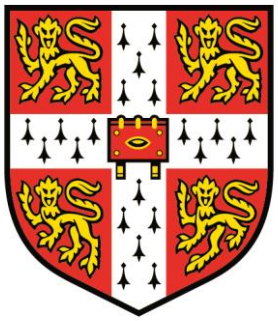
Study of 3D Genome Organisation in Budding Yeast by Heterogeneous Polymer Simulations



Zahra Fahmi

Babraham Institute

University of Cambridge

This dissertation is submitted for the degree of Doctor of Philosophy

Murray Edwards College May 2018

Abstract

Study of 3D genome organisation in budding yeast by heterogeneous polymer simulations

Zahra Fahmi

Investigating the arrangement of the packed DNA inside the nucleus has revealed the essential role of genome organisation in controlling genome function. Furthermore, genome architecture is highly dynamic and significant chromatin re-organisation occurs in response to environmental changes. However, the mechanisms that drive the 3D organisation of the genome remain largely unknown. To understand the effect of biophysical properties of chromatin on the dynamics and structure of chromosomes, I developed a 3D computational model of the nucleus of the yeast *S. cerevisiae* during interphase. In the model, each chromosome was a hetero-polymer informed by our bioinformatics analysis for heterogeneous occupancy of chromatin-associated proteins across the genome. Two different conditions were modelled, normal growth (25°C) and heat shock (37°C), where a concerted redistribution of proteins was observed upon transition from one temperature to the other. Movement of chromatin segments was based on Langevin dynamics and each segment had a mobility according to their protein occupancy and the expression level of their corresponding genes. The model provides a significantly improved match with quantitative microscopy measurements of telomere positions, the distributions of 3D distances between pairs of different loci, and the mean squared displacement of a labelled locus. The quantified contacts between chromosomal segments were similar to the observed Hi-C data. At both 25°C and 37°C conditions, the segments that were highly occupied by proteins had high number of interactions with each other, and the highly transcribed genes had lower contacts with other segments. In addition, similar to the experimental observations, heat-shock genes were found to be located closer to the nuclear periphery upon activation in the simulations. It was also shown that the determined distribution of proteins along the genome is crucial to achieve the correct genome organisation. Hence, the heterogeneous binding of proteins, which results in differential mobility of chromatin segments, leads to 3D self-organisation.

Declaration

I hereby declare that this dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration, except where specifically indicated in the text. I further state that no substantial part of my dissertation is the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution. This dissertation does not exceed the 60,000-word limit. A part of my Introduction Chapter, i.e. sections 1.1 and 1.3, has been accepted to be published as a review paper in a Springer book on Systems Biology. The Materials and Methods has been rewritten with slight changes in our research paper, which is archived in bioRxiv.

Zahra Fahmi
May 2018

Acknowledgement

First of all, I am especially grateful to my supervisor Dr. Karen Lipkow for her invaluable guidance, full support, and encouragement through all steps of my PhD. I am very thankful to Dr. Sven Sewitz for his continuous advice and generous assistance during this project on a daily basis. Karen and Sven should also be thanked for providing a pleasant and inspiring environment to work.

I am indebted to Dr. Csilla Várnai, who has helped me with the Hi-C and simulation analysis. Thanks are also due to various people for their contribution to this project: Ms. Latifa Aljebali for performing Hi-C experiments; Dr. Hicham Saad, Isabelle Goiffon, and Prof. Kerstin Bystricky for providing their unpublished microscopy data on 3D distances between labelled loci; Dr. Hua Wong, Dr. Pascal Carrivain, and Dr. Christophe Zimmer for sharing the code of their polymer model and advice on the details of the model; Dr. Steven Wingett for his advice on Hi-C data analysis; Dr. Simon Andrews for his support with the use of Babraham computing cluster; Dr. Jonathan Cairns for his help with statistical tests; Dr. Jurij Kotar and Dr. Avelino Javer for help with MSD measurements. In addition, I would like to thank Dr. Nicolas Le Novère, my mentor, and Prof. Pietro Cicuta, my graduate advisor, for their kind support.

I am grateful to project students in our group, Raphaël Forquet, Samuel Kidman, and Ema Etchegaray, for their input into this project, and to members of Nuclear Dynamics for useful discussions and feedback. I would also like to thank friends and colleagues at the Babraham Institute – both present and past. This experience would have been impossible without the financial support of the IDB, who has provided a full scholarship for my PhD.

Last but not least, I am extremely thankful to my parents for their continuous encouragement and support and to my husband for being incredibly supportive, patient, and interested in my research project, which led to helpful scientific discussions. I also owe thanks to my brother and my friends for their caring thoughts and advice throughout my PhD project.

Table of acknowledgement of assistance received during course of thesis

1) Initial training in techniques and laboratory practice and subsequent mentoring: Karen Lipkow and Sven Sewitz guided me on a daily basis and step by step helped me through all stages of the project. Csilla Várnai helped me through understanding the Langevin Thermostat and normalisation of Hi-C data. She also advised me on how to use Homer and VMD software. Pascal Carrivain and Hua Wong helped and advised me via email on their published polymer model. Hua Wong has also provided extra information about the physics of the model. Steven Wingett has helped me to understand the filtering of Hi-C data in HiCUP software. He has also advised me on Homer and SeqMonk software and cluster commands. Jonathan Cairns helped with complicated statistical comparison of simulated and experimental data.
2) Data obtained from a technical service provider (e.g. DNA sequencing, illustrations, simple bioinformatics information etc.) Simon Andrews has provided guidance and support for using the cluster and SeqMonk software.
3) Data produced jointly (e.g. where it was necessary or desirable to have two pairs of hands)
4) Data/materials provided by someone else (e.g. one-off analysis, bioinformatics analysis, where parallel data or technical provision in a very different area is needed to provide a connected account in the thesis) Karen Lipkow prepared the figures for our research paper in Illustrator software. I used these figures to make the figures that are displayed in my thesis. Sven Sewitz and Jeremy Bancroft determined chromatin states in budding yeast and Sven Sewitz did much of their further analysis. Latifa Aljebali produced Hi-C data at 25°C and 37°C. Isabelle Goiffon, Hicham Saad, and Kerstin Bystricky provided unpublished microscopy data on 3D distances of pairs of loci. Kerstin Bystricky provided strains for MSD measurements and Sven Sewitz, Avelino Javer, Juri Kotar, Pietro Cicuta measured and plotted the MSD figures. Raphaël Forquet performed network analysis on Hi-C data. Samuel Kidman analysed the 3D organisation of chromosomes using Hi-C data. Sven Sewitz and Shubham Tripathi performed bioinformatics analysis of protein complexes.

Table of Contents

List of Figures

List of Tables

1. Introduction

Major breakthroughs in the field of chromatin biology have revolutionised our understanding of genome organisation and the links to gene regulation. The first section (1.1) reviews the fundamental properties of the dynamic chromosome structure. Extensive experimental and computational methods have been developed to dissect the features of the genome structure. These methods are reviewed in sections 1.2 and 1.3. In this project, I have adopted a computational approach to investigate the mechanisms that drive the 3D self-organisation of chromosomes in *S. cerevisiae* (section 1.4).

1.1. Genome organisation

Intensive studies over the past decades have revealed multiple levels of organisation in eukaryotic genomes. On the most basic level, the DNA wraps around the core histones, made up of eight histone proteins, to make a nucleosome, the fundamental subunit of the chromatin fibre (Ramakrishnan 1997; van Holde 1989). In higher eukaryotes, the chromatin then folds to build higher order genomic structures of different scales such as sub-megabase topologically associated domains (TADs), megabase A and B compartments, and chromosomal territories (Bonev & Cavalli 2016; Sewitz et al. 2017b). The nucleus is a highly crowded environment with efficiently packed and organised chromatin and hundreds to thousands of protein species engaged in various types of interactions, such as protein-protein, DNA-protein, chromatin-chromatin and chromatin-lamina interactions. It is now known that these interactions play an important role in controlling the organised structure and the transcriptional activity of the genome (Long et al. 2016; Gómez-Díaz & Corces 2014; Flavahan et al. 2016), which changes upon differentiation, internal and external conditions (Javierre et al. 2016; Sewitz et al. 2017a; Lazar-Stefanita et al. 2017; Guidi et al. 2015). However, a comprehensive view of the mechanisms that drive organisation and dynamics of this highly complex system remains elusive.

Many research projects have investigated the linear arrangement of DNA, identifying the local regulatory elements that modulate transcription, such as transcription factor binding sites and their consensus sequences (Levine & Tjian 2003), enhancers (Long et al. 2016), histone modifications (Smolle & Workman 2013), and sites of DNA methylation (Schübeler 2015). Activator and repressor proteins recruit enzymes, such as histone acetyltransferase or histone deacetylase, that modify histones. Histone modifications control gene expression by altering the local chromatin

structure and inhibiting or attracting DNA binding factors (Dindot & Cohen 2013). In addition, DNA methylation can repress transcription through blocking the binding of transcription factors or mediating the binding of repressors (Jaenisch & Bird 2003).

More recently, it has become possible to quantitatively investigate the 3D genome architecture. This has greatly enhanced our understanding of gene regulatory mechanisms, by showing how the three-dimensional organization of the genome influences gene regulation (Cavalli & Misteli 2013; Babu et al. 2008; Zuin et al. 2014; Dixon et al. 2016; Schmitt, Hu, Jung, et al. 2016; Lupiáñez et al. 2015). Many genes occupy preferred, non-random positions within the nucleus: In mammals, gene-poor or transcriptionally inactive regions are located close to the nuclear envelope in most cell types, whereas gene-rich or transcriptionally active regions prefer to localize at the borders of chromosome territories, away from the nuclear periphery (Foster & Bridger 2005; Nagano et al. 2013). Manipulating the position of genes can also affect their activity; for human and mouse cells, it has been shown that relocating genes from their normal position to regions close to the nuclear periphery results in gene silencing (Reddy et al. 2008; Finlan et al. 2008). On the other hand, the single-celled eukaryote *S. cerevisiae* displays a mosaic arrangement of heterochromatin and euchromatin at the nuclear periphery, with active genes located close to the nuclear pores (Casolari et al. 2004), and inactive genes associated with other parts of the nuclear periphery and the nuclear centre (Zimmer & Fabre 2011).

This organisation is achieved within a highly dynamic nucleoplasm (Lanctôt et al. 2007; Vazquez et al. 2001; Misteli 2001). For example, in mammalian cells, GFP-tagged proteins were measured to diffuse with diffusion coefficients of 0.24 - 0.53 $\mu m^2 s^{-1}$, taking 24 - 54 seconds to travel 5 μm, a distance almost equal to the radius of the nucleus (Phair & Misteli 2000). Tagged chromosomal loci in living *S. cerevisiae* cells move more than 0.5 μm, equivalent to half of the nuclear radius, within a few seconds (Heun et al. 2001).

1.2. Experimental methods for studying genome organisation

Investigation of 3D chromatin organisation by novel experimental techniques has unravelled some of the key features of this intricate system of how genome structure relates to the function of the genome. A subset of these techniques are described in more detail in sections 1.2.1, 1.2.2, and 1.2.3.

1.2.1. Microscopy

Fundamental discoveries in the field of nuclear structure and dynamics have been facilitated by microscopy techniques (Pederson 2011; Lakadamyali & Cosma 2015). Initially, electron microscopy has aided the investigation of important nuclear elements, such as nucleosomes (Oudet et al. 1975; Olins & Olins 1974; Wilson & Costa 2017), the nuclear pore complexes (Callan & Tomlin 1950; Gall 1964; Gall 1967; von Appen & Beck 2016), and the nuclear lamina (Fawcett 1966; Zwerger & Medalia 2013). More recently, the structure of transcriptionally silent and active chromatin domains (Machida et al. 2018; Wang et al. 2014; Boettiger et al. 2016; Fussner et al. 2010), and the interactions of chromatin with lamina (van Steensel & Belmont 2017; Kind et al. 2013; Kind et al. 2015) have been explored by both electron and super resolution microscopy methods.

Different fluorescence microscopy techniques, such as fluorescence recovery after photobleaching (FRAP) (Liebman & Entine 1974), fluorescent in situ hybridisation (FISH) (Pinkel et al. 1988; Lichter et al. 1988; Cremer et al. 1988), and fluorescent repressor operator system (FROS) (Robinett et al. 1996; Straight et al. 1996) have been adopted to detect the location and movement of proteins and DNA loci. By analysing the recovery of fluorescent signal over time, the binding rate of transcription factors (Mueller et al. 2008) and the binding of histone H1 protein were investigated (Stasevich et al. 2010). FISH experiments, which use fluorescent RNA probes to detect target DNA sequences, have shown that certain genes loop out of their chromosomal territory upon activation (Ferrai et al. 2010; Chambeyron & Bickmore 2004). In addition, they have indicated that the occupied volume by a chromosome can increase in the presence of transcription in human cells (Croft et al. 1999). In the operator/repressor system (FROS), bacterial repressors fused to fluorescent proteins bind to the repeats of operators that are integrated at target chromatin loci (Lassadi & Bystricky 2011) (Fig. 1.1). Using particle tracking, the mean square displacement and the spatial distribution of specific chromatin regions were measured in *S. cerevisiae* cells (Hajjoul et al. 2013; Dieppois et al. 2006; Miele et al. 2009).

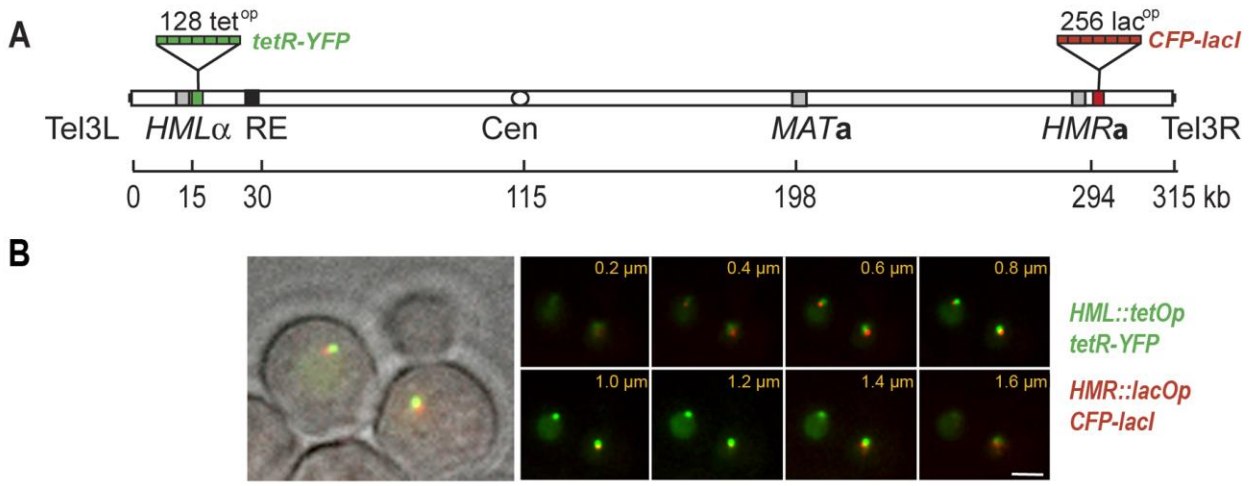


Figure 1.1. Fluorescent repressor operator system (FROS) (modified from Miele et al. 2009)**. (A)** The *lac* and *tet* operator arrays were inserted into chromosomal loci and bound by fluorescently labelled *lac* and *tet* repressor proteins. **(B)** The spatial localisation of tagged loci in live *S. cerevisiae* cells, calculated from measurements of stacks of images that were acquired at different focal depths. Bar is 2 µm.

1.2.2. Chromosome Conformation Capture Methods

The chromosome conformation capture (3C) technique (Dekker et al. 2002) and its derivatives, such as 4C, 5C, Hi-C, and capture Hi-C (de Laat & Dekker 2012; Sajan & Hawkins 2012; Han et al. 2018; Schmitt, Hu & Ren 2016), were developed to investigate the 3D structure of chromosomes by analysing the interactions between chromatin regions. These techniques rely on the re-ligation of digested DNA in fixed cells, which leads to a list of ligation products consisting of fragments in close spatial proximity. In high-throughput chromosome conformation capture (Hi-C) assay (Lieberman-Aiden et al. 2009) the library of ligation products is sequenced and mapped to the reference genome to generate the genome-wide chromatin interaction maps (Fig. 1.2A,B). The analysis of Hi-C contact maps has suggested the fractal globule structure for chromosomes in human cells (Lieberman-Aiden et al. 2009; Mirny 2011) (Fig. 1.2C), which was previously proposed by Grosberg et al (Yu. Grosberg et al. 1988; Grosberg et al. 1993). Hi-C experiments in *S. cerevisiae* cells have confirmed the Rabl configuration of their chromosomes, i.e. showing that centromeres cluster at one pole of the nucleus as well as the co-localisation of telomeres and their proximity to the nuclear membrane (Duan et al. 2010) (Fig. 1.2D). Hi-C data have also allowed the discovery (Dixon et al. 2012; Sexton et al. 2012; Nora et al. 2012) and extensive study of topologically associated domains (TAD) (Xia 2018; Fraser et al. 2015) (Fig. 1.2B). As Hi-C is performed on a population of cells, the results show the average of an ensemble of structures of several thousands or millions of cells. To overcome this averaging effect, single-cell Hi-C has been

developed. This has revealed the robust structure of domains at megabase level across the cell population, despite cell-to-cell variable inter-domain interactions (Nagano et al. 2013).

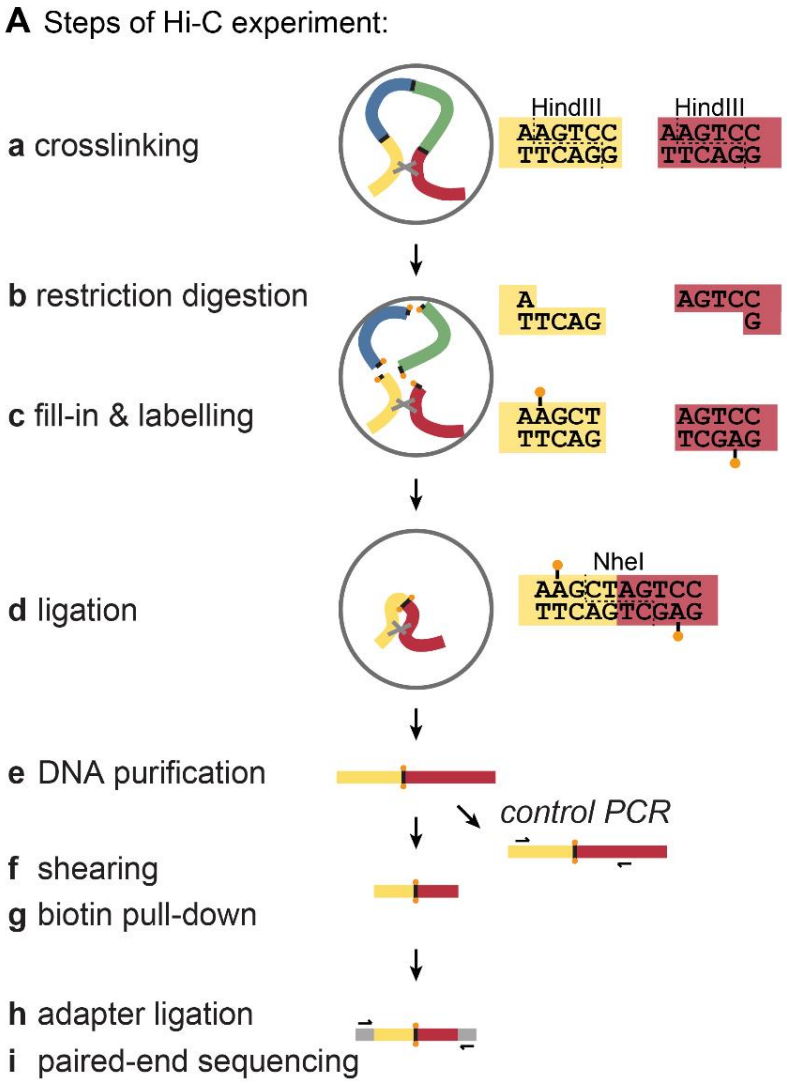


Figure 1.2B,C,D removed for copyright reasons. The copyright holders are 2015 Elsevier Inc., 2009 American Association for the Advancement of Science, and 2010 Springer Nature.

Figure 1.2. Hi-C experiment. (A) A schematic overview of the Hi-C protocol (Sewitz et al. 2017a). The result of the experiment is a list of pairwise interactions. **(B)** A part of a Hi-C contact map, which visualises the pairwise interactions between genomic loci (modified from (Ulianov et al. 2015)). A self-interacting genomic region, which has lower interactions with neighbouring regions, is called a TAD. **(C)** A schematic view of a fractal globule (Lieberman-Aiden et al. 2009). **(D)** The Rabl-like structure of chromosomes in *S. cerevisiae* was inferred from Hi-C data (Duan et al. 2010).

Hi-C data are affected by various systematic biases, such as spurious ligation products and artefacts from PCR amplification, which need to be filtered out (Schmitt, Hu & Ren 2016; Osborne &

Mifsud 2017). In addition, other biases could arise from differential length and mappability of interacting fragments and GC content at ligation junctions (Yaffe & Tanay 2011). Different software tools and algorithms have been developed to remove these biases and normalise the contact maps (Imakaev et al. 2012; Servant et al. 2015). The comparison of these tools (Forcato et al. 2017) has shown that there is not any one pipeline that outperforms the others in all different analyses. Moreover, the characteristics of output contact maps could be dependent on the selected pipeline.

1.2.3. Chromatin states

How a specific type of protein (Bina 2013; González-Mariscal et al. 2014; Lüscher & Vervoorts 2012; Bondos et al. 2015) or a histone modification (Zhang et al. 2015; Creyghton et al. 2010; Barrera et al. 2008; Deckert & Struhl 2001; Liang et al. 2004; Schübeler et al. 2004) can regulate the transcriptional activity of genes has been studied extensively for different cell types. Over the recent years, the regulatory role of combinatorial patterns of protein binding and histone marks has been investigated (Strahl & Allis 2000; Sparmann & van Lohuizen 2006; Hediger & Gasser 2006). This has led to the determination of chromatin states based on the recurring combinations of chromatin-associated proteins (Filion et al. 2010), histone marks (Ernst & Kellis 2010; Ernst et al. 2011; Kharchenko et al. 2010; Fiziev et al. 2017), or both (Liu et al. 2011; Riddle et al. 2011; Kasowski et al. 2013). In most chromatin state analyses, the chromatin immunoprecipitation (ChIP) technique has been adopted to identify protein-DNA interactions (Baker 2011). This technique relies on antibodies targeted against specific histone modifications or chromatin-associated proteins to find the bound DNA fragments (Das et al. 2004; Kuo & Allis 1999).

Different studies have shown that chromatin states correspond to distinct functional regions (modENCODE Consortium et al. 2010), which have characteristic spatial organisation (Boettiger et al. 2016; de Graaf & van Steensel 2013; Fraser et al. 2015). The determined chromatin states in human cells have allowed the study of chromatin regulatory elements, their function and their associated loci and histone modifications (Ram et al. 2011). In addition, it has been revealed that different differentiation states coincide accurately with distinct chromatin states determined by histone modifications (Larson & Yuan 2012). Our group was the first to determine chromatin states for *S. cerevisiae* (Sewitz & Lipkow 2016), by analysing the genome-wide binding pattern of 201 chromatin-associated proteins obtained from ChIP datasets (Venters et al. 2011) (section 1.4.2).

These chromatin states indicate close correlations between the protein occupancy and the transcriptional activity of genes (Sewitz et al. 2017a).

1.3. Computational models for studying genome organisation

Computational models provide the most direct way to explore mechanisms, as all components, interactions, reactions and forces are defined, and any observed behaviour must be a consequence of these. During recent years, a wide range of models of the full or partial genome has been developed to analyse the interplay of genome structure and function. In this section, I have categorised these models into three major groups: models of epigenetic modification dynamics, protein-DNA models, and polymer-based models.

1.3.1. Models of epigenetic modification dynamics

Histone proteins can be covalently modified on several residues after translation (Allfrey et al. 1964), which leads to the recruitment of transcriptional regulatory proteins and structural proteins over a local chromatin region. For example, the combined deacetylation and methylation of the lysine at position 9 of Histone H3 (H3K9) is required to create a binding site for the Swi6/HP1 silencing factor (Shankaranarayana et al. 2003; Nakayama et al. 2001). Binding of silencing factors facilitates the modification of histones on adjacent nucleosomes, and sequential rounds of epigenetic modification and protein binding leads to the spreading of heterochromatin over a chromatin region (Grewal & Moazed 2003). Specialised boundary elements inhibit the heterochromatin extension and therefore separate silent and active chromatin domains (Labrador & Corces 2002; West et al. 2002).

To understand the mechanisms behind the epigenetic memory of monostable domains, predictive models have investigated the behaviour of H3K9 methylation domains (Hodges & Crabtree 2012; Hathaway et al. 2012; Müller-Ott et al. 2014; Erdel & Greene 2016). Simulations at single nucleosome resolution showed that confined and heritable steady states of histone marks can be achieved by modelling linear propagation of histone modifications from nucleation sites to adjacent nucleosomes. Turnover of modified nucleosomes could also happen simultaneously (Hodges & Crabtree 2012; Hathaway et al. 2012). In contrast, another model assumed loop-driven spreading of histone marks with sparse nucleation sites. By adjusting parameters such as modification rates, the model was shown to be robust against replication (Erdel & Greene 2016) and the response towards transient perturbations was in line with experimental data (Müller-Ott et al. 2014).

Genomic regions of high epigenetic dynamics are bistable states, characterised by the presence of both activating and repressive histone marks (Bernstein et al. 2006) (Fig. 1.3). They have been observed for confined chromatin domains in various cell types (Rohlf et al. 2012; Tee et al. 2014). To study the features and dynamics of these states, several computational models have been developed (Dodd et al. 2007; Micheelsen et al. 2010; Dodd & Sneppen 2011; Angel et al. 2011; David-Rus et al. 2009; Sedighi & Sengupta 2007; Mukhopadhyay et al. 2010; Berry et al. 2017). In these models, a region of chromatin is represented as a sequence of nucleosomes. At every time step, each nucleosome has a state or a rate of histone modification based on its histone marks, with rules that govern state transitions or changes in rates. These models have shown that nonlinear positive feedback loops are required for robust and heritable bistable epigenetic states. Positive feedback loops arise when modifications of one nucleosome stimulate the modifications of other nucleosomes. The required non-linearity can be achieved in different ways: 1) via the cooperativity of two or more nucleosomes with the same histone marks, which recruit histone modifiers on other nucleosomes (Micheelsen et al. 2010; Dodd & Sneppen 2011; Angel et al. 2011; David-Rus et al. 2009; Dodd et al. 2007; Sedighi & Sengupta 2007; Mukhopadhyay et al. 2010); 2) through two-step feedback loops, where the switch of histone modification states of nucleosomes occurs via an intermediate state, i.e. the state first changes to the intermediate state and then to the favoured state (Dodd et al. 2007; Berry et al. 2017; Angel et al. 2011); 3) through the local transcription rate, which can be affected by silencing, in turn leading to a change in the local modification rate (Sedighi & Sengupta 2007); 4) through interactions between non-neighbour nucleosomes (Dodd et al. 2007). The role of long-range interactions in the regulation of epigenetic dynamics has been investigated using 3D polymer models of chromatin, where monomers of the same epigenetic state have self-attractive interactions (Michieletto et al. 2016; Jost & Vaillant 2018). These models have shown how the interplay between the folding of chromatin and the linear spread of epigenetic marks results in stable confined domains. Another mathematical model with a 1D array of nucleosomes has been formulated to study the dynamics of histone modification in bivalent domains, where active and repressive histone marks coexist on nucleosomes (Ku et al. 2013). These domains are important elements in stem cells, and according to the model's prediction, their formation process is generally slow. The model also suggested that a coordinated set of parameters, such as recruitment and exchange rates of marks, leads to established and maintained bivalent domains over several cell cycles.

Figure 1.3 removed for copyright reasons. The copyright holder is 2011 Elsevier Ltd.

Figure 1.3. Overview of a bistable domain (Dodd & Sneppen 2011)**.** A chain of nucleosomes with activating (blue) and silencing (red) histone marks is depicted. Readers are proteins that form bridges between nucleosomes of the same epigenetic mark. Readers can recruit writers, i.e. histone modifying enzymes, to propagate their histone marks. Barriers, such as specific DNA-bound proteins, stop the propagation of epigenetic marks.

1.3.2. Protein-DNA models

Transcription factors (TF) affect the transcriptional activity of specific genes through binding to specific DNA sequences (Ptashne & Gann 2002). It has been proposed that these proteins search for their target sequences through facilitated diffusion (Berg et al. 1981; Berg et al. 1982; Berg & von Hippel 1985), i.e. alternating rounds of 3D diffusion in the solution, sliding along the DNA, short-range excursions called hopping, and intersegmental transfer between DNA segments (Schmidt et al. 2014). The characteristics of this search mechanism have been widely studied and computational models of different scales have brought new insights into its dynamics. All models discussed in this section have focused on facilitated diffusion of TFs.

At the most detailed, atomistic level, molecular dynamics (MD) simulations have been used to explain how e.g. the *lac* repressor protein (LacI) moves along DNA (Marklund et al. 2013) and how it identifies its target site (Furini et al. 2013). LacI is modelled to take a helical path to probe the DNA, with its DNA binding interface being insensitive to modest bends in DNA conformation. The hydrogen bonds formed between the DNA and the LacI interface are dynamic and flexible, allowing fast sliding of the protein (Marklund et al. 2013). This was found to enable the protein to probe the DNA quickly and reach the proximity of the target site. Once the specific DNA sequence is bound, it becomes significantly slower, resulting in the formation of a stable protein-DNA structure and a drop in enthalpy (Furini et al. 2013; Iwahara & Levy 2013). Another fine-grained

MD simulation has proposed that binding of the CSL (CBF1/Suppressor of Hairless/LAG-1) protein to the DNA can transmit a signal through the protein structure according to the bound sequences. This influences the inter-domain dynamics of the protein and consequently its functional activities (Torella et al. 2014).

The effects of DNA conformation on the dynamics of TF proteins probing the DNA were explored via coarse-grained MD simulations, where proteins interact with the DNA via electrostatic interactions (Bhattacherjee & Levy 2014a; Bhattacherjee & Levy 2014b). The geometry of DNA was tuned by two factors, curvature and the degree of helical twisting. Highly curved or highly twisted DNA was seen to lead to a decrease in sliding frequencies and an increase in hopping events (Bhattacherjee & Levy 2014a) (Fig. 1.4A). In addition, introducing curvatures in the DNA conformation was found to increase the frequency of jumping events of a multidomain protein between distant DNA sites. However, curvature does not necessarily result in faster search kinetics as sliding happens less often (Bhattacherjee & Levy 2014b). Hence, an optimal DNA conformation can lead to a balanced number of searching events and maximal probing of DNA.

Figure 1.4A,B removed for copyright reasons. The copyright holders are 2014 Oxford University Press and 2013 American Physical Society.

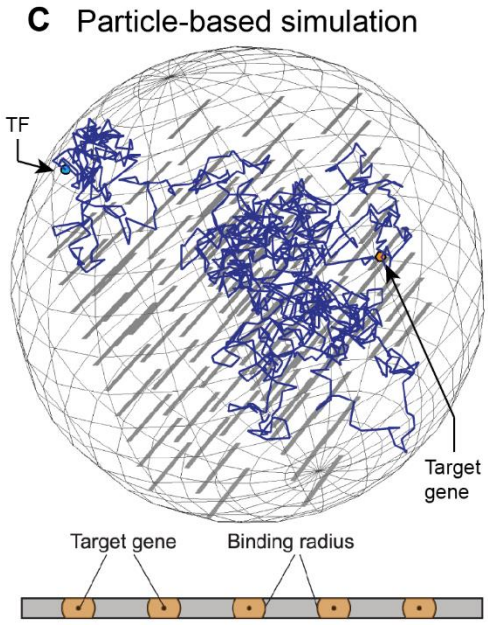


Figure 1.4. Protein-DNA models. (A) The search path of Sap1 protein along 100 bp circular DNA (modified from (Bhattacherjee & Levy 2014a)). The DNA, which is coloured in orange, was twisted to different extents. The twist of DNA affected the sliding (cyan) and the hopping (green) dynamics of the protein. **(B)** The DNA was modelled as a bead-spring chain (size of each bead = 2.5 nm) and three protein types, crowders, blockers, and searchers, were modelled as spheres (Brackley et al. 2013). The crowders had 3D diffusion, blockers had 1D diffusion along the DNA, and searchers had facilitated diffusion. **(C)**

The nucleus of *S. cerevisiae* with diffusing transcription factor (TF) (light blue particle), target genes (orange region), and DNA segments modelled as grey stacks (modified from (Schmidt et al. 2014)). The dark blue line shows the trace of the search path of the TF. To bind to the target gene, the TF had to enter the orange region, which was specified by the binding radius.

To investigate the role of nonspecific DNA-protein interactions during the search for specific target sites, Monte Carlo simulations were adopted (Mahmutovic et al. 2015; Tabaka et al. 2014; Das & Kolomeisky 2010). It was argued that the binding of the LacI repressor to nonspecific DNA is controlled by either activation or steric effects instead of being limited by diffusion (Mahmutovic et al. 2015; Tabaka et al. 2014). Furthermore, it was shown that for efficient and fast probing of DNA, moderate ranges of nonspecific binding energies and protein concentrations are required (Das & Kolomeisky 2010). The necessity for moderate DNA-protein binding strength has been indicated for proteins with different subdiffusive motions using simulations based on Brownian dynamics (Liu et al. 2017).

Large-scale computer simulations have been performed to study the search kinetics of transcription factors both in prokaryotic and eukaryotic cells. Software called GRiP (Gene Regulation in Prokaryotes) (Zabet & Adryan 2012b) provides a simulation framework for analysing the stochastic target search process of TF proteins. In GRiP the DNA is modelled as a string of base-pairs and TFs are highly diffusing components that interact with DNA sequences or with each other. This framework has been utilised to build a detailed model of facilitated diffusion, where TF orientation on the DNA, cooperativity of TFs, and crowding were incorporated (Zabet & Adryan 2012a). A similar model was adopted to dissect the effects of biologically relevant levels of mobile and immobile crowding on TF performance in a bacterial cell (Zabet & Adryan 2013): Immobile crowding fixed on the DNA raises the occupancy of target sites significantly, whereas both mobile and immobile crowding have negligible impacts on the mean search time. Another model of the bacterial genome has taken two types of crowding molecules into account (Brackley et al. 2013) (Fig. 1.4B). Proteins that bind to and move along DNA (1D crowding), do not change the search time significantly, even at very high densities. However, crowding molecules diffusing freely in 3D space increase the frequency of 1D sliding of TFs along DNA, while they enhance the robustness of the search time against any change in protein-DNA affinity.

How chromatin folding affects the dynamics of TFs has been studied by a polymer model of chromatin, where protein species could bind at both specific and non-specific sites (Cortini & Filion 2018). The results of the model were in line with experimental data: highly compact chromosome conformation decreases the concentration of TFs, while TFs colocalise at chromatin loop anchors within open chromatin regions. A different approach based on the Gillespie stochastic simulation algorithm has been developed to analyse the influence of macromolecular crowding on gene expression in stem cells (Golkaram et al. 2017). The crowding was assumed to be correlated with the local chromatin density, which was calculated using Hi-C data. Diffusive TFs and RNA polymerases were only moving in the proximity of promoters, as crowding would not allow them to diffuse to other regions between rebindings. The model predicted that an increase in chromatin density during development leads to a rise in transcriptional bursting and subsequently heterogeneous expression of genes in a cell population.

Our group has developed a computational model of TF motions in eukaryotes (Schmidt et al. 2014; Sewitz & Lipkow 2016) using the particle-based simulator Smoldyn (Andrews et al. 2010) (Fig. 1.4C). This model has considered different types of movements for TFs: 3D diffusion, sliding, hopping and intersegmental transfer. Among others, it showed the importance of inter-segmental transfer, and it provided an explanation for the size of nucleosome free regions on the DNA, which improve the process of TFs binding to their targets. Similar to a prokaryotic model (Tabaka et al. 2014) inclusion of 1D diffusion reduced the time to find the target sites by 1 - 2 orders of magnitude.

Finally, the complexity of gene regulation in higher eukaryotes has motivated the study of evolutionary dynamics of the TF repertoire and their binding preferences. A stochastic model based on duplication and mutation of genes suggested that more complex organisms with higher number of genes have higher levels of redundancy of TF binding (Rosanova et al. 2017).

1.3.3. Polymer-based models

The dynamic nature of the chromatin fibre lends itself to simulating chromatin as an extended, highly mobile polymer. Several studies have extended concepts developed in physics and applied them to the analysis of chromatin (Tark-Dame et al. 2011; Shukron & Holcman 2017; Koslover & Spakowitz 2014)**.** This has led to an understanding of genome-wide data of chromosome folding,

and their interactions with each other and with other nuclear elements. In all models presented here, the chromatin fibre is a diffusing and self-avoiding chain of beads arranged in 3D space (Fig. 1.5).

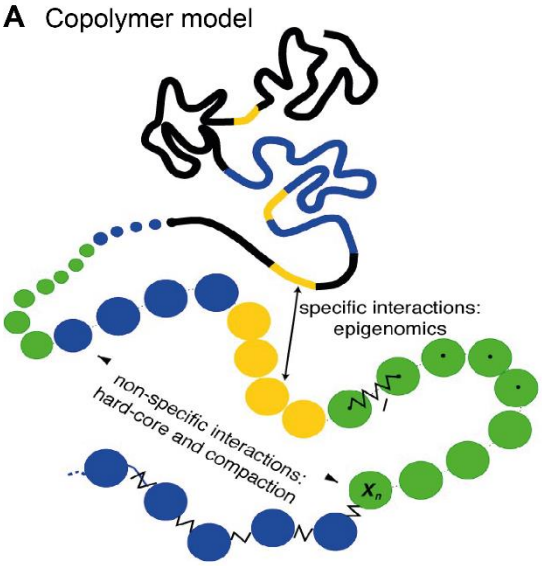
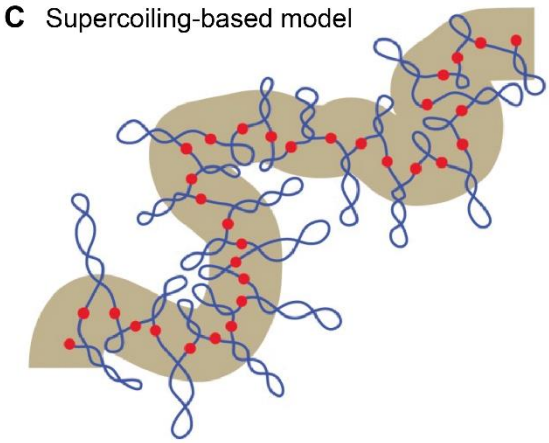


Figure 1.5B removed for copyright reasons. The copyright holder is 2015 John Wiley and Sons.



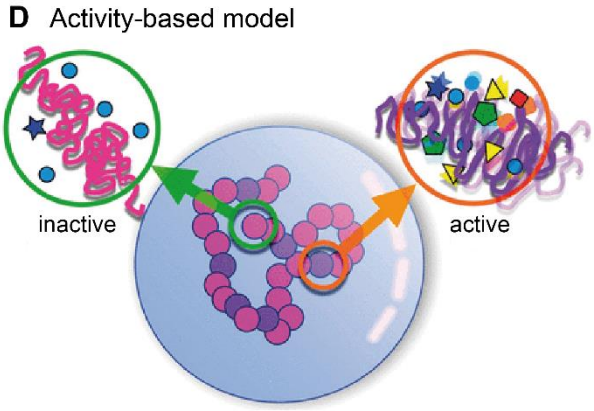


Figure 1.5. Polymer-based models. (A) In the *Drosophila* co-polymer model, the chromatin was modelled as a chain of beads connected by springs (Jost et al. 2014). Each monomer had an epigenetic state: yellow (active chromatin), green (HP1-like heterochromatin), blue (Polycomb-like heterochromatin), or black (repressive chromatin). There were specific interactions between monomers of the same epigenetic state and non-specific interactions between each pair of beads. **(B)** A model based on loop extrusion, where loop extruding factors bind to chromatin and extrude the DNA progressively until they reach the boundary elements (modified from (Imakaev et al. 2015)). This model explained the formation of TADs and the inward orientation of CTCF sites at TAD boundaries. **(C)** A schematic view of a section of chromatin composed of supercoiled domains of various sizes and different supercoiling levels (Benedetti et al. 2014). The grey tube represents the coarse-grained view of chromatin, which would behave like an elastic polymer. **(D)** A chain of active and inactive spherical monomers, which has modelled the chromatin in a confined nucleus (Ganai et al. 2014). Active and inactive states were assigned according to the gene density of

chromatin segments. The coloured polygons and circles symbolise the chromatin remodelling enzymes that increase the exerted stochastic forces on chromatin segments.

1.3.3.1. Models based on protein interactions

The important role of protein interactions in shaping the 3D structure of the genome has been investigated by various polymer models. These models can be categorised into four main groups: binders, copolymers, top-down, and loop extrusion models.

Binders: chromatin loops have been observed in both eukaryotes and prokaryotes (Hofmann & Heermann 2015) and their vital regulatory impact has been demonstrated. Models based on binder proteins have suggested that chromatin loops are formed mainly by interactions between specific protein complexes like condensin (Cheng et al. 2015) or CTCF (Tark-Dame et al. 2014). These models have successfully reproduced general aspects of the experimentally observed genome compaction. In addition, the importance of balance between short-range and long-range loops for controlling the changes in chromosomes structure has been revealed (Tark-Dame et al. 2014; Mifsud et al. 2015). It has furthermore been indicated that the dynamic bridges between condensin complexes bring about the intrachromosomal interactions during both interphase and mitosis in budding yeast (Cheng et al. 2015). Other models have relied on generic protein bridges to study the population of possible 3D genome conformations and predict the 3D distances between selected genomic sites (Barbieri et al. 2012; Brackley et al. 2016). One model has incorporated the regulatory DNA regions that are binding sites of bridging proteins (Brackley et al. 2016): the location of DNase1 hypersensitive sites, CTCF binding cites, and histone modifications were mapped onto chromatin segments. On the other hand, the strings and binders switch (SBS) model (Barbieri et al. 2012) was not informed by experimentally determined binding site of proteins and these sites were evenly distributed along the chromatin fibre.

Copolymers: other models have explored the general effects of protein interactions on chromatin structure without having protein particles in the model. A heteropolymer model incorporated proteins implicitly, by mapping different epigenetic states onto the beads. Specific interactions between beads of the same state were differentiated from non-specific interactions between any pair of beads (Jost et al. 2014) (Fig. 1.5A). This copolymer model predicted that inter-TAD interactions are highly dynamic, which was in line with Hi-C results. It also predicted the fast formation of TADs, followed by a slow and long process of compaction (Jost et al. 2014). The

lattice version of this model (Olarte-Plata et al. 2016), and another heteropolymer model (Ulianov et al. 2016) with active or inactive epigenomic states for beads, confirmed stronger self-attraction for inactive domains (Olarte-Plata et al. 2016; Ulianov et al. 2016) and an increase in their compaction as the domain size grows (Olarte-Plata et al. 2016). Other models based their assignment on levels of gene activity, with highly active or less active states assigned according to their expression levels (Jerabek & Heermann 2012). Highly active chromatin sections had low interaction strength while less active ones had higher interaction affinity. The average distances between genomic loci, the average volume ratio between highly active and less active regions, and the positioning of highly active loci close to the boundary of chromosome territories were all in line with experimental measurements.

Top-down models: polymer models based on protein interactions and without relying on pre-determined information for the state of chromatin beads were developed (Chiariello et al. 2016; Giorgetti et al. 2014; Tiana et al. 2016). Using iterative Monte Carlo simulations and comparisons to the measured contact frequencies, the parameters of the models were optimised and ensembles of chromatin configurations were achieved (Chiariello et al. 2016; Giorgetti et al. 2014; Tiana et al. 2016). These models correctly estimated the contact frequencies of TADs (Giorgetti et al. 2014; Chiariello et al. 2016) and the mean 3D distances between labelled loci upon perturbations of specific sites (Giorgetti et al. 2014). Combined with live-cell measurements, it has been suggested that changes in TAD conformations happen fast enough (in a much shorter time frame than the cell cycle) to facilitate dynamic interactions between regulatory elements, such as enhancer-promoter interactions (Tiana et al. 2016). A homopolymer model (Doyle et al. 2014), which implemented chromatin loops in the proximity of enhancer and promoter elements, indicated that the loops can either facilitate or insulate the enhancer-promoter interactions significantly. It was shown that the regulatory effect of the loop was dependent on the relative positions of loop anchors. To minimise the reliance on specific biological data, a heteropolymer model was built based on hierarchical folding and statistical physics of disordered systems (Nazarov et al. 2015). This model has two types of monomers that can interact with each other. By tuning the 1D sequence of monomers and the temperature controlling the folding, the simulated contact maps achieved a resemblance to Hi-C data.

Loop extrusion model: besides the notion that direct interactions between bound proteins shape chromatin loops, another mechanism, called loop extrusion has been proposed (Nasmyth 2001;

Alipour & Marko 2012; Fudenberg et al. 2016; Sanborn et al. 2015; Terakawa et al. 2017). This model calls for the action of extruding machines, possibly condensin or cohesin complexes, to bind and move along the DNA in opposite directions (Nasmyth 2001; Alipour & Marko 2012). This leads to the extrusion of DNA loops until domain boundaries, occupied by CTCF proteins, are reached (Fudenberg et al. 2016; Sanborn et al. 2015) (Fig. 1.5B). This mechanism can account for the compaction and folding of mitotic chromosomes (Nasmyth 2001; Alipour & Marko 2012). Furthermore, in combination with polymer physics, the model reproduced the observed decay of contact probabilities with increasing genomic distance, leading to simulated contact maps consistent with Hi-C data. It also predicted the changes in contact frequencies and 3D distances between loci due to CTCF and cohesin perturbations (Sanborn et al. 2015; Fudenberg et al. 2016).

