thus even starting from a very different configuration, the IBD algorithm converges to the target contact probability matrix, establishing the robustness of the method (see section S2 for more details). Having validated the IBD algorithm, the next section applies this technique to experimentally obtained contact probabilities of a chromatin, on the length scale of a gene.

It is interesting to examine how the polymer chain converges from different initial configurations (swollen or collapsed) to the final reference state of the prototype model. Note that the reference contact probabilities for the prototype model were produced with the interaction strength $\epsilon_{\mu\nu} = 7$ for eight bead-pairs, and with $\epsilon_{\mu\nu} = 0$ for all the remaining 982 bead-pairs. On the other hand, we set $\epsilon_{\mu\nu} = 0$ for all bead-pairs in the swollen state, and $\epsilon_{\mu\nu} = 1$ for all bead-pairs in the collapsed state. The value of 7 chosen for the interaction parameter $\epsilon_{\mu\nu}$ of the 8 bead-pairs in the prototype model has no particular significance. It has been chosen arbitrarily such that a high contact probability is obtained between some bead-pairs, as displayed in Fig. 4.4(a). Snapshots of the initial swollen and collapsed states are shown in Figs. 4.5(a) and 4.5(b), while a snapshot of the final converged state corresponding to the reference state is shown in Fig. 4.5(c). It is clear
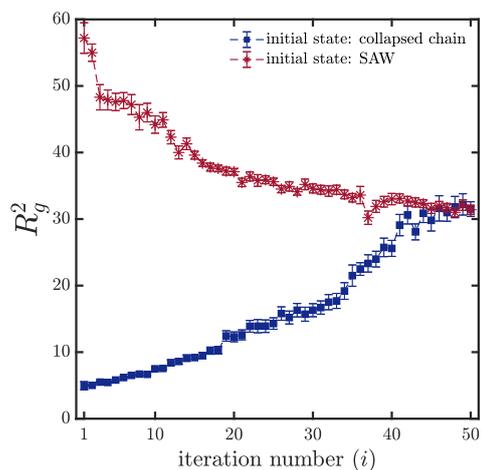
(a) initial state: swollen
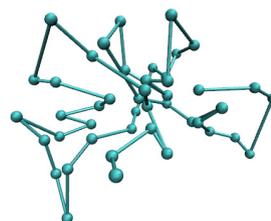
(b) initial state: collapsed
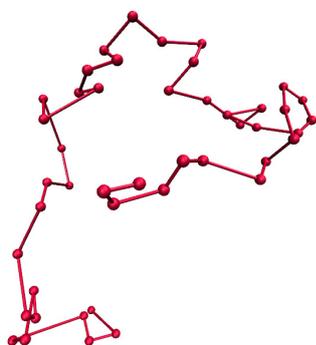
(c) reference state
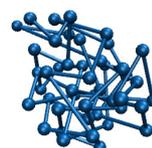
(d) radius of gyration

**Figure 4.5:** Randomly selected snapshots of (a) the swollen (self-avoiding walk) initial state (b) the collapsed initial state, and (c) the converged reference state. (d) Change in $R_g^2$ with the number of iterations in the IBD algorithm, as the system goes from initial (swollen or collapsed) state to the final (reference) state.

that while some beads in the reference state appear tightly bound, most of the others repel each other. Figure 4.5(d) displays the radius of gyration as a function of number of iterations in the IBD algorithm, as the system goes from the initial (swollen or collapsed) state to the final (reference) state. Starting from an initial swollen state, the radius of gyration decreases monotonically to the final state, defined by the contact probabilities in Fig. 4.4(a). When the starting state is the collapsed polymer, the radius of gyration increases monotonically until it reaches the final state. As is clear from the snapshot of the reference state, since the majority of the bead-pairs have no attraction, the reference state has a higher radius of gyration than the collapsed state. It is possible that some "strand passage" occurs in the neighbourhood of the tightly bound beads in the reference state as the chain conformation evolves from the initial collapsed state (where all the beads are attracted to each other) to the final reference state with a few tightly bound beads, since the current algorithm permits such passage (as discussed in section 2.1). The overall conformational evolution, however, as demonstrated by the radius of gyration in Fig. 4.5(d), is likely to be dominated by expanding chain conformations since the majority of the beads repel each other.

## 4.4    Summary

We have developed a novel IBD method to compute the optimal interaction strength parameters, given the contact probabilities. We discussed the IBD first in general terms and then generalised it for a chromatin system. We validated the IBD method for a bead-spring chain comprising of 45 beads. Starting from two different initial states of a bead spring polymer – SAW and collapsed state – the IBD recovers the contact probabilities and the interaction strengths (within 5% of error), reflecting its reliability. We presented the recovered contact probability map and the difference between the reference and the recovered contact probabilities. After this successful implementation of IBD on a prototype, we will proceed to estimate the interaction strength parameters for a real chromatin domain.

# Chapter 5

# Computing 3D chromatin configurations from a contact probability map

In this chapter, we go beyond the prototype and study the 3D organization of real chromatin. To do this, the $\alpha$-globin gene locus (ENCODE region ENm008) is chosen for which Bau et al. [9] have experimentally determined the contact counts using the 5C technique. This is a 500 kbp long region on human chromosome 16 containing the $\alpha$-globin gene and a few other genes like LUC7L. Since 5C data does not interrogate the contact counts between all feasible 10 kbp segment pairs, many elements in the heat map have no information. This is in contrast with typical Hi-C experiments where information on all possible contact pairs are obtained. In principle, this method can be applied to Hi-C data; however, in this instance, we chose the 5C data since it has sufficiently good resolution. In sec. 5.1, we present the coarse-graining procedure for the experimentally obtained contact probability data from 5C, followed by the normalization procedure for the conversion of contact counts to contact probabilities in sec. 5.2. We compute the spatial organization of $\alpha$-globin gene locus as given in sec. 5.3. We then analysed the relation between the contact probabilities and the spatial distance between different chromatin segments in sec. 5.4. Finally, we compared our findings of spatial distance with the experimental data and is shown in sec. 5.5.

## 5.1    The coarse-graining procedure

For simulation purpose, the $\alpha$-globin locus is coarse-grained to a bead-spring chain of 50 beads. That is, the experimental 5C data (contact count matrix of size $70 \times 70$) for the EMn008 region was converted to a contact count matrix of size $50 \times 50$. The coarse graining procedure is as follows: 500kb of the gene locus was divided into 50 beads, each comprising 10 kb equal-sized fragments. The midpoint of each restriction fragment was located and was assigned to the corresponding bead in the coarse-grained polymer. There are cases where two or more restriction fragments (each of size less than 10kbp) get mapped to the same bead. For example, consider restriction fragments $r_1$ and $r_2$ being mapped on to a single coarse-grained bead $\mu$, and fragments $r_3$ and $r_4$ being mapped on to another bead $\nu$. The contact counts of the coarse-grained bead-pair $C_{\mu\nu}$ can then be computed in at least three different ways, namely *independent*, *dependent* and *average* coarse-graining procedures, as described below.

- *Independent coarse graining*: Take the sum of all contact counts for the four restriction fragment combinations ($C_{\mu\nu} = C_{r_1r_3} + C_{r_1r_4} + C_{r_2r_3} + C_{r_2r_4}$) — i.e., assume that all contacts occur independently of each other, in other words not more than one of the contact pairs occurs in the same cell.

- *Dependent coarse graining*: Take the maximum contact count amongst all the four restriction fragment combinations ($C_{\mu\nu} = \max\{C_{r_1r_3}, C_{r_1r_4}, C_{r_2r_3}, C_{r_2r_4}\}$). This assumes that whenever the pairs having small contact counts are in contact, the pair with the largest contact count is also in contact. These are the two extreme cases and the reality could be somewhere in between.

- *Average coarse graining*: The third option is then to choose some such intermediate value. Here, we use the approximation that the coarse grained contact count is equal to the average of the two extreme contact counts mentioned earlier, namely $C_{\mu\nu} = \frac{1}{2}[(C_{r_1r_3} + C_{r_1r_4} + C_{r_2r_3} + C_{r_2r_4}) + \max\{C_{r_1r_3}, C_{r_1r_4}, C_{r_2r_3}, C_{r_2r_4}\}]$.

## 5.2    Conversion of contact counts to contact
##           probabilities: the normalization problem

The contact counts obtained from the Chromosome Conformation Capture experiments are not normalized. That is, the contact count values can vary from experiment to experi-
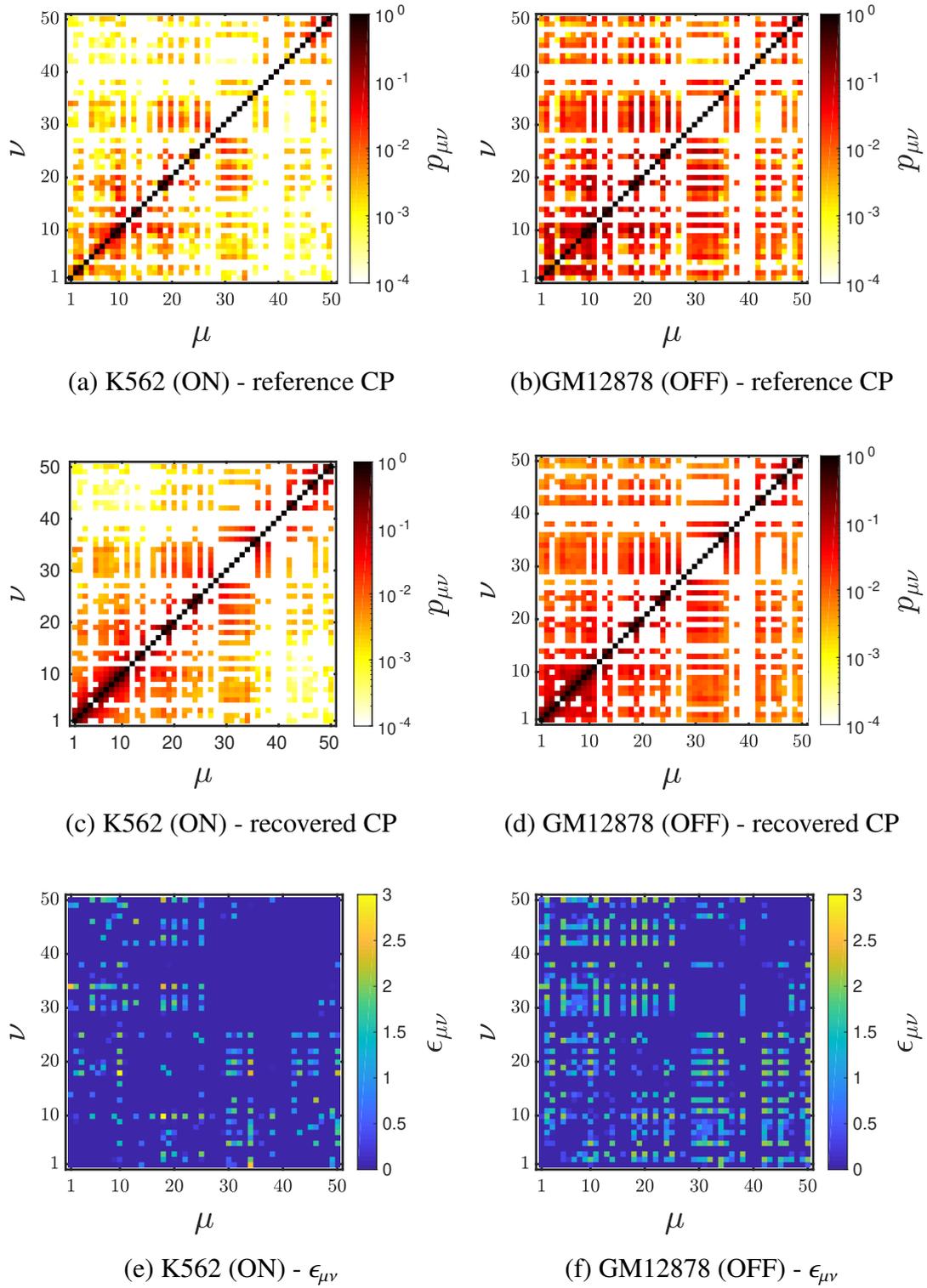
**Figure 5.1:** Comparison of the reference normalized contact probabilities ((a) and (b)) with the recovered contact probabilities ((c) and (d)), obtained with the IBD method for K562 and GM12878, respectively, at $N_f = 0$. The value of interaction strength parameter $\epsilon_{\mu\nu}$ for (e) K562 (ON state) and (f) GM12878 (OFF state) cell lines, respectively, at the converged state.

ment and total number of contacts are not quantified. This data cannot be compared across cell lines or across different experimental sets. While several normalization techniques exist, the ICE method is one of the more widely used techniques, where through an iterative process biases are removed and equal "visibility" are provided to each bins/segments of the polymer. The resulting contact count matrix is a normalized matrix where $\sum_\mu C_{\mu\nu} = 1$. While the existing normalization techniques help in removing biases, they still only give relative contact probabilities and not the absolute values. To accurately predict the distance between any two segments in chromatin, it is essential to know their absolute contact probabilities. Since the total number of genome equivalent (number of cells) cannot be estimated in a chromosome conformation capture experiment, the calculation of absolute contact probability from the contact count is highly challenging. A simple technique to normalize these counts is described here. The contact count matrix can be normalized by imposing the following constraint, namely, that the sum of times any segment pairs $(\mu\nu)$ are in contact $(C_{\mu\nu}^{\text{c}})$ and the number of times they are not in contact $(C_{\mu\nu}^{\text{nc}})$ must be equal to the total number of samples $N_s$. This is true for all bead-pairs i.e. $C_{\mu\nu}^{\text{c}} + C_{\mu\nu}^{\text{nc}} = N_s$, for all $\mu\nu$. Since only $C_{\mu\nu}^{\text{c}}$ is known, two limiting values of $N_s$ are estimated using the following scenarios. In one scenario, it is assumed that for the segment pairs $(\mu\nu)$ which has the largest contact count in the matrix, $\mu$ and $\nu$ are always in contact in all cells. In other words $C_{\mu\nu}^{\text{nc}} = 0$; in this case $N_s$ is simply equal to the largest element of the contact count matrix. Since this is the smallest value of $N_s$ possible, it is denoted by $(C_{\mu\nu}^{\text{c}})_{\text{max}} = N_{\text{min}}$. The other scenario estimates the sample size from the row $\mu$ for which the sum over all contact counts is the largest i.e., $N_s = $ maximum of $(\sum_\nu C_{\mu\nu}^{\text{c}})$. This assumes that $\nu$ is always in contact with only one other segment in a cell and there is no situation when it is not in contact with any segment. This case is denoted as $N_{\text{max}}$. However, in a real system, there might be situations where segment $\nu$ is not in contact with any of the remaining segments. In such a case, $N_s$ could be greater than $N_{\text{max}}$. We have investigated this question in the context of simulations, where we know the exact ensemble size, and can normalize the contact count matrix with the exact ensemble size, i.e., $N_s$. From this analysis, it was observed that there are very few samples where the bead $\nu$ is not in contact with any of the remaining beads. It supports our hypothesis that $N_{\text{max}}$ could be considered to be the upper limit in estimating the ensemble size $N_s$. Since the precise value of $N_s$ is not known in experiments, $N_s$ is varied as a parameter from $N_{\text{min}}$ to $N_{\text{max}}$. To systematically vary $N_s$,

for convenience, a parameter $N_f$ is defined,

$$N_f = \frac{N_s - N_{\min}}{N_{\max} - N_{\min}} \tag{5.1}$$

in the range of $[0, 1]$. Clearly, $N_f = 0$ implies $N_s = N_{\min}$, which is the lower bound for $N_s$ and $N_f = 1$ implies $N_s = N_{\max}$, which is the upper bound. The contact probabilities at various $N_f$ values are calculated as $p_{\mu\nu} = (C^c_{\mu\nu}/N_s)$ where $N_s = N_{\min} + N_f(N_{\max} - N_{\min})$.

For several values of $N_f$, the contact count matrices are normalized and IBD is carried out to obtain the optimal interaction strengths between the bead-pairs. Fig. 5.1(a) and 5.1(b) show the normalized contact probabilities at $N_f = 0$ for cell lines K562 (ON state) and GM12878 (OFF state), respectively (reference contact probabilities), when they are coarse-grained to 50 segments of length 10 kbp each, as per the procedure described above and the corresponding recovered contact probability matrices for both the cell lines from simulation are shown in Fig. 5.1(c) and 5.1(d). The corresponding optimized interaction energies ($\epsilon_{\mu\nu}$) are plotted in Figs. 5.1(e) and 5.1(f). The values range approximately from 0 to $3k_BT$. Given that typical contact probability numbers are very small, the optimized energies are just above thermal energy and are comparable to interaction energies of certain proteins.

### 5.2.1    ICE normalization

ICED package (http://projects.cbio.mines-paristech.fr/iced) was used to normalize the contact count matrix of $\alpha$-globin gene locus (both K562 and GM12878 cell line). The whole matrix was then divided by the summation of the row (summation of each row is a constant after the ICE normalization) so that the contact probability for each bin sums to 1.

Fig. 5.2(a) and 5.2(b) shows the ICE normalized matrix for K562 (ON state) and GM12878 (OFF state), respectively. To study the 3D configurations, we performed IBD on the ICE normalized contact counts matrix. The contact probabilities recovered in this process are represented in Fig. 5.2(c) and 5.2(d) for K562 and GM12878 cell line. It is evident from these figures that IBD algorithm has successfully recovered the ICE normalized contact matrix for both the states.

In order to compare the normalization method introduced in the current work with the normalization ICE procedure that is commonly used, we have also carried out the IBD procedure on an ICE normalized matrix. Clearly, the IBD method also recovers the

(a)K562 (ON) - reference

(b)GM12878 (OFF) - reference

(c)K562 (ON) - recovered

(d)GM12878 (OFF) - recovered
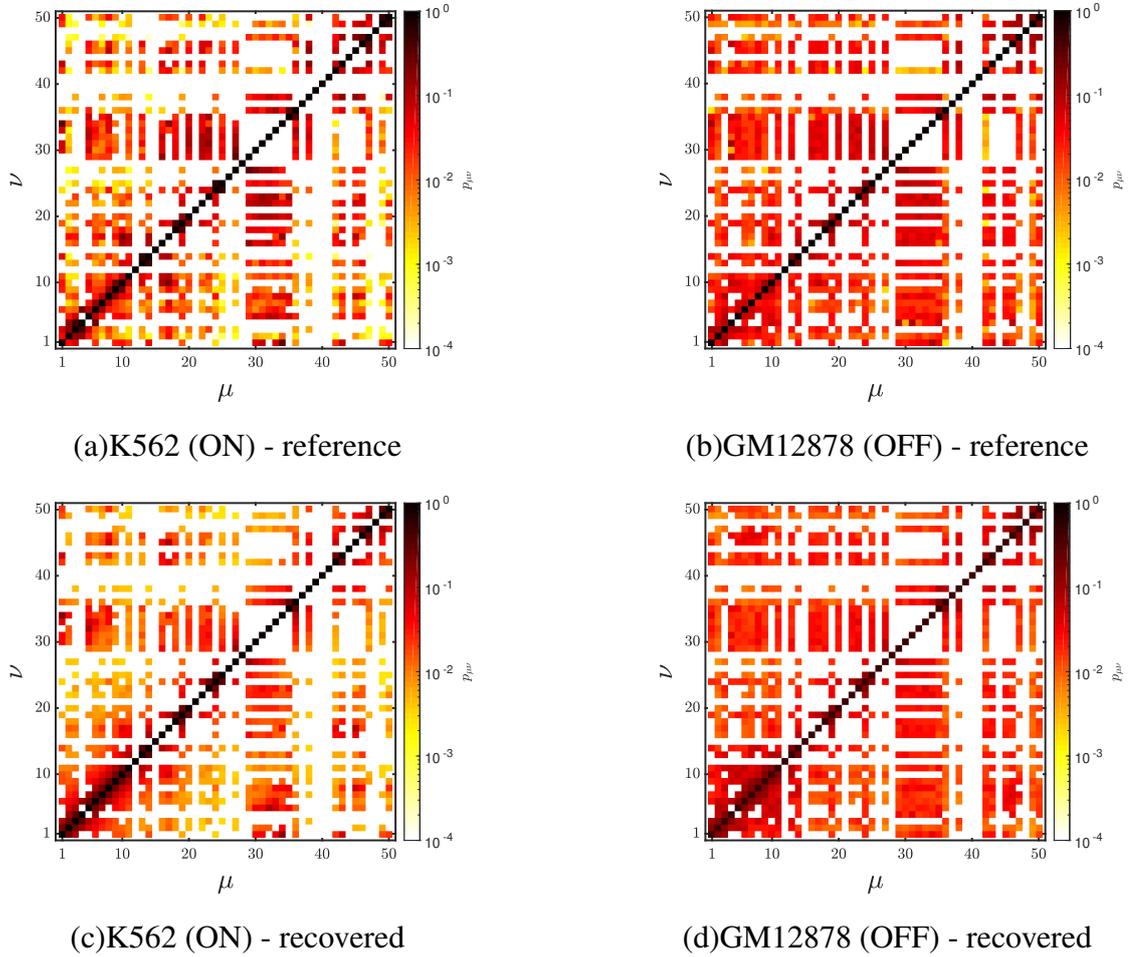
**Figure 5.2:** Comparison of the reference ICE normalized contact probabilities ((a) and (b)) with the recovered contact probabilities ((c) and (d)), obtained with the IBD method for K562 and GM12878, respectively.

contact probability matrix obtained with the ICE normalization. As will be discussed in further detail below, the normalization method has a significant effect on all the structural properties that have been evaluated in the current work.

## 5.3 Three-dimensional configuration of the $\alpha$-globin gene locus

The spatial extent of the chromatin polymer, as quantified by the square radius of gyration $R_g^2$, for different values of $N_f$ is presented in Fig. 5.3. In the case of the cell line where the
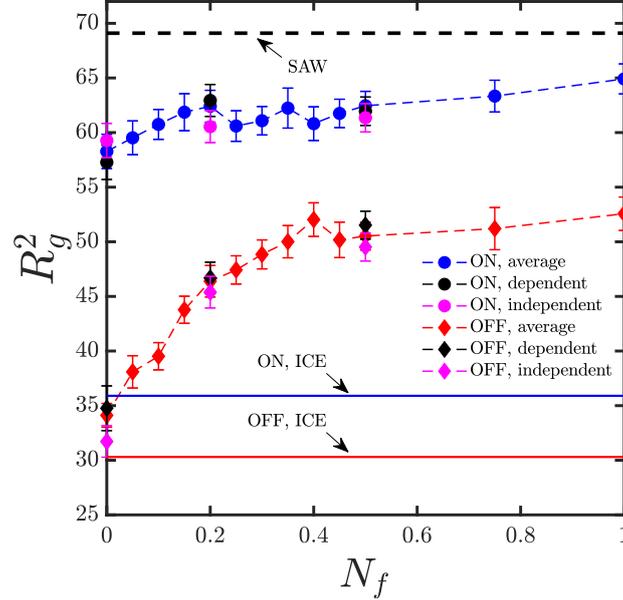
**Figure 5.3:** Spatial extension of the polymer chain, quantified by the radius of gyration, $R_g^2$, computed at various values of the normalization parameter $N_f$ (see Eq. 5.1 for definition), for both K562 (ON state) and GM12878 (OFF state) cell lines. All three coarse-graining techniques, i.e., dependent, independent and average, have been used. The black dashed line represents the value of $R_g^2$ for a chain executing self-avoiding walk statistics. Blue and red lines indicate the $R_g$ for ICE normalized ON and OFF state, respectively.

gene is ON (K562), the increase in $R_g^2$ for small values of $N_f$ is relatively less prominent and becomes nearly independent of $N_f$ as $N_f$ approaches one. It is clear that contact probabilities decrease with increasing $N_f$, since $N_s$ increases with $N_f$. It is consequently expected that with sufficiently large $N_f$, $R_g^2$ should approach the value for a self-avoiding walk. We have simulated a self-avoiding walk using the SDK potential with $\epsilon_{\mu\nu} = 0$; this represents a purely repulsive potential, and the result is shown as a black dashed line in Fig. 5.3. In the cell line where the gene is OFF (GM12878), the value of $R_g^2$ increases relatively rapidly for small values of $N_f$ and reaches a nearly constant value for $N_f \gtrsim 0.4$. However, the limiting value is significantly smaller than that of a self-avoiding walk. This suggests that some significant interactions are still present amongst the bead-pairs, even for $N_f$ approaching one. The influence of the different coarse-graining procedures was examined and it was found that the value of $R_g^2$ from all the three coarse-graining

procedures agreed with each other within error bars (as seen from the data at $N_f = 0, 0.2$ and 0.5, for both the cell lines). This suggests that, at least as far as $R_g^2$ is concerned, the choice of coarse-graining method is not vitally important.

However, the IBD results for ICE-normalized reference contact probability predicts a very different value for $R_g$ of the ON state (blue line) and OFF (red line) state. As can been seen, the $R_g^2$ for ON state using ICE normalization is close to the $R_g^2$ obtained here for OFF state at $N_f = 0$. Interestingly this similarity is observed for many of the properties considered here, as will be discussed in more detail below.

### 5.3.1 Shape functions

Since chromatin folded in 3D can have spatial organization that is beyond simple spherically symmetric packing, various non-globular 3D shape properties have been analysed here. Eigenvalues of the radius of gyration tensor for polymer chains are usually reported in terms of ratios, either between individual eigenvalues, or with the mean square radius of gyration. For a chain with a spherically symmetric shape about the centre of mass, we expect $\langle \lambda_i^2 \rangle / \langle R_g^2 \rangle = 1/3$, for $i = 1, 2, 3$, and $\langle \lambda_i^2 \rangle / \langle \lambda_j^2 \rangle = 1$ for all combinations $i$ and $j$. For chain shapes with tetrahedral or greater symmetry, the asphericity $B = 0$, otherwise $B > 0$. For chain shapes with cylindrical symmetry, the acylindricity $C = 0$, otherwise $C > 0$. With regard to the degree of prolateness, its sign determines whether chain shapes are preponderantly oblate ($S \in [-0.25, 0]$) or prolate ($S, \in [0, 2]$). The relative anisotropy ($\kappa^2$), on the other hand, lies between 0 (for spheres) and 1 (for rods).

