# MODELING NUCLEOSOMAL DNA IN LIVING YEAST

Caitlin Suzanne Hult

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Mathematics in the College of Arts and Sciences.

Chapel Hill
2017

Approved by:

M. Gregory Forest

David Adalsteinsson

Kerry Bloom

Boyce Griffith

Paula Vasquez

## ABSTRACT

Caitlin Suzanne Hult: Modeling Nucleosomal DNA in Living Yeast
(Under the direction of M. Gregory Forest)

The genome in living yeast cells is a highly dynamic system where entropic interactions and nuclear confinement drive the formation of domains of high chromosomal interaction, known as topologically associating domains. We investigate dynamic organization and territory formation of 16 chromosomes in living yeast cells during interphase, using coarse-grained, entropic polymer chain models. We are interested in determining the mechanisms, such as packaging molecules that create loops within chromatin fibers, that govern inter- and intra-chromatin fluctuations and induce global features of the entire genome as well as more localized features of the nucleolus. The Bloom Lab measures specific DNA sites in specific chromosomes using live cell fluorescence microscopy. Our goal is to identify the sufficient biological and biophysical assumptions necessary to reproduce the experimental data, from which we aim to shed insights into dynamics and structure that are beyond current experimental resolution. In this dissertation, we show through a reductionist 3D mathematical model that entropic interactions are the driving mechanisms of the formation of domains of high chromosomal interaction and that enzymatic processes modulate the spatiotemporal dynamics of these domains. We present novel modeling and visualization techniques for nucleolus spatiotemporal dynamics. We explore the minimal physical requirements to segregate this region of repeated DNA from the rest of the genome, with a particular emphasis on the role of transient loop formation. Our findings support and significantly expand upon previous work.

To life, to love, to me.

## ACKNOWLEDGEMENTS

This work would not have been possible without the guidance, support, and mentorship of my advisor Greg Forest, my co-advisor David Adalsteinsson, and Paula Vasquez. I would like to thank them for introducing me to this research area and encouraging me as I learned the ropes, and for both challenging and motivating me to be the best researcher I could be. I would also like to thank the members of the Bloom Lab, particularly Kerry Bloom and Josh Lawrimore, for the valuable perspective and knowledge that they have brought to my work over the years, and for their insightful comments and perpetual reminders to keep my work grounded in the biology. And lastly, I would like to thank my family and friends for their unwavering support, encouragement, and humor over the years.

## TABLE OF CONTENTS

## LIST OF FIGURES

## LIST OF TABLES

CHAPTER 1

# Introduction

Although densely compacted, the yeast genome has a dynamic, non-random spatial organization. Acquiring a better understanding of the spatiotemporal dynamics of DNA, particularly with regard to how cells both effectively package and preserve the functionality of DNA, would lend valuable insight to many areas of cell biology, including gene expression, transcription, and DNA repair [1]. Previous work suggests that the genome “is far from being a static information warehouse" and is instead “a mechanically active entity that is constantly altering its shape", causing “on average, any given gene [to] be exposed within the population" [3]. Comparison of the organizational principles that emerge from the statistics of population studies of fixed cells with those of an individual living cell reveals the dynamic nature of DNA and further suggests that the forces on chromosomes are spatially and temporally regulated throughout the cell [4]. That cells must balance both gene accessibility and long polymer compaction raises many interesting biological questions: how do chromosomes explore space within the nucleus? to what degree do they form territories, and do those territories persist over time? do specific, predictable sections of the genome remain in close proximity to one another over time? how do enzymatic processes such as binding proteins influence chromatin movement?

In this dissertation, we look to further our understanding of chromosome dynamics and spatial organization in living yeast cells through a thorough, comparative analysis of simulated data produced from mathematical models and experimental data gathered by the Bloom Lab. The primary interests of the Bloom Lab lie in the study of the mechanisms of chromosome segregation, distribution, and organization at varying stages of the yeast cell cycle, the construction of artificial chromosomes to study functionality, and the development of microscopy techniques and simulator programs such as ChromoShake [5]. Yeast are eukaryotic cells, and thus experience four primary stages during replication, known as G1, S, G2, and Mitosis. The first three stages – G1, S, and G2 – compose approximately 90% of the cell cycle and are known as interphase, which is broadly defined as the period during which the cell grows and chromosomes are replicated in preparation for cell division.

Although the Bloom Lab has also done very interesting work focusing on tension-sensing, microtubule dynamics, and the spindle checkpoint in metaphase – and although some of my initial research focused in this area–, we confine the studies presented in this dissertation to an investigation of live yeast cells whose chromosomes are currently in the G1 stage. At this stage in the cell cycle, the 16 chromosomes have not yet replicated and are tethered to the nuclear wall at their respective centromeres and telomeres. In a study looking to track chromosome spatiotemporal dynamics, the ability to keep cells alive and thereby observe actual chromosome movement when studying genome organization is crucial. The work of Verdaasdonk et al. [1] and Vasquez and Bloom [4] stresses the importance of working with live cells and interpreting chromosome territorial organization as a reflection of spatiotemporal dynamic fluctuations of the chromatin polymer, and these studies couple "in vivo chromatin motion analysis with mathematical modeling to elucidate the physical properties that underlie the formation and fluctuations of territories" [1]. Furthermore, the importance of a close collaboration between experimental biologists and mathematical modelers cannot be overstated. From the math modeling perspective, experimental data informs and limits how we fine-tune the mathematical model; namely, which physical parameters are accepted in the field, which parameters we should consider to be variables, what timescales are realistic, etc. Comparison of simulated data with experimental data also provides a method of assessing the accuracy of our model. From the biologist perspective, our mathematical model enables biologists to ask questions that they cannot resolve experimentally, and in the best case scenarios, the model can suggest experiments to confirm unexpected model predictions.

A more specific motivation for the work presented here is the desire to better understand the mechanisms governing the formation and dynamics of topologically associating domains, or TADs. We define TADs as “fundamental structural and functional building blocks of...interphase chromosomes" [6], or more generally, as regions of the genome that exhibit enhanced proximity and self-association. Multiple studies have suggested that TADs play important functional roles in the control of gene expression and development [6, 7, 8, 9, 2], and consequently their potential biological relevance is extremely significant. A central hypothesis is that TADs act as a mechanism for gene expression within a compact environment, allowing a cell to effectively tune which genes it expresses by a non-random organizational “tuning" of the chromatin. Their importance and existence have been recognized and studied by groups using a variety of techniques, perhaps most

notably, the Hi-C method championed by the Dekker and Mirny labs [10]. Chromosome conformation capture methods, such as Hi-C, rely on deep sequencing to detect and quantify pairwise chromatin interactions, thereby providing not only information "about spatial contacts between pairs of genetic loci, but also simultaneously their genetic coordinates along the chromatin fiber" [10]. Through cross-linking with formaldehyde and DNA ligation, Hi-C experimentally measures spatial contacts in 3D, and remarkably, can do so across the entire genome, averaging chromosome conformations from millions of nuclei simultaneously. Mirny et al. define TADs in the context of Hi-C — as "contiguous regions of enriched contact frequency that appear as squares in a Hi-C map, which are relatively insulated from neighboring regions...[many] TADs have homogeneous interiors, while others have particularly enriched boundaries, or even more complex features" [6].

Although chromosome conformation capture methods are certainly valuable and powerful, they also have significant limitations. Hi-C measures interaction frequencies over a population of cells, and this raises an important question: are population-averaged patterns reflective of single cell dynamics? In other words, there must be an inherent assumption when using chromosome conformation capture methods that averaging of a population over space is the same as averaging of a single cell over time. Hi-C only provides static spatial information, representative of the organization of a population of dead cells at a specific point in time. Our work has shown that TADs in living cells may evolve over time and depend on the timescale of observation. By studying live cells over time windows, both experimentally and through modeling, we are able to learn about the dynamics of the living genome. This approach is likely more valuable if the goal is to learn about cellular processes, which by definition occur over time. Furthermore, as will be discussed in-depth later in this dissertation, population averaging can result in "lost" information, whereas analyzing single cell dynamics over time does not.

The primary focus of this dissertation is to explore the spatiotemporal dynamics and organization of DNA in living yeast cells during interphase through the development of a 3D mathematical model, large-scale model simulations, data analysis, and visualization. This work involved writing codes in C++ and Matlab, numerical analysis of stochastic differential equations (SDEs), statistical analysis of big data, extensive work in DataTank (an object-oriented programming environment offering large set data support and visualization tools), and developing ways to effectively visualize simulated and experimental chromosome dynamics. We use coarse-grained models to capture the

dynamics of the chromatin fiber as an effective entropic bead-spring chain. The model we present here is a reductionist, predictive model that is fully reflective of biological parameters and has successfully recapitulated experimental results and at times served as a predictive tool for system behavior. In this dissertation, we introduce the mathematical model representation of chromatin force by force, beginning with the thermodynamic force created by Brownian motion, and adding in spring forces, excluded volume, geometric confinement, tethering, and basic enzymatic processes. We present a comprehensive tutorial detailing the use of coarse-grained bead-spring chain models to describe chromatin fiber movement during interphase in yeast cells. We demonstrate ways to not only quantify but to effectively visualize chromosome 3D dynamics and organization. We aim to improve our understanding of the forces driving TADs formation. Much of my work focuses on development of modeling approaches and visualization techniques for nucleolus dynamics, as well as quantification of nucleolus shape and movement, especially with regard to its level of interaction with other sections of the yeast genome. We present novel results from our published paper [2] and submitted paper [11], as well as additional results and analysis not included in those works. Future work is multifaceted and ongoing. From a general standpoint, we plan to continue adding complexity to the 3D model, with a particular focus on how we simulate binding proteins, phase separations pertaining to the nucleolus, hydrodynamic interactions, and other nuclear bodies. In addition to this very interesting, more experimentally-guided and biological work, we are also investigating nucleolus dynamics through the lens of networks, as well as through the lens of microrheology and the local viscoelastic environment. Through this dissertation, we hope to improve our understanding of how chromosomes explore space in the nucleus and to develop a concrete foundation and reference point from which future, more complex models can expand.

CHAPTER 2

# The Rouse Bead-Spring Model of Chromatin

[1]

## 2.1 Modeling long-chain polymers

Due to the inherent complexity of biological systems, the main challenge of the applied mathematician often lies in developing models that both effectively manage and simplify this complexity, and yet remain realistic enough to allow one to deduce meaningful trends and patterns. Ideally, such a model should distinguish among, quantify, and accurately reflect the differing contributions of the many variables present in the system, as well as sources of variable uncertainty and potential error. A mathematical model aims to emulate but will never entirely capture the full complexity and parameter space of a DNA polymer, and thus great importance lies in how we choose to simplify the polymer system to a computationally manageable state. Long-chain polymers such as DNA present interesting modeling challenges, not only in the number of forces contributing to their movement and in the way we define those forces, but in the "wide range of length and time scales involved in their dynamics" [4]. In this thesis, we rely on past models to validate the starting point of our polymer model. A recent example includes that described in Verdaasdonk et al. [1], in which they "found that coarse graining of the dynamics of each chromosome arm into a double tethered bead-spring chain, with the addition of excluded volume interactions and confinement by the cell wall, accurately recapitulates the thermal, and random ATP-dependent motion of experimentally captured chromosome spots," although similar bead-spring models have been used for many years. Further discussion of and references supporting the use of polymer bead-spring chain models in describing the dynamics of interphase chromosomes can be found in later chapters. In this dissertation, we use an effective entropic bead-spring chain model to simulate the movement of 16 double-tethered

[1] Sections of this chapter were adapted from Caitlin Hult's Master's project from May 2015, entitled, "Modeling of living yeast chromosomes".

chromosomes during interphase. Note that two chromosome arms compose each chromosome; thus, a model that simulates the movement of 16 chromosomes contains 32 bead-spring chains. By representing each chromosome arm as a bead-spring chain composed of $N$ beads and $N-1$ springs, we can track DNA spot fluctuations over time and thereby simulate the movement of both untethered and double-tethered chromosomes during interphase. When doubly-tethered to the nuclear wall, the two end beads on each chain are fixed and experience no forces, while the remaining $N-2$ beads move in response to four primary forces, including Brownian motion, spring forces, excluded volume, and nuclear wall interactions. When the chain is untethered, all beads experience all forces.

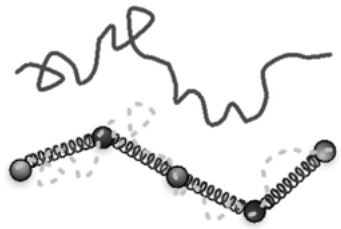

Figure 2.1: *The chromatin polymer (solid line) can be modeled as a bead-spring polymer chain (dotted line indicates original chromatin polymer chain). Diagram of a bead-spring chain composed of 100 beads tethered at both ends and confined within a circle of radius 1um. Figure taken from [1].*

### 2.1.1 Discretization

As we study mechanics and gene expression, we are modeling up quite a few levels from protein structure; namely, we model and fluoresce on the order of several thousand base pairs, whereas models at the protein structure level are on the order of 50 base pairs. As described in Chapter 4, we consider chromosomes to be De Gennes blobs linked by nonlinear springs, where the size of the blobs represents the spatial extent over which the motion of the chromatin fiber is uncorrelated [2]. From the biologist perspective, this translates to the DNA living in a given bead (blob) with some frequency, and making excursions out of this blob with some lesser frequency.

The number of beads composing each of the 32 chromosome arms in our model directly correlates with their respective experimentally-determined lengths, given in number of kilobase pairs. We chose a discretization level of 5kb because genes are typically several thousand base pairs and because the size of the spot that the biologists fluoresce is 5kb. This level of discretization produces a system

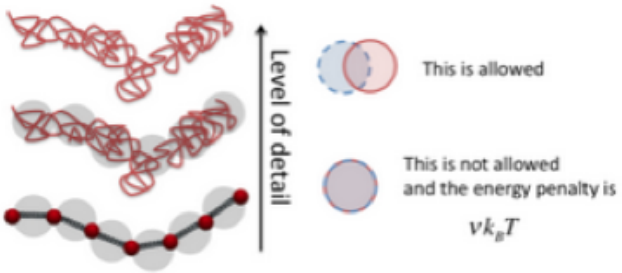


Figure 2.2: *Cartoon depicting discretization and energy penalty that prevents bead overlap.*

of 2443, 2803, or 2804 total beads (depending on whether the system does not contain a nucleolus, contains a single nucleolus, or contains a split nucleolus, respectively), which is manageable from a computational standpoint. See Chapter 4 for further discussion.

### 2.1.2 Comparison of 2D and 3D models

The initial model was written in Matlab, first in 2D and later in 3D. The 2D Matlab model simulated four double-tethered chromosome arms, each 52 beads in length, whereas the 3D Matlab model simulated 32 double-tethered chromosome arms, each 52 beads in length. The current model is written in C++ and DataTank, is 3D, and is fully reflective of experimental parameters. Therefore, as noted in the previous section, the number of beads composing each of the 32 arms in the system varies and reflects the experimental lengths of the 32 arms. Unless otherwise noted, all work in this dissertation corresponds to the 3D C++/DataTank model. The 2D model was integral to Vasquez et al. [2], and thus will be discussed further in Chapter 4. Additionally, we note that the 3D model discussed in [2] corresponds to the C++/DataTank model version. The 2D model remains useful [2]; however, as the 3D model provides a much more realistic representation of the biology, enables quantitative comparison with experimental data, has more options for model variability, and is much more computationally powerful, we restrict current and future work to the 3D model. Although the 3D C++ code can be run locally on one's computer for short times, I always submit runs to KillDevil, often through the application DataTask, which enables easy submission of parameter sweeps or groups of runs.

Regarding my role in model development, I joined this project when the model was a 2D code, written in Matlab by Paula Vasquez. I helped adapt this code to allow for simulation of damaged DNA and different boundary conditions. David Adalsteinsson ported the code to C++ and

DataTank, in the process adapting it to 3D and to a more complicated model that is fully reflective of experimental parameter values. I worked closely with David on further modifications of the code; namely, how to incorporate dynamic looping into the code, the option to simulate a single or split nucleolus, and particularly on development of the DataTank template document associated with the code.

## 2.2 The mathematical model

As stated in [4], we require the balance of forces in the zero-mass limit for bead $i$ to satisfy,

$$\mathbf{F}^{Thermal} + \mathbf{F}^{Spring} + \mathbf{F}^{Drag} + \mathbf{F}^{ExcludedVolume} + \mathbf{F}^{WallInteractions} = 0. \tag{2.1}$$

Through the use of linear spring and drag laws, as well as an excluded volume potential described by a Gaussian distribution, we can express the evolution equation for the position of bead $i$ as follows:

$$\zeta \frac{d\mathbf{X}_i}{dt} = \mathbf{F}_i^B(t) + \mathbf{F}_i^S(t) + \mathbf{F}_i^{EV}(t) + \mathbf{F}_i^W(t), \tag{2.2}$$

where $\zeta$ represents the bead drag coefficient, $\mathbf{F}_i^B(t) = \sqrt{2k_BT\zeta}d\mathbf{W}_i$ represents the Brownian motion term, $\mathbf{F}_i^S(t) = -k_s(\mathbf{X}_{i-1} - 2\mathbf{X}_i + \mathbf{X}_{i+1})dt$ represents the intra-molecular forces term captured by a Hookean spring, $\mathbf{F}_i^{EV}(t) = \frac{zk_s}{2d^5}\left[\sum_{j=1,j\neq i}^{N_b}(\mathbf{X}_i - \mathbf{X}_j)\exp\left(-\frac{k_s}{k_BT}\frac{(\mathbf{X}_i-\mathbf{X}_j)^2}{2d^2}\right)\right]dt$ represents excluded volume interactions among the beads, and $\mathbf{F}_i^W(t)$ represents the influence of the nuclear wall on bead movement. Note that since $\mathbf{X}_i$ and $\mathbf{X}_j$ are vectors, the expression $(\mathbf{X}_i - \mathbf{X}_j)^2$ denotes the scalar dot product. A more in-depth description of these forces and a list of parameter definitions follow in Sections 2.3 and 2.4. Verdaasdonk et al [1] cites Equation 2.2 as that given by Doi and Edwards in "The Theory of Polymer Dynamics" [12]. Written out in its entirety, the full dimensional evolution equation is that given by [4] as:

$$\zeta d\mathbf{X}_i = -k_s(\mathbf{X}_{i-1} - 2\mathbf{X}_i + \mathbf{X}_{i+1})dt + \sqrt{2k_BT\zeta}d\mathbf{W}_i + \frac{zk_s}{2d^5}\left[\sum_{j=1,j\neq i}^{N_b}(\mathbf{X}_i - \mathbf{X}_j)\exp\left(-\frac{k_s}{k_BT}\frac{(\mathbf{X}_i - \mathbf{X}_j)^2}{2d^2}\right)\right]dt. \tag{2.3}$$

## 2.3 Description of primary component forces

### 2.3.1 Brownian Motion

$$\mathbf{F}_i^B(t) = \sqrt{2k_BT\zeta}d\mathbf{W}_i \tag{2.4}$$

Brownian motion is defined as the random motion of particles suspended in a fluid resulting from their collision with the quick atoms or molecules in the fluid. In yeast cells, the chromatin fiber is constantly moving around due to both thermal and non-thermal or ATP dependent fluctuations. We capture this random forcing through the use of a Brownian motion force term in our model, and it is this randomness that defines our evolution equation as a stochastic differential equation. Larson [13] notes that over a time scale of perhaps microseconds, "there are so many random [thermal bombardments of the polymer by the solvent] that the random force imparted by them averages toward zero, except for a small drift force whose magnitude gets smaller with the inverse square root of the averaging time." $\mathbf{W}_i$ is a Wiener process, ie., the stochastic process from which fluctuations of the variable $\mathbf{X}$ originate. $\mathbf{W}_i$ is more fully discussed in the section on normal mode calculations. We use a random number generator in our code to create the vector of random numbers in $\mathbf{W}_i$. Regarding the drag parameter $\zeta$, [13] notes that a "relationship must exist between Brownian motion and drag because the rate of Brownian motion determines the diffusion coefficient, which is also reflected in the magnitude of the drag coefficient."

### 2.3.2 Spring Forces

$$\mathbf{F}_i^S(t) = -k_s(\mathbf{X}_{i-1} - 2\mathbf{X}_i + \mathbf{X}_{i+1})dt \tag{2.5}$$

In early versions of this model, we assumed that all springs between beads were Hookean and approximately one Kuhn length long. We could thereby describe the spring force relationship between neighboring beads on a chain through a connectivity matrix (discussed further in normal modes section). We can visualize the 2D spring force on bead i as follows:

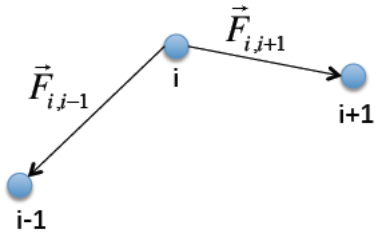


Figure 2.3: *Spring forces on bead i. Figure created by Paula Vasquez.*

In the current version of the model, we use the wormlike chain (WLC) force law proposed by Marko and Siggia [14, 15]. This nonlinear, attractive potential prevents springs from stretching indefinitely and captures intra-chain interactions in a much more biologically realistic way, and it has been shown to describe the dynamics of naked DNA and chromatin *in vivo* well. We define the WLC force as follows:

$$\mathbf{F}_i^S = \mathbf{F}_{i,j-1} + \mathbf{F}_{i,j+1}, \tag{2.6}$$

where

$$\mathbf{F}_{i,j} = \frac{k_B T}{4L_p}\left[\left(1 - \frac{R_{ij}}{R_0}\right)^{-2} - 1 + 4\frac{R_{ij}}{R_0}\right]\frac{\mathbf{R}_{ij}}{R_{ij}} \tag{2.7}$$

and

$$\mathbf{R}_{ij} = \mathbf{X}_i - \mathbf{X}_j, \qquad R_{ij} = \sqrt{\mathbf{R}_{ij}\cdot\mathbf{R}_{ij}}, \qquad R_0 = N_{k,s}\,(2L_p) \tag{2.8}$$

### 2.3.3 Excluded Volume

$$\mathbf{F}_i^{EV}(t) = \frac{z k_s}{2d^5}\left[\sum_{j=1, j\neq i}^{N_b}(\mathbf{X}_i - \mathbf{X}_j)\exp\left(-\frac{k_s}{k_B T}\frac{(\mathbf{X}_i - \mathbf{X}_j)^2}{2d^2}\right)\right]dt \tag{2.9}$$

Although conceptually straightforward, the excluded volume forces are perhaps the most difficult to explicitly define. This is in part because there is no analytical solution to compare model output with, in part because excluded volume potentials often lump together many different repulsive forces under one heading. Various excluded volume potentials have been proposed in the literature — some are designed to rigorously prevent chain-crossing, some impose spring-spring repulsions instead of bead-bead repulsions, some prioritize the ability to vary with molecular discretization, etc. We conducted a thorough literature review of ways to define excluded volume potential, and Eqn 2.9 is

that presented in [13]. At its core, the physical basis for excluded volume is a repulsive force, a way to ensure that different beads do not occupy the same space at the same time. Larson describes excluded volume interactions as "the repulsive forces between monomers that prevent their overlap, leading to a tendency of the chains to expand beyond the ideal random-walk conformations they have in the melt" [13]. $z$ and $d$ are dimensionless parameters, where $d$ describes the range of repulsion (or width of the Gaussian) and $z$ describes the strength of repulsion (or height of the Gaussian). Note that in [2] and [11], we use the excluded volume potential proposed by Jenderjack et al. [16]. Larson shows that these two potentials are identical if certain condition for $z$ and $d$ are met [13].

**Parameter sensitivity study.** Since we do not have an analytical solution or biological correlate for excluded volume, we performed a parameter sensitivity study in order to "tune" the excluded volume parameters. I ran a 121-run parameter sweep of possible $z$ and $d$ values using the application DataTask, where the scaling factors for $z$ ranged from 0.5 to 1.5 and for $d$ from 0.5 to 1. I generated 16001 total position values for each of the 2803 beads in the system, for which positions were saved every 0.1 seconds, for ∼26.6 minutes total. I converted each of these position values to 3-bead blob center of mass position values, e.g. $x_{i(centerofmass)} = (x_{i-1(original)} + x_{i(original)} + x_{i+1(original)})/3$. There are experimental results suggesting that when the 240kb spot (which corresponds to chain 4, bead 49 in our model) falls within a 400nm z-band, the percentages of time that it is found in zones 1, 2, or 3 are approximately 5%, 25%, and 70%, respectively (Figure 2.4). Thus, we are interested in determining how often our model's correlate for the 240kb spot is found in a given section of the nucleus. I wrote a code in C++/DataTank that computes these percentages for all of the runs in the parameter sweep. In order to determine the location of the 400nm z-band that would encompass the most possible position values, I found the mean and standard deviation of the bead's center of mass z-position values, for each run. Centering the 400nm wide z-strip at the mean saves between 9000 and 11000 values (out of the 16001), so we chose to use the z-band $(mean - 200) \leq z \leq (mean + 200)$. For each z-value in each run:

1. Check is z is within the z-band. If yes, go to step 2. If no, discard point.

2. Determine which zone the corresponding 2D point is in by converting the point to polar coordinates and observing whether $0 \leq r \leq \frac{1000}{3}$, $\frac{1000}{3} \leq r \leq \frac{2000}{3}$, $\frac{2000}{3} \leq r \leq 1000$.

3. Add one to the count for the appropriate zone.

Based on Figure 2.5, we concluded that the scaling factors should be z=0.9 and d=0.9. This is because: 1) it gives us 70% in zone 3, and this is the experimental measurement that the biologists trust the most, since it is easier to verify if a spot is close to or far away from the wall than if they are in zone 1 or zone 2, and 2) we do not have biological justification for choosing a combination of scaling factors in which one is bigger than the other. Note that all results in chapters 4 and 6 use position values that were generated using these scaling factors.

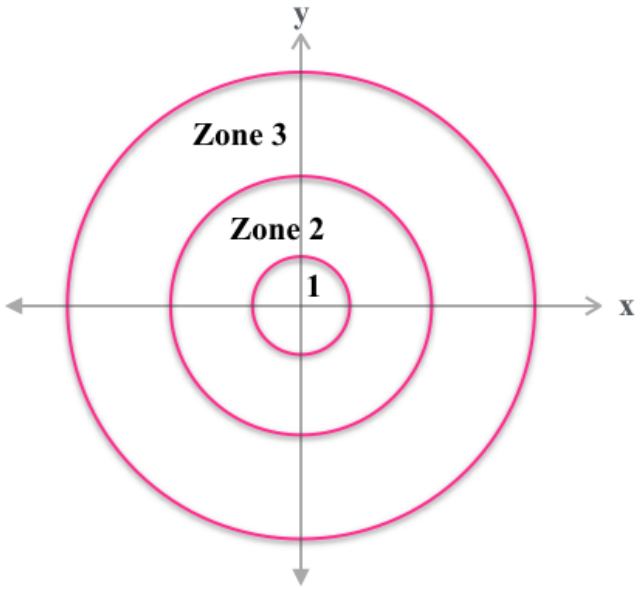


Figure 2.4: *Picture depicting three zones used in excluded volume sensitivity study.* Zone 1 radius: [0,1000/3], Zone 2 radius: [1000/3, 2000/3], Zone 3 radius: [2000/3, 1000], where all values are in nm.

| All have z-band centered at the (rounded) mean of z-positions | Zone 1 | % | Zone 2 | % | Zone 3 | % | Total # Positions Considered | Mean of z-positions | StDev of z-positions |
|---|---|---|---|---|---|---|---|---|---|
| **Range = 1, Strength = 0.9** | 0 | 0.0 | 1845 | 20.2 | 7284 | 79.8 | 9129 | 267.78 | 244.59 |
| **Range = 0.95, Strength = 0.9** | 0 | 0.0 | 2898 | 30.6 | 6585 | 69.4 | 9483 | 261.41 | 241.93 |
| **Range = 0.9, Strength = 0.9** | 0 | 0.0 | 2712 | 30.0 | 6342 | 70.0 | 9054 | 259.65 | 247.2 |
| **Range = 0.85, Strength = 0.9** | 0 | 0.0 | 3311 | 35.4 | 6033 | 64.6 | 9344 | 259.03 | 236.23 |
| **Range = 0.8, Strength = 0.9** | 67 | 0.7 | 4004 | 40.1 | 5922 | 59.3 | 9993 | 299.39 | 230.47 |
| **Range = 0.75, Strength = 0.9** | 83 | 0.8 | 4507 | 43.8 | 5698 | 55.4 | 10288 | 321.075 | 226.36 |
| **Range = 0.7, Strength = 0.9** | 114 | 1.1 | 5245 | 48.8 | 5381 | 50.1 | 10740 | 324.23 | 220.74 |
| **Range = 0.65, Strength = 0.9** | 114 | 1.1 | 5536 | 53.7 | 4662 | 45.2 | 10312 | 332.57 | 227.5 |
| **Range = 0.6, Strength = 0.9** | 128 | 1.3 | 5914 | 58.0 | 4146 | 40.7 | 10188 | 325.53 | 228.31 |

Figure 2.5: *Selected values from parameter sweep. We chose z=0.9 and d=0.9 as scaling factors.*

### 2.3.4 Geometric Confinement

$$\mathbf{F}_i^W(t) \tag{2.10}$$

During interphase, yeast chromatin fibers remain inside the nucleus, trapped by the nuclear wall. In our mathematical model, we therefore need to account for this geometric confinement and ensure that none of the beads in the bead-spring chains “step outside" the nuclear wall as they move in space over time. We can incorporate this geometric confinement into our code under the guise of “nuclear wall interactions", in which we specify boundary conditions to describe beads’ interactions with the nuclear wall. In this thesis, we define our boundary as a circle of radius 1000nm that is fixed in place with a hard wall. After every time step, we calculate the new bead positions and perform a check to see if any of them have stepped outside of our defined boundary. If the beads lie inside the

boundary, we record their positions. If the beads lie outside the boundary — in other words, outside the domain —, the hard wall confinement condition relocates the errant bead back to the boundary, taking the shortest path possible. Although not discussed in this thesis, several other definitions of wall confinement forces have been or are currently being considered in the bead-spring model. In the biology, the nuclear wall is dynamic and the nucleus itself is able to move around within the cell. The choice of a hard wall confinement, although effective in imposing a confining force on the beads, is therefore not the most realistic representation of the nuclear wall. As described in Vasquez and Bloom [4], an alternative boundary condition is a "soft-wall" confinement, in which we assume that beads can leave the domain with a given probability governed by a normal distribution (Gaussian) with mean in the boundary and a given standard deviation. Future work may allow the boundary to fluctuate as a sort of "Fourier" wall (in which we allow the wall to pulse according to an imposed Fourier spectrum), to rotate, and to slowly move through the surrounding cytoplasm.

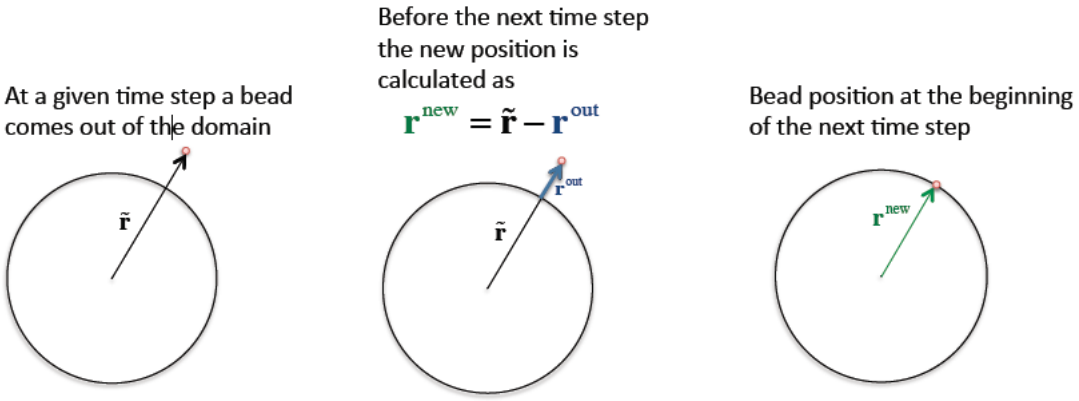


Figure 2.6: *Cartoon depiction of hard wall boundary condition. For the soft wall boundary condition, the new bead position is calculated as* $r^{new} = \tilde{r} - r^{out} + \sigma \mathbf{W}_1$*, ie., normally distributed noise with a given standard deviation. Figure created by Paula Vasquez.*

### 2.3.5 Hydrodynamic Interactions

$$\mathbf{F}_i^{HI}(t)$$

In this model, we do not consider hydrodynamic interactions (HI). Defining HI as "the disturbance to the flow field produced by one part of the chain that influences the drag on another part," Larson [13] notes that although "HI is always important for long, flexible, polymer molecules," its effect

can sometimes be “subsumed into the effective drag properties of the chain." Multiple groups have compared results from models that do not consider hydrodynamic interactions, called free-draining models (FD), with those that consider hydrodynamic interactions. Although HI effects are certainly not inconsequential, the necessity of incorporating an HI force into a bead-spring model varies. Groups such as Jendrejack et al. [16], Schroeder et al. [17], and Izmitli et al. [18] all suggest that the magnitude of HI effects depends greatly on factors such as translocation time, chain length, molecular weight, and whether the chain is a coiled state. In particular, Jenderjack et al. [16] showed that it is possible for both models (HI and FD) to predict the configurational properties of 21 $\mu$m DNA in shear and planar extension. (Vasquez report, 2015). Thus, at this stage in the development of our model, we choose to omit explicit consideration of hydrodynamic interactions.

It is worth mentioning that we have done some preliminary work towards incorporating hydrodynamic interactions into the existing free-draining model. I worked with Wenhua Guan, an August 2016 applied mathematics PhD graduate of The University of North Carolina at Chapel Hill (UNC), in the summer of 2016 to investigate the extent to which HIs affects simulated chromosome dynamics. Wenhua was the primary driver of the new code, as her dissertation work focused on using the Fast Rotne-Prager-Yamakawa Solver with Laplace FMM and developing fast algorithms for the hydrodynamic interactions of rigid bodies in suspension. The Rotne-Prager-Yamakama solver is a common method for modeling colloidal suspensions and polymer solutions. As expected, adding in the hydrodynamics greatly increased the required run time, and thus we have temporarily shelved this project.

## 2.4 Values chosen for parameters

This section summarizes some of the input values we use in generating the position data. Figure 2.7 defines some of the parameters that remain unchanged from simulation to simulation. See my DataTank document “Tethered Setup" for the full, much more comprehensive list and description of parameters. In particular, the options for dynamic looping are extensive and require careful attention when submitting a new run.

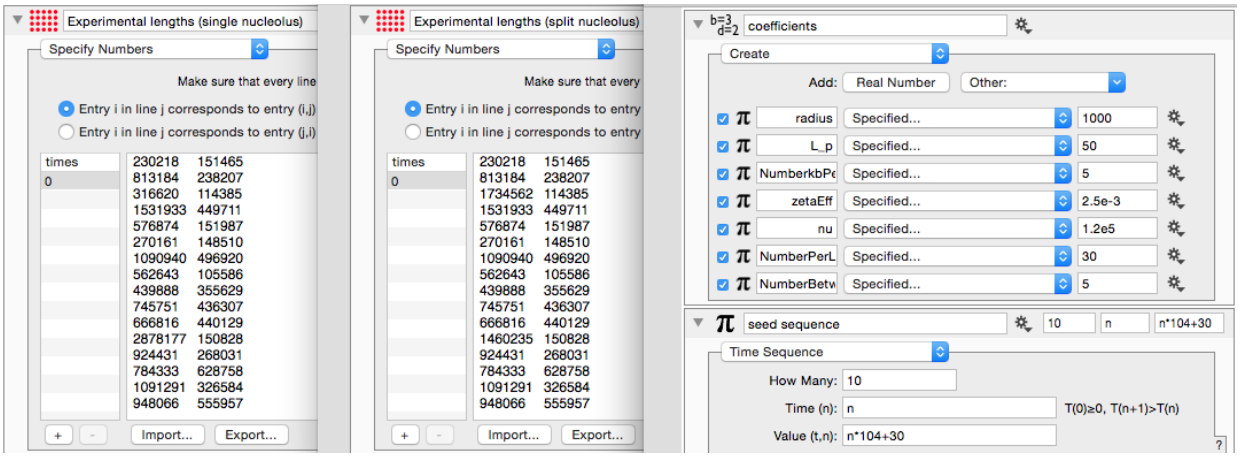


Figure 2.7: *Parameter assignments that remain unchanged for all simulations.*

## 2.5 Equilibration and Initial Conditions

The location of the centromere, the locations of the six telomere sites, and the particular chromosome arms attached to each telomere remained the same for all simulations. The initial configuration of the beads composing the 32 chromosome arms was also the same for all simulations, as all interior beads in a given arm (i.e. all beads except the two tethered beads on each end of the arm) were equidistantly spaced in a straight line at the start of the simulation (Figure 2.8). Experimental results show that each of the 32 chromosome arms in yeast cells are tethered at the same centromere location on one end and at one of typically 5-7 telomeres on the other, and that the 32 chromosome arms are fairly evenly distributed among those telomere sites. Thus, we chose six telomere locations, and assigned 4-6 arms to each of those locations. The location of the six telomere sites were chosen arbitrarily, since telomere sites can vary from cell to cell and the biologists do not yet have a firm understanding of the importance (if any) of this variation (see Figure 2.11 for their polar coordinates). It is noteworthy that in the split nucleolus case the total numbers of beads on arms 6 and 24 differ from those in the single nucleolus case, see Chapter 6 for more details.

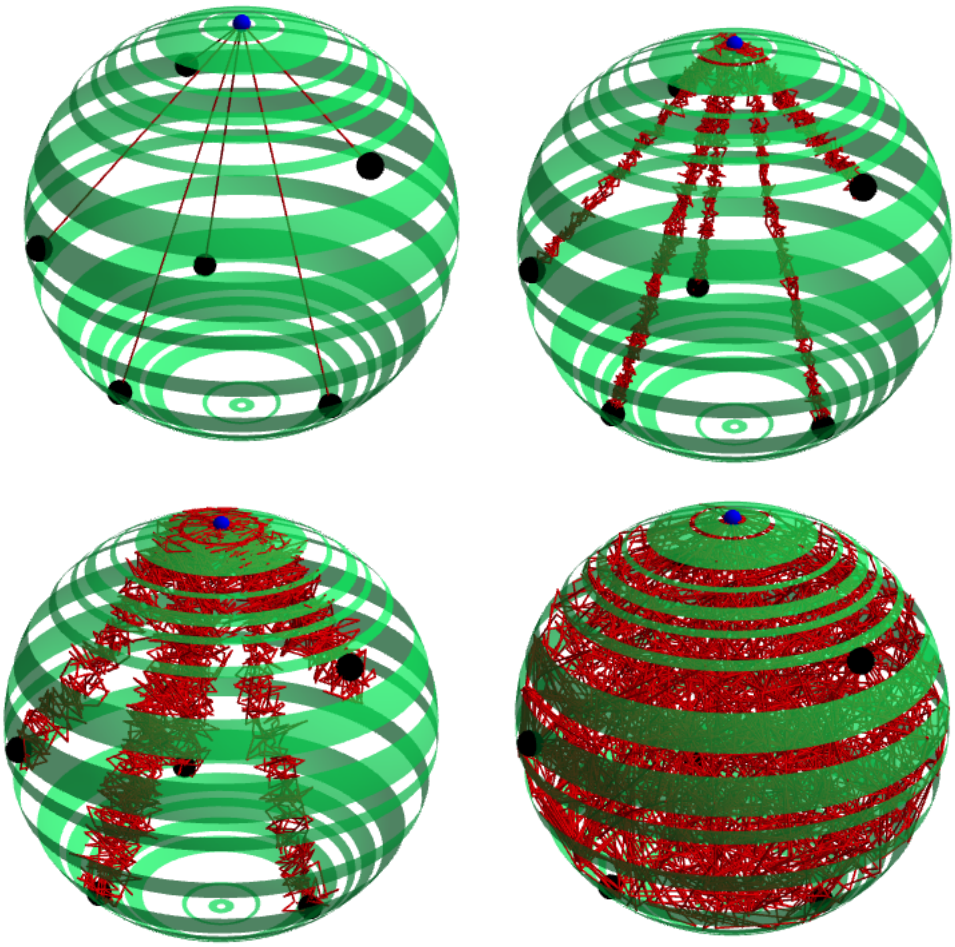

Figure 2.8: *Equilibration of the 32 arms. Black sphere=telomere, blue sphere=centromere, red=bead-spring chains. Telomeres and centromeres have been enlarged for visualization purposes.* Top left: initial configuration. Top right: 0.1 seconds after start of simulation. Bottom left: 0.5 seconds after start of simulation. Bottom right: 3700 seconds after start of simulation.