1.3.3.2. Models based on supercoiling

Different levels of unconstrained supercoiling have been observed for chromatin (Kouzine et al. 2013; Naughton et al. 2013), and it is been reported that transcription leads to supercoiling (Papantonis & Cook 2011; Kouzine et al. 2008; Wu et al. 1988). To explore the effects of supercoiling on genome organisation in both eukaryotic (Benedetti et al. 2014) and prokaryotic (Le et al. 2013) cells, detailed polymer models have been employed. In a eukaryotic model, borders of TADs were mapped to the chromatin fibre, and strong supercoiling was imposed to the intervening chromatin (Benedetti et al. 2014) (Fig. 1.5C). This led to the formation of TADs and contact maps broadly consistent with 3C data. In a bacterial model, chromatin was simulated as a dense array of plectonemes that were attached to a back bone (Le et al. 2013). By inserting plectoneme free regions in the model at the positions of highly expressed genes, the contact frequencies observed for chromosomal interaction domains were reproduced. Overall, supercoiling is essential for creating chromosomal interaction domains (Le et al. 2013) and topologically associated domains (Benedetti et al. 2014). Intriguingly, a recent model investigated the role of supercoiling introduced by the transcribing RNA polymerase (Racko et al. 2017): When both CTCF and cohesin were included in the simulation, cohesin rings were seen to accumulate at CTCF sites demarking TAD borders. These observations are also seen experimentally (Uusküla-Reimand et al. 2016). Under these conditions, supercoiled DNA loops were extruded, and the supercoiling was the driving force for extruding the DNA loops. This is interesting because until now it was unclear how the energetically expensive loop extrusion could be achieved. Now, supercoiling generated by RNA polymerase provides a credible and testable hypothesis.

1.3.3.3. Models based on dynamical heterogeneity and self-organisation

Biological systems, such as chromatin fibre, are active matter (Agrawal et al. 2017): a collection of active units (i.e. self-propelled), which take energy from ATP-dependent internal sources (Menon 2010). Therefore, they are far from thermal equilibrium. Physical models have studied the phase-separation of active and passive particles in non-equilibrium systems (Grosberg & Joanny 2015; Stenhammar et al. 2015; Weber et al. 2016; McCandlish et al. 2012), using high local (effective) temperature (Loi et al. 2008), high diffusion speed, or self-propulsion force to drive active particles. A significant challenge in this area is to develop physical models of heteropolymeric motion applicable to chromatin. Taking into account the heterogeneous distribution of ATP-dependent activities across the genome, a heteropolymer model of chromatin was developed, in which chromatin segments that harboured more active genes were given a higher temperature (Fig. 1.5D). This model reproduced the experimentally observed chromosomal territories (Ganai et al. 2014), but only if an unphysiological temperature difference of 20-fold between active and inactive segments was assumed. A later version of this model tested 6- and 12-fold changes in temperature (Agrawal et al. 2017). The results were almost identical for different temperature settings, while the proportion of active monomers had a stronger effect on the spatial distribution of chromosomes. Using much longer chromosomal segments, similar phase separations could already be observed with much smaller differences in temperature, bringing the model in closer proximity to real life biological systems (Smrek & Kremer 2017). Still, current models are not yet fully able to deal with the structural complexity that is the hallmark of chromatin.

1.4. The aim of my PhD project

The budding yeast is a simple yet powerful model organism to study chromatin organisation (section 1.4.1). Our group has analysed the genome-wide binding pattern of chromatin-associated proteins in *S. cerevisiae* cells (section 1.4.2), which has revealed the differential protein occupancy of chromatin segments. As these differences in protein occupancy are large, and affect a large number of protein types, amounting to nearly 50% of all proteins, we asked the question of how these changes would affect chromatin mobility and consequently drive the 3D organisation of chromosomes. To test this, I have built a 3D computational polymer model, which is informed by the bioinformatics analysis of the protein occupancy across the genome (section 1.4.3).

1.4.1. *Saccharomyces cerevisiae* as a model organism

The budding yeast *Saccharomyces cerevisiae* is a single-celled eukaryotic microorganism (Fig. 1.6A). It has a unique set of properties that has made it an ideal model to study various biological processes. Firstly, there is a high degree of homology between the fundamental cellular mechanisms that govern regulation of gene expression, replication, and many other nuclear processes in yeast and higher eukaryotes. Many eukaryotic proteins like basal transcription factors (Hahn 2004), chromatin remodelling factors, and histone modifiers, are highly conserved between yeast and higher organisms (van Heusden & Steensma 2006). In addition, common features of chromatin structure and dynamics have been observed in this organism. While chromosomes fold to form higher-order chromatin structures (Lowary & Widom 1989), they have non-random interactions with each other (Rodley et al. 2009). They have Rabl-like conformations (Jin et al. 2000) (Fig. 1.6B), meaning that their centromeres are tightly clustered near the spindle pole body (Jin et al. 2000; Jin et al. 1998) (Fig. 1.6C,D) and telomeres co-localization occurs close to the nuclear periphery (Gotta et al. 1996; Trelles-Sticken et al. 2000) (Fig. 1.6E). Furthermore, *S. cerevisiae* has a small genome size, of about 12 million base-pairs on 16 chromosomes, compared to the 3 billion base-pairs of the human genome. Importantly, this makes it computationally feasible to model the dynamics of the whole genome structure in a single simulation, at gene level resolution. Experimentally, Hi-C methods give rise to accurate and high-resolution DNA-DNA contact data due to its small genome size. Finally, extensive research on budding yeast has led to a large number of databases containing experimental data including complete sequencing of the genome (Goffeau et al. 1996), genome-wide expression analysis (Holstege et al. 1998), histone modifications (Liu et al. 2005), DNA-protein binding (Venters et al. 2011) and chromatin states analysis (Sewitz & Lipkow 2016).

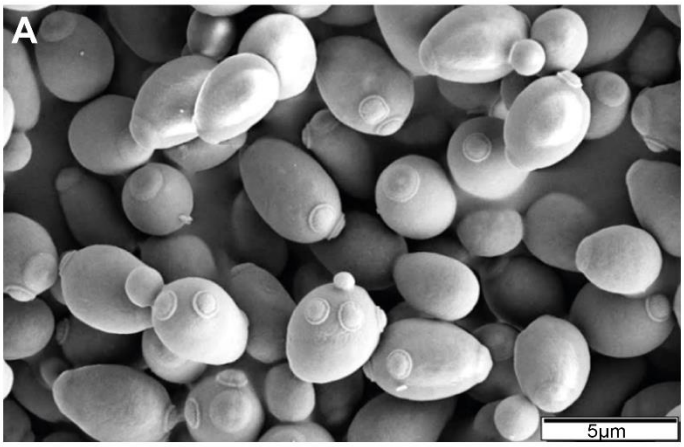


Figure 1.6B removed for copyright reasons. The copyright holder is 2012 by the Genetics Society of America.

Figure 1.6C,D,E removed for copyright reasons. The copyright holder is 2001 Elsevier Science Ltd.

Figure 1.6. Rabl-like configuration of chromosomes in S. cerevisiae. (A) Electron micrographs of a culture of *S. cerevisiae* cells (Murtey & Ramasamy 2016). **(B)** A schematic view of the Rabl-like conformation of chromosomes during interphase (Taddei & Gasser 2012). **(C)** The visualisation of centromeres (in green), which are not confined by nuclear pore proteins (in red) (Heun et al. 2001). **(D)** The localisation of centromeres close to the spindle pole body (SPB) during late G1 phase (Heun et al. 2001). Centromeres, DNA, and the SPB are visualised in green, blue, and red respectively. **(E)** Peripheral position of telomeres in a cell during G1 phase (Heun et al. 2001). The telomere repeat binding protein, Rap1p, the nucleolar protein, Nop1p, and the DNA were stained in green, blue, and red respectively.

1.4.2. Chromatin states in *Saccharomyces cerevisiae*

Our group started with an extensive set of protein-DNA-binding data, which were obtained by the chromatin immunoprecipitation (ChIP) technique from cells that were grown at 25°C, and 15 minutes after shifting the culture to 37°C (Venters et al. 2011) (Fig. 1.1A). Principal component analysis (PCA) and k-means clustering were applied to determine the chromatin states, and subsequently a multivariate Hidden Markov Model was used to map the chromatin states onto the genes (Fig. 1.1B,C). This analysis was performed separately for each temperature condition, revealing that 80% of the genes have the same chromatin state at both temperatures (Sewitz et al.

2017a). The states were sorted by the number of contained genes, and named S1 to S5, from most to least frequent (Fig. 1.1D).

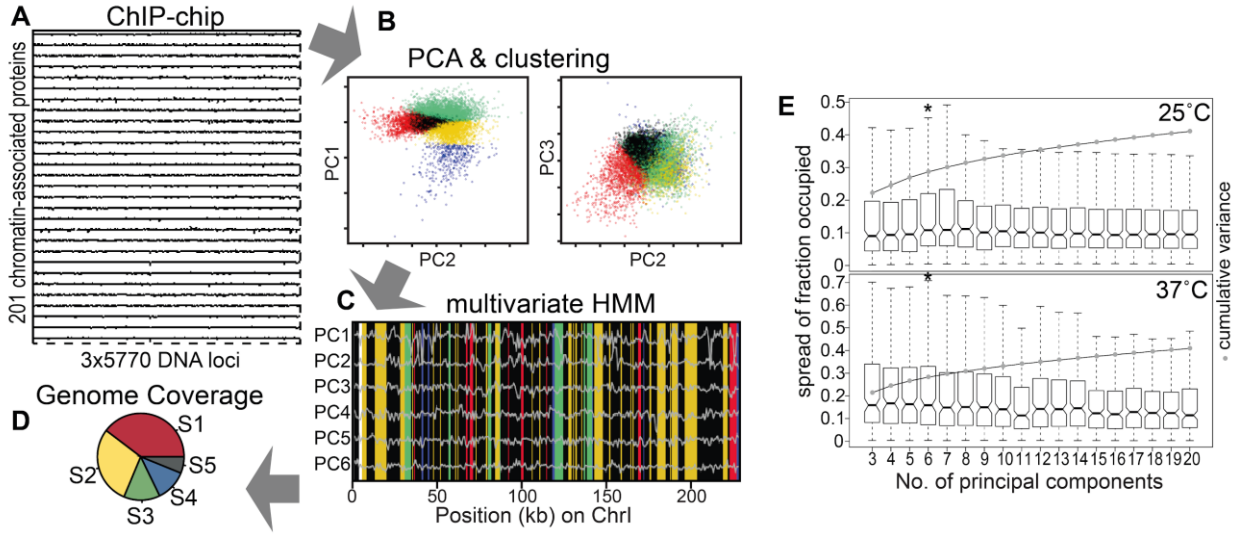


Figure 1.7. Determination of chromatin states for S. cerevisiae (Sewitz & Lipkow 2016)**. (A)** The ChIP method was used to determine the binding pattern of 201 chromatin-associated proteins. Three ChIP-probes were used per gene (Venters et al. 2011). **(B)** The protein binding data was analysed by principal component analysis (PCA); the data points along PCs are k-means clustered. **(C)** Then, a specific chromatin state was assigned to each gene using six principal component scores and a multivariate hidden Markov model (HMM). **(D)** The genome-wide coverage of chromatin states (S1-S5) was determined. **(E)** Binary HMM was employed to determine the fraction of genes in different chromatin states that are bound by a specific protein. The fractions were quantified and compared for different choices of the number of components in PCA. By selecting the first 6 principal components, the widest spread and 29% variance were achieved.

Gene ontology (GO) enrichment analysis (Sewitz et al. 2017a) (Fig. 1.2A) has shown that the genes associated to S1 are essential for basic maintenance of cellular functions. The genes in S2 cover a broad range of functions and consequently have a very limited GO enrichment. S3 and S4 genes are essential for translation during exponential growth and response to heat-shock, respectively. The genes mapped to S5 mainly regulate protein phosphorylation, amino acid metabolism, and nucleolus activity. The GO analysis is in agreement with the gene expression levels: S3 genes are highly expressed at 25°C, and they get repressed at 37°C, whereas S4 genes get highly upregulated at 37˚C (heat-shock) (Sewitz et al. 2017a) (Fig. 1.2B,C).

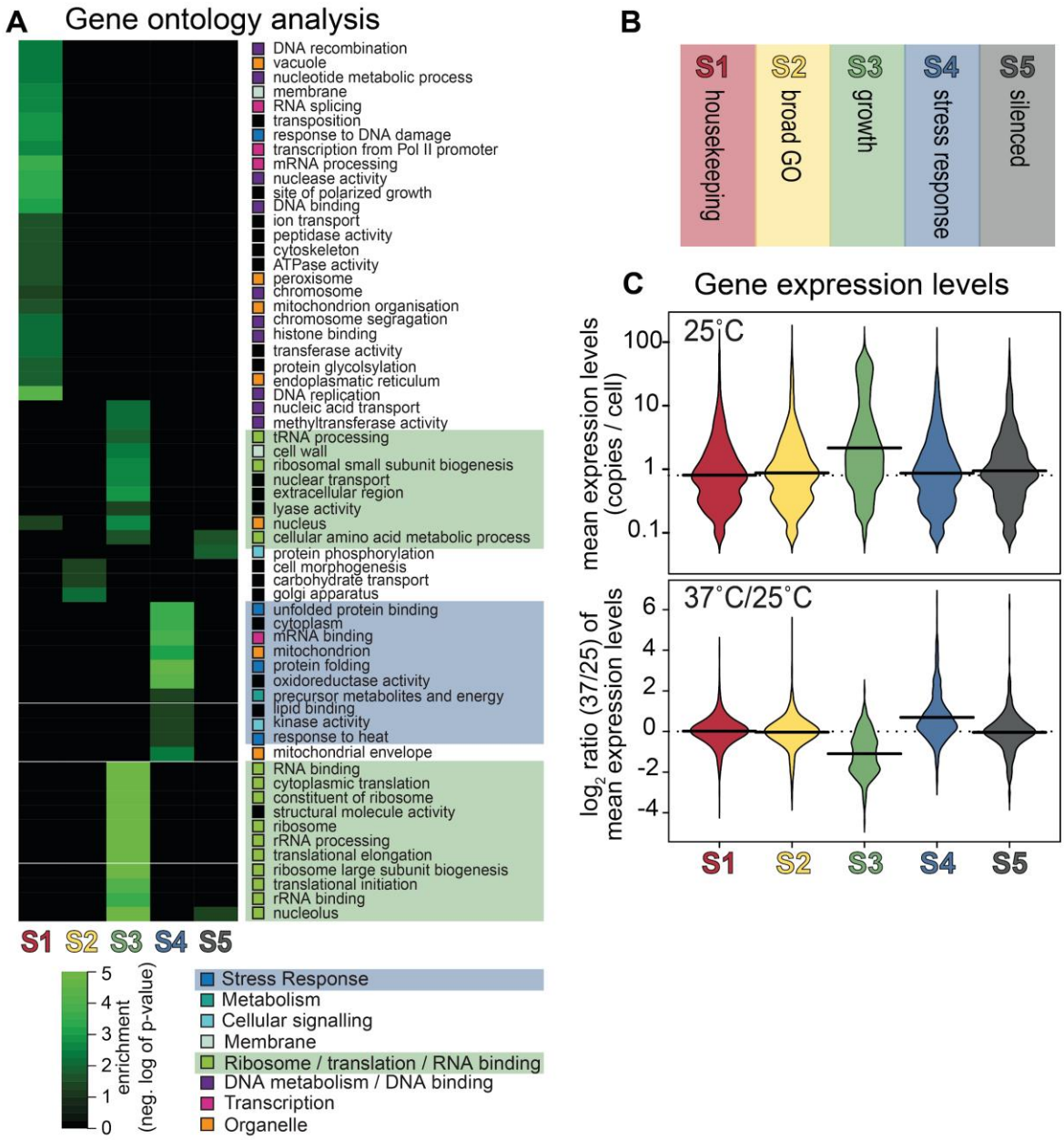


Figure 1.8. Gene ontology analysis and expression level of genes in different chromatin states (Sewitz et al. 2017a)**. (A)** Gene ontology (GO) enrichment analysis for the determined chromatin states, using Ontologizer tool (Bauer et al. 2008). **(B)** Summary of GO analysis per chromatin state. **(C)** The expression level of genes in different chromatin states, for both 25°C and 37°C temperatures. The expression array data was obtained from (Holstege et al. 1998; Zanton & Pugh 2004).

In addition, the occupancy of individual proteins at different chromatin states were measured (Sewitz et al. 2017a). The occupancy of three subunits of the RNA polymerase II complex is shown in Fig. 1.3 as an example. Rpb2p and Rpb3p are subunits of the core subassembly (Kimura et al. 1997), while a small fraction of RNA polymerases II complexes contain Rpb7p, which is essential

for heat-shock and stress response (Choder & Young 1993; Jensen et al. 1998). The analysis for the occupied fraction of the genome shows that the Rpb7p mostly binds to highly expressed genes, i.e. S3 genes at 25°C and S4 genes at 37°C, while the Rpb2p and the Rpb3p have reversed patterns (Fig. 1.3B,C). They mostly bind to the genes that will get highly expressed if the condition changes and they have the least binding to the genes with the highest expression levels. Therefore, they are in a 'poised' state, i.e. the highest binding to S4 genes at 25°C and to S3 genes at 37°C. The top rank occupancy of all proteins in the dataset (Fig 1.3D) has shown that 85 proteins (42%) bind mostly to S4 genes at 25°C. At 37°C, 68 of these proteins relocate to have the highest binding at S2 (30 proteins) and S3 (38 proteins) genes. Therefore, a significant proportion of proteins, 38 out of 201 proteins (19%), are in a poised state, and they have a concerted movement upon changes in temperature (Fig. 1.3E).

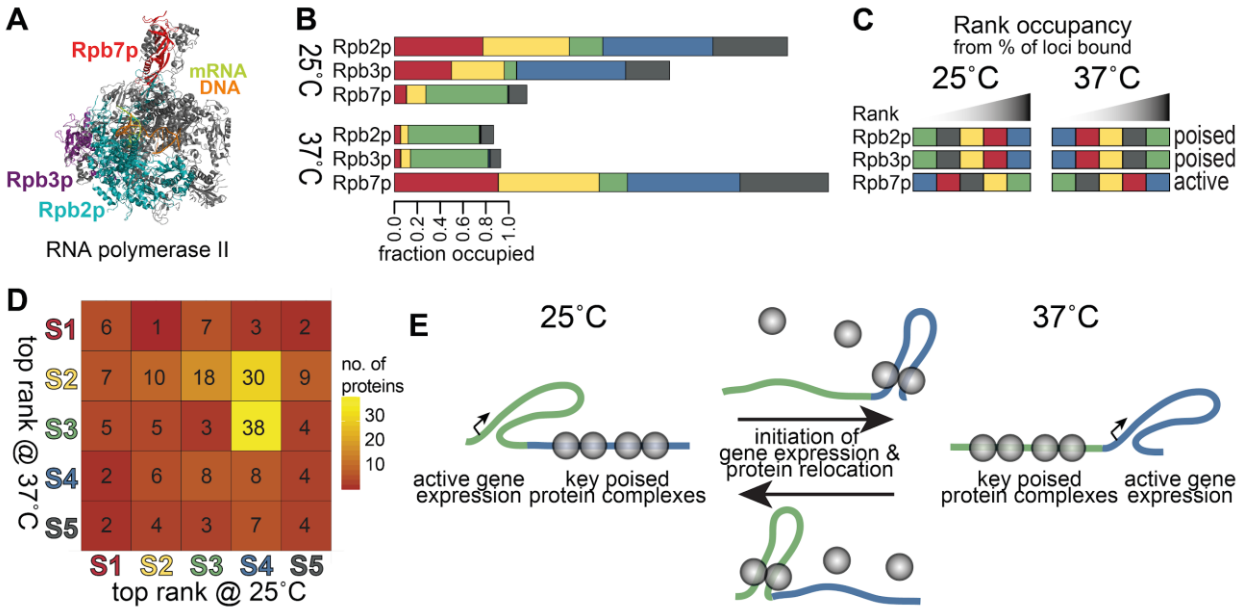


Figure 1.9. Poised proteins at normal growth and heat-shock conditions. (A) Structure of RNA polymerase II complex, which consists of 12 subunits. Three subunits, i.e. Rpb2p, Rpb3p, and Rpb7p, were in the protein binding dataset (Venters et al. 2011). **(B)** The fraction of the genome that is bound by RNA polymerase II subunits was quantified and plotted as stacked bar plots. **(C)** The fraction data was sorted to show the occupancy rank from the lowest to the highest. The Rpb2p and Rpb3p subunits are in poised state: the highest occupancy rank to S4 genes at 25°C and to S3 genes at 37°C. The Rpb7p is in an active state as it has the highest level of binding to highly transcribed genes. **(D)** The top rank occupancy for all proteins at both 25°C and 37°C temperatures. **(E)** A schematic view of the coordinated movement of poised proteins from S4 genes at 25°C to S3 genes at 37°C. The dissociation of poised proteins upon changes in temperature coincides with the high transcription of genes that were formerly bound by those proteins.

1.4.3. Aim and approach

Protein occupancy can affect the local physical properties of chromatin segments by means of a range of parameters such as changes in mass, diameter, local viscosity (Oldfield & Dunker 2014; Jirgensons 1958), diffusion speed (Jerabek & Heermann 2012; Phillip & Schreiber 2013; Wollman et al. 2017), and local electrical charge. The overall impact of these changes could hypothetically influence the local mobility of chromatin fibre. It has been shown that the transcriptional activity of genes and chromatin remodelling are associated with ATP-driven activities (Talwar et al. 2013; Hameed et al. 2012), which induce stochastic fluctuations of DNA (Weber et al. 2012). The analysis of chromatin states and the expression levels of genes (section 1.4.2) has indicated the unequal protein occupancy of chromatin segments, which is correlated with the differential transcriptional activity of corresponding genes. Highly expressed genes, which have lower protein occupancy, would have faster movements, while chromatin segments with higher protein occupancy would move slower. In this project, we hypothesized that the heterogeneous mobility of chromatin segments drives the 3D genome organisation in budding yeast during interphase. To test this hypothesis, I have built a 3D polymer model of chromosomes, in which the determined chromatin states were mapped onto the chromosomal segments. Then, the stochastic and differential mobility of segments was incorporated. Experimental data for the position of telomeres (Hajjoul et al. 2013), the distances between pairs of genomic loci, and the peripheral location of the *HSP104* gene (Dieppois et al. 2006) were employed to validate the results of the polymer model. I have also analysed our Hi-C data and compared the results with the simulated contact maps. Different predictions of the model for positioning and movement of genes could be tested by new experiments in the future. In addition, various features could be added to the model, such as mobile proteins and nuclear pore complexes.

2. Materials and Methods

2.1. Building the polymer model

2.1.1. Polymer simulations using Open Dynamics Engine

To simulate the dynamic structure of chromosomes in the *S.cerevisiae* nucleus during interphase, I developed a genome-wide heteropolymer model. The model was built by modifying a homopolymer model (Wong et al. 2012), which used a physics engine, called Open Dynamics Engine™ (ODE, http://www.ode.org/). The results of this model for the spatial conformation of genome and chromosomal contacts were consistent with microscopy and Hi-C data. In this model, each chromosome was a self-avoiding chain of rigid cylinders, connected by ball-joints, with dimensions based on those of compacted DNA as found in the yeast nucleus. Ribosomal DNA (rDNA) repeats were modelled as an insert of thicker, shorter segments. All chromosomes were tethered to the Spindle Pole Body (SPB) by attaching a cylinder representing a microtubule to the centromere-containing chromosomal segment. All cylindrical segments were moving at each time step in the x, y, and z dimensions based on Langevin Thermostat dynamics (Eq. 2.1-2.4), while their movements were constrained by ball-joints connecting them to their neighbouring segments and by the nuclear envelope. Telomeres were pulled towards the nuclear envelope by a separate force, which had a magnitude of 0.4 µN. The Langevin thermostat equations are as follows:

$$F_{LC}^{x} = c_1\sqrt{\frac{2\gamma k_B T}{\Delta t}}R(t) - c_2\gamma v^{x}(t) \quad \text{(Eq. 2.1)},$$

$$F_{LC}^{y} = c_1\sqrt{\frac{2\gamma k_B T}{\Delta t}}R(t) - c_2\gamma v^{y}(t) \quad \text{(Eq. 2.2)},$$

$$F_{LC}^{z} = c_1\sqrt{\frac{2\gamma k_B T}{\Delta t}}R(t) - c_2\gamma v^{z}(t) \quad \text{(Eq. 2.3)},$$

$$F_{LC} = (F_{LC}^{x}, F_{LC}^{y}, F_{LC}^{z}) \quad \text{(Eq. 2.4)},$$

where $\gamma = 6\pi\eta d$ and $\Delta t = \frac{m}{\gamma}$. The $R(t)$ is a random number between -1 and 1 with a uniform distribution at time t. The k_B, T, and η are the Boltzmann constant, temperature, and viscosity, respectively. The mass, diameter, and velocity of the segment at time t are denoted by m, d, and $v(t)$. In addition, c_1 and c_2 are the speeding up factors, which equal 1000 and 0.02 respectively (Wong et al. 2012).

I increased the resolution of the model by reducing the size of every segment to 2 kb, which is the approximate average length of a gene in *S. cerevisiae* (Goffeau et al. 1996), including 5' and 3'

flanking regions (see Table 2.1 for segment properties). As reported before, an increase in rDNA radius achieves spatial separation of the nucleolus from the remaining genome (Wong et al. 2012). I reduced the radius of the rDNA segments from 100 nm (Wong et al. 2012) to 50 nm, to compensate for the lower persistence length of segments. I also ran simulations with rDNA segments of 100 nm, 75 nm, and 20 nm radii. I mapped a chromatin state to each segment, as described in section 2.1.2. For all segments of a given chromatin state, the compound Langevin force, F_{LC} (Eq. 2.4), or other parameters in short test simulations, was then scaled by a factor, as described in section 3.1.

	Standard DNA segments	rDNA segments	Telomere segments	Microtubules
Number	5997	375	32	16
Length (nm)	25	8	0.25	300
Radius (nm)	10	50	12.5	12.5
Mass (kg)	4.42e-21	4.42e-21	4.42e-21	4.42e-21
Chromatin State	S1, S2, S3, S4, S5, S6 (‘gap state’), or S7 (‘NA’ state)	S7 (‘NA’ state)	S1, S2, S3, S4, S5, or S7 (‘NA’ state)	none

Table 2.1. The properties of segments in the simulations. Measurements based on (Cui & Denis 2003; Schalch et al. 2005; Dekker 2008).

2.1.2. Introducing chromatin states to the model

I used the 25°C state assignment for main simulations (Chapter 3-5), because the state assignments for 25°C and 37°C are highly similar (section 1.4.2), and because this ensured that any observed changes were only due to changes in F_{LC}. I mapped genes to segments using a length-agnostic approach, i.e., regardless of its length, each gene was associated to one 2 kb cylindrical segment. Each segment was assigned the same chromatin state as its mapped gene. The number of segments was greater than the number of genes for all chromosomes except chromosome XIV, and after mapping genes to segments, an average of 14 segments per chromosome remained unmapped. I associated those segments to the largest gaps between genes and considered them as ‘gap state’ (S6) segments (Fig. 2.1). Chromosome XIV had one more gene than the number of segments. Here, one of the genes with unknown state was ignored and all the other genes were mapped as explained above. All rDNA segments had S7 state.

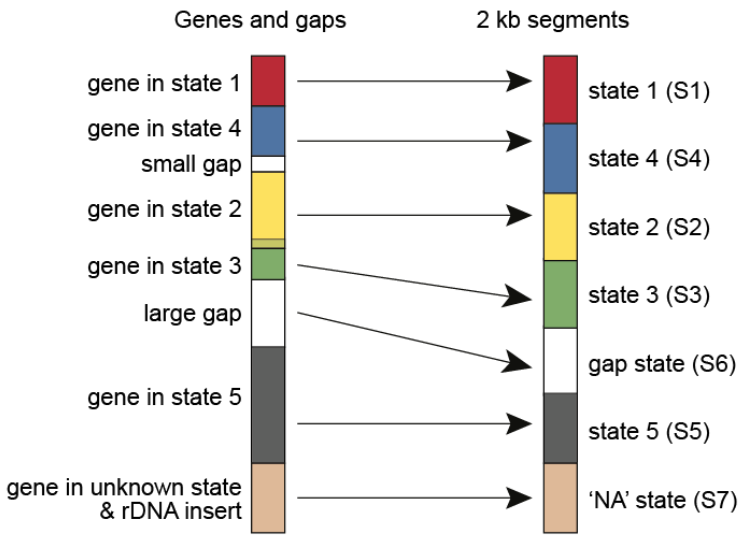


Figure 2.1. Mapping of states onto simulation segments. A schematic view of mapping the genes and the gaps between them to the 2 kb-segments in simulations. Chromatin states were mapped accordingly.

For simulations with 37°C state assignment (Chapter 6), I mapped the chromatin state of genes during heat-shock onto 2 kb-segments as explained above. In addition, for simulations with randomised and clustered states (Chapter 6), I used the 25°C state assignment to build the new lists of chromatin states. For the randomised states, the *randperm* function of MATLAB was employed to shuffle the order of states. In one set of lists, "random states (A)", all states (S1-S7) were shuffled, whereas in the other set, "random states (B)", S6 and S7 were kept at their original position, i.e. only S1-S5 were randomised. The number of states for each chromosome was preserved. For the clustered states, clusters of different chromatin states were arranged randomly along each chromosome. The size of all clusters was 10, except a few that had less than 10 repeats of states to preserve the total number of each chromatin state per chromosome. There were two types of lists for clustered states: In "clustered states (A)", S6 and S7 states were positioned at borders of S1-S5 clusters; if the number of these states was more than the number of borders, one S6 or S7 state was placed at each border and the rest were placed randomly. In "clustered states (B)", S6 and S7 states were held at their original position, which means they could appear inside S1-S5 clusters. Thus, there were two types of random states and two types of clustered states. Three different lists, per type, were generated to run three stochastic replicates of simulations. For all simulations with 37°C, random, and clustered states, the rDNA segments had the same state (S7) and the same diameter (50 nm) as in the main simulations.

2.2. Simulation process

2.2.1. Simulation input and initialisation

Each main simulation started from a configuration, in which all chromosomes were contained within the spherical nuclear envelope, without knotting. To determine the starting position of the chromosomes, I ran simulations without any state assignment (a homopolymer similar to the [uniform] model) to initialise the position of segments. At the beginning, chromosomes were vertical fibres, randomly positioned around the spindle pole body inside a capsule (as described in (Wong et al. 2012)). During the simulation, the two hemispheres at the ends of the capsule were gradually moved towards each other. After c. 350,000 time steps, the capsule converted to a complete sphere and then maintained its shape. The simulations were terminated after one million time steps and the final structure of chromosomes was used as initial conformations for all following main simulations (Chapter 3-6). Each main simulation was run in triplicate, starting from three independently obtained chromosome configurations (input 1-3).

The main simulation requires a set of input parameters to be run:

- the initial position, orientation, length and diameter for each segment;
- the position of the centromere for all chromosomes;
- the chromatin state of each chromatin segment;
- the state-specific scaling factors for changing the F_{LC}, applied to each segment of S1, S2, S3, S4 and S5.

2.2.2. Simulation process and output

Each simulation had three million time steps. To be within the limits, in which the Langevin equation is valid, the time step length was calculated from the mass of a segment and the frictional drag coefficient: $\Delta t = \frac{m}{\gamma}$ (Lemons & Gythiel 1997). At each time step, the current positions, linear and angular velocities of all segments were noted and the collisions between them were detected by the collision detection engine of ODE. Subsequently, the compound Langevin force was calculated stochastically for each segment and the resulting vector was multiplied by a state-specific scaling factor. To keep telomeres close to the periphery, additional forces were applied to push them towards the nuclear envelope. The compound Langevin and telomeric forces were passed to the ODE to be summed with the constraint forces computed internally for each segment. The constraint forces had three different sources: a) forces that kept segments connected to each

other at ball-joints; b) forces that prevented the segments to inter-penetrate at contact points; c) forces that confined the segments inside the nucleus. Next, according to the Euler semi-implicit integration scheme and using the Succesive-Over-Relaxation (SOR) method, the position and the orientation of segments at the next time step were calculated (Carrivain et al. 2014). At the end of the time step, the simulation proceeded for another step. Every one hundred steps, the collision data for the last one hundred steps, and the position and orientation of all segments were written to output files. In the text, time step may refer to the time unit, Δt, or to a specific time point in a simulation (Fig. 2.2). For example, n time steps (time units) means $n * \Delta t$ seconds, however, time step n refers to a specific point in time, i.e. at $n * \Delta t$ seconds. This point could be called a time point in the simulation as well (Figure 2.2).

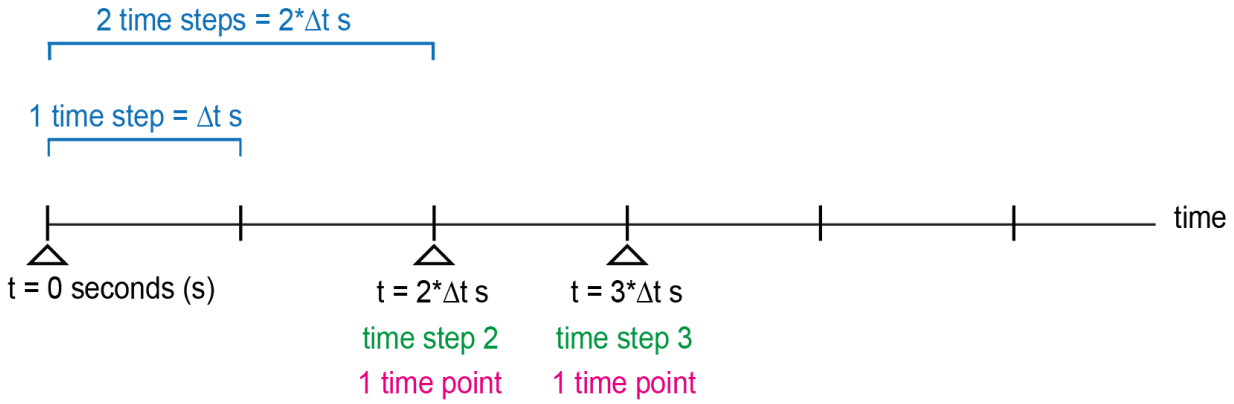


Figure 2.2. Time step and time point in simulations. The horizontal line shows time. Different examples are given to explain the notion of time step and time point in simulations.

2.3. Statistical methods

To compare the peripherality rate of telomeres positions (section 3.3.1), first the confidence intervals of the *in vivo* measurements were estimated as follows:

$$confidence\ interval = [p + Z_{0.95}\sqrt{\frac{p(1-p)}{N}} + \frac{0.5}{N}, p - Z_{0.95}\sqrt{\frac{p(1-p)}{N}} - \frac{0.5}{N}] \qquad \text{(Eq. 2.5)},$$

where p is the rate of peripheral positions of telomere and N is the total number of measurements (Tel3R: $p = 0.6876, N = 80$; Tel14L: $p = 0.6081, N = 74$). The $Z_{0.95}$ is the z-score to estimate the middle 95% of a normal distribution, and it equals 1.96. The $\frac{0.5}{N}$ term is used to correct for the fact that the binomial distribution of experimental data is approximated by a continuous normal distribution (http://onlinestatbook.com/2/estimation/proportion_ci.html).

The binomial test was also used to compare the position of telomeres in simulations with experimental data. A two-sided version of the test was employed using the *myBinomTest* function in MATLAB (input arguments: the number of occurrences of telomere within the peripheral region as measured *in vivo*, the total number of *in vivo* measurements, the rate of peripheral positions of telomere in simulations).

To compare the 3D distances between loci, we consulted the statistician Jonathan Cairns (Babraham Institute) and according to his advice, I first fitted linear regression models to estimate the relationship between the simulation and microscopy data. Then, I used the Vuong test to compare the fit of linear models. The linear models were computed by the *lm* function from the 'lmtest' package and Vuong tests were performed using the *vuongtest* function from the 'nonnest2' package in R.

Our 25°C and 37°C Hi-C data were analysed by SeqMonk (http://www.bioinformatics.babraham.ac.uk/projects/seqmonk/) at single restriction fragment resolution. The output list of interacting fragments was built by setting the following options: Min interaction strength (Obs/Exp) 1.0, Max P-value 1.0, Min absolute count 1, and correction for physical linkage. To correct for the linear proximity effect (physical linkage), SeqMonk defines a correction parameter for distance d, which equals to the average number of contacts between two fragments separated by a genomic distance d. Then, it normalises the expected contact frequencies of two fragments using the calculated correction parameters. Finally, the binomial test is adopted to determine the probability of getting more or equal the observed number of interactions, for each pair of fragments.

2.4. Hardware and software

Simulations were implemented in Python 2.7.3, calling Open Dynamics Engine v. 0.13 (https://sourceforge.net/projects/opende/) through the PyODE module v. 1.2.0 (https://sourceforge.net/projects/pyode/files/pyode/1.2.0/). Configuration files, which had defined the physical properties and the chromatin states of segments, the scaling factors for changing the F_{LC} (or other parameters), and the position of centromeres, were passed to the simulation code. Simulations were run on the Babraham Institute Computing Cluster, on Dell Intel Xeon CPU E5-4620 0, 2.20 GHz, using an average of 1.1 GB RAM. The $3x10^9$ time step simulations took c. two months.

The simulations converged within the first 10^5 time steps, as shown in Fig. 5.7. Unless stated otherwise, the figures were created from the first 10^6 time steps after convergence, to be precise from time step 99,901 to 1,099,900. In Fig. 5.1 and Fig. 5.2, the data from all 10^6 time steps were plotted, however, in other figures (Chapter 3-6) 10^4 non-correlated time points were taken by sampling every 100 time steps. To analyse the results of [uniform], [25°C], and [37°C] simulations, the data of three stochastic replicate simulations were combined. In Fig. 5.5, Fig.5.6, and Fig. A.4, the results of only one replicate are shown. In Fig. 7D, a randomly selected sample of $1x10^6$ out of the recorded $12.9x10^6$ data points, which were obtained from the triplicate simulations, was plotted.

Fig. 3.1A and Movie Vid.1 were created from the simulation output files using VMD version 1.9.1 (Humphrey et al. 1996). Chromosomes were drawn as smooth tubes, connecting the centres of segments. Data analysis of simulation output, microscopy data and Hi-C contact maps was performed with MATLAB_R2013b, MATLAB_R2017b, R 3.5.0, and Python 2.7.3.

3. Test and validation of the polymer model

Our bioinformatics analysis for binding of chromatin-associated proteins has revealed a heterogeneous pattern of protein occupancy across the genome. This heterogeneity has been determined for two temperature conditions: 25°C (normal growth) and 37°C (heat-shock). To study the effects of differential protein occupancy on the dynamic chromosome structure of *S. cerevisiae* cells, I developed a heteropolymer model of the complete yeast genome, consisting of 16 chromosomes, which contain between 230kb (Chr I) to 1531 kb of DNA (Chr IV). The first section of this chapter describes how the model was built (section 3.1). As discussed below, the mobility of each chromatin segment was altered according to its protein occupancy. The mobilities of segments were quantified and tested in both short test simulations and main simulations (section 3.2). To evaluate the results of simulations, the 3D distance distributions between various loci, locations of telomeres, and the mean square displacements of segments, were measured. Then, they were compared to analogous experimental quantifications (section 3.3).

3.1. Polymer model

To model the chromatin organisation of *S. cerevisiae* during interphase, I modified a polymer model (Wong et al. 2012) that uses Open Dynamics Engine (ODE, http://www.ode.org/). The model simulates chromosomes as chains of rigid cylindrical segments, which move based on Langevin thermostat dynamics (Lemons & Gythiel 1997; Berg 1993; Wong et al. 2012; Carrivain et al. 2014). I set the size of segments to 2 kb and mapped the determined chromatin states on to the segments (see the definition of chromatin states in section 1.4.2 and the mapping in section 2.1.2) (Fig. 3.1). At each time step, the collisions of segments are detected and recorded. To expedite simulations and to get enough collisions between segments, I used the speeded up version of compound Langevin force (Eq. 2.4), as it was introduced in the original model (Wong et al. 2012). This force is called speeded up F_{LC} or just F_{LC} in the text, and the standard version of Langevin force is referred to as standard F_{LC}. As a result of the speeding up factors, the time unit in the simulations are arbitrary units and not real seconds.

Similar to the previous model, the x, y, and z coordinates of all segments at every 100th time step are recorded in output files. To get more detailed information, I extended the code to write the movement profile of segments, i.e. linear velocity, angular velocity, linear acceleration, and angular

acceleration, to files at every 100th time steps. In addition, for three short intervals of time steps (49 time steps long), the coordinates and the movement profiles are stored at each time step.

I developed two variations of the model: [25°C] and [37°C] models, which simulate the normal growth (25°C) and heat-shock (37°C) conditions, respectively. Our determination of chromatin states has revealed the heterogeneous pattern of protein binding across the genome. Genes at state 4 (S4) are highly occupied by poised proteins at 25°C, while S3 genes are highly expressed and less occupied. At 37°C, there is a reversed situation: S4 genes are highly transcribed and less bound by proteins than S3 genes, which are heavily occupied. Protein binding changes various physical properties of chromatin segments and hypothetically affects their mobility. In addition, high expression levels coincide with more ATP-dependent activities and it has been proposed that these activities increase the thermal fluctuations of segments (see section 1.4.3). As the amount of each change is not known, I scaled the F_{LC} exerted on segments at each time step to change the mobility of segments according to their protein occupancy and expression level. Different factors were tested to scale the F_{LC}: for both [25°C] and [37°C] models for highly occupied segments, the force was decreased 2, 5, or 10 times. For less occupied segments, i.e. highly expressed genes, the force was increased by the same scaling factors at both temperatures (Fig. 3.1B). I also set a control simulation called [uniform] model, where the F_{LC} was not scaled. The heteropolymer models with 5x changes provide the best match to experimental data, and they are represented as [25°C] (F_{LC}: S3= 5x; S4=0.2x) and [37°C] (F_{LC}: S3= 0.2x; S4=5x). The radius of rDNA segments was set to 20, 50, 75, or 100 nm. The results reported in the main text (Chapter 3-6) are obtained from simulations with 50 nm rDNA segments, as they produced a good match to the experimental data (see Appendix A: Fig. A.1, Fig. A.2, Fig. A.3) and a smaller nucleolus (Fig. 3.1A), which is closer to its reported volume (⅓ of the nucleus) (Therizols et al. 2010; Léger-Silvestre et al. 1999).

Fig. 3.1A illustrates a snapshot of the modelled chromosomes at 25°C, where chromatin segments are coloured according to their chromatin state and F_{LC} is scaled by a factor of 5. Centromeres are attached to the Spindle Pole Body (SPB) by microtubules and telomeres occupy positions close to the nuclear envelope. The folding of chromatin fibre and the segregation of DNA segments and rDNA segments, which fill a significant part of the nuclear volume, are visible. Vid. 1 demonstrates a movie of the same simulation over 49 time steps.

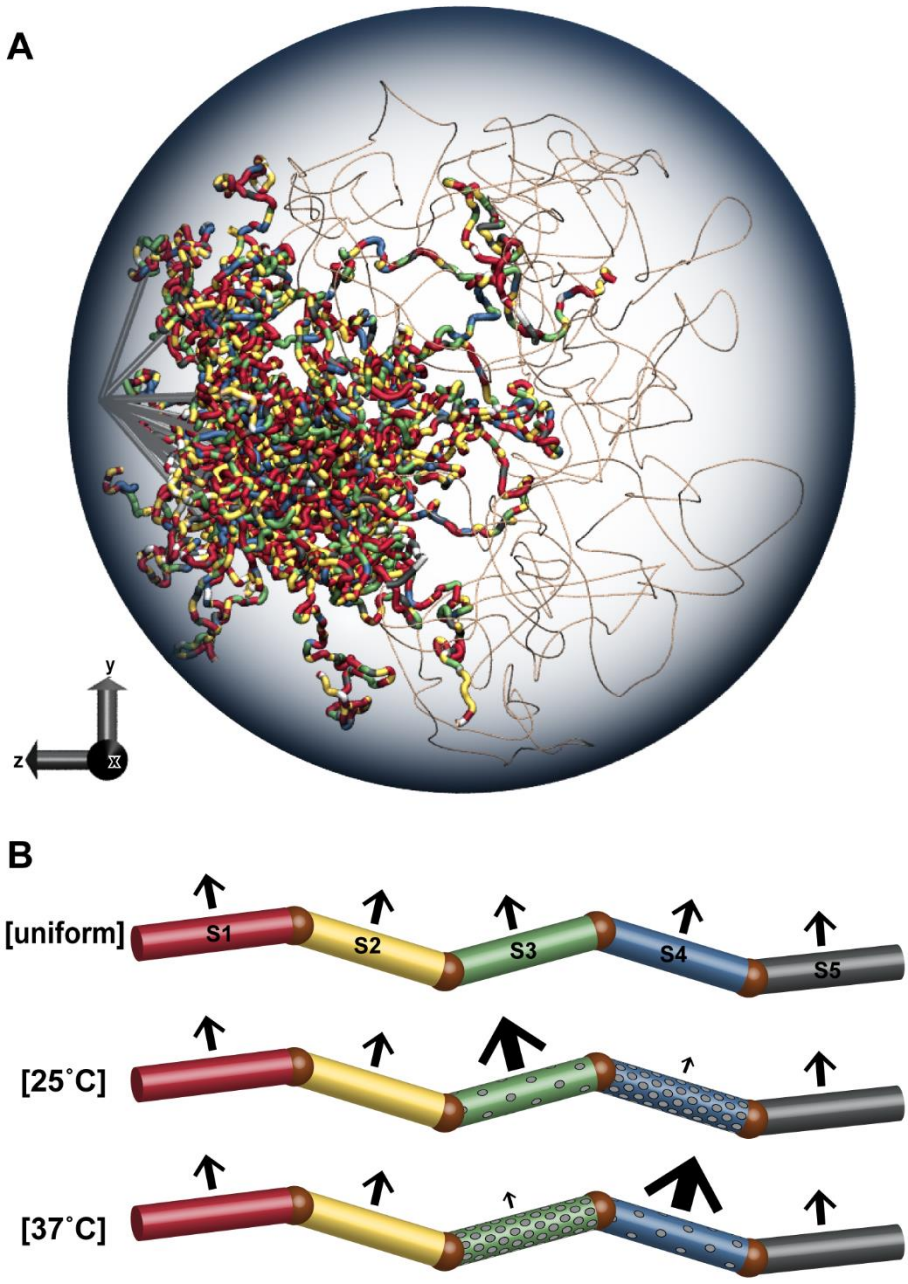


Figure 3.1. Polymer model of the S. cerevisiae genome. (A) A snapshot of dynamic chromosome structure in 25°C model, where 5x changes was applied to the F_{LC}. Chromatin segments of different states are represented by different colours. Grey coloured microtubule segments attach centromeres to the SPB and beige coloured rDNA segments model the nucleolus. The thicknesses of the segments are not drawn to scale. **(B)** The exerted F_{LC} on chromatin segments in [uniform], [25°C] and [37°C] models.