All these properties are investigated for $N_f = 0, 1$ and for the ICE normalization, and compared in the ON and OFF states, as displayed in Table 5.1. It is clear that the while the chain is highly non-spherical in both states, it appears to be slightly more spherical in the OFF than in the ON state. The biggest difference is observed at $N_f = 0$ between ON and OFF states. As we approach $N_f = 1$, the difference between ON and OFF states is not so significant. With ICE, there is not much difference between the two states. As previously observed with the radius of gyration, ICE values are very close to the OFF state at $N_f = 0$.

### 5.3.2 Density profiles

To get a different prospective on the 3D organization of the gene, the density distribution about the centre of mass was considered. In order to do this, all polymer configurations were aligned along the major axis of the radius of gyration tensor **G** and each bead
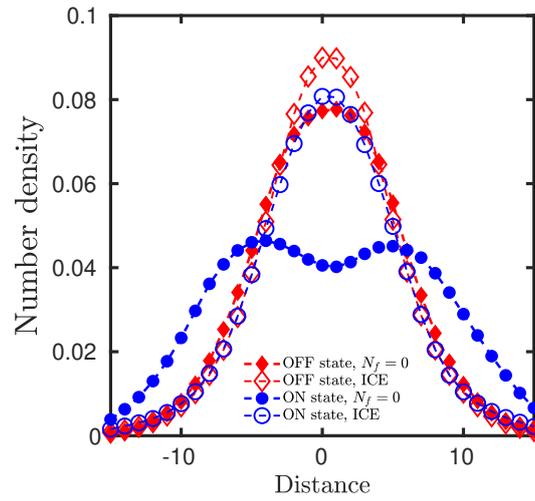
(a)
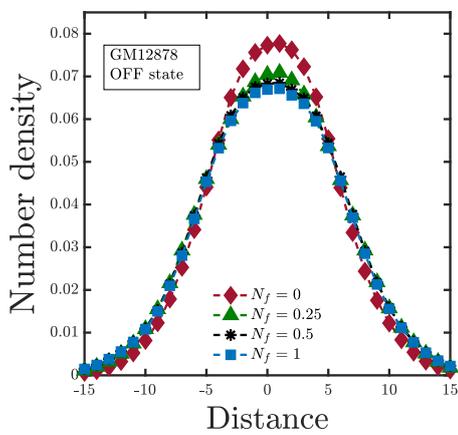


(b)
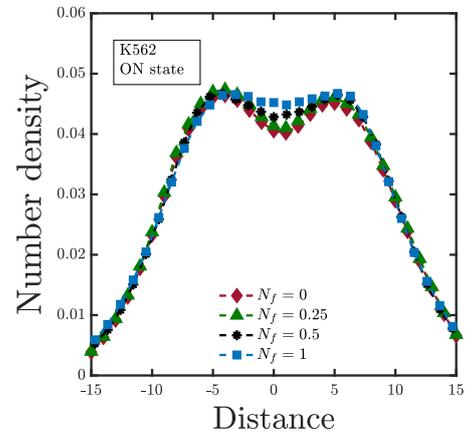


(c)

**Figure 5.4:**   Comparison of the number density of beads along the major axis of the radius of gyration tensor, for various values of the normalization parameter $N_f$ (see Eq. 5.1 for definition), (a) ON and OFF states at $N_f = 0$, (b) the OFF state, and (c) the ON state for various values of $N_f$.
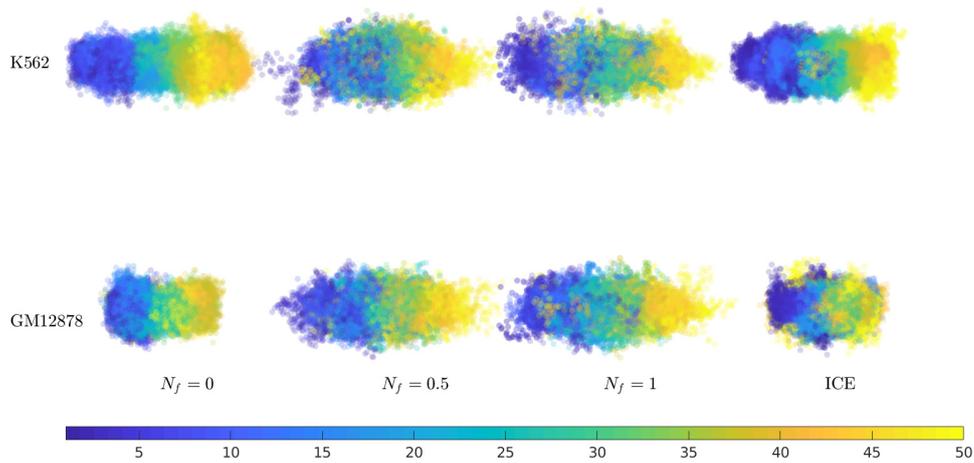
**Figure 5.5:** Snapshots of 3D configurations, obtained by aligning chains along the major axis of the radius of gyration tensor and superimposing them on top of each other with transparency. Configurations at different values of the normalization parameter $N_f$ (see Eq. 5.1 for definition) are displayed for cell lines K562 and GM12878. The colour assigned to each marker (blue to yellow) represents the bead number along the contour length (bead 1 to bead 50) of the polymer chain.

**Table 5.1:** Various shape property based on the eigen values of gyration tensor **G** are defined here for $N_f = 0$, $N_f = 1$ and ICE normalized contact matrix for K562 (ON state) and GM12878 (OFF state) cell line.

| Shape properties | K562 (ON state) | | | GM12878 (OFF state) | | |
|:---:|:---:|:---:|:---:|:---:|:---:|:---:|
| | $N_f = 0$ | $N_f = 1$ | ICE | $N_f = 0$ | $N_f = 1$ | ICE |
| $\langle \lambda_1^2 \rangle / R_g^2$ | 0.058 | 0.057 | 0.078 | 0.081 | 0.066 | 0.083 |
| $\langle \lambda_2^2 \rangle / R_g^2$ | 0.164 | 0.175 | 0.189 | 0.201 | 0.177 | 0.195 |
| $\langle \lambda_3^2 \rangle / R_g^2$ | 0.778 | 0.768 | 0.732 | 0.718 | 0.757 | 0.722 |
| $\langle \lambda_2^2 \rangle / \langle \lambda_1^2 \rangle$ | 2.828 | 3.054 | 2.417 | 2.479 | 2.703 | 2.357 |
| $\langle \lambda_3^2 \rangle / \langle \lambda_1^2 \rangle$ | 13.412 | 13.422 | 9.356 | 8.874 | 11.563 | 8.727 |
| $B/R_g^2$ | 0.667 | 0.652 | 0.599 | 0.578 | 0.636 | 0.583 |
| $C/R_g^2$ | 0.106 | 0.118 | 0.111 | 0.120 | 0.112 | 0.112 |
| $S$ | 0.913 | 0.816 | 0.988 | 0.772 | 0.926 | 0.867 |
| $\kappa^2$ | 0.545 | 0.513 | 0.537 | 0.452 | 0.525 | 0.497 |

position was binned and the number density of beads along the major axis was computed. As displayed in Fig. 5.4(a), in GM12878 (OFF state) cells, the number density shows a single peak at the center of mass position suggesting a symmetric organization around the centre of mass along the major axis. In the case of K562 (ON state) cells, the number density is seen to have a double peak, implying a bimodal distribution of polymer beads around the centre of mass along the major axis (Fig. 5.4(a)), as suggested by earlier 3D models for the α-globin gene [9, 99]. With an increase in $N_f$, a slight decrease in the number density at the core of the α-globin gene in the OFF state is observed (Fig. 5.4(b)), while a decrease in extent of bimodality is observed in the ON state (Fig. 5.4(c)). However, the differences for different $N_f$ values are less prominent at the peripherial regions of the globule.

We have also compared the density profile corresponding to the ICE-normalized matrix, displayed in Fig. 5.4(a) along with $N_f = 0$. With the ICE normalization, both states (ON and OFF) show a single peak at the centre of mass. The bimodal nature of the ON state is no longer observed. This is a clear prediction that distinguishes the ICE-normalized result from the other results and can be tested in future experiments.

### 5.3.3   3D conformations

To obtain a snapshot of the 3D structure of the $\alpha$-globin gene locus, 1000 different configurations from the ensemble were aligned along its major axis and then superimposed on top of each other, as displayed in Fig. 5.5, for both the cell lines at different values of $N_f$ and with the ICE normalization. Each dot represents a bead and to make them visible, they have been made transparent to some degree. Different colors in the plot represent the bead number along the contour length of the polymer chain. As indicated from the shape functions and the density profiles, the snapshot shows that the structure is highly non-spherical in both cases, In particular, the K562 (ON state) cell line chromatin has a more extended configuration, with slightly higher density away from the centre of mass. As can be seen in Fig. 5.5, the snapshot for $N_f = 0$ has some differences with snapshots for larger $N_f$ values. The value of $N_f$ was seen earlier to affect average properties like $R_g^2$ (Fig. 4). The snapshots in Fig. 5.5 show a similar behaviour as $R_g$ reflecting the variation for small $N_f$ and saturation for larger $N_f$.

## 5.4   3D spatial distances and contact probabilities

The 3D conformation of the $\alpha$-globin gene locus has been investigated earlier [9, 100]. These studies differ from the current work in some important aspects. Firstly, they assume that the contact counts between any two pairs can be converted to an equilibrium distance between those pairs through a certain pre-determined functional form. Secondly, instead of optimizing the interaction strengths to recover the contact counts, their simulations attempt to recover the equilibrium distances that have been derived from contact matrices. It is not clear in these cases whether the experimentally observed contact counts will be recovered by simulations. In this work, no assumptions have been made about the relationship between spatial distance and contact probability for any pair of beads. On the contrary in the present case, we can compute the mean spatial distances $\langle r_{\mu\nu} \rangle$ that are consistent with the contact probability matrix. Further, no configuration from the ensemble is discarded.

The spatial distances calculated in the current work for the contact probabilities in the ON and OFF state are shown in Fig. 5.6(a) for K562 (ON state) and in Fig. 5.6(b) for GM12878 (OFF state) cell lines for $N_f = 0$. Each point in these figures represents the ensemble-averaged 3D distance between a given pair of beads ($y$-axis) having a contact

(a)



(b)
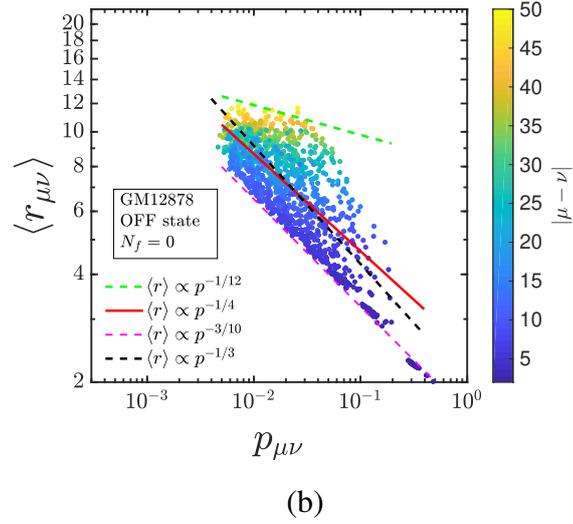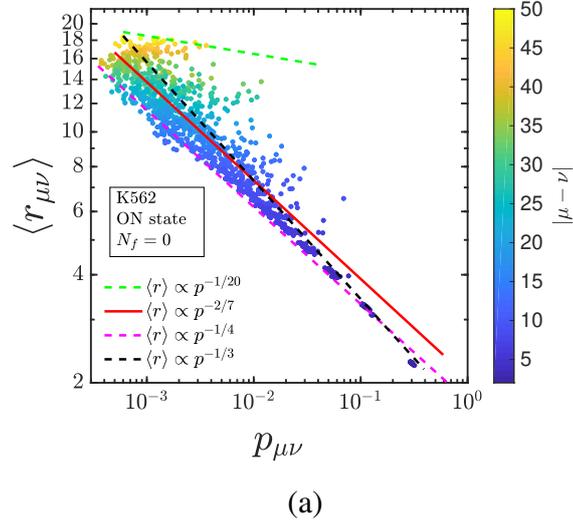
**Figure 5.6:**  Dependence of mean 3D distances $\langle r_{\mu\nu} \rangle$ on contact probabilities $p_{\mu\nu}$ for (a) K562 (ON state) and (b) GM12878 (OFF state) cell lines, respectively for $N_f = 0$. For the K562 (ON state) cell line, the contact probabilities are bounded by power laws, $\langle r_{\mu\nu} \rangle \propto p_{\mu\nu}^{\tau}$, where $\tau$ varies from $-1/20$ (upper bound) to $-1/4$ (lower bound) as indicated by the green and magenta dashed line. Similarly, in the GM12878 (OFF state), $\tau$ varies from $-1/12$ to $-3/10$. The red line indicates the power law fitted to the simulation data points. The black dashed line represent the analytical relation between the contact probability and a spatial distance for an ideal polymer chain.
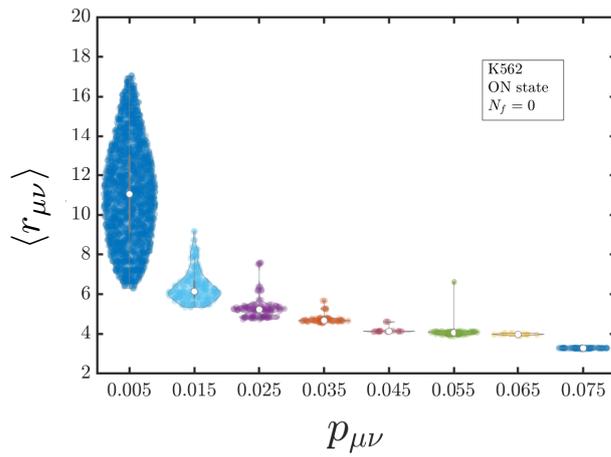
(a)
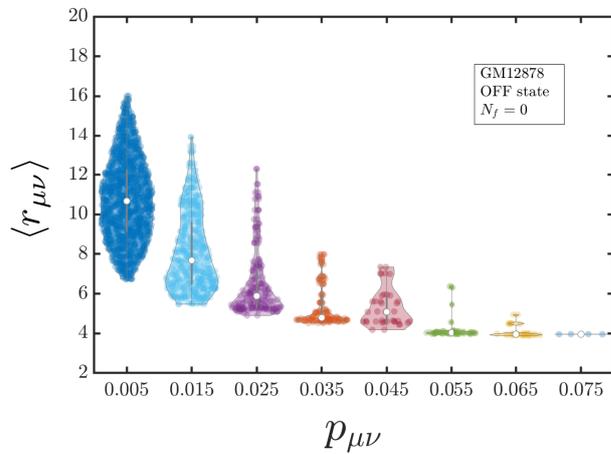


(b)

**Figure 5.7:** Violin plots which display the probability distribution of mean 3D distances for selected ranges of contact probabilities in (a) the K562 (ON state) cell lines, and (b) the GM12878 (OFF state) cell lines for $N_f = 0$.
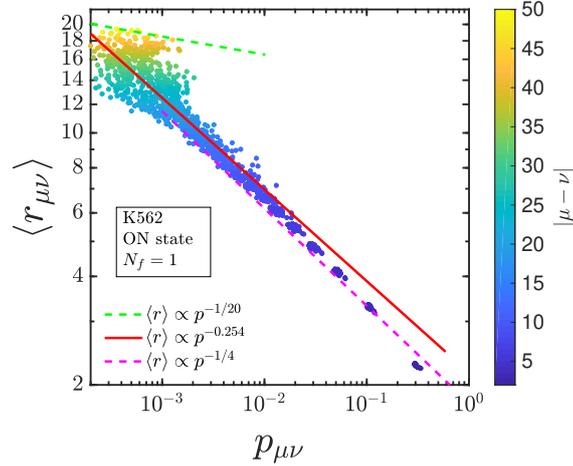
(a)
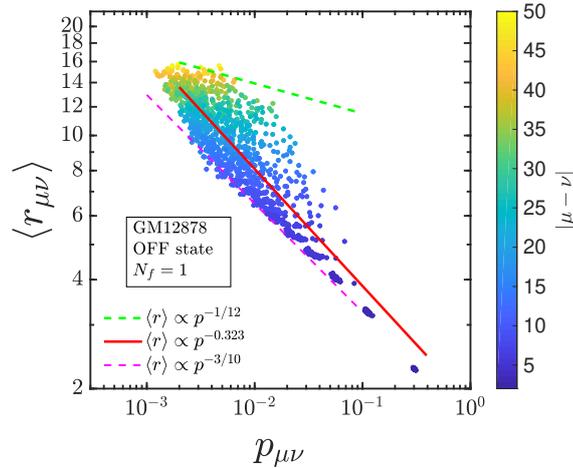


(b)

**Figure 5.8:** Dependence of mean 3D distances $\langle r_{\mu\nu} \rangle$ on contact probabilities $p_{\mu\nu}$ for (a) K562 (ON state) and (b) GM12878 (OFF state) cell lines, respectively at $N_f = 1$. For the K562 (ON state) cell line, the contact probabilities are bounded by power laws, $\langle r_{\mu\nu} \rangle \propto p_{\mu\nu}^{\tau}$, where $\tau$ varies from $-1/20$ (upper bound) to $-1/4$ (lower bound) as indicated by the green and magenta dashed line. Similarly, in the GM12878 (OFF state), $\tau$ varies from $-1/12$ to $-3/10$. The red line is fitted to the data points.
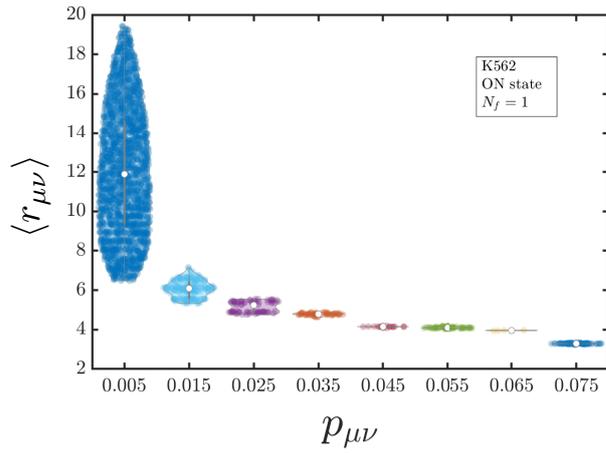
(a)
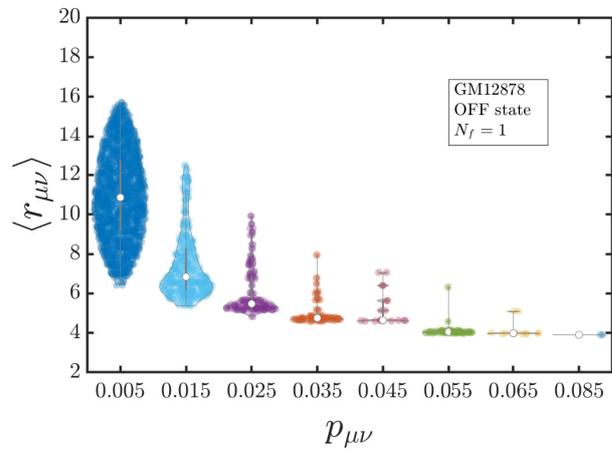


(b)

**Figure 5.9:** Violin plots showing the probability distribution of mean 3D distances for selected ranges of contact probabilities in (a) the K562 (ON state) cell lines, and (b) the GM12878 (OFF state) at normalization factor $N_f = 1$.
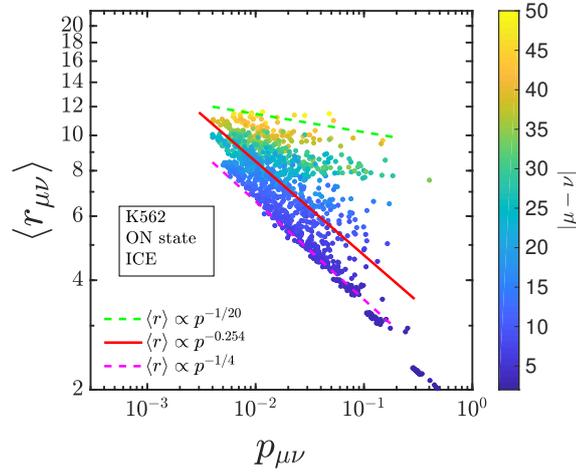
(a)
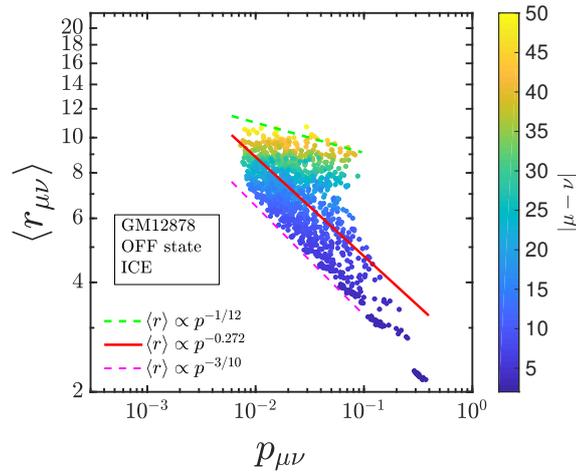


(b)

**Figure 5.10:** Dependence of mean 3D distances $\langle r_{\mu\nu} \rangle$ on contact probabilities $p_{\mu\nu}$ for (a) K562 (ON state) and (b) GM12878 (OFF state) cell lines, respectively for ICE normalization. For the K562 (ON state) cell line, the contact probabilities are bounded by power laws, $\langle r_{\mu\nu} \rangle \propto p_{\mu\nu}^{\tau}$, where $\tau$ varies from $-1/20$ (upper bound) to $-1/4$ (lower bound) as indicated by the green and magenta dashed line. Similarly, in the GM12878 (OFF state), $\tau$ varies from $-1/12$ to $-3/10$. The red line indicates the power law fitted to the simulation data points.
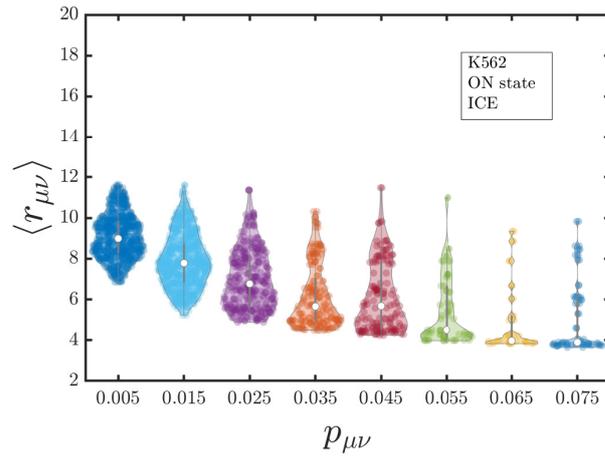
(a)
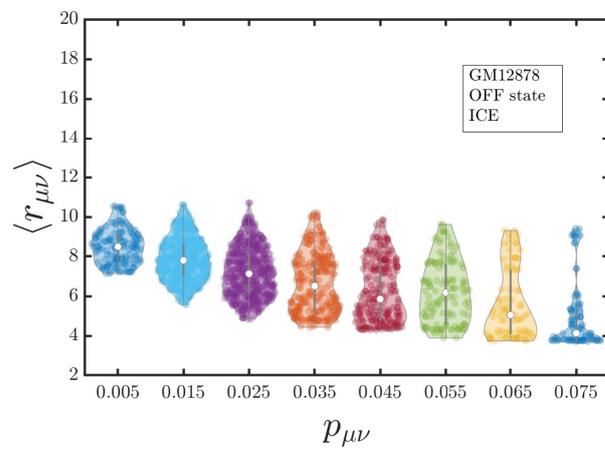


(b)

**Figure 5.11:** Violin plots which display the probability distribution of mean 3D distances for selected ranges of contact probabilities in (a) the K562 (ON state) cell lines, and (b) the GM12878 (OFF state) cell lines normalized with ICE.

probability as indicated in the *x*-axis. As is immediately apparent, a wide range of 3D distances is possible, unlike what was assumed in earlier studies. It appears that the average 3D distance is not just a function of contact probability $p_{\mu\nu}$ (where the interaction between the beads plays a role), but is also a function of the distance along the contour between the beads ($|\mu - \nu|$) – the color variation in Figs. 5.6(a) and (b) indicates the influence of contour length. The red line in both the figures are fitted power-laws to the data.In both cases, the exponents are close to $-1/4$. But the interesting element here is the variability (scatter) in the data which shows that for a given contact probability value, there can be multiple values of 3D distances, with deviation of many units.

