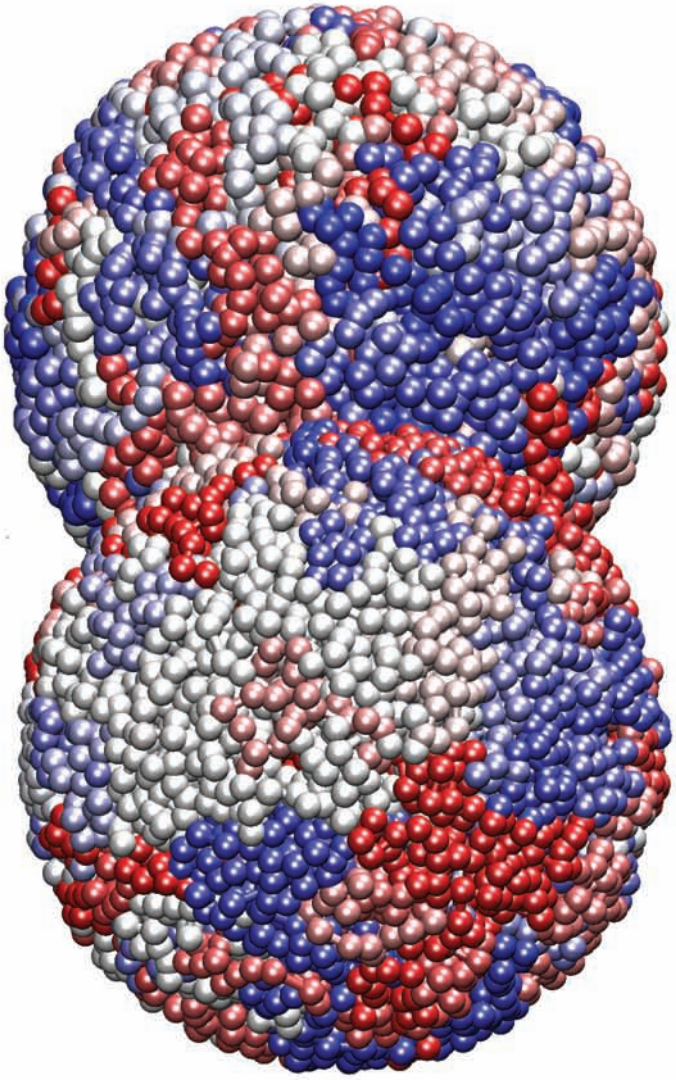
# The role of non-specific interactions in nuclear organization



Silvester de Nooijer

# The role of non-specific interactions in nuclear organization

Silvester de Nooijer

**Thesis committee**

**Thesis supervisors**
Prof. dr. Ton H.J. Bisseling
Professor of Molecular Biology (Development Biology of Plants)
Wageningen University

Prof. dr. Bela M. Mulder
Professor of Theoretical Cell Physics
Wageningen University

**Thesis co-supervisor**
Dr. ir. Joan E. Wellink
Assistant professor, Laboratory of Molecular Biology
Wageningen University

**Other members**
Dr. ir. Paul F. Fransz, University of Amsterdam
Prof. dr. Roel van Driel, University of Amsterdam
Prof. dr. ir. Frans A.M. Leermakers, Wageningen University
Prof. dr. Peter R. Cook, University of Oxford, UK

This research was conducted under the auspices of the Graduate School of Experimental Plant Sciences.

# The role of non-specific interactions in nuclear organization

Silvester de Nooijer

**Thesis**
submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. dr. M.J. Kropff,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Tuesday 26 October 2010
at 4 p.m. in the Aula.

Silvester de Nooijer
The role of non-specific interactions in nuclear organization,
120 pages

Thesis Wageningen University, Wageningen, NL (2010)
With references, with summary in Dutch

ISBN: 978-90-8585-792-1

# Summary

The most important organelle in eukaryotic cells is the nucleus. Within the nucleus, gene transcription occurs: the first step towards protein production. Regulation of gene expression determines which proteins are available to the cell and therefore its function. Errors in gene transcription are the basis of many diseases, and also for many other biological and biotechnological applications a thorough understanding is required of the molecular mechanisms behind regulation of gene expression. Regulation of gene expression is a complicated process in which many factors determine whether, and to what extent, a gene is transcribed. One of these is the spatial position of the transcribed gene in the nucleus. Therefore, it is important to understand the mechanisms behind the spatial organization of the nucleus.

The spatial organization of the eukaryotic nucleus derives from interactions between its constituents. Many of these interactions are specific, for instance the interactions between a DNA binding protein and its target DNA sequence. Apart from specific interactions, non-specific interactions also influence nuclear organization. Non-specific interactions stem from physical encounters between molecules or particles, which can favour particular organizations, i.e. the ones that have the lowest entropy. The role of non-specific interactions in nuclear organization is so far not extensively studied. It is the aim of the research described in this thesis to increase our understanding of their contribution to nuclear organization.

Non-specific interactions are significant for nuclear organization because the nucleoplasm contains 10-40% (v/v) macromolecules. 'Excluded volume' interactions such as depletion attraction are particularly important. Polymer effects also greatly affect the nucleus because interphase chromosomes consist of DNA bound by proteins, the chromatin fibre, which has a very large length-to-thickness ratio. The primary structure of chromatin is linear, but during processes such as regulation of transcription and the formation of heterochromatin, looped chromatin structures are formed. The presence of chromatin loops affects the folding of chromosomes into chromosome territories, which in turn affects nuclear organization. To study the effects of non-specific interactions on nuclear organization, we have used molecular dynamics (MD) simulation techniques from the field of statistical mechanics. Various chromatin loop models can be implemented in these simulations as chains of monomers, which can form loops, branches or networks. In these simulations, the effects of chromosome loops on the dynamics and spatial distribution of the model chromosomes can be evaluated. Through a comparison of simulation results with experimental data, these models can be verified or falsified.

In chapter 2 MD simulations of previously published models for Arabidopsis chromatin organisation are used to show that non-specific interactions can explain the *in vivo*

localisation of nucleoli and chromocenters. Specifically, we show that there is no need for specific interactions between chromocenters and the periphery. Also, we quantitatively demonstrate that chromatin looping contributes to the formation of chromosome territories. The results are consistent with the previously published rosette model for Arabidopsis chromatin organisation and suggest that chromocenter-associated loops limit chromocenter clustering. We show how nuclear organization depends on parameter variations in this rosette model, such as the sizes of chromocenters and nucleoli. Focussing on the forces driving nuclear organization in the rosette model, we derive effective interaction potentials for rosette-loop interactions. These potentials are weak, but nevertheless drive chromocenters and nucleoli to the nuclear periphery and away from each other, providing further proof that chromocenter-associated loops limit clustering. Effective interactions can also be implemented in simulations, leading to a significant improvement in simulation speed as well as further understanding of the interplay of forces in Arabidopsis nuclear organization.

