HOW CHROMOSOME-NUCLEAR ENVELOPE ATTACHMENTS AFFECT 3D GENOME ORGANIZATION

Nicholas Allen Kinney

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Genetics, Bioinformatics, and Computational Biology

Alexey Onufriev, Co-chair
Igor Sharakhov, Co-chair
Michel Pleimling
Daniela Cimini

February 12, 2015
Blacksburg, VA

Keywords: chromosome organization and dynamics, Drosophila melanogaster

HOW CHROMOSOME-NUCLEAR ENVELOPE ATTACHMENTS AFFECT 3D GENOME ORGANIZATION

Nicholas Allen Kinney

ABSTRACT

The length of eukaryotic chromosomes is many times longer than the nucleus diameter in most cells; thus, their confinement depends on adopting highly folded configurations. Remarkably, these configurations are non-random and may be important for gene expression and regulation. Thus, genome sequences must be understood in the context of their 3D organization which critically influences the flow of information. The effort to understand this added complexity now encompasses an entire field of chromosome biology and is reshaping the traditional concept of the central dogma. Although little is known about the principles which govern chromosome folding and influence gene regulation, the nuclear envelope is expected to play a significant role since it serves as the physical boundary preventing chromosome from freely diffusing in the cell cytosol. Moreover, experiments suggest that the nuclear envelope engages chromosomes actively by anchoring specific loci and limiting their range of motion. The broad goal of the research presented in this dissertation is to advance our understanding of 3D genome organization with an emphasis on determining the role of the nuclear envelope.

GENERAL AUDIENCE ABSTRACT

The vast majority of cells in the human body contain the exact same genetic material; indeed, this is true for all organisms. Typically, this genetic material is encoded by several long pieces of DNA: each individual piece is called a chromosome. In eukaryotic organisms, such as Humans and Fruit Flies, these chromosomes are confined to the interior of the cell nucleus. However, each chromosome is typically longer than the diameter of a spherical cell nucleus. Thus, chromosomes must fold to physically fit within the boundary of the nucleus. This important boundary is called the nuclear envelope. Remarkably, the folded chromosome configurations are non-random and may be important for accessing the encoded genetic material. The effort to understand chromosome folding now encompasses an entire field of chromosome biology. Although little is known about the principles which govern chromosome folding, the nuclear envelope is expected to play a significant role. In fact, experiments suggest that the nuclear envelope engages chromosomes actively by forming attachments which limit the range of 3D motion for particular regions of the genome. The broad goal of the research presented in this dissertation is to advance our understanding of 3D genome organization with an emphasis on determining the role of the nuclear envelope.

Acknowledgements

To Dr. Sharakhov and Dr. Onufriev: Thank you for providing the opportunity to begin my career in your research lab. Thank you for providing positive feedback. There have been times where I have felt a bit lost or confused about the direction of my work, but your encouraging words of wisdom that have put me back on track to finish my projects.

To members of the Sharakhov lab and Onufriev lab: Thank you for being a source of optimism. It has been a very demanding four and a half years, and all of you certainly have made those years much more manageable and fulfilling.

To Abhishiek: Thank you for sharing my cubicle and dark sense of humor.

To Ramu and Igor: Thank you for all the curiosity and excitement about my work which is why I got into science in the first place. Without your outside interest I'd just be writing manuscripts.

To Casey: Thank you for nostalgic trips to Starbucks, thermal problem sets, and enthusiasm for science. Your old cubicle served me well the past four years

To Ashley, Atashi, and Jiyoung: your optimism is contagious. Thank you for letting me vent about the frustrations of research science. All of your support has alleviated a lot of the academic tension that I've had. I'll miss the coffee breaks and cheez-its that helped to brighten my days. I'm quite lucky to be able to call you all friends.

To Phil: Thank you for having the patience to listen to me rant about baseball and sports in general. Time flies when you have some pleasant distractions. Thank you for assistance with the work not present in this dissertation. Working in collaboration really does make science more enjoyable; in a few years we might even have a paper.

To Nicole: During my third and fourth year your violin lessons were essentially all I had to look forward to. Thanks for making me feel good about my accomplishments.

To Vladimir Timoshevskiy, Anastasia Noumenko, and Maria Sharakhova: I cannot thank you enough for all of the help you have provided me. I have learned so much from all of you.

To Kyle, Ross, Omair, and any students I have taught: You guys embody the reasons I want to continue teaching.

To my parents: thank you for my education. I'm still not sure graduate school was a good idea but I could not have done it without your support.

ATTRIBUTIONS

Several collegues aided in the writing and data collection which support the chapters presented as part of this dissertation. A brief description of the contributions is included here.

Chapter 2: Alexey V. Onufriev and Igor V. Sharakhov helped conceive and designed the experiments. Alexey V. Onufriev and Igor V. Sharakhov helped performed the experiments. Igor V. Sharakhov assisted with data analysis. Alexey V. Onufriev and Igor V. Sharakhov contributed reagents, materials, and analysis tools. Alexey V. Onufriev and Igor V. Sharakhov helped write the paper.

Chapter 3: Alexey V. Onufriev and Igor V. Sharakhov helped conceive and designed the experiments. Alexey V. Onufriev assisted with data analysis. Alexey V. Onufriev and Igor V. Sharakhov contributed reagents, materials, and analysis tools. Alexey V. Onufriev and Igor V. Sharakhov helped write the paper.

Chapter 4: Alexey V. Onufriev and Igor V. Sharakhov contributed reagents, materials, and analysis tools.

Chapter 5: Alexey V. Onufriev and Igor V. Sharakhov helped conceive and designed the experiments. Alexey V. Onufriev assisted with data analysis. Alexey V. Onufriev and Igor V. Sharakhov contributed analysis tools. Alexey V. Onufriev and Igor V. Sharakhov helped write the paper.

Appendix B: Alexey V. Onufriev and Igor V. Sharakhov helped conceive and designed the experiments. Alexey V. Onufriev and Igor V. Sharakhov helped performed the experiments. Igor V. Sharakhov assisted with data analysis. Alexey V. Onufriev and Igor V. Sharakhov contributed reagents, materials, and analysis tools. Alexey V. Onufriev and Igor V. Sharakhov helped write the paper.

Appendix C: Alexey V. Onufriev and Igor V. Sharakhov helped conceive and designed the experiments. Alexey V. Onufriev assisted with data analysis. Alexey V. Onufriev and Igor V. Sharakhov contributed reagents, materials, and analysis tools. Alexey V. Onufriev and Igor V. Sharakhov helped write the paper.

Igor V. Sharakhov, PhD, is currently an Associate Professor in the Department of Entomology at Virginia Tech.

Alexey V. Onufriev, PhD, is currently an Associate Professor in the Departments of Computer Science and Physics at Virginia Tech.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

APPENDIX A

APPENDIX B

APPENDIX C

LIST OF TABLES

CHAPTER 1

LITERATURE REVIEW

1.1 BACKGROUND

The central dogma of modern biology links the DNA sequence of the genome with its downstream expression in the form of mRNA and protein, with few exceptions, it is regarded as non-reversible process. In this scheme, information within the DNA polymer is 2-dimensional (2D) and is represented by the sequence of deoxyribonucleic acids polymerized to form the DNA backbone (1); however, DNA is not freely floating within the nucleus. Instead, the 3-dimensional (3D) DNA packaging within the nucleus is highly organized and has recently been linked with gene expression (2,3). Thus, the genome sequence must be understood in the context of 3D chromosome folding which critically influences the flow of information. The effort to understand this added complexity now encompasses an entire field of chromosome biology and is reshaping the traditional concept of the central dogma.

Recently, 3D chromosome folding has been likened to proteins which possess primary, secondary, tertiary, and quaternary structure (figure 1.1) (4,5). In the case of chromosomes, primary structure arises from the sequence of linear features along polymer backbone. These features include the sequence of base pairs, epigenetic modifications, and DNA-protein interactions (4). Local folding of chromosomes gives rise to secondary structures such as the 10nm or 30nm fiber. The stability of these fibers may be dictated by histone acetylation, phosphorylation, and methylation (4). Tertiary structure emerges from the long range interactions within and between chromosomes. The pattern of these interaction may be

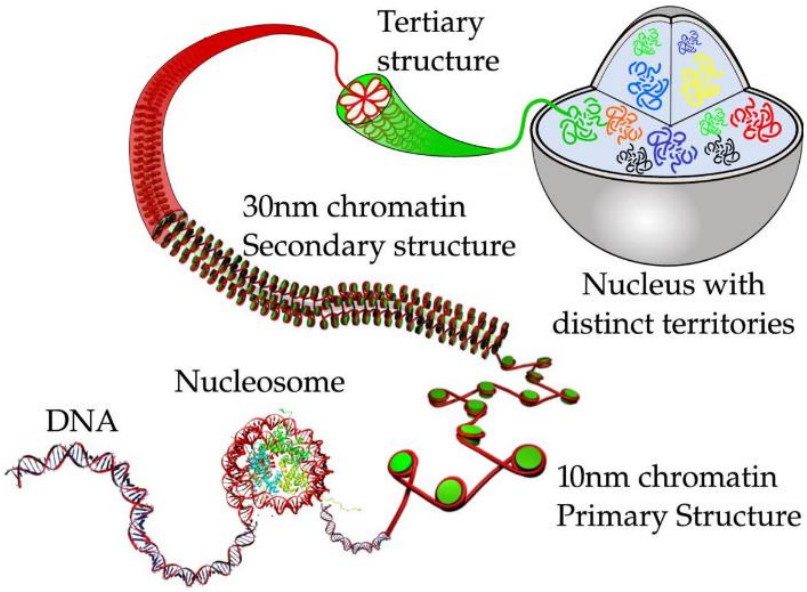


Figure 1.1 – Depiction of the hierarchy of organization in the interphase nucleus taken from Ref. 5

influenced by the positioning of chromosome territories. These territories allow chromosomes to occupy distinct regions of the nucleus but permit a degree of intermingling (6,7). The highest level of genome organization dictates the arrangement of chromosome territories (8,9). Remarkably this hierarchy has been linked directly or indirectly with genome function at every level; however, its governing principles are largely unknown.

The principles that underlie the hierarchy of chromosome organization remain elusive for several reasons; perhaps the greatest challenge stems from the uncertainty in genome organization. Unlike most proteins which possess predetermined folding, the conformational states of the chromatin fiber are stochastic to some degree (10). This element of uncertainty means that single cells cannot reveal the tendencies of genome organization; instead, populations of cells must be considered which increases experimental and computational complexity. It's not surprising that the experimental techniques designed to mitigate this obstacle, such as Hi-C and DamID, have been remarkably successful. The Hi-C technique reveals chromosome interactions by crosslinking, fragmenting, and ligating the genome (11). The DamID technique uses a similar crosslinking strategy to reveal chromosomal interactions with the nuclear envelope (12).

Probing the various levels of chromosome organization has been impeded by additional obstacles. In particular, the resolving power of the light microscope places limitations on the structures that are directly observable in the interphase nucleus. In general, the fine scale folding of chromosomes into 10nm and 30nm fibers is not directly visible; however, whole chromosome territories are visible with fluorescent in-situ hybridization (FISH) (13). This technique has been used to directly characterize the size and shape of chromosome territories (7,9,14), as well as their positioning within the nucleus. Chromosome folding has been indirectly studied with FISH by measuring the spatial proximity of probes with known genomic separation (15-17). For folded chromosomes in the nucleus the relationship between spatial proximity and genomic separation is not linear and may be unique in different organisms. Unlike Hi-C and DamID, FISH is generally limited to small numbers of nuclei.

Older approaches to studying 3D genome organization used the superior visibility of polytene chromosomes (figure 1.2). Each polytene chromosomes replicates without cell division and

forms a bundled fiber of approximately 1024 individual strands aligned in parallel (figure 1.2) (18); consequently, the chromosomes in the nucleus become visible with a light microscope. This is a critical advantage over ‘‘regular’’ interphase chromosomes because it becomes possible to obtain full spatial information about the position of each chromosome in 3D (19). The original studies of Rabl and Boveri used polytene chromosome to detect a segregation of chromosome centromeres and telomeres in the polytene nucleus of Ascaris (20). This motif is now known as the Rabl configuration and has since been detected in multiple species of eukaryotes including fruit fly (21), yeast (22), and wheat (23). Although the Rabl configuration has been established for nearly a century its origin is largely unknown. The characteristic polarization is speculated to be a vestige of the previous anaphase (24,25) but is conspicuously absent in humans (26) and rice (27).

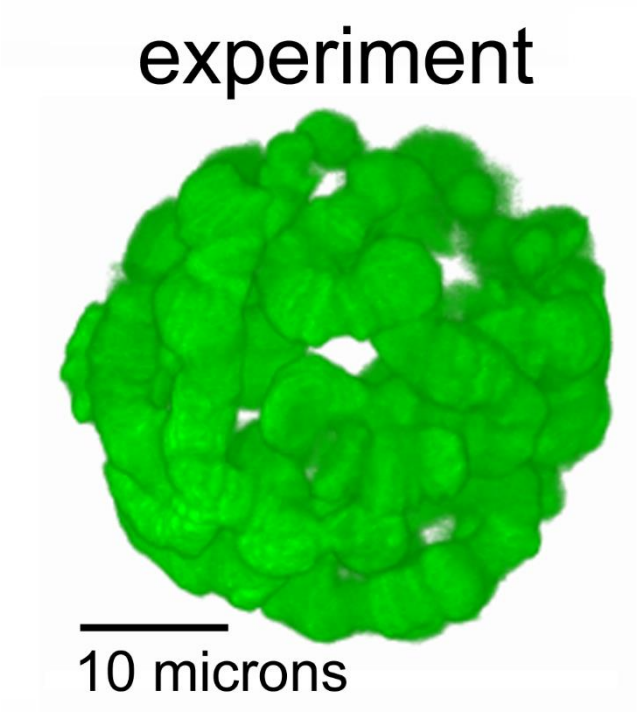


Figure 1.2 – image of a polytene nucleus in D. melanogaster salivary gland.

A series of studies in the 1980’s used confocal microscopy to reconstruct optically sectioned images of D. melanogaster polytene chromosomes in salivary gland cells, midget cells, prothoracic cells, and various mutants (28-33). Several aspects of genome organization were revisited in each study. In most cases chromosomes were found occupying distinct territories within the nuclear volume (30). In salivary gland cells, it was found that territories partition the nucleus into wedge shaped sections with right and left chromosomal arms frequently juxtaposed (30). In all cases chromosomes were seen to interact with other chromosomes and with the nuclear envelope (32). The complexity and uniqueness of chromosome-nuclear envelope

interactions in each cell type was emphasized. In particular, these interactions included both permanent and highly frequent attachments between certain chromosomal loci and the nuclear envelope (32).

The principles that may link chromosome territories, chromosome folding, and chromosome-nuclear enevelope attachments are poorly understood. Is the formation of chromosome territories controlled by the chromosome folding or chromosome-nuclear envelope attachments? Does the number and distribution of attachments influence the patterns of 3D chromosome folding and intra-/inter- chromosomal interactions? Do specific folding patterns and defined chromosome territories allow only certain chromosomal sites to interact with the nuclear envelope? Both chromosome-nuclear envelope attachments and chromosomal interactions appear stochastic to some degree, as in Dam ID and Hi-C experiments (34,35). The stochastic nature of genome organization means computational models will be crucial tools for complementing experiments. Computation will be essential if the goal is to understand the mechanisms that establish and maintain nuclear architecture.

Several recent experiments have renewed an interest in the 3D genome organization of polytene chromosomes. This revival stems in large part from an emerging correspondence between 3D genome organization in polytene chromosomes and 3D genome organization in their non-polytene counterparts. In the case of *D. melanogaster*, the correspondence is supported by multiple lines of evidence. First, polytene and non-polytene cell line chromosomes have a highly similar pattern of bands and inter-bands which likely reflects their basic level of organization. This identity has been revealed by comparing localization data of interband- and band-specific proteins from modENCODE and the genomic position of interbands and bands (36-39). Second the profile of chromosome nuclear envelope attachments is highly similar in polytene and non-polytene chromosomes (12). This conclusion has been reached by comparing DamID data obtained from embryonic cells with light microscopy mapping studies of the nuclear envelope-polytene chromosome contacts. Third, the polytene chromosomes occupy about a third of the nuclear volume in fruit fly (31); this chromosome to nucleus volume ratio is the same in regular non-polytene nuclei (40) and suggests a similarity of general folding principles.

1.2 THEORETICAL BACKGROUND

The importance of chromosome volume in relation to nucleus volume can be demonstrated by a simple polymer physics argument. In general, there are two concepts of distance between loci belonging to the same chromosome (41,42). The genomic distance, *s*, is a measure of linear separation along the polymer backbone; the spatial distance, R, is a measure of the 3D separation. These distances may differ considerably for the highly folded chromosomes in the nucleus. In polymer physics this relationship is captured by the expression $R(s) = s^{\alpha}$ with the scaling exponent α reflecting the polymer folding. For a completely straight chromosome R and s should coincide which implies a value of α close to 1. However, loci confined within the boundaries of the nucleus must have spatial separation less than the diameter, D, of the nucleus. This basic requirement implies the following inequality: $s^{\alpha} \leq D$. Since the length of each chromosome is greater than the diameter of the nucleus, α takes a value less than one. Thus, the value and behavior of the exponent α quantifies one of the primary scaling concepts in polymer physics (41).

Determining the precise value of the scaling exponent α is an experimental and computational challenge. Computational studies of chromosome folding can predict its value by simulating chromosomes in confinement. For chromosomes in equilibrium these simulations suggest that α ranges from 1/3 to 1/2 depending on the degree of polymer confinement, monomer attraction, and excluded volume repulsion (43). Chromosomes in equilibrium often possess the additional characteristic that α vanishes for large genome distances (43). This scaling behavior indicated that a chromosome can explore the entire nuclear volume and generally does not predict the presence of chromosome territories. A model that does predict the presence of chromosome territories was originally proposed by Grosberg in 1988 based on the concept of the crumpled globule (44). The crumpled globule is one of the key theoretical shapes in the field of genome organization (11).

A crumpled globule can be achieved from an elongated polymer by the recursive introduction of crumples (44). First, crumples are introduced locally along the polymer backbone forming a "fiber of crumples". Next, larger crumples are introduced along the thicker fiber formed from the

initial polymer collapse. Reiterating this process results in a structure that possess self-similarity on all length scales (44). In other words, every linear segment of a crumpled globule is itself a crumpled globule. On account of this self-similarity the configuration has earned the sobriquet "fractal globule" (44,45). The crumpled globule has a distinct scaling exponent α and several attractive properties that suggest it is a biologically favorable configuration. In particular, α does not vanish for large genomic distances which reflects the configuration's territorial organization (45). This territorial organization in turn limits chromosome entanglement and makes the crumpled globule easy to fold and unfold. Most importantly, the configuration is consistent with experiments that probe the fine structure of human chromosomes (11). Specifically, the primary scaling law ($R = s^{\alpha}$) of the fractal globule reiterates the pattern of real chromosomes revealed by Hi-C experiments.

In the field of genome organization, a second scaling concept relates the volume of the nucleus and surface area of its envelope. Several theoretical studies of polymer brushes illustrate why this relation matters (46-48). These brushes consist of polymer chains end-attached to an interface and recently they have been likened to the genome organization in yeast (49). Analytical treatment of brush systems in various geometries has been developed using lattice models with differences in curved and planar geometries underpinning key theoretical results (46,47). In particular, it was noted that the number of lattice sites increases moving away from the center of a curved geometry (47). Although these analytical studies were not specific to genome organization, they highlight the importance of volume and space. Consider the salivary gland cells of D. melanogaster. The nuclei of these cells are approximately spherical with 15 micron (μ) radius amounting to a volume on the order of 10^4 μ^3 (30). However, the surface area of the nucleus is on the order of 10^3 μ^2 which severely restricts the placement of loci attached to the envelope. To be clear, the restriction arises because the volume of the nuclear periphery is small in comparison to entire volume of the nuclear enterior. Thus, the relation between volume and surface area may be important for understanding the effects of chromosome-nuclear envelope attachments. This is discussed in chapter 3.

This interplay between volume, space, and chromosome nuclear envelope attachments has an interesting corollary. In particular the volume of the nucleus and surface area of its envelope do

not scale in a linear relation; thus, the ratio of these two measurements may differ greatly in different cell types. In D. melanogaster three of these ratios are known. In cells of the salivary gland, nuclei have a volume to surface area ratio of ~6.0 μ (31). In prothoracic cells the same ratio is ~3.8 μ (31). Midgut cells have a ratio of ~5.0 μ (31). Moreover, the number and position of chromosome nuclear envelope attachments in each of these cells types is known to vary (31,32). These observations indicate that chromosome-nuclear envelope attachments must be understood in the context of specific cell types. Since the effects of chromosome-nuclear envelope attachments are probabilistic, it's essential to model entire computational ensembles with realistic parameter sets. The design of ensembles to represent different cells types is discussed in chapter 4.

In a more general sense, chromosome–nuclear envelope attachments, along with the size and shape of the nucleus, can be thought of as "boundary conditions", which may affect multiple layers of genome organization in interphase. Nucleus size and shape were thoroughly investigated in a computational study of human chromosomes (50). By simulating chromosomes under confinement it was demonstrated that territories become centralized by increasing the eccentricity of their confining volume (50). This prediction has even received some experimental support in plants. Essentially, the comparison of genome size and nucleus size in a number of plant species revealed that the two are related isometrically (51). This observation is thought to reflect the importance of geometrical constraints and rule out alternative constraints based on membrane transport (51). However, chromosome-nuclear envelope attachments were not considered as part of the boundary in these studies. This omission is important since chromosome nuclear envelope attachments add a great deal of complexity to the boundary conditions of the nucleus.

1.3 COMPOSITION OF ATTACHMENTS

The complexity of chromosome nuclear envelope interactions was first demonstrated in studies of polytene chromosomes in salivary gland cells of D. melanogaster. Optical reconstruction of each polytene nucleus revealed the specific points of contact between each chromosome in the nuclear envelope (30). Although the number and positioning of contacts varied in individual

nuclei, enumerating contacts in an ensemble of nuclei revealed their overall distribution. In the aggregate, these positions were used to infer the probability of nuclear envelope contact along each chromosome. A statistical guideline was used to identify 15 positions in the genome that frequently juxtapose the nuclear envelope (30). These 15 attachments coincide almost exclusively with regions of intercalary heterochromatin – gene poor, dark staining, late replicating regions of the genome. The original set of 15 attachments was later confirmed and augmented by multiple studies using a variety of experimental approaches. In 1988, Polytene reconstructions were again used to identify sites of chromosome nuclear envelope attachment in *D. melanogaster* with an inverted X chromosome, ring X chromosome, or compound XY chromosome (33). This study confirmed 12 of the original 15 attachments and identified four additional nuclear envelope-contacting sites at 97A, 19DE, 60EF, and 61AB (33).

Two studies used a fluorescent in situ hybridization (FISH) approach to identify chromosome-nuclear envelope attachments along the left arm of chromosome 2 (2L) in *D. melanogaster* embryos. In 1996, a set of 42 FISH probes reiterated the peripheral localization of 3 chromosome regions (22A-B, 32F-33A, and 34F-35C) from the original set of 2L attachments and identified at least 3 others (23A, 34A, and 87B) (52). It was speculated that these three additional attachments were simply lost during polytenization in the salivary gland nuclei. In fact, a total of 16 probes were found in close proximity to the nuclear envelope along 2L demarcating 1-2Mb chromatin loops (52); five of these coincided with regions euchromatin – gene rich, light staining, early replicating regions of the genome. It was concluded that intercalary heterochromatin is not necessary or sufficient for nuclear envelope association (52).

In 2004, 2L was successfully reconstructed using a smaller set of 13 color coded FISH probes (53). This study found only 1 probe frequently in the nuclear periphery with 5 occupying the nuclear interior. Although 2L possessed 4 attachments sites in the original set of 15, it was noted that the 13 FISH probes did not completely overlap with the collection of sites identified in 1986 and 1996 (30,52,53). It's more surprising that 5 chromosome regions preferentially occupying the nuclear interior went unidentified in the original 1986 study. Were these 5 regions undetectable in the original study, overlooked, or simply absent in nuclei of the salivary gland? Furthermore, the discovery of euchromatin regions interacting with the nuclear envelope in 1996

revised the composition of nuclear envelope attachments. Were these euchromatin regions overlooked in the original study? These questions are addressed in chapter 2 by reassessing the original 1986 data.

In 2007 FISH probes were used to study the localization of *D. melanogaster* telomeres inside salivary gland nuclei (54). These telomere regions are composed of three retrotransposons (HeT-A, TAHRE, and TART) and satellite DNA repeats. Probes designed to hybridize to the HeT-A retrotransposons revealed that telomeres predominantly occupy the nuclear interior. Furthermore, HeT-A hybridization did not reveal clustering of telomeres (54); it was concluded that ectopic contacts between telomeres may vary randomly. These results are surprising since three different nuclear envelope contacts identified in 1986 are positioned near the telomere regions of chromosomes 2, 3, and X respectively (30). It's hard to reconcile these two studies of salivary gland nuclei. The statistical threshold used to identify chromosome-nuclear envelope contacts in the 1986 study is clearly stronger. In fact, no statistical threshold was used to evaluate nuclear envelope proximity in the 2007 study and the conclusions were mainly subjective (54).

Some FISH studies have directed focus on particular sequences of DNA thought to interact with the nuclear matrix. This matrix encompasses three structures thought to directly interact with interphase chromosome: the nuclear lamina, the cytoskeleton protein network, and nucleoli (55). FISH Studies of the AT-rich DNA fragment λ20p1.4 in *D. melanogaster* polytene chromosomes revealed localization at the chromocenter and to the 49CD region on the right arm of chromosome 2 (2R) (56); both of these region were juxtoposed the nuclear envelope in 3D images (56). Separate assays were used to demonstrate that λ20p1.4 binds major lamina polypeptides that form the meshwork of protein forming the nuclear lamina. This was the first demonstration that DNA sequences capable of binding lamins may also be found in proximity to the nuclear envelope. Ironically this finding only added to the complexity of nuclear envelope attachments. No homology was found between λ20p1.4 and previously identified matrix/scaffold attached sequence motifs (56,57). Thus, no single family of sequences can fully predict the interactions between chromosome and the nuclear matrix. In addition, the 49 CD region was not a member of the original 15 nuclear envelope attachments identified in 1986 (30).

The common thread in the various studies of chromosome nuclear envelope attachments is their complexity. There seems to be no consistent way of detecting attachments based on chromatin type or sequence motifs. The most complete and rigorously identified attachments continue to be the set of 15 identified in 1986; however, the statistical threshold defining their identity was developed without the aid of polymer model (30). It's unclear if the same 15 attachments attachment would be identified by a biologically realistic model of confined chromosomes; conceivably, different positions along chromosome have an intrinsically different chance to contact the nuclear envelope. Thus, a suitably designed model could confirm the 15 established regions of attachments and potentially discover new ones. This is important since the intercalary heterochromatic regions in D. melanogaster number more than 100 (58) and the complete rules for chromosome positioning with respect to the nuclear envelope remain undiscovered. These topics are addressed in chapter 2.

The 15 chromosome nuclear envelope attachments in D. melanogaster salivary gland have embraced even greater significance since the discovery of a correspondence with non-polytene interphase chromosome. This discovery was made with DamID - a molecular mapping approach that identifies nuclear envelope attachments as they occur *in vivo* (12,59). It has been found that lamin binding is linked to a combination of several features including late replication, large size of intergenic regions, low gene expression status, and the lack of active histone marks (12). Another study (36) compared localization of lamina-associated domains (LADs) and 60 regions of intercalary heterochromatin (IH), i.e., heterochromatin located inside chromosomal arms. Complete overlap was observed for four regions of IH. Most of the IH regions partially overlapped with LADs, six regions did not overlap with any of the LADs, and one region of IH encompassed five separate LADs.

1.4 EXPERIMENTAL APPROACHES

The frequent association between heterochromatin and the nuclear envelope has led to the general notion that the nuclear periphery is a region of transcriptional inactivity (60,61). Groundbreaking work in *D. melanogaster* demonstrated that this idea links the specific sites of chromosome-nuclear envelope attachment with gene expression. In wild type D. melanogaster,

the brown gene is required for pteridine pigment in the eye (62). Expression of the gene is disrupted when placed in the proximity of heterochromatin (62). This phenomenon is called the position effect variegation. The brown gene is unusual because a single variegated allele also silences its partner in trans. In the 1990's several explanations of this "trans-inactivation' were proposed that shared a common thread. Essentially, the variegated gene was thought to drag its paired homolog into physical association with centromeric heterochromatin. This results was later verified experimentally with the corollary that variegation places both copies of the brown allele in the vicinity of the nuclear envelope (63).

Even earlier studies in yeast succeeded in directly measuring the transcriptional activity of genes artificially anchored to the nuclear periphery. Normally, the yeast telomeric loci and silent mating loci (HML and HMR) are flanked regions that bind various proteins and recruit a set of 4 silent information regulator (SIR) proteins (64,65). These SIR proteins interact with histones to transcriptionally deactivate contiguous regions of chromatin (64,65). Interestingly, telomeres tend to cluster together in yeast at the nuclear periphery and sequester the pool of SIR proteins. It has been demonstrated that tethering an HMR locus with defective silencer at the nuclear periphery near telomeric pools of SIR proteins restores its state of transcriptional inactivity (64,66,67). On the other hand, the silent state of HMR cannot be restored if its E silencer is completely removed. These studies suggest that chromosome-nuclear envelope attachments are part of a collections of factors that determine transcriptional activity.

Experiments that aim to study the effects of chromosome-nuclear envelope attachments on large scale genome organization have used both mutant and drug based approaches. Mutant approaches typically investigate the spectrum of disease states caused by defective lamin proteins (68). These lamins in large part support the meshwork of protein forming the nuclear lamina; consequently, the nuclei in mutants become amorphous and lack chromosome nuclear envelope attachments (69). In particular, heterochromatin in effected nuclei often detaches from the nuclear envelope. These diseases are commonly called "laminopathies". Drug based approaches typically investigate the outcomes of chemically induced laminopathies (70). Drugs that inhibit farnesylation have been widely used to prevent maturation pre-lamin A (70). The common thread in these approaches is that they alter the boundary conditions of the nucleus.

Evidence suggest that chromosome-nuclear envelope attachments may even affect the dynamics of chromosomes in interphase (71-73). Typically, the dynamics of chromosomes are studied experimentally by tagging individual loci with GFP and tracking their motion over several minutes (72,73). Most experiments agree that chromosome motion is in the Brownian regime. This type of motion has been identified by plotting the mean squared deviation (MSD) in position of individual loci over time. In free space a Brownian MSD curve is linear with a diffusive coefficient equated to its slope by the Einstein– Smoluchowski relation (72). If diffusion is constrained, the MSD response curve is concave down and reaches its plateau at the confinement radius. Indeed, most experiments reveal a concave down MSD curve for chromosomes in interphase (72-74). However, the origin of this constraint is unclear since the confinement radius of chromosomes in interphase does not correspond to the radius of the nucleus. This is surprising since it means that individual chromosomal loci are not able to explore the full interior of the nucleus. It has been suggested that this additional constraint could be achieved by chromosome nuclear envelope attachments (72).

Unfortunately, experimentally tracking the dynamics of individual chromosome loci is generally limited to several minutes due to photobleaching of florescent markers. Given these experimental limitations, computational simulation has emerged as a pivotal tool in the study of chromosome folding and dynamics in the interphase nucleus. In simulation, no temporal compromises are necessary and accelerated computational search can produce simulations that represent several days in reality (75). This absence of temporal compromise is important because experiments suggest that chromosomes possess distinct motions at different time scales. In *D. melanogaster* short time scale Brownian motion on the order of minutes coexists with the gradual loss of the Rabl configuration over the course of several hours (76).

1.5 COMPUTATIONAL APPROACHES

Computational models have the added advantage that the number and position of chromosome nuclear envelope attachments can be precisely manipulated. Development of a computational models generally occurs in three phases: parameterization, validation, and prediction. Ideally the

parameters of computation model can be taken directly from experiment. For models of genome organization these may include the size of the nucleus and length of each chromosome. Properties of model that arise without the assistance of specific parameters can be validated against experiment or interpreted as theoretical predictions. However, the effort to identify model parameters is often challenging. Although most studies treat chromosome nuclear envelope attachments as model parameters (22,77-80); some argue that they form without specification (81,82).

A recent study demonstrated that non-specific forces alone may be sufficient to localize centromeric heterochromatin to the nuclear envelope in Arabidopsis (81). Although this study was specific to Arabidopsis, it was emphasized that the result was robust to chromosome topology. Centromeric heterochromatin localized to the nuclear periphery with chromosomes arranged as linear chains, loops, and rosettes. Another study demonstrated that non-specific forces can position compact thick fibers mimicking heterochromatin at the nuclear periphery and long thin polymers to the interior (82). On the other hand, most studies in yeast specifically include centromere and telomere nuclear envelope attachments as model parameters. These studies have demonstrated that the 3D position of a gene can be altered due to the presence of an nuclear envelope tether positioned within 10 kb (80) and that removal of chromosomal tethers at the centromere increases chromosome mobility (71). Possibly the positioning of centromeric heterochromatin and intercalary heterochromatin arise from unique circumstances; however, this question has not been explored previously. Chapter 2 uses computational modeling to test the possibility that geometrical constraints alone position intercalary heterochromatin in the nuclear periphery.

Including chromosome-nuclear envelope attachments in computation models is difficult for two reasons. First, attachments are to some degree stochastic meaning that they vary in number and position in individual cells (30); consequently, their statistical effects can only be determined by modeling entire ensembles of chromosome configurations. Second, live imaging of yeast and fruit fly chromosomes have revealed that chromosome motion in interphase is Brownian (72,73). The diffusion constant of individual loci depends on nuclear localization; loci attached to or adjacent to the nuclear periphery are less mobile than other loci (83). Modeling these dynamics

generally involves integrating the Langevin equation of motion. These two requirements are incompatible. It's computational infeasible to model dynamic simulation for a large ensemble of chromosome configuration. Fortunately these two requirements are separable. Large computational ensembles of nuclei that represent "snapshots" can be used to study the statistical effects of chromosome-nuclear envelope attachments. Much smaller ensembles of nucleus can be time evolved with the Langevin equation to predict dynamics.

1.6 COMPUTATIONAL ENSEMBLES

The starting points for most computational ensembles is the random walk (42). In its simplest one-dimensional form, the random walk can be conceptualized as a series of steps beginning at the origin and randomly moving in positive or negative units along the real line with equal probability. Since the probability of moving in either direction is ½, any n-step sequence has a probability of $(½)^n$. In this simple form various properties of the random walk can be analytically derived; for example, it's well known that the collection of these n-step sequences form a binomial distribution (40). Thus, the average displacement, *d*, of these walks is related to the number of steps: $d^2 \sim n$. However, even simply modifications of the random walk quickly render its statistical properties analytically unsolvable. In this case, statistics can be recovered by generating walk configurations evenly sampled from the space of all configurations. Acceptable models of genome organization are expected to recapitulate short range and long range chromosome interactions which requires sampling a large number of configuration.

In the field of genome organization, early use of the random walk introduced the random walk giant-loop model (RWGL) (16) and multi-loop sub-compartment model (MLS) (84). In the RWGL, loops averaging 3Mbp were linked along a random walk backbone. This model was analytical and using a minimal set of three parameters demonstrated good agreement with fluorescence in situ hybridization data. However, the use of fitting parameters was later criticized. The MLS model placed a greater emphases on chromosome territories using smaller 120kbp loops to form sub-compartments dictated by the banding pattern of human chromosomes (84). A computational ensemble of 200 configuration was used to compute statistics of the

model. It was emphasized that the basic loop size was not used as a fitting parameter yet distance scaling indicative of a globular state matched the experimental data.

However, the basic random walk makes a critical simplifying assumption. Specifically, the path taken by the walk is allowed to self-intersect. In reality, excluded volume interactions between monomers prohibits polymer self-intersection. These interactions can be incorporated by representing the polymer as a coarse grained series of non-overlapping beads colloquially called a "beads-on-string" model or self-avoiding walk (SAW) model (41,42,85). The diameter of each bead is typically equated with the Kuhn length of polymer. Since the Kuhn length is twice the polymer persistence length, the flexibility of the polymer is automatically captured by the bead excluded volume (42). Variations of SAW approach have been successfully used for decades to generate computational ensembles of 3D chromatin organizations. The random loop model (RL) attempted to improve on the RWGL model and MLS model by incorporating multiple loop sizes with attachment points allows to vary from cell to cell (17). Statistics of the RL model averaged over large conformational ensembles demonstrated good agreement with experimental results, particularly for large genomic distance.

Nuclear envelope attachments were largely absent from the early computational ensemble models of genome organization proposed in the early 1990's. This omission is somewhat surprising since chromosome-nuclear envelope attachments and their role in nuclear organization had been well established decades earlier. Indeed review articles on the topic of chromosome-nuclear envelope attachments as early as 1974 outline conclusive evidence that "DNA is stably attached to the nuclear envelope of the interphase nucleus with multiple points of attachment per chromosome" (86). By this time, chromosome-nuclear envelope attachment had already been included in schematic models of interphase nuclear organization (87). These early models were based on the hypothesis that multiple sites of chromosome nuclear envelope attachment guide the non-random chromosome configurations of the interphase nucleus. Paradoxically, these early schematic models were largely overlooked when computational modeling of ensembles became feasible and routine.

Recent works in yeast have incorporated limited sets of chromosome nuclear envelope attachments in ensemble based models of nuclear organization (22,77-80). In most models these attachments are critical members of minimal parameter sets that dictate genome organization. Casting attachment in this important role tacitly assumes that they effect genome organization; however, these effects have not been fully characterized. So far it has been demonstrated that 3D gene positioning is altered due to the presence of nearby attachments (80). Furthermore, removal of chromosomal tethers at the centromere increases chromosome mobility (71). It's somewhat less clear how attachments affect the proximity of chromosome telomeres. Some studies argue that the distribution of telomere-telomere distances does not depend on the number of attachments (88). Other studies suggest that telomere tethering tends to increase their spatial separation (77). Possibly, the effects of chromosome nuclear envelope attachments depend on their number as well as their placement.

Compared to yeast, fruit fly possess a larger set of chromosome nuclear envelope attachments (30). This larger set of attachments biologically grounds an investigation of attachment number and distribution. On top of this, two lines of evidence suggest that the effects of chromosome-nuclear envelope attachments differ in yeast (22) and fruit fly (89). First, the folding parameter, α, which characterizes the scaling of chromosomal interactions differs greatly in *D. melanogaster* and yeast (22,89). In *D. melanogaster* the magnitude of α is smaller and non-vanishing for large genomic distance; this implies that chromosomes in *D. melanogaster* are more likely to interact over large genomic distances. Consequently, chromosome nuclear envelope attachments could have a broader impact on genome organization in *D. melanogaster* compared to yeast. Second, the yeast nucleus likely lacks well-defined chromosome territories found in fruit fly or human (90-92). Thus, the effect of attachments on this aspect of genome organization cannot be studied in the yeast genome. These ideas are modeled and made quantitative in chapter 3.

Possibly, more numerous sets of chromosome nuclear-envelope attachments have been omitted from ensemble based models due to computational complexity. In general, generating a computational ensemble requires even sampling of "configuration space". For SAW models this task is notoriously tricky and requires a carefully formulated Monte-Carlo procedure (93-95).

Typically, these procedures pivot and rotate segments of an initial configuration or use dynamic simulations to reach unique equilibrium configurations. The Rosenbluth algorithm is a computationally less expensive approach that adds beads sequentially to one end of the growing polymer chain (96). For short chains this approach is a good approximation of self-repelling chains which are true equilibrium states of polymers (97). Although this approach has seen some use in the field of protein folding (98), its computational efficiency has been mostly overlooked. Chapter 2 demonstrates that this overlooked efficiency is critical for polymer models that enforce a large number of constraints such as those associated with the placement of numerous chromosome-nuclear envelope attachments.

1.7 DYNAMICS

Dynamic models of chromosome organization possess their own unique set of challenges. First, dynamics polymer models must contend with many degrees of freedom such as fluctuating bond lengths and bond angles that join the individual monomers (99). Often computational simulations are made tractable by eliminating degrees of freedom using a process called "coarse graining". In this scheme the behavior of a polymer is preserved in the simplified representation of soft beads connected by springs (99). Typically, the diameter of each bead is identified with the polymer Kuhn length (42). This reduces the number of fluctuating bond length and permits complete removal of bending potentials that otherwise increase computational complexity. However, determining the force extension of the connected springs is highly non-trivial and continues to be a subject of ongoing research (100). Many coarse grained simulations of chromatin simply use Hook springs (81) with a force constants sufficiently high to preclude crossing of individual strands.

The simulation of a bead-spring polymer model is typically implemented with the Langevin equation (101-104). This is well justified empirically and theoretically. Theoretically, the dynamics of each bead depend on the connecting springs, bead-bead interactions, friction, and collisions with energetic molecules of the solvent. Since the polymer is coarse grained, it's safe to assume that the period of the connecting springs and non-bonded interactions is much longer than the elapsed time between consecutive collisions of solvent molecules. This allows solvent

collisions to be modeled as random Gaussian noise uncorrelated in time and space. This is exactly the approach implicit in the original Langevin model of Brownian motion (103,104). Indeed, empirical observations of GFP tagged loci have confirmed that chromosome dynamics in interphase is Brownian (73). However, the Langevin equation uses a key parameter, λ, to balance friction and solvent collisions; determining the appropriate value of this parameter is challenging.

In Langevin's original 1908 paper (104) the viscous resistance constant, λ, was determined by the Stokes' formula ($\lambda = 6\pi\mu a$). Here μ is solvent viscosity and a is the radius of the Brownian particle. In most coarse grained polymer models this approach implies a large value of λ which severely limits the simulation time step. Essentially, any increase in the value of λ decreases the characteristic relaxation time of the viscous resistance force which quickly becomes the simulation's limiting time scale. Often, artificial lowering of λ is used to increase the simulation time step without affecting the thermodynamic sampling of polymer configuration space. Indeed many polymer simulations simply set λ low enough that non-bonded interactions limit the simulation time step; usually this is done by setting λ to the inverse of the Lennard-Jones (simulation) time (105-108).

Simulations that aimlessly reduce λ may also reduce their biological relevance. Essentially, artificially reducing λ implies abandoning realistic time scales in a simulation in favor of rapidly exploring configuration space. In fact, a simulation that combines coarse graining and λ reduction can exceed the finite lifetime of most cells in interphase (75). Without knowledge of how such a simulation maps to reality it's impossible to predict what happens during the cell's lifetime. Fortunately, experimental data can be used to restore the realistic time scales. Usually this is done by comparing mean squared displacement (MSD) of loci tracked in experiment and simulation (74,87). The problem reemerges in long simulations that greatly exceed the range of experimental data. Chapter 4 explores a new way of assigning realistic time scales to long simulations. The result demonstrates that simulations not only reproduce the initial slope of the MSD curve, which would be trivial, but also reproduce the more complex experimental diffusive motion of interphase chromosomes in the nucleus.

1.8 THERMODYNAMIC CONSIDERATINS

Thermodynamic considerations are important for computational models of all types; however, it's often difficult to determine when equilibrium conditions are appropriate. In the case of *D. melanogaster*, polytene chromosomes adopt the Rabl type configuration with chromosome centromeres and telomeres clustering in opposite poles of the nucleus (30). It has been speculated that the Rabl configuration of chromosomes is a remnant of anaphase, which in the presence of chromosome-nuclear envelope attachments, may be a long lived non-equilibrium state. However, the nature of Rabl configuration is not completely clear; an alternative possibility is that formation of chromosome-nuclear envelope attachments trap chromosomes in a polarized configurations within the nucleus which remain polarized even after reaching equilibrium. Added to this is the observation that polytene chromosomes in D. melanogaster possess a distinct right handed chirality (30). Thus, it's possible that polytene chromosomes are not in equilibrium which adds complexity to the formulation of ensemble based and dynamic models of genome organization.

Greater consensus has been reached regarding the thermodynamic state of non-polytene interphase chromosomes in human (109) and yeast (22). In yeast, experimental measurements and theoretical estimates suggest that interphase chromosomes are indeed in equilibrium (110-114). The scaling law of yeast chromosomes determined by FISH plateaus in a manner consistent with the equilibrium globule or swollen coil (43) and it has been suggested that the relatively low density of chromatin within the nucleus accommodates these shapes. Indeed, a number of successful theoretical models of the yeast genome assume equilibrium (71,77-79,88). In contrast, equilibrium models cannot explain genome organization in humans and fruit fly (11,89). Non-equilibrium conditions must be invoked to reproduce the experimentally verified chromosome territories and pattern of chromosomal interactions.

The design of equilibrium based ensemble models is well understood because thermodynamics can be completely encoded by the Hamiltonian, H. In this case the probability of each microstate is proportional its Boltzmann factor: $\exp(-H/K_bT)$ (42). For random walks and homogeneous self-avoiding walks every microstate has the same probability and simple Monte Carlo

procedures ensure even sampling of configuration space. Outside of equilibrium the probability of each microstate is highly non-trivial and not simply proportional to its Boltzmann factor. Thus, sampling the configuration space outside of equilibrium depends on the initial conditions and history of the evolving system. In computational models, methods to sample such a complex configuration space generally depend on costly time-evolving of individual microstates (115-118). This is a problem because ensemble models of genome organization require a large number of configurations to make meaningful statistical predictions. Chapter 2 introduces new strategies to generate chromosome configurations when equilibrium cannot be assumed.

Interestingly, the crumpled globule is a non-equilibrium state which inevitably transitions to equilibrium (45); however, the time scales necessary to reach the equilibrium state are highly debated. It was originally conjectured that the equilibrium state is reached by reptation of the polymer ends through the entire globule configuration (105,119). This implies that the time necessary to transition to the equilibrium state is highly dependent on the polymer length. Indeed computational simulations have found that transition to equilibrium is on the order of N^3 where N represents the number of monomers in the polymer chain. In 2008 this protracted transition to equilibrium was reiterated in computational study of human chromosome (75): it was estimated that the entanglement time may exceed 500 years and that chromosomes in interphase never reach the equilibrium state (75).

Other studies seem to disagree. It was recently suggest that fractal-like configurations exist along a spectrum connecting open chromatin at one extreme to compact chromatin at the other (120). In the strings and binders switch model (SBS), this spectrum of configurations is explored by altering the affinity and concentration of binder molecules that mimic the cell's DNA binding machinery. The SBS model suggests that the fractal-like configurations occurring during the transition from open to compact chromatin states may be fleeting in the presence of topoisomerases (120). These topoisomerases allow double stranded DNA to freely cross and operate efficiently on bare DNA and nucleosome bound DNA (121,122). Simulations of the crumpled globule that allow strand crossing rapidly transition to the equilibrium state (43). Thus, the duration and stability of fractal-like configurations as seen in experiment remain largely unknown.

Although the native stability of the crumpled globule continues to be debated, it's generally agreed that the configuration is not stable when substantial strand crossing is allowed (43,120). It's this point of agreement that will likely resolve the debated lifetime of the crumpled globule. Although it's known that topoisomerase has the capability of passing doubled stranded nucleosome bound DNA, little is known about the extent of topoisomerase activity in the interphase nucleus. Nevertheless, computational studies have investigated ways to stabilize the crumpled globule in scenarios where strand crossing is allowed. These studies have demonstrated that crosslinks can significantly prolong the lifetime of the fractal globule (43). These crosslinks could represent reversible DNA-protein interactions that topologically preserve the fractal globule while maintaining its biologically attractive properties.

It has been hypothesized that interactions between chromosomes and nuclear substructures may also stabilize the crumpled globule (43). These interactions possibly include anchoring to the nuclear matrix or chromosome-nuclear envelope attachments. Despite this speculation, experimental and computational support is lacking. This omission is surprising for two reasons. First, a recent experiment identified ~500 D. melanogaster genes in close proximity to the nuclear envelope *in vivo* (12). Since the position of these genes is known, computational studies can design a realistic heteropolymer to represent the profile of chromosome-nuclear envelope interactions alone each chromosome. Second, Hi-C studies suggest that D. melanogaster chromosomes fold into a configuration that has scaling properties similar to the crumpled globule. Specifically, the probability of chromosomal contact scales inversely with genomic distance: $P(s) \sim s^{-1}$ (89). Regardless of topoisomerase activity in the nucleus, these chromosome-nuclear envelope interactions likely affect the dynamics of the crumpled globule. Thus, simulation of the D. melanogaster genome is well motivated biologically. This topic is discussed extensively in chapter 5.

The notion that chromosome-nuclear envelope attachments preserve the topology of the genome is consistent with experimental studies of chromosome motion in interphase. Live imaging of yeast and *D. melanogaster* chromosomes have revealed short time-scale Brownian movements confined within variably sized domains (72,73). The movement of chromatin loci depends on

their nuclear localization; loci attached to or adjacent to the nuclear periphery are less mobile than other loci. Consequently, attachments to the nuclear envelope could constrain chromatin folding and prevent chromosome territories from freely diffusing in the entire nuclear volume. This idea received some support in 2001 with live imaging of *D. melanogaster* G2 spermatocytes (72). It was demonstrated that transition through the G2 phase of the cell cycle coincides with the complete arrest of long range chromosomal motion. This abrupt change was speculated to involve interactions between the chromosomes and the nuclear envelope (72).

1.9 METRICS

Quantifying the effects of chromosome nuclear envelope attachments in theoretical models and experiments has been impeded by additional challenges. Conceivably, these effects could influence the statistics of structural ensembles or the dynamics of individual configurations; however, the existing metrics designed to quantify chromosome territories, intertwining, and orientation have limitations. In some cases these features are only given qualitative assessment (30). Addressing this issue is important for two reasons. First, qualitative accounts are difficult to incorporate in computational models. This limits the predictive power of models designed to guide experiment. Second, robust metrics can add sophistication to our understanding of genome organization since establishing statistical significance, correlations, and rates of change are virtually impossible using qualitative assessment alone.

Well defined metrics have added sophistication to the notion of chromosome territories. In human, procrustes analysis of territory size and shape has revealed degree of intermingling on their boundaries (6). The volume of intermingling has in turn correlated with the pattern of chromosomal rearrangements. In general, this type of analysis involves constructing a shape that tightly enclose the volume of a chromosome territory; well-designed shapes limit complexity and have biologically meaningful parameters. In simple schemes the radius of a sphere (123) is used to quantify the volume and position of a chromosome territories. In more complex schemes, the three semi-axes of an ellipsoid together quantify the eccentricity (124), volume, and orientation of a chromosome territory. Although these approaches can quantify territory morphology, they

are less effective for quantifying the mutual exclusion of territories. This fundamental feature of chromosome territories is made quantitative in chapter 2 using the concept of a convex hull.

The concept of polymer intertwining has been quantified previously using Alexander polynomials (125-129). These polynomials are knot invariants and useful for identifying equivalent knots and quantifying knot complexity. Computing the Alexander polynomial of protein structures has led to biologically meaningful results. In particular, the presence of knots may stabilize a protein's native folded state (130). Recent computational studies of genome organization have adopted use of Alexander polynomials for detecting intertwined chromosome configurations; however, the biological interpretation in this context is less clear. Specifically, knot complexity does not directly quantify the ability to segregate chromosomes which is necessary for passage through the cell cycle. Chapter 2 introduces a metric designed to reflect this special biology of chromosomes based on putative translation in 3D.

1.10 FUTURE DIRECTIONS

The denouement of any computational prediction is reached with experimental support or rebuke. These experimental conclusions may in turn generate model refinements; thus, a tandem cycle of computation and experiment arises where each builds on the other. This cycle will be critical for understanding how chromosome-nuclear envelope attachments affect the hierarchy of 3D genome organization. This dissertation in particular explores how attachments affect the highest level this hierarcy: chromosome territories, chromosome intertwining, and chromosomal interactions. However, 3D genome organization is remarkably complex and entire dissertations could focus exclusively on the lower levels of its structure. Moreover, chromosome attachment to the nuclear envelope is only one phenomenon that shapes 3D genome organization. Indeed, 3D genome organization may also be shaped by chromosome looping (123), condensing (131), and strand crossing (43). Even more principles are likely undiscovered. How many dissertations could be written on these phenomena and and their layered effects on the entire hierarcy of 3D genome organization? Although 3D genome organization can now be studied with sophisticated models and high performance computing, many cycles of computation and experiment will be needed to reach a basic understanding its principles.