To understand this variability better, we bin the same data and plot it as violin plots that display the mean 3D distance for a given small range of contact probabilities, as shown in Figs. 5.7 (a) and (b). It is clear that the distribution of points around the mean is very diverse – bimodal in a few cases and with an extended tail in many cases – suggesting that a simple functional form between the mean 3D distance and the contact probability may not be feasible. It must be reiterated here that many previous studies have assumed power law relations such as $r_{\mu\nu} \propto p_{\mu\nu}^{\tau}$ can be used, with exponents $\tau = -1$ [47, 38] and $\tau = -1/2$ [108], independent of $|\mu - \nu|$. Some groups have also assumed exponential [132] and logarithmic decay of distance with probability [9]. As shown above, the results reported here do not support the usage of such simple functional forms. However, for an ideal chain, we know that contact probability $p \propto s^{-3/2}$ and the average 3D distance scales as $r \propto s^{1/2}$ where $s$ is the contour length between any two polymer beads. Combining these two, we get $r \propto p^{-1/3}$. This is shown by the black dashed line in Fig. 5.6. Clearly, the relation between mean 3D distance and the contact probability is significantly more complex than for a simple ideal chain. The relationship and its variability for $N_f = 1$ and ICE normalization are in Fig 5.8 to 5.11 respectively. In these instances as well, the mean 3D distance is observed to be a function of both the contact probability and contour distance $|\mu - \nu|$.

The spatial distances calculated for the ICE normalised contact probabilities in the ON and OFF state are shown in Fig. 5.10(a) for K562 (ON state) and in Fig. 5.10(b) for GM12878 (OFF state) cell lines. Similar to the $N_f = 0$ case, a wide range of 3D distances is possible and is also dependent on the distance along the contour between the beads ($|\mu - \nu|$). The same data has been binned and plotted as violin plots that display the mean 3D distance for a given small range of contact probabilities, as shown in Figs. 5.11 (a) and (b).

This distribution is observed to be very different from the other simple normalization studied in the present thesis.

## 5.5 Comparison with experimental observations

To establish a connection between our simulations and experiments, we computed the average 3D spatial distance ($r_{\mu\nu}^* = |\mathbf{r}_\mu^* - \mathbf{r}_\nu^*|$) between different pairs of chromatin segments as a function of the corresponding 1D genomic separation ($s_{\mu\nu} = |\mu - \nu|$). As shown in Fig. 5.12, $\langle r_{\mu\nu}^* \rangle \sim s_{\mu\nu}^\nu$ with $\nu = 0.38$ in the OFF state, suggesting near close packing within the chromatin domain (red symbols). As a "control", we also simulated a self-avoiding walk (SAW) polymer with no crosslinking ($\epsilon_{\mu\nu} = 0$) which results in $\nu = 0.6$ as expected (black symbols) [21]. Our chromatin model predicts a spread (variability) in the 3D distance (red symbols) which is absent in the control revealing the implications of heterogeneous intra-chromatin interactions. A recent microscopy study [129] on a mouse ESC chromatin domain also showed a similar behavior – both the scaling (slope) and the variability in the experimental and simulation data are comparable without any fitting parameter. This is a validation that macroscopic polymer properties of chromatin domain in our simulation accurately represent what is observed in realistic systems. The $y$-intercept of the experimental data gives us the size ($\sigma = l_H$) of the 10kb chromatin (a single bead in our simulation). For this experimental system (mouse chromosome 6, 1.2MB in Szabo et al. [129]) we get $l_H = 22$nm. Even though we do not have such extensive spatial distance data for $\alpha$ globin, we compared the available FISH data for $\alpha$ globin and deduced the $l_H = 36$nm (see Appendix A.2). Throughout this paper, $l_H = 36$nm and $\lambda_H = 0.1$s are used to convert all non-dimensional lengths and times into standard units and we will present quantities in both units. The reasons for the choice of both these specific values are discussed in greater detail in the Appendix A.2.

## 5.6 Role of adjacent fragments

In our simulation of the $\alpha$-globin gene locus, we modeled only the chromatin segment representing the locus and not the whole chromosome. It is worthwhile studying the changes in 3D configurations due to chromatin segments attached at both the ends of the domain. In this context, one can ask the following questions: Do the domain structure and properties remain unaffected by the chromatin segments on both sides or is there is
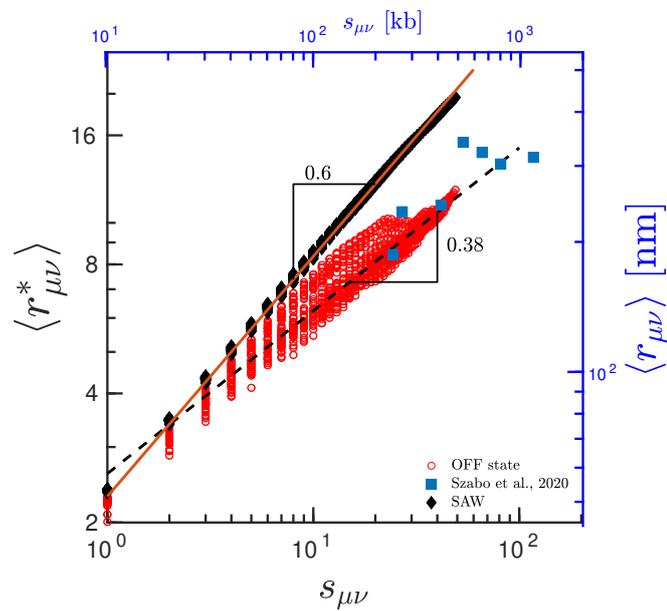
**Figure 5.12:** Average 3D distances between all bead-pairs as a function of corresponding genomic distances for the control simulations (SAW, black symbols), chromatin domain that we simulated (red symbols) and comparison with experimental data from [129] (blue symbols). The major axes (lower $x$ and left $y$) represent quantities in dimensionless units (see methods) while the other axes (upper $x$ and right $y$) represent the same in standard units.

a drastic change due to the attached fragments at the ends? To probe this, one can add short chain segments to both the ends of the domain. These short chains could be either a simple polymer with steric hinderance (SAW) or a collapsed chain. An ideal study would consider all possible combinations of SAW and collapse chains at the ends, which would be:

1. SAW chains on both ends of the $\alpha$-globin domain.

2. A SAW chain on one side and a collapsed chain on the other.

3. Collapsed chains on both sides.

In our study we have considered only the first scenario of a SAW chain of 30 beads attached to both ends of the domain. The interaction of the middle portion of the chain representing the $\alpha$-globin domain remains the same. We performed Brownian dynamics on a 110 beads chain for both cell lines, ON and OFF states. The interaction strength parameters used for this simulation are the ones recovered from the IBD for 50 beads chain. We equilibrated the chain for 8 Rouse relaxation times and sampled at every 1000 time steps. A total of $10^5$ samples were collected.

Fig. 5.13(a) and (b) show the contact probability obtained from simulation for OFF and ON state of the elongated chain ($N = 110$), respectively. As seen in Fig. 5.13(a), significant contacts are only present at the interior of the chain. This region accounts for the interaction of the OFF state of $\alpha$-globin gene locus. As expected from the SAW polymer, end portion of the chains are devoid of any significant far away contacts. Only neighbouring beads are observed to be in contact. A similar pattern is seen for the ON state of elongated chain (see Fig. 5.13(b)), where most of the significant contact is for the middle portion of the chain, corresponding to the ON state of $\alpha$-globin gene locus.

We then looked at the size of the elongated polymer chain. Fig, 5.14 shows the radius of gyration ($R_g$) in the two cases of elongated chains and compared with our earlier results. This shows that indeed with the addition of chains on both side, the $R_g$ of the polymer increases in both the ON and OFF states. We then investigated the effect of chains at the end on the density profile. Fig. 5.15 shows the plot for the density profile. In the OFF state (red symbols in Fig. 5.15), with the addition of chain at the end, we observe the decrease in peak as increase in the spread (standard deviation of the curve). Both curves show a single peak. In the ON state (blue symbols in Fig. 5.15), the nature of the curve
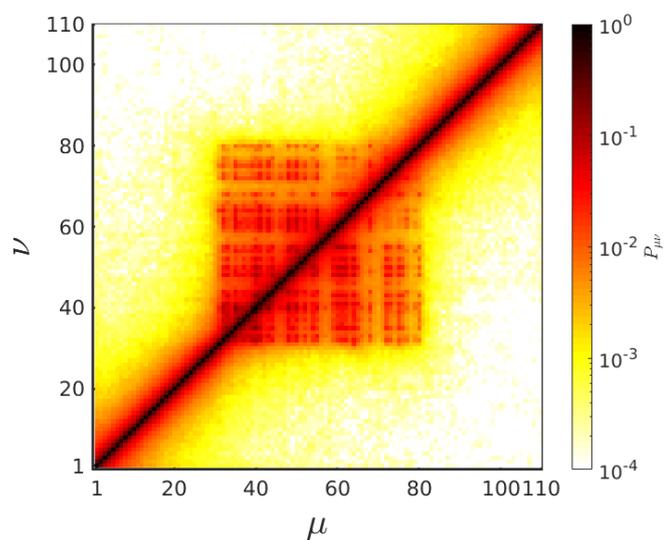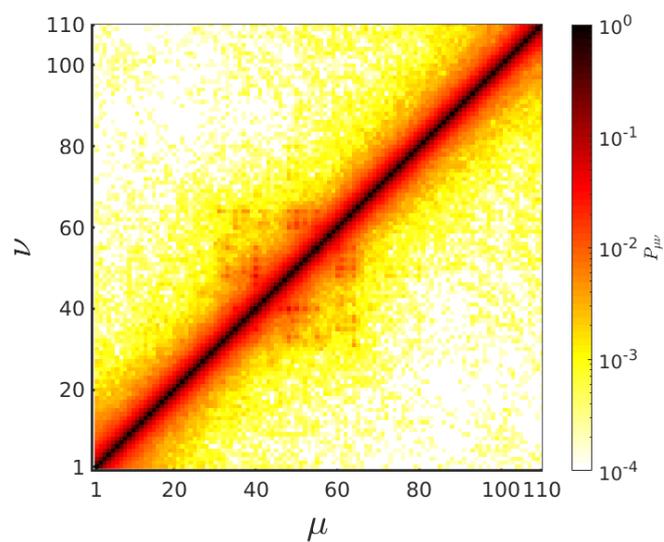
**Figure 5.13:** Contact probability matrix for the 110 beads long chain where the middle portion (bead 31-80) represent the $\alpha$-globin gene locus (a) for the OFF (GM12878) state and (b) for the ON (K562) state.
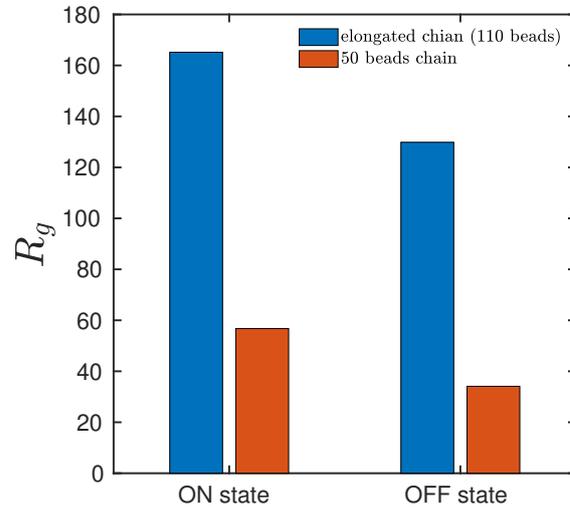
**Figure 5.14:** Comparison between $R_g$ of the previous 50 beads chain and the elongated 110 beads chain. The $R_g$ of the polymer increases with the addition of extra piece of chain on both sides.
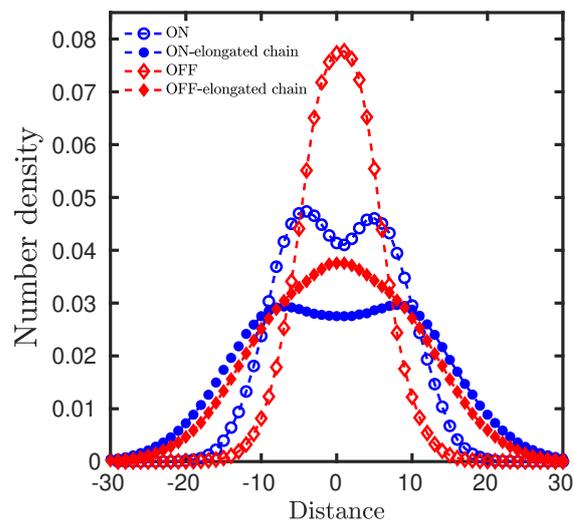


**Figure 5.15:** Density profile for 50 beads chain and the elongated 110 beads chain for the ON (K562) and OFF (GM12878) state.

(a)                                                                      (b)

(c)                                                                      (d)
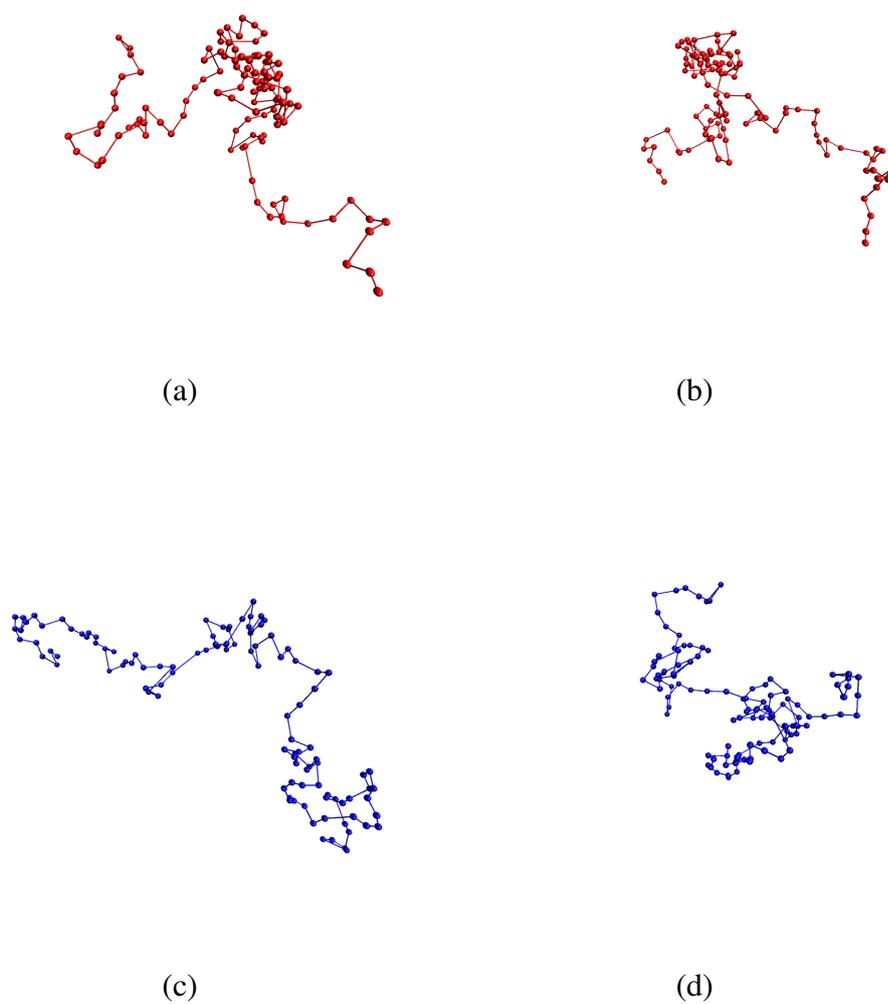
**Figure 5.16:** (a) & (b) Randomly chosen configuration for the OFF (GM12878) state. It can be observed that the middle portion of the chain remains in the collapsed state, while the ends stay in open configuration. (c) & (d) Randomly chosen configuration for the ON (K562) state. As seen earlier in this thesis, ON state is close to the SAW state, the whole polymer seems to exist in open configuration.

in both the cases ($N = 50$ and $N = 110$) remains the same. Both show a double peak in density profile. Similar to the OFF state, here also, we observe the decrease in height and increase in the spread of the density profile for the elongated chain. This observation is consistent with the observation in $R_g$. To understand the configurations of the elongated chain, we looked at some of the random configurations as shown in Fig. 5.16. In the OFF state, we observe the bead clustering for the middle portion of the chain only, while the segments of the chain at the ends remains in open state. For the ON state, we observe the whole chain in an open configuration. This is consistent with our earlier observation that ON state is close to the SAW and exists in an open configuration.

## 5.7   Summary

In the present chapeter, the 3-dimensional organisation of chromatin based on publicly available chromatin conformation capture experimental data is investigated using the IBD method. We used three coarse-graining procedures – *independent*, *dependent* and *average* were used to map between the experimental and coarse-grained contact matrices. For the gene locus studied in this work ($\alpha$-globin gene), no significant differences between the three cases was observed both for the gene extension and density profile. A procedure for normalizing the contact count matrix was introduced with a parameter $N_f$ varying from 0 to 1 between two different scenario for the sample size. For GM12878 (OFF state), the gene extension rapidly increases initially. while for K562 (ON state), on the other hand, is already in the extended form, it has a very small scope of further extension with increase in the normalization. Since there is relationship between the normalization factor $N_f$ and the physically measured properties such as the radius of gyration, It is conceivable that the $N_f$ can be estimated from experiments such as FISH, Chip-seq. The structural properties of the $\alpha$-globin gene locus were investigated in terms of shape function, density distribution and 3D snapshots. In K562 (ON state), alpha-globin lacks any prominent interaction and exists in an extended structure. Whereas in the case of GM12878 (OFF state), the gene is in a folded state. This is also consistent with the theory, as in ON state, the transcription factors need to access the gene in ON state (K562) while the structural status of OFF state (GM12878) should be avert the transcription factor resulting in gene silencing. The density profile along the major axis also supports the extended structure in cell line K562 (ON state) and a sharpe cluster of monomers at the core of GM12878 (OFF state). The dependency of spatial distance on contact probability is also investigated

here showing a broad distribution. Finally, we compared our prediction of spatial distance between chromatin segments with the published experimental distance data. The scaling and variability of our prediction is in agreement with the experimental data without any fitting parameter.

# Chapter 6

# Fluctuations and dynamics of a chromatin domain

So far, we have studied the 3-dimensional configurations and static properties of a chromatin domain. While these domains are often imagined as static structures, they are highly dynamic and show cell-to-cell variability. Since processes such as gene regulation and DNA replication occur in the context of these domains, it is important to understand not only their organization but fluctuation and dynamics as well. In this chapter, we go beyond the average static properties and compute the fluctuations of chromatin segments and their folding dynamics. In sec. 6.1, we examine the full distance distribution between different pairs of segments. To investigate the epigenetic effect, we altered the interaction strength and studied the size and shape changes in sec. 6.2. We then study the dynamics of the domain, compute relaxation times, stiffness and viscous drag experienced by the domain in sec 6.3. In sec. 6.4, we computed the loop formation times and contact times between different segments of chromatin. The probability distribution of temporal quantities is provided in sec. 6.5, followed by the summary of our findings.

## 6.1    Distance distributions and cooperative nature of chromatin folding

Even though the average distance between two chromatin segments is often used to represent chromatin organization, this may not describe the accurate biological picture in a dynamic, heterogeneous context. We compute the distribution of 3D distance between dif-
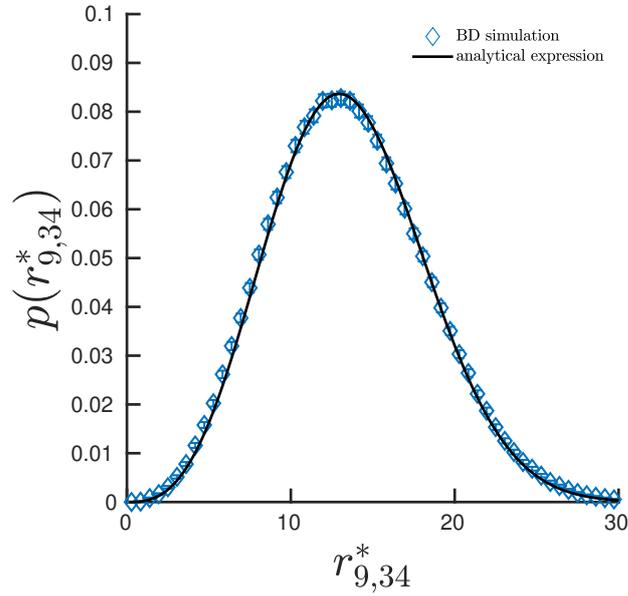
**Figure 6.1:** Distance probability distribution $p(r^*_{\mu\nu})$ from simulations compared with the analytical expression [32] for a pair of beads 9 and 34 of a SAW polymer.

ferent segments, $p(r^*)$, as it captures the maximum information about variability and fluctuations of chromatin. As a control, we computed the $p(r^*)$ for a SAW and it agrees well with the known analytical expression of des Cloizeaux, $p(r^*) = C(r^*)^{\theta+2}e^{-(Kr^*)^{1/(1-\nu)}}$ [32]. Here $\nu$ is the Flory exponent, $\theta$ is a geometrical exponent and the coefficients $C$ and $K$ are functions of $\theta$ [32]. Recent work has led to an accurate estimation for these constant which are discussed in Appendix A.3. Fig. 6.1 shows the validation with intermediate beads of a SAW (with $\nu = 0.6, \theta = 0.81, K = 1.17, C = 2.05$).

We then studied the $p(r^*)$ for the $\alpha$-globin gene locus in GM12878 cell type. Examining various segments 250kb apart along the chain backbone (25 beads), we find that all the distributions have a broad peak near their respective average distances (Fig. 6.2(a)). However, for bead pairs having high $\epsilon$ values, a sharp peak emerges near $r^* \approx r_c$ — we call this an "attraction-driven peak" as it is within the attractive range of the potential. The height of the peak is correlated with the strength of attraction ($\epsilon$). However, the average distances (vertical lines in Fig. 6.2(a)) appear independent of $\epsilon_{\mu\nu}$. This difference is also reflected in the cumulative distribution function as shown in Fig. 6.2(b). Together, these results imply that average distances between bead-pairs may not represent the complete picture of chromatin organisation; understanding the whole distribution is necessary.
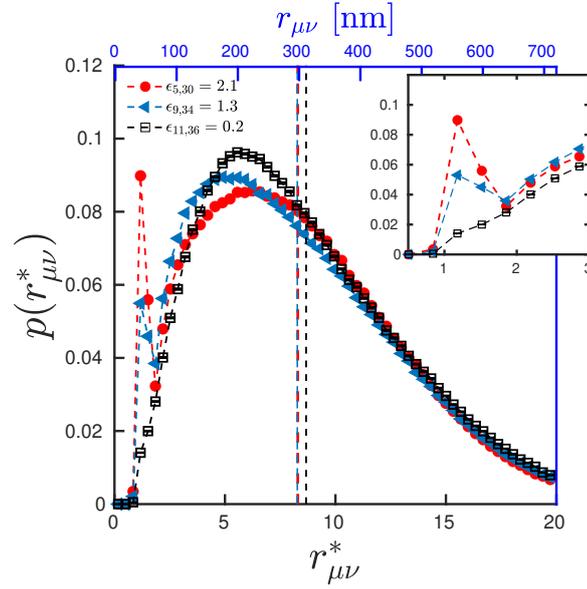
(a)
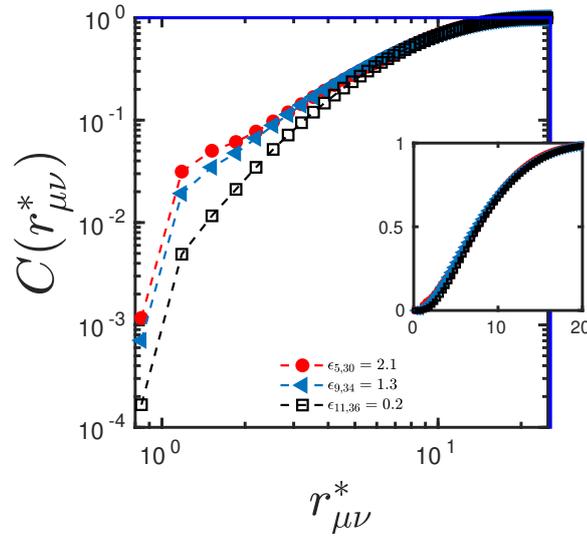


(b)

**Figure 6.2:** (a) $p(r^*_{\mu\nu})$ for various bead pairs with different $\epsilon_{\mu\nu}$, but same $s_{\mu\nu} = 25$ in the OFF state of $\alpha$-globin gene. The interaction-driven peak is highlighted in the inset. Vertical dashed lines represent $\langle r^*_{\mu\nu} \rangle$ corresponding to each distribution. (b) Cumulative distance distribution $C(r^*_{\mu\nu})$ for various bead-pair with the same genomic separation $s_{\mu\nu} = 25$ in the OFF state of $\alpha$-globin gene in log-log scale. The same is indicated in linear scale in the inset.