3.2. Testing the model

To analyse the changes in segments mobility in response to force alterations, short test simulations were run (section 3.2.1). Then the movements of segments were examined in main simulations

(section 3.2.2) and the autocorrelation of segment velocities were calculated (section 3.2.3). To test the main simulations, 10,000 time steps (from time step 99,901 to 1,099,900) were analysed from [uniform], [25°C], and [37°C] models.

3.2.1. Mobility of segments in short test simulations

To investigate the effect of changing F_{LC} on the mobility of segments and to examine its relationship with protein occupancy, I ran short test simulations. I made a short chromosome composed of 18 segments, where the 4th segment was connected to the SPB and the properties of 15th and 16th segments were changed (Fig. 3.2A). Simulations ran for 5000 time steps and the coordinate and movement profile of segments were stored at each time step. In one set of simulations, the applied force to segments was scaled, and in the other set, the radius (volume) and mass of segments were scaled. To keep the density of cylindrical segments constant, the radius was scaled by the square root of the scaling factor of mass. First, simulations were run using the standard F_{LC} (Fig. 3.2), then the speeded up F_{LC} was tested (Fig. 3.3), similar to the main simulations. In both Fig. 3.2 and Fig. 3.3, each pair of box plots shows the results of one simulation, the left and right plots depict the distribution of data for 15th and 16th segments, respectively.

Increasing the force leads to a clear rise in acceleration and displacement of the segment. It also raises the mobility of the neighbouring segment, even when this adjacent segment is subject to a scaled down force (Fig. 3.2B,C, Fig. 3.3A,B). Reducing the force has an opposite effect on the mobility of segments, however, the effect is less noticeable. In the other set of simulations, scaled up radius and mass make segments move more slowly, while decreased radius and mass lead to faster movements (Fig. 3.2D,E, Fig. 3.3C,D). Two neighbours of different volume and mass compete to overwrite the effect of each other on their movements, and it is not clear which one is more dominant. Overall, increasing the force has the same qualitative effect on the mobility of segments as decreasing the radius and mass and vice versa. Therefore, changing the force has the desired effect on the mobility of segments in line with their protein occupancy.

Furthermore, in simulations with speeded up F_{LC}, 5x changes in force or in volume and mass of segments result in larger differences in mobilities than 2x alterations (Fig. 3.3A,C vs. Fig 3.3. B,D). In addition, segments subject to speeded up F_{LC} move faster compared to the ones with standard F_{LC}; however, the same qualitative pattern of changes in mobility is achieved (Fig. 3.2 vs. Fig. 3.3). Therefore, speeded up simulations provide internally consistent results.

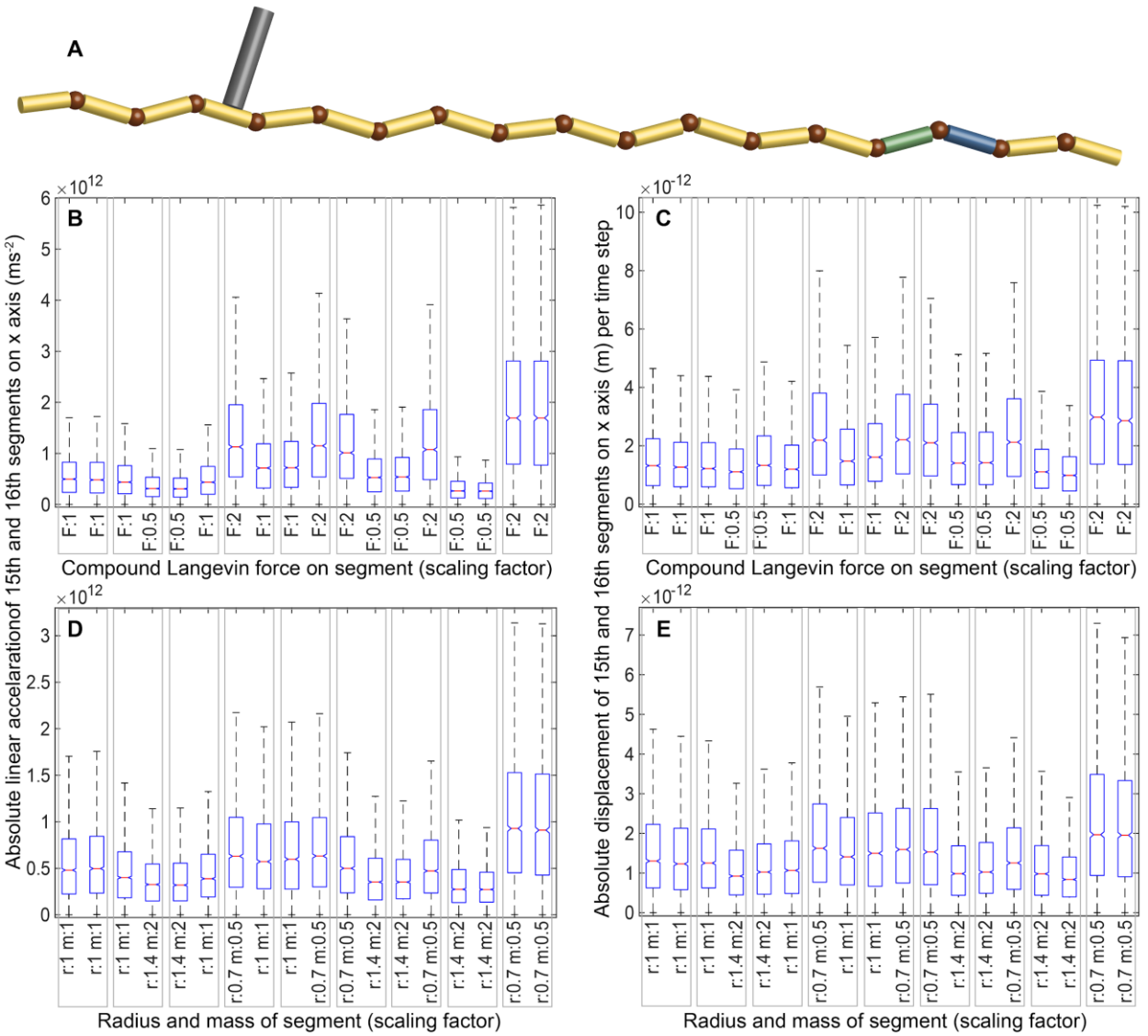


Figure 3.2. Movement of segments in short test simulations with standard F_{LC}. (A) A short chromosome with 18 segments. The 4th segment is modelling a centromere and is attached to a fixed point. The properties of 15th and 16th segments were altered. **(B,D)** The linear instantaneous acceleration of segments along the x-axis, measured at each time step. The acceleration changes by scaling the force (upper plot) or the radius and mass (lower plot). **(C,E)** The displacement of segments along the x-axis over one time step were quantified. In the upper plot, the force is scaled and in the lower one the radius and mass are changed by the given scaling factors. For all boxplots, the red line shows the median, and the outliers are not shown. The edges of boxes represent the lower and upper quartile values and the maximum whisker length is set to 1.5 times the interquartile range.

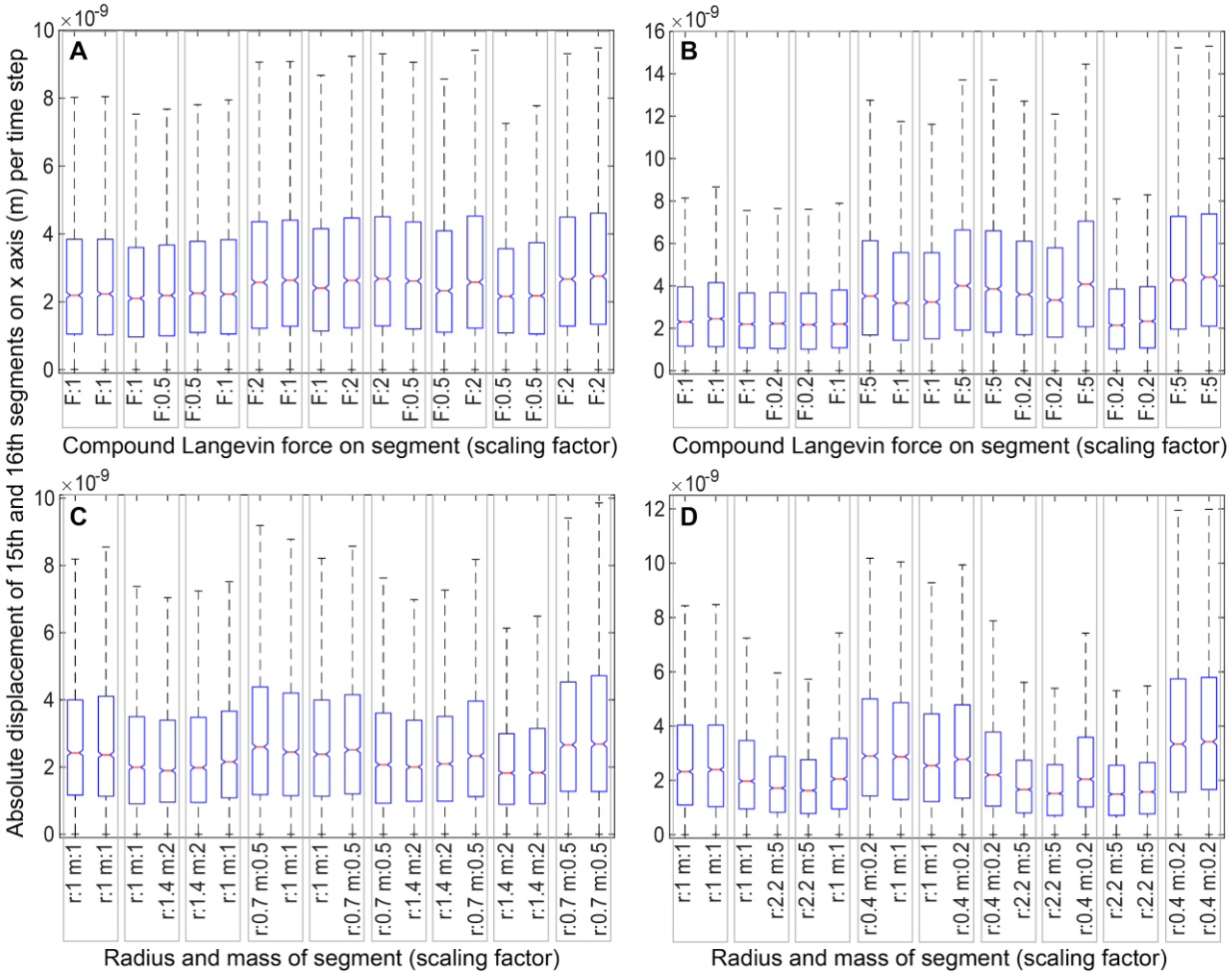


Figure 3.3. Movement of segments in short test simulations with speeded up F_{LC}. The displacement of segments along the x-axis over one time step were calculated. **(A)** The force was scaled by a factor of 2. **(B)** The force was scaled by a factor of 5. **(C)** The radius and mass of segments were scaled by $\sqrt{2}$ and 2 respectively. **(D)** The radius and mass were changed by $\sqrt{5}$ and 5 factors respectively. For all boxplots, the red line shows the median, and the outliers are not shown. The edges of boxes represent the lower and upper quartile values and the maximum whisker length is set to 1.5 times the interquartile range.

3.2.2. Mobility of segments in main simulations

To confirm that segments are moving with different speeds according to their chromatin state, the displacement of segments of chromosome XIV were measured in uniform and heteropolymer simulations. The displacement equals to the magnitude of the distance between two consecutive recorded positions. Positions were stored every 100 time steps, and the displacement was calculated from the data obtained from 10,000 time points. The distributions of mean displacement of segments reveal that segments move faster in 25°C and 37°C simulations than in uniform simulation (Fig. 3.4A). This suggests that, as observed in short test simulations, the effect of

increased force on the mobility of segments and their neighbours is more dominant than the impact of decreased force. Therefore, on average segments have faster movements in heteropolymer models, while the segments subject to scaled down forces are relatively slow. In the [25°C] model, the median of the mean displacement of all S3 segments is higher than that of other segments. In particular, the median of S3 segments is about 20 percent higher than that of S4 segments (Fig. 3.4A(b)). Similarly, the mean displacement of S4 segments has the highest median in [37°C] simulations and it is greater than the median of S3 segments by almost 19% (Fig. 3.4A(c)). In the [uniform] model, S4 and S5 segments have lower medians than S1-S3 segments (Fig. 3.4A(a)). This predicts that the chromatin states are not randomly spread along chromosome XIV, e.g. S4 and S5 segments might be closer to the immobile centromere. To confirm this prediction, I plotted the relative genomic distances of segments of each chromatin state from the centromere of chromosome XIV (Fig. 3.4B). As Fig. 3.4.B(a) demonstrates, the median of the genomic distances of S4 and S5 segments are lower than the other states. To calculate the relative distance, the linear distance of a segment from the centromere (CEN) was divided by the length of the corresponding chromosomal arm (Fig. 3.4B(b)).

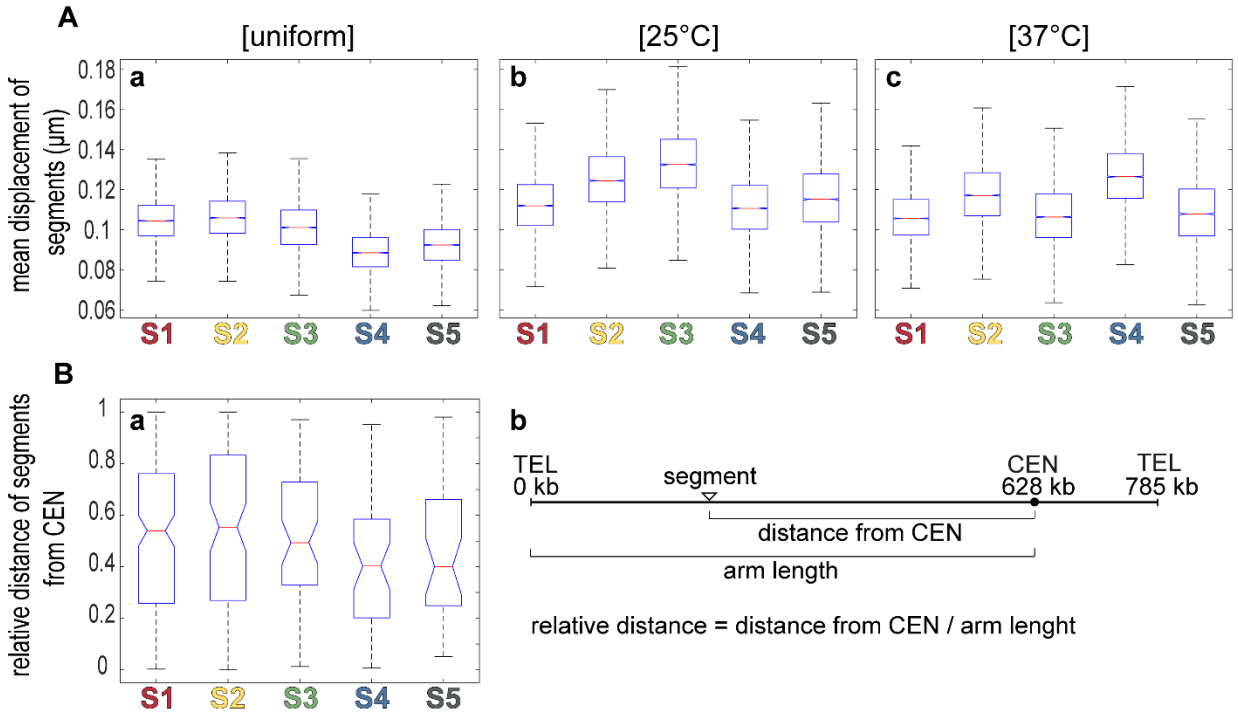


Figure 3.4. Mean displacement of segments of chromosome XIV. (A) The distributions of mean displacements were calculated for 9,999 time periods in [uniform], [25°C], and [37°C] simulations. **(B)** The relative genomic distances of segments from the centromere (CEN) was calculated for each chromatin states. For all box plots, the red line is the median. The lower and the upper edges of boxes represent the lower and

upper quartile values, respectively. The maximum whisker length is 1.5 times the interquartile range, and all points out of this range are outliers, which are not shown.

To further analyse the movement of chromatin segments, the mean square displacements (MSDs) were calculated for the same time periods as analysed for Fig. 3.4A. For all chromatin states, the lowest MSDs are seen for the segments closest to the centromere and the highest MSDs are for telomeres, which are pulled toward the nuclear periphery by an additional force (Fig. 3.5A). In the mean MSD plots, S2 has the highest mean MSD for almost all Δt values (Fig. 3.5B). For [25°C] simulations, the effect of changes to the F_{LC} of S3 and S4 segments is evident (Fig. 3.5B(b)), where the green trace is constantly higher than the blue one. However, for [37°C] simulations, the green trace starts lower than the blue one and towards the middle (after about 0.01 time units), it crosses the blue line and remains higher for the rest of Δt values (Fig. 3.5B(c)). Thus, the effect of F_{LC} alteration on movement speed is detectable at small time delays only. For larger time lags (Δt) the MSD curves behave more similar to the uniform model, where S4 and S5 segments have smaller displacements. Therefore, F_{LC} alterations drive differential movement of segments for smaller time delays. These results also suggest that the order of states along the chromosome determines their movement for larger time lags.

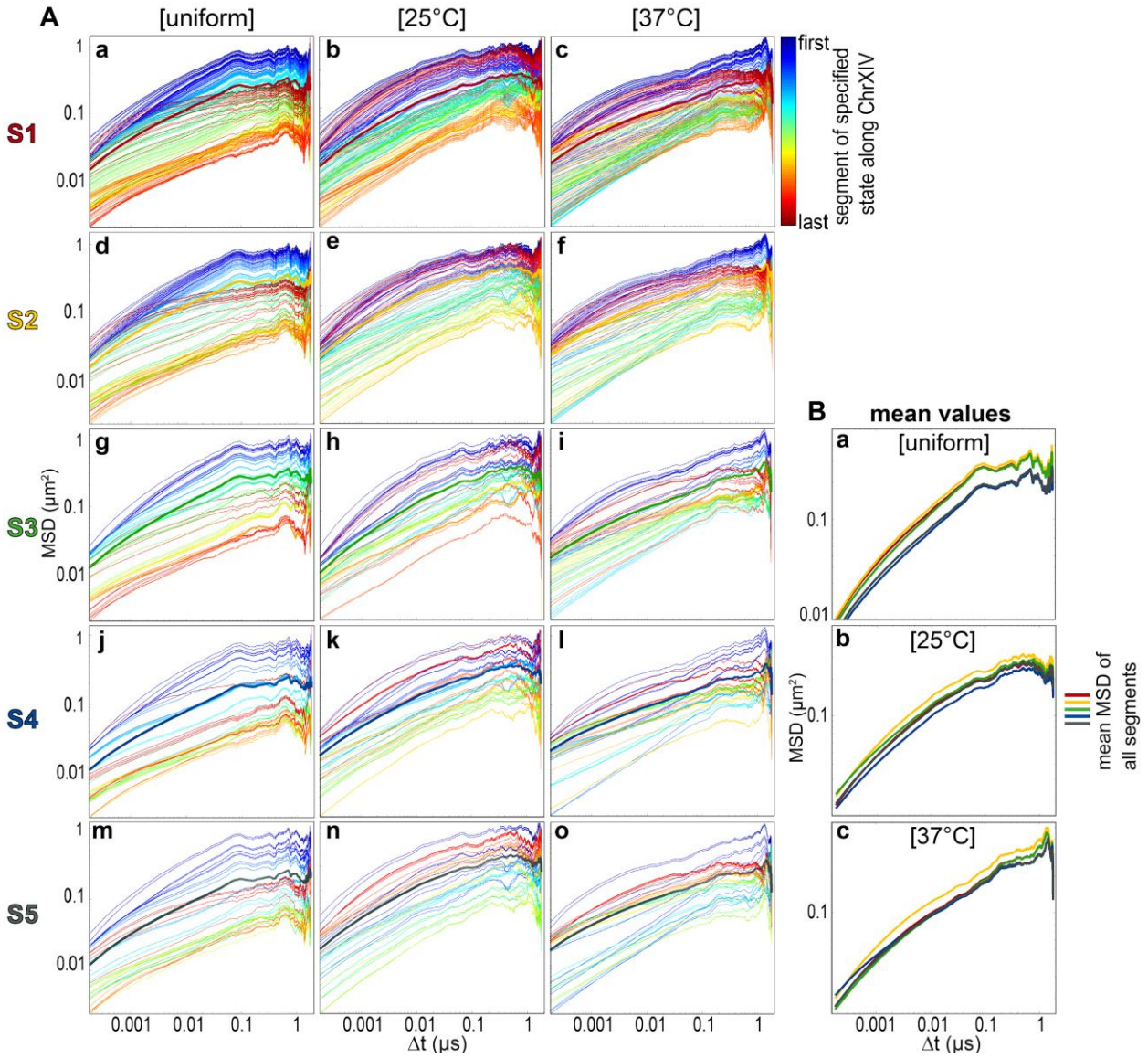


Figure 3.5. Mean square displacement (MSD) plot of chromosome XIV. (A) The MSD plots for all segments of chromosome XIV for [uniform], [25°C], and [37°C] simulations. The colours from red to blue are associated to the segments from the start to the end of the chromosome. The lowest MSDs were measured for the segments closest to the centromere, which is tethered to the spindle pole body. The thick lines represent the average MSDs and are drawn in the colour of the chromatin state. **(B)** Mean MSDs for all chromatin states of chromosome XIV. Simulations have been speeded up; therefore, the time unit of simulations is not real seconds.

3.2.3. Autocorrelation of segment velocity in main simulations

To check if the movement of a segment is not correlated at analysed time steps, i.e. the velocity of a segment at one time step is not dependent on its velocity at next analysed time step, I calculated the autocorrelation of the x-axis component of the velocity vector. From uniform and

heteropolymeric models, the velocity of two individual segments of chromosome XIV were quantified for 49 consecutive time steps and for 49 sampled time steps (sampled every 100 time steps). Then, I used the *autocorr* function in MATLAB to calculate the autocorrelation of velocity data for 48 time intervals (Fig. 3.6). There is a sharp peak at interval = 0 in all plots as the velocity vector is highly correlated with itself. The red dots at interval = 1 are higher than the blue lines in Fig. 3.6A(a-f), indicating a high degree of correlation between velocities at consecutive time steps. In sampled data, there is no autocorrelation of velocities (Fig. 3.6B(a-f)), hence, sampling the data every 100 time steps is sufficient to provide independent stochastic movement data.

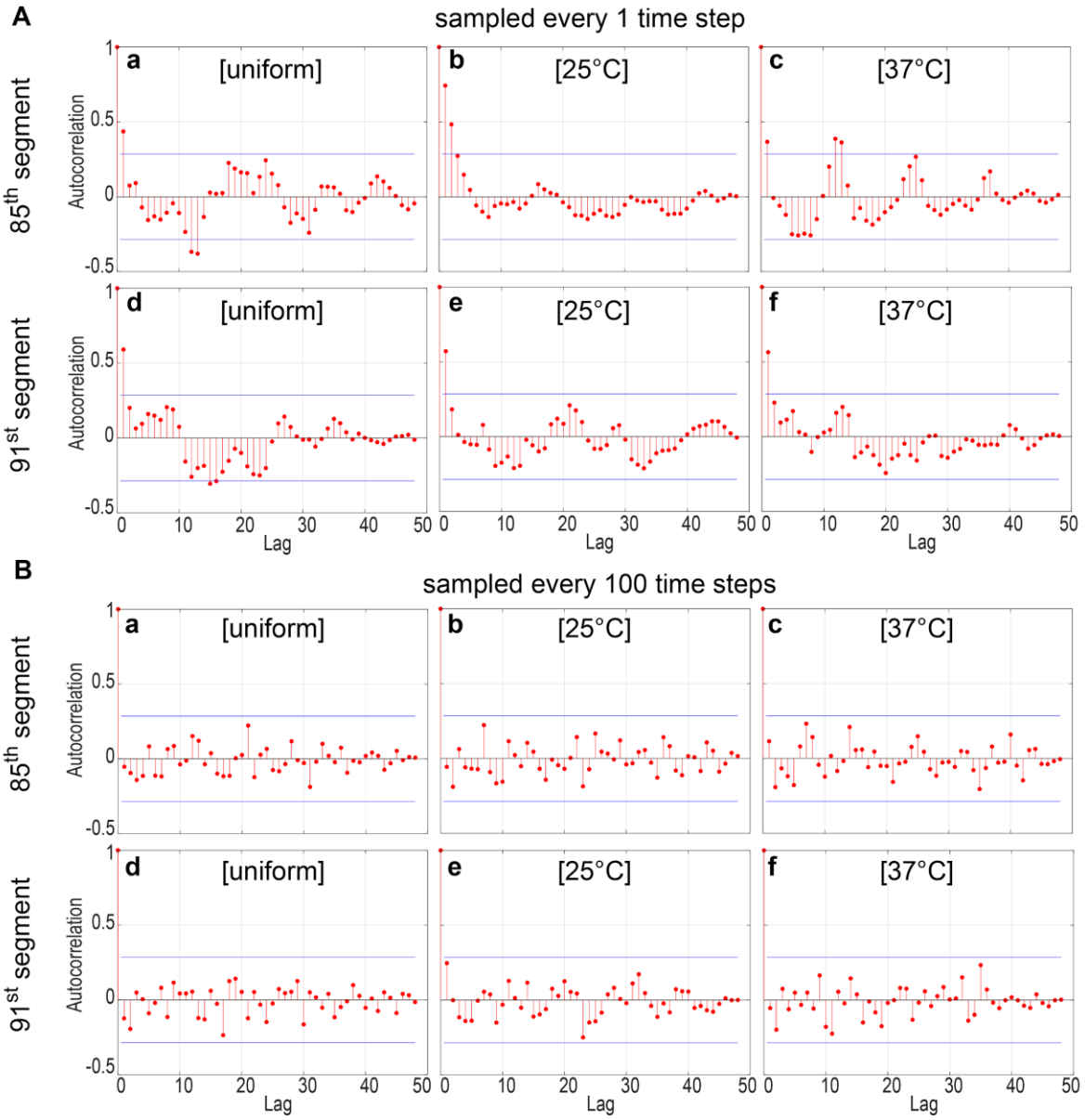


Figure 3.6. Autocorrelation of velocity data. The autocorrelation plots were calculated for the x-axis component of velocity vectors for the 85th and 91st segments of chromosome XIV. The number of time lags, i.e. temporal differences, was set to 48. The blue lines show the lower and upper 95% confidence bounds.

(A) The velocities at 49 consecutive time steps were used to compute the autocorrelation profile. **(B)** The velocities at 49 sampled time steps were tested for correlation.

3.3. Validation of the model by microscopy data

To validate the polymer model, I compared the results of simulations with experimental data from different sources: telomere positions, MSD measurements, and inter-loci distances. All computational results are obtained from 10,000 time steps per simulation, and were sampled between time steps 99,901 and 1,099,900. The position of telomeres and the distribution of inter-loci distances between pairs of loci were calculated for uniform and heteropolymeric models, where F_{LC} was scaled by a factor of 2, 5, or 10 (sections 3.3.1. and 3.3.2). For uniform, [25°C] and [37°C] models, the data of 3 independent replicate simulations were combined. Then, I measured MSD plots for two individual segments (section 3.3.3).

3.3.1. Telomere positions

To measure the peripherality rate of telomere positions, the nucleus was divided into a central and a peripheral region, such that they had equal areas in 2D projection of the nucleus. Microscopy measurements reported in (Hajjoul et al. 2013) revealed that the right telomere of chromosome III and the left telomere of chromosome XIV locate in the peripheral region with frequencies of 68.8% and 60.8%, respectively. To compare these results with simulation data, the 95% confidence intervals of *in vivo* data were determined. P-values were calculated using the binomial test for comparisons (Table 3.1). In the uniform model, telomeres are in the peripheral region most of the time (Fig. 3.7F: 94.0% and 85.3%), similar to simulations with 2x changes in their F_{LC} (Fig. 3.7C,G). The results of these simulations are outside the experimentally determined confidence intervals, indicating that they are significantly different from the experimental data (p-value<4.0e-05). Heteropolymeric models with 5x and 10x changes have less peripheral telomeres. The results of [25°C] and [37°C] simulations with 5x changes, and 25°C simulations with 10x changes are within the confidence intervals and consequently are similar to *in vivo* data. Binomial tests also showed that they are not significantly different (Table 3.1, p-value>0.1). In summary, a 5x change in F_{LC} provides a consistently good match to experimental data (see also section 3.3.2), and is more conservative and realistic than a 10x alteration. The 5x reciprocal change was therefore adopted as the standard for following analyses. The corresponding simulations are labelled with square brackets: [25˚C] and [37˚C].

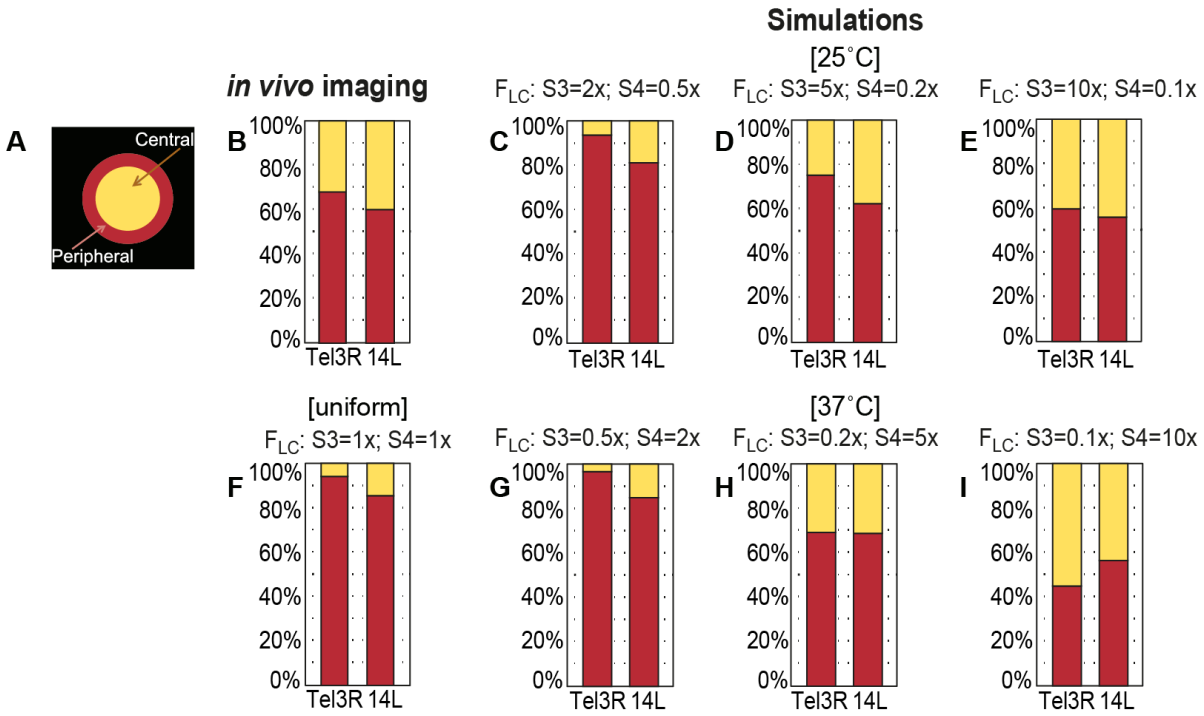


Figure 3.7. Telomere positions. (A) The peripheral and central regions with equal areas. **(B)** The positions of labelled telomeres, i.e. right telomere of chromosome III (Tel3R) and left telomere of chromosome XIV (Tel14L), were measured at room temperature. For Tel3R and Tel14L, 80 and 74 measurements were carried out, respectively. The drawn plot is modified from (Hajjoul et al. 2013). **(C-I)** Telomere positions were calculated from 10,000 time points per simulation. The data of the [uniform] model and heteropolymer models with 2x changes are significantly different from microscopy data (binomial test: $p<1.6e-11$ (Tel3R) and $p<4.0e-5$ (Tel14L)). The results of [25°C] and [37°C] simulations and 25°C simulation with 10x changes are not significantly different from the *in vivo* data (binomial test: $p>0.1$ (Tel3R) and $p>0.2$ (Tel14L)), and are within the confidence intervals (Table 3.1).

	Tel3R			Tel14L		
	n	% peripheral	p-value	n	% peripheral	p-value
experiments	80	68.7500 (55/80)	n/a	74	60.8108 (45/74)	n/a
Confidence interval (%)	[57.97 ≤ π ≤ 79.53]			[49.01 ≤ π ≤ 72.61]		
[uniform] F_{LC}: S1-S5=1x	30,000	94.0067	3.8739e-12	30,000	85.3433	2.1858e-07
F_{LC}: S3=2x; S4=0.5x	10,000	93.6000	1.5932e-11	10,000	81.1900	3.9920e-05

F_{LC}: S3= 0.5x; S4=2x	10,000	96.5500	1.5894e-17	10,000	84.8100	4.7146e-07
[25˚C] F_{LC}: S3= 5x; S4=0.2x	30,000	75.1933	0.1952	30,000	62.3233	0.8109
[37˚C] F_{LC}: S3= 0.2x; S4= 5x	30,000	69.0900	1	30,000	68.7100	0.1672
F_{LC}: S3=10x; S4=0.1x	10,000	59.6000	0.1104	10,000	55.8600	0.4146
F_{LC}: S3=0.1x; S4=10x	10,000	44.8400	2.2908e-05	10,000	56.3000	0.4828

Table 3.1. Statistical comparison of telomere positions quantified in vivo and in simulations. The binomial confidence intervals for the probability of peripheral positions were calculated from the *in vivo* data (http://onlinestatbook.com/2/estimation/proportion_ci.html). The confidence level was set to 95% and continuity correction was applied. Simulations with results within the computed confidence intervals are the best match to the experimental data. To further compare the simulation and microscopy data, the binomial tests were performed using the MATLAB file exchange function *myBinomTest*. H_0: The experimental data is taken from a population with the same distribution as the simulation data. Small p-values indicate that the distributions are significantly different.

3.3.2. 3D distance distribution of pairs of labelled loci

To gain a better insight into chromosome organisation, I analysed the physical distances between pairs of loci on chromosome XIV. The *in vivo* data was obtained using confocal microscopy (strains, images and analysis by the group of Kerstin Bystricky, Toulouse), and it was compared to the simulation data from my model. In the Bystricky group, a series of yeast strains, each with a pair of fluorescently labelled loci at different genomic distances, were created (Fig. 3.8A,B). Next, using live cell imaging the 3D distances between loci were measured. I plotted their comprehensive dataset, where the number of distance measurements ranged from 498 to 1233 per labelled strain (Fig. 3.8D).

In simulations, I identified the segments that correspond to the closest genes of the sites of the labelled insertions. Then, the 3D distances between pairs of segments were quantified for 10,000 time points, for each of the triplicate simulation runs. (Fig. 3.8E-K). In all figures, for small linear

gaps between loci (27-79 kb), increasing the linear distance results in longer 3D distances. However, there is a drop at 150 kb and then a plateau for larger genomic distances (220-495 kb). In uniform and heteropolymeric models with 2x and 10x changes in F_{LC}, the medians of box plots increase monotonically (Fig. 3.8E,F,H,I,K). The [25°C] and [37°C] models (with 5x changes) behave differently, as increasing the 1D distance from 220 kb to 319 kb does not change the distribution of 3D distances. Comparing the *in vivo* and simulation datasets for smaller genomic gaps demonstrates that experimental data have consistently higher 3D distances. This difference arises from a noise in experimental measurements (Fig. 3.8C). To statistically determine which simulation provides a better match to the experimental distance profile, I first fitted linear regression models. For each simulation, a linear model, which related the ln-transformed experimental distances to the mean of ln-transformed simulation data, was calculated. Then, to compare the fit of linear models, the Vuong test (Vuong 1989), a likelihood ratio test for closeness, was performed. The test compares the Kullback–Leibler distance of the linear models from the true model, i.e. data-generating model (Merkle et al. 2014). The results indicate that the statistical linear model derived from microscopy and [25°C] simulation data is significantly closer to the experimental data (Table 3.2) than the linear model derived from microscopy and the uniform simulation data. The 25°C simulation with 10x changes in F_{LC} is the only simulation that has a comparable fit to the fit of [25°C] simulation. However, p-values suggest that the [25°C] simulation has a better fit. Moreover, as discussed in the previous section, changing the F_{LC} by a factor of 5 is more realistic than a 10x change. Therefore, the simulations with 5x reciprocal changes, labelled [25°C] and [37˚C], were adopted as the standard for all following analyses.

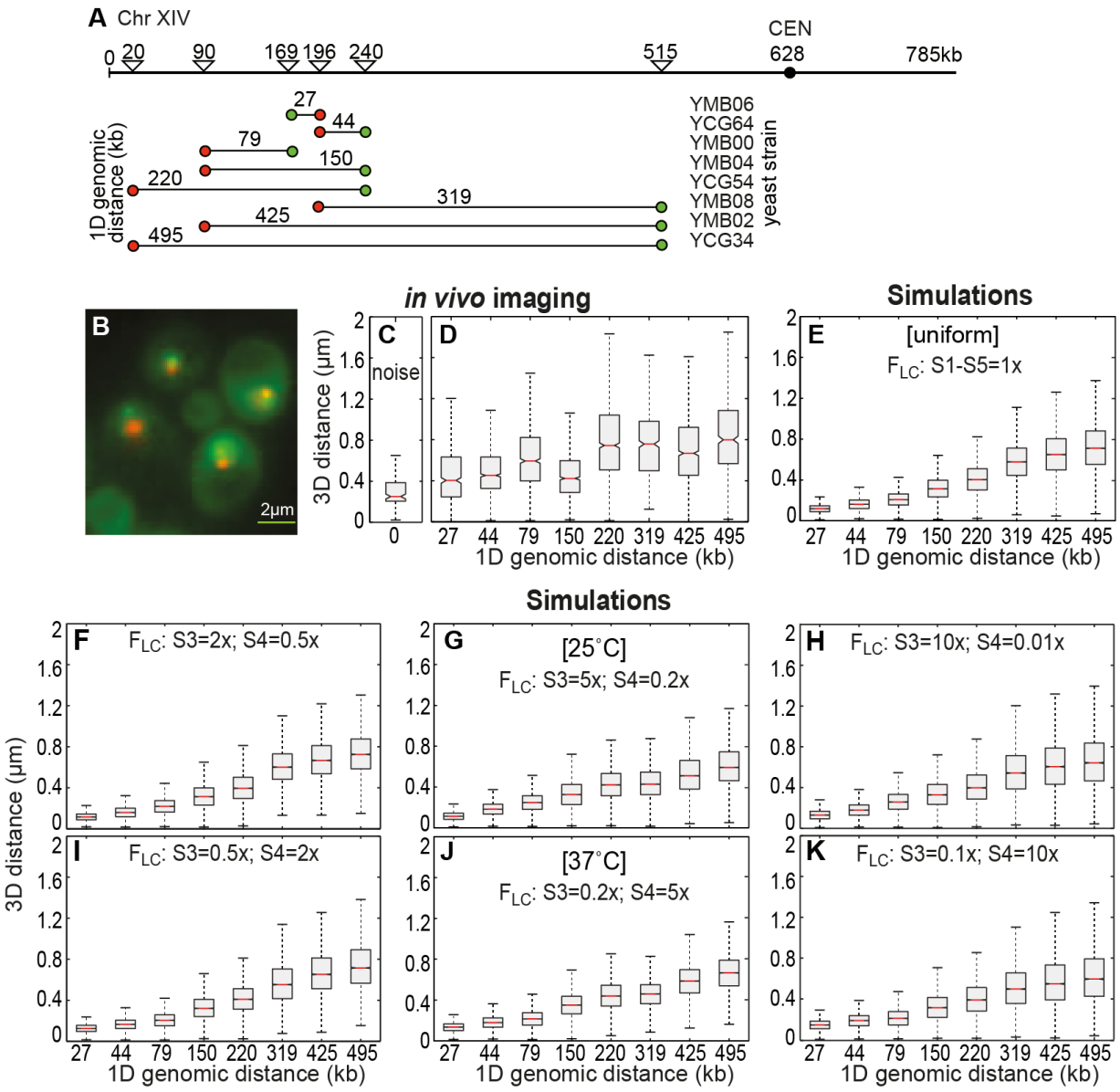


Figure 3.8. 3D distance distribution between pairs of labelled loci. (A) The positions of inserted operator arrays with fluorescent repressor operator system (FROS) on chromosome XIV for different yeast strains, and the distances (in kb) between the genomic positions of the insertions sites. **(B)** Microscopy image of strain YCG54 ("220 kb"), red and green fluorescent spots resulted from the insertion of mRFP- and GFP-labelled fusion proteins, respectively. Scale bar: 2 µm. **(C)** Control for experimental noise: The 3D distances were measured in strain YPT237 (Therizols et al. 2010), where a single array, tagged by both mRFP- and GFP-labelled fusion proteins, was inserted. **(D)** The distributions of physical distances between labelled loci for different strains; the data were measured for about 1000 frames. **(E-K)** The 3D distance distributions between pairs of marked segments for different simulations. The distances were calculated for 10,000 time steps for each simulation. For all box plots, the central red line is the median. The bottom and the top lines

of the boxes represent the 25th and the 75th percentiles, respectively. The maximum whisker length is 1.5 times the interquartile range, and all points beyond this range are outliers, which are not shown.

Linear model A:	Linear model B:	Vuong test assumption (H1):			
derived from experimental data and		Model A fits better than model B		Model B fits better than model A	
simulation dataset A	simulation dataset B	p-value	significance	p-value	significance
[25˚C] F_{LC}: S3= 5x; S4=0.2x	[uniform] F_{LC}: S1-S5=1x	1.76e-5	****	1	ns
	[37˚C] F_{LC}: S3= 0.2x; S4= 5x	1.25e-15	****	1	ns
	F_{LC}: S3=2x; S4=0.5x	0.0595	*	0.9405	ns
	F_{LC}: S3= 0.5x; S4=2x	0.000548	***	0.9995	ns
	F_{LC}: S3=10x; S4=0.1x	0.264	ns	0.7356	ns
	F_{LC}: S3=0.1x; S4=10x	0.00485	**	0.9952	ns

Table 3.2. Statistical comparison of 3D distances measured in vivo and in simulations. Linear models were derived from ln-transformed experimental data, defined as predicted variable, and the means of ln-transformed simulation data, defined as predictor variable (10,000 non-correlated data points per distance for each independent simulation). Next, the Vuong test was performed to compare the goodness of the the fit of the calculated models. Hypotheses for the Vuong test: H0: Model fits are equal; H1A: Model A fits better than Model B; H1B: Model B fits better than Model A. As small p-values indicate a better fit, according to the columns 3 and 5, the linear model derived from the [25˚C] simulation data is a better fit to the experimental data than the linear models derived from any other simulation. To run the Vuong test, the R package nonnest2, v. 0.4-1 (https://CRAN.R-project.org/package=nonnest2) (Merkle et al. 2014), was used.

3.3.3. Mean square displacement of chromatin segments

To further compare the results of simulations with microscopy data, I used MSD plots calculated by our group and that of our collaborator, Pietro Cicuta, for the 169 kb locus labelled by LacI-GFP fusion on chromosome XIV. The MSD was measured for cells held at 25°C, 30°C, and 37°C (Fig. 3.9A). I found the closest S3 and S4 segments to the 169 kb locus on chromosome XIV, and then I measured the MSD for these two segments for the [uniform], [25°C], and [37°C] simulations (Fig. 3.9C,D). Interestingly, the MSDs of both the S3 segment (169 kb) and the S4 segment (181 kb) show a similar pattern to the *in vivo* data: The same order of MSDs (for 30°C, 25°C and 37°C)

is preserved for the [uniform], [25°C], and [37°C]. For smaller time delays, the S3 segment has faster movements in the [25°C] simulation, which is expected, as S3 segments are subjected to a stronger force in this condition (Fig. 3.9C). However, the S4 segment moves slower in [37°C] simulations, despite having a scaled up force towards it (Fig. 3.9D). This suggests that the S4 segment has neighbours with S3 chromatin state that affect the magnitude of its displacements. Thus, the local arrangement of chromatin states is an important determinant of movement speed. The observation that two segments, which are in close linear proximity but belong to different chromatin states, have a very similar MSD profile, suggests that the overall arrangement of the genome has a crucial role in regulating the mobility of segments at larger time scales.