CHAPTER 2

INVESTIGATION OF THE CHROMOSOME REGIONS WITH SIGNIFICANT AFFINITY FOR THE NUCLEAR ENVELOPE IN FRUIT FLY – A MODEL BASED APPROACH

Nicholas Allen Kinney[1], Igor V. Sharakhov*[2] and Alexey V. Onufriev*[3,4]

[1] Genetics Bioinformatics and Computational Biology, Virginia Tech, Blacksburg, VA, USA

[2] Department of Entomology, Virginia Tech, Blacksburg, VA, USA

[3] Department of Physics, Virginia Tech, Blacksburg, VA, USA

[4] Department of Computer Science and Applications, Virginia Tech, Blacksburg, VA, USA

* Igor Sharakhov Tel: (540)-231-7316 Fax: (540)-231-7126 Email: igor@vt.edu

* Alexey Onufriev Tel: (540) - 231 - 4237 Email: alexey@cs.vt.edu

2.1 ABSTRACT

Three dimensional nuclear architecture is important for genome function, but is still poorly understood. In particular, little is known about the role of the "boundary conditions" – points of attachment between chromosomes and the nuclear envelope. We describe a method for modeling the 3D organization of the interphase nucleus, and its application to analysis of chromosome-nuclear envelope (Chr-NE) attachments of polytene (giant) chromosomes in *Drosophila melanogaster* salivary glands. The model represents chromosomes as self-avoiding polymer chains confined within the nucleus; parameters of the model are taken directly from experiment, no fitting parameters are introduced. Methods are developed to objectively quantify chromosome territories and intertwining, which are discussed in the context of corresponding experimental observations. In particular, a mathematically rigorous definition of a territory based on convex hull is proposed. The self-avoiding polymer model is used to re-analyze previous experimental data; the analysis suggests 33 additional Chr-NE attachments in addition to the 15 already explored Chr-NE attachments. Most of these new Chr-NE attachments correspond to intercalary heterochromatin – gene poor, dark staining, late replicating regions of the genome; however, three correspond to euchromatin – gene rich, light staining, early replicating regions of the genome. The analysis also suggests 5 regions of anti-contact, characterized by aversion for the NE, only two of these correspond to euchromatin. This composition of chromatin suggests that heterochromatin may not be necessary or sufficient for the formation of a Chr-NE attachment. To the extent that the proposed model represents reality, the confinement of the polytene chromosomes in a spherical nucleus alone does not favor the positioning of specific chromosome regions at the NE as seen in experiment; consequently, the 15 experimentally known Chr-NE attachment positions do not appear to arise due to non-specific (entropic) forces. Robustness of the key conclusions to model assumptions is thoroughly checked.

2.2 INTRODUCTION

Unlike enzyme proteins, which usually adopt the same unique three-dimensional (3D) shapes in all cells, the conformational states of chromatin fibers are not nearly as compact and ordered; the basic principles governing these conformational states are only beginning to emerge through computation and experiment (11,30,43,75,77,78,132-135). Just like in the case of many polymers, the states of folded chromatin in the cell nucleus are expected to depend on the

"boundary conditions", in this case the location and properties of the nuclear envelope (NE). For example, if an "unrestricted" random coil were the same length as the human genome, it would occupy a 3D space many times greater than the volume of a typical cell nucleus, implying that in reality boundary conditions restrict the polymer to a much smaller actual volume matter. General polymer physics arguments suggest that the conformational state of chromatin across cell types depends strongly on the chromosome to nucleus volume ratio, and thus, there may be different folding principles in different lineages e.g. human and yeast cells (75,77,78,132,136,137). Indeed, recent computational studies have demonstrated that chromosome organization in the nucleus may strongly depend on the degree of (spherical) confinement (138): increasing the degree of confinement mimicked the effect of increasing chromosome looping probability, reinforcing the idea that the boundary conditions of the nucleus matter. These (138), and the results of other studies (17,22,75,81,82,139-141), have raised the possibility that the boundary conditions of the nucleus, chromosome topology, and non-specific (entropic) forces may be sufficient to account for the organization of chromosomes in the nucleus of some Metazoans. Furthermore, chromosome looping, potentially brought about by the degree of confinement (138), has been linked to gene expression levels; specifically, higher chromosome looping probability was associated with higher local chromosome density and lower transcriptional activity in a recent study (142).

In addition to including the NE in computational chromosome models; many studies now take into consideration the specific interactions of the chromosomes with the NE (77,78). For example, a recent model of the yeast nucleus recapitulated key features of 3D chromosome organization and incorporated both centromere and telomere attachment to the NE (77,78). However, the nature of these interactions with the NE remains unclear; other studies have suggested that non-specific and specific forces acting together position chromosomes in the nucleus (82), and a recent study demonstrated that non-specific forces alone may be sufficient to localize chromocenter and heterochromatin to the NE in *Arabidopsis* (81). Regardless of mechanism, identifying regions of chromosome-nuclear envelope (Chr-NE) contacts and "anti-contacts" (regions which statistically avoid the NE) is important for their inclusion in future modeling studies and for determining the types of chromatin typically found at or away from the nuclear periphery, which is in turn important for better understanding of 3D-chromosome

organization. Stated simply, the main goal of our study is to objectify the finding of Chr-NE attachments and characterization of their composition.

Earlier experiments (30) discovered 15 Chr-NE attachments, identified by their high probability of contact with the NE exceeding 66% in an ensemble of 24 nuclei. These 15 known Chr-NE attachments coincide almost exclusively with regions of intercalary heterochromatin – gene poor, dark staining, late replicating regions of the genome (58). The seminal study has clarified the character of the most frequent NE attachments, but left several important questions unanswered. Does the 66% ad-hoc threshold used previously for discovering Chr-NE attachments reveal all of the Chr-NE attachments in *Drosophila* polytene chromosomes, too many, or too few? Using a more objective threshold here is important because the composition of chromatin inferred from the analysis of the Chr-NE attachments may change if too many or too few Chr-NE attachments are identified. The use of a more objective threshold may help reveal previously uncharacterized NE attachments in the old experimental data; if those attachments are indeed found, then what is their heterochromatic character? Finally, could pure geometric effects, such as confinement in a spherical nucleus, specific chromocenter arrangement, and the excluded volume of the chromosomes and nucleolus favor the placement of specific chromosome positions at the NE, and could these non-specific (entropic) forces alone position the 15 most significant Chr-NE attachments?

Our study is designed to address these and several other questions, while delivering to the community a computational model that can be used to complement experiments that study the 3D architecture of chromosomes. Here we use polytene chromosome from salivary gland nuclei of *D. melanogaster*, which is a well-established model for studying organization and function of the eukaryotic genome (143-146). Each of the polytene chromosomes contains approximately 1024 chromosome replicas bundled together in parallel; thus the entire genome organization in a single nucleus becomes visible under a light microscope. This is a critical advantage over "regular" interphase chromosomes because it becomes possible to obtain full spatial information about the position of each individual polytene chromosome – its complete trace in 3D space. The study of polytene chromosomes has significant potential for general understanding of 3D genome organization because recent experiments revealed identical structural and functional

organization of non-polytene and polytene chromosomes in fruit fly (36,147-149). Moreover, the polytene chromosomes are estimated to occupy about a third of the nuclear volume (31); this chromosome to nuclear volume ratio, which critically affects the over all 3D nuclear architecture (43), is the same in regular non-polytene nuclei (40), and is likely similar to the values characterizing human nuclei (75).

Experimental studies have identified several plausible biological roles and effects of Chr-NE contacts, such as maintenance of nuclear architecture and separation of the chromosome territories (73,74,83,150,151). Despite their importance, experimental validation and analysis of Chr-NE contacts in most non-polytene interphase nuclei is difficult since regular interphase chromosomes and their NE contact sites cannot be visualized directly by standard techniques of light microscopy. Instead, Chr-NE contacts in non-polytene interphase nuclei are often identified by indirect methods with fluorescence *in situ* hybridization (52) or inferred using a DamID approach – a method based on detecting DNA methylation by a chimeric protein consisting of a chromatin protein fused with methyltransferase (12,152,153). The drawback of fluorescence *in situ* hybridization is that only a small number of chromosome positions can be labeled; consequently, determining the complete folding pattern of the chromosomes in a single nucleus is nearly impossible. The drawback of using a DamID (153) approach is that methylation via methyltransferase can only be detected using an entire ensemble of cells; consequently, the stochasticity and cell-to-cell variability of the Chr-NE contacts is lost. In polytene chromosomes, 3D tracing experiments have been used (30,31) to directly visualize chromatin folding and to identify Chr-NE attachments, but these studies typically involve small numbers of nuclei, which makes establishing statistical significance difficult. The model described in this study is used to improve the criteria for identifying statistically significant Chr-NE attachments, and consequently improve our knowledge regarding the type of chromatin found at or away from the NE.

The polytene chromosomes from *D. melanogaster* salivary glands have been extensively characterized in previous experiments (28,30,33). We model each of the five largest chromosome arms of *D. melanogaster* as a random self-avoiding walk (SAW) under confinement; parameters of the model come from available experimental data. We validate our

method of model building by quantifying the experimentally observed presence of chromosome territories and the absence of chromosome intertwining. The model answers three questions: Are there additional statistically significant Chr-NE attachment regions? If there are additional Chr-NE attachment regions, do they also correspond to heterochromatin? Does confinement of the polytene chromosomes in a spherical nucleus alone favor the positioning of specific chromosome regions at the NE as seen in experiment? Our model demonstrates that the geometric effects of chromosome confinement inside a spherical nucleus alone do not bring about specific Chr-NE attachments. We use our model to improve criteria for locating Chr-NE attachments. By applying our criteria to the data available from previous tracing experiments (28,30,33) we identify 33 new, previously unreported, but statistically significant Chr-NE contacts and 5 regions of anti-contact. The composition of these new Chr-NE attachments is discussed.

2.3 MATERIALS AND METHODS

2.3.1 Model Building

Motivation. Our model incorporates *all* experimentally known parameters *D. melanogaster* polytene chromosomes *with the exception* of introducing specific Chr-NE attachments; in other words, the model is a Null model with respect to Chr-NE attachment. Essentially, the deviations between our Null model and experiment then reveal the positions of Chr-NE attachment from experimental data (this is the focus of the paper and is discussed extensively in what follows). We construct the Null model using an equilibrium based self-avoiding walk approach and introduce several modifications in order to recapitulate experiment. Some of these modifications likely introduce non-equilibrium features into our model; however, we stress that the fully modified model contains all the known features of the polytene nucleus from experiment except for specific Chr-NE attachments. For any other model the deviations from experiment would arise from multiple factors, not just the Chr-NE attachments. Regardless, we check that all model conclusions are robust to the non-equilibrium features that our model contains (discussed below).

Approach. The five largest chromosome arms of *D. melanogaster* salivary glands are modeled as beads-on-string (41-43,82,85,132) and are represented as five random self-avoiding walks (SAWs) (15,84,154,155) (Figure 2.1). This approach is common in theoretical studies of 3D

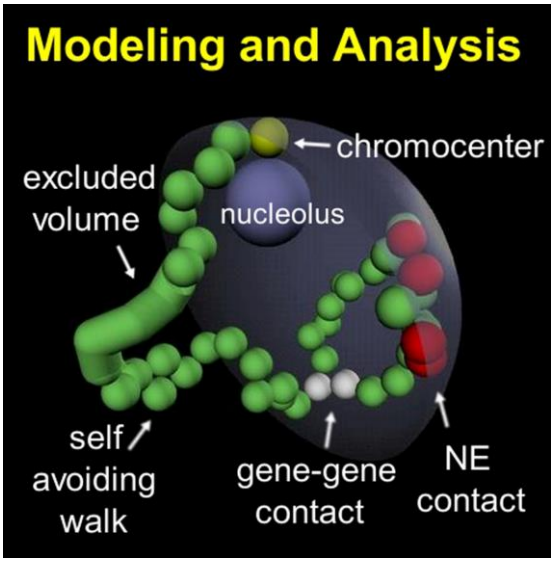


Figure 2.1 - Computational model of equilibrium states of a *Drosophila* polytene nucleus. The five chromosome arms are represented by five random SAW chains under confinement (one SAW chain is shown). The chains are built simultaneously starting from the chromocenter (yellow). Contacts are beads within one micron of the NE (red); locus-locus contacts are beads within two microns of each other (white). Full excluded volume includes the cylinder connecting two adjacent beads.

chromosome architecture, and has already been shown to recapitulate some properties of experimental ensembles of polytene chromosomes (53). The sixth arm, chromosome 4, is not considered due to its negligible length. Experimental data for the chromosomes and the nucleus become realistic model parameters and constraints imposed during the construction of SAWs (see text S1 for a complete derivation of all model parameters and constraints).

Modeling procedure. One bead representing the chromocenter is placed adjacent to the NE (yellow bead in Figure 2.2) at the "north pole" of the nucleus. Then, five initial beads are placed, without overlapping, at random angular positions around the chromocenter (Figure 2.2); these five beads touch the chromocenter and NE, mimicking the experimental configuration of *D. melanogaster* chromocenter with the five chromosome arms extending outward. The arrangement

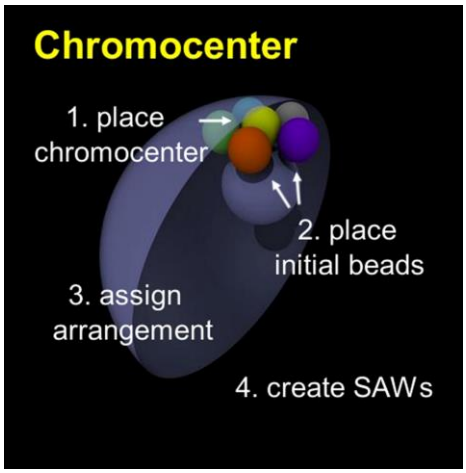


Figure 2.2 - The first four steps of constructing the model nuclei. SAW = self-avoiding random walk.

of the five initial beads around the chromocenter bead is designed to match the relative proportion of chromocenter spatial arrangements seen in experiment (30) (Figure 2.3, details in text S1). After assigning the chromocenter arrangement, SAWs are constructed using Rosenbluth algorithm (96) (i.e. SAW chains grow by addition of monomers in a "true" SAW fashion). We use a Rosenbluth algorithm for computational efficiency; for short chains this approach is a good approximation of self-repelling chains which are true equilibrium states of polymers (97). This approach was recently used to generate densely packed SAWs in a study of protein folding (98). Although our model is based on a SAW model, which is equilibrium by construction (to the extent that it approximates self-repelling chains), two non-equilibrium features are introduced to better represent experiment; these include the Rabl configuration of chromosomes and right-handed chromosome chirality.

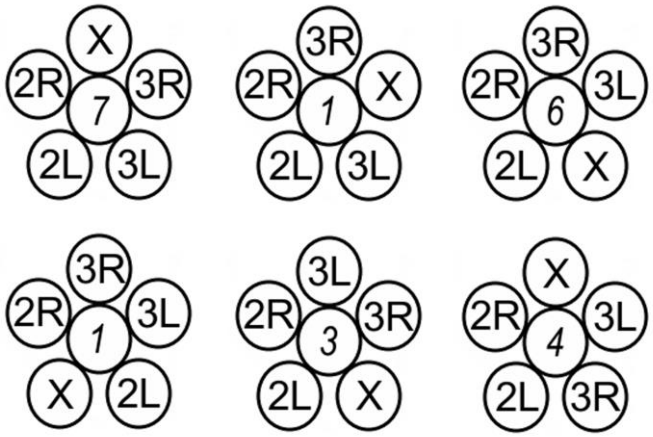


Figure 2.3 - Relative number of chromocenter arrangements in 22 experimental nuclei. Numbers in italic represent the total nuclei found with the corresponding arrangement of the chromosome arms.

It is known that most *D. melanogaster* polytene chromosomes conform to the Rabl type configuration (30). This configuration is characterized by the predominant (80%) presence of the chromosome telomeres in the nuclear hemisphere opposite the chromocenter. It has been speculated that the Rabl configuration of chromosomes is a remnant of anaphase, which upon formation of Chr-NE attachments, may trap chromosomes in non-equilibrium configurations within the nucleus. However, the nature of Rabl configuration is not completely clear; an alternative possibility is that formation of Chr-NE envelope attachments trap chromosomes in a polarized configurations within the nucleus which remain polarized after reaching equilibrium. Rabl configuration was enforced in our models by *a posteriori* filtering of the generated ensembles of nuclei to achieve Rabl configurations in the final ensemble (details in text S1). This *a posteriori* filtering introduces a non-equilibrium modification of the SAW's forming the basis

of our model and is intended to reproduce the Rabl configuration seen in experiment (see figure B.1); but, this does not necessarily imply that the experimental polytene chromosomes in the nucleus are non-equilibrium for the reasons stated above.

Studies that trace the path of each chromosome arm in *D. melanogaster* salivary gland nuclei have observed a disproportionate amount (2:1) of right handed twist compared to left handed twist (30). We enforce right handedness in our simulated chromosomes during construction of the SAWs: it is twice as likely for a new bead to be accepted if it forms right handed chirality rather than left handed chirality (details in text S1). This introduces a second non-equilibrium modification of the SAWs that form the basis of our model; the modification is intended to reproduce the chirality seen in experiment. This modification also does not imply that the experimental polytene chromosomes in the nucleus are non-equilibrium; the right-handed chirality seen in experiment may be equilibrium with dihedral potentials that are currently unknown. To address the question, one has to go beyond the current model.

A single step in growing the SAWs consists of simultaneously picking a random direction in 3D space to extend each model chromosome arm, adding the five new beads, and checking for violation of model constraints such as excluded volume (no bead overlap) and right-handed chromosome chirality. If no model constraints (see below) are violated, then the new beads are accepted and the model chromosome arms continue growing. In the case of rejecting the new beads, the step is repeated with new random directions in 3D space. The avoidance of perpetual SAW rejections is accomplished with two backtracking parameters, BT_1 and BT_2, that tally the number of SAW rejections. A single bead backtrack is made after $BT_1 = 2000$ failed SAW additions, followed by its resetting; a 5 bead backtrack is made after $BT_2 = 6000$ failed SAW additions, followed by its resetting. The above process is repeated until model completion (see Figure 2.4). This simple technique allows us to easily impose preferred right handed twist, chromocenter arrangement, and chromosome confinement which would be more difficult to enforce using alternative model building techniques (43,75,156). The main conclusions of this work are insensitive to the choice of BT_1 and BT_2, see below. We chose the manifestly symmetric SAW construction procedure (at each step the beads for all the chromosomes are added simultaneously) because there is no biological evidence that suggests a spatial symmetry

breaking between the chromosomes. That is a conceivable alternative procedure in which a certain chromosome is fully built first, followed by other(s) would be less justified biologically.

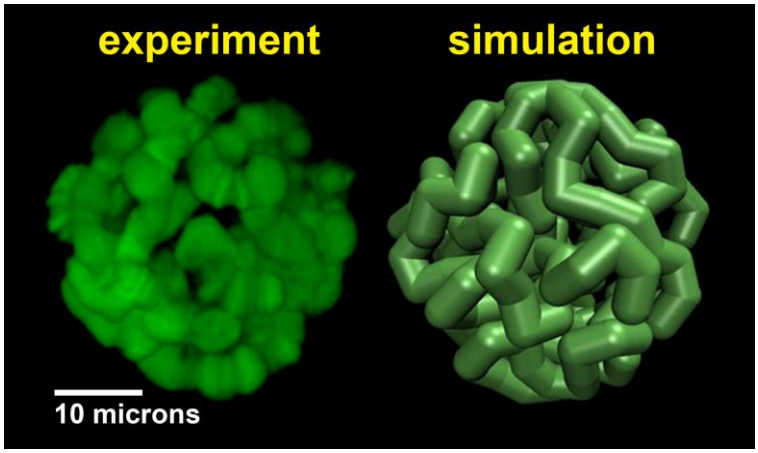


Figure 2.4 - Representation of a simulated model nucleus (right) compared to experiment (left).

Robustness of major conclusions to model details. The general SAW approach introduced here to model polytene chromsomes was validated in several previous studies in similar contexts (43,53,75,77,78). To improve biological realism of our model, we have introduced several additional features to the basic procedure. For simplicity we construct our SAW's using an unweighted Rosenbluth algorithm (96). It is reiterated that our final ensemble reproduces all known features of experimental polytene chromsomes in the nucleus without enforcing specific Chr-NE attachments; consequently, the deviations between our model and experiment must stem from Chr-NE attachments alone. Regardless, we have checked explicitly that the key model conclusions are robust to all non-equilibrium SAW modifications introduced in our model. To this end, two variations of our SAW were considered: (1) *fully modified SAW* – with Rabl configuration, right-handed chirality, and chromocenter arrangement designed to recapitulate all features of experimental nuclei with the exception of Chr-NE attachments. This is the main model used in this work. In addition, we have considered: (2) *unmodified SAW* - which does *not* introduce Rabl configuration, right handed chirality, or chromocenter arrangement, and so is equilibrium to the extent that our chain growing algorithm approximates self-repelling chains. Both variations of our SAW model lead to the same main conclusions (see table S1 and table S2). To enforce constraints in our models (spherical boundary and excluded volume) we use a simple backtracking procedure controlled by two parameters, BT_1 and BT_2 (see methods) that determine when a single bead and 5 bead backtrack is made respectively during the construction of the SAW. Unlike all other parameters of the model, the values of these two parameters do not come

from experiment. To verify robustness of the key conclusions to the specific choice of $BT_1 \gg 1$ and $BT_2 \gg 1$, we used a third variation of our SAW approach: *fully modified SAW with BT_1 = 1000 and BT_2 = 3000* – with backtracking parameters reduced by a factor of 2. This variation of our model also yielded the same main conclusions (see table S1); computational complexity prohibited testing model robustness in the entire BT_1-BT_2 parameter space.

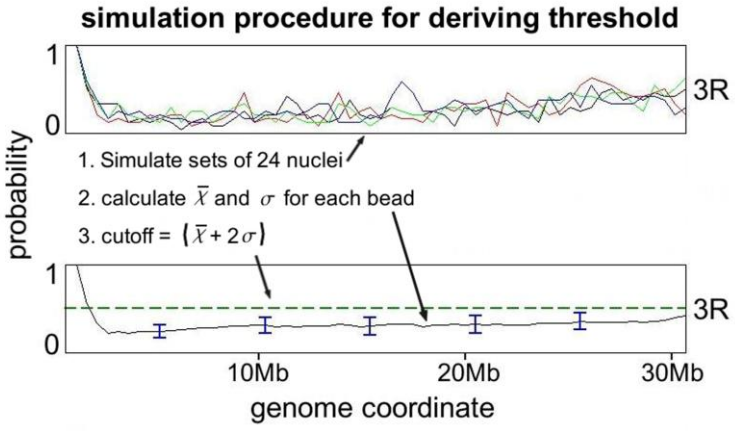


Figure 2.5 - Procedure for deriving statistically significant threshold for identifying Chr – NE contacts. Only chromosome 3R is shown for clarity. 96 sets of 24 nuclei were simulated (without enforcement of Chr-NE contacts). NE contact frequency for each chromosome position is plotted as a "contact frequency profile"; profiles from 4 independent simulations are exemplified in the top panel. The mean ($\bar{\chi}$) contact frequency and the standard deviation obtained from these simulated tracing experiments are used to set a threshold for identifying statistically significant Chr-NE contacts ($\langle \bar{\chi} + 2\sigma \rangle$) and anti-contacts ($\langle \bar{\chi} - 2\sigma \rangle$).

Derivation of model parameters and constraints from biological data. See text S1

2.3.2 Simulations

A previous experiment (30) estimated Chr-NE attachment probability for each chromosome position in 24 nuclei; each nucleus represented a single snapshot of the true state of the chromatin – a conformational ensemble of the five chromosomes. In this previous experiment (1) 15 Chr-NE attachments were defined by setting an ad hoc threshold of >66% probability of observed contact with the NE. We use our model to essentially simulate a large, statistically significant number of these same tracing experiments also with 24 nuclei, but without specific Chr-NE attachments. Upon simulating 96 repeated tracings of 24 experimental nuclei (four shown in Figure 2.5), we calculate the mean, $\bar{\chi}$, and the standard deviation, σ, in contact frequency for each bead. It is unlikely to observe beads in our simulations (24 nuclei) with

frequency of NE contact greater than $\langle \bar{\chi} + 2\sigma \rangle$. We identify 48 Chr-NE contact frequencies above this threshold (green line Figure 2.5 and 2.6) in previous experimental data (30). The only difference between our model and experiment is the presence of specific Chr-NE attachments in the latter, and so it is statistically highly unlikely that the 48 experimentally determined Chr-NE contact frequencies are above the $\langle \bar{\chi} + 2\sigma \rangle$ threshold due to pure chance (black and red arrows in Figure 2.6). By definition, approximately 2.5% of Chr-NE contact frequencies were above this threshold in our Null model that contains no specific Chr-NE attachments; thus, a lower statistical threshold would run the risk of identifying more "false positives" in experimental data while higher levels of significance would overlook the true Chr-NE attachments in experiment. We checked that 96 repeated simulations are enough to yield a reproducible $\langle \bar{\chi} + 2\sigma \rangle$ threshold, see table S1.

Using this same analysis, a threshold set at $\langle \bar{\chi} - 2\sigma \rangle$ was used to establish statistical significance for regions of *anti-contact* - regions which statistically avoid the NE. With this definition, we identified statistically significant Chr-NE anti-contacts in previously published experimental data (blue line in Figure 2.6) (30). A large number of model nuclei (96x24=2304 model nuclei) were needed to simulate these repeated tracing experiments because a new set of 24 nuclei

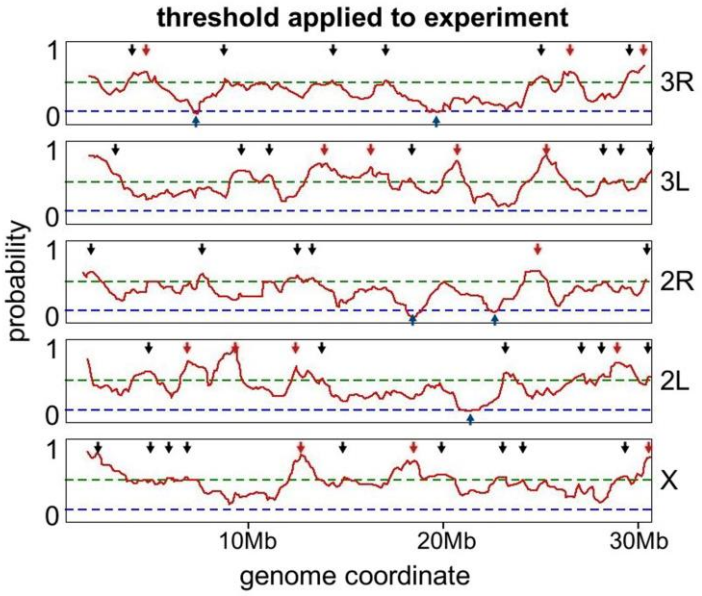


Figure 2.6 - High frequency and sub-high-frequency NE-contacts at a new threshold. Green dashed line - 50.5% (2σ) significance; red line – experiment. Red arrows are the original 15 contacts, Black arrows are the additional contacts which are statistically significant according to our simulations. Blue arrows are significant regions of anti-contact – contacts that occur below the threshold (blue dashed line) of 14.3% (2σ) significance.

was used for each simulated experiment.

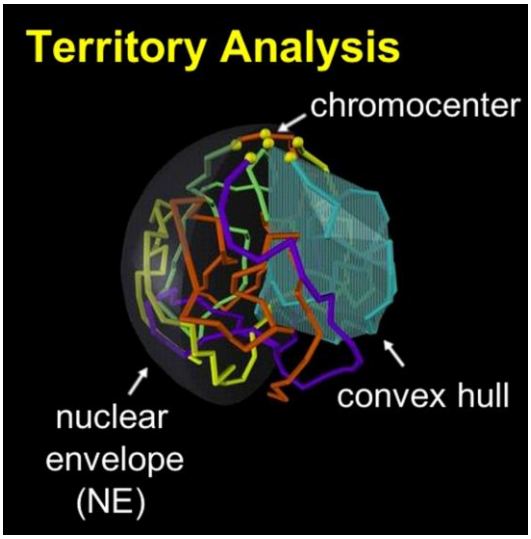


Figure 2.7 - The territory index. This index is defined as the percent of a chromosome found inside its own convex hull. Example: light blue chromosome.

2.3.3 Analysis of the simulations

Simulated tracing experiments. For a single chromosome arm in a model nucleus, an array is formed with entries for each bead in the chromosome arm. For every bead an entry of 1 is recorded in the array if contact occurs with the NE. A frequency profile (Figure 2.5) is formed by averaging corresponding entries in 24 arrays, this being the same number of nuclei that was used in a previous experiment (30). In our study this procedure is repeated 96 times, simulating the outcome of 96 chromosome tracing experiments involving 24 nuclei each; the average contact frequency, $\bar{\chi}$, and standard deviation, σ, for each bead in these 96 simulated tracing experiments is calculated. The standard deviation of these simulated chromosome tracing experiments provides a measure of how contact frequencies for a single set of 24 nuclei *may* change for repeated experiments.

Chromosome territory index. Chromosome territories (11,43,157) are assessed by quantifying how effectively one chromosome excludes other chromosomes from the volume it occupies in the 3D space. There is no universally accepted definition of chromosome territory, and, to the best of our knowledge, there is no mathematically rigorous one either. Our definition of the territory is similar in spirit to the construct used to define Kolmogorov-Sinai entropy. We begin by calculating the convex hull for a single chromosome, Figure 2.7 (we use MATLAB (158)), this is the minimum volume that includes all the chromosome's points (bead centers) inside a convex polyhedron. In general, each convex hull contains its own chromosome, and may also encompass some points belonging to other chromosomes. A fully "territorial" chromosome is one whose convex hull does not contain points from any other chromosomes while a less "territorial" chromosome is one whose convex hull contains some points from other

chromosome. We define the chromosome territory index as the fraction of points inside a convex hull that belong to the chromosome used for its construction; for example, the fraction of light blue chromosome points inside the light blue convex hull shown in Figure 2.7. Under this definition the maximum territory index is 1. The minimum territory index for a chromosome depends on how many beads the chromosome's convex hull can possibly accommodate; different chromosomes have a different minimum territory index. To establish this minimum territory index for a chromosome arm having n_{arm} beads, the 3D chromosome configuration having a global maximum convex hull volume under spherical confinement, $H_{\max}$, is found (Figure B.2). The minimum territory index for the chromosome is then given by $T_{\min} = n_{arm}/n_{\max}$, where $n_{\max}$ is the maximum number of beads that $H_{\max}$ can accommodate.

We approximate $T_{\min}$ for each chromosome of fruit fly by finding the chromosome configuration that corresponds to $H_{\max}$ (see Figure B.2 for details). For each chromosome, the volume of this $H_{\max}$ (Figure B.2 and Figure B.3) was found to exceed the total volume of all 248 beads in our model nucleus ($248 \cdot V_{bead}$) implying that a fully "anti-territorial" chromosome is one whose convex hull contains all 248 beads from itself and all other chromosomes in our model; thus, for our model $T_{\min} = n_{arm}/248$. Using these definitions, the lowest territory index ranges from .18 to .24 depending on the chromosome.

Test for chromosome intertwining. Chromosome intertwining is an intuitive concept that can be rigorously assessed by attempting to separate model chromosomes by a translation in 3D space: if the two chromosomes can be separated in this manner, we call them non-intertwining. We begin by selecting the backbone of two chromosomes (including centromere) from a model nucleus; the backbone of a model chromosome consists of the line segments connecting the centers of each bead. A 35 micron long direction vector is then chosen to translate the backbone of one model chromosome; (the translation vector is longer than the diameter of the nucleus). If the two backbones cross during this translation then a new direction vector is picked. A total of 162 different direction vectors are tested in this manner, with unit vectors that uniformly cover the S^2 space (spherical surface). If it is possible by one of these translations to separate the

chromosome backbones, then the two chromosomes do not intertwine. The amount of intertwining in an ensemble is quantified by calculating the percent of all chromosome pairs in the ensemble that intertwine.

2.4 RESULTS

2.4.1 Validation of the model

Chromosome to nucleus volume ratio. The calculated chromosome to nuclear volume ratio is .30 in our model, close to the experimentally measured ratio of .34 in *D. melanogaster* (31). The difference which arises from coarse graining of the chromosomes is likely to be within the margin of error of the experiment.

Chromosome territories. Experimental, qualitative descriptions of *D. melanogaster* polytene nuclei have established that chromosomes form territories, "analogous to the sections of a grapefruit" (29,30) (Figure 2.7). The average territory index per chromosome in our model is .650 out of the highest possible value of 1.0 (see the precise definitions in "Methods"); the computed territory indexes are significantly higher than the smallest possible territory index in fruit fly, which ranges from .18 to .24, depending on the chromosome. The comparison confirms that our model chromosomes are indeed "territorial". Thus, no additional, territory-specific *a posteriori* filtering was needed within our model to recapitulate this critical feature of chromosomes seen in experiment. We interpret this as validation of our modeling approach; we further checked the robustness of our modeled chromosome territories to the non-equilibrium modifications of our SAWs. Specifically, an unmodified SAW model (also described in robustness section) without right-handedness, Rabl orientation, or preferred chromocenter arrangement had an average territory index per chromosome of .651. Thus, non-equilibrium considerations may not be needed to account for the territorial property of polytene chromosomes in *D. melanogaster* salivary gland nuclei. Incidentally, we noted that a subjectively (visually) "territorial" model chromosome does not imply a chromosome territory index of 1; for example, a territory index of .650 has a qualitative description similar to the qualitative descriptions of previous experiments (30), (see Figure B.4). The degree of objectivity

and rigor that we have introduced by our definition of chromosome territory may therefore be useful in analysis of both experimental and modeled chromosomes.

Intertwining. Experimental, qualitative descriptions of *D. melanogaster* polytene nuclei have established that salivary gland chromosome arms do not intertwine (29-31). We calculated the percent of non-intertwining chromosome arms (see methods) in our models. This analysis suggests that the percent of non-intertwining chromosome arms approaches 95% using our modeling method, (see details in Figure B.5). We interpret this agreement with experiment as an indication of the strength of our model. The virtual absence of chromosome intertwining within our model was also robust to the non-equilibrium modifications of the model; approximately 95% of chromosomes in an unmodified model (described in robustness section) were also non-intertwining. In addition, we noted that our test for chromosome intertwining is highly sensitive; some chromosomes which may subjectively (visually) be identified as "non-intertwining" still narrowly failed the rigorous test (see Figure B.6).

Scaling properties of the generated SAWs. The end-to-end length of our simulated SAWs in free space is described by $\langle r^2 \rangle \sim n^{1.18}$, where *r* is chain end-end length and *n* is number of beads; this is in good quantitative agreement with theoretical results that give a range of $\langle r^2 \rangle \sim n^{1.172}$ to $\langle r^2 \rangle \sim n^{1.2}$ (94,159-162). When we exclude the volume of the bond between nearest neighbor beads (Figure 2.1), the end-end length of our SAW's in free space is described by $\langle r^2 \rangle \sim n^{1.25}$ (Figure B.7).

2.4.2 Additional high frequency contact positions are suggested by simulation.

Improved criterion for identifying chromosome – NE contacts. A contact frequency of .505 was on average (not including the centromere) two standard deviations above the mean for beads in our model (Figure 2.5 bottom panel), this value defined an objective threshold used to identify additional Chr-NE contact positions in the experimental data for a single set of 24 nuclei. Although this amounts to a lowering of the .66 frequency threshold originally used to identify the 15 Chr-NE attachments in (1), we stress that our model is intended to *improve* the threshold used

to identify Chr-NE attachments not to simply lower it; an *ad hoc* threshold that is too high or too low could lead to an altered composition of chromatin at the NE and influence our understanding of Chr-NE attachment formation.

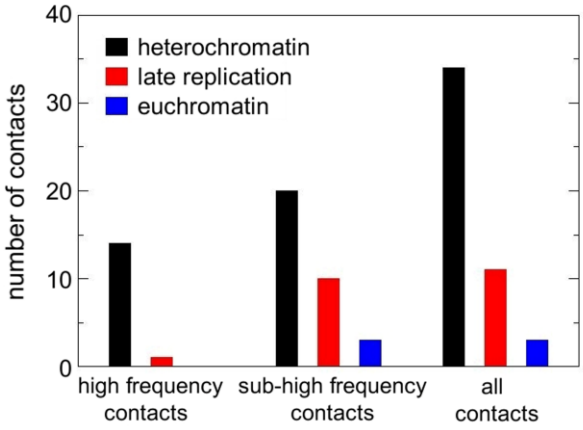


Figure 2.8 - Composition of Chr-NE contacts in interphase polytene chromosomes of *D. melanogaster*. High frequency contacts were identified previously in experiment (left set of bars). New sub-high frequency contacts identified by our model (center). All contacts combined (right).

All peaks above our improved threshold were identified in the experimental data (Figure 2.6); nearby peaks also above the threshold were only considered if they were further away than the Kuhn length (3.1 microns) from neighboring peaks, this being the length over which there is no directional correlation in the chromosome fiber. We refer to the new Chr-NE contacts revealed as "sub-high frequency" to distinguish from the 15 previously reported "high frequency" Chr-NE contacts from (30).

Composition of newly identified chromosome–NE contact positions by chromatin type. Most of the positions we identify are located in regions of the chromosome corresponding to intercalary heterochromatin—dark staining compact regions of the chromosome (163). In total, we identify 33 additional Chr-NE contacts, 20 of which are intercalary heterochromatin and 3 euchromatin; the additional 10 Chr-NE contacts display some properties of heterochromatin by being late replicating regions (58) (Table S3). We classified a chromosome region as intercalary heterochromatin if it contained a site of late replication and localization of antibodies against SuUR (Suppressor of UnderReplication) protein in wild-type flies (58). We classified a chromosome region as a region of late replication if it contained a site of late replication in wild-type flies or a site of localization of antibodies against SuUR protein in SuUR 4x flies, which have two additional $SuUR^{+}$ doses (58). Of the 20 intercalary heterochromatin regions, 6 are regions of under-replication, which is typical to the large bands of intercalary heterochromatin

(36). Experiments have already demonstrated that the 15 most significant Chr-NE contacts in *D. melanogaster* are almost exclusively heterochromatic. Our results (Figure 2.8) suggest that affinity for the NE can change gradually, with the highest affinity for the NE almost *exclusively* possessed by intercalary heterochromatin, and the next highest affinity for the NE *mostly* a property of intercalary heterochromatin. The presence of 3 euchromatic regions in our set of 33 sub-high frequency contacts suggests that it is not necessary for a chromosome region to be heterochromatic in order to possess *some* affinity for the NE. We stress that this result is based on the known biological parameters of *D. melanogaster* polytene chromosomes.

Composition of contact positions with strong aversion for the NE. We applied the same analysis to identify positions of anti-contacts, which we define as chromosome regions that have significantly low probability to form a Chr-NE contact. A contact frequency of .143 was on average two standard deviations below the mean for any bead in our model; this defined a threshold used to identify anti-contacts. Five anti-contact regions were found below this threshold: 27B, 55E, 52F/53A, 85E, 93F (Figure 2.6). Three of these regions are euchromatic (55E, 52F, 85E), two are regions of late replication (27B, 93F), and one (52F/53A) is at the boundary between euchromatin and heterochromatin. The highly significant anti-contact at the border of regions 52F/53A may suggest that heterochromatin is not sufficient for formation of a Chr-NE contact. Thus at 1Mb resolution, identifying a chromosome region as late replication, and possibly heterochromatin alone may not be completely sufficient to determine if a chromosome region will form a contact with the NE or if it will likely avoid it.

2.4.3 Experimental chromosome-nuclear envelope contact positions appear non-random.

A simple corollary follows from our computational modeling used primarily to reveal statistically significant Chr-NE attachments in experimental tracing data: because Chr-NE attachments are *not enforced* in our model nuclei (except for the chromocenter) these models should also reveal whether geometric effects alone (spherical confinement, presence of nucleolus, chromocenter arrangement, etc) predetermine which chromosome positions are in contact with the NE. When we average over 96 simulated chromosome tracing experiments, each bead in each chromosome in our model has approximately an equal chance to contact the NE (exemplified for 3R in Figure 2.5 bottom panel). This result shows that a bead's purely

geometric positioning in a model chromosome under spherical confinement has no effect on the affinity or aversion of that bead for the NE. Experiments (30) suggest, however, that 15 regions on the fruit fly chromosomes have a much higher affinity for the NE than the average. Statistically, it is virtually impossible for these 15 experimental Chr-NE contacts to arise in their corresponding beads due to pure chance in 24 model nuclei; thus, we conclude that, to the extent that our model represents reality, the 15 experimental Chr-NE contacts must have intrinsic affinity for the NE, unrelated to their pure geometric position along the chromosome.

2.5 DISCUSSION

2.5.1 The Model. The random self-avoiding walk (SAW) is a classic, widely used approach to modeling polymers (77,78,105,132,164-166). Variations of the SAW were employed in studies that model 3D chromatin organization, and were shown to accurately capture average locus-to-locus distances (43,53,132,155,167). Several models based on variations of a SAW have already been used to successfully explain the structural features of chromosome 2L in *D. melanogaster* polytene chromosomes (53). These previous models used a variety of strategies to explain 3D chromosome structure: strategies which included incorporating the Rabl configuration of chromosomes and generating the SAWs under confinement (53,132). In this work we capitalize on the earlier successes of the SAW-based approaches to build a more realistic model of the 3D architecture of chromosomes in interphase nuclei - our model incorporates only the known parameters on *D. melanogaster* polytene nuclei, no fitting parameters are introduced. In addition, our model creates entire ensembles of nuclei which realistically describe the cell-to-cell variations in chromatin folding. The application of our model to the analysis of chromosome tracing experiments offers a new valuable tool. We have applied the model to a previous tracing experiment of 24 *D. melanogaster* salivary gland nuclei (30); however, our model can easily be extended to the analysis of tracing experiments involving a different number of nuclei. In addition, our model can easily be applied to polytene chromosome of different cell types by reconfiguring the model with the corresponding parameters from experiment.

Chromosome territories, in which each chromosome occupies a distinct sub-volume of the cell nucleus, have been observed in many experiments including both polytene and non-polytene

chromosomes (7,14,30,31,157,168-170). Recent simulations have demonstrated that non-specific entropic forces may play a significant role in establishing and maintaining chromosome territories (139,141,171), and it has been suggested (139) that this entropic effect stems from long flexible polymers having access to more chain configurations if they remain separate in distinct domains, rather than tangling together. This entropic effect has been shown to depend on the degree of confinement (172) and the presence of chromosome loops (139), which may also arise due to non-specific forces (139). These arguments essentially assume that the chromatin reaches the state of thermodynamic equilibrium on the experimentally relevant time-scales. On the other hand, it has been argued that equilibrium configurations of human interphase chromosomes would not display territories and that territory formation is best explained by a non-equilibrium fractal globule (11,43). The prediction of chromosome territories in our model is robust to the non-equilibrium modifications we make to the underlying, essentially equilibrium, SAW model. This robustness suggests that non-equilibrium considerations may not be necessary to explain territories seen in polytene chromosome in fruit fly nuclei. Given that chromosome territories appear to be a generic feature of many genomes including human, our intuitive, yet mathematically rigorous and easily computable definition of the territory should be of interest as well.

2.5.2 New Chr-NE contact positions. Previous experiments (30,31) identified the 15 polytene chromosome positions with the highest probability to contact the NE; 14 of these corresponded to regions of intercalary heterochromatin. With the aid of our computational model we re-analyzed the experimental data and presented several important new results. First, our model provides a method to objectively define a Chr-NE contact or anti-contacts; these objective criteria are based only on robust statistical properties of polymer ensembles, the known parameters of *D. melanogaster* polytene chromosomes, and geometric dimensions of the *D. melanogaster* polytene nucleus. This analysis has led to identification of 33 new Chr-NE contacts, of which 20 are heterochromatic, 10 are late replicating, and 3 are euchromatic. This result suggests that affinity for the NE is not a discrete property; the most significant Chr-NE contacts may be *exclusively* heterochromatin (30), with less prominent contacts composed of *mostly* heterochromatin. We put forward a testable hypothesis that it is local density of heterochromatin that may determine the propensity to form Chr-NE contacts. Three of the Chr-

NE contacts we identify are euchromatin suggesting that it may not be necessary for a chromosome region to be heterochromatic in order to have some affinity for the NE. We found 5 regions of anti-contact (avoiding NE): 2 euchromatic, 2 late replicating, and 1 at the boundary between a euchromatin and heterochromatin region. This shows that late replication and possibly heterochromatin may not be sufficient to place a chromosome region in contact with the NE.

In non-polytene interphase chromosomes, pericentric and intercalary heterochromatin has been shown experimentally to possess a mechanism of localization to the NE, specifically, by lamin (12,36,56,173-175). A previous study of 24 polytene nuclei found that 14 NE contacts are composed of intercalary heterochromatin and one is a late replicating region (30). A following study confirmed 12 contacts (at 9A, 12DEF, 22A, 33A, 35A, 36C, 57A, 64D, 67D, 83D-84A, 98C, 100AF) and identified four more NE-contacting sites at 97A, 19DE, 60EF, and 61AB (33). Interestingly, these four contacts were also identified as sub-high frequency contacts in our study, and all four include regions of intercalary heterochromatin. Our study identified a total of 48 significant contact sites, 45 possessing properties of heterochromatin/late replication regions and 3 possessing properties of true euchromatin (Table S3). Thus, our results are consistent with previous experiments, but also suggest that intercalary heterochromatin (at 1Mb resolution) is not completely necessary or sufficient for the formation of a Chr-NE contact; however, it may be necessary for formation of Chr-NE contacts at the highest level of significance (30).

A genome-wide study of DNA-lamin binding in embryonic cells using DamID has shown significant correspondence to polytene Chr-NE contacts in larvae (12). This study has also indicated that lamin binding is linked to a combination of several features including late replication, large size of intergenic regions, low gene expression status, and the lack of active histone marks, suggesting that a combination of cell-type dependent and independent factors may influence NE association. Furthermore, this study (12) reported that when the c-terminal, nuclear membrane binding portion of lamin protein was deleted (referred to as Lam Δ^{CaaX}) there was a negative correlation between Lam Δ^{CaaX} and polytene chromosome NE association; consequently, it was suggested that Lam Δ^{CaaX} may co-localize with genes that have an aversion for the NE. Both of these results are consistent with our findings: that late replication or

heterochromatin alone may not be sufficient to bind a chromosome locus to the NE and that some regions of the chromosome can be preferentially located at the nuclear interior.

Another study (36) has compared localization of lamin-associated domains identified in DamID experiments (176) and 60 regions of intercalary heterochromatin. Interestingly, the overlap was far from complete: 6 regions of intercalary heterochromatin showed no overlap with any of the lamin-associated domains, and one region of intercalary heterochromatin encompassed five separate lamin-associated domains. Complete overlap was observed for 4 regions of intercalary heterochromatin (36) supporting our conclusion that intercalary heterochromatin is not completely necessary or sufficient for the formation of a Chr-NE contact.

2.5.3 Overall conclusions. The recently discovered correspondence between the organization of polytene and non-polytene chromosomes of *D. melanogaster* (148) has revived interest in using polytene chromosomes to study the 3D organization of the genome. Chromosome tracing experiments, as demonstrated in several classic studies (28,30,31,177), can be used to reconstruct the 3D organization of polytene chromosomes; however, these types of experiments still remain bottlenecked by the labor required to trace even a small ensemble of nuclei. Our study shows that experimentally parameterized computational models can assist studies of experimentally reconstructed nuclei. Our computational models complement a previous experiment (30) by revealing 33 new Chr-NE contacts and 5 anti-contacts; most of the 33 new contacts have properties of heterochromatin. However, the intercalary heterochromatic regions in *D. melanogaster* number more than 100 (58) and the complete rules for chromosome positioning with respect to the NE remain undiscovered. Further experiments may reveal additional Chr-NE contacts corresponding to the remaining regions of heterochromatin, or perhaps, additional contacts composed of both heterochromatin and euchromatin. The composition of contacts and anti-contacts in this study suggest a conclusion similar to that in previous studies (12): that the placement of chromosomes at or away from the NE does not depend exclusively on chromatin type and a more complicated set of rules governs the formation of Chr-NE contacts. Importantly, our computational modeling indicates that confinement of chromosomes in a spherical nucleus alone does not favor the positioning of specific chromosome regions at the NE as seen in experiment.

CHAPTER 3

QUANTIFIED EFFECTS OF CHROMOSOME-NUCLEAR ENVELOPE ATTACHMENTS ON 3D ORGANIZATION OF CHROMOSOMES

Short Title: Effects of chromosome-nuclear envelope attachments

Nicholas Allen Kinney[1], Alexey V. Onufriev*[1,2,3] and Igor V. Sharakhov*[1,4]

Used with permission of *Nucl Austin Tex*

Kinney NA, Onufriev AV, Sharakhov IV (2015) Quantified effects of chromosome-nuclear envelope attachments on 3D organization of chromosomes. *Nucl Austin Tex* 6: 212–224.

[1] Genetics Bioinformatics and Computational Biology, Virginia Tech, Blacksburg, VA, 24061

[2] Department of Physics, Virginia Tech, Blacksburg, VA, 24060

[3] Department of Computer Science and Applications, Virginia Tech, Blacksburg, VA, 24061

[4] Department of Entomology, Virginia Tech, Blacksburg, VA, 24061

***Corresponding authors:**

Alexey Onufriev Tel: (540) - 231 - 4237 Email: alexey@cs.vt.edu

Igor Sharakhov Tel: (540)-231-7316 Fax: (540)-231-7126 Email: igor@vt.edu

Keywords: chromosome-chromosome contacts, chromosome-nuclear envelope attachments, chromosome territories, polymer models.

Abbreviations: Chr-NE, chromosome-nuclear envelope; DAPI, 4',6-diamidino-2-phenylindole; MSD, mean squared deviation; PBS, phosphate buffered saline; SAW, self-avoiding walks.

3.1 ABSTRACT

We use a combined experimental and computational approach to study the effects of chromosome-nuclear envelope (Chr-NE) attachments on the 3D genome organization of *Drosophila melanogaster* (fruit fly) salivary gland nuclei. We consider 3 distinct models: a Null model – without specific Chr-NE attachments, a 15-attachment model – with 15 previously known Chr-NE attachments, and a 48-attachment model – with 15 original and 33 recently identified Chr-NE attachments. The radial densities of chromosomes in the models are compared to the densities observed in 100 experimental images of optically sectioned salivary gland nuclei forming "z-stacks." Most of the experimental z-stacks support the Chr-NE 48-attachment model suggesting that as many as 48 chromosome loci with appreciable affinity for the NE are necessary to reproduce the experimentally observed distribution of chromosome density in fruit fly nuclei. Next, we investigate if and how the presence and the number of Chr-NE attachments affect several key characteristics of 3D genome organization; this analysis leads to novel insight about the possible role of Chr-NE attachments. Specifically, we find that model nuclei with more numerous Chr-NE attachments form more distinct chromosome territories and their chromosomes intertwine less frequently. Intra-chromosome and intra-arm contacts are more common in model nuclei with Chr-NE attachments compared to the Null model, while inter-chromosome and inter-arm contacts are less common in nuclei with Chr-NE attachments. We demonstrate that Chr-NE attachments increase the specificity of long-range inter-chromosome and inter-arm contacts. The predicted effects of Chr-NE attachments are rationalized by intuitive volume *vs.* surface accessibility arguments.

3.2 INTRODUCTION

Experimental studies of polytene nuclei in fruit flies and nonpolytene nuclei in other eukaryotes have revealed several common principles of nuclear architecture. Chromosomes fold within the confined nuclear volume to occupy well-defined spatial domains called "territories." These domains are mutually exclusive in the following sense: although different chromosomes may be in physical contact, they never interweave (28,178) (as do, for example, strands of DNA in the double-helix). The 3D chromosome folding is manifested in specific intra- and inter-chromosomal interactions. These interactions have been observed microscopically (28) and, more recently, via cross-linking Hi-C experiments (109). The locations of inter-chromosomal

contacts correlate with fragile sites where chromosomal breakpoints occur in evolution (22) and tumorigenesis (179). In addition to chromosome-chromosome interactions, there are permanent (centromeric) and statistically high-frequency (but non-permanent) attachments between certain chromosomal loci and the nuclear envelope (NE). Several lines of evidence, including genome-wide mapping studies in *Drosophila* and human nuclei using the DamID method show that the gene-poor and transcriptionally repressed regions tend to form high-frequency Chr-NE attachments (12,180-182). Still, the interplay among these principles of 3D nuclear organization is poorly understood (183). Is the formation of chromosome territories controlled by the Chr-NE attachments, or vice versa? Does the number of chromosome-NE attachments influence the patterns of intra- and inter- chromosomal interactions? Is it necessary for computational models to include all of the Chr-NE attachments in order to recapitulate experimental data? Our study aims to address these and several related questions using a combined experimental and computational approach.