(a) OFF state (GM12878)



(b) OFF:GT1



(c) OFF:GT2
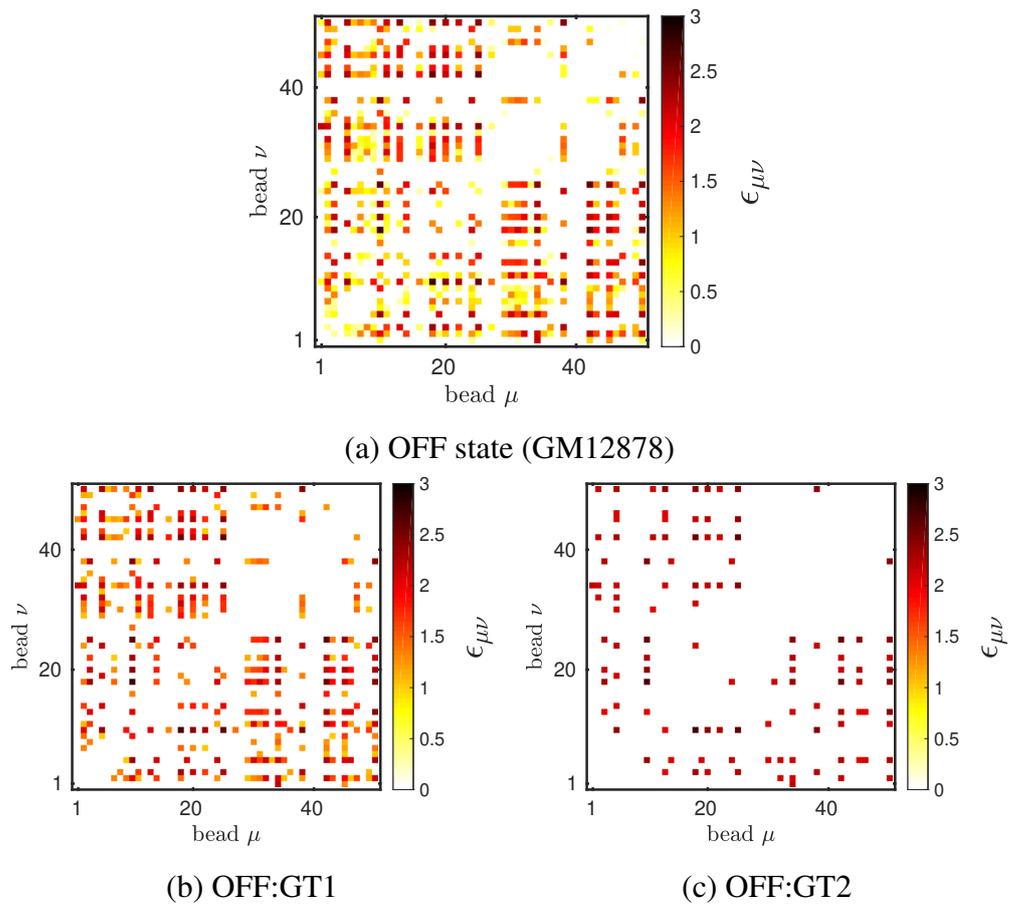
**Figure 6.3:** (a) The interaction strength obtained form the IBD for GM12878 $\alpha$-globin gene. (b) A subset of (a) where only the interaction strengths greater than $1k_BT$ are considered. (c) A subset of (a) where the interaction strengths greater than $2k_BT$ are considered.
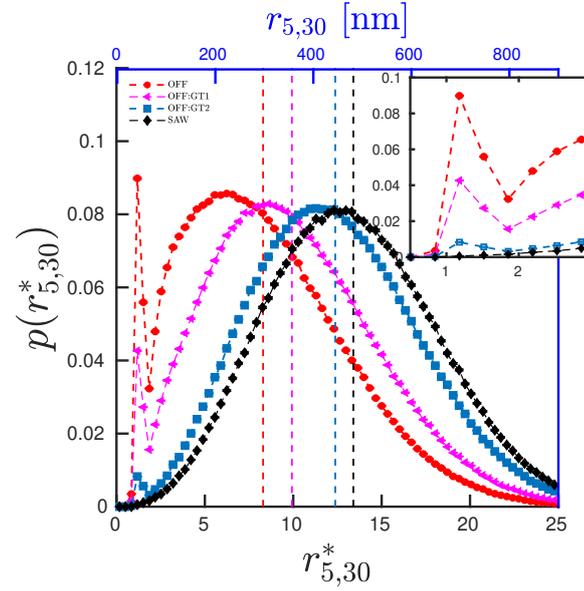
(a)
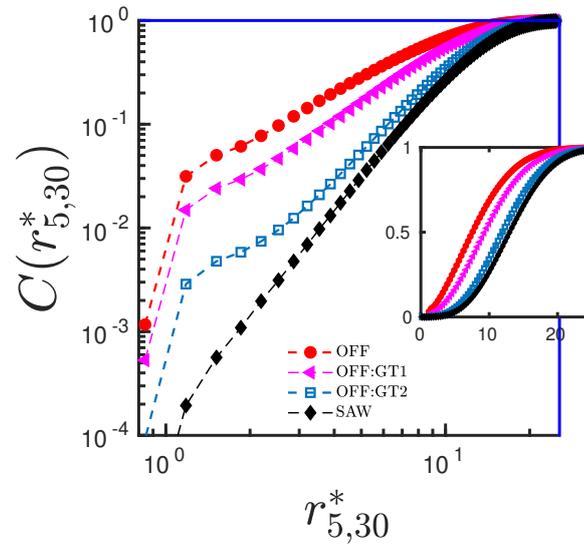


(b)

**Figure 6.4:** (a) Comparison of $p(r^*_{\mu\nu})$ between different "epigenetic states". OFF: state with all WT interactions in GM12878 (red), OFF GT1: when weak interactions are ignored; only with $\epsilon_{\mu\nu} > 1k_BT$ interactions (pink), OFF GT2: when only very high interactions ($\epsilon_{\mu\nu} > 2k_BT$) are accounted (blue), SAW: control simulation with no crosslinking (black). Vertical dashed lines represent $\langle r^*_{\mu\nu} \rangle$ corresponding to each distribution. (b) Comparison of $C(r^*_{\mu\nu})$ for the chromatin domain under different epigenetic states.

Given that we have the optimal interaction strengths that satisfy the experimentally known contact probability constraints [74], we can answer the following important question: Are the measurable properties of a given bead-pair (e.g. $r_{5,30}$) solely determined by the interaction between those two particular beads ($\epsilon_{5,30}$) or are they influenced by the interactions among other bead-pairs as well? To answer this, we adopted the following strategy: we systematically switched off the attractive interaction among certain bead pairs and computed probability distributions and other polymer properties. We simulated polymers for the following four cases: (i) all interactions are considered – GM12878 (OFF), as shown in Fig. 6.3(a), (ii) only those interactions above $1k_BT$ are considered – we call it OFF:GT1 – all weak interactions ($< 1k_BT$) are switched off here (as shown in Fig. 6.3(b)), (iii) only strong interactions above $2k_BT$ are considered – OFF:GT2 – all weak and medium interactions ($< 2k_BT$) are switched off (as shown in Fig. 6.3(c)), (iv) all attractive interactions are switched off – the SAW polymer. These four cases can be cosidered as four different epigenetic states – states having different interaction strengths due to underlying epigenetic variations. Fig. 6.4(a) shows $p(r_{5,30})$ for all the four cases. When we switch off the weak interactions below $1k_BT$ (OFF:GT1), compared to the OFF state, the height of the interaction-driven peak of the distribution decreases and overall the polymer swells resulting in the shift of the second peak (compare pink and red curves in Fig. 6.4(a)). This implies that weak interactions having strengths comparable to thermal fluctuations can also influence the contact probability and polymer configurations. If we keep only the highly prominent interactions and neglect all interactions below $2k_BT$ (OFF:GT2), the interaction-driven peak further diminishes and the distribution function approaches the SAW distribution (compare blue with other curves in Fig. 6.4(a)). Note that the interaction between beads 5 and 30 is present ($\epsilon_{5,30} = 2.09$) in all the cases except in the SAW case. The same behaviour can be observed in the cumulative plot as shown in Fig. 6.4(b). These results suggest that the measurable properties for a given bead-pair (e.g. $r_{5,30}$) depends not only on the attraction strength of that particular bead pair but also on the interactions of the whole polymer chain. This result implies that all bead-pairs collectively/cooperativily contribute in determining the relative position for a particular bead-pair.

## 6.2   Epigenetic changes alter the volume and shape of the chromatin domains

The above picture suggests that the chromatin folding is influenced by collective behavior of all beads having different interaction strengths. To examine the nature of collective behavior, we probed a property of the whole polymer namely the radius of gyration ($R_g$). To understand how folding is affected by different epigenetic states, we did the following. We started with a polymer having no interactions (SAW), added weak interactions (small $\epsilon$) that exist between beads in the OFF state as the first step, equilibrated, computed $R_g$ and sequentially added stronger interactions between beads step by step ($\epsilon < 0.5, \epsilon < 1.0, ..., \epsilon < 2$ and so on, denoted as LT1, LT2 etc), until the OFF state (GM12878) is reached. Each step was equilibrated and $R_g$ was computed (see Fig:6.5, top panel). From $R_g$, we have also computed the volume $V = (4/3)\pi R_g^3$ of the chromatin domain as shown in the right side $y$-axis. As seen from the figure, adding very weak interactions does not change the $R_g$ much. However, adding intermediate interactions significantly reduces the $R_g$ and it saturates as the interactions gets stronger, resulting in a sigmoidal-like curve showing signatures of cooperative/collective behavior. Since, we have equilibrated the polymer for each set of $\epsilon$ values, the LHS of the curve can be interpreted in two ways: folding the polymer by adding stronger and stronger interaction starting with a completely unfolded state or equivalently unfolding the polymer by removing the stronger interaction starting with a completely folded OFF state. One can also ask how the polymer would fold if one adds strong interactions (larger $\epsilon$) as the first step, starting with SAW, and then add weaker interaction sequentially step by step (denoted as GT1, GT2 etc). This is shown in the RHS of Fig. 6.5 (blue symbols). The whole curve suggests that having prominent interactions alone or weaker interactions alone may not take the system closer to its full equilibrium state. We also show typical snapshots of 3D chromatin configurations corresponding to different epigenetic states. As expected, the OFF state is compact and the volume of the domain increases as we go towards the SAW state. The fold-change in volume we observe between the two extreme states are roughly the same order as the density change observed experimentally [65].

To quantify how the shape of the chromatin domain changes with epigenetic states, we computed the asphericity (B) and the acylindricity (C) parameters (see Chapter 2 for definition). Asphericity quantifies the extent of deviation from a spherical shape. If a
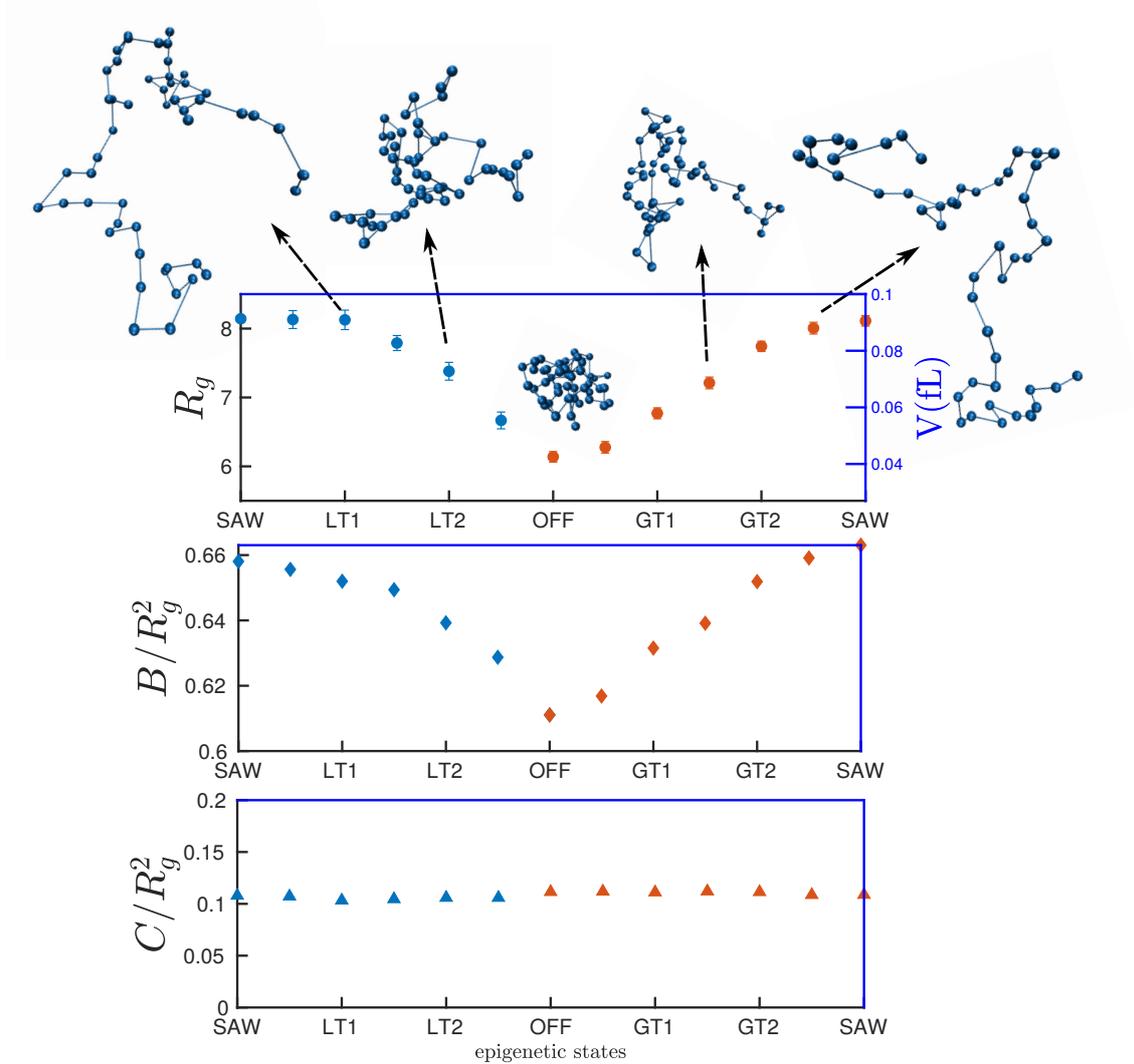
**Figure 6.5: Interaction strengths determining shape properties of chromatin domain:** Upper panel: Size of the chromatin domain ($R_g$) as we perturb interaction strengths. $x$-axis represents different interaction states with extreme ends representing the control (SAW) polymer and the OFF state (WT) in the middle. LT1 (LT$x$) indicates that all interactions below $1k_BT$ ($xk_BT$) are present in the polymer. Similarly GT1 (GT$x$) indicates that all interactions above $1k_BT$ ($xk_BT$) are present. The right $y$-axis indicates volume of the chromatin domain in femtolitre. Snapshots from simulations at various epigenetic states are shown around the perimeter of the graph. Bottom two panels represent the normalized asphericity ($B/R_g^2$) and acylindricity ($C/R_g^2$), respectively. The $x$-axis is the same in all panels.
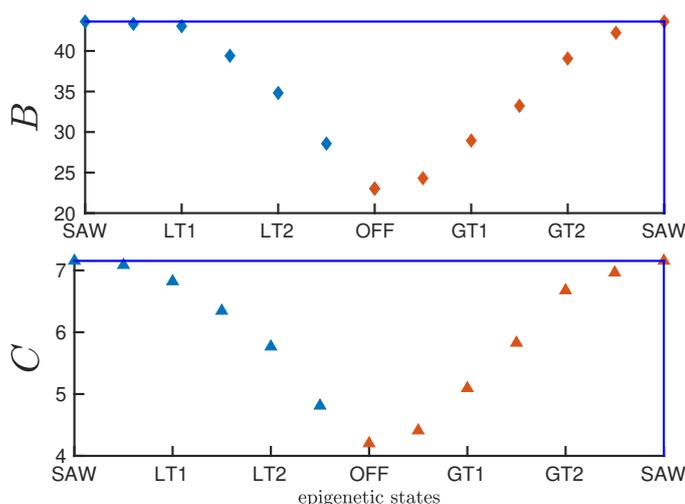
**Figure 6.6:** (a) and (b) represent the asphericity and acylindricity, respectively. *x*-axis represents different interaction states with extreme ends representing the control (SAW) polymer and the OFF state (WT) in the middle. LT1 (LT*x*) indicate that all interactions below $1k_BT$ ($xk_BT$)are present in the polymer. Similarly GT1 (GT*x*) indicate that all interactions above $1k_BT$ ($xk_BT$) are present.

polymer is coiled with the average shape of a sphere, $B = 0$. Here a positive B value suggests that even in the OFF state, the chromatin domain is not a perfect sphere. As we go from OFF to SAW, the asphericity increases by $\approx 65\%$ as shown in Fig. 6.6. However, the asphericity scaled with the polymer size ($B/R_g^2$) changes by $\approx 10\%$. Similar to $R_g$, we have shown the GT (RHS, orange symbols) and LT (LHS, blue symbols) cases for the asphericity too. Even though both sides are monotonically increasing, note that LT cases are not equivalent to the GT cases. We also compute the acylindricity parameter that quantifies the extent of the deviation from a perfect cylinder. Here too, $C > 0$ values suggest that the chromatin domain is not a perfect cylinder (see the lower panel of Fig. 6.5 and Fig. 6.6). Even though the acylindricity is monotonically increasing (shown in Fig. 6.6) from the OFF state to a SAW, it is increasing in proportion to the size of the polymer. Hence the scaled acylindricity ($C/R_g^2$) is nearly a constant as shown in the lower panel of Fig. 6.5.

## 6.3 Estimation of solid-like (stiffness) and liquid-like (drag) properties from domain relaxation times and fluctuations

Whether chromatin is liquid-like, solid-like or gel-like has been a matter of intense discussion in the recent literature [82, 93, 55, 128]. In the phase separation picture, chromatin segments are thought to be "liquid-like", dynamically exploring various configurations. Given that our model can study the stochastic nature of formation and breakage of bonds, and polymer dynamics, consistent with what is observed in Hi-C experiments, below we compute relaxation times and fluctuations of the chromatin domain and estimate effective elastic and drag properties.

First, we computed the end-to-end autocorrelation function $\left\langle \mathbf{R}_{\mathrm{E}}^*(0) \cdot \mathbf{R}_{\mathrm{E}}^*(t^*) \right\rangle / \left\langle \mathbf{R}_{\mathrm{E}}^{*2}(0) \right\rangle$ where $\mathbf{R}_{\mathrm{E}}^* = |r_1^* - r_{50}^*|$ and extracted the longest relaxation time $\tau^*$ with and without HI. The autocorrelation decay computed with HI is shown in Fig. 6.7(a) and no-HI case is shown in Fig. 6.7(b). Fig. 6.8(b) shows that the relaxation times for all the epigenetic states are lower with HI, as observed previously for the protein folding simulations [102, 103]. All results presented in this thesis are computed with HI, unless stated otherwise. The chromatin in the OFF state has a lower relaxation time compared to a crosslinking-free chromatin (SAW). This can be counter-intuitive as one would naively expect that the more the crosslinking the slower the chromatin will relax. A similar puzzle has also been observed in recent experiments [54, 93] where the repressed chromatin domain diffuses faster than the active one. To understand this apparent contradiction, we investigate the elastic and drag properties of the chromatin domain. From the measurement of fluctuations of each bead-pair we can compute an effective stiffness defined as $K_{\mu\nu} = k_{\mathrm{B}}T / \langle |r_\mu - r_\nu|^2 \rangle$ (Fig. 6.9(a)). As expected, highly cross-linked OFF state is more stiff than the other epigenetic states including SAW. This can also be understood from the free energy as a function of bead pair distance $F^*(r^*) = -\ln(p(r^*)/4\pi r^{*2})$ (see inset). The above behaviour is consistent with $K_{\mu\nu}^* \sim \dfrac{\partial^2 F_{\mu\nu}^*}{\partial r_{\mu\nu}^2}$ and stiffness ($K_{\mu\nu}^*$) of different epigenetic states do show similar behaviour. For the known stiffness and relaxation times, we can compute an effective drag coefficient defined as $\zeta_{\mathrm{eff}}^* = \tau^* \times K^*$. Taking the effective stiffness of the end beads ($K_{1,50}^*$), we find that the drag for the OFF state is higher than the other states suggesting that higher cross-linking reduces its ability to reorganize (see Fig. 6.9(b)). Both the stiffness and drag are greater for the OFF state than the SAW,
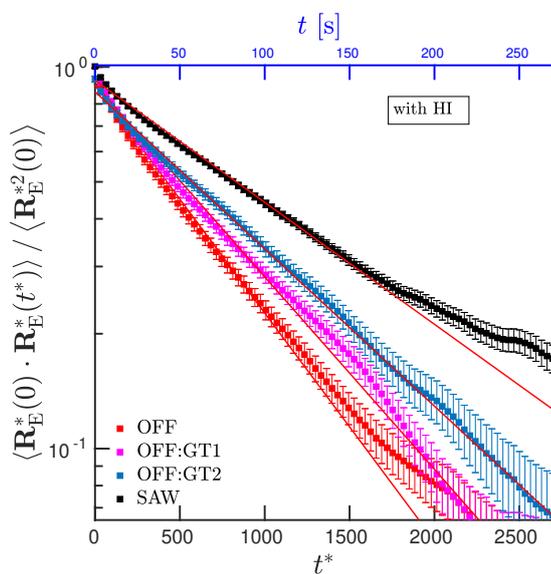
(a)
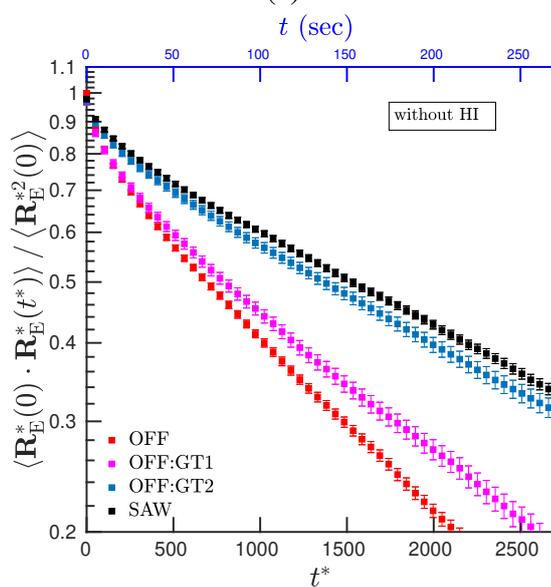


(b)

**Figure 6.7:** Exponential decay of end-to-end auto-correlation function with time for four epige-netic states computed (a) with HI and (b) without HI.
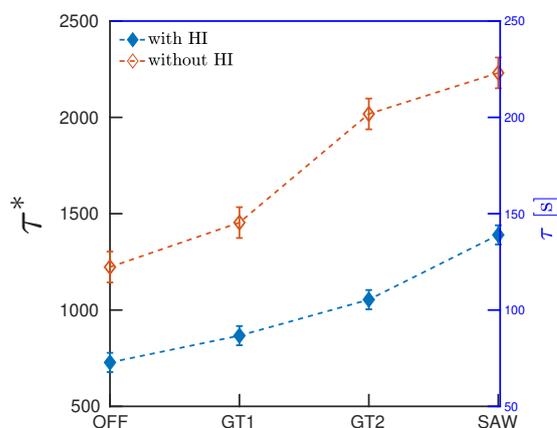
**Figure 6.8:** Relaxation times ($\tau$) with and without HI reveal that HI helps in relaxing the polymer faster.

but they combine to lead to a faster relaxation time for the OFF state. Our findings agree with the recent experimental report that crosslinked chromatin shows less FRAP revealing gel-like nature of chromatin [128, 55].

## 6.4 Interplay between interaction energy and polymer entropy influences the dynamics of chromatin domain

While we gained insights into steady state fluctuations and distance distributions, how the interactions would affect chromatin dynamics can be further probed. We know that contacts between chromatin segments are dynamic; proteins that form contacts bind and dissociate resulting in stochastic formation and breakage of contacts. This opens up interesting questions: How long do two beads remain in contact (looped)? When loops break and beads diffuse away, how long does it take for the bead pairs to come back in contact? What are the factors (interaction strengths, polymer entropy etc.) dictating the phenomena of dynamic contacts?

To study the temporal nature of chromatin, we define loop formation time ($t_L^*$) and contact time ($t_C^*$) for all bead-pairs. $t_L^*$ is defined as the time taken for a pair of beads to meet ($r_{\mu\nu}^* < r_C^*$) for the first time, starting from a random equilibrium configuration. $t_C^*$ is defined as the duration that the bead-pairs remain looped/in contact. A schematic

(a)



(b)

**Figure 6.9:** (a) Effective stiffness between all the bead-pairs; OFF state is more stiff compared to less interacting states and SAW. Inset: Free energy as a function of bead pair distance $r_{5,30}$. (b) Effective viscous drag felt by different chromatin states. OFF state chromatin (with stronger interactions) is more stiff and has a higher viscous drag.
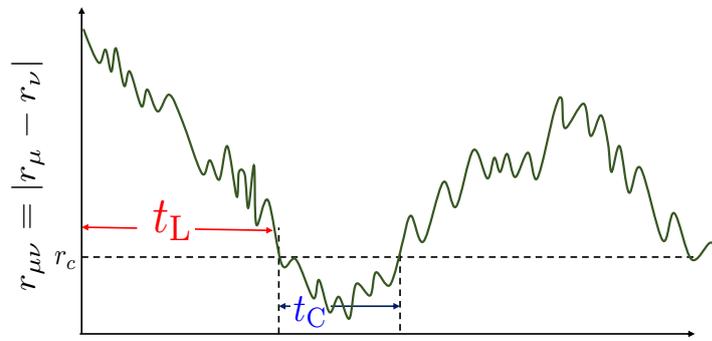
**Figure 6.10:** schematic representation of the distance between two beads in a single trajectory showing $t_L^*$ and $t_c^*$.
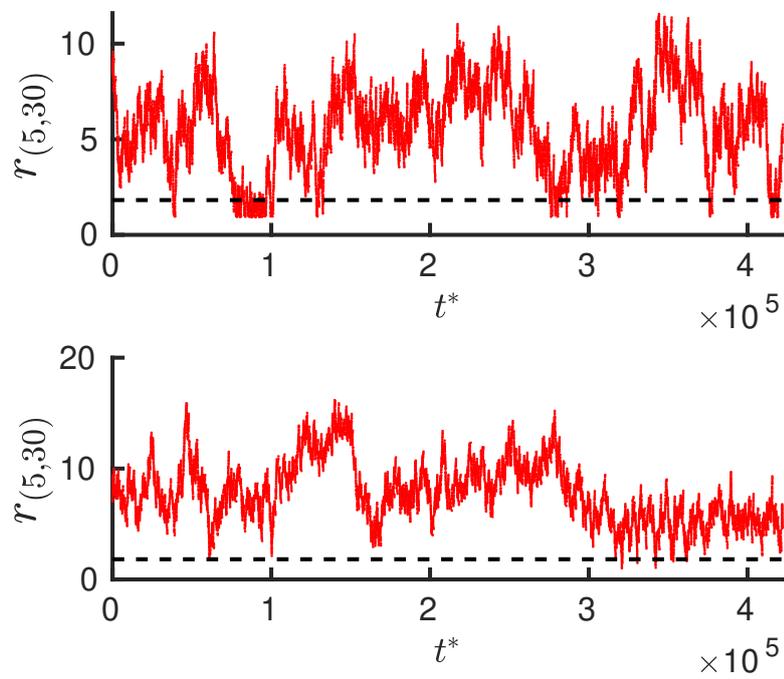


**Figure 6.11:** Distance data from our simulation for a particular pair of beads for a three randomly chosen realisation.