In chapter 4 we have used MD simulations to study the folding of a single human chromosome within its territory and the effect of chromatin loops on its folding,. The results of our simulations are analysed using a virtual confocal microscope algorithm which has the same limitations as a real confocal microscope, notably a limited spatial resolution and a signal detection threshold. Thus we show that chromatin looping increases the volume occupied by a 10Mbp chromosomal sub-domain, but decreases the overlap between two neighbouring sub-domains. Our results furthermore show that the measured amount of overlap is highly dependent on both spatial resolution and signal detection threshold of the confocal microscope, and that in typical fluorescence in situ hybridisation experiments these two factors contribute to a gross underestimation of the real overlap. Zooming out to whole nucleus organization, we investigate which features of human nuclear organization can be explained through interactions between heterochromatin and lamina. A full complement of human chromosomes inside a nucleus is simulated, with each chromosome represented by a block polymer containing two monomer types: eu- and heterochromatin. We added two short-ranged attractive interactions: an interaction between heterochromatic monomers, and an interaction between these and the nuclear periphery. Results show that an interplay between these interactions generates a wide variety of nuclear organizations, with those occurring in nature requiring a fine balance between both interactions. The radial distribution of chromosomes in these simulations correlated with the distribution in real human nuclei, showing that the heterochromatin-heterochromatin and heterochromatin-lamina interactions provide a feasible mechanism for human nuclear organization.

Finally, we discuss the implications of the results presented in this thesis for models of chromosome folding. The differences between chromosome folding in human and Arabidopsis can be explained through differences in genomic structure and chromosome loops, but the underlying mechanisms and forces that organize the nucleus are very similar. The insight how specific and non-specific forces cooperate to shape nuclear organization, is therefore the most important contribution of this thesis to scientific progress.

# Table of Contents

# 1. Thesis introduction

## The importance of nuclear organization

The most important organelle in every eukaryotic cell is the nucleus. It contains the genome and is the setting in which gene transcription takes place, which determines the proteins available to the cell and thus its function and identity. Gene transcription is a tightly regulated process which involves many proteins, often present in complexes and associating with DNA. Apart from gene transcription and the regulation thereof, the nucleus also has important functions in DNA replication and maintenance, mRNA processing and ribosome production.

Far from being a disorganized bag of macromolecules, the nucleus is divided into distinct sub-compartments which have separate functions (Spector, 2003). Such spatial separation is important because different processes require different proteins. Physical separation thus serves to ensure that proteins and nucleic acid sequences interact with their correct targets. For instance, there are many examples (Dernburg *et al*., 1996; Brown *et al*., 1999; Skok *et al*., 2001) of genes that move to different sub-nuclear locations when transcribed, and genes of which transcription is altered when they are artificially moved to another nuclear location (Finlan *et al*., 2008; Reddy *et al*., 2008; Kumaran & Spector, 2008). Thus the sub-nuclear localization of genes into functional compartments fine-tunes levels of transcription (Deniaud & Bickmore, 2009). Therefore, it is important to understand how nuclear organization is established. Although nuclear organization has been studied for a long time (Flemming, 1880) and nuclear organization has been described in great detail (summarized below), we still know little of the mechanisms by which nuclear organization is achieved.

Until recently, nuclear organization was thought to be relatively static and organized by a skeleton-like nuclear matrix analogous to the cytoskeleton (Davie, 1995; Bode *et al*. 2000). However, in recent work (Cremer *et al*., 2000; Soutoglou & Misteli, 2007) the nucleus is described as a highly dynamic structure in which most components, including those previously described as a rigid nuclear matrix (Castano *et al*., 2010), are free to diffuse and encounter each other. In contrast to cellular organelles, which are delineated by membranes and positioned by the cytoskeleton, the nucleus has no physical barriers between compartments. Instead, organization arises from associations between proteins and DNA into complexes, and then association of these complexes into ever larger complexes (Misteli, 2005). Because each protein has binding affinity only for specific other proteins or specific DNA sequences, clusters are formed of interacting proteins and DNA, and different clusters are spatially separated. These spatially separated clusters of proteins, DNA and other

molecules are the functional sub-compartments of the nucleus, of which the largest, e.g. the nucleolus, can even be discerned through a light microscope. Such organization through interactions between constituents is called self-organization (Misteli, 2001; Cook, 2002; Rajapakse *et al*., 2009). While there is thus a gradual transition between large protein-DNA complexes and small compartments, we here reserve the latter term for those protein/DNA aggregates that are too large and too structurally diverse to be described as complexes.

Many of the interactions that organize the nucleus are specific, for instance interactions between a DNA binding protein and its target DNA sequence. Apart from these, non-specific interactions inevitably also influences nuclear organization (Rivas *et al*., 2004; Richter *et al*., 2008). Non-specific interactions are the result of apparent forces caused by physical encounters between molecules or particles. The role of non-specific interactions in nuclear organization is so far poorly studied. It is the aim of this thesis to contribute towards understanding of non-specific interactions in nuclear organization. Before discussing these interactions, nuclear organization will be discussed in detail.

## Organization of the eukaryotic nucleus

The largest of the functional compartments of the nucleus is the chromatin fraction, which contains the DNA. To form chromatin, chromosomal DNA is wrapped around a complex of 8 proteins, 2 copies of each of the histone proteins 2a, 2b, 3 and 4. One histone complex associates with 146bp of DNA, forming a nucleosome. Successive nucleosomes bind every ~160-220bp on each chromosome, resulting in a "beads on a string" configuration consisting of alternating nucleosomes and spacer stretches of DNA devoid of nucleosomes. The resulting fibre has a thickness of approximately 10nm. In actively transcribed parts of chromatin, no further compaction occurs. In the promoter region of many genes nucleosomes even dissociate entirely, resulting in chromatin that has a low nucleosome density (Leimgruber *et al*. 2009). In transcriptionally inactive parts of chromatin however, the 10nm fibre folds through association between nucleosomes, leading to the formation of higher order structures through the action of proteins such as histone 1 (Li *et al*., 2010). *In vitro*, formation of a 30nm fibre has been demonstrated (Finch & Klug, 1976), for which many different structures have been proposed and observed (Woodcock *et al*., 1984; Dorigo *et al*., 2004; Schalch *et al*. 2005). These include solenoid, crossed-linker and irregular structures. The 30nm fibre has so far not been demonstrated in interphase nuclei (Eltsov *et al*., 2008). And while during mitosis chromosomes are compacted further through formation of higher-order structures, it is unclear whether compaction beyond a 30nm fibre takes place during interphase. Thousands of other proteins bind to the chromatin fibre, performing many functions, such as regulating DNA methylation (Teixeira & Colot, 2010) and modifying histones (Bartova *et al*., 2008), which in turn influence chromatin compactness and ultimately contribute to regulation of gene expression and cell identity (Geiman & Robertson, 2002).