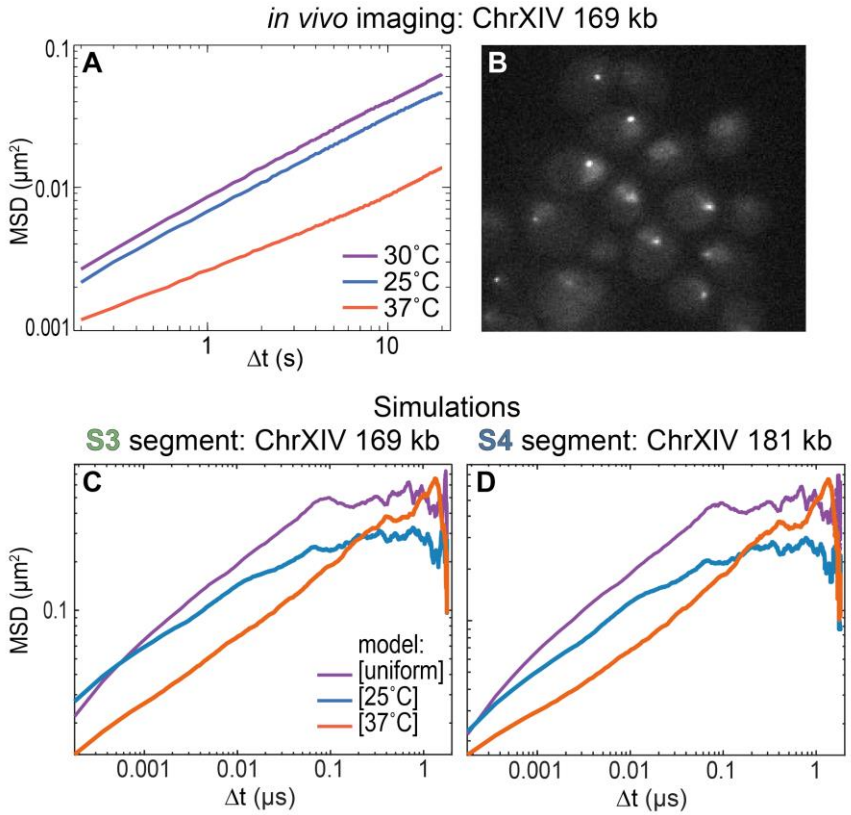


Figure 3.9. MSD plots for 169 kb locus on chromosome XIV. (A) The mean squared displacements were measured for the 169 kb locus at different temperatures (Avelino Javer and Sven Sewitz, unpublished); the data were captured from 400 frames. **(B)** This locus was labelled by LacI-GFP fusion in *S. cerevisiae* cells. The strain was provided by Kerstin Bystricky. **(C,D)** The MSDs for S3 segment at 169 kb and for S4 segment at 181 kb, for three different simulations. The trajectories were measured from 10,000 time steps. As simulations have been speeded up, the time unit of simulations is not real seconds.

4. Hi-C and simulation contact maps

To understand how chromosomes are arranged inside the nucleus, I analysed the results of Hi-C experiments performed by our group. The first step was to build contact maps, which show which pairs of chromosome fragments had been in close proximity in the yeast nucleus. Subsequently, these contacts were quantified according to the chromatin state of interacting fragments.

A common practice in Hi-C analysis is to divide the genome into equal size bins and report the number of interactions between each pair of bins (Lajoie et al. 2015). This reduces the noise in sparse datasets and provides a measure of normalisation (e.g. interactions per Megabase). However, it also reduces the resolution. In *S. cerevisiae,* the average size of a gene (2 kb) is of the same order of magnitude as an average *Hin*dIII restriction fragment (2.7 kb), the fragments of our Hi-C experiments. As our lab is interested in studying the effects of differential protein occupancy of genes, I carried out the Hi-C analysis at single restriction fragment resolution (section 4.1). As a result, the resolution of the interaction maps was comparable to the resolution of the protein binding data. Alternatively, I could have binned the protein occupancy data to make it comparable with binned interaction maps, however, this would have led to bins with an ambiguous and ineffective mixture of binding data.

For comparison, full contact maps of 2 kb segments were quantified for different simulations. These results are discussed in section 4.2. To dissect how the protein occupancy of chromatin segments influences the genome structure, I calculated the state-wise contact maps for both Hi-C and simulation data (section 4.3).

4.1. Full Hi-C contact maps

To explore the chromosome conformation during normal growth (25°C) and heat-shock (37°C) conditions, our group performed Hi-C experiments in duplicate on *S. cerevisiae* cells. To this end, we replicated the growth conditions for the protein binding dataset we had based our chromatin states analysis on: One set of cells was grown at 25°C, and the other set was grown at 25°C and then shifted to 37°C for 15 minutes. The biological replicates are distinguished by “A” or “B” affixes: 25.A, 25.B, 37.A, 37.B. During Hi-C, the DNA was digested with the *Hin*dIII restriction enzyme, which resulted in restriction fragments of an average length of almost 2.7 kb. To process the sequencing reads generated by the experiments, a bioinformatics tool called HiCUP (Wingett

et al. 2015) was used by Steven Wingett (Babraham Institute): The reads were mapped to the reference genome, then the artefacts of experiments, such as PCR duplicates and products of self-ligation or re-ligation, were filtered out. Finally, the processed data was reported as pairs of reads in BAM files. To analyse the read pairs and calculate the number of contacts between chromatin regions, two software tools were employed: HOMER (Heinz et al. 2010) and SeqMonk (http://www.bioinformatics.babraham.ac.uk/projects/seqmonk/). The results of HOMER and SeqMonk are reported in sections 4.1.1 and 4.1.2, respectively.

4.1.1. Hi-C analysis by HOMER

HOMER (Hypergeometric Optimization of Motif EnRichment) is an extensive package for processing different experimental data, such as expression arrays, ChIP-seq, or Hi-C (Heinz et al. 2010), and trusted in the field. It is capable of aligning the pair of reads, captured by Hi-C, with the genome. It can normalise the data for the genomic distance between interacting chromatin loci to distinguish the chromatin regions that are close to each other in 3D, despite their long linear distance. HOMER assumes that chromatin regions would have equal read coverage if the experiment did not have any technical biases. Thus, it corrects the data by iterative matrix balancing to reach equal coverage for chromatin sections. To quantify the number of interactions, HOMER splits the genome into bins with equal sizes.

In order to implement the analysis at the resolution of single restriction fragments, I modified the HOMER source code (written in C++). To apply the required changes, first I had to learn how the normalisation algorithms were designed and implemented. Therefore, I used the debugger tool of Microsoft Visual Studio to navigate through the extensive source code of Hi-C analysis and find the right lines to edit. Specifically, I completed the partial implementation of the *getCoverageRestrictionFragments* method of *TagLibrary* class in *SeqTag.cpp* file. This method has the main role of defining the interacting regions as restriction fragments. Other methods in the *SeqTag.cpp* file were tweaked accordingly (Table 4.1).

Method	Class	Functionality
getCoverageRestrictionFragments	*TagLibrary*	Defining chromatin regions as fragments
adjustPETagTotalsWithModel	*PeakLibray*	Iterative matrix balancing
adjustTotalsThread	*ChrPeaks*	Part of iterative matrix balancing
loadSequence	*ChrTags*	Loading the reference genome
getPETagDistribution		Part of distance normalisation
findRestrictionSites		Finding restriction fragments
makeHiCMatrix		Calculation of contact map
Initialize	*HiCBgModelChr*	Initialisation of parameters for the analysis

Table 4.1. Updated methods in HOMER. The name of methods, their class, and their functionality. They are all in *SeqTag.cpp* file and they were modified to implement the Hi-C data analysis at single restriction fragment resolution.

Next, the source code was compiled and the binary executable files, *makeTagDirectory* and *analyzeHiC*, were produced. The *makeTagDirectory* preprocesses the data and the *analyzeHiC* command executes the main analysis and normalision. The BAM files were converted to Hi-C summary files by a Python script, provided by Csilla Várnai (Babraham Institute), and they were given as input files to the *makeTagDirectory* command:

```
makeTagDirectory <output tag directory> -restrictionSite AAGCTT -genome
<reference genome file> -format HiCsummary <input read file>
```

Other input parameters were the sequence of the reference genome and the sequence that is recognised by the restriction enzyme (AAGCTT). The output was a directory of processed data, which was used by the next command:

```
analyzeHiC <input tag directory> -res site -superRes site -restrictionSite
AAGCTT <normalisation option> > <output matrix file>
```

Both -res and -superRes options set the resolution of analysis; -res defines the size of genomic bins and -superRes defines the size of the window for counting the number of interactions for each bin. I set these parameters to “site”, which means single restriction site. Two different normalisation

options were tested: -norm, which normalises for both read coverage and distance; -simpleNorm, which normalises only for the read coverage. The output file was the normalised interaction matrix.

The commands were run for all Hi-C data. A section of a full contact map of 25°C data (25.A), which is normalised for both coverage and distance, is shown in Fig. 4.1B. To demonstrate the increased resolution of Hi-C map, the contact map was calculated and plotted for 10 kb bins as well (Fig. 4.1A). The clusters of red dots indicate that the centromeric region of chromosome II has a high number of contacts with that of chromosome I. In contrasts, this region avoids contacts with arms of chromosome II, which have moderately high contact frequencies with each other (Fig. 4.1).

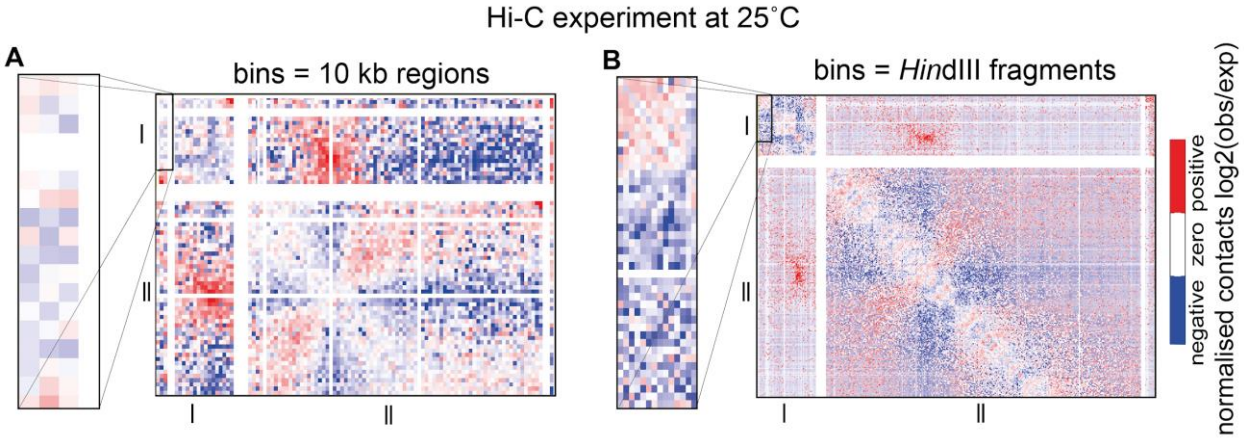


Figure 4.1. Hi-C contact map by HOMER. Sections of normalised contact maps for Hi-C 25°C data (25.A). The data is normalised for both read coverage and distance. The red dots present contacts that happened more than expected and the blue ones display the interactions that were less than expected. **(A)** The bins have equal size of 10 kb. The analysis was done by Csilla Várnai using the original version of HOMER and I plotted the output matrix. **(B)** The bins have varied size and each bin is a single restriction fragment. Comparing the zoomed-in figures on the left side of each panel shows the significant increase in the resolution by using the modified version of HOMER. TreeView program was used to visualise the matrices (https://sourceforge.net/projects/jtreeview/).

The fraction of read pairs with different distances between them was calculated by the *makeTagDirectory* command for all Hi-C results. To confirm that there is no bias in Hi-C experiments due to different temperatures, I plotted the quantified fractions (Fig. 4.2). The patterns are nearly identical; thus, different temperatures did not introduce any systematic bias in the data.

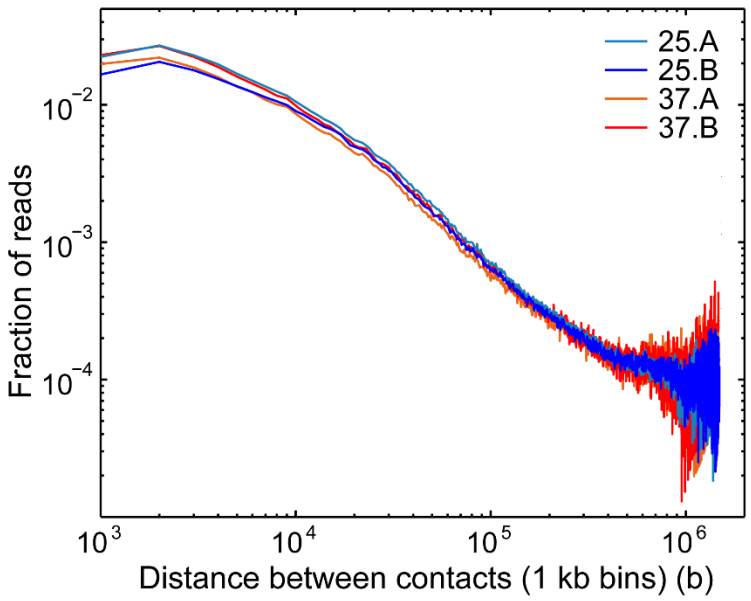


Figure 4.2. Comparison of Hi-C data for different temperatures. The fraction of Hi-C read pairs for binned distances was quantified for 25°C and 37°C data, per biological replicate. The HOMER command *makeTagDirectory* was used to analyse the Hi-C data. The results were stored in petag.FreqDistribution_1000.txt file and plotted in MATLAB.

HOMER's assumption of equal read coverage for chromatin regions holds true if chromosomes have fractal globule conformation, as observed for human cells (Lieberman-Aiden et al. 2009). However, the yeast genome has a Rabl-like configuration, and as became clear to us in the course of this analysis, the same read coverage for all chromatin regions, such as distant telomeric and clustered centromeric segments, is not expected. Thus, the read coverage normalisation is not desirable for yeast, and we decided to employ SeqMonk to repeat the Hi-C analysis (section 4.1.2). The results of HOMER analysis are still reported in this chapter to demonstrate how incorrect normalisation can affect the output data (section 4.3.4).

4.1.2. Hi-C analysis by SeqMonk

SeqMonk is a program to study high throughput sequencing data, and is capable of quantifying Hi-C read counts for pairs of restriction fragments (http://www.bioinformatics.babraham.ac.uk/projects/seqmonk/). Our Hi-C data was processed by SeqMonk and was corrected for linear proximity. This analysis was carried out by Sven Sewitz, then I received the output lists of interacting fragments. I combined the lists of two biological replicates and I plotted the calculated interaction matrices in MATLAB. Full contact map of 25°C and 37°C conditions are shown in Fig. 4.3. The high number of interactions between centromeric

regions are clearly visible at both temperatures as bright dots. During heat-shock, chromosomes tend to have more intra-chromosomal interactions, whereas at 25°C, the frequencies of intra- and inter-chromosomal contacts are comparable.

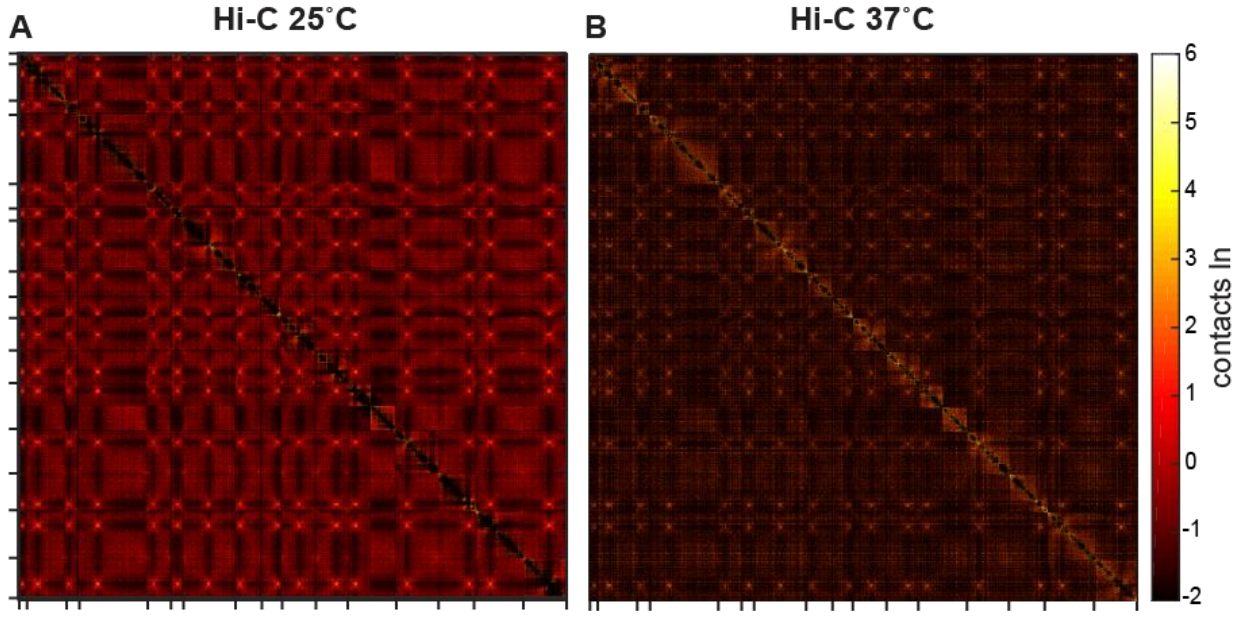


Figure 4.3. Full contact maps analysed by SeqMonk. For each dataset, the read pairs were analysed and corrected for linear proximity effect by SeqMonk. Then, for each temperature condition, the data from biological replicates were added up. Next, to make the 25°C and 37°C contact maps comparable to each other, the calculated frequencies were scaled up to reach 10 million read pairs in total. The results were log-transformed and plotted in MATLAB. In both plots, all 16 chromosomes are shown, in order from top to bottom, and left to right.

4.2. Full simulation contact maps

The contacts between chromatin segments were recorded at each time step in the simulations. I quantified the contact frequencies for 10,000 sampled time steps, per simulation, and plotted them in MATLAB (Fig. 4.4). As segments have excluded volume, adjacent segments cannot collide and consequently there are 3 diagonal lines of zeros in calculated contact matrices. The [uniform] model has the least number of total contacts, suggesting that the homopolymer is not as compact as heteropolymeric models. The total number of contacts in [37°C] is 1.2 times higher than that of [25°C], indicating that the model has a more compacted genome conformation at heat-shock conditions. Clusters of bright dots illustrate high contact frequencies for certain pairs of segments, such as centromeres and linearly proximal segments. To be consistent with Hi-C contact maps,

which do not contain the repeats of rDNA, I have omitted the rDNA segments from the full contact maps.

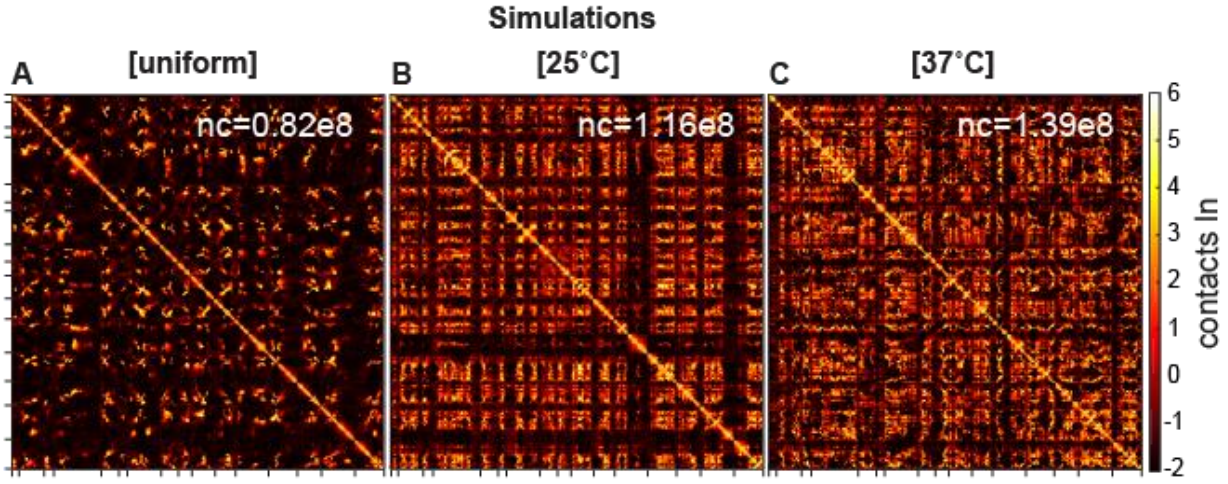


Figure 4.4. Full contact maps of simulations. The contact frequencies were calculated in [uniform], [25°C], and [37°C] simulations, every 100 time steps, from time step 99,901 to 1,099,900. Then, the contact maps of three simulation replicates were added up, per condition. The ln-transformed matrices were plotted in MATLAB. The total number of quantified contacts is written at the top of each plot. For all plots, the rDNA segments are excluded.

4.3. State-wise contact maps

To investigate how contacts are distributed across different chromatin states, state-wise contact maps were calculated. For Hi-C data, first the chromatin states were mapped to restriction fragments (section 4.3.1). Then, the normalised state-wise contact maps were computed for both Hi-C and simulation data (section 4.3.2). To analyse the Hi-C 25°C and 37°C data, the 25°C and 37°C state assignments were used respectively. For all simulation data, the state assignment of 25°C was used. The results are compared in section 4.3.3. In addition, the state-wise contact maps calculated from HOMER reports are discussed in section 4.3.4.

4.3.1. Mapping the chromatin states onto *Hin*dIII fragments

To map the chromatin states of genes onto the *Hin*dIII restriction fragments, I first had to resolve the overlaps between genes. Therefore, I assigned different influence levels to various sections of genes according to the transcriptional start site (TSS), upstream activator sequence (UAS), and open reading frame (ORF) probes. Since the TSS probe was used to determine the state of a gene, it was given the highest priority for state assignment. The UAS probe was given the second highest priority, and the ORF probe the lowest priority (Fig. 4.5A). Thus, for a gene on the Watson strand,

the sequence between the start of the TSS and the start of the ORF was given the highest influence level. The stretch of DNA from the start of the UAS to the start of the ORF had the medium influence level. And the lowest level was assigned to the section between start of the UAS and the end of the ORF (Fig. 4.5A). Then, the states were assigned according to the following rules: Overlaps between sections of the same level of influence resulted in assigning the state of both sections to the relevant fragment (rule 1). If a higher-level section and a lower-level section overlapped with the same *Hin*dIII fragment, the fragment would get the state of the higher-level section (rule 2). To implement these rules, I used the coordinates of probes provided by (Venters et al. 2011). I updated the coordinates according to the reference genome sequence R.64.1.1 (released on 03/02/2011) (Engel et al. 2014).

The algorithm of assigning chromatin states to fragments, as described below, led to a list of percentages of chromatin states for each restriction fragment (Fig. 4.5B). For gaps between genes, state 6 (S6) was assigned.

1. Mapping of the lowest influence level sections for each chromosome:

1.1. First the overlaps between these sections were identified.

1.2. Then states were mapped to fragments. Where there was an overlap, the states of both overlapping sections were mapped (rule 1).

2. Mapping of the medium influence level sections for each chromosome:

2.1. Similar to the previous step, overlaps of regions composed of UAS and TSS probes were detected.

2.2. Where there was a higher-level region on the top of a lower level one, the states of corresponding fragments were updated (rule 2).

3. Mapping of the highest influence level sections for each chromosome:

3.1. Overlaps between TSS probes were located.

3.2. Similar to 2.2.

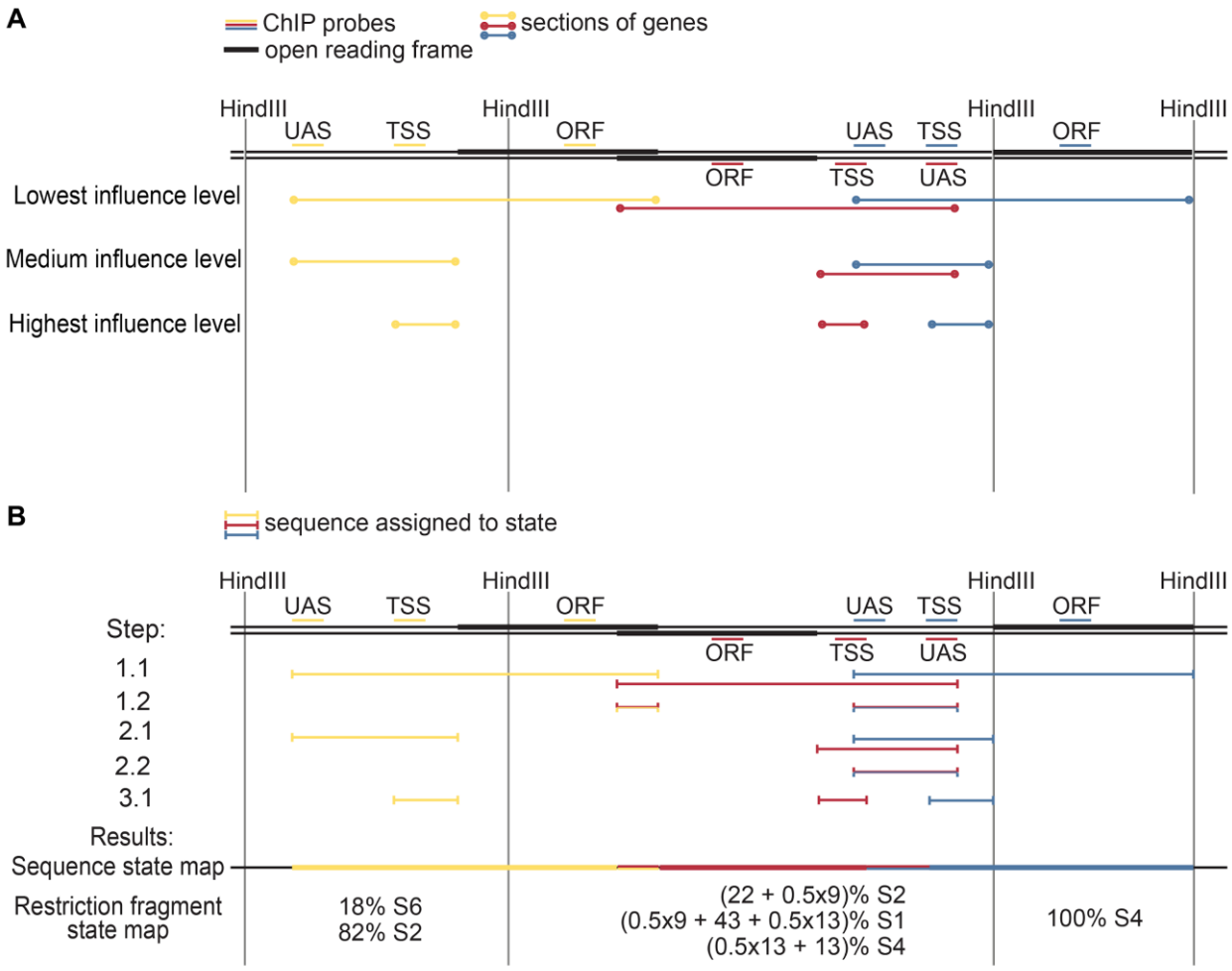


Figure 4.5. Mapping chromatin states to restriction fragments. (A) A schematic view of how probes are prioritised for resolving the overlaps of genes. **(B)** Mapping chromatin states to the corresponding genome sequence and restriction fragments.

4.3.2. Normalisation of state-wise contact maps

To investigate the interactions of different chromatin states, I calculated the normalised state-wise contact frequencies for both Hi-C and simulation data. First, the frequencies of observed contacts between different chromatin states were quantified as follows:

$$f_{obs}(contact(S_i, S_j)) = \frac{N(contact(S_i,S_j))}{\sum_{i,j=1}^{5} N(contact(S_i,S_j))} \quad \text{(Eq. 4.1),}$$

where $N(contact(S_i, S_j))$ is the number of observed interactions between chromatin states S_i and S_i. For Hi-C, the observed interactions of restriction fragments were obtained from SeqMonk or HOMER reports. I used the chromatin states of interacting fragments to count the number of observed contacts between different chromatin states. Since a fragment could have more than one assigned chromatin state, the number of contacts was divided according to the proportions of

mapped states (Fig. 4.5B). For simulations, I calculated the observed state-wise interactions using the recorded contacts of 2 kb segments and their mapped chromatin states.

Next, the frequency of expected contacts between two states was calculated under the assumption that any pair of genomic regions had an equal chance of interacting with each other. Each interaction involves two interacting regions (IR), which were *Hin*dIII fragments for Hi-C experiments and 2 kb segments for simulations. IR_1 and IR_2 represent the first and the second interaction regions respectively:

$$f_{exp}(contact(S_i, S_j)) = P((IR_1, IR_2) \in (S_i, S_j), IR_1 \neq IR_2) \quad \text{(Eq. 4.2),}$$

which leads to:

$$f_{exp}\left(contact\left(S_i, S_j\right)\right) = P\left(IR_1 \in S_i, IR_2 \in S_j\right) + P\left(IR_1 \in S_j, IR_2 \in S_i\right) - P(IR_1 = IR_2) \quad \text{(Eq. 4.3).}$$

This implies:

$$f_{exp}\left(contact\left(S_i, S_j\right)\right) = \left[P\left(IR_1 \in S_i\right) * P(IR_2 \in S_j)\right] + \left[P\left(IR_1 \in S_j\right) * P(IR_2 \in S_i)\right] -$$

$$P(IR_1 = IR_2) = 2 * \frac{\sum_{k=1}^{n} prop_{s_i}^{R_k}}{n} * \frac{\sum_{k=1}^{n} prop_{s_j}^{R_k}}{n} - \frac{\sum_{k=1}^{n} prop_{s_i}^{R_k} * prop_{s_j}^{R_k}}{n^2} \quad \text{(Eq. 4.4).}$$

Where n equals to the number of all genomic regions, and $prop_{s_i}^{R_k}$ is the proportion of S_i in the kth region (R_k). The proportion of S_i ($1 \leq i \leq 5$) for a restriction fragment was determined in the previous section (4.3.1). Each 2 kb segment was assigned to only one chromatin state, thus the proportion was 1 for the mapped chromatin state and 0 for the others.

Then, the normalised contact frequencies were calculated as follows:

$$f_{norm}\left(contact\left(S_i, S_j\right)\right) = \ln \frac{f_{obs}\left(contact(S_i, S_j)\right)}{f_{exp}\left(contact(S_i, S_j)\right)} \quad 1 \leq i, j \leq 5 \quad \text{(Eq. 4.5).}$$

For Hi-C experiments, the weighted average of normalised contact maps of biological replicates was computed. The total number of read pairs in each experiment was used as the weight of each data set. Similarly, the weighted average of the state-wise contacts maps of simulations was calculated using the total number of recorded contacts as the weight of each replicate simulation.

4.3.3. Comparison of the state-wise contact maps

The state-wise contact maps were quantified and normalised for Hi-C data using SeqMonk reports (Fig. 4.6A,B). In an analogous manner, the state-wise contact maps were analysed for [uniform], [25°C], and [37°C] simulations (Fig. 4.6C,D). Strikingly [25°C] and [37°C] simulations have similar results to 25°C and 37°C Hi-C data, respectively (see correlation coefficients and p-values in Table 4.2). The most important features are as follows: (1) The genes that are highly occupied by poised proteins, i.e. S4 at 25°C and S3 at 37°C, have the highest numbers of intra-state interactions (Fig. 4.6). (2) The genes that are highly expressed, i.e. S3 genes at 25°C and S4 genes at 37°C, have moderate numbers of intra-state contacts in Hi-C and simulations. (3) These highly expressed genes have lower number of contacts with other states. The observed patterns suggest that there is a spatial separation between the poised genes with lower mobility and the highly expressed genes with higher mobility. Other comparisons between these contact maps, e.g. Hi-C 25°C vs. Hi-C 37°C or Hi-C 25°C vs. [37°C] model, show that the patterns are not correlated at all (see Table 4.2).

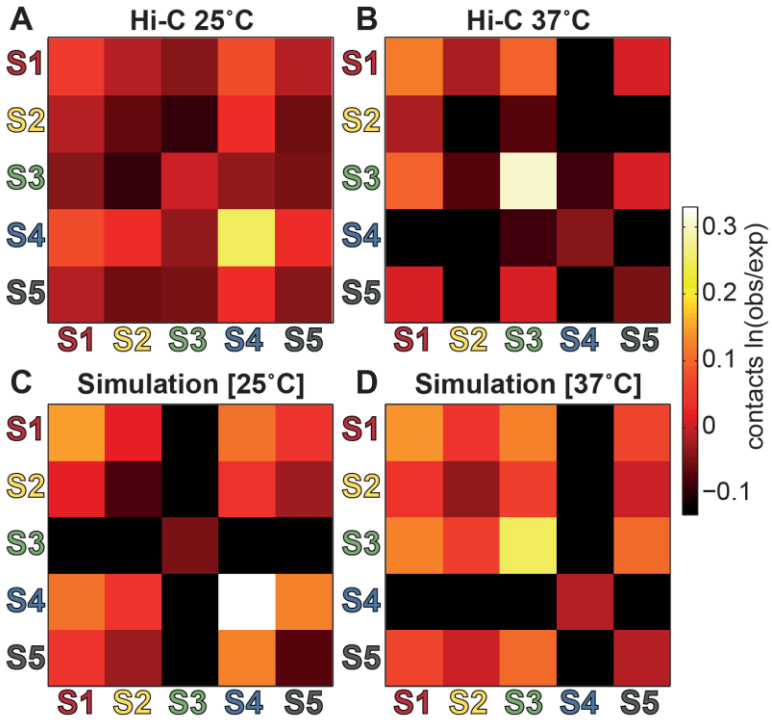


Figure 4.6. State-wise contact maps for Hi-C, [25°C], and [37°C] simulations. (A,B) The number of interactions between chromatin segments were analysed by SeqMonk. The data was corrected for physical linkage and a cut-off of ≥3 contacts was applied. Then, the normalised state-wise contact maps were calculated for each temperature condition. **(C,D)** The number of interactions were obtained from 30,000

sampled time steps (10,000 per simulation replicate). To calculate the Hi-C and simulation contact maps, the interaction of S6 and S7 segments were not counted. The same colour map range was used for all plots.

State-wise contact maps		Correlation coefficient	p-value	Significance
Hi-C 25˚C	Simulation [25˚C]	0.8662	2.9712e-5	****
Hi-C 37˚C	Simulation [37˚C]	0.7409	0.0016	**
Simulation [25˚C]	Simulation [37˚C]	-0.0614	0.8280	ns
Hi-C 25˚C	Hi-C 37˚C	-0.0078	0.9781	ns
Hi-C 25˚C	Simulation [37˚C]	-0.1734	0.5367	ns
Hi-C 37˚C	Simulation [25˚C]	-0.1057	0.7078	ns

Table 4.2. Correlation of state-wise contact maps of Hi-C, [25°C], and [37°C] simulations. Pearson correlation coefficients calculated in MATLAB between two vectors of size 15 (the numerical values of the upper-right triangular part of matrices depicted in Fig. 4.6). ****:p-value<0.0001; **:0.001<p-value<0.01; ns: not significant.

The [25°C] and [37°C] models, which have provided the best matches to the microscopy data, also produced the best match to the Hi-C state-wise contact maps. To explore the state-wise contact maps of simulations that were not in line with microscopy measurements, the normalised contact frequencies were quantified. When the F_{LC} is scaled by a factor of 2 to simulate the 25°C condition (Fig. 4.7B), the generated contact pattern is similar to the 25°C Hi-C data (Fig. 4.6A(a)) (see correlation coefficients and p-values in Table 4.3). In addition, the [uniform] model and the 25°C simulation with 10x scaled F_{LC} are correlated with the 25°C Hi-C data (Table 4.3). However, they do not bear all the hallmarks: Unlike the Hi-C data, S3 segments have a high number of intra-state interactions, and S4 segments have fewer interactions with other states (Fig. 4.7A,D). Furthermore, the contact maps of [37°C] simulations with 2x and 10x changes in F_{LC} do not provide good matches to the experimental data (lower correlation coefficients in Table 4.3). Although, similar to the Hi-C data, S4 segments avoid contacts with other chromatin states, they have the highest number of intra-state interactions (Fig. 4.7C,E).

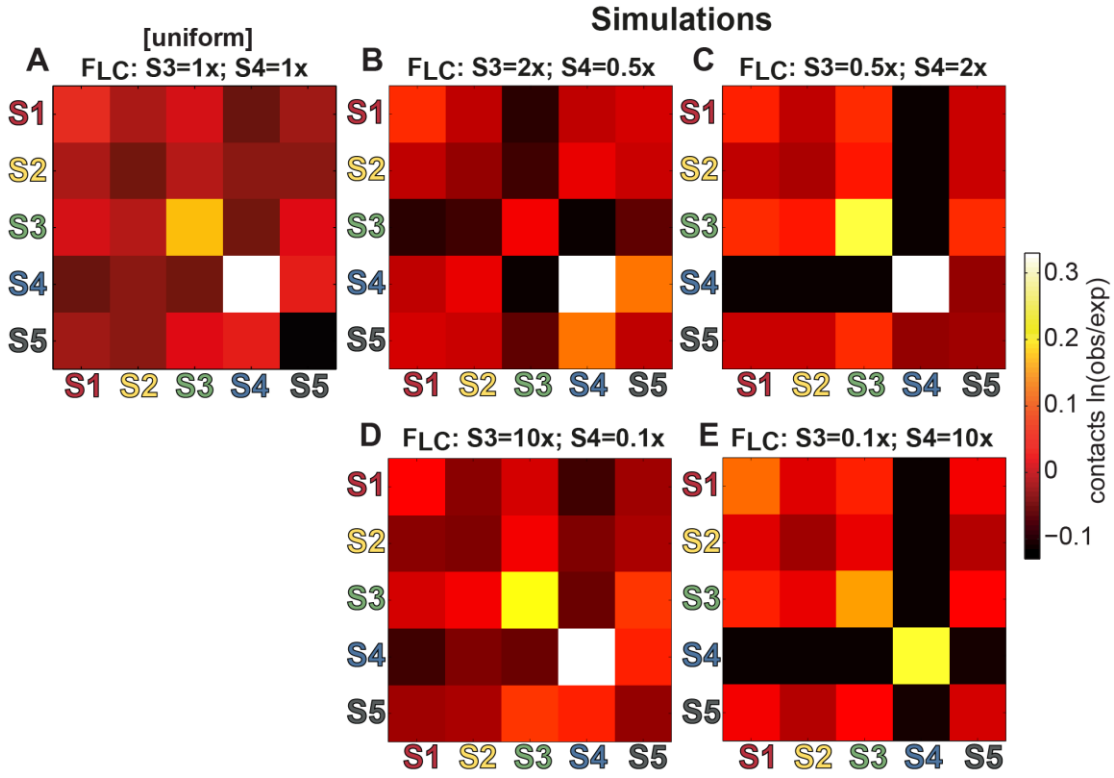


Figure 4.7. State-wise contact maps for models with 1x, 2x, and 10x changes in F_{LC}. (A) The contact maps were calculated by analysing 30,000 sampled time steps from 3 replicate simulations. **(B-E)** The data was obtained from 10,000 sampled time steps for each plot. For all contact maps, the interactions of S6 and S7 segments were discarded. The same colour map range was used for all plots.

State-wise contact maps		Correlation coefficient	p-value	Significance
Hi-C 25˚C	F_{LC}: S3=2x; S4=0.5x	0.8938	7.1010e-6	****
Hi-C 25˚C	Simulation [uniform]	0.7309	0.0020	**
Hi-C 25˚C	F_{LC}: S3=10x; S4=0.1x	0.6807	0.0052	*
Hi-C 37˚C	F_{LC}: S3=0.1x; S4=10x	0.6255	0.0126	*
Hi-C 37˚C	F_{LC}: S3=0.5x; S4=2x	0.5723	0.0258	*
Hi-C 37˚C	Simulation [uniform]	0.4150	0.1240	ns

Table 4.3. Correlation of state-wise contact maps of Hi-C and simulations. To compare the contact maps, the Pearson correlation coefficients were calculated in MATLAB. The rows are sorted by correlation coefficient. ****:p-value<0.0001; **:0.001<p-value<0.01; *:0.01<p-value<0.05; ns: not significant.

Whilst the focus is on the comparison of Hi-C and simulation contact maps, it is crucial to not lose sight of differences in the nature of experimental and simulation data sets. The Hi-C data has been obtained from a culture of almost two billion fixed cells, whereas, the simulation data was captured from three independently modelled dynamic genomes over one million time steps.

4.3.4. State-wise contact maps calculated from HOMER reports

HOMER was employed to analyse our Hi-C data, as explained in section 4.1.1. I used the output matrices that were normalised for read coverage, to quantify the contact frequencies of chromatin states (Fig. 4.8). As discussed in section 4.1.1, the read coverage normalisation is not appropriate for our data. Therefore, the calculated state-wise contact maps illustrate the effect of incorrect normalisation on the results. Different types of filters were tested and applied after the read coverage normalisation and prior to the quantification of state-wise contact frequencies (Fig. 4.8).

As genes can spread over adjacent restriction fragments, the chromatin states of neighbouring fragments can be similar and correlated. Thus, filtering out the interactions of adjacent fragments resulted in less prominent intra-state contacts (Fig. 4.8G-L). Filtering out rDNA data or S6 and S7 interactions did not change the pattern of contacts (Fig. 4.8C-F vs. Fig. 4.8A,B and Fig. 4.8I-L vs. Fig. 4.8G,H). As Fig. 4.8 illustrates for all plots, the intra-state contacts are stronger than the inter-state interactions. In contrast to the SeqMonk and Hi-C results at both temperature conditions, the highly occupied segments have a low number of interactions with other states. However, in all three types of state-wise contact map, the highly occupied segments have the highest number of intra-state contacts (Fig. 4.6A,B).

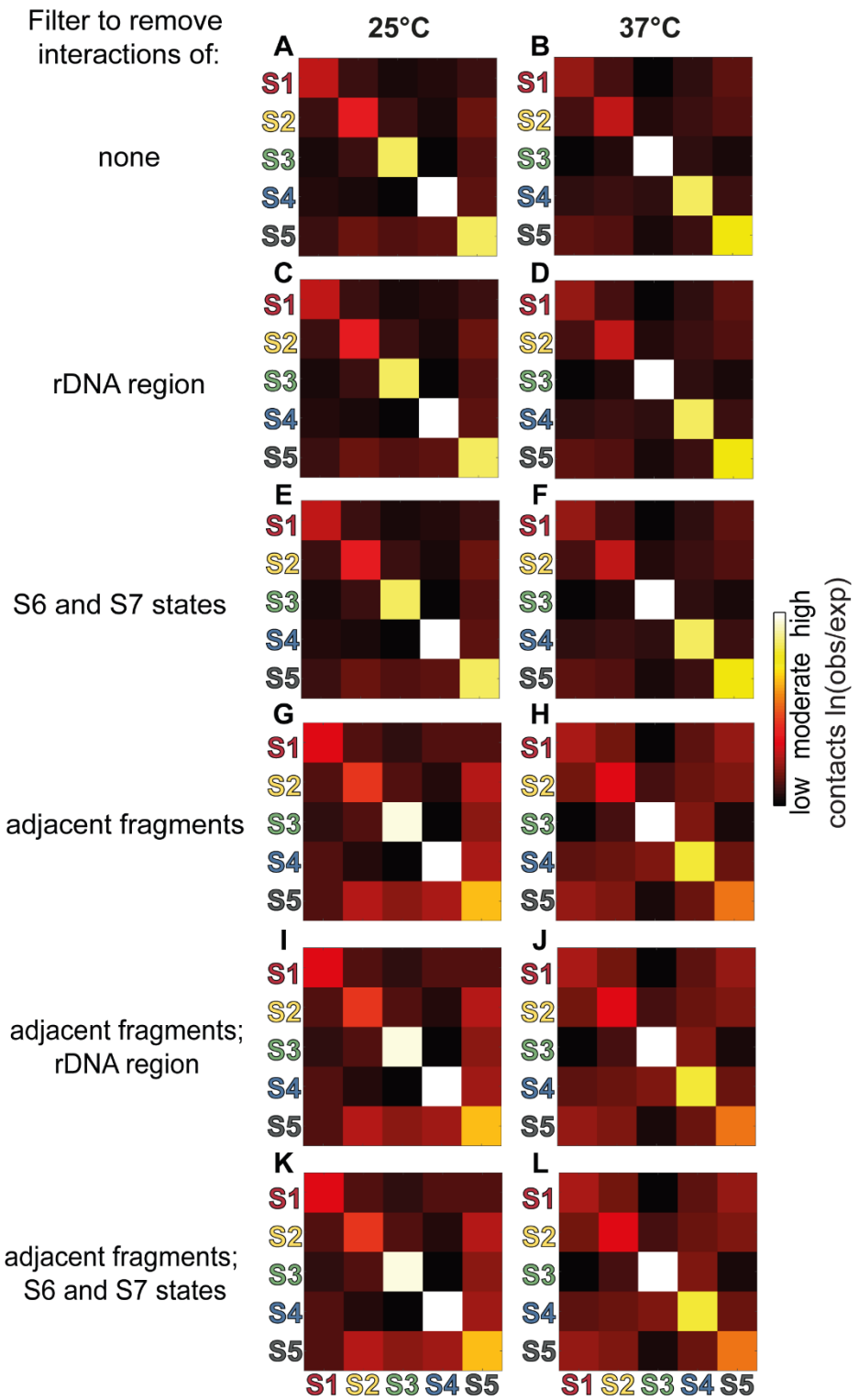


Figure 4.8. State-wise contact maps for Hi-C data analysed by HOMER. The Hi-C data was normalised by HOMER for read coverage, different filters applied as described, and then the number of state-wise interactions was calculated. The weighted averages of replicates were calculated using the number of total read pairs as the weight of each dataset. **(A,B)** No filter is applied. **(C,D)** The interactions of the rDNA region were discarded. **(E,F)** The interactions of S6 and S7 states were removed. **(G,H)** The contacts of

immediate neighbours were removed. **(I,J)** The interactions of adjacent fragments and rDNA region were discarded. **(K,L)**. The interactions of adjacent fragments, S6, and S7 states were not counted.

5. Positions of contacts and relocation of genes

Changes in the transcriptional activity of genes can affect their spatial positioning in different cell types (Shachar & Misteli 2017). Other factors, such as gene-density of the corresponding chromatin region and environmental signals, could also determine the 3D localisation of genes (Meaburn et al. 2007; Misteli 2007; Cremer et al. 2006). To determine the spatial distribution of segments corresponding to the poised and highly expressed genes, the locations of contacts between these segments were recorded during simulations. The distributions of contacts along different axes were used to calculate the degree of co-localisation between segments. The 2D projections of contacts and their clustering degrees are reported in section 5.1.