(a)



(b)

**Figure 6.12:** (a) $\langle t_L^* \rangle$ has a power law scaling with genomic length ($\langle t_L^* \rangle \sim s^\mu$) with exponent varying from 1.4 (OFF state) to 2.3 (SAW) for different chromatin states. The exponents are shown in lower inset. The upper inset shows $\langle t_L^* \rangle$ for various $N$ values for a random walk polymer. (b) $\langle t_L^* \rangle$ as a function of interaction strength with each point representing a bead pair. Note the huge spread in $\langle t_L^* \rangle$. Inset: $\langle t_L^* \rangle$ binned and averaged over all bead pairs having same $\epsilon$ showing minimal influence of $\epsilon$.

(a)



(b)
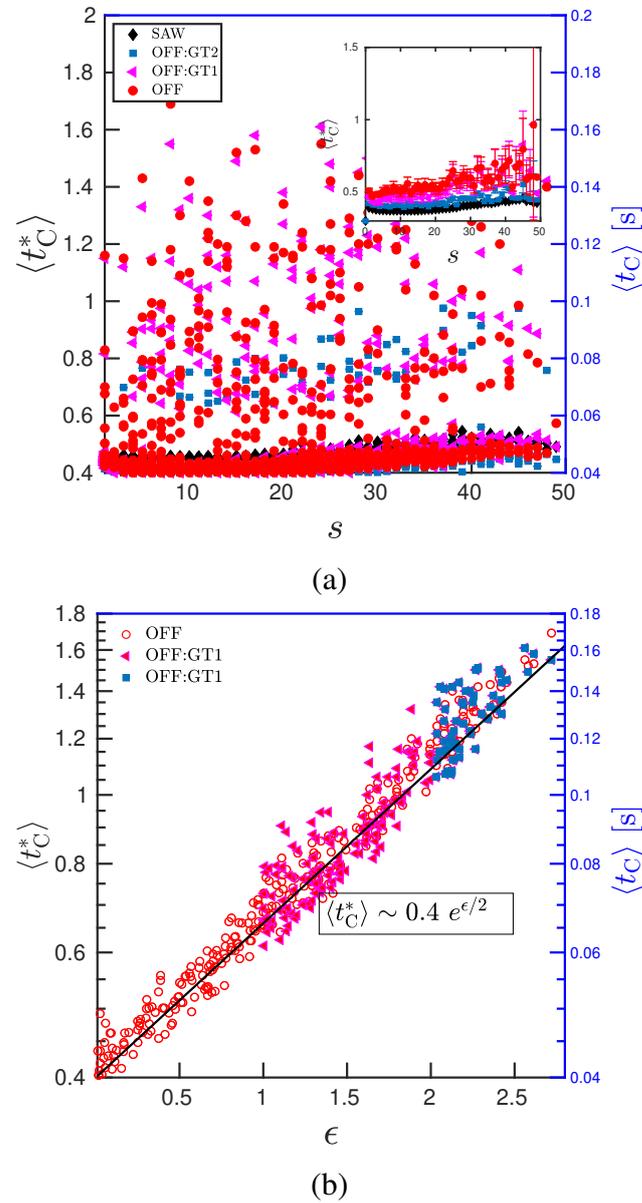
**Figure 6.13:** (a) $\langle t_{\mathrm{C}}^{*} \rangle$ as a function of $s$ with each point representing a bead pair. Here too, note the spread. Inset: $\langle t_{\mathrm{C}}^{*} \rangle$ binned and averaged over all bead pairs having the same $s$ showing minimal dependence on the segment length. (b) $\langle t_{\mathrm{C}}^{*} \rangle$ increases exponentially with the interaction strength.

(a)



(b)
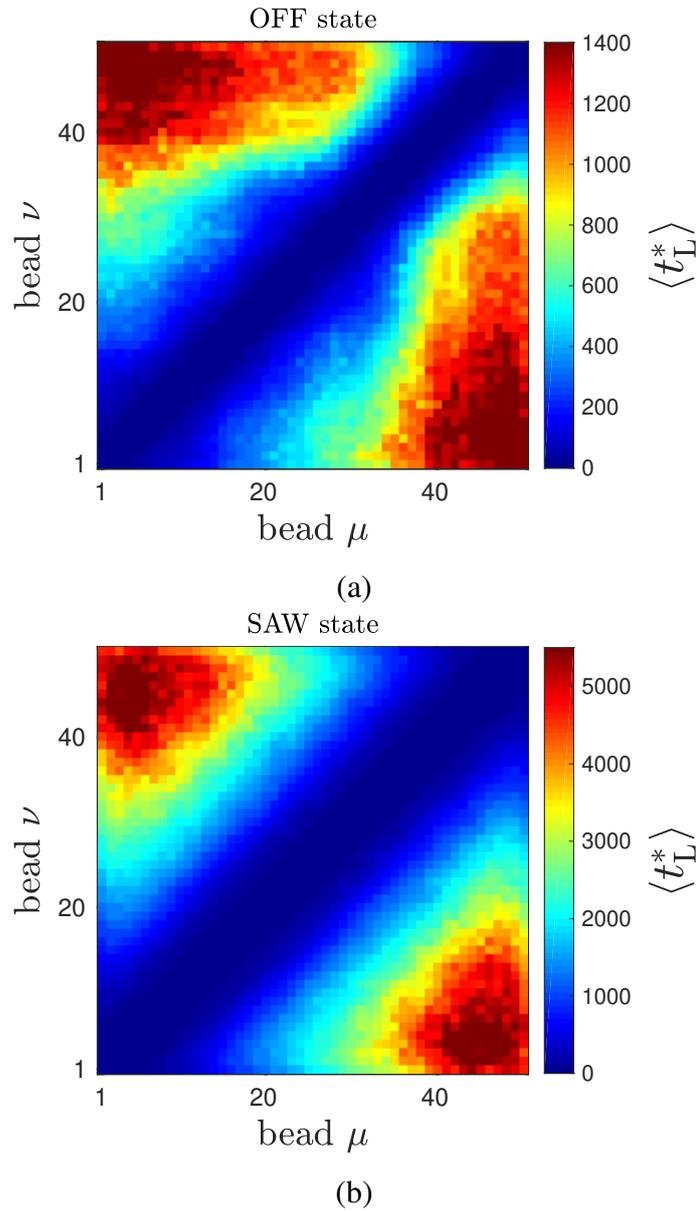
**Figure 6.14:** $\langle t_L^* \rangle$ for all the bead pairs as a heatmap for (a) OFF and (b) SAW states.

representation of a typical time trajectory of 3D distance indicating $t_L^*$ and $t_C^*$ is shown in Fig. 6.10 and the actual data from our simulation, as an example, is shown in Fig. 6.11. Corresponding average quantities are defined by $\langle t_L^* \rangle$ and $\langle t_C^* \rangle$, respectively.

Two possible factors that can influence these temporal quantities are interaction strengths ($\epsilon$) and polymer entropy. Since two beads having a larger segment length between them will have a higher entropy, it is expected that the time to come into contact is longer. In other words, the time of looping is expected to be dictated by polymer entropy. To validate this hypothesis, we looked at $\langle t_L^* \rangle$ as a function of the genomic length with and without HI.

As shown in Fig. 6.12(a) $\langle t_L^* \rangle$ monotonically increases with $s$ showing a power law behavior $\langle t_L^* \rangle \sim s^\mu$. As a control, we matched our $\langle t_L^* \rangle$ results with the previously known exponents $\mu \approx 2.3$ for SAW and $\mu = 2.0$ for a random polymer (see top inset) [135]. By simulating various chain lengths ($N = 10, 20, ...$) we can infer that the deviation from the power law for large $s$ is due to finite chain effects (top inset). We have also computed $\langle t_L^* \rangle$ for all the other epigenetic states revealing $2.3 < \mu \leq 1.4$. The OFF state having all interactions shows the smallest exponent of 1.4. As we remove interactions from the system, $\mu$ gradually approaches the SAW limit. The change in power law may also be understood by looking at the free energy plotted in Fig. 6.9(c) inset. One can see that the free energy has a higher tilt in the OFF state compared to the other states, implying that the bead-pairs can move along the landscape quicker in the OFF state. The results for $\langle t_L^* \rangle$ suggests that even in the absence of loop extrusion, the looping time is not too long (seconds to minutes). This also indicates that the micro phase-separation could be a viable mechanism for bringing together chromatin segments and possibly explains the experimentally observed fact that chromatin is functional even in the absence of loop extruding factors [13, 18, 69]. We then examined how the interaction strength influences $\langle t_L^* \rangle$, and found that there is a huge spread in the $\langle t_L^* \rangle$ values, for a given $\epsilon$ (Fig. 6.12(b)), with the average showing a mild dependence on $\epsilon$ (inset).

Interestingly the values of $\langle t_C^* \rangle$ are nearly independent of genomic separation (Fig. 6.13(a)). Here too there is a huge variability among different bead pairs with the inset showing the behaviour when the segment length is averaged over all pairs having the same $s$. However, the interaction strength significantly alters the $\langle t_C^* \rangle$ (Fig. 6.13(b)) showing an exponential increase. This suggests that the $\langle t_L^* \rangle$ is determined by the interplay between entropy (resulting from genomic separation) and energy (interaction strength). Once bead-pairs come

in contact $\langle t_C^* \rangle$ is dominated by the interaction strength.

For the OFF and SAW states, we also show $\langle t_L^* \rangle$ between all pairs of beads as a heatmap (see Fig. 6.14(a) & (b)). One can quickly note that the range of SAW time scales is much higher than that of the OFF state. This is the consequence of higher $\mu$ for the SAW compared to the OFF state. In the SAW, one can observe that the times are similar for all points having the same distance away from the diagonal (a line parallel to diagonal axis), suggesting that what matters in this case is the inter bead distance ($s$). In contrast, in the OFF state, there is a heterogeneity and curvy color contours suggesting that the time values are not just a function of segment length alone but also the identity (interaction strength) of the individual bead pairs. This once again points to the interplay between entropy and energy. The phase-space exploration of a specific bead-pair in a bead-spring chain can be imagined as a single bead moving in an effective potential energy landscape given by the free energy $F^*(r^*) = -\ln(p(r^*)/4\pi r^{*2})$ as a function of the 3D distance ($r^*$) between the corresponding bead-pair. Details on this is provided in Appendix A.4.

## 6.5    Nature of loop formation and contact time distributions

So far we have studied the average loop formation times and contact times; however, should one assume that the average values describe these quantities completely? To answer this, similar to $p(r^*)$, here we have investigated the nature of the distribution of the temporal quantities. In Fig. 6.15(a) and (b), we present the probability distributions of contact ($p(t_C^*)$) and loop formation ($p(t_L^*)$) times, respectively. We observe that $p(t_C^*) \sim \exp(-t_C^*/\tau_c)$ with the average time $\tau_c$ that depends on the epigenetic state (SAW: $\tau_c = 1/1.6$, OFF: $\tau_c = 1/1.25$). $\tau_c$ is small for the SAW and it increases as we add interactions to the system. However, interestingly, the probability of loop formation time ($t_L^*$) has a power law decay ($p(t_L^*) \sim (t_L^*)^{-\gamma}$). This suggests that there is a huge diversity in loop formation times, and the average looping time alone may not be sufficient to describe the loop formation phonomena. We find that the epigenetic states alter the slope of the distribution (SAW: $\gamma = 0.4$, OFF: $\gamma = 1.0$) keeping the overall nature the same. Comparison of these two distributions reveals that quantitatively the $t_L^*$ is much larger than $\tau_c^*$, indicating that chromatin segments take longer to come into contact but stay in contact for a short time.
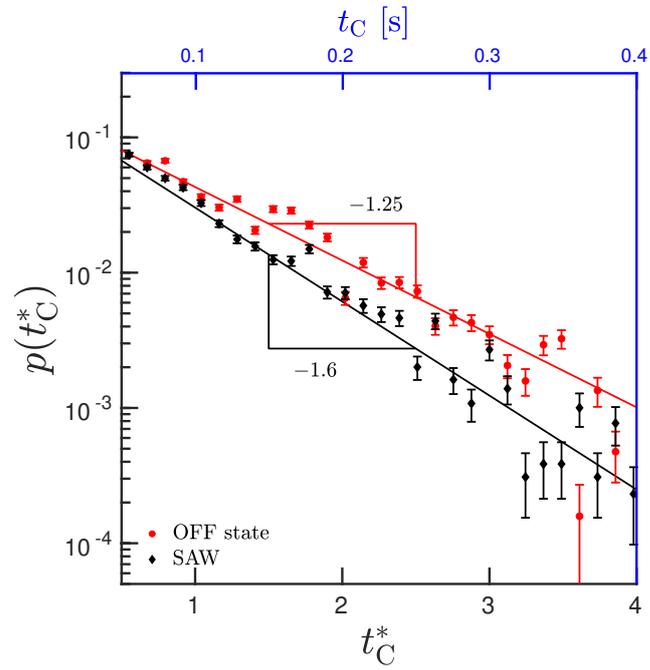
(a)
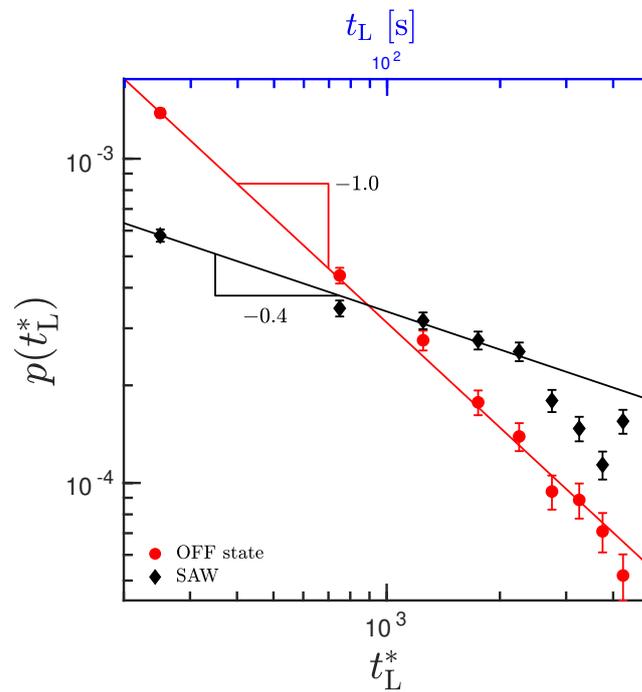


(b)

**Figure 6.15:** Distribution function for contact time ($t_C^*$) and loop formation time ($t_L^*$) for a specific bead-pair (bead 5 and bead 30) in SAW and OFF state are shown in (a) and (b), respectively.

## 6.6    Summary

Even though there is a great improvement in our understanding of static nature of chromatin organization, very little is known about the dynamics, which is a crucial aspect of *in vivo* chromatin. Here, in this chapter, we investigated the fluctuations and dynamics of chromatin domain. Going beyond the average properties, we computed the distance probability distribution and it shows two peaks – an interaction-driven peak and an entropy-dominated peak. Assuming that interactions are arising from epigenetic states, we show how perturbations in epigenetic states would alter $p(r)$; the distance distribution between a given bead pair depends on the interaction strength of all other pairs suggesting the cooperative nature of chromatin folding. Volume and the shape properties of the chromatin domain depends on the epigenetic state. The OFF state is highly collapsed/compact, more spherical compared to the extended, less spherical SAW. The relaxation time of the domain is dependent on the epigenetic state of the domain. Counter-intuitively, the relaxation time of a highly crosslinked OFF state is much smaller than that of a crosslink-free SAW polymer. We explain this phenomenon by computing effective stiffness of the domain, from polymer fluctuations. We also show that the OFF state has a higher effective drag. We study dynamics accounting for crucial hydrodynamic interactions; we show that HI has a significant influence on the relaxation time of the chromatin domain. With HI, the domain takes half the time to relax as compared to the no-HI case. We compute the loop formation time and the time for the looped bead pairs to remain in contact. We show that average looping time has different scaling with genomic separation, depending on the epigenetic nature of the chromatin states. The looping times show a power law distribution indicating multiple timescales that might be involved with looping. On the other hand, the contact time has an exponential distribution.

Apart from understanding of the spatiotemporal nature of chromatin domains, quantities calculated here have immense biological significance too. There is an ongoing debate in the field about whether the gene regulation requires actual physical contact between two regulatory segments or only the proximity would suffice. Cellular processes such as transport of proteins from one region to another (eg. enhancer-promoter), spreading of histone modifications in the 3D space etc would crucially depend on $p(r)$. For example, given $r$, one can compute the time ($\tau_p$) for proteins/enzymes to diffuse from location $r_\mu$ to $r_\nu$. The mean time would depend on the distribution as $\langle \tau_p \rangle = \int \tau_p p(r) dr$. However, apart from the distance among segments, the accessibility would depend on the local compact-

ness and diffusivity too. That is, compactness of the domain and effective viscous drag together with $p(r)$ would be crucial for understanding how physics of chromatin would affect biological function. Given that phase separation is argued to be one of the important factors determining the domain formation, our study also reveals how interplay between epigenetic states and polymer dynamics would affect loop formation and contact times.

# Chapter 7

# Conclusion

In this thesis, we set out to study the 3D organization and dynamics of chromatin. Recent advances in experimental studies have generated a large amount of experimental data giving us information about contact frequencies between chromatin segments (3C, 5C, HiC etc.) as well as the spatial location of certain genes (microscopy). The chromatin conformation capture experiments provide us partial information about the population-averaged contacts counts between certain segments. Computation/theoretical studies can complement these experiments, go beyond the contact counts and generate the complete 3D organization and study dynamics.

The main challenge to simulate the dynamics of chromatin is that we do not know the interaction strength parameters among different chromatin segments. To overcome this challenge, in this thesis, we have developed an inverse technique and obtained optimal interaction strengths between all chromatin segments and used it to investigate the dynamics of a chromatin domain. We investigated the 3D organization of chromatin-based on publicly available chromatin conformation capture experimental data.

In this thesis, we have presented a coarse-grained model for chromatin, considering it as a bead-spring chain. We used the Fraenkel spring for the linear connectivity and a novel SDK potential for excluded volume and intra-chromatin interaction. We combined the inverse method with the Brownian dynamics simulation to present an inverse Brownian dynamics (IBD) method. Using IBD, we optimised the interaction strength parameters, consistent with experimental contact count data. Our IBD algorithm has the following advantages: i) unlike some of the existing models in the literature, we do not assume any *à priori* relation between spatial distance and contact probability, ii) we optimize

the interaction strength between the chromatin segments to reproduce the experimentally known contact probability, and iii) our simulations have accounted for hydrodynamic interactions; therefore, our model is capable of investigating the dynamics of the chromatin polymer in detail.

To conclude, the contribution of the current work can be summarized as follows:

1. We introduced the Soddemann-Dunweg-Kremer (SDK) potential to model the excluded volume and intra-chromatin interactions. We investigated this in detail and estimated the appropriate potential energy parameters such that we could reproduce the universal scaling laws ($R_g \sim N^\nu$) known in polymer physics.

2. We developed an Inverse Brownian Dynamics method to compute the optimal interaction strengths between different segments of chromatin such that the experimentally measured contact count probability constraints are satisfied.

3. We validated the IBD method for a prototype bead-spring chain. The IBD reproduced the contact probabilities and the interaction strengths (within 5% error), reflecting its reliability.

4. Applying this method to the $\alpha$-globin gene locus in two different cell types, we predicted the 3D organization corresponding to active and repressed states of the locus. The structural properties of the $\alpha$-globin gene locus were investigated in terms of shape functions, bead number density distributions, and 3D snapshots. In the ON state, the $\alpha$-globin exists in an extended structure, whereas in the case of the OFF state, the gene appears to be in a folded/collapsed state. This is also consistent with the notion that in the ON state, the chromatin must be more open for transcription factors to access the gene, whereas the structural status of the OFF state should be more compact to prevent transcription factors from accessing the gene.

5. A procedure for normalizing the contact count matrix was introduced with a parameter $N_f$ varying from 0 to 1 that reflected the two different extreme scenarios for estimating the sample size. We also simulated the $\alpha$-globin gene with well-known ICE normalization. We studied the differences in structural features such as shape properties, density profile, and 3D configuration between ICE and the other normalization procedure.

6. Our simulations could predict the average spatial distance ($r_{\mu\nu}$) between any two segments. We have shown how the $r_{\mu\nu}$ varies as a function of contact probability($p_{\mu\nu}$). We found that, unlike the prevalent notion, there is no simple functional relation between $r_{\mu\nu}$ and $p_{\mu\nu}$. Given a $p_{\mu\nu}$ value, one cannot uniquely predict the mean spatial distance as there is huge variability arising from heterogenous intra-chromatin interactions.

7. We also computed the mean spatial distance ($r_{\mu\nu}$) as a function of genomic distance($s$) for alpha-globin in the OFF state and compared it with similar known experimental observation. It is interesting to note that the slope as well as the spread of $r_{\mu\nu}$ vs $s$ relation is consistent with what is observed experimentally.

8. We computed the distance probability distribution, and it shows two peaks. The first peak is determined by the interaction energy between the bead pair (interaction-driven peak), and the second peak is dominated by polymer entropy (genomic separation).

9. We studied the chromatin dynamics accounting for crucial hydrodynamic interactions; we showed that the HI has a significant influence on the relaxation time of the chromatin domain. With HI, the domain takes half the time to relax as compared to the no-HI case.

10. We investigated the role of epigenetic modification by altering the interaction strength of the system.

11. We computed the loop formation time and the time for the looped bead pairs to remain in contact. We have shown that the average looping time has a different scaling with each epigenetic state. The looping times show a power-law distribution indicating multiple timescales that might be involved with looping. Contrary to this, the contact time has an exponential distribution.

12. We demonstrated the cooperative nature of chromatin folding by examining various static and temporal quantities such as distance probability distribution and loop formation time.

In a nutshell, taking chromatin conformation capture data as input and using a novel

inverse Brownian dynamics method, we have investigated various crucial features of chromatin organization, its fluctuations and dynamics.

## 7.1   Suggestions for experiments to test our predictions

Most of the results in this work are predictions that may be tested in suitably designed experiments. Since we take HiC-like data as input and predict average properties, positional fluctuations and dynamics of chromatin segments, microscopy is the ideal method to test our predictions [54, 18, 93]. We predict that the spatial segmental distance depends not only on the contact probability but also on the segment length along the contour. One of the ways to test our prediction is to perform 3D FISH on segment pairs having the same contact probability but different segment lengths. The distance obtained from the FISH experiment will validate the predictions made in this work. So far 3D FISH techniques have measured distances [18, 129] above 100nm (sizes equivalent of $\approx$ 3 beads in our simulation). As you can see from Figures 5.6 - 5.12, our predicted 3D distances are of the order of 100nm - 700nm (size of 3 beads to 20 beads, i.e. $3\ell_H$ - $20\ell_H$). Therefore, 3D FISH techniques can measures distances in the same range as what we predict in our simulations. Shape properties and density profiles of the $\alpha$-globin locus are also predicted and can be tested using techniques like super-resolution microscopy and electron microscopy. We require these additional experiments to determine the appropriate normalization. Our work predicts that 3D distances, shape properties, density profile, etc., will depend on the precise nature of normalization. Hence, the appropriate normalization methodology may be determined by comparing our results with future experiments that measure these quantities. It will be very interesting to experimentally test how the average 3D distance as a function of contact probability varies. Certain experiments that probe structural features (e.g., electron microscopy) can also investigate the broad shape properties. All $p(r)$ predictions may be tested either via live (without fixing) microscopy experiments or by collecting a large number of frozen snapshots of segment-locations via FISH or equivalent methods. Imaging experiments may also estimate the volume occupied by a domain. From the positional fluctuation data, one can also obtain the effective stiffness as described earlier in this paper. To measure the time-dependent quantities, apart from live microscopy experiments, one may also design appropriate FRET pairs that can probe quantities like the contact time [143]. Obtaining all of these quantities for different epigenetic states would facilitate comparison with our predictions.

## 7.2   Limitation of the study

We make an effort to understand and explain the phenomena of chromatin folding by the use of a polymer model. Considering a simple model undoubtedly expands our understanding of the complex system but suffers from certain limitations. The coarse-grained polymer model lacks certain chemical details below the resolution of the monomer of the model. For example, we do not have nucleosome level information considered here.

One of the concerns regarding our work could be that this study simulates only a short segment. However, most of the biologically relevant processes happen on the length scale of a gene (or a few genes). Hence, it is essential to zoom in and study the organization and dynamics of short segments. Given that chromatin is organized into small local domains (topologically associated/chromatin domains) having only local interactions predominantly, it may be reasonable to analyze one locus or domain at a time. The IBD algorithm can also be used to study the static and dynamics properties of the whole genome by considering a longer polymer chain. Several sampling techniques can be utilized to sample the phase space efficiently, such as parallel tempering techniques (64). This method can be used to check the validity of the simplest model for a given contact probability matrix. In other words, if a model does not converge to the desired probabilities even after proper sampling, it implies that the model (as represented by the Hamiltonian or the included physics) may require modification and a more sophisticated model may be required. For instance, we have chosen the simplest model that can reproduce the experimentally observed contact probability map. A lack of convergence (even after proper sampling) may imply the need for adding additional physics into the model. For example, certain faraway contacts may require the addition of nonequilibrium processes like loop extrusion.