We need to allow our bead-spring chains to equilibrate before we begin recording bead positions. We considered the “total length of all chromosomes" parameter in determining equilibration time (Figure 2.12). The amount of required equilibration varied depending on whether dynamic looping was imposed. For simulations without dynamic looping, i.e. the no loops or uniform looping cases, equilibration was reached after approximately 100 seconds in real time. The position data obtained from these simulations corresponds to 15 total minutes worth, from 400 seconds after the start of the simulation to 1300 seconds after the start of the simulation. For simulations with dynamic looping, we found that a longer equilibration was necessary in order for the portions of the genome experiencing dynamic looping to be able to thoroughly “mix", as exemplified by contact maps of

**Telomere #1**
- 3D Distance b/t T3 and centromere = 1084.88 nm
- Connected chains: 3, 14, 15, 19, 22, 26

**Telomere #2**
- 3D Distance b/t T6 and centromere = 1333.72 nm
- Connected chains: 11, 17, 23, 28, 30

**Telomere #3**
- 3D Distance b/t T2 and centromere = 1944.04 nm
- Connected chains: 2, 16, 21, 25, 29, 32

**Telomere #4**
- 3D Distance b/t T1 and centromere = 898.64 nm
- Connected chains: 1, 4, 8, 9, 31

**Telomere #5**
- 3D Distance b/t T4 and centromere = 1893.30 nm
- Connected chains: 5, 6, 12, 18, 24, 27

**Telomere #6**
- 3D Distance b/t T5 and centromere = 1811.88 nm
- Connected chains: 7, 10, 13, 20

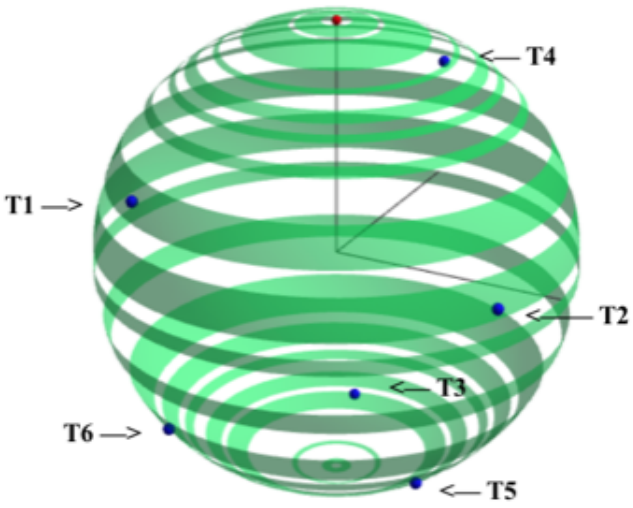


Figure 2.9: *Visualization of the 6 telomere sites and chromosome arm assignments to telomere locations.* The axis is xyz-positive. Note that the telomeres are indexed differently in Figure 4.7. Note as well that with the exception of the data generated for [2] (Chapter 4), we choose to assign chromosome arm 6 to the same telomere as chromosome 24.

| PARAMETERS | | DIMENSIONAL VALUES |
|---|---|---|
| **Time step, dt** | | 1e-3 seconds |
| **How often time steps are saved** | | Every 100 time steps |
| **Frequency of recorded position values** | | Every 0.1 seconds |
| **Time allowed for equilibration** | **No loops/uniform loops** | 100 seconds |
| | **Dynamic loops** | 2800 seconds |
| **Total # of position values generated in simulation** | **No loops/uniform loops** | 13000 |
| | **Dynamic loops** | 37000 |
| **Time range from which we draw position values** | **No loops/uniform loops** | 400 to 1300 seconds |
| | **Dynamic loops** | 2800 to 3700 seconds |
| **Total # of position values to work with (after equilibration)** | **No loops/uniform loops** | 9000 |
| | **Dynamic loops** | 9000 |

Figure 2.10: *Table of parameter values for 3D model. Times given are "real time", not simulation time.*

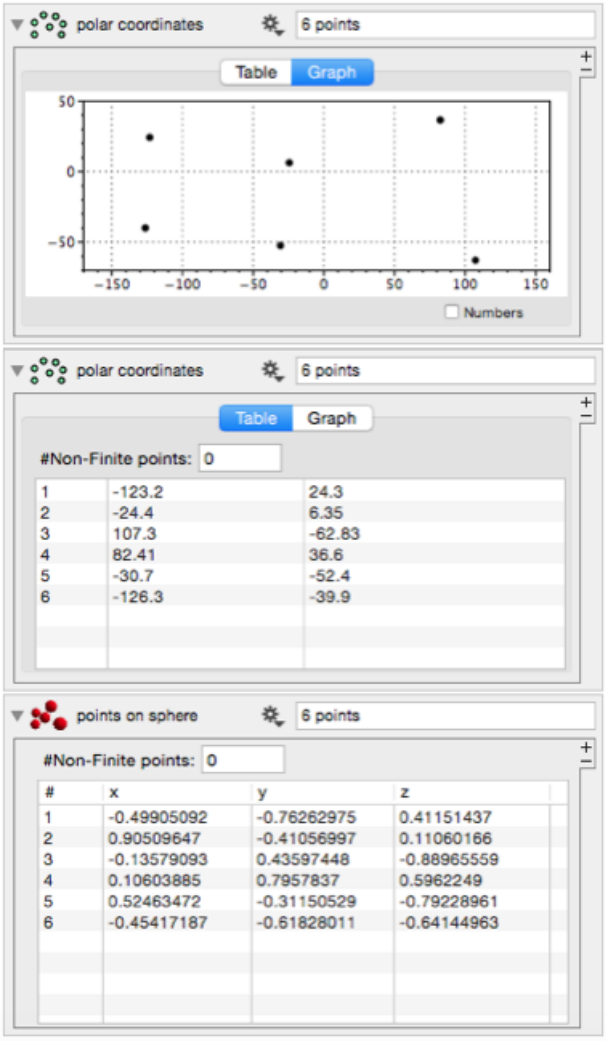


Figure 2.11: *Polar and cartesian coordinates for the six telomere locations.*

bead-bead distance (Figure 2.13). Thus, all simulations considering dynamic looping correspond to the 15 minutes that correspond to between 2800 to 3700 seconds after the start of the simulation.

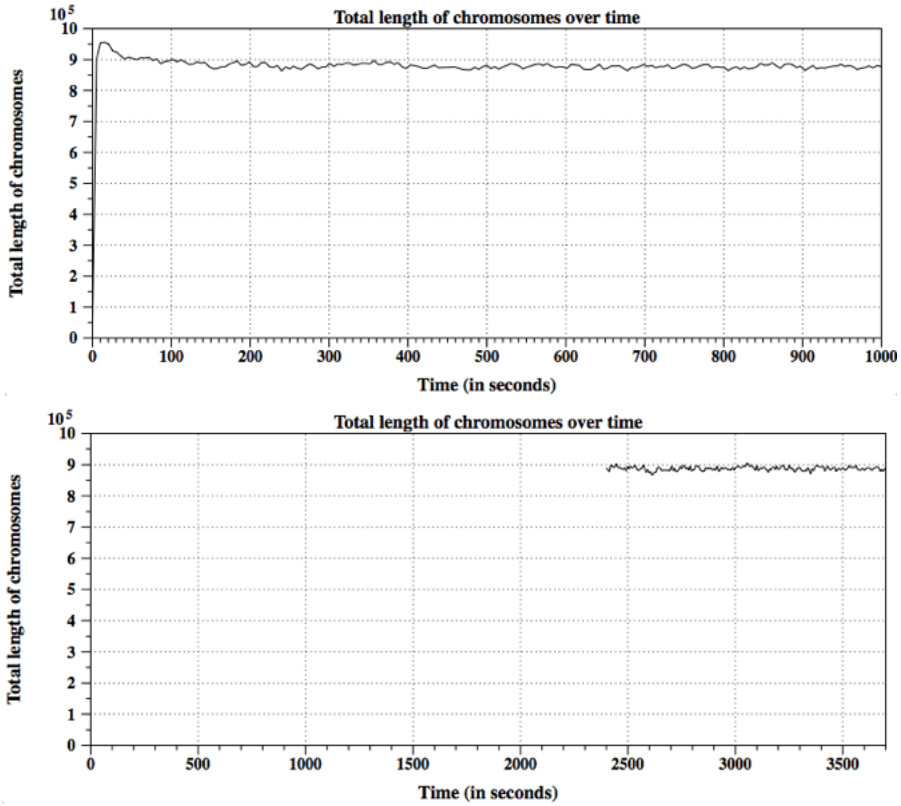


Figure 2.12: *Using total length of all chromosomes as a metric for determining equilibration time. Data shown corresponds to a simulation in which dynamic looping has been imposed.*

As discussed earlier, one challenge of modeling long chain polymers such as chromosomes is effectively managing the considerable range of length and time scales present in the system. As modelers, we must always be cognizant of what time scales are biologically realistic, and attempt to ensure that our models reflect those realities as well as possible. Although a direct correlation between biological and model time scales is not always possible, knowledge of the biology can help validate and guide the formation of a model.

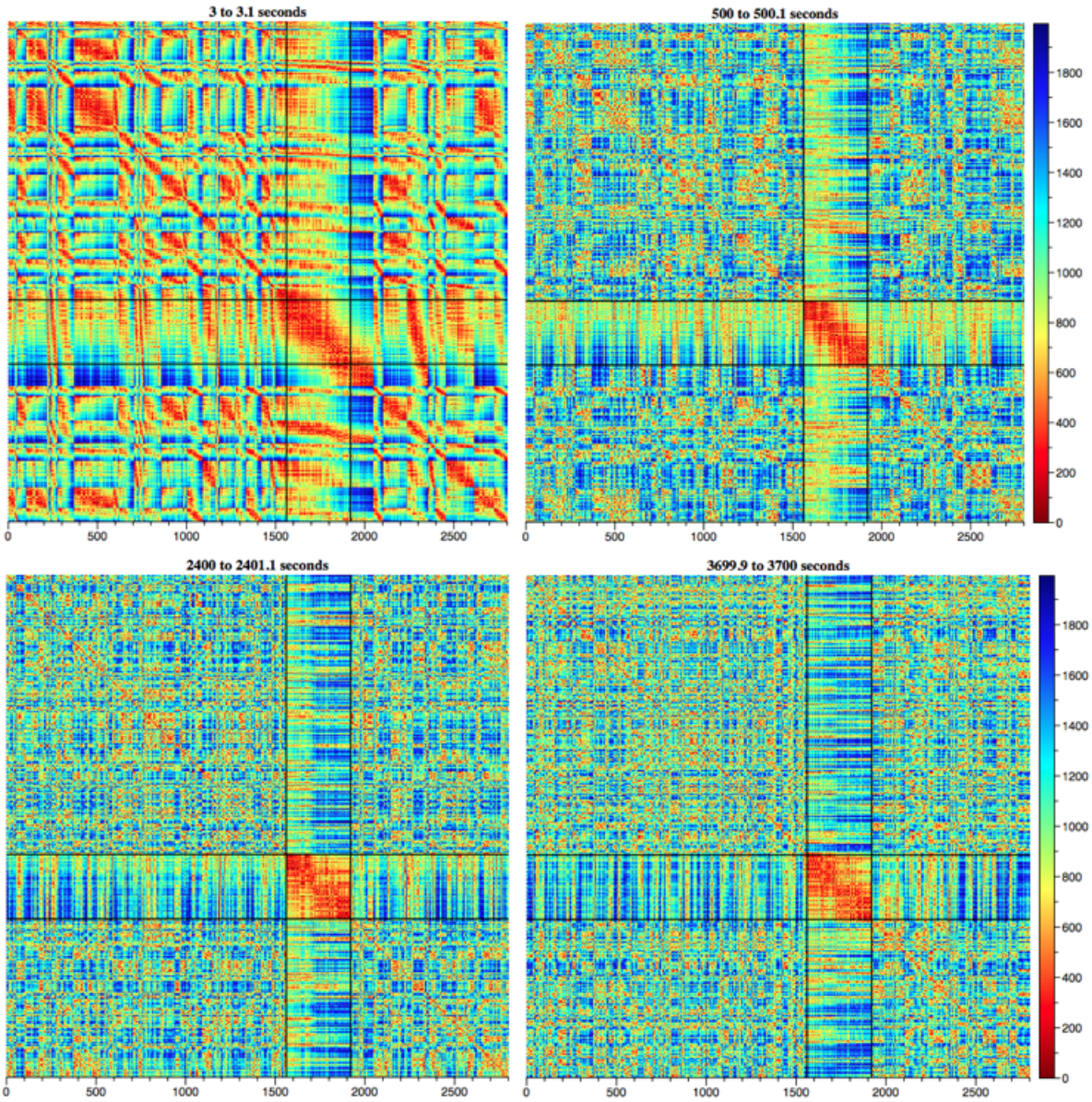


Figure 2.13: *Contact maps depicting bead-bead distances at varying points in time after simulation start.*

## 2.6 Analytical solution of Rouse chain and statistics

Often considered the cornerstone of polymer dynamics, the Rouse model serves as the foundation for many polymer models, including our own, and has a well-known analytic solution. The existence of an analytic solution is a useful tool for modelers, as it allows us to directly compare the statistics of simulated data produced by computational models to the statistics required by the analytic solution, and thereby to better test the accuracy of our model. The starting point for a Rouse model is a

Gaussian chain, often represented by a mechanical model of beads connected by harmonic springs. The beads in such a “Rouse chain" move only in response to thermal forces (Brownian motion) and spring forces (such as those described by our connectivity matrix in sections 2.3 and 2.7). Thus, if our chain is untethered and allowed to freely diffuse through a medium over time, we expect the mean displacement of any given bead to be zero and the variance of any given bead to reflect the viscosity of the medium, ie., we would expect a step-size distribution histogram for bead i to be a Gaussian distribution centered at zero.

### 2.6.1 Sample Rouse chain analyses

The following figures show several basic types of Rouse chain analyses and highlight differences we would expect to see between untethered and double-tethered Rouse chains. In all figures (Figures 2.14, 2.15, 2.16, 2.17, 2.18, 2.19), the step-size $dt = 1e-3$, we save every 100 time steps, the total runtime is 5000 seconds, and the total number of position values saved is 50000.

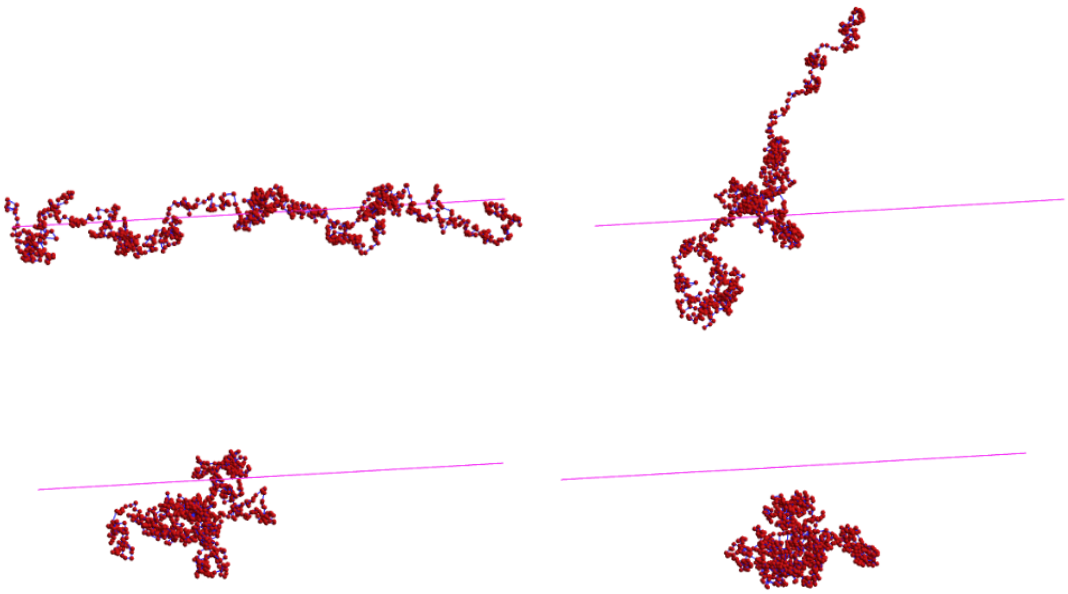

Figure 2.14: *Untethered Rouse chain at various points in time after simulation start. Time increases as we move from top left, to top right, to bottom left, to bottom right. dt=1e-3.*

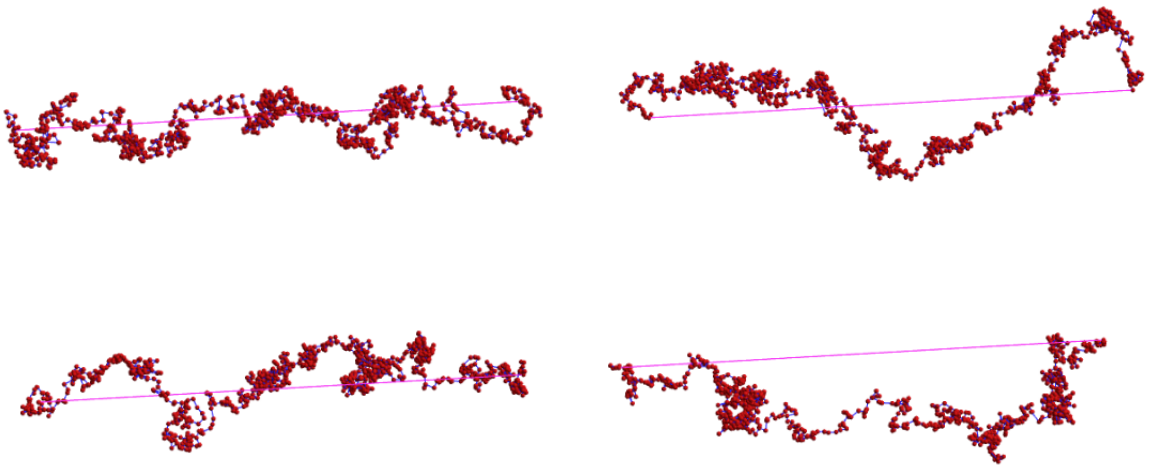

Figure 2.15: *Double-tethered Rouse chain at various points in time after simulation start. Time increases as we move from top left, to top right, to bottom left, to bottom right. dt=1e-3.*

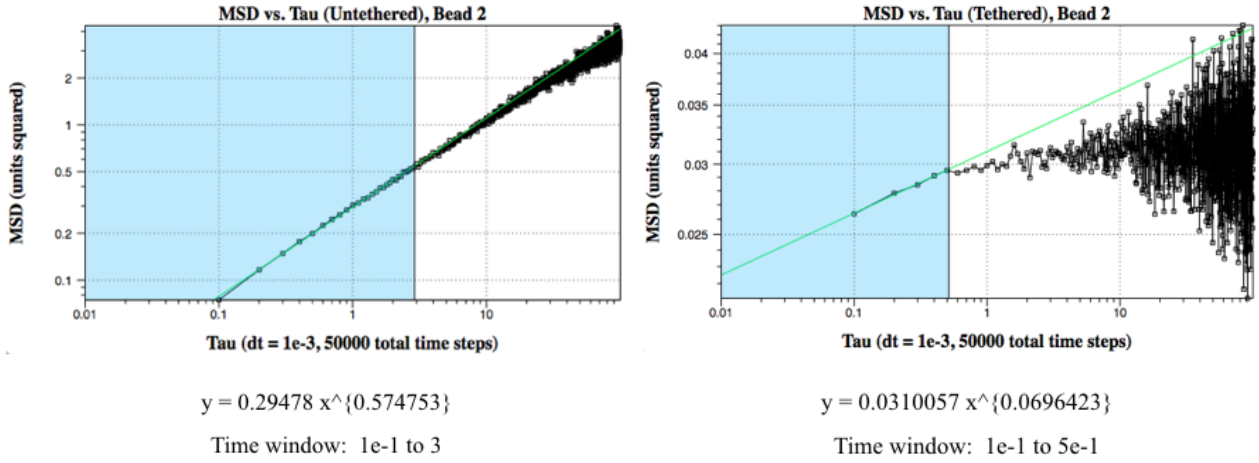


Figure 2.16: *MSD averaged over a single run, estimating alpha using powerfit, for an untethered and double-tethered Rouse chain.* We look at Bead 2 in a 1002 bead chain, i.e. a bead located near the end of the chain. Only values inside the light blue box are considered in the powerfit.

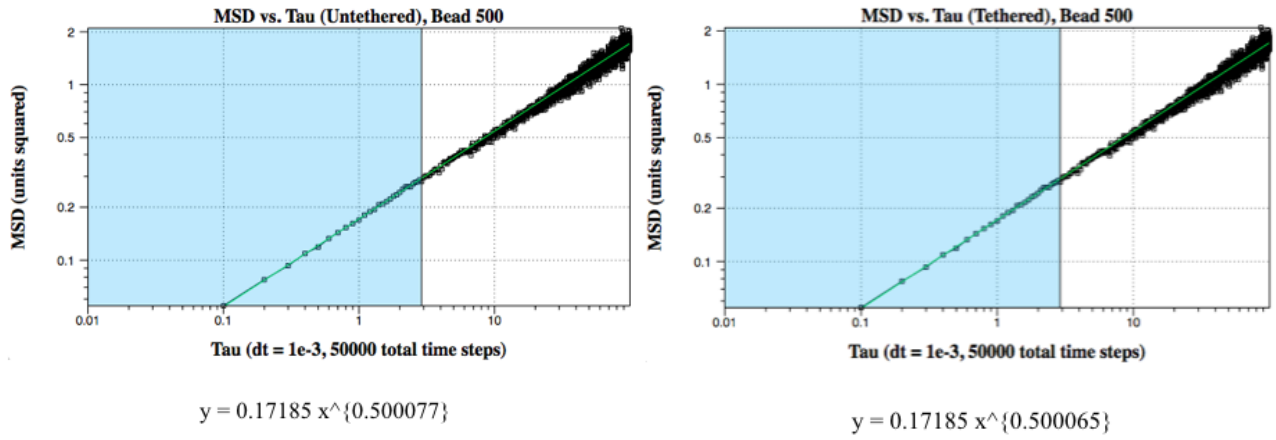


Figure 2.17: *MSD averaged over a single run, estimating alpha using powerfit, for an untethered and double-tethered Rouse chain.* We look at Bead 500 in a 1002 bead chain, i.e. a bead located near the middle of the chain. Only values inside the light blue box are considered in the powerfit.

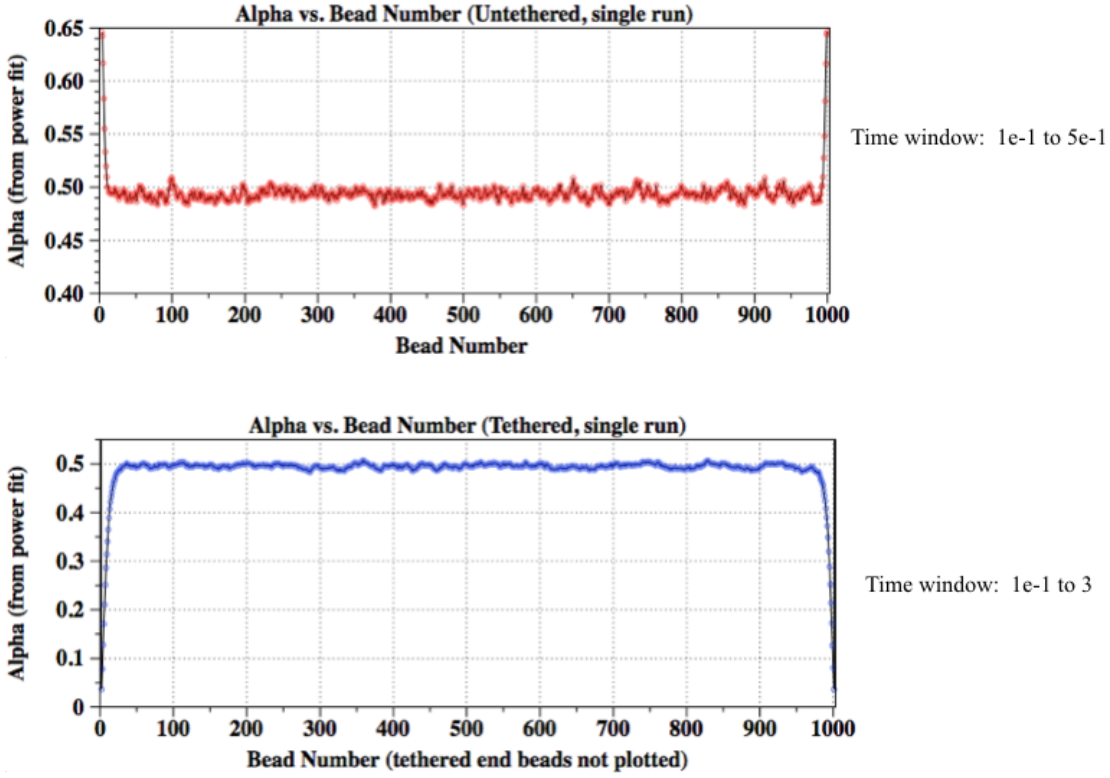


Figure 2.18: *Plot of alpha versus bead number for a 1002 bead chain, where alpha is computed using powerfit as in Figures 2.16 & 2.17.*

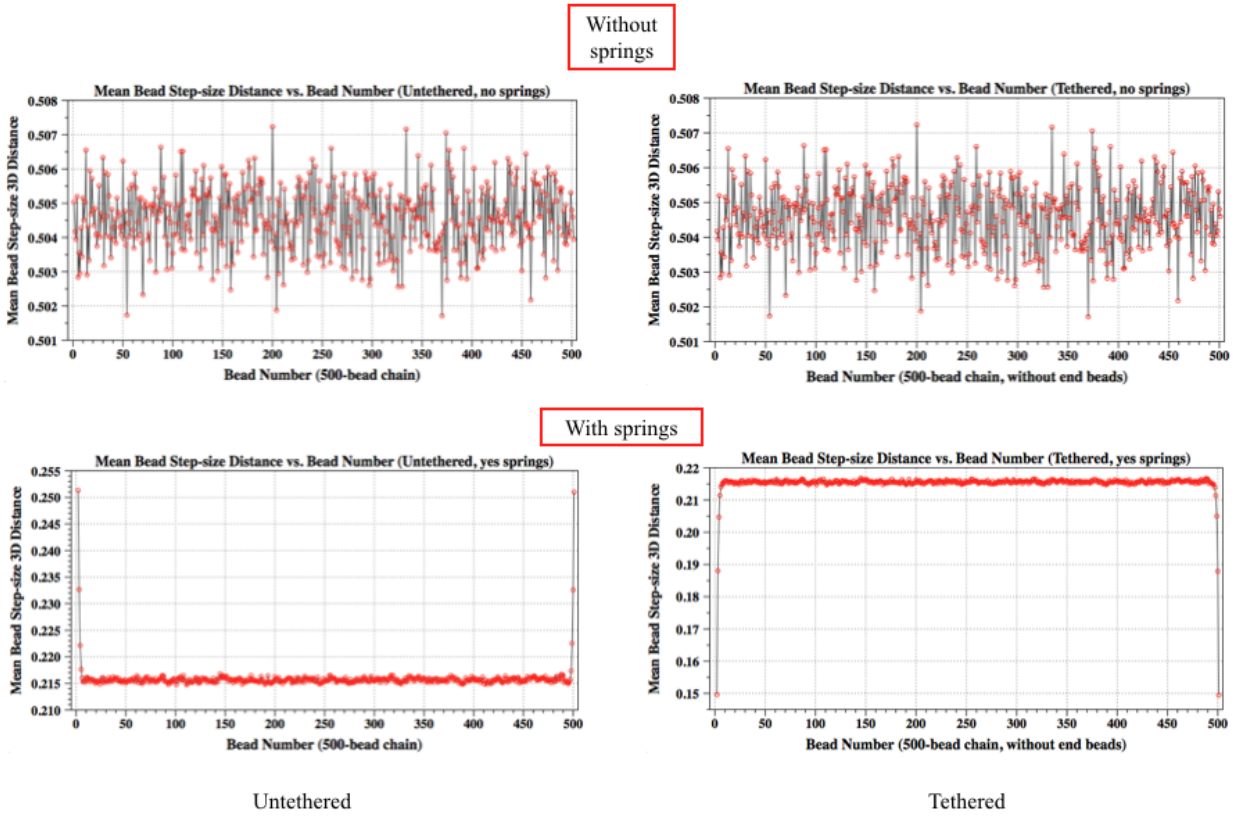


Figure 2.19: *Mean bead step-size distance over time for untethered vs. double-tethered Rouse chain, comparing cases with and without springs*

We also note that the MSD is the same as the variance, at each lag time (Figure 2.20). This is a useful check, both conceptually and as a way to check the accuracy of our code.

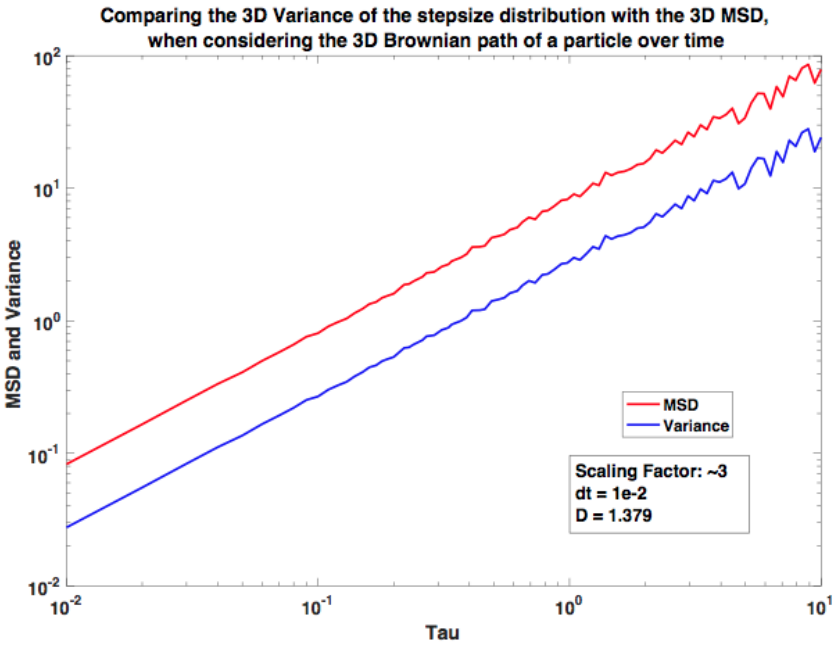


Figure 2.20: *Sample plot showing that the MSD is the same as the variance.* At each lag time, $\tau$, the MSD is the variance of the corresponding van hove correlation function.

### 2.6.2 Mean Squared Displacement

One common way of analyzing polymer dynamics is through calculating mean-squared displacement, a measure of the spatial distribution and potential area of exploration of particles over time. Vasquez and Bloom [4] give an excellent introduction to and motivation for mean-squared displacement, as they state: "The chromosome is floppy and constantly stepping forward, backward, up and down due to thermal and non-thermal, ATP dependent fluctuations. Mean-square displacement measurements are a quantitative tool to deduce the nature of the driving force (random or directed) and the area of exploration (Rc). The mean-squared displacement (MSD) for a given lag time, tau, is defined as $MSD_\tau = \left\langle (x_{t+\tau} - x_t)^2 + (y_{t+\tau} - y_t)^2 \right\rangle$," where $\langle\rangle$ denotes the ensemble average. They further note that "if the movement is a random walk, a log-log plot of MSD vs. lag time is a straight line slope 1, MSD $\sim \tau$." Movement that results in a MSD vs. tau plot with a slope less than 1, ie., MSD $\sim \tau^\alpha$ where $\alpha < 1$, is called subdiffusive. Live cell experiments suggest that tagged DNA spots move subdiffusively, as $\alpha \sim 0.4$. Molecular crowding makes it harder for beads to explore space, and thus contributes to the smaller $\alpha$ value, and confinement forces imposed by cell walls and tethering additionally restrict bead movement, "resulting in a MSD plot that 'plateaus' at large lag times" and estimates the total area that can be explored [1]. In the case of a doubly-tethered chain, we expect beads

located near the end of the chain to explore less space than beads located near the middle of the chain.

To compute the mean-squared displacement, we must first choose a lag time, denoted in our code by the tau variable. If our position vector for bead i over time is defined as $[x_1, x_2, x_3, x_4, x_5, ..., x_N]$, where $N$ is the total number of recorded positions, then a $\tau$ value of 1 means that we will find the displacements vector $[(x_2 - x_1), (x_3 - x_2), (x_4 - x_3), ..., (x_N - x_{N-1})]$, a $\tau$ value of 2 means that we will find the displacements vector $[(x_3 - x_1), (x_5 - x_3), (x_7 - x_5), ..., (x_N - x_{N-2})]$, a $\tau$ value of 3 means that we will find the displacements vector $[(x_4 - x_1), (x_7 - x_4), ..., (x_N - x_{N-3})]$, etc. To complete the MSD calculation, we then square the displacements individually, and take the average. This is our MSD value for a given $\tau$ value. Generating many paths for bead $i$, calculating the MSD each time, and then averaging all of the resulting MSD values improves the accuracy of the calculation. We can compute the MSD for different variables separately, or generalize the calculation to 2D or 3D. MSD can be averaged over many different beads or compared among many different beads. Note that the size of the lag time, $\tau$, is bounded above by the total number of recorded position values; if $\tau$ is too large, we will only be able to average over a few squared displacement values, which would produce unrealistic effects and noise.

## 2.7 Normal mode analysis of a double-tethered Rouse chain

This section addresses and shows explicitly through mathematics how we determine the normal modes of a Rouse chain. This section was originally written by Paula Vasquez, although I have added a substantial amount of comments and more fully written out some of the steps.

We start with a Rouse chain with $N + 2$ beads, the equation of motion for bead $i$ is

$$\zeta d\mathbf{X}_i = -k_s A_{i,j} \mathbf{X}_j dt + \sqrt{2k_B T \zeta} d\mathbf{W}_i \tag{2.11}$$

where each $\mathbf{W}$ satisfies

$$\begin{aligned} \langle \mathbf{W}(t) \rangle &= \mathbf{0}, \\ \langle \mathbf{W}(t_1), \mathbf{W}(t_2) \rangle &= \boldsymbol{\delta}(t_1 - t_2). \end{aligned}$$

Since we have $N+2$ beads in our bead-spring chain, $i$ ranges from 1 to 52, and thus Eqn.(2.11) actually defines $N+2$ total differential equations. (We will see shortly that these equations are coupled.) The notation $\mathbf{X}_i$ is a $(N+2)\times 1$ vector, in which the entries denote the x-position coordinate of each of the $N+2$ beads at a specific time:

$$\mathbf{X_i(t')} = \begin{pmatrix} \text{x-position of Bead 1 at time } t' \\ \text{x-position of Bead 2 at time } t' \\ \text{x-position of Bead 3 at time } t' \\ \vdots \\ \text{x-position of Bead N+1 at time } t' \\ \text{x-position of Bead N+2 at time } t' \end{pmatrix}_{(N+2)\times 1} = \begin{pmatrix} X_1(t') \\ X_2(t') \\ X_3(t') \\ \vdots \\ X_{N+1}(t') \\ X_{N+2}(t') \end{pmatrix}_{(N+2)\times 1} .$$

We abbreviate this notation as:

$$\mathbf{X_i} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_{N+1} \\ X_{N+2} \end{pmatrix}_{(N+2)\times 1} .$$

The delta function, $\boldsymbol{\delta}(t_1 - t_2)$, helps define the random noise term. We can think of the delta function as a way to express the "impulse forces" of bead collisions; it says that there is no correlation between the direction a bead moves at one time step and the direction it moves at the next time step, unless $t_1 = t_2$ (ie. time does not change). Note that $t_1$, $t_2$, and $t'$ just represent two points in time.

The matrix $\mathbf{A}$ is

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & \dots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \end{pmatrix}_{(N+2)\times(N+2)} .$$

We refer to this matrix as the connectivity matrix, $A$. This is how we describe the spring force relationship between neighboring beads on a chain. Conceptually, we are thinking about the forces pointing in towards whatever bead we are currently considering. Note that rows 1 and $N+2$ represent the stationary end beads, 1 and $N+2$. Note as well that there are three nonzero values associated with each middle bead.

Thus, we can express Eqn.(2.11) in the following expanded form (purely for clarity purposes):

$$\begin{pmatrix} dX_1 \\ dX_2 \\ dX_3 \\ \vdots \\ dX_{N+1} \\ dX_{N+2} \end{pmatrix}_{(N+2)\times 1} = \frac{-k_s dt}{\zeta} \begin{pmatrix} 0 \\ \text{-}X_1\text{+}2X_2\text{-}X_3 \\ \text{-}X_2\text{+}2X_3\text{-}X_4 \\ \text{-}X_3\text{+}2X_4\text{-}X_5 \\ \vdots \\ \text{-}X_N\text{+}2X_{N+1}\text{-}X_{N+2} \\ 0 \end{pmatrix}_{(N+2)\times 1} + \sqrt{\frac{2k_B T}{\zeta}} \begin{pmatrix} dW_1 \\ dW_2 \\ dW_3 \\ \vdots \\ dW_{N+1} \\ dW_{N+2} \end{pmatrix}_{(N+2)\times 1} .$$

In order to integrate the coupled system of equations defined in Eqn.(2.11), we first need to decouple the equations, meaning we need to rewrite them in a way that allows us to express $\mathbf{X}_i$ in terms of only $\mathbf{X}_i$ and not $\mathbf{X}_j$. Thinking ahead, we hypothesize that writing A as an eigenvalue decomposition might be useful, as doing so might allow us to make a substitution in the equation of motion that might in turn enable us to effectively rewrite the $\mathbf{X}$ term. In order to write A

as an eigenvalue decomposition, we need to first calculate the eigenvectors and eigenvalues of A. Since we are calculating this by hand, it would be *very* helpful if A was a symmetric tridiagonal matrix.

If we consider the following change of variables:

$$\tilde{\mathbf{X}}_i = \mathbf{X}_i - \mathbf{X}_1 \tag{2.12}$$

$$\mathbf{R}_i = \tilde{\mathbf{X}}_i - \left(\frac{i-1}{N+1}\right)\tilde{\mathbf{X}}_{N+2}, \tag{2.13}$$

the system reduces to

$$\zeta d\mathbf{R}_i = -k_s A_{i,j}\mathbf{R}_j dt + \sqrt{2k_B T\zeta}\, d\mathbf{W}_i. \tag{2.14}$$

To break down this change of variables a bit:

$$\mathbf{R}_i = \begin{pmatrix} 0 \\ X_2\text{-}X_1\text{-}\frac{1}{N+1}(X_{N+2}\text{-}X_1) \\ \\ X_3\text{-}X_1\text{-}\frac{2}{N+1}(X_{N+2}\text{-}X_1) \\ \vdots \\ X_{N+1}\text{-}X_1\text{-}\frac{N}{N+1}(X_{N+2}\text{-}X_1) \\ \\ 0 \end{pmatrix}_{(N+2)\times 1}.$$

Since $\mathbf{R}_1 = 0$ and $\mathbf{R}_{N+2} = 0$, the system becomes $N \times N$ and

$$\mathbf{A} = \begin{pmatrix} 2 & -1 & 0 & \dots & 0 & 0 & 0 \\ -1 & 2 & -1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -1 & 2 & -1 \\ 0 & 0 & 0 & \dots & 0 & -1 & 2 \end{pmatrix}_{N \times N} .$$

The matrix A is now a tridiagonal symmetric matrix, and we can now determine its eigenvalues

$$\lambda_p = 4 \sin^2 \left( \frac{p\pi}{2(N+1)} \right). \tag{2.15}$$

*Note: In Matlab, there is a built-in eigenvalue decomposition function. Thus, there we do not need to introduce a change of variables in Eqns (2.12) and (2.13) and rewrite Eqn.(2.11) — instead we can directly evaluate* $A_{(N+2)x(N+2)}$.