On the scale of entire interphase nuclei, two states of chromatin can be distinguished: euchromatin and heterochromatin (Heitz, 1928). Euchromatin has a lower amount of DNA per volume unit, contains relatively gene-rich chromosome segments, and is more readily accessed by binding proteins. Heterochromatin in contrast contains mainly repeat-rich sequences and is more condensed. These states also differ in relative enrichments of DNA methylation (Yan *et al*., 2010) and various histone modifications (Rosenfeld *et al*., 2009), and in protein composition (Cheutin *et al*., 2003). Euchromatin and heterochromatin occupy spatially separate compartments, but the distribution of these compartments over the nucleus varies widely per species and sometimes even between cell-types within one species (e.g. Bartova *et al*., 2002; Solovei *et al*., 2009). In human nuclei, heterochromatin occurs mostly dispersed at the nuclear periphery and around the nucleolus. In Arabidopsis and to a lesser extent in mouse, heterochromatin is more concentrated, forming chromocenters; micrometer-sized compact sub-compartments surrounded by euchromatin (Soppe *et al*., 2002). Like dispersed heterochromatin, chromocenters often localize at the nuclear periphery. In animals, heterochromatin interacts with lamina, a network of type V intermediate filaments consisting of Lamin A, B1, B2 and C proteins. Lamina are bound to the inner surface of the nuclear membrane through interactions with several proteins such as LAP1/2 and LBR (Furukawa *et al*., 1995; Worman *et al*., 1988). Chromatin interacts directly with lamina (Yuan *et al*., 1991) and indirectly through proteins that bind to lamin proteins, such as HP1 (Ye & Worman, 1996), MeCP2 (Guarda *et al*., 2009), LBR and LAP2 (Mattioli *et al*., 2008). These interactions are thought to play a role both in the formation and the peripheral localization of heterochromatin (Fiserova & Goldberg, 2010). However, in plants no lamina proteins have been found (Fiserova & Goldberg, 2010).

Each chromosome in higher eukaryotes occupies its own spatial domain within the interphase nucleus. Spatial separation of chromosomes is thought to facilitate chromosome compaction and subsequent diakinesis during mitosis by keeping chromosomes disentangled. The extent to which intermingling between territories occurs is debated: while some experiments show that very little intermingling occurs between adjacent territories (Zorn *et al*., 1979; Munkel *et al*., 1999), other experiments show that DNA belonging to other chromosomes occurs deep within chromosome territories (Pombo *et al*., 2006) and that extensive contacts occur between DNA from different chromosomes (Lieberman-Aiden *et al*., 2009). While the positions of territories within the nucleus vary from cell to cell, in some species territories have a specific radial localization which correlates with heterochromatin content (Branco *et al*., 2008), and some territories co-localize more frequently than others (Lieberman-Aiden *et al*., 2009). Chromosome territories are themselves further sub-compartmentalized, with chromosome arms (Dietzel *et al*., 1998) as well as 10Mbp-sections of these arms (Goetze *et al*., 2007) occupying separate sub-territories in a structure resembling a fractal globule (Liebermann-Aiden *et al*., 2009).

Besides the chromatin fraction, the nucleus contains many functional sub-compartments that interact with chromatin. The most conspicuous by size is the nucleolus, an often micrometer-sized compartment of which the primary function is the synthesis of rRNA required for the production of ribosomes. The nucleolus contains little DNA apart from the rDNA gene repeats (also known as nucleolar organizing regions, NOR) which are highly transcribed inside the nucleolus when production of additional ribosomes is required. Other nuclear compartments include Cajal bodies (Morris, 2008), which are involved in assembly of the transcriptional machinery of the cell, splicing-factor compartments (Vyakarnam *et al*., 1998), which contain many components of mRNA processing pathways, and many others such as paraspeckles (Fox *et al*., 2002) and PML bodies (Lallemand-Breitenback & de The, 2010) of which the function is unclear. Much smaller still are transcription factories (Jackson & Cook 1995; Osborne *et al*., 2004), which contain RNA polymerase 2 and associated transcription factors. None of the compartments described here have a fixed position in every nucleus, but nevertheless their formation and localization are controlled by interactions between them.

## Interactions organize the nucleus

As mentioned before, the interactions driving nuclear self-organization can be separated into two groups: specific and non-specific interactions. Specific interactions occur between proteins and specific binding sites within other proteins or in DNA. Such specific binding is often an essential step in protein activity. At the same time, the specific association of proteins with DNA and with each other imposes structure on the interphase nucleus, for instance through the formation of chromatin, and specifically chromatin loops. Specific interactions between transcription factors bound to distant regulatory elements and RNA polymerase 2 complexes bound at a transcription initiation site, lead to transient formation of a loop in chromatin containing all the elements in between. Similarly, loops can be formed through the action of boundary elements. These loops involve a single gene and its regulator elements and are thus generally small, although in human they may span up to hundreds of kilobases of DNA (Palstra *et al*., 2008). Larger loops occur when genomically distant genes that share regulatory elements and thus require to bind the same transcription factors, co-localize into a transcription factory (Jackson *et al*., 1998). Other large loops are formed by co-localization of distant heterochromatin domains (Fischer *et al*., 2004). All of these loops are transient, but may be present for considerable amounts of time.

Non-specific interactions are, as their name implies, ubiquitous; they occur between all molecules everywhere. While their effects are usually small and easily overcome by thermal fluctuations, there are two reasons why they contribute significantly to nuclear organization. First, the nucleus is a crowded solution of macromolecules, which associate into the complexes and particles described above, and which occupy 10-40% of the nuclear volume (Bohrmann *et al*., 1994; Hancock, 2004). Therefore excluded volume interactions, particularly

depletion attraction, are strong organizing forces inside the nucleus (Marenduzzo *et al*., 2006a; Cook & Marenduzzo, 2009). Crowding also increases interaction rates between proteins and changes the kinetics of such interactions (Zimmerman & Harrison, 1987; Richter *et al*., 2008), which may affect nuclear organization. For instance, the maintenance of the heterochromatic state has been proposed to be influenced by molecular crowding (Bancaud *et al*., 2009). Second, the nucleus contains multiple DNA strands: polymers with a linear length of up to decimetres, which are nevertheless confined into the nucleus, which has dimensions in the range of micrometers. Therefore polymer behaviour, including effects due to excluded volume interactions (Dorier & Stasiak, 2009) and length-to-thickness ratio (Rosa & Everaers, 2008), greatly affects the organization of the nucleus. However, despite these theoretical considerations the contributions of non-specific interactions to nuclear organization remain both little studied and understood. An overview of our current knowledge is presented in the introduction of chapter 2. This thesis aims to elucidate the role of non-specific interactions in nuclear organization using polymer models in two species: Arabidopsis and human. Both species have well-studied chromatin organization and for both systems polymer models of their chromosomes have been described in literature.