Experimental evidence indicates that Chr-NE attachments are present in diverse organisms including fruit fly (28), yeast (184), and human (185). In recognition of this growing body of evidence, many computational studies of genome organization now incorporate the specific sites of Chr-NE attachment as model parameters (71,77-80,123,132,186). Remarkably, these attachments are emerging as key components of 3D genome organization, which enhance agreement with experiment. In yeast, computational studies have considered a range of models differing in the number of attachments they consider (71,77-80,132). Homogeneous interaction models assume all chromosome sites interact equally with the NE due to the complete presence or absence of attachments at all sites along the chromosome fiber (132). Heterogeneous interaction models allow affinity for the NE to vary along the chromosome fiber; several models have specifically investigated the effects of chromosome tethers positioned at the centromeres and telomeres (71,77-80,113,132). These studies in yeast have led to several predictions: the 3D distribution of a gene is altered due to the presence of a NE tether positioned within 10kb (80); removal of a Chr-NE tether increases chromosome mobility as quantified by its confinement radius (71); the presence of Chr-NE tethers affects the distribution of telomere-telomere distances (132); the position of chromosomes within the nucleus may be altered due to a combination of Chr-NE tethering and volume exclusion (79); the distribution of distances

between the spindle pole body and silent mating locus depends on tethering at the telomere (80). These results have motivated more complicated heterogeneous models to consider more numerous sets of experimentally identified Chr-NE attachments (22). Although the distribution of telomere-telomere distances did not depend appreciably on attachment number (132), the distance distribution between the spindle pole body and silent mating locus was unique in the presence of zero, one, or two attachments (80).

We consider *D. melanogaster* (fruit fly) which, compared to yeast, possesses a more complex set of chromosome nuclear envelope attachments and different chromosome organization in interphase. In *Drosophila*, chromosome interactions decay more gradually with genomic distance (22,89), a change that may stem from the underlying difference in the chromosome volume to nucleus volume ratio (45) of these two organisms. In addition Chr-NE attachments are more numerous in *Drosophila* and predominately correspond to sites of intercalary heterochromatin (30). These additional attachments cannot be adequately modeled with centromere and telomere tethers alone. For example, seminal experimental work identified 15 chromosome regions frequently in contact with the NE in *D. melanogaster* polytene chromosomes (30); most of these contacts were located in regions of intercalary heterochromatin. A follow-up work identified 48 attachment sites (187), 45 located in regions of heterochromatin or late replication regions and 3 possessing properties of true euchromatin. What is the effect of Chr-NE attachments when they are positioned within these regions and coupled with the unique parameters of the *Drosophila* nucleus? We investigate how fundamental characteristics of the 3D chromosome organization - chromosome territories, chromosome intertwining, and gene-gene contact probability - change under the influence of different sets of Chr-NE attachments. In addition to quantifying these changes, we develop a simple and robust mechanistic understanding of the underlying fundamental reasons for the changes. Furthermore, which is the right number of attachments that one needs to include in computational models? To address these issues within a model, one needs to vary the number of Chr-NE attachments in a computational model, including all previously known and recently identified Chr-NE attachments, and verify the outcomes directly against experiment. These questions are addressed using rigorous definitions of chromosome territories (187), chromosome intertwining, gene-gene contact probability, and experimental density profiles of chromosomes in the nucleus.

Including all of the Chr-NE attachments could be necessary because experimental evidence suggests that they play a significant role in determining 3D genome organization (52,71,74,79,80,83,188). For instance, live imaging of yeast (74) and fruit fly (188) chromosomes have revealed short time-scale Brownian movements confined within variably sized domains. The movement of chromatin loci depends on their nuclear localization; loci attached to or adjacent to the nuclear periphery are less mobile than other loci (83). Thus, it is likely that attachments to the NE could constrain chromatin folding and prevent chromosome territories from freely diffusing in the entire nuclear volume. Similar conclusions have been reached in computational studies of yeast where each gene's genomic separation from the telomeric tether directly affected its spatial distribution (80). Alterations such as these may in turn be linked to chromosomal rearrangements since cytogenetic and genomic studies performed on various organisms have provided evidence for non-uniform distribution of chromosomal rearrangements and breakpoints (189,190). Analysis of the architecture of these nuclei revealed a general principle: because of non-random nuclear organization, certain loci are non-randomly closer together in the 3D space than others, and thus are more likely to interact and generate these rearrangements (191,192). Therefore, Chr-NE attachments may play a role in non-uniform distribution of chromosomal rearrangements. Because their possible effects on chromosome topology is largely unknown when they are numerous in D. *melanogaster*, this study aims to provide insights into this important issue.

Polytene chromosomes from *D. melanogaster* salivary gland nuclei are a well-established model for studying organization and function of the eukaryotic genome (143-145). Each polytene chromosome contains approximately 1024 DNA replicas bundled together in parallel; thus the genome organization in a single nucleus becomes visible under a light microscope. This is a critical advantage over "regular" interphase chromosomes because it is possible to obtain detail spatial information about the position of each individual polytene chromosome. The study of polytene chromosomes has significant potential for general understanding of 3D genome organization because recent experiments revealed very similar structural and functional organization of non-polytene and polytene chromosomes in fruit fly (36,147,148). Moreover, the polytene chromosomes are estimated to occupy about a third of the nuclear volume (31); this

chromosome to nuclear volume ratio, which critically affects the over-all 3D nuclear architecture (43), is the same in regular non-polytene nuclei (40), and is likely similar to the values characterizing human nuclei (75).

Our current work aims to determine the *effect* of Chr-NE attachments on the 3D organization of the genome in *D. melanogaster*. This addresses a fundamentally new question compared to our previous work aimed at merely *identifying* and characterizing Chr-NE attachments (187). In total, here we consider three computational models of the nucleus; all three models contain the same relevant measured parameters of the polytene nucleus from *D. melanogaster*, and differ only by the number and type of Chr-NE attachments incorporated. (I) the Null model – zero specific Chr-NE attachments; the only model considered in a previous work (187), which here serves as a reference for determining the effects of Chr-NE attachments. (II) a 15-attachment model, containing the same set of 15 Chr-NE attachment identified in (30). (III) a 48-attachment model, containing the same set of 48 attachments identified in (187). We emphasize that Chr-NE attachments are not randomly positioned; rather, the position of attachments in each model has been determined experimentally. Here, the Null model mimics a hypothetical mutant that would completely abolish all Chr-NE interactions, while the 48-attachment model corresponds to the experimentally accessible wild type. The 15-attachment model is an intermediate case. These computational models are in turn used to determine the effect of Chr-NE attachments on the other principles of the 3D genome organization.

3.3 MATERIAL AND METHODS

3.3.1 Salivary gland preparation

The preparation of salivary glands has been described elsewhere (28,193). Briefly, salivary glands are dissected in phosphate buffered saline (PBS) and then fixed in 3.7% formaldehyde and 1% triton. Glands were removed from the fixative solution after 15 minutes, washed, and stained for 45 minutes in 3 µg/ml 4',6-diamidino-2-phenylindole (DAPI). Glands are whole mounted on slides and nuclei were optically sectioned in 312x312x30 voxel images (z-stacks) with an inverted Zeiss LSM 510 Laser Scanning Microscope (Carl Zeiss MicroImaging, Inc.,

Thornwood, NY, USA). The microscope resolution should have little effect on radial density measurements (described below) since each voxel is smaller than the chromosome radius.

3.3.2 Radial density measurement

To determine the radial density of chromosomes from z-stack images we use MATLAB® (158). Each z-stack image is composed of voxels containing quantitative fluorescent intensity data, which depends on the position of chromosomes in the nucleus. Sobel edge detection is used to detect the NE in each z-stack image (194). Z-stack images are normalized such that $\sum_{\vec{r}} I(\vec{r})V = 0.3$, where 0.3 is the known chromosome volume to nucleus volume ratio (31).

Radial density and fraction of chromosome at the nucleus periphery follow directly from the normalized z-stack images. The radial density of chromosomes measured from each z-stack is compared to that of computational models.

3.3.3 Modeling approach

The five largest chromosome arms of *D. melanogaster* salivary glands are modeled as beads-on-string (41-43,82,85,132) and are represented as five random self-avoiding walks (SAWs) (15,84,154,155) (**Fig. 1**). The sixth arm, chromosome 4, is not considered due to its negligible length. Experimental data (28,30,31) for the chromosomes and the nucleus become realistic model parameters and constraints imposed during the construction of SAWs. The sampling protocol for generating conformational ensembles has been described previously [2]. Briefly, one bead representing the chromocenter is placed adjacent to the NE at the "north pole" of the nucleus. The arrangement of the five initial beads placed around the chromocenter bead is designed to match the relative proportion of chromocenter spatial arrangements seen in experiment (30). After assigning the chromocenter arrangement, SAWs are constructed using Rosenbluth's algorithm (96) (i.e. SAW chains grow by addition of monomers in a "true" SAW fashion). We use a Rosenbluth algorithm for computational efficiency. Complete details of model parameters are described in **Text S1** and **Table S1**.

Bead size and chromosome thickness

The diameter of *D. melanogaster* polytene chromosome can range from 3.1-3.2 microns (31). Our model uses beads with a diameter of 3.1 microns. Each bead represents approximately 0.7 Mb of genomic material. To fully capture the thickness of the chromosome fiber we place a cylinder of excluded volume around the bond between nearest neighbor beads. This detail was important for achieving the right nucleus volume to chromosome volume ratio, but was found to have little effect on the scaling of our self-avoiding walks in free space (187).

Polytene chromosome persistence length

A 1.5 micron persistence length for *D. melanogaster* salivary gland polytene chromosomes has been inferred from experimental data (195). A 1.5 micron persistence length means the effective Kuhn length of our model is about twice the persistence length, meeting the condition necessary to build our models as a SAW (85,196). We recognize that recent studies have questioned the standard definitions of persistence length (197) for chains with self-avoidance. Despite the debate, we emphasize that our approach based on the polymer Kuhn length is de rigueur in many current polymer models (43,75,78,156,187) and the concept has worked well for several decades.

Three models incorporate specific sets of Chr-NE attachments

Specific (non-random) chromosome regions have been shown experimentally to associate with the NE in a significant fraction of nuclei (30); these "Chr-NE attachments" have typically been determined in chromosome tracing studies that reconstruct the 3D organization of the chromosomes in the nucleus (28,30). One study (30) identified 15 Chr-NE attachments each present in 66% of experimentally reconstructed nuclei. These 15 experimentally identified (30) Chr-NE attachments are mapped to the corresponding beads in our computational model with the same frequency (0.66) of NE attachment; the computational nuclei constructed with this set of Chr-NE attachments belong to our "15-attachment model". In our previous work (187) we identified 33 additional Chr-NE attachments present in 51% of experimental nuclei. Here, these Chr-NE attachments are mapped to the corresponding beads in our computational model with the same frequency (0.51) of NE attachment. We thus construct a "48-attachment model" that has the 15 Chr-NE attachments from (30) plus the additional 33 Chr-NE attachments from (187). These two are novel models in the sense that, for the first time, they incorporate the full complement of previously known and recently identified Chr-NE attachments. We emphasize

that the number of attachments differs in each model since they enforce different numbers of Chr-NE attachments. The specific sets of Chr-NE attachment and their mapping onto our computational models is provided in **Table S1**. Our "Null model" contains no Chr-NE attachments and has been considered previously (187). The 3D genome organization of each ensemble is discussed in the context of our experimental results below.

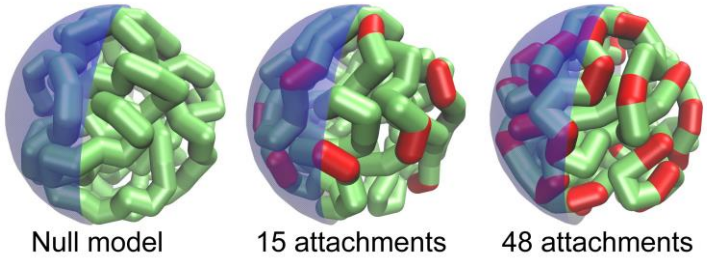


Figure 3.1 - Visual summary of the three models of nuclear architecture in *D. melanogaster* polytene nuclei. Two identified sets of Chr-NE attachments (red) are mapped onto a computational model of *D. melanogaster* polytene nucleus. The effects of Chr-NE attachments are compared to a previously developed Null model lacking Chr-NE attachments.

Robustness of conclusions to model details

The SAW approach used here to model polytene chromosomes has been utilized in multiple studies in similar contexts (43,53,75,77,78,187). Briefly, we construct our SAW's using an unweighted Rosenbluth algorithm (96). For short chains this approach is a good approximation of self-repelling chains that are true equilibrium states of polymers (97). Although we use a SAW approach, which is equilibrium by construction (to the extent that it approximates self-repelling chains), each of our three models contain non-equilibrium features introduced to better represent experiment. These include: (1) Rabl configuration of chromosomes, (2) right-handed chromosome chirality, and (3) asymmetric chromocenter arrangement. We refer to the use of these modifications as a *modified self-avoiding walk approach.* The key model conclusions in this study are the effects of Chr-NE attachments on the 3D organization of the genome (see main text); these effects are investigated by comparing simulated ensembles of models possessing differing sets of chromosome nuclear envelope attachments (these results are discussed thoroughly in the main text). We check that the crucial model conclusions are robust to the non-equilibrium features of our modified self-avoiding walk approach. For robustness checking of the model conclusions we use an *unmodified self-avoiding walk* approach - does *not* introduce (1) Rabl chromosome configuration, (2) right-handed chromosome chirality, or (3) non-random

chromocenter arrangement and is equilibrium to the extent that our chain growing algorithm approximates self-repelling chains (see main text). The same conclusions are reached with these models (**Text S5** and **Table S4**). Additional discussion of these parameters and their effect on the simulation outcomes can be found in a previous work (187). We emphasize that robustness checking was specifically designed to assess the modeling algorithm and not every conceivable parameter of the model itself. In particular, we did not test robustness of the model conclusions against the size of the nucleus, chromosome confinement, or aspect ratio of the nucleus. Although these test are conceivable, testing robustness against a range of parameters that have not been measured biologically is impractical.

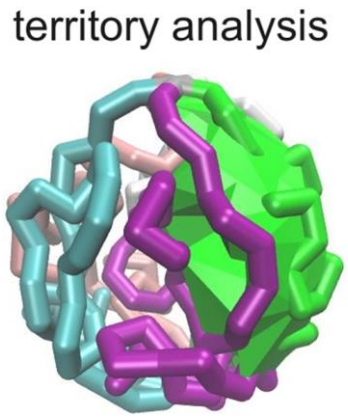


Figure 3.2 - The convex hull is used to quantify chromosome territories. The convex hull of the green chromosome above may encompass beads belonging to other chromosomes. The ratio of native beads compared to all (including foreign) beads in a hull is the territory index.

3.3.4 Computational simulations of the models

Large conformational ensembles (i.e. snapshots) of nuclei were simulated for each model of Chr-NE attachment: Null model – 3,193 simulated nuclei, 15-attachment model – 3,623 simulated nuclei, and 48-attachment model – 3,477 simulated nuclei. The generating of conformational ensembles was performed on 240 parallel CPUs; approximately 40,000 CPU hours total. Sampling protocol for generation of computational ensembles is provided (**Text S2**).

Analysis of the simulated ensembles of nuclei

The simulated ensembles of nuclei are analyzed at three spatial resolutions including whole chromosome resolution, whole arm resolution, and single bead resolution (the highest resolution of our model, about 0.7 Mb of linear sequence). To quantify bead-bead interactions (single bead resolution), we calculate how often specific pairs of beads are in contact in an ensemble of computational nuclei; a contact is defined as two microns or less separation between the surfaces of two beads. There are 248 beads in a single computational nucleus (including all of the 5

chromosome arms); a 248x248 contact map has entries for each possible pair of beads. The "(i,j)" entry of this contact map gives the probability that bead "i" and bead "j" in our computational nuclei form a contact. We also compute contact probability at two lower resolutions: chromosome arm resolution – bins of beads belonging to the same chromosome arm, whole chromosome resolution – bins of beads belonging to the same chromosome (i.e. chromosome 2 right and left arms). These interaction types are normalized against the total number of interactions in each model, e.g. the number of intra-arm interactions in our Null model out of all interactions in our Null model. Since our simulations generate equilibrium conformational ensembles of model nuclei (187), we do not calculate time dependent observables such as auto-correlation times. Furthermore, we do not consider the proposed molecular clamp forces that may be provided by cohesins, which are thought to play an important role in chromosome territory formation and maintenance (71). In addition, we do not consider compaction forces provided by condensins (131) or the movement of Chr-NE attachments along the NE surface. Each of these considerations would necessarily require a fully dynamical model capable of describing time-evolution, which would be out of scope of this study. Here, we model mature polytene chromosomes that have reached quasi-equilibrium, with essentially no time-evolution of the relevant experimental time-scales.

Chromosome territories

We quantify chromosome territories by addressing how each chromosome excludes other chromosomes from the volume it occupies in the 3D space. We begin by calculating the convex hull for a single chromosome -- this is the minimum volume that includes all the chromosome's points (bead centers) inside a convex polyhedron. In general, each convex hull contains its own chromosome, and may also encompass some beads belonging to other chromosomes. A fully "territorial" chromosome is one whose convex hull does not contain points from any other chromosomes, while a less "territorial" chromosome is one whose convex hull contains some points from other chromosomes (**Fig. 2**). We define the chromosome territory index as the fraction of points inside a convex hull that belong to the chromosome used for its construction. A detailed description of chromosome territories is given in our recent work. (187)

Chromosome intertwining

Chromosome intertwining is rigorously assessed by attempting to separate model chromosomes by putative translations in 3D space (**Fig. 3**). Chromosomes that can be spatially separated by at least one translation are called non-intertwining. We begin by selecting the backbone of two chromosomes (excluding centromere) from a model nucleus; the backbone of a model chromosome consists of the line segments connecting the centers of each bead. A direction vector is then chosen to translate the backbone of one model chromosome (**Fig. 3**). If the two backbones cross during this translation then a new direction vector is picked. A total of 162 different direction vectors are tested in this manner, the tips of the vectors uniformly cover the surface of a unit sphere, as detailed in our recent work. (187)

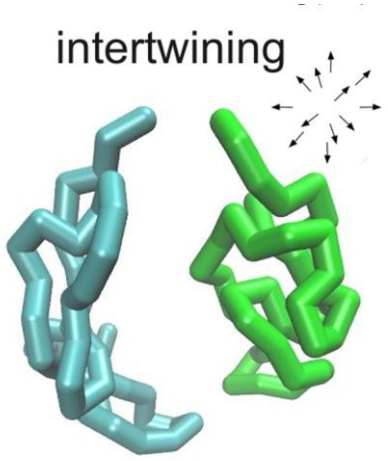


Figure 3.3 - Spatial separation of chromosomes with respect to putative translations. This technique is used to quantify chromosome intertwining. The blue and green chromosomes above do not intertwine because a translation can be found that separates the chromosomes without crossing.

3.4 RESULTS

3.4.1 The presence of Chr-NE attachments enhances chromosome territories and reduces chromosome intertwining

We use recently developed rigorous metrics (187) (briefly summarized in Material and Methods) to quantify the effect of Chr-NE attachments on chromosome territories and chromosome intertwining in our computational models. We can easily predict the effects of Chr-NE attachments by comparing simulated nuclei, which have all or some Chr-NE attachments, against the Null model nuclei (0 attachments). The results are summarized in **Fig. 4**. Overall, the presence of Chr-NE attachments leads to chromosomes that are more territorial and intertwine less frequently. Compared to the Null model the territory index of the 15 and 48 attachment nuclei ensembles increased by 1.8% and 4.6%, respectively. The chromosomes in the 15 and 48

attachment nuclei ensembles intertwine 1.6%, and 3.0% less frequently, respectively, which also means that the chromosomes become more territorial in the presence of Chr-NE attachments. The statistical significance in the mean territory index and intertwining frequency was tested for each model using the Null model as a Null hypothesis. P-values for the 15 and 48 attachment models support the statistical difference compared to the Null model (figure 3.4).

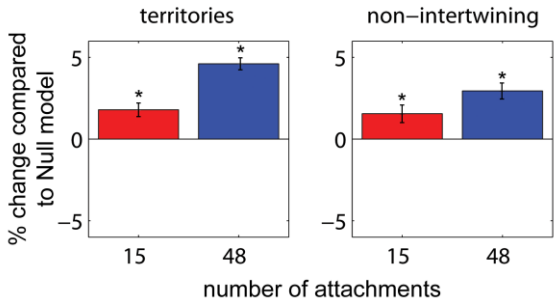


Figure 3.4 - Effects of "turning on" Chr-NE attachments on chromosome territories and intertwining. At whole genome resolution, chromosomes become more territorial and intertwine less frequently. (*) indicates P-value < 0.001. Bars indicate 5% confidence interval.

3.4.2 Chr-NE attachments affect intra- and inter- chromosomal contacts in opposite ways

In addition to the global organizational effects of Chr-NE attachments, we predict the more local changes induced by introducing Chr-NE attachments into our computational models. We analyzed bead-bead contact probability in our models at two progressively increasing resolutions (see Material and Methods): whole chromosome (40 Mb) and chromosome arm (20 Mb). Interestingly, Chr-NE attachments did not equally affect all types of chromosome-chromosome contacts. Specifically, intra-chromosome and intra-arm interactions are more common in the Chr-NE attachments models compared to the Null-model, while inter-chromosome and inter-arm interactions are less common in the Chr-NE attachment models compared to our Null model (**Fig. 5**).

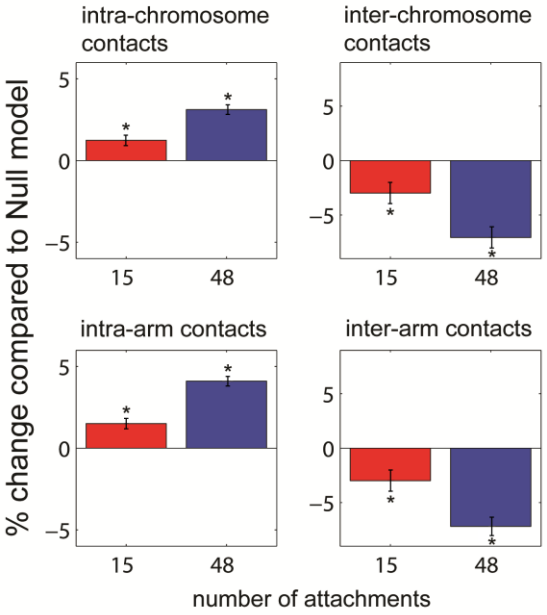


Figure 3.5 - Effects of "turning on" Chr-NE attachments on chromosomal interactions. The relative number of intra-chromosome and intra-arm contacts increases because of the "turning on" the Chr-NE attachments, while inter-chromosome and inter-arm contacts decrease. (*) indicates P-value < 0.001. Bars indicate 5% confidence interval.

This observation is consistent with the more territorial chromosomes present in our Chr-NE attachment models (**Fig. 4**). Overall, the changes in pairwise interaction probability (compared to our Null model) are proportional to the total number of Chr-NE attachments. For example, normalized intra-arm contacts increased by 1.5% in the 15 Chr-NE attachment ensemble compared to the Null model and increased again (by 4.1%) in the 48 Chr-NE attachment ensemble compared to our Null model (**Fig. 5**). The statistical significance in the mean number of intra/inter chromosome and intra/inter arm interactions was tested for each model using the Null model as a Null hypothesis. P-values for the 15 and 48 attachment models support the statistical difference compared to the Null model (figure 3.5).

3.4.3 Chr-NE attachments increase the specificity of long-range inter-arm and inter-chromosome contacts at the highest resolution of the model (0.7Mb)

We considered bead-bead contact frequencies at the highest resolution (bead size=0.7 Mb) of our model. Changes in contact probability at bead-bead resolution for each model were compared to the Null model (**Fig. 6**). The result shows that the predicted organizational changes brought about by Chr-NE attachments are non-random. For example, the likelihood of specific inter-chromosome contacts and inter-arm contacts may increase due to Chr-NE attachments (two examples are indicated by black arrows in **Fig. 6,** the matrix is symmetric) despite the overall decreased probability of inter-chromosome and inter-arm contacts (**Fig. 5**). In biological terms, our model predicts that chromosomal regions 93D-94E and 52F-53F (top arrow in **Fig. 6**) interact more frequently in the presence of Chr-NE attachments than expected in the absence of Chr-NE contacts. The rightmost arrow in **Fig. 6** corresponds to chromosomal regions 31E-32D interacting with regions 94E-96A more frequently in the presence of Chr-NE attachments than expected in the absence of Chr-NE contacts. Interestingly, these hotspots of increased bead-bead interaction correspond to the chromosomal regions that void NE attachments (187), suggesting that the interacting beads may loop into the nuclear interior, which increases the probability of contact.

3.4.4 Experimental radial density of polytene chromosomes changes with the distance from the NE

To date, two sets of Chr-NE attachments in *D. melanogaster* have been identified: a smaller set of 15 "high frequency only" attachments (30), and a larger set that includes the additional 33 attachments recently identified (187); these sets are enforced in our 15 and 48 Chr-NE attachment ensembles, respectively. Here we preform new experiments to determine which set better represents reality. This is important since our simulations clearly suggest that the number of Chr-NE attachments may affect several key parameters of the nuclear architecture (**see results above**).

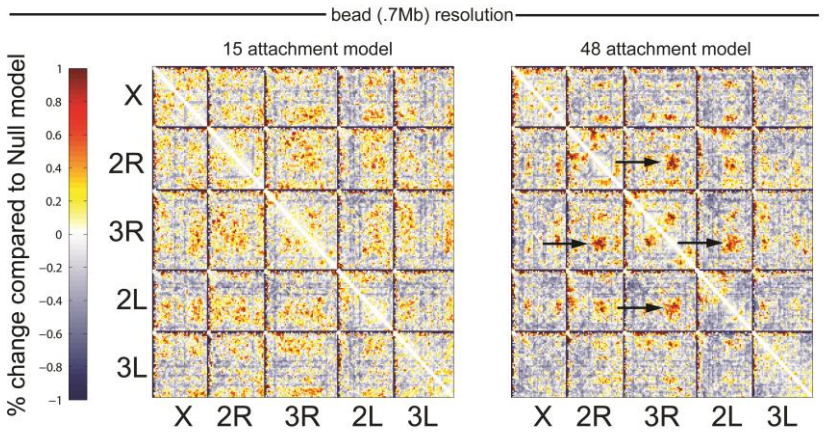


Figure 3.6 - Effects of "turning on" Chr-NE attachments, quantified at the bead resolution. Chr-NE attachments increase the specificity of long-distance inter-arm and inter-chromosome contacts in the simulations. Examples are indicated by black arrows.

We collected fluorescence images of optically sectioned fruit fly salivary gland nuclei forming "z-stacks" for an ensemble of nuclei - 100 in this study. The radial density of chromosome in each z-stack image is inferred from fluorescence intensity (see Material and Methods for details). **Fig. 7** shows the average radial density, $\langle \rho(r) \rangle_{n=100}$, of chromosomes over the ensemble of 100 experimental nuclei. The peak density of chromatin occurs near the periphery of the nucleus, approximately 2 µm from the NE. Chromatin density steadily decreases

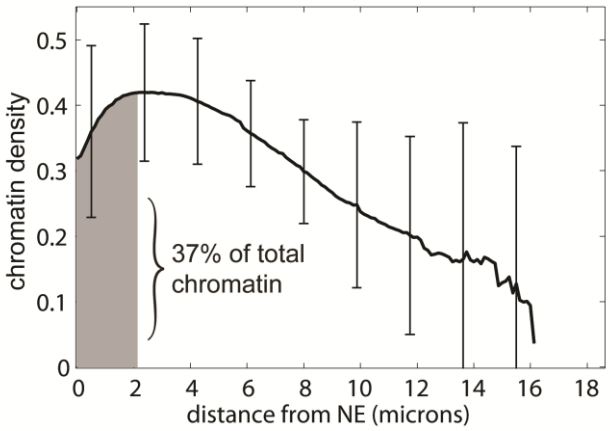


Figure 3.7 - Experimentally measured radial density of chromatin. The data are obtained from ensemble of 100 experimental nuclei from *D. melanogaster* salivary glands.

towards the nuclear center (**Fig. 7**). The integrated density of chromatin (total chromatin) within 2 μm of the NE accounts for about 37% of all chromatin in our experimental nuclei. This result is consistent with earlier complete tracing experiments (30), in which approximately 40% of each chromosome arm was located at the nuclear periphery. Bars indicate the standard deviation in experimental radial chromatin density.

3.4.5 The 48 Chr-NE attachment model better represents the real nucleus than the 15 Chr-NE attachment model or the Null model with no attachments

To compare our computational models with the experimentally measured radial density of chromosomes, we first computed the average radial density of chromosome in the Null model **(green curve in Fig. 8, top panel)**, 15 Chr-NE attachment model **(red curve in Fig. 8, top panel)**, and 48 Chr-NE attachments model **(blue curve in Fig. 8, top panel)**. Bars indicate the standard deviation. Then for each of the 100 experimental z-stack images, the radial density of chromosomes was computed and compared to that of each model. To quantify the ability of each model to recapitulate the experimental data we used the mean squared deviation (MSD). The MSD is designed to *objectively quantify* the difference between the model and experimental curves; the total variation is not used because we are not interested in the arc-length of the chromatin density curves. We did not test every metric that could in principle be applied to the curves; for our purposes the MSD provides a sufficient measure of the difference between the model and experiment. The minimum MSD determines the "best fit" model. The number of best fits indicates how well each model recapitulates the experimental data. In

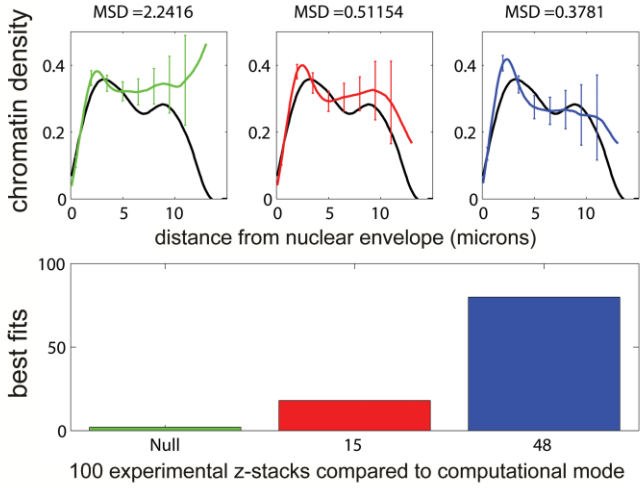


Figure 3.8 - Models of nuclear organization compared to experimental nuclei. Example of a single experimental measurement (black) that best fits the 48 Chr-NE attachment ensemble (blue). Most of the experimental measurements fit the 48 Chr-NE attachment ensemble.

general comparing orthogonal chromatin density is difficult because both the measured density curves and those present in computational models varies from nucleus to nucleus. We make this problem tractable by comparing the individual experimental z-stacks rather than simply the average of all experimental z-stack that would eliminate the intrinsic variability. Thus, the approach is designed to test how many individual experimental nuclei map to each model, not just the population mean which may be somewhat different from its constituent parts. The number of best fits for each model is shown in **Fig. 8, bottom panel**. The 48 Chr-NE attachment model scores the greatest number of best fits, while the Null model and the 15 Chr-NE attachment model score a considerably smaller number of best fits. The fact that not every experimental nucleus has 3D distribution of chromosomes best described by a single model is not surprising: the folding of chromosomes and placement of Chr-NE attachments varies from nucleus to nucleus. Therefore, the radial density of chromosomes measured in z-stack images varies for individual nuclei. What is important is that the entire ensemble of experimental images favors the 48 Chr-NE attachment model. We conclude that, within the assumptions made by our computational models, the 48 Chr-NE attachment model better represents reality than the 48 Chr-NE attachment model or the Null model. It is worth mentioning that in contrast to polytene chromosomes considered here, non-polytene chromosome experiments suggest that Ch-NE attachments may number in the hundreds (12,198). These studies used a DamID approach, in which DNA adenine methyltransferase (Dam) fused to lamin leaves a stable adenine-methylation 'footprint' *in vivo* at the interaction sites. One DAM-ID study identified 412 *Drosophila* Lamina Associated Domains (LADs) with a median size of ~90 kb (198). In contrast, the Chr-NE attachment regions in polytene chromosomes corresponds to one or few subdivisions on a cytogenetic map (30), which typically span several hundred kilobase pairs. The difference between polytene and regular interphase chromosome models may stem from the necessary coarse graining of the polytene model. Since each bead represents 0.7Mb, clusters of LADs along the bundle of polytene strands may be impossible to resolve using a polytene chromosome approach. This seems likely since it has been demonstrated that smoothing of the LAD profile seen in interphase chromosomes (to simulate the microscopy resolution) makes it correspond well to the Chr-NE attachment profile seen in polytene chromosomes (12).

We stress that the fact that all 48 attachments appear necessary for the closest agreement between model and reality, does not completely resolve the still debated number of attachments present in real nuclei as each model necessarily makes simplifying assumptions. To name a few, the nucleus shape, nucleus size, and nucleolus position are fixed in our models to improve computational efficiency. It is therefore conceivable that the chromosome density profile and number of best fits may change slightly if these parameters varied as they may do in real nuclei. In order to incorporate complete sets of Chr-NE attachments (for the first time) these compromises simply had to be made to make our models computationally tractable. Along those same lines, we did not test what would hypothetically occur if the probability of high-frequency chromosome-NE attachments was allowed to vary from its fixed value of 66% and explore the full range of NE contact probability. Likewise the probability of sub-high frequency contacts was kept fixed at 51% just like experiment. Exploring the full range of conceivable parameters is computationally intractable and potentially irrelevant in cases where parameters very considerably from biological realism. We also recognize that a complete investigation of the effects of Chr-NE attachments in the context of many other potentially relevant chromatin properties has not been made in this work. For example, we have not investigated how changes in chromosome rigidity may alter the predicted effects of Chr-NE attachments. Also, we have not made a complete investigation of how the chromosome distribution of Chr-NE attachments may affect the quantified observables. These are interesting "what if" questions for a future comprehensive investigation. At the same time we note that all of the key model parameters (such as chromosome rigidity) currently used in this work are not arbitrary, but are taken directly from experiment. Thus, the evidence presented here demonstrates the importance of the additional recently identified 33 Chr-NE attachments (187).

3.4.6 Predicted effects of Chr-NE attachments are rationalized by simple volume *vs.* surface accessibility arguments

In what follows we explain why inter-arm and inter-chromosome contacts are less likely to

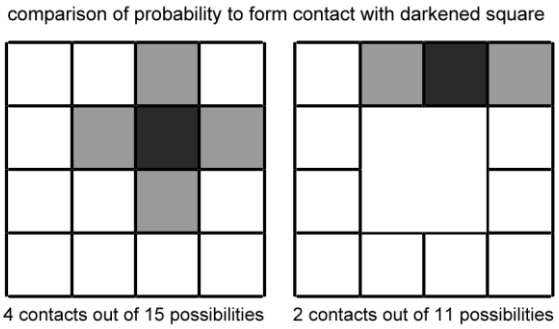


Figure 3.9 - Conceptual schematic of volume *vs.* surface accessibility argument.

occur in our non-zero Chr-NE attachment models compared to the reference Null model with no Chr-NE attachments; that is we explain why the "turning on" of the Chr-NE attachments affects the chromosome-chromosome contacts in the specific way predicted by the model. The demonstration is based on very general volume *vs.* surface space accessibility arguments. A schematic, intended only to illustrate the concept of the argument is shown in **Fig. 9**. First, consider a centrally located gene represented by the shaded square. Next, assume that a second gene randomly occupies one of the remaining squares. Four of these possibilities (grey squares) represent locus-locus contact; thus, locus-locus contact probability in this case is 4 out of 15, or ≈0.27 (**Fig. 9, left panel**). Consider another limiting case where all the genes are moved to the nuclear periphery. In this case only 2 possibilities (grey squares) represent locus-locus contact; thus, locus-locus contact probability in this case is 2 out of 11, or ≈0.18 (**Fig. 9, right panel**). An analogous "volume *vs.* surface" effect in our simulated nuclei reduces the inter-chromosome and inter-arm contacts in our models with Chr-NE attachments. The same basic argument is made quantitative in the next paragraph and rationalizes the increases and decreases in chromosomal contact probabilities as seen in our models.

To this end, we relate the specific contact probabilities to the following specific volumes: accessible volume for a bead – volume in which a bead is confined (**Fig. 10**), contact sub-volume for a bead – the small volume within which positioning of nearby beads form bead-bead contacts (cyan in **Fig. 10**). The three types of pairwise contacts in our models are: attachment-attachment - contacts between two "attachment" beads which form Chr-NE attachment, bulk-bulk - contacts between two "bulk" beads which are not attached to the NE (i.e. no constraints), and bulk-attachment – contacts of bulk beads with

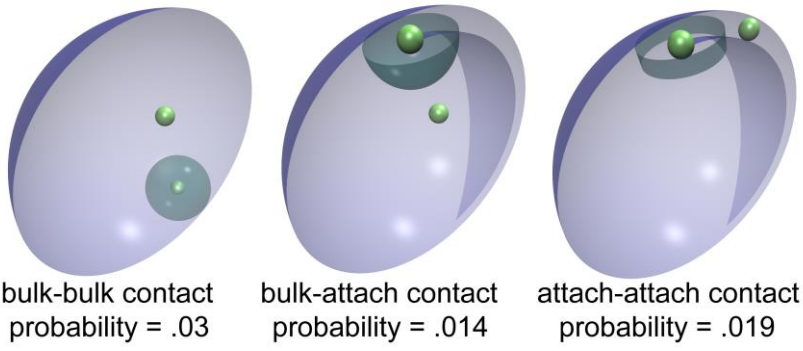


Figure 3.10 - Three types of pairwise interactions in our computational models. Two beads (green) are illustrated in each panel. The probability for these two beads to interact is approximated by dividing its contact sub-volume (cyan) by the accessible volume - entire nucleus for bulk beads or volume at the nuclear periphery for attachment beads.

attachment beads (**Fig. 10**). The accessible volume of bulk beads and attachment beads clearly differs; bulk beads are found anywhere in the nucleus while attachment beads are only at the periphery. In addition, the contact sub-volume (defined above and indicated with cyan regions) depends on the contact type. To the extent that most beads are randomly positioned in the nucleus relative to each other – a simplifying assumption of this demonstration – we can approximate the probability of contact for each interaction type by dividing its contact sub-volume by the accessible volume of the bead. **Fig. 10** summarizes contact probabilities by interactions type. Predicted derivation for these contact probabilities are given in the **Text S3, Fig. S1 and Table S2**. Note that $P_{bulk}^{bulk} > P_{attach}^{attach} > P_{attach}^{bulk}$, where variables represent the contact probability of bulk beads with bulk bead, attachment beads with attachment bead, and bulk beads with attachment bead respectively. Clearly all contacts in the Null model are bulk-bulk contacts. The "turning on" of each Chr-NE attachment essentially replaces bulk-bulk interactions with one of the two other interaction types, each having lower contact probability than the bulk-bulk. A total of 15 Chr-NE attachments are "turned on" in our 15-attachment model; therefore, fewer total inter-chromosomal interactions are realized compared to the Null model. The effect is even more pronounced in the 48 Chr-NE attachment model leading to fewer total inter-chromosomal interactions compared to the 15 Chr-NE attachment model. We uses these contact probabilities to predict the average number of interactions in each of our three models (Null model, 15-attachment model, and 48-attachment model) (**Table 1**). We arrive are these prediction by simply multiplying the number of possible interactions (see table headings) of each given type by the contact probability predicted by the volume vs surface accessibility argument and summing the results. Despite the number of simplifying assumptions made by our volume *vs*. surface accessibility argument, the resulting predictions are in relatively good agreement with our full SAW simulations. More importantly, our simple accessibility argument predicts a decrease in the number of inter chromosome-chromosome interactions in our Null model, 15 Chr-NE attachments model, and 48 Chr-NE attachment model, respectively (**Table 1**). Thus, the effects of Chr-NE attachments as seen in our simulations are indeed likely to stem in part from the very general volume *vs.* surface accessibility effects, which are very robust. In the case of intra-chromosomal interactions, we can no longer assume that beads are randomly positioned relative to each other due to the linking of beads along the polymer backbone. As a result, the relative ordering of contact probabilities, $P_{bulk}^{bulk}, P_{attach}^{attach}, P_{attach}^{bulk}$, changes, with P_{attach}^{attach} becoming the largest of

the three. That change leads to in an increase in the number of intra-arm and intra-chromsome contacts upon turning on of the Chr-NE attachments. See Text S4 and Table S3 for argument rationalizing the increase in intra-chromosome contacts due to the turning on of Chr-NE attachments.

Model	Possible bulk-bulk interactions	Possible bulk-attach interactions	Possible attach-attach interactions	Predicted number of interactions from volume-surface argument	Interactions seen in simulations
Null	248^2	0	0	$.03 \cdot 248^2 = 1845$	1625
15 attachment	233^2	$2 \cdot 233 \cdot 15$	15^2	$.03 \cdot 233^2 + .014 \cdot 2 \cdot 233 \cdot 15 + .019 \cdot 15^2 = 1731$	1591
48 attachment	200^2	$2 \cdot 200 \cdot 48$	48^2	$.03 \cdot 200^2 + .014 \cdot 2 \cdot 200 \cdot 48 + .019 \cdot 48^2 = 1513$	1538

Table 3.1 – Predicted number of interactions based on volume *vs.* surface accessibility argument compared to simulation. Although each of our computational models contains a total of 248*248 possible bead-bead interactions, the types of interactions differ. Since our volume *vs.* surface accessibility argument predicts a different contact probability for each interaction type, a different number of interactions in each model is expected.

3.5 DISCUSSION

3.5.1 Key outcomes

The close integration of experimental data analysis and computational modeling had already led to significant breakthroughs in deciphering some basic principles of the 3D architecture of interphase chromosomes (11,30,77,78,123,138,142,187,199). Despite this progress the Chr-NE attachments – a potentially critical component of 3D organization – have received limited attention in computational models. In particular, Chr-NE attachments were an essential feature of several recent computational models (77,78,123,138,156) and their inclusion was necessary to recapitulate experimental results. Our model described here was designed in an effort to simulate the effects of chromosome-NE attachments on nuclear architecture; it incorporates all known chromosome interactions with the NE. The model makes several key predictions, in two categories: effect on territories and effect on chromosome-chromosome interactions.

Specifically, we show that (1) chromosomes with more numerous NE attachments are more territorial; (2a) intra-arm and intra-chromosome interactions are more common in nuclei with more numerous Chr-NE attachments, (2b) inter-arm and inter-chromosome interactions are less common in nuclei with more numerous Chr-NE attachments, and (2c) Chr-NE attachments increase the specificity of long-range inter-arm and inter-chromosome interactions. We show how these conclusions can be rationalized by simple and robust volume *vs.* surface arguments, which further supports robustness of the conclusions themselves. Although our biological system is the polytene chromosome of *D. melanogaster*, recent experiments have demonstrated that polytene chromosome have very similar structural and functional organization compared to their nonpolytene counterparts (36,147,148).

3.5.2 Chromosomes with more numerous NE attachments are more territorial

Chromosome territories have become a focus of a number of experiments (7,14,30,31,157,168-170) and computational simulations (139,141,171). Recently, simulations that addressed the question of territory formation demonstrated that non-specific entropic forces may play a significant role in establishing and maintaining chromosome territories (139,141,171). However, simulations have also shown that other factors may be involved in addition to the dominant role played by entropic forces. In one study, this entropic effect has been shown to depend on the presence of chromosome loops, which may also arise due to non-specific forces (139). Our simulations suggest that chromosomes are largely territorial regardless of Chr-NE attachments. However, the simulated nuclei with more numerous Chr-NE attachments were consistently more territorial; therefore, Chr-NE attachments may play a certain role in maintaining the integrity of chromosome territories. This effect may be universal to some extent. Computational studies in yeast reveal that territorial organization of chromosomes is partially governed by chromosome tethering at the centromere (71). The similarity with our results is significant since the composition of Chr-NE attachments differs in fruit fly and yeast. Yeasts have fewer and predominately heterochromatic attachments, while *Drosophila* possesses more numerous attachments in regions of intercalary heterochromatin. Despite this progress, the influence of attachment number may be more elusive. In our models increasing attachment number from 15 to 48 likely increases the chromosome territory index and pattern of chromosome-chromosome interactions (see results). In contrast, altering attachment number in yeast models had little effect

on the chromosome-chromosome distance measured between telomeres at experimentally relevant resolution (132). The unique effects of attachment number in yeast and *Drosophila* may be coupled with the unique genome organizations in these two organisms (45,200); for example, the effect of attachments may change when the chromosome volume to nucleus volume ratio is altered. A combination of higher resolution experimental and computational studies focused on Chr-NE attachments will be needed to address these questions.

3.5.3 Chr-NE attachments affect whole chromosome and chromosome arm interactions

Experimental Hi-C data have recently mapped the global chromosome-chromosome interactions of *D. melanogaster* (non-polytene) nuclei (89). The pattern of global chromosome-chromosome interactions was characterized by the following trends: an abundance of intra-arm contacts and lack of inter-arm contacts. Interestingly, we observe this pattern (intra arm > inter arm) in all three of our computational models; however, the ratio of intra-arm/inter-arm contacts progressively increased by 4% and 11% with the addition of 15 and 48 Chr-NE attachments respectively. This suggests that intra-arm interactions may be more common than inter-arm interactions regardless of Chr-NE attachments with the added caveat that attachments may increase the ratio of the two. The abundance of intra-arm interactions in *Drosophila* may also be dictated by robust homolog pairing that competes away non-homologous contacts (201). Thus, Chr-NE attachments likely affect nuclear organization to a degree (see results) but they may not be the dominant factor. This result is consistent with recent computational studies of the yeast genome that showed that chromosome organization in the nucleus can be explained exclusively by the confinement of chromosomes and tethering of chromosome centromere and telomere to the NE (77,78). A number of other studies have also emphasized that non-specific (entropic) forces drive the self-organization of polymers (82,139,140,156,202). However, we do see that the Chr-NE attachments significantly increase intra-arm and intra-chromosome contacts and decrease inter-arm and inter-chromosome contacts. This observation is consistent with the effect of the Chr-NE attachments inducing more territorial organization of chromosomes in our attachment models.

3.5.4 Chr-NE attachments increase the specificity of long-range inter-chromosome and inter-arm interactions

Experiments have suggested that actively transcribed genes co-localize in order to share sites of transcription (203); a result that persists over several genomic length scales (203,204). For example, one study demonstrated frequent 3D co-localization of two genes (Hbb-b1 and Eraf) which are separated by ~25 Mb on mouse chromosome 7 (203). Another study, focusing on a 2.9 Mb region of *D. melanogaster*, demonstrated that transcription factor co-localization hotspots range in size from 1-5 kb and are separated by 50 kb (204). Unlike experiment (89), our Null model shows no specificity in inter-arm and inter-chromosome contacts, which are random; however, we do find that specific chromosome regions have the capacity to increase their likelihood to co-localize in the presence of Chr-NE attachments. Thus, the "turning on" of all 48 Chr-NE attachments in our model increases the specificity of inter-arm and inter-chromosome contacts. Likewise, specific long range chromosome interactions are a key feature of experimental Hi-C contact maps (89). We avoid directly comparing the co-localized regions in our 48 attachment model with Hi-C experimental data because specific chromosome-chromosome interactions are not featured as parameters in our computational models. These parameters are absent because specific chromosome-chromosome interactions have not been quantified experimentally in D. melanogaster salivary gland nuclei. It's conceivable that these interaction have a greater impact on chromosome-chromosome contacts than the presence of Chr-NE attachments and both factors superimpose their relative influence on chromosome-chromosome interactions; we have not conducted a study of this possibility. However, the fact that we clearly observe chromosome-chromosome interactions stemming from Chr-NE contacts leads to speculation about their effects at higher resolution. Specifically, If the effect of Chr-NE attachments as seen in our models persists at higher resolutions they may play a role in establishing these long-range chromosome interactions; this possibility will be investigated in a future study using models of nonpolytene interphase nuclei. On the other hand, a recent epigenetic based copolymer model in Drosophila reproduced the chromosome folding inferred from Hi-C data without depending on Chr-NE attachments (205). However, the regions of chromatin modeled in that study were on the order of 1Mb (205). Chr-NE attachments may have limited influence within these specific regions or the effect of attachments may be altogether limited within 1Mb regions. Indeed, the coarse grained beads in our model do span one or more subdivisions of the *D. melanogaster* polytene chromosome map and, therefore, the co-localization of beads, and the genes represented by those beads, may be relevant specifically to

polytene chromosomes. Interestingly, a recent computational study of gene co-localization demonstrated that simultaneously co-localizing gene pairs depends on a low rate of chromosome intertwining (206). This result is consistent with the correlation between the number or Chr-NE attachments and a low rate of chromosome intertwining. We hypothesize that the density of Chr-NE attachments is the source of these localized increases in inter-arm contact frequency. This result shows the importance of considering NE attachments for predicting key details of 3D nuclear architecture.

3.5.5 Methodology

In the pioneering study of genome organization, reconstructions of the 3D chromosome folding were obtained from optically sectioned polytene chromosomes from *D. melanogaster* (30). This approach provides a detailed picture of chromosome folding in the nucleus; however, it is typically limited to small ensembles of nuclei. Our approach essentially depends less on complete chromosome reconstruction while depending more on comparison with a computational model. In this study, we compared the radial density of chromosomes from optically sectioned polytene chromosomes to three computational models; each of the three models having a different pattern of radial chromosome density due to specific Chr-NE attachments. In principle, this strategy could be extended to additional features of chromosomal organization, non-polytene chromosome, and other organisms. In general our approach involves comparing a Null model to a model that differs in a single organizational feature; in our case we compare a Null model (which contains no Chr-NE attachments) to models containing 15 or 48 Chr-NE attachments. Comparison with experiment is then used to support or refute either of these models; we compare the radial density of chromosome in each of our models with experimental z-stack images (see results). Despite the reliance on computation, sophisticated modeling is now possible on desktop computers, which may make this approach feasible for future studies.

3.5.6 Limitations

We acknowledge the limited resolution of our computational models. The bead radius, a key parameter in our computational models, is determined by the Kuhn length of the polytene chromosomes of *D. melanogaster*, which in turn limits the highest resolution of our

computational models to ~0.7Mb. Therefore, our model is not suited for determining the effects of Chr-NE attachments at higher resolutions. It is also reiterated that our model is a course grained representation of the polytene chromosomes and not every conceivable parameter of the nuclear interior was modeled or investigated for robustness; to maintain biological realism, only the parameter values known from experiment have been included in this study. Regardless, Chr-NE attachments affect chromosome-chromosome interactions at all resolutions in our computational models and we hypothesize that at least some of the effects will persist at higher resolutions as well. Due to quasi-equilibrium nature of the models considered, we cannot comment on possible time-dependence of these effects, including variations with cell age.

3.6 Brief Summary

It is now well recognized that the 3D organization of chromosomes plays an important role in key cellular processes such as DNA replication, repair, transcription, and epigenetic inheritance, i.e., inheritance that is not encoded by the DNA sequence. Interestingly, chromosomes are not freely floating within the nucleus but they form attachments between certain loci and the NE. The role of these Chr-NE attachments in the 3D genome organization is only beginning to emerge. Using a combined experimental and computational approach, we demonstrate that attachments of chromosomes to the nuclear periphery affect the 3D organization of the fruit fly genome in a number of ways. Specifically, several key features differ between chromosomes with Chr-NE attachments compared to those without Chr-NE attachments: they form more distinct territories, they intertwine less frequently, intra-chromosomal interactions are more common, and long-distance inter-chromosomal interactions are more specific. These results have biological significance: the Chr-NE attachments may facilitate specific chromosome-chromosome contacts, where actively transcribed genes co-localize and share sites of transcription.

3.7 Acknowledgements

This work was partially supported by the graduate program in Genomics Bioinformatics and Computational Biology (GBCB) of Virginia Tech to Nicholas Kinney, by NIH grant R01 GM099450 to Alexey V Onufriev, by the Fralin Life Science Institute, the USDA National Institute of Food and Agriculture Hatch project 223822 and NIH grant R21 AI099528 to Igor V

Sharakhov. This work was also supported in part by the NSF grant CNS-0960081 and the HokieSpeed supercomputer at Virginia Tech. The funders had no role in study design, data collection, and analysis, decision to publish, or preparation of the manuscript.