The matrix A is diagonalizable as,

$$\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{P},$$

where $\mathbf{P}$ is the matrix whose column vectors are composed of the eigenvectors of A, and $\mathbf{P} = \mathbf{Q}^{-1}$, and thus

$$\begin{aligned} Q_{i,j} &= 2 \sin \left( \frac{ij\pi}{N+1} \right), \\ P_{i,j} &= \frac{1}{N+1} \sin \left( \frac{ij\pi}{N+1} \right), \\ \Lambda_{i,i} &= \lambda_i. \end{aligned}$$

Substituting in our decomposition, Eqn.(2.14) becomes

$$\begin{aligned}
\zeta d\mathbf{R}_i &= -k_s Q_{i,j}\Lambda_{i,j}P_{i,j}\mathbf{R}_j dt + \sqrt{2k_BT\zeta}d\mathbf{W}_i \\
\zeta P_{i,j}d\mathbf{R}_i &= -k_s \Lambda_{i,j}P_{i,j}\mathbf{R}_j dt + \sqrt{2k_BT\zeta}P_{i,j}d\mathbf{W}_i \\
\zeta d\mathbf{Y}_i &= -k_s \Lambda_{i,j}\mathbf{Y}_i dt + \sqrt{2k_BT\zeta}P_{i,j}d\mathbf{W}_i \\
d\mathbf{Y}_i &= \frac{-k_s}{\zeta}\lambda_i \mathbf{Y}_i dt + \sqrt{\frac{2k_BT}{\zeta}}P_{i,j}d\mathbf{W}_i.
\end{aligned}$$

Observe that this equation is now uncoupled! We are expressing $\mathbf{Y}_i$ only in terms of itself. We want to rewrite this equation as,

$$d\mathbf{Y}_p = \frac{k_p}{\zeta_p}\mathbf{Y}_p dt + \sqrt{\frac{2k_BT}{\zeta_p}}d\mathbf{W}_p \tag{2.16}$$

because we want to investigate the effect that multiplying $P_{i,j}$ has on the random noise, $d\mathbf{W}_i$. We require that

$$\begin{aligned}
\left\langle \sqrt{\frac{2k_BT}{\zeta}}P_{p,i}, \sqrt{\frac{2k_BT}{\zeta}}P_{q,i} \right\rangle &= \frac{2k_BT}{\zeta_p}\delta_{p,q} \\
\left(\sqrt{\frac{2k_BT}{\zeta}}\right)^2 \langle P_{p,i}, P_{q,i}\rangle &= \frac{2k_BT}{\zeta_p}\delta_{p,q} \\
\langle P_{p,i}, P_{q,i}\rangle &= \frac{\zeta}{\zeta_p}\delta_{p,q} \\
\frac{1}{2(N+1)}\delta_{p,q} &= \frac{\zeta}{\zeta_p}\delta_{p,q} \\
\frac{1}{2(N+1)} &= \frac{\zeta}{\zeta_p},
\end{aligned}$$

and $k_p = k_s\lambda_p(\zeta_p/\zeta) = 2(N+1)k_s\lambda_p$. (We let $p$ subscripts denote normal modes, while $i$ subscripts denote regular coordinates.)

We nondimensionalize with (these appear as our scaling factors ts and xs in the code)

$$\tilde{t} = \frac{2k_s}{\zeta}t, \qquad \tilde{\mathbf{Y}} = \sqrt{\frac{ks}{k_BT}}\mathbf{Y},$$

dropping the tildes, the resulting dimensionless uncoupled equation is

$$d\mathbf{Y}_p = -\frac{\lambda_p}{2}\mathbf{Y}_p dt + \sqrt{\frac{1}{2(N+1)}}d\mathbf{W}_p. \tag{2.17}$$

$$\begin{aligned}
y\left(t+\Delta t\right) &= \left(1-\frac{\lambda_p}{2}\Delta t\right)y(t) + \sqrt{\frac{dt}{2(N+1)}}r_1 \\
y\left(t+2\Delta t\right) &= \left(1-\frac{\lambda_p}{2}\Delta t\right)y(t+\Delta t) + \sqrt{\frac{dt}{2(N+1)}}r_2 \\
&= \left(1-\frac{\lambda_p}{2}\Delta t\right)\left[\left(1-\frac{\lambda_p}{2}\Delta t\right)y(t) + \sqrt{\frac{dt}{2(N+1)}}r_1\right] + \sqrt{\frac{dt}{2(N+1)}}r_2 \\
&= \left(1-\frac{\lambda_p}{2}\Delta t\right)^2 y(t) + \left(1-\frac{\lambda_p}{2}\Delta t\right)\sqrt{\frac{dt}{2(N+1)}}r_1 + sqrt\frac{dt}{2(N+1)}r_2 \\
&\vdots
\end{aligned}$$

so in general

$$\mathbf{Y}_p\left(t+n\Delta t\right) = \left(1-\frac{\lambda_p}{2}\Delta t\right)^n \mathbf{Y}_p(t) + \sqrt{\frac{dt}{2(N+1)}}\sum_{k=1}^{n}\left(1-\frac{\lambda_p}{2}\Delta t\right)^{n-k}\mathbf{r}_k, \tag{2.18}$$

where the $r_k$'s are normally distributed random numbers.

Finally, we need to put our variables back in terms of $\mathbf{R}$. The mapping between $\mathbf{R}$ and $\mathbf{Y}$ is

$$\begin{aligned}
\mathbf{Y}_p &= P_{p,j}\mathbf{R}_j \\
&= \frac{1}{N+1}\sum_{k=1}^{N}\mathbf{R}_k \sin\left(\frac{pk\pi}{N+1}\right),
\end{aligned} \tag{2.19}$$

$$\begin{aligned}
\mathbf{R}_i &= Q_{i,p}\mathbf{Y}_p \\
&= 2\sum_{k=1}^{N}\mathbf{Y}_p \sin\left(\frac{ik\pi}{N+1}\right).
\end{aligned} \tag{2.20}$$

We have now solved for the x-positions of all of the beads over time by integrating Eqn.(2.11). If bead positions are desired in dimensions besides x, then this analysis must be repeated using the appropriate variable. For example, if we wish to compute positions over time for our beads in 3D, we must also solve the equations:

$$\zeta d\mathbf{Y}_i = -k_s A_{i,j}\mathbf{Y}_j dt + \sqrt{2k_B T\zeta}d\mathbf{W}_i$$

and

$$\zeta d\mathbf{Z}_i = -k_s A_{i,j}\mathbf{Z}_j dt + \sqrt{2k_B T\zeta}d\mathbf{W}_i.$$

All of the data generated for use in this thesis was produced using the C++ version of the code, in which we implement forward Euler to numerically solve the SDEs. Most of the data generated for use in my Master's project was produced using a version of the Matlab code that computes the classical solution (or full solution) of a Rouse chain with N+2 beads. In that code, we "evolve" the bead positions by integrating using the forward Euler method, and the importance of the normal modes is primarily apparent in the chosen size of our time step $dt$ and in the amount of time we choose to equilibrate the code. Paula Vasquez has also written a Matlab position-generating code that more fully utilizes the normal mode analysis. As long as we know the random numbers that were used in a given simulation, we can check for accuracy by computing the normal mode solution and comparing to our model's results.

Note that solving the normal modes for an untethered chain or a single-tethered chain follows the same process, the only differences being in the definition of the transformations defined in Eqns 2.10 and 2.11, due to the consequent changes in several of the values of the connectivity matrix. Further detail regarding the calculations for these two cases are described in "Viscoelasticity and molecular rheology" by Alexei E. Likhtman [19].

## 2.8 In vivo experiments

In order to obtain the experimental position data, the Bloom Lab used light microscopy to track DNA movement in live yeast cells. They struck the cells with a laser at a designated spot on the DNA previously tagged with GFP (green fluorescent protein). At each time step, they took a series of images of the cell at multiple z-positions and chose the most in-focus image for consideration in our data analysis. As they stepped along in time, the resulting images showed the tagged DNA spot at 3D different positions, thus illustrating DNA movement. Using microscope simulation programs like Metamorph to obtain xy-coordinates for these tagged spots, they obtained position data over time for the tagged spots. This program determines spot coordinates by searching for the brightest pixels, a method that is much more efficient than manually going through and locating the marked spot on each image. Each tagged spot is approximately 100-5000 base pairs long and can be thought of as representing one gene. The beads in our model's bead-spring chains are proxies for these tagged DNA spots, and allow us to track and to directly compare simulated and experimental position data for specific locations on specific chromosomes. In addition to the figures and statistics that we can produce through our computational model, the Bloom Lab can perform their own analysis of our simulated data by inputing it into the same microscope simulator program that they use for their experimental data. Importantly, this constant two-way discussion comparing simulated data from the mathematical model and experimental lab data enables the biologists to ask questions that they cannot see experimentally. Although there is a 0.25 micro resolution limit in light microscopy, we are able to get motion analysis on the nanometer length scale through understanding how light is defracted.

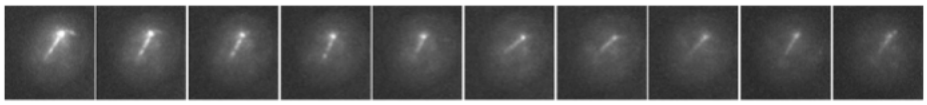

Figure 2.21: *Yeast cell in G1 phase with DNA tagged at a certain spot. Images 1-3 are taken consecutively. Images 4-10 are taken in 1 minute intervals.*

CHAPTER 3

## Introducing Chapter 4

This chapter serves as an introduction to the paper presented in the following chapter, “Entropy Gives Rise to Topologically Associating Domains". As stated in the abstract for that paper, here we investigate chromosome organization within the nucleus using polymer models whose formulation is closely guided by experiments in live yeast cells. Polymer models have been well-shown to successfully recapitulate chromatin behavior *in vivo*. We employ bead-spring chromosome models together with loop formation within the chains and the presence of nuclear bodies to quantify the extent to which these mechanisms shape the topological landscape in the interphase nucleus. By investigating the genome as a dynamical system, we show that domains of high chromosomal interactions can arise solely from the polymeric nature of the chromosome arms due to entropic interactions and nuclear confinement. In this view, the role of bio-chemical related processes is to modulate and extend the duration of the interacting domains, rather than to be essential to their creation. This prediction of the role of enzymatic processes is not fully expected by the biologists and represents an important step towards better understanding how proteins influence chromosome dynamics. Furthermore, we present findings for both 2D and 3D versions of our model and discuss how the 2D model effectively captures the qualitative results of the 3D model, and thus although is not useful from a quantitative standpoint, remains valuable for illustrating patterns and visualizing results.

CHAPTER 4

# Entropy Gives Rise to Topologically Associating Domains

[1]

## 4.1 Introduction

The genome comprises the entire genetic information that makes up an organism. This information is encoded in DNA and stored, spatially and dynamically, in the nucleus of every cell in that organism. Packaging while preserving functionality of DNA in the nucleus is one of the most remarkable tasks performed by cells. To accomplish this task, the cell employs hierarchical levels of compaction and organization. Understanding the spatial dynamics of the genome is a crucial step in characterizing how DNA adopts and transitions between different functional states over the course of the cell cycle, facilitating vital functions such as gene expression, DNA replication, recombination and repair. Today, advances in instrumentation, experimental techniques, theory and computation are poised for integration toward a predictive description of the structural organization and dynamics of the living genome, and for understanding the entropy-dominated statistical mechanics underpinning different cellular functions. Toward this goal, we hypothesize that in the nucleus, thermodynamics, in particular entropy, dominates the spatial organization of chromosome arms while active kinetic processes modulate this organization. This view represents a shift from that of enzymatic, biochemical cellular processes playing the leading role in these cell functions. In short, enzymes do not create a new topological and energetic landscape in the nucleus; rather they bias the entropy-dominated stochastic dynamics into cycle-specific states. Here we show that entropy and confinement dictate the leading order structure and dynamics of the genome, whereas the role of enzymes is to guide, stabilize and sustain cycle-specific genome states.

[1]This chapter previously appeared as an article in *Nucleic Acids Research*. The original citation is as follows: Vasquez, P.A., Hult, C., Adalsteinsson, D., Lawrimore, J., Forest, M.G., and Bloom, K. "Entropy gives rise to topologically associating domains." *Nucleic Acids Research*, **44**, 5540-5549. June 2 2016.

The organization of the genome in the nucleus can be divided into three length scales [20]. At the lower scale, between 1 base pair (bp) and several thousand base pairs (kb), the genome adopts a beads-on-a-string structure, ∼11 nm in diameter, formed by nucleosomes and their linking DNA. This is the building block of the chromatin fiber [21, 22, 23, 24, 25, 26]. Above the 11-nm fiber, the structure is irregular and dynamically samples a variety of morphologies. Experiments *in vitro*, using DNA templates with uniform arrays of repeating sequences, have shown that chromatin condenses under specific salt conditions to a more compact structure of 30 nm [23, 27, 28]. However, the regularity and existence of this 30-nm fiber *in vivo* continue to be debated [29, 30, 31, 32]. At length scales on the order of the nuclear domain (micron scale $>$ 100 Mb or mega base pairs), the genome is divided into chromosomes, each of which occupies statistically defined territories. Previous work has shown that polymer models, coupled with confinement in the nucleus and centromere/telomere tethering, capture the essential features of experimental chromosome territories [1, 33, 34]. These territories are defined statistically, indicative of dynamics of the chromatin chain, implying that the chromosomes are not spatially confined to a given region, rather there is a high probability that they will be found in a specific region (territory) in the nucleus [1, 33, 34, 4]. At intermediate and large length scales (∼Mb), it is known that the genome is not randomly organized in the nucleus and this spatial organization plays a key role in the execution of a variety of nuclear functions, e.g. coordinately regulated genes and DNA repair factories [35, 36, 37]. A key question is to understand the organizational principles that guide the hierarchical organization at this level of compaction. Loops along the polymer chain are one such guiding principle. It has been shown that chromatin loops influence the size and dynamic features of chromosome territories [38, 39, 40, 41, 42, 43, 44, 45]. Similarly, sub-nuclear compartments and sub-chromosomal interacting domains, such as the nucleolus, gene bodies and topologically associating domains (TADs), play important roles in the organization and dynamic interactions of the genome [46, 47, 8, 48, 49]; however much remains unknown at this level.

Here, we explore the formation of subdomains in individual live cells during interphase that arise solely from entropy-driven polymeric properties of the nuclear-confined chromatin fiber. We show that entropic fluctuations and excluded volume interactions of confined, tethered polymer chains are sufficient to represent a genome organization that is dynamic, with genes and chromosomal interacting domains varying considerably in time and space and from cell to cell as observed experimentally

[50, 51, 52, 53]. Addition of loops to the model provides a mechanism to tune contact frequencies, also observed experimentally [41]. Finally, sub-nuclear compartments offer a mechanism to control interactions at a local level.

Our modeling is based on experimental data from individual, live yeast cells. By quantifying cellular phenomena at this level, we gain a unique perspective on different biological processes that is distinct from the concepts gained by averaging data from populations of cells. Our studies complement those obtained through high-throughput data by extrapolating single cell observations to population averages. The use of dynamic imaging of sub-diffraction chromosomal loci in live cells together with mathematical modeling offers a powerful approach to dissect chromosome interactions in living cells. This approach yields insight into the mechanisms that govern the establishment and maintenance of diverse functional chromatin states in a dynamic cell.

## 4.2 Materials and Methods

### 4.2.1 Modeling approach

Recent results on the diffusive behavior of proteins and DNA in vivo suggests that chromatin motion obeys the dynamics of a polymer network [1, 54, 55, 56, 57, 58]. It has also been shown that purely random chain behavior of isolated chromosomes cannot explain many of the specific patterns observed in experiments [1, 4, 59, 60]. The question remains whether these patterns arise from generic polymer physics, specific processes such as proteins binding to different sites of the chromatin or higher order organizational principles. Computational simulations show that many of the organizational features observed experimentally emerge from the coupling of the polymer nature of chromatin and simplified geometric constraints [1, 33, 4, 51, 52, 61, 62, 63, 64]. These studies investigate different numbers and sizes of chromosomes, confinement to the nuclear space, centromere/telomere tethering and excluded volume effects. We extend these studies to include the formation of loops within chromatin fibers and the inclusion of sub-nuclear compartments, and explore their impact on global organization as well as local interactions of the genome.

The dynamics of chromatin loci are consistent with those of highly flexible polymers [1, 34, 4, 65, 66, 67]; for a review of polymer models of interphase chromosomes we refer the reader to [34, 58]. Accordingly, modeling the four-dimensional behavior of a complex system such as multiple

chromatin fibers in the yeast nucleus requires the exploration of a large conformational space. The goal herein is to study the consequences of entropic fluctuations and polymer conformational constraints (confinement, tethering, looping) on the formation of zones of local intra-chromosomal interactions. To gain insights into the statistical features of this complex system, we use a simplified, low dimensional, model that captures the essential physics and geometry: a two-dimensional (2D), round nuclear domain and four entropic chains representing four chromosome arms. Results from this simple model are consistent with a three-dimensional (3D) model of the yeast genome (32 chains), as discussed in the supplemental material, thereby providing a simplified context to convey the spatial and dynamic features of chromatin organization.

The dynamics of the chains are modeled using a bead-spring polymer model where the chromosomes are represented by interacting beads connected via springs described by a worm-like chain (WLC) force law [14, 15]. Each chain is tethered at both ends as shown in Figure 4.1, representing the tethering of the telomeres to the nuclear membrane and the centromeres to the spindle pole body. This tethering resembles the Rabl configuration observed experimentally [50, 63, 68, 69, 70]. In addition to tethering, chains are confined within the nuclear (circular in 2D, spherical in 3D) domain; both constraints reflect in vivo observations of yeast chromosomes [1, 70]. Within this description a segment in a chain represents a 'blob' [71] of chromatin, rather than linear segments. Each blob consists of $N_{k,s}$ Kuhn lengths of length $2L_p$, where $L_p$ is the persistence length of the chromatin fiber. Although a disparate set of values has been reported in the literature for the persistence length of chromatin in vivo, ranging from 30 to 220 nm [33, 66, 72, 73, 74, 75]; here we consider a persistence length of $L_p = 50$ nm, which corresponds to the persistence length of naked DNA and lies within the reported range for yeast chromosomes. From this perspective it is assumed that the dynamics of the $N_{k,s}$ segments inside each blob follow those of a random walk, forming what is called a Brownian bridge. In this sense, the size of the blobs represents the spatial extent over which the motion of the chromatin fiber is uncorrelated. A detailed description of the model and parameter selection is discussed in the next section.

### 4.2.2 Model formulation

In the model, the evolution of bead positions is governed by a force balance. In the nucleus, forces controlling molecular interactions include van der Waals (attractive or repulsive), electrostatic (attractive or repulsive), steric (repulsive), hydrodynamic and forces due to thermal fluctuations.

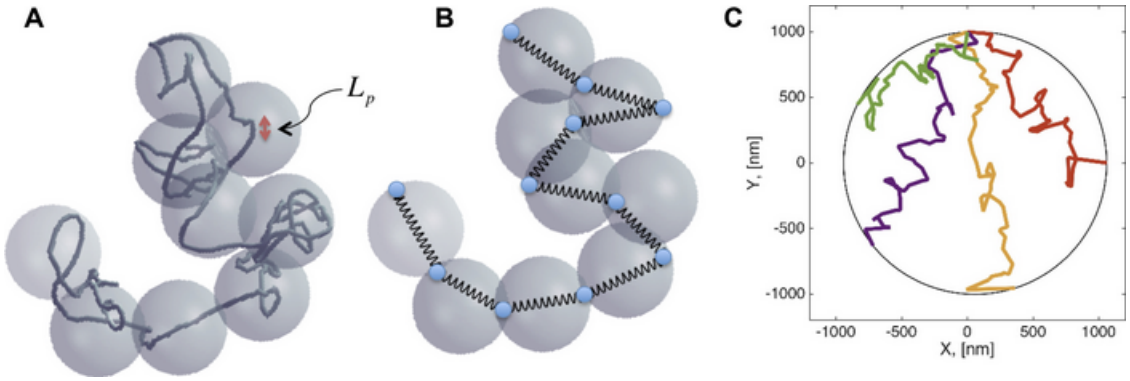


Figure 4.1: *Bead spring representation of chromosome arms.* **(A-B)** Each segment, composed by a spring connecting two beads, capture the dynamics of a 'blob'. **(C)** The computational domain consists of four chains, tethered at both ends to a circular domain of radius 1 $\mu$m.

In our coarse-grained representation, all attractive forces are captured through a spring-like force and all repulsive effects by an excluded volume potential. Since the chromatin network is embedded in a highly viscous environment, hydrodynamic interactions (HIs) are also considered. Here, we incorporate viscous drag on the beads and the so-called free draining approximation, as discussed below.

With these simplifications, our formulation is reduced to the balance of five forces acting on each bead,

$$\mathbf{F}_i^D + \mathbf{F}_i^S + \mathbf{F}_i^{EV} + \mathbf{F}_i^W + \mathbf{F}_i^B = 0. \tag{4.1}$$

**Drag force.** The drag force is assumed to follow a linear relation,

$$\mathbf{F}_i^D = -\zeta \frac{d\mathbf{X}_i}{dt}, \tag{4.2}$$

where $X_i$ is the vector position of bead $i$, and $\zeta$ is the effective drag coefficient. To determine the drag coefficient we consider the relaxation of a single chain following experiments by Fisher *et al.* [76]. In that study the authors found that it takes ∼40 times longer for the chromatin to relax *in vivo*, compared to naked DNA *in vitro*. From this, and based on data from Jendrejack *et al.* [16], we obtain an effective drag coefficient of $\zeta = 2.5 \times 10^{-3}$ pN s/nm.

For simplicity, we omit other types of HIs. Inclusion of HIs in bead-spring models has been shown to be negligible in weak flows [13, 77], whereas the flows inside the nucleus are extremely weak.

For this reason, and for computational cost considerations, we impose the so-called free draining approximation.

**Spring force.** This force captures intramolecular interactions via an attractive potential. We use the WLC force proposed by Marko and Siggia [14, 15], shown to describe well the dynamics of naked DNA [14, 15, 16] and chromatin *in vivo* [76],

$$\mathbf{F}_i^S = \mathbf{F}_{i,j-1} + \mathbf{F}_{i,j+1},$$

$$\mathbf{F}_{i,j} = \frac{k_B T}{4L_p}\left[\left(1 - \frac{R_{ij}}{R_0}\right)^{-2} - 1 + 4\frac{R_{ij}}{R_0}\right]\frac{\mathbf{R}_{ij}}{R_{ij}} \tag{4.3}$$

where,

$$\mathbf{R}_{ij} = \mathbf{X}_i - \mathbf{X}_j, \qquad R_{ij} = \sqrt{\mathbf{R}_{ij}\cdot\mathbf{R}_{ij}}, \qquad R_0 = N_{k,s}\left(2L_p\right). \tag{4.4}$$

**Excluded volume force.** This force imposes an energy penalty on the overlapping of two blobs. In static polymer theory the excluded volume potential is assumed to be a $\delta$-function of the form [12],

$$\mathbf{E}\left(\mathbf{r}_j - \mathbf{r}_i\right) = \nu k_B T \delta\left(\mathbf{r}_j - \mathbf{r}_i\right), \tag{4.5}$$

so that overlapping between two beads has an energy penalty of $\nu k_B T$, where $\nu$ is the excluded volume parameter, $k_B$ is the Boltzmann constant and $T$ is the absolute temperature. For our excluded volume potential we use the form proposed by Jendrejack *et al.* [16], where this function is approximated by a narrow Gaussian function:

$$U_i^{EV} = \frac{\nu k_B T}{2}\left(\frac{3}{2\pi S_k^2}\right)^{3/2}\sum_{i,j=1;i\neq j}^{N}\exp\left[-\frac{3\mathbf{R}_{ij}^2}{2S_k^2}\right], \tag{4.6}$$

$$\mathbf{F}_i^{EV} = -\nabla U_i^{EV},$$

where $S_k^2 = N_{k,s}(2L_p)^2/6$.

**Interaction with the cell wall ($\mathbf{F}_i^W$).** To account for confinement due to the nuclear envelope, whenever a bead moves outside the domain during a time step, it is moved to the nearest point on the domain boundary as follows,

$$\mathbf{X}_i^{new} = \mathbf{X}_i^{out} + \Delta\mathbf{X}_i^{HM}, \tag{4.7}$$

$$\Delta\mathbf{X}_i^{HM} = \Delta p_i \mathbf{H}\left(\Delta p_i\right)$$

where $\Delta\mathbf{X}_i^{HM}$ is the displacement vector due to the Heyes-Melrose algorithm [78], $\Delta p_i$ is the vector from the position of the bead outside the domain boundary $\left(\mathbf{X}_i^{out}\right)$ to the nearest boundary point and $\mathbf{H}\left(\cdot\right)$ is the Heaviside step function.

**Brownian force.** This force captures the random motion due to thermal fluctuations,

$$\mathbf{F}_i^B = \sqrt{2\delta k_B T} d\mathbf{W}_i, \tag{4.8}$$

where $d\mathbf{W}_i$ is a standard normal distribution.

### 4.2.3 Other model parameters

**Validation with experimental data** To validate the model parameters we developed a 3D model, consisting of 32 chains corresponding to the arms of the 16 yeast chromosomes. Model settings were selected by matching 3D simulations to experimental data, as discussed in the Supplementary Material. While the 2D model is not appropriate for direct comparison to experimental data, qualitative agreement between the 2D and 3D models, as shown in Figure 4.2 and Figure 4.3 and Figure 4.13, demonstrates that even a simplified polymer model can emulate the domains of high chromosomal interaction seen in experimental data. Given the ability of the 2D model to display these results with experimentally derived parameters, the 2D model provides a simplified framework to explore the effects of a large range of parameters on the formation and maintenance of domains of high chromosomal interaction.

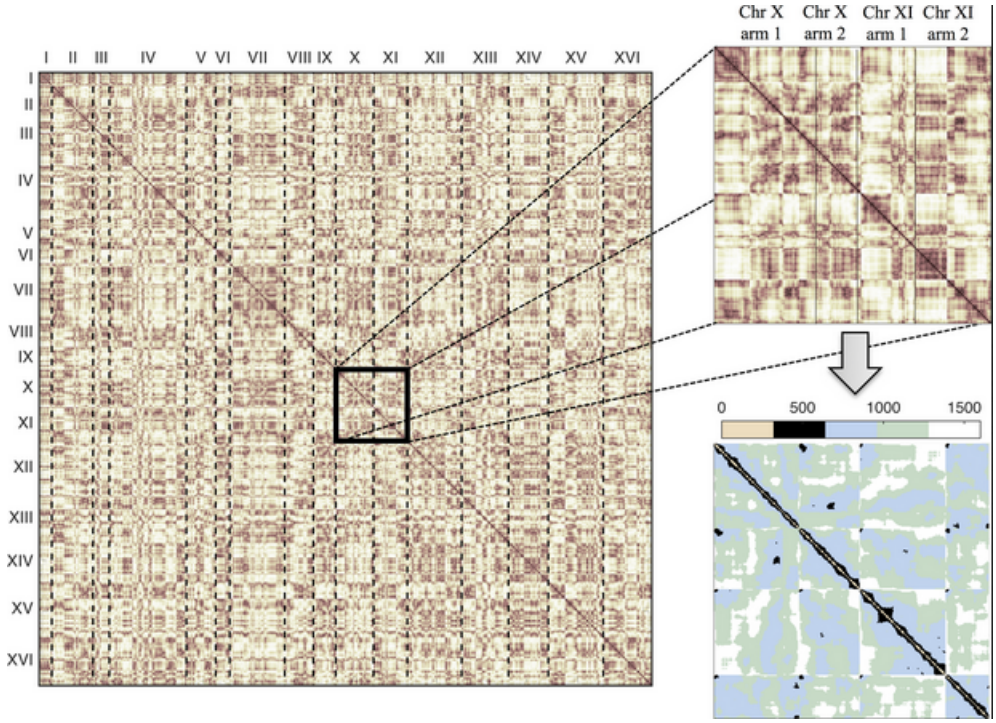


Figure 4.2: Average separation between beads in the 3D model (2443 beads). The behavior of four selected chains agrees *qualitatively* with the behavior obtained for the 2D, four-chains model (see Figure 4.3 and Figure 4.13)

**Number of beads** The number of beads is related to the level of discretization used to describe a chromosome arm. If we consider the number of kilo base pairs (kb) represented by each spring, $N_{kb} = 5$ and the number of Kuhn lengths per spring, $N_{k,s} = 17$, the number of beads representing a chromosome arm is determined by its length in kb. In our 2D model, we consider chains with 52 beads or equivalently 255 kb. For the 3D model, as explained in the supplemental material, the given length of each chromosome arm determines the number of beads for the different chains, see Table 4.1.

**Loops** Chromatin loops impose constraints on chromosome conformation; however, the detailed mechanisms and driving forces of looping are not clear [43, 45, 79, 80]. In this work we assume that protein complexes at the base of loops create a spring-like force, dictated by the binding constants within the complex and the DNA. To form loops, we employ a linear spring force and assume that the spring constant has the same order of magnitude as the springs used to form the chains. Lacking experimental information regarding the distribution of co-localized chromatin loci, our

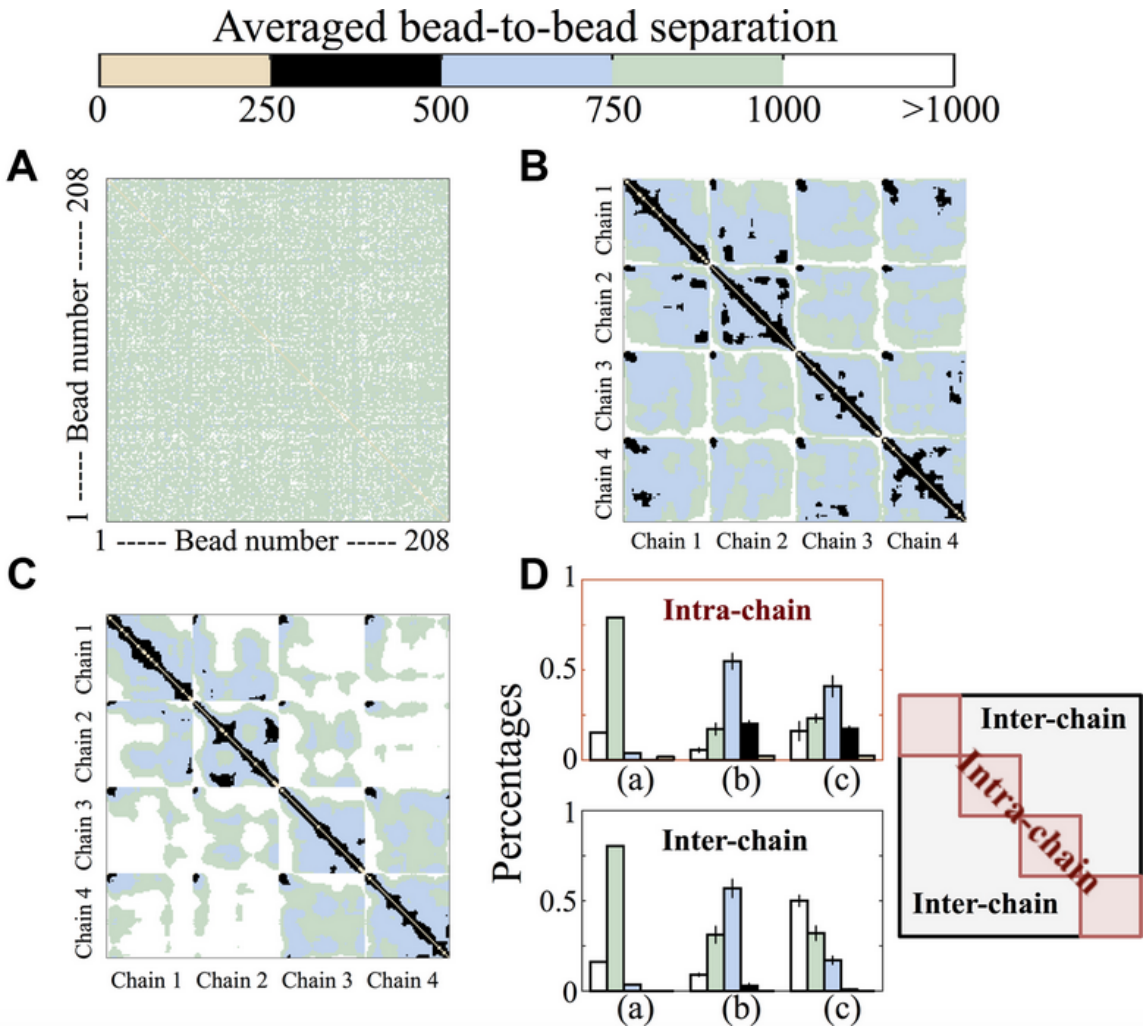


Figure 4.3: Heat maps of distances between beads averaged over 1 h and divided into five regions according to the mean bead-to-bead separation. (A) Beads confined to a circle of radius 1 micron and moving through Brownian motion. (B) An attractive potential (worm-like spring) is added to the same noise history as in (A). The addition of this potential also allows the division of the 208 beads into four chains as shown in Figure 4.1C. (C) Inter-molecular interactions are added to the same simulation through an excluded volume potential. In addition to steric effects, this force also captures any type of repulsion between different beads. (D) Percentage of the five bead-to-bead separation areas for figures (A), (B), and (C), separated according to intra- and inter-chain regions as shown in the inset.

first approximation is to distribute them in a uniform manner along the chain. As pointed out by Fudenberg and Mirny [58] this approach falls short of quantitatively capturing experimental observations; however here we show that they suffice, as a first order approximation, to understand the qualitative role of loops in the organization of chromosomes. Furthermore, in spite of this uniform localization, the distribution of loops within the chain is neither fixed nor uniform in time. At a

given time, some of the 'looping beads' can be near each other (closed loops) while others can be farther from each other (open loops); this is a consequence of the coupling between the spring forces at the base of the loops and the global dynamics within the domain. An example of this dynamic behavior is shown in Figure 4.17.

**Tethering** The first and last beads of each chain correspond to the centromere and telomere sites of a chromosome arm. The position of these beads remain fixed to the boundary throughout the simulations to mimic the attachment of the centromeres to the spindle pole body and telomeres to the nuclear membrane in budding yeast.

## 4.3 Results and Discussion

### 4.3.1 Entropic forces acting on chromosome arms generate domains of high intra-chromosomal interactions

In our coarse-grained representation of chromosome arms, the movement of individual beads captures the dynamics of groups of genes. In this sense, analysis of the distance between beads provides the means to make quantitative conclusions about the contact dynamics between genes. We bin the range of interactions represented by averaged separation between beads as shown at the top of Figure 4.3. We start with the simplest approximation of 208 beads inside a circular domain (representing the nuclear envelope) moving via thermal fluctuations. The beads are independent of each other, meaning the only forces acting on the beads are the confinement force, viscous drag and Brownian forces; these last two are coupled according to the fluctuation-dissipation principle. Figure 4.3A shows the separation of the beads averaged over 1 h. As expected, there are no preferential interactions between any groups of beads. A quantitative description is given in Figure 4.3D, which shows the corresponding percentage for the five ranges of separation. In the absence of linkage between beads, the majority of beads remain, on average, separated by distances greater than 750 nm (green and white regions).

To simulate chromosome arms, non-linear springs are introduced between beads (4 chains of 52 beads each) and each chain is attached at both ends (centromere and telomere) to the nuclear envelope as shown in Figure 4.1C. Figure 4.3B shows the resulting contact map obtained by averaging bead- to-bead distances. The attractive potential, represented by the spring law, captures intra-

chain interactions; for instance an increase in histone occupancy is reproduced by stiffer springs [31]. Introduction of chain configurations into the model allows differentiation between intra-chain and inter-chain interactions in the contact maps (Figure 4.3D). The addition of springs effectively decreases the number of possible configurations for the 208 beads, which in turn decreases the entropy of the system. In other words, a decrease in the system's entropy results in an increase in local interactions. This is represented in Figure 4.3D by an increase in the percentage of blue (500-750 nm separation) and black (250-500 nm separation) regions. We note that 500 nm is the radius of gyration ($R_g$) of a 20 kb random coil of DNA: $R_g = \sqrt{N(2L_p)^2/6}$, validating the estimate of bead separation and the black regions as domains with high bead-bead interactions mimicking TADs.

Finally, excluded volume interactions are incorporated into the model and the resulting contact map is shown in Figure 4.3C. Excluded volume is a mathematical construct that captures the physical basis for molecular exclusion, which is about 8× the intrinsic volume of the polymer chain [81]. Addition of this repulsive potential further reduces the degrees of freedom in chain configurations and accordingly the entropy of the system. This additional reduction in entropy results in a decrease of inter-chain interactions, as shown by the increase in the percentage of white regions (>1000 nm separation) within the inter-chain plot of Figure 4.3D.

Numerical simulations of polymer models have shown that, in addition to tethering, the spatial organization of chromosomes is regulated by excluded volume effects. The latter becomes crucial when confinement is introduced, as chromosomes 'compete' for the limited nuclear space [1, 33, 34, 4]. In this context, the changes in inter- and intra-chromosomal interactions shown in Figure 4.3 are not surprising. The main conclusion from Figure 4.3 is the formation of high-frequency intra-chain interactions, as the black regions within the contact maps imply that those beads remain on average separated by <500 nm over the course of 1 h. By modeling chromosome arms as polymer chains (bead-spring model) interacting via entropic potentials, such as the spring and excluded volume forces, we have shown that domains of high interaction arise naturally in the system without the need for other mechanical or chemical potentials. However, these regions are not static and vary from cell to cell (Supplementary Movie 1).

Finally, we note that previous work argues that loops are responsible for both the formation of chromosome territories and TADs [43, 67]. Here, we argue that loops, or any other type of contact

dynamics, are not necessary to create these domains, as they arise from first principles. Below we show that the role of loops is to modulate these domains by reducing the conformational degrees of freedom of the chains.

### 4.3.2 Changes in chain configuration through the addition of loops result in compartmentalization of chain interactions

A distinguishing feature of eukaryotic chromosomes is an underlying scaffold from which a series of chromatin loops emanate [38, 40, 42]. Loops represent mechanisms to confine topologies to sub-regions of the chromosome providing structural differentiation for processes such as transcription, replication and repair. To examine the consequences of loops in the entropic model, we implemented linear springs connecting specific pairs of beads along each chain. The resulting contact maps are shown for two different loop configurations in Figure 4.4. To assess the changes imposed in the system, simulations were run using the same thermal noise with and without loops. In this manner, subtraction of bead distances, as shown in Figure 4.7, exposes changes arising solely from loop dynamics. These changes are divided in three main categories: distances between beads in the backbone, distances between beads in the outer loops and distances between beads in the backbone and beads in the outer loops; each is shown in the bar plots of Figure 4.4. We further consider intra- and inter-chain distances dividing each bar plot into these two categories.

Figure 4.4 shows that the majority of the interactions remain unchanged after the addition of loops (gray bars). The next largest change comes from beads in the backbone of the same chain (blue bars, intra-chain), since loops in the chain bring the backbone beads closer. In addition, distances between beads in different chains increase (red bars, inter-chain). The fact that this change is the same in both outer-loop and backbone beads suggests an enhancement in chromosome territories. To further illustrate this, we plot the positions of beads with and without loops and show that indeed chromosome territories are more defined with loops (Figure 4.8). Thus chromosome loops increase a chromosome arm's individuality, in agreement with previous studies [43, 67].

Loops arise from several different sources *in vivo*. Cohesin together with CTCF (CCCTC-binding factor) are found in many organisms to tether distant sites and promote expression of the intervening genes [49, 82, 83, 84]. Condensin along with tDNA (DNA encoding transfer RNA (tRNA) genes) and rDNA (ribosomal DNA) transcription factors function to bridge distant sequences in the nucleolus of most organisms [85, 86], as well as pericentric chromatin in yeast [56, 87]. We refer to these

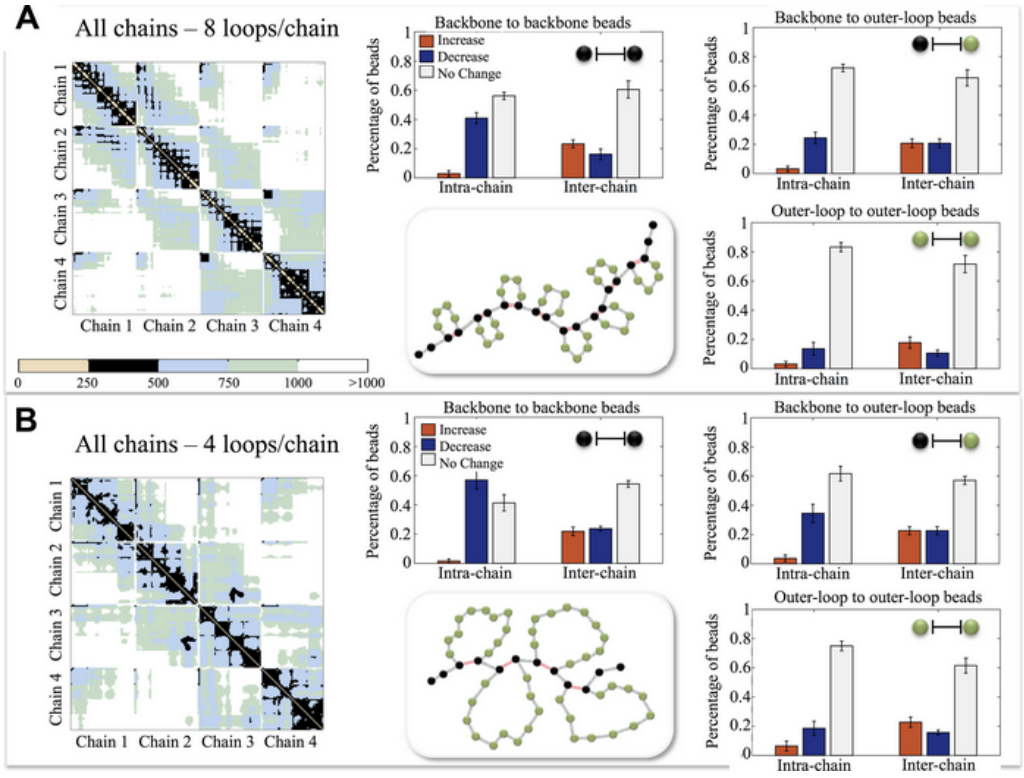


Figure 4.4: Contact maps for different conformations that include loops and changes in the average distance between beads. bead interactions are divided into backbone-to-backbone (dark beads), outer loop-to-outer loop (light beads) and backbone-to-outer loop. Intra-chain interactions are given y the 'same chain' bars, while inter-chain interactions correspond to 'different chain' bars. Shades of the bars correspond to increase, decrease, and no change in the distance between two beads due to the presence of loops. These bars correspond to the regions discussed in Figure 4.14. **(A)** Chain conformation consists of eight loops per chain, each loop comprising six beads. **(B)** Chains are organized into four loops per chain, each loop with 13 beads. Calculation of the changes in the average distance is described in Figure 4.14.

complexes in general as *topology adjusters* [88]. The finding herein indicates that rather than *de novo* creation of subdomain with high interactions, these complexes enhance and extend the duration of domains that stochastically arise from the entropic interactions of the chromosome arms in a confined domain. It will be important in future extensions of the model to explicitly include the stochastic dynamics of these complexes and their interactions with chromosome arms as their biophysical properties become better understood. However, it is clear that such topology adjusters such as cohesin and condensin bias the longevity of naturally 'transient' loops, thus influencing the overall genome architecture (Figure 4.9 & 4.10).