In budding yeast, several genes have been reported to move towards the nuclear envelope upon activation (Egecioglu & Brickner 2011). This recruitment leads to physical interactions with the nuclear pore complexes and thought to facilitate the transport of messenger RNA into the cytoplasm (Brickner & Walter 2004; Guet et al. 2015; Casolari et al. 2005; Drubin et al. 2006; Cabal et al. 2006). As an example, the gene that encodes the stress-inducible protein HSP104p (Sanchez et al. 1992; Parsell et al. 1994), relocates to the nuclear periphery upon induction (Dieppois et al. 2006). To explain these observations, different theories have been proposed (Steglich et al. 2013; Egecioglu & Brickner 2011; Drubin et al. 2006; Dieppois et al. 2006; Randise-Hinchliff & Brickner 2016; Ahmed et al. 2010); however, the underlying mechanisms are not fully understood (Randise-Hinchliff et al. 2016). To investigate the relocation of heat-shock genes to the nuclear periphery at 37°C, the positions of a heat-shock gene (*HSP104*) and all S4 genes were quantified in simulations (section 5.2). An enrichment analysis of S3 and S4 states was also carried out to identify the genes that have neighbours with similar chromatin states. These genes are more likely to undergo spatial relocation within the nucleus upon temperature alterations (section 5.2). Finally, to assess the robustness of simulation results, the peripherality of S4 genes, the distribution of contacts, and the state-wise contact maps were plotted for each simulation replicate (section 5.3).

5.1. Contact projection

Segments with the highest protein occupancy are in a poised state and have the highest number of intra-state interactions (S4 at 25°C, S3 at 37°C), as described in section 4.3.3. To evaluate the positions of these interactions in simulations, the middle point between any pair of colliding segments was calculated. Then the coordinates were projected onto 2D planes and the density plots were computed using the Python Seaborn library (Fig. 5.1, Fig. 5.2). The projections onto the yz

and xz planes demonstrate that S3 and S4 segments are separated from the aggregated rDNA segments on the spindle pole body distal side of the nucleus, as expected (see Fig. 3.1A). In the uniform model, these segments have comparable and moderate levels of compaction. However, in the heteropolymeric models, the heavily occupied segments are tightly clustered (Fig. 5.1C,E, Fig. 5.2C,F) and the less occupied ones are loosely distributed (Fig. 5.1B,F, Fig. 5.2B,G). The distributions of contacts along the z-axis confirm the spatial clustering of slow moving poised segments (Fig. 5.1G). These segments also have a higher number of intra-state contacts, as the area under the corresponding curve is larger: high numbers of S4:S4 and S3:S3 contacts at 25°C and 37°C respectively. This is in agreement with the state-wise contact maps (see Fig 4.6).

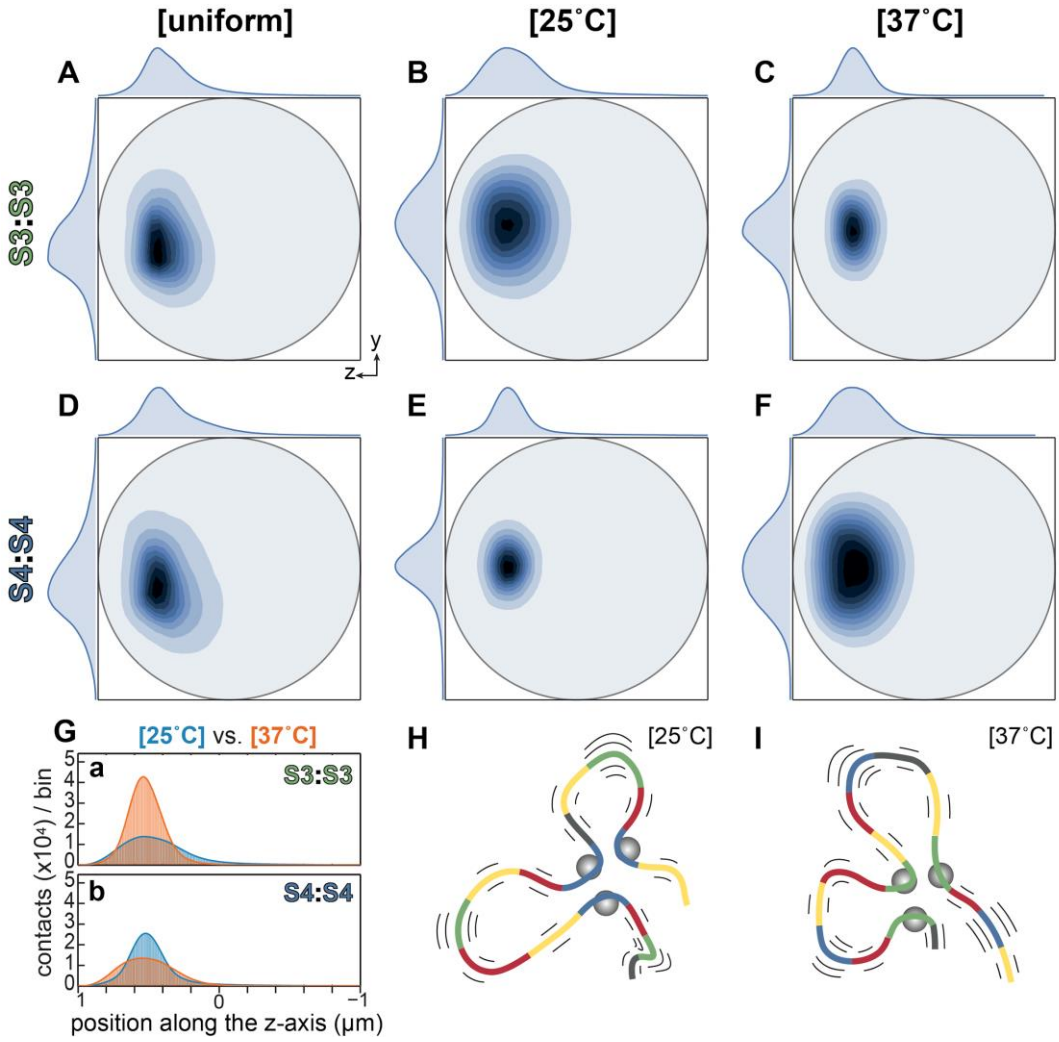


Figure 5.1. Positions of intra-state contacts in simulations. (A-F) Density plots for 2D projections of simulated contacts between segments of the same chromatin state. The data was obtained from 3 replicate simulations; 1 million time steps (not sampled) per simulation. The plots are normalised to the highest density in each panel. The nucleus is oriented as Fig. 3.1A. (A-C) Intra-state contacts of S3 segments; (D-F) intra-state contacts of S4 segments. (A,D) homopolymer model; (B,E) heteropolymer model representing

[25°C]; (C,F) heteropolymer model representing [37°C]. **(G)** The numbers of contacts were quantified for 500 bins (each with 4 nm width) along the z dimension. The distributions of [25˚C] (cyan) and [37˚C] (orange) conditions are overlaid. **(H,I)** A schematic view of poised proteins (grey spheres) binding to poised segments: S4 segments at 25˚C (H) and S3 segments at 37˚C (I). Hypothetically, the poised genes have slower movements as a result of increased protein binding. The length and the number of flanking black lines symbolise the degree of segment mobility.

To quantify the clustering level of contacts, we defined a formula based on the distribution of contacts along the z-axis, as follows:

$$non\ normalised\ clustering\ degree\ along\ z\ axis \coloneqq h_z/w_z \quad \text{(Eq. 5.1)},$$

where h_z is the height of the curve and w_z is the width of the distribution at half height (Fig. 5.2D,H). Since the segments are confined by the limited nuclear volume, a higher number of segments leads to more compact conformations. Thus, the normalised clustering degree is defined as follows:

$$normalised\ clustering\ degree\ along\ z\ axis \coloneqq (h_z/w_z)/[n * (n-1)] \quad \text{(Eq. 5.2)},$$

where n is the number of segments and $n * (n-1)$ is equal to the number of possible segment pairs. The measured clustering levels indicate a drastic change in the compaction of poised and highly transcribed genes: S3 segments are compacted 5.7-fold upon change from [25°C] to [37°C], while S4 segments have a 3.4-fold compaction upon the reverse change, from [37°C] to [25°C] (Table 5.1).

To examine the compaction of the whole genome, I calculated the positions of all contacts that had occurred in the simulations (i.e. both intra- and inter-state contacts). Then, the clustering levels were measured for S1-S5, S1-S6, and S1-S7 segments by analysing the relevant contact distributions. The results predict that chromosomes have more compact structures during heat-shock (cf. Table 5.1: [25°C]/[37°C] and [37°C]/[25°C]). Excluding S6 or S7 segments from the calculations did not change the outcome.

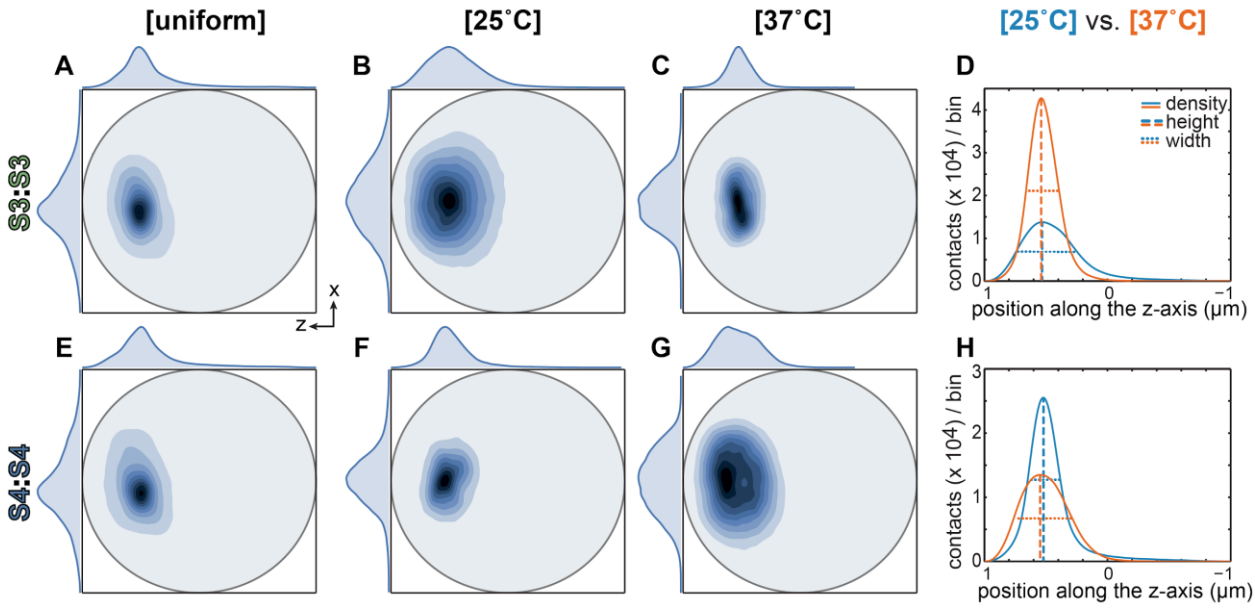


Figure 5.2. Positions of intra-state contacts in the zx plane. (A-C, E-G) Similar to Fig. 5.1, the density plots were calculated for the zx projection of contact positions. They are normalised to the highest density in each panel. **(D,H)** The number of contacts for each bin on the z-axis was quantified; there were 500 bins of 4 nm width. To calculate the clustering strength of segments, these shape parameters of the contact distributions were used: the height (the vertical dashed lines), and the width at half height (the horizontal dashed lines). For each simulation type, 3 million time points from three replicate simulations (1 million per replicate) were analysed.

State	Clustering along the z-axis				
	[uniform]	[25˚C]	[37˚C]	[25˚C]/[37˚C]	[37˚C]/[25˚C]
intra-state S3	2.0275e2	1.1125e2	6.3814e2	0.17434	5.7359
	3.2364e-4	1.7759e-4	1.0186e-3		
intra-state S4	1.4823e2	3.8656e2	1.1339e2	3.4091	0.29333
	3.4497e-4	8.9964e-4	2.6389e-4		
all contacts S1-S5	1.0122e4	1.95e4	2.6713e4	0.72999	1.3699
	5.8198e-4	7.9724e-4	3.0209e-4		
all contacts S1-S6	1.0718e4	2.0695e4	2.8181e4	0.73437	1.3617
	2.9719e-4	5.7381e-4	7.8136e-4		
all contacts S1-S7	1.0755e4	2.0773e4	2.8299e4	0.73408	1.3623
	2.6229e-4	5.0661e-4	6.9013e-4		

Table 5.1. Degree of clustering of contacts along the z-axis in simulations. Height / width of distributions along the z-axis. In columns 2-4, the first row for each type of contacts is not normalised (Eq. 5.1); the second row is normalised by the number of possible segment pairs (Eq. 5.2).

The distributions of contacts along the x and y axes are in line with the strong clusters of poised genes (Fig. 5.3). To incorporate these data into the calculation of the clustering degree, I expanded the equations as follows:

$$non\ normalised\ clustering\ degree := (h_z/w_z) * (h_y/w_y) * (h_x/w_x) \quad \text{(Eq. 5.3)},$$

$$normalised\ clustering\ degree := \frac{(h_z/w_z)*(h_y/w_y)*(h_x/w_x)}{[n*(n-1)]^3} \quad \text{(Eq. 5.4)}.$$

The results of the 3D clustering calculations are more significant than the clustering levels along the z-axis (Table 5.2 vs. Table 5.1). The whole genome is ~1.5 times more condensed at 37°C, while the co-localisation strength of S3 segments has a 62.6-fold increase. The clustering degree of S4 segments changes by a factor of 42.6.

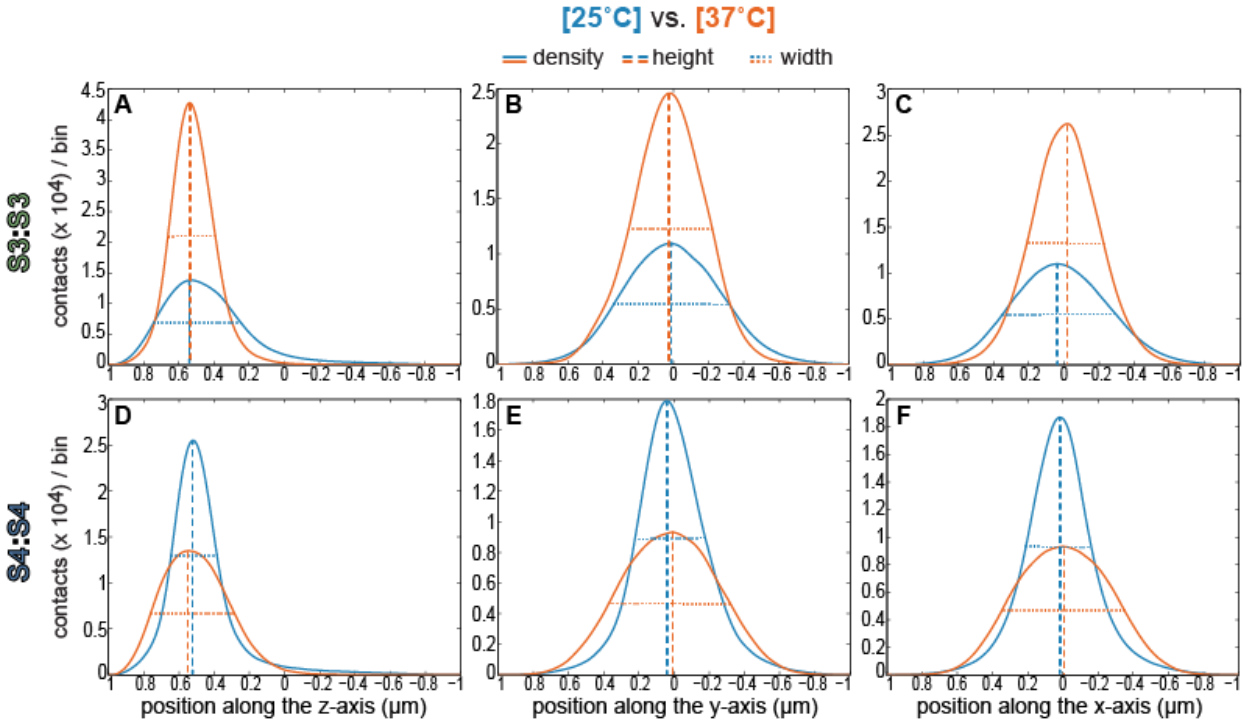


Figure 5.3. Distributions of contacts along x, y, and z dimensions. The numbers of contacts were calculated for bins of 4 nm width. Each horizontal axis extends from -1 to 1 μm as the centre of the nucleus is (0, 0, 0) and its radius is 1μm. **(A-C)** Contacts between S3 segments; **(D-F)** Contacts between S4 segments. For each plot, the dashed vertical and horizontal lines are the height and the width at half height of distributions respectively.

State	Clustering in 3D				
	[uniform]	[25˚C]	[37˚C]	[25˚C]/[37˚C]	[37˚C]/[25˚C]
intra-state S3	1.2886e6	5.0706e5	3.1732e7	0.015979	62.58
	5.2411e-12	2.0623e-12	1.2906e-10		
intra-state S4	4.376e5	1.3901e7	3.2642e5	42.586	0.023482
	5.5162e-12	1.7523e-10	4.1147e-12		
all contacts S1-S5	1.6173e11	1.3632e12	2.1358e12	0.63826	1.5668
	4.2994e-12	3.6238e-11	5.6777e-11		
all contacts S1-S6	1.9292e11	1.6138e12	2.4511e12	0.65839	1.5189
	4.1124e-12	3.4399e-11	5.2247e-11		
all contacts S1-S7	1.952e11	1.6398e12	2.4869e12	0.65938	1.5166
	2.8312e-12	2.3785e-11	3.6071e-11		

Table 5.2. Degree of clustering of contacts in the simulations. Height / width of distributions along the x, y, and z axes are multiplied and normalised by the number of possible segment pairs. In columns 2-4, the first row for each type of contacts is not normalised (Eq. 5.3); the second row is normalised (Eq. 5.4).

5.2. Gene relocation

In our simulations, a change in the temperature condition leads to an alteration in the spatial distribution of segments. This led us to speculate that the model might be capable of reproducing the relocation of inducible and highly transcribed genes to the nuclear periphery. To test this hypothesis, I analysed the positions of S4 segments, which are enriched for heat-shock genes. As 20% of genes change their chromatin state upon changes in temperature (section 1.4.2), the list of S4 segments is not identical in 25°C and 37°C state assignments. Therefore, segments that are assigned to S4 state at both 25°C and 37°C conditions have been analysed in this section. First, the x, y, z coordinates of these segments were determined at each time step. Then, the distances of segments from the z-axis were calculated in [25°C] and [37°C] simulations as follows.

$$distance\ of\ i^{th} segment\ from\ z\ axis = \sqrt{{x_i}^2 + {y_i}^2} \qquad \text{(Eq. 5.5).}$$

Next, the calculated distances were drawn as density plots for the segment corresponding to the *HSP104* gene (Fig. 5.4A) and for all S4 segments (Fig. 5.4B). The 3D distances of segments from the centre of the nucleus were also quantified (Fig. 5.4C,D). The dashed red line demonstrates the border of the peripheral zone, which covers the outermost 33% of the nuclear area. The *HSP104* gene has a broader distribution at 37°C (Fig. 5.4A) and it moves towards the nuclear membrane upon activation (Fig. 5.4A,C). In more detail, the number of occurrences of this gene within the

peripheral zone has more than doubled (Fig. 5.4C,F) ($p<2.2e-16$, Wilcoxon rank sum test). This is very similar to the results of confocal microscopy measurements for the peripheral positions of the *HSP104* gene at room temperature and upon induction (Dieppois et al. 2006) (Fig. 5.4E). The distribution of all S4 genes also shows a significant shift towards the nuclear periphery at 37°C (Fig. 5.4B,D): the number of their peripheral locations has increased by 50 percent (Fig. 5.4G).

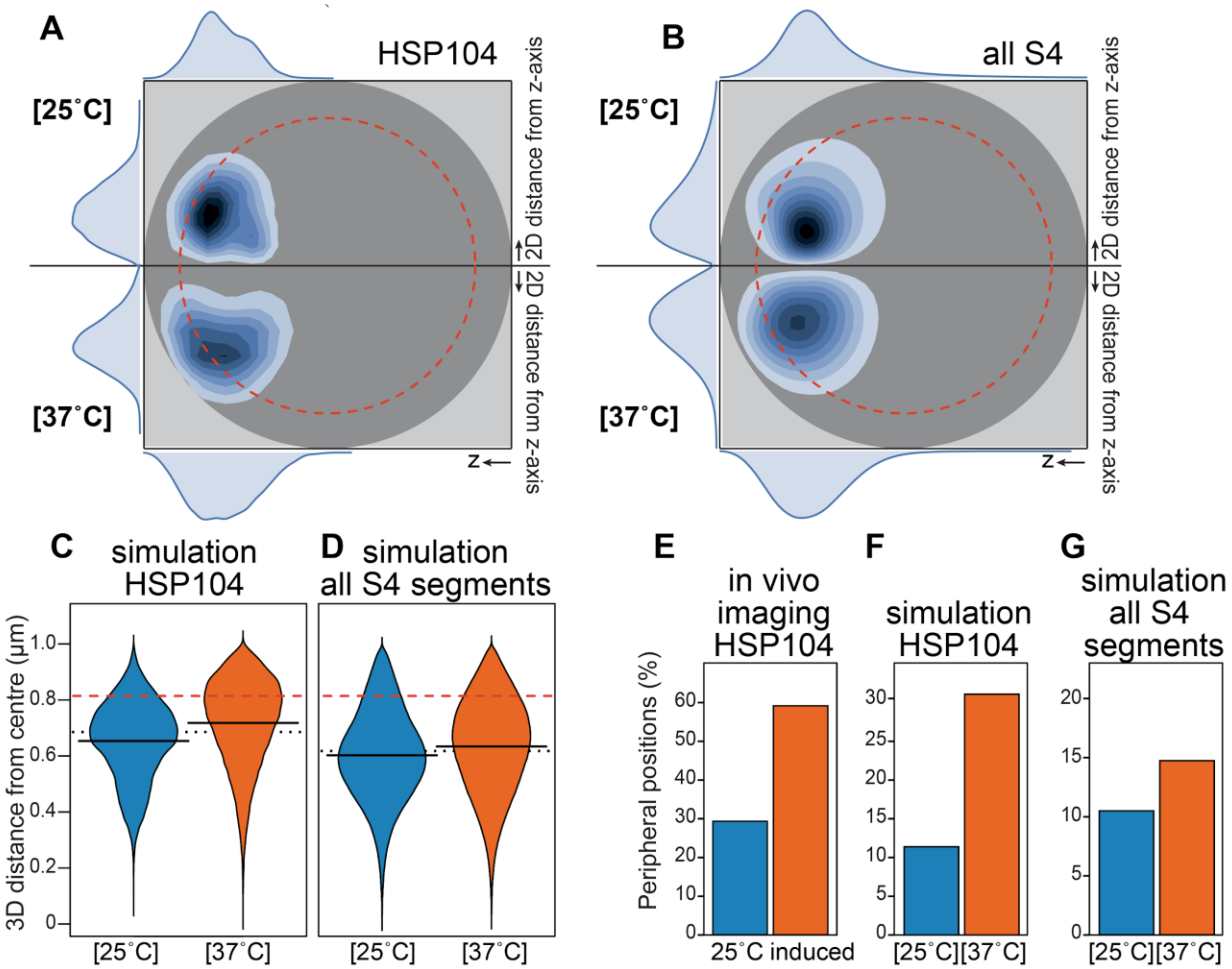


Figure 5.4. Peripheral positions of heat-shock genes before and after activation. (A) The 2D distance of the *HSP104* gene from the z-axis in [25°C] and [37°C] conditions (see Eq. 5.5). **(B)** The distance distribution of all S4 segments from the z-axis. For both (A) and (B), $3x10^4$ sampled time steps were analysed and the data was plotted using the Python Seaborn package. **(C,D)** The 3D distances of the *HSP104* gene and all S4 segments from the centre of the nucleus. The red dashed line specifies the boundary of the peripheral region, which covers 33% of the nuclear area in 2D. The solid black lines show the mean of 3D distances, per condition, and the dashed black lines indicate the total averages. **(E)** The percentage of peripheral positions for the *HSP104* gene (replotted from (Dieppois et al. 2006)). The position of the gene locus was tracked by microscopy imaging before and after induction. **(F,G)** The frequencies of peripheral positions for the *HSP104* gene and for all heat-shock genes. For (C,F,G), $3x10^4$ sampled time steps from

three simulation replicates were analysed. For (D) a random sample of $1x10^6$ out of the recorded $12.9x10^6$ data points was plotted.

The relocation of all S4 segments shows the average behaviour of heat-shock genes, which is less significant than that of the *HSP104* gene (Fig. 5.4F,G). This suggests that there are some S4 genes with limited relocation or reversed relocation towards the centre of the nucleus. The 3D distances of individual S4 genes from the centre of the nucleus were analysed and confirmed that some S4 genes prefer central positions when the condition changes from [25°C] to [37°C] (Fig. A.4A). This could have two possible reasons: a) The S4 segment has S3 neighbours, which move more slowly at 37°C and limit the mobility of the activated heat-shock gene. Therefore, the S4 segment might have slower movements at 37°C despite being less occupied by poised proteins. b) The folding of chromosomes prevent the segment from moving fast enough to get closer to the nuclear membrane. Focusing on the first assumption (a), I performed preliminary analysis of S4 segments. We hypothesized that having facilitator neighbours is a necessary condition for a gene to be able to relocate to the nuclear periphery upon activation, i.e. S4 genes with S4 neighbours can relocate at 37°C. To find these genes, I defined the enrichment score of segments for S4 as follows:

$score_{S_4}^{i} = enrichment_{S_4}^{i} + \sum_{j=1}^{5}[(enrichment_{S_4}^{i+j} + enrichment_{S_4}^{i-j})/(j+1)]$ (Eq. 5.6),

where $score_{S_4}^{i}$ is the S4 enrichment score of the ith segment. The $i+j$ and $i-j$ are the indices of jth neighbours on the right and left hand sides of ith segment, respectively. The $enrichment_{S_4}^{i\pm j}$ equals one if the state of the jth neighbour is S4 and zero if the jth neighbour is in any other state. Closer neighbours have more influence on the score of the segment, and distant neighbours (>10 kb from the segment) are not taken into account. The scores of segments were computed using the state assignments at 25°C and 37°C (section 1.4.2). The S4 segments with high enrichment scores have facilitator neighbours. Therefore, they could have faster movements at 37°C as expected and relocate to the proximity of nuclear membrane. However, segments with low scores would probably not be fast enough to be able to shift to the nuclear periphery. The position of loci with the highest scores could be analysed in simulations using their 2D distances from the z-axis and their 3D distances from the centre of the nucleus (similar to the analysis of *HSP104* gene). In addition, these genes are good candidates to be further studied and compared by microscopy imaging before and after activation.

I repeated the enrichment analysis of segments for S3 to identify the S3 genes with facilitator neighbours:

$score^{i}_{S_3} = enrichment^{i}_{S_3} + \sum_{j=1}^{5}[(enrichment^{i+j}_{S_3} + enrichment^{i-j}_{S_3})/(j+1)]$ (Eq. 5.7).

The S3 enrichment score of segments were measured using the 25°C and 37°C state assignments. The relocation of S3 genes with the highest scores could be also studied by our polymer model and microscopy experiments upon activation at 25°C.

To verify if the enrichment scores have been able to identify the S4 and S3 genes with faster movements at [37°C] and [25°C] conditions, respectively, I analysed the displacement of segments with high scores. First, I selected the S4 segments with the highest S4 enrichment scores (>2.5) at both temperature conditions and plotted their MSD curves from the [uniform], [25°C], and [37°C] models (Fig. 5.5). The MSDs of the *HSP104* were also measured (Fig. 5.5Q). The *HSP104* gene and all chosen loci, except for the last three (Fig. 5.5N-P), have faster mobilities during heat-shock, i.e. the orange curves are higher than the cyan ones.

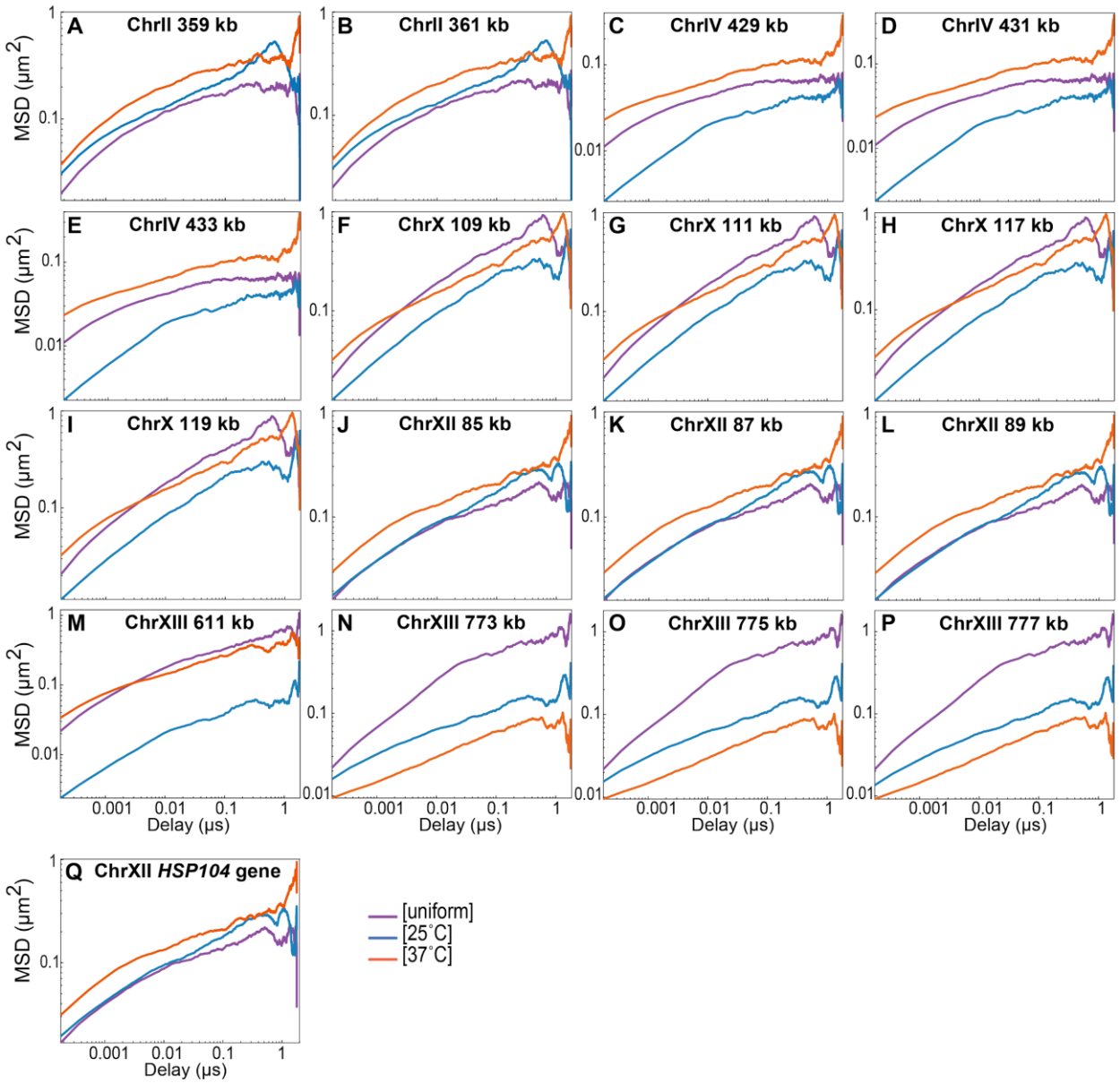


Figure 5.5. MSDs of S4 segments with high S4 enrichment score. (A-P) Segments with high S4 enrichment score (>2.5) were identified. All these segments are assigned to S4 and they have neighbours with similar state assignments according to both 25°C and 37°C state assignments. The displacement of segments in [25°C], [37°C], and uniform simulations were plotted in MATLAB. The specified loci in kb are the centre of analysed segments. **(Q)** The MSDs of the corresponding segment to the *HSP104* gene (chromosome XII, 41st segment) were quantified. The score of this segment is 2.78 and 2.45 according to 25°C and 37°C state assignments, respectively. For all plots, 10,000 sampled time steps were analysed from the first simulation replicate. The time unit of simulations is not real seconds.

Next, S3 segments with high S3 enrichment scores (>2.5) were selected and their displacements were quantified (Fig. 5.6). All segments, except the loci on chromosome VI (Fig. 5.6 D-F), have

faster movements in [25°C] simulations (the cyan curves are higher than the orange ones for all time lags). Therefore, the enrichment scores have been able to find the genes that change their movements upon activation in simulations, as expected. In both Fig. 5.5 and Fig. 5.6, the enriched loci that do not move faster when they are highly expressed (Fig. 5.5N-P, Fig. 5.6 D-F) could be confined by other constraints, such as the folding of chromatin that affects the displacement of engaged loci.

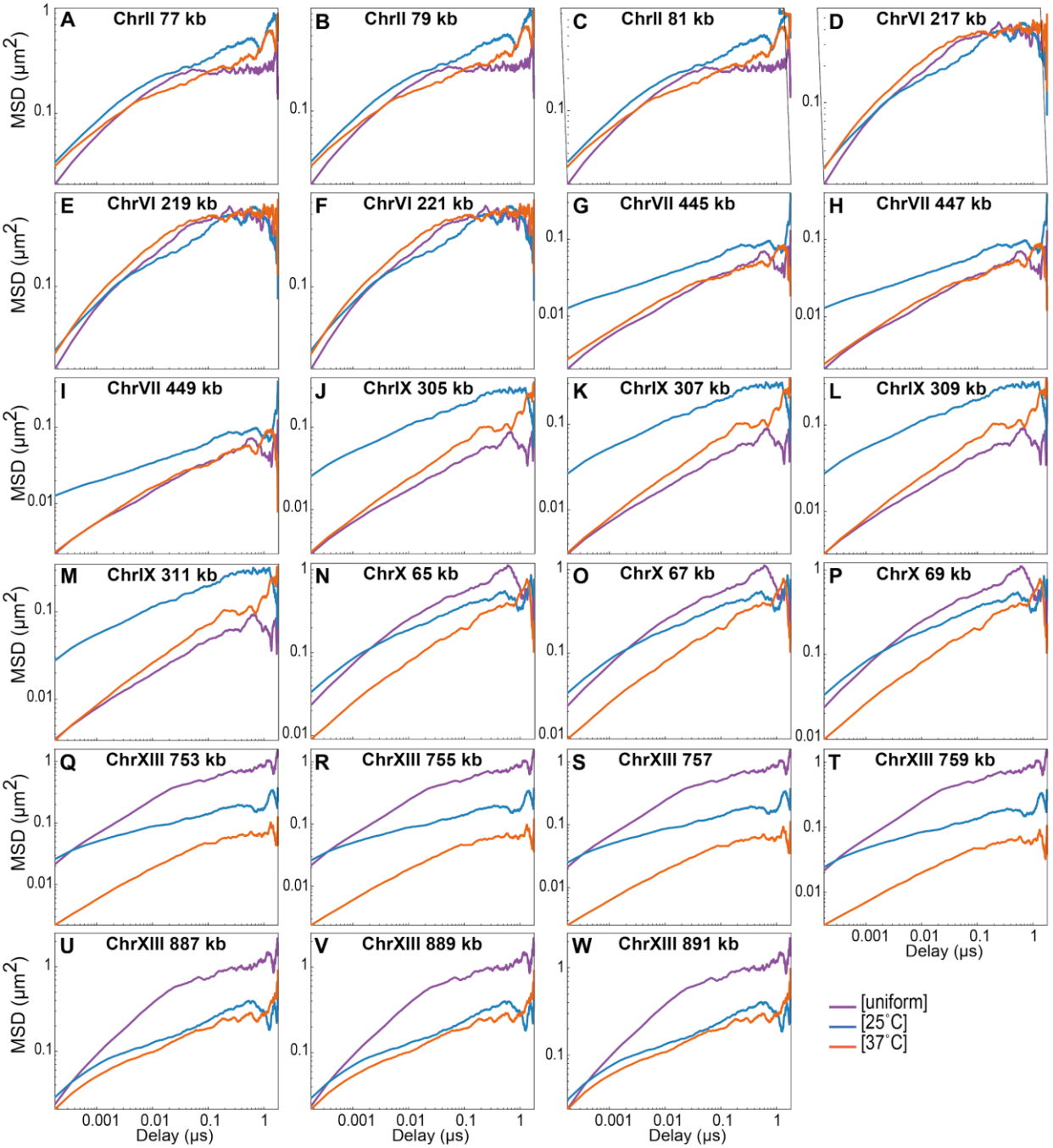


Figure 5.6. MSDs of S3 segments with high S3 enrichment score. (A-W) The score of segments for S3 enrichment was computed. Then, the MSDs of segments with scores higher than 2.7 at both 25°C and 37°C

conditions were plotted in MATLAB. The indicated locus in each plot is the centre of the identified segment. The time unit is simulations is not real seconds. For all plots, 10,000 non-correlated time points were analysed from the first simulation replicate.

5.3. Convergence of simulations

To test when the simulation results converge, I calculated the peripherality of all S4 segments for shorter time intervals and for six independent simulation runs (Fig. 5.7A(a,b)). The fluctuations of peripherality at different time intervals demonstrate the stochasticity of the system. The simulations reach the convergence point within the first 10^5 time steps, and after that point, there is no clear trend in the peripherality data. We used these results to choose the analysed simulation period: From time step 99,901 to 1,099,900 (the transparent bar in the background of Fig. 5.7A). This time interval provides a reliable and robust representation of the system's behaviour, as demonstrated by the results of three key types of analysis, for each of the three simulation replicates. The peripherality rate (Fig. 5.7A(c)), the distribution of contacts (Fig. 5.7B), and the state-wise contact maps (Fig. 5.7C) have stable patterns across independent replicates.

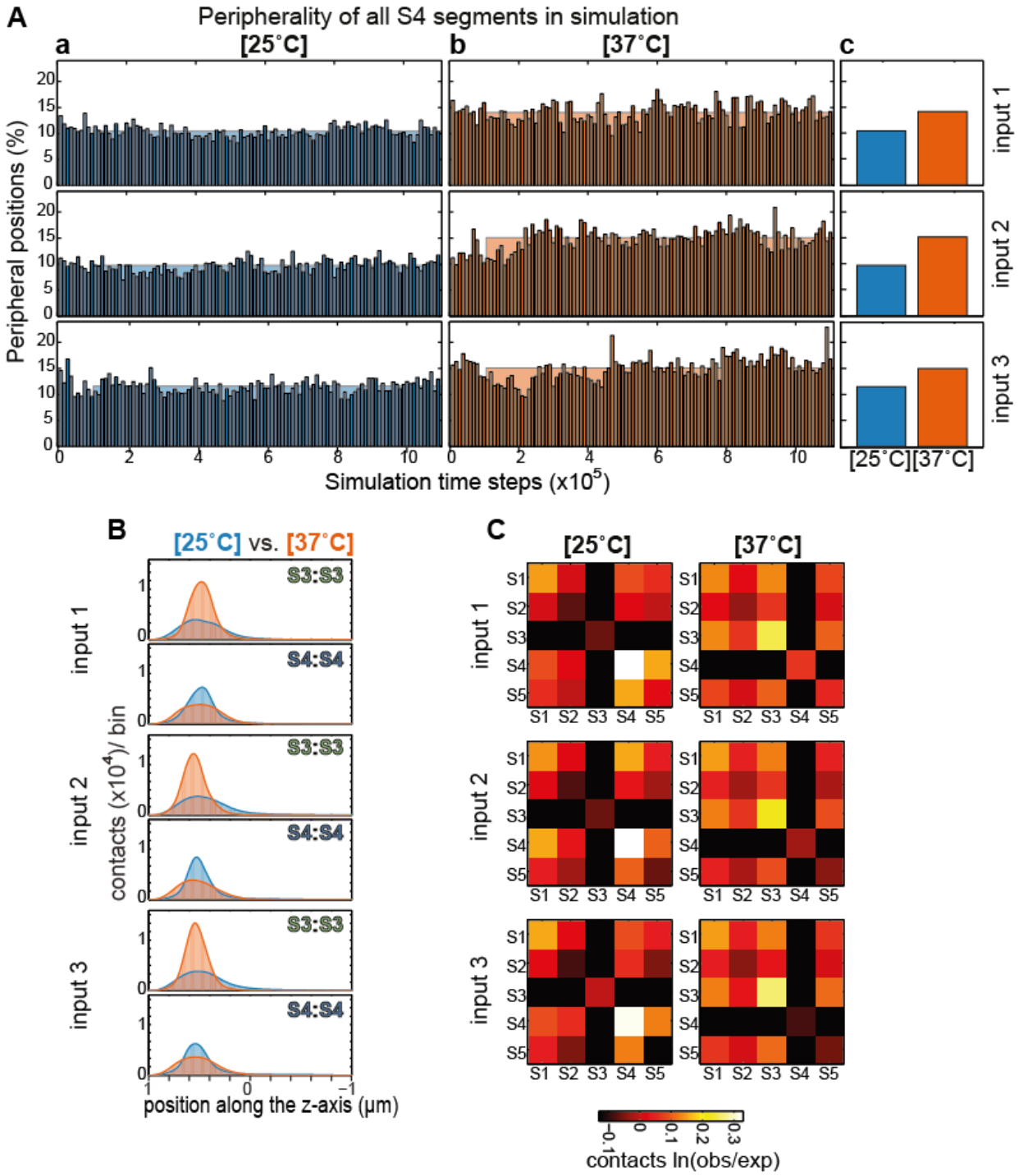


Figure 5.7. Convergence, stochasticity and robustness of the simulations. (A) Peripherality of all S4 segments calculated for smaller time intervals. Each interval comprises 10,000 time steps, with data obtained every 100 time steps. The percentage of S4 positions in the peripheral zone (as defined for Fig. 5.4: outermost 33% of the nuclear area) is plotted for **(a)** three independent [25˚C] simulations, and **(b)** three independent [37˚C] simulations. The transparent bars in the background mark the peripherality percentages of the entire period from time step 99,901 to 1,099,900. These data are plotted separately in **(c). (B)** The distribution of contacts along the z-axis, calculated as in Fig. 5.1G, for six individual simulations. **(C)** The

state-wise contact maps, as measured in Fig. 4.6, for each simulation replicate. For each panel in (B) and (C), the plots were created from $1x10^4$ data points, sampled from time step 99,901 to 1,099,900.

6. Effects of the one-dimensional arrangement of chromatin states on the three-dimensional genome structure

Our lab has determined the chromatin states of genes based on the binding pattern of chromatin-associated proteins across the genome. The analysis was carried out independently for two growth conditions, 25°C and after a shift to 37°C. In 80% of all loci, the state assignment is stable between the two temperature conditions (section 1.4.2). To investigate how the overall protein binding across the genome determines the 3D genome structure, I have adopted polymer models simulating the 25°C and 37°C conditions. For all these simulations (Chapters 3-5), the 25°C state assignment was used to remove any confounding effects by the slightly different state assignments and focus on the effects of segment mobility changes. Now, to refine the [37°C] model and make it closer to 37°C condition, I mapped the 37°C state assignment onto 2 kb segments and analysed the results of new simulations (section 6.1).

The non-random order of co-regulated and co-functional genes on chromosomes have been studied in different eukaryotic cell types (Michalak 2008; Hurst et al. 2004). The 1D proximity of functionally related genes (Zhang & Smith 1998), such as well-known DAL and GAL clusters (Cooper 1996; Hittinger et al. 2004), and groups of adjacent co-expressed genes (Cohen et al. 2000) have been observed in *S. cerevisiae.* However, only about 2 percent of annotated genes in *S. cerevisiae* are found to be in linear proximity of the genes that are functionally related to them (Yi et al. 2007). Our analysis of chromatin states has shown that the distribution of genes of the same chromatin state is nearly random across the genome. However, the occurrence of small clusters (<=3 gene long) is more often than expected by chance (Sewitz et al. 2017a). A preliminary analysis in our lab has compared the impact of natural and artificial order of genes (natural order: experimentally determined chromatin states; artificial order: randomised or linearly clustered genes along chromosomes) on the concerted redistribution of poised proteins upon changes in temperature condition (not shown; Fahmi, Sewitz & Lipkow, *in preparation*). In simulations with natural order of genes, the required time for proteins to relocate to the target genes is shorter than simulations with clustered arrangement of genes. However, the required time for the movement of proteins and the differences between the overall protein occupancy of poised and active genes in simulations with random order of genes is similar to simulations with natural order of genes. To investigate the role of the one-dimensional arrangement of chromatin states along the genome on the 3D chromosome organisation, I ran and analysed simulations with random and 1D clustered patterns of chromatin states (section 6.2).

6.1. Simulations with 37°C state assignment

To make the [37°C] simulations resemble better the 37°C condition, I mapped the determined states at 37°C onto 2 kb segments (section 2.1.2) and ran the [37°C] simulations. I also ran the [25°C] simulations with 37°C state assignment to evaluate if the slight differences between 25°C and 37°C state assignments affect the telomere positions and the state-wise contact maps. In all new simulations the mobility of highly occupied segments, S4 at 25°C and S3 at 37°C, were decreased. In contrast, the segments corresponding to the highly transcribed gene, i.e. S3 at 25°C and S4 at 37°C, had faster mobilities. To achieve the differential mobility of segments, the F_{LC} was scaled by a factor of 5, similar to the simulations that provided the best results in Chapter 3-5. In all simulations, the data was obtained from time step 99,901 to 1,099,900 and combined for three replicates, per condition

6.1.1. Telomere positions

I measured the positions of the right telomere of chromosome III and the left telomere of chromosome XIV. In the simulations with 37˚C state assignment, the telomeres have more peripheral positions (Fig. 1.6E,F) compared to both *in vivo* measurements (Fig. 6.1B) and simulations with 25°C state mapping (Fig. 1.6C,D). In addition, the calculated frequencies for the peripheral location of telomeres in [37°C] simulations (Fig. 1.6F) are outside the 95% confidence interval of *in vivo* data (Table 6.1). The p-values also confirm that the results are not similar to the microscopy data. This divergence was expected as the microscopy measurements were carried out at room temperature, 25°C, and the simulations were informed by the protein occupancy of chromatin segments during heat-shock, 37°C. The observed increase in the peripherality of telomeres in refined [37°C] simulations could be further investigated by performing microscopy experiments at 37°C. It is noteworthy that switching from 25°C to 37°C state assignment has slightly changed the position of telomeres in [25°C] simulations, although the chromatin states are highly similar at two temperatures (Fig. 6.1C,E).