*Arabidopsis thaliana* is a small plant favoured as a model system for the study of nuclear organization because of its small (150Mbp), fully sequenced genome and low repetitive sequence content of ~15% (Arabidopsis genome initiative, 2000). It is a diploid with 5 chromosome pairs, each of which contains a heterochromatic central block consisting of a centromere surrounded by pericentromeric repeats, which are largely transposon derived. The chromosome arms are largely euchromatic, with a very high gene density of approximately 1 gene per 5Kbp. NORs containing 45S rDNA sequences are localized at the telomeric ends of the short arms of chromosomes 2 and 4, and 5S rDNA is located at the centromeric end of the short arms of chromosomes 2 and 5. During interphase the centromere, pericentromere and NORs associate into 6 to 10 chromocenters, which are localized at the nuclear periphery. Fluorescent in situ hybridisation studies have shown that euchromatin emanates from the chromocenters in loops of 0.1-1.5Mbp. Based on these studies a model has been proposed for interphase organization where sequences from the euchromatic arms associate with the chromocenter. This would create a rosette-like structure with approximately 15-50 loops emanating from the chromocenter (Fransz *et al*., 2002).

Human is favoured for the study of nuclear organization because of the important role chromatin organization plays in physiology and pathology (e.g. Kohler & Hurt, 2010; Marella *et al*., 2009). Human nuclei contain 23 chromosome pairs, and the genome size is ~3Gbp. Compared to Arabidopsis, interspersed repetitive sequences are far more abundant in human (45%, international human genome sequencing consortium, 2001), and genes contain larger regulatory elements and introns, resulting in a gene density of approximately

1 gene per 100Kbp. Like Arabidopsis, human chromosomes have gene-poor and repeat-rich centromeres. Within the chromosome arms gene-rich and gene-poor areas can be distinguished which at a length-scale of ~10Mbp are called ridges and anti-ridges, and at a scale <1Mbp are called gene islands and deserts, respectively. Eu- and heterochromatin occur mostly dispersed throughout the nucleus, with heterochromatin concentrated at the nuclear periphery.

A succession of models has been proposed that would describe the structure of human chromosomes during interphase. The Random Walk-Giant Loop model (RW-GL, Sachs *et al*., 1995) models chromosomes as chains following a random walk path, tethered at matrix attachment regions to the nuclear matrix, thus forming large loops. Later, a model was proposed that does not depend on the existence of a cytoskeleton-like nuclear framework to which DNA attaches: the multi-loop-subcompartment model (MLS, Munkel *et al*., 1999). In this model, the chromosome consists of a string of ~1Mbp rosettes formed by multiple loops in the size range of ~100Kbp. However, such rosettes have never been demonstrated, and an alternative model was proposed: the Random Loop model (RL, Bohn *et al*., 2007; Mateos-Langerak *et al*., 2009), in which associations occur between loci at random positions in the chromosome, leading to a network structure. Unlike in previous models, in the RL model chromosome territories are formed due to the cross-links which hold the chromosome together.

## Research methods

The described models for chromosome folding can be studied using simulation techniques from the field of statistical mechanics. Models for chromosome structure can be implemented in these simulations as chains of monomers, which can form loops, branches or networks. By running simulations of various models and evaluating the effects of the various substructures on the dynamics and spatial distribution of the model chromosomes, and comparing these to experimental data, models can be verified or falsified. In this way the RW-GL model was rejected because data on physical versus genomic distance between markers obtained from simulations of this model did not match experimental data (Muenkel *et al*., 1999). In contrast, data obtained from simulations implementing the MLS- (Muenkel *et al*., 1999) or RL-models (Bohn *et al*., 2007) did match experimental data. In this thesis, we test the rosette model for Arabidopsis chromosomes against biological data, and also examine whether the RL model for human chromosomes can explain experimental data on sub-chromosomal domain formation.

Another way in which simulations can help us understand chromatin folding, is through examination of the forces underlying chromatin folding. Simulations have already long been used to evaluate chromatin folding at the molecular scale, for instance folding into 10- (Langowski & Heermann, 2007, Sereda & Bishop, 2010) and 30-nanometre fibres (van Holde

& Zlatanova, 2007). At larger length-scales, simulation techniques have been used to investigate forces involved in transcription factory formation (Cook & Marenduzzo, 2009). Other applications have been to the question whether chromosomes form territories, where it was predicted that unlooped chromosomes will eventually mix (Muenkel *et al*., 1999) but may not have sufficient time to do so within the lifetime of the cell (Rosa & Everaers, 2008). In this thesis, we examine the implications that non-specific interactions have on chromatin organization in both human and Arabidopsis using coarse-grained simulations.

In our simulations, we model chromosomes as chains of monomers, sometimes including loops or particles of various sizes to represent nuclear sub-compartments such as nucleoli, and then perform simulations in which the particles and polymers interact based on the laws of mechanics. The spatial configurations of polymers and particles are stored at intervals. On these, statistical analysis is performed to determine equilibrium states of the system. These are the states most favoured by the organizing effects of non-specific interactions, operating within the constraints (e.g. polymer structure, confinement) of the simulated system.

Such simulations allow analysis and perturbation of the studied system to a much greater degree than real biological systems. Simulations of nuclear organization can contribute to understanding of the non-specific forces at play in a real biological nucleus. Although the modelled systems necessarily include only a small fraction of the total biological and chemical complexity of real biology and biochemistry, their behaviour can be compared to that of real biological model systems and thus non-specific forces that drive organization can be identified.

Two main simulation techniques exist for obtaining configuration samples. Molecular Dynamics (MD) emulates the behaviour of chromosomes and nuclear sub-compartments by applying Newton's laws to interacting particles. This technique only allows particles to move in ways that are physically feasible. By integrating the equations of motion for these particles for many thousands of cycles, sample equilibrium configurations are directly obtained. In contrast, Monte Carlo (MC) methods allow non-physical moves such as particles disappearing and appearing in other places and reorganizing entire groups of particles, but at a cost: only those moves that lead to a net decrease of energy are accepted directly. All other moves are accepted with a probability inversely dependent on the gain of energy they cause.

Which of these techniques equilibrates and samples the equilibrium fastest depends on the system simulated and the computational resources that are available. MD algorithms are limited because of the complex calculations required to integrate Newtonian equations and because only physical moves are permitted. MC calculations can be simpler but are performed in vain when a move is subsequently rejected. This especially becomes a problem

in high density systems, because many moves will generate high overlap between particles and thus have high rejection rates. In contrast, MD algorithms become more efficient in high density systems because of shielding: each particle only influences the particles around it and therefore only the local neighbourhood of each particle needs to be considered. In this thesis, we have therefore chosen to use MD simulations for sampling.

Although MD simulations emulate the dynamics that occur within a real nucleus, their goal is not to recreate the actual dynamics of the system, but rather to serve as a method for obtaining samples from the thermodynamic equilibrium of the system. Statistical analysis is then required to calculate parameters of interest from these samples. Therefore it is important to provide proof of equilibration and of sufficient sampling.