To test whether loops alone can modify the territory of a given chromosome, we introduce loops

in only one of the four chains. The data is shown in Figure 4.5. Addition of loops in only one chain results in a decrease of the intra-chain distances, but only for the chain with loops. This implies that the extent of the territory for that given chain is reduced, while other territories remain unchanged. Furthermore, increase in inter-chain distances also implies that loops serve as a mechanism to isolate individual chains to increase interactions within the chain and to decrease interaction with other chains. These changes in distances are quantified in Figure 4.9.

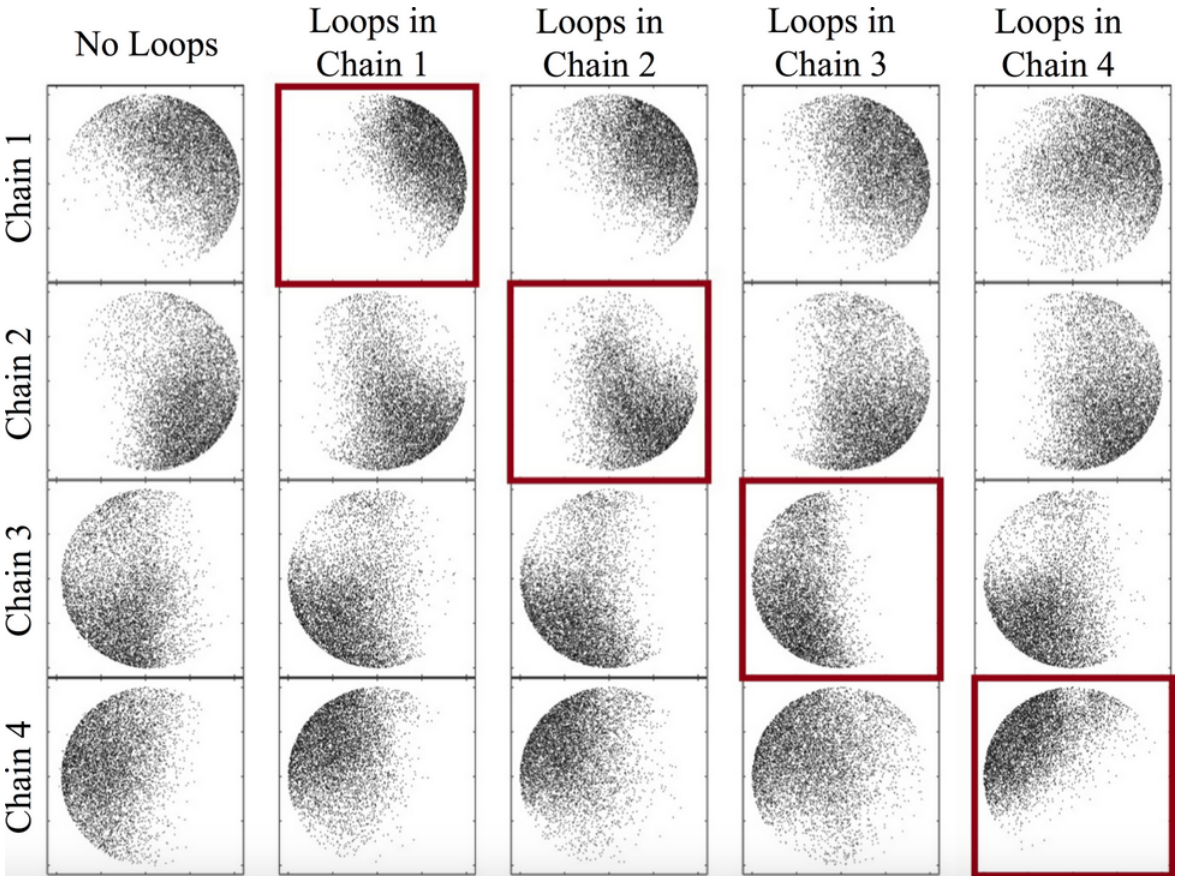


Figure 4.5: Chain's territories as a function of chain configuration. Rows correspond to territories of the same chain, columns correspond to different configurations, from left to right: no loops, loops only in chain 1, loops only in chain 2, loops only in chain 3 and loops only in chain 4. A quantitative analysis of the change in bead-to-bead distances is given in Figure 4.16.

### 4.3.3 Nuclear bodies are secondary levels of regulation of subdomains of high intra-chromosomal interactions

In addition to loops, the cell can employ other secondary mechanisms to bias genome architecture toward a given state. Figure 4.6 explores two such mechanisms: the first one introduces a nuclear body by defining a circular sub-domain within the cell that beads cannot penetrate

(Figure 4.6B); the second one is similar to the previous nuclear body, but, in addition, a set of beads belonging to one chain is confined to the sub-domain (Figure 4.6C). Comparison of the resulting contact maps clearly shows that these nuclear bodies only affect genome interactions at a local level, independent of whether or not the chains have a linear or a loop configuration. As discussed in the previous section, the addition of loops results in enhanced isolation for the regions of high bead-interactions and smaller modifications to the interactions in the rest of the nucleus.

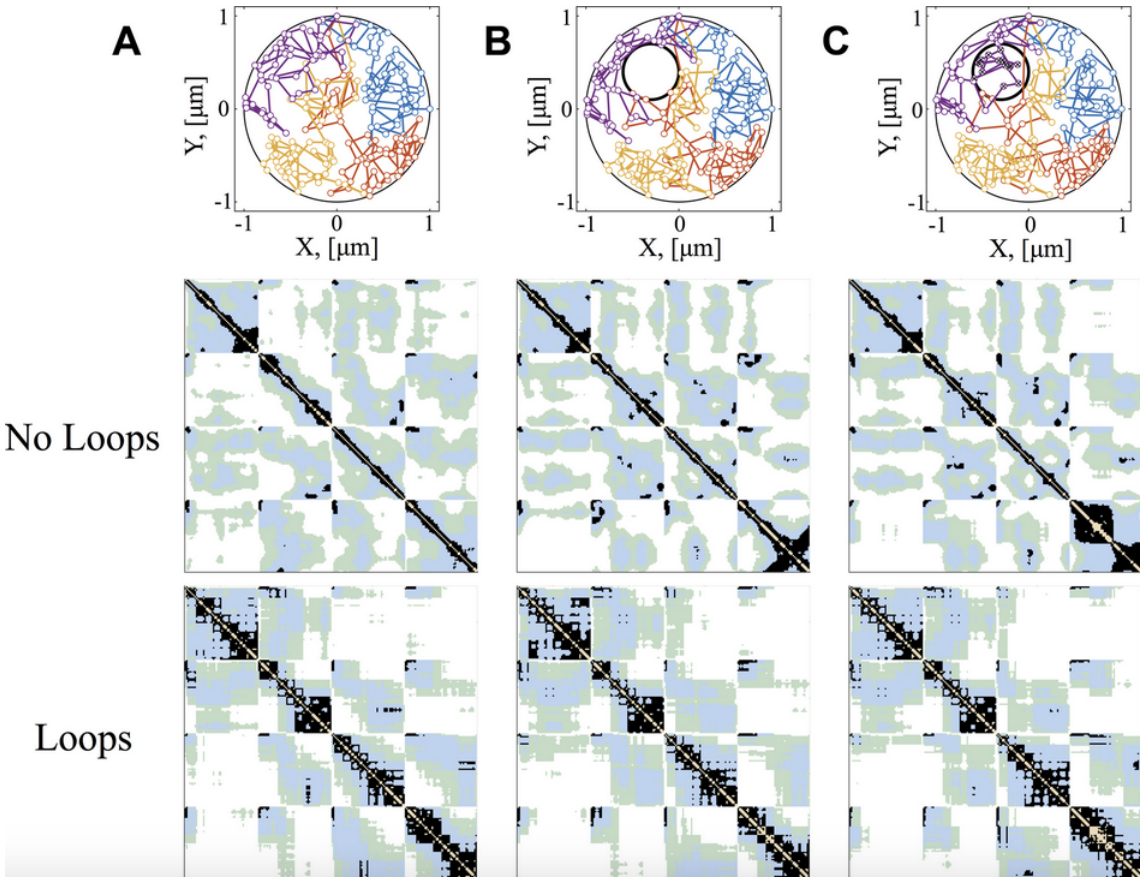


Figure 4.6: Secondary levels of regulation of intra-chain interactions. **(A)** Simulation without of nuclear bodes and resulting contact maps for chains without and with loops. **(B)** Simulations where beads are excluded from a circular region of radius 300 nm placed inside the nucleus (radius 1000 nm). **(C)** Simulations where most beads are excluded from a circular region while beads 15-30 from chain 4 (purple) are confined to the sub-domain.

## 4.4 Conclusion

Critical chromosomal functions have been linked to the spatial conformation of chromosomes. While sequencing-based techniques such as Hi-C provide an overall impression of genome organization,

as pointed out in [58], a large number of contacts observed in Hi-C data does not imply a specific physical or functional interaction between the contacting loci. Rather, mechanistic models of chromatin are necessary to deduce the major physical influences on genome organization. Using mathematical models, we investigated the role that entropic forces, confinement and chromatin looping play in the formation of domains of higher order chromosomal interactions. A persistent view is that this spatial organization is primarily established by enzymatic, protein-guided cellular processes. Here we showed that entropy and the fundamental polymeric nature of chromatin dictate the structure and dynamics of the genome. The role of proteins is to guide, stabilize and sustain cycle-specific genome states.

An important feature of protein-guided chromatin organization is chromatin looping. In addition to their role in the organization of interphase chromatin, loop formation has been shown to play an important role in models of meiotic and mitotic chromosomes [89, 90]. In contrast to the work of Bohn and Heermann [43], the loops in our model are mediated by long-range interactions (linear springs). In [43] the authors showed that 'the formation of loops can be accomplished solely on the basis of diffusional motion.' A next step in the refinement of genome models is to introduce the dynamics for the formation and modulation of loops and to investigate their role in the spatial organization and interaction of the yeast genome. Furthermore, as the biological understanding of specific DNA-protein interactions increase, future models of dynamic genome organization will need to incorporate these interactions to better understand how proteins adjust an entropically-organized genome. In this way, mathematical models of genome organization, together with live cell experiments, will identify genomic signatures that define cell states and protein functions.

## 4.5 Supplementary Data

Supplementary Data are also available at NAR Online.

### 4.5.1 Non-dimensional equations of motion

We use the following relations to scale the model's variables, (prime denotes dimensional quantities)

$$t = t' \left( \frac{k_B T}{N_{k,s}^2 \left(2L_p\right)^2 \zeta} \right), \quad \mathbf{X}_i = \mathbf{X}_i' \left( \frac{1}{2N_{k,s}L_p} \right), \quad \nu = \nu' \left( \frac{1}{\left(2N_{k,s}L_p\right)^3} \right). \tag{4.9}$$

**Dimensional equation of motion for bead i,**

$$\begin{aligned}\zeta d\mathbf{X}'_i = \frac{k_B T}{4L_p} \sum_{j=i-1;i\neq j}^{i+1} \left( \left[ \left(1 - \frac{R'_{ij}}{R_0}\right)^{-2} - 1 + 4\frac{R'_{ij}}{R_0} \right] \frac{\mathbf{R}'_{ij}}{R_0} \right) dt' \\ - \frac{\nu' k_B T}{\pi^{3/2}} \left(\frac{3}{2S_k^2}\right)^{5/2} \sum_{j=1;j\neq i}^{N} \left( \exp\left[ -\frac{3R'^2_{ij}}{2S_k^2} \right] \mathbf{R}'_{ij} \right) dt' + \sqrt{2k_B T\zeta} d\mathbf{W}_i,\end{aligned} \tag{4.10}$$

where

$$\mathbf{R}'_{ij} = \mathbf{X}'_i - \mathbf{X}'_j, \quad R'_{ij} = \sqrt{\mathbf{R}'_{ij} \cdot \mathbf{R}'_{ij}}, \quad R_0 = 2N_{k,s}L_p, \quad S_k^2 = N_{k,s}\left(2L_p\right)^2/6.$$

**Non-dimensional equation of motion for bead i** ,

$$\begin{aligned}d\mathbf{X}_i = \frac{N_{k,s}}{2} \sum_{j=i-1;i\neq j}^{i+1} \left( \left[ (1 - R_{ij})^{-2} - 1 + 4R_{ij} \right] \mathbf{R}ij \right) dt \\ - \nu N_{k,s}^{5/2} c_0 \sum_{j=1;j\neq i}^{N} \left( \exp\left[ -9R_{ij}^2 \right] \mathbf{R}_{ij} \right) dt + \sqrt{2} d\mathbf{W}_i.\end{aligned} \tag{4.11}$$

Here $c_0$ is a constant and N is the total number of beads, i.e., 208 for 2D model and 2443 for the 3D model.

Inspection of Equation (3) shows that the only parameters that can change the *qualitative* behavior of the model are $N_{k,s}$ and $\nu$. Other parameters only affect the magnitude of the dimensional variables. With an increasing level of discretization, $N_{k,s}$, both the spring force and the excluded volume force increase. However, a change in the value of $N_{k,s}$ has a larger effect on the strength of the excluded volume interaction. Note that an increase in $N_{k,s}$ corresponds to a more coarse-grained representation of the chromosomes, i.e., more kb per springs or equivalent less springs in a given chain.

### 4.5.2 Simulation parameters

For the 2D calculations, we run 25 simulations per parameter set, for a total dimensionless time corresponding to 2 hours. We assume equilibration after 1000 non-dimensional time steps and data between 0-1000 is not considered in our analysis. The non-dimensional time step used is 0.001. Other model parameters are given in the table below.

| Parameter | Description | Current Values |
|---|---|---|
| $N$ | Number of beads per chain | 52 |
| $R_n$ | Mean radius of nucleus wall | 1000 nm |
| $N_{k,s}$ | Number of Kuhn lengths per spring | 5 |
| $L_p$ | Persistence length | 50 nm |
| $\zeta$ | Effective drag coefficient | $2.5\times10^{-3}$ pN s/nm |
| $v$ | Excluded volume parameter | $1.2\times10^{-5}$ nm$^3$ |

### 4.5.3 Three-dimensional model

We model the dynamics of the yeast genome during interphase using a bead-spring model as explained in the main text. Assuming a discretization of 5kb per spring, the number of beads per chromosome arms is shown in Table 4.1.

**Centromere and telomere positions.** Both the centromere and telomere sites are kept fixed throughout the simulations. Bead 1 for each chain corresponds to the centromere site and it is shown in Figure 4.7 by a red dot. The telomeres of the 32 chains are distributed among 6 sites as shown in blue dots in Figure 4.7. The corresponding telomere site for each chain is given in Table 4.1.

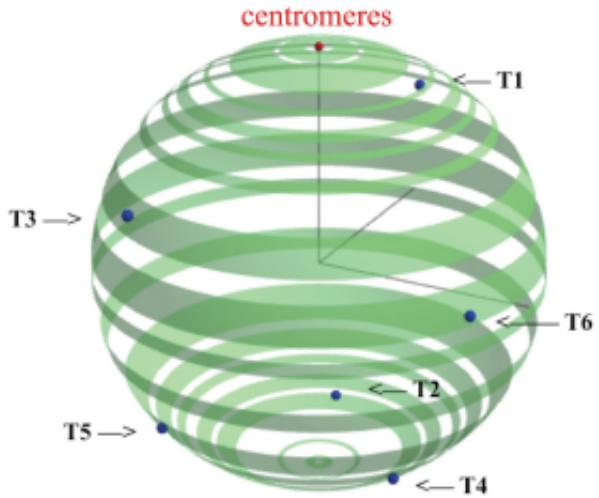


Figure 4.7: Telomere and centromere locations. The first bead of each chain is attached to the centromere location (red), the last bead of each chain is attached to one of the six telomere locations (blue) as stated in Table 4.7.

| Chromosome | Arm | Chain number | Length (kb) | Number of beads | Telomere Site (Fig. S1) |
|---|---|---|---|---|---|
| **I** | 1 | 1 | 151.465 | 31 | T1 |
| | 2 | 2 | 78.752 | 17 | T2 |
| **II** | 1 | 3 | 238.207 | 49 | T3 |
| | 2 | 4 | 574.976 | 116 | T1 |
| **III** | 1 | 5 | 114.385 | 24 | T4 |
| | 2 | 6 | 202.234 | 41 | T5 |
| **IV** | 1 | 7 | 449.711 | 91 | T5 |
| | 2 | 8 | 1082.221 | 217 | T1 |
| **V** | 1 | 9 | 151.987 | 31 | T1 |
| | 2 | 10 | 424.886 | 86 | T5 |
| **VI** | 1 | 11 | 148.51 | 31 | T6 |
| | 2 | 12 | 121.65 | 25 | T4 |
| **VII** | 1 | 13 | 496.92 | 100 | T5 |
| | 2 | 14 | 594.019 | 120 | T3 |
| **VIII** | 1 | 15 | 105.586 | 22 | T3 |
| | 2 | 16 | 457.056 | 92 | T2 |
| **IX** | 1 | 17 | 355.629 | 72 | T6 |
| | 2 | 18 | 84.258 | 18 | T4 |
| **X** | 1 | 19 | 436.307 | 88 | T3 |
| | 2 | 20 | 309.443 | 63 | T5 |
| **XI** | 1 | 21 | 440.129 | 89 | T2 |
| | 2 | 22 | 226.686 | 46 | T3 |
| **XII** | 1 | 23 | 150.828 | 31 | T6 |
| | 2 | 24 | 927.348 | 186 | T4 |
| **XIII** | 1 | 25 | 268.031 | 55 | T2 |
| | 2 | 26 | 656.399 | 132 | T3 |
| **XIV** | 1 | 27 | 628.758 | 127 | T4 |
| | 2 | 28 | 155.574 | 32 | T6 |
| **XV** | 1 | 29 | 326.584 | 66 | T2 |
| | 2 | 30 | 764.706 | 154 | T6 |
| **XVI** | 1 | 31 | 555.957 | 112 | T1 |
| | 2 | 32 | 392.108 | 79 | T2 |

Table 4.1: Number of beads used to simulate the arms of the 16 yeast chromosomes.

**Comparison to experimental data** The experimental and simulated dynamics of the 240kb LacO array are shown in Sup Movie 1. Figure 4.8 below the poster frame from the movie. The DNA is in green, the spindle pole body is in red, and the scale bar is one micron. The simulated image was created using Microscope Simulator 2.3.0 (http://cismm.cs.unc.edu/downloads/, [91]).

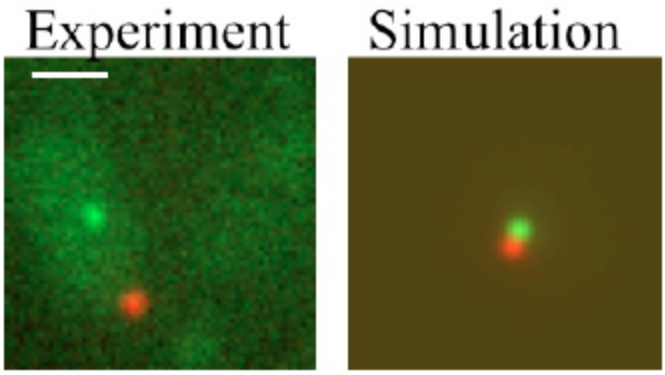


Figure 4.8: Experimental and simulated dynamics of 240 kb LacO.

**Radius of confinement.** Using the temporal distribution of the spot position with respect to the centromere, on can calculate the radius of confinement as explained elsewhere [1]. Results from the simulations are compared to experimental data in Table 4.2.

| Chromosome | Spot location | Chain #, Bead # | Experimental Rc, (nm) | Simulated Rc, (nm) |
|---|---|---|---|---|
| Chr XV, arm 1 | 6.8 kb | Ch 29, Bead 2 | 396 | $425.35 \pm 31.37$ |
| Chr III, arm 2 | 8.8 kb | Ch 6, Bead 3 | 274 | $534.11 \pm 40.95$ |
| Chr III, arm 2 | 24 kb | Ch 6, Bead 6 | 441 | $695.79 \pm 89.30$ |
| Chr II, arm 2 | 240 kb | Ch 4, Bead 49 | 705 | $718.23 \pm 98.89$ |

Table 4.2: Radius of confinement of chromosome loci for simulated and experimental data. Experimental values are obtained from [1]. Simulated values are averaged over 10 runs for 6200 total time steps, each time step is 0.5 s.

**Mean distances.** Following experimental results from Cheng *et al.* [67] we look at the mean separation between two loci in the long arm of chromosome 5. The spots are toto 242 and lacO 386, which correspond to beads 18 and 47 of chain 10 in our model.

The mean distances from the model are 988, 828, and 877 nm for the linear, 5-beads per loop,

and 10-beads per loop configurations. The mean distance reported in Cheng *et al.* [67] is 670 nm. We note that our values are obtained assuming a persistence length of 50 nm, as explained in the previous section, changes in the persistence length only change the quantitative results of the model. In other words, if a persistence length of 30 nm is assumed, the resulting mean distances would be 592, 497, and 526, respectively. Figure S3 shows the statistics for the distance obtained with the 3D model for beads 17-48 and beads 48-77; both of these have separations of 145 kb.

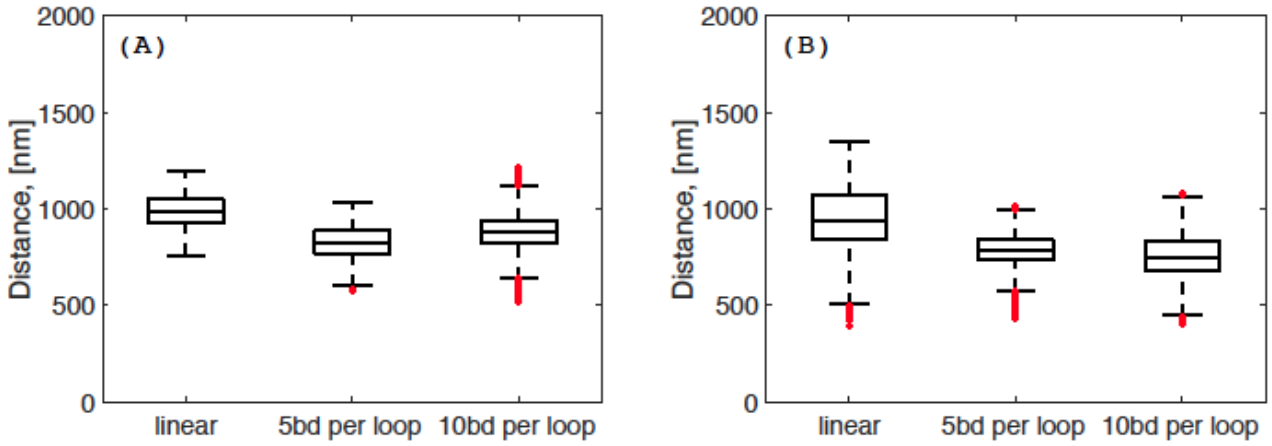


Figure 4.9: Distribution of mean distance between two beads separated 145 kb in arm 2 of chromosome 5, **(A)** beads 17-48 and **(B)** beads 48-77.

In addition to the mean distance, we calculate the kurtosis, following results from Cheng *et al.* [67] who found kurtosis for the experimental data between 2.5 and 3. Figure 4.10 displays the calculated kurtosis for the three types of chain configurations considered in our simulations. The kurtosis is obtained from the distribution of distances between beads 17 and 48 of chromosome 5, arm 2. All reported values are averaged over 6200 time steps ($\Delta t = 0.5s$), correspond to ∼58 min, and 10 repeats of our model simulations. Kurtosis is a non-dimensional parameter and as such is independent of our choice of persistence length.

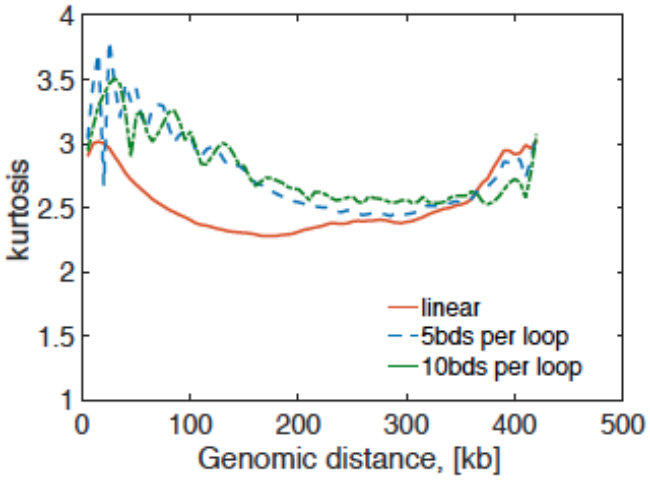


Figure 4.10: Kurtosis values calculated from simulations for three chain configurations and beads 17-48 of arm 2 – chromosome 5.

**Contact probabilities.** The contact probability as a function of genomic separation is shown in Figure 4.11. The probability decays with a power law of 1.1. Duan *et al.* [69] reported an experimental value for the power law exponent in yeast cells of -1.5.

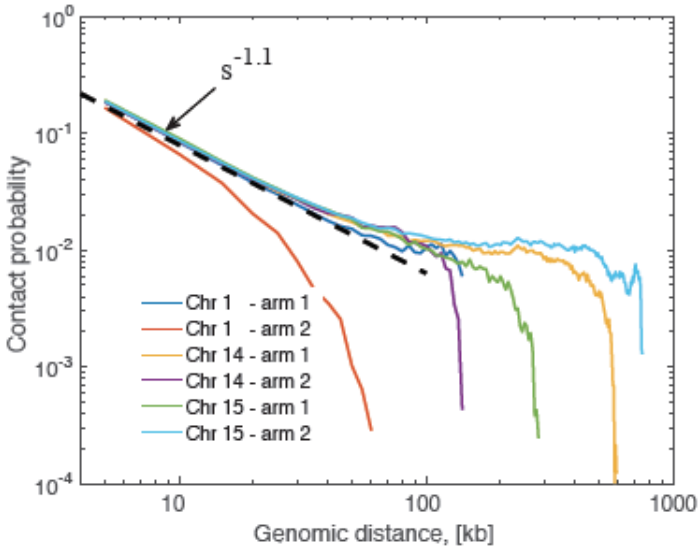


Figure 4.11: Contact probability for 3 chromosomes. At small genomic distances the probability decays with a power law exponent of -1.1. Shorter arms (See Table S1) deviate from this trend at shorted distances.

**Chromosome spot occupancy.** Next we look at the percentage a given chromosome locus is found in different regions within the nucleus. For these, we project the 3D data into the 2D plane as shown in Figure 4.12. In the figure, this projection is color-coded to show the number of times a beads was found in a given point in the xy-plane, but a different positions in the z-direction.

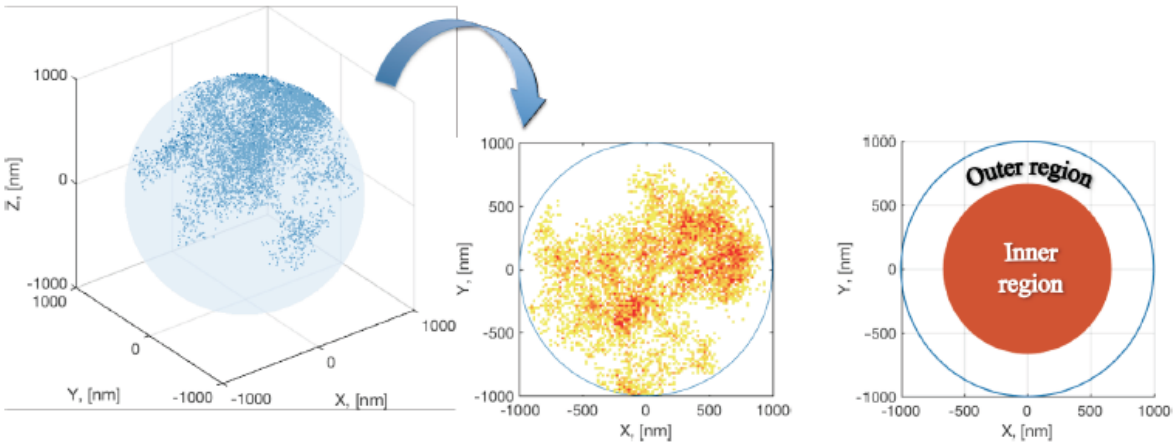


Figure 4.12: Chromosome occupancy. Three-dimensional data is projected into the xy-plane. The percentage of times a bead is found in either the inner or outer regions is reported.

From simulations with linear chains, bead 43 in arm 2 of chromosome 5 remains in the inner region 25.6% and 74.4% in the outer region; for chains with 5-beads-per-loop configurations the percentages are 34.1% - 65.9%; and for chains with 10-beads-per-loop they are 31.3% - 68.7%. Experimental values are 30% - 70% (Barry et al., In Press).

### 4.5.4 Two-dimensional model

**Comparison with 3D model.** To compare the results from the 2D model to the 3D model we choose 4 arms in the 3D model. here we use chromosomes X and XI, since their arms have comparable sizes. Figure 4.13 shows the reduction of intra-chain interactions due to the addition of excluded volume as described in Figure 4.3.

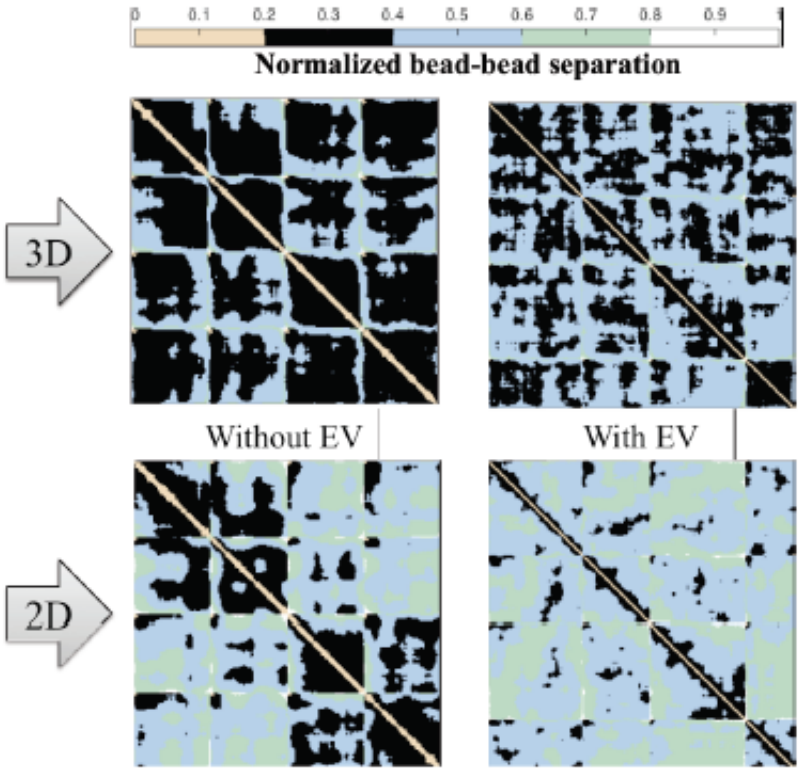


Figure 4.13: Normalized separation between beads for 4 chains. The average bead-to-bead separation resulting from the 2D and the 3D model are shown for simulations with and without excluded volume.

**Quantification of differences in contact maps.** by controlling the parameters in the model, including the random noise, differences arising from a given condition can be easily quantified. Figure 4.14 represents such quantification for chains with and without loops.

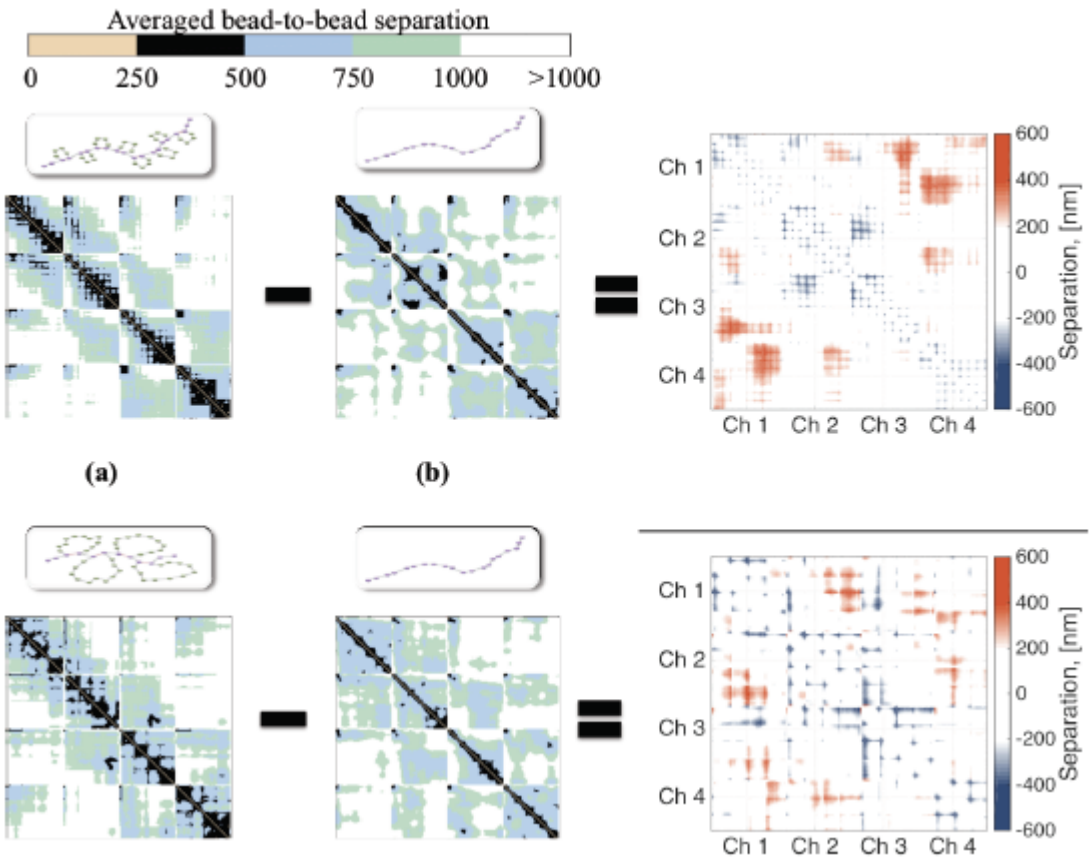


Figure 4.14: Calculated differences in bead-to-bead averaged separation. Since the noise history in **(a)** and **(b)** are the same, the differences arise solely from the inclusion of loops in the chains. In the right hand side, red represents regions in which separation between beads has increased, blue beads for which the separation has decreased and white are regions where there are not changes in mean separation.

### 4.5.5 Loops Dynamics

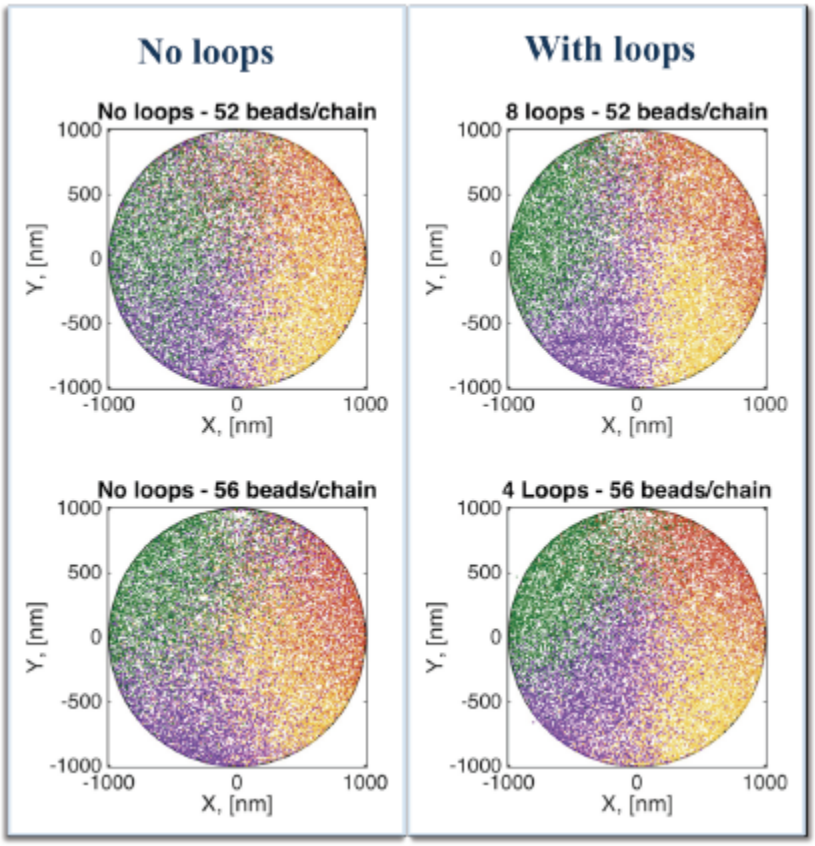


Figure 4.15: Chromosome territories for different looping configurations corresponding to Figure 4.3.

### 4.5.6 Captions for supplemental movies

**Sup. Movie 1.** Temporal evolution of bead-to-bead distances for two different runs with the same model parameters.

**Sup. Movie 2.** Comparison between experimental and simulated data.

## 4.6 Acknowledgements

Computational resources were provided by the Maxwell high performance computer cluster at the University of South Carolina and the KillDevil cluster at the University of North Carolina at Chapel Hill.

## 4.7 Funding

NSF - Division of Mathematical Sciences (DMS) [1410047 to P.A.V.]; NIH [R37 GM32238 to K.B.]; NSF [DMS-1100281 to C.H., DMS-1412844, DMS-1462992 to M.G.F.]; NIH [T32CA201159-01 to J.L.]. Funding for open access charge: NIH [R37 GM32238].