CHAPTER 4

A MODEL OF NUCLEAR ORGANIZATION DEMONSTRATES THE EFFECT OF NUCLEAR ENVELOPE – CHROMOSOME ATTACHMENTS IN THREE CELL TYPES

Nicholas Allen Kinney[1], Igor V. Sharakhov*[2] and Alexey V. Onufriev*[3,4]

[1] Genetics Bioinformatics and Computational Biology, Virginia Tech, Blacksburg, VA, 24061, USA

[2] Department of Entomology, Virginia Tech, Blacksburg, VA, 24061, USA

[3] Department of Physics, Virginia Tech, Blacksburg, VA, 24060, USA

[4] Department of Computer Science and Applications, Virginia Tech, Blacksburg, VA, 24061

* Igor Sharakhov Tel: (540)-231-7316 Fax: (540)-231-7126 Email: igor@vt.edu

* Alexey Onufriev Tel: (540) - 231 - 4237 Email: alexey@cs.vt.edu

4.1 ABSTRACT

The chromosomes in all eukaryotic organisms are confined within the boundary of the nuclear envelope. Two fundamental scaling laws forged from this confinement critically affect the 3D organization of the genome. The first fundamental scaling law relates the 2D genomic distance between genes, s, and the 3D spatial distance between genes, R. This relationship is often quantified by the parameter α in the expression $R(s) = s^{-\alpha}$. For completely straight chromosomes the 2D genomic and 3D spatial distances coincide which implies $\alpha = 1$. For highly folded chromosomes in the nucleus the value of α may range from ½ - ⅓. The second, often overlooked, scaling law relates volume of the nucleus and surface area of its envelope. This relationship matters because chromosome-nuclear envelope attachments may sequester particular loci in the nuclear periphery. This additional layer of confinement may be significant when the surface area of the envelope is small in relation to the volume of the nucleus. Possibly, the significance of peripheral confinement diminishes when the surface area of the envelope increases in relation to the volume of the nucleus. This hypothesis was tested by constructing computational models of three cells types in D. melanogaster. The parameters of each model were taken directly from experiment; no fitting parameters were introduced. We find that chromosome confinement partially dictates the effects of chromosome-nuclear envelope attachments. Chromosomes are more territories with chromosome-nuclear envelope attachments present; however, this effect dimishes when chromosomes are highly confined as in salivary gland nuclei. These results highlight the importance of volume and space.

4.2 INTRODUCTION

Experimental studies of polytene nuclei in fruit flies and nonpolytene nuclei in other eukaryotes have begun to characterize the three-dimensional (3D) genome organization. Several features appear universal to some degree. Chromosome territories, in which each chromosome occupies a distinct region of the nucleus, have been observed in numerous organisms and cell types, such as yeast (207), human (7), fruit fly (28-30), mouse (124), and Arabidopsis (81). These territories are mutually exclusive in the following sense: although different chromosomes may be in physical contact, they never interweave (28,178,208) (as do, for example, strands of DNA in the double-helix). Chromosome interactions, both within (intra) chromosomes and between (inter) chromosomes, have been observed microscopically (28,30) and inferred using cross-linking

techniques (109) such as the Hi-C method. Recently, Hi-C experiments in D. melanogaster (fruit fly) have revealed an abundance of intra-chromosomal interactions compared to inter-chromosomal interactions (89). Both the presence and location of these interactions appear to be important since they correlate with fragile sites where chromosomal breakpoints occur in evolution (22). Chromosomal entanglement – characterized by knots which hamper chromosome folding and unfolding – appears to occur infrequently based on direct observations in D. melanogaster (30) and both experimental and computational studies in human (11,43).

Since some features of 3D genome organization appear universal (previous paragraph), many computational and experimental studies have turned to the question of their origin. In other words, are there common principles which bring about these features? Indeed, several candidates have been identified. First, general polymer physics arguments suggest that the conformational state of chromatin depends on the chromosome to nucleus volume ratio (degree of chromosomal confinement) (43,75); thus, there may be different folding principles in difference lineages and cell types where this ratio differs greatly, e.g. human and yeast cells (43). In addition to the degree of confinement, non-specific (entropic) forces have been implicated in many recent studies of genome organization (81,82,123,138,139,171,209). Simulations have demonstrated that non-specific forces play a significant role in establishing and maintaining chromosome territories (81,139,171,210), and it has been suggested (139) that this entropic effect stems from long flexible polymers having access to more chain configurations if they remain separate in distinct domains, rather than tangling together. There are permanent (centromeric) and statistically high-frequency (but non-permanent) contacts between certain chromosomal loci and the nuclear envelope (28-30). The movement of chromatin loci depends on their nuclear localization; loci attached to or adjacent to the nuclear periphery are less mobile than other loci (83). Consequently, chromosome attachments to the nuclear envelope could constrain chromatin folding and prevent chromosome territories from freely diffusing in the entire nuclear volume (73). Chromosome loops of various sizes, in which long-range regions of chromosome interact, are now common in computational studies (123,138,155,211,212) and have also been linked with the formation of chromosome territories. Additional principles have been identified and it is likely that many principles are yet to be discovered.

The interplay among these principles and their effect on the hierarchy of 3D genome organization appears highly complex. For instance, multiple studies have suggested that non-specific (entropic) forces help establish and maintaining chromosome territories (81,82,140,171,209,210); however, this entropic effect has been shown to depend on the degree of confinement (172) and the presence of chromosome loops (139), which may also arise due to non-specific forces (139). Specifically, in one study increasing the degree of (spherical) confinement (123) mimicked the effect of increasing chromosome looping probability. The emerging importance of chromosome-nuclear envelope attachments have added to the complexity. A recent computational study demonstrated that chromosome-nuclear envelope attachments may contribute to the formation of chromosome territories in D. melanogaster (187); however, it has been suggest that the specific interactions between the chromosomes and the nuclear envelope arise due to non-specific forces in Arabidopsis (156). In contrast, a recent computational study suggested that activity based chromosome segregation (213), stemming from an inhomogeneous distribution of chromatin remodeling and transcriptional machinery, can more effectively position chromosome territories than non-specific forces alone. These studies support the notion of complex interplay and overlapping effect among the principles that govern 3D genome organization.

Due the recognized complexity and interdependence of the principles governing 3D genome organization (previous paragraph), an increasing number of computational studies now model the entire nucleus rather than single chromosomes (77,78,81,187). This strategy often takes into consideration the degree of chromosome confinement, boundary interactions (such as chromosome-nuclear envelope attachments), Rabl (polarized) configuration of chromosomes, and the excluded volume of bodies such as the nucleolus. This promise of this approach has already been shown in diverse organisms. A Recent model of the entire yeast nucleus (77,78) not only recapitulated key experimental features of 3D chromosome organization but also made predictions regarding chromosomal breakpoints and evolution. Recent studies of D. Melanogaster (187) used entire nucleus simulations to identify points of frequent attachment between chromosomes and the nuclear envelope and predict their effect on the 3D genome organization. Simulations of the entire human nucleus recently revealed a mechanism of chromosome territory segregation (213).

Parameter	Salivary	Midgut	Prothoracic
Chromosome width	3.1μ	2.1μ	1.5μ
X length	140μ	114μ	85μ
2L length	142μ	112μ	81μ
2R length	144μ	112μ	74μ
3L length	154μ	116μ	81μ
3R length	185μ	142μ	103μ
Nucleus aspect ratio	1.0	2.3±.4	1.0
Chrom vol / nucleus vol	.34±.03	.18±.02	.13±.02
Nucleolus vol	$200\mu^3$	NA	$343\mu^3$
+/- chromosome chirality	2:1	NA	2:1
Rabl configuration	present	absent	present
Chr-NE attachments	15	12	23

Table 4.1 – parameter sets for computational models of three cell types in D. melanogaster.

In a previous study we reported the effects of chromosome-nuclear envelope attachments on 3D genome organization in polytene nuclei of D. melanogaster salivary gland (187). Will the effects of chromosome-nuclear envelope attachments change in different cell types which have different sets of attachments or different chromosomal confinement, which critically affects the over-all 3D nuclear architecture (43)? The chromosome volume to nucleus volume ratio of salivary gland nuclei, midgut nuclei, and prothoracic nuclei of D. melanogaster ranges from .11-.34; do the effect of chromosome nuclear envelope attachments change in this biologically relevant range? In addition, salivary gland nuclei, midgut nuclei, and prothoracic nuclei each have a different set of chromosome-nuclear envelope attachments. Does the set of chromosome-nuclear envelope attachments specific to each cell type lead to distinct effects on the 3D genome organization? It is known that the 3D genome organization in each of these cell types differs: what is the relative importance of the factors that cause the difference, e.g. degree of chromosome confinement vs chromosome-nuclear envelope attachments? This study aims to address these questions and for the first time construct comprehensive models of three cell types.

Polytene chromsomes of D. melanogaster is a well-established model for studying organization and function of the eukaryotic genome (19,28-32,177). Each polytene chromosome contains approximately 1024 DNA replicas bundled together in parallel; thus the genome organization in a single nucleus becomes visible under a light microscope. In addition, the study of polytene chromosomes has significant potential for general understanding of 3D genome organization because recent experiments revealed similar structural and functional organization of non-polytene and polytene chromosomes in fruit fly (147,148). In this work we consider three computational models which correspond to three different cell types of D. melanogaster: salivary

gland model – contains all parameters of the polytene nucleus from D. melanogaster salivary gland nuclei; midgut model – contains all parameters of the polytene nucleus from D. melanogaster midgut nuclei; prothoracic model – contains all parameters of the polytene nucleus from D. melanogaster prothoracic nuclei. Parameters of each model are taken directly from experimental data (31,32), no fitting parameters are introduced. These computational models are in turn used to determine the effect of chromosome-nuclear envelope attachments on the 3D genome organization of each cell type.

4.3 MATERIALS AND METHODS

Modeling approach

The five largest chromosome arms of *D. melanogaster* salivary gland, midgut, and prothoracic cell nuclei are modeled as beads-on-string (41-43,82,85,132) and are represented as five random self-avoiding walks (SAWs) (15,84,154,155) (Figure 4.1). The sixth arm, chromosome 4, is not considered due to its negligible length. Experimental data (30-33) for the chromosomes and the nucleus become realistic model parameters and constraints imposed during the construction of SAWs (see table 1 for parameter sets). Complete details of model parameters are described in the supplementary material.

Model parameters

In this work we consider computational models of three different D. melanogaster cell types: salivary gland model – contains all parameters of the polytene nucleus from D. melanogaster salivary gland nuclei; midgut model – contains all parameters of the polytene nucleus from D. melanogaster midgut nuclei; prothoracic model – contains all parameters of the polytene nucleus from D. melanogaster prothoracic nuclei. Parameters of each model are summarized in table 1; for a complete discussion of parameters see supplementary material.

Nucleus heterogeneity

The degree of chromosome confinement and number of nuclear envelope attachments are unique to each cell type modeled in this study; however, these parameters also vary (to a lesser extent) between the individual nuclei of a single cell type, e.g. not every salivary gland cell nucleus has

the exact same degree of chromosome confinement and not every chromosome nuclear envelope attachment is realized in every cell. The variation in these parameters is known from experiment and this variation is incorporated into each modeled cell type. This is important because a recent model of cell mobility improved when cell heterogeneity was taken into consideration (214). See also table 1.

Robustness of conclusions to model details

The SAW approach used here to model polytene chromsomes was validated in several previous studies in similar contexts (43,53,75,77,78). Briefly, we construct our SAW's using an unweighted Rosenbluth algorithm (96). For short chains this approach is a good approximation of self-repelling chains which are true equilibrium states of polymers (97). Although we use a SAW approach, which is equilibrium by construction (to the extent that it approximates self-repelling chains), each of our three models contain non-equilibrium features which are introduced to better represent experiment; these include the Rabl configuration of chromosomes, right-handed chromosome chirality, and asymmetric chromocenter arrangement (see supplementary material). These parameters and the affect on the simulation outcomes are thoroughly discussed in a previous work (187). There we provided validation of our Null model by checking explicitly that key model conclusions are robust to these non-equilibrium SAW modifications. The key model conclusions in this study are the effects of chromosome-nuclear envelope attachments on the 3D organization of the genome (see results). We thoroughly check that these conclusions are also robust to the above non-equilibrium modifications introduced into our SAW approach. This is accomplished by considering versions of our three models (Null model, 15 attachment model, and 48 attachment model) constructed without Rabl chromosome configuration, right handed chromosome chirality, and asymmetric chromocenter

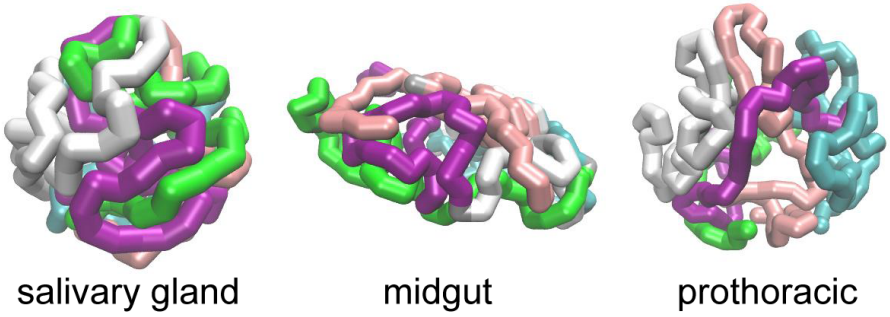


Figure 4.1 – representative simulated nuclei of the three models we consider

arrangement. The same conclusions are reached with these models (see supplementary material for details).

Simulations preformed

Two conformational ensembles of nuclei were simulated for each of the three computational models we consider (6 conformational ensembles total): Null ensemble – contains simulated nuclei constructed with all known parameters specific to the modeled cell type except for chromosome-nuclear envelope attachments; attachment ensemble – constructed with the set of chromosome-nuclear envelope attachments specific to the modeled cell type (all other parameters identical to the Null ensemble). Computationally, we compare the 3D organization of each modeled cell type. In addition, we investigate the effect of turning on the chromosome-nuclear envelope attachments in each modeled cell type by comparing the attachment ensemble to the Null ensemble (extensively discussed in results).

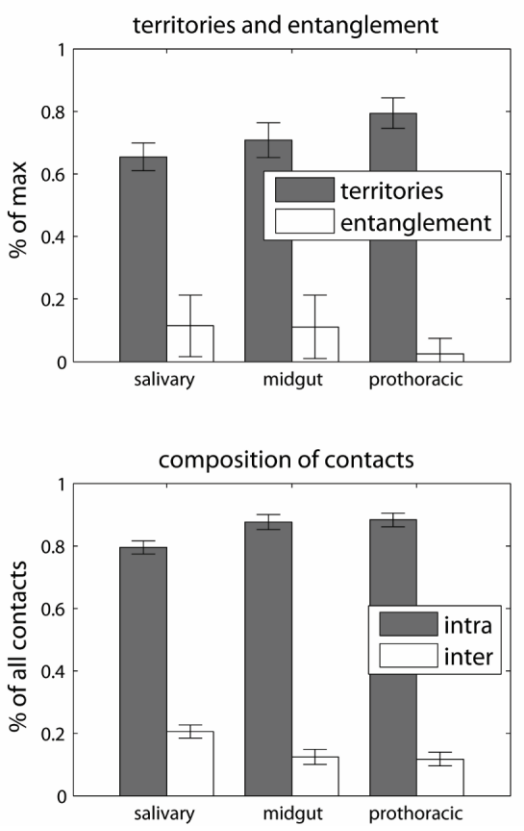


Figure 4.2 – chromosome territory index and intra-chromosome contact probability in the three modeled cell types.

Analysis of the simulations

To quantify bead-bead interactions (the highest resolution of our model), we calculate how often specific pairs of beads are in contact in an ensemble of computational nuclei; a contact is defined as two microns or less separation between the surfaces of two beads. There are 248 beads in single computational nucleus (including all of the 5 chromosome arms); a 248x248 contact map has entries for each possible pair of beads. The "i,j" entry of this contact map gives the probability that bead "i" and bead "j" in our computational nuclei form a contact. The average of all entries of a contact map represents the overall average bead-bead interaction probability

(lowest resolution of our model). We also compute contact probability at two lower resolutions: chromosome arm resolution – bins beads belonging to the same chromosome arm, whole chromosome resolution – bins beads belonging to the same chromosome (i.e. chromosome 2 right and left arms). These interactions types are normalized against the total number of interactions in each model, e.g. the number of intra-arm interaction in our Null model out of all interactions in our Null model. Metrics used to quantify chromosome territories and chromosome intertwining have been described previously (187).

Chromosome territories

We quantify chromosome territories using a previously described metric that quantifies how each chromosome excludes other chromosomes from the volume it occupies in the 3D space. We begin by calculating the convex hull for a single chromosome. This is the minimum volume that includes all the chromosome's points (bead centers) inside a convex polyhedron. In general, each convex hull contains its own chromosome and may also encompass some beads belonging to other chromosomes. A fully "territorial" chromosome is one whose convex hull does not contain points from any other chromosomes, while a less "territorial" chromosome is one whose convex hull contains some points from other chromosomes (**Fig. 2**). We define the chromosome territory index as the fraction of points inside a convex hull that belongs to the chromosome used for its construction. A detailed description of chromosome territories is given in our recent work (187).

4.4 RESULTS

4.4.1 Model predicts pattern of chromosome territories and whole chromosome interactions in three cell types.

We use recently developed metrics

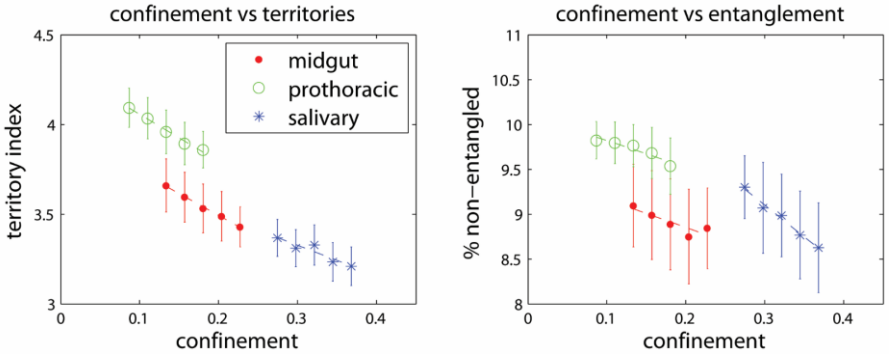


Figure 4.3 – The degree of chromosomal confinement inversely correlates with the chromosome territory index. An inverse correlation also realtes to the % of non-entangled chromosomes in the three modeled cell types.

(187) to quantify chromosome territories and chromosome intertwining in our computational models of three cell types. The results are summarized in figure 4.4 (see also supplementary material) . Overall chromosome territories are present in all three computational ensembles despite significant differences in chromosome confinement (which ranges from .08-.37), chromosome nuclear envelope attachments, and cell shape (spherical vs oblate spheroid). This suggests that chromosome territories may be a robust feature of polytene chromosome organization in D. melanogaster. This result is also consistent with a growing body of experimental and computation evidence alluding to a universal presence of chromosome territories. Although chromosome territories were present in each of our computational ensembles; the territory index differed for each cell type we modeled (187). Chromosome in the simulated prothoracic nuclei had the highest territory index followed by midgut nuclei and salivary gland cell nuclei respectively. Interestingly, the territory index for chromosomes in the midgut cell nuclei, was higher than salivary gland cell nuclei yet many of the midgut nuclei intertwined more frequently (discussed below). This

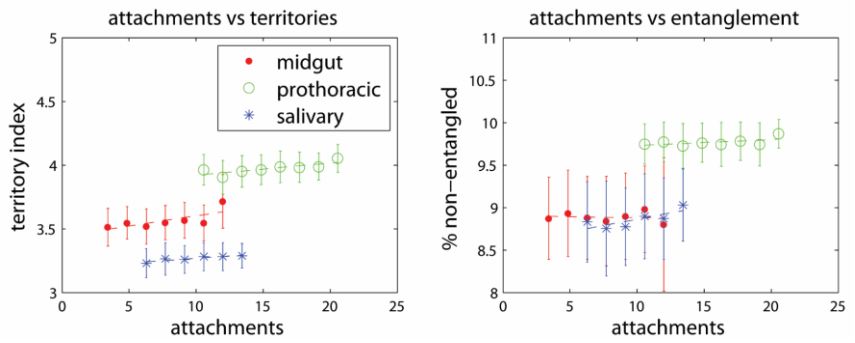


Figure 4.4 – the number of chromosome-nuclear envelope attachments positively correlates with the chromosome territory index. A positive correlation also relates to the % of non-entangled chromosomes in the three modeled cell types.

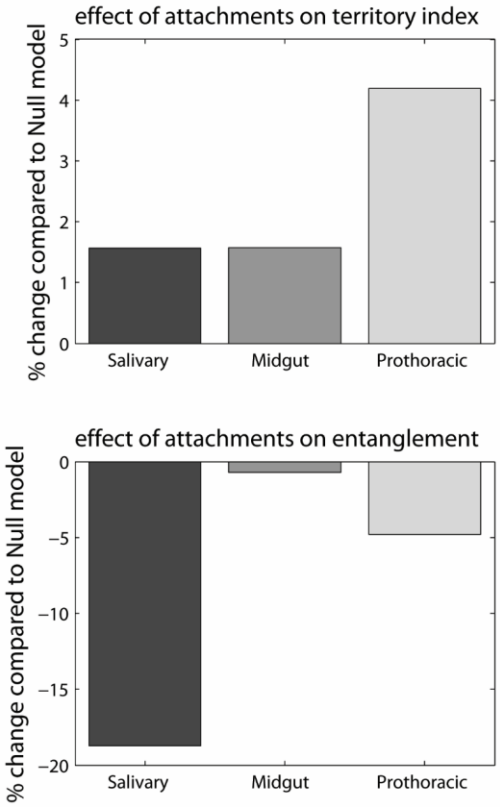


Figure 4.5 – effects of "turning on" chromosome–nuclear envelope attachments in three cell tyes. Higher resolutions reveal that intra-chromosome and intra-arm contacts are more

recapitulates a key difference in the genome organization of these cells from experiment (32).

We hypothesized that the robust presence of chromosome territories in each computational ensemble would also indicate an abundance of intra-arm chromosome interactions compared to inter-arm chromosome interactions. To test this hypothesis we analyzed bead-bead contact probability in our models at two progressively increasing resolutions (see methods): whole chromosome (60 Mb) and chromosome arm (30 Mb). In all three modeled cell types intra-chromosome interactions outnumber inter-chromosome interactions (figure 4.2 bottom); similar results were found at chromosome arm resolution. Noted is the consistency with the territory index in each modeled cell type; for example, simulated prothoracic cell nuclei have the highest territory index and the highest probability of intra-chromosome interactions. We conclude that the presence of chromosome territories may be indicative of the relative number of inter and intra chromosomal interactions in the polytene chromosomes of D. melanogaster. The chromosome territory index did not hold the same consistency with chromosomal intertwining (see supplementary material), e.g. simulated midgut cells have a higher territory index than simulated salivary gland cells but intertwine slightly more frequently.

4.4.2 Chromosome territories and intertwining correlate with chromosome confinement and number of chromosome-nuclear envelope attachments three cell types.

Given the predictive power of chromosome territories in our computational ensembles (see previous paragraph) and the ubiquitous presence of chromosome territories in recent computational studies, we next tested which cellular parameters accommodate chromosome territories in our computational ensembles. The results in figure 4.3 (left) show that the territory index correlates inversely with the degree of chromosomal confinement. This makes sense if we consider the following limiting case: in free space (zero confinement) chromosome would be completely separated and thus highly territorial. On the contrary, a weaker positive correlation exists between the number of chromosome nuclear envelope attachments and the territory index (figure 4.4 left), suggesting that chromosomal confinement may be a stronger indicator of chromosome territories than chromosome-nuclear envelope attachments. Interestingly, we also find an inverse correlation between the degree of chromosome confinement and chromosome intertwining and again a positive correlation between chromosome-nuclear envelope attachments

and chromosome intertwining. These results may suggest a simple biological result: a low degree of chromosomal confinement and numerous chromosome-nuclear envelope attachment will result in territorial chromosomes that do not intertwine. Territorial chromosomes will in turn have numerous intra-chromosome interactions compared to inter-chromosome interactions.

4.4.3 Complex interplay between chromosome-nuclear envelope attachments and chromosome confinement demonstrated in three cell types.

A recent computational study demonstrated that chromosomes with numerous nuclear envelope attachments are more territorial than chromosomes with few nuclear envelope attachments in Drosophila melanogaster salivary gland nuclei. Since we demonstrate (above) that chromosome confinement also effects chromosome territories, which in turn effect inter/intra chromosomal attachments (see results above), a complex interplay must exists between chromosome territories, chromosome-nuclear envelope attachments, and their effects. Computationally, we can easily investigate the effect of turning on the chromosome-nuclear envelope attachments in our 3 modeled cell types (salivary, midgut, and prothoracic) by comparing simulated nuclei with attachments to the corresponding Null model nuclei (0 attachments). The results are summarized in figure 4.5. The qualitative effects of chromosome-nuclear envelope attachments are similar in each cell type: the simulated nuclei which possess chromosome-nuclear envelope attachment are more territorial and intertwine less frequently than the corresponding nuclei which lack chromosome-nuclear envelope attachments. Our conclusion is that the "turning on" of chromosome nuclear envelope attachments reinforces chromosome territories (increase the chromosome territory index) in the modeled cell types.

Since chromosome territories are indicative of the number of intra-chromosome contacts compared to inter-chromosome contacts (see above results) we made the following prediction: the "turning on" of chromosome nuclear envelope attachments should in turn increase the composition of intra-chromosome contacts compared to inter-chromosome contacts. We tested this prediction by investigating how the composition of intra and inter chromosome contacts changes in each model cell type due to the turning on of chromosome-nuclear envelope attachments; our results (bottom figure 4.5) support our prediction. The composition of chromosome contacts indeed changes upon placement of chromosome-nuclear envelope contacts

in each model; specifically, a larger percent of all contacts are intra-chromosomal in the presence of chromosome-nuclear envelope attachments. This result is consistent with the predicted effect of chromosome nuclear envelope attachments (that they increase the territory index).

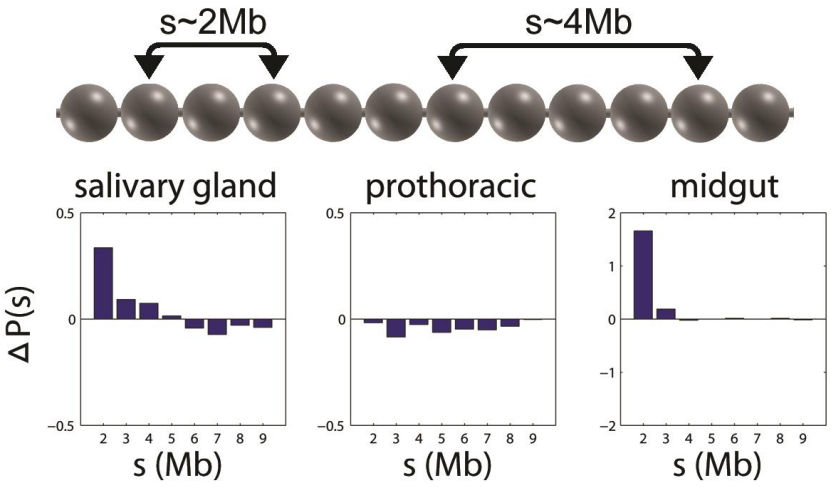


Figure 4.6 - ΔP(s) for intrachromosomal contacts in three modeled cell types. ΔP(s) is non-zero at all genomic distances indicating that chromosome nuclear envelope attachments have long-range effects on genome organization.

4.4.4 Chromosome-nuclear envelope attachments alter the pattern of chromosome contacts.

We investigated how chromosome-nuclear envelope attachments affect the pattern of chromosome contacts in each model; this resolution corresponds to bead-bead interactions within each model. Each bead in the computational models represent approximately 1Mb of DNA. These high resolution effects were made quantitative by first computing the probability, P(s), of contact as a function of bead separation, s, within each modeled cell type. Next, we computed how these probabilities change in the presence of chromosome nuclear envelope contacts by comparing attachment ensembles (which possess chromosome-nuclear envelope attachments) to Null ensembles (which lack chromosome-nuclear envelope attachments) in the three modeled cell types. Thus, high resolution changes in chromosome interactions are assessed by computing ΔP(s) in the presence and absence of chromosome-nuclear envelope attachments. Remarkably, ΔP(s) is non-zero at all genomic distances indicating that chromosome nuclear envelope attachments have long-range effects on genome organization. In salivary gland models, ΔP(s) is positive when s is small indicating that chromosome-nuclear envelope attachments increase the probability of interaction between nearby (<5Mb) chromosomal loci (see figure). On the other hand, ΔP(s) is negative for large values of s (>5Mb) indicating a decreased probability of

chromosome interactions. This pattern differs in models of prothoracic and midgut nuclei. In models of the midgut nuclei ΔP(s) is positive for small values of s (< 3Mb) and vanishes for large values of s (> 3Mb). Thus the effect of chromosome-nuclear envelope attachments on P(s) is unique in different cell types. Interestingly, ΔP(s) is negative for all values of *s* in models of the prothoracic nuclei; this has two interpretations. Possibly, P(s) decreases for all values of s in prothoracic nuclei when chromosome nuclear envelope attachments are "turned on". On the other hand, it's possible that the probability of chromosome interactions increases for values of *s* less than 1Mb; however, these interactions are beyond the resolution of our model in which each bead represents ~1Mb.

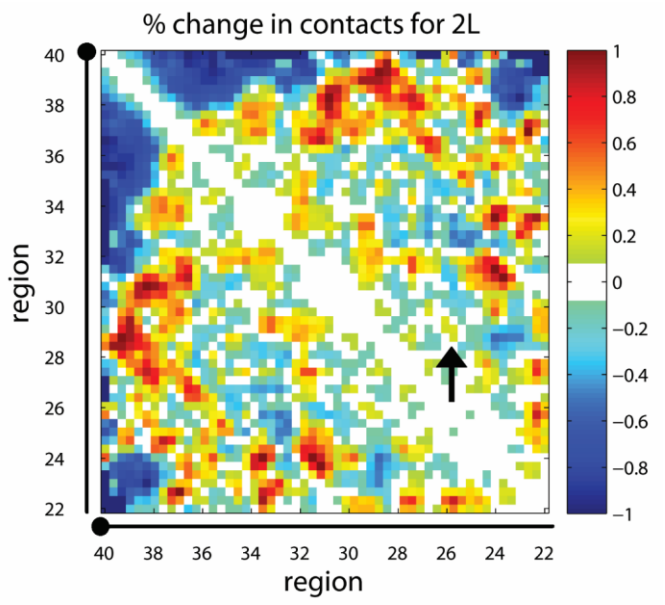


Figure 4.7 - A contour map for chromosome 2L in our modeled prothoracic nuclei. The map shows the % change in co-localization due to the turning on of chromosome-nuclear envelope attachments.

4.4.5 Chromosome-nuclear envelope attachments may not be necessary to co-localize regions 25C and 27C on the cytogenetic map.

In a previous study, regions 25C and 27C on the cytogenetic map co-localized in 5 out of 11 (45%) experimental prothoracic nuclei. We assess the role (if any) of chromosome-nuclear envelope attachments in establishing the co-localization of these regions by considering bead-bead contact frequencies at the highest resolution (.5Mb) of our prothoracic nuclei models. We find that the beads corresponding to regions 25C and 27C co-localize in 38% of our prothoracic nucleus models (with chromosome nuclear envelope attachment turned on) in reasonable agreement with experiment. However, in the corresponding Null model (with chromosome nuclear envelope attachments turned off) the same beads co-localize in 37% of our model nuclei. The change in co-localization frequency was not statistically significant (p value < 10^{-4}). A

contour map for chromosome 2L in our modeled prothoracic nuclei (figure 4.6) shows the % change in co-localization due to the turning on of chromosome-nuclear envelope attachments, the plotted region corresponds to ~2500 bead-bead interactions. The black arrow indicates the interaction between regions 25C and 27C. Although chromosome nuclear envelope attachments do not appear to influence the interaction of these two regions, it was emphasized in a previous study that their interaction may be frequent simply due to their proximity on chromosome 2L (31,32). We suggest several conclusions. First, chromosome-nuclear envelope attachments may not influence the interaction of regions 25C and 27C if they indeed co-localize as speculated in experiment. Thus, some other mechanism may be necessary to establish their frequent interaction. However, our computational models clearly indicate that some regions do interact more frequently in the presence of nuclear envelope attachment than in their absence (black arrow figure). It's likely that not all of these attachments in prothoracic nuclei have been discovered; possibly, a complete set of chromosome-nuclear envelope attachments would influence the interaction of regions 25C and 27C in our computational models. In other words, our computational models could be an indications that some chromosome-nuclear envelope attachments have yet to be discovered.

4.5 DISCUSSION

4.5.1 Modeling of multiple cell types can be used to address new questions regarding 3D genome organization. Computational studies of 3D genome organization now routinely simulate the entire cell nucleus (77,78,81,187,213). The motivation for this approach often stems from a need to model the multiple factors affecting 3D genome organization, such as entropic forces, spherical chromosome confinement, boundary interactions, and chromosomal looping. The promise of this approach has been show in recent computational studies which recapitulate key features of genome organization form experiment. Indeed, a computational model of chromatin as a fractal globule improved when interactions between the chromatin fiber and the nuclear periphery were taken into account (215). Our study builds on this approach by simulating the entire cell nucleus of three different cell types from D. melanogaster. This advance can address new question regarding 3D genome organization not possible in studies of a single cell type. What is the unique pattern of chromosomal interactions in different cell types of a single organism? Is there interplay among the factors that affect 3D genome organization? What

features of genome organization are common among different cell types and which differ? Our approach demonstrates how to address these questions and is applied to the salivary gland, midgut, and prothoracic cell nuclei of fruit fly; however, we speculate that this approach may be a valuable if applied to multiple cell types of other organisms as well.

4.5.2 The role of chromosome nuclear envelope attachments differs in three cell types. Multiple computational and experimental studies have alluded to a complex interplay among the factors that govern 3D genome organization. For example, it is widely accepted that entropic forces play a significant role in the formation of chromosome territories; however, this entropic effect has been shown to depend on the degree of chromosome confinement (123) suggesting an interplay between the degree of chromosomal confinement and entropic forces which govern 3D genome organization. A recent computational study of D. melanogaster salivary gland chromosomes added to the complexity by suggesting that the effects of chromosome-nuclear envelope attachments (187), which affect chromosome territories and chromosome folding (123), may also depend on the degree of chromosomal confinement (123). Our study confirms the prediction of this previous study in biologically relevant context by simulating the salivary gland, midgut, and prothoracic nuclei nucleus of D. melanogaster using only known parameters from experiment. Thus, our study adds to a growing body of evidence that chromosomal confinement is critically for determining the 3D organization of the genome. Studies have already shown that chromosome looping is potentially affected by the degree of chromosome confinement (123); our study shows chromosome nuclear envelope attachments may also be affected by the degree of that chromosome confinement.

4.5.3 Chromosome territories are a robust feature of simulated nuclei but may be affected by multiple factors. Chromosome territories have been observed in numerous organisms and cell types, such as yeast (207), human (7), fruit fly (28-30), mouse (124), and Arabidopsis (81); as such, chromosome territories appear to be a ubiquitous feature of 3D genome organization. In light of these experimental results, it is not surprising that chromosome territories are a robust feature of our simulations. We observe chromosome territories in all three simulated cell types regardless of cell shape (spherical vs ellipsoid), presence of chromosome-nuclear envelope attachments, and positioning of chromosome-nuclear envelope attachments. In particular, our

simulations also suggest that chromosome territories are robust to substantial changes in the chromosome confinement; in our simulated nuclei this parameter ranges from .11-.30. Although the chromosomes in all three of our simulated cell types are clearly territorial they differ in detail; specifically, several factors reinforced the chromosome territories present in our simulations. Both increasing the number of chromosome-nuclear envelope attachments and decreasing the degree of chromosomal confinement corresponded to more territorial chromosome in our simulation. This result appears consistent with a number of computational and experimental results. For instance, live imaging of yeast (74) and fruit fly (73) chromosome territories have revealed that movement of chromatin loci attached to or adjacent to the nuclear periphery are less mobile than other loci (83), supporting the idea that chromosome-nuclear envelope attachments reinforce chromosome territories.

4.5.4 Low resolution pattern of chromosome contacts depends on multiple factors.

Experimental Hi-C data (89) has recently mapped the global chromosome-chromosome interactions of D. melanogaster (non-polytene) nuclei. The pattern of global chromosome-chromosome interactions was characterized by the following trends: an abundance of intra-arm contacts and an absence of inter-arm contacts (89). In addition, a predominance of intrachromosome recombination has been demonstrated by experimentally irradiating embryos which supports the notion of abundant intra-chromosome interactions since recombination implies physical proximity (216). We observe the same trends in each of our three simulated cell types; this consistency is expected because several studies have demonstrated a similarity of functional organization between polytene and non-polytene chromsomes from *D Melanogaster* (31,78-80). Despite the abundance of intrachromosomal interactions in all three of our simulated cell types, several factors affect the degree to which they outnumber interchromosome interactions. Intuitively, a decrease in chromosomal confinement corresponds to a decrease in interchromosome contacts. The effect of chromosome-nuclear envelope attachments is more complex; the “turning on” of attachments in the highly confined salivary gland nuclei modestly increases the proportion of intra vs inter chromosome interactions, but does so to a much lesser extent in nuclei with a lesser degree of chromosome confinement. The combined effect of these factors may be even more complex in the presence of chromosome loops which were not a focus of this study.

4.5.5 Chromosome-nuclear envelope attachments increase the specificity of inter-chromosome and inter-arm interactions. A previous study of D. melanogaster salivary gland nuclei suggested that chromosome-nuclear envelope attachments may increase the specificity of chromosomal interactions in the following sense: predictable chromosome regions interact more frequently in the presence of chromosome-nuclear envelope attachment than in their absence. Our results are consistent with this previous result and establish biological significance by focusing on regions 25C and 27C which are known to colocalize in experimental nuclei; specifically, we demonstrate that chromosome nuclear envelope attachments increase the probability of colocalization between these two regions and improve agreement with experiment. This results is important because most computational models to date have not included the full complement of known chromosome-nuclear envelope attachments. Although a recent model of the yeast genome demonstrated a correspondence between inter-chromosomal contacts and chromosomal breakpoints (22), this model included a limited set of chromosome-nuclear envelope attachments which seems to reiterate their role in establishing specific chromosome-chromosome interactions. The significance of chromosome co-localization has been recognized in a number of additional contexts; for example, it is known that actively transcribed genes often co-localize in 3D space (203). These transcription "hotspots" may arise due to a sharing of the cell's transcriptional machienry. Although the highest resolution in our models is .5Mb, chromosome-nuclear envelope attachments clearly affect how often specific bead pairs colocalize. We speculate that this result persists at higher resolutions (not a focus of this study), in which case chromosome-nuclear envelope attachments may contribute to mechanisms that colocalize genes in 3D.

4.5.6 Chromosome resolution higher than .5Mb and chromosome dynamic chromosome loops are not explored in this study. We acknowledge several limitations of our computational models. In particular, the bead radius, a key parameter in our computational models, is determined by the Kuhn length of the polytene chromosomes of D. melanogaster, which in turn limits the highest resolution of our computational models to ~.7Mb. Therefore, our model is not suited for determining the effects of chromosome-nuclear envelope attachments at higher resolutions. Regardless, chromosome-nuclear envelope attachments affect chromosome-chromosome interactions at all resolutions in our computational models and we hypothesize an

affect at higher resolutions as well. An increasing body of evidence suggests that chromosome loops may play a role in the functioning and organization of the genome. In particular, computational models of chromosome loops have established possible their possible role in the formation and maintaining of chromosome territories; these studies typically model chromosome loops by assigning dimerization probabilities to the monomers with a polymer chain. Chromosome loops were not a focus of this study; the chromosome loops present in our model nuclei form due to chance alone without influence of assigned dimerization probability. This is justified in the context of experimental studies of polytene chromsomes which do not suggest chromosome looping probabilities beyond chance alone; we reiterate that all of the known experimental parameters of polytene chromosome from each cell type (and no others) are included in our simulated nuclei.

CHAPTER 5

CHROMOSOME NUCLEAR ENVELOPE ATTACHMENTS AFFECT CHROMOSOME TERRITORIES AND ENTANGLEMENT IN SIMULATION

Nicholas Allen Kinney[1], Igor V. Sharakhov*[2] and Alexey V. Onufriev*[3,4]

[1] Genetics Bioinformatics and Computational Biology, Virginia Tech, Blacksburg, VA, 24061

[2] Department of Entomology, Virginia Tech, Blacksburg, VA, 24061

[3] Department of Physics, Virginia Tech, Blacksburg, VA, 24060

[4] Department of Computer Science and Applications, Virginia Tech, Blacksburg, VA, 24061

* Igor Sharakhov Tel: (540)-231-7316 Fax: (540)-231-7126 Email: igor@vt.edu

* Alexey Onufriev Tel: (540) - 231 - 4237 Email: alexey@cs.vt.edu

5.1 ABSTRACT

It is well recognized that the chromosomes of eukaryotes fold into non-random configurations within the nucleus – these configurations have biologically significant properties, but may deteriorate with time. We use coarse-grained molecular dynamics simulations to study the effects of chromosome-nuclear envelope (Chr-NE) interactions on the dynamics of chromosomes within a model of *Drosophila melanogaster* (fruit fly) regular interphase chromosomes. The computational model simulates the dynamics of chromosomes bounded by the nuclear envelope (NE) on time scales comparable to life time of the cell. Initially, the chromosomes in the model are prearranged in fractal-like configurations with physical parameters such as nucleus size and chromosome persistence length taken directly from experiment. Time-evolution of several key observables is quantified during each simulation: chromosome territories, chromosome entanglement, intra-chromosomal interaction probability, and presence of the Rabl (polarized) chromosome arrangement. We compare the outcome of simulations with and without Chr-NE interactions. We find that Chr-NE interactions help maintain chromosome territories and limit chromosome entanglement on biologically relevant time scales. Results suggest that chromosome-nuclear envelope attachments may prevent a critical amount of entanglement that would otherwise interfere with proper cell division. At the same time, Chr-NE interactions have little effect on intra-chromosome interaction probability. This result is rationalized by a simple dimensionality argument. All results are robust to the simulated activity of topoisomerase which may be present in the interphase cell nucleus. Based on these results we conclude that the presence of Chr-NE interactions reinforces some properties of fractal-like chromosome configurations.

5.2 INTRODUCTION

The three-dimensional (3D) organization of the genome (chromatin) plays an important role in key cellular processes such as DNA replication, repair, transcription (217), and epigenetic inheritance, *i.e.*, inheritance that is not encoded by the DNA sequence (218). Links between chromatin architecture and diseases such as cancer are being established (219). However, unlike most proteins that adopt the same unique 3D shapes in all cells, the conformational states of the chromatin fiber are not nearly as compact or ordered and are stochastic to some degree. Remarkably, several features of chromatin folding appear to be universal to some degree.

Chromosome territories, in which each chromosome occupies a distinct region of the nucleus, have been observed in numerous organisms and cell types, such as yeast (207), human (7), fruit fly (28-30), mouse (124), and Arabidopsis (81). Chromosome interactions, both within (intra) chromosomes and between (inter) chromosomes, have been observed microscopically (28,30) and inferred using cross-linking techniques (109) such as the Hi-C method; intra-chromosomal interactions in particular are often characterized by their power law decay which may differ in different organisms (89,109). Chromosomal entanglement, characterized by knots which hamper chromosome folding and unfolding, appears to occur infrequently based on direct observations in D. melanogaster (30) and both experimental and computational studies in human (11,43). Chromosomes in yeast (207), fruit fly (28-30), and Arabidopsis (81) possess a distinctly polarized (Rabl) chromosome arrangement characterized by a separation of chromosome centromeres and telomeres; the arrangement is thought to be a remnant of anaphase.

Computational approaches are now routinely used to predict genome wide folding based on the collection of features revealed by a given experiment. For example, close integration of computation and experiment was recently used to suggest that the human genome folds into a shape called the fractal globule (FG) (11,43). This shape correctly predicts three key features of experiment: a presence of chromosome territories, a lack of chromosome knots, and the power law $P(s) = s^{-1}$ relating intra-chromosomal contact probability (P) and genomic separation (s) (45). Indeed the intra-chromosomal contact probability revealed by Hi-C experiment in human suggests a power law with -1 exponent. Interestingly, the FG is a non-equilibrium state which may imply that the true chromosome configurations suggested by experimental Hi-C maps are also out of equilibrium. Nevertheless, computational approaches have used both equilibrium and non-equilibrium approaches to generate fractal-like chromosome configurations. Equilibrium approaches have used pseudo-Boltzmann distributions to simulate the non-equilibrium properties of fractal configurations (220). In a recent study, the chromosome configurations generated by this thermodynamic based approach reiterated many properties of the fractal globule. It was demonstrated that ideal chromosome configurations are largely free of knots and tend to form fibrils of fibrils reminiscent of the crumples that recursively form the FG (220). On the other hand, true (fully) equilibrium based folding models have been rejected due to an absence of

chromosome territories, high degree of knotting, and power law $P(s) = s^{-3/2}$ which deviates from experiment.

In D. melanogaster (fruit fly) both fractal and hierarchical (modular) shapes have been proposed to explain the folding of interphase chromosome (89). In the hierarchical model, multiple genes cluster into domains bounded by epigenetic markers positioned along the chromosome fiber. These domains in turn form their own clusters: inactive domains tend to aggregate while the less compact active domains tend to facilitate inter-chromosomal interactions (89). In contrast to the FG, hierarchical chromosome folding is more compartmentalized, reflecting the known epigenetic profiles of the chromosomes; however, both recapitulate the overall pattern of chromosome interactions revealed by Hi-C – a cross linking technique used to infer chromosome interactions. A recent study has introduced two additional models: the "tension globule" model and the chromatin extrusion model (221). The tension globule is formed during polymer condensation by inter-monomer attraction forces (221). The chromatin extrusion model proposes that CCCTC-binding factor and cohesin partitions unknotted loops of chromatin in a manner consistent with experiment (221). Each of these models possess characteristics of the FG while being distinct from it (221). Thus, chromosome folding predicted in most recent studies (220,221) has been likened to the theoretical fractal globule (11,43) with the qualification that chromosome folding may not be strictly fractal (221).

Although most studies now agree that chromosomes in higher eukaryotes fold into a non-equilibrium state which inevitably transitions to equilibrium, there is no consensus at present on the time scales necessary to reach the equilibrium. In human it has been suggested that the FG is a long-lived state and transition to equilibrium is simply longer than the lifetime of the cell. However, other recent studies argue that fractal-like configurations exist along a spectrum connecting open chromatin at one extreme to compact chromatin at the other (120). In the strings and binders switch model (SBS) (120), this spectrum of configurations is explored by altering the affinity and concentration of binder molecules that mimic the cell's DNA binding machinery. The SBS model suggests that the fractal-like configurations occurring during the transition from open to compact chromatin states may be fleeting in the presence of topoisomerases (120). Thus,

the duration and stability of fractal-like configurations as seen in experiment remain largely unknown.

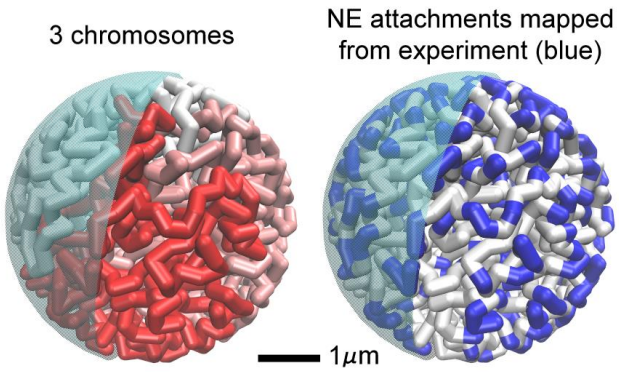


Figure 5.1 – computation "beads-on-string" model of D. melanogaster interphase chromosomes. On the left, the beads are colored by the three chromosomes in wild type D. melanogaster. On the right, the coloring is by NE-attachments mapped from experiment, shown in blue.

Here, we investigate the duration and stability of fractal-like configurations for the 3D organization of the fruit fly genome. A recent experiment identified ~500 D. melanogaster genes in close proximity to the nuclear envelope *in vivo (12)*; the genes were identified using a DamID approach – a method based on detecting DNA methylation by a chimeric protein consisting of a chromatin protein fused with methyltransferase (153). These genes correlate with sites of chromosome-nuclear envelope attachment in polytene chromosomes: a correspondence which has important implications. Since nuclear envelope attachments in polytene chromosomes are known to affect the folding of polytene chromosomes (187,222), it is speculated that they may play a similar role in non-polytene chromosomes. However, little is known about the *specifics* of this hypothesis. For instance, could the presence of chromosome nuclear envelope attachments prolong fractal-like configurations, which given sufficient time will transition to equilibrium? Are chromosome nuclear envelope attachments necessary to maintain the Rabl configuration of chromosomes, which is estimated to last over two hours in the interphase nucleus of D. melanogaster (76)? Chromosome-nuclear envelope attachments in the polytene nucleus are known to reinforce chromosome territories and mitigate chromosome entanglement (187,222); is this also the case in regular non-polytene interphase chromosomes? Our study aims to answer these questions using a computational model of the drosophila interphase nucleus.

Our study is based on interphase nucleus of D. melanogaster cycle 12 or 13 embryos which is well-established for studying the organization and function of the eukaryotic genome (19,28-32,177). As a model organism, D. melanogaster has several critical advantages over others. First,

the chromosome interactions with the nuclear envelope have been comprehensively mapped in Dam-ID experiments (12). So far the full complement of these interactions have been mapped for a limited number of organisms. In fruit fly, the complex pattern of these Chr-NE interactions includes over 500 genes identified by their capacity to bind with B-type lamin. It was demonstrated that these genes form approximately 52 Lam target clusters (223). Experimental mapping of these interactions in humans has revealed their clustering into more than 1,300 well defined lamin associated domains (180). Just like fruit fly, clustered regions of chromosome nuclear envelope interaction are generally transcriptionally inactive. Having access to the experimentally determined Chr-NE interactions enables the comprehensive mapping of each interaction onto a computational model of chromosomes in interphase. Second, since the dynamics of specific chromosomal loci are known from experiment (73) we are able to directly validate the time scale of chromosome dynamics seen in simulations. Third, the initial configuration of chromosomes in the model can be designed to match the fractal-like configurations suggested by experiment (89). Fourth, physical parameters such as nucleus diameter and genome size can be incorporated directly from experimental measurements. Using these data we consider two models of the D. melanogaster interphase nucleus. A wild type model possesses all known parameters of D. melanogaster nucleus including the experimentally identified chromosome-nuclear envelope attachments; a control (Null) model is identical to the wild type model but lacks specific sites of attachment between chromosomes and the nuclear envelope. The effects of chromosome-nuclear envelope attachments are studied by comparing the dynamics of the wild type and control models (see results).

5.3 MATERIALS AND METHODS

5.3.1 Modeling approach

The five largest chromosome arms of *D. melanogaster* are modeled as beads-on-

Parameter	Value
X chromosome	22,422,827bp = 321 beads
2L arm	23,011,544bp = 329 beads
2R arm	21,146,708bp = 302 beads
3L arm	24,543,557bp = 351 beads
3R arm	27,905,053bp = 399 beads
Bead mass	77M daltons
Bead radius (R_{bead})	.2 μm
Nucleus radius (R_{nuc})	3.5 μm
% confinement	$\left(1702 \cdot {R_{bead}}^3\right) / {R_{nuc}}^3 = .32$

Table 5.1 – essential model parameters.

string (41-43,82,85,132). The sixth arm, chromosome 4, is not considered due to its negligible length. Each beads-on-string chromosome consists of particles interacting as soft spheres bonded by a harmonic spring potentials; a detailed description of potentials is provided below. We additionally consider the presence of a nucleolus in simulations by excluding the volume of a spherical region 1μ in diameter positioned near the X-chromosome centromere. Experimental data (12,73,89) for the chromosomes and the nucleus become realistic model parameters and constraints imposed during simulations (see table 1 for parameters). Fractal-like initial configurations of chromosomes are assembled on a simple cubic lattice and transferred to free space during warmup integration. Complete details of model assembly and warmup protocol are provided in supplementary material. Simulations are performed in Espresso (224). The "beads-on-string" model of D. melanogaster interphase chromosomes is depicted in figure 5.1.