## Contributions of this thesis to the study of nuclear organization

In chapter 2, we study the contributions of non-specific interactions to the nuclear organization of Arabidopsis. We use MD simulations of previously published models for Arabidopsis chromatin organisation to show that non-specific interactions can explain the *in vivo* localisation of nucleoli and chromocenters. Specifically, we show that there is no need for specific interactions between chromocenters and the periphery. Also, we quantitatively demonstrate that chromatin looping contributes to the formation of chromosome territories. These results are consistent with the previously published rosette model for Arabidopsis chromatin organisation and suggest that chromocenter-associated loops limit chromocenter clustering.

In chapter 3, we further examine the rosette model of chapter 2, studying the underlying physical principles. We show how nuclear organization depends on parameter variations such as the sizes of chromocenters and nucleoli. We also replace the rosette loops with effective interaction potentials. These potentials are remarkably weak, but nevertheless drive chromocenters and nucleoli to the nuclear periphery and away from each other, providing further proof that chromocenter-associated loops limit clustering.

In chapter 4, we use MD simulations to study the folding of a single human chromosome within its territory and the effect of chromatin loops on this folding, through simulations of both a linear and a randomly looped polymer in confinement. The results of our simulations are analysed using a virtual confocal microscope algorithm which has the same limitations as a real confocal microscope, notably a limited spatial resolution and a signal detection threshold. Thus we show that chromatin looping increases the volume occupied by a 10Mbp chromosomal sub-domain, but decreases the overlap between two such sub-domains. Our results furthermore show that the measured amount of overlap is highly dependent on both spatial resolution and detection threshold of the confocal microscope, and that in typical FISH experiments these two factors contribute to an underestimation of the real overlap.

In chapter 5, we study the organization of human radial chromosome territories. Our research question is whether interactions between heterochromatin and lamina can explain the observed radial distribution of human chromosome territories. We simulated a full complement of human chromosomes inside a nucleus, with each chromosome represented by a block polymer containing two monomer types: eu- and heterochromatin. Apart from excluded volume interactions, we added two short-ranged attractive interactions: an interaction between beads of heterochromatin, and an interaction between these and the nuclear periphery (representing the lamina). Results show that an interplay between these interactions generates a wide variety of nuclear organizations, with those occurring in nature requiring a fine balance between both interactions. We then selected the organizations that most resemble human nuclei, and studied the radial distribution of chromosome territories. The distribution in our simulations markedly resembled the measured organization of real human nuclei, showing that the heterochromatin-heterochromatin and heterochromatin-lamina interactions provide a feasible mechanism for human nuclear organization.

# 2. The role of non-specific interactions in Arabidopsis nuclear organization

S. de Nooijer[1], J. Wellink[1], B. Mulder[2,3], T. Bisseling[1]

1: Laboratory for Molecular Biology, Wageningen University, Droevendaalsesteeg 1, Wageningen, The Netherlands
2: FOM institute for atomic and molecular physics (AMOLF), Science Park 104, Amsterdam, The Netherlands
3: Laboratory for Plant Cell Biology, Wageningen University, Droevendaalsesteeg 1, Wageningen, The Netherlands

Published as "Non-specific interactions are sufficient to explain the position of heterochromatic chromocenters and nucleoli in interphase nuclei" in Nucleic Acids Research 37, 3558-3568 (2009).

## Abstract

The organisation of the eukaryotic nucleus into functional compartments arises by self-organisation both through specific protein-protein and protein-DNA interactions and non-specific interactions that lead to entropic effects, such as e.g. depletion attraction. While many specific interactions have so far been demonstrated, the contributions of non-specific interactions are still unclear. We used coarse-grained molecular dynamics simulations of previously published models for Arabidopsis chromatin organisation to show that non-specific interactions can explain the *in vivo* localisation of nucleoli and chromocenters. Also, we quantitatively demonstrate that chromatin looping contributes to the formation of chromosome territories. Our results are consistent with the previously published rosette model for Arabidopsis chromatin organisation and suggest that chromocenter-associated loops play a role in suppressing chromocenter clustering.

## Introduction

The eukaryotic interphase nucleus is organised into many functionally specialized regions or substructures such as chromosome territories (Cremer *et al*., 1982), nucleoli, Cajal bodies and speckles (Handwerger & Gall, 2006). The chromosomes are composed of chromatin, a complex of DNA and proteins. The gene-rich euchromatin and repeat-rich heterochromatin, which differ significantly in sequence content, volumetric DNA density, transcriptional activity and epigenetic modifications (Bartova *et al*., 2008) form distinct substructures that occupy spatially separate nuclear regions. Heterochromatin is mostly localised at the nuclear periphery and around the nucleolus, and euchromatin at the interior of the nucleus (Fang & Spector, 2005; Ye *et al*., 1997; Solovei *et al*., 2004). While this nuclear organisation may appear static, the spatial organisation of these substructures most likely involves specific as well as non-specific interactions between dynamic constituent (Chubb & Bickmore, 2003).

Specific interactions, such as protein-DNA and protein-protein interactions have been well described (Lorson *et al*., 1998; Ishov *et al*., 1999; Hebert *et al*., 2000). For instance, heterochromatin has been suggested to localise to the nuclear periphery through interactions with lamina proteins in animal nuclei (Fedorova & Zink, 2008). However, in plants, for which no lamina homologues have been described, heterochromatin still localises peripherally. Apart from such specific interactions, inevitably non-specific interactions also occur. The architecture of the interphase nucleus is thought to arise by self-organisation through both types of interactions (Misteli, 2001; Hancock, 2004). However, the contribution of non-specific interactions to nuclear organisation has so far not been well characterised. Here we study this by identifying properties of nuclear organisation that can be explained through non-specific interactions. Using molecular dynamics simulations of chromatin and comparing the results with microscopy data, we specifically focus on the question where nuclear substructures, especially heterochromatin and nucleoli, will localise due to the effect of non-specific interactions.

Non-specific interactions, and the entropic effects they give rise to, can, based on arguments derived from statistical mechanics, be expected to play a role in this localisation. The nucleus is a crowded environment containing up to 0.4 g/ml of macromolecule (Bohrmann *et al*., 1993), which can be regarded as a mixture of large and small particles in a dense solution (Marenduzzo *et al*., 2006a). Entropy plays an important role in determining the localisation of such particles through depletion attraction (Marenduzzo *et al*., 2006a). This attraction occurs when the translational and rotational degrees of freedom of each particle are limited by all other particles in a crowded environment. Around each particle a zone of excluded volume exists which is inaccessible to the centres of mass of other particles. When less numerous large particles (such as nucleoli and heterochromatic regions) coexist with

more numerous small particles, the total entropy gain of the small particles may outweigh the entropy loss of the large particles when the latter aggregate, thus minimising their excluded volumes. This leads to an apparent force between the large particles, the depletion attraction. Depletion attraction can also occur between large particles and a confinement wall, since a wall too is lined by a large excluded volume. Depletion attraction has in a nuclear context been implicated to be responsible for grouping DNA polymerases together into replication factories (Marenduzzo *et al*., 2006a) and RNA polymerases into transcription factories (Marenduzzo *et al*., 2006b), but can be expected to affect many other structures or functional compartments as well. We here study the effects of depletion interaction on the position of nucleoli and of heterochromatin in interphase nuclei.