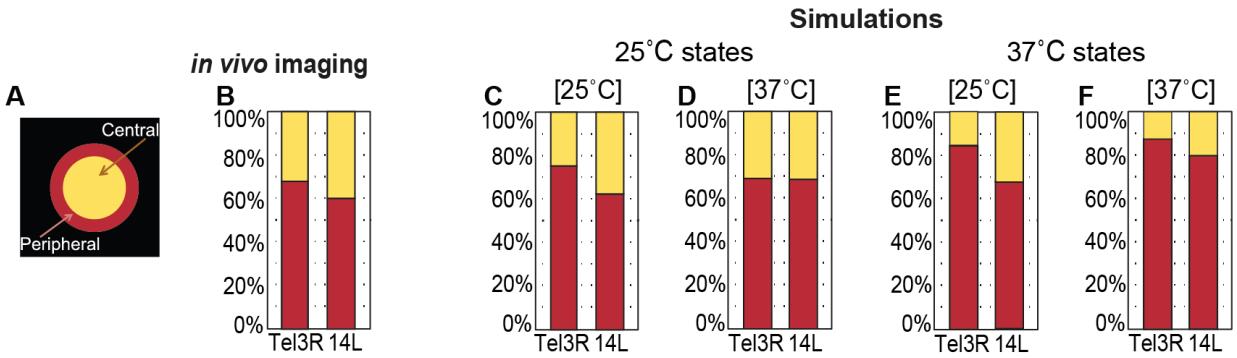


Figure 6.1. Telomere positions in heteropolymeric simulations with 25°C and 37°C state assignments. **(A)** Schematic of separating the nucleus into a peripheral and a central region of equal areas (same as Fig. 3.7A). **(B)** The positions of the labelled right telomeres of chromosome III (Tel3R) and left telomeres of chromosome XIV (Tel14L), measured at room temperature by live-cell microscopy. Number of measurements: n=80 for Tel3R and n=74 for Tel14L. (Plot modified from (Hajjoul et al. 2013), same as Fig. 3.7B). **(C,D)** The quantified peripherality rates for telomeres in simulations representing [25°C] and [37°C] conditions, with 25˚C state assignment (same as Fig. 3.7D,H). **(E,F)** The peripheral position of telomeres were measured for [25°C] and [37°C] simulations with 37°C state assignment. These results are not in agreement with the *in vivo* data (see Table 6.1. for p-values). For each simulation plot (C-F), n=30,000, i.e. 10,000 non-correlated time points were analysed per simulation replicate.

	Tel3R			Tel14L		
	n	% peripheral	p-value	n	% peripheral	p-value
experiments	80	68.7500 (55/80)	n/a	74	60.8108 (45/74)	n/a
Confidence interval (%)	[57.97 ≤ π ≤ 79.53]			[49.01 ≤ π ≤ 72.61]		
[25˚C] 25˚C states	30,000	75.1933	0.1952	30,000	62.3233	0.8109
[37˚C] 25˚C states	30,000	69.0900	1	30,000	68.7100	0.1672
[25˚C] 37˚C states	30,000	84.3933	5.1010e-4	30,000	67.7400	0.2142
[37˚C] 37˚C states	30,000	87.4867	8.9937e-6	30,000	79.8433	2.0477e-4

Table 6.1. Statistical comparison of telomere positions quantified in vivo and in simulations with 25°C and 37°C state assignments. The binomial confidence intervals were calculated for the experimental data of the peripherality rate of telomeres (95% confidence level and continuity correction were applied). The

binomial tests were also used to compare the microscopy and simulation data. MATLAB file exchange function *myBinomTest* was employed. H_0: The experimental data is taken from a population with the same distribution as the simulation data. Small p-values indicate that the distributions are significantly different. Both confidence intervals and p-values show that only the results of simulations with 25˚C chromatin state assignment (rows 5-6) match the microscopy data, but not most of those with 37˚C state assignment (rows 7-8).

6.1.2. State-wise contact maps

The normalised state-wise contact maps were calculated for simulations with 37°C state assignment. The number of contacts between different chromatin states were quantified and the normalisation procedure was performed as explained in section 4.3.2. There are slight differences between the corresponding maps, such as higher intra-S4 and lower intra-S3 interactions in [37°C] simulations with 37°C state assignment (Fig. 6.2H) compared to both 37°C Hi-C (Fig. 6.2B) and [37°C] simulations with 25°C state mapping (Fig. 6.2E). However, the overall pattern of contact maps for each temperature condition are very similar. The correlation coefficients and p-values (Table 6.2) confirm that the results from [25°C] and [37°C] simulations with 37°C state mapping (Fig. 6.2G,H) are highly correlated with the Hi-C contact maps (Fig. 6.2A,B). Interestingly, [25°C] simulations with 25°C state assignment and [37°C] simulations with 37°C state mapping have the highest correlations with Hi-C maps at 25°C and 37°C, respectively (Table 6.2). I used the 37°C state assignment to calculate the contact map of [uniform] simulations (Fig. 6.2F). It shows high intra-state contact frequencies for both S3 and S4 segments, which is not in agreement with the Hi-C contact maps (lower correlation coefficients in Table 6.2).

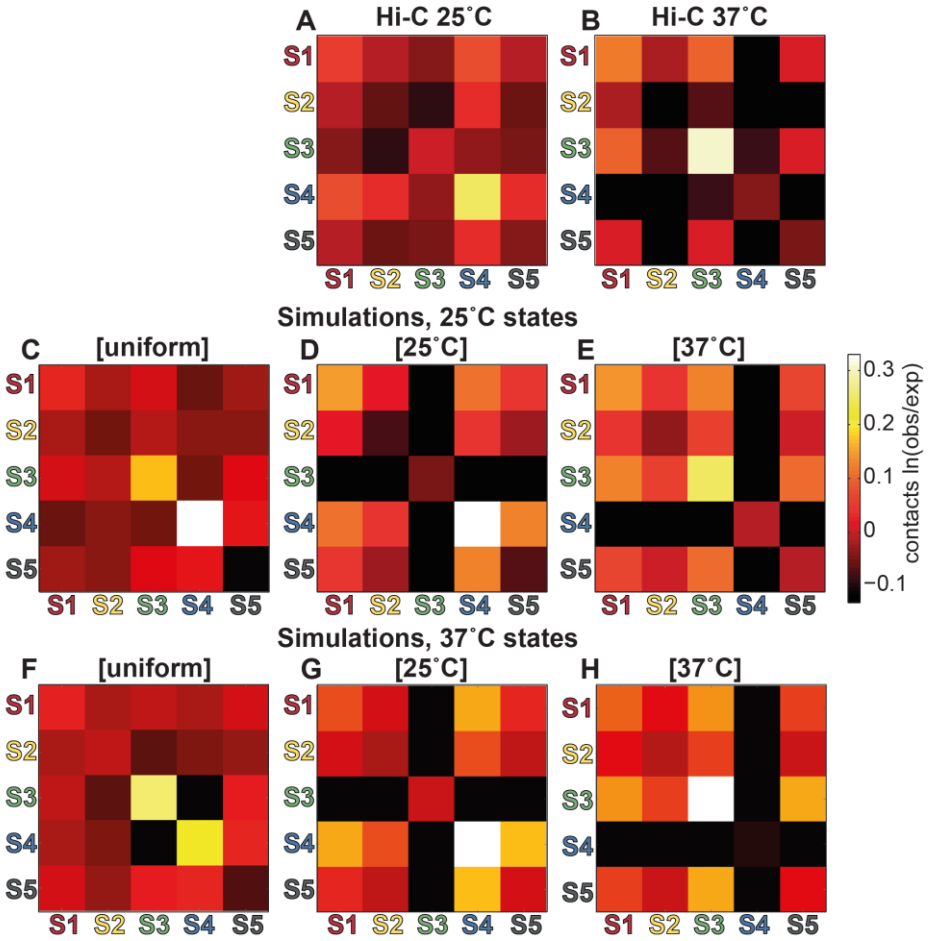


Figure 6.2. State-wise contact maps of [uniform] and heteropolymeric simulations with 25°C and 37°C state assignments. All maps are drawn by the same range of colours. **(A,B)** Normalised contact frequencies of 25°C and 37°C Hi-C data. **(C,D,E)** The analysed state-wise interactions in [uniform], [25°C], and [37°C] simulations with 25°C state assignment for segments. **(F,G,H)** The normalised contact frequencies in [uniform], [25°C], and [37°C] simulations, where 37°C chromatin states were mapped onto chromatin segments. The S6 and S7 segments were not included in the analysis. For each simulation contact map, 10,000 non-correlated time points were analysed, per simulation replicate.

State-wise contact maps		Correlation coefficient	p-value	Significance
Hi-C 25˚C	Simulation [25˚C] 25˚C states	0.8662	2.9712e-5	****
Hi-C 37˚C	Simulation [37˚C] 37˚C states	0.7677	8.3196e-4	***
Hi-C 25˚C	Simulation [25˚C] 37˚C states	0.7520	0.0012	**

Hi-C 37˚C	Simulation [37˚C] 25˚C states	0.7409	0.0016	**
Hi-C 25˚C	Simulation [uniform] 25˚C states	0.7309	0.0020	**
Hi-C 25˚C	Simulation [uniform] 37˚C states	0.6037	0.0172	*
Hi-C 37˚C	Simulation [uniform] 37˚C states	0.5461	0.0352	*
Hi-C 37˚C	Simulation [uniform] 25˚C states	0.4150	0.1240	ns
Simulation [25˚C] 37˚C states	Simulation [37˚C] 37˚C states	-0.1202	0.6695	ns
Hi-C 37˚C	Simulation [25˚C] 37˚C states	-0.1416	0.6148	ns
Hi-C 25˚C	Simulation [37˚C] 37˚C states	-0.2640	0.3418	ns

Table 6.2. Correlation of state-wise contact maps of Hi-C and simulations with 25˚C and 37˚C state assignments. For clarity, the two significant correlations from simulations with 25˚C state mapping are included (from Table 4.1). Pearson correlation coefficients calculated in MATLAB between two vectors of size 15 (the numerical values of the upper-right triangular part of matrices depicted in Fig. 6.2.A,B,F-H). The rows are sorted by correlation coefficient. ***:0.0001<p-value<0.001; **:0.001<p-value<0.01; *:0.01<p-value<0.05; ns: not significant.

6.1.3. Comparison of inter- and intra-state interactions

To further investigate the distribution of state-wise interactions, I compared the intra- and inter-state contacts (obs/exp) of S3 and S4 segments. To make the analysis easier, for each simulation, I first divided the contact frequencies of these segments by the median of all contact frequencies in that simulation (15 values of the upper-right triangular part of the contact matrix (Fig. 6.2), prior to log transformation). Then, I plotted the results for Hi-C and simulation data (Fig. 6.3). As shown by the state-wise contact maps, the highly occupied genes, S4 at 25°C and S3 at 37°C, have very high intra-state contact frequencies in both Hi-C and heteropolymeric simulations. This is also visible in Fig. 6.3A: indicated by the blue and green circles in 25°C and 37°C conditions, respectively. These segments have moderately high (above median) inter-state interactions, as the blue and green stars show in the corresponding plots (Fig. 6.3B). For highly expressed genes, S3

at 25°C and S4 at 37°C, the pattern is completely different: they have moderate intra-state interactions and low inter-state interactions. The distributions of data in [uniform] simulations with 25°C and 37°C state assignments are also interesting. Both S3 and S4 segments have very high intra-state interactions and moderate inter-state contacts. They seem to be in an intermediate state, which could change towards the 25°C or 37°C pattern upon the introduction of differential mobility of segments.

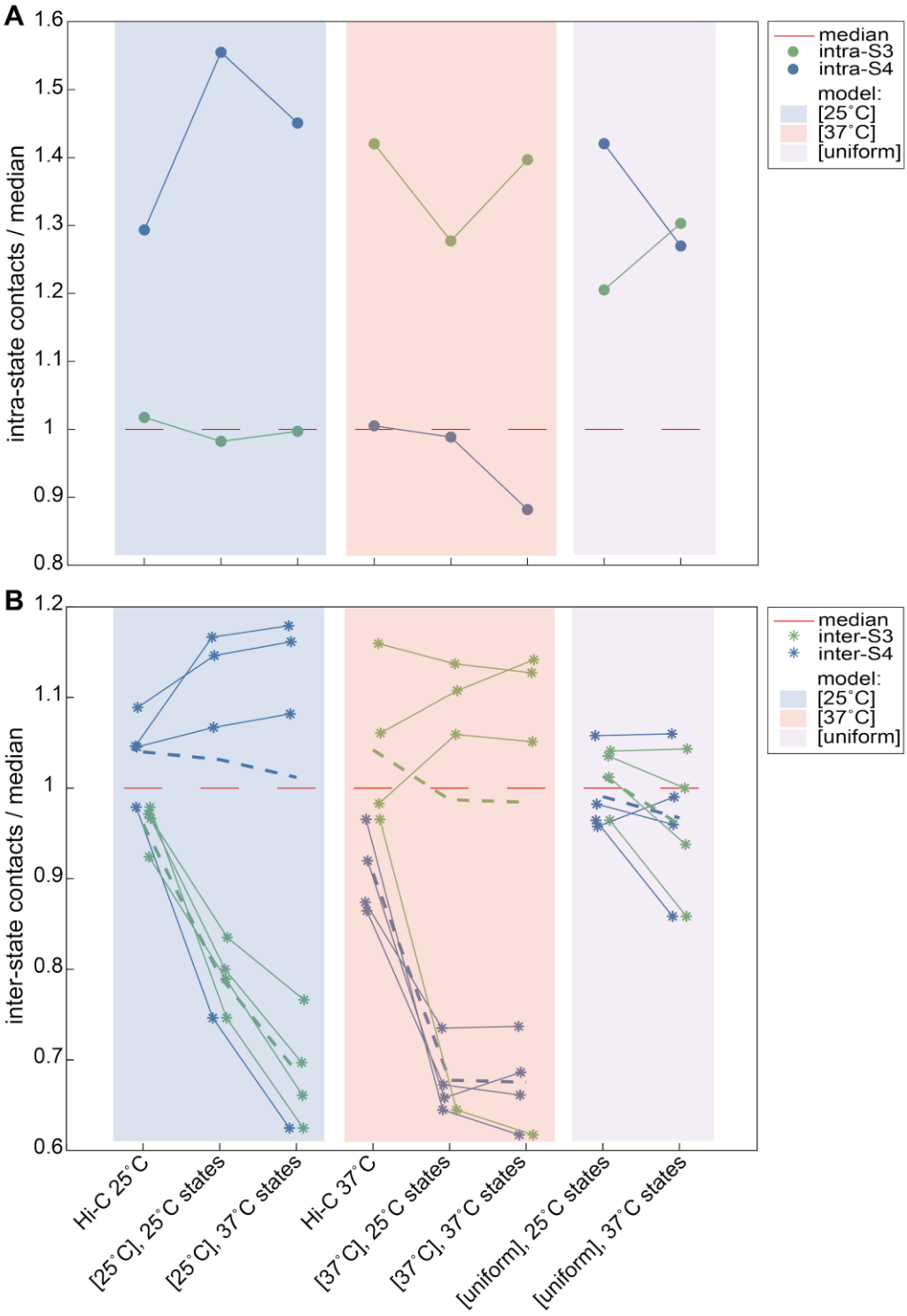


Figure 6.3. Distribution of intra- and inter-state contact frequencies. The intra- and inter-state contacts (observed/expected) of S3 and S4 genes were divided by the median of the calculated contact frequencies (the numerical values of the upper-right triangular part of the contact matrices in Fig. 6.2, prior to log transformation). The results were plotted in MATLAB (S3: green-coloured values, S4: blue-coloured values). The analysis was carried out for Hi-C data and simulations with 25°C and 37°C state assignments. **(A)** Intra-state interactions of S3 and S4 genes measured by Hi-C experiments and simulations of [25°C],

[37°C], and [uniform] conditions. **(B)** Inter-state interactions of S3 segments (4 green asterisks) with S1, S2, S4, and S5 segments; and inter-state interactions of S4 segments (4 blue asterisks) with S1, S2, S3, and S5 segments. The asterisks corresponding to the same pair of states in different datasets are connected to each other by lines to show how their values change. The green and blue dashed lines illustrate the changes in the average values of inter-state interactions of S3 and S4 segments, respectively. To better visualise the asterisks that have close values, each is offset by a distinct x value. The frequency of S3-S4 contacts is drawn twice per dataset. The red lines show the normalised median of each full dataset.

6.2. Simulations with 1D clustered and random state assignment

To investigate the influence of the order of states along the chromosome on the overall genome organisation, I built lists of either 1D clustered or randomly distributed chromatin states. In each case, the total number of segments of each chromatin state per chromosome was kept the same as in the 25°C chromatin state determination (section 2.1.2). For the first type of random lists (A), I randomised all S1-S7 states. For the second type (B), I kept S6 and S7 at their original positions, where they were in 25°C state assignment, and I shuffled the S1-S5 states. For the linearly clustered lists, 10-segment long clusters of S1-S5 states were generated and they were randomly placed along chromosomes. In the first type of clustered lists (A), S6 and S7 were positioned at the boundaries of S1-S5 clusters, while in the second type (B), S6 and S7 had their original position. These random and clustered lists are abbreviated in figures as follows: random states (A), random states (B), clustered states (A), and clustered states (B). For each type of random and clustered states, three distinct lists were generated stochastically, to be used in three simulation replicates. Similar to previous simulations, the mobility of highly occupied segments were decreased by reducing the F_{LC} by a factor 5. In contrast, the mobility of less occupied segments, i.e. highly expressed genes, were increased by increasing the F_{LC} by a factor of 5. For each plot, 30,000 non-correlated time points (10,000 from each of triplicate simulations) were analysed from time step 99,901 to 1,099,900.

6.2.1. Telomere positions

To determine how the linear order of chromatin states affected telomere positioning, I measured the position of telomeres from clustered and randomised states (Fig. 6.4) and compared the results with the microscopy data (Fig. 6.1A,B). In all simulations, the quantified peripherality rates are within the calculated confidence interval for the experimental data (Table 6.3). In addition, high p-values for binomial tests show that the simulated data are not significantly different from the *in*

vivo measurements. It is interesting that despite the changed order of chromatin states, the position of telomeres has not diverged significantly from the confidence interval. We can conclude that, in our heteropolymeric chromosome model, the peripheral position of telomeres is more strongly affected by the quantitative difference in segment mobility (Fig. 3.7) than by the order of segments (Fig. 6.4).

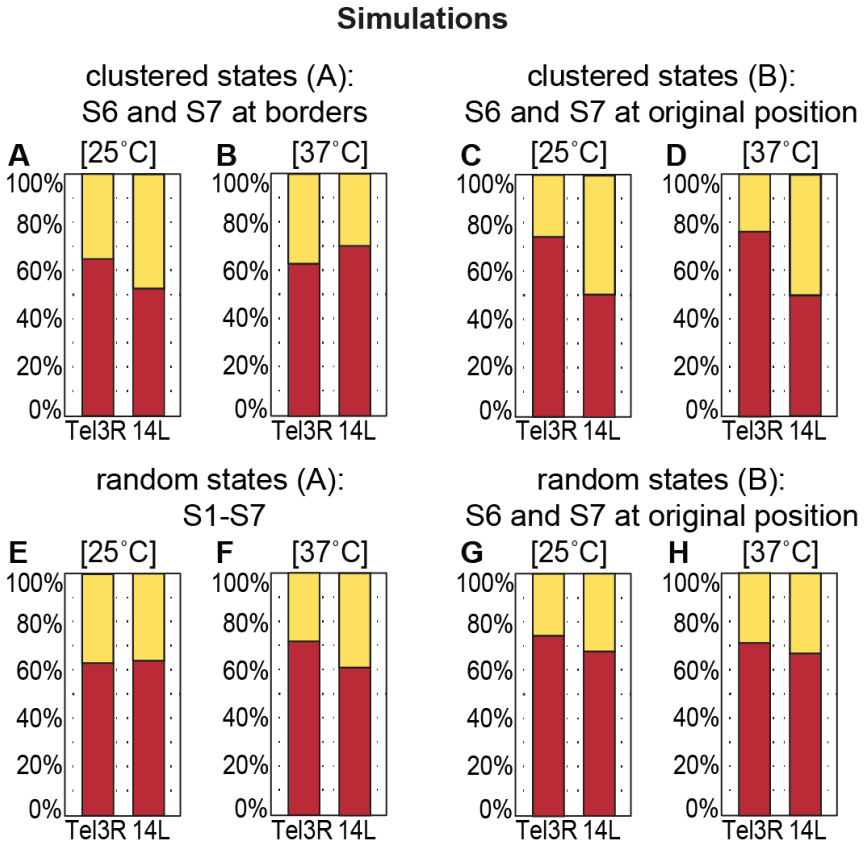


Figure 6.4. Telomere positions in simulations with clustered and randomised states. The positions of the right telomere of chromosome III (Tel3R) and the left telomere of chromosome XIV (Tel14L) in [25°C] and [37°C] simulations with clustered or random chromatin states: clusters of S1-S5 states, where S6 and S7 segments were placed at the borders the clusters **(A,B)**; clusters of S1-S5 states, where S6 and S7 were at their original positions **(C,D)**; randomised S1-S7 states **(E,F)**; randomised S1-S5 states with S6 and S7 at their original positions **(G,H)**.

	Tel3R			Tel14L		
	n	% peripheral	p-value	n	% peripheral	p-value
experiments	80	68.7500 (55/80)	n/a	74	60.8108 (45/74)	n/a
Confidence interval (%)	[57.97 ≤ π ≤ 79.53]			[49.01 ≤ π ≤ 72.61]		
Simulation [25˚C] clustered states (A): S6 and S7 at borders	30,000	64.7133	0.4847	30,000	52.7767	0.2000
Simulation [37˚C] clustered states (A): S6 and S7 at borders	30,000	62.8600	0.2993	30,000	70.2967	0.0761
Simulation [25˚C] clustered states (B): S6 and S7 at original position	30,000	74.2033	0.2531	30,000	50.8300	0.1031
Simulation [37˚C] clustered states (B): S6 and S7 at original position	30,000	76.3867	0.1143	30,000	50.1933	0.0807
Simulation [25˚C] random states (A): S1-S7	30,000	62.9133	0.2996	30,000	64.0733	0.5473
Simulation [37˚C] random states (A): S1-S7	30,000	71.7167	0.5371	30,000	60.9700	1
Simulation [25˚C] random states (B): S6 and S7 at original position	30,000	74.3867	0.2501	30,000	67.8733	0.2131
Simulation [37˚C] random states (B): S6 and S7 at original position	30,000	71.1733	0.6229	30,000	66.9633	0.2673

Table 6.3. Statistical comparison of telomere positions quantified in vivo and in simulations with clustered and randomised states. The 95% confidence intervals of the *in vivo* data were calculated. The

peripherality rates of telomeres in simulations with clustered and random states are within the calculated confidence intervals. The similarity of the microscopy and simulation data was confirmed by high p-values obtained for the binomial tests (MATLAB file exchange function *myBinomTest*). H_0: The experimental data is taken from a population with the same distribution as the simulation data. Small p-values indicate that the distributions are significantly different.

6.2.2. State-wise contact maps

To investigate the effect of 1D arrangement of chromatin states on the spatial structure of chromatin states, I quantified the normalised state-wise contact frequencies for simulations with 1D clustered and random states (Fig. 6.5) (see section 4.3.2 for the normalisation analysis). In all [25°C] and [37°C] simulations, similar to the Hi-C data (Fig. 6.2A,B), the less occupied segments, i.e. S3 at 25°C and S4 at 37°C, have low interactions with segments of other states (visible as a black cross in the state-wise contact maps). In addition, the intra-state contact frequencies of segments with high protein occupancy, i.e. slow moving segments, are high and comparable to the observed pattern in the Hi-C maps. However, other calculated interactions are very different from the observed Hi-C frequencies (Fig. 6.2A,B). All of them have higher obs/exp values, except for the intra-state interactions of less occupied segments in random simulations, which are lower than the corresponding Hi-C interactions. The calculated Pearson correlations confirm that the simulated contact maps of clustered and randomly arranged heteropolymeric simulations are not well correlated with the Hi-C contact maps (Table 6.3).

The state-wise contact maps of [uniform] simulations were calculated with the linearly clustered and random lists of states mapped onto the interacting chromatin segments (Fig. 6.5A,D,G,J). The results show prominently high intra-state interactions for simulations with clustered states (Fig. 6.5A,D). In simulations with random states, the arbitrary pattern of interactions is evident (Fig. 6.5G,J). In addition, the position of S6 and S7 segments change the distribution of contacts between S1-S5 chromatin states (Fig. 6.5G,J). The [uniform] simulation with random states (A) has a moderately high correlation with the Hi-C 25°C state-wise contact map (Table 6.3), although the patterns are not visually similar.

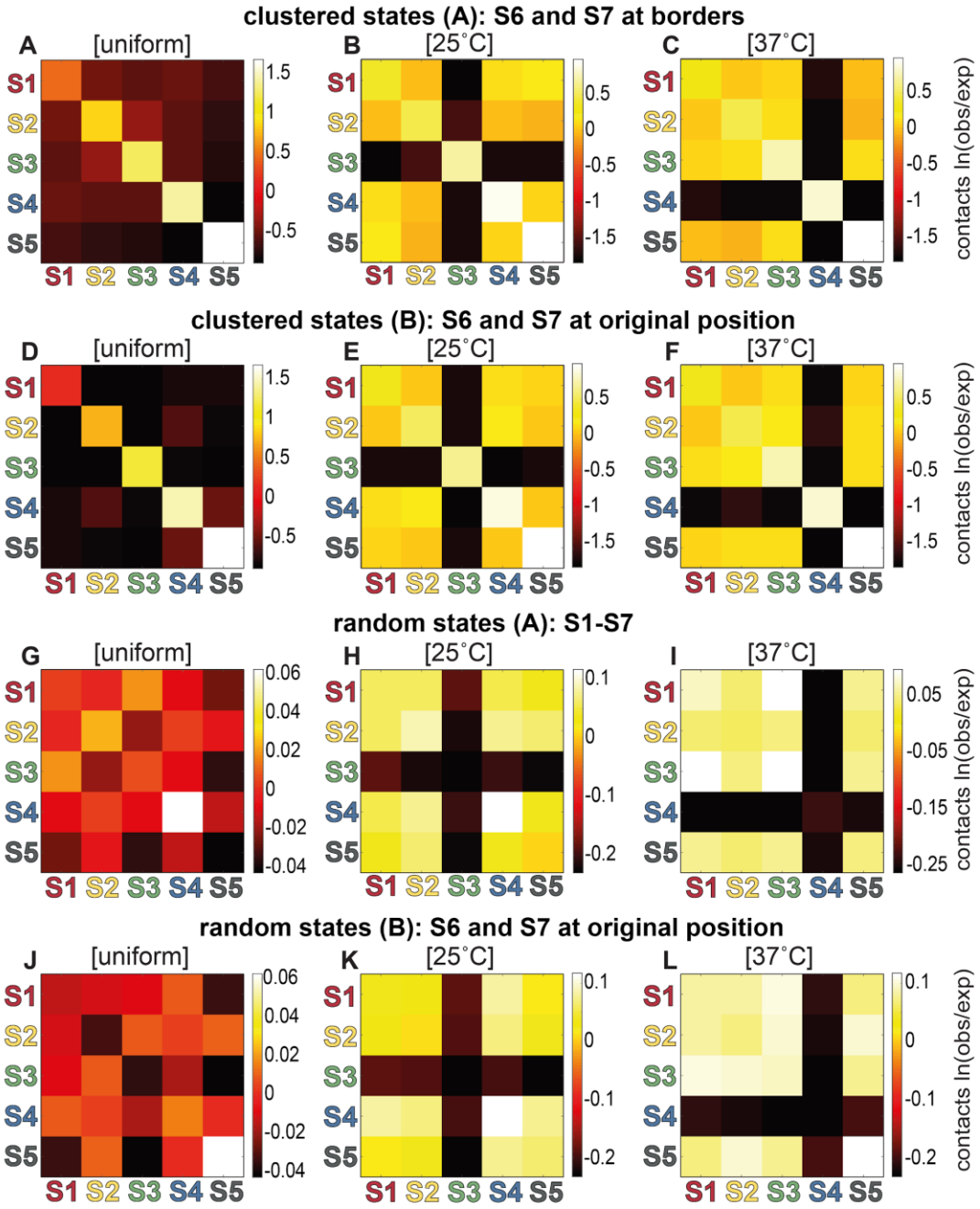


Figure 6.5. State-wise contact maps of simulations with clustered and randomised states. The normalised contact frequencies were calculated for the [uniform], [25°C], and [37°C] simulations. Different state assignments were employed for these simulations: clustered S1-S5 states with S6 and S7 at the borders of clusters **(A-C)**; clustered S1-S5 states with S6 and S7 at their original positions **(D-F)**; randomised S1-S7 states **(G-I)**; randomised S1-S5 states, where S6 and S7 were placed at their original positions **(J-L)**.

State-wise contact maps		Correlation coefficient	p-value	Significance
Hi-C 25˚C	Simulation [25˚C] clustered states (A): S6 and S7 at borders	0.5018	0.0567	ns
Hi-C 37˚C	Simulation [37˚C] clustered states (A): S6 and S7 at borders	0.5617	0.0293	*
Hi-C 25˚C	Simulation [25˚C] clustered states (B): S6 and S7 at original position	0.4893	0.0642	ns
Hi-C 37˚C	Simulation [37˚C] clustered states (B): S6 and S7 at original position	0.5589	0.0303	*
Hi-C 25˚C	Simulation [25˚C] random states (A): S1-S7	0.4909	0.0632	ns
Hi-C 37˚C	Simulation [37˚C] random states (A): S1-S7	0.6006	0.0179	*
Hi-C 25˚C	Simulation [25˚C] random states (B): S6 and S7 at original position	0.5347	0.0400	*
Hi-C 37˚C	Simulation [37˚C] random states (B): S6 and S7 at original position	0.5087	0.0528	ns
Hi-C 25˚C	Simulation [uniform] clustered states (A): S6 and S7 at borders	0.3328	0.2255	ns
Hi-C 37˚C	Simulation [uniform] clustered states (A): S6 and S7 at borders	0.3745	0.1691	ns
Hi-C 25˚C	Simulation [uniform] clustered states (B): S6 and S7 at original position	0.3948	0.1453	ns
Hi-C 37˚C	Simulation [uniform] clustered states (B):	0.2980	0.2806	ns

	S6 and S7 at original position			
Hi-C 25˚C	Simulation [uniform] random states (A): S1-S7	0.6397	0.0102	*
Hi-C 37˚C	Simulation [uniform] random states (A): S1-S7	0.0799	0.7772	ns
Hi-C 25˚C	Simulation [uniform] random states (B): S6 and S7 at original positions	0.1885	0.5011	ns
Hi-C 37˚C	Simulation [uniform] random states (B): S6 and S7 at original positions	-0.3901	0.1506	ns

Table 6.4. Correlation of state-wise contact maps of Hi-C and simulations with clustered and randomised states. Pearson correlation coefficients were calculated in MATLAB between two vectors of size 15 (the numerical values of the upper-right triangular part of matrices depicted in Fig. 6.2.A,B and Fig. 6.5). *:0.01<p-value<0.05; ns: not significant.

6.2.3. Comparison of inter- and intra-state interactions

The ranges of the calculated state-wise contact frequencies of simulations are different from the Hi-C data (see colour bar values in Fig. 6.5 vs Fig. 6.2A,B). To better visualise and assess the resemblance between the simulation and experimental datasets, I plotted the intra- and inter-state interactions of S3 and S4 segments (Fig. 6.6). Similar to Fig. 6.3, the interactions (obs/exp) are divided by the median of all contact frequencies in each dataset. The positions of S6 and S7 states have slight influences on both intra- and inter-state interactions in [25°C] and [37°C] simulations with clustered and randomised states (Fig. 6.6A,B,D,E). Heteropolymeric simulations with 1D clustered states have wider ranges of contact frequencies compared to the Hi-C data. The S3-S3 and S4-S4 interactions are much higher in these simulations compared to the Hi-C data (Fig. 6.6A,B), while inter-state contacts of S3 segments in [25°C] and those of S4 segments in [37°C] simulations are very low (Fig. 6.6D,E). The heteropolymeric simulations with random chromatin states have some similarities with the experimental data. The inter-state interactions of fast and slow moving segments are low and moderately high, respectively (fast: S3 at 25°C and S4 and 37°C; slow: S4 at 25°C and S3 at 37°C) (Fig. 6.6D,E). However, unlike the observed Hi-C data, the intra-state contacts of S3 segments at [25°C] and those of S4 segments at [37°C] are lower than the median (Fig. 6.6A,B). In addition, the highly occupied segments do not have high enough

number of interactions with each other (Fig. 6.6A,B). In the [uniform] simulations with clustered states, the intra-state contacts are much higher than the median, which indicates that in the absence of heterogeneous mobility of segments, the 3D clustering of segments is stronger (Fig. 6.6C). In [uniform] simulations with random states, the spread of the data points is small and the inter-state interactions of S3 and S4 segments have moderate values, which are not matched with the experimental distributions (Fig. 6.6F).

Therefore, all simulations with clustered chromatin states have very different patterns compared to the experimental data and other simulations. Heteropolymeric simulations with randomised state mappings have some similar features to the Hi-C data and simulations with 25°C state assignment. The comparisons between the output of different simulations and experimental data re-confirms that the [25°C] and [37°C] simulations with 25°C state assignment provide the best match to the Hi-C data.

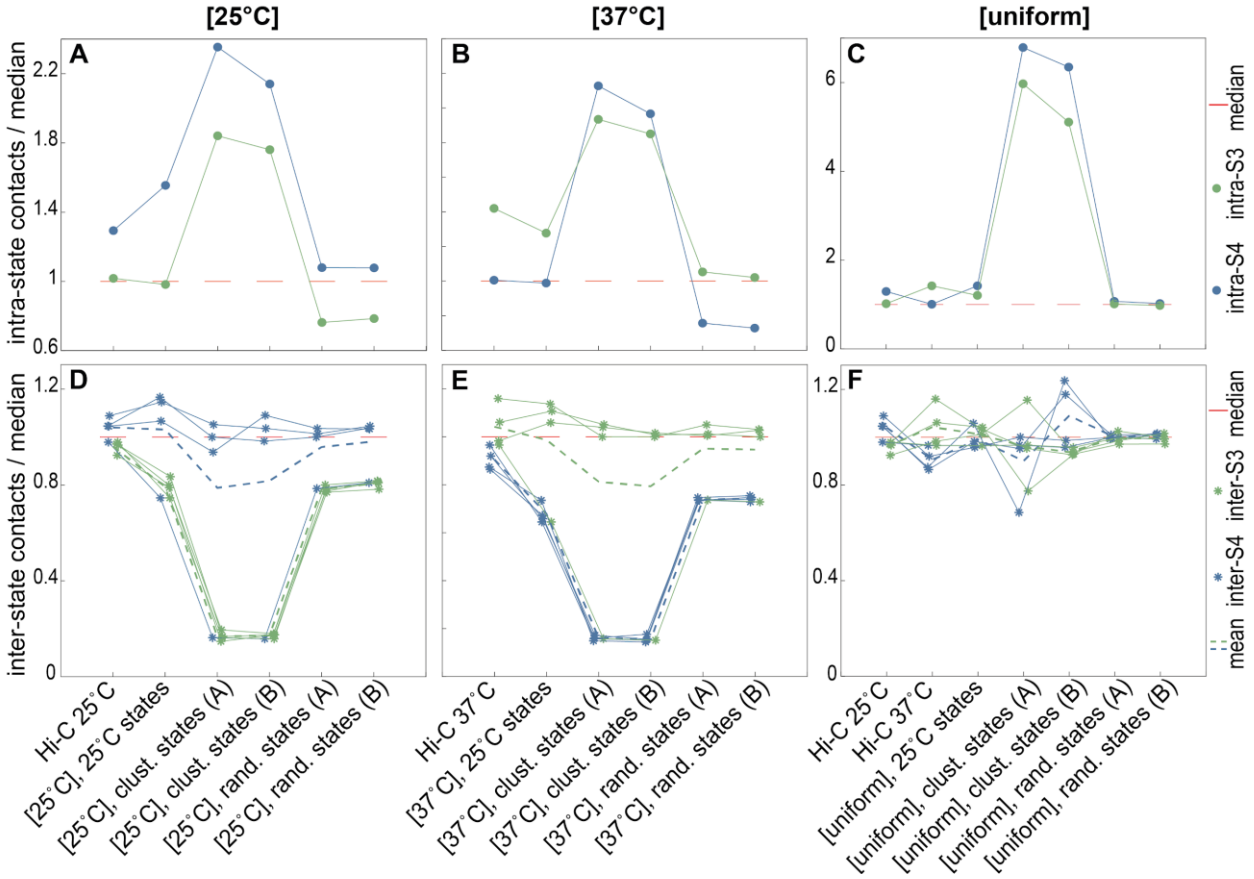


Figure 6.6. Distribution of intra- and inter-state contact frequencies for simulations with clustered and randomised states. The contact frequencies of S3 and S4 segments were divided by the median value of all contact frequencies in that simulation (the numerical values of the upper-right triangular part of the

contact matrices in Fig. 6.2, prior to log transformation). Then, the results were plotted in MATLAB for Hi-C and simulation data. The results of [25°C] (A,D), [37°C] (B,E), and [uniform] (C,F) simulations with 25°C state assignment, 1D clustered and randomised states are represented. **(A-C)** The intra-state contacts of S3 and S4 segments: green and blue circles respectively. **(D-F)** The inter-state interactions of S3 and S4 segments: 4 green asterisks for S1-S3, S1-S2, S3-S4, S3-S5 and 4 blue asterisks for S1-S4, S2-S4, S3-S4, S4-S5. The corresponding asterisks in different datasets are connected to each other by lines. The green and blue dashed lines illustrate the changes in the average values of intra-state interactions of S3 and S4 segments, respectively. To better visualise the asterisks that have close values, each star has a distinct x value. The red lines show the medians. (clust. states (A): S6 and S7 at borders; clust. states (B): S6 and S7 at original position; rand. states (A): S1-S7; rand. states (B): S6 and S7 at original positions)

6.2.4. Total number of interactions

For the final analysis, I quantified the number of total interactions between S1-S5 segments for different simulations. The data is obtained from three simulation replicates per simulation type and the mean values are plotted as bar graphs with error bars (Fig. 6.7). The results show that the heteropolymeric simulations have higher numbers of total interactions than the [uniform] simulations. In addition, the [37°C] simulations have about 12 percent more interactions than the [25°C] simulations on average. Similar observations were made in Fig. 4.4, where the total number of S1-S7 contacts were quantified for simulations with 25°C state assignment. It is notable that changing the state assignment of chromatin segments has not affected this pattern. Switching from the 25°C state assignment to the 37°C states has led to more S1-S5 contacts in [25°C] simulations. Linearly clustered chromatin states have raised the number of interactions for heteropolymeric simulations, while the random state assignments have slightly decreased those numbers. Thus, the 1D order of the chromatin states that were mapped onto the modelled segments has an impact on the total number of interactions between the chromatin segments.

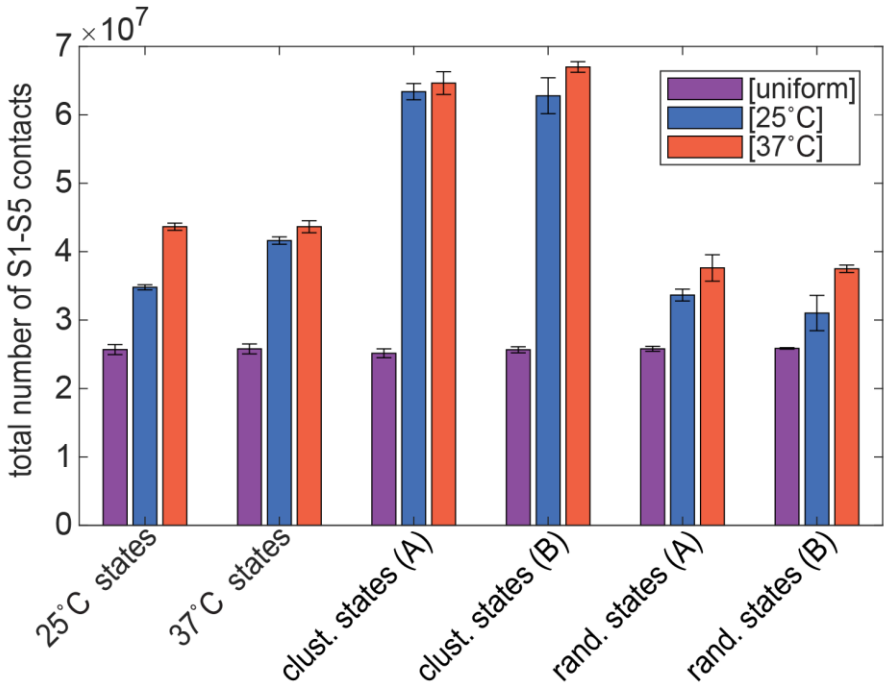


Figure 6.7. Total number of interactions in simulations. The total number of contacts between S1-S5 chromatin segments were quantified for different [uniform], [25°C], and [37°C] simulations. The mean and the standard deviation (error bar) of three replicate simulations are represented, per simulation type. (clust. states (A): S6 and S7 at borders; clust. states (B): S6 and S7 at original position; rand. states (A): S1-S7; rand. states (B): S6 and S7 at original positions)

7. Discussion

7.1. Heterogeneous mobility of chromatin segments leads to self-organisation of chromosomes

To understand the three dimensional organisation of chromatin in *S. cerevisiae*, I employed a heteropolymer model informed by the genome-wide DNA-binding data of chromatin associated proteins. Analysis of this data has shown that the distribution of proteins across the genome at two growth conditions, normal growth (25°C) and heat-shock (37°C), is highly heterogeneous in each case. In addition, the transition from one temperature to another changes the protein occupancy of chromatin segments and their transcriptional activity. We propose and provide a computational model to show how this affects the mobility of corresponding genes and overall genome organisation. In this model of [25°C] and [37°C] simulations, chromatin segments with higher protein occupancy are modelled as slow-moving segments, while the highly expressed genes, which are less occupied, have higher mobilities. This heteropolymeric model has reproduced the experimentally observed patterns for the positioning of telomeres (Fig. 3.7), spatial distances between loci (Fig. 3.8), contact frequencies between chromatin segments (Fig. 4.6), and relocation of heat-shock genes upon activation (Fig. 5.4). Hence, our proposed mechanism, the differential mobility of segments, is an important determinant of the 3D chromatin organisation.

7.2. Different mechanisms drive 3D genome organisation

Different computational models have investigated how the organised structure of chromosomes is derived by protein-mediated interactions (Giorgetti et al. 2014; Chiariello et al. 2016; Jost et al. 2014; Brackley et al. 2016; Cheng et al. 2015; Barbieri et al. 2012; Haddad et al. 2017), loop extrusion (Sanborn et al. 2015; Fudenberg et al. 2016), and supercoiling of DNA (Le et al. 2013; Benedetti et al. 2014; Racko et al. 2017). These models successfully simulated the folding and compartmentalization of chromatin structure, as reported by microscopy and Hi-C data. Therefore, it could be speculated that these suggested mechanisms, including the heterogeneous mobility of chromatin fibre, work together to form and maintain the hierarchical organisation of the genome. Each of these factors could influence the various scales of genomic structure in different ways. For example, the depletion of CTCF, which works with cohesin to facilitate the formation of TADs (Yuen & Gerton 2018), does not interrupt the A and B compartments (Nora et al. 2017; Kubo et al. 2017), however it slightly weakens the insulation of most TADs in mouse embryonic stem cells (Kubo et al. 2017). These mechanisms could also affect each other: the formation of protein bridges

may slow down the mobility of chromatin regions and vice versa, i.e. the slow moving segments might have a higher chance of being involved in protein-protein interactions, while the fast moving ones might avoid those interactions. Therefore, models based only on effective interactions between loci could take the differential mobility of loci into account implicitly; and the other way around. Computational models, which combine different mechanistic explanations for the organised genome structure, could lead to a comprehensive view of how they influence and complement each other at different structural levels.

7.3. Genes of different chromatin states are spatially separated

An important feature of living systems is that they have internal sources of energy, which leads them to diverge from thermal equilibrium. Due to several nuclear processes that consume or release energy and the heterogeneous nature of chromatin, the forces that drive the movement of chromatin regions will not be statistically homogenous (Ganai et al. 2014). Some studies have focused on ATP-consuming processes, such as transcription, and changed the local temperature (effective temperature) (Cugliandolo 2011; Loi et al. 2008) of chromatin regions according to their gene density (Ganai et al. 2014; Agrawal et al. 2017). This temperature difference has led to the separation of gene-dense and gene-poor chromosomes, which is analogous to the activity-induced separation of cold (inactive) and hot (active) polymers (Smrek & Kremer 2017) or particles (Grosberg & Joanny 2015).

In our polymer model, statistically heterogeneous forces have been applied to chromatin segments to slow down or speed up their movement according to their protein occupancy. The co-localisation of slowly moving chromatin segments, which are highly occupied by poised chromatin-associated proteins, and the loose, peripheral distribution of fast moving segments, i.e. highly expressed genes, are predicted by our polymer model, Hi-C data, and microscopy measurements (Fig. 4.6, Fig. 5.1, Fig. 5.2, Fig. 5.4). This is similar to the spatial separation of active (self-propelled) and passive spherical particles, where a cluster of less diffusive particles is surrounded by highly diffusive ones (Stenhammar et al. 2015; Weber et al. 2016). The driving force behind this separation is a positive feedback loop triggered by the aggregation of particles with slower movements, which leads to more collisions between them and further slows down their mobility (Gonnella et al. 2015).

The aggregation of poised genes could be mechanistically comparable to the assembly of membraneless compartments, such as super enhancers (Hnisz et al. 2017) and stress granules,

which are derived by phase separation (Boeynaems et al. 2018). It has been shown that some of these structures are mediated by interactions between intrinsically disordered proteins (Dao et al. 2018; Burke et al. 2015). This is consistent with the protein disorder analysis performed in our lab, which indicated that the poised chromatin segments are enriched for proteins with high levels of disordered regions (unpublished).