Our simulation does not account for nonequilibrium activity explicitly. We neglect the effects of active processes like transcription and chromatin remodelling. However, the kind of interactions that we consider in our model ($\epsilon_{\mu\nu}$) are maintained via nonequilibrium activity. The effect of activity is probably buried in $\epsilon_{\mu\nu}$ such that we get back the experimentally seen contacts.

We have also not considered the loop extrusion explicitly. This may not affect the static properties since we have accounted for all intra-chromatin interactions. However, the precise dynamics may depend on the amount of loop extrusion. While the loop extrusion may be crucial for understanding chromatin compaction during metaphase [57],

it is not clear precisely how extrusion works in the interphase. For example, given that there are many intra-chromatin interactions, how the loop extruding factors would move through these dense contacts is not clear. Moreover, there are large number of domains where the loop extrusion may not be the dominating factor. Recent experiments have shown that domains with transcription are less dynamic compared to the repressed domains, opening up new questions on the role of extrusion like dynamics in the interphase [54, 65, 6]. Since there are many unanswered questions on how to implement loop extrusion, we restricted ourselves to studying the role of intra-chromatin interaction energy and polymer entropy. In the current thesis, we have only accounted for intra-chromatin interactions that are relevant for $\alpha$-globin domain. However, for some domains, there might be additional interactions with the nuclear periphery/nuclear lamina. For example, some heterochromatin regions are known to be lamina-associated domains (LADs). In future, when we will simulate such LADs, we will incorporate bead-lamina interactions in an appropriate manner. In addition, for some domains, there would be interchromosomal interactions. For example, in a recent work, it has been shown that chromatin intermingling regions are regulatory hotspots for transcription [12]. While studying such domains in future, it will be crucial to incorporate relevant inter-chromosomal interactions.

## 7.3   Future perspectives

Our model can be extended to incorporate more data (histone modification data, CHIP-Seq data of certain proteins) and address chromatin organization on the length scale of genes in more detail. Recent experiments suggest that 3D chromatin organization is driven by two different dynamic processes, namely, phase separation and loop extrusion. Since our model is capable of studying dynamics, the model may be extended to investigate the interplay between different dynamic processes in determining chromatin organization. This study can be further extended genomewide to examine various gene loci and investigate the fluctuations and dynamics of all domains in the genome. Such polymer models are useful for examining aspects like the spread of histone modifications and accessibility of the domains. Recent polymer studies have highlighted the role of crowding under good/poor solvent conditions [136]. Our work may also be extended to incorporate the effect of crowding with the parameters relevant to the nucleus. With the capability of analyzing the 3D configuration along with chromatin dynamics, IBD can complement

experimental research and also provide deeper and more useful insights.  We hope that this study would catalyze new experimental and computational studies examining the interplay between epigenetics and polymer dynamics.

# Chapter 8

# Research outcomes

## 8.1 Publications

1. **K. Kumari**, J. R. Prakash, R. Padinhateeri, *Spatiotemporal organization of chromatin domains: role of interaction energy and polymer entropy* (under review), **bioRxiv (2021)**, doi: 10.1101/2021.02.17.431616.

2. **K. Kumari**, B. Duenweg, R. Padinhateeri, J. R. Prakash, *Computing 3D chromatin configurations from contact probability maps by Inverse Brownian Dynamics*, **Biophys J. 118: 2193-2208 (2020)**, doi: 10.1016/j.bpj.2020.02.017.

3. A. Santra, **K. Kumari**, R. Padinhateeri, B. Duenweg, J. R. Prakash, *Universality of the collapse transition of sticky polymers*, **Soft Matter, 15, 7876-7887 (2019)**, doi: 10.1039/C9SM01361J.

## 8.2 Conferences

1. **K. Kumari**, J. R. Prakash, R. Padinhateeri, *Computing the spatial organization and dynamics of chromatin domains*, Statistical Mechanics of Soft Matter. Organized jointly by the Griffith University and the University of Queensland, Brisbane, Australia (held virtually) 14-15 December, 2020 (Talk).

2. **K. Kumari**, J. R. Prakash, R. Padinhateeri, *Computing the spatial organization and dynamics of chromatin domains*, Biophysics Paschim, CSIR-NCL, Pune, India (held virtually) 3 October 2020 (Poster).

3. **K. Kumari**, B. Duenweg, R. Padinhateeri, J. R. Prakash, *Computing three dimensional chromatin configurations from two dimensional contact probability maps: An inverse Brownian dynamics algorithm*, Diamond Jubilee Symposium held at IIT Bombay, India from 15-16 February 2019 (Talk).

4. **K. Kumari**, B. Duenweg, R. Padinhateeri, J. R. Prakash, *Computing three dimensional chromatin configurations from two dimensional contact probability maps: An inverse Brownian dynamics algorithm*, EMBO Symposium: Regulatory epigenomics: From large data to useful models. Organized by IMSc Chennai, India from 10 – 13 March 2019 (Poster).

5. **K. Kumari**, B. Duenweg, R. Padinhateeri, J. R. Prakash,*Computing three dimensional chromatin configurations from two dimensional contact probability maps: An inverse Brownian dynamics algorithm*, International Union of Theoretical and Applied Mechanics (IUTAM) held at IIT Kanpur, India from 17-20 December 2018 (Poster).

6. **K. Kumari**, B. Duenweg, R. Padinhateeri, J. R. Prakash, *Computing three dimensional chromatin configurations from two dimensional contact probability maps: An inverse Brownian dynamics algorithm*, 7th Meeting of the Asian Forum of Chromosome and Chromatin Biology held at JNCASR, Bangalore, India from 15-17 November 2018 (Poster).

7. **K. Kumari**, B. Duenweg, R. Padinhateeri, J. R. Prakash, *Computing three dimensional chromatin configurations from two dimensional contact probability maps: An inverse Brownian dynamics algorithm*, Chemical Engineering Postgraduate Association (CEPA) conference held at Monash University, Australia on 26 October 2017. (Talk)

# Bibliography

[1] Agrawal, Ankit, Nirmalendu Ganai, Surajit Sengupta, and Gautam I Menon, 2020, "Nonequilibrium biophysical processes influence the large-scale architecture of the cell nucleus," *Biophysical journal* **118**, 2229–2244.

[2] Alberts, Bruce, 2014, *Garl. Sci.*, 6th ed. (Garland Science, Taylor and Francis Group, New York).

[3] Alipour, Elnaz, and John F Marko, 2012, "Self-organization of domain structures by DNA-loop-extruding enzymes," *Nucleic Acids Res.* **40**, 11202–11212.

[4] Annunziatella, Carlo, Andrea M Chiariello, Andrea Esposito, Simona Bianco, Luca Fiorillo, and Mario Nicodemi, 2018, "Molecular dynamics simulations of the strings and binders switch model of chromatin," *Methods* **142**, 81–88.

[5] Arya, Gaurav, and Tamar Schlick, 2006, "Role of histone tails in chromatin folding revealed by a mesoscopic oligonucleosome model," *Proceedings of the National Academy of Sciences* **103**, 16236–16241.

[6] Ashwin, SS, Tadasu Nozaki, Kazuhiro Maeshima, and Masaki Sasai, 2019, "Organization of fast and slow chromatin revealed by single-nucleosome dynamics," *PNAS* **116**, 19939–19944.

[7] Bajpai, Gaurav, Ishutesh Jain, Mandar M Inamdar, Dibyendu Das, and Ranjith Padinhateeri, 2017, "Binding of dna-bending non-histone proteins destabilizes regular 30-nm chromatin structure," *PLoS computational biology* **13**, e1005365.

[8] Bascom, Gavin D, Christopher G Myers, and Tamar Schlick, 2019, "Mesoscale modeling reveals formation of an epigenetically driven hoxc gene hub," *PNAS* **116**, 4955–4962.

[9] Bau, Davide, Amartya Sanyal, Bryan R Lajoie, Emidio Capriotti, Meg Byron, Jeanne B Lawrence, Job Dekker, and Marc A Marti-Renom, 2011 jan, "The three-dimensional folding of the alpha-globin gene domain reveals formation of chromatin globules.." *Nat. Struct. Mol. Biol.* **18**, 107–114.

[10] Baù, Davide, Amartya Sanyal, Bryan R Lajoie, Emidio Capriotti, Meg Byron, Jeanne B Lawrence, Job Dekker, and Marc A Marti-Renom, 2011, "The three-dimensional folding of the $\alpha$-globin gene domain reveals formation of chromatin globules," *Nat. Struct. Mol. Biol.* **18**, 107.

[11] Bednar, Jan, Rachel A Horowitz, Sergei A Grigoryev, Lenny M Carruthers, Jeffrey C Hansen, Abraham J Koster, and Christopher L Woodcock, 1998, "Nucleosomes, linker dna, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin," *Proceedings of the National Academy of Sciences* **95**, 14173–14178.

[12] Belyaeva, Anastasiya, Saradha Venkatachalapathy, Mallika Nagarajan, GV Shivashankar, and Caroline Uhler, 2017, "Network analysis identifies chromosome intermingling regions as regulatory hotspots for transcription," *Proceedings of the National Academy of Sciences* **114**, 13714–13719.

[13] Benabdallah, Nezha S, Iain Williamson, Robert S Illingworth, Lauren Kane, Shelagh Boyle, Dipta Sengupta, Graeme R Grimes, Pierre Therizols, and Wendy A Bickmore, 2019, "Decreased enhancer-promoter proximity accompanying enhancer activation," *Mol. Cell* **76**, 473–484.

[14] Berger, Axel B, Ghislain G Cabal, Emmanuelle Fabre, Tarn Duong, Henri Buc, Ulf Nehrbass, Jean-Christophe Olivo-Marin, Olivier Gadal, and Christophe Zimmer, 2008, "High-resolution statistical mapping reveals gene territories in live yeast," *Nature methods* **5**, 1031–1037.

[15] Bhattacharya, Dipanjan, Shefali Talwar, Aprotim Mazumder, and GV Shivashankar, 2009, "Spatio-temporal plasticity in chromatin organization in mouse cell differentiation and during drosophila embryogenesis," *Biophysical journal* **96**, 3832–3839.

[16] Bianco, Simona, Andrea M Chiariello, Carlo Annunziatella, Andrea Esposito, and Mario Nicodemi, 2017, "Predicting chromatin architecture from models of polymer physics," *Chromosome Research* **25**, 25–34.

[17] Bickmore, Wendy A, 2013, "The spatial organization of the human genome," *Annu. Rev. Genom. Hum. G.* **14**, 67–84.

[18] Bintu, Bogdan, Leslie J Mateo, Jun-Han Su, Nicholas A Sinnott-Armstrong, Mirae Parker, Seon Kinrot, Kei Yamaya, Alistair N Boettiger, and Xiaowei Zhuang, 2018, "Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells," *Science* **362**.

[19] Bird, R. B., C. F. Curtiss, R. C. Armstrong, and O. Hassager, 1987, *Dynamics of polymeric liquids*, Vol. 2 (John Wiley and Sons, New York).

[20] Bird, R B, C F Curtiss, R C Armstrong, and O Hassager, 1989, *Dynamics of polymeric liquids. Volume 2: Kinetic Theory* (John Wiley & Sons, Ltd).

[21] Bird, RB, CF Curtiss, RC Armstrong, and O Hassager, 1987, "Dynamics of polymeric liquids, kinetic theory (volume 2)," .

[22] Bishop, Marvin, and J P J Michels, 1986 nov, "Polymer shapes in three dimensions," *J. Chem. Phys.* **85**, 5961–5962.

[23] Briand, Nolwenn, and Philippe Collas, 2020, "Lamina-associated domains: peripheral matters and internal affairs," *Genome biology* **21**, 1–25.

[24] Brunet, Annaël, Nicolas Destainville, and Philippe Collas, 2021, "Physical constraints in polymer modeling of chromatin associations with the nuclear periphery at kilobase scale," *Nucleus* **12**, 6–20.

[25] Caragine, Christina M, Shannon C Haley, and Alexandra Zidovska, 2018, "Surface fluctuations and coalescence of nucleolar droplets in the human cell nucleus," *Phys. Rev. Lett.* **121**, 148101.

[26] Collepardo-Guevara, Rosana, and Tamar Schlick, 2014, "Chromatin fiber polymorphism triggered by variations of dna linker lengths," *Proceedings of the National Academy of Sciences* **111**, 8061–8066.

[27] Conte, Mattia, Luca Fiorillo, Simona Bianco, Andrea M Chiariello, Andrea Esposito, and Mario Nicodemi, 2020, "Polymer physics indicates chromatin folding variability across single-cells results from state degeneracy in phase separation," *Nat. Commun.* **11**, 1–13.

[28] Cremer, Thomas, and Christoph Cremer, 2001, "Chromosome territories, nuclear architecture and gene regulation in mammalian cells," *Nature reviews genetics* **2**, 292–301.

[29] Cremer, Thomas, and Marion Cremer, 2010, "Chromosome territories," *Cold Spring Harbor perspectives in biology* **2**, a003889.

[30] Dans, Pablo D, Jürgen Walther, Hansel Gómez, and Modesto Orozco, 2016, "Multiscale simulation of dna," *Curr. Opin. Struct. Biol.* **37**, 29–45.

[31] Dellino, Gaetano Ivan, Fernando Palluzzi, Andrea Maria Chiariello, Rossana Piccioni, Simona Bianco, Laura Furia, Giulia De Conti, Britta AM Bouwman, Giorgio Melloni, Davide Guido, *et al.*, 2019, "Release of paused rna polymerase ii at specific loci favors dna double-strand-break formation and promotes cancer translocations," *Nature genetics* **51**, 1011–1023.

[32] Des Cloizeaux, J, 1980, "Short range correlation between elements of a long polymer in a good solvent," *J. Phys* **41**, 223–238.

[33] Di Pierro, Michele, Bin Zhang, Erez Lieberman Aiden, Peter G Wolynes, and Jose N Onuchic, 2016 oct, "Transferable model for chromosome architecture.." *PNAS* **113**, 12168–12173.

[34] Di Stefano, Marco, Hans-Wilhelm Nützmann, Marc A Marti-Renom, and Daniel Jost, 2020, "Polymer modelling unveils the roles of heterochromatin and nucleolar organizing regions in shaping 3d genome organization in arabidopsis thaliana," *BioRxiv*.

[35] Di Stefano, Marco, Jonas Paulsen, Daniel Jost, and Marc A Marti-Renom, 2021, "4d nucleome modeling," *Curr. Opin. Genet. Dev.* **67**, 25–32.

[36] Diesinger, Philipp M, Susanne Kunkel, Jörg Langowski, and Dieter W Heermann, 2010, "Histone depletion facilitates chromatin loops on the kilobasepair scale," *Biophysical journal* **99**, 2995–3001.

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

[38] Duan, Zhijun, Mirela Andronescu, Kevin Schutz, S McIlwain, Sean McIlwain, Yoo Jung Kim, Choli Lee, Jay Shendure, Stanley Fields, C Anthony Blau, and S William, 2010, "A Three-Dimensional Model of the Yeast Genome," *Nature* **465**, 363–367.

[39] Dünweg, Burkhard, and Wolfgang Paul, 1991, "Brownian dynamics simulations without gaussian random numbers," *International Journal of Modern Physics C* **2**, 817–827.

[40] Duplantier, Bertrand, 1989, "Statistical mechanics of polymer networks of any topology," *J. Stat. Phys.* **54**, 581–680.

[41] Ecker, Joseph R, Wendy A Bickmore, Inês Barroso, Jonathan K Pritchard, Yoav Gilad, and Eran Segal, 2012, "Genomics: Encode explained," *Nature* **489**, 52.

[42] Emanuel, Marc, Nima Hamedani Radja, Andreas Henriksson, and Helmut Schiessel, 2009 jul, "The physics behind the larger scale organization of DNA in eukaryotes," *Phys. Biol.* **6**, 025008.

[43] Erdel, Fabian, Michael Baum, and Karsten Rippe, 2015, "The viscoelastic properties of chromatin and the nucleoplasm revealed by scale-dependent protein mobility," *J. Phys.: Condens. Matter* **27**, 064115.

[44] Fill, James Allen, and Donniell E Fishkind, 2000, "The Moore – Penrose Generalized Inverse for Sums of Matrices," *SIAM J. Matrix Anal. Appl.*, 629–635.

[45] Finch, John T, and Aaron Klug, 1976, "Solenoidal model for superstructure in chromatin," *Proceedings of the National Academy of Sciences* **73**, 1897–1901.

[46] Forsberg, Frida, Annaël Brunet, Tharvesh M Liyakat Ali, and Philippe Collas, 2019, "Interplay of lamin a and lamin b lads on the radial positioning of chromatin," *Nucleus* **10**, 7–20.

[47] Fraser, James, Mathieu Rousseau, Solomon Shenker, Maria A Ferraiuolo, Yoshihide Hayashizaki, Mathieu Blanchette, and Josée Dostie, 2009, "Chromatin conformation signatures of cellular differentiation.." *Genome Biol.* **10**, R37.

[48] Fraser, Peter, and Wendy Bickmore, 2007, "Nuclear organization of the genome and the potential for gene regulation," *Nature* **447**, 413.

[49] Fudenberg, Geoffrey, Maxim Imakaev, Carolyn Lu, Anton Goloborodko, Nezar Abdennur, and Leonid A Mirny, 2016, "Formation of chromosomal domains by loop extrusion," *Cell Rep.* **15**, 2038–2049.

[50] Ganai, Nirmalendu, Surajit Sengupta, and Gautam I Menon, 2014, "Chromosome positioning from activity-based segregation," *Nucleic Acids Res.* **42**, 4145–4159.

[51] Ganji, Mahipal, Indra A Shaltiel, Shveta Bisht, Eugene Kim, Ana Kalichava, Christian H Haering, and Cees Dekker, 2018, "Real-time imaging of DNA loop extrusion by condensin," *Science* **360**, 102–105.

[52] Gehlen, Lutz R, Gerd Gruenert, M Beatrix Jones, Chris D Rodley, Jorg Langowski, and J M O'Sullivan, 2012 jul, "Chromosome positioning and the clustering of functionally related loci in yeast is driven by chromosomal interactions.." *Nucleus* **3**, 370–383.

[53] Gerchman, SE, and V Ramakrishnan, 1987, "Chromatin higher-order structure studied by neutron scattering and scanning transmission electron microscopy," *Proceedings of the National Academy of Sciences* **84**, 7802–7806.

[54] Germier, Thomas, Silvia Kocanova, Nike Walther, Aurélien Bancaud, Haitham Ahmed Shaban, Hafida Sellou, Antonio Zaccaria Politi, Jan Ellenberg, Franck Gallardo, and Kerstin Bystricky, 2017, "Real-time imaging of a single gene reveals transcription-initiated local confinement," *Biophys. J.* **113**, 1383–1394.

[55] Gibson, Bryan A, Lynda K Doolittle, Maximillian WG Schneider, Liv E Jensen, Nathan Gamarra, Lisa Henry, Daniel W Gerlich, Sy Redding, and Michael K

Rosen, 2019, "Organization of chromatin by intrinsic and regulated phase separation," *Cell* **179**, 470–484.

[56] Gilbert, Nick, Susan Gilchrist, and Wendy A Bickmore, 2005, "Chromatin organization in the mammalian nucleus," *Int. Rev. Cytol.* **242**, 283–336.

[57] Goloborodko, Anton, John F Marko, and Leonid A Mirny, 2016, "Chromosome compaction by active loop extrusion," *Biophys. J.* **110**, 2162–2168.

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

[59] Haber, C, S A Ruiz, and D Wirtz, 2000 sep, "Shape anisotropy of a single random-walk polymer.." *PNAS* **97**, 10792–10795.

[60] Hänggi, Peter, Peter Talkner, and Michal Borkovec, 1990, "Reaction-rate theory: fifty years after kramers," *Reviews of modern physics* **62**, 251.

[61] Heard, Edith, and Wendy Bickmore, 2007, "The ins and outs of gene regulation and chromosome territory organisation," *Current opinion in cell biology* **19**, 311–316.

[62] Hnisz, Denes, Krishna Shrinivas, Richard A Young, Arup K Chakraborty, and Phillip A Sharp, 2017, "A phase separation model for transcriptional control," *Cell* **169**, 13–23.

[63] Hsieh, Tsung-Han S, Assaf Weiner, Bryan Lajoie, Job Dekker, Nir Friedman, and Oliver J Rando, 2015, "Mapping nucleosome resolution chromosome folding in yeast by micro-c," *Cell* **162**, 108–119.

[64] Hsu, Hsiao-Ping, Walter Nadler, and Peter Grassberger, 2004, "Scaling of star polymers with 1- 80 arms," *Macromolecules* **37**, 4658–4663.

[65] Imai, Ryosuke, Tadasu Nozaki, Tomomi Tani, Kazunari Kaizu, Kayo Hibino, Satoru Ide, Sachiko Tamura, Koichi Takahashi, Michael Shribak, and Kazuhiro Maeshima, 2017, "Density imaging of heterochromatin in live cells using orientation-independent-dic microscopy," *Mol. Biol. Cell* **28**, 3349–3359.

[66] Jost, Daniel, Pascal Carrivain, Giacomo Cavalli, and Cédric Vaillant, 2014, "Modeling epigenome folding: formation and dynamics of topologically associated chromatin domains," *Nucleic Acids Res.* **42**, 9553–9561.

[67] Jost, Daniel, Cédric Vaillant, and Peter Meister, 2017, "Coupling 1d modifications and 3d nuclear organization: data, models and function," *Current opinion in cell biology* **44**, 20–27.

[68] Kaplan, Noam, Irene K Moore, Yvonne Fondufe-Mittendorf, Andrea J Gossett, Desiree Tillo, Yair Field, Emily M LeProust, Timothy R Hughes, Jason D Lieb, Jonathan Widom, *et al.*, 2009, "The dna-encoded nucleosome organization of a eukaryotic genome," *Nature* **458**, 362–366.

[69] Kaushal, Anjali, Giriram Mohana, Julien Dorier, Isa Özdemir, Arina Omer, Pascal Cousin, Anastasiia Semenova, Michael Taschner, Oleksandr Dergai, Flavia Marzetta, *et al.*, 2021, "Ctcf loss has limited effects on global genome architecture in drosophila despite critical regulatory functions," *Nat. Commun.* **12**, 1–16.

[70] Kornberg, Roger D, 1974, "Chromatin structure: a repeating unit of histones and dna," *Science* **184**, 868–871.

[71] Kornberg, Roger D, and Yahli Lorch, 1999, "Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome," *Cell* **98**, 285–294.

[72] Kramers, Hendrik Anthony, 1940, "Brownian motion in a field of force and the diffusion model of chemical reactions," *Physica* **7**, 284–304.

[73] Kuhn, Werner, 1934, "Über die Gestalt fadenförmiger Moleküle in Lösungen," *Kolloid-Zeitschrift* **68**, 2–15.

[74] Kumari, Kiran, Burkhard Duenweg, Ranjith Padinhateeri, and J Ravi Prakash, 2020, "Computing 3D chromatin configurations from contact probability maps by inverse brownian dynamics," *Biophys. J.* **118**, 2193–2208.

[75] Larson, Adam G, and Geeta J Narlikar, 2018, "The role of phase separation in heterochromatin formation, function, and regulation," *Biochemistry* **57**, 2540–2548.

[76] Lieberman-Aiden, Erez, 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, 2009, "Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome," *Science* **326**, 289–293.

[77] Lieberman-Aiden, Erez, 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.*, 2009, "Comprehensive mapping of long-range interactions reveals folding principles of the human genome," *Science* **326**, 289–293.

[78] Lorch, Yahli, Bradley R Cairns, Mincheng Zhang, and Roger D Kornberg, 1998, "Activated rsc–nucleosome complex and persistently altered form of the nucleosome," *Cell* **94**, 29–34.

[79] Luger, Karolin, Mekonnen L Dechassa, and David J Tremethick, 2012, "New insights into nucleosome and chromatin structure: an ordered state or a disordered affair?." *Nature reviews Molecular cell biology* **13**, 436–447.

[80] Luger, Karolin, Armin W Mäder, Robin K Richmond, David F Sargent, and Timothy J Richmond, 1997, "Crystal structure of the nucleosome core particle at 2.8 å resolution," *Nature* **389**, 251–260.

[81] MacPherson, Quinn, Bruno Beltran, and Andrew J Spakowitz, 2020, "Chromatin compaction leads to a preference for peripheral heterochromatin," *Biophysical journal* **118**, 1479–1488.

[82] Maeshima, Kazuhiro, Sachiko Tamura, Jeffrey C Hansen, and Yuji Itoh, 2020, "Fluid-like chromatin: Toward understanding the real chromatin organization present in the cell," *Curr. Opin. Cell Biol.* **64**, 77–89.

[83] Mehta, Ishita S, Mugdha Kulashreshtha, Sandeep Chakraborty, Ullas Kolthur-Seetharam, and Basuthkar J Rao, 2013, "Chromosome territories reposition during dna damage-repair response," *Genome biology* **14**, 1–15.

[84] Meluzzi, Dario, and Gaurav Arya, 2013, "Recovering ensembles of chromatin conformations from contact probabilities," *Nucleic Acids Res.* **41**, 63–75.

[85] Meluzzi, Dario, and Gaurav Arya, 2013, "Recovering ensembles of chromatin conformations from contact probabilities," *Nucleic Acids Res.* **41**, 63–75.

[86] Mir, Mustafa, Wendy Bickmore, Eileen EM Furlong, and Geeta Narlikar, 2019, "Chromatin topology, condensates and gene regulation: shifting paradigms or just a phase?." *Development* **146**, dev182766.