Other non-specific interactions arise from the polymer nature of chromosomes. Each chromosome consists of a DNA chain which is compacted by association with proteins to form a chromatin fibre. The first level of compaction occurs through the formation of nucleosomes, consisting of histone proteins, which associate with DNA forming a fibre of approximately 10nm thickness. In more condensed chromatin the nucleosomes form a 30 nm diameter fibre, the exact structure of which is still debated (Tremethick, 2007). Histone 1 and other proteins stabilize this and other higher-order chromatin structures. Chromosomes have a high length to thickness ratio and many internal translational and rotational degrees of freedom and therefore are expected to show behaviour typical for confined polymers in solution. For instance, the chromatin chains can be expected to resist intermingling and will influence the localisation of other functional compartments in the nucleus through exclusion interactions. However, the behaviour of confined polymers in general and the effect of the polymer nature of chromatin on nuclear organisation in particular are difficult to predict theoretically. Therefore behaviour of confined polymers has been investigated using soft matter computer simulation approaches (Cacciutto & Luijten, 2006). These methods have already been applied in a biological context to show that the combination of a physically confined genome inside a rod-shaped bacterium and conformational entropy could fully explain the spatial segregation of duplicated circular chromosomes (Jun & Mulder, 2006).

In interphase nuclei chromatin is not a purely linear chain. Instead it forms loops through specific interactions, for instance during regulation of gene expression (Engel & Tanimoto, 2000), at boundary elements (Wallace & Felsenfeld, 2007) and in transcription factories (Jackson *et al*., 1993), and non-specific interactions (Toan *et al*., 2006) creating in effect a network polymer. The geometry of network polymers influences their localisation and mixing properties (Jun & Mulder ,2006). The geometry of chromatin has been studied initially in human cells where several models have been proposed to describe how the linear 10 to 30 nm fibre folds into a higher order structure through loop formation. An early approach was the random walk/giant loop (RW/GL) model (Sachs *et al*., 1995) which proposes a highly flexible backbone to which giant loops, each comprising several Mbp of DNA, are attached. Monte Carlo simulations based on this model, however, showed that it

cannot explain the spatial distance distribution between chromatin markers in interphase nuclei. Therefore the multi-loop sub-compartment (MLS) model was introduced which does predict the interphase distances rather well (Münkel *et al*., 1999). Based on the MLS model the spherical chromatin domain (SCD) model (Cremer *et al*., 2000) was developed which proposes that multiple loops form rosette-like domains of approximately 1 Mbp in size that are linked by DNA stretches of 120Kbp. Monte Carlo simulations based on the SCD model have been used to show that chromosomes do not have preferred association with any other chromosome. This was done by comparing theoretical association rates between chromosome territories with experimental data on chromosome territory positions in for instance human (Cremer *et al*., 2001) and Arabidopsis nuclei (Pecinka *et al*., 2004). Here we use a similar approach based on molecular dynamics to determine by simulation the localisation of heterochromatin and nucleoli, and the level of chromosome mixing.

Most previous modelling studies have simulated human nuclei or chromosomes. However, the large size of the human genome and the dispersed localization of its heterochromatin make it less suitable for whole-genome simulations. Our study is based on the model plant *Arabidopsis thaliana*, which has a small completely sequenced genome of ~150Mbp (Arabidopsis Genome Initiative, 2001). The 5 chromosome pairs occur in territories that are distributed randomly except for chromosomes 2 and 4 which bear the nucleolar organizing regions (NOR) and therefore associate more frequently with each other and the nucleolus (Pecinka *et al*., 2004). On each of these chromosomes the NOR is located close to the telomere of the short chromosome arm. Fifteen percent of the Arabidopsis genome consists of heterochromatin which is concentrated around the centromeres and at the NORs. The chromosome arms are predominantly euchromatic. In the interphase nucleus the heterochromatic (peri-)centromere and NOR regions localize to 6 to 10 chromocenters, which are preferentially located at the periphery of the nucleus (Fang & Spector, 2005; Fransz *et al*., 2002). A model has been proposed in which the euchromatic arms form loops consisting of 0.1-1.5 Mbp protruding out of the heterochromatic chromocenter, resulting in a rosette-like structure (Fransz *et al*., 2002).

We have implemented this chromatin organization model (and for comparison other models in which loops are not associated with the chromocenter) using self-avoiding polymer chains and performed computer simulations based on Molecular Dynamics on these models. Parameter settings such as chromosome lengths and nuclear size, density and heterochromatin content were based on experimental data. We then compared the predictions derived from our simulations regarding the stability of chromosome territories, the localization of heterochromatic chromocenters and the nucleolus with microscopy data of Arabidopsis leaf mesophyll nuclei and found that entropic forces can by themselves explain the localization of the nucleolus and heterochromatin *in vivo*.

## Methods

**molecular dynamics setup**

Chromatin was modelled as a polymer chain consisting of particles and simulations were performed using the ESPResSo (Limbach *et al*., 2006) software package as described in e.g. in (Arnold and Jun, 2007). Particles are modelled as point centres of a repulsive force, with an interaction range between particles establishing an excluded volume around each particle. The radius of this excluded volume is hereafter treated as the particle radius. For each combination of particle types a Weeks-Chandler-Andersen potential (a Lennard-Jones potential that is shifted and truncated to only include the repulsive part of the potential

$$U(r) = 4\epsilon \left( \left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} + shift \right) \qquad \text{Formula 2.1}$$

(Weeks *et al*., 1971; formula 2.1) was used to define non-bonded particle-particle interactions. The basic length scale in our simulation is the effective distance between the centres of two of the smallest particles in the simulation, set by the parameter σ. The energy scale ε is chosen such that at room temperature (T=298K) $k_BT/\ \varepsilon = 1$. Analogous to (Arnold and Jun, 2007) the time unit is $\tau = \sigma\sqrt{(m/\varepsilon)} = 1$, where m is the (irrelevant) particle mass. The attractive part of the potential was cut off at 1.12246σ by setting shift to 0.25ε and setting the remainder of the potential to 0. Bonded interactions were defined by a harmonic spring potential between two particles with distance r according to formula 2.2. R was set to the

$$U(r) = \tfrac{1}{2}\,(r - R)^2 \qquad \text{Formula 2.2}$$

sum of the radii of the particles involved, K was used as a configurable parameter to control the elasticity of the bonds. To prevent bond extensions that would allow chains to pass through each other, $K_{harmonic}$ was set to 50.0.