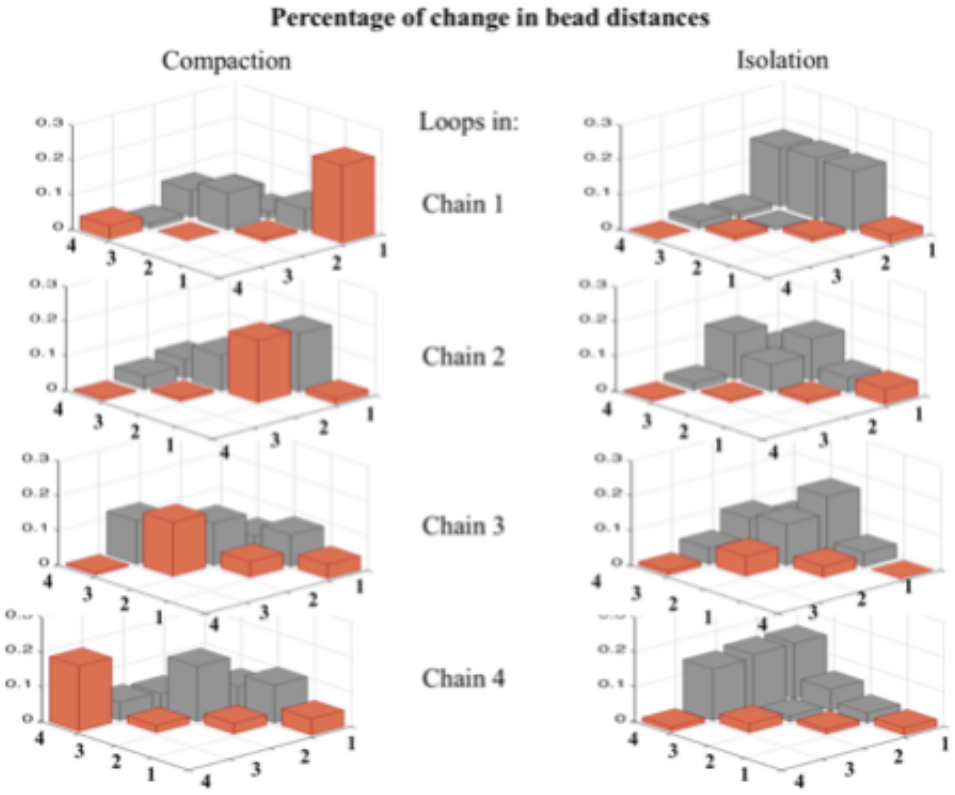


Figure 4.16: Changes in the average distance between beads from simulations where only one chain has loops compared to simulations where all chains have linear configurations. Red bars indicate changes in intra-chain distances and gray bars give changes in the inter-chain distances.

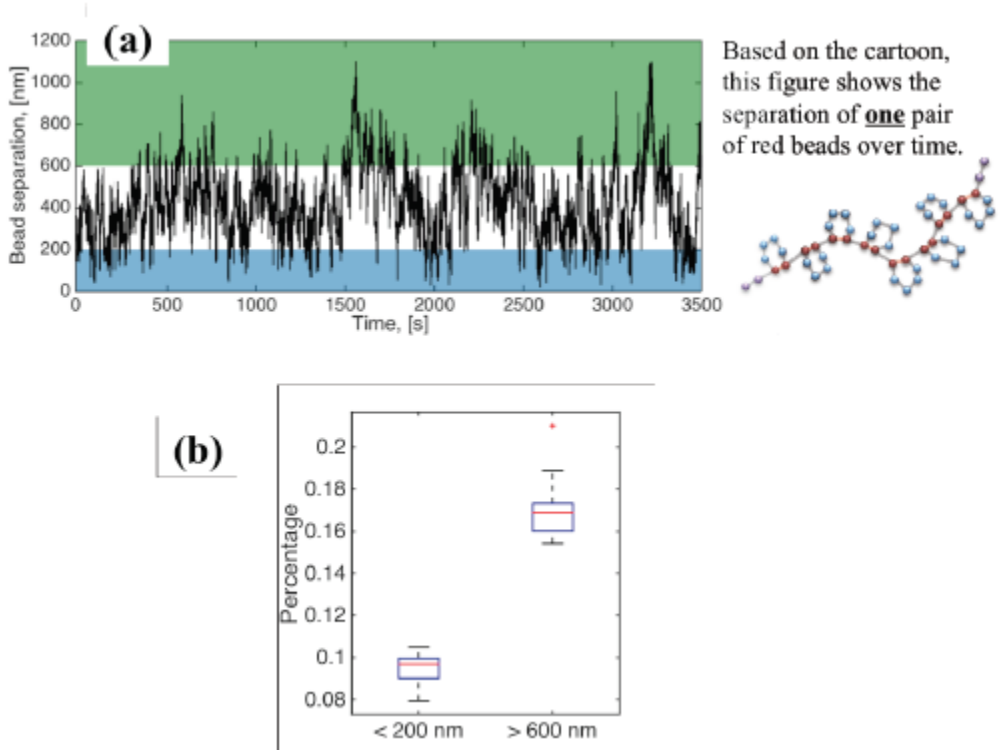


Figure 4.17: Temporal behavior of distance between beads at the base of loops. **(a)** Separation between one pair of red beads as a function of time. **(b)** Percentage of times two red beads were within 200 nm of each other (closed loops) and farther than 600 nm from each other (open loops). Data obtained from all pairs of connecting (red) beads in a chain over 7000 time steps ($\Delta t = 0.5s$) and 10 different simulations. The chain corresponds to the longest arm of chromosome II in the 3D simulations.

CHAPTER 5

# Introducing Chapter 6

This chapter serves as an introduction to the work presented in the following chapter, “Enrichment of dynamic chromosomal crosslinks drive phase separation of the nucleolus". As stated in the abstract for that paper, here we discuss how regions of highly repetitive DNA, such as those found in the nucleolus, show a self-organization that is marked by spatial segregation and frequent self-interaction. The mechanisms that underlie the sequestration of these sub-domains are largely unknown, in part due to the difficulty of distinguishing among DNA repeats experimentally. Using a stochastic, bead-spring representation of chromatin in budding yeast, we find enrichment of dynamic chromosomal cross-links recapitulates the segregation, morphology, and self-interaction of the nucleolus. Rates and enrichment of crosslinking have profound consequences on domain morphology. Our model demonstrates that the nucleolus is phase separated from other chromatin in the nucleus, and predicts that multiple rDNA loci would form a single nucleolus independent of their location within the genome. Fluorescent labeling of budding yeast nucleoli with CDC14-GFP revealed that a split rDNA locus indeed forms a single nucleolus. We propose that nuclear sub-domains, such as the nucleolus, result from phase separations within the nucleus, which are driven by the enrichment of protein-mediated, chromosomal crosslinks. We show through visualization and statistical analyses that weak dynamic binding kinetics in the nucleolus leads to the formation of clusters, whereas strong dynamic binding kinetics promotes a more equilibrated and uniformly mixed spatial organization of the nucleolus beads. This clustering, or lack thereof, in response to different dynamic binding kinetics is robust; namely, clusters form when we change from strong to weak dynamic binding kinetics partway through the simulation, and clusters break apart and uniformly mix with other nucleolus beads when we change from weak to strong dynamic binding kinetics partway through. As will be discussed in the following chapter, these predictions regarding nucleolus dynamics have significant potential biological implications.

CHAPTER 6

# Enrichment of Dynamic Chromosomal Crosslinks Drive Phase Separation of the Nucleolus

[1]

## 6.1 Introduction

The eukaryotic nucleus is a complex three-dimensional environment in which genome function depends not only on the linear arrangement of regulatory sequence elements but also on their spatial organization for effective control of gene expression and nucleic acid metabolism [92, 93, 94]. The spatial organization is in constant flux; individual genes can reposition within the nucleus in response to environmental or developmental cues, and the genome can be mobilized in times of genotoxic stress. The nucleus is composed of a variety of sub-domains or different compartments, each of them with a distinct structure and function. The mechanisms by which subnuclear compartments are formed and maintained as well as what determines their composition, size, shape, and number at various stages of the cell cycle remain largely unknown. Integration of current advances in microscopy, chromosome engineering, theory and computation enable exploration and validation of the statistical mechanical underpinnings that account for formation and maintenance of different subnuclear compartments.

In this paper we investigate the dynamic organization of the nucleolus, the site of ribosomal RNA synthesis, within the nucleus of budding yeast. The nucleolus is composed of repeat DNA sequences and, as a result, is often not included in large-scale sequencing or chromosome conformation studies [64]. The nucleolus can readily be identified as an approximately crescent-shaped structure adjacent to the nuclear envelope and typically opposed to the spindle pole body [95, 96]. In addition to housing the ribosomal DNA (rDNA), the nucleolus is a reservoir for cell cycle regulatory factors

[1]This chapter has been submitted as an article in *Nucleic Acids Research*. The projected citation is as follows: Hult, C., Adalsteinsson, D., Vasquez, P.A., Lawrimore, J., Bennett, M., York, A.C., Forest, M.G., and Bloom, K. "Enrichment of dynamic chromosomal crosslinks drive phase separation of the nucleolus." *Nucleic Acids Research*. Submitted 11 March, 2017.

such as FEAR and MEN. Upon anaphase onset, several of the regulatory proteins are modified and released from the nucleolus to carry out anaphase. In yeast, the nucleolus is found in arm 2 of chromosome XII. Albert et al. [65] performed a comprehensive investigation of this chromosome and found that the dynamics of non-rDNA loci consistently followed those of homogeneous, tethered polymer chains. In contrast, the dynamics of the rDNA loci showed a distinctive deviation from such behavior; namely, larger separation with respect to the nuclear center and slower movement. In agreement with these results, Wong et al., [64, 53] used polymer models to simulate chromosome XII as a heteropolymer assigning a 10x larger size to the rDNA segments of the chain. This size was used so that the resulting nucleolus occupied roughly one third of the nuclear volume. Using this minimalistic dynamic model the authors were able to explain a large set of quantitative data reported on yeast nuclear architecture including locus positions, contact frequencies, and motion characteristics [53]. Although the nucleolus plays a unique role in the organization of the nuclear architecture, to our knowledge no other modeling efforts have included its dynamics in an explicit manner. Few exceptions remain where the nucleolus is included as a topological constraint within the computational domain [33, 97, 98]. However, in these works the nucleolus was assumed to be static and uniform.

Stochastic simulations of entropy-driven, bead-spring polymer chain models account for many features of the dynamic properties of chromatin fibers confined within the yeast nucleus [2, 53, 65, 1, 4]. In this study we explore key parameters that endow a sub-domain of the genome with characteristics that define the nucleolus. As discussed above, previous work has modeled the nucleolus as a chain of increased diameter [53]. This approach provides a physical basis for chain thickening and manifests experimental findings. Here we take a different approach and introduce crosslinking within or between chains representing chromatin interactions with condensin or other SMC proteins. The model herein is fully reflective of experimentally based biological parameters including chromosomal DNA properties, nuclear confinement, tethering of chromosome arms to the centromere and telomere sites, and the relative lengths of all 32 chromosome arms (16 chromosomes, 32 arms roughly metacentric). We translate simulated data from the 3D computational models into equivalent microscope images, to view and analyze experimental images obtained from live cell microscopy. Through these visualization tools and comparison of simulation results and experimental data, we report the statistical mechanics sufficient to account for nucleolar dynamics and conformation due to molec-

ular mechanisms (SMC protein, crosslinking kinetics) that are beyond current experimental resolution.

## 6.2 Materials and Methods

### 6.2.1 Strains and Imaging

Budding yeast strains EMS219 (*Mat alpha, his5 leu2-3,212 ura3-50 CAN1 asp5 gal2 (form I1 rDNA::leu2 URA3+)*), intact rDNA, and EMS60-UVR-12 (*LEU2+, URA3+ CANS form I1 rDNA), translocated rDNA [99]*, were transformed with CDC14-GFP:KAN to label the nucleolus to generate DCY1021.1 and DCY1017.2, respectively. Cells were grown in YPD with excess adenine. Strains were grown until mid-log phase prior to imaging. Images were acquired at room temperature (25 degrees C) using an Eclipse E600FN microscope (Nikon) with a 100x Plan Apo TIRF 1.45 NA objective (Nikon) and ImagEM EM-CCD digital camera (Hamamatsu) with a custom Lumencor LED illumination system (Lumencor Inc.) using MetaMorph 7.7 imaging software (Molecular Devices). Single stacks contained 7 Z-planes sections with 300 nm step-size. Image stacks were cropped and maximum intensity projections were created using ImageJ.

### 6.2.2 Modeling approach

The observed motion of chromatin loci is consistent with those of highly flexible polymers. As a consequence, polymer models have proved valuable in the understanding of chromatin dynamics [2, 67, 34, 4, 66, 1, 53, 58, 33, 63, 62, 89]. In our approach, chromosome chains are modeled using a bead-spring polymer model where each arm is represented by interacting beads connected via springs following a worm-like chain (WLC) force law [14, 15]. Each chain is tethered at both ends representing the tethering of the telomeres to the nuclear membrane and the centromeres to the spindle pole body. This tethering resembles the Rabl configuration observed experimentally [50, 63, 68, 69, 70]. In addition to tethering, chains are confined within the nuclear domain represented by a sphere of radius 1 $\mu$m; both constraints reflect in vivo observations of yeast chromosomes [1, 70]. The basis of the model is a balance of forces acting on each bead,

$$\mathbf{F}_i^D + \mathbf{F}_i^S + \mathbf{F}_i^{EV} + \mathbf{F}_i^W + \mathbf{F}_i^B = 0. \tag{6.1}$$

The forces considered in our model are the drag force $\mathbf{F}_i^D$ opposing the movement of the bead; the spring force $\mathbf{F}_i^S$ capturing bead-bead interactions via an attractive potential; the excluded volume

force $\mathbf{F}_i^{EV}$ that opposes overlapping of two beads; interactions with the cell wall $\mathbf{F}_i^{W}$ that ensures that beads remain within the spherical domain and that beads corresponding to centromere and telomere sites remain fixed to the domain wall; and the Brownian force $\mathbf{F}_i^{B}$ that captures the random motion of the beads due to thermal fluctuations. For details of the functional form of these forces and the model parameters, we refer the reader to [2].

### 6.2.3 Nucleolus modeling

The discretization in our model corresponds to 5kb of DNA per spring. The total number of beads composing each chromosome arm are determined based on this discretization; for specific numbers see [2]. The nucleolus is approximately 1.8 million bp in length, corresponding to 361 beads. Here, we simulate the nucleolus by increasing the size of arm 2 in chromosome XII by $\sim 361$ beads. The position of these beads is in agreement with experimental observations; however the qualitative observations presented in this study are independent of the location of these 361 beads within arm 2 of chromosome XII. In general, nucleolus beads follow the same force laws and have the same parameter values as all other beads in our computational domain, with the exception that crosslinking molecules (condensin or other SMC proteins) are introduced that bind and unbind to the nucleolus beads. For simulations where the nucleolus is split the nucleolus beads are divided between arm 2 of chromosome III and arm 2 of chromosome XII, consistent with experiments. Table 6.1 summarizes modeling assumptions regarding number of beads.

| | | **Number of beads** | | |
|---|---|---|---|---|
| **Chromosome** | **Arm** | No nucleolus | Single nucleolus | Split nucleolus |
| III | 1 | 24 | 24 | 24 |
| | 2 | 41 | 41 | 325 |
| XII | 1 | 31 | 31 | 31 |
| | 2 | 186 | 546 | 263 |

Table 6.1: *Summary of number of beads used in each type of simulation.* All other chromosome arms are discretized as in (Vasquez et al. 2016).

### 6.2.4 Crosslinking

As in [2], we assume a pair-wise binding of beads by introducing a spring-like force between them. This spring force obeys a WLC law, just like the one between neighboring beads in the bead-spring chromatin chain. However, in the crosslinks, the springs are 50x stronger. Note that, when these crosslinks occur intrachain, loops are formed. We explore several assumptions about crosslinking in the nucleolus and external to the nucleolus.

(a). *Absence of crosslinks throughout the whole domain.* This set of data is used as baseline comparison with all other simulations.

(b). *Fixed loops in the nucleolus only.* Here we assume that the 361 beads composing the nucleolus are arranged such that strong WLC springs connect every third, fifth, or seventh bead. This results in chain configurations of permanent 3-bead, 5-bead, and 7-bead loops, respectively.

(c). *Dynamic cross linking in the nucleolus only.* We assume stochastic dynamic crosslink formation within the nucleolus, with all beads within the nucleolus available for binding-unbinding kinetics, and assume no crosslinks outside the nucleolus. We tune dynamic cross linking through the use of six parameters, defined below. Unlike single molecule analysis or imaging, this approach enables us to discern how the dynamics of entropic chain fluctuations, together with on and off timescale distributions of the binding protein, influence the strength of interactions within the nucleolus.

(d). *Preferential dynamic cross linking interval vs. external to the nucleolus.* We assume dynamic cross linking throughout the genome, but with a fraction (1/3, 1/10) of active binding sites external to the fully active nucleolus.

### 6.2.5 Dynamic looping parameters

(a). *Barrier.* The distance (90 nm) between which two beads must fall in order to be eligible to form a crosslink. Experimental work suggests that the binding protein condensin can reach and bind portions of DNA that are at most 90 nm apart [100].

(b). *Maximum loop distance.* The distance where the WLC spring has a singularity. the bond might start at a larger distance when they form the crosslink, but will quickly move within

this distance. This value was chosen base on experimental work suggesting that condensin can stretch to 45 nm [100].

(c). *Loop force scale.* Applicable in both the uniform looping and dynamic looping cases, this parameter is a coefficient that scales the linear part of the WLC spring; namely, it makes the spring stiffer at shorter length scales.

(d). *Mean on, Mean off, Standard Deviation on, Standard Deviation off.* These parameters are times in seconds. They are the mean and standard deviation for how long a crosslink is active or inactive. In our simulations, we have varied the values for “Mean On" and “Mean Off" (often choosing values so that the ratio Mean On:Mean Off = 9) but have kept both “Standard Deviation On" and “Standard Deviation Off" set to 20% of their respective mean value for all simulations. For example, if “Mean On" = 0.09 sec and “Mean Off" = 0.01 sec, then “Standard Deviation On" = 0.018 sec and “Standard Deviation Off" = 0.002 sec.

(e). *Formation and destruction of links.* Due to the dynamic nature of the loop formation and the constantly changing spatial organization of the beads, the pair bonds need to be updated each time step. Bonds might break because one of the beads in the pair became inactive. Bonds can form because two active and available beads might be close enough to form a bond. To do this, we compute the pairwise distance between all beads that are active and not currently bonded, and put the distance in a strictly upper triangular matrix. If all of those distances are larger than the barrier then no more bonds need to be created. Otherwise the smallest value in this matrix means that there will be a dynamic bond formed between the beads corresponding to the row and column of this entry. The bond is made and the corresponding row and column is removed from the upper triangular matrix since a bead can at most dynamically link with one other. This process is repeated until the smallest distance is above the barrier distance. In the end this matrix does not have to be empty, since the number of available beads might be odd and there are often lone beads that are too far away from the other unattached beads. We also note that, in the case of dynamic looping, it is likely that two beads will be more than the maximum loop distance apart from one another when they first join to form a loop, since the barrier value is larger than the maximum loop distance value. Thus, at every subsequent time step after the initial loop connection is made, the distance between the two beads is halved

until that distance is less than the maximum loop distance value.

### 6.2.6 Microscope simulator

We have created two distinct pipelines —one for experimental images, and the other for simulated bead position data. Simulated data goes through a program, which we call the microscope simulator, to create images that can be post-processed the same way as images that come from experiments. This allows us to analyze the entire simulation run. This is done by using and extending DataTank, an object-oriented programming environment offering large data set support and visualization tools made by Adalsteinsson http://www.visualdatatools.com/DataTank. The metric we use is nucleolar area.

(a). *Experimental input.* Images from light fluorescence microscopy are input into the DataTank script. The script takes the max projection of the z-stack to produce a single image for a given time. We use the Otsu thresholding method (OTSU 1979) to determine the appropriate nucleolar contour threshold for each frame. This gives us a polygonal representation of the boundary, and allows us to compute the area for the cell at each frame and compute statistics from the collection of images.

(b). *Simulated input.* When analyzing simulated data, the process is similar, but we must first transform the simulated data into an image. This is the program we refer to as the Microscope simulator and was created as a module in DataTank. This module contains information about how a single point of light spreads and is visualized in 3D; namely, it utilizes a point spread function (PSF) to generate a simulated 3D image. This PSF was captured using the same microscope that the Bloom Lab uses in their experimental work. In the DataTank script, the simulator module takes as input a three-dimensional stack of images coming from a single fluorescent bead along with the x,y,z coordinates from the simulation. The module goes over each simulated data point and shifts the three dimensional input image to that center and adds together all of those images to form a three dimensional image. At this point it computes a max projection to create a single two-dimensional image. This is done for every frame of the simulation. After this, we use the same process as for the experimental images, except that the threshold is kept the same for all of the images, at 68% of the maximum light intensity.

## 6.3 Results

### 6.3.1 The nucleolus is a heterogeneous and dynamic nuclear sub-domain.

In the yeast cell, 1.8 Mb of a tandem repeat of 9.1 kb (about 180 repeats) resides on the right arm of chromosome XII. This repeat contains the genes for 5S, 5.8S, 18S and 25S rRNA and several transcribed and non-transcribed spacer regions. The repeat array occupies a specific nuclear domain throughout the cell cycle in living yeast. To visualize the nucleolus we fused several protein components to GFP including Cdc14 (protein phosphatase), Cbf5 (pseudouridine synthase), and a multiply integrated array of lacO/lacI-GFP [101]. Nucleolar protein fusions and integrated lacO arrays occupy a distinct region of the nucleus adjacent to the nuclear envelope and typically opposed to the spindle pole body. We have developed analytical tools to quantitate changes in nucleolar morphology and distribution (Methods). Our analysis shows that the distribution of areas occupied by a nucleolar protein (Cdc14-GFP) display a lognormal distribution in wild-type cells in G1 of the cell cycle (Figure 6.1A & Figure 6.2). For each of these distributions we calculated the expected value (E[X]) and coefficient of variation (CV) as measures of central tendency and dispersion of the data (Figure 6.1B).

### 6.3.2 Nucleolar position and morphology is not dependent on rDNA continuity along a single chromosome.

In yeast cells, rDNA can be experimentally manipulated through chromosome translocation to split the locus among different chromosomes. We have utilized a strain where the rDNA has been split between chromosome XII and III (Figure 6.2). Through an engineered chromosome translocation, Mikus and Petes [99] have generated a yeast strain where a translocation between two chromosomes results in splitting the rDNA. Remarkably, the split nucleolus is indistinguishable from the nucleolus in wild-type cells (Figure 6.2), similar to that found in cells with rDNA at ectopic sites (34). Figure 6.1A-B show that the area within the nucleolus, as well as the quantitative analysis of central tendency and dispersion, are comparable between the two biological situations. This finding provides the impetus to explore models that account for the biological merger of rDNA when the gene arrays lie on disparate chromosomes.

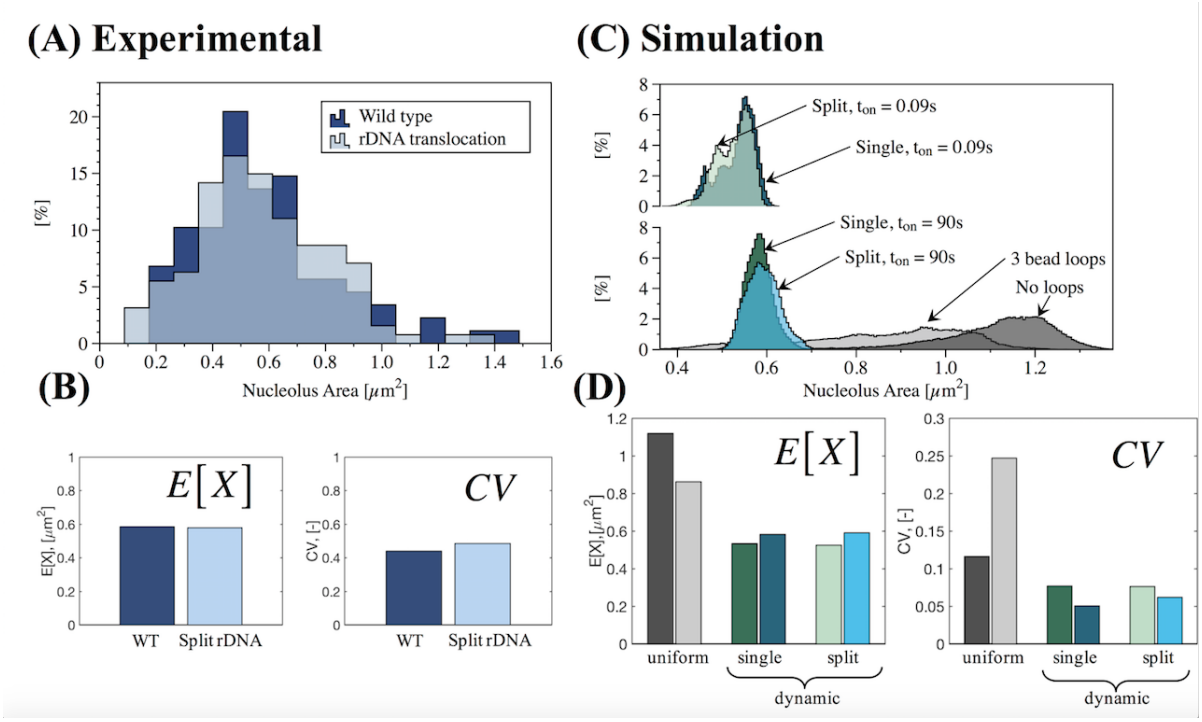


Figure 6.1: *Quantitative analysis of nucleolus area in wild-type and rDNA translocation strains.* **(A)** Histogram of experimental results for wild type (WT, 88 cells) and rDNA translocation (127 cells). Y-axis is percentage in each bin. **(B)** Measures of central tendency (Expected value, E[X]) and dispersion (Geometric coefficient of variation, CV) obtained from best fit to lognormal distributions to experimental data in (A). **(C)** Histogram of simulations results for no crosslinks, uniform crosslinks and dynamic crosslinks, with $K_{eq} = 9$ and two different (slow and fast) binding times $t_{on}$ for single and split nucleolus. Y-axis is percentage in each bin. **(D)** Measures of central tendency (E[X]) and dispersion (CV) obtained from best fit to lognormal distributions of simulated data in (C).

### 6.3.3 Implementation of crosslinks through molecular springs to simulate the nucleolus

To directly compare bead-spring models to experimental data we convolved beads that occupy the position comparable to the rDNA repeats in chromosome XII with a point-spread function from a fluorescence microscope used to image the nucleolus (Methods). The strategy was implemented in DataTank (http://www.visualdatatools.com). Analyses of nucleolus area from simulations also display a lognormal distribution (Figure 6.1C). Figure 6.1D shows the resulting measures of central tendency and dispersion. Examples of microscope-simulator images and area thresholds are shown in Figure 6.3, for different modeling assumptions.

Structural maintenance of chromosomes (SMC) proteins, including condensin (SMC2,4) and cohesin (SMC1,3), are major structural components of eukaryotic chromosomes and are enriched

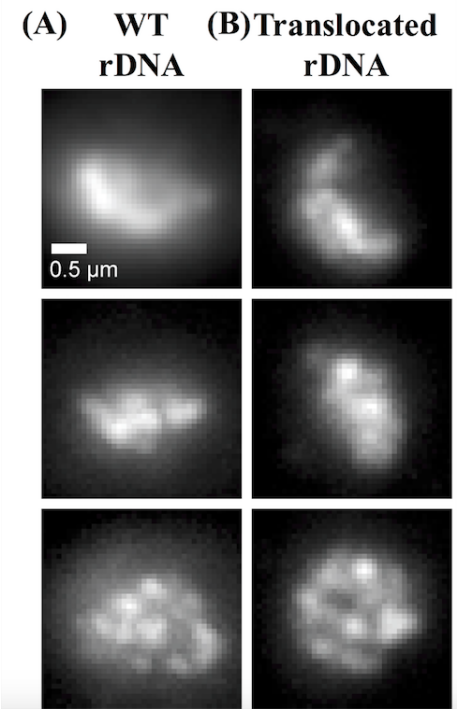


Figure 6.2: *Experimental results for CDC14-GFP of intact and translocated rDNA.*Maximum intensity projections of CDC14-GFP in strains with **(A)** WT (DCY102.1) and **B** translocated (DCY1017.2) rDNA.

in the nucleolus throughout phylogeny. The proteins comprise small rings (25-50 nm dia.) that function to build chromatin loops, hold sister chromatids together, and are the basis for the 3D organization of the chromatin fiber into TADs ([2]). To simulate rings that physically aggregate remote chromatin domains, we implemented molecular springs in the model that bridge and hold pairs of beads for prescribed statistical durations. This spring force obeys a worm-like chain law, based on polymer models of random coils [14, 15] and is the same as the one implemented in the springs between neighboring beads of the bead-spring chromatin chains. These crosslinking springs are proxies for a ∼45 nm diameter condensin ring complex. In addition, the crosslinking springs are 50x stronger than those connecting neighboring beads. In this way, crosslinking springs represent relatively stiff protein complexes (small persistence length, $L_p$) while springs between neighboring beads represent chromatin tension blobs (large $L_p$).

To explore whether introducing crosslinking via simple springs is sufficient to segregate sub-domains in the genome, we implemented simulations with no crosslinks (Figure 6.4A), permanent crosslinks formed by springs located at fixed locations (uniform crosslinks —Figure 6.4B), and

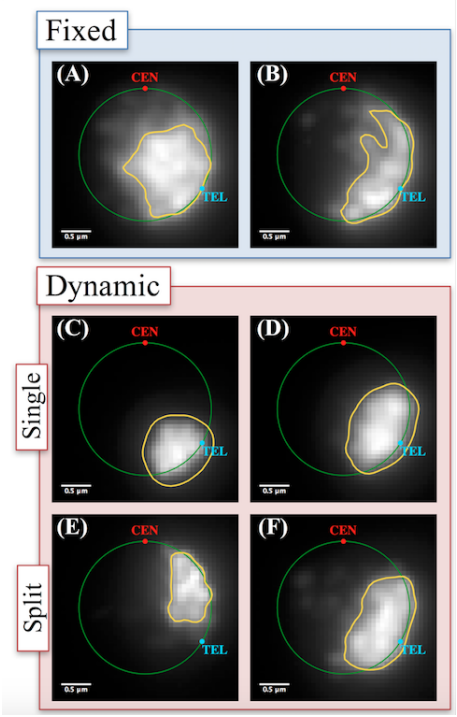


Figure 6.3: *Qualitative analysis of nucleolus area.* Simulation results converted to microscope images (Methods). **(A)** Without crosslinks, **(B)** uniform, fixed cross-links, **(C)** dynamic crosslinks for single nucleolus with $t_{on} = 0.09$ s, **(D)** dynamic crosslinks for single nucleolus with $t_{on} = 90$ s, **(E)** dynamic crosslinks for split nucleolus with $t_{on} = 0.09$ s, **(F)** dynamic crosslinks for split nucleolus with $t_{on} = 90$ s.

crosslinks formed by springs whose locations are changed dynamically (dynamic crosslinks —Figure 6.4C), as described in the Methods section. In Figure 6.4, for uniform crosslinks we impose chain configurations with 3-bead loops: two looping beads separated by one bead, as described in Methods. For dynamic crosslinks we assume stochastic dynamic crosslink formation within the nucleolus, with all beads within the nucleolus available for binding-unbinding kinetics, and assume no crosslinks outside the nucleolus (see below for inclusion of crosslinks outside the nucleolus). Several noteworthy differences distinguish the dynamic crosslinks case from the uniform, fixed-loops case. For instance, dynamic crosslinking allows for the possibility of a transient, fluctuating "loops within loops" structure. We tune dynamic crosslinking through the use of six parameters, defined in detail in Methods. The parameters are the distance (90 nm) between which two beads must fall in order to be eligible to form a crosslink, maximum separation (45 nm) that beads cannot breach when they are bound by a dynamic crosslinking spring, spring force that makes the spring stiffer at shorter length scales (50x), and mean $t_{on}$, mean $t_{off}$, standard deviation on, and standard deviation off. The

last four parameters are in seconds and they define the distribution of times at which beads can turn “on" and “off", i.e., are eligible for binding with other active beads. In the model, we have varied the values for ton and toff, keeping the equilibrium rate constant, $K_{eq} = t_{on}/t_{off} = 9$, and have set the standard deviations of both distributions equal to 20% of the mean value for all simulations.

### 6.3.4 Crosslinks in the model promote compaction and spatial segregation while dynamic cross linking promotes connectivity and substructure within the nucleolus

Unlike single molecule analysis or imaging, simulation enables us to discern how the dynamics of entropic chain fluctuations, together with permanent versus dynamic crosslinking, influence the strength of interactions within the nucleolus. Figure 6.4 shows contact maps consist of average bead-bead distances over a 15-minute period. In the absence of crosslinking, the maps in Figure 6.4A reveal a fairly uniform distribution of bead-bead distances across the genome. Although the average distance is about 1 micron, there is heterogeneity across the genome and different simulations show different distributions (top-bottom Figure 6.4). Upon introducing static 3-bead looping, Figure 6.4B shows that the major change is a decrease in average bead-bead distances, from 1 micron to 600 nm, in the region where the looping was implemented (denoted as the rDNA containing 1.8Mb of ribosomal DNA gene repeats). However the remaining beads exhibit similar contact statistics as in the non-crosslinking configuration, except for their interactions with beads belonging to the nucleolus region, which show an increase in bead separation (lighter “cross" in the contact maps), indicating enhanced sequestration of the nucleolus. In contrast, fast dynamic crosslinking has a major impact on the behavior of the rDNA cluster (Figure 6.4C). There is a predominant compaction of bead-bead distance in the dynamic case, down to 300-400 nm among a subset of beads (further explored below) within the nucleolar region and less dramatic yet still much closer bead-bead proximity within the rest of the nucleolus relative to the rest of the genome. Furthermore, the distance between the rDNA beads and the rest of the genome increases, corresponding to more enhanced sequestration of the nucleolus. This implies that the rDNA beads are interacting more frequently with themselves, and therefore less frequently with the remainder of the genome. This is notable in the perpendicular cross, lighter regions vertical and horizontal from the rDNA, signifying the rDNA bead interactions with unlinked chromosomes. This somewhat non-intuitive impact of weak binding relative to tight binding kinetics has been explored in other biological contexts (Newby et al., Nature Communications, revisions submitted), both experimentally and theoretically.

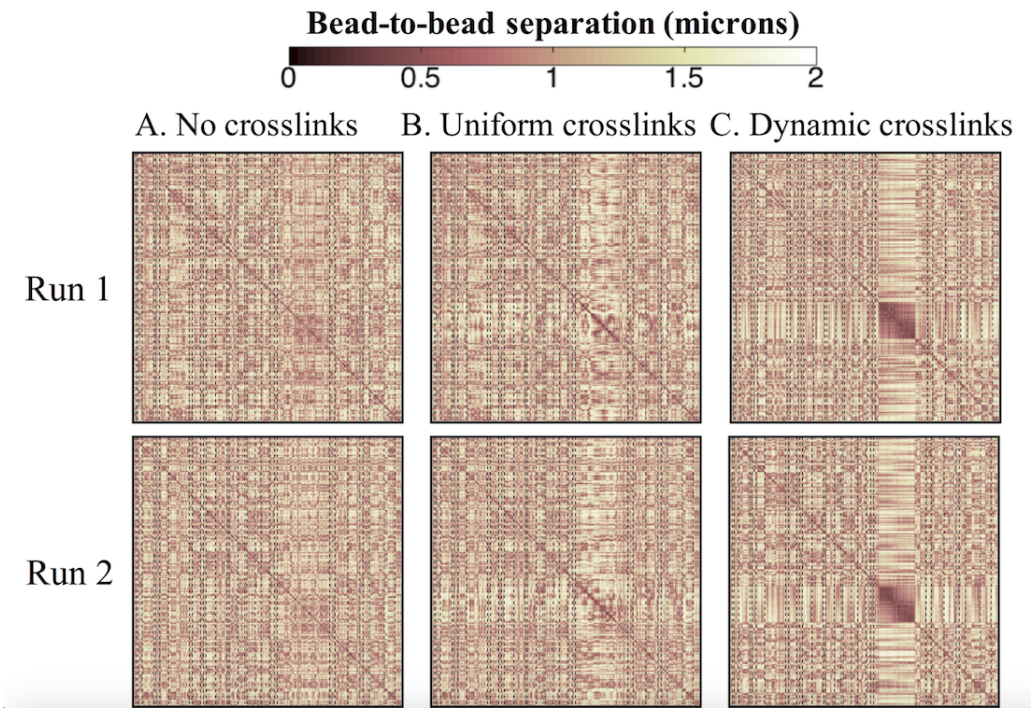


Figure 6.4: *Contact Maps depicting the average 3D bead-to-bead distance over 15 minutes.* Each plot corresponds to a different data set; however data sets in the same row (top or bottom) share the same random seed and only differ in the type of crosslinking imposed in the nucleolus. Fast dynamic crosslinking dramatically increases nucleolus connectivity, whereas static crosslinking has a far lesser effect. **(A)** No crosslinks, **(B)** 3-bead fixed loops, **(C)** Fast dynamic crosslinking ($K_{eq} = t_{on}/t_{off} = 9$, $t_{on} = 0.09$ s).

### 6.3.5 Binding kinetics timescales tune segregation and heterogeneity of the nucleolus

In the previous section we showed that fast dynamic crosslinking enhances interactions within the active binding domain, leading to enhanced compaction and segregation of the nucleolus relative to permanent or no crosslinking. Here we explore the critical biological parameters involved in dynamic crosslinking, the on and off timescales of binding kinetics between nucleolar beads via the molecular springs, and their influence on nucleolar morphology and nucleus-wide genome organization. We simulated 3 dynamic kinetic regimes, each with $K_{eq} = t_{on}/t_{off} = 9$, spanning fast to slow (weak to tight) binding with $t_{on} = 0.09$, 0.9, and 90 seconds, shown from left to right in Figure 6.5. These bead-bead contact maps reveal that faster binding kinetics induces closer bead-bead contacts (progressively darker shades) within the nucleolus. Further, upon closer inspection of Figure 6.5 left column, fast kinetics induces heterogeneity of bead-bead contacts within the nucleolus, indicative of structure-within-structure in nucleolus morphology, which homogenizes with slower binding (left to

right in Figure 6.5).

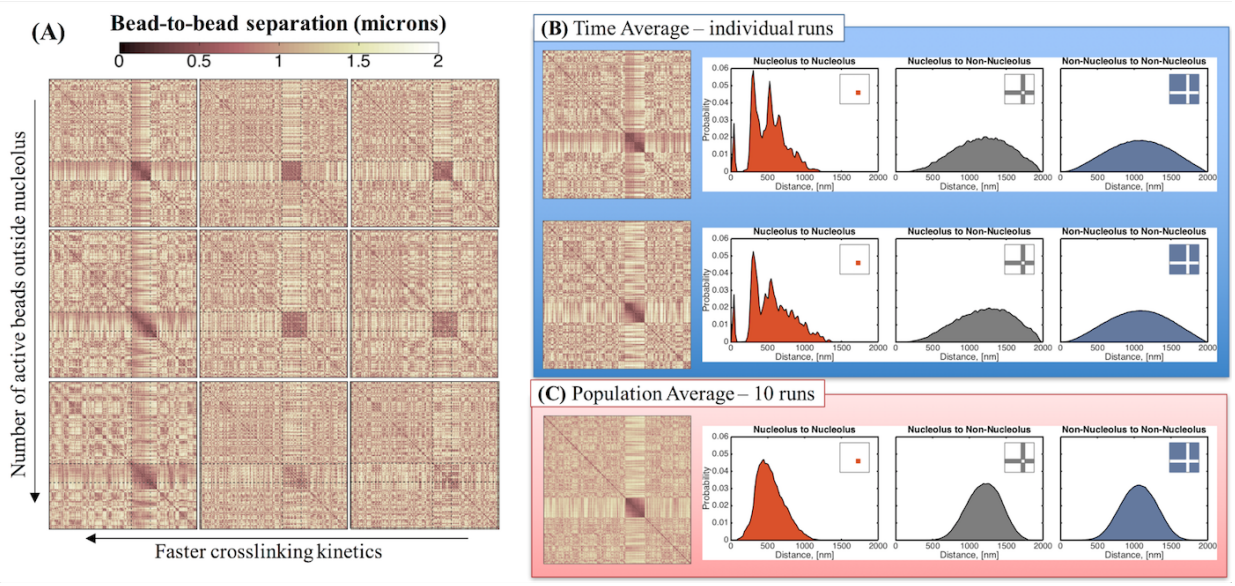


Figure 6.5: *Nucleolus and genome-wide interactions as functions of crosslinking parameters.* **(A)** Faster kinetics (lower $t_{on}$) result in more interactions within the nucleolus. From left to right $t_{on}$ = 0.09, 0.9, 90 s. Dynamics outside the nucleolus are controlled by the number of beads that are actively crosslinking. From top to bottom: all beads outside the nucleolus are inactive, every 10th bead is active and every 3rd bead is active. All contact maps correspond to an average over 15 min for a single run. All data shown are simulated using the same random noise. (B) Distribution of bead-to-bead distances for dynamic crosslinking with $K_{eq}$=9 and $t_{on}$=0.09s. Substructures within the nucleolus are formed by groups of beads that interact more frequently, leading to the darker regions in the contact maps and the peaks in the distribution functions. Although dynamic, these substructures persist over time for a given run (see Sup. Movie 1) and vary from run to run. (C) The substructures are lost in the population averages resulting in a more uniform distribution of bead-to-bead distances within the nucleolus.