7.4. Clusters of poised genes facilitate their rapid activation

The association of silent or not highly expressed genes with inactive transcription factories, which contain RNA polymerase II phosphorylated on Ser5 but not on Ser2, has been shown by ChIP experiments (Ferrai et al. 2010; Stock et al. 2007; Wu & Snyder 2008). It has been suggested that the presence of these poised transcription factories allows rapid activation of genes in response to inducing signals (Sutherland & Bickmore 2009; Rieder et al. 2012). Our group's genome-wide analysis of protein binding data and gene expression levels have shown that a large proportion of chromatin-associated proteins, including main subunits of RNA polymerase II, are in a poised state in *S. cerevisiae* cells (Sewitz et al. 2017a). The poised genes, which are highly occupied by poised proteins, are seen to co-localise in our polymer models. The high concentration of these genes together with their slow movement could facilitate efficient interactions between proteins, which could facilitate the formation of inactive transcription factories. The high occupancy of Rbp2p and Rpb3p subunits at poised genes could be a sign for the presence of these complexes. In addition, it could be envisioned that these pre-assembled factories are already associated with multiple genes that would become co-expressed in response to changes in condition. This would lead to fast activation of these genes as they are already close to each other and to the transcription factories. An alternative possibility is that the poised genes will relocate to the proximity of transcription factories upon activation, as shown in mammalian cells (Osborne et al. 2007). These genes could aggregate in their poised state, analogous to chromatin hub structures which are formed prior to transcription (Mitchell & Fraser 2008), and enter the factories together. Therefore, the co-localisation of poised genes could contribute to the prompt activation of genes through the assembly of inactive transcription factories, aggregation of co-regulated genes before activation, or association of poised genes with inactive factories.

7.5. A loose conformation of active genes allows efficient transcription

Although compact aggregation of poised genes could provide the required framework for preparation of transcription, the genes need to have more space to be able to slide through the

transcription machinery while they are transcribed (Sutherland & Bickmore 2009). In addition, a more loosely packed chromatin conformation would allow easier access of regulatory elements, such as transcription factors and upstream activating sequences (UAS) (Petrascheck et al. 2005), to DNA. Hence, the predicted loose distribution of highly expressed chromatin segments would be in favour of transcription. This conformation could accommodate discrete clusters of co-expressed genes, such as co-localised target genes of the same transcription factor (Ben-Elazar et al. 2013; Janga et al. 2008), interchromosomal clusters of genes that are associated with the NPC (nuclear pore complex) (Brickner 2017; Brickner et al. 2012), and the nucleolar cluster of tRNA genes (Thompson et al. 2003; Haeusler et al. 2008).

7.6. The heterogeneous mobility of chromatin segments results in compact genome structure, which is comparable to mammalian chromosomal domains

As was shown by (Wong et al. 2012), a homopolymer model of chromosomes with only the rDNA treated differently can achieve key characteristics of genome organisation, such as the spatial positioning of the genome and inter-chromosomal contacts. This model is identical to our [uniform] model, except that the latter has a higher resolution (2 kb vs. 5 kb segments) and thinner rDNA segments. The conformation of chromosomes in our homopolymeric model is not as compact as in our heteropolymeric models. This is indicated by the position of telomeres very close to the nuclear periphery (Fig. 3.7F), higher 3D distances of loci (Fig. 3.8E), lower number of total contacts between chromatin segments (Fig. 4.4A, Fig. 6.7), and a lower degree of clustering for the whole genome (Table 5.2). These results have also shown that the modelled genome structure in our [uniform] simulations is not in quantitative agreement with the microscopy and Hi-C data. However, a heteropolymeric model with differential mobility of segments has led to a more compact conformation of chromosomes in [25°C] and [37°C] simulations. Higher number of contacts between chromatin regions and a higher degree of clustering show that the [37°C] model has adopted an even more condensed structure than the [25°C] simulation.

The overall genome structure in budding yeast has different features from the organisation of chromosomes in higher eukaryotes. For example, the structure of human chromosomes has a fractal globule shape, while chromosomes in yeast are arranged in the Rabl configuration. However, the size of the yeast genome (~12 Mbp) has a comparable scale to a single mammalian chromosome, which occupies a distinct territory inside the nucleus. Therefore, similar mechanisms might govern their organisation and some structural similarities could be noted in comparisons. In human and

mouse cells, both active and silent genes have been observed on the interior region of chromosomal domains by microscopy experiments (Mahy, Perry, Gilchrist, et al. 2002; Küpper et al. 2007; Cremer & Cremer 2010). In other studies, highly transcribed genes were seen to be enriched on the outer layer of their chromosomal territory or on extended loops away from the main mass of the territory (Volpi et al. 2000; Williams et al. 2002; Mahy, Perry & Bickmore 2002; Chambeyron & Bickmore 2004; Nagano et al. 2013). In *S. cerevisiae*, as the 2D projection of simulated contacts has shown, the poised genes are located in the internal zone of the genome and the highly expressed genes are present at both interior and exterior layers (Fig. 5.1, Fig, 5.2). Active genes have also adopted a less condensed conformation in our polymer models, and some genes are seen to shift to the nuclear periphery upon activation, by both microscopy measurements (Taddei et al. 2006; Sarma et al. 2007; Dieppois et al. 2006; Brickner & Walter 2004; Casolari et al. 2004) and our heteropolymeric models (Fig. 5.4). This indicates that the mechanisms we describe for the entire yeast genome could operate on the scale of individual mammalian chromosomes, leading to active genes being located at the periphery of chromosome territories, analogous to the nuclear periphery in yeast.

7.7. Heat-shock genes relocate to the nuclear periphery upon activation

Strikingly, our simple heteropolymeric chromosome model has been able to reproduce the relocation of activated *HSP104* gene to the proximity of nuclear membrane. Different mechanisms have been proposed to explain the relocation of activated genes to the nuclear periphery in budding yeast. Some studies have suggested that the mRNA export proteins (Vinciguerra & Stutz 2004; Casolari et al. 2005; Dieppois et al. 2006) or transcription factors (Randise-Hinchliff & Brickner 2016) establish links between the expressed genes and the nuclear pore complex (NPC), however, it has been shown that the recruitment of genes to the nuclear periphery happens even in the absence of active transcription (Schmid et al. 2006; Brickner et al. 2007) and specific nuclear pore complex proteins (Guet et al. 2015). Therefore, it can be concluded that different mechanisms, including the heterogeneous mobility of segments, contribute to the positioning of activated genes to the nuclear periphery: the mobility effect leads to a rapid shift of activated genes towards the nuclear membrane, followed by transcription- or NPC-dependent interactions to maintain the peripheral localisation of the gene. The lower absolute values of the peripherality rates of the *HSP104* gene in our model compared to the microscopy data could indicate that the model is missing factors, which stabilise the peripheral location of the gene, such as interactions with proteins of the NPC.

In our model, other heat-shock genes shift to the nuclear periphery at 37°C as well. However, a group of genes shifts towards the centre of the nucleus upon activation (Fig. A.4). This has led us to propose that the positioning of a gene upon activation could be affected by the activity of its adjacent genes. An example of such genes are the α-globin genes in mammalians. The difference between the position of α-globin genes in mouse and human cells has been explained by their distinct local chromatin environment. In mouse, these genes reside on a gene-poor region and consequently remain inside their chromosomal territory irrespective of their transcriptional activity (Brown et al. 2006). On the other hand, in human cells, the α-globin genes are located on a gene-rich region, which is enriched for housekeeping genes, and loop out of their chromosomal territory even when they are not expressed (Mahy, Perry & Bickmore 2002).

7.8. The linear arrangement of genes on chromosomes affects the 3D genome structure

The determined lists of chromatin states at 25°C and 37°C conditions are very similar to each other. To investigate how the slight differences between these lists influence the chromosome conformation, I ran the [25°C] and [37°C] simulation with both 25°C and 37°C state assignments. The results have shown that the different state assignments result in some differences in the position of telomeres and contact frequencies of chromatin segments. The [25°C] simulations with 25°C state assignment provide the best match to the microscopy and Hi-C data at 25°C (Fig. 6.1, Fig. 6.2). The [37°C] simulations with 37°C state assignment result in contact maps, which have the highest correlation with contact maps of Hi-C experiments at 37°C (Fig. 6.2). Therefore, the telomeric positions and the state-wise contact frequencies are sensitive to changes in the chromatin state of genes, even if 80% of genes remain unchanged.

To further investigate the effect of linear order of genes on 3D genome organisation, I mapped the 1D clustered and randomised lists of chromatin states to the modelled chromatin segments. In these lists, the total number of genes of each chromatin state is the same as the 25°C state assignment. The analysed telomeric positions in simulations with clustered and randomised states are in agreement with the microscopy data (Fig. 6.4). This suggests that the heterogeneous mobility of genes and the preserved number of genes of each chromatin state, per chromosome, is sufficient to lead to the observed patterns of telomere positions in living cells, whereas for these features the order of states is less important. On the other hand, the normalised state-wise contact frequencies that are obtained from simulations with clustered states do not match the Hi-C data (Fig. 6.5, Fig.

6.6). However, the heteropolymeric simulations with randomised chromatin states do reproduce some of the observed features in Hi-C contact maps. These results are in agreement with our analysis that the natural order of genes is nearly random, as was shown in (Sewitz et al. 2017a), and not composed of linear clusters of genes that have similar functions or are in the same chromatin state.

Different studies have reported the presence and the importance of linear proximity of genes controlled by the same transcription factor (Janga et al. 2008) or genes with common function (Yi et al. 2007), such as genes encoding stably interacting proteins (Teichmann & Veitia 2004) or genes of the same metabolic pathway (Lee & Sonnhammer 2003). The proximity of these genes can aid the efficient and coordinated regulation of expression (Osbourn & Field 2009). However, we found in particle-based simulations that large clusters of co-regulated genes can actually slow down the rapid, coordinated movement of regulatory and poised proteins between different sets of genes, which is crucial upon changes in cellular environment (not shown; Fahmi, Sewitz & Lipkow, *in preparation*). In addition, having gaps between co-regulated genes might help them to come closer to each other in 3D when chromatin folds, this is consistent with the observed periodic position of genes that are regulated by the same transcription factor (Képès 2003). The randomised order of genes in both particle-based and heteropolymeric simulations has achieved similar results to the experimental data. Therefore, it could be postulated that the nearly random spread of genes, where small clusters of 2 or 3 genes are more probable than expected by chance (Sewitz et al. 2017a), would facilitate both efficient regulation and rapid redistribution of chromatin-associated proteins.

7.9. Hi-C data analysis

A growing number of bioinformatics pipelines is being developed to analyse Hi-C data obtained from different organisms (Shavit et al. 2016; Lazaris et al. 2017; Servant et al. 2015; Paulsen et al. 2018). These tools adopt distinct methods to correct the experimental biases and artifacts. Therefore, they can generate different results for chromatin interactions from the same input Hi-C data (Forcato et al. 2017). The normalisation of our Hi-C data by HOMER and SeqMonk resulted in different patterns for the contact frequencies of chromatin states (Fig. 4.6, Fig. 4.8). Similar to other Hi-C tools, such as hiclib (Imakaev et al. 2012) and HiC-Pro (Servant et al. 2015), HOMER uses iterative matrix balancing, which assumes that all chromatin regions would have a similar total number of interactions with other regions if there were no biases. We have learned that this method should not be used for the yeast genome as it disregards the Rabl configuration of

chromosomes, which results in the centromeric regions to be highly spatially clustered and telomeric loci to be closer to the nuclear membrane, which therefore have significantly less contacts with the rest of the genome. Hence, this approach could artificially distort expected contact frequencies for these regions. This is evident in state-wise contact maps calculated from HOMER reports, where the high number of intra-state interactions of poised genes (S4 at 25°C and S3 at 37°C) has been balanced out by reducing the inter-state contact frequencies of those genes (Fig. 4.8). Therefore, choosing the right normalisation approach requires careful attention to the genomic features of the organism under investigation as they greatly influence the analysis of the data.

7.10. Future work

7.10.1. The effect of 1D arrangement of chromatin states on clustering degree of genes and gene relocation upon activation

To study how the 3D genome organisation is influenced by the heterogeneous distribution of chromatin-associated proteins and the linear order of genes on chromosomes, I have developed and analysed a variety of whole-genome polymer simulations. Additional insights into the organised structure of chromosomes can be gained from further analysis of my existing simulation data.

One of the most straightforward steps to take is to investigate the spatial distribution of segments in heteropolymeric simulations with 37°C state assignment, and to compare these to the observed patterns in simulations with 25°C state mapping. This would clarify whether slight differences in 25°C and 37°C state assignments affect the degree of 3D clustering of genes and the relocation of heat-shock genes to the nuclear periphery upon activation. This analysis can be repeated for simulations with artificial orders of genes, i.e. randomised or 1D clustered arrangements of genes. The quantified contact frequencies (Fig. 6.5, Fig. 6.6) have predicted a higher co-localisation level for genes of the same chromatin state in simulations with 1D clustered states and a lower co-localisation degree in simulations with randomised states, compared to simulations with natural order of genes. The 2D projection of contacts and the quantification of clustering levels would specify the magnitude of these differences. According to the proposed hypothesis in section 5.2, the chromatin state of neighbouring segments affects the relocation of an activated gene. The position of heat-shock genes in simulations with random or clustered states would provide a verification of this assumption.

7.10.2. Relocation of genes with facilitator and inhibitor neighbours

Another interesting analysis to undertake is the inspection of the relocation of genes with facilitator neighbours, which are identified by high enrichment scores (section 5.2). This could be conducted by tracking the position of genes before and after activation using the simulation results and live cell imaging. The simulation analysis is not complicated, as previously developed code for the relocation study only needs to be tweaked slightly. However, microscopy imaging requires more time to be set up, and yeast strains with specific labelled chromosomal loci need to be created or sourced from other groups.

In addition, the relocation of genes could be compared to their displacement (MSD curves in Fig. 5.5 and Fig. 5.6) to test whether the fast movement of activated genes and their relocation to the nuclear periphery are correlated in simulations. The MSD analysis and its comparison to gene relocation could be repeated for the tracked loci by microscopy experiments.

My current calculation of enrichment score of a gene only takes the effect of neighbouring segments with the same chromatin state as the gene into account (Eq. 5.6, Eq. 5.7). However, the movement of an active gene could be inhibited by the slow movements of poised adjacent segments. For example, an S4 gene with S3 neighbours lacks enhancing neighbours and its movement upon heat-shock would be affected by the slow movement of S3 segments. Therefore, the calculation of enrichment score could be updated to include the impact of inhibiting neighbours. Then, the position of genes with the highest and lowest scores could be studied by our polymer model and microscopy experiments. To understand which kind of neighbour has a stronger effect on the movement, the relocation of genes with moderate scores, which have both types of neighbouring genes, could be analysed. Other factors, such as chromatin loops, that might also determine the repositioning of activated genes could be envisaged and tested by both simulations and experiments (microscopy and Hi-C).

7.10.3. Interaction networks and 3D chromosome conformation

The co-localisation of poised genes and the loose distribution of active genes are shown by our polymer model. In addition, the genome is predicted to be more compact during heat-shock. The clustering degree of genes and the compaction level of the genome could be further explored by calculating the radius of their semicircular distribution in 2D projections (Fig. 5.1, Fig. 5.2). Furthermore, to investigate how chromatin is folded at different levels, the Hi-C contact maps can

be analysed to build interaction networks of chromatin segments and to infer the 3D structure of chromosomes. A project student in our group, who was co-supervised by me, performed the network analysis for our 25°C and 37°C Hi-C data (Forquet 2016). Further inspection of these networks can identify chromatin regions with distinct structural features, such as small clusters of co-regulated active genes. In addition, interaction networks could be built for our simulation data and the results could be compared to the Hi-C data (Kidman 2016). This would indicate how the heterogeneous mobility of genes and the determined order of chromatin states affect the formation and the interaction of chromatin domains at various scales.

Another project student in our group tested different bioinformatics tools, such as MOGEN (Trieu & Cheng 2016) and TADbit (Serra et al. 2017), to build consensus 3D genome structures using our ensemble Hi-C data (Kidman 2016). This work could be extended by using new pipelines, like miniMDS (Rieber & Mahony 2017) and chromosome3D (Adhikari et al. 2016), and the results could be compared to the chromosome conformation in our polymer models. The analysis of interaction networks and 3D structure of chromosome require the development of new code and exploration of different comparison metrics.

7.10.4. Detailed comparison of 25°C and 37°C conditions

Poised chromatin-associated proteins display a concerted relocation across the genome upon changes in temperature. It has been shown that this relocation coincides with changes in the state-wise interactions between chromatin segments (Fig. 4.6) and the overall genome conformation (Fig. 5.1). However, we have not yet analysed our Hi-C datasets for more subtle changes in chromatin structure, and in gene-by-gene interactions. This could be addressed by identifying the similarities and differences of full contact maps, networks interactions, and 3D chromosome structures at 25°C and 37°C using our Hi-C data. For each type of comparison, the most suitable and efficient method should be selected among the many tools that are developed in the field. Then, an analogous analysis could be repeated for the results of our polymer model to reveal the effect of segment mobility and order of genes on structural changes of the genome when the temperature condition changes.

7.10.5. Model improvement

Our polymer model could be improved in several aspects:

Shorter simulation times could be achieved by parallelising the code. To this end, we recently collaborated with a group of computer scientists in Edinburgh, who ported my short test simulations to another physics engine. Running this code on Graphics Processing Units (GPUs), the Honors student achieved a two-fold speedup for 1000 segments, with a fairly flat relationship between runtime and number of segments (Decova 2018). Extrapolating to the 6000 segments of our whole genome simulations, the speedup could be as high as 12-fold. We would now port our main simulations to their system, to achieve a significant reduction in runtime from two months to five days.

The current version of simulation compares the position of all segments with each other to identify the interacting segments. This is the slowest part of the code, which could get better by implementing a virtual lattice to divide the nuclear space into small pieces and assess the position of segments in the same or adjacent pieces. A wider range of input parameters, such as scaling factors for the F_{LC}, could be tested with faster simulations.

Mobile and explicit protein species and their concerted movement upon changes in temperature can be introduced. The first version could have only poised protein molecules, however, later versions could incorporate various types of proteins, once more data is revealed about all recruited proteins and their biophysical properties at different states.

The properties of the rDNA and microtubule segments, the joints that connect chromatin segments to each other, and the attractive forces between telomeres and the nuclear membrane could be adjusted to achieve a more realistic representation of the yeast genome.

7.10.6. Statistical analysis improvement

There are common statistical tests to compare two data distributions, however there is not any well-known test that compares a whole set of distributions with another set of distributions. For the measured 3D distances between loci, we have eight distance distributions for each simulation type (Fig. 3.8). To compare each set of eight distributions as a whole with the entire eight distance distributions measured by live cell microscopy (not in a pair-wise manner), a two-step approach with some estimates was adopted. In the future, we could collaborate with statisticians to come up with other methods to compare the data with more accuracy and fewer approximations.

7.11. Conclusion

In this work, the dynamic 3D organisation of the budding yeast genome was studied using heteropolymer simulations, which were informed and verified by quantitative experimental data. The principal finding is that biologically meaningful 3D self-organisation can be achieved through differential mobility of differentially occupied chromatin segments. These results contribute a fundamentally new mechanism and complement recent advances in the field of genome organisation and biophysics. They provide the foundation for an emerging field, and studies that seek to understand how the different mechanisms influence each other.

References

Adhikari, B., Trieu, T. & Cheng, J., 2016. Chromosome3D: reconstructing three-dimensional chromosomal structures from Hi-C interaction frequency data using distance geometry simulated annealing. *BMC genomics*, 17(1), p.886.

Agrawal, A. et al., 2017. Chromatin as active matter. *Journal of statistical mechanics*, 2017(1), p.014001.

Ahmed, S. et al., 2010. DNA zip codes control an ancient mechanism for gene targeting to the nuclear periphery. *Nature cell biology*, 12(2), pp.111–118.

Alipour, E. & Marko, J.F., 2012. Self-organization of domain structures by DNA-loop-extruding enzymes. *Nucleic acids research*, 40(22), pp.11202–11212.

Allfrey, V.G., Faulkner, R. & Mirsky, A.E., 1964. ACETYLATION AND METHYLATION OF HISTONES AND THEIR POSSIBLE ROLE IN THE REGULATION OF RNA SYNTHESIS. *Proceedings of the National Academy of Sciences of the United States of America*, 51, pp.786–794.

Andrews, S.S. et al., 2010. Detailed simulations of cell biology with Smoldyn 2.1. *PLoS computational biology*, 6(3), p.e1000705.

Angel, A. et al., 2011. A Polycomb-based switch underlying quantitative epigenetic memory. *Nature*, 476(7358), pp.105–108.

von Appen, A. & Beck, M., 2016. Structure Determination of the Nuclear Pore Complex with Three-Dimensional Cryo electron Microscopy. *Journal of molecular biology*, 428(10 Pt A), pp.2001–2010.

Babu, M.M. et al., 2008. Eukaryotic gene regulation in three dimensions and its impact on genome evolution. *Current opinion in genetics & development*, 18(6), pp.571–582.

Baker, M., 2011. Making sense of chromatin states. *Nature methods*, 8(9), pp.717–722.

Barbieri, M. et al., 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(40), pp.16173–16178.

Barrera, L.O. et al., 2008. Genome-wide mapping and analysis of active promoters in mouse embryonic stem cells and adult organs. *Genome research*, 18(1), pp.46–59.

Bauer, S. et al., 2008. Ontologizer 2.0—a multifunctional tool for GO term enrichment analysis and data exploration. *Bioinformatics*, 24(14), pp.1650–1651.

Benedetti, F. et al., 2014. Models that include supercoiling of topological domains reproduce several known features of interphase chromosomes. *Nucleic acids research*, 42(5), pp.2848–2855.

Ben-Elazar, S., Yakhini, Z. & Yanai, I., 2013. Spatial localization of co-regulated genes exceeds genomic gene clustering in the Saccharomyces cerevisiae genome. *Nucleic acids research*, 41(4), pp.2191–2201.

Berg, H.C., 1993. *Random Walks in Biology*, Princeton University Press.

Berg, O.G. & von Hippel, P.H., 1985. Diffusion-controlled macromolecular interactions. *Annual review of biophysics and biophysical chemistry*, 14, pp.131–160.

Berg, O.G., Winter, R.B. & von Hippel, P.H., 1981. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry*, 20(24), pp.6929–6948.

Berg, O.G., Winter, R.B. & von Hippel, P.H., 1982. How do genome-regulatory proteins locate their DNA target sites? *Trends in biochemical sciences*, 7(2), pp.52–55.

Bernstein, B.E. et al., 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*, 125(2), pp.315–326.

Berry, S., Dean, C. & Howard, M., 2017. Slow Chromatin Dynamics Allow Polycomb Target Genes to Filter Fluctuations in Transcription Factor Activity. *Cell systems*, 4(4), pp.445–457.e8.

Bhattacherjee, A. & Levy, Y., 2014a. Search by proteins for their DNA target site: 1. The effect of DNA conformation on protein sliding. *Nucleic acids research*, 42(20), pp.12404–12414.

Bhattacherjee, A. & Levy, Y., 2014b. Search by proteins for their DNA target site: 2. The effect of DNA conformation on the dynamics of multidomain proteins. *Nucleic acids research*, 42(20), pp.12415–12424.

Bina, M., 2013. Gene regulation. *Methods in molecular biology*, 977, pp.1–11.

Boettiger, A.N. et al., 2016. Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature*, 529, p.418.

Boeynaems, S. et al., 2018. Protein Phase Separation: A New Phase in Cell Biology. *Trends in*

cell biology, 28(6), pp.420–435.

Bondos, S.E., Swint-Kruse, L. & Matthews, K.S., 2015. Flexibility and Disorder in Gene Regulation: LacI/GalR and Hox Proteins. *The Journal of biological chemistry*, 290(41), pp.24669–24677.

Bonev, B. & Cavalli, G., 2016. Organization and function of the 3D genome. *Nature reviews. Genetics*, 17(11), pp.661–678.

Brackley, C.A. et al., 2016. Predicting the three-dimensional folding of cis-regulatory regions in mammalian genomes using bioinformatic data and polymer models. *Genome biology*, 17(1), p.59.

Brackley, C.A., Cates, M.E. & Marenduzzo, D., 2013. Intracellular facilitated diffusion: searchers, crowders, and blockers. *Physical review letters*, 111(10), p.108101.

Brickner, D.G. et al., 2007. H2A.Z-mediated localization of genes at the nuclear periphery confers epigenetic memory of previous transcriptional state. *PLoS biology*, 5(4), p.e81.

Brickner, D.G. et al., 2012. Transcription factor binding to a DNA zip code controls interchromosomal clustering at the nuclear periphery. *Developmental cell*, 22(6), pp.1234–1246.

Brickner, J., 2017. Genetic and epigenetic control of the spatial organization of the genome. *Molecular biology of the cell*, 28(3), pp.364–369.

Brickner, J.H. & Walter, P., 2004. Gene recruitment of the activated INO1 locus to the nuclear membrane. *PLoS biology*, 2(11), p.e342.

Brown, J.M. et al., 2006. Coregulated human globin genes are frequently in spatial proximity when active. *The Journal of cell biology*, 172(2), pp.177–187.

Burke, K.A. et al., 2015. Residue-by-Residue View of In Vitro FUS Granules that Bind the C-Terminal Domain of RNA Polymerase II. *Molecular cell*, 60(2), pp.231–241.

Cabal, G.G. et al., 2006. SAGA interacting factors confine sub-diffusion of transcribed genes to the nuclear envelope. *Nature*, 441(7094), pp.770–773.

Callan, H.G. & Tomlin, S.G., 1950. Experimental studies on amphibian oocyte nuclei. I. Investigation of the structure of the nuclear membrane by means of the electron microscope. *Proceedings of the Royal Society of London. Series B, Containing papers of a Biological*

character. Royal Society, 137(888), pp.367–378.

Carrivain, P., Barbi, M. & Victor, J.-M., 2014. In silico single-molecule manipulation of DNA with rigid body dynamics. *PLoS computational biology*, 10(2), p.e1003456.

Casolari, J.M. et al., 2005. Developmentally induced changes in transcriptional program alter spatial organization across chromosomes. *Genes & development*, 19(10), pp.1188–1198.

Casolari, J.M. et al., 2004. Genome-wide localization of the nuclear transport machinery couples transcriptional status and nuclear organization. *Cell*, 117(4), pp.427–439.

Cavalli, G. & Misteli, T., 2013. Functional implications of genome topology. *Nature structural & molecular biology*, 20(3), pp.290–299.

Chambeyron, S. & Bickmore, W.A., 2004. Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. *Genes & development*, 18(10), pp.1119–1130.

Cheng, T.M.K. et al., 2015. A simple biophysical model emulates budding yeast chromosome condensation. *eLife*, 4, p.e05565.

Chiariello, A.M. et al., 2016. Polymer physics of chromosome large-scale 3D organisation. *Scientific reports*, 6, p.29775.

Choder, M. & Young, R.A., 1993. A portion of RNA polymerase II molecules has a component essential for stress responses and stress survival. *Molecular and cellular biology*, 13(11), pp.6984–6991.

Cohen, B.A. et al., 2000. A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nature genetics*, 26(2), pp.183–186.

Cooper, T.G., 1996. Regulation of Allantoin Catabolism in Saccharomyces cerevisiae. In R. Brambl & G. A. Marzluf, eds. *Biochemistry and Molecular Biology*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 139–169.

Cortini, R. & Filion, G.J., 2018. Theoretical principles of transcription factor traffic on folded chromatin. *Nature communications*, 9(1), p.1740.

Cremer, T. et al., 2006. Chromosome territories--a functional nuclear landscape. *Current opinion in cell biology*, 18(3), pp.307–316.

Cremer, T. et al., 1988. Detection of chromosome aberrations in metaphase and interphase tumor cells by in situ hybridization using chromosome-specific library probes. *Human genetics*, 80(3), pp.235–246.

Cremer, T. & Cremer, M., 2010. Chromosome territories. *Cold Spring Harbor perspectives in biology*, 2(3), p.a003889.

Creyghton, M.P. et al., 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America*, 107(50), pp.21931–21936.

Croft, J.A. et al., 1999. Differences in the localization and morphology of chromosomes in the human nucleus. *The Journal of cell biology*, 145(6), pp.1119–1131.

Cugliandolo, L.F., 2011. The effective temperature. *Journal of Physics A: Mathematical and Theoretical*, 44(48), p.483001.

Cui, Y. & Denis, C.L., 2003. In vivo evidence that defects in the transcriptional elongation factors RPB2, TFIIS, and SPT5 enhance upstream poly (A) site utilization. *Molecular and cellular biology*, 23(21), pp.7887–7901.

Dao, T.P. et al., 2018. Ubiquitin Modulates Liquid-Liquid Phase Separation of UBQLN2 via Disruption of Multivalent Interactions. *Molecular cell*, 69(6), pp.965–978.e6.

Das, P.M. et al., 2004. Chromatin immunoprecipitation assay. *BioTechniques*, 37(6), pp.961–969.

Das, R.K. & Kolomeisky, A.B., 2010. Facilitated search of proteins on DNA: correlations are important. *Physical chemistry chemical physics: PCCP*, 12(12), pp.2999–3004.

David-Rus, D. et al., 2009. Inheritance of epigenetic chromatin silencing. *Journal of theoretical biology*, 258(1), pp.112–120.

Deckert, J. & Struhl, K., 2001. Histone acetylation at promoters is differentially affected by specific activators and repressors. *Molecular and cellular biology*, 21(8), pp.2726–2735.

Decova, S., 2018. Efficient rigid-body simulations in systems biology. *Bachelor Honors project*, University of Edinburgh.

Dekker, J. et al., 2002. Capturing chromosome conformation. *Science*, 295(5558), pp.1306–1311.

Dekker, J., 2008. Mapping in vivo chromatin interactions in yeast suggests an extended chromatin

fiber with regional variation in compaction. *The Journal of biological chemistry*, 283(50), pp.34532–34540.

Dieppois, G., Iglesias, N. & Stutz, F., 2006. Cotranscriptional Recruitment to the mRNA Export Receptor Mex67p Contributes to Nuclear Pore Anchoring of Activated Genes. *Molecular and cellular biology*, 26(21), pp.7858–7870.

Dindot, S.V. & Cohen, N.D., 2013. Epigenetic Regulation of Gene Expression: Emerging Applications for Horses. *Journal of equine veterinary science*, 33(5), pp.288–294.

Dixon, J.R. et al., 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485(7398), pp.376–380.

Dixon, J.R., Gorkin, D.U. & Ren, B., 2016. Chromatin Domains: The Unit of Chromosome Organization. *Molecular cell*, 62(5), pp.668–680.

Dodd, I.B. et al., 2007. Theoretical analysis of epigenetic cell memory by nucleosome modification. *Cell*, 129(4), pp.813–822.

Dodd, I.B. & Sneppen, K., 2011. Barriers and silencers: a theoretical toolkit for control and containment of nucleosome-based epigenetic states. *Journal of molecular biology*, 414(4), pp.624–637.

Doyle, B. et al., 2014. Chromatin loops as allosteric modulators of enhancer-promoter interactions. *PLoS computational biology*, 10(10), p.e1003867.

Drubin, D.A., Garakani, A.M. & Silver, P.A., 2006. Motion as a phenotype: the use of live-cell imaging and machine visual screening to characterize transcription-dependent chromosome dynamics. *BMC cell biology*, 7(1), p.19.

Duan, Z. et al., 2010. A three-dimensional model of the yeast genome. *Nature*, 465(7296), pp.363–367.

Egecioglu, D. & Brickner, J.H., 2011. Gene positioning and expression. *Current opinion in cell biology*, 23(3), pp.338–345.

Engel, S.R. et al., 2014. The reference genome sequence of Saccharomyces cerevisiae: then and now. *G3*, 4(3), pp.389–398.

Erdel, F. & Greene, E.C., 2016. Generalized nucleation and looping model for epigenetic memory of histone modifications. *Proceedings of the National Academy of Sciences*, 113(29),

pp.E4180–E4189.

Ernst, J. et al., 2011. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, 473(7345), pp.43–49.

Ernst, J. & Kellis, M., 2010. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nature biotechnology*, 28(8), pp.817–825.

Fawcett, D.W., 1966. On the occurrence of a fibrous lamina on the inner aspect of the nuclear envelope in certain cells of vertebrates. *The American journal of anatomy*, 119(1), pp.129–145.

Ferrai, C. et al., 2010. Poised transcription factories prime silent uPA gene prior to activation. *PLoS biology*, 8(1), p.e1000270.

Filion, G.J. et al., 2010. Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. *Cell*, 143(2), pp.212–224.

Finlan, L.E. et al., 2008. Recruitment to the nuclear periphery can alter expression of genes in human cells. *PLoS genetics*, 4(3), p.e1000039.

Fiziev, P. et al., 2017. Systematic Epigenomic Analysis Reveals Chromatin States Associated with Melanoma Progression. *Cell reports*, 19(4), pp.875–889.

Flavahan, W.A. et al., 2016. Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature*, 529(7584), pp.110–114.

Forcato, M. et al., 2017. Comparison of computational methods for Hi-C data analysis. *Nature methods*, 14(7), pp.679–685.

Forquet, R., 2016. Chromatin state 3D organisation and dynamics in response to heat shock in Yeast: from Hi-C data to network analysis. *4th year internship report*, Polytech Nice-Sophia.

Foster, H.A. & Bridger, J.M., 2005. The genome and the nucleus: a marriage made by evolution. Genome organisation and nuclear architecture. *Chromosoma*, 114(4), pp.212–229.

Fraser, J. et al., 2015. Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation. *Molecular systems biology*, 11(12), p.852.

Fudenberg, G. et al., 2016. Formation of Chromosomal Domains by Loop Extrusion. *Cell reports*, 15(9), pp.2038–2049.

Furini, S., Barbini, P. & Domene, C., 2013. DNA-recognition process described by MD simulations of the lactose repressor protein on a specific and a non-specific DNA sequence. *Nucleic acids research*, 41(7), pp.3963–3972.

Fussner, E. et al., 2010. Changes in chromatin fiber density as a marker for pluripotency. *Cold Spring Harbor symposia on quantitative biology*, 75, pp.245–249.

Gall, J.G., 1964. Electron Microscopy of the Nuclear Envelope. In C. M. Feldherr et al., eds. *The Nuclear Membrane and Nucleocytoplasmic Interchange*. Vienna: Springer Vienna, pp. 4–25.

Gall, J.G., 1967. Octagonal nuclear pores. *The Journal of cell biology*, 32(2), pp.391–399.

Ganai, N., Sengupta, S. & Menon, G.I., 2014. Chromosome positioning from activity-based segregation. *Nucleic acids research*, 42(7), pp.4145–4159.

Giorgetti, L. et al., 2014. Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell*, 157(4), pp.950–963.

Goffeau, A. et al., 1996. Life with 6000 genes. *Science*, 274(5287), pp.546, 563–7.

Golkaram, M. et al., 2017. The Role of Chromatin Density in Cell Population Heterogeneity during Stem Cell Differentiation. *Scientific reports*, 7(1), p.13307.

Gómez-Díaz, E. & Corces, V.G., 2014. Architectural proteins: regulators of 3D genome organization in cell fate. *Trends in cell biology*, 24(11), pp.703–711.

Gonnella, G. et al., 2015. Motility-induced phase separation and coarsening in active matter. *Comptes Rendus Physique*, 16(3), pp.316–331.

González-Mariscal, L. et al., 2014. Tight junctions and the regulation of gene expression. *Seminars in cell & developmental biology*, 36, pp.213–223.

Gotta, M. et al., 1996. The clustering of telomeres and colocalization with Rap1, Sir3, and Sir4 proteins in wild-type Saccharomyces cerevisiae. *The Journal of cell biology*, 134(6), pp.1349–1363.

de Graaf, C.A. & van Steensel, B., 2013. Chromatin organization: form to function. *Current opinion in genetics & development*, 23(2), pp.185–190.

Grewal, S.I.S. & Moazed, D., 2003. Heterochromatin and epigenetic control of gene expression. *Science*, 301(5634), pp.798–802.

Grosberg, A. et al., 1993. Crumpled globule model of the three-dimensional structure of DNA. *EPL*, 23(5), p.373.

Grosberg, A.Y. & Joanny, J.-F., 2015. Nonequilibrium statistical mechanics of mixtures of particles in contact with different thermostats. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 92(3), p.032118.

Guet, D. et al., 2015. Combining Spinach-tagged RNA and gene localization to image gene expression in live yeast. *Nature communications*, 6, p.8882.

Guidi, M. et al., 2015. Spatial reorganization of telomeres in long-lived quiescent cells. *Genome biology*, 16, p.206.

Haddad, N., Jost, D. & Vaillant, C., 2017. Perspectives: using polymer modeling to understand the formation and function of nuclear compartments. *Chromosome research: an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology*, 25(1), pp.35–50.

Haeusler, R.A. et al., 2008. Clustering of yeast tRNA genes is mediated by specific association of condensin with tRNA gene transcription complexes. *Genes & development*, 22(16), pp.2204–2214.

Hahn, S., 2004. Structure and mechanism of the RNA polymerase II transcription machinery. *Nature structural & molecular biology*, 11(5), pp.394–403.

Hajjoul, H. et al., 2013. High-throughput chromatin motion tracking in living yeast reveals the flexibility of the fiber throughout the genome. *Genome research*, 23(11), pp.1829–1838.

Hameed, F.M., Rao, M. & Shivashankar, G.V., 2012. Dynamics of passive and active particles in the cell nucleus. *PloS one*, 7(10), p.e45843.

Han, J., Zhang, Z. & Wang, K., 2018. 3C and 3C-based techniques: the powerful tools for spatial genome organization deciphering. *Molecular cytogenetics*, 11, p.21.

Hathaway, N.A. et al., 2012. Dynamics and memory of heterochromatin in living cells. *Cell*, 149(7), pp.1447–1460.

Hediger, F. & Gasser, S.M., 2006. Heterochromatin protein 1: don't judge the book by its cover! *Current opinion in genetics & development*, 16(2), pp.143–150.

Heinz, S. et al., 2010. Simple combinations of lineage-determining transcription factors prime cis-

regulatory elements required for macrophage and B cell identities. *Molecular cell*, 38(4), pp.576–589.

Heun, P., Taddei, A. & Gasser, S.M., 2001. From snapshots to moving pictures: new perspectives on nuclear organization. *Trends in cell biology*, 11(12), pp.519–525.

van Heusden, G.P. & Steensma, H.Y., 2006. Yeast 14-3-3 proteins. *Yeast*, 23(3), pp.159–171.

Hittinger, C.T., Rokas, A. & Carroll, S.B., 2004. Parallel inactivation of multiple GAL pathway genes and ecological diversification in yeasts. *Proceedings of the National Academy of Sciences of the United States of America*, 101(39), pp.14144–14149.

Hnisz, D. et al., 2017. A Phase Separation Model for Transcriptional Control. *Cell*, 169(1), pp.13–23.

Hodges, C. & Crabtree, G.R., 2012. Dynamics of inherently bounded histone modification domains. *Proceedings of the National Academy of Sciences of the United States of America*, 109(33), pp.13296–13301.

Hofmann, A. & Heermann, D.W., 2015. The role of loops on the order of eukaryotes and prokaryotes. *FEBS letters*, 589(20, Part A), pp.2958–2965.

van Holde, K.E., 1989. *Chromatin*, Springer.

Holstege, F.C. et al., 1998. Dissecting the regulatory circuitry of a eukaryotic genome. *Cell*, 95(5), pp.717–728.

Humphrey, W., Dalke, A. & Schulten, K., 1996. VMD: visual molecular dynamics. *Journal of molecular graphics*, 14(1), pp.33–8, 27–8.

Hurst, L.D., Pál, C. & Lercher, M.J., 2004. The evolutionary dynamics of eukaryotic gene order. *Nature reviews. Genetics*, 5(4), pp.299–310.

Imakaev, M. et al., 2012. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nature methods*, 9(10), pp.999–1003.

Imakaev, M.V., Fudenberg, G. & Mirny, L.A., 2015. Modeling chromosomes: Beyond pretty pictures. *FEBS letters*, 589(20 Pt A), pp.3031–3036.

Iwahara, J. & Levy, Y., 2013. Speed-stability paradox in DNA-scanning by zinc-finger proteins. *Transcription*, 4(2), pp.58–61.

Jaenisch, R. & Bird, A., 2003. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature genetics*, 33 Suppl, pp.245–254.

Janga, S.C., Collado-Vides, J. & Babu, M.M., 2008. Transcriptional regulation constrains the organization of genes on eukaryotic chromosomes. *Proceedings of the National Academy of Sciences of the United States of America*, 105(41), pp.15761–15766.

Javierre, B.M. et al., 2016. Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell*, 167(5), pp.1369–1384.

Jensen, G.J. et al., 1998. Structure of wild-type yeast RNA polymerase II and location of Rpb4 and Rpb7. *The EMBO journal*, 17(8), pp.2353–2358.

Jerabek, H. & Heermann, D.W., 2012. Expression-dependent folding of interphase chromatin. *PloS one*, 7(5), p.e37525.

Jin, Q. et al., 1998. Yeast nuclei display prominent centromere clustering that is reduced in nondividing cells and in meiotic prophase. *The Journal of cell biology*, 141(1), pp.21–29.

Jin, Q.W., Fuchs, J. & Loidl, J., 2000. Centromere clustering is a major determinant of yeast interphase nuclear organization. *Journal of cell science*, 113 (Pt 11), pp.1903–1912.

Jirgensons, B., 1958. Optical rotation and viscosity of native and denatured proteins. X. Further studies on optical rotatory dispersion. *Archives of biochemistry and biophysics*, 74(1), pp.57–69.

Jost, D. et al., 2014. Modeling epigenome folding: formation and dynamics of topologically associated chromatin domains. *Nucleic acids research*, 42(15), pp.9553–9561.

Jost, D. & Vaillant, C., 2018. Epigenomics in 3D: importance of long-range spreading and specific interactions in epigenomic maintenance. *Nucleic acids research*, 46(5), pp.2252–2264.

Kasowski, M. et al., 2013. Extensive variation in chromatin states across humans. *Science*, 342(6159), pp.750–752.

Képès, F., 2003. Periodic epi-organization of the yeast genome revealed by the distribution of promoter sites. *Journal of molecular biology*, 329(5), pp.859–865.

Kharchenko, P.V. et al., 2010. Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. *Nature*, 471, p.480.

Kidman, S., 2016. Investigation of programmes able to model the 3D structure of the *S. cerevisiae* genome from Hi-C data. *BBSRC Doctoral Training Partnership Rotation Project*, Babraham Institute.

Kimura, M., Ishiguro, A. & Ishihama, A., 1997. RNA polymerase II subunits 2, 3, and 11 form a core subassembly with DNA binding activity. *The Journal of biological chemistry*, 272(41), pp.25851–25855.

Kind, J. et al., 2015. Genome-wide maps of nuclear lamina interactions in single human cells. *Cell*, 163(1), pp.134–147.

Kind, J. et al., 2013. Single-Cell Dynamics of Genome-Nuclear Lamina Interactions. *Cell*, 153(1), pp.178–192.

Koslover, E.F. & Spakowitz, A.J., 2014. Multiscale dynamics of semiflexible polymers from a universal coarse-graining procedure. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 90(1), p.013304.

Kouzine, F. et al., 2008. The functional response of upstream DNA to dynamic supercoiling in vivo. *Nature structural & molecular biology*, 15(2), pp.146–154.

Kouzine, F. et al., 2013. Transcription-dependent dynamic supercoiling is a short-range genomic force. *Nature structural & molecular biology*, 20(3), pp.396–403.

Kubo, N. et al., 2017. Preservation of Chromatin Organization after Acute Loss of CTCF in Mouse Embryonic Stem Cells. *bioRxiv*, p.118737. Available at: https://www.biorxiv.org/content/early/2017/03/20/118737.abstract [Accessed May 9, 2018].

Kuo, M.H. & Allis, C.D., 1999. In vivo cross-linking and immunoprecipitation for studying dynamic Protein:DNA associations in a chromatin environment. *Methods*, 19(3), pp.425–433.

Küpper, K. et al., 2007. Radial chromatin positioning is shaped by local gene density, not by gene expression. *Chromosoma*, 116(3), pp.285–306.

Ku, W.L. et al., 2013. Modeling the dynamics of bivalent histone modifications. *PloS one*, 8(11), p.e77944.

de Laat, W. & Dekker, J., 2012. 3C-based technologies to study the shape of the genome. *Methods*, 58(3), pp.189–191.

Labrador, M. & Corces, V.G., 2002. Setting the boundaries of chromatin domains and nuclear

organization. *Cell*, 111(2), pp.151–154.

Lajoie, B.R., Dekker, J. & Kaplan, N., 2015. The Hitchhiker's guide to Hi-C analysis: practical guidelines. *Methods*, 72, pp.65–75.

Lakadamyali, M. & Cosma, M.P., 2015. Advanced microscopy methods for visualizing chromatin structure. *FEBS letters*, 589(20 Pt A), pp.3023–3030.

Lanctôt, C. et al., 2007. Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nature reviews. Genetics*, 8(2), pp.104–115.

Larson, J.L. & Yuan, G.-C., 2012. Chromatin states accurately classify cell differentiation stages. *PloS one*, 7(2), p.e31414.

Lassadi, I. & Bystricky, K., 2011. Tracking of single and multiple genomic loci in living yeast cells. *Methods in molecular biology*, 745, pp.499–522.

Lazaris, C. et al., 2017. HiC-bench: comprehensive and reproducible Hi-C data analysis designed for parameter exploration and benchmarking. *BMC genomics*, 18(1), p.22.

Lazar-Stefanita, L. et al., 2017. Cohesins and condensins orchestrate the 4D dynamics of yeast chromosomes during the cell cycle. *The EMBO journal*. Available at: http://dx.doi.org/10.15252/embj.201797342.

Lee, J.M. & Sonnhammer, E.L.L., 2003. Genomic gene clustering analysis of pathways in eukaryotes. *Genome research*, 13(5), pp.875–882.

Léger-Silvestre, I. et al., 1999. Functional compartmentalization of the nucleus in the budding yeast Saccharomyces cerevisiae. *Chromosoma*, 108(2), pp.103–113.