Starting configurations for simulations were obtained in several ways. For models involving only linear chains, ESPResSo's pseudo self avoiding walk algorithm was used to create the chains. For models with internal loops, geometrical arrangements were used as starting configurations. Similarly, the initial spatial distribution of chains through the nucleus was either determined randomly or followed a geometrical arrangement.

The simulation was performed by successive velocity-Verlet MD integrations with a time step set to 0.01 in the natural units determined by the (arbitrary) mass of the particles, with the temperature controlled by a Langevin thermostat. The simulations took place within a spherical confinement defined by an LJ potential similar to formula 1 between the virtual perimeter of the confining sphere and all other particle. Initially the radius of the confining sphere was between 3 to 10 times larger than the eventual radius depending on the spatial size of the initial configuration, which had to fit within the confinement. In all cases,

simulations started with a equilibration phase in which the initial configuration was allowed to relax while the LJ potential was 'capped' in order to avoid excessive repulsive potentials due to high degrees of particle overlap in the original configuration. During the equilibration phase the LJ cap was slowly increased, until the cap exceeded all LJ interactions in the simulation and the cap was removed. Simultaneously the confining sphere was slowly shrunk until it reached its eventual size. After the confining sphere reached its final size and the LJ cap was removed, all simulations were further equilibrated until parameters such as total energy, end-to-end distance and contour length, and mixing stabilised (supplementary data figure 2.1).

**number of monomers**

The maximum computationally feasible monomer resolution for a single simulation was ~10Kbp/monomer (30000 monomers in total in the simulation) but at this resolution each experiment takes 6 weeks to complete (on 4 AMD Opteron type 248 processors). Since large series of parameter settings had to be evaluated, a lower monomer resolution of ~75Kbp/monomer (4000 monomers in total) was selected at which simulations (each of which lasted at least $10^7$ integration time steps after equilibration) completed within 1-7 days. We note that if one assumes that the chromosomal DNA is well modelled by a yeast-like 30nm fibre, which has an estimated Kuhn length of about 400nm (Bystricky *et al*., 2004), the most appropriate monomer size would be estimated as 400nm /(10 μm/Mbp) ~ 40Kbp. Unfortunately there is not enough data available on chromatin structure in Arabidopsis to more accurately determine the proper coarse-graining length. However, polymer physics suggests that the global phenomena discussed in this work i.e. the relative positioning of the different chromosomes are relatively insensitive to the actual number of monomers (de Gennes, 1971). The number of monomers in the simulation did prove to have some effect on one of our observables, the mixing score (defined below). In an unmixed state, longer polymers have a higher ratio of internal monomers (which only interact with other monomers belonging to the same chain) to external monomers (interacting with monomers belonging to other chains), resulting in a lower score. However, since scores are normalised through determining the extreme scores of no mixing and perfect mixing, the mixing parameter is corrected for the influence of monomer numbers. To make sure no unexpected monomer number dependence occurs in specific chromatin conformations, we only compared simulations in which the number of monomers was equal (4000 monomers). At a typical monomer volume fraction of 15% and assuming a nucleus 5μm in diameter, the resulting monomer diameter is 169nm.

**implementing chromatin chains, chromocenters and loops**

Linear Chains (LC) and Linear Chains with Chromocenters (LCC) models were implemented by linear chains of monomers with relative lengths as in Arabidopsis (Arabidopsis Genome Initiative, 2001), with the LCC model including a chromocenter at the genomically appropriate position. Chromocenters were implemented as single monomers with size as measured in Arabidopsis leaf mesophyll cells (see below). The rosette model

was implemented by dividing the monomers belonging to the euchromatic chromosome arms equally over the desired number of loops. In the standard situations, 15 loops per chromosome were included because this resulted in loops of 1-2 Mbp. This loop size corresponds to the upper range of sizes suggested in (Fransz *et al.*, 2002). In smaller loops the monomer resolution becomes a limiting factor. In loop size variation experiments the distribution of monomers over loops was changed but the total number of monomers stayed the same. Each loop consists of a linear stretch of regular polymer chain, of which both ends are joined to the chromocenter by harmonic springs of which the length has been adjusted to reflect the centre-centre distance between a chromocenter and a euchromatic monomer. Upon these bonds no rotational constraints are placed. The loops and chromocenter (LAC) model was implemented by adding a bond between two monomers that are one loop size away in sequence. This creates a loop since the monomers between these two monomers no longer participate in the main polymer chain. All loops were of the same size as the bonds in the rosette model with 15 loops per chromocenter, for comparison. In the LAC 100% model, bonds were added in all positions that qualify according to the loop size criterion, creating a very short main chain bristling with loops. In the LAC 50% model, there was a 50% chance that a bond was actually created at each of these locations, leading to loops being on average of equal size as the linear stretches in between the loops, and in the LAC 10% model only a 10% chance. For both rosette and LAC models, geometrical arrangements were employed as starting configurations. No difference was observed after equilibration between starting configurations in which these geometrical arrangements were randomly distributed through space and configurations in which these models were oriented on a spiral as described in the section on starting configurations below.

**the influence of initial starting configurations on mixing parameter**

Two types of starting configurations were employed. In the first, linear chains were introduced at random positions in the confining sphere before compression, and the monomers of each chain were positioned by the pseudo random walk algorithm of the ESPResSo software. The other starting configuration was designed to achieve initial spatial separation of all chains and consisted of a geometrical arrangement in which the positions of the linear chains' middle monomers were distributed over a sphere surface as described by the following equations in polar coordinates: $r_{center} = 2/3\ r_{confining\ sphere}$, $\phi = 6\pi\ i_{chain} / n_{chain}$, $\theta = -9/20\ \pi + \pi\ i_{chain} / n_{chain}$, with $i_{chain}$ the $i^{th}$ chain and $n_{chain}$ the total number of chains. The other monomers were positioned on a cylindrical spiral with a periodicity of 12 monomers, a dislocation of one monomer diameter per winding, and a radius of $12/\pi$ times the monomer radius. No significant difference was observed in the value of the mixing parameter after $10^4$ time units between simulations with these two starting configurations, or by visual inspection of the configurations.

**sampling of ensemble averages of parameters from simulations**

Sample configurations were stored every $10^3$ integration steps (1.6 MD time units) after equilibration was achieved. These stored configurations were then used for calculation of

ensemble averages of all measured parameters. Since all simulations were performed for at least $10^7$ cycles after equilibration, at least $10^4$ configurations were used to calculate ensemble averages of the parameters, except for mixing, which for computational reasons was, in all but one simulation, calculated only once every $10^4$ integration steps (for the simulation on which supplementary data figure 2.1 is based, mixing was calculated every $10^3$ integration steps).