### 6.3.6 Dynamic crosslinking is a segregation mechanism within the nucleus.

By allowing a preferential interaction of beads in a specific part of the genome, a crosslink-induced phase separation occurs within the nucleus. In our simulations, one phase consists of beads in the nucleolus and this region is characterized by being denser than the rest of the nucleus. This compaction comes from enhanced interactions among these beads together with reduced interactions with other beads in the genome. To take into account different binding kinetics *outside* the nucleolus we define the number of 'actively crosslinking' beads outside. Figure 6.5 (top to bottom) shows that as this number decreases (less externally active beads) the nucleolus becomes more compact. As more beads outside the nucleolus participate in binding kinetics the identity of the nucleolus is weakened: less separation between bead-bead contacts within and external, and less compaction and

segregation of the nucleolus (Fig. 6.12).

Figure 6.5A shows that substructures are formed within the nucleolus exclusively in the fast kinetics case, $t_{on}$ = 0.09 s. These substructures are formed by clusters of nucleolar beads that reside closer and interact more frequently with members of their cluster, as suggested by the nonuniformity in the contact maps (left column of Figure 6.4). To amplify this feature, in Figure 6.5B we show histograms of bead-bead distances for two individual simulations and for a population average over 10 individual simulations, all with $t_{on}$ = 0.09 s and identical initial data. Note the remarkable similarity of histograms for both individual runs shown (indeed for all 10 individual runs), with striking peaks within the nucleolus bead-bead distances yet with smooth proximity distributions of the population average and all bead-bead statistics involving extra-nucleolar beads. The first peak at ∼10 nm identifies subsets of beads tightly and persistently packed (forming clusters), and the second peak at ∼300 nm implies a robust separation between the persistent clusters. To visualize this implied morphology, Figure 6.6 is a 3D snapshot of all nucleolus beads for both individual runs, revealing ∼25 clusters of 10-25 beads each. Distances within these clusters correspond to the leftmost peak in the graphs for individual runs in Figure 6.5B, while inter-cluster distances correspond to the second and third peaks in the same histograms. Although dynamic, these substructures persist in time for each individual run and vary from run to run (see Figure6.16). We note that these peaks are not present in runs with slower dynamics (see Fig. 6.11 for $t_{on}$ = 90 s and Figure 6.17). Furthermore, these substructures are lost in the population averages that show uniformly distributed bead-to-bead distances within the nucleolus as shown in Figure 6.5C). That these peaks do not survive population averaging, despite the remarkable similarity in the cluster morphology of both runs in Figure 6.5A (indeed all ten runs), is due to the fact that nucleolar beads are randomly distributed in the clusters for each run, so the tight associations of beads in any one cluster for any individual realization are different in every other realization. This result indicates that population averages might obscure a rich set of dynamic structures, which can only be appreciated at the single, live cell level.

Dynamic crosslinking within the nucleolus imposes itself on the remainder of the chromosomes as seen in the contact maps. By taking out a large segment of potential interactions (rDNA), contacts between the remainder of the genome occur with greater frequency. This indicates that despite a large block of the chromosome segregated in its own territory, the ability of remaining chromosomes to independently interact, e.g., by different SMC proteins, is not impeded. There is a small increase

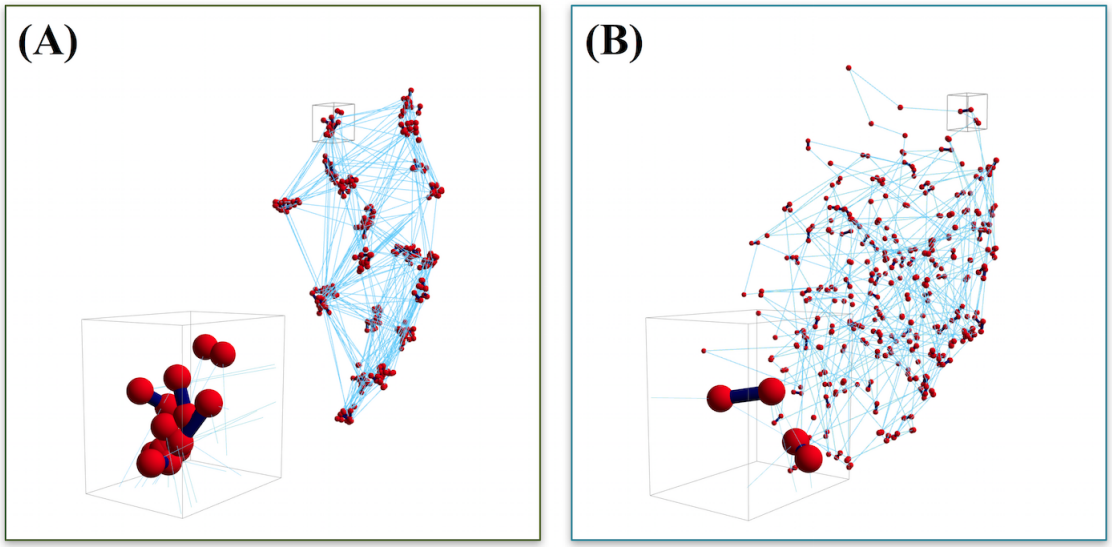


Figure 6.6: *Snapshots of 3D nucleolus simulations.* Bead distributions for (A) $t_{on}$ = 0.09 s and (B) $t_{on}$ = 90 s. Red symbols represent bead positions, dark blue segments represent transient crosslinks between beads both inter- and intra-chain, light blue lines represent intra-chain neighboring bead connections. Inserts in A and B are blow-ups of small volumes around bead clusters.

in the average distance between nucleolus beads and beads outside the nucleolus (Figure 6.5B-S3). These results suggest that dynamic crosslinking, although segregating nucleolar beads, still allows them to explore the nuclear state. There is physiological evidence for nucleolar DNA exiting the nucleus under conditions of DNA damage [102]. The simulation indicates that rDNA is constantly able to “explore" the nucleus, and perhaps DNA damage takes advantage of this naturally occurring process.

We can divide our simulations along 3 main conditions: 1) static/uniform vs. dynamic crosslinking, 2) magnitude of dynamic binding-unbinding timescales, and 3) percentage of active binding domains throughout the nucleus. A summary of these conditions is shown in Fig. 6.9 where we compare the statistics of bead-bead separation using best fits to lognormal distributions and their measures of central tendency (expected value, E[X]) and dispersion with respect to the mean (geometric coefficient of variation, CV). These statistics allow us to quantitate the degree of nucleolar structure. Imposing uniform looping in the nucleolus has little effect in the average bead-bead distance for beads both inside and outside of the nucleolus. Dynamic crosslinking, on the other hand, reduces the average distance between intra-nucleolar beads, while an increase of active beads outside the nucleolus has the opposite effect, i.e., increasing the average intranucleolar distance. As more beads outside the

nucleolus participate in the crosslinking dynamics, the average separation between nucleolus beads increases (Fig. 6.9). This effect is more pronounced in the slow kinetics ($t_{on}$ = 90 s), which shows bead-bead distances that are comparable with those found in the case where crosslinking was absent throughout the whole nucleus. With more active beads outside the nucleolus, the nucleolar beads are more spread out, indicating that the area increases as nucleolar beads spend more time outside the nucleolus.

### 6.3.7 Split nucleolus: Co-linearity is not required to form a single domain.

A critical prediction of the model is that the coalescence of beads into a distinct phase is not dependent on their co-linearity, (i.e., contiguous on the same chromosome arm) rather it is a consequence of binding kinetics. To test this in the model we simulate a split nucleolus by dividing "nucleolar beads" into two different chains as described in the Methods section. The objective of this numerical experiment is to assess whether dynamic crosslinking suffices to create different regions within the nucleus or if co-linearity of such beads is required. Our results indicate that the split chains occupy a single region within the nucleus as shown in the contact maps in Figure 6.7 & 6.13. In addition, the distance between all nucleolar beads are comparable, consistent with them being in close proximity (Figure 6.6B). These results suggest that the main driving mechanism in the formation of nucleolar-like regions is a "preferential crosslinking" within that region. In our case, this preferential crosslinking is controlled by the on/off binding dynamics and the number of active beads outside the nucleolus.

Remarkably, *the dynamics observed in our single nucleolus simulations are mirrored by those in the split nucleolus.* Namely, faster binding kinetics result in a decrease in the nucleolar area (compare Figs. 6.9 and 6.14) arising from beads within the nucleolus being closer to each other and interacting less with beads outside the nucleolus (Figs. 6.10 & 6.15 ). The introduction of crosslinking dynamics outside the nucleolus results in an increase in nucleolar area and increased interactions of nucleolus beads with non-nucleolus beads. Finally, for slow crosslinking dynamics and large number of active beads outside the nucleolus, the resulting distribution of bead-bead distances shows a complete mixing throughout the nucleus; evident in the contact maps shown in Figure 6.7. We note that this "mixing" effect is more pronounced in the split nucleolus case than in the single nucleolus (compare bottom right of Figure 6.5A & Figure 6.7A).

Results discussed above point toward a different understanding of the formation of nucleolar-

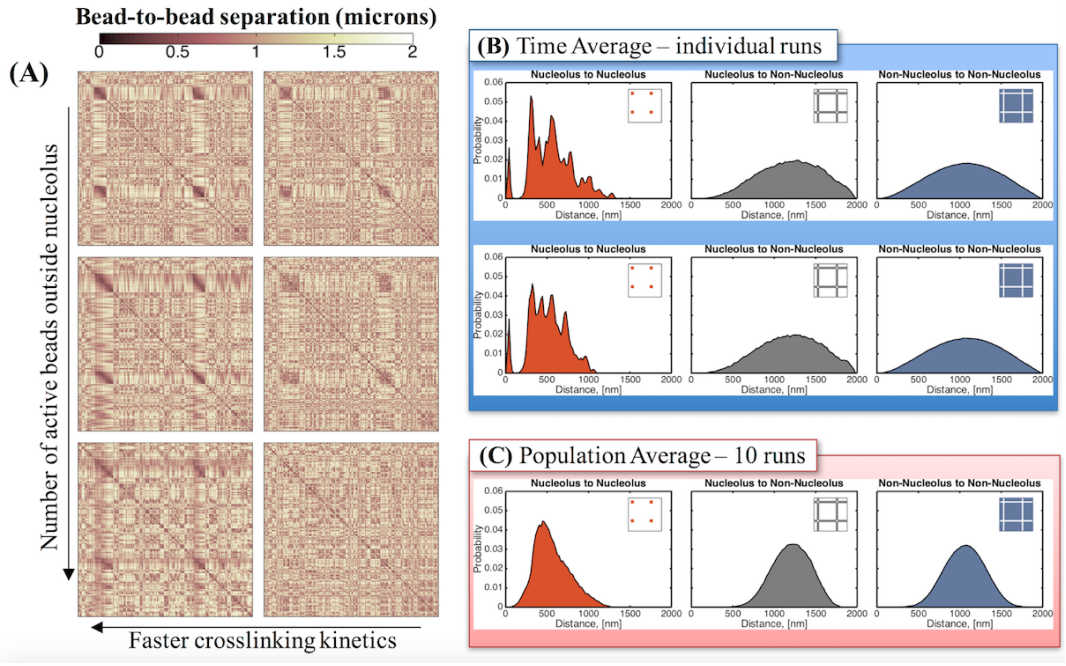


Figure 6.7: *Distribution of bead-to-bead distances for split nucleolus.* (A) Contact maps show that there is a marked increased in bead interactions within the nucleolus as a function of decreasing the on/off times of binding (left to right). Faster kinetics result in more interactions. Left column $t_{on}$=0.09 s, right column $t_{on}$=90 s. Active beads outside the nucleolus are introduced to account for binding interactions throughout the nucleus. In the non-nucleolar chromatin, interactive beads are inactive outside the nucleolus (stride zero —top row), every 10th bead is active (stride 10) and every 3rd bead is active (stride 3 —bottom row). (B) Substructures within the nucleolus are formed by clusters of beads that are closer, interact more frequently, and maintain some separation between clusters. These fluctuating substructures create darker regions in the contact maps and peaks in the bead-bead proximity histograms within the nucleolus. Although dynamic, these substructures persist over time for a given run (see Fig. 6.18) and are robust from run to run although the nucleolar bead cluster assignments are random. (C) The substructures are lost in the population averages, even though the cluster morphology is robust, because of random cluster assignments, resulting in a more uniform distribution of bead-to-bead distances within the nucleolus.

type regions in the nucleus. As previously discussed, assuming different polymeric properties for the nucleolar beads results in segregation and slower dynamics of the nucleolus, consistent with experimental results [53]. However, this assumption alone cannot describe the results observed experimentally for a split nucleolus. Our approach of distinguishing the nucleolar beads not by their polymeric properties but by their affinity to certain proteins (binding kinetics), recapitulates segregation, slower dynamics *and* split nucleolus observations. We note, however, that *in vivo* the most probable scenario is a combination of different non-specific (entropy, polymeric properties, macromolecular crowding, etc.) and specific interactions (e.g., crosslinking).

## 6.4 Discussion

The numerical study presented here suggests that preferential crosslinking in different regions of the genome results in the self-segregation, compaction, and morphology of these regions, with each feature tunable by the timescales of crosslinking. Using bead-spring polymer models, we have shown that the only necessary condition to “create" a nucleolus is that certain beads (which we labeled “nucleolar beads") have binding affinity to each other in a dynamic way. Whether or not the nucleolar beads belong to the same chain, has little effect in the formation of the nucleolus or its dynamics. The important parameters are their interaction kinetics and their ability to interact with non-nucleolar beads. We control the interaction kinetics via on/off binding timescale distributions and the interactions with other beads through the fraction of active beads outside the nucleolar region. We have summarized the parameter space explored in this study in Figure 6.8, where the radius of gyration of the resulting nucleolus is used as a metric for nucleolus size.

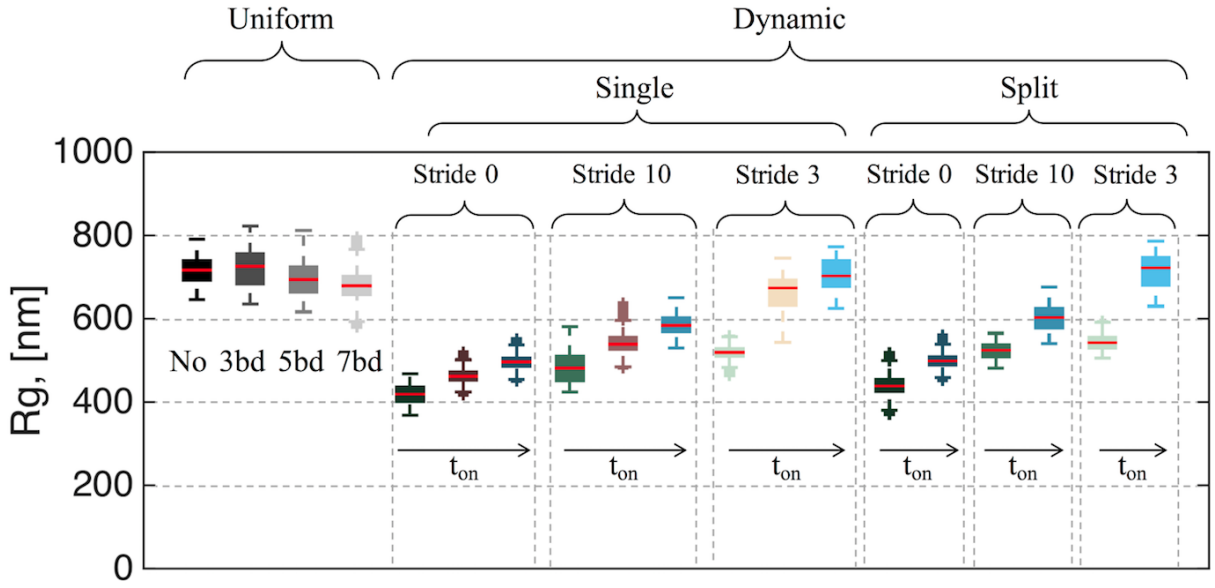


Figure 6.8: *Summary of nucleolus size with respect to crosslinking parameters.* Radius of gyration, Rg, for each of the parameter values investigated in this study. For single nucleolus dynamics the values for $t_{on}$ are 0.09, 0.9 and 90 s. For split nucleolus dynamics $t_{on}$ = 0.09 and 90 s. Increase in “on" time leads to an increase in the nucleolus size. Similarly, increase in the fraction of beads that are active outside the nucleolus (stride) increases the size of the nucleolus. For slow kinetics and comparable number of active beads inside and outside the nucleolus (light blue), there is little difference with respect to the case where no crosslinks are included (black). All dynamic crosslinking simulations have $K_{eq} = 9$.

The kinetics of the crosslinking agent relative to the motion of the substrate is a major driver of sub-nuclear organization. If the crosslinkers bind and release more rapidly than the chains can relocate, the non-intuitive consequence is that the chains explore less space. When nodes of protein/binding sites arise, they persist for extended time periods. The result is evident in Figure 6.5 as peaks and valleys in the histogram of bead-bead distances and as a 3D array of bead clusters in the snapshots of nucleolus beads in Figure 6.6, with persistence revealed in Supplemental movies. The primary peak in the fast-binding histograms represents an increased number of beads interacting within small areas, the array of clusters in Figure 6.5C. The valleys are a consequence of cluster-cluster separation statistics. Importantly, this heterogeneity is lost if one averages across a population (compare population average vs. individual runs, Figure 6.5). A second consequence of the rapid kinetics relative to chain motion is the increase in distance for a portion of beads (tail toward increasing distance). This can be seen in the increase in coefficient of variation for dynamic looping. This may have important biological consequences in terms of a mechanism for extending the DNA, thereby facilitating RNA transcription.

In previous studies we described some general principles driving nuclear organization during interphase [2]. We showed that entropic and/or other non-specific interactions dictate, to first order, the structure and dynamics of the yeast genome. We hypothesized that the role of biochemical and other specific interactions is to modulate this organization by guiding, stabilizing and sustaining cycle-dependent genomic states. One objective of the present work is to validate such hypotheses and indeed we have shown how certain specific interactions (modeled as crosslinks) drive the rearrangement into different functional phases within the nucleus. While entropic and other non-specific interactions such as macromolecular crowding [103, 104, 105, 106] and viscoelastic phase separation [107, 108] promote the association of different regions or compartments, specific interactions are ultimately responsible for shaping and maintaining such compartmentalization.

One can view the nuclear environment as a complex, cross-linked network formed by chromosomes, proteins and other macromolecular components. This cross-linked network exhibits characteristics observed in gel solutions such as subdiffusive behavior [109, 110, 1]. By assuming “preferential" crosslinking kinetics we had effectively modified the local gel properties of the network. In this view, one could consider the nucleolus formation to be driven by a gel-gel transition, where one gel phase, the nucleolus, exhibits properties corresponding to a “stiffer" gel compared to the rest of the

genome. In addition, the formation of different phases within the nucleus via crosslinks favors the formation of nuclear compartments through self-organization allowing the development of diverse dynamic structures [103], which gives the genome the ability to rearrange into different functional states according to modifications of the crosslinking mechanisms. These mechanisms can arise from the fast exchange of proteins across different regions in the genome, which has been observed in the nucleolus [111], Cajal [112] and PML bodies [113].

## 6.5 Supplementary Data

Supplementary figures and screenshots from supplementary movies are included below. Supplementary Data is also available at NAR Online.

## 6.6 Acknowledgements

Computational resources were provided by the KillDevil cluster at the University of North Carolina at Chapel Hill and the Maxwell and Boltzmann high performance computer clusters at the University of South Carolina. We would like to thank Dr. Thomas D. Petes (Duke University) for sharing the split nucleolus strain with us.

## 6.7 Funding

This work was partially supported by National Science Foundation - Division of Mathematical Sciences (DMS) [1410047 to P.A.V.], National Institutes of Health [R37 GM32238 to K.B.], National Science Foundation [DMS-1412844, DMS-1462992 to M.G.F.], National Institutes of Health [T32CA201159-01 to J.L.].

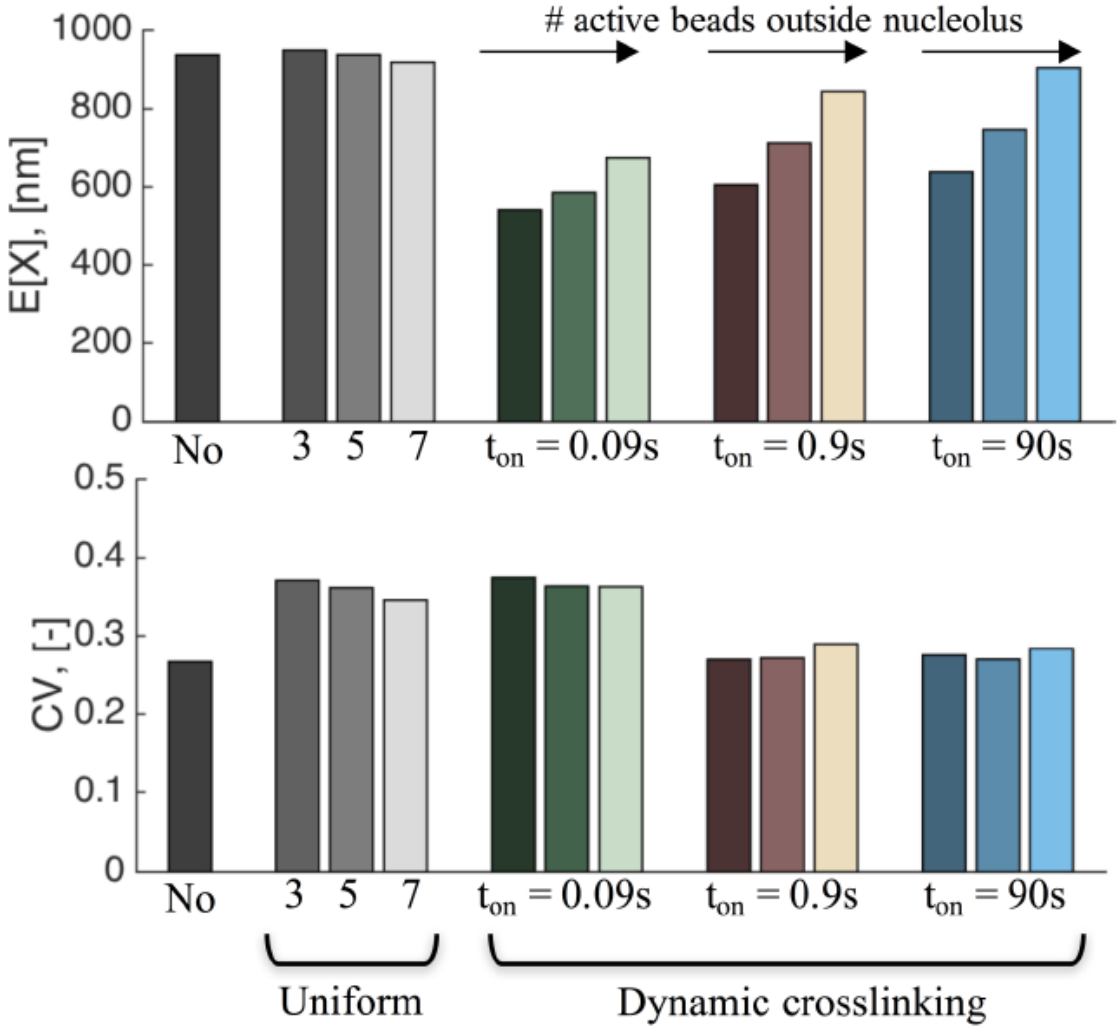


Figure 6.9: *Quantitative analysis of bead-bead distances within the nucleolus for different crosslinking parameters.* Expected value (E[X]) and geometric coefficient of variation (CV) obtained from best fit to lognormal distribution of the population averages of bead-to-bead distances. Dark gray bar corresponds to the no loops case. In the uniform loops group, 3, 5, 7 correspond, respectively, to 3-bead, 5-bead, and 7-bead loops. In the dynamic crosslink case, for a given value of $t_{on}$, the three bars correspond, respectively, to 0, every 10th, and every 3rd bead is active outside the nucleolus, i.e., from left to right, within each group of bars, there are more active beads outside the nucleolus.

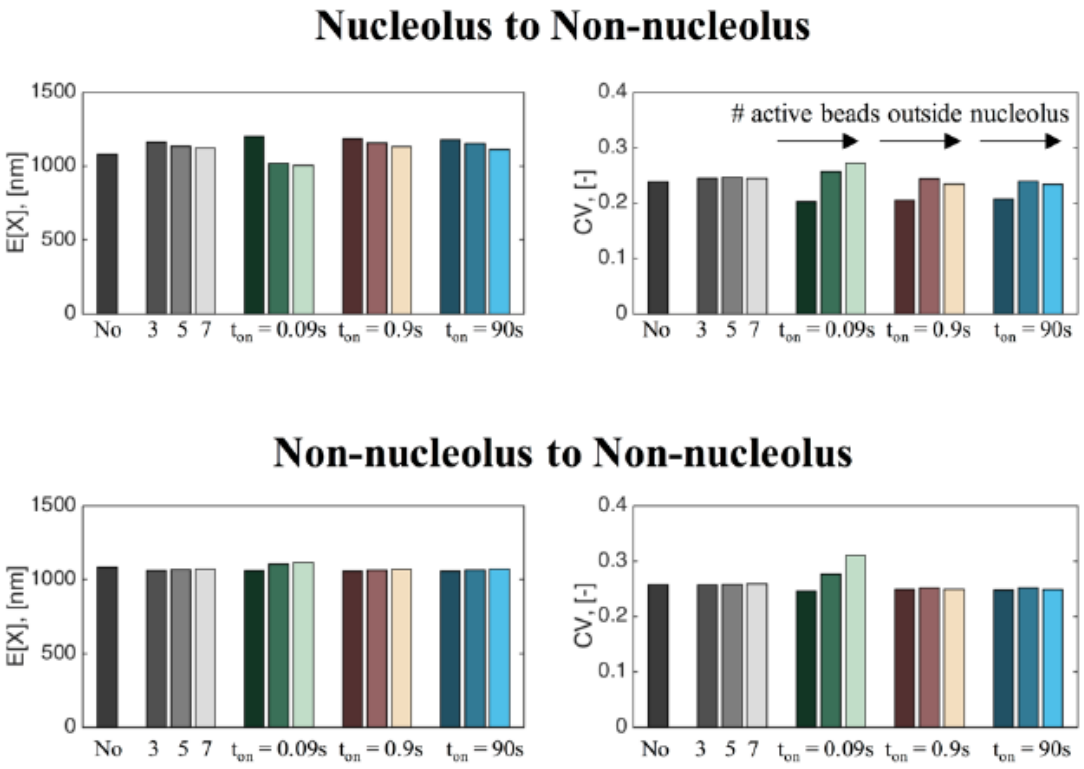


Figure 6.10: *Quantitative analysis of bead-bead distances outside the nucleolus for different crosslinking parameters.* Results for single nucleolus dynamics.

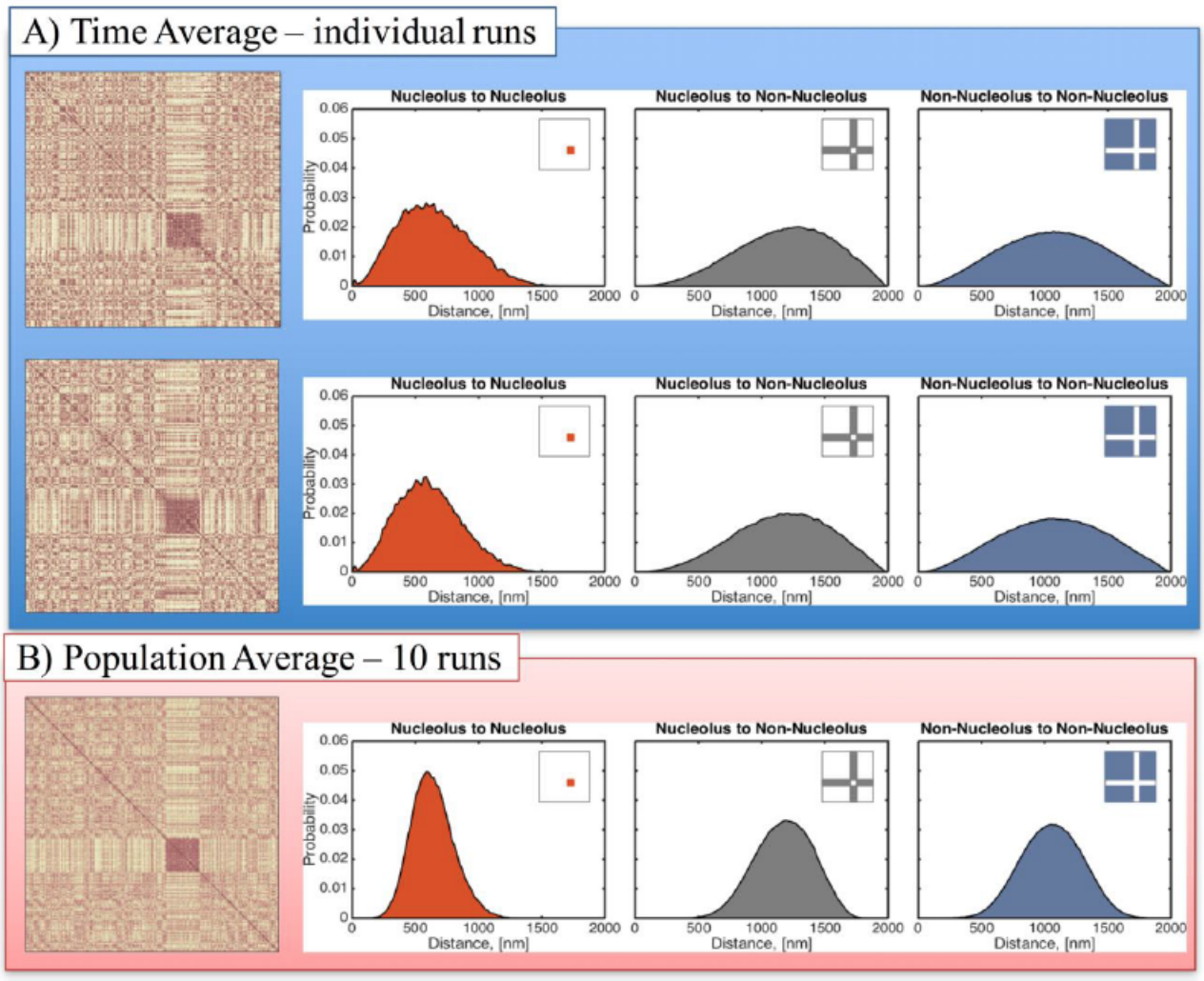


Figure 6.11: *Distribution of bead-to-bead distances for dynamic crosslinking with single nucleolus,* $K_{eq} = 9$ *and* $t_{on} = 90s$.

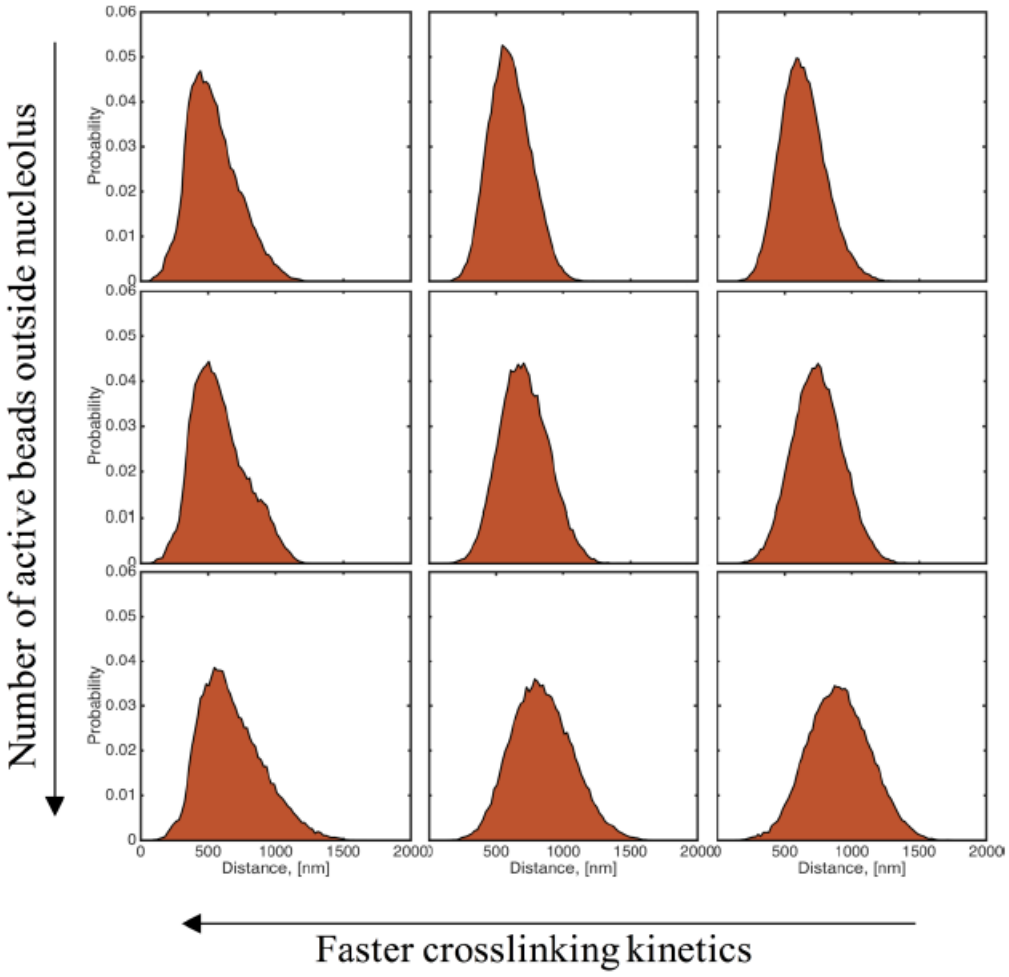


Figure 6.12: *Distribution of bead-to-bead distances for different model parameters.* Top to bottom, change n stride outside the nucleolus. Right to left decrease in $t_{on}$.

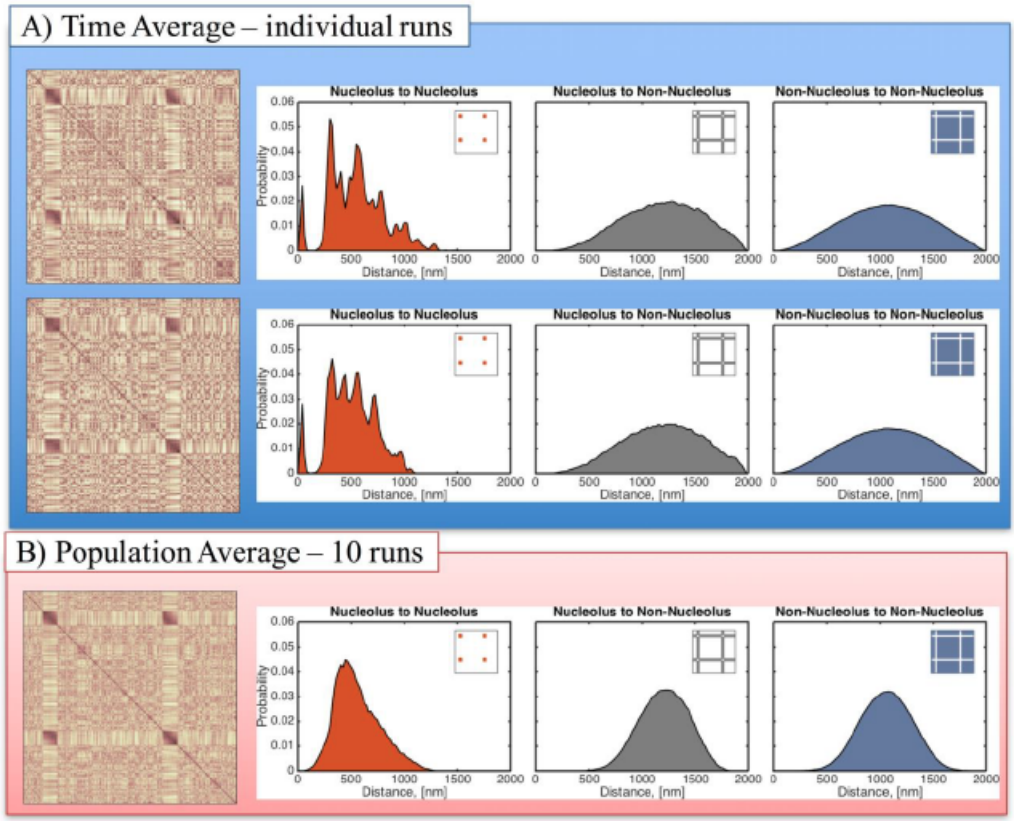


Figure 6.13: *Distribution of bead-to-bead distances for dynamic crosslinking with SPLIT nucleolus,* $K_{eq} = 9$ *and* $t_{on} = 0.09s$.

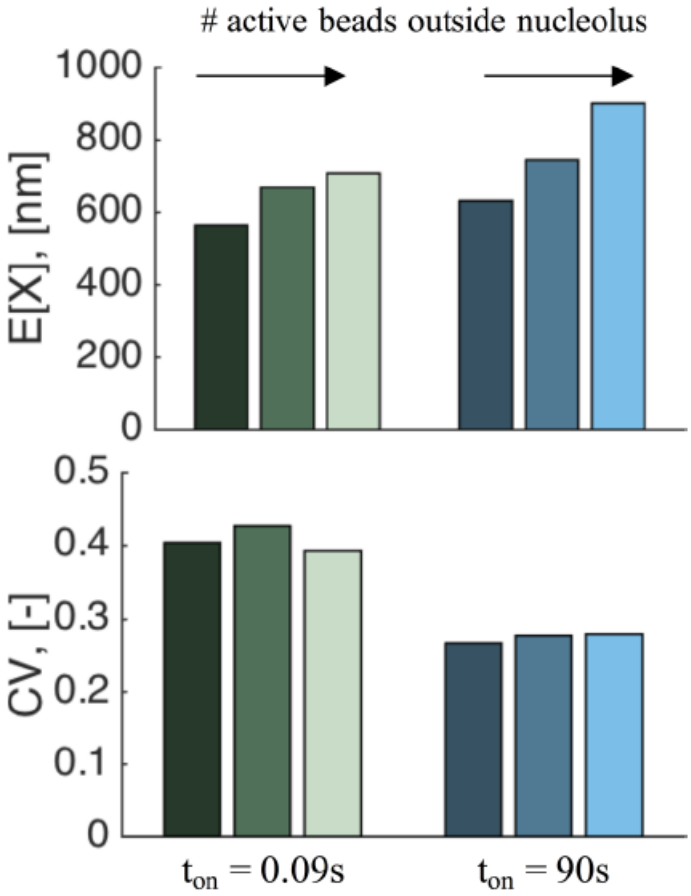


Figure 6.14: *Quantitative analysis of bead-bead distances for different cross linking parameters in SPLIT nucleolus simulations.* Expected value (E[x]) and geometric coefficient of variation (CV) obtained from best fit to lognormal distribution of the population averages of bead-to-bead distances. For a given value of $t_{on}$, the three bars correspond, respectively, to stride 0, 10, and 3 outside the nucleolus.

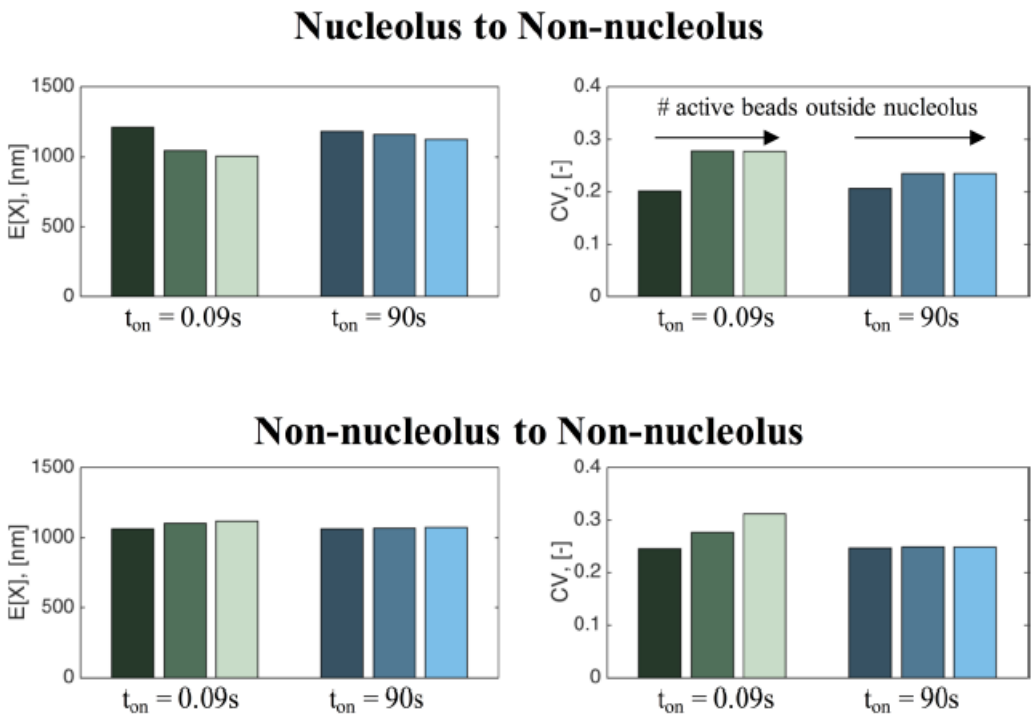


Figure 6.15: *Quantitative analysis of bead-bead distances outside the nucleolus for different crosslinking parameters.* Results for split nucleolus dynamics.