[87] Mirny, Leonid A, 2011, "The fractal globule as a model of chromatin architecture in the cell," *Chromosome Res.* **19**, 37–51.

[88] Nair, Sreejith J, Lu Yang, Dario Meluzzi, Soohwan Oh, Feng Yang, Meyer J Friedman, Susan Wang, Tom Suter, Ibraheem Alshareedah, Amir Gamliel, *et al.*, 2019, "Phase separation of ligand-activated enhancers licenses cooperative chromosomal enhancer assembly," *Nat. Struct. Mol. Biol.* **26**, 193–203.

[89] Netz, RR, 2003, "Nonequilibrium unfolding of polyelectrolyte condensates in electric fields," *Physical review letters* **90**, 128104.

[90] Nir, Guy, Irene Farabella, Cynthia Perez Estrada, Carl G Ebeling, Brian J Beliveau, Hiroshi M Sasaki, S Dean Lee, Son C Nguyen, Ruth B McCole, Shyamtanu Chattoraj, Jelena Erceg, Jumana AlHaj Abed, Nuno M C Martins, Huy Q Nguyen, Mohammed A Hannan, Sheikh Russell, Neva C Durand, Suhas S P Rao, Jocelyn Y Kishi, Paula Soler-Vila, Michele Di Pierro, Jose N Onuchic, Steven P Callahan, John M Schreiner, Jeff A Stuckey, Peng Yin, Erez Lieberman Aiden, Marc A Marti-Renom, and C-Ting Wu, 2018 dec, "Walking along chromosomes with super-resolution imaging, contact maps, and integrative modeling.." *PLos Genet.* **14**, 1–35.

[91] Nishino, Yoshinori, Mikhail Eltsov, Yasumasa Joti, Kazuki Ito, Hideaki Takata, Yukio Takahashi, Saera Hihara, Achilleas S Frangakis, Naoko Imamoto, Tetsuya Ishikawa, *et al.*, 2012, "Human mitotic chromosomes consist predominantly of irregularly folded nucleosome fibres without a 30-nm chromatin structure," *The EMBO journal* **31**, 1644–1653.

[92] Nora, Elphège P, Job Dekker, and Edith Heard, 2013, "Segmental folding of chromosomes: A basis for structural and regulatory chromosomal neighborhoods?." *BioEssays* **35**, 818–828.

[93] Nozaki, Tadasu, Ryosuke Imai, Mai Tanbo, Ryosuke Nagashima, Sachiko Tamura, Tomomi Tani, Yasumasa Joti, Masaru Tomita, Kayo Hibino, Masato T Kanemaki, *et al.*, 2017, "Dynamic organization of chromatin domains revealed by super-resolution live-cell imaging," *Mol. Cell* **67**, 282–293.

[94] Öttinger, H C, 1996, *Stochastic Processes in Polymeric Fluids* (Springer).

[95] Ou, Horng D, Sébastien Phan, Thomas J Deerinck, Andrea Thor, Mark H Ellisman, and Clodagh C O'shea, 2017, "Chromemt: Visualizing 3d chromatin structure and compaction in interphase and mitotic cells," *Science* **357**.

[96] Parmar, Jyotsana J, and Ranjith Padinhateeri, 2020, "Nucleosome positioning and chromatin organization," *Curr. Opin. Struct. Biol.* **64**, 111–118.

[97] Parmar, Jyotsana J, Maxime Woringer, and Christophe Zimmer, 2019, "How the genome folds: the biophysics of four-dimensional chromatin organization," *Annual review of biophysics*.

[98] Paulsen, Jonas, Tharvesh M Liyakat Ali, Maxim Nekrasov, Erwan Delbarre, Marie-Odile Baudement, Sebastian Kurscheid, David Tremethick, and Philippe Collas, 2019, "Long-range interactions between topologically associating domains shape the four-dimensional genome during differentiation," *Nature genetics* **51**, 835–843.

[99] Paulsen, Jonas, Tharvesh Moideen Liyakat Ali, and Philippe Collas, 2018 apr, "Computational 3D genome modeling using Chrom3D," *Nat. Protoc.* **13**, 1137.

[100] Paulsen, Jonas, Monika Sekelja, Anja R Oldenburg, Alice Barateau, Nolwenn Briand, Erwan Delbarre, Akshay Shah, Anita L Sørensen, Corinne Vigouroux, Brigitte Buendia, and Philippe Collas, 2017 jan, "Chrom3D: three-dimensional genome modeling from Hi-C and nuclear lamin-genome contacts," *Genome Biol.* **18**, 21.

[101] Paulsen, Jonas, Monika Sekelja, Anja R Oldenburg, Alice Barateau, Nolwenn Briand, Erwan Delbarre, Akshay Shah, Anita L Sørensen, Corinne Vigouroux, Brigitte Buendia, *et al.*, 2017, "Chrom3d: three-dimensional genome modeling from hi-c and nuclear lamin-genome contacts," *Genome biology* **18**, 1–15.

[102] Pham, Tri Thanh, Mohit Bajaj, and J Ravi Prakash, 2008, "Brownian dynamics simulation of polymer collapse in a poor solvent: Influence of implicit hydrodynamic interactions," *Soft Matter* **4**, 1196–1207.

[103] Pham, Tri Thanh, Burkhard Duenweg, and J Ravi Prakash, 2010, "Collapse dynamics of copolymers in a poor solvent: Influence of hydrodynamic interactions and chain sequence," *Macromolecules* **43**, 10084–10095.

[104] Poirier, Michael G, and John F Marko, 2002, "Mitotic chromosomes are chromatin networks without a mechanically contiguous protein scaffold," *Proceedings of the National Academy of Sciences* **99**, 15393–15397.

[105] Prabhakar, R, and J.Ravi Prakash, 2004, "Multiplicative separation of the influences of excluded volume, hydrodynamic interactions and finite extensibility on the rheological properties of dilute polymer solutions," *J. Non-Newtonian Fluid Mech.* **116**, 163–182.

[106] Press, William H, Saul A Teukolsky, William T Vetterling, and Brian P Flannery, 1992, *Numerical Recipes in C*, 2nd ed. (Cambridge University Press, Cambridge, USA).

[107] Rao, Suhas S.P., Miriam H. Huntley, Neva C. Durand, Elena K. Stamenova, Ivan D. Bochkov, James T. Robinson, Adrian L. Sanborn, Ido Machol, Arina D. Omer, Eric S. Lander, and Erez Lieberman Aiden, ????

[108] Rousseau, Mathieu, James Fraser, Maria A Ferraiuolo, Josée Dostie, and Mathieu Blanchette, 2011, "Three-dimensional modeling of chromatin structure from interaction frequency data using markov chain monte carlo sampling," *BMC Bioinf.* **12**, 414.

[109] Rowley, M Jordan, and Victor G Corces, 2018, "Organizational principles of 3d genome architecture," *Nat. Rev. Genet.* **19**, 789–800.

[110] Rubinstein, M., and R. H. Colby, 2003, *Polymer Physics* (Oxford University Press).

[111] Rubinstein, Michael, Ralph H Colby, *et al.*, 2003, *Polymer physics*, Vol. 23 (Oxford university press New York).

[112] Ryu, Je-Kyung, Céline Bouchoux, Hon Wing Liu, Eugene Kim, Masashi Mi-
namino, Ralph de Groot, Allard J Katan, Andrea Bonato, Davide Marenduzzo,
Davide Michieletto, *et al.*, 2021, "Bridging-induced phase separation induced by
cohesin smc protein complexes," *Science Advances* **7**, eabe5905.

[113] Sandholtz, Sarah H, Quinn MacPherson, and Andrew J Spakowitz, 2020, "Physical
modeling of the heritability and maintenance of epigenetic modifications," *PNAS*
**117**, 20423–20429.

[114] Santra, Aritra, B Dünweg, and J Ravi Prakash, ????, "Universal scaling and char-
acterisation of gelation in associative polymer solutions," *arXiv:2008.05641*.

[115] Santra, Aritra, Kiran Kumari, Ranjith Padinhateeri, B. Dünweg, and J. Ravi
Prakash, 2019, "Universality of the collapse transition of sticky polymers," *Soft
Matter* **15**, 7876–7887.

[116] Schiessel, Helmut, 2003 may, "The physics of chromatin," *J. Phys.: Condens. Mat-
ter* **15**, R699–R774.

[117] Schiessel, Helmut, William M Gelbart, and Robijn Bruinsma, 2001, "Dna fold-
ing: structural and mechanical properties of the two-angle model for chromatin,"
*Biophysical Journal* **80**, 1940–1956.

[118] Schlick, Tamar, Jeff Hayes, and Sergei Grigoryev, 2012, "Toward convergence of
experimental studies and theoretical modeling of the chromatin fiber," *Journal of
Biological Chemistry* **287**, 5183–5191.

[119] Segal, Eran, Yvonne Fondufe-Mittendorf, Lingyi Chen, AnnChristine Thåström,
Yair Field, Irene K Moore, Ji-Ping Z Wang, and Jonathan Widom, 2006, "A ge-
nomic code for nucleosome positioning," *Nature* **442**, 772–778.

[120] Shi, Guang, Lei Liu, Changbong Hyeon, and Dave Thirumalai, 2018, "Interphase
human chromosome exhibits out of equilibrium glassy dynamics," *Nat. Commun.*
**9**, 1–13.

[121] Shrinivas, Krishna, Benjamin R Sabari, Eliot L Coffey, Isaac A Klein, Ann Boija,
Alicia V Zamudio, Jurian Schuijers, Nancy M Hannett, Phillip A Sharp, Richard A

Young, *et al.*, 2019, "Enhancer features that drive formation of transcriptional condensates," *Mol. Cell* **75**, 549–561.

[122] Sicard, François, 2018, "Computing transition rates for rare events: When kramers theory meets the free-energy landscape," *Physical Review E* **98**, 052408.

[123] Sinha, Bidisha, Dipanjan Bhattacharya, Deepak Kumar Sinha, Shefali Talwar, Shovamayee Maharana, Soumya Gupta, and GV Shivashankar, 2010, "Dynamic organization of chromatin assembly and transcription factories in living cells," *Methods in cell biology* **98**, 57–78.

[124] Soddemann, T, B Dünweg, and K Kremer, 2001 dec, "A generic computer model for amphiphilic systems," *Eur. Phys. J. E: Soft Matter Biol. Phys.* **6**, 409–419.

[125] Šolc, Karel, 1971 jul, "Shape of a Random-Flight Chain," *J. Chem. Phys.* **55**, 335–344.

[126] Soysa, W Chamath, B Dunweg, and J Ravi Prakash, 2015 aug, "Size, shape, and diffusivity of a single Debye-Huckel polyelectrolyte chain in solution.." *J. Chem. Phys.* **143**, 64906.

[127] Steinhauser, Martin Oliver, 2005 mar, "A molecular dynamics study on universal properties of polymer chains in different solvent qualities. Part I. A review of linear chain properties.." *J. Chem. Phys.* **122**, 94901.

[128] Strickfaden, Hilmar, Thomas O Tolsma, Ajit Sharma, D Alan Underhill, Jeffrey C Hansen, and Michael J Hendzel, 2020, "Condensed chromatin behaves like a solid on the mesoscale in vitro and in living cells," *Cell*.

[129] Szabo, Quentin, Axelle Donjon, Ivana Jerković, Giorgio L Papadopoulos, Thierry Cheutin, Boyan Bonev, Elphège P Nora, Benoit G Bruneau, Frederic Bantignies, and Giacomo Cavalli, 2020, "Regulation of single-cell genome organization into tads and chromatin nanodomains," *Nat. Genet.* **52**, 1151–1157.

[130] Szabo, Quentin, Daniel Jost, Jia-Ming Chang, Diego I Cattoni, Giorgio L Papadopoulos, Boyan Bonev, Tom Sexton, Julian Gurgo, Caroline Jacquier, Marcelo Nollmann, *et al.*, 2018, "Tads are 3d structural units of higher-order chromosome organization in drosophila," *Science advances* **4**, eaar8082.

[131] Talwar, Shefali, Abhishek Kumar, Madan Rao, Gautam I Menon, and GV Shivashankar, 2013, "Correlated spatio-temporal fluctuations in chromatin compaction states characterize stem cells," *Biophysical journal* **104**, 553–564.

[132] Tanizawa, Hideki, Osamu Iwasaki, Atsunari Tanaka, Joseph R Capizzi, Priyankara Wickramasinghe, Mihee Lee, Zhiyan Fu, and Ken Ichi Noma, 2010, "Mapping of long-range associations throughout the fission yeast genome reveals global genome organization linked to transcriptional regulation," *Nucleic Acids Res.* **38**, 8164–8177.

[133] Teif, Vladimir B, and Klemen Bohinc, 2011, "Condensed dna: condensing the concepts," *Prog. Biophys. Mol. Biol.* **105**, 208–222.

[134] Theodorou, Doros N, and Ulrich W Suter, 1985 jun, "Shape of unperturbed linear polymers: polypropylene," *Macromolecules* **18**, 1206–1214

> KEY: Theodorou1985
> ANNOTATION: doi: 10.1021/ma00148a028.

.

[135] Toan, Ngo Minh, Greg Morrison, Changbong Hyeon, and D Thirumalai, 2008, "Kinetics of loop formation in polymer chains," *J. Phys. Chem. B* **112**, 6094–6106.

[136] Tripathi, Kamal, Gautam I Menon, and Satyavani Vemparala, 2019, "Confined crowded polymers near attractive surfaces," *The Journal of chemical physics* **151**, 244901.

[137] Uhler, Caroline, and GV Shivashankar, 2017, "Regulation of genome organization and gene expression by nuclear mechanotransduction," *Nature reviews Molecular cell biology* **18**, 717–727.

[138] Wei, Pei-Chi, Cheng-Sheng Lee, Zhou Du, Bjoern Schwer, Yuxiang Zhang, Jennifer Kao, Jeffrey Zurita, and Frederick W Alt, 2018, "Three classes of recurrent DNA break clusters in brain progenitors identified by 3d proximity-based break joining assay," *PNAS* **115**, 1919–1924.

[139] Widom, J, and A Klug, 1985, "Structure of the 3000å chromatin filament: X-ray diffraction from oriented samples," *Cell* **43**, 207–213.

[140] de Wit, Elzo, 2020, "Tads as the caller calls them," *Journal of molecular biology* **432**, 638–642.

[141] Witten Jr, TA, and JJ Prentis, 1982, "The interpenetration of two chain polymers in a good solvent," *J. Chem. Phys.* **77**, 4247–4253.

[142] Yan, Jie, Ryo Kawamura, and John F Marko, 2005, "Statistics of loop formation along double helix dnas," *Physical Review E* **71**, 061905.

[143] Yang, Janet G, Tina Shahian Madrid, Elena Sevastopoulos, and Geeta J Narlikar, 2006, "The chromatin-remodeling enzyme acf is an atp-dependent DNA length sensor that regulates nucleosome spacing," *Nat. Struct. Mol. Biol.* **13**, 1078–1083.

[144] Zhang, Yu, Rachel Patton McCord, Yu-Jui Ho, Bryan R Lajoie, Dominic G Hildebrand, Aline C Simon, Michael S Becker, Frederick W Alt, and Job Dekker, 2012, "Spatial organization of the mouse genome and its role in recurrent chromosomal translocations," *Cell* **148**, 908–921.

[145] Zifferer, Gerhard, 1999 sep, "Monte Carlo simulation studies of the size and shape of linear and star-branched polymers embedded in the tetrahedral lattice," *Macromol. Theory Simul.* **8**, 433–462.

# Appendix A

## A.1 Non-dimensionalization techniques

In this section, we will compare two different schemes of non-dimensionalization and show that the simulation results are unaffected by the non-dimensional scheme. Throughout the thesis, we used the non-dimensional scheme inspired from the Bird *et al.* [21]. Here in this section, we compare this scheme with another scheme (inspired by Netz [89]) and establish the consistent results, from both the simulations.

**The Fokker-Planck equation**

The Fokker-Planck equation for the density distribution function ($\psi$) of a single bead under the force F can be written as

$$\frac{\partial \psi}{\partial t} = -\frac{1}{\zeta} \frac{\partial}{\partial r} F \psi + \frac{k_B T}{\zeta} \frac{\partial^2 \psi}{\partial r^2} \tag{A.1}$$

where $\zeta = 6\pi\eta a$ is the Stokes friction coefficient of a spherical bead, $\eta$ is the solvent viscosity, and $a$ is the bead radius. $k_B$ is the Boltzmann constant, $T$ is temperature, $r$ is the bead position and $t$ is the time.

**Bird's non-dimensional FP equation**

In this section, we will use the Bird's non-dimension formulation. All lengths are made dimensionless by using the characteristic length scale $l_H = \sqrt{k_B T/H}$. Here, $H$ is coefficient of the harmonic force ($F$). Characteristic time scale $\lambda_H = \zeta/4H$ is used to make time dimensionless. Non-dimensional quantities (indicated with an asterisk " $*$ ") can be written as

$$r^* = \frac{r}{l_H}; \quad F^* = \frac{F}{\sqrt{k_B T H}}; \quad \psi^* = \psi l_H^3; \quad t^* = \frac{t}{\lambda_H}; \tag{A.2}$$

Using the above dimensionless parameters, the FP equation in Eq. 1 can be written as

$$\frac{\partial \psi^*}{\partial t^*} \frac{1}{l_H^3 \lambda_H} = -\frac{1}{\zeta} \frac{\partial}{\partial r^* l_H} F^* \sqrt{k_B T H} \frac{\psi^*}{l_H^3} + \frac{k_B T}{\zeta} \frac{\partial^2 \psi^*}{\partial r^{*2}} \frac{1}{l_H^2 l_H^3} \tag{A.3}$$

$$\frac{\partial \psi^*}{\partial t^*} \frac{1}{\lambda_H} = -\frac{1}{\zeta} \frac{\partial}{\partial r^* l_H} F^* \sqrt{k_B T H} \psi^* + \frac{k_B T}{\zeta} \frac{\partial^2 \psi^*}{\partial r^{*2}} \frac{1}{l_H^2} \tag{A.4}$$

$$\frac{\partial \psi^*}{\partial t^*} = -\left[ \frac{1}{\zeta} \frac{1}{l_H} \sqrt{k_B T H} \lambda_H \right] \frac{\partial}{\partial r^*} F^* \psi^* + \left[ \frac{k_B T}{\zeta} \frac{1}{l_H^2} \lambda_H \right] \frac{\partial^2 \psi^*}{\partial r^{*2}} \tag{A.5}$$

Further simplifying the terms in brackets by substituting the value of $l_H = \sqrt{k_B T / H}$

$$\frac{\partial \psi^*}{\partial t^*} = -\left[ \frac{1}{\zeta} \sqrt{\frac{H}{k_B T}} \sqrt{k_B T H} \lambda_H \right] \frac{\partial}{\partial r^*} F^* \psi^* + \left[ \frac{k_B T}{\zeta} \frac{H}{k_B T} \lambda_H \right] \frac{\partial^2 \psi^*}{\partial r^{*2}} \tag{A.6}$$

$$\frac{\partial \psi^*}{\partial t^*} = -\left[ \frac{1}{\zeta} H \lambda_H \right] \frac{\partial}{\partial r^*} F^* \psi^* + \left[ \frac{H}{\zeta} \lambda_H \right] \frac{\partial^2 \psi^*}{\partial r^{*2}} \tag{A.7}$$

Now, substituting the value of $\lambda_H = \zeta / 4H$

$$\frac{\partial \psi^*}{\partial t^*} = -\left[ \frac{1}{\zeta} H \frac{\zeta}{4H} \right] \frac{\partial}{\partial r^*} F^* \psi^* + \left[ \frac{H}{\zeta} \frac{\zeta}{4H} \right] \frac{\partial^2 \psi^*}{\partial r^{*2}} \tag{A.8}$$

$$\frac{\partial \psi^*}{\partial t^*} = -\frac{1}{4} \frac{\partial}{\partial r^*} F^* \psi^* + \frac{1}{4} \frac{\partial^2 \psi^*}{\partial r^{*2}} \tag{A.9}$$

The Stochastic Differential Equation (SDE) from FP in the form

$$\frac{\partial \psi}{\partial t} = -\frac{\partial}{\partial r} (A\psi) + \frac{1}{2} \frac{\partial^2}{\partial r^2} (D\psi) \tag{A.10}$$

using Ito's formulation is

$$dr = A\, dt + B\, dW \tag{A.11}$$

where $D = B.B^T$.

Following the Ito's formulation, the equivalent SDE for FP in Eq. A.9 is

$$dr^* = \frac{1}{4} F^* dt^* + \frac{1}{\sqrt{2}} dW^* \tag{A.12}$$

$$r^*(t^* + dt^*) = r^*(t^*) + \frac{1}{4} F^* dt^* + \frac{1}{\sqrt{2}} dW^* \tag{A.13}$$

time in dimension can be written as

$$t = t^* \lambda_H \tag{A.14}$$

$$t = t^* \frac{\zeta}{4H} \tag{A.15}$$

**Another non-dimensionalization scheme**

All the length scale are made dimensionless by the monomer radius $a$ and the time scales by the time step $dt$. Dimensionless quantities (indicated by "ˆ") can then be written as

$$\hat{r} = \frac{r}{a}; \quad \hat{F} = \frac{F}{k_B T / a}; \quad \hat{\psi} = \psi a^3; \quad \hat{t} = \frac{t}{\partial t}; \quad \hat{\partial t} = \frac{\partial t}{\partial t} = 1 \tag{A.16}$$

Eq. 1 can be written as

$$\frac{\partial \hat{\psi}}{\partial \hat{t}} \frac{1}{a^3 \partial t} = -\frac{1}{\zeta} \frac{\partial}{\partial \hat{r} a} \hat{F} \frac{k_B T}{a} \frac{\hat{\psi}}{a^3} + \frac{k_B T}{\zeta} \frac{\partial^2 \hat{\psi}}{\partial \hat{r}^2} \frac{1}{a^2 a^3} \tag{A.17}$$

$$\frac{\partial \hat{\psi}}{\partial \hat{t}} = -\left[\frac{k_B T \partial t}{a^2 \zeta}\right] \frac{\partial}{\partial \hat{r}} \hat{F} \hat{\psi} + \left[\frac{k_B T \partial t}{a^2 \zeta}\right] \frac{\partial^2 \hat{\psi}}{\partial \hat{r}^2} \tag{A.18}$$

$$\frac{\partial \hat{\psi}}{\partial \hat{t}} = -\frac{1}{\hat{\zeta}} \frac{\partial}{\partial \hat{r}} \hat{F} \hat{\psi} + \frac{1}{\hat{\zeta}} \frac{\partial^2 \hat{\psi}}{\partial \hat{r}^2} \tag{A.19}$$

where

$$\frac{1}{\hat{\zeta}} = \frac{k_B T \partial t}{a^2 \zeta} \tag{A.20}$$

The dimensionless friction coefficient $\frac{1}{\hat{\zeta}}$ is the only parameter in the simulation. Please note that the $\partial t$ can be obtained from the above equation.

Following the Itô formulation, the equivalent SDE for FP in Eq. A.19 is

$$d\hat{r} = \frac{1}{\hat{\zeta}} \hat{F} \partial \hat{t} + \sqrt{\frac{2}{\hat{\zeta}}} d\hat{W} \tag{A.21}$$

where $\hat{W}$ is the Wiener process. Now, substituting $d\hat{r} = \hat{r}(\hat{t} + 1) - \hat{r}(\hat{t})$, the above equation becomes

$$\hat{r}(\hat{t} + 1) = \hat{r}(\hat{t}) + \frac{1}{\hat{\zeta}} \hat{F} \partial \hat{t} + \sqrt{\frac{2}{\hat{\zeta}}} d\hat{W} \tag{A.22}$$

Substituting the value of $\partial t$ from Eq. A.20 in Eq. 16, we can write the time in dimension ($t$) as

$$t = \frac{\hat{t} \zeta a^2}{k_B T \hat{\zeta}} \tag{A.23}$$

**Comparing both the non-dimensional equations**

For the comparison purpose, we can write the SDE given in Eq. A.21 as

$$\frac{dr}{a} = \frac{1}{\hat{\zeta}}\hat{F}\partial\hat{t} + \sqrt{\frac{2}{\hat{\zeta}}}d\hat{W} \qquad (A.24)$$

$$dr = \frac{a}{\hat{\zeta}}\hat{F}\partial\hat{t} + a\sqrt{\frac{2}{\hat{\zeta}}}d\hat{W} \qquad (A.25)$$

Note that the *dr* in LHS is in dimensional units. Similarly, the SDE given in Eq. A.12 can be written as

$$\frac{dr}{l_H} = \frac{1}{4}F^*dt^* + \frac{1}{\sqrt{2}}dW^* \qquad (A.26)$$

$$dr = \frac{l_H}{4}F^*dt^* + \frac{l_H}{\sqrt{2}}dW^* \qquad (A.27)$$

The LHS of Eq. A.25 and A.27 are the same. Hence, the diffusion and the drift terms should be the same. Comparing the diffusion term Eq. A.25 and A.27, we get

$$ \qquad (A.28)$$

$$a\sqrt{\frac{2}{\hat{\zeta}}}d\hat{W} = \frac{l_H}{\sqrt{2}}dW^* \qquad (A.29)$$

$$ \qquad (A.30)$$

$$a\sqrt{\frac{2}{\hat{\zeta}}}\sqrt{d\hat{t}}dW = \frac{l_H}{\sqrt{2}}\sqrt{dt^*}dW \qquad (A.31)$$

since $\sqrt{d\hat{t}} = 1$

$$ \qquad (A.32)$$

$$\sqrt{\frac{1}{\hat{\zeta}}} = \frac{1}{2}\frac{l_H}{a}\sqrt{dt^*} \qquad (A.33)$$

$$ \qquad (A.34)$$

$$\frac{1}{\hat{\zeta}} = \frac{1}{4}\frac{l_H^2}{a^2}dt^* \qquad (A.35)$$
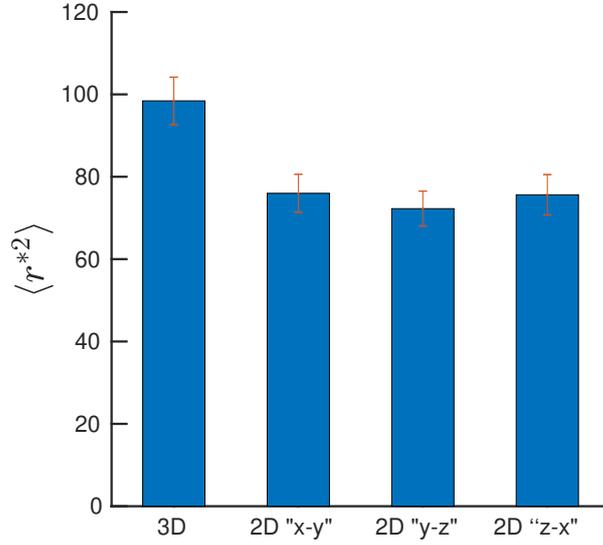
**Figure A.1:** Depiction of the calculation of 2D distance between bead 5 and bead 40 from simulation.