5.3.2 Bead size and chromosome persistence length

It is well known from polymer physics that a chain with persistence length, l_p, may be modeled as a self-avoiding walk (SAW) segmented by the Kuhn length, $l_k = 2l_p$. Although the Kuhn length of interphase chromatin has not been directly measured, several previous studies (11) estimate the Kuhn length based on the following argument which we briefly reiterate. The persistence length of double stranded DNA is known to be 150bp and the linker DNA between histones is on average 50bp (225-228). Since the histone bound DNA consists of 150bp and does not contribute to the flexibility the chromatin, the Kuhn length of the chromatin corresponds to about 6 histone/linker segments amounting to 1,200bp. The estimated persistence length of 600bp is a lower bound for the following reason: protein bound to DNA and possible higher order structure of chromatin will increase persistence length. Indeed, persistence length estimates of the yeast 30nm fiber range as high as 40,000bp (229). The aggregate of multiple other experiments (230,231) suggest the persistence length of chromatin ranges from 3,000bp-20,000bp. The details of chromatin packaging specific to D. Melanogaster are limited and complicated by evidence of chromatin remodeling which can affect the chromatin flexibility (232); therefore, we conservatively take each bead in our model to represent 70,000bp (which is more than twice the persistence length measurements of most experiments) and model the chromatin as freely jointed beads-on-string. Next, we calculate the mass and volume of each bead, which is important for establishing the simulation time step (discussed below). The

70,000bp represented by each bead is associated with approximately 350 nucleosomes. Since the mass of each nucleosome is 100,000Da and each base pair 600Da, we assign each bead in our model a mass, m_{bead}, of 77MDa. To establish the diameter of each bead, D_{bead}, we first approximate each nucleosome as a cylinder with radius and length of 5nm (109); its volume, V_n, is therefore $\pi \cdot 5^3$ cubic nanometers. The diameter of each bead depends on the volume and arrangement of the 350 nucleosomes it represents. Under the assumption that the 350 nucleosomes represented by each bead are structured as a 30nm fiber the diameter of each bead becomes .2 microns; assuming a 10nm fiber implies a bead diameter of 1.8 microns (see detailed calculation in supplementary material). We set the diameter of each bead to .4μm as a compromise between the debated structures of chromatin in interphase.

5.3.3 Volume of nucleus and chromatin

The dimensions of D. melanogaster interphase nucleus are known experimentally (12); however, it was noted previously that D. melanogaster chromosomes are frequently confined inside rapidly growing nuclei (75,233). We take the nucleus to be approximately spherical with a diameter of 7μm; thus the 1702 beads in our model occupy 32% of the nuclear volume. A summary of essential model parameters is shown is table 1.

5.3.4 Details of potentials and simulation

The simulation of a bead-spring polymer model is typically implemented with the Langevin equation: $m\ddot{x} = -\nabla U - \gamma \dot{x} + \Gamma$ (101-104). In this scheme a viscous friction, controlled by the value of γ, is balanced by uncorrelated gaussian noise, Γ, which represents collisions with energetic solvent molecules. The Langevin approach is well justified empirically and theoretically. Theoretically speaking, the dynamics of each bead is governed by the connecting springs, bead-bead interactions, friction, and solvent. Since the polymer is coarse grained, it is safe to assume that the time-scales corresponding to the oscillation of the bead in the potential wells of either the connecting springs or the non-bonded interactions are much longer than the time between consecutive collisions of solvent molecules. This is what allows solvent collisions to be modeled as random uncorrelated noise, Γ. Indeed, empirical observations of GFP tagged loci have confirmed that chromosome dynamics in interphase is Brownian (73).

The energy scale, K_bT , uses 300K in all simulations. We use a purely repulsive Lennard-Jones (Weeks-Chandler-Andersen) potential between non-bonded beads: $U(r) = 4\varepsilon\left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} + \left(\frac{\varepsilon}{4}\right)\right)$. In this scheme the traditional Lennard-Jones potential is cut (set to zero) at distances greater than $\sigma\sqrt[6]{2} = 2r_{bead}$ and shifted by $\varepsilon/4$ to represent the short range repulsion of spherical monomers. We use a well depth, ε, is set to $3K_bT$ (see figure 5.1), guided by the expectation that thermal fluctuations should not cause significant bead overlap. We use a harmonic potential between bonded beads. To minimize chain crossings we set the spring constant to $10k_bT$; this choice prohibits the significant bond fluctuations that would otherwise permit one link in the bead spring chain from extending sufficiently to cross over another. We use the integration time step $t_{step} = \tau/100$; here $\tau = \sigma\sqrt{m_{bead}/\varepsilon}$ is the Lennard-Jones (LJ) time scale (105). Temperature is maintained using a Langevin thermostat with friction term, γ, set to τ^{-1} as in (108,234). This choice of γ is discussed in detail below. Each bead in our model has mass, m_{bead}, of 77MDa (see derivation above).

5.3.5 Simulation Time Scales

In Langevin's original 1908 paper (104) the viscous resistance constant, γ, was determined by the Stokes' formula ($\gamma = 6\pi\mu a$) instead of the Lennard-Jones (simulation) time. In the Stokes formula μ is solvent viscosity and a is the radius of the Brownian particle. Often in coarse grained polymer models this approach implies a large value of γ and thus large viscous resistance forces compared to the bonded and non-bonded interactions in simulation. In our case, this would severely limit the simulation time step because the large viscous resistance forces would require a small time step on the order of $1/\gamma$. The modern approach (105-108) is to allow the Lennard-Jones forces to dictate the effective simulation times scales by artificially setting γ to the inverse of the LJ time scale τ, (see above). This artificial lowering of γ grants a larger time step without affecting the thermodynamic sampling of polymer configuration space - -a technique often used in MD simulations of protein folding (235-237). Indeed, simply setting γ to the inverse of the Lennard-Jones (simulation) time is common in polymer simulations (105-108).

Essentially, artificially reducing γ implies abandoning realistic time scales in a simulation in favor of rapidly exploring the available configuration space. In fact, a simulation that combines coarse graining and γ reduction can exceed the finite lifetime of most cells in interphase (75). However, without knowing of how time-scales of such a simulation map to reality it's impossible to predict what happens on biologically relevant time scales.. Fortunately, experimental data can be used to restore the realistic time scales. Usually this is done by comparing mean squared displacement (MSD) of loci tracked in experiment and simulation (75,88). We use this proven approach to establish, *a-posteriori,* a correspondence between the simulation and experimental time-scales. We begin with a plot of $\left\langle \Delta\vec{r}^{\,2}(t) \right\rangle$ for our wild type model and determine its initial slope, $6D_{sim}$ (see figure 5.5). Next, we define a dimensionless parameter λ such that $\lambda D_{sim} = D_{\exp}$. Thus, if we re-scale the simulation time with the fitting parameter λ, the experimental rate of diffusion is reproduced. Our results (see below) demonstrate that the simulation not only reproduces the initial slope, which would be trivial, but also reproduces the more complex experimental diffusive motion of interphase chromosomes in the nucleus. Thus, we are reasonably confident that the simulation time rescaled by λ corresponds to the realistic experimental time.

Since the dynamics of Brownian particles are intrinsically stochastic, their motion is best described by the average of squared particle displacement $\left\langle \Delta\vec{r}^{\,2} \right\rangle$. For free Brownian particles, the plot of $\left\langle \Delta\vec{r}^{\,2} \right\rangle$ is linear with a slope of $6D$ where D is the coefficient of diffusion; on the other hand, confined Brownian particles generally possess sub-linear $\left\langle \Delta\vec{r}^{\,2} \right\rangle$ plots which reach a plateau corresponding to the confinement radius. For fluorescently tagged chromosomal loci, both of these features have a biological interpretation. The plot's initial slope reflects the rate of diffusion where displacements are small and confinement effects are minimum; the plateau height reflects the radius of accessible volume within the nucleus. Thus, we calculate the diffusion constant in simulation, D_{sim}, from the initial slope of the $\left\langle \Delta\vec{r}^{\,2} \right\rangle$ plot generated from the simulation trajectory. The radius of confinement in simulation is approximated by observing the height at which the $\left\langle \Delta\vec{r}^{\,2} \right\rangle$ plot levels off. In the case of wild type D. melanogaster,

experimental $\left\langle \Delta \vec{r}^{\,2} \right\rangle$ plots (73) suggest that chromosome motion is diffusive at a rate of $D_{\exp} = 2 \cdot 10^{-11} \frac{cm^2}{s}$. This parameter in particular is used to map simulation trajectories to experimental time scales.

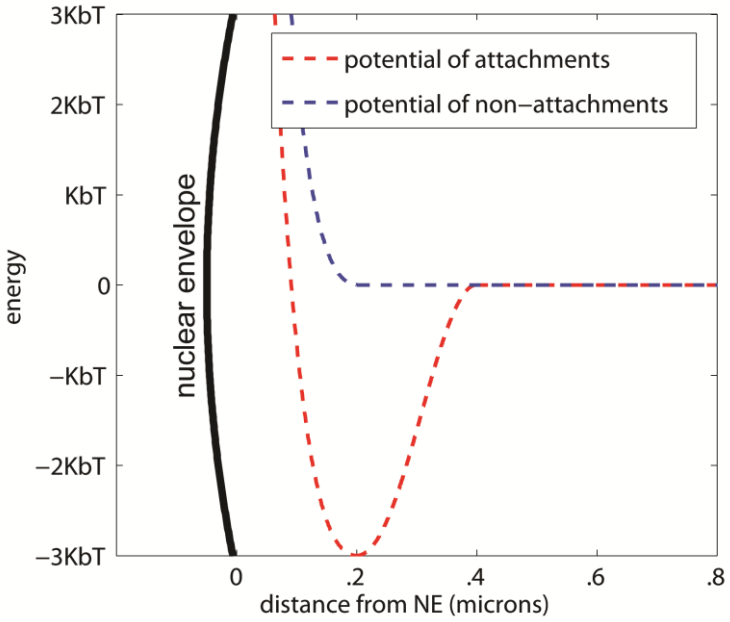


Figure 5.2 – Specific beads are attached to the nuclear envelope using a Lennard-Jones cosine interaction. Beads lacking affinity for the nuclear envelope use a shifted Lennard-Jones potential.

5.3.6 Chromosome nuclear envelope interactions

Experimental Lam data (12) for D. melanogaster interphase chromosomes is mapped to the corresponding beads in our model. One of two potential energy functions below is then assigned to each bead in the model based on the mapping outcome. Beads representing Lam associating regions identified in (12) interact with the NE in simulation according to a short range attractive potential. We use a Lennard-Jones cosine interaction: $U(r) = \frac{1}{2}\varepsilon\left(\cos\left[\alpha r^2 + \beta\right] - 1\right)$ where $\alpha = \frac{1}{\pi\left(r_{cut}^2 - \sqrt[3]{2}\sigma^2\right)}$ and $\beta = \pi - \sqrt[3]{2}\alpha\sigma^2$, with the well depth set to $3K_bT$ (red in figure 5.2). In this scheme the minimum of a traditional Lennard-Jones interaction is smoothly stitched to zero to maintain the function's differentiability (224). Beads lacking any mapped Lam associating regions interact with the nuclear envelope according to a shifted Lennard Jones potential (Weeks-Chandler-Andersen) as described above, see also figure 5.2. As a corollary, beads representing Lam associating regions anchor to the nuclear envelope in a potential well (red in figure 5.2) with passive confinement of non-attachment beads (blue in figure 5.2).

5.3.7 Models we consider

We consider two models of the D. melanogaster nucleus. Wild type model – possesses all known parameters of D. melanogaster nucleus and includes the chromosome-nuclear envelope attachments identified experimentally; this model corresponds to the experimentally accessibly wild type. Control model – chromosome do not possess specific sites of attachments to the nuclear envelope, their confinement is maintained passively (see above). This model possess all other features of the wild type model and represents a hypothetical mutant in which chromosomes do not anchor to the nuclear envelope.

5.3.8 Definition of fractal-like configurations

All simulations are initialized in "fractal-like" configurations. These fractal-like configurations are designed to match two key features possessed by experimental D. melanogaster chromosomes. First, the decay of intra-chromosomal contacts is described by the power law, $P(s) = s^{-1}$, where P is the probability of contact between two beads belonging to the same chromosome and s is their separation along the polymer backbone. Second, chromosomes are territorial and lack entanglement. In addition to these two key fractal-like signatures, chromosomes in fruit fly (28-30) possess a distinctly polarized (Rabl) chromosome arrangement characterized by clustering of chromosome centromeres and telomeres at opposite ends of the nucleus. This characteristic arrangements is not encoded automatically in the initial "fractal-like" configurations. Nonetheless, the Rabl (polarized) chromosome configuration is considered in our model as an additional constraint that initially positions chromosome centromeres and telomeres at opposite nuclear poles. The effects of chromosome-nuclear envelope attachments are then studied by comparing the dynamics of the wild type and control models as each transition to the equilibrium state. This aim is made quantitative by computing four observables during simulation trajectories: the scaling of intra-chromosome contact probability, the chromosome territory index, and the rates of chromosomal diffusion, and chromosome entanglement. Persistence of the Rabl chromosome configuration is compared to the 2 hour relaxation time suggested by experiment (76). Simulations are mapped to biologically relevant time scales (see results) and checked for robustness to the specifics of chain crossing (see results).

5.3.9 Robustness of results to initial conditions

Each computational model comprises 1703 monomers confined within a boundary representing the NE. Warm-up integration consists of $\sim 10^3$ integration steps designed to control introduction of the 300K Langevin thermostat. Subsequent simulation of the models consists of $\sim 10^6$ integration time steps preformed on Intel® core i7 type CPUs: each simulation completed within several hours of real time. We checked that the conclusions in this study (see results) are robust to several key details of the computational model. Robustness to initial conditions was checked by changing the random seed value used during warmup integration. Each random seed value generates a unique initial configurations prior to simulation. All model conclusions were reproduced using pairs of the wild type model and control model stemming from different random seeds: a total of 8 initial configurations were tested.

Topoisomerase II (topo II), an enzyme that facilitates strand crossing of the DNA, may increases the relaxation time of fractal-like chromatin configurations by allowing strands of dsDNA to cross. The activity of topo II activity in interphase is unclear: some studies indicate that topo II is present in the interphase nucleus (73) while other suggest that most of the topo II is degraded upon exit of mitosis (238). Thus, we considered the possibility that Topo II in the interphase nucleus may allow strands to cross by preforming simulations with a reduced barrier to strand crossing; simulations were recast with neighboring beads bonded harmonically with a spring constant set to $1k_bT$ (a tenfold decrease). A reduction of the harmonic spring constant serves as a simple model of topoisomerase II activity by allowing large bond fluctuations in the bead-spring model of chromatin; consequently, simulated chromosomes are allowed to cross. This simple approach to modeling topoisomerase II activity has been used previously (43). This outcome is detailed in results (see below) since chain crossing is a biologically significant phenomenon. We did not extensively investigate robustness of conclusions to the effects of model resolution; however, it has been noted in previous computational studies (81) and in theory (41) that polymer models are insensitive to coarse graining schemes above the Kuhn resolution.

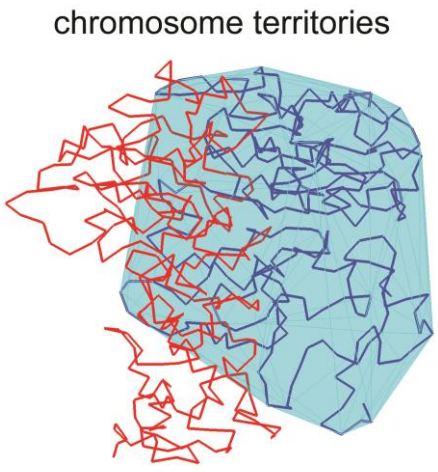


Figure 5.3 - The territory index of a chromosome is defined as the percent of its beads found inside the chromosome's own convex hull. Example: light blue chromosome.

5.3.10 Chromosome territory index

Our definition of the territory had been used previously in similar contexts (187). Briefly we begin by calculating the convex hull for a single chromosome (blue chromosome in figure 5.3); this is the minimum volume that includes all the chromosome beads inside a convex polyhedron. In general, each convex hull contains its own chromosome, and may also encompass some points belonging to other chromosomes (red chromosome in figure 5.3). A fully ‘‘territorial’’ chromosome is one whose convex hull does not contain beads from any other chromosomes while a less ‘‘territorial’’ chromosome is one whose convex hull contains some beads from other chromosome. We define the chromosome territory index as the fraction of beads inside a convex hull that belong to the chromosome used for its construction (figure 5.3).

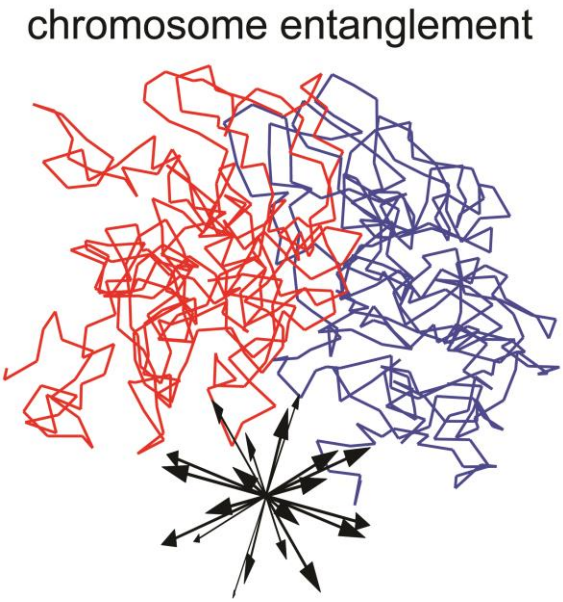


Figure 5.4 - Spatial separation of chromosomes with respect to putative translations (inset) is used to quantify chromosome entanglement. The minimum number of crossings enumerated in 20 directions is used as a quantitative measure of chromosome entanglement.

5.3.11 Chromosome entanglement

The concept of polymer intertwining has been quantified previously using Alexander polynomials. These polynomials are knot invariants and useful for identifying equivalent knots and quantifying knot complexity. Computing the Alexander polynomial of protein structures has led to biologically meaningful results. In particular, the presence of knots may stabilize a protein’s native folded state (130). Recent computational studies of genome organization have adopted use of Alexander polynomials for detecting intertwined chromosome configurations (43,220); however, the biological interpretation in this context is less clear. Specifically, knot complexity does not directly quantify the ability to segregate two different chromosomes which is necessary for passage through the cell cycle. To quantify how entangled two chromosomes are in free space we separate two model chromosome by putative translation in 3D space and enumerate the chain crossings along the direction of the applied

translation (see figure 5.4). In general, the number of chain crossings in different directions will differ; therefore, we test 20 directions (figure 5.4) that uniformly cover the S2 space (spherical surface). From the 20 directions tested, the minimum number of crossings quantifies the entanglement of a pair of different chromosomes. Biologically speaking, this number is intended to represent how easily to chromosomes separate in free space.

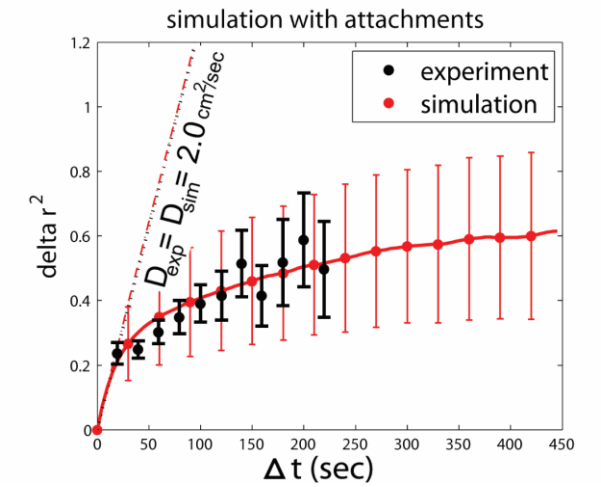


Figure 5.5 - By matching just one model parameter to experiment (effective diffusion coefficient, Dsim), simulation reproduces the complexity of experimental diffusive motion of non-polytene interphase chromosomes in the nucleus. Trivial unconfined diffusive motion would correspond to a straight line r2 = 6Dt. Error bars represent 1 standard deviation calculated from n=8 simulation trajectories.

5.4 RESULTS

5.4.1 Simulation time rescaling recapitulates the complex dynamics of chromosomes in interphase

To re-establish realistic time scales in each simulation we use a fitting parameter, λ, designed to match the chromosomal diffusion constant in experiment ($D_{\exp}$) and simulation (D_{sim}) (see methods). These diffusion constants are related only to the initial slope of the $\left\langle \Delta \vec{r}^{\,2} \right\rangle$ plot (see figure 5.5). However, we find that the simulation not only reproduces the initial slope, which would be trivial, but also apparently reproduces the more complex experimental diffusive motion of interphase chromosomes in the nucleus, Fig. 5**.** This means that by multiplying the elapsed time in each simulation trajectory by λ we recover realistic time scales (see figure 5.5).

The result of adjusting λ (figure 5.5) demonstrates that by varying just one parameter of the model (diffusion coefficient) to match experiment we automatically match the fairly complex, non-trivial dynamic behavior of regular interphase chromosomes under nuclear confinement**.** The plateau height from simulations suggests that chromosome motion is confined in sub-nuclear regions with .6 micron radius; the suggested confinement radius in experiment is .9 microns (73). Thus, experiment and simulation suggest that chromosomal loci do not freely

diffuse in the entire volume of the nucleus; instead their motion is confined within much smaller sub-regions of the nucleus. It has been speculated that loci confinement to these sub-regions stems from tethering to a nuclear structure such as the nuclear envelope (73). In what follows we test the effects of chromosome-NE attachments by comparing simulations of our wild type model (which possess chromosome-nuclear envelope attachments) to our control model (which lacks chromosome-nuclear envelope attachments).

5.4.2 Chromosome NE attachments may reinforce chromosome territories.

Theoretical studies suggest that fractal-like polymer configurations are highly territorial in the sense that chromosomes occupy distinct mutually exclusive domains without entangling (45,109,220,221). On the contrary, equilibrium configurations are expected to be less organized and highly entangled (45). Thus, transition from fractal-like to equilibrium configurations should be accompanied by the deterioration of chromosome territories. We compare chromosome territories in our wild type model (with Chr-NE attachment) to our control model (without attachments) using an established metric called the "territory index" based on the convex hull, (see methods). As expected, chromosomes in all simulations are initially highly territorial since each begins in a fractal-like configuration. We observe that the average territory index decreases with time in both the wild type model and control model (see figure 5.6). Thus, chromosome-nuclear envelope attachments are not sufficient to prevent some decline in chromosome territory index for fractal-like configurations. However, the absence of attachments in the control model simulations leads to a faster decay in the territory index. Indeed, from Fig. 5.6 it is immediately clear that Chr-NE attchments slow down territory deterioration; to test if

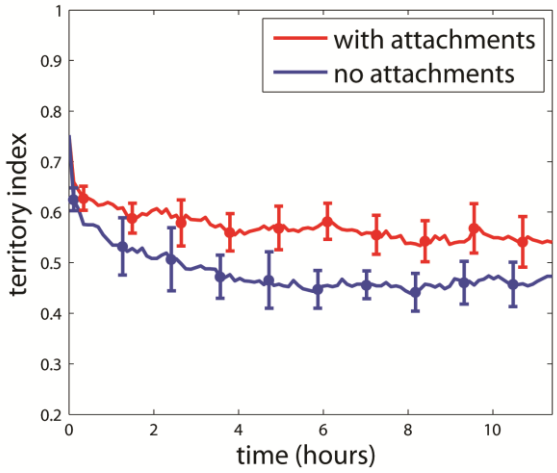


Figure 5.6 – effect of Chr-NE attachments on chromosome territories. Error bars represent 1 standard deviation calculated from n=8 simulation trajectories. Red line – mean with attachments; blue line – mean without attachments. Biologically speaking, the territory index (y-axis) represents the fraction of chromatin inside its native convex hull (see methods).

these also affect the end values on biologically relevant scales, the simulations were extended by a factor of 10 (not shown). The end values with and w/o the attachments were still different. In conclusion, attachments appear to stabilize but do not prevent inevitable declines in the territory index.

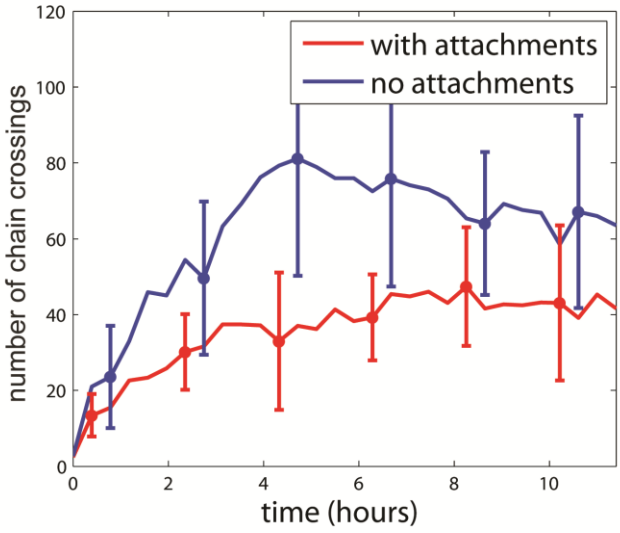


Figure 5.7 - effect of Chr-NE attachments on chromosome entanglement. Red line – mean with attachments; blue line – mean without attachments. Error bars represent 1 standard deviation calculated from n=8 simulation trajectories.

We add intuitive explanation for the observation that chromosome-NE attachments reinforce chromosome territories with robust dimension-based arguments that we used perviosuly to explain how chromosome-NE attachments affect intra-chromosomal interactions (222). We consider two limiting cases: (i) a completely spherical chromosome territory lacking any chromosome-NE attachments and (ii) a chromosome territory completely anchored to the NE possessing many attachments. We propose that in (i) the convex hull representing a chromosome territory (blue chromosome in figure 5.3) is easily invaded (red chromosome figure 5.3) due to its relatively large surface area in contact with neighboring territories. For instance, a completely spherical chromosome territory would have surface area $4\pi r^2$ and surface area to volume ratio 3/r. Next consider fully anchoring each chromosome to the nuclear envelope, corresponding to the limiting case (ii). In

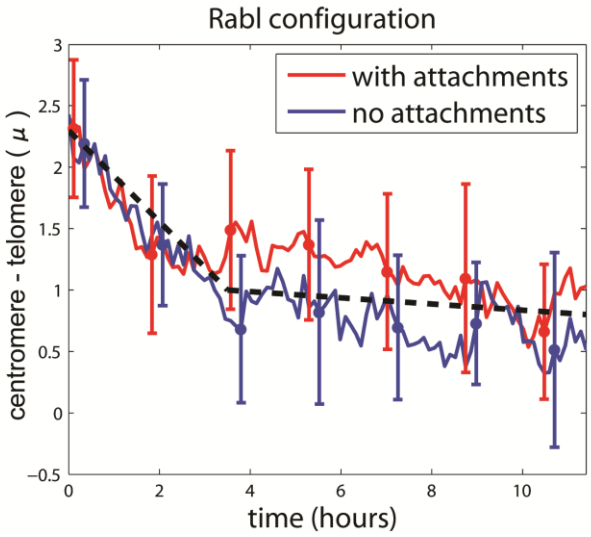


Figure 5.8 – effect of Chr-NE attachments on the degree of chromosome polarization (Rabl configuration). Error bars represent 1 standard deviation calculated from n=8 simulation trajectories. Red line – mean with attachments; blue line – mean without attachments.

this case, each chromosome territory occupies a thin layer annealed to the 2D interior of the nuclear envelope. In this limiting case the convex hull representing each chromosome territory would resemble a flattened surface that can be invaded by other chromosomes only along its 1D perimeter. In other words, there is less opportunity for chromosomes to invade neighboring chromosome territories when confined to the 2D surface of the nuclear envelope. Although chromosome territories in simulation and in reality adopt far more complex shapes, we speculate that a chromosome partially "annealed" to the nuclear envelope by Chr-NE attachments is less likely to be invaded by its neighbor chromosomes.

5.4.3 Chromosome nuclear envelope attachments limit but do not prevent chromosome entanglement.

Theoretical studies suggest that fractal-like polymer configurations are unentangled (43). In previous studies this lack of entanglement has been made quantitative with the concept of knot complexity which is computed by identifying knot invariant Alexander polynomials (45). The absence of entanglement and knots within fractal-like configurations in turn facilitates chromosome folding, unfolding, and loop opening (43,220). Each of these properties make fractal-like configurations biologically attractive; however, chromosomes tend to acquire knots as they transition from fractal-like configurations to equilibrium (43). We determine if chromosome-NE attachments prevent or at least delay the onset of chromosome entanglement during this transition; however, we *do not* quantify entanglement by computing knot invariant Alexander polynomials. Instead, we use a simple metric designed to more closely resemble the biological separation of chromosomes preceding passage through the cell cycle. Specifically, we apply putative translations to determine the minimum number of chain crossings (i.e. DNA double strand crossings) to fully dissociate two chromosomes in 3D space (see methods).

Our results suggest that chromosome nuclear envelope attachments delay but do not prevent the chromosome entanglement that arises during transition to equilibrium (figure 5.7). We note that some chromosome entangle arises quickly (faster than <1 hour) in the wild type model (which possess attachments) and control model (which lacks attachments). Due to this rapid accumulation of entanglement (regardless of chromosome-NE attachments) it's possible that some chromosomes entanglement is inevitable before physical segregation of chromosomes in mitosis; indeed, our quantitative metric of chromosome entanglement is designed to represent

this physical separation (see methods). We propose several conclusions. Chromosome nuclear envelope attachments may delay chromosome entangling long enough to ensure chromosome separability during cell division; possibly, this delay prevents a critical amount of entanglement that would otherwise interfere with proper cell division.

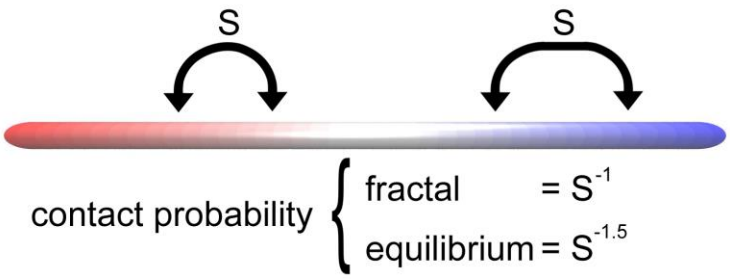


Figure 5.9 – Probability of contact, P, between loci belonging to the same chromosome depends on their separation, s, along the polymer backbone. In general this relation is captured by the parameter α in the expression, $P(s) = s^{\alpha}$

However, if a minimal amount of chromosome entanglement interferes with cell division then additional mechanisms must be enlisted to prevent entanglement more effectively than chromosome-NE attachments alone. These additional mechanisms could include DNA crosslinks (45). Theoretical studies demonstrate that crosslinks – which represent reversible protein bound DNA interactions - can significantly prolong the lifetime of the fractal chromosome configurations (43). On the contrary, limiting chromosome entanglement may not be the primary or even necessary role of chromosome-NE attachments if chromosome entanglement is not a significant obstacle during the cell cycle.

5.4.4 Chromosome nuclear envelope attachments do not maintain the Rabl configuration.

All simulated chromosome were initially configured in polarized arrangements consistent with the Rabl chromosome configuration present in D. melanogaster. In general the Rabl configuration is not a signature of fractal-like chromosome configurations; this additional property was specifically included in the initial configurations of our model (see methods). In the case of D. melanogaster, the dynamics of the Rabl configuration is well-studied experimentally; nuclei display a Rabl configuration only temporarily after mitosis (76). Breaking down of the Rabl configuration generally occurs after 2 hours with apposition of telomeres and pericentric heterochromatin often occurring after 5 hours (76). For each simulation the Rabl chromosome configuration was made quantitative by exploring the axial distance (Δr_z) between centromeres and telomeres of each chromosome arm. Our simulations suggest that apposition of telomeres

and pericentric heterochromatin occurs after 2-4 hours (figure 5.8) in relatively good agreement with the experimental value; the black dashed line (figure 5.8) is a guide to the eye which emphasizes the breaking down of the Rabl configuration occurring at 2-4h. There was no significant difference between models the wild type model (which includes chromosome-NE attachments) and control model (which lacks chromosome-NE attachments). Thus, chromosome-NE attachment may not prevent the breaking down of the Rabl configuration.

5.4.5 Chromosome-nuclear envelope attachments do not inhibit the evolution of the FG scaling exponent towards the equilibrium value.

Naturally, the probability of contact, P, between loci belonging to the same chromosome depends on their separation, s, along the polymer backbone (figure 5.9). In general this relation is captured by the parameter α in the expression, $P(s) = s^{\alpha}$. For fractal-like chromosome folding $\alpha = -1$ (45); for equilibrium chromosome folding $\alpha = -3/2$ (45). To determine if Chr-NE attachments prolong fractal-like chromosome folding we plot α during long simulations of our wild type model (which possesses chromosome-NE attachments) and control model (which lacks chromosome-NE attachments) (figure 5.6). These long simulations represent approximately ~11h in reality (see simulation time rescaling in methods). Since all simulations are initialized in fractal-like configurations with $\alpha \sim -1$; we expect α to gradually approach the equilibrium value, $-3/2$. We observe that α decays at a similar rate in all simulations regardless of the presence or absence of chromosome nuclear envelope attachments (figure 5.10). Consequently, we conclude that chromosome nuclear envelope

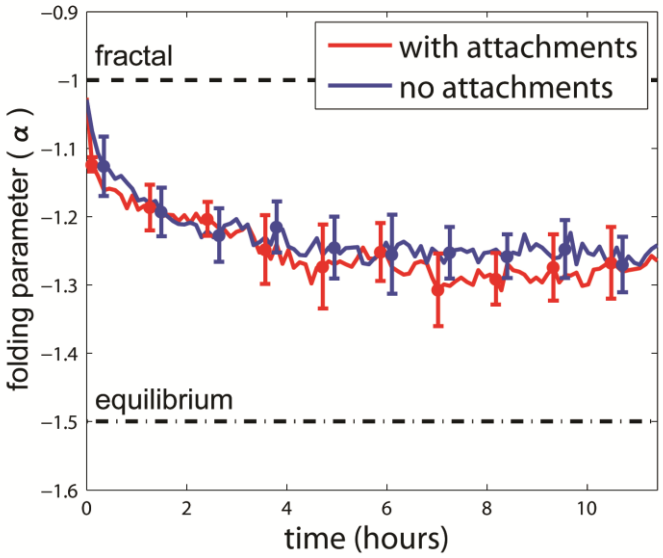


Figure 5.10 – scaling of chromosome contacts in the presence of attachments (top panel) and absence of attachment (bottom panel). Error bars represent 1 standard deviation calculated from n=8 simulation trajectories. Red line – mean with attachments; blue line – mean without attachments.

interactions may not affect fractal-like chromosome scaling. This conclusion suggests that the effects of chromosome-NE attachments on chromosomes folding are distinct from the effects chromosome crosslinks (45). Specifically, chromosome crosslinks are known to greatly delay the transition to equilibrium (45). To some extent chromosome-nuclear envelope attachments can be thought of as a type of crosslink – one that tethers chromosomes to a nuclear structure. Thus, it may be surprising that the presence of chromosome-NE attachments do not affect the decay of the fractal-like scaling exponent.

The relatively insensitive of the fractal-like scaling exponent can be rationalized by a simple dimensionality argument. The argument proposes that chromosome folding in the nucleus possess two limiting cases: a 2D case and a 3D case. Chromosomes with numerous NE-attachments are essentially annealed to the inner NE surface and represent the 2D case; meanwhile, chromosomes without attachments can explore the interior of the nucleus and represent the 3D case. Our simulations which lack attachments correspond to the 3D case. Meanwhile, simulations which possess attachments anchor a portion of each chromosome to the NE and represent an intermediate case. However, a previous study has shown that that for fractal-like curves $P(s) = S^{-1}$ regardless of dimension (11). In other words the scaling exponent is the same for the 3D case and the 2D case (11). Thus, the "turning on" of chromosome nuclear envelope attachments is expected to have little effect on the scaling exponent in simulation.

5.4.6 Effect of turning on Chr-NE attachments is robust to topoisomerase II activity

By design, all simulations virtually eliminated chain crossing by bonding neighboring beads harmonically with a spring constant set to $10k_bT$ (see methods); this choice prohibits the significant bond fluctuations that would otherwise permit one link in the bead spring chain from extending sufficiently to cross another. However, topoisomerase II (topo II), an enzyme that facilitates strand crossing of the DNA, is a known presence in the interphase nucleus (73). Thus, we considered the possibility that Topo II in the interphase nucleus may allow strands to cross by preforming simulations with a reduced barrier to strand crossing; simulations were recast with neighboring beads bonded harmonically with a spring constant set to $1k_bT$ (a tenfold decrease). A reduction of the harmonic spring constant serves as a simple model of topoisomerase II activity by allowing large bond fluctuations in the bead-spring model of chromatin;

consequently, simulated chromosomes are allowed to cross. This simple approach to modeling topoisomerase II activity has been used previously (43).

We reassessed the effects of chromosome-nuclear envelope attachments by comparing simulations of the wild type and control model *both possessing the lowered barrier to strand crossing*. All previously stated results (above) were recapitulated with the reduced strand crossing barrier. We conclude that the presence of chromosome-nuclear envelope attachments has two key effects regardless of simulated Topo II activity: chromosome territories are reinforced and chromosome entanglement is reduced.

5.5 DISCUSSION

5.5.1 Overall conclusions and limitations. Overall our simulations of fruit-fly interphase chromosomes suggest an important role of chromosome-nuclear envelope interactions in preserving nuclear architecture in higher eukaryotes. We emphasize four key results: 1) chromosome-nuclear envelope interactions assist in prolonging chromosome territories, 2) chromosome-nuclear envelope interactions limit chromosome entanglement 3) Chromosome-nuclear envelope interactions do not inhibit the evolution of the FG scaling exponent towards the equilibrium value, and 4) Chromosome-nuclear envelope interactions do not prevent the breaking down of the Rabl configuration which occurs after 2h. Each of these results was reiterated in the presence of simulated topoisomerase II activity. In addition, the effects of chromosome-nuclear envelope interactions on chromosome territories is explained by a simple volume accessibility argument.

5.5.2 Nuclear envelope attachments limit chromosome entanglement. Previous studies have quantified chromosome entanglement using Alexander polynomials (45). These polynomials are knot invariants useful for identifying knotted chromosomes, identifying equivalent knots, and quantifying knot complexity. It has been speculated that knot free chromosome configurations are biologically favorable and accommodate chromosome loop opening and closing (45). We develop a new simple measure of chromosome entanglement reminiscent of chromosome separation during the cell cycle. The metric enumerates chromosome strand crossings upon putative translation in free space. Our results suggest that chromosome-nuclear envelope attachments do not completely prevent chromosome entanglement in simulation. However, the

accumulation of chromosome entanglement is clearly delayed in the presence of Chr-NE attachment. Importantly, our simulations suggest that the time scale of this delay is on the order of the lifetime of the cell interphase. Thus, Chr-NE attachment not only limit chromosome entanglement, they do so on biologically relevant time scales. This result is consistent with previous studies of chromosome knot complexity but offers a potentially more tangible biological interpretation. In particular, the enumeration of strand crossings in the absence of Chr-NE attachment may exceed the capacity of Topo II necessary to fully separate two chromosomes. Unfortunately, little is known about Topo II activity in the interphase nucleus. An analysis of chromosome strand crossings and Topo II in the interphase nucleus will be pursued in a future study.

5.5.3 Unique effect of chromosome-nuclear envelope attachments on different fractal-like signatures. Support for the fractal-like configurations in human and drosophila stems the chromosome interaction probability described by the scaling law $P(s) = s^{-1}$. Physically speaking, a scaling law of this form means that chromosome loops form on all lengths which rules out many equilibrium based chromosome folding models (45). However, this unique scaling is lost upon transition to equilibrium. If these this change arises exclusively due to reptation of the polymer ends then transition to equilibrium may be on the order of $\sim N^3$ leading some computational studies to suggest that fractal configurations are stable for ~500 years (75). On the contrary, other studies suggest that fractal configurations quickly transition to a semi-entangled state and that crosslinks within the fractal globule are necessary to maintain its native shape (45). The situation is even less clear in D. melanogaster in which a fractal configurations are expected to reach equilibrium much faster due to its smaller genome size (and thus smaller $\sim N^3$). Although the endpoint of our simulations have unlikely reached equilibrium, we clearly see changes in chromosome contacts (scaling), chromosome territories, and chromosome entanglement on biologically relevant time scales. In addition, our model clearly suggests that Chr-NE interactions have unique effects on the signatures typically co-associated with fractal chromosome configurations. In particular, turning on Chr-NE attachments in our simulations appears to stabilize chromosome territories with little effect on chromosomal contacts.

5.5.4 Chromosome-nuclear envelope interactions do not affect the Rabl configuration. The field of research interested in 3D genome organization can perhaps be traced back to the original studies of Rabl and Boveri, which described a polarized configuration of chromosomes (now known as the Rabl configuration) and suggested the possibility of a highly organized nucleus. Surprisingly, the configuration is present in multiple lineages of metazoans such as fruit fly (30), yeast (22), and wheat (23); yet the details of the Rabl configuration are largely unknown. The characteristic polarization, with centromeres and telomeres at apogee within the nucleus, is speculated to be a vestige of the previous anaphase but has not been confirmed. In the case of D. melanogaster, the dynamics of the Rabl configuration are well-studied; its known lifetime is on the order of 2 hours in interphase with apposition of telomeres and pericentric heterochromatin often occurring after 5 hours (76). Our simulations suggest that the lifetime of Rabl configuration does not depend on the presence or absence of chromosome-nuclear envelope interactions. In addition, chromosome motion in our simulations is guided only by the dynamics of Brownian motion. Thus, the underlying Brownian motion of chromosomes may be sufficient to dictate large scale chromosome motions that evolve over the course of several hours.

5.5.5 Nuclear envelope attachments prolong chromosome territories regardless of simulated Topo II activity. The question of stabilizing territorial chromosome configurations was raised in a recent computational study which noted a rapid transition to the equilibrium state in the presence of active Topoisomerase II (topo II), an enzyme that facilitates strand crossing of the DNA (45). Although, knowledge of Topo II activity in interphase is limited, as is the ability of topo II to pass whole strands of chromosome; current evidence suggests that topo II does act efficiently on nucleosome bound DNA. Therefore, factors may be necessary to topologically constrain territorial chromosome configurations and prolong their lifetime in the cell. One proposal is the cross-linking of distant chromosomal loci, which prolonged the fractal globule in recent simulations (45); our proposal involves the interactions between chromosomes and the nuclear envelope. These interactions have already been shown to constrain chromatin motion in vivo (73) and reinforce chromosome territories in silico (187); both of these results allude to a topologically preserving role of chromosome-nuclear envelope interactions. The results presented here are consistent with these previous studies and uncover an important detail: chromosome-nuclear envelope interactions may stabilize chromosome territories and do so

regardless of simulated Topo II activity. This detail may be critically important if future experiments confirm the activity of topo II in the interphase nucleus which would otherwise induce rapid transition to equilibrium.

5.5.6 Limitations and future work. These results may be specific to D. melanogaster. For example, we cannot conclude that chromosome-nuclear envelope attachments also prolong territories in coil-like configurations which may exist in yeast; therefore, more work is needed to determine the role of chromosome-nuclear envelope interactions in organisms and cell types where chromosome folding principles differ. We acknowledge several limitations of our computational models. In particular, the bead radius, a key parameter in our computational models, is determined by the Kuhn length of D. melanogaster chromosomes, which in turn limits the highest resolution of our computational models. For simplicity we use short range potentials to represent chromosome nuclear envelope attachment but concede that potentials of a different form may be as realistic. For example, dynamically forming bonds between chromosomes and the nuclear envelope could be used to represent the protein dependent mechanisms that physically anchor chromosomes to the nuclear envelope. Regardless, our effort to incorporate these interactions is likely an improvement over their absence.

REFERENCES

1. Crick, F. (1970) Central dogma of molecular biology. *Nature*, **227**, 561-563.
2. Woodcock, C.L. and Dimitrov, S. (2001) Higher-order structure of chromatin and chromosomes. *Current opinion in genetics & development*, **11**, 130-135.
3. Horn, P.J. and Peterson, C.L. (2002) Molecular biology. Chromatin higher order folding--wrapping up transcription. *Science*, **297**, 1824-1827.
4. Risca, V.I. and Greenleaf, W.J. (2015) Unraveling the 3D genome: genomics tools for multiscale exploration. *Trends in genetics : TIG*, **31**, 357-372.
5. Iyer, B.V., Kenward, M. and Arya, G. (2011) Hierarchies in eukaryotic genome organization: Insights from polymer theory and simulations. *BMC biophysics*, **4**, 8.
6. Branco, M.R. and Pombo, A. (2006) Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *Plos Biol*, **4**, 780-788.
7. Dietzel, S., Jauch, A., Kienle, D., Qu, G.Q., Holtgreve-Grez, H., Eils, R., Munkel, C., Bittner, M., Meltzer, P.S., Trent, J.M. *et al.* (1998) Separate and variably shaped chromosome arm domains are disclosed by chromosome arm painting in human cell nuclei. *Chromosome Research*, **6**, 25-33.
8. Foster, H.A., Estrada-Girona, G., Themis, M., Garimberti, E., Hill, M.A., Bridger, J.M. and Anderson, R.M. (2013) Relative proximity of chromosome territories influences chromosome exchange partners in radiation-induced chromosome rearrangements in primary human bronchial epithelial cells. *Mutation research*, **756**, 66-77.
9. Khalil, A., Grant, J.L., Caddle, L.B., Atzema, E., Mills, K.D. and Arneodo, A. (2007) Chromosome territories have a highly nonspherical morphology and nonrandom positioning. *Chromosome Res*, **15**, 899-916.
10. Gibcus, J.H. and Dekker, J. (2013) The Hierarchy of the 3D Genome. *Mol Cell*, **49**, 773-782.
11. Lieberman-Aiden, E., van Berkum, N., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B., Sabo, P., Dorschner, M. *et al.* (2009) Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science*, **326**, 289-293.
12. Pickersgill, H., Kalverda, B., de Wit, E., Talhout, W., Fornerod, M. and van Steensel, B. (2006) Characterization of the Drosophila melanogaster genome at the nuclear lamina. *Nat Genet*, **38**, 1005-1014.
13. Ried, T., Schrock, E., Ning, Y. and Wienberg, J. (1998) Chromosome painting: a useful art. *Hum Mol Genet*, **7**, 1619-1626.
14. Zink, D., Cremer, T., Saffrich, R., Fischer, R., Trendelenburg, M.F., Ansorge, W. and Stelzer, E.H.K. (1998) Structure and dynamics of human interphase chromosome territories in vivo. *Hum Genet*, **102**, 241-251.
15. van den Engh, G., Sachs, R. and Trask, B.J. (1992) Estimating genomic distance from DNA sequence location in cell nuclei by a random walk model. *Science*, **257**, 1410-1412.
16. Sachs, R.K., Vandenengh, G., Trask, B., Yokota, H. and Hearst, J.E. (1995) A Random-Walk Giant-Loop Model for Interphase Chromosomes. *P Natl Acad Sci USA*, **92**, 2710-2714.
17. Mateos-Langerak, J., Bohn, M., de Leeuw, W., Giromus, O., Manders, E., Verschure, P., Indemans, M., Gierman, H., Heermann, D., van Driel, R. *et al.* (2009) Spatially confined folding of chromatin in the interphase nucleus. *Proceedings of the National Academy of Sciences*, **106**, 3812-3817.
18. SWIFT, H. (1962) In ALLEN, J. (ed.), *The Molecular Control of Cellular Activity*. McGraw-Hill, New York.

19. Mathog, D., Hochstrasser, M., Gruenbaum, Y., Saumweber, H. and Sedat, J. (1984) Characteristic Folding Pattern of Polytene Chromosomes in Drosophila Salivary-Gland Nuclei. *Nature*, **308**, 414-421.
20. Cremer, T., Cremer, C., Baumann, H., Luedtke, E.K., Sperling, K., Teuber, V. and Zorn, C. (1982) Rabl's model of the interphase chromosome arrangement tested in Chinese hamster cells by premature chromosome condensation and laser-UV-microbeam experiments. *Human genetics*, **60**, 46-56.
21. Hochstrasser, M., Mathog, D., Gruenbaum, Y., Saumweber, H. and Sedat, J.W. (1986) Spatial-Organization of Chromosomes in the Salivary-Gland Nuclei of Drosophila-Melanogaster. *Journal of Cell Biology*, **102**, 112-123.
22. Duan, Z., Andronescu, M., Schutz, K., McIlwain, S., Kim, Y.J., Lee, C., Shendure, J., Fields, S., Blau, C.A. and Noble, W.S. (2010) A three-dimensional model of the yeast genome. *Nature*, **465**, 363-367.
23. Abranches, R., Beven, A.F., Aragon-Alcaide, L. and Shaw, P.J. (1998) Transcription sites are not correlated with chromosome territories in wheat nuclei. *Journal of Cell Biology*, **143**, 5-12.
24. Dernburg, A.F., Sedat, J.W., Cande, W.Z. and Bass, H.W. (1995) 11 Cytology of Telomeres. *Cold Spring Harbor Monograph Archive*, **29**, 295-338.
25. Cowan, C.R., Carlton, P.M. and Cande, W.Z. (2001) The polar arrangement of telomeres in interphase and meiosis. Rabl organization and the bouquet. *Plant Physiol*, **125**, 532-538.
26. Spector, D.L. (2003) The dynamics of chromosome organization and gene regulation. *Annu Rev Biochem*, **72**, 573-608.
27. Santos, A.P. and Shaw, P. (2004) Interphase chromosomes and the Rabl configuration: does genome size matter? *Journal of microscopy*, **214**, 201-206.
28. Mathog, D., Hochstrasser, M., Gruenbaum, Y., Saumweber, H. and Sedat, J. (1984) Characteristic folding pattern of polytene chromosomes in Drosophila salivary gland nuclei. *Nature*, **308**, 414-421.
29. Gruenbaum, Y., Hochstrasser, M., Mathog, D., Saumweber, H., Agard, D.A. and Sedat, J.W. (1984) Spatial organization of the Drosophila nucleus: a three-dimensional cytogenetic study. *Journal of cell science. Supplement*, **1**, 223-234.
30. Hochstrasser, M., Mathog, D., Gruenbaum, Y., Saumweber, H. and Sedat, J.W. (1986) Spatial organization of chromosomes in the salivary gland nuclei of Drosophila melanogaster. *The Journal of Cell Biology*, **102**, 112-123.
31. Hochstrasser, M. and Sedat, J.W. (1987) Three-dimensional organization of Drosophila melanogaster interphase nuclei. I. Tissue-specific aspects of polytene nuclear architecture. *J Cell Biol*, **104**, 1455-1470.
32. Hochstrasser, M. and Sedat, J.W. (1987) 3-Dimensional Organization of Drosophila-Melanogaster Interphase Nuclei .2. Chromosome Spatial-Organization and Gene-Regulation. *Journal of Cell Biology*, **104**, 1471-1483.
33. Mathog, D. and Sedat, J.W. (1989) The Three-Dimensional Organization of Polytene Nuclei in Male Drosophila melanogaster With Compound XY or Ring X Chromosomes. *Genetics*, **121**, 293-311.
34. Cheutin, T., Bantignies, F., Leblanc, B. and Cavalli, G. (2010) Chromatin folding: from linear chromosomes to the 4D nucleus. *Cold Spring Harb Symp Quant Biol*, **75**, 461-473.
35. Peric-Hupkes, D. and van Steensel, B. (2010) Role of the nuclear lamina in genome organization and gene expression. *Cold Spring Harb Symp Quant Biol*, **75**, 517-524.
36. Belyaeva, E.S., Goncharov, F.P., Demakova, O.V., Kolesnikova, T.D., Boldyreva, L.V., Semeshin, V.F. and Zhimulev, I.F. (2012) Late Replication Domains in Polytene and Non-Polytene Cells of Drosophila melanogaster. *PloS one*, **7**, e30035.