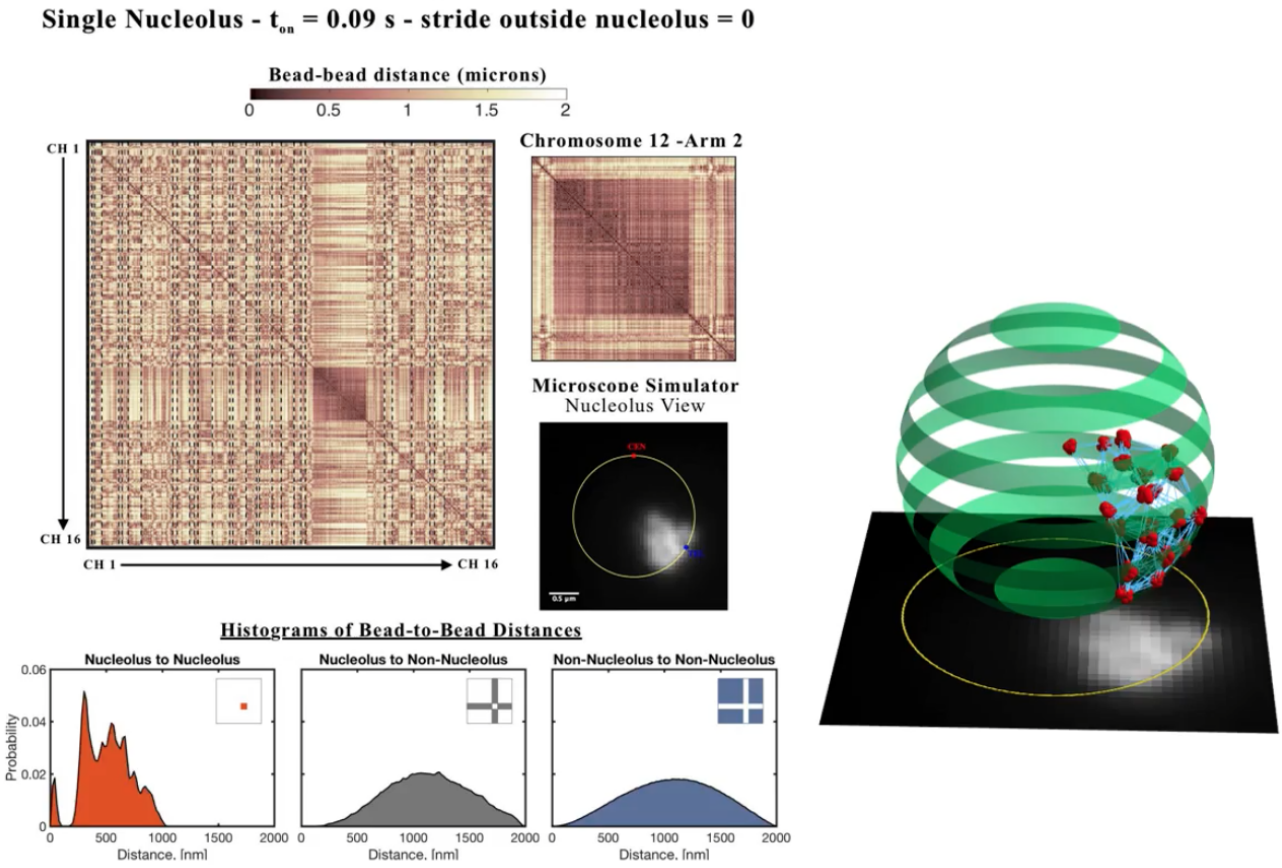


Figure 6.16: *Various forms of visualization and analysis for the single nucleolus, with dynamic looping parameter $t_{on} = 0.09s$. One frame of the movie is shown.*

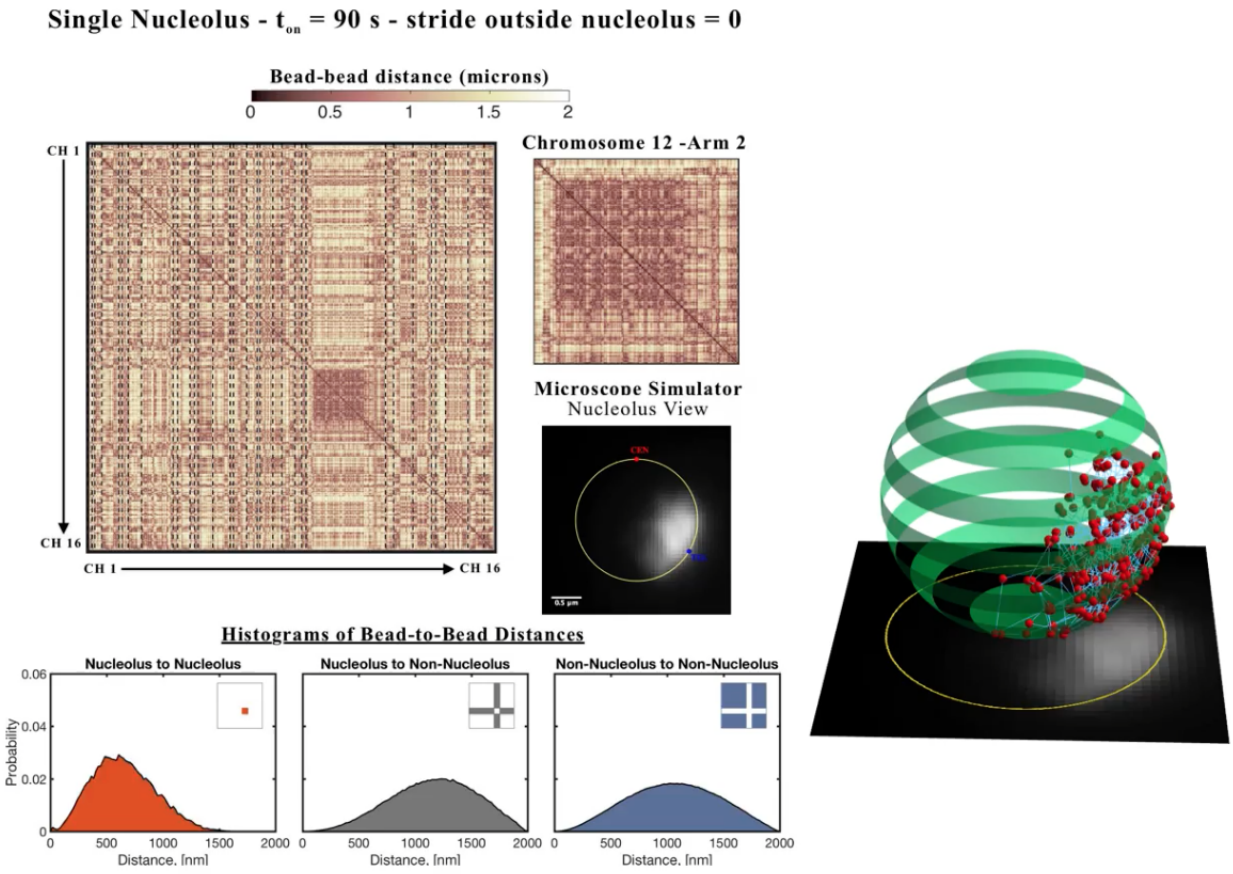


Figure 6.17: *Various forms of visualization and analysis for the single nucleolus, with dynamic looping parameter $t_{on} = 90s$. One frame of the movie is shown.*

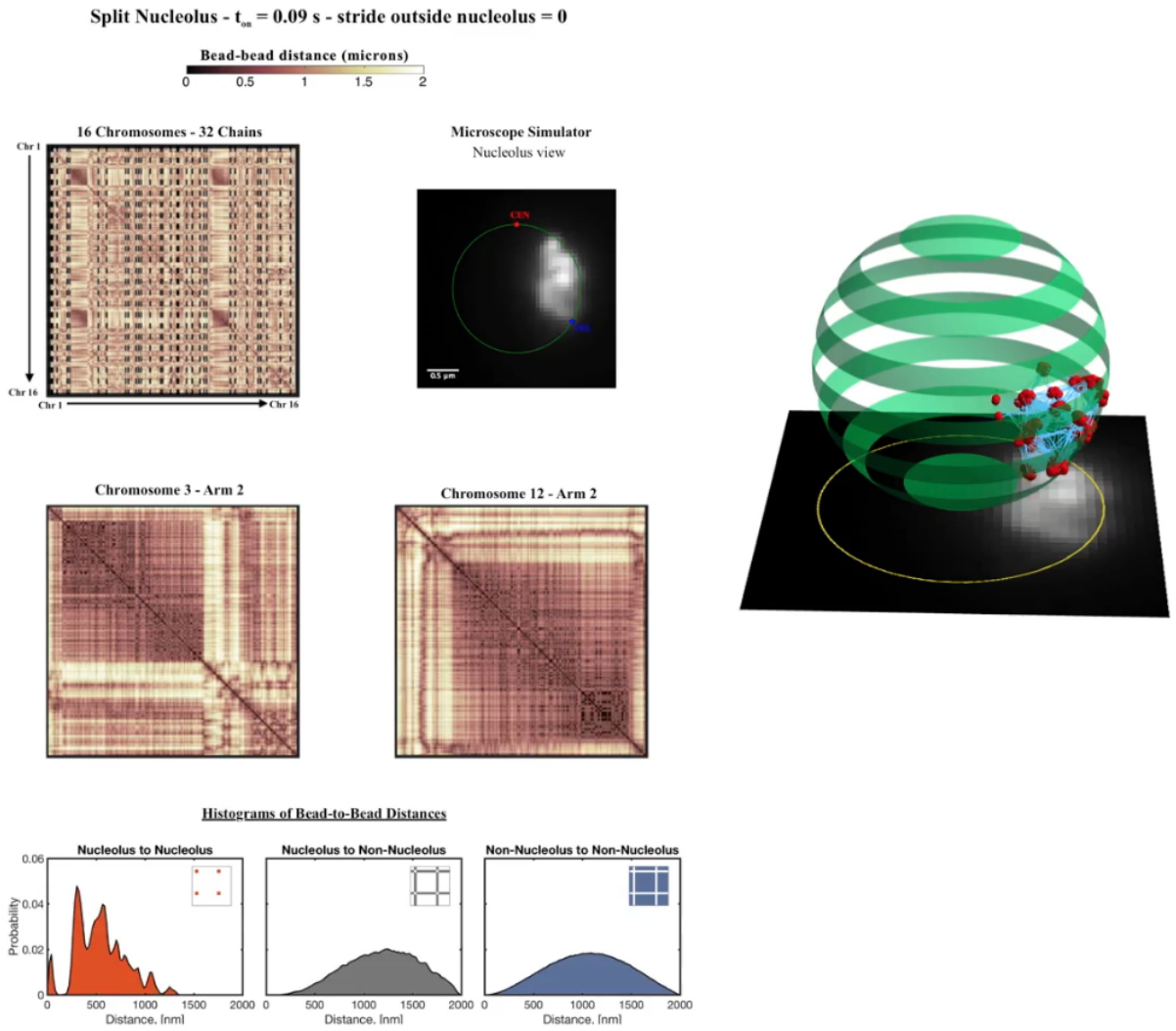


Figure 6.18: *Various forms of visualization and analysis for the split nucleolus, with dynamic looping parameter $t_{on}$ = 0.09s. One frame of the movie is shown.*

CHAPTER 7

# Additional Analyses & Interpretation of Results

## 7.1 Visualization

In mathematical modeling, appropriate visualization of a system can prove extremely valuable for advancing understanding. Model visualization may offer a means of comparison with experimental work or a quick check that the system is behaving as we expect. In the case of yeast spatiotemporal dynamics, since our model contains easily tunable parameters and discretizes to a much higher resolution than what can be seen experimentally, visualization of the simulated data provides valuable insight into the system dynamics. A combination of visualization and other forms of analyses, such as the investigation of persistence length discussed later in this chapter, has resulted in insights to system behavior that would otherwise have likely gone unnoticed or remained somewhat unclear. A central focus of my dissertation work has been on the development of visualization techniques. In my dissertation defense I will show many movies that I have made illustrating chromatin dynamics, with a particular focus on nucleolus dynamics. Selected frames from selected movies are presented in this chapter.

## 7.2 Additional nucleolus results and analyses

### 7.2.1 Microscope simulator: further work and details

As discussed in the previous chapter, in order to better assess and compare simulated and experimental data in a consistent and systematic way, we developed a microscope simulator computational program using DataTank. This program is more versatile than indicated in the previous chapter; namely, there are two types of valid experimental input. Experimental input option i corresponds to a population of cells (∼120, depending on the dataset), where each cell has 1 time value and 11 z-slices. Note that the number of z-slices corresponds to the number of different “heights" along the z-axis at which the 2D image of the cell is taken by the microscope. This is the input type used in the submitted paper [11]. Experimental input option ii corresponds to 41 time values of a single cell,

where each cell has 7 z-slices. Thus, instead of looking at a single time value for many cells, we are looking at a time-lapse of a single cell.

If we are using input option ii, then we need to account for the fact that the experimental images bleach over time, to the point where it is no longer possible to obtain an accurate area measurement for the nucleolus. Therefore, we need to use an extended version of the pipeline described in Chapter 6. At each time step, following the computation of the nucleolus area using the Otsu thresholding method previously described, we do an n-to-n ratio computation to assess whether the bleaching has progressed to the point at which the data has become too noisy to use. In other words, as we move forward in time, once the "n-to-n" ratio drops below 0.04 for a given time value, I only consider the data from the time steps before that point when computing the nucleolus area over time. I have automated the entire microscope simulator process in DataTank for both experimental input options, including but not limited to the computation of specific nucleolus thresholds over different cells or time values, nucleolus areas computation, and the calculations determining which time values should be discarded due to noise. 8 "toggle" options allow users to alter the pipeline as they see fit. Users can also easily adjust the number of desired passes through the Otsu thresholding scheme or the n-to-n ratio.

**Identifying and discarding noisy data:** The goal is to obtain the background-corrected signal with respect to integrated intensity. This is a way of dealing with the background noise, to be used instead of superimposing an arbitrarily-created bleaching function over the image using DataTank. Josh Lawrimore suggested this procedure, and I wrote and implemented it in DataTank. In the steps below, we let Total area = "total area of all pixels in contour", and Integrated intensity ("II") = "sum of pixel intensity values within contour".

Steps:

(a). $F_{back}$ = "the background noise for something of this particular area"

$$F_{back} = \frac{(II_{nucleus} - II_{nucleolus}) * (TotalArea_{nucleolus})}{(TotalArea_{nucleus} - TotalArea_{nucleolus})}$$

(b). $F_{kinetochore}$ = "background-corrected signal with respect to integrated intensity"

$$F_{kinetochore} = II_{nucleolus} - F_{back}$$

(c). $Ratio = \dfrac{Signal}{Noise} = \dfrac{F_{kinetochore}}{F_{back}}$

⋆ The condition required for keeping a time value is: $\dfrac{F_{kinetochore}}{F_{back}} \geq 0.04$ (in other words, the nucleolus is at least 4% brighter than the nucleus).

Note that this process requires the choice of a nucleus contour threshold. We chose this value through a somewhat arbitrary eyeballing of the image, for lack of a more experimentally-verified method. Currently, this threshold value remains constant over time and across cells. Figure 7.1 shows selected sequential images of a G1 cell over time. Note how bleaching increases over time.

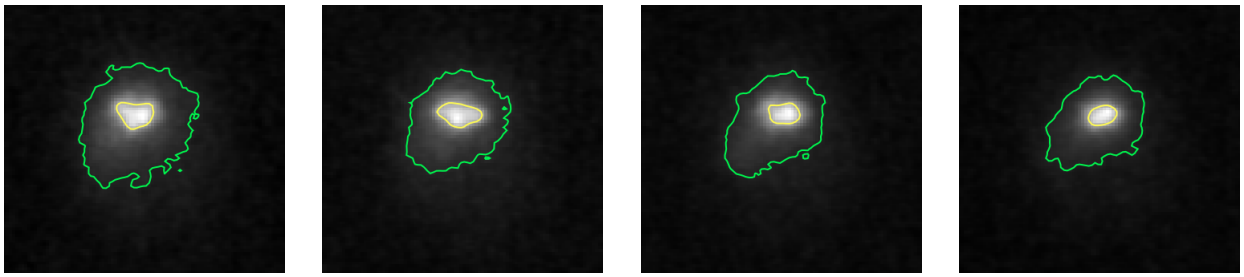

Figure 7.1: *Microscope simulator output of the first four sequentially taken experimental images of a single cell. Images were taken every 30 seconds.*

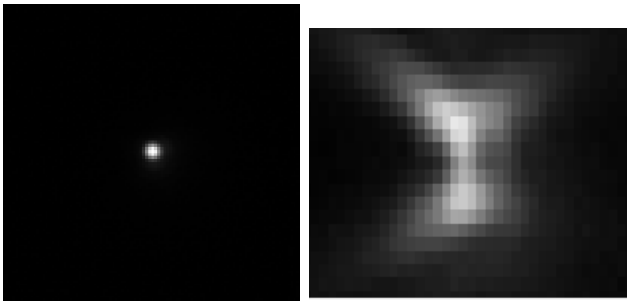

Figure 7.2: *Visualization of how the point spread function we define within our microscope simulator module influences the way light from a simulated fluorescent bead spreads. This PSF is used in the simulated input pipeline to translate simulated position data to simulated fluorescent beads.* Left: View from xy-plane. Right: View from xz-plane.

**Otsu thresholding code.** Sample Matlab code for Otsu thresholding algorithm (Figure 7.3). The microscope simulator pipeline uses a C++ version of the code that I wrote.

```
%We have 22500 total pixels to work with in our image. Let's bin them into a histogram
%of 4096 total bins (since 12-bit), and call the count values of the bins
%"histogramCounts"and the corresponding intensity values of those bins "imageintensities".

[histogramCounts,imageintensities] = hist(ListOfVals_0,4096);
```

```
function level = otsu(histogramCounts)
total = sum(histogramCounts); % '''total''' is the number of pixels in the given image.
%% OTSU automatic thresholding method
sumB = 0;
wB = 0;
maximum = 0.0;
sum1 = dot( (0:4095), histogramCounts);  %0:255 if 8-bit image
for ii=1:4096                            %1:256 if 8-bit image
    wB = wB + histogramCounts(ii);
    if (wB == 0)
        continue;
    end
    wF = total - wB;
    if (wF == 0)
        break;
    end
    sumB = sumB +  (ii-1) * histogramCounts(ii);
    mB = sumB / wB;
    mF = (sum1 - sumB) / wF;
    between = wB * wF * (mB - mF) * (mB - mF);
    if ( between >= maximum )
        level = ii;
        maximum = between;
    end
end
```

```
chosenthreshold = imageintensities(level);
```

Figure 7.3: *Sample Otsu thresholding algorithm, written in Matlab.*

### 7.2.2 Loop size variability and persistence over time.

**Nucleolus beads dynamically form loops of varying sizes.** In the previous chapter we showed that dynamic looping promotes increased compaction, associativity, and substructure formation among nucleolus beads; however, we did not explicitly discuss which nucleolus beads bind with one another over the course of the simulation. Figures 7.5 & 7.4 show how often two nucleolus beads form a loop during the course of the simulation; in other words, how many of the 9000 total time

steps does nucleolus bead $i$ spend bound to other nucleolus beads, and what are the indices of those other nucleolus beads? Parts A and B from Figure 7.5 show a single frame from two movies that I created – using the data sets for slow and fast kinetics, respectively –, in which each frame of the movie shows the amount of time steps that a single nucleolus bead (index specified by the blue dotted line) spends bound to one of the other 360 nucleolus beads during the 15-minute simulation.

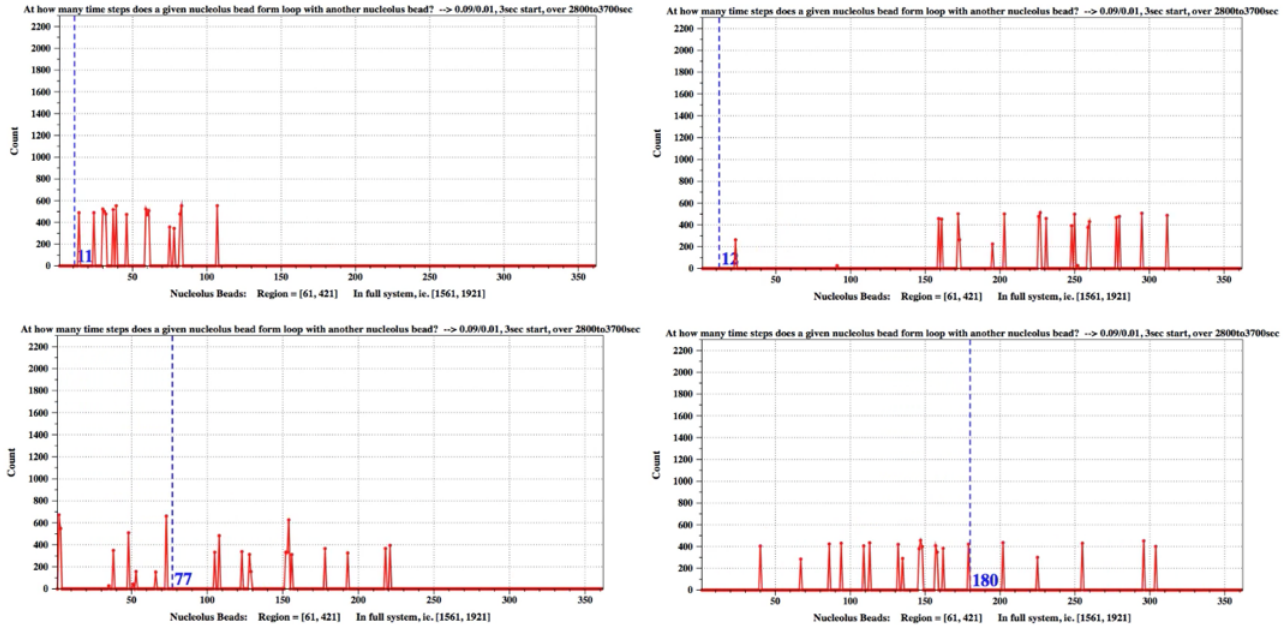


Figure 7.4: *Four selected frames corresponding to the dataset for $t_{on}$=0.09 and a single nucleolus. Dynamic loop partners are shown for nucleolus beads 11, 12, 77, and 180, as indicated by the dotted blue line.*

We note that dynamic looping allows for loops of differing sizes to form and that a single nucleolus bead will dynamically bind with multiple other nucleolus beads during the course of the simulation. Thus, loops truly are forming and breaking during the simulation. Furthermore, beads do not appear to preferentially form loops based on location within the chain; in other words, which two beads form a loop does not appear to be closely correlated with those beads' indices. This behavior is also evident in the split nucleolus case (Figures 7.6 & 7.7). It is particularly noteworthy that beads dynamically bind with beads in both regions of the split nucleolus, not just their own, indicating that the two split nucleolus regions are spatially intermixing and in close proximity.

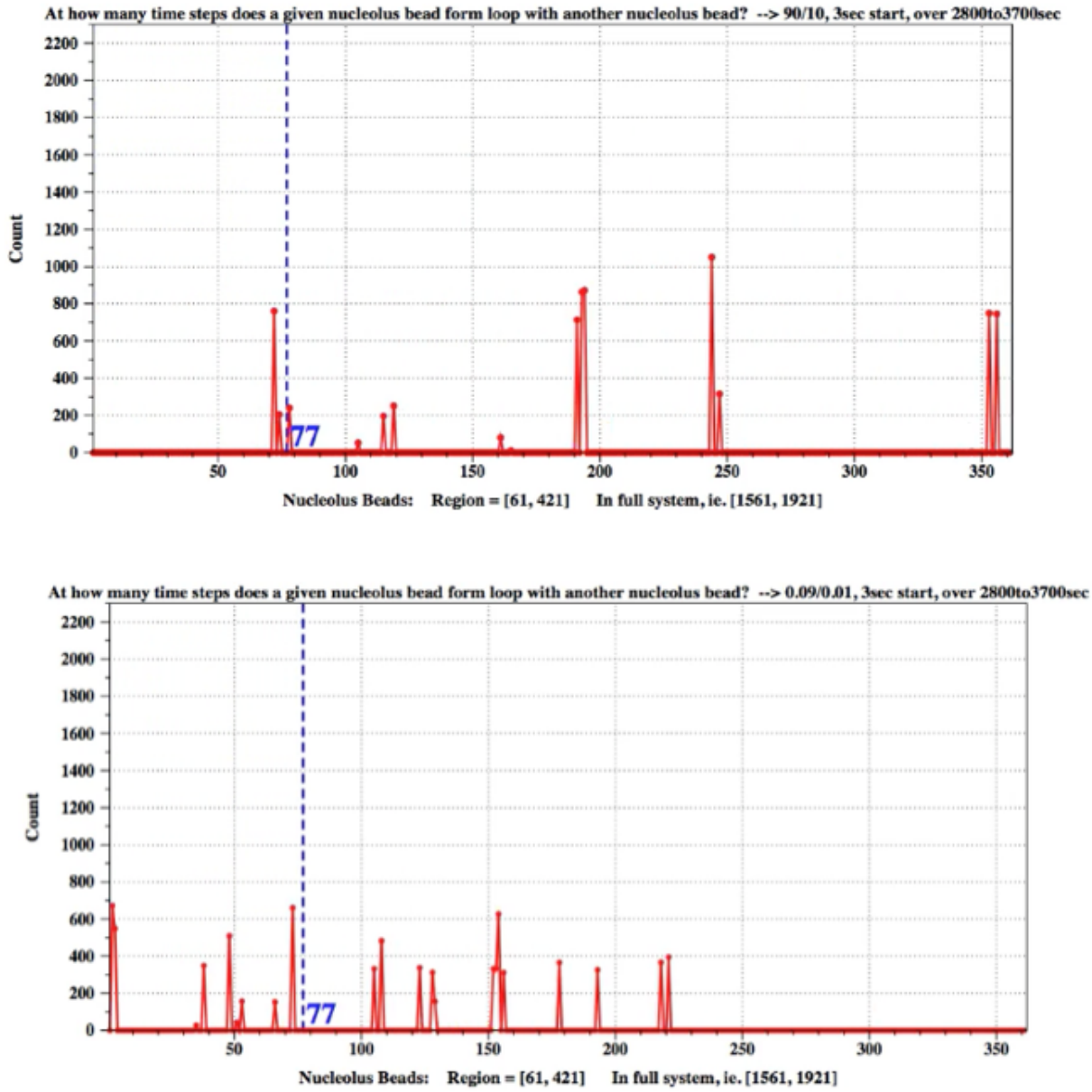


Figure 7.5: *Frequency with which a given nucleolus bead forms a loop with other nucleolus beads. In the single nucleolus there are 361 total beads; here we look at nucleolus bead 77.* **(A)** Slower kinetics, i.e. $t_{on}$=90. **(B)** Faster kinetics, i.e. $t_{on}$=0.09.

Figures 7.8 & 7.7 similarly highlight the fact that "loop size" (ie. the number of beads composing a loop) is variable, through the lens of contact maps.

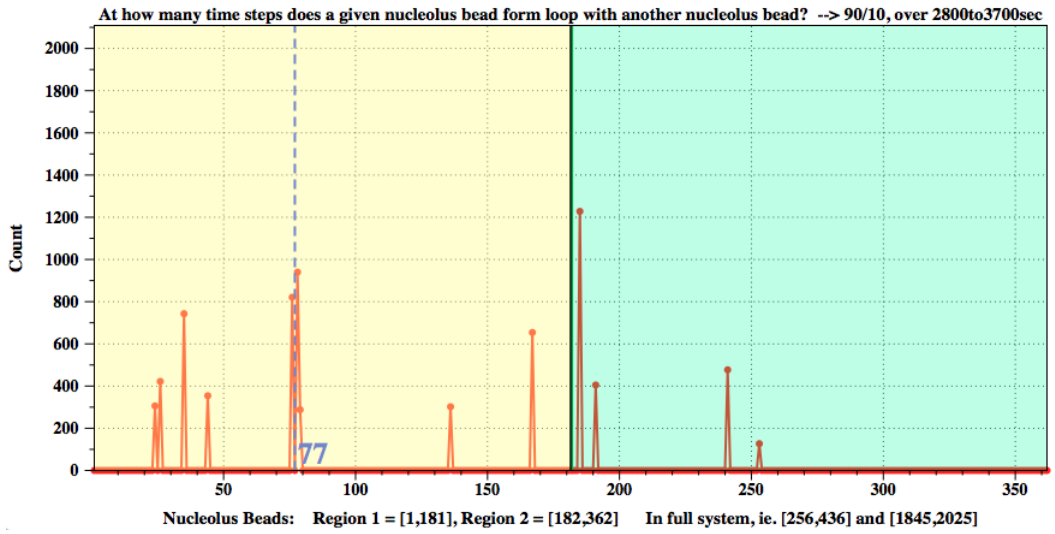


Figure 7.6: *Frequency with which a given nucleolus bead forms a loop with other nucleolus beads. In the split nucleolus there are 362 total beads; here we look at nucleolus bead 77.* The yellow region incorporates the beads that compose region 1 of the nucleolus, the green region incorporates the beads that compose region 2 of the nucleolus. We see that bead 77 binds to beads in both nucleolus regions. Dynamic looping, $t_{on}$=90.

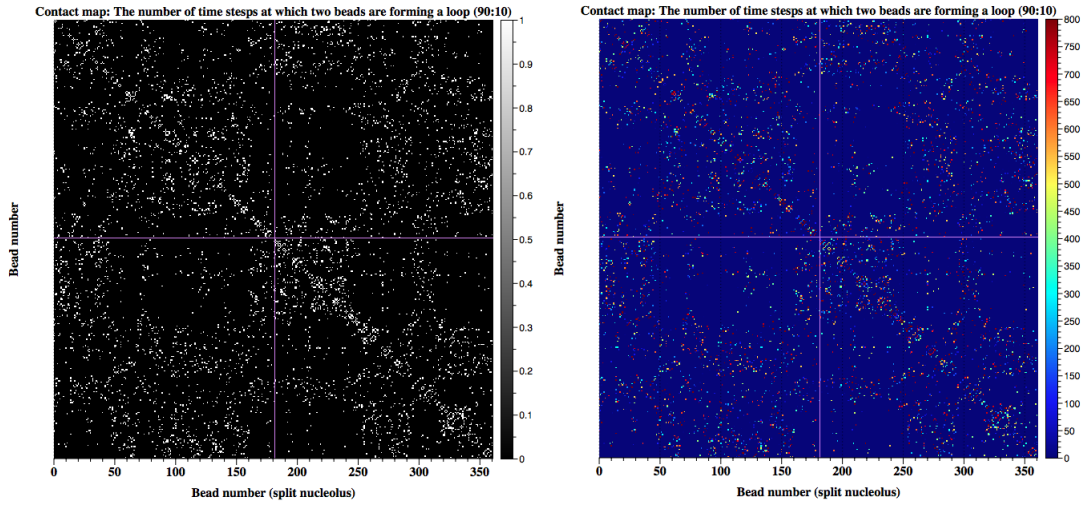


Figure 7.7: *Contact map representation of nucleolus loop binding frequency, corresponding to split nucleolus.*

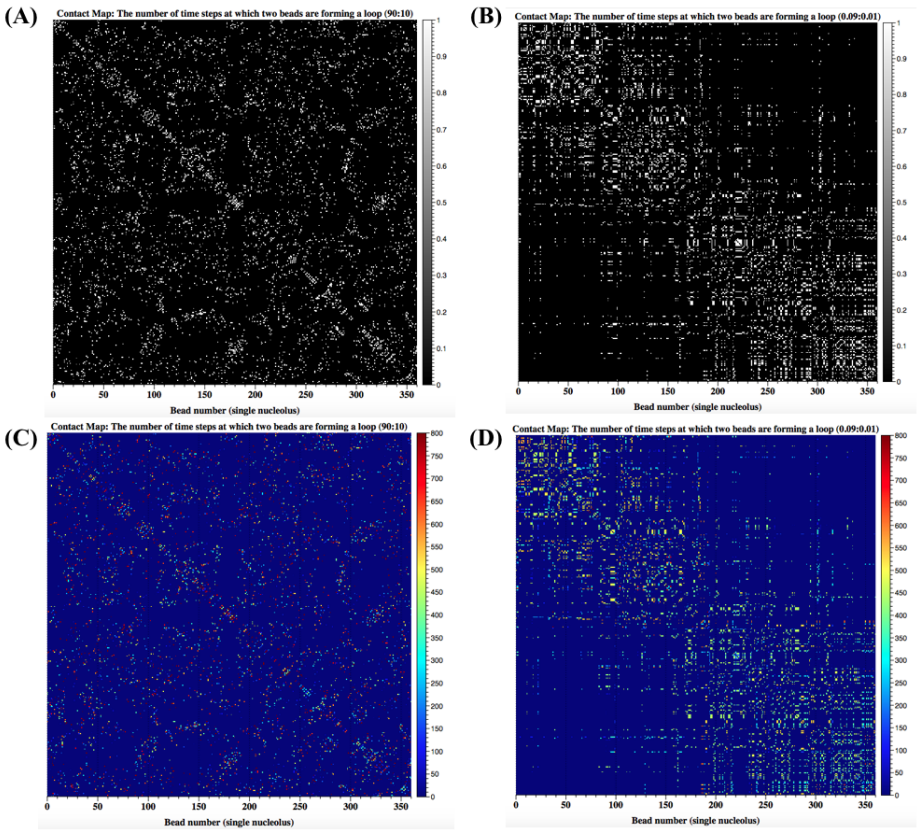


Figure 7.8: *Contact map representation of nucleolus loop binding frequency, corresponding to single nucleolus.*

**Persistence of looping.** In addition to learning which nucleolus beads bind to one another, I also wanted to investigate how long these dynamic loops persisted over the course of the simulation. For example, Figure 7.5 shows that nucleolus bead 77 spends 761 of 9000 time steps bound to nucleolus bead 72, but it does not provide information about the persistence of that bond – did beads 77 and 72 bind and unbind several times, or did they remain bound for all 761 time steps? Figures 7.9& 7.10 show that the magnitude of the on/off binding rates significantly influences the longevity of a dynamic crosslink. As hypothesized, slow dynamic looping enables dynamically formed loops to remain unbroken for a longer period of time than fast dynamic looping, in which loops are breaking and forming at almost every time step. *Very interestingly, however,* we see that, when viewed in conjunction with the 3D visualizations of the nucleolus shown in Figure 6.6, weak (i.e fast) dynamic looping enables sections of the genome to remain in close proximity for an extended period of time, longer than what is seen when strong (i.e. slow) dynamic looping is imposed. This might suggest a mecha-

nism for phase separation, or provide a long enough period of proximity for cellular processes to occur.

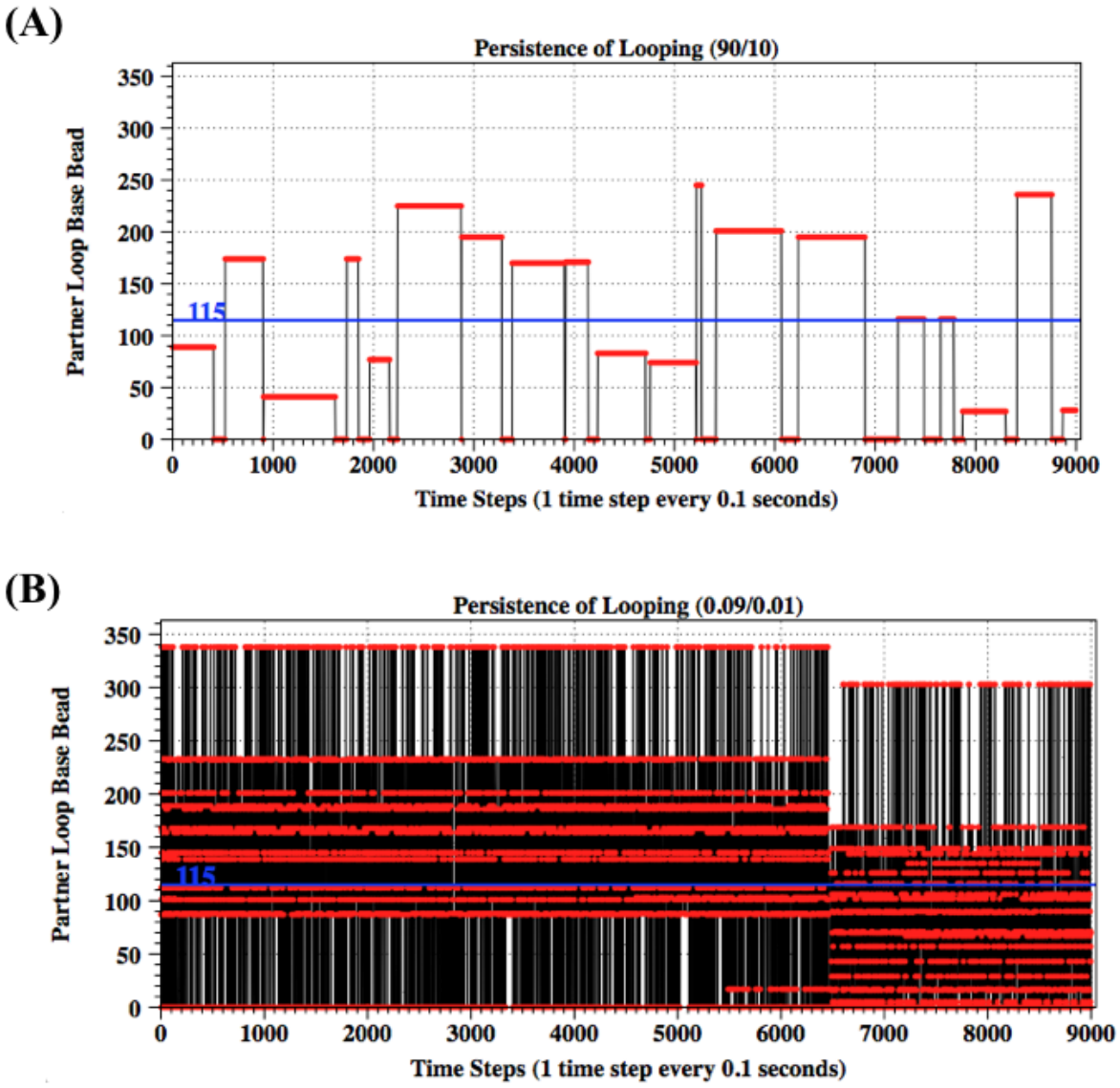


Figure 7.9: *Persistence of looping for nucleolus bead 115.* **(A)** Slow dynamic looping, $t_{on}$=90 s. **(B)** Fast dynamic looping, $t_{on}$=0.09s .