Above equation is the relation between the $\hat{\zeta}$ and $dt^*$. For simplicity, we consider $l_H = a$. Hence,

$$\tag{A.36}$$

$$\frac{1}{\hat{\zeta}} = \frac{1}{4}dt^* \tag{A.37}$$

This indicates that the $\frac{1}{4}dt^*$ in Bird's non-dimensional scheme is similar to $\frac{1}{\hat{\zeta}}$ in the other non-dimensional scheme. This establishes the fact the simulation results when converted to dimensional form are independent of the non-dimensional scheme used.

## A.2    Conversion of non-dimensional length and time to standard units

In our simulations, all quantities are computed in dimensionless units as described earlier. To convert these dimensionless numbers to standard units having appropriate dimensions, we need to determine a lengthscale and a timescale. By comparing our simulations with appropriate experimental observations, we deduce values of characteristic length and time scales that can be used for the unit conversion as follows:

**Length scale:** Even though the 3D distances between the genomic segments of $\alpha$−globin are not available, the 2D distance between the two probes located at $34, 512$ - $77, 058$ bp and $386, 139$ - $425, 502$ bp for GM12878 (OFF state) was found to be $318.8 \pm 17.0$ nm from the 2D FISH [10]. We computed the average 2D distance (= 8.81) between the corresponding bead pair (bead 5 and bead 40) from our simulation by averaging it over all the three 2D planes ($xy$, $yz$, $zx$) as depicted in Fig. A.1. By comparing 2D distance values obtained from simulation and experiment, we estimate the characteristic lengthscale in our simulation as $l_H = 318.8/8.81 \approx 36$nm. We use this value of $l_H$ to convert all non-dimensional lengths to standard units.

**Time scale:** The timescale in our simulation is given by:

$$\lambda_H = \frac{\zeta}{4H} = \frac{6\pi\eta_s a^3}{4k_B T} \tag{A.38}$$

where $H$ is the spring constant, $T$ is the absolute temperature, $k_B$ is the Boltzmann constant, and $\zeta = 6\pi\eta_s a$ is the Stokes friction coefficient of a spherical bead of radius $a$ where $\eta_s$ is the solvent viscosity. For our problem $a = h^* l_H \sqrt{\pi} \approx 16$nm. However, we do not know the precise viscosity of the solvent in the nucleus. There are many estimates ranging over several orders of magnitude from $10^{-3}$ Pa.s to $10^3$ Pa.s. [43, 25]. Given this degree of variability, we decided to use a simple method to estimate time, based on recent experimental reports of chromatin dynamics. Chromatin segments under microscope seems to "diffuse" around in a region having the size of the order of $\approx 0.1(\mu m)^2$ within a timescale of $\approx 50$ seconds [54]. This leads to a diffusion coefficient ($D$) of the order of 500 nm$^2$/s, and a timescale

$$\lambda_H = \frac{a^2}{4D} = \frac{(16 \text{ nm})^2}{4 \times 500 \text{ nm}^2/\text{s}} = 0.12 \text{ s} \tag{A.39}$$

Since the calculation is to estimate the order of magnitude number, throughout this work, we use $\lambda_H = 0.1$ s. Interestingly, this also corresponds to an effective viscosity roughly in the middle of the wide range estimated previously.

## A.3 Distance probability distributions

The analytical expression given by des Cloizeaux [32] for the distance probability distribution for a self-avoiding walk (SAW) polymer is

$$p(r^*) = C[r^*]^{\theta+2} e^{-[Kr^*]^{\frac{1}{1-\nu}}} \tag{A.40}$$

Here $\nu$ is the Flory exponent, $\theta$ is a geometrical exponent and the coefficients $C$ and $K$ are given by

$$K^2 = \frac{\Gamma([\theta + d + 2][1 - \nu])}{\Gamma([\theta + d][1 - \nu])}$$

$$C = 4\pi \frac{[\Gamma([\theta + d + 2][1 - \nu])]^{\frac{\theta + d}{2}}}{[\Gamma([\theta + d][1 - \nu])]^{\frac{\theta + d + 2}{2}}}$$

where $d$ is the dimension. Since our simulations are in 3D, $d = 3$. The geometrical exponent $\theta$ takes different values in the following three cases

1. Case 1: When both beads are the end beads of the polymer ($\theta = \theta_0$),

2. Case 2: When one of the beads is at the end and the other bead is an intermediate bead within the chain ($\theta = \theta_1$),

3. Case 3: When both the beads are intermediate beads ($\theta = \theta_2$)

As the coefficients $C$ and $K$ depend on $\theta$, they take different values in each of the above cases. Following the findings of des Cloizeaux [32], Witten and Prentis [141], Duplantier [40] and Hsu et al. [64], one can determine that $\theta_0 = 0.267$, $\theta_1 = 0.461$ and $\theta_2 = 0.814$ [114]. Simulating a SAW polymer, we compared the probability distribution for all the three cases with the corresponding analytical expressions using the appropriate values of $\theta$. Fig. A.2 show the validation for case 1 and 2, while the validation for case 3 has been shown in the main text. As can be seen, the simulations are in excellent agreement with the analytical expression. To the best of our knowledge, this is the first comparison of exact numerical results with the analytical expression proposed by des Cloizeaux.

We studied the distance probability distributions of various bead-pairs revealing a distribution with two peaks where one of the peaks is dominated by entropy of the polymer (genomic separation). The other peak emerges with an increase in the interaction strength. Fig. 6.4(a) depicts the Cumulative distance distribution $C(r^*)$ for various bead-pairs at same genomic separation ($s_{ij} = 25$) experiencing different interaction strengths. Differences in these plot can be easily noticed only at the small $r_{ij}$ while they looks similar overall (see inset). The same has been depicted for a specific bead-pair $(5, 30)$ in different epigenetic states in Fig. 6.4(b). The difference in this case is not only observable for small $r_{ij}$, but for the whole regime of $r_{ij}$ as can be seen in the inset.
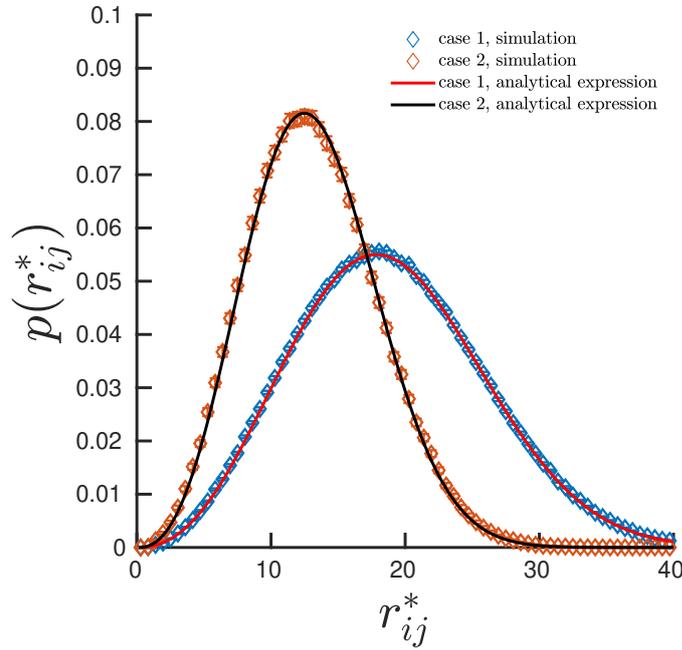
**Figure A.2:** Comparison of distance probability distributions obtained from the simulation of a SAW chain with the analytical expression for case 1 - where both beads are end beads, and case 2 - where one bead is at the chain end and the other bead is an intermediate bead.

## A.4    Temporal quantities from equivalent 1D simulation

The phase-space exploration of a specific bead-pair in a bead-spring chain can be imagined as a single bead moving in an effective potential energy landscape given by the free energy $F^*(r^*) = -\ln(p(r^*)/4\pi r^{*2})$ as a function of the 3D distance ($r^*$) between the corresponding beads in the pair of beads. In this section, we will present the dynamics of a single bead in this free energy landscape.

However, before studying the problem in a new free energy landscape, we reproduced known standard results for single particle dynamics in the following three cases

**Wiener process**

For a Wiener process, the probability distribution function obeys the Fokker-Planck equation:

$$\frac{\partial \psi}{\partial t} = \frac{1}{2}\frac{\partial^2 \psi}{\partial x^2} \tag{A.41}$$
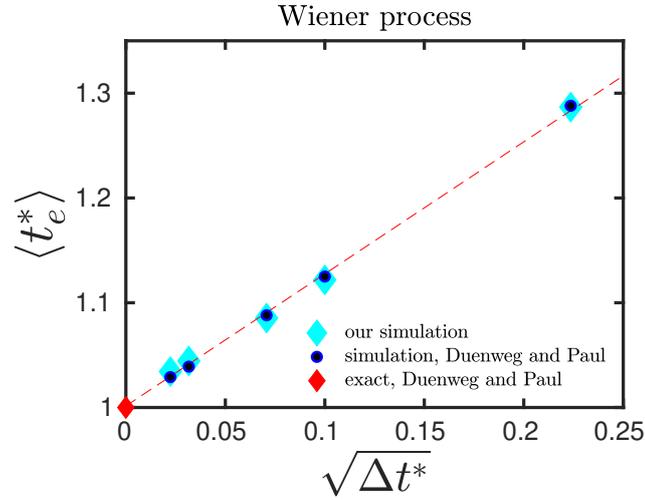
**Figure A.3:** The mean first exit time as a function of time step of the simulation for the Wiener process.

This describes a randomly diffusing particle along the spatial coordinate $x$ as a function of time $t$. We simulated the corresponding stochastic differential equation and calculated the first exit time ($t_e$) for the Wiener process. $t_e$ is defined as the time it takes for the position of the particle $x$ to exit a region $I$ for the first time, starting for a fixed position $x = x_0$ at time $t = 0$. The mean first exit time $\langle t_e \rangle$ was computed for an ensemble of trajectories. We performed Brownian dynamics for five different time steps ($\Delta t = 0.0005, 0.001, 0.005, 0.01, 0.05$) and generated $10^6$ trajectories to calculate the $\langle t_e \rangle$ for the interval $[0, 1]$ . We compared our results with the simulation results of Dünweg and Paul [39]. We also compared the extrapolated result to the $\Delta t = 0$ with the exact solution of Dünweg and Paul [39]. Fig. A.3 shows the agreement of our simulation results with the exact and simulation results of Dünweg and Paul [39].

**Ornstein-Uhlenbeck process**

The Ornstein-Uhlenbeck process can be described using the equation

$$\frac{\partial \psi}{\partial t} = \left( \frac{\partial}{\partial x} x + \frac{1}{2} \frac{\partial^2}{\partial x^2} \right) \psi \tag{A.42}$$

Similar to the Wiener process, we simulated the Ornstein-Uhlenbeck process too using Brownian dynamics for five different time steps ($\Delta t = 0.0005, 0.001, 0.005, 0.01, 0.05$) and generated $10^6$ trajectories to calculate the average quantities. Here too the first exit
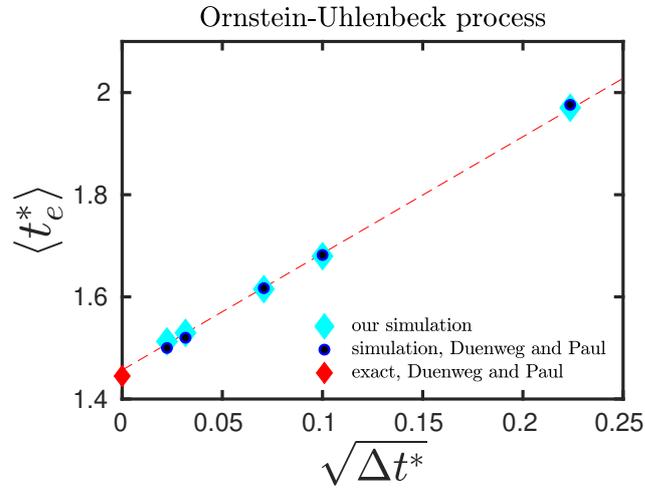
**Figure A.4:** The mean first exit time as a function of time step of the simulation for the Ornstein-Uhlenbeck process.

time $\langle t_e \rangle$ is calculated to exit from the interval $[-1, 1]$ (on either side), starting from $x = 0$. Fig. A.4 shows the comparison of our simulation and extrapolated value at $\Delta t = 0$ with the simulation and exact solution of Dünweg and Paul [39].

**Ginzburg-Landau process**

$$\frac{\partial \psi}{\partial t} = \left( \frac{\partial}{\partial x}(x^3 - x) + D\frac{\partial^2}{\partial x^2} \right)\psi \qquad (A.43)$$

In the Ginzburg-Landau case, the particle is moving in double-well potential given by

$$U = \frac{1}{4}x^4 - \frac{1}{2}x^2 \qquad (A.44)$$

See Fig. A.5(a) for the plot of the potential. $t_e$ is defined as the escape time from one of the symmetric minima ($x = \pm 1$) over the barrier at $x = 0$. Fig. A.5(b) shows the comparison of our simulation and extrapolated value at $\Delta t = 0$ with the simulation and exact solution of Dünweg and Paul [39].

As we are computing the escape time of a particle from the double-well potential, the escape time can also be compared with the Kramers' rate theory [72, 60, 122]. The Kramers' rate of escape ($t_e^{-1}$) is given by

$$k = \left( \frac{\omega_0 \, \omega_1}{2\pi\zeta} \right) e^{-\Delta U/K_B T}. \qquad (A.45)$$

(a)



(b)
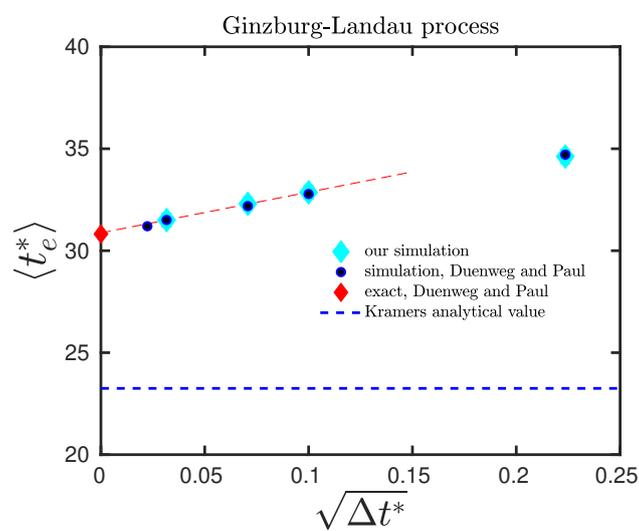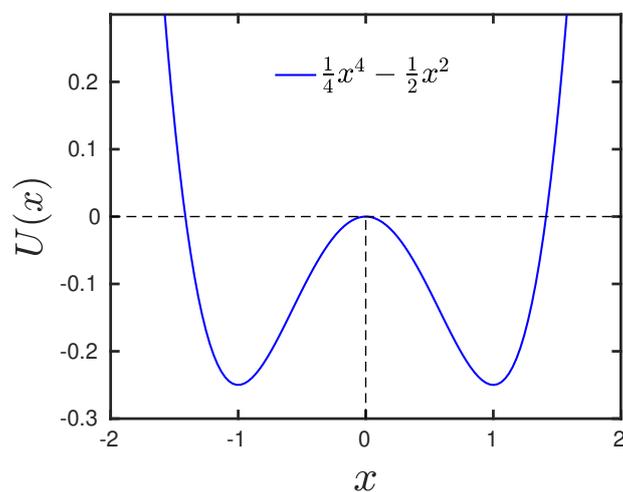
**Figure A.5:** (a) The double-well potential used in the Ginzburg-Landau process. (b) The mean first exit time as a function of time step of the simulation for the Ginzburg-Landau process. The analytical value from Kramer's first passage time theory is also indicated here.

Here, $\omega_0^2 = \left|\frac{\partial^2 U}{\partial x^2}\right|_{x=0}$ and $\omega_1^2 = \left|\frac{\partial^2 U}{\partial x^2}\right|_{x=1}$ are the magnitudes of curvature values of the potential evaluated at the barrier crossing point ($x = 0$) and at the minima ($x = 1$), respectively. $\Delta U = |U(x = 1) - U(x = 0)|$ is the height of the barrier. Evaluating Eq. A.45 using the Ginzburg-Landau potential given in Eg. A.44, we get

$$k = \left(\frac{\sqrt{2}\,\sqrt{1}\,[K_{\mathrm{B}}T/a^2]}{2\pi\zeta}\right)e^{-1/4} \tag{A.46}$$

$$k = \left(\frac{1}{\sqrt{2}\pi\zeta}\frac{K_{\mathrm{B}}T}{a^2}\right) \times 0.778 \tag{A.47}$$

$$k = \left(\frac{1}{4\sqrt{2}\pi}\frac{4K_{\mathrm{B}}T}{\zeta a^2}\right) \times 0.778 \tag{A.48}$$

$$k = \left(\frac{1}{4\sqrt{2}\pi}\frac{1}{\lambda_H}\right) \times 0.778 \tag{A.49}$$

$$k = \left(\frac{1}{17.74}\frac{1}{\lambda_H}\right) \times 0.778 \tag{A.50}$$

$$k = 0.043\frac{1}{\lambda_H} \tag{A.51}$$

Hence the escape time is

$$t_e = k^{-1} = 23.25\lambda_H \tag{A.52}$$

The value from Kramers' analytical formulae is indicated by the blue dashed line in Fig. A.5(b). It can be easily seen from the Fig. A.5(b) that there is a difference between the first passage time and the Kramers' solution. It is expected that this difference will become negligible when $\Delta U \gg K_{\mathrm{B}}T$.

In all the above cases, it can be observed that the simulation results are dependent on the time step used in the simulation. Therefore, to estimate the exact value (independent of the time step), one needs to extrapolate the simulation results to the zero time step limit ($\Delta t \to 0$).
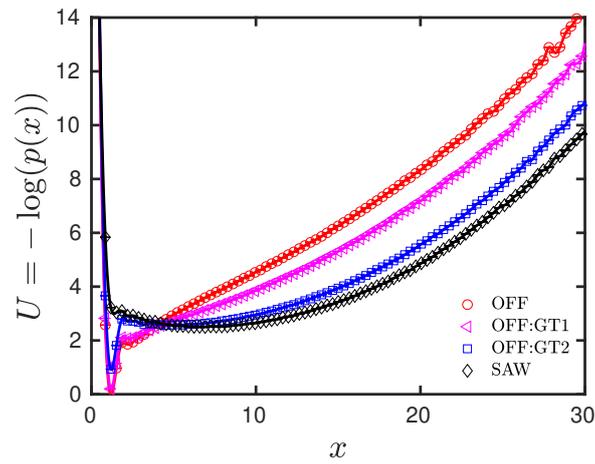
(a)
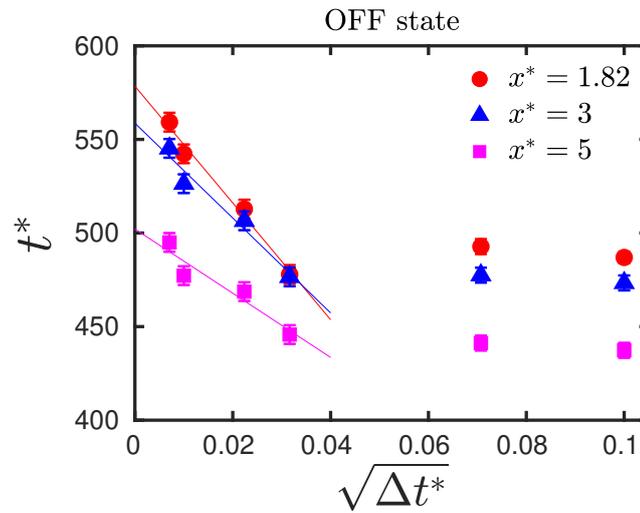


(b)

**Figure A.6:** (a) Effective potential energy landscape for the bead calculated as $F^*(r^*) = -\ln(p(r^*)/4\pi r^{*2})$. (b) Time step extrapolation result for the first passage time of a single bead under the effective potential landscape given by the free energy. Different colour indicated different destination position.
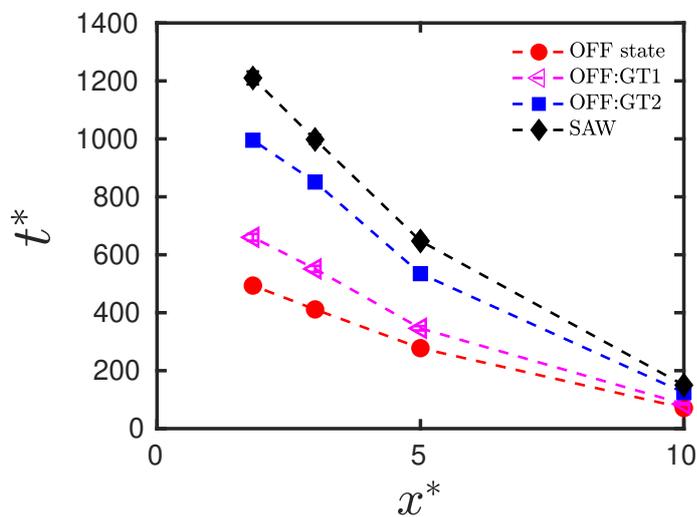
(a)
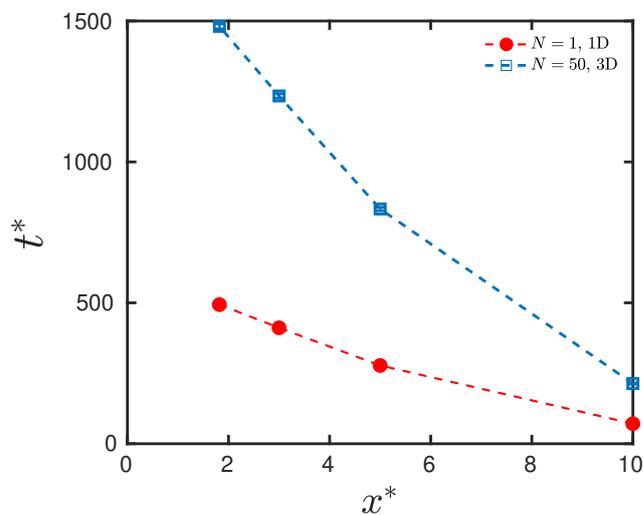


(b)

**Figure A.7:** (a) Time to reach different positions starting at $x^* = 15$. Different curves are for different epigenetic states. (b) Comparison of the first passage time result from full blown 3D polymer simulation with the equivalent case of a single particle simulation in the effective 1D free energy landscape.

After validation of single bead behaviour in three different potentials, we simulated the bead in the effective potential energy landscape given by the free energy of a specific bead-pair (bead 5 and 30) $U = F^*(r^*) = -\ln(p(r^*)/4\pi r^{*2})$, where $r^* = r_5^* - r_{30}^*$. Fig. A.6(a) shows the potential energy landscape for all the four epigenetic states namely; OFF, OFF:GT1, OFF:GT2 and SAW state. We took the cubic spline through the data points to generate a smooth free energy landscape for simulation. Starting from $x = 15$, we computed the exit time for the bead to cross various positions such as $x = 5, 3, 1.8$. This was done for various time steps and the results for the OFF state are plotted in Fig. A.6(b). We then extrapolated the simulation results to $\Delta t \to 0$ to estimate the exact values. Solid line in Fig. A.6(b) indicates fitting to the simulation data and the $y$-intercept of solid lines represents the extrapolated value for $\Delta t = 0$.

We repeated this study for various epigenetic states and simulated the bead in the corresponding free energy landscape for OFF:GT1, OFF:GT2 and SAW states. Fig. A.7(a) shows the extrapolated exit time (time at $\Delta t \to 0$). It can be observed that, starting from $x = 15$, it takes more time to reach the farther distance, from the starting position. For example, in the SAW (black curve in Fig. A.7(a)), bead reaches the position $x = 5$ quickly compared to the position $x = 1.82$, which is far from the starting point $x = 15$. In the epigenetic context, OFF state takes the smallest time to reach as compared to the other states and SAW. This is because, in the OFF state, chromatin interactions have tilted the free energy landscape (see Fig. A.6(a)) to facilitate the contact between different segments.

Finally, we compared the results obtained from simulating a single bead in the effective landscape (1D) to the simulation of the 50 bead polymer chain (3D). Fig. A.6(b) shows this comparison for the OFF state. It was observed that the first passage time in the polymer is higher compared to the results obtained in the 1D single bead case. It should be noted here that the results presented in this section are only preliminary, and a much deeper understanding and simulations are required to map the 3D polymer simulation result in simple single bead simulation.