37. Vatolina, T.Y., Boldyreva, L.V., Demakova, O.V., Demakov, S.A., Kokoza, E.B., Semeshin, V.F., Babenko, V.N., Goncharov, F.P., Belyaeva, E.S. and Zhimulev, I.F. (2011) Identical Functional Organization of Nonpolytene and Polytene Chromosomes in Drosophila melanogaster. *PloS one*, **6**.
38. Demakov, S.A., Yu Vatolina, T., Babenko, V.N., Semeshin, V.F., Belyaeva, E.S. and Zhimulev, I.F. (2011) Protein composition of interband regions in polytene and cell line chromosomes of Drosophila melanogaster. *BMC Genomics*, **12**, 566.
39. Zhimulev, I.F., Zykova, T.Y., Goncharov, F.P., Khoroshko, V.A., Demakova, O.V., Semeshin, V.F., Pokholkova, G.V., Boldyreva, L.V., Demidova, D.S., Babenko, V.N. *et al.* (2014) Genetic organization of interphase chromosome bands and interbands in Drosophila melanogaster. *PloS one*, **9**, e101631.
40. Wako, T. and Fukui, K. (2010) Higher organization and histone modification of the plant nucleus and chromosome. *Cytogenetic and genome research*, **129**, 55-63.
41. PGD, G. (1979) *Scaling concepts in polymer physics*. Cornell University Press, Ithaca.
42. Rubinstein M, C.R. (2003) *Polymer physics*. Oxford University Press, Oxford.
43. Mirny, L. (2011) The fractal globule as a model of chromatin architecture in the cell. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology*, **19**, 37-51.
44. Grosberg, A.Y., Nechaev, S.K. and Shakhnovich, E.I. (1988) The Role of Topological Constraints in the Kinetics of Collapse of Macromolecules. *J Phys-Paris*, **49**, 2095-2100.
45. Mirny, L.A. (2011) The fractal globule as a model of chromatin architecture in the cell. *Chromosome Res*, **19**, 37-51.
46. Dan, N. and Tirrell, M. (1992) Polymers Tethered to Curved Interfaces - a Self-Consistent-Field Analysis. *Macromolecules*, **25**, 2890-2895.
47. Wijmans, C.M. and Zhulina, E.B. (1993) Polymer Brushes at Curved Surfaces. *Macromolecules*, **26**, 7214-7224.
48. Mondescu, R.P. and Muthukumar, M. (1998) Brownian motion and polymer statistics on certain curved manifolds. *Physical Review E*, **57**, 4411-4419.
49. Goloborodko. (2013) S cerevisiae genome as a confined equilibrium polymer brush. *Epigentics & Chromatin*, **6 (Suppl 1)**, P 129.
50. Christian Munkel, R.E., Jorg Imhoff, Steffen and Dietzel, C.C.a.T.C. (1995) Simulation of the distribution of chromosome targets in cell nuclei under topological constraints. *Bioimaging*, **3**, 108–120.
51. Simova, I. and Herben, T. (2012) Geometrical constraints in the scaling relationships between genome size, cell size and cell cycle length in herbaceous plants. *P Roy Soc B-Biol Sci*, **279**, 867-875.
52. Marshall, W.F., Dernburg, A.F., Harmon, B., Agard, D.A. and Sedat, J.W. (1996) Specific interactions of chromatin with the nuclear envelope: positional determination within the nucleus in Drosophila melanogaster. *Molecular biology of the cell*, **7**, 825-842.
53. Lowenstein, M.G., Goddard, T.D. and Sedat, J.W. (2004) Long-range interphase chromosome organization in Drosophila: a study using color barcoded fluorescence in situ hybridization and structural clustering analysis. *Molecular biology of the cell*, **15**, 5678-5692.
54. Zakharenko, L.P., Perepelkina, M.P. and Mai, S. (2008) [Three-dimensional organization of telomeres in the Drosophila melanogaster salivary glands nuclei]. *Tsitologiia*, **50**, 585-589.
55. Capco, D.G., Wan, K.M. and Penman, S. (1982) The nuclear matrix: three-dimensional architecture and protein composition. *Cell*, **29**, 847-858.

56. Baricheva, E.A., Berrios, M., Bogachev, S.S., Borisevich, I.V., Lapik, E.R., Sharakhov, I.V., Stuurman, N. and Fisher, P.A. (1996) DNA from Drosophila melanogaster beta-heterochromatin binds specifically to nuclear lamins in vitro and the nuclear envelope in situ. *Gene*, **171**, 171-176.
57. Mirkovitch, J., Mirault, M.E. and Laemmli, U.K. (1984) Organization of the higher-order chromatin loop: specific DNA attachment sites on nuclear scaffold. *Cell*, **39**, 223-232.
58. Zhimulev, I.F., Belyaeva, E.S., Makunin, I.V., Pirrotta, V., Volkova, E.I., Alekseyenko, A.A., Andreyeva, E.N., Makarevich, G.F., Boldyreva, L.V., Nanayev, R.A. *et al.* (2003) Influence of the SuUR gene on intercalary heterochromatin in Drosophila melanogaster polytene chromosomes. *Chromosoma*, **111**, 377-398.
59. van Steensel, B. and Henikoff, S. (2000) Identification of in vivo DNA targets of chromatin proteins using tethered Dam methyltransferase. *Nature Biotechnology*, **18**, 424-428.
60. Cremer, T. and Cremer, M. (2010) Chromosome Territories. *Csh Perspect Biol*, **2**.
61. Akhtar, A. and Gasser, S.M. (2007) The nuclear envelope and transcriptional control. *Nat Rev Genet*, **8**, 507-517.
62. Henikoff, S. and Dreesen, T.D. (1989) Trans-Inactivation of the Drosophila Brown Gene - Evidence for Transcriptional Repression and Somatic Pairing Dependence. *Proceedings of the National Academy of Sciences of the United States of America*, **86**, 6704-6708.
63. Dernburg, A.F., Broman, K.W., Fung, J.C., Marshall, W.F., Philips, J., Agard, D.A. and Sedat, J.W. (1996) Perturbation of nuclear architecture by long-distance chromosome interactions. *Cell*, **85**, 745-759.
64. Chien, C.T., Buck, S., Sternglanz, R. and Shore, D. (1993) Targeting of Sir1 Protein Establishes Transcriptional Silencing at Hm Loci and Telomeres in Yeast. *Cell*, **75**, 531-541.
65. Gross, D.S. (2001) Sir proteins as transcriptional silencers. *Trends Biochem Sci*, **26**, 685-686.
66. Maillet, L., Boscheron, C., Gotta, M., Marcand, S., Gilson, E. and Gasser, S.M. (1996) Evidence for silencing compartments within the yeast nucleus: A role for telomere proximity and SIR protein concentration in silencer-mediated repression. *Gene Dev*, **10**, 1796-1811.
67. Andrulis, E.D., Neiman, A.M., Zappulla, D.C. and Sternglanz, R. (1998) Perinuclear localization of chromatin facilitates transcriptional silencing. *Nature*, **394**, 592-595.
68. Munoz-Alarcon, A., Pavlovic, M., Wismar, J., Schmitt, B., Eriksson, M., Kylsten, P. and Dushay, M.S. (2007) Characterization of lamin mutation phenotypes in Drosophila and comparison to human laminopathies. *PloS one*, **2**, e532.
69. Beard, G.S., Bridger, J.M., Kill, I.R. and Tree, D.R. (2008) Towards a Drosophila model of Hutchinson-Gilford progeria syndrome. *Biochem Soc Trans*, **36**, 1389-1392.
70. Mattioli, E., Columbaro, M., Capanni, C., Santi, S., Maraldi, N.M., D'Apice, M.R., Novelli, G., Riccio, M., Squarzoni, S., Foisner, R. *et al.* (2008) Drugs affecting prelamin A processing: effects on heterochromatin organization. *Exp Cell Res*, **314**, 453-462.
71. Verdaasdonk, J.S., Vasquez, P.A., Barry, R.M., Barry, T., Goodwin, S., Forest, M.G. and Bloom, K. (2013) Centromere Tethering Confines Chromosome Domains. *Mol Cell*, **52**, 819-831.
72. Vazquez, J., Belmont, A.S. and Sedat, J.W. (2001) Multiple regimes of constrained chromosome motion are regulated in the interphase Drosophila nucleus. *Curr Biol*, **11**, 1227-1239.
73. Marshall, W.F., Straight, A., Marko, J.F., Swedlow, J., Dernburg, A., Belmont, A., Murray, A.W., Agard, D.A. and Sedat, J.W. (1997) Interphase chromosomes undergo constrained diffusional motion in living cells. *Current biology : CB*, **7**, 930-939.
74. Heun, P., Laroche, T., Shimada, K., Furrer, P. and Gasser, S.M. (2001) Chromosome dynamics in the yeast interphase nucleus. *Science*, **294**, 2181-2186.
75. Rosa, A. and Everaers, R. (2008) Structure and Dynamics of Interphase Chromosomes. *PLoS Comput Biol*, **4**, e1000153.

76. Csink, A.K. and Henikoff, S. (1998) Large-scale chromosomal movements during interphase progression in Drosophila. *J Cell Biol*, **143**, 13-22.
77. Wong, H., Marie-Nelly, H., Herbert, S., Carrivain, P., Blanc, H., Koszul, R., Fabre, E. and Zimmer, C. (2012) A predictive computational model of the dynamic 3D interphase yeast nucleus. *Curr Biol*, **22**, 1881-1890.
78. Wong, H., Arbona, J.M. and Zimmer, C. (2013) How to build a yeast nucleus. *Nucleus*, **4**.
79. Tjong, H., Gong, K., Chen, L. and Alber, F. (2012) Physical tethering and volume exclusion determine higher-order genome organization in budding yeast. *Genome research*, **22**, 1295-1305.
80. Avsaroglu, B., Bronk, G., Gordon-Messer, S., Ham, J., Bressan, D.A., Haber, J.E. and Kondev, J. (2014) Effect of chromosome tethering on nuclear organization in yeast. *PloS one*, **9**, e102474.
81. de Nooijer, S., Wellink, J., Mulder, B. and Bisseling, T. (2009) Non-specific interactions are sufficient to explain the position of heterochromatic chromocenters and nucleoli in interphase nuclei. *Nucleic acids research*, **37**, 3558-3568.
82. Cook, P.R. and Marenduzzo, D. (2009) Entropic organization of interphase chromosomes. *J Cell Biol*, **186**, 825-834.
83. Chubb, J.R., Boyle, S., Perry, P. and Bickmore, W.A. (2002) Chromatin motion is constrained by association with nuclear compartments in human cells. *Curr Biol*, **12**, 439-445.
84. Münkel, C. and Langowski, J. (1998) Chromosome structure predicted by a polymer model. *Physical Review E*, **57**, 5888-5896.
85. Grosberg AY, K.A. (1994) *Statistical physics of macromolecules*. AIP, New York.
86. Franke, W.W. (1974) Structure, biochemistry, and functions of the nuclear envelope. *International review of cytology*, **Suppl 4**, 71-236.
87. Comings, D.E. and Okada, T.A. (1970) Association of Chromatin Fibers with Annuli of Nuclear Membrane. *Exp Cell Res*, **62**, 293-&.
88. Tokuda, N., Terada, T.P. and Sasai, M. (2012) Dynamical Modeling of Three-Dimensional Genome Organization in Interphase Budding Yeast. *Biophysical journal*, **102**, 296-304.
89. Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A. and Cavalli, G. (2012) Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome. *Cell*, **148**, 458-472.
90. Bystricky, K., Laroche, T., van Houwe, G., Blaszczyk, M. and Gasser, S.M. (2005) Chromosome looping in yeast: telomere pairing and coordinated movement reflect anchoring efficiency and territorial organization. *Journal of Cell Biology*, **168**, 375-387.
91. Rodley, C.D.M., Bertels, F., Jones, B. and O'Sullivan, J.M. (2009) Global identification of yeast chromosome interactions using Genome conformation capture. *Fungal Genet Biol*, **46**, 879-886.
92. Haber, J.E. and Leung, W.Y. (1996) Lack of chromosome territoriality in yeast: Promiscuous rejoining of broken chromosome ends. *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 13949-13954.
93. Bishop, M. and Saltiel, C.J. (1988) Application of the Pivot Algorithm for Investigating the Shapes of Two-Dimensional and 3-Dimensional Lattice Polymers. *J Chem Phys*, **88**, 6594-6596.
94. Madras, N. and Sokal, A.D. (1988) The Pivot Algorithm - a Highly Efficient Monte-Carlo Method for the Self-Avoiding Walk. *J Stat Phys*, **50**, 109-186.
95. Siepmann, J.I. and Frenkel, D. (1992) Configurational Bias Monte-Carlo - a New Sampling Scheme for Flexible Chains. *Mol Phys*, **75**, 59-70.
96. Rosenbluth, M.N.R.a.A.W. (1955) Monte Carlo Calculation of the Average Extension of Molecular Chains. *The Journal of Chemical Physics*, **23**, 4.
97. Michael Plischke, B.B. (2006) *Equilibrium Statistical Physics*. World Scientific Publishing Co. Pte. Ltd., Singapore.

98. Jinfeng Zhang, R.C., Chao Tang, and Jie Liang1. (2003) Origin of Scaling Behavior of Protein Packing Density: A Sequential Monte Carlo Study of Compact Long Chain Polymers. *J. Chem. Phys.*, **118**, 7.
99. Underhill, P.T. and Doyle, P.S. (2004) On the coarse-graining of polymers into bead-spring chains. *J Non-Newton Fluid*, **122**, 3-31.
100. Underhill, P.T. and Doyle, P.S. (2006) Alternative spring force law for bead-spring chain models of the worm-like chain. *J Rheol*, **50**, 513-529.
101. Ornstein, L.S. and Burger, H.C. (1919) On the theory of the Brownian motion. *P K Akad Wet-Amsterd*, **21**, 922-931.
102. Uhlenbeck, G.E. and Ornstein, L.S. (1930) On the theory of the Brownian motion. *Phys Rev*, **36**, 0823-0841.
103. Lemons, D.S. and Gythiel, A. (1997) Paul Langevin's 1908 paper ''On the theory of Brownian motion''. *Am J Phys*, **65**, 1079-1081.
104. Langevin, P. (1908) The theory of brownian movement. *Cr Hebd Acad Sci*, **146**, 530-533.
105. Kremer, K. and Grest, G. (1990) Dynamics of entangled linear polymer melts: A molecular-dynamics simulation. *The Journal of Chemical Physics*, **92**, 5057-5086.
106. Arnold, A. and Jun, S. (2007) Time scale of entropic segregation of flexible polymers in confinement: Implications for chromosome segregation in filamentous bacteria. *Physical Review E*, **76**.
107. Winkler, R.G., Gold, M. and Reineker, P. (1998) Collapse of polyelectrolyte macromolecules by counterion condensation and ion pair formation: A molecular dynamics simulation study. *Physical Review Letters*, **80**, 3731-3734.
108. Harmandaris, V.A., Reith, D., Van der Vegt, N.F.A. and Kremer, K. (2007) Comparison between coarse-graining models for polymer systems: Two mapping schemes for polystyrene. *Macromol Chem Physic*, **208**, 2109-2120.
109. Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O. *et al.* (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, **326**, 289-293.
110. Weber, S.C., Spakowitz, A.J. and Theriot, J.A. (2012) Nonthermal ATP-dependent fluctuations contribute to the in vivo motion of chromosomal loci. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 7338-7343.
111. Gehlen, L.R., Rosa, A., Klenin, K., Langowski, J., Gasser, S.M. and Bystricky, K. (2006) Spatially confined polymer chains: implications of chromatin fibre flexibility and peripheral anchoring on telomere-telomere interaction. *J Phys-Condens Mat*, **18**, S245-S252.
112. Gehlen, L.R., Gruenert, G., Jones, M.B., Rodley, C.D., Langowski, J. and O'Sullivan, J.M. (2012) Chromosome positioning and the clustering of functionally related loci in yeast is driven by chromosomal interactions. *Nucleus-Austin*, **3**, 370-383.
113. Hajjoul, H., Mathon, J., Ranchon, H., Goiffon, I., Mozziconacci, J., Albert, B., Carrivain, P., Victor, J.M., Gadal, O., Bystricky, K. *et al.* (2013) High-throughput chromatin motion tracking in living yeast reveals the flexibility of the fiber throughout the genome. *Genome research*, **23**, 1829-1838.
114. Hoze, N., Ruault, M., Amoruso, C., Taddei, A. and Holcman, D. (2013) Spatial telomere organization and clustering in yeast Saccharomyces cerevisiae nucleus is generated by a random dynamics of aggregation-dissociation. *Molecular biology of the cell*, **24**, 1791-1800.
115. Hoover, W.G. and Hoover, C.G. (2005) Nonequilibrium molecular dynamics. *Condens Matter Phys*, **8**, 247-260.

116. Hoover, W.G. (2005) Nonequilibrium molecular dynamics: Reversible irreversibility from symmetry breaking, thermostats, entropy production, and fractals. *Statistical Physics and Beyond*, **757**, 16-28.
117. Hoover, W.G. (1993) Nonequilibrium Molecular-Dynamics - the 1st 25 Years. *Physica A*, **194**, 450-461.
118. Hoover, W.G. (1992) Nonequilibrium Molecular-Dynamics. *Nucl Phys A*, **545**, C523-C536.
119. Paul, W., Binder, K., Heermann, D.W. and Kremer, K. (1991) Dynamics of Polymer-Solutions and Melts - Reptation Predictions and Scaling of Relaxation-Times. *J Chem Phys*, **95**, 7726-7740.
120. Barbieri, M., Chotalia, M., Fraser, J., Lavitas, L.M., Dostie, J., Pombo, A. and Nicodemi, M. (2012) Complexity of chromatin folding is captured by the strings and binders switch model. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 16173-16178.
121. Salceda, J., Fernandez, X. and Roca, J. (2006) Topoisomerase II, not topoisomerase I, is the proficient relaxase of nucleosomal DNA. *Embo Journal*, **25**, 2575-2583.
122. Trigueros, S., Salceda, J., Bermudez, I., Fernandez, X. and Roca, J. (2004) Asymmetric removal of supercolls suggests how topoisomerase II simplifies DNA topology. *Journal of Molecular Biology*, **335**, 723-731.
123. Jerabek, H. and Heermann, D.W. (2014) How chromatin looping and nuclear envelope attachment affect genome organization in eukaryotic cell nuclei. *International review of cell and molecular biology*, **307**, 351-381.
124. Khalil, A., Grant, J.L., Caddle, L.B., Atzema, E., Mills, K.D. and Arneodo, A. (2007) Chromosome territories have a highly nonspherical morphology and nonrandom positioning. *Chromosome Res*, **15**, 899-916.
125. Koniaris, K. and Muthukumar, M. (1991) Knottedness in Ring Polymers. *Physical Review Letters*, **66**, 2211-2214.
126. Muthukumar, M. and Koniaris, K.G. (1994) Knots in Polymeric Systems. *Abstr Pap Am Chem S*, **208**, 344-PMSE.
127. Taylor, W.R. (2000) A deeply knotted protein structure and how it might fold. *Nature*, **406**, 916-919.
128. Virnau, P., Mirny, L.A. and Kardar, M. (2006) Intricate knots in proteins: Function and evolution. *Plos Computational Biology*, **2**, 1074-1079.
129. Kolesov, G., Virnau, P., Kardar, M. and Mirny, L.A. (2007) Protein knot server: detection of knots in protein structures. *Nucleic acids research*, **35**, W425-W428.
130. Yeates, T., Norcross, T.S. and King, N.P. (2007) Knotted and topologically complex proteins as models for studying folding and stability. *Curr Opin Chem Biol*, **11**, 595-603.
131. Bauer, C.R., Hartl, T.A. and Bosco, G. (2012) Condensin II Promotes the Formation of Chromosome Territories by Inducing Axial Compaction of Polyploid Interphase Chromosomes. *Plos Genet*, **8**.
132. Tokuda, N., Terada, T.R. and Sasai, M. (2012) Dynamical modeling of three-dimensional genome organization in interphase budding yeast (vol 102, pg 296, 2012). *Biophys J*, **102**, 719-719.
133. Albert, B., Mathon, J., Shukla, A., Saad, H., Normand, C., Leger-Silvestre, I., Villa, D., Kamgoue, A., Mozziconacci, J., Wong, H. *et al.* (2013) Systematic characterization of the conformation and dynamics of budding yeast chromosome XII. *J Cell Biol*, **202**, 201-210.
134. Wong, H., Winn, P.J. and Mozziconacci, J. (2009) A molecular model of chromatin organisation and transcription: how a multi-RNA polymerase II machine transcribes and remodels the beta-globin locus during development. *Bioessays*, **31**, 1357-1366.
135. Agmon, N., Liefshitz, B., Zimmer, C., Fabre, E. and Kupiec, M. (2013) Effect of nuclear architecture on the efficiency of double-strand break repair. *Nature cell biology*, **15**, 694-699.

136. Marti-Renom, M. and Mirny, L. (2011) Bridging the resolution gap in structural modeling of 3D genome organization. *PLoS computational biology*, **7**, e1002125.
137. Wong, H., Victor, J.-M. and Mozziconacci, J. (2007) An All-Atom Model of the Chromatin Fiber Containing Linker Histones Reveals a Versatile Structure Tuned by the Nucleosomal Repeat Length. *PloS one*, **436**, e877.
138. Heermann, D.W., Jerabek, H., Liu, L. and Li, Y. (2012) A model for the 3D chromatin architecture of pro and eukaryotes. *Methods*, **58**, 307-314.
139. Finan, K., Cook, P.R. and Marenduzzo, D. (2011) Non-specific (entropic) forces as major determinants of the structure of mammalian chromosomes. *Chromosome Research*, **19**, 53-61.
140. Bohn, M. and Heermann, D.W. (2011) Repulsive forces between looping chromosomes induce entropy-driven segregation. *PloS one*, **6**, e14428.
141. Munkel, C., Eils, R., Dietzel, S., Zink, D., Mehring, C., Wedemann, G., Cremer, T. and Langowski, J. (1999) Compartmentalization of interphase chromosomes observed in simulation and experiment. *Journal of Molecular Biology*, **285**, 1053-1065.
142. Jerabek, H. and Heermann, D.W. (2012) Expression-dependent folding of interphase chromatin. *PloS one*, **7**, e37525.
143. Lis, J.T. (2007) Imaging Drosophila gene activation and polymerase pausing in vivo. *Nature*, **450**, 198-202.
144. Gilbert, L.I. (2008) Drosophila is an inclusive model for human diseases, growth and development. *Molecular and Cellular Endocrinology*, **293**, 25-31.
145. Riddle, N.C., Shaffer, C.D. and Elgin, S.C.R. (2009) A lot about a little dot - lessons learned from Drosophila melanogaster chromosome 4. *Biochemistry and Cell Biology-Biochimie Et Biologie Cellulaire*, **87**, 229-241.
146. Zhimolev, I.F. (1996) Morphology and structure of polytene chromosomes. *Adv Genet*, **34**, 1-490.
147. Zhimulev, I.F., Belyaeva, E.S., Vatolina, T.Y. and Demakov, S.A. (2012) Banding patterns in Drosophila melanogaster polytene chromosomes correlate with DNA-binding protein occupancy. *Bioessays*, **34**, 498-508.
148. Vatolina, T., Boldyreva, L., Demakova, O., Demakov, S., Kokoza, E., Semeshin, V., Babenko, V., Goncharov, F., Belyaeva, E. and Zhimulev, I. (2011) Identical Functional Organization of Nonpolytene and Polytene Chromosomes in Drosophila melanogaster. *PloS one*, **6**, e25960.
149. Demakov, S.A., Vatolina, T.Y., Babenko, V.N., Semeshin, V.F., Belyaeva, E.S. and Zhimulev, I.F. (2011) Protein composition of interband regions in polytene and cell line chromosomes of Drosophila melanogaster. *BMC Genomics*, **12**.
150. Hübner, M. and Spector, D. (2010) Chromatin Dynamics. *Annual Review of Biophysics*, **39**, 471-489.
151. Finlan, L.E., Sproul, D., Thomson, I., Boyle, S., Kerr, E., Perry, P., Ylstra, B., Chubb, J.R. and Bickmore, W.A. (2008) Recruitment to the nuclear periphery can alter expression of genes in human cells. *Plos Genet*, **4**.
152. Filion, G.J., van Bemmel, J.G., Braunschweig, U., Talhout, W., Kind, J., Ward, L.D., Brugman, W., de Castro, I.J., Kerkhoven, R.M., Bussemaker, H.J. *et al.* (2010) Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. *Cell*, **143**, 212-224.
153. Vogel, M.J., Peric-Hupkes, D. and van Steensel, B. (2007) Detection of in vivo protein-DNA interactions using DamID in mammalian cells. *Nat Protoc*, **2**, 1467-1478.
154. Hahnfeldt, P., Hearst, J.E., Brenner, D.J., Sachs, R.K. and Hlatky, L.R. (1993) Polymer models for interphase chromosomes. *Proceedings of the National Academy of Sciences*, **90**, 7854-7858.

155. Sachs, R.K., van den Engh, G., Trask, B., Yokota, H. and Hearst, J.E. (1995) A random-walk/giant-loop model for interphase chromosomes. *Proceedings of the National Academy of Sciences*, **92**, 2710-2714.
156. de Nooijer, S., Wellink, J., Mulder, B. and Bisseling, T. (2009) Non-specific interactions are sufficient to explain the position of heterochromatic chromocenters and nucleoli in interphase nuclei. *Nucleic Acids Res*, **37**, 3558-3568.
157. Cremer, T. and Cremer, C. (2001) Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nature Reviews Genetics*, **2**, 292-301.
158. MATLAB. (2010) *MATLAB:2010 version 7.10.0 (R2010a)*. The MathWorks Inc., Natick, Massachusetts.
159. Dayantis, J. and Palierne, J.F. (1994) Scaling exponents of the self-avoiding-walk-problem in three dimensions. *Physical review. B, Condensed matter*, **49**, 3217-3225.
160. Sergio Caracciolo, M.S.C., Andrea Pelissetto. (1997) High-precision determination of the critical exponent γ for self-avoiding walks. *Phys. Rev.*, **E**, R1215-R1218.
161. Guttmann, A.J. (1989) On the Critical-Behavior of Self-Avoiding Walks .2. *J Phys a-Math Gen*, **22**, 2807-2813.
162. Alexander Y. Grosberg, A.R.K., Pierre-Gilles de Genne. (2011) *Giant molecules: Here, there and everywhere*. World Scientific Publishing, Singapore.
163. Daban, J.R. (2003) High concentration of DNA in condensed chromatin. *Biochemistry and Cell Biology-Biochimie Et Biologie Cellulaire*, **81**, 91-99.
164. Neal Madras, G.S. (1993) *The Self Avoiding Walk*. Birkhauser, Boston, Boston.
165. Cacciuto, A. and Luijten, E. (2006) Self-Avoiding Flexible Polymers under Spherical Confinement. *Nano Letters*, **6**, 901-905.
166. Marshall, W., Marko, J., Agard, D. and Sedat, J. (2001) Chromosome elasticity and mitotic polar ejection force measured in living Drosophila embryos by four-dimensional microscopy-based motion analysis. *Current Biology*, **11**, 569-578.
167. Mateos-Langerak, J., Bohn, M., de Leeuw, W., Giromus, O., Manders, E.M., Verschure, P.J., Indemans, M.H., Gierman, H.J., Heermann, D.W., van Driel, R. *et al.* (2009) Spatially confined folding of chromatin in the interphase nucleus. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 3812-3817.
168. Gonzalez-Melendi, P., Beven, A., Boudonck, K., Abranches, R., Wells, B., Dolan, L. and Shaw, P. (2000) The nucleus: a highly organized but dynamic structure. *J Microsc-Oxford*, **198**, 199-207.
169. Ferreira, J., Paolella, G., Ramos, C. and Lamond, A.I. (1997) Spatial organization of large-scale chromatin domains in the nucleus: A magnified view of single chromosome territories. *Journal of Cell Biology*, **139**, 1597-1610.
170. Bau, D. and Marti-Renom, M.A. (2011) Structure determination of genomic domains by satisfaction of spatial restraints. *Chromosome Research*, **19**, 25-35.
171. Jun, S. and Mulder, B. (2006) Entropy-driven spatial organization of highly confined polymers: Lessons for the bacterial chromosome. *Proceedings of the National Academy of Sciences*, **103**, 12388-12393.
172. Jun, S. (2008) Can entropy save bacteria? *eprint arXiv:0808.2646*.
173. Rzepecki, R., Bogachev, S.S., Kokoza, E., Stuurman, N. and Fisher, P.A. (1998) In vivo association of lamins with nucleic acids in Drosophila melanogaster. *J Cell Sci*, **111 (Pt 1)**, 121-129.
174. Zhao, K., Harel, A., Stuurman, N., Guedalia, D. and Gruenbaum, Y. (1996) Binding of matrix attachment regions to nuclear lamin is mediated by the rod domain and depends on the lamin polymerization state. *FEBS letters*, **380**, 161-164.
175. Shevelyov, Y.Y., Lavrov, S.A., Mikhaylova, L.M., Nurminsky, I.D., Kulathinal, R.J., Egorova, K.S., Rozovsky, Y.M. and Nurminsky, D.I. (2009) The B-type lamin is required for somatic repression of

testis-specific gene clusters. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 3282-3287.
176. van Bemmel, J.G., Pagie, L., Braunschweig, U., Brugman, W., Meuleman, W., Kerkhoven, R.M. and van Steensel, B. (2010) The insulator protein SU(HW) fine-tunes nuclear lamina interactions of the Drosophila genome. *PLoS ONE*, **5**, e15013.
177. Hochstrasser, M. (1987) Chromosome structure in four wild-type polytene tissues of <i>Drosophila melanogaster</i&gt. *Chromosoma*, **95**, 197-208.
178. Cremer, T. and Cremer, C. (2001) Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet*, **2**, 292-301.
179. Gandhi, M., Evdokimova, V. and Nikiforov, Y.E. (2010) Mechanisms of chromosomal rearrangements in solid tumors: the model of papillary thyroid carcinoma. *Mol Cell Endocrinol*, **321**, 36-43.
180. Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M.B., Talhout, W., Eussen, B.H., de Klein, A., Wessels, L., de Laat, W. *et al.* (2008) Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature*, **453**, 948-951.
181. Kind, J. and van Steensel, B. (2014) Stochastic genome-nuclear lamina interactions: Modulating roles of Lamin A and BAF. *Nucleus*, **5**, 124-130.
182. Amendola, M. and van Steensel, B. (2014) Mechanisms and dynamics of nuclear lamina-genome interactions. *Curr Opin Cell Biol*, **28C**, 61-68.
183. Meaburn, K.J. and Misteli, T. (2007) Cell biology: chromosome territories. *Nature*, **445**, 379-781.
184. Dekker, J. (2008) Mapping in vivo chromatin interactions in yeast suggests an extended chromatin fiber with regional variation in compaction. *J Biol Chem*, **283**, 34532-34540.
185. Crabbe, L., Cesare, A.J., Kasuboski, J.M., Fitzpatrick, J.A. and Karlseder, J. (2012) Human telomeres are tethered to the nuclear envelope during postmitotic nuclear assembly. *Cell reports*, **2**, 1521-1529.
186. Laachi, N. and Dorfman, K. (2010) Statistics of tethered self-avoiding chains under spherical confinement and an external force. *The Journal of Chemical Physics*, **132**, 084108.
187. Kinney, N.A., Sharakhov, I.V. and Onufriev, A.V. (2014) Investigation of the chromosome regions with significant affinity for the nuclear envelope in fruit fly - a model based approach. *PloS one*, **9**, e91943.
188. Marshall, W.F., Straight, A., Marko, J.F., Swedlow, J., Dernburg, A., Belmont, A., Murray, A.W., Agard, D.A. and Sedat, J.W. (1997) Interphase chromosomes undergo constrained diffusional motion in living cells. *Curr Biol*, **7**, 930-939.
189. Pombi, M., Caputo, B., Simard, F., Di Deco, M.A., Coluzzi, M., della Torre, A., Costantini, C., Besansky, N.J. and Petrarca, V. (2008) Chromosomal plasticity and evolutionary potential in the malaria vector Anopheles gambiae sensu stricto: insights from three decades of rare paracentric inversions. *BMC Evol Biol*, **8**, 309.
190. Alekseyev, M.A. (2008) Multi-Break Rearrangements and Breakpoint Re-Uses: From Circular to Linear Genomes. *J Comput Biol*, **15**, 1117-1131.
191. Marshall, W.F. (2002) Order and disorder in the nucleus. *Curr Biol*, **12**, R185-192.
192. Folle, G.A. (2008) Nuclear architecture, chromosome domains and genetic damage. *Mutat Res*, **658**, 172-183.
193. Sedat, J. and Manuelidis, L. (1978) A direct approach to the structure of eukaryotic chromosomes. *Cold Spring Harb Symp Quant Biol*, **42 Pt 1**, 331-350.
194. Sobel, I. (1968) An Isotropic 3 3 Image Gradient Operator. *Presentation at Stanford A.I. Project*
195. Ostashevsky, J. (2002) A polymer model for large-scale chromatin organization in lower eukaryotes. *Molecular biology of the cell*, **13**, 2157-2169.

196. Tark-Dame, M., van Driel, R. and Heermann, D. (2011) Chromatin folding--from biology to polymer models and back. *Journal of cell science*, **124**, 839-845.
197. Hsu, H.P., Paul, W. and Binder, K. (2010) Standard Definitions of Persistence Length Do Not Describe the Local "Intrinsic" Stiffness of Real Polymer Chains. *Macromolecules*, **43**, 3094-3102.
198. van Bemmel, J.G., Pagie, L., Braunschweig, U., Brugman, W., Meuleman, W., Kerkhoven, R.M. and van Steensel, B. (2010) The Insulator Protein SU(HW) Fine-Tunes Nuclear Lamina Interactions of the Drosophila Genome. *PLoS ONE*, **5**.
199. Vasquez, P.A. and Bloom, K. (2014) Polymer models of interphase chromosomes. *Nucleus-Austin*, **5**, 376-390.
200. Halverson, J.D., Smrek, J., Kremer, K. and Grosberg, A.Y. (2014) From a melt of rings to chromosome territories: the role of topological constraints in genome folding. *Reports on progress in physics. Physical Society*, **77**, 022601.
201. Duncan, I.W. (2002) Transvection effects in Drosophila. *Annual review of genetics*, **36**, 521-556.
202. Muthukumar, M. and Baumgartner, A. (1989) Effects of Entropic Barriers on Polymer Dynamics. *Macromolecules*, **22**, 1937-1941.
203. Osborne, C.S., Chakalova, L., Brown, K.E., Carter, D., Horton, A., Debrand, E., Goyenechea, B., Mitchell, J.A., Lopes, S., Reik, W. *et al.* (2004) Active genes dynamically colocalize to shared sites of ongoing transcription. *Nature Genetics*, **36**, 1065-1071.
204. Moorman, C., Sun, L.V., Wang, J., de Wit, E., Talhout, W., Ward, L.D., Greil, F., Lu, X.J., White, K.P., Bussemaker, H.J. *et al.* (2006) Hotspots of transcription factor colocalization in the genome of Drosophila melanogaster. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 12027-12032.
205. Jost, D., Carrivain, P., Cavalli, G. and Vaillant, C. (2014) Modeling epigenome folding: formation and dynamics of topologically associated chromatin domains. *Nucleic acids research*, **42**, 9553-9561.
206. Di Stefano, M., Rosa, A., Belcastro, V., di Bernardo, D. and Micheletti, C. (2013) Colocalization of coregulated genes: a steered molecular dynamics study of human chromosome 19. *PLoS Comput Biol*, **9**, e1003019.
207. Tanizawa, H., Iwasaki, O., Tanaka, A., Capizzi, J.R., Wickramasinghe, P., Lee, M., Fu, Z. and Noma, K. (2010) Mapping of long-range associations throughout the fission yeast genome reveals global genome organization linked to transcriptional regulation. *Nucleic acids research*, **38**, 8164-8177.
208. Hochstrasser, M., Mathog, D., Gruenbaum, Y., Saumweber, H. and Sedat, J.W. (1986) Spatial organization of chromosomes in the salivary gland nuclei of Drosophila melanogaster. *J Cell Biol*, **102**, 112-123.
209. Edwards, S.F. and Freed, K.F. (1969) The entropy of a confined polymer. I. *Journal of Physics A: General Physics*, **2**, 145-150.
210. Heermann, D.W. (2011) Physical nuclear organization: loops and entropy. *Curr Opin Cell Biol*, **23**, 332-337.
211. Bohn, M. and Heermann, D.W. (2010) Diffusion-driven looping provides a consistent framework for chromatin organization. *PloS one*, **5**, e12218.
212. Bohn, M., Heermann, D.W. and van Driel, R. (2007) Random loop model for long polymers. *Physical review. E, Statistical, nonlinear, and soft matter physics*, **76**, 051805.
213. Ganai, N., Sengupta, S. and Menon, G.I. (2014) Chromosome positioning from activity-based segregation. *Nucleic acids research*, **42**, 4145-4159.
214. Wu, P.H., Giri, A., Sun, S.X. and Wirtz, D. (2014) Three-dimensional cell migration does not follow a random walk. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 3949-3954.

215. Leonid Mirny. personal communication.
216. Hilliker, A.J. (1985) Assaying chromosome rearrangement in embryonic interphase nuclei of Drosophila melanogaster by radiation-induced interchanges. *Genet. Res.*, **47**, 13-18.
217. Cao, J., Luo, Z.Y., Cheng, Q.Y., Xu, Q.L., Zhang, Y., Wang, F., Wu, Y. and Song, X.Y. (2015) Three-dimensional regulation of transcription. *Protein Cell*, **6**, 241-253.
218. Pombo, A. and Dillon, N. (2015) Three-dimensional genome architecture: players and mechanisms. *Nat Rev Mol Cell Bio*, **16**, 245-257.
219. Fudenberg, G., Getz, G., Meyerson, M. and Mirny, L. (2011) High order chromatin architecture shapes the landscape of chromosomal alterations in cancer. *Nature Biotechnology*, **29**, 1109-1113.
220. Zhang, B. and Wolynes, P.G. (2015) Topology, structures, and energy landscapes of human chromosomes. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, 6062-6067.
221. Sanborn, A.L., Rao, S.S., Huang, S.C., Durand, N.C., Huntley, M.H., Jewett, A.I., Bochkov, I.D., Chinnappan, D., Cutkosky, A., Li, J. *et al.* (2015) Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proceedings of the National Academy of Sciences of the United States of America*.
222. Kinney, N.A., Onufriev, A.V. and Sharakhov, I.V. (2015) Quantified effects of chromosome-nuclear envelope attachments on 3D organization of chromosomes. *Nucleus*, **6**, 212-224.
223. Pickersgill, H., Kalverda, B., de Wit, E., Talhout, W., Fornerod, M. and van Steensel, B. (2006) Characterization of the Drosophila melanogaster genome at the nuclear lamina. *Nat Genet*, **38**, 1005-1014.
224. Limbach, H.J., Arnold, A., Mann, B.A. and Holm, C. (2006) ESPResSo - an extensible simulation package for research on soft matter systems. *Comput Phys Commun*, **174**, 704-727.
225. Triebel, H., Reinert, K.E. and Strassburger, J. (1971) Persistence length of DNA from hydrodynamic measurements. *Biopolymers*, **10**, 2619-2621.
226. Porschke, D. (1991) Persistence length and bending dynamics of DNA from electrooptical measurements at high salt concentrations. *Biophysical chemistry*, **40**, 169-179.
227. Bednar, J., Furrer, P., Katritch, V., Stasiak, A.Z., Dubochet, J. and Stasiak, A. (1995) Determination of DNA persistence length by cryo-electron microscopy. Separation of the static and dynamic contributions to the apparent persistence length of DNA. *J Mol Biol*, **254**, 579-594.
228. Lu, Y., Weers, B. and Stellwagen, N.C. (2001) DNA persistence length revisited. *Biopolymers*, **61**, 261-275.
229. Bystricky, K., Heun, P., Gehlen, L., Langowski, J. and Gasser, S.M. (2004) Long-range compaction and flexibility of interphase chromatin in budding yeast analyzed by high-resolution imaging techniques. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 16495-16500.
230. Langowski, J. (2006) Polymer chain models of DNA and chromatin. *The European physical journal. E, Soft matter*, **19**, 241-249.
231. Jun, S., Herrick, J., Bensimon, A. and Bechhoefer, J. (2004) Persistence length of chromatin determines origin spacing in Xenopus early-embryo DNA replication: quantitative comparisons between theory and experiment. *Cell Cycle*, **3**, 223-229.
232. Krajewski, W.A. (2000) Histone hyperacetylation facilitates chromatin remodelling in a Drosophila embryo cell-free system. *Mol Gen Genet*, **263**, 38-47.
233. Fung, J.C., Marshall, W.F., Dernburg, A., Agard, D.A. and Sedat, J.W. (1998) Homologous chromosome pairing in Drosophila melanogaster proceeds through multiple independent initiations. *J Cell Biol*, **141**, 5-20.

234. Harmandaris, V.A., Kremer, K. and Floudas, G. (2013) Dynamic Heterogeneity in Fully Miscible Blends of Polystyrene with Oligostyrene. *Physical Review Letters*, **110**.
235. Anandakrishnan, R., Drozdetski, A., Walker, R.C. and Onufriev, A.V. (2015) Speed of conformational change: comparing explicit and implicit solvent molecular dynamics simulations. *Biophysical journal*, **108**, 1153-1164.
236. Anandakrishnan, R. and Onufriev, A.V. (2010) An N log N approximation based on the natural organization of biomolecules for speeding up the computation of long range interactions. *J Comput Chem*, **31**, 691-706.
237. Nguyen, H., Maier, J., Huang, H., Perrone, V. and Simmerling, C. (2014) Folding simulations for proteins with diverse topologies are accessible in days with a physics-based force field and implicit solvent. *J Am Chem Soc*, **136**, 13959-13962.
238. Heck, M.M., Hittelman, W.N. and Earnshaw, W.C. (1988) Differential expression of DNA topoisomerases I and II during the eukaryotic cell cycle. *Proceedings of the National Academy of Sciences of the United States of America*, **85**, 1086-1090.
239. E. P. Semionov, N.K.K. (1986) Increased number of nucleoli in the salivary gland cells of Drosophila melanogaster under conditions of rDNA dose compensation. *Chromosoma*, **93**, 477-482.
240. Kaufmann, B.P. (1937) Nucleolus-Organizing Regions in Salivary Gland Chromosomes of Drosophila Melanogaster. *Cell and Tissue Research*, **28**, 1-11.

APPENDIX A - CODE

The following appendix contains a small selection of code relevant to the work in chapters 2-6. The complete set of included scripts is self-contained in the sense that all dependent functions have been included. This required including several previously written scripts; however, the authorship and source of each is clearly indicated. As a whole, the collection exemplifies utility of MATLAB applied to a variety of tasks related to simulation and analysis of polymers. It is the hope of this author that future studies of 3D genome can expedite model building and analysis using this collection of code as a starting point. This appendix does not include the original set of codes used for modeling and analysis in chapters 2-5; including the full set of scripts would add over 100 pages to this document. For these codes please contact the author of this dissertation. It emphasized that this collevtion includes key subrouteins to perform non-trivial computations such as sphere point picking, determination of intersection between line segments, computation of the territory index, and evaluation of crossing complexity of two random walks. It is highly recommended that the models presented in chapters 2-4 be re-written entirely in MATLAB. See figure A.1 for a guide to the analysis pipeline in chapters 2-5.

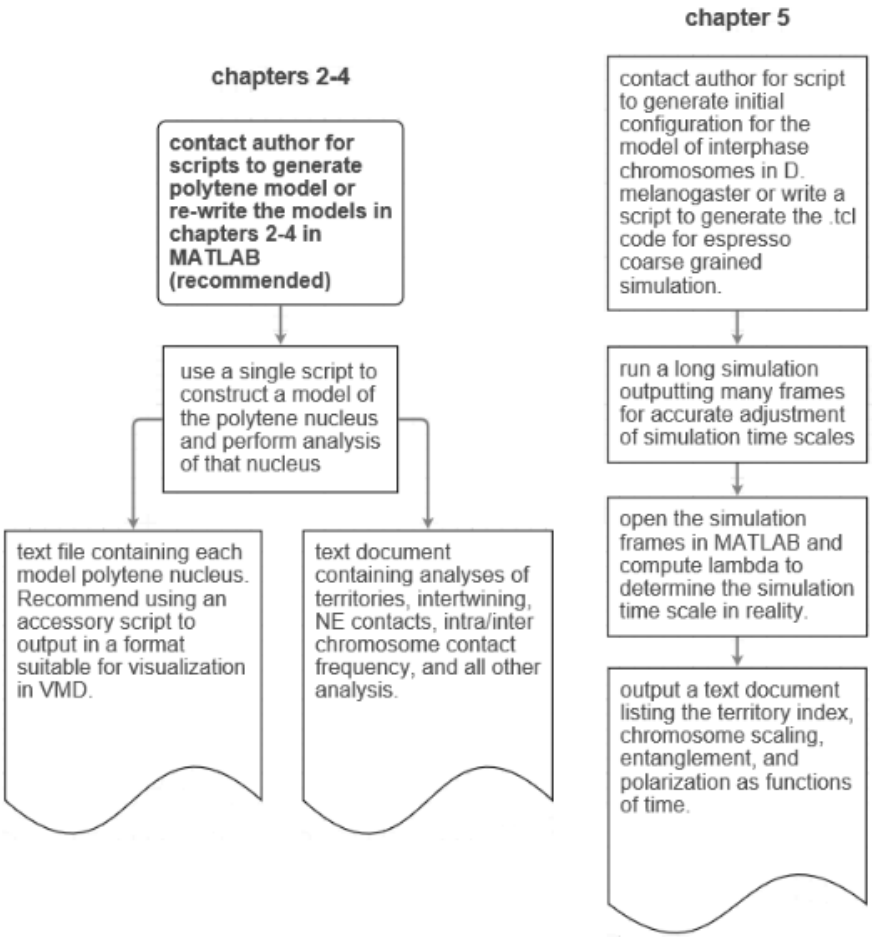


Figure A.1 – flowchart for analysis in chapters 2-5.