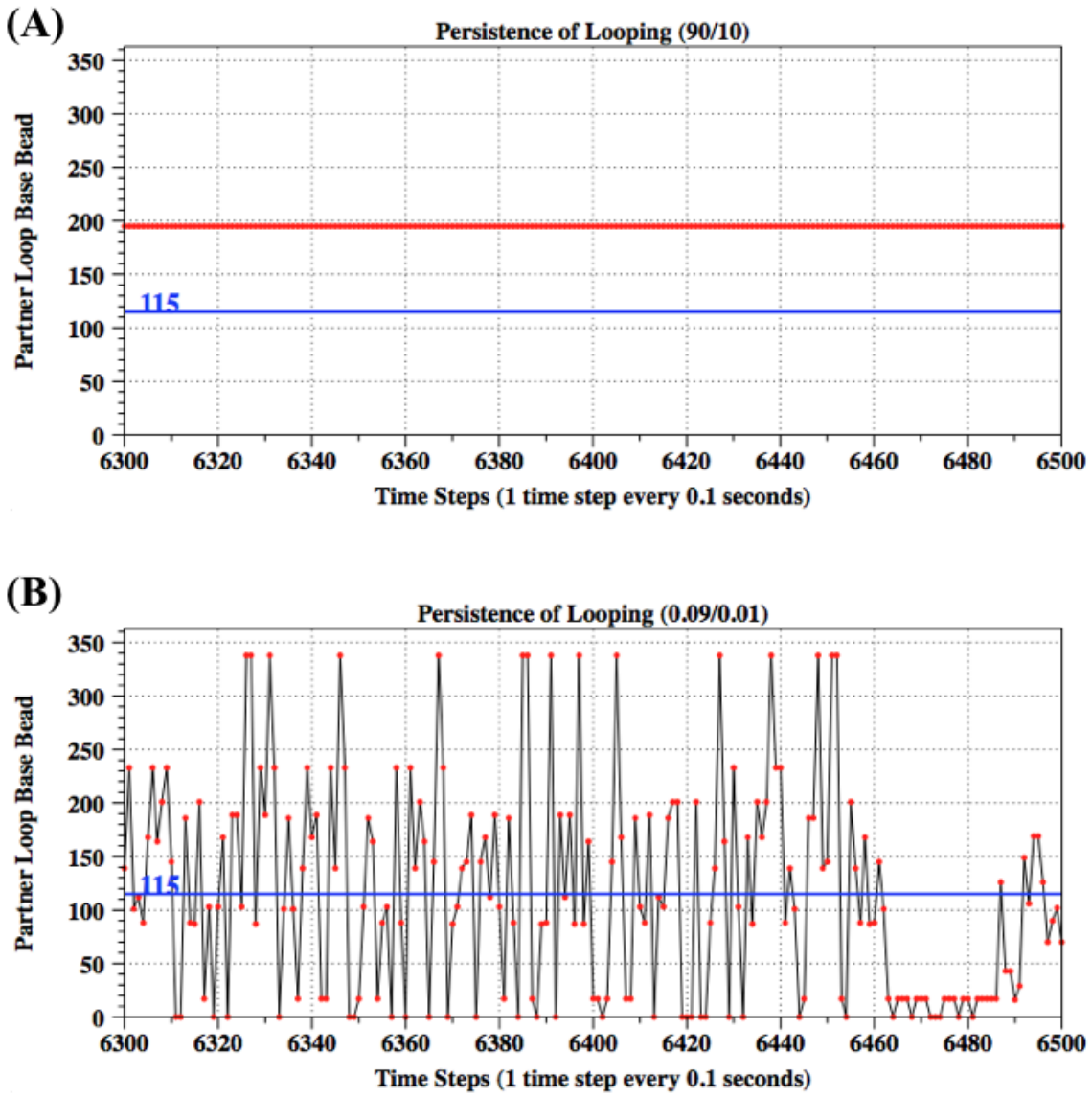


Figure 7.10: *Same as Figure 7.9, just zoomed in (shorter range in x-axis).*

### 7.2.3 3D Visualization within 3D Nucleus.

The figures in this section are snapshots of the system in time, corresponding to varying parameter choices. I include them here to help readers better visualize the nucleolus within the context of the entire nucleus. All figures (Figures 7.11, 7.12, 7.13, 7.14) are frames from movies that I created in which the beads move over time.

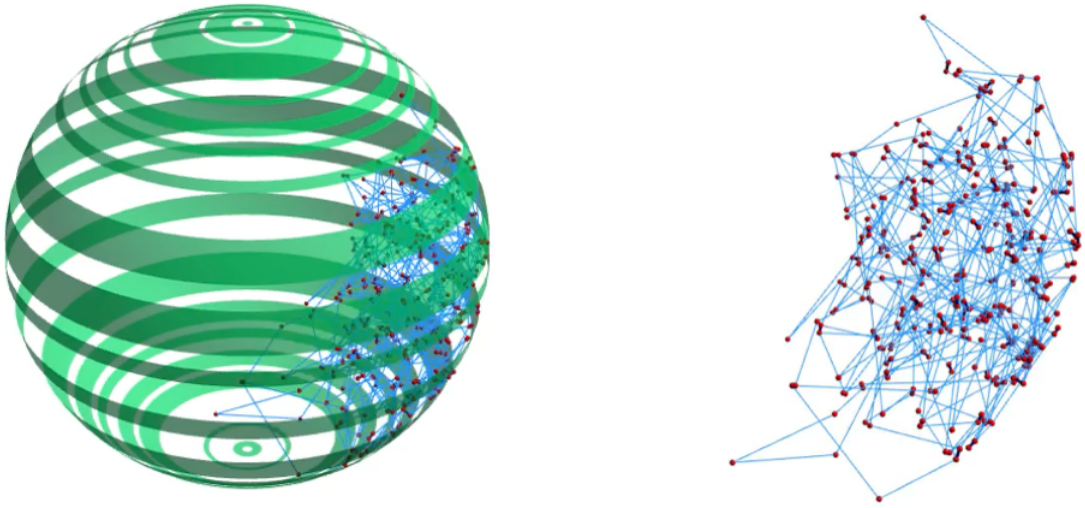

Figure 7.11: *The nucleolus at a single point in time, when there is dynamic looping in the nucleolus with* $t_{on}$*=90.*

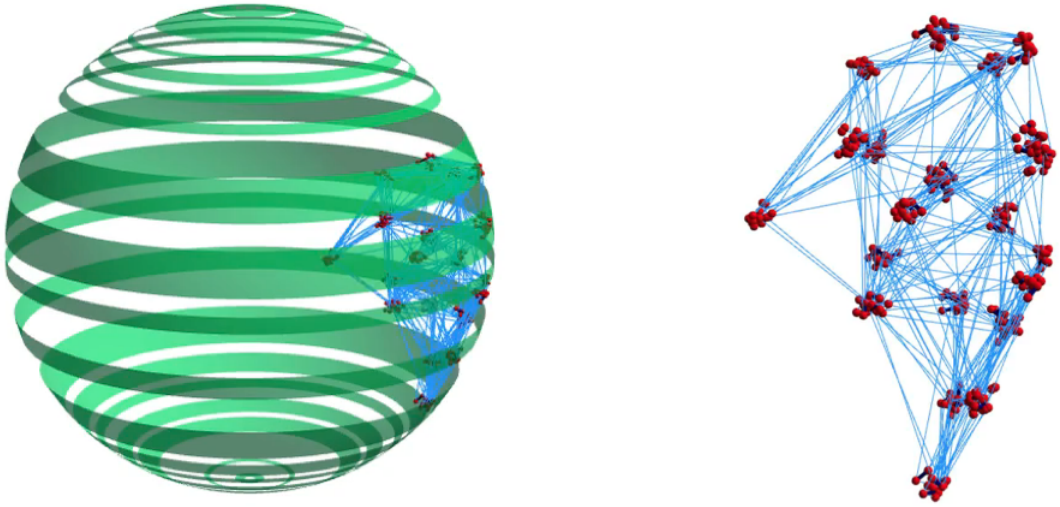

Figure 7.12: *The nucleolus at a single point in time, when there is dynamic looping in the nucleolus with* $t_{on}$*=0.09.*

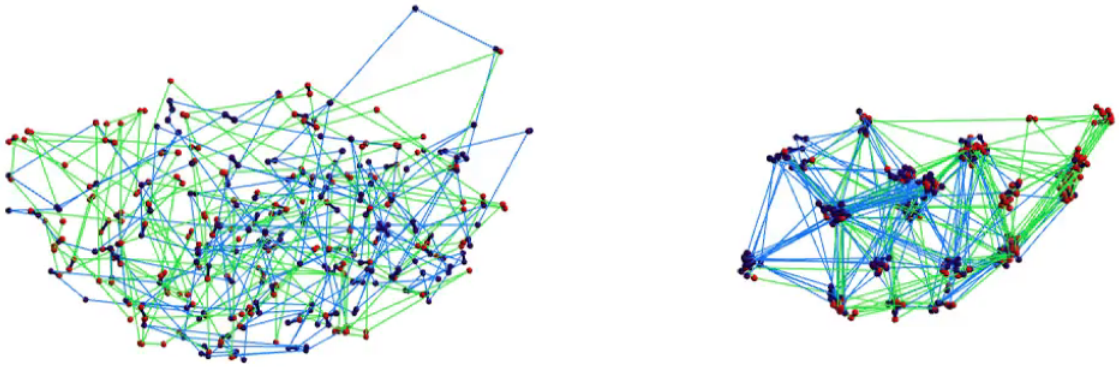

Figure 7.13: *The nucleolus at a single point in time, when the nucleolus is split into two regions.* ***Left)*** *Dynamic looping in the nucleolus with* $t_{on}$*=90.* ***Right)*** *Dynamic looping in the nucleolus with* $t_{on}$*=0.09.*

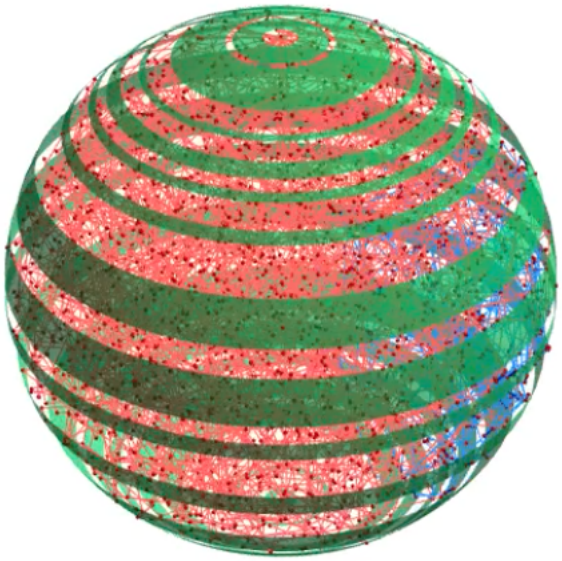

Figure 7.14: *The nucleolus at a single point in time, when there is dynamic looping in the nucleolus with* $t_{on}$*=90. All 2803 beads in the system are shown as red spheres. Springs between non-nucleolus neighbor beads are light red, springs between nucleolus neighbor beads are light blue.*

## 7.3 Changes in cell nucleus viscoelasticity in response to DNA damage

In yeast cells, damaged DNA will attempt to repair itself through homology search, or through non-homologous mechanisms. The collective environment within the cell nucleus exhibits viscoelastic properties, meaning it experiences both viscous and elastic fluctuations. We are interested in determining how the mechanical properties of the cell environment change after a DNA strand is cut, and how the resulting changes to the viscous or elastic nature of the environment may facilitate the repair pathway. Taking into consideration varying boundary conditions for the nuclear walls, we use mathematical analysis to investigate how the movement of damaged bead-spring chains and subsequent interactions among all chains influence the viscoelastic moduli of the cell nucleus and offer insight into the collective fluctuations of chromatin spots. Are effects from DNA damage felt both locally and globally? As the lead mentor for an REU on this topic, I guided three undergraduate students as we implemented the Mason-Weitz Protocol [114, 115], as well as other analysis techniques such as mean squared displacement, to study how the movement of simulated damaged DNA and the stiffness of the nucleus wall influence the environment within the nucleus. We found that the local viscoelastic signature of a chosen DNA spot is position- and chain-dependent.

It would be interesting to expand this project to a study of colored noise, and of how colored noise may be facilitating repair processes. For example, through implementing the differences power law scaling of the mean squared displacement of beads located near a damaged site or away from a damaged site, we could investigate how DNA damage translates to different colored noise signatures of the local environment.

CHAPTER 8

# Conclusions

## 8.1 Conclusions

In this dissertation, we present a comprehensive 3D model describing the spatiotemporal dynamics of interphase chromosomes in yeast cells that is fully reflective of experimental parameters. Our findings support and expand on past results. We show that a model consisting of first principles – thermal fluctuations, entropic forces, repulsive forces, tethering, and confinement – is enough to produce and see evidence of topologically associating domains. We show that nuclear bodies are secondary levels of regulation of subdomains of high intra-chromosomal interactions and that enzyme-related processes modulate these domains. In particular, we present novel modeling and visualization techniques for the nucleolus. We explore the minimal physical requirements to segregate this region of repeated DNA from the remainder of the genome. Through an investigation of different types of looping, we found that imposing cross-links between non-adjacent beads in a bead-spring polymer model promotes compaction, physical segregation, and heterogeneity of the nucleolus. Through tuning the magnitude of the binding kinetics, we can further tune the degree of segregation and connectivity; namely, fast dynamic crosslinking enhances interactions within the active binding domain. We find that fast dynamic crosslinking promotes substructure within the nucleolus, as evidenced by the formation of clusters of beads that persist over time as well as analyses of average distance between nucleolus beads with other nucleolus beads over time. Notably, cluster formation in response to fast kinetics but not in response to slow kinetics is robust, as clusters disappear when we slow down the binding kinetics partway through a simulation and clusters form when we speed up the binding kinetics partway through a simulation. From a biological standpoint, such results have potentially very significant implications regarding our understanding of the factors driving and mechanisms behind phase separation and various cellular processes such as gene expression.

Through development of analysis and visualization tools such as the microscope simulator, in which we transform simulated position data into equivalent microscope images, we show that our

model reproduces the nucleolar compaction, morphology, and statistics seen in live cells. Furthermore, these distinct features persist when the nucleolus is split between two chromosome arms. Significantly, our model reproduces the experimental result that co-linearity is not required to form a single domain; preferential dynamic crosslinking is a sufficient segregation mechanism.

Perhaps most significantly, we highlight the importance of studying chromosome dynamics on a single, live cell level. As stated in Hult et al. [11], we find that analyses based on population averages are likely insufficient and perhaps misleading, as population averages might obscure a variety of dynamic structures that are appreciable only when considering single, living cells. This also underscores the value of continuing to develop comprehensive, time-dependent, experimentally-guided mathematical models.

## 8.2 Future Work and Directions

As discussed earlier, there are many interesting future directions for this work. Of particular interest and focus is further development of how we model the nucleolus. We plan to incorporate additional levels of complexity to the dynamic looping parameters, i.e., additional types of binding proteins, additional binding sites on proteins, variation of binding rates throughout the genome. We also hope to develop better ways of quantifying the nucleolus, beyond the nucleolus area measurements we can obtain through our microscope simulator. For instance, we would like to write a code to describe the point cloud of the nucleolus over time. This would give us a volume metric and help us to better analyze the dynamic shape of the nucleolus over time.

Another research focuses on developing a better understanding of the formation of clusters at fast kinetics. We are working with Dane Taylor, Sam Heroy, and Peter Mucha to use network techniques such as k-means clustering to study these clusters. Central goals of this study include determining how long clusters persist, how system size or binding kinetics influence the number and size of clusters that form, and developing a metric for measuring cluster robustness. Additional studies include analyzing the local viscoelastic environment within and outside of clusters, as well as a parameter sensitivity study of binding on/off rates to determine the optimal parameters for cluster formation.

An underlying motivational factor for further development of the mathematical model introduced in this thesis is the desire to better understand the movement of damaged DNA. In yeast cells, damaged DNA will attempt to repair itself through homology search, or through non-homologous

mechanisms. We are interested in determining how the mechanical properties of the cell environment change after a DNA strand is cut, and how the resulting change to the viscous or elastic nature of the environment may facilitate the repair pathway. It is known that DNA explores more space when it is cut; through our future work, we hope to learn how much more space it explores, why it explores more space when damaged, and how to modify our parameters and equations to describe this movement.

## REFERENCES


[1] J. S. Verdaasdonk, P. A. Vasquez, R. M. Barry, T. Barry, S. Goodwin, M. G. Forest, and K. Bloom, "Centromere tethering confines chromosome domains," *Molecular cell*, vol. 52, no. 6, pp. 819–831, 2013.

[2] P. A. Vasquez, C. Hult, D. Adalsteinsson, J. Lawrimore, M. G. Forest, and K. Bloom, "Entropy gives rise to topologically associating domains," *Nucleic acids research*, vol. 44, no. 12, pp. 5540–5549, 2016.

[3] K. Bloom, "A close look at wiggly chromosomes," *Developmental cell*, vol. 25, no. 4, pp. 330–332, 2013.

[4] P. A. Vasquez and K. Bloom, "Polymer models of interphase chromosomes," *Nucleus*, vol. 5, no. 5, pp. 376–390, 2014.

[5] J. Lawrimore, J. K. Aicher, P. Hahn, A. Fulp, B. Kompa, L. Vicci, M. Falvo, R. M. Taylor II, and K. Bloom, "Chromoshake: a chromosome dynamics simulator reveals that chromatin loops stiffen centromeric chromatin," *Molecular Biology of the Cell*, vol. 27, no. 1, pp. 153–166, 2016.

[6] G. Fudenberg, M. Imakaev, C. Lu, A. Goloborodko, N. Abdennur, and L. A. Mirny, "Formation of chromosomal domains by loop extrusion," *Cell Reports*, vol. 15, pp. 2038–2049, 2016.

[7] G. Andrey, T. Montavon, B. Mascrez, F. Gonzalez, D. Noordermeer, M. Leleu, D. Trono, F. Spitz, and D. Duboule, "A switch between topological domains underlies hoxd genes collinearity in mouse limbs," *Science*, vol. 340, p. 1234167, 2013.

[8] D. Lupiáñez, K. Kraft, V. Heinrich, P. Krawitz, F. Brancati, E. Klopocki, D. Horn, H. Kayserili, J. Optiz, and R. Laxova, "Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions," *Cell*, vol. 161, pp. 1012–1025, 2015.

[9] O. Symmons, V. Uslu, T. Tsujimura, S. Ruf, S. Nassari, W. Schwarzer, L. Ettwiller, and F. Spitz, "Functional and topological characteristics of mammalian regulatory domains," *Genome Res*, vol. 24, pp. 390–400, 2014.

[10] A. Y. Grosberg, "Extruding loops to make loopy globules?," *Biophysical Journal*, vol. 110, pp. 2133–2135, 2016.

[11] C. Hult, D. Adalsteinsson, P. Vasquez, J. Lawrimore, M. Bennett, A. York, M. G. Forest, and K. Bloom, "Enrichment of dynamic chromosomal crosslinks drive phase separation of the nucleolus," *Nucleic Acids Research, submitted*.

[12] M. Doi and S. Edwards, *The theory of polymer dynamics*. Oxford: Oxford University Press, 1986.

[13] R. Larson, "The rheology of dilute solutions of flexible polymers: progress and problems," *J. Rheol.*, vol. 49, pp. 1–70, 2005.

[14] J. F. Marko and E. D. Siggia, "Bending and twisting elasticity of dna," *Macromolecules*, vol. 27, no. 4, pp. 981–988, 1994.

[15] J. F. Marko and E. D. Siggia, "Stretching dna," *Macromolecules*, vol. 28, no. 26, pp. 8759–8770, 1995.

[16] R. Jendrejack, J. de Pablo, and M. Graham, "Stochastic simulations of dna in flow: dynamics and the effects of hydrodynamic interactions," *J. Chem. Phys.*, vol. 116, pp. 7752–7759, 2002.

[17] D. Schroeder, E. Shaqfeh, and S. Chu, "Effect of hydrodynamic interactions on dna dynamics in extensional flow: Simulation and single molecule experiment," *Macromolecules*, vol. 37, no. 24, pp. 9242–9256, 2004.

[18] A. Izmitli and et al, "The effect of hydrodynamic interactions on the dynamics of dna translocation through pores," *The Journal of Chemical Physics*, vol. 128, no. 8, p. 085102, 2008.

[19] A. E. Likhtman, "Viscoelasticity and molecular rheology," *Polymer Science: A Comprehensive Reference*, vol. 1, pp. 133–179.

[20] V. Ea, M.-O. Baudement, A. Lesne, and T. Forné, "Contribution of topological domains and loop formation to 3d chromatin organization," *Genes*, vol. 6, no. 3, pp. 734–750, 2015.

[21] S. Chakravarthy, Y. Park, J. Chodaparambil, R. Edayathumangalam, and K. Luger, "Structure and dynamic properties of nucleosome core particles," *FEBS Lett.*, vol. 579, no. 4, pp. 895–898.

[22] J. Chodaparambil, R. Edayathumangalam, Y. Bao, Y. Park, and K. Luger, "Nucleosome structure and function. In: *The Histone Code and Beyond*," vol. 57, (Berlin Heidelberg), pp. 29–46, Ernst Schering Res Found Workshop, Springer, 2006.

[23] J. D. McGhee, D. C. Rau, E. Charney, and G. Felsenfeld, "Orientation of the nucleosome within the higher order structure of chromatin," *Cell*, vol. 22, no. 1, pp. 87–96, 1980.

[24] R. McGinty and S. Tan, "Nucleosome structure and function," *Chem. Rev.*, vol. 115, pp. 2255–2273, 2014.

[25] H. Schiessel, "The physics of chromatin," *J. Phys. Condens. Matter*, vol. 25, pp. R699–R774, 2003.

[26] A. Wolffe, *Chromatin: structure and function.* London: Academic press, 1998.

[27] A. Allahverdi, Q. Chen, N. Korolev, and L. Nordenski$\acute{\text{o}}$ld, "Chromatin compaction under mixed salt conditions: opposite effects of sodium and potassium ions on nucleosome array folding," *Scientific Reports*, vol. 5, pp. 1–7, 2015.

[28] K. Van Holde and J. Zlatanova, "What determines the folding of the chromatin fiber?," *Proc. Natl. Acad. Sci.*, vol. 93, pp. 10548–10555, 1996.

[29] K. Van Holde and J. Zlatanova, "Chromatin higher order structure: chasing a mirage?," *J. Biol. Chem.*, vol. 270, pp. 8373–8376, 1995.

[30] C. Woodcock, L. Frado, and J. Rattner, "The higher-order structure of chromatin: evidence for a helical ribbon arrangement," *J. Cell Biol.*, vol. 99, pp. 42–52, 1984.

[31] E. Fussner, R. Ching, and D. Bazett-Jones, "Living without 30nm chromatin fibers," *Trends Biochem. Sci.*, vol. 36, pp. 1–6, 2011.

[32] M. Eltsov, K. MacLellan, K. Maeshima, A. Frangakis, and J. Dubochet, "Analysis of cryo-electron microscopy images does not support the existence of 30-nm chromatin fibers in mitotic chromosomes in situ," *Proc. Natl. Acad. Sci.*, vol. 105, pp. 19732–19737, 2008.

[33] H. Tjong, K. Gong, L. Chen, and F. Alber, "Physical tethering and volume exclusion determine higher-order genome organization in budding yeast," *Genome research*, vol. 22, no. 7, pp. 1295–1305, 2012.

[34] R. Wang, J. Mozziconacci, A. Bancaud, and O. Gadal, "Principles of chromatin organization in yeast: relevance of polymer models to describe nuclear organization and dynamics," *Current opinion in cell biology*, vol. 34, pp. 54–60, 2015.

[35] T. Misteli and E. Soutoglou, "The emerging role of nuclear architecture in dna repair and genome maintenance," *Nat. Rev. Mol. Cell Biol.*, vol. 10, pp. 243–254, 2009.

[36] A. Taddei, H. Schober, and S. Gasser, "The budding yeast nucleus," *Cold Spring Harb. Perspect. Biol.*, vol. 2, p. a000612.

[37] T. Takizawa, K. Meaburn, and T. Misteli, "The meaning of gene positioning," *Cell*, vol. 135, pp. 9–13, 2008.

[38] E. Alipour and J. Marko, "Self-organization of domain structures by dna-loop-extruding enzymes," *Nucleic Acids Research*, vol. 40, pp. 11202–11212, 2012.

[39] K. Giles, H. Gowher, R. Ghirlando, C. Jin, and G. Felsenfeld, "Chromatin boundaries, insulators, and long-range interactions in the nucleus," vol. 75, (New York), p. 7985, Cold Spring Harbor Symposia on Quantitative Biology, Cold spring Harbor Laboratory Press, 2010.

[40] M. H. Kagey, J. J. Newman, S. Bilodeau, Y. Zhan, D. A. Orlando, N. L. van Berkum, C. C. Ebmeier, J. Goossens, P. B. Rahl, S. S. Levine, D. J. Taatjes, J. Dekker, and R. A. Young, "Mediator and cohesin connect gene expression and chromatin architecture," *Nature*, vol. 467, pp. 430–435, 2010.

[41] S. Rao, M. Huntley, N. Durand, E. Stamenova, I. Bochkov, J. Robinson, A. Sanborn, I. Machol, A. Omer, and E. Lander, "A 3d map of the human genome at kilobase resolution reveals principles of chromatin looping," *Cell*, vol. 159, pp. 1665–1680, 2014.

[42] R. Schleif, "Dna looping," *Annu. Rev. Biochem.*, vol. 61, pp. 199–223, 1992.

[43] M. Bohn and D. Heermann, "Diffusion-driven looping provides a consistent framework for chromatin organization," *PLoS One*, vol. 5, p. e12218.

[44] M. Tark-Dame, H. Jerabek, E. Manders, D. Heermann, and R. van Driel, "Depletion of the chromatin looping proteins ctcf and cohesion causes chromatin compaction: insight into chromatin folding by polymer modelling," *PLoS Comput. Biol.*, vol. 10, p. e1003877.

[45] B. Doyle, G. Fudenberg, M. Imakaev, and L. Mirny, "Chromatin loops as allosteric modulators of enhancer-promoter interactions," *PLoS Comput. Biol.*, vol. 10, p. e1003867.

[46] J. Dekker, M. Marti-Renom, and L. Mirny, "Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data," *Nat. Rev. Genet.*, vol. 14, pp. 390–403, 2013.

[47] J. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. Liu, and B. Ren, "Topological domains in mammalian genomes identified by analysis of chromatin interactions," *Nature*, vol. 485, pp. 376–380, 2012.

[48] E. Nora, J. Dekker, and E. Heard, "Segmental folding of chromosomes: a basis for structural and regulatory chromosomal neighborhoods?," *Bioessays*, vol. 35, pp. 818–828, 2013.

[49] C. Ong and V. Corces, "Ctcf: an architectural protein bridging genome topology and function," *Nat. Rev. Genet.*, vol. 15, pp. 234–246, 2014.

[50] A. B. Berger, G. G. Cabal, E. Fabre, T. Duong, H. Buc, U. Nehrbass, J.-C. Olivo-Marin, O. Gadal, and C. Zimmer, "High-resolution statistical mapping reveals gene territories in live yeast," *Nature Methods*, vol. 5, no. 12, pp. 1031–1037, 2008.

[51] P. Heun, T. Laroche, K. Shimada, P. Furrer, and S. Gasser, "Chromosome dynamics in the yeast interphase nucleus," *Science*, vol. 294, pp. 2181–2186, 2001.

[52] W. Marshall, A. Straight, J. Marko, J. Swedlow, A. Dernburg, A. Belmont, A. Murray, D. Agard, and J. Sedat, "Interphase chromosomes undergo constrained diffusional motion in living cells," *Curr. Biol.*, vol. 7, pp. 930–939, 1997.

[53] H. Wong, J.-M. Arbona, and C. Zimmer, "How to build a yeast nucleus," *Nucleus*, vol. 4, no. 5, pp. 361–366, 2013.

[54] F. Erdel, M. Baum, and K. Rippe, "The viscoelastic properties of chromatin and the nucleoplasm revealed by scale-dependent protein mobility," *J. Phys. Condens. Matter*, vol. 27, p. 064115.

[55] J. Lawrimore, P. A. Vasquez, M. Falvo, R. Taylor, L. Vicci, E. Yeh, M. G. Forest, and K. Bloom, "Dna loops generate intracentromere tension in mitosis," *J. Cell Biol.*, vol. 210, pp. 553–564, 2015.

[56] A. D. Stephens, R. A. Haggerty, P. A. Vasquez, L. Vicci, C. E. Snider, F. Shi, C. Quammen, C. Mullins, J. Haase, R. M. Taylor, J. S. Verdaasdonk, M. R. Falvo, Y. Jin, M. G. Forest, and K. Bloom, "Pericentric chromatin loops function as a nonlinear spring in mitotic force balance," *J Cell Biol*, vol. 200, pp. 757–772, 2013.

[57] A. D. Stephens, C. Snider, J. Haase, R. Haggerty, P. A. Vasquez, M. G. Forest, and K. Bloom, "Individual pericentromeres display coordinated motion and stretching in the yeast spindle," *J Cell Biol*, vol. 203, pp. 407–416, 2013.

[58] G. Fudenberg and L. A. Mirny, "Higher-order chromatin structure: bridging physics and biology," *Current opinion in genetics & development*, vol. 22, no. 2, pp. 115–124, 2012.

[59] A. Rosa, N. Becker, and R. Everaers, "Looping probabilities in model interphase chromosomes," *Biophys. J.*, vol. 98, pp. 2410–2419, 2010.

[60] T. Perkins, D. Smith, and S. Chu, "Single polymer dynamics in an elongational flow," *Science*, vol. 276, pp. 2016–2021, 1997.

[61] K. Bystricky, T. Laroche, G. van Houwe, M. Blaszczyk, and S. Gasser, "Chromosome looping in yeast telomere pairing and coordinated movement reflect anchoring efficiency and territorial organization," *J. Cell Biol.*, vol. 168, pp. 375–387, 2005.

[62] A. Rosa and R. Everaers, "Structure and dynamics of interphase chromosomes," *PLoS Comput Biol*, vol. 4, no. 8, pp. e1000153–e1000153, 2008.

[63] P. Therizols, T. Duong, B. Dujon, C. Zimmer, and E. Fabre, "Chromosome arm length and nuclear constraints determine the dynamic relationship of yeast subtelomeres," *Proceedings of the National Academy of Sciences*, vol. 107, no. 5, pp. 2025–2030, 2010.

[64] H. Wong, H. Marie-Nelly, S. Herbert, P. Carrivain, H. Blanc, R. Koszul, E. Fabre, and C. Zimmer, "A predictive computational model of the dynamic 3d interphase yeast nucleus," *Current Biology*, vol. 22, no. 20, pp. 1881–1890, 2012.

[65] B. Albert, J. Mathon, A. Shukla, H. Saad, C. Normand, I. Léger-Silvestre, D. Villa, A. Kamgoue, J. Mozziconacci, and H. Wong, "Systematic characterization of the conformation and dynamics of budding yeast chromosome xii," *The Journal of cell biology*, vol. 202, no. 2, pp. 201–210, 2013.

[66] H. Hajjoul, J. Mathon, H. Ranchon, I. Goiffon, J. Mozziconacci, B. Albert, P. Carrivain, J.-M. Victor, O. Gadal, and K. Bystricky, "High-throughput chromatin motion tracking in living yeast reveals the flexibility of the fiber throughout the genome," *Genome research*, vol. 23, no. 11, pp. 1829–1838, 2013.

[67] T. M. Cheng, S. Heeger, R. A. Chaleil, N. Matthews, A. Stewart, J. Wright, C. Lim, P. A. Bates, and F. Uhlmann, "A simple biophysical model emulates budding yeast chromosome condensation," *Elife*, vol. 4, p. e05565, 2015.

[68] J. Dekker, K. Rippe, M. Dekker, and N. Kleckner, "Capturing chromosome conformation," *Science*, vol. 295, no. 5558, pp. 1306–1311, 2002.

[69] Z. Duan, M. Andronescu, K. Schutz, S. McIlwain, Y. J. Kim, C. Lee, J. Shendure, S. Fields, C. A. Blau, and W. S. Noble, "A three-dimensional model of the yeast genome," *Nature*, vol. 465, no. 7296, pp. 363–367, 2010.

[70] C. Zimmer and E. Fabre, "Principles of chromosomal organization: lessons from yeast," *The Journal of cell biology*, vol. 192, no. 5, pp. 723–733, 2011.

[71] P. De Gennes, *Scaling concepts in polymer physics*. New York: Cornell University Press, 1979.

[72] K. Bystricky, P. Heun, L. Gehlen, J. Langowski, and S. Gasser, "Long-range compaction and flexibility of interphase chromatin in budding yeast analyzed by high-resolution imaging techniques," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 101, pp. 16495–16500, 2004.

[73] Y. Cui and C. Bustamante, "Pulling a single chromatin fiber reveals the forces that maintain its higher-order structure," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 97, pp. 127–132, 2000.

[74] J. Langowski, "Polymer chain models of dna and chromatin," *Euro. Phys. J. E.*, vol. 19, pp. 241–249, 2006.

[75] J. Dekker, "Mapping in vivo chromatin interactions in yeast suggests an extended chromatin fiber with regional variation in compaction," *J. Biol. Chem.*, vol. 283, pp. 34532–34540, 2008.

[76] J. Fisher, M. Ballenger, E. O'Brien, J. Haase, R. Superfine, and K. Bloom, "Dna relaxation dynamics as a probe for the intracellular environment," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 106, pp. 9250–9255, 2009.

[77] P. Underhill and P. Doyle, "Accuracy of bead-spring chains in strong flows," *J. Non-Newtonian Fluid Mech.*, vol. 145, pp. 109–123, 2007.

[78] D. Heyes and J. Melrose, "Brownian dynamics simulations of model hard-sphere suspensions," *J. Non-Newtonian Fluid Mech.*, vol. 46, pp. 1–28, 1993.

[79] K. Rippe, "Making contacts on a nucleic acid polymer," *Trends Biochem. Sci.*, vol. 26, pp. 733–740, 2001.

[80] R. Sachs, G. Van Den Engh, B. Trask, H. Yokota, and J. Hearst, "A random-walk/giant-loop model for interphase chromosomes," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 92, pp. 2710–2714, 1995.

[81] R. Hancock, *The crowded nucleus. In: New models of the cell nucleus: crowding, entropic forces, phase separation, and fractals*, vol. 307. Cambridge: Academic Press, 2014.

[82] R. Ghirlando and G. Felsenfeld, "Ctcf: making the right connections," *Genes Dev.*, vol. 30, pp. 881–891, 2016.

[83] C. Hou, R. Dale, and A. Dean, "Cell type specificity of chromatin organization mediated by ctcf and cohesin," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 107, pp. 3651–3656, 2010.

[84] J. Phillips and V. Corces, "Ctcf: master weaver of the genome," *Cell*, vol. 137, pp. 1194–1211, 2009.

[85] M. Chen and M. Gartenberg, "Coordination of trna transcription with export at nuclear pore complexes in budding yeast," *Genes Dev.*, vol. 28, pp. 959–970, 2014.

[86] R. Haeusler, M. Pratt-Hyatt, P. Good, T. Gipson, and D. Engelke, "Clustering of yeast trna genes is mediated by specific association of condensin with trna gene transcription complexes," *Genes Dev.*, vol. 22, pp. 2204–2214, 2008.

[87] C. Snider, A. Stephens, J. Kirkland, O. Hamdani, R. Kamakaka, and B. Kerry, "Dyskerin, trna genes, and condensin tether pericentric chromatin to the spindle axis in mitosis," *J. Cell Biol.*, vol. 207, pp. 189–199, 2014.

[88] K. Bloom and A. Joglekar, "Towards building a chromosome segregation machine," *Nature*, vol. 463, pp. 446–456, 2010.

[89] J. F. Marko and E. D. Siggia, "Polymer models of meiotic and mitotic chromosomes," *Molecular biology of the cell*, vol. 8, no. 11, pp. 2217–2231, 1997.

[90] W. Marshall and J. Fung, "Modeling meiotic chromosome pairing: nuclear envelope attachment, telomere-led active random motion, and anomalous diffusion," *Phys. Biol.*, vol. 13, p. 026003.

[91] C. Quammen, A. Richardson, J. Haase, B. Harrison, and R. Taylor, "Eurographics workshop on visual computing for biomedicine," *NIH Public Access*, vol. 2008, p. 151, 2008.

[92] C. Cubeñas-Potts and V. G. Corces, "Architectural proteins, transcription, and the three-dimensional organization of the genome," *FEBS letters*, vol. 589, no. 20PartA, pp. 2923–2930, 2015.

[93] J. Dekker and T. Misteli, "Long-range chromatin interactions," *Cold Spring Harbor perspectives in biology*, vol. 7, no. 10, p. a019356, 2015.

[94] J. S. Verdaasdonk and K. Bloom, "Centromeres: unique chromatin structures that drive chromosome segregation," *Nature reviews Molecular cell biology*, vol. 12, no. 5, pp. 320–332, 2011.

[95] C. H. Yang, E. J. Lambie, J. Hardin, J. Craft, and M. Snyder, "Higher order structure is present in the yeast nucleus: autoantibody probes demonstrate that the nucleolus lies opposite the spindle pole body," *Chromosoma*, vol. 98, no. 2, pp. 123–128, 1989.

[96] I. Léger-Silvestre, S. Trumtel, J. Noaillac-Depeyre, and N. Gas, "Functional compartmentalization of the nucleus in the budding yeast saccharomyces cerevisiae," *Chromosoma*, vol. 108, no. 2, pp. 103–113, 1999.

[97] L. R. Gehlen, G. Gruenert, M. B. Jones, C. D. Rodley, J. Langowski, and J. O'Sullivan, "Chromosome positioning and the clustering of functionally related loci in yeast is driven by chromosomal interactions," *Nucleus*, vol. 3, no. 4, pp. 370–383, 2012.

[98] B. Avşaroʻglu, G. Bronk, S. Gordon-Messer, J. Ham, D. A. Bressan, J. E. Haber, and J. Kondev, "Effect of chromosome tethering on nuclear organization in yeast," *PloS one*, vol. 9, no. 7, p. e102474, 2014.

[99] M. D. Mikus and T. D. Petes, "Recombination between genes located on nonhomologous chromosomes in saccharomyces cerevisiae," *Genetics*, vol. 101, no. 3-4, pp. 369–404, 1982.

[100] V. F. Holmes and N. R. Cozzarelli, "Closing the ring: links between smc proteins and chromosome partitioning, condensation, and supercoiling," *Proceedings of the National Academy of Sciences*, vol. 97, no. 4, pp. 1322–1324, 2000.

[101] B. Harrison, M. Hoang, and K. Bloom, "Persistent mechanical linkage between sister chromatids throughout anaphase," *Chromosoma*, vol. 118, no. 5, pp. 633–645, 2009.

[102] J. Torres-Rosell, I. Sunjevaric, G. De Piccoli, M. Sacher, N. Eckert-Boulet, R. Reid, S. Jentsch, R. Rothstein, L. Aragon, and M. Lisby, "The smc5-smc6 complex and sumo modification of rad52 regulates recombinational repair at the ribosomal gene locus," *Nature cell biology*, vol. 9, no. 8, pp. 923–31, 2007.

[103] K. Rippe, "Dynamic organization of the cell nucleus," *Current opinion in genetics & development*, vol. 17, no. 5, pp. 373–380, 2007.

[104] M. C. Konopka, I. A. Shkel, S. Cayley, M. T. Record, and J. C. Weisshaar, "Crowding and confinement effects on protein diffusion in vivo," *Journal of bacteriology*, vol. 188, no. 17, pp. 6115–6123, 2006.

[105] R. Hancock, "A role for macromolecular crowding effects in the assembly and function of compartments in the nucleus," *Journal of structural biology*, vol. 146, no. 3, pp. 281–290, 2004.

[106] R. Hancock, "Structure of metaphase chromosomes: a role for effects of macromolecular crowding," *PLoS One*, vol. 7, no. 4, p. e36045, 2012.

[107] F. J. Iborra, "Can visco-elastic phase separation, macromolecular crowding and colloidal physics explain nuclear organisation?," *Theoretical Biology and Medical Modelling*, vol. 4, no. 1, p. 15, 2007.

[108] H. Tanaka, "Universality of viscoelastic phase separation in dynamically asymmetric fluid mixtures," *Physical review letters*, vol. 76, no. 5, p. 787, 1996.

[109] M. Wachsmuth, W. Waldeck, and J. Langowski, "Anomalous diffusion of fluorescent probes inside living cell nuclei investigated by spatially-resolved fluorescence correlation spectroscopy," *Journal of molecular biology*, vol. 298, no. 4, pp. 677–689, 2000.

[110] J. Mine-Hattab, V. Recamier, I. Izeddin, R. Rothstein, and X. Darzacq, "Fast imaging of dna motion reveals distinct sub-diffusion regimes at the site of dna damage," *bioRxiv*, p. 042051, 2016.

[111] F.-M. Boisvert, S. van Koningsbruggen, J. Navascués, and A. I. Lamond, "The multifunctional nucleolus," *Nature reviews Molecular cell biology*, vol. 8, no. 7, pp. 574–585, 2007.

[112] K. E. Handwerger, C. Murphy, and J. G. Gall, "Steady-state dynamics of cajal body components in the xenopus germinal vesicle," *J Cell Biol*, vol. 160, no. 4, pp. 495–504, 2003.

[113] G. Dellaire and D. P. Bazett-Jones, "Pml nuclear bodies: dynamic sensors of dna damage and cellular stress," *Bioessays*, vol. 26, no. 9, pp. 963–977, 2004.

[114] T. G. Mason, "Estimating the viscoelastic moduli of complex fluids using the generalized stokes-einstein equation," *Rheologica Acta*, vol. 39, no. 4, pp. 371–378, 2000.

[115] T. G. Mason and D. Weitz, "Optical measurements of frequency-dependent linear viscoelastic moduli of complex fluids," *Physical Review Letters*, vol. 74, no. 7, pp. 1250–1253, 1995.