Lemons, D.S. & Gythiel, A., 1997. Paul Langevin's 1908 paper "On the Theory of Brownian Motion" ["Sur la théorie du mouvement brownien," C. R. Acad. Sci. (Paris) 146, 530–533 (1908)]. *American journal of physics*, 65(11), pp.1079–1081.

Le, T.B.K. et al., 2013. High-resolution mapping of the spatial organization of a bacterial chromosome. *Science*, 342(6159), pp.731–734.

Levine, M. & Tjian, R., 2003. Transcription regulation and animal diversity. *Nature*, 424(6945), pp.147–151.

Liang, G. et al., 2004. Distinct localization of histone H3 acetylation and H3-K4 methylation to the

transcription start sites in the human genome. *Proceedings of the National Academy of Sciences of the United States of America*, 101(19), pp.7357–7362.

Lichter, P. et al., 1988. Delineation of individual human chromosomes in metaphase and interphase cells by in situ suppression hybridization using recombinant DNA libraries. *Human genetics*, 80(3), pp.224–234.

Lieberman-Aiden, E. et al., 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326(5950), pp.289–293.

Liebman, P.A. & Entine, G., 1974. Lateral diffusion of visual pigment in photorecptor disk membranes. *Science*, 185(4149), pp.457–459.

Liu, C.L. et al., 2005. Single-nucleosome mapping of histone modifications in S. cerevisiae. *PLoS biology*, 3(10), p.e328.

Liu, L., Cherstvy, A.G. & Metzler, R., 2017. Facilitated Diffusion of Transcription Factor Proteins with Anomalous Bulk Diffusion. *The journal of physical chemistry. B*, 121(6), pp.1284–1289.

Liu, T. et al., 2011. Broad chromosomal domains of histone modification patterns in C. elegans. *Genome research*, 21(2), pp.227–236.

Loi, D., Mossa, S. & Cugliandolo, L.F., 2008. Effective temperature of active matter. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 77(5 Pt 1), p.051111.

Long, H.K., Prescott, S.L. & Wysocka, J., 2016. Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell*, 167(5), pp.1170–1187.

Lowary, P.T. & Widom, J., 1989. Higher-order structure of Saccharomyces cerevisiae chromatin. *Proceedings of the National Academy of Sciences of the United States of America*, 86(21), pp.8266–8270.

Lupiáñez, D.G. et al., 2015. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*, 161(5), pp.1012–1025.

Lüscher, B. & Vervoorts, J., 2012. Regulation of gene transcription by the oncoprotein MYC. *Gene*, 494(2), pp.145–160.

Machida, S. et al., 2018. Structural Basis of Heterochromatin Formation by Human HP1. *Molecular cell*, 69(3), pp.385–397.e8.

Mahmutovic, A., Berg, O.G. & Elf, J., 2015. What matters for lac repressor search in vivo--sliding, hopping, intersegment transfer, crowding on DNA or recognition? *Nucleic acids research*, 43(7), pp.3454–3464.

Mahy, N.L., Perry, P.E., Gilchrist, S., et al., 2002. Spatial organization of active and inactive genes and noncoding DNA within chromosome territories. *The Journal of cell biology*, 157(4), pp.579–589.

Mahy, N.L., Perry, P.E. & Bickmore, W.A., 2002. Gene density and transcription influence the localization of chromatin outside of chromosome territories detectable by FISH. *The Journal of cell biology*, 159(5), pp.753–763.

Marklund, E.G. et al., 2013. Transcription-factor binding and sliding on DNA studied using micro- and macroscopic models. *Proceedings of the National Academy of Sciences of the United States of America*, 110(49), pp.19796–19801.

McCandlish, S.R., Baskaran, A. & Hagan, M.F., 2012. Spontaneous segregation of self-propelled particles with different motilities. *Soft matter*, 8(8), pp.2527–2534.

Meaburn, K.J., Misteli, T. & Soutoglou, E., 2007. Spatial genome organization in the formation of chromosomal translocations. *Seminars in cancer biology*, 17(1), pp.80–90.

Menon, G.I., 2010. Active Matter. In J. M. Krishnan, A. P. Deshpande, & P. B. S. Kumar, eds. *Rheology of Complex Fluids*. New York, NY: Springer New York, pp. 193–218.

Merkle, E.C., You, D. & Preacher, K.J., 2014. Testing non-nested structural equation models. *arXiv [stat.AP]*. Available at: http://arxiv.org/abs/1402.6720.

Michalak, P., 2008. Coexpression, coregulation, and cofunctionality of neighboring genes in eukaryotic genomes. *Genomics*, 91(3), pp.243–248.

Micheelsen, M.A. et al., 2010. Theory for the stability and regulation of epigenetic landscapes. *Physical biology*, 7(2), p.026010.

Michieletto, D., Orlandini, E. & Marenduzzo, D., 2016. Polymer model with Epigenetic Recoloring Reveals a Pathway for the de novo Establishment and 3D Organization of Chromatin Domains. *Physical Review X*, 6(4), p.041047.

Miele, A., Bystricky, K. & Dekker, J., 2009. Yeast silent mating type loci form heterochromatic clusters through silencer protein-dependent long-range interactions. *PLoS genetics*, 5(5),

p.e1000478.

Mifsud, B. et al., 2015. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nature genetics*, 47(6), pp.598–606.

Mirny, L.A., 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(1), pp.37–51.

Misteli, T., 2007. Beyond the sequence: cellular organization of genome function. *Cell*, 128(4), pp.787–800.

Misteli, T., 2001. Protein dynamics: implications for nuclear architecture and gene expression. *Science*, 291(5505), pp.843–847.

Mitchell, J.A. & Fraser, P., 2008. Transcription factories are nuclear subcompartments that remain in the absence of transcription. *Genes & development*, 22(1), pp.20–25.

modENCODE Consortium et al., 2010. Identification of functional elements and regulatory circuits by Drosophila modENCODE. *Science*, 330(6012), pp.1787–1797.

Mueller, F., Wach, P. & McNally, J.G., 2008. Evidence for a common mode of transcription factor interaction with chromatin as revealed by improved quantitative fluorescence recovery after photobleaching. *Biophysical journal*, 94(8), pp.3323–3339.

Mukhopadhyay, S., Nagaraj, V.H. & Sengupta, A.M., 2010. Locus dependence in epigenetic chromatin silencing. *Bio Systems*, 102(1), pp.49–54.

Müller-Ott, K. et al., 2014. Specificity, propagation, and memory of pericentric heterochromatin. *Molecular systems biology*, 10, p.746.

Murtey, M.D. & Ramasamy, P., 2016. Sample Preparations for Scanning Electron Microscopy--Life Sciences. In *Modern Electron Microscopy in Physical and Life Sciences*. InTech.

Nagano, T. et al., 2013. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature*, 502(7469), pp.59–64.

Nakayama, J. et al., 2001. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science*, 292(5514), pp.110–113.

Nasmyth, K., 2001. Disseminating the genome: joining, resolving, and separating sister

chromatids during mitosis and meiosis. *Annual review of genetics*, 35, pp.673–745.

Naughton, C. et al., 2013. Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures. *Nature structural & molecular biology*, 20(3), pp.387–395.

Nazarov, L.I. et al., 2015. A statistical model of intra-chromosome contact maps. *Soft matter*, 11(5), pp.1019–1025.

Nora, E.P. et al., 2012. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, 485(7398), pp.381–385.

Nora, E.P. et al., 2017. Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell*, 169(5), pp.930–944.e22.

Olarte-Plata, J.D. et al., 2016. The folding landscape of the epigenome. *Physical biology*, 13(2), p.026001.

Oldfield, C.J. & Dunker, A.K., 2014. Intrinsically disordered proteins and intrinsically disordered protein regions. *Annual review of biochemistry*, 83, pp.553–584.

Olins, A.L. & Olins, D.E., 1974. Spheroid chromatin units (v bodies). *Science*, 183(4122), pp.330–332.

Osborne, C.S. et al., 2007. Myc dynamically and preferentially relocates to a transcription factory occupied by Igh. *PLoS biology*, 5(8), p.e192.

Osborne, C.S. & Mifsud, B., 2017. Capturing genomic relationships that matter. *Chromosome research: an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology*, 25(1), pp.15–24.

Osbourn, A.E. & Field, B., 2009. Operons. *Cellular and molecular life sciences: CMLS*, 66(23), pp.3755–3775.

Oudet, P., Gross-Bellard, M. & Chambon, P., 1975. Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. *Cell*, 4(4), pp.281–300.

Papantonis, A. & Cook, P.R., 2011. Fixing the model for transcription: the DNA moves, not the polymerase. *Transcription*, 2(1), pp.41–44.

Parsell, D.A. et al., 1994. Protein disaggregation mediated by heat-shock protein Hsp104. *Nature*, 372(6505), pp.475–478.

Paulsen, J., Liyakat Ali, T.M. & Collas, P., 2018. Computational 3D genome modeling using Chrom3D. *Nature protocols*, 13(5), pp.1137–1152.

Pederson, T., 2011. The nucleus introduced. *Cold Spring Harbor perspectives in biology*, 3(5). Available at: http://dx.doi.org/10.1101/cshperspect.a000521.

Petrascheck, M. et al., 2005. DNA looping induced by a transcriptional enhancer in vivo. *Nucleic acids research*, 33(12), pp.3743–3750.

Phair, R.D. & Misteli, T., 2000. High mobility of proteins in the mammalian cell nucleus. *Nature*, 404(6778), pp.604–609.

Phillip, Y. & Schreiber, G., 2013. Formation of protein complexes in crowded environments--from in vitro to in vivo. *FEBS letters*, 587(8), pp.1046–1052.

Pinkel, D. et al., 1988. Fluorescence in situ hybridization with human chromosome-specific libraries: detection of trisomy 21 and translocations of chromosome 4. *Proceedings of the National Academy of Sciences of the United States of America*, 85(23), pp.9138–9142.

Ptashne, M. & Gann, A., 2002. *Genes and Signals*, Cold Spring Harbor Laboratory Press.

Racko, D. et al., 2017. Transcription-induced supercoiling as the driving force of chromatin loop extrusion during formation of TADs in interphase chromosomes. *Nucleic acids research*. Available at: http://dx.doi.org/10.1093/nar/gkx1123.

Ramakrishnan, V., 1997. Histone structure and the organization of the nucleosome. *Annual review of biophysics and biomolecular structure*, 26, pp.83–112.

Ram, O. et al., 2011. Combinatorial patterning of chromatin regulators uncovered by genome-wide location analysis in human cells. *Cell*, 147(7), pp.1628–1639.

Randise-Hinchliff, C. et al., 2016. Strategies to regulate transcription factor-mediated gene positioning and interchromosomal clustering at the nuclear periphery. *The Journal of cell biology*, 212(6), pp.633–646.

Randise-Hinchliff, C. & Brickner, J.H., 2016. Transcription factors dynamically control the spatial organization of the yeast genome. *Nucleus*, 7(4), pp.369–374.

Reddy, K.L. et al., 2008. Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature*, 452(7184), pp.243–247.

Riddle, N.C. et al., 2011. Plasticity in patterns of histone modifications and chromosomal proteins in Drosophila heterochromatin. *Genome research*, 21(2), pp.147–163.

Rieber, L. & Mahony, S., 2017. miniMDS: 3D structural inference from high-resolution Hi-C data. *Bioinformatics*, 33(14), pp.i261–i266.

Rieder, D., Trajanoski, Z. & McNally, J.G., 2012. Transcription factories. *Frontiers in genetics*, 3, p.221.

Robinett, C.C. et al., 1996. In vivo localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition. *The Journal of cell biology*, 135(6 Pt 2), pp.1685–1700.

Rodley, C.D. et al., 2009. Global identification of yeast chromosome interactions using Genome conformation capture. *Fungal genetics and biology: FG & B*, 46(11), pp.879–886.

Rohlf, T. et al., 2012. Modeling the dynamic epigenome: from histone modifications towards self-organizing chromatin. *Epigenomics*, 4(2), pp.205–219.

Rosanova, A. et al., 2017. Modelling the evolution of transcription factor binding preferences in complex eukaryotes. *Scientific reports*, 7(1), p.7596.

Sajan, S.A. & Hawkins, R.D., 2012. Methods for identifying higher-order chromatin structure. *Annual review of genomics and human genetics*, 13, pp.59–82.

Sanborn, A.L. 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*, 112(47), pp.E6456–65.

Sanchez, Y. et al., 1992. Hsp104 is required for tolerance to many forms of stress. *The EMBO journal*, 11(6), pp.2357–2364.

Sarma, N.J. et al., 2007. Glucose-responsive regulators of gene expression in Saccharomyces cerevisiae function at the nuclear periphery via a reverse recruitment mechanism. *Genetics*, 175(3), pp.1127–1135.

Schalch, T. et al., 2005. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature*, 436(7047), pp.138–141.

Schmid, M. et al., 2006. Nup-PI: the nucleopore-promoter interaction of genes in yeast. *Molecular cell*, 21(3), pp.379–391.

Schmidt, H.G. et al., 2014. An Integrated Model of Transcription Factor Diffusion Shows the Importance of Intersegmental Transfer and Quaternary Protein Structure for Target Site Finding. *PloS one*, 9(10), p.e108575.

Schmitt, A.D., Hu, M., Jung, I., et al., 2016. A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. *Cell reports*, 17(8), pp.2042–2059.

Schmitt, A.D., Hu, M. & Ren, B., 2016. Genome-wide mapping and analysis of chromosome architecture. *Nature reviews. Molecular cell biology*, 17(12), pp.743–755.

Schübeler, D., 2015. Function and information content of DNA methylation. *Nature*, 517(7534), pp.321–326.

Schübeler, D. et al., 2004. The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. *Genes & development*, 18(11), pp.1263–1271.

Sedighi, M. & Sengupta, A.M., 2007. Epigenetic chromatin silencing: bistability and front propagation. *Physical biology*, 4(4), pp.246–255.

Serra, F. et al., 2017. Automatic analysis and 3D-modelling of Hi-C data using TADbit reveals structural features of the fly chromatin colors. *PLoS computational biology*, 13(7), p.e1005665.

Servant, N. et al., 2015. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome biology*, 16, p.259.

Sewitz, S.A. et al., 2017a. Heterogeneous chromatin mobility derived from chromatin states is a determinant of genome organisation in S. cerevisiae. *bioRxiv*, p.106344. Available at: https://www.biorxiv.org/content/early/2017/03/02/106344 [Accessed November 1, 2017].

Sewitz, S., Fahmi, Z. & Lipkow, K., 2017b. Higher order assembly: Folding the Chromosome. *Current opinion in structural biology*, 42, pp.162–168.

Sewitz, S. & Lipkow, K., 2016. Systems Biology Approaches for Understanding Genome Architecture. *Methods in molecular biology*, 1431, pp.109–126.

Sexton, T. et al., 2012. Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell*, 148(3), pp.458–472.

Shachar, S. & Misteli, T., 2017. Causes and consequences of nuclear gene positioning. *Journal*

of cell science, 130(9), pp.1501–1508.

Shankaranarayana, G.D. et al., 2003. Sir2 regulates histone H3 lysine 9 methylation and heterochromatin assembly in fission yeast. *Current biology: CB*, 13(14), pp.1240–1246.

Shavit, Y. et al., 2016. How computer science can help in understanding the 3D genome architecture. *Briefings in bioinformatics*, 17(5), pp.733–744.

Shukron, O. & Holcman, D., 2017. Transient chromatin properties revealed by polymer models and stochastic simulations constructed from Chromosomal Capture data. *PLoS computational biology*, 13(4), p.e1005469.

Smolle, M. & Workman, J.L., 2013. Transcription-associated histone modifications and cryptic transcription. *Biochimica et biophysica acta*, 1829(1), pp.84–97.

Smrek, J. & Kremer, K., 2017. Small activity differences drive phase separation in active-passive polymer mixtures. *arXiv [cond-mat.soft]*. Available at: http://arxiv.org/abs/1701.07362.

Sparmann, A. & van Lohuizen, M., 2006. Polycomb silencers control cell fate, development and cancer. *Nature reviews. Cancer*, 6, p.846.

Stasevich, T.J. et al., 2010. Dissecting the binding mechanism of the linker histone in live cells: an integrated FRAP analysis. *The EMBO journal*, 29(7), pp.1225–1234.

van Steensel, B. & Belmont, A.S., 2017. Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell*, 169(5), pp.780–791.

Steglich, B., Sazer, S. & Ekwall, K., 2013. Transcriptional regulation at the yeast nuclear envelope. *Nucleus*, 4(5), pp.379–389.

Stenhammar, J. et al., 2015. Activity-induced phase separation and self-assembly in mixtures of active and passive particles. *Physical review letters*, 114(1), p.018301.

Stock, J.K. et al., 2007. Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nature cell biology*, 9(12), pp.1428–1435.

Strahl, B.D. & Allis, C.D., 2000. The language of covalent histone modifications. *Nature*, 403(6765), pp.41–45.

Straight, A.F. et al., 1996. GFP tagging of budding yeast chromosomes reveals that protein-protein interactions can mediate sister chromatid cohesion. *Current biology: CB*, 6(12), pp.1599–

1608.

Sutherland, H. & Bickmore, W.A., 2009. Transcription factories: gene expression in unions? *Nature reviews. Genetics*, 10(7), pp.457–466.

Tabaka, M., Kalwarczyk, T. & Hołyst, R., 2014. Quantitative influence of macromolecular crowding on gene regulation kinetics. *Nucleic acids research*, 42(2), pp.727–738.

Taddei, A. et al., 2006. Nuclear pore association confers optimal expression levels for an inducible yeast gene. *Nature*, 441(7094), pp.774–778.

Taddei, A. & Gasser, S.M., 2012. Structure and function in the budding yeast nucleus. *Genetics*, 192(1), pp.107–129.

Talwar, S. et al., 2013. Correlated spatio-temporal fluctuations in chromatin compaction states characterize stem cells. *Biophysical journal*, 104(3), pp.553–564.

Tark-Dame, M. et al., 2014. Depletion of the Chromatin Looping Proteins CTCF and Cohesin Causes Chromatin Compaction: Insight into Chromatin Folding by Polymer Modelling. *PLoS computational biology*, 10(10), p.e1003877.

Tark-Dame, M., van Driel, R. & Heermann, D.W., 2011. Chromatin folding--from biology to polymer models and back. *Journal of cell science*, 124(Pt 6), pp.839–845.

Tee, W.-W. et al., 2014. Erk1/2 activity promotes chromatin features and RNAPII phosphorylation at developmental promoters in mouse ESCs. *Cell*, 156(4), pp.678–690.

Teichmann, S.A. & Veitia, R.A., 2004. Genes encoding subunits of stable complexes are clustered on the yeast chromosomes: an interpretation from a dosage balance perspective. *Genetics*, 167(4), pp.2121–2125.

Terakawa, T. et al., 2017. The condensin complex is a mechanochemical motor that translocates along DNA. *Science*, 358(6363), pp.672–676.

Therizols, P. et al., 2010. Chromosome arm length and nuclear constraints determine the dynamic relationship of yeast subtelomeres. *Proceedings of the National Academy of Sciences of the United States of America*, 107(5), pp.2025–2030.

Thompson, M. et al., 2003. Nucleolar clustering of dispersed tRNA genes. *Science*, 302(5649), pp.1399–1401.

Tiana, G. et al., 2016. Structural Fluctuations of the Chromatin Fiber within Topologically Associating Domains. *Biophysical journal*, 110(6), pp.1234–1245.

Torella, R. et al., 2014. A combination of computational and experimental approaches identifies DNA sequence constraints associated with target site binding specificity of the transcription factor CSL. *Nucleic acids research*, 42(16), pp.10550–10563.

Trelles-Sticken, E., Dresser, M.E. & Scherthan, H., 2000. Meiotic telomere protein Ndj1p is required for meiosis-specific telomere distribution, bouquet formation and efficient homologue pairing. *The Journal of cell biology*, 151(1), pp.95–106.

Trieu, T. & Cheng, J., 2016. MOGEN: a tool for reconstructing 3D models of genomes from chromosomal conformation capturing data. *Bioinformatics*, 32(9), pp.1286–1292.

Ulianov, S.V. et al., 2016. Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. *Genome research*, 26(1), pp.70–84.

Ulianov, S.V., Gavrilov, A.A. & Razin, S.V., 2015. Nuclear compartments, genome folding, and enhancer-promoter communication. *International review of cell and molecular biology*, 315, pp.183–244.

Uusküla-Reimand, L. et al., 2016. Topoisomerase II beta interacts with cohesin and CTCF at topological domain borders. *Genome biology*, 17(1), p.182.

Vazquez, J., Belmont, A.S. & Sedat, J.W., 2001. Multiple regimes of constrained chromosome motion are regulated in the interphase Drosophila nucleus. *Current biology: CB*, 11(16), pp.1227–1239.

Venters, B.J. et al., 2011. A comprehensive genomic binding map of gene and chromatin regulatory proteins in Saccharomyces. *Molecular cell*, 41(4), pp.480–492.

Vinciguerra, P. & Stutz, F., 2004. mRNA export: an assembly line from genes to nuclear pores. *Current opinion in cell biology*, 16(3), pp.285–292.

Volpi, E.V. et al., 2000. Large-scale chromatin organization of the major histocompatibility complex and other regions of human chromosome 6 and its response to interferon in interphase nuclei. *Journal of cell science*, 113 (Pt 9), pp.1565–1576.

Vuong, Q.H., 1989. Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica: journal of the Econometric Society*, 57(2), pp.307–333.

Wang, Y. et al., 2014. Super-resolution microscopy reveals decondensed chromatin structure at transcription sites. *Scientific reports*, 4, p.4477.

Weber, S.C., Spakowitz, A.J. & 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(19), pp.7338–7343.

Weber, S.N., Weber, C.A. & Frey, E., 2016. Binary Mixtures of Particles with Different Diffusivities Demix. *Physical review letters*, 116(5), p.058301.

West, A.G., Gaszner, M. & Felsenfeld, G., 2002. Insulators: many functions, many mechanisms. *Genes & development*, 16(3), pp.271–288.

Williams, R.R.E. et al., 2002. Subchromosomal positioning of the epidermal differentiation complex (EDC) in keratinocyte and lymphoblast interphase nuclei. *Experimental cell research*, 272(2), pp.163–175.

Wilson, M.D. & Costa, A., 2017. Cryo-electron microscopy of chromatin biology. *Acta crystallographica. Section D, Structural biology*, 73(Pt 6), pp.541–548.

Wingett, S. et al., 2015. HiCUP: pipeline for mapping and processing Hi-C data. *F1000Research*, 4, p.1310.

Wollman, A.J. et al., 2017. Transcription factor clusters regulate genes in eukaryotic cells. *eLife*, 6. Available at: http://dx.doi.org/10.7554/eLife.27451.

Wong, H. et al., 2012. A predictive computational model of the dynamic 3D interphase yeast nucleus. *Current biology: CB*, 22(20), pp.1881–1890.

Wu, H.Y. et al., 1988. Transcription generates positively and negatively supercoiled domains in the template. *Cell*, 53(3), pp.433–440.

Wu, J.Q. & Snyder, M., 2008. RNA polymerase II stalling: loading at the start prepares genes for a sprint. *Genome biology*, 9(5), p.220.

Xia, K., 2018. Sequence-based multiscale modeling for high-throughput chromosome conformation capture (Hi-C) data analysis. *PloS one*, 13(2), p.e0191899.

Yaffe, E. & Tanay, A., 2011. Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. *Nature genetics*, 43(11), pp.1059–1065.

Yi, G., Sze, S.-H. & Thon, M.R., 2007. Identifying clusters of functionally related genes in genomes. *Bioinformatics*, 23(9), pp.1053–1060.

Yuen, K.C. & Gerton, J.L., 2018. Taking cohesin and condensin in context. *PLoS genetics*, 14(1), p.e1007118.

Yu. Grosberg, A., Nechaev, S.K. & Shakhnovich, E.I., 1988. The role of topological constraints in the kinetics of collapse of macromolecules. *Journal de Physique*, 49(12), pp.2095–2100.

Zabet, N.R. & Adryan, B., 2012a. A comprehensive computational model of facilitated diffusion in prokaryotes. *Bioinformatics*, 28(11), pp.1517–1524.

Zabet, N.R. & Adryan, B., 2012b. GRiP: a computational tool to simulate transcription factor binding in prokaryotes. *Bioinformatics*, 28(9), pp.1287–1289.

Zabet, N.R. & Adryan, B., 2013. The effects of transcription factor competition on gene regulation. *Frontiers in genetics*, 4, p.197.

Zanton, S.J. & Pugh, B.F., 2004. Changes in genomewide occupancy of core transcriptional regulators during heat stress. *Proceedings of the National Academy of Sciences of the United States of America*, 101(48), pp.16843–16848.

Zhang, T., Cooper, S. & Brockdorff, N., 2015. The interplay of histone modifications - writers that read. *EMBO reports*, 16(11), pp.1467–1481.

Zhang, X. & Smith, T.F., 1998. Yeast “operons”. *Microbial & comparative genomics*, 3(2), pp.133–140.

Zimmer, C. & Fabre, E., 2011. Principles of chromosomal organization: lessons from yeast. *The Journal of cell biology*, 192(5), pp.723–733.

Zuin, J. et al., 2014. Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proceedings of the National Academy of Sciences of the United States of America*, 111(3), pp.996–1001.

Zwerger, M. & Medalia, O., 2013. From lamins to lamina: a structural perspective. *Histochemistry and cell biology*, 140(1), pp.3–12.

Appendix

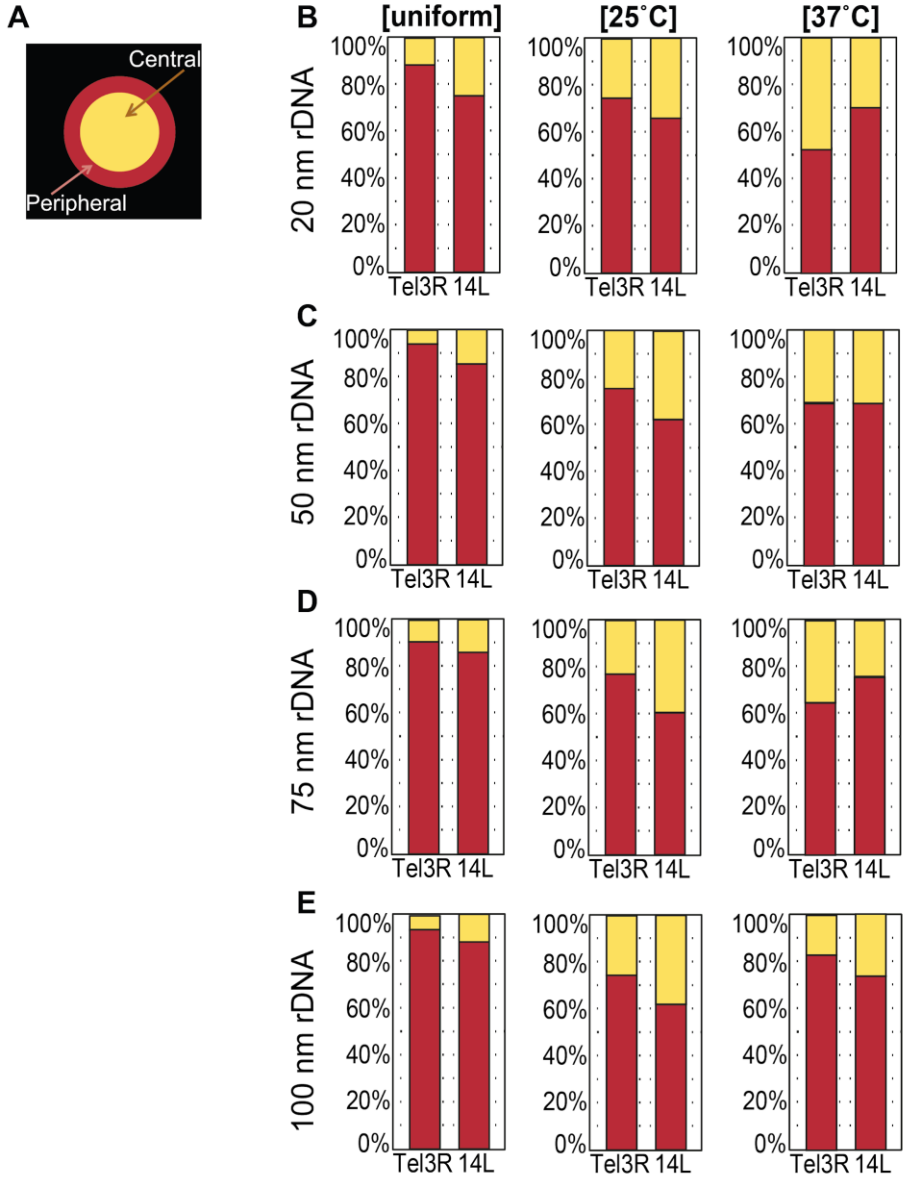


Figure A.1. Telomere positions in simulations with different radius rDNA thickness. (A) A Schematic view of the peripheral and central regions with equal areas (same as Fig. 3.7A). **(B-E)** The positions of labelled telomeres, i.e. right telomere of chromosome III (Tel3R) and left telomere of chromosome XIV (Tel14L), in [uniform], [25°C], and [37°C] simulations with different radius for rDNA segments: 20 nm (B); 50 nm (C); 75 nm (D); 100 nm (E). For all plots (B-E), the data was combined from three simulation replicates. For simulations with 100 nm rDNA segments (E), 64,100 sampled time points (5,900 from the first replicate and 29,100 from the other two replicates) were analysed, whereas for other simulations (B-D), the data were taken from 30,000 non-correlated time steps (10,000 per simulation). These data were compared to the microscopy measurements (Fig. 3.7B, Table A.1).

		Tel3R			Tel14L		
		n	% peripheral	p-value	n	% peripheral	p-value
	experiments	80	68.7500 (55/80)	n/a	74y	60.8108 (45/74)	n/a
	Confidence interval (%)	[57.97 ≤ π ≤ 79.53]			[49.01 ≤ π ≤ 72.61]		
20 nm rDNA	[uniform] F_{LC}: S1-S5=1x	30,000	87.8500	5.3358e-06	30,000	75.2167	0.0066
	[25˚C] F_{LC}: S3= 5x; S4=0.2x	30,000	74.7433	0.2460	30,000	65.9100	0.3907
	[37˚C] F_{LC}: S3= 0.2x; S4= 5x	30,000	52.3500	0.0034	30,000	66.3167	0.3264
50 nm rDNA	[uniform] F_{LC}: S1-S5=1x	30,000	94.0067	3.8739e-12	30,000	85.3433	2.1858e-07
	[25˚C] F_{LC}: S3= 5x; S4=0.2x	30,000	75.1933	0.1952	30,000	62.3233	0.8109
	[37˚C] F_{LC}: S3= 0.2x; S4= 5x	30,000	69.0900	1	30,000	68.7100	0.1672
75 nm rDNA	[uniform] F_{LC}: S1-S5=1x	30,000	90.7533	3.1350e-08	30,000	86.1033	6.8183e-08
	[25˚C] F_{LC}: S3= 5x; S4=0.2x	30,000	77.0867	0.0835	30,000	60.6533	1
	[37˚C] F_{LC}: S3= 0.2x; S4= 5x	30,000	64.9900	0.5582	30,000	76.1667	0.0037
100 nm rDNA	[uniform] F_{LC}: S1-S5=1x	64,100	93.8050	7.9179e-12	64,100	88.1342	1.8961e-09
	[25˚C] F_{LC}: S3= 5x;	64,100	74.6708	0.2466	64,100	62.1092	0.8117

	S4=0.2x						
	[37˚C] F_{LC}: S3= 0.2x; S4= 5x	64,100	83.0109	0.0016	64,100	73.5819	0.0170

Table A.1. Statistical comparison of telomere positions quantified in vivo and in simulations with different rDNA thickness. The binomial confidence intervals for the probability of peripheral positions were calculated from the *in vivo* data (http://onlinestatbook.com/2/estimation/proportion_ci.html). The confidence level was 95% and continuity correction was applied. Simulations with results within the calculated confidence intervals are the best match to the experimental data. The simulation and microscopy data were further compared using binomial tests (MATLAB file exchange function *myBinomTest*). H_0: The experimental data is taken from a population with the same distribution as the simulation data. Small p-values indicate that the distributions are significantly different. According to the confidence interval and the p-values, the [25°C] and [37°C] simulations with 50 nm rDNA segments have provided the best match to the experimental measurements.

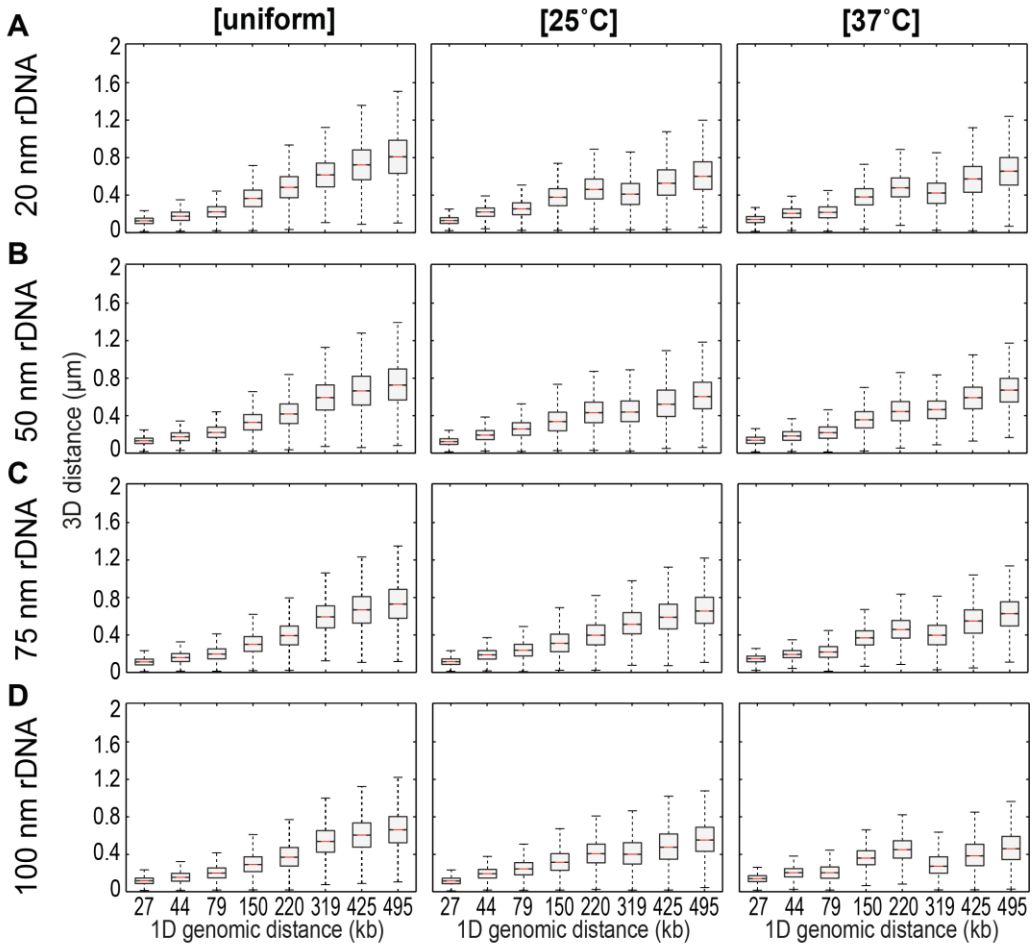


Figure A.2. 3D distance distribution between pairs of labelled loci in simulations with different rDNA thickness. (A-D) The 3D distances between loci separated by different genomic distances were quantified in simulations. The rDNA radius equals 20 nm in A, 50 nm in B, 75 nm in C, and 100 nm in D. For each type of simulation, the data from triplicate simulation runs were combined. For A-C, 30,000 sampled time steps and for simulations with 100 nm rDNA segments, 64,100 non-correlated time points were analysed. The results of simulations were compared to the *in vivo* measurements (Fig. 3.8D, Table A.2).

	Linear model A:	Linear model B:	Vuong test assumption (H1):			
	derived from experimental data and		Model A fits better than model B		Model B fits better than model A	
	simulation dataset A	simulation dataset B	p-value	significance	p-value	significance
20 nm rDNA	[25˚C] F_{LC}: S3= 5x; S4=0.2x	[uniform] F_{LC}: S1-S5=1x	0.99	ns	1.7e-4	***
		[37˚C] F_{LC}: S3= 0.2x; S4= 5x	5.75e-5	****	0.99	ns
50 nm rDNA	[25˚C] F_{LC}: S3= 5x; S4=0.2x	[uniform] F_{LC}: S1-S5=1x	1.76e-5	****	1	ns
		[37˚C] F_{LC}: S3= 0.2x; S4= 5x	1.25e-15	****	1	ns
75 nm rDNA	[25˚C] F_{LC}: S3= 5x; S4=0.2x	[uniform] F_{LC}: S1-S5=1x	1.021e-9	****	1	ns
		[37˚C] F_{LC}: S3= 0.2x; S4= 5x	p=<2e-16	****	1	ns
100 nm rDNA	[25˚C] F_{LC}: S3= 5x; S4=0.2x	[uniform] F_{LC}: S1-S5=1x	p=1.251e-3	**	0.999	ns
		[37˚C] F_{LC}: S3= 0.2x; S4= 5x	p=<2e-16	****	1	ns

Table A.2. Statistical comparison of 3D distances measured in vivo and in simulations with different rDNA thickness. Linear models were derived from ln-transformed experimental data, defined as predicted variable, and the means of ln-transformed simulation data, defined as predictor variable. Then, the Vuong test was performed to compare the fit of the calculated linear models. Hypotheses of the Vuong test are as follows: H0: Model fits are equal; H1A: Model A fits better than Model B; H1B: Model B fits better than Model A. Small p-values indicate a better fit. To run the Vuong test, the R package nonnest2, v. 0.4-1

(https://CRAN.R-project.org/package=nonnest2) (Merkle et al. 2014), was used. ****:p-value<0.0001; ***:0.0001<p-value<0.001; **:0.001<p-value<0.01; ns: not significant. According to the p-values, the linear models derived from the [25˚C] simulations with ≥50 nm rDNA segments have provided better fits to the experimental data than the linear models derived from [uniform] and [37˚C] simulations.

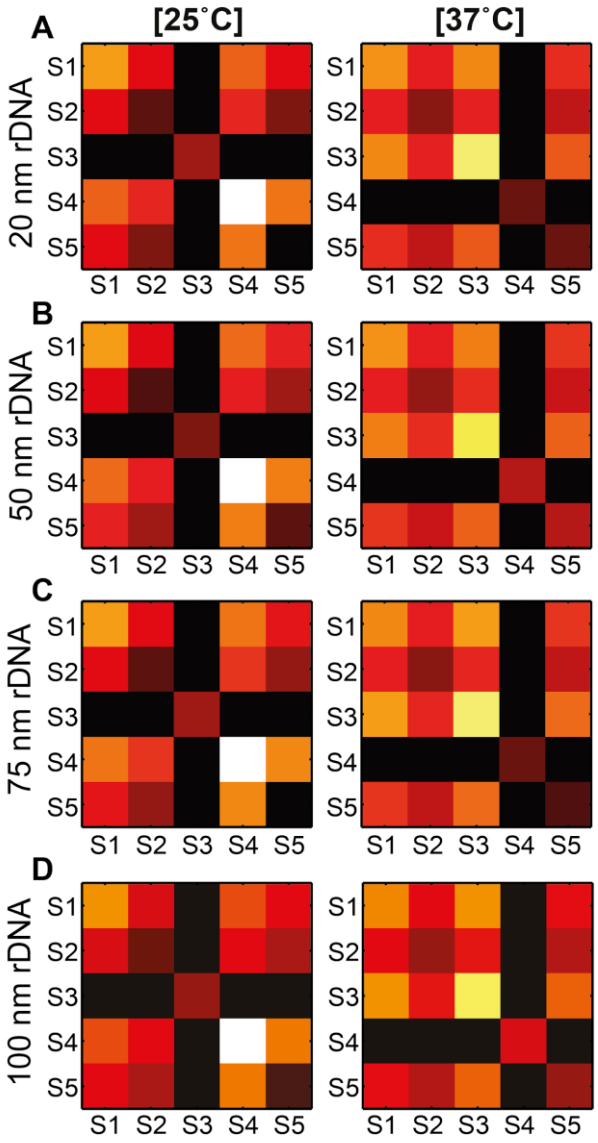


Figure A.3. State-wise contact maps of [25˚C] and [37˚C] simulations with different rDNA thickness. (A-D) The normalised state-wise contact maps were quantified from three simulation replicates as explained in section 4.3.2. For simulations with 20 (A), 50 (B), and 75 (C) nm rDNA segments, the data was taken from 10,000 non-correlated data points, per independent simulation. For simulations with 100 nm rDNA segments (D), 5,900 sampled time points were analysed, per simulation replicate. These results were compared to the Hi-C contact maps (Fig. 4.6A,B, Table A.3).

	State-wise contact maps		Correlation coefficient	p-value	Significance
20 nm rDNA	Hi-C 25˚C	Simulation [25˚C]	0.8630	3.4383e-5	****
	Hi-C 37˚C	Simulation [37˚C]	0.7561	0.0011	**
50 nm rDNA	Hi-C 25˚C	Simulation [25˚C]	0.8662	2.9712e-5	****
	Hi-C 37˚C	Simulation [37˚C]	0.7409	0.0016	**
75 nm rDNA	Hi-C 25˚C	Simulation [25˚C]	0.8543	5.0069e-5	****
	Hi-C 37˚C	Simulation [37˚C]	0.7570	0.0011	**
100 nm rDNA	Hi-C 25˚C	Simulation [25˚C]	0.8634	3.3810e-5	****
	Hi-C 37˚C	Simulation [37˚C]	0.7495	0.0013	**

Table A.3. Correlation of state-wise contact maps of Hi-C and heteropolymeric simulations with different rDNA thickness. Pearson correlation coefficients calculated in MATLAB between two vectors of size 15 (the numerical values of the upper-right triangular part of matrices depicted in Fig. 4.6A,B and Fig. A.3). ****:p-value<0.0001; **:0.001<p-value<0.01; ns: not significant.

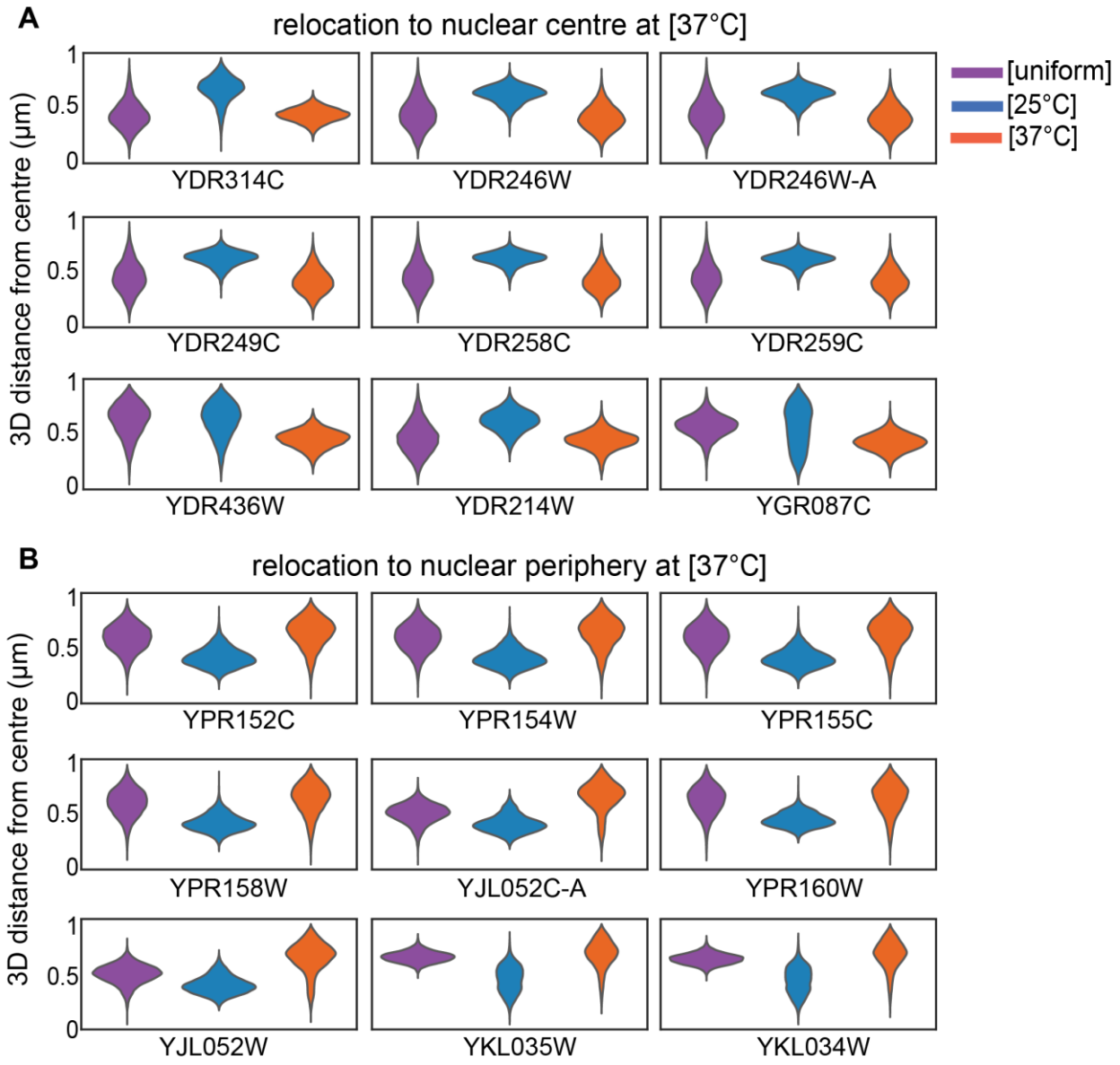


Figure A.4. Relocation of S4 segments upon changes in temperature condition. The 3D distances of genes from the centre of the nucleus and their median values were quantified in [25˚C] and [37˚C] simulations. Then, the genes with the highest differences in their median values at [25˚C] and [37˚C] conditions were analysed and their distances were plotted using the Python Seaborn library function *violinplot*. The first group of genes have more peripheral position at [25˚C] **(A)**, whereas, the second group has more peripheral positions at [37˚C] **(B)**. The data was obtained from one simulation replicate, i.e. 10,000 time points (rDNA = 50 nm).