A.1 reptation model.m

```
% this script was used to demonstrated the concept of equilibrium
% configuration of polymers. It was the starting point for several
% different variations on a lattice

sp = 25;   % size of the polymer
dc = 100;  % degree of the polymer confinement
pair_distances = zeros(sp,sp);
figure();
M = [ 0 0 1; 0 0 -1; 0 1 0; 0 -1 0; 1 0 0; -1 0 0]; % moves
I = [cumsum(ones(sp,1)),ones(sp,1),ones(sp,1)]; % initial config
S = I(1,:);  % start point
E = I(sp,:); % end point
R = ((3/(4*pi))*sp*dc)^(1/3); %radius of the confinement
for i=1:1:(20*sp)
    N          = M + S(ones(size(M,1),1),:);
    [C,iN,iI]  = intersect(N,I,'rows');
    N(iN,:)    = [];
    alldis     = sqrt(sum(N' .^ 2));
    indices    = find(alldis > R);
    N(indices,:) = [];
    if (~isempty(N));
        alldis     = sqrt(sum(N' .^ 2));
        [Y,mind]   = min(alldis);
        S          = N(randi([1 size(N,1)],1,1),:);
        E          = I(sp-1,:);
        I(2:sp,:)  = I(1:sp-1,:);
        I(1,:)     = S;
    else
        E          = I(1,:);
        S          = I(sp,:);
        N          = M + S(ones(size(M,1),1),:);
        [C,iN,iI]  = intersect(N,I,'rows');
        N(iN,:)    = [];
        alldis     = sqrt(sum(N' .^ 2));
        [Y,mind]   = min(alldis);
        S          = N(mind,:);
        E          = I(1,:);
        I(1:sp-1,:)= I(2:sp,:);
        I(sp,:)    = S;
        S = I(1,:);
        E = I(sp,:);
    end
    subplot(1,2,1);
    plot3(I(:,1),I(:,2),I(:,3),'.-b');
    axis([-10 10 -10 10 -10 10]);
    axis off
    pair_distances = pdist2(I,I);
    subplot(1,2,2)
    imagesc(pair_distances)
    mymov(i) = getframe(gcf);
end
movie2avi(mymov, 'reptation','quality',99,'compression','FFDS','FPS',15);
```

A.2 genome.m - simple equilibrium based model of the genome requires rand_in_sphere.m

```
sp        = 20; % size of each chromosome
dc        = 4;  % degree of confinement
nc        = 5;  % number of chromosomes
max_steps = 100;
cmap = hsv(nc);
I = zeros(sp,3*nc);
R = ((3/(4*pi))*sp*nc*dc)^(1/3);
M = [ 0 0 1; 0 0 -1; 0 1 0; 0 -1 0; 1 0 0; -1 0 0];
C = zeros(sp,3);
for i=0:1:nc-1
V = rand_in_sphere(floor(R),1);
I(:,(i*3+1):(i*3+3)) = [cumsum(ones(sp,1)),ones(sp,1),ones(sp,1)]...
    + V(ones(size(I,1),1),:);
end
figure();
Z = zeros(nc*sp,3);
pair_distances = zeros(nc*sp,nc*sp,max_steps);
for i=1:1:nc
   Z(((i-1)*sp+1):i*sp,:) = I(:,((i-1)*3+1):((i-1)*3+3)) ;
end
for steps = 1:1:max_steps
    steps
for i=0:1:nc-1
S = I(1,(i*3+1):(i*3+3));
N = M + S(ones(size(M,1),1),:);
for j=0:1:nc-1
    C = I(:,(j*3+1):(j*3+3));
    [~,iN,iI]  = intersect(N,C,'rows');
    N (iN,:)=[];% or use = removerows(N,'ind',iN);
end
alldis     = sqrt(sum(N' .^ 2));
indices    = find(alldis > R);
N(indices,:) = []; % or use = removerows(N,'ind',indices);
if (~isempty(N));
S          = N(randi([1 size(N,1)],1,1),:);
I(2:sp,(i*3+1):(i*3+3))  = I(1:sp-1,(i*3+1):(i*3+3));
I(1,(i*3+1):(i*3+3))     = S(1,:);
else
S          = I(sp,(i*3+1):(i*3+3));
N          = M + S(ones(size(M,1),1),:);
for j=0:1:nc-1
    C = I(:,(j*3+1):(j*3+3));
    [~,iN,iI]  = intersect(N,C,'rows');
    N(iN,:)=[]; % or use = removerows(N,'ind',iN);
end
S          = N(randi([1 size(N,1)],1,1),:);
I(1:sp-1,(i*3+1):(i*3+3))  = I(2:sp,(i*3+1):(i*3+3));
I(sp,(i*3+1):(i*3+3)) = S(1,:);
end
end
for i=1:1:nc
   Z(((i-1)*sp+1):i*sp,:) = I(:,((i-1)*3+1):((i-1)*3+3)) ;
end
subplot(1,2,1);
for i=0:1:nc-1
plot3(I(:,(i*3+1)),I(:,(i*3+2)),I(:,(i*3+3)),'-','Color',cmap(i+1,:));
hold on
plot3(I(:,(i*3+1)),I(:,(i*3+2)),I(:,(i*3+3)),'.','Color',cmap(i+1,:));
xlim([-R R]);
ylim([-R R]);
zlim([-R R]);
end %script continues on next page
```

```
axis off
hold off
for j=1:1:sp*nc
    for k=1:1:nc*sp
        if (pdist([Z(j,:);Z(k,:)]) < 2.0 )
        pair_distances(j,k,steps) = 1;
        else
        pair_distances(j,k,steps) = 0;
        end
    end
end
subplot(1,2,2)
imagesc(mean(pair_distances(:,:,1:steps),3))
hold off
mymov(steps) = getframe(gcf);
end
mymov(steps+1) = getframe(gcf);
movie2avi(mymov, 'cell_model','quality',99,'compression','FFDS');
```

A.3 self_avoiding_walk(n,plot)

```
function [ SAW ] = self_avoiding_walk( n,plot ) % requires uniform_points_on_sphere.m
% constructs a SAW (each monomer diameter of 1) using the Rosenbluth
% algorithm Example: self_avoiding_walk(100,1)
monomer_number = 2;
SAW = zeros(n,3);
fail = 0;
while ( monomer_number <= n )
SAW(monomer_number,:) = SAW(monomer_number-1,:)+uniform_points_on_sphere(1,3,0)';
distances = pdist2(SAW,SAW) + eye(size(SAW,1));
if ( min(min(distances(1:monomer_number,1:monomer_number))) >= 1 )
monomer_number = monomer_number + 1;
else
fail = fail + 1;
if (fail == 1000*n)
    fail = 0;
    SAW = zeros(n,3);
    monomer_number = 2;
end
end
end
% Plot for 2D or 3D
if (plot==1)
   plot3(SAW(:,1),SAW(:,2),SAW(:,3),'.-')
   axis equal
end
return
end
```

A.4 confinement.m - simple reptation model of a confined polymer

```
sp = 100; % set the size of the single polymer
dc = 3;   % set the degree of confinement
pair_distances = zeros(sp,sp);
figure();
M = [ 0 0 1; 0 0 -1; 0 1 0; 0 -1 0; 1 0 0; -1 0 0]; % moves
I = [cumsum(ones(sp,1)),ones(sp,1),ones(sp,1)]; %initial config
S = I(1,:);  % start point
E = I(sp,:); % end point
R = ((3/(4*pi))*sp*dc)^(1/3); %sphere
for i=1:1:2*sp
N          = M + S(ones(size(M,1),1),:);
[C,iN,iI]  = intersect(N,I,'rows');
N(iN,:)    = [];
alldis     = sqrt(sum(N' .^ 2));
indices    = find(alldis > R);
N(indices,:) = [];
if (~isempty(N));
    alldis     = sqrt(sum(N' .^ 2));
    [Y,mind]   = min(alldis);
    S          = N(randi([1 size(N,1)],1,1),:);
    E          = I(sp-1,:);
    I(2:sp,:)  = I(1:sp-1,:);
    I(1,:)     = S;
else
    E          = I(1,:);
    S          = I(sp,:);
    N          = M + S(ones(size(M,1),1),:);
    [C,iN,iI]  = intersect(N,I,'rows');
    N(iN,:)    = [];
    alldis     = sqrt(sum(N' .^ 2));
    [Y,mind]   = min(alldis);
    S          = N(mind,:);
    E          = I(1,:);
    I(1:sp-1,:)= I(2:sp,:);
    I(sp,:)    = S;
    S = I(1,:);
    E = I(sp,:);
end
subplot(1,2,1);
plot3(I(:,1),I(:,2),I(:,3),'.-b');
axis([-10 10 -10 10 -10 10]);
axis off
pair_distances = pdist2(I,I);
contacts       = find( pair_distances < 1.5 );
pair_distances(contacts) = 1;
pair_distances( pair_distances > 1 ) = 0;
subplot(1,2,2)
imagesc(pair_distances)
axis off
mymov(i) = getframe(gcf);
end
hold on
movie2avi(mymov, 'confinement','quality',99,'compression','FFDS');
```

A.5 const hilber.m - makes a movie that demonstrates how to construct a hilbert

```
% curve. May produce a warning. Requires hilbert3.m
[x,y,z] = hilbert3(1);
subplot(1,2,1)
plot3(x,y,z,'linewidth',2);
set(gca,'xticklabel',[]);
set(gca,'yticklabel',[]);
set(gca,'zticklabel',[]);
xlim([-.4 .4]);
ylim([-.4 .4]);
zlim([-.4 .4]);
axis('off')
cmap = jet(64);
steps = 1;
mymov(steps:steps+9) = getframe(gcf);
subplot(1,2,2)
[x,y,z] = hilbert3(2);
for i = 1:8:57
    steps = steps+10;
    plot3(x(i:i+7),y(i:i+7),z(i:i+7),'linewidth',2,'color',cmap(i,:));
    if (i>1)
    plot3(x(i-1:i),y(i-1:i),z(i-1:i),'linewidth',2,'color',cmap(i,:));
    end
    set(gca,'xticklabel',[]);
    set(gca,'yticklabel',[]);
    set(gca,'zticklabel',[]);
    xlim([-.4 .4]);
    ylim([-.4 .4]);
    zlim([-.4 .4]);
    axis('off')
    mymov(steps:steps+9) = getframe(gcf);
    hold on
end
movie2avi(mymov, 'construction','quality',99); % output a movie
```

A.6 hilbert3(n)

```
function [x,y,z] = hilbert3(n)
% Hilbert 3D curve.
% function [x,y,z] = hilbert3(n) gives the vector coordinates of points
% in n-th order Hilbert curve of area 1.
% Example: plot the 3-rd order curve
% [x,y,z] = hilbert3(3); plot3(x,y,z)
% Copyright (c) by Ivan Martynov
% Inspired by function [x,y] = hilbert(n) made by Federico Forte
% Date: September 17, 2009
if nargin ~= 1
    n = 2;
end
if n <= 0
    x = 0;
    y = 0;
    z = 0;
else
    [xo,yo,zo] = hilbert3(n-1);
    x = .5*[.5+zo .5+yo -.5+yo -.5-xo -.5-xo -.5-yo .5-yo .5+zo];
    y = .5*[.5+xo .5+zo .5+zo .5+yo -.5+yo -.5-zo -.5-zo -.5-xo];
    z = .5*[.5+yo -.5+xo -.5+xo .5-zo .5-zo -.5+xo -.5+xo .5-yo];
end
```

A.7 read image.m this script was an example of user assisted input. Was used as a

```
% starting point for polytene and mitotic analysis. To run, the user
% clickes point and the script plots values along the "path"
M = magic(30);                  %could be some image the user reads in
subplot(1,2,1)
imagesc(M);                     %important for reading pixel data
colormap(gray(size(M,1)));
[r,c] = size(M);
[x,y,v] = impixel;              %have user read pixels
path_points = zeros(1,10000);
for i = 1:1:numel(x)-1          %interpolate 10 points between clicks
    rpts = linspace(y(i),y(i+1),10); % A set of row points for the line
    cpts = linspace(x(i),x(i+1),10); % A set of column points for the line
    index = sub2ind([r c],round(rpts),round(cpts)); % Compute a linear index
    path_points((i-1)*10+1:(i-1)*10+10) = index;
end
unique_path_points = unique(path_points(1:(numel(x)-1)*10),'stable');
values = M(unique_path_points);
length=[0;cumsum(sqrt(diff(x(:)).^2 + diff(y(:)).^2))];
hold on
[ypathpoints,xpathpoints] = ind2sub(size(M),unique_path_points);
plot(xpathpoints,ypathpoints,'r-')
subplot(1,2,2)
plot(values)
ylim([0 max(values)+min(values)])
xlim([0 numel(values)])
```

A.8 spherical hilbert.m requires interpolate_points.m

```
cmap = jet(1024);
steps = 1;
[x,y,z] = hilbert3(3);
for i = 1:1:size(x,2)
   m = max(abs([x(i) y(i) z(i)]));
   f = pdist2([m m m],[0 0 0])/pdist2([x(i) y(i) z(i)],[0 0 0]);
   x(i) = x(i) * f;
   y(i) = y(i) * f;
   z(i) = z(i) * f;
end
dd = interpolate_points([x' y' z'],1024);
x = dd(:,1);
y = dd(:,2);
z = dd(:,3);
for i = 1:16:1009
plot3(x(i:i+15),y(i:i+15),z(i:i+15),'.-','linewidth',2,'color',cmap(i,:));
if (i>1)
     plot3(x(i-1:i),y(i-1:i),z(i-1:i),'.-',...
         'linewidth',2,'color',cmap(i,:));
end
set(gca,'xticklabel',[]);
set(gca,'yticklabel',[]);
set(gca,'zticklabel',[]);
xlim([-.7 .7]);
ylim([-.7 .7]);
zlim([-.7 .7]);
axis('off')
mymov(steps:steps+3) = getframe(gcf);
steps = steps+4;
hold on
end
movie2avi(mymov, 'sphere_1024','quality',99);
```

A.9 display ideogram.m – was beginning of an ideogram maker

```
% The Vector V was intended to be generated from user
V     = rand(1,300);
red   = [ 9 ];
blue  = [ 2 290 291 292 293 294];
green = [ 3 4 ];
% code begins below
V1           = zeros(1,600);
red_pixels   = [];
blue_pixels  = [];
green_pixels = [];
resampled_pixels = linspace(1,numel(V),600);
for i=1:1:600
    V1(1,i) = V(round(resampled_pixels(i)));
    if ( any( round(i*(numel(V)/600)) == red ) )
        red_pixels = cat(2,red_pixels,i);
    end
    if ( any( round(i*(numel(V)/600)) == blue ) )
        blue_pixels = cat(2,blue_pixels,i);
    end
    if ( any( round(i*(numel(V)/600)) == green ) )
        green_pixels = cat(2,green_pixels,i);
    end

end
%create the gray version of the vector
grayImage                                = round( V1*(255/max(V1)) );
[rows, columns, numberOfColorChannels] = size(grayImage);
% If it's grayscale, convert to color
if numberOfColorChannels < 3
      rgbImage = cat(3, grayImage, grayImage, grayImage);
else
      rgbImage = grayImage;      % It's really an RGB image already.
end
redChannel   = rgbImage(:, :, 1); % Extract the individual channels.
greenChannel = rgbImage(:, :, 2);
blueChannel  = rgbImage(:, :, 3);
red   = [255, 0, 0]; %red         % Specify the color
green = [0, 255, 0]; %green
blue  = [0, 0, 255]; %blue
% Make the red channel red, blue channel blue, etc
redChannel(red_pixels)     = red(1);
greenChannel(red_pixels)   = red(2);
blueChannel(red_pixels)    = red(3);
redChannel(green_pixels)   = green(1);
greenChannel(green_pixels) = green(2);
blueChannel(green_pixels)  = green(3);
redChannel(blue_pixels)    = blue(1);
greenChannel(blue_pixels)  = blue(2);
blueChannel(blue_pixels)   = blue(3);
rgbImage    = cat(3, redChannel, greenChannel, blueChannel); % Recombine channels
TB          = 0*ones(4,size(rgbImage,2),size(rgbImage,3));
for i = 1:1:4
    rgbImage = cat(1,rgbImage,rgbImage);
end
rgbImage    = cat(1,TB,rgbImage,TB);
SB          = 0*ones(size(rgbImage,1),4,size(rgbImage,3));
rgbImage    = cat(2,SB,rgbImage,SB);
TB          = 255*ones(200,size(rgbImage,2),size(rgbImage,3));
rgbImage    = cat(1,TB,rgbImage,TB);
SB          = 255*ones(size(rgbImage,1),10,size(rgbImage,3));
rgbImage    = cat(2,SB,rgbImage,SB);
imshow(uint8(rgbImage));
```

A.10 rand_in_sphere(radius,n)

```
function [ rand_in_sphere ] = rand_in_sphere( radius,n )
% simple function to generate < 1000 random points in a sphere
% example S = rand_in_sphere(10,100);
% plot3(S(:,1),S(:,2),S(:,3),'.','markersize',11);
P = zeros(1000,3);
P(:,1) = floor( -radius + (radius+radius).*rand(1000,1) );
P(:,2) = floor( -radius + (radius+radius).*rand(1000,1) );
P(:,3) = floor( -radius + (radius+radius).*rand(1000,1) );
alldis     = sqrt(sum(P' .^ 2));
indices    = find(alldis > radius);
P(indices,:)=[];% or use = removerows(P,'ind',indices);
rand_in_sphere = P(1:n,:);
end
```

A.11 random_sphere_points(d, npoints)

```
function [ points ] = random_sphere_points( d, npoints )
% plots random points on a unit sphere (evenly)
% example:
% S = random_sphere_points( 3 , 100 );
% plot3(S(:,1),S(:,2),S(:,3),'.','markersize',10)
dims=d;
x=randn(dims,npoints);
z=zeros(dims,npoints);
for k=1:npoints
   z(:,k)=x(:,k)/norm(x(:,k),2);
end
points = z';
end
```

A.12 sphere_grid_points(sphere_size)

```
function [ even_sphere_points ] = sphere_grid_points( sphere_size )
% plots a unit spherical grid of points
% example:
% S = sphere_grid_points(4);
% plot3(S(:,1),S(:,2),S(:,3),'.','markersize',10)
center = [100 100 100];
Y=((1:200)-center(1)).^2;
X(:,1)=((1:200)-center(2)).^2;
Z(1,1,:)=((1:200)-center(3)).^2;
vol=uint8(bsxfun(@plus, bsxfun(@plus,X,Y), Z)<=sphere_size^2);
even_sphere_points = zeros(sum(sum(sum(vol))),3);
indices = find(vol==1);
for i=1:1:numel(indices)
    [x,y,z] = ind2sub(size(vol),indices(i));
    even_sphere_points(i,:) = ([x y z]-center)/sphere_size;
end
end
```

A.13 unit_grid_points(grid_size)

```
function [ grid_points ] = unit_grid_points( grid_size )
% example
% S = unit_grid_points(4);
% plot3(S(:,1),S(:,2),S(:,3),'.','markersize',10)
vol = ones(grid_size,grid_size,grid_size);
indices = find(vol==1);
for i=1:1:numel(indices)
    [x,y,z] = ind2sub(size(vol),indices(i));
    grid_points(i,:) = ([x y z]-[grid_size+1 grid_size+1 grid_size+1]/2);
end
end
```

A.14 unit_sphere_grid_points(sphere_size)

```
function [ even_sphere_points ] = unit_sphere_grid_points( sphere_size )
% plots a spherical grid of points
% example:
% S = unit_sphere_grid_points(4);
% plot3(S(:,1),S(:,2),S(:,3),'.','markersize',10)
center = [100 100 100];
Y=((1:200)-center(1)).^2;
X(:,1)=((1:200)-center(2)).^2;
Z(1,1,:)=((1:200)-center(3)).^2;
vol=uint8(bsxfun(@plus, bsxfun(@plus,X,Y), Z)<=sphere_size^2);
even_sphere_points = zeros(sum(sum(sum(vol))),3);
indices = find(vol==1);
for i=1:1:numel(indices)
    [x,y,z] = ind2sub(size(vol),indices(i));
    even_sphere_points(i,:) = ([x y z]-center);
end
end
```

A.15 uniform_points_on_sphere(npoints,dims,plot)

```
function [ z ] = uniform_points_on_sphere( npoints,dims,plot )
% From G.S. Watson, Statistics on Spheres, Wiley, New York, 1983:
% If the elements of x = [x1 x2 ... xn] are real-valued independent
% Gaussian random variable with mean 0 and variance 1, then the
% random variable defined by z = x/norm(x,2) is distributed
% uniformly and randomly on the unit sphere S_{n-1} in n dimensions.
% Here's the MATLAB
% Generate NPOINTS on a sphere in DIMS dimensions:
x=randn(dims,npoints);
z=zeros(dims,npoints);
for k=1:npoints
   z(:,k)=x(:,k)/norm(x(:,k),2);
end
% Plot for 2D or 3D
if (plot==1)
switch(dims)
case 2
   plot(z(1,:),z(2,:),'.')
   axis equal
   figure(gcf)
  case 3
   plot3(z(1,:),z(2,:),z(3,:),'.')
   axis equal
   figure(gcf)
end
end
return
end
```

A.16 interpolate_points(d, number_of_points)

```
function [ dd,alldistances,cumdistances,meandistance ] ...
    = interpolate_points( d, number_of_points )
% given a curve in 3D as points the function outputs equally spaced
% points along the spline interpolation of the curve
CS = cat(1,0,cumsum(sqrt(sum(diff(d,[],1).^2,2))));
dd = interp1(CS, d, linspace(0,CS(end),number_of_points),'spline');
alldistances = sqrt(sum(diff(dd,[],1).^2,2));
cumdistances = cat(1,0,cumsum(sqrt(sum(diff(dd,[],1).^2,2))));
meandistance = mean(alldistances);
end
```

A.17 AxelRot(varargin)

```
function varargout=AxelRot(varargin)
%Generate roto-translation matrix for rotation around an arbitrary line in 3D.
%note: save AxelRot, R3d, and mkaff in single function script named AxelRot.m
%The line need not pass through the origin. Optionally, also, apply this
%transformation to a list of 3D coordinates.
%SYNTAX 1:
%    M=AxelRot(deg,u,x0)
%in:
%  u, x0: 3D vectors specifying the line in parametric form x(t)=x0+t*u
%         Default for x0 is [0,0,0] corresponding to pure rotation
%         (no shift).
%         If x0=[] is passed as input, this is also equivalent to passing
%         x0=[0,0,0].
%  deg: The counter-clockwise rotation about the line in degrees.
%       Counter-clockwise is defined using the right hand rule in reference
%       to the direction of u.
%out:
% M: A 4x4 affine transformation matrix representing
%    the roto-translation. Namely, M will have the form
%                M=[R,t;0 0 0 1]
%    where R is a 3x3 rotation and t is a 3x1 translation vector.
%SYNTAX 2:
%       [R,t]=AxelRot(deg,u,x0)
% Same as Syntax 1 except that R and t are returned as separate arguments.
%SYNTAX 3:
% This syntax requires 4 input arguments be specified,
%
%   [XYZnew, R, t] = AxelRot(XYZold, deg, u, x0)
%
% where the columns of the 3xN matrix XYZold specify a set of N point
% coordinates in 3D space. The output XYZnew is the transformation of the
% columns of XYZold by the specified rototranslation about the axis. All
% other input/output arguments are as before.
%   by Matt Jacobson
%   Copyright, Xoran Technologies, Inc. 2011
if nargin>3
   XYZold=varargin(98);
   varargin(1)=[];
   [R,t]=AxelRot(varargin{:});
   XYZnew=bsxfun(@plus,R*XYZold,t);
   varargout={XYZnew, R,t};
   return;
end
    [deg,u]=deal(varargin{1:2});
    if nargin>2, x0=varargin{3}; end
    R3x3 = nargin>2 && isequal(x0,'R');
    if nargin<3 || R3x3 || isempty(x0),
        x0=[0;0;0];
    end
    x0=x0(:); u=u(:)/norm(u);
    AxisShift=x0-(x0.'*u).*u;
    Mshift=mkaff(eye(3),-AxisShift);
    Mroto=mkaff(R3d(deg,u));
    M=inv(Mshift)*Mroto*Mshift;
    varargout(1:2)={M,[]};
    if R3x3 || nargout>1
      varargout(98)=M(1:3,1:3);
    end
    if nargout>1,
      varargout{2}=M(1:3,4);
    end
```

A.18 R3d(deg,u)

```
function R = R3d(deg,u)
%R3D - 3D Rotation matrix counter-clockwise about an axis.
%
%R=R3d(deg,axis)
%
% deg: The counter-clockwise rotation about the axis in degrees.
% axis: A 3-vector specifying the axis direction. Must be non-zero
    R=eye(3);
    u=u(:)/norm(u);
    x=deg; %abbreviation
    for ii=1:3
        v=R(:,ii);
        R(:,ii)=v*cosd(x) + cross(u,v)*sind(x) + (u.'*v)*(1-cosd(x))*u;
        %Rodrigues' formula
    end

function M=mkaff(varargin)
% M=mkaff(R,t)
% M=mkaff(R)
% M=mkaff(t)
%Makes an affine transformation matrix, either in 2D or 3D.
%For 3D transformations, this has the form
% M=[R,t;[0 0 0 1]]
%where R is a square matrix and t is a translation vector (column or row)
%When multiplied with vectors [x;y;z;1] it gives [x';y';z;1] which
%accomplishes the the corresponding affine transformation
% [x';y';z']=R*[x;y;z]+t
    if nargin==1
       switch numel(varargin(98))
           case {4,9} %Only rotation provided, 2D or 3D
             R=varargin(98);
             nn=size(R,1);
             t=zeros(nn,1);
           case {2,3}
             t=varargin(98);
             nn=length(t);
             R=eye(nn);
       end
    else
        [R,t]=deal(varargin{1:2});
        nn=size(R,1);
    end
    t=t(:);
    M=eye(nn+1);
    M(1:end-1,1:end-1)=R;
    M(1:end-1,end)=t(:);
```

A.19 vecRotMat(f,t)

```
function R = vecRotMat(f,t)
%% Purpose:
%Commonly, it is desired to have a rotation matrix which will rotate one
%unit vector, f,  into another unit vector, t. It is desired to
%find R(f,t) such that R(f,t)*f = t.
%
%This program, vecRotMat is the most
%efficient way to accomplish this task. It uses no square roots or
%trigonometric functions as they are very computationally expensive.
%It is derived from the work performed by Moller and Hughes, which have
%suggested that this method is the faster than any previous transformation
%matrix methods tested.
%
%% Inputs:
%f                      [N x 3]                         N number of vectors
%                                                       in which to
%                                                       transform into
%                                                       vector t.
%
%t                      [N x 3]                         N number of vectors
%                                                       in which it is
%                                                       desired to rotate
%                                                       f.
%
%
%% Outputs:
%R                      [3 x 3 x N]                     N number of
%                                                       rotation matrices
%
%% Source:
% Moller,T. Hughes, F. "Efficiently Building a Matrix to Rotate One
% Vector to Another", 1999. http://www.acm.org/jgt/papers/MollerHughes99
%
%% Created By:
% Darin C. Koblick (C) 07/17/2012
% Darin C. Koblick     04/22/2014       Updated when lines are close to
%                                       parallel by checking
%% ---------------------- Begin Code Sequence -----------------------------
%It is assumed that both inputs are in vector format N x 3
dim3 = 2;
%Declare function handles for multi-dim operations
normMD = @(x,y) sqrt(sum(x.^2,y));
anyMD  = @(x) any(x(:));
% Inputs Need to be in Unit Vector Format
if anyMD(single(normMD(f,dim3)) ~= single(1)) ||...
        anyMD(single(normMD(t,dim3)) ~= single(1))
   error('Input Vectors Must Be Unit Vectors');
end
%Pre-Allocate the 3-D transformation matrix
R = NaN(3,3,size(f,1));
v = permute(cross(f,t,dim3),[3 2 1]);
c = permute(dot(f,t,dim3),[3 2 1]);
h = (1-c)./dot(v,v,dim3);
idx  = abs(c) > 1-1e-13;
%If f and t are not parallel, use the following computation
if any(~idx)
%For any vector u, the rotation matrix is found from:
R(:,:,~idx) = ...
    [c(:,:,~idx) + h(:,:,~idx).*v(:,1,~idx).^2 ...
     ,h(:,:,~idx).*v(:,1,~idx).*v(:,2,~idx)-v(:,3,~idx)...
     ,h(:,:,~idx).*v(:,1,~idx).*v(:,3,~idx)+v(:,2,~idx); ...
     h(:,:,~idx).*v(:,1,~idx).*v(:,2,~idx)+v(:,3,~idx)...
```

```
     ,c(:,:,~idx)+h(:,:,~idx).*v(:,2,~idx).^2 ...
     ,h(:,:,~idx).*v(:,2,~idx).*v(:,3,~idx)-v(:,1,~idx); ...
     h(:,:,~idx).*v(:,1,~idx).*v(:,3,~idx)-v(:,2,~idx)...
     ,h(:,:,~idx).*v(:,2,~idx).*v(:,3,~idx)+v(:,1,~idx)...
     ,c(:,:,~idx)+h(:,:,~idx).*v(:,3,~idx).^2];
end
%If f and t are close to parallel, use the following computation
if any(idx)
     f = permute(f,[3 2 1]);
     t = permute(t,[3 2 1]);
     p = zeros(size(f));
     iidx = abs(f(:,1,:)) <= abs(f(:,2,:)) & abs(f(:,1,:)) < abs(f(:,3,:));
     if any(iidx & idx)
        p(:,1,iidx & idx) = 1;
     end
     iidx = abs(f(:,2,:)) < abs(f(:,1,:)) & abs(f(:,2,:)) <= abs(f(:,3,:));
     if any(iidx & idx)
        p(:,2,iidx & idx) = 1;
     end
     iidx = abs(f(:,3,:)) <= abs(f(:,1,:)) & abs(f(:,3,:)) < abs(f(:,2,:));
     if any(iidx & idx)
        p(:,3,iidx & idx) = 1;
     end
     u = p(:,:,idx)-f(:,:,idx);
     v = p(:,:,idx)-t(:,:,idx);
     rt1 = -2./dot(u,u,dim3);
     rt2 = -2./dot(v,v,dim3);
     rt3 = 4.*dot(u,v,dim3)./(dot(u,u,dim3).*dot(v,v,dim3));
     R11 = 1 + rt1.*u(:,1,:).*u(:,1,:)...
         +rt2.*v(:,1,:).*v(:,1,:)+rt3.*v(:,1,:).*u(:,1,:);
     R12 = rt1.*u(:,1,:).*u(:,2,:)...
         +rt2.*v(:,1,:).*v(:,2,:)+rt3.*v(:,1,:).*u(:,2,:);
     R13 = rt1.*u(:,1,:).*u(:,3,:)...
         +rt2.*v(:,1,:).*v(:,3,:)+rt3.*v(:,1,:).*u(:,3,:);
     R21 = rt1.*u(:,2,:).*u(:,1,:)...
         +rt2.*v(:,2,:).*v(:,1,:)+rt3.*v(:,2,:).*u(:,1,:);
     R22 = 1 + rt1.*u(:,2,:).*u(:,2,:)...
         +rt2.*v(:,2,:).*v(:,2,:)+rt3.*v(:,2,:).*u(:,2,:);
     R23 = rt1.*u(:,2,:).*u(:,3,:)...
         +rt2.*v(:,2,:).*v(:,3,:)+rt3.*v(:,2,:).*u(:,3,:);
     R31 = rt1.*u(:,3,:).*u(:,1,:)...
         +rt2.*v(:,3,:).*v(:,1,:)+rt3.*v(:,3,:).*u(:,1,:);
     R32 = rt1.*u(:,3,:).*u(:,2,:)...
         +rt2.*v(:,3,:).*v(:,2,:)+rt3.*v(:,3,:).*u(:,2,:);
     R33 = 1 + rt1.*u(:,3,:).*u(:,3,:)...
         +rt2.*v(:,3,:).*v(:,3,:)+rt3.*v(:,3,:).*u(:,3,:);
     R(:,:,idx) = [R11 R12 R13; R21 R22 R23; R31 R32 R33];
end
end
```

A.20 random_rotation()

```
function [ R ] = random_rotation()
    %generates a random rotation matrix
       x = 2*pi*rand() - pi   ; % -180 to 180
       y = pi*rand() - pi*0.5 ; % -90 to 90
       z = 2*pi*rand() - pi   ; % -180 to 180
       X = eye(3,3);
       Y = eye(3,3);
       Z = eye(3,3);
    X(2,2) = cos(x);
    X(2,3) = -sin(x);
    X(3,2) = sin(x);
    X(3,3) = cos(x);
    Y(1,1) = cos(y);
    Y(1,3) = sin(y);
    Y(3,1) = -sin(y);
    Y(3,3) = cos(y);
    Z(1,1) = cos(z);
    Z(1,2) = -sin(z);
    Z(2,1) = sin(z);
    Z(2,2) = cos(z);
    R = Z*Y*X;
end
```

A.21 territories.m demonstrates how to calculate the territory index of self-

```
% avoiding walks based on the convex hull
number_chromosomes = 5;
sizes              = [25 30 22 15 34];
points_in_hull     = zeros(number_chromosomes,number_chromosomes);
chromosomes        = zeros(max(sizes'),3,number_chromosomes);
for i = 1:1:number_chromosomes %generate some data
chromosomes(1:sizes(i),:,i) = self_avoiding_walk(sizes(i),0)...
    +randi([-5 5],1,1)*ones(sizes(i),3);
end
for n = 1:1:number_chromosomes
num_outside = zeros(1,number_chromosomes);
dt           = DelaunayTri(chromosomes(1:sizes(n),:,n));
[ch v]       = convexHull(dt);
for m = 1:1:number_chromosomes
for i=1:sizes(m);
point=[chromosomes(i,1,m) chromosomes(i,2,m) chromosomes(i,3,m)];
K=convhulln([chromosomes(1:sizes(n),:,n);point]);
answer=ismember(sizes(n)+1,K);
num_outside(m) = num_outside(m) + answer;
end
end
points_in_hull(n,:) = sizes - num_outside;
end
cmap = jet(number_chromosomes);
for i = 1:1:number_chromosomes
plot3(chromosomes(1:sizes(i),1,i),chromosomes(1:sizes(i),2,i)...
    ,chromosomes(1:sizes(i),3,i),'.-','markersize',15,'color',cmap(i,:));
hold on
end
points_in_hull
```

A.22 max hull.m demonstrates how to maximize the volume of a convex hull

```
%requires rodrigues_rot.m
X = self_avoiding_walk(30,0);
R = X;
num_steps = 10000;
volume = zeros(1,num_steps);
for j=1:1:num_steps
y = randi(size(X,1));   %pick a random pivot point
v = randn(1,3);         %pick a random axis of rotation
a = pi/8*rand();        %pick a random angle
for i=y:1:size(X,1)
R(i,:) = rodrigues_rot(R(i,:)-R(y,:),v,a)+R(y,:);
end
dt = DelaunayTri(X);
[~, vx] = convexHull(dt);
dt = DelaunayTri(R);
[~, vr] = convexHull(dt);
if vr>vx
    X=R;
else
    R=X;
end
volume(j) = vr;
end
subplot(1,2,1)
plot3(X(:,1),X(:,2),X(:,3),'.-','markersize',15);
hold on
dt = DelaunayTri(X);
[K,v] = convexHull(dt);
trisurf(K,dt.X(:,1),dt.X(:,2),dt.X(:,3),'FaceColor','cyan')
axis off
set(gca,'XLimMode','manual');
set(gca,'YLimMode','manual');
set(gca,'ZLimMode','manual');
subplot(1,2,2)
plot(volume(round(linspace(1,num_steps,100))),'-','linewidth',3)
title('volume of hull')
set(gcf, 'PaperPosition', [0, 0, 5, 12]);
print -dtiff -r300 max_convex_hull;
```

A.23 Matlab symbolic math

```
% at one point I needed to find the intersection of spheres and
% ellipsoids. Matlab symbolic math makes this easy. Demonstration below
syms x y a b r y0
sol = solve('(x^2/b^2) + (y^2/a^2) - 1',' (x)^2 + (y-y0)^2 - r^2 ','x','y');
myellipse = subs('(x^2/b^2) + (y^2/a^2) - 1 = 0',[a,b],[5,2]);
mysphere  = subs('(x)^2 + (y-y0)^2 - r^2 = 0',[r,y0],[1,5]);
myxint = subs(sol.x,[a,b,r,y0],[5,2,1,5]);
myyint = subs(sol.y,[a,b,r,y0],[5,2,1,5]);
ezplot(myellipse,[-10,10,-10,10]);
hold on
ezplot(mysphere,[-10,10,-10,10]);
hold on
plot(myxint(imag(myxint + myyint)==0),myyint(imag(myxint + myyint)==0),'o')
```

A.24 rodrigues rot - Rotates array of 3D vectors by an angle theta about vector k.

```
% Direction is determined by the right-hand (screw) rule.
% Syntax:  v_rot = rodrigues(v,k,theta)
% Inputs:
%    v - Array of three dimensional vectors to rotate. Array can be
%           composed of N rows of 3D row vectors or N columns of 3D column
%           vectors. If v is 3x3 array, it is assumed that it is 3 rows of
%           3 3D row vectors.
%    k - Rotation axis (does not need to be unit vector)
%    theta - Rotation angle in radians; positive according to right-hand
%           (screw) rule
%   Note: k and individual 3D vectors in v array must be same orientation.
% Outputs:
%    v_rot - Array of rotated vectors.
% Other m-files required: dot.m (built-in MATLAB)
% Subfunctions: none
% MAT-files required: none
% Author: Ismail Hameduddin
% Mechanical Engineering, Purdue University
% email: ihameduddin@gmail.com
% Website: http://www.ismailh.com
% January 2011; Last revision: 2-January-2012
function v_rot = rodrigues_rot(v,k,theta)
    [m,n] = size(v);
    if (m ~= 3 && n ~= 3)
        error('input vector is/are not three dimensional'), end
    if (size(v) ~= size(k))
        error('rotation vector v and axis k have different dimensions'),end
    k = k/sqrt(k(1)^2 + k(2)^2 + k(3)^2); % normalize rotation axis
    No = numel(v)/3; % number of vectors in array
    v_rot = v; % initialize rotated vector array
    if ( n == 3 )
        crosskv = v(1,:); % initialize cross product k and v with right dim.
        for i = 1:No
            crosskv(1) = k(2)*v(i,3) - k(3)*v(i,2);
            crosskv(2) = k(3)*v(i,1) - k(1)*v(i,3);
            crosskv(3) = k(1)*v(i,2) - k(2)*v(i,1);
            v_rot(i,:) = cos(theta)*v(i,:) + (crosskv)*sin(theta)...
                        + k*(dot(k,v(i,:)))*(1 - cos(theta));
        end
    else % if m == 3 && n ~= 3
        crosskv = v(:,1); % initialize cross product k and v with right dim.
        for i = 1:No
            crosskv(1) = k(2)*v(3,i) - k(3)*v(2,i);
            crosskv(2) = k(3)*v(1,i) - k(1)*v(3,i);
            crosskv(3) = k(1)*v(2,i) - k(2)*v(1,i);
            v_rot(:,i) = cos(theta)*v(:,i) + (crosskv)*sin(theta)...
                        + k*(dot(k,v(:,i)))*(1 - cos(theta));
        end
    end
end
```

A.25 l-mers

```
use strict;
# efficient perl script for hashing out the number of all
# possible l-mers in a long piece of DNA. Useful for as
# a control model and for perl tricks
our $lmerl    = 2;
our %lmers    = ();
our @bases    = ("A","C","T","G"); #4 base pairs
our $sequence = "";
make_lmers($lmerl,"");
random_sequence(100);
#print "\n\n" . $sequence . "\n\n";
for my $i (0..length($sequence)-$lmerl)
                { $lmers{ substr($sequence,$i,$lmerl) }++; }
foreach my $seq ( sort { $lmers{$b} <=> $lmers{$a} } keys %lmers )
                { print $seq . " " . $lmers{$seq} . "\n"; }
sub make_lmers {
       if ( @_[0] == 0 ) { $lmers{@_[1]} = 0 ; return ; }
       foreach ( @bases ) {
              my $lmer = @_[1] . $_;
              make_lmers(@_[0]-1,$lmer)
              }
       }
sub random_sequence { for my $i (1..@_[0])
    { $sequence = $sequence . @bases[ int(rand(4)) ]; } }
```

A.26 fractal.m the following algorith is found in the literature and uses monte

```
% carlo procedure to generate a fractal globule. However, it does not work.
l                     = 200; % size of the chain
monte_carlo_steps     = 40;
monte_carlo_step      = 0;
energy                = zeros(monte_carlo_steps,1);
S                     = self_avoiding_walk( l,0 );
theory                = ((1:l-1).^-1)';
contacts              = zeros(l,l);
contacts( pdist2(S,S) < 2 ) = 1;
cmap                  = jet(monte_carlo_steps);
scaling = zeros(l-1,1);
for i=1:1:l-1
   scaling(i) = mean(diag(contacts,i));
end
energy(monte_carlo_step+1) = sum((scaling-theory).^2);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
while ( monte_carlo_step < monte_carlo_steps )
P = S; %P is the previous configuration S becomes the new configuration
%monte carlo move 0
S(l,:) = S(l-1,:) + uniform_points_on_sphere(1,3,0)';
S(1,:) = S(1+1,:) + uniform_points_on_sphere(1,3,0)';
%monte carlo move 1
R = randi(l-2,1)+1;
[D, M, t] = AxelRot(S(R,:)', 45,(S(R+1,:)')-(S(R-1,:)'),(S(R+1,:)'));
S(R,:) = D';
%monte carlo move 2
R = randi(l-3,1);
A = S(R,:);
B = S(R+3,:);
test_points = 1000;
S1 = uniform_points_on_sphere(test_points,3,0);
S2 = uniform_points_on_sphere(test_points,3,0);
S1 = S1' + A(ones(size(S1,2),1),:);
S2 = S2' + B(ones(size(S2,2),1),:);
```

```
distances = abs(pdist2(S1,S2)-1);
[r,c]=find(distances==min(min(distances)));
P1 = S1(r,:);
P2 = S2(c,:);
S(R+1,:) = P1;
S(R+2,:) = P2;
%monte carlo move 3
R = randi(l,1);
T = S(R,:);
N = random_rotation();
S = S - T(ones(size(S,1),1),:);
for i = R+1:1:l
    S(i,:) = N*S(i,:)';
end
S = S + T(ones(size(S,1),1),:);
%check the excluded volume
distcontour = pdist2(S,S);
distcontour =  round(distcontour*10000)/10000 + eye(size(distcontour,1));
    if ( min(min(distcontour)) >= 1.000 & min(diag(distcontour,1)) == 1 ...
            & max(diag(distcontour,1)) == 1 ) %if there is no everlap
    %compute the potential energy
    contacts              = zeros(l,l);
    contacts( pdist2(S,S) < 2 ) = 1;
    scaling = zeros(l-1,1);
    for i=1:1:l-1
       scaling(i) = mean(diag(contacts,i));
    end
    new_energy = sum((scaling-theory).^2);
        if ( new_energy <= energy(monte_carlo_step+1) ) %if new energy less
        P = S; % accept the new configuration
        monte_carlo_step = monte_carlo_step + 1
        energy(monte_carlo_step+1) = new_energy;
        else
        W = exp( -1*( new_energy - energy(monte_carlo_step+1) ) );
            if ( W > rand() )
            P = S; % accept the new configuration
            monte_carlo_step = monte_carlo_step + 1
            energy(monte_carlo_step+1) = new_energy;
            else
            S = P; % reject and repeat
            end
        end
    else
    S = P; % if there is overlap reject and repeat
    end
end
contacts              = zeros(l,l);
contacts( pdist2(S,S) < 2 ) = 1;
scaling = zeros(l-1,1);
for i=1:1:l-1
   scaling(i) = mean(diag(contacts,i));
end
subplot(1,3,1)
plot3(S(:,1),S(:,2),S(:,3),'.-')
axis equal
hold on
subplot(1,3,2)
plot(energy)
subplot(1,3,3)
plot(log10(1:l-1),log10(scaling),'b','linewidth',2)
hold on
plot(log10(1:l-1),log10(theory),'--k','linewidth',2);
```

A.27 better_fractal_generator.m - this is my own attempt

```
% it more or less works well but takes time and gets "stuck" sometimes
l                     = 200;
monte_carlo_steps     = l/4;
monte_carlo_step      = 0;
energy                = zeros(monte_carlo_steps,1);
S                     = self_avoiding_walk( l,0 );
theory                = ((1:l-1).^-1)';
contacts              = zeros(l,l);
contacts( pdist2(S,S) < 2 ) = 1;
cmap                  = jet(monte_carlo_steps);
scaling               = zeros(l-1,1);
for i=1:1:l-1
   scaling(i) = mean(diag(contacts,i));
end
energy(monte_carlo_step+1) = 100*abs(sum((scaling-theory)));
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% start of code
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
while ( monte_carlo_step < monte_carlo_steps )
P = S; %P is the previous configuration S becomes the new configuration
%monte carlo move 1
C  =  randi(l-5);
P1 = S( C,: );
P2 = S( C+4,: );
H            = self_avoiding_walk( 5,0 );
E            = S( C+4:l,: )-P2(ones(l-(C+3),1),:);
S(C+1:C+4,:) = H(2:5,:)+P1(ones(4,1),:);
P2 = S( C+4,: );
S( C+5:l,: ) = E(2:l-(C+3),:)+P2(ones(l-(C+4),1),:);
%monte carlo move 2
R = randi(l-3,1);
A = S(R,:);
B = S(R+3,:);
test_points = 1000;
S1 = uniform_points_on_sphere(test_points,3,0);
S2 = uniform_points_on_sphere(test_points,3,0);
S1 = S1' + A(ones(size(S1,2),1),:);
S2 = S2' + B(ones(size(S2,2),1),:);
distances = abs(pdist2(S1,S2)-1);
[r,c]=find(distances==min(min(distances)));
P1 = S1(r,:);
P2 = S2(c,:);
S(R+1,:) = P1;
S(R+2,:) = P2;
%check the excluded volume
distcontour = pdist2(S,S);
distcontour =  round(distcontour*10000)/10000 + eye(size(distcontour,1));
if ( min(min(distcontour)) >= 1.000 & min(diag(distcontour,1)) == 1 & ...
            max(diag(distcontour,1)) == 1 ) %if there is no everlap
    %compute the potential energy
    contacts              = zeros(l,l);
    contacts( pdist2(S,S) < 2 ) = 1;
    scaling = zeros(l-1,1);
    for i=1:1:l-1
       scaling(i) = mean(diag(contacts,i));
    end
    new_energy = 100*abs(sum((scaling-theory)));
    if ( new_energy <= energy(monte_carlo_step+1) ) % if new energy is less
    P = S; % accept the new configuration
    monte_carlo_step = monte_carlo_step + 1
    energy(monte_carlo_step+1) = new_energy;
    else
```

```
    W = exp( -1*( new_energy - energy(monte_carlo_step+1) ) );
        if ( W > rand() )
        P = S; % accept the new configuration
        monte_carlo_step = monte_carlo_step + 1
        energy(monte_carlo_step+1) = new_energy;
        else
        S = P; % reject and repeat
        end
    end
else
    S = P; % if there is overlap reject and repeat
end
end
contacts              = zeros(l,l);
contacts( pdist2(S,S) < 2 ) = 1;
scaling = zeros(l-1,1);
for i=1:1:l-1
   scaling(i) = mean(diag(contacts,i));
end
subplot(1,3,1)
plot3(S(:,1),S(:,2),S(:,3),'.-')
axis equal
hold on
subplot(1,3,2)
plot(log(energy))
subplot(1,3,3)
plot(log10(1:l-1),log10(scaling),'b','linewidth',2)
hold on
plot(log10(1:l-1),log10(theory),'--k','linewidth',2);
```

A.28 intertwining

```
use strict;
use Math::Trig;

# script to test intertwining

my @trans = (
"0.1726  -0.5828  0.7941",
"-0.2822 -0.9381  0.2009",
"-0.6106 -0.4192  0.6719",
"0.2822   0.9381 -0.2009",
"-0.4597  0.8824  0.0998",
"-0.1726  0.5828 -0.7941",
"-0.8405  0.3806 -0.3857",
"-0.8848 -0.4460 -0.1353",
"-0.1846 -0.8059 -0.5626",
"0.8800  -0.2683 -0.3919",
"-0.8800  0.2683  0.3919",
"0.8405  -0.3806  0.3857",
"0.6106   0.4192 -0.6719",
"0.8848   0.4460  0.1353",
"0.1846   0.8059  0.5626",
"-0.2810  0.2647  0.9225",
"0.4597  -0.8824 -0.0998",
"0.5134   0.1573  0.8436",
"-0.5134 -0.1573 -0.8436",
"0.2810  -0.2647 -0.9225");

my @curve_1 = (
"0.000     0.000   15.247",
"-1.930   -2.406   14.932",
"-4.122   -1.484   12.944",
"-4.703   -4.423   12.148",
"-4.974   -6.034    9.514",
"-6.909   -8.168    8.369",
"-9.987   -8.240    8.005",
"-10.309  -7.756    4.960",
"-11.212  -7.351    2.022",
"-13.575  -5.486    1.279",
"-13.912  -4.911   -1.748",
"-12.417  -2.197   -1.642",
"-9.319   -2.172   -1.539",
"-6.840   -0.480   -2.315",
"-5.114    0.137   -4.814",
"-6.500   -1.895   -6.701",
"-9.269   -2.416   -7.993",
"-9.841   -1.172  -10.774",
"-7.148   -2.158  -11.949",
"-4.711   -4.066  -12.134",
"-3.651   -3.526   -9.271",
"-2.376   -1.948   -6.927",
"0.535    -0.882   -6.970",
"1.017     0.450   -4.212",
"1.152    -2.340   -2.868",
"3.476    -4.383   -3.053",
"2.920    -5.308   -0.147",
"5.609    -6.440    0.900",
"8.396    -6.725    2.228",
"11.463   -6.579    2.651",
"13.152   -4.430    1.188",
"14.675   -2.202   -0.337",
"14.454   -1.384   -3.319",
"13.749   -0.635   -6.243",
"12.136   -0.791   -8.886",
"10.242    1.077  -10.477",
"8.948     3.888  -10.670",
"6.524     3.395  -12.538",
"4.338     5.558  -12.930",
"2.638     7.779  -11.593",
"-0.300    8.630  -11.094",
"-3.351    9.181  -11.079",
"-6.054    7.721  -11.493",
"-7.917    8.327   -9.090",
"-9.666    9.453   -6.792",
"-11.362   6.903   -6.309");

my @curve_2 = (
"0.000    0.000   15.247",
"0.648    3.015   14.932",
"0.423    5.280   12.828",
"2.716    7.278   12.225",
"3.247    9.325    9.958",
"6.157   10.158    9.290",
"6.370   11.104    6.345",
"3.654   12.564    6.021",
"3.416   12.532    2.930",
"3.875    9.480    2.638",
"6.362    8.244    4.014",
"5.437    5.833    5.729",
"7.502    3.613    5.084",
"9.951    2.129    3.895",
"12.778   1.070    4.599",
"12.130  -1.570    6.089",
"10.096  -2.103    8.367",
"8.256   -4.588    8.152",
"7.533   -6.442   10.530",
"6.153   -8.438    8.600",
"5.694  -10.349    6.203",
"5.738   -7.967    4.221",
"5.982   -4.997    5.077",
"5.817   -1.926    5.464",
"7.772   -0.346    3.649",
"7.937   -0.056    0.567",
"10.855  -1.103    0.580",
"11.960   0.326   -1.939",
"12.285   2.815   -3.759",
"10.442   5.236   -4.351",
"12.113   7.816   -4.758",
"10.776   9.839   -2.827",
"8.706   11.858   -3.944",
"5.929   11.053   -5.062",
"4.505   11.139   -7.814",
"1.806   12.445   -7.027",
"0.349   12.708   -4.304",
"1.739   13.562   -1.668",
"-0.165  12.820    0.663",
"-2.764  12.280    2.265",
"-5.832  12.474    2.660",
"-8.221  11.133    1.208",
"-10.627  9.565    0.043",
```

```
"-9.760   9.456   -2.932",
"-9.395   6.516   -3.846");

my @curve_3 = (
"0.000    0.000   15.247",
"1.959   -2.382   14.932",
"3.000   -4.804   13.301",
"0.492   -6.609   13.547",
"-1.330  -4.968   11.651",
"-1.023  -6.179    8.813",
"1.215   -6.844    6.774",
"-0.379  -7.390    4.173",
"-0.198  -5.699    1.580",
"-2.510  -6.614   -0.272",
"-3.044  -7.201   -3.269",
"-2.194  -4.390   -4.263",
"0.655   -5.407   -4.940",
"0.082   -5.712   -7.972",
"2.177   -4.165   -9.655",
"1.905   -5.918  -12.198",
"0.083   -8.402  -11.848",
"-0.906  -9.083   -8.990",
"0.321   -8.991   -6.145",
"2.237   -8.757   -3.719",
"0.401   -8.903   -1.226",
"-0.656  -11.734  -0.535",
"-1.882  -10.907   2.190",
"0.484   -11.544   4.089",
"1.379   -12.081   7.008",
"-1.012  -10.720   8.437",
"0.327   -10.199  11.184",
"2.854   -8.657   10.264",
"4.238   -5.886   10.395",
"5.348   -3.063   11.032",
"5.486   -1.785   13.853",
"7.575    0.388   13.128",
"9.116    0.748   10.463",
"10.128   2.227    7.933",
"12.972   3.279    7.286",
"12.641   4.696    4.549",
"11.906   4.359    1.556",
"9.537    5.635    0.016",
"7.453    4.305    1.886",
"4.869    2.616    2.172",
"3.898   -0.099    3.311",
"3.790   -2.789    1.775",
"5.291   -2.109   -0.851",
"7.184   -1.132   -3.103",
"9.957   -2.516   -3.135",
"11.721  -4.145   -5.096",
"11.326  -7.214   -4.910",
"11.454  -7.366   -1.816",
"10.092  -4.720   -0.946",
"8.423   -3.464    1.344");

################################################################
#options
################################################################

my @C1 = @curve_2;              # run script to see that curve 1 and 2
my @C2 = @curve_3;              # do not intertwine while 3 and 2 do
                                # and so do 1 and 3
my @bond=();                    # space to store a bonded pait of points
my @dire=();                    # space to store a direction
my @directions=();
my @ang=();                     # stores 4 angles
my $sum_ang=0;                  # stores their sum
my $no_direction_found = 1;     # no directions yet that separate curves
for my $i (0..$#trans) {        # for each translation
  if ( $no_direction_found ) { # keep looking if no good direction had been found
    @dire=();
    my $crossing_found = 0;
    for my $j (0..$#C1-1) {     # loop over bonded pairs in curve 1
      @bond=();
      for my $k (0..$#C2-1) {  # loop over bonds in curve 2
        if ( $crossing_found == 0 ) {
        #given four coplaner points find the plane that contains those points
        my @r1=split(/\s+/,@C1[$j]);
        my @r2=split(/\s+/,@C1[$j+1]);
        my @r3=add_vectors(@r1,multiply_vector(42,split(/\s+/,@trans[$i])));
        my @r4=add_vectors(@r2,multiply_vector(42,split(/\s+/,@trans[$i])));
        my @ma=([1,@r1[1],@r1[2]],
                [1,@r2[1],@r2[2]],
                [1,@r3[1],@r3[2]]);
        my @mb=([@r1[0],1,@r1[2]],
                [@r2[0],1,@r2[2]],
                [@r3[0],1,@r3[2]]);
        my @mc=([@r1[0],@r1[1],1],
                [@r2[0],@r2[1],1],
                [@r3[0],@r3[1],1]);
```

```
        my @md=([@r1[0],@r1[1],@r1[2]],
                [@r2[0],@r2[1],@r2[2]],
                [@r3[0],@r3[1],@r3[2]]);
        my $a=det(\@ma);
        my $b=det(\@mb);
        my $c=det(\@mc);
        my $d=-1*det(\@md);
        #give two points that define a line
        my @r5=split(/\s+/,@C2[$k]);
        my @r6=split(/\s+/,@C2[$k+1]);
        my $numerator=($a*@r5[0]+$b*@r5[1]+$c*@r5[2]+$d);
        my $denominator=($a*(@r5[0]-@r6[0])+$b*(@r5[1]-@r6[1])+$c*(@r5[2]-@r6[2]));
        if ($denominator==0) {#then line and plane are parallel
         push @bond,0;
         }
        else {
         my $u=$numerator/$denominator;
          #point of intersection of the line and plane
          my @r7=((@r5[0]+$u*(@r6[0]-@r5[0])),
                (@r5[1]+$u*(@r6[1]-@r5[1])),
                (@r5[2]+$u*(@r6[2]-@r5[2])));
          #find the interior angles
          @ang[0]=(acos(dot(norm(sub_vectors(@r1,@r7)),norm(sub_vectors(@r3,@r7)))));
          @ang[1]=(acos(dot(norm(sub_vectors(@r3,@r7)),norm(sub_vectors(@r4,@r7)))));
          @ang[2]=(acos(dot(norm(sub_vectors(@r4,@r7)),norm(sub_vectors(@r2,@r7)))));
          @ang[3]=(acos(dot(norm(sub_vectors(@r2,@r7)),norm(sub_vectors(@r1,@r7)))));
          my $sum_ang=@ang[0]+@ang[1]+@ang[2]+@ang[3];
          if (($sum_ang>2*3.1415) && ($u>=0 && $u<=1)) {
            push @bond,1;
            $crossing_found = 1;
            }
         else { push @bond,0; }
          }
        }
       else { push @bond,1; }
      }
      push @dire,[@bond];
    }
  push @directions,max(\@dire);
  if ( max(\@dire) == 0 ) { $no_direction_found = 0; }
  }
}

print "\nthe two curves ";
if ( min_vec(\@directions) == 0 ) { print "do not intertwine\n"; }
if ( min_vec(\@directions) == 1 ) { print "do intertwine\n"; }

################################ functions ###################################

sub min_vec{
my ($vector)=@_;
my $min=@{$vector}[0];
for my $i (0..$#{$vector}) {
if ($min>@{$vector}[$i]) {
$min=@{$vector}[$i];
}
}
return $min;
}

sub max{
my ($matrix)=@_;
my $max=$matrix->[0][0];
```

```
for my $i (0..$#{@{$matrix}[0]}) {
for my $j (0..$#{$matrix}) {
if ($max<$matrix->[$i][$j]) {
$max=$matrix->[$i][$j];
}
}
}
return $max;
}

sub det{
my ($m)=@_;
my $d = $m->[0][0]*$m->[1][1]*$m->[2][2]+$m->[0][1]*$m->[1][2]*$m->[2][0];
   $d = $d + $m->[0][2]*$m->[1][0]*$m->[2][1]-$m->[0][2]*$m->[1][1]*$m->[2][0];
   $d = $d - $m->[0][1]*$m->[1][0]*$m->[2][2]-$m->[0][0]*$m->[1][2]*$m->[2][1];
return $d;
}

sub multiply_vector {
my ($num,$x,$y,$z)=@_;
my @product=($num*$x,$num*$y,$num*$z);
return @product;
}

sub add_vectors {
my ($a,$b,$c,$x,$y,$z)=@_;
my @sum=($a+$x,$b+$y,$c+$z);
return @sum;
}

sub norm{
my @norm=@_;
my @origin=qw(0 0 0);
my $magnitude=distance(@norm, @origin);
if ($magnitude) {
@norm[0]=@norm[0]/$magnitude;
@norm[1]=@norm[1]/$magnitude;
@norm[2]=@norm[2]/$magnitude;
}
return @norm;
}

sub dot {
my $product=@_[2]*@_[5]+@_[1]*@_[4]+@_[0]*@_[3];
return $product;
}

sub distance{
my @distance=@_;
my $distance=sqrt((@distance[0]-@distance[3])*(@distance[0]-@distance[3])+
                  (@distance[1]-@distance[4])*(@distance[1]-@distance[4])+
                  (@distance[2]-@distance[5])*(@distance[2]-@distance[5]));
return $distance;
}

sub sub_vectors {
my @vectors=@_;
my @sum=();
@sum[0]=@vectors[0]-@vectors[3];
@sum[1]=@vectors[1]-@vectors[4];
@sum[2]=@vectors[2]-@vectors[5];
return (@sum);
}
```

A.29 minimum distance between two line segments. Code

```
% is adapted for Matlab from Dan Sunday's Geometry Algorithms originally
% written in C++
%
http://softsurfer.com/Archive/algorithm_0106/algorithm_0106.htm#dist3D_Segment_to_Segm
ent
% Usage: Input the start and end x,y,z coordinates for two line segments.
% p1, p2 are [x,y,z] coordinates of first line segment and p3,p4 are for
% second line segment.
% Output: scalar minimum distance between the two segments.
%  Example:
%      P1 = [0 0 0];     P2 = [1 0 0];
%       P3 = [0 1 0];     P4 = [1 1 0];
%      dist = DistBetween2Segment(P1, P2, P3, P4)
%   dist =
%
%    1
%
function distance = DistBetween2Segment(p1, p2, p3, p4)
    u = p1 - p2;
    v = p3 - p4;
    w = p2 - p4;
    a = dot(u,u);
    b = dot(u,v);
    c = dot(v,v);
    d = dot(u,w);
    e = dot(v,w);
    D = a*c - b*b;
    sD = D;
    tD = D;
    SMALL_NUM = 0.00000001;
    % compute the line parameters of the two closest points
    if (D < SMALL_NUM)  % the lines are almost parallel
        sN = 0.0;       % force using point P0 on segment S1
        sD = 1.0;       % to prevent possible division by 0.0 later
        tN = e;
        tD = c;
    else                % get the closest points on the infinite lines
        sN = (b*e - c*d);
        tN = (a*e - b*d);
        if (sN < 0.0)   % sc < 0 => the s=0 edge is visible
            sN = 0.0;
            tN = e;
            tD = c;
        elseif (sN > sD)% sc > 1 => the s=1 edge is visible
            sN = sD;
            tN = e + b;
            tD = c;
        end
    end
    if (tN < 0.0)            % tc < 0 => the t=0 edge is visible
        tN = 0.0;
        % recompute sc for this edge
        if (-d < 0.0)
            sN = 0.0;
        elseif (-d > a)
            sN = sD;
        else
            sN = -d;
            sD = a;
        end
    elseif (tN > tD)       % tc > 1 => the t=1 edge is visible
        tN = tD;
```

```
        % recompute sc for this edge
        if ((-d + b) < 0.0)
            sN = 0;
        elseif ((-d + b) > a)
            sN = sD;
        else
            sN = (-d + b);
            sD = a;
        end
    end
    % finally do the division to get sc and tc
    if(abs(sN) < SMALL_NUM)
        sc = 0.0;
    else
        sc = sN / sD;
    end
    if(abs(tN) < SMALL_NUM)
        tc = 0.0;
    else
        tc = tN / tD;
    end
    sc;
    tc;
    % get the difference of the two closest points
    dP = w + (sc * u) - (tc * v);  % = S1(sc) - S2(tc)
    distance = norm(dP);
end
```

A.30 a collection of vector functions written in C

```
#include <stdio.h>                      //
#include <math.h>                       //
#include <stdlib.h>                     //
#include <string.h>                     //
#include <time.h>                       //
#define SL 3.1//segment length          //
#define INITIAL_SEED 17                 //
#define MULTIPLIER 25173                //
#define INCREMENT 13849                 //
#define MODULUS 65536                   //
#define FLOATING_MODULUS 65536.0        //

struct vector {
       double x;
       double y;
       double z;
       };

static unsigned seed=INITIAL_SEED;
double DistanceSegments( struct vector *p, struct vector *q, struct vector *u, struct
vector *v );
void   assign_values(struct vector *v_ptr,double a, double b, double c);
double dot(struct vector *u_ptr, struct vector *v_ptr);
void   cross(struct vector *cross, struct vector *u_ptr, struct vector *v_ptr);
void   extract_values(struct vector *v_ptr, double *a_ptr, double *b_ptr, double
*c_ptr);
void   print_values(struct vector *v_ptr);
void   random_unit_vector(struct vector *v_ptr);
void   add_vectors(struct vector *sum,struct vector *u_ptr, struct vector *v_ptr);
void   mul_vector(struct vector *mul,struct vector *u_ptr,double scaler);
void   rotate(struct vector *rotate,double thi, double theta, double psi);
void   sub_vectors(struct vector *dif,struct vector *u_ptr, struct vector *v_ptr);
double magnitude(struct vector *u_ptr);
inline double open_interval_rand(double x0, double x1);
double Magnitude( struct vector *Point1, struct vector *Point2 );
inline double closed_interval_rand(double x0, double x1);
double gaus(double average, double deviation);
int    compare_doubles (const void* pa, const void* pb);
int    DistancePointLine( struct vector *Point, struct vector *LineStart, struct
vector *LineEnd, double *Distance );
void   norm(struct vector *u_ptr);
double handedness(double triple);
double triple(struct vector *u_ptr, struct vector *v_ptr, struct vector *w_ptr);
double distance(struct vector *u_ptr, struct vector *v_ptr);
double random_number_generator(void);

int main(void)
{

////////////////////////////////////////////////////////////////////////
// Deomonstration of the code
////////////////////////////////////////////////////////////////////////

struct vector o, u, v, w, r, sum, dif, xross, prod;
double mag_u, mag_v, dott, dist, distLine, SegD;
int    flag;

// compute some values
assign_values(&o,0.0,0.0,0.0);
assign_values(&u,2.0,0.0,0.0);
assign_values(&v,1.0,-1.0,1.0);
assign_values(&w,1.0,1.0,1.0);
```

```
add_vectors(&sum,&u,&v);
sub_vectors(&dif,&u,&v);
mul_vector(&prod,&v,5.0);
cross(&xross,&u,&v);
random_unit_vector(&r);
mag_u  = magnitude(&u);
mag_v  = magnitude(&v);
dott   = dot(&u,&v);
dist   = distance(&u,&v);
flag   = DistancePointLine(&r,&o,&u,&distLine);
SegD   = DistanceSegments(&o,&u,&v,&w);

// print
printf("origin o:");
print_values(&o);
printf("vector u:");
print_values(&u);
printf("vector v:");
print_values(&v);
printf("sum     :");
print_values(&sum);
printf("diff    :");
print_values(&dif);
printf("product :");
print_values(&prod);
printf("cross   :");
print_values(&xross);
printf("random  :");
print_values(&r);
printf("mag  u  :%8.3f\n",mag_u);
printf("mag  v  :%8.3f\n",mag_v);
printf("dot u v :%8.3f\n",dott);
printf("segDist :%8.3f\n",SegD);

if ( flag ) { // perpendicular of r does fall between origin and v
       printf("distLine:%8.3f\n",distLine);
              }
else { printf("r does not fall on line\n"); }

norm(&v);
printf("norm   v:");
print_values(&v);

return 0;
}



////////////////////////////////////////////////////////////////////////////
//functions
////////////////////////////////////////////////////////////////////////////

double DistanceSegments( struct vector *p, struct vector *q, struct vector *u, struct
vector *v )
{
double a, b, c, d, e, Di, sD, tD, sN, tN, sc, tc, small_num, distance;
struct vector x, y, z, dP;
sub_vectors( &x, p, q );
sub_vectors( &y, u, v );
sub_vectors( &z, q, v );
a = dot( &x, &x );
b = dot( &x, &y );
c = dot( &y, &y );
```

```
d = dot( &x, &z );
e = dot( &y, &z );
Di = a*c - b*b;
sD = Di;
tD = Di;
small_num = .00000001;
if ( Di < small_num ) { sN = 0.0; sD = 1.0; tN = e; tD = c; }
else {  sN = ( b*e - c*d ); tN = ( a*e - b*d );
        if ( sN < 0.0 ) { sN = 0.0; tN = e; tD = c; }
        else if ( sN > sD ) { sN = sD; tN = e + b; tD = c; }
        }
if ( tN < 0.0 ) { tN = 0.0;
        if ( -1*d < 0.0 ) { sN = 0.0; }
        else if ( -1*d > a ) { sN = sD; }
        else {  sN = -1*d; sD = a; }
        }
else if ( tN > tD ) { tN = tD;
        if ( ( -1*d + b ) < 0.0 ) { sN = 0.0; }
        else if ( ( -1*d + b ) > a ) { sN = sD; }
        else {  sN = ( -1*d + b ); sD = a; }
        }
if ( fabs(sN) < small_num ) { sc = 0.0; }
else { sc = sN / sD; }
if ( fabs(tN) < small_num ) { tc = 0.0; }
else { tc = tN / tD; }
mul_vector( &x, &x, sc );
mul_vector( &y, &y, tc );
sub_vectors( &x, &x, &y);
add_vectors( &z, &z, &x);
distance = magnitude( &z );
return distance;
}

double Magnitude( struct vector *Point1, struct vector *Point2 )
{
    struct vector Vector;
    Vector.x = Point2->x - Point1->x;
    Vector.y = Point2->y - Point1->y;
    Vector.z = Point2->z - Point1->z;
    return (double)sqrt( Vector.x * Vector.x + Vector.y * Vector.y + Vector.z *
Vector.z );
}

int DistancePointLine( struct vector *Point, struct vector *LineStart, struct vector
*LineEnd, double *Distance )
{
    double LineMag;
    double U;
    struct vector Intersection;
    double Magnitude( struct vector *Point1, struct vector *Point2 );

    LineMag=Magnitude(LineStart,LineEnd);

    U = ( ( ( Point->x - LineStart->x ) * ( LineEnd->x - LineStart->x ) ) +
        ( ( Point->y - LineStart->y ) * ( LineEnd->y - LineStart->y ) ) +
        ( ( Point->z - LineStart->z ) * ( LineEnd->z - LineStart->z ) ) ) /
        ( LineMag * LineMag );

    if( U < 0.0f || U > 1.0f )
        return 0;   // closest point does not fall within the line segment

    Intersection.x = LineStart->x + U * ( LineEnd->x - LineStart->x );
    Intersection.y = LineStart->y + U * ( LineEnd->y - LineStart->y );
```

```
    Intersection.z = LineStart->z + U * ( LineEnd->z - LineStart->z );

    *Distance = Magnitude( Point, &Intersection );

    return 1;
}

void norm(struct vector *u_ptr)
{
double x, y, z, m;
void extract_values(struct vector *u_ptr,double *a_ptr, double *b_ptr, double *c_ptr);
void assign_values(struct vector *v_ptr,double a, double b, double c);
extract_values(u_ptr, &x, &y, &z);
m=sqrt(pow(x,2)+pow(y,2)+pow(z,2));
assign_values(u_ptr,x/m,y/m,z/m);
}

void rotate(struct vector *rotate,double thi, double theta, double psi)
{
double x, y, z;
void extract_values(struct vector *u_ptr,double *a_ptr, double *b_ptr, double *c_ptr);
void assign_values(struct vector *v_ptr,double a, double b, double c);
extract_values(rotate, &x, &y, &z);
assign_values(rotate, x*cos(theta)*cos(psi)-y*cos(thi)*sin(psi)+

y*sin(thi)*sin(theta)*cos(psi)+z*sin(thi)*sin(psi)+z*cos(thi)*sin(theta)*cos(psi),

x*cos(theta)*sin(psi)+y*cos(thi)*cos(psi)+y*sin(thi)*sin(theta)*sin(psi)-
                      z*sin(thi)*cos(psi)+z*cos(thi)*sin(theta)*sin(psi),
                      x*-sin(theta)+y*sin(thi)*cos(theta)+z*cos(thi)*cos(theta));
}

double magnitude(struct vector *u_ptr)
{
double a, b, c;
void extract_values(struct vector *u_ptr,double *a_ptr, double *b_ptr, double *c_ptr);
extract_values(u_ptr, &a, &b, &c);
return sqrt(pow(a,2)+pow(b,2)+pow(c,2));
}

double distance(struct vector *u_ptr, struct vector *v_ptr)
{
double a, b, c, x, y, z;
void extract_values(struct vector *u_ptr,double *a_ptr, double *b_ptr, double *c_ptr);
extract_values(u_ptr, &a, &b, &c);
extract_values(v_ptr, &x, &y, &z);
return sqrt(pow(a-x,2)+pow(b-y,2)+pow(c-z,2));
}

void sub_vectors(struct vector *dif,struct vector *u_ptr, struct vector *v_ptr)
{
double a, b, c, x, y, z;
void assign_values(struct vector *v_ptr,double a, double b, double c);
void extract_values(struct vector *u_ptr,double *a_ptr, double *b_ptr, double *c_ptr);
extract_values(u_ptr, &a, &b, &c);
extract_values(v_ptr, &x, &y, &z);
assign_values(dif,a-x,b-y,c-z);
}

void add_vectors(struct vector *sum,struct vector *u_ptr, struct vector *v_ptr)
{
double a, b, c, x, y, z;
void assign_values(struct vector *v_ptr,double a, double b, double c);
```

```
void extract_values(struct vector *u_ptr,double *a_ptr, double *b_ptr, double *c_ptr);
extract_values(u_ptr, &a, &b, &c);
extract_values(v_ptr, &x, &y, &z);
assign_values(sum,a+x,b+y,c+z);
}

void mul_vector(struct vector *mul,struct vector *u_ptr,double scaler)
{
double a, b, c;
void assign_values(struct vector *v_ptr,double a, double b, double c);
void extract_values(struct vector *u_ptr,double *a_ptr, double *b_ptr, double *c_ptr);
extract_values(u_ptr, &a, &b, &c);
assign_values(mul,a*scaler,b*scaler,c*scaler);
}

void random_unit_vector(struct vector *v_ptr)
{
double u=open_interval_rand(0.0,1.0), v=open_interval_rand(0.0,1.0);
double angle_y=2*M_PI*u, angle_z=acos(2*v-1);
double z=cos(angle_z);
double y=sqrt(1-pow(z,2))*sin(angle_y);
double x=sqrt(1-pow(z,2)-pow(y,2));
if (rand()%2==1)
x=x*-1;
if (rand()%2==1)
y=y*-1;
if (rand()%2==1)
z=z*-1;
v_ptr -> z = z;
v_ptr -> y = y;
v_ptr -> x = x;
}

double dot(struct vector *u_ptr, struct vector *v_ptr)
{
double a, b, c, x, y, z;
void extract_values(struct vector *u_ptr,double *a_ptr, double *b_ptr, double *c_ptr);
extract_values(u_ptr, &a, &b, &c);
extract_values(v_ptr, &x, &y, &z);
return (a*x+b*y+c*z);
}

void cross(struct vector *cross,struct vector *u_ptr, struct vector *v_ptr)
{
double a, b, c, x, y, z;
void assign_values(struct vector *v_ptr,double a, double b, double c);
void extract_values(struct vector *u_ptr,double *a_ptr, double *b_ptr, double *c_ptr);
extract_values(u_ptr, &a, &b, &c);
extract_values(v_ptr, &x, &y, &z);
assign_values(cross,b*z-y*c,c*x-z*a,a*y-x*b);
}

double triple(struct vector *u_ptr, struct vector *v_ptr, struct vector *w_ptr)
{
double triple, a, b, c, x, y, z, q, p, r;
void extract_values(struct vector *u_ptr,double *a_ptr, double *b_ptr, double *c_ptr);
extract_values(u_ptr, &a, &b, &c);
extract_values(v_ptr, &x, &y, &z);
extract_values(w_ptr, &q, &p, &r);
return q*(b*z-y*c)+p*(c*x-z*a)+r*(a*y-x*b);
}

void assign_values(struct vector *v_ptr,double a, double b, double c)
```

```
{
v_ptr -> x = a;
v_ptr -> y = b;
v_ptr -> z = c;
}

void extract_values(struct vector *v_ptr, double *a_ptr, double *b_ptr, double *c_ptr)
{
*a_ptr = v_ptr -> x;
*b_ptr = v_ptr -> y;
*c_ptr = v_ptr -> z;
}

void print_values(struct vector *v_ptr)
{
double a, b, c;
void extract_values(struct vector *v_ptr,double *a_ptr, double *b_ptr, double *c_ptr);
extract_values(v_ptr, &a, &b, &c);
printf("%8.3lf %8.3lf %8.3lf\n", a, b, c);
}

inline double open_interval_rand(double x0, double x1)
{
return x0 + (x1 - x0) * (rand()+1.0) / ((double)(RAND_MAX)+2.0);
}

inline double closed_interval_rand(double x0, double x1)
{
return x0 + (x1 - x0) * rand() / ((double) RAND_MAX);
}

double random_number_generator(void)
{
double double_number = closed_interval_rand(0.0,1.0);
return double_number;
//seed=(MULTIPLIER * seed + INCREMENT) % MODULUS;
//return (seed / FLOATING_MODULUS);
}
```

APPENDIX B - SUPPLEMENT FOR CHAPTER 2

B.1 Derivation of model parameters and constraints from biological data

Number of chromocenters. It has been established experimentally (30) that the five chromosome arms of *D. melanogaster* salivary gland share a single common chromocenter in most nuclei (31,239); our models are constructed with a single chromocenter (Figures 2.1 and 2.2 main text).

Chromocenter position. The chromocenter in *D. melanogaster* salivary gland is always positioned at the nuclear periphery (30,31). One bead representing the chromocenter touches the NE in all models (Figure 2.1 and 2.2 main text).

Chromocenter arrangement. The configuration of the chromocenter, Figure 2.3, has been described experimentally (30) by recording the order of chromosome arms around the chromocenter. Six different chromocenter arrangements were observed in an experimental set of 22 nuclei (30); the number of nuclei satisfying the given arrangement (out of 22) was recorded for each type. In our models the experimental numbers from Figure 2.3 become frequencies for assigning the corresponding chromocenter arrangement.

Bead size and chromosome thickness. The diameter of *D. melanogaster* polytene chromosome can range from 3.1-3.2 microns (31). Our model uses beads with a diameter of 3.1 microns. To fully capture the thickness of the chromosome fiber we place a cylinder of excluded volume around the bond between nearest neighbor beads (Figure 2.1 main text). This detail was important for achieving the right nucleus volume to chromosome volume ratio, but was found to have little effect on the scaling of our self-avoiding walks in free space, see below.

Chromosome length. The length of the five major chromosome arms have been measured experimentally for *D. melanogaster* salivary gland (31). Our model incorporates the measured arm lengths by assigning the nearest whole number of beads to each model chromosome arm; X - 45 beads, 2R - 47 beads, 3R – 60 beads, 2L – 46 beads, 3L – 50 beads.

Chromosome right handedness. Studies that trace the path of each chromosome arm in *D. melanogaster* salivary gland nuclei have observed a disproportionate amount of right handed twist (30). This preferred right handedness has been quantified by measuring the triple products of 3 unit vectors (*a, b,* and *c*), tangent to the chromosome path, each spaced 7 microns apart; the triple product, defined as $a\cdot(bxc)$, produces a positive scalar for right handed segments of chromosome and a negative scalar for left-handed segments of chromosome. The distribution of chromosome twist has been measured by calculating the triple product for every set of three vectors formed within a 14 micron window that slides along the path of the chromosome arm. We perform the same analysis during the construction of our SAWs; however, unit vectors that point from the center of one bead to the next nearest neighbor are used

in place of unit vectors tangent to the chromosome path. 3.1 microns separate neighboring bead centers; consequently we calculate the triple product of three vectors formed within a 12.4 micron window, rather than a 14 micron window. We enforce right handedness in our simulated chromosomes: during construction of the SAWs, it is twice as likely for a new bead to be accepted if the new triple product formed with this bead is right handed rather than left handed.

Polytene chromosome persistence length. Polymer models have estimated a 1.5 micron persistence length for *D. melanogaster* salivary gland polytene chromosomes (195). A 1.5 micron persistence length means the effective Kuhn length of our model is about twice the persistence length, meeting the condition necessary to build our models as a SAW (85,196).

Nucleus size. The diameter of *D. melanogaster* salivary gland nuclei has a range of 30–35 microns (146) with an average volume around 19,000 cubic microns. All of our model nuclei have a diameter of 33.5 microns.

Nucleolus size. The nucleolus size in our model is based on the measured volume of the nucleolus in *D. melanogaster* salivary glands (33). Our models use a nucleolus diameter of 7.26 microns (Figure 2.1 and 2.4 main text).

Nucleolus number. Although some nuclei have more than one nucleolus (31,33,177,239), we only include one in our models.

Nucleolus position. The nucleolus is known to (permanently) associate with the nucleolar organizing region at the base of the X chromosome (240). In our model the nucleolus position is fixed in space, it touches the base of the single shared chromocenter (Figure 2.1 and 2.4 main text).

Rabl configuration. It is known that 80% of *D. melanogaster* polytene chromosomes conform to the Rabl type configuration (30). This configuration is characterized by the predominant presence of the chromosome telomeres in the nuclear hemisphere opposite the chromocenter. Rabl configuration was enforced in our models by filtering the generated ensembles of nuclei to achieve, in the final ensemble, 80% of telomeres per nucleus in the hemisphere opposite the chromocenter. Specifically, each model nucleus from our ensemble contains 5 chromosome arms. A score (0-5) was assigned to each (unfiltered) model nucleus corresponding to its number of chromsomes in Rabl configuration. All models with a score of 4 or 5 were included in our final (filtered) ensemble. In addition, we have also included a fraction of models with a score of 3, such that the model nuclei in the final ensemble have 80% of telomeres in the hemisphere opposite the chromocenter just like experiment (see figure). The final ensemble contained 96 sets of 24 (2304 total) model nuclei.

B.2 Robustness of threshold used to identify Chr-NE attachments

Our model incorporates *all* experimentally known parameters from *D. melanogaster* polytene chromosomes *with the exception* of introducing specific Chr-NE attachments; in other words, the model is a Null model with respect to chromosome-NE attachment. A threshold, $\langle \overline{\chi} + 2\sigma \rangle$, is then used to identify 48 statistically significant deviations between the null model and experiment which correspond to the regions of Chr-NE attachment (this is described extensively in the main text). We construct the Null model using an equilibrium based self-avoiding walk approach and introduce several modifications in order to recapitulate experiment. Some of these modifications likely introduce non-equilibrium features into our model; however, we stress that the fully modified model contains all the known features of the polytene nucleus from experiment except for specific Chr-NE attachments. For any other model the deviations from experiment would arise from multiple factors, not just the Chr-NE attachments. Regardless, we check that the crucial model conclusion, the statistical thresholds $\langle \overline{\chi} \pm 2\sigma \rangle$, are robust to the non-equilibrium features that our model contains. Three variations of our SAW approach (also described in the main text) were used for robustness checking of the statistical thresholds, chromosome territories, and chromosome intertwining: *fully modified SAW* – with Rabl configuration, right-handed chirality, and chromocenter arrangement designed to recapitulate all features of experimental nuclei with the exception of Chr-NE attachments; *unmodified SAW* – does *not* introduce Rabl configuration, right-handed chirality, or chromocenter arrangement and is equilibrium to the extent that our chain growing algorithm approximates self-repelling chains (see main text); *fully modified SAW with BT1 = 1000 and BT2 = 3000* – designed to check robustness of model conclusions to the backtracking parameters used in our chain growing algorithm (also see main text). We also checked robustness of the statistical thresholds to the number of models used in the calculation. Results are summarized in table S1 and S2.

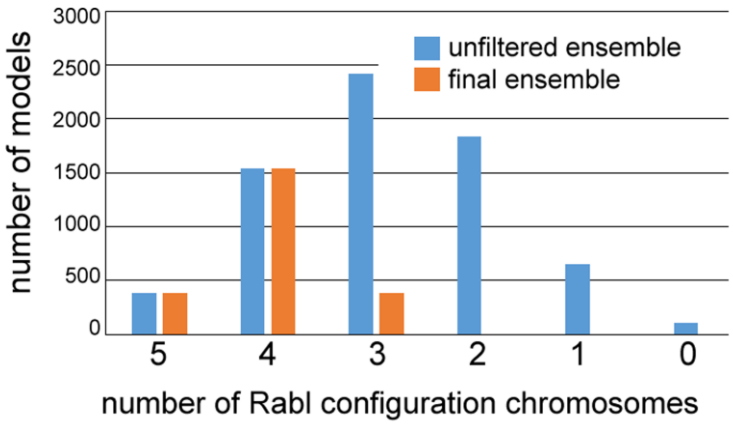


Figure B.1 - A posteriori filtering to achieve in the final ensemble 80% of telomeres in the hemisphere opposite the chromocenter as seen in experiment.

Model description	$\langle \chi + 2\sigma \rangle$	$\langle \chi - 2\sigma \rangle$
96 sets of 24 model nuclei; fully modified SAW (focus of paper)	**.505**	**.143**
48 sets of 24 model nuclei; fully modified SAW	**.506**	**.151**
24 sets of 24 model nuclei; fully modified SAW	**.505**	**.154**
96 sets of 24 model nuclei; unmodified SAW	**.502**	**.148**
fully modified SAW with BT1 = 1000 and BT2 = 3000	**.503**	**.143**

Table B.1 – **Robustness of thresholds to model details.**

Model description	average territory index	Percent non-
fully modified SAW (focus of paper)	**.650**	**95%**
unmodified SAW	**.651**	**95%**

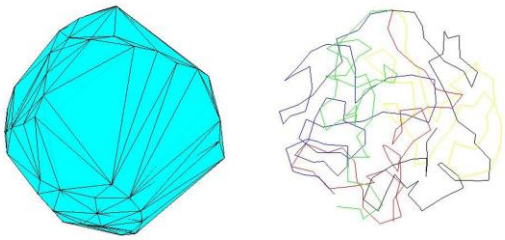
Table B.2 – **Robustness of territories and intertwining to model details.**

Chromosome arm	NE contact regions	LR - Sites of late replication (239 regions)	SUUR antibody binding sites in SuUR 4x (280 regions)	SUUR antibody binding sites in SuUR 2x (113 regions)	Region classification
X	1AB*	1AB	1AB	1AB	IH
X	1EF	1EF	1E		LR
X	5C	5C	5C		LR
X	6AB	6A	6A	6A	IH
X	8AB	8AB	8B	8B	IH
X	9A*	9A	9A	9A	IH
X	11C	11C	11C		LR
X	12E*	12E	12E	12E	IH
X	16D				E
X	17B	17B		17B	IH
X	18A	18A	18A	18A	IH
X	19DE^	19E	19E	19E	IH
2L	21A	21A	21A	21A	IH
2L	22AB*	22AB	22A	22A	IH
2L	22D				E
2L	23C				E
2L	25EF	25EF	25E	25E	IH
2L	32A	32A	32A	32A	IH
2L	32F-33A*	32F-33A	32F-33A	32F-33A	IH

2L	35AC*	35BC	35C	35C	IH
2L	36D*	36D		36D	IH
2L	37D	37D	37D	37D	IH
2R	42BC	42B	42B	42B	IH
2R	45F-46A	46A	46A		LR
2R	48EF	48E	48EF	48E	IH
2R	49CD	49C	49D		LR
2R	56F-57A*	56F-57A	56F-57A	56F-57A	IH
2R	60EF^	60F	60F	60F	IH
3L	61AB^	61A	61A	61A	IH
3L	62AB		62AB		LR
3L	62D	62D	62D		LR
3L	64C*	64C	64C	64C	IH
3L	67D*	67D	67D	67D	IH
3L	69BC	69B			LR
3L	70CD*x	70CD	70C	70C	IH
3L	72AB*x	72A	72A		LR
3L	73F-74A	74A	74A	74A	IH
3L	74E-75A	75A	75A	75A	IH
3L	79DE	79DE	79E	79E	IH
3R	83DE*	83DE	83DE	83D	IH
3R	84A	84A	84A	84A	IH
3R	86D	86D	86D	86D	IH
3R	90B		90B		LR
3R	92A	92A	92A		LR
3R	97AB^	97AB	97A	97AB	IH
3R	98C*	98C	98C	98C	IH
3R	100BC	100BC	100BC	100BC	IH
3R	100F*	100F	100F	100F	IH

* - high frequency NE-associated regions identified at thresholds 0.66-0.63 in Hochstrasser et al 1986 [1] (15 loci)
Unmarked - sub-high frequency NE-contacts at threshold 0.5 (33 loci)
^ - high frequency NE-associated regions identified ALSO in Mathog and Sedat 1989 [9] (4 loci)
x - high frequency NE-associated regions found in Hochstrasser et al 1986 [1] but not in Mathog and Sedat 1989 [9] (2 loci)
E – Euchromatin
LR - Late replicated region (transition stage b/w E and IH)
IH - Intercalary heterochromatin (small-medium size IH bands)
UR IH - Underreplicated intercalary heterochromatin (the largest IH bands)

Table B.3 - **Classification of chromosome-nuclear envelope contacts by chromatin type.**



Fiure B.2 - The maximum volume of chromosome convex hull under confinement. The convex hull volume of a chromosome is maximized using a pivot algorithm. Random rotations of chromosome segments are preformed, rejecting those that do not increase the convex hull volume. Iterations are preformed until numerical convergence is achieved. A maximum convex hull for chromosome 3R under confinement (left) is shown next to a model nucleus (right).

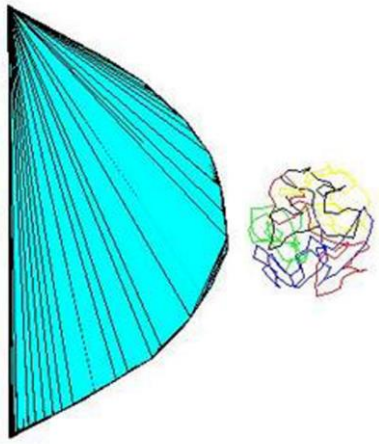


Figure B.3 - The maximum volume of chromosome convex hull in free space. In free space the maximum convex hull is larger than the entire nucleus.

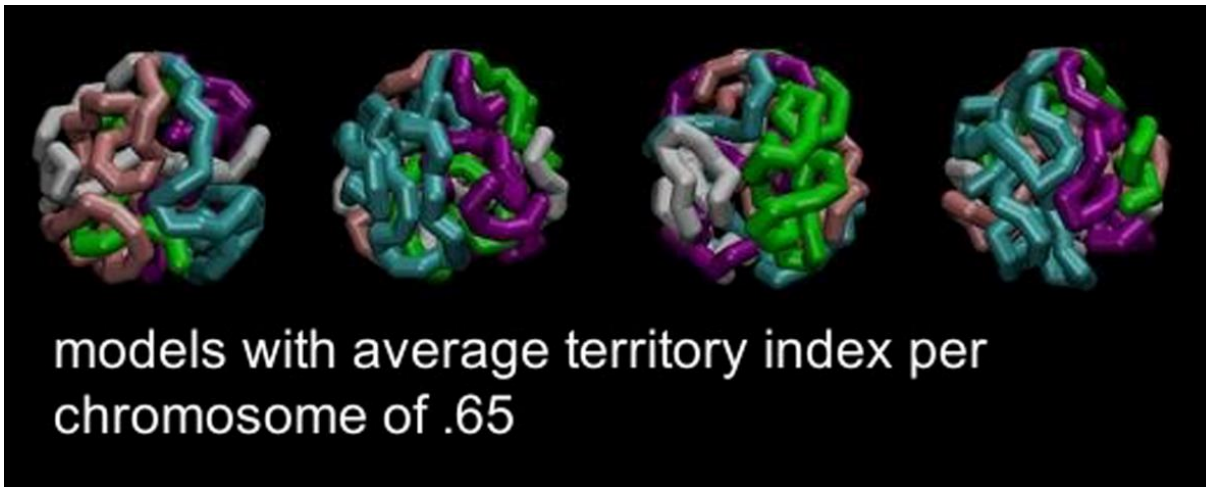


Figure B.4 - Simulated nuclei with average territory index per chromosome of .65. The average territory index per chromosome over all simulated nuclei we generated was .65 (see methods), examples of single model nuclei with this territory index are shown above. The standard deviation of the territory index per chromosome was .04.

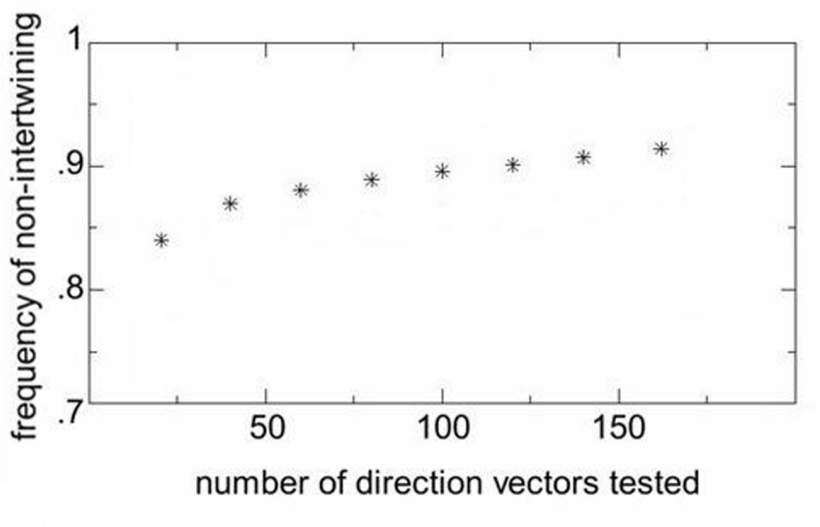


Figure B.5 - Convergence of the non-intertwining frequency between pairs of chromosomes as the number of test directions for spatial separation is increased. Shown is frequency of non-intertwining depending on the number of direction vectors tested (methods); this suggests that as the number of test directions increases the frequency on non-intertwining chromosomes in our models approaches 95%.

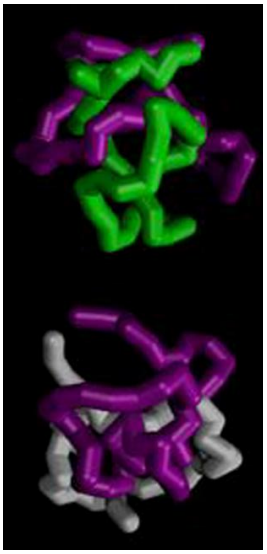


Figure B.6 - Examples of model chromosomes that intertwine.

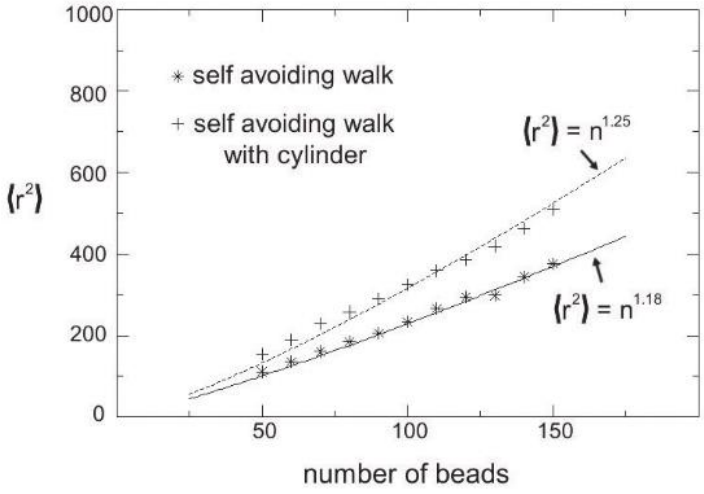


Figure B.7 - Scaling of self avoiding walks. Each data point represents the square end-end length averaged over 1000 self avoiding walks. This averaging was repeated for self avoiding walks ranging from 50 monomers to 150 monomers. To capture the thickness of the chromosomes a cylinder of excluded volume was placed around the bond between nearest neighbor beads. Scaling with and without this extra excluded volume is shown above. Least square regression lines are shown for each set of points.

APPENDIX C - SUPPLEMENT FOR CHAPTER 3

C.1 Detailed Model Parameters

Number of chromocenters. It has been established experimentally (30) that the five chromosome arms of *D. melanogaster* salivary gland share a single common chromocenter in most nuclei (31,239); our models are constructed with a single chromocenter (Figure 3.1 main text).

Chromocenter position. The chromocenter in *D. melanogaster* salivary gland is always positioned at the nuclear periphery (30,31). One bead representing the chromocenter touches the NE in all models (Figure 3.1 main text).

Chromocenter arrangement. The configuration of the chromocenter has been described experimentally (30) by recording the order of chromosome arms around the chromocenter. Six different chromocenter arrangements were observed in an experimental set of 22 nuclei (30); the number of nuclei satisfying the given arrangement (out of 22) was recorded for each type. In our models the experimental numbers become frequencies for assigning the corresponding chromocenter arrangement, see reference (187) for details.

Bead size and chromosome thickness. The diameter of *D. melanogaster* polytene chromosome can range from 3.1-3.2 microns (31). Our model uses beads with a diameter of 3.1 microns. To fully capture the thickness of the chromosome fiber we place a cylinder of excluded volume around the bond between nearest neighbor beads (Figure 3.1 main text). This detail was important for achieving the right nucleus volume to chromosome volume ratio, but was found to have little effect on the scaling of our self-avoiding walks in free space. See reference (187) for details.

Chromosome length. The length of the five major chromosome arms have been measured experimentally for *D. melanogaster* salivary gland (31). Our model incorporates the measured arm lengths by assigning the nearest whole number of beads to each model chromosome arm; X - 45 beads, 2R - 47 beads, 3R – 60 beads, 2L – 46 beads, 3L – 50 beads.

Chromosome right handedness. Studies that trace the path of each chromosome arm in *D. melanogaster* salivary gland nuclei have observed a disproportionate amount of right handed twist (30). This preferred right handedness has been quantified by measuring the triple products of 3 unit vectors (*a, b,* and *c*), tangent to the chromosome path, each spaced 7 microns apart; the triple product, defined as $a\cdot(bxc)$, produces a positive scalar for right handed segments of chromosome and a negative scalar for left-handed segments of chromosome. The distribution of chromosome twist has been measured by calculating the triple product for every set of three vectors formed within a 14 micron window that slides along the path of the chromosome arm. We perform the same analysis during the construction of our SAWs; however, unit vectors that point from the center of one bead to the next nearest neighbor are used

in place of unit vectors tangent to the chromosome path. 3.1 microns separate neighboring bead centers; consequently we calculate the triple product of three vectors formed within a 12.4 micron window, rather than a 14 micron window. We enforce right handedness in our simulated chromosomes: during construction of the SAWs, it is twice as likely for a new bead to be accepted if the new triple product formed with this bead is right handed rather than left handed.

Polytene chromosome persistence length. Polymer models have estimated a 1.5 micron persistence length for *D. melanogaster* salivary gland polytene chromosomes (195). A 1.5 micron persistence length means the effective Kuhn length of our model is about twice the persistence length, meeting the condition necessary to build our models as a SAW (85,196).

Nucleus size. The diameter of *D. melanogaster* salivary gland nuclei has a range of 30–35 microns (146) with an average volume around 19,000 cubic microns. All of our model nuclei have a diameter of 33.5 microns.

Nucleolus size. The nucleolus size in our model is based on the measured volume of the nucleolus in *D. melanogaster* salivary glands (33). Our models use a nucleolus diameter of 7.26 microns.

Nucleolus number. Although some nuclei have more than one nucleolus (31,33,177,239), we only include one in our models.

Nucleolus position. The nucleolus is known to (permanently) associate with the nucleolar organizing region at the base of the X chromosome (240). In our model the nucleolus position is fixed in space, it touches the base of the single shared chromocenter.

Rabl configuration. It is known that 80% of *D. melanogaster* polytene chromosomes conform to the Rabl type configuration (30). This configuration is characterized by the predominant presence of the chromosome telomeres in the nuclear hemisphere opposite the chromocenter. Rabl configuration was enforced in our models by filtering the generated ensembles of nuclei to achieve, in the final ensemble, 80% of telomeres per nucleus in the hemisphere opposite the chromocenter, details are thoroughly discussed in a previous work (187).

Mapping of Chromosome nuclear envelope attachments onto the model. In this work considers two sets of chromosome-nuclear envelope attachments. The complete mapping of attachments and their origin from experiment are summarized in table S1. Each Chr-NE attachment is imposed as a constraint during the construction of SAW's (see text S2).

Three models considered in this study. In total, we consider three computational models of the nucleus; all three models contain the same relevant measured parameters of the polytene nucleus from *D. melanogaster,* and differ only by the number and type of Chr-NE attachments incorporated. (I) the Null model – zero specific Chr-NE attachments; the only model considered in a previous work which here serves as a reference for determining the effects of Chr-NE attachments. (II) a 15-attachment model, containing the same set of 15 Chr-NE attachment identified previously. (III) a 48-attachment model, containing the same set of 48 attachments identified in. We emphasize that the only difference between the Null models and the attachment models is the absence of Chr-NE attachments.

Chromosome arm	NE contact regions	corresponding bead in model	included in Null model	Identified in Hochstrasser et al 1986 [1] and included in 15 attachment model	Identified in Kinney et al 2013 [4] and present in 48 attachment model
X	1AB	2		✓	✓
X	1EF	6			✓
X	5C	8			✓
X	6AB	9			✓
X	8AB	18			✓
X	9A	21		✓	✓
X	11C	26			✓
X	12E	28		✓	✓
X	16D	33			✓
X	17B	34			✓
X	18A	42			✓
X	19DE	44			✓
2L	21A	6			✓
2L	22AB	10		✓	✓
2L	22D	13			✓
2L	23C	18			✓
2L	25EF	20			✓
2L	32A	34			✓
2L	32F-33A	40		✓	✓
2L	35AC	41		✓	✓
2L	36D	43		✓	✓
2L	37D	45			✓
2R	42BC	2			✓
2R	45F-46A	11			✓
2R	48EF	18			✓
2R	49CD	19			✓
2R	56F-57A	37		✓	✓
2R	60EF	46			✓
3L	61AB	4			✓
3L	62AB	15			✓
3L	62D	17			✓
3L	64C	22		✓	✓
3L	67D	26		✓	✓

3L	69BC	29			✓
3L	70CD	33		✓	✓
3L	72AB	40		✓	✓
3L	73F-74A	45			✓
3L	74E-75A	46			✓
3L	79DE	49			✓
3R	83DE	7		✓	✓
3R	84A	8			✓
3R	86D	16			✓
3R	90B	27			✓
3R	92A	32			✓
3R	97AB	48			✓
3R	98C	51		✓	✓
3R	100BC	57			✓
3R	100F	58		✓	✓

Table C.1 – mapping of chromosome nuclear envelope contact regions.

C.2 Sampling protocol for generation of computational ensembles

Modeling procedure. A single step in growing the SAWs consists of simultaneously picking a random direction in 3D space to extend each model chromosome arm, adding the five new beads, and checking for violation of model constraints such as excluded volume (no bead overlap) and right-handed chromosome chirality. If no model constraints are violated, then the new beads are accepted and the model chromosome arms continue growing. In the case of rejecting the new beads, the step is repeated with new random directions in 3D space. The avoidance of perpetual SAW rejections is accomplished with two backtracking parameters, BT_1 and BT_2, that tally the number of SAW rejections. A single bead backtrack is made after $BT_1 = 2000$ failed SAW additions, followed by its resetting; a 5 bead backtrack is made after $BT_2 = 6000$ failed SAW additions, followed by its resetting. The above process is repeated until model completion. We chose the manifestly symmetric SAW construction procedure (at each step the beads for all the chromosomes are added simultaneously) because there is no biological evidence that suggests a spatial symmetry breaking between the chromosomes. That is a conceivable alternative procedure in which a certain chromosome is fully built first, followed by other(s) would be less justified biologically.

C.3 Derivation of contact probabilities by interaction type for volume vs surface accessibility argument. Quantitatively, the accessible/contact volumes and the corresponding contact probabilities can be expressed in terms of the 4 fundamental distances in our model: the distance between beads forming pairwise contact (α), the nucleus radius (β), the bead radius (r), and the distance between attachment beads and the NE (γ). See figure C.1 for illustration of these fundamental distances. Table C.2 summarizes the derivation of contact probabilities based on these 4 fundamental distances in our model.

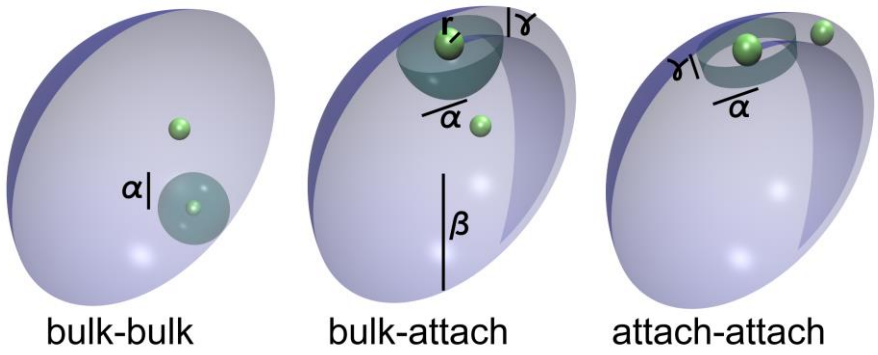


Figure C.1 - Derivation of contact probabilities by interaction type – the volume vs surface accessibility argument (see main text) is made quantitative by comparing the contact probability of thee interactions types. The contact probability is computed by considering two volumes: accessible volume for a bead – volume in which a bead is confined, contact sub-volume for a bead – all nearby beads within this small volume form a bead-bead contact (cyan sub-volume above). These volumes are computed explicitly in table S1 using the 4 fundamental distances in our models. The contact probability for the interaction type is the contact sub-volume divided by the accessible volume of the bead (computed in table S1).

Interaction	Contact sub-volume	Bead accessible volume	contact probability
Bulk-bulk	$(4\pi/3)\alpha^3$ *(In our case* $\alpha = 5.1$*)*	$(4\pi/3)\beta^3$ (In our case $\beta = 16.5$)	$P_{bulk}^{bulk} = \frac{\alpha^3}{\beta^3} = .03$
Bulk-attach.	$(2\pi/3)\alpha^3$ *(half-sphere)*	$(4\pi/3)\beta^3$	$P_{attach}^{bulk} = \frac{\alpha^3}{2\beta^3} = .014$
Attach.--attach.	$\gamma\pi\alpha^2$ *(In our case* $\gamma = 1.0$*)*	$\gamma 4\pi\beta^2$	$P_{attach}^{attach} = \frac{\alpha^2}{4\beta^2} = .019$

Table C.2 – Predicted inter-arm/inter-chromosome bead-bead contact frequencies.

C.4 volume vs surface accessibility argument for intra-arm chromosome-chromosome contacts. Intra-arm and intra-chromosome contacts are more common in our 48 and 15 attachment models compared to our Null model (Figure 3.3 main text). The same general volume *vs.* surface accessibility arguments are used here to rationalize this increase as seen in our computational models. The "contact sub-volume" around each bead is the same as above (Table S2). However, we no longer assume that beads are randomly positioned relative to each other due to the linking of beads along the polymer backbone; instead, we note that the largest distance between intra-arm beads is $2nr$, where *n* is an integer and *r* is the radius of a bead. For example, nearest neighbor beads are separated by a distance of $2r$ and the accessible volume with respect its neighbor is $(4\pi/3)(2r)^3$ (for bulk-bulk beads). Table S3 summarizes intra-arm contact probabilities by interactions type for second nearest neighbors only (nearest neighbors are always in contact). Note from table S3 that $P_{bulk}^{bulk} = P_{attach}^{bulk} < P_{attach}^{attach}$. All second-neighbor contacts in our Null model are bulk-bulk and the turning on of each chromosome nuclear envelope attachments essentially replaces bulk-bulk interactions with attachment-attachment interactions having higher contact probability. Therefore more second neighbor interactions are realized in the 15 Chr-NE attachment model compared to our Null model. The effect is even more pronounced in the 48 Chr-NE attachment model. A more comprehensive prediction of intra-arm interactions would involve an approach including all *n* orders of neighbors, which is beyond the motivation of this demonstration of the basic argument.

Interaction	Contact sub-volume	Bead accessible volume	**second neighbors**
Bulk-bulk	$(4\pi/3)\alpha^3$	$(4\pi/3)(2nr)^3$ (n=2 for second neighbors)	$P_{bulk}^{bulk} = \frac{\alpha^3}{64r^3} = .56$
Bulk-attach.	$(2\pi/3)\alpha^3$	$(2\pi/3)(2nr)^3$ (a ½ sphere)	$P_{attach}^{bulk} = \frac{\alpha^3}{64r^3} = .56$
Attach.-attach.	$\gamma\pi\alpha^2$	$\gamma\pi(2nr)^2$ (a disk)	$P_{attach}^{attach} = \frac{\alpha^2}{16r^2} = .66$

Table C.3 - Predicted intra-arm/intra-chromosome bead-bead contact frequencies.

C.5 Robustness of major conclusions to models details

We construct our computational models using an equilibrium based self-avoiding walk approach and introduce three modifications in order to recapitulate experiment. These modifications include, (1) Rabl chromosome configuration, (2) right-handed chromosome chirality, and (3) non-random chromocenter arrangement (see text S1). We refer to the use of these modifications as a *modified self-avoiding walk approach.* These modifications and the effect on the simulation outcomes are thoroughly discussed in a previous work. There we provided validation of our model by checking explicitly that key model conclusions are robust to these non-equilibrium SAW modifications. The key model conclusions in this study are the effects of Chr-NE attachments on the 3D organization of the genome (see main text); these effects are investigated by comparing simulated ensembles of models possessing differing sets of chromosome nuclear envelope attachments (these results are discussed thoroughly in the main text). We check that the crucial model conclusions here are also robust to the non-equilibrium features of our modified self-avoiding walk approach. For robustness checking of the model conclusions we use an *unmodified self-avoiding walk* approach - does *not* introduce (1) Rabl chromosome configuration, (2) right-handed chromosome chirality, or (3) non-random chromocenter arrangement and is equilibrium to the extent that our chain growing algorithm approximates self-repelling chains (see main text). Results of robustness checking are summarized in table S4.

observable	SAW approach used	% change due to turning on chromosome-nuclear envelope attachments
chromosome territory index (15 attachment model vs Null model)	modified	1.8
chromosome territory index (15 attachment model vs Null model)	unmodified	2.0
chromosome territory index (48 attachment model vs Null model)	modified	4.6
chromosome territory index (48 attachment model vs Null model)	unmodified	4.9
non-intertwining chromosomes (15 attachment model vs Null model)	modified	1.6
non-intertwining chromosomes (15 attachment model vs Null model)	unmodified	1.5
non-intertwining chromosomes (48 attachment model vs Null model)	modified	3.0
non-intertwining chromosomes (48 attachment model vs Null model)	unmodified	4.1
normalized inter-chromosome contacts (15 attachment model vs Null model)	modified	-2.7
normalized inter-chromosome contacts (15 attachment model vs Null model)	unmodified	-2.8

normalized inter-chromosome contacts (48 attachment model vs Null model)	modified	-8.4
normalized inter-chromosome contacts (48 attachment model vs Null model)	unmodified	-8.1
normalized inter-arm contacts (15 attachment model vs Null model)	modified	-2.6
normalized inter-arm contacts (15 attachment model vs Null model)	unmodified	-2.5
normalized inter-arm contacts (48 attachment model vs Null model)	modified	-7.3
normalized inter-arm contacts (48 attachment model vs Null model)	unmodified	-7.2
normalized intra-chromosome contacts (15 attachment model vs Null model)	modified	1.2
normalized intra-chromosome contacts (15 attachment model vs Null model)	unmodified	1.1
normalized intra-chromosome contacts (48 attachment model vs Null model)	modified	3.1
normalized intra-chromosome contacts (48 attachment model vs Null model)	unmodified	3.3
normalized intra-arm contacts (15 attachment model vs Null model)	modified	1.5
normalized intra-arm contacts (15 attachment model vs Null model)	unmodified	1.5
normalized intra-arm contacts (48 attachment model vs Null model)	modified	4.1
normalized intra-arm contacts (48 attachment model vs Null model)	unmodified	4.3

Table C.4 – Robustness of conclusions to model details.