**derivation of model/simulation parameters from biological data**

calculation of chromatin density in Arabidopsis nuclei

The volume density of all chromatin (the DNA and all attached proteins/nucleic acids) in Arabidopsis nuclei is unknown. However, a lower and upper limit for this density can be estimated. An upper limit derives from the observation that the nucleoplasm is a molecularly crowded solution containing 0.1-0.4 g/ml of macromolecules (Chubb & Bickmore, 2003). This would roughly translate to a 10-40% volume density of macromolecules, providing an upper limit to chromatin density. However, not all macromolecules in the nucleus are bound to chromatin.

A lower limit can be derived from estimations of known components of chromatin. DNA itself can be modelled as a cylinder of 300 Mbp long, 0.33 nm per nucleotide high and 2.4nm wide. This results in a total volume of 1.8 $\mu m^3$, or 2.7% of total nuclear volume assuming a spherical nucleus with a diameter of 5μm (volume 65 $\mu m^3$). In Arabidopsis chromatin, one nucleosome is associated with ~200 base-pairs of DNA. The molecular weight of each nucleotide is approximately 340 Dalton, and therefore 200bp of DNA weighs roughly 135KDa. These are associated with ~120KDa of core histones. Assuming that the specific weight of DNA and proteins is similar, the DNA-histone complex roughly has a double volume compared to naked DNA. Therefore chromatin occupies at least 5% of the nuclear volume. This lower estimate ignores many components of chromatin, such as other chromatin proteins and attached RNA and small molecules. As a compromise between the lower and upper limits, we use a density of 15 percent of the nuclear volume for all simulations.

**measurements of chromocenter size**

Using a Zeiss LSM 510 confocal microscope z-stack image series were produced of 45 Arabidopsis ecotype Colombia leaf mesophyll nuclei expressing H2B-YFP fusion protein (Willemse, 2006). Of these, 24 were of sufficient quality to be used for chromocenter size analysis. The nuclear volume occupied by euchromatin and heterochromatin fractions were measured based on a semi-automated thresholding algorithm. Pixels belonging to the nucleus were separated from background pixels by thresholding based on a smoothed background mask. The threshold levels used to distinguish between background, euchromatin and heterochromatin were set by a human operator. The relative volume of the heterochromatic fraction was determined by adding up the volumes represented by each pixel above the threshold and dividing by the total nuclear volume. To calculate the radius of chromocenters relative to total nuclear radius for use in simulations, this value was

divided by 10 (chromocenter number in simulations) and subsequently the cubic root was taken.

**calculation of nucleolus position and size**

In the same dataset as used for the measurement of chromocenter size, all images containing more than one nucleolus were discarded, leaving 39 images for analysis. Pixels belonging to the nucleus were separated from background pixels by thresholding based on a smoothed background mask. The threshold levels were determined by a human operator. The centre of volume of the nucleus was determined as the intersection point of the medial planes of the nucleus in x, y and z dimensions. Thresholding methods were unable to consistently determine the nucleolar radius and centre position because noise dominated the signal. Therefore the centre and diameter of the nucleolus were determined by a human operator using the measure function of the ImageJ software package in the z-stack image containing the median nucleolar section. The nucleolar radial position was measured as relative distance of the nuclear centre to the nucleolar centre compared to the distance from the nuclear centre to the nuclear periphery along the line running through both centres. The resulting relative nuclear eccentricity of the nucleolus was binned into 10 subsequent spherical shells of equal volume (taking into account the peripheral excluded volume).

**calculation of chromocenter clustering**

Distances were measured between each combination of 2 chromocenters. A randomly chosen initial chromocenter was assigned to the first cluster and all other chromocenters close (distances < 2.2 times the chromocenter radius) to the first chromocenter were assigned to its cluster. This analysis was repeated for each new chromocenter in the cluster. When no more chromocenters could be assigned to the first cluster, a so far unclustered chromocenter was picked and assigned to the next cluster, after which clustering proceeded in the same way as for the first cluster until all chromocenters were assigned to a cluster.

**mixing parameter calculation**

To quantify the amount of mixing in any single simulation state a mixing score was assigned to each monomer. The ten closest neighbours to the monomer were determined and the mixing score was equal to the amount of unique/different chains these ten neighbour monomers belonged to. A mixing value was obtained by averaging all monomer scores for the configuration, and to obtain the score for a model, sample configurations were drawn from a simulation at fixed MD time intervals after equilibration and their mixing scores averaged. Error bars show the standard deviation of the averaged mixing scores. To determine the range of MP values that can be encountered in simulation states, the mixing algorithm was applied to two situations representing the extremes of perfect mixing and no mixing. To simulate perfect mixing, all repulsive Lennard-Jones potentials between monomers were switched off, leaving only bond interactions, which in effect meant monomers belonging to different polymer chains did not interact at all. This resulted in an upper limit on the mixing parameter. Minimal mixing was achieved by assigning a 10 times

stronger repulsive potential between monomers of different chains than between monomers in the same chain (but with the same interaction range cut-off), enforcing the formation of (stable) chromosome territories. This resulted in a lower limit for the mixing value. Based on these values all scores are linearly re-normalised to a 0 to 1 range. In this way the mixing parameter becomes independent of monomer number. For alternative ways to quantify mixing, we refer to supplementary data.

## Results

**Computer simulations of interphase chromosomes**

To determine the contribution of non-specific interactions on the localization of heterochromatic regions in Arabidopsis, chromatin was modelled as polymer chains and implemented in simulations using the ESPResSo software package (Limbach *et al*., 2006). The chromosomes were modelled as 10 self-avoiding chains of monomers, with the relative lengths (in base pairs) of the 5 Arabidopsis chromosomes, and inserted into a confining sphere representing the nuclear envelope. Monomers were scaled to occupy 15% of the confinement volume based on a chromatin density estimation of 15 volume percent (see Methods). As a first approximation, the chromosomes were composed of identical monomers. As a compromise between resolution and performance, all simulations were performed at a resolution of 75Kbp per monomer. Since 75Kbp is larger than the experimentally observed persistence length of a 30 nm chromatin fiber (3-20Kbp, (Langowski, 2006)) there is no correlation of chain orientation over stretches of 75Kbp. We therefore connected adjacent monomers by simple harmonic springs without rotational constraints (figure 2.1, LC model). For a more detailed description of the simulation system and the way simulations were carried out, we refer to Methods.

Most of the heterochromatin of an Arabidopsis chromosome is present in the centromeric and flanking pericentromeric regions and form a compact chromocenter. The chromocenters were modelled as a single spherical monomer and positioned in the linear chain at the position of the centromere, resulting in the linear chromosome chromocenter (LCC) model (figure 2.1). Since chromocenters in Arabidopsis nuclei vary in shape from near spherical to elongated shapes, we measured their total volume relative to the total nuclear volume in intact leaf mesophyll cells using H2B-YFP expressing seedlings (Methods). In these nuclei chromocenters occupy 6.4% (+/- 0.02%, n=24) of the nuclear volume. Dividing this volume equally over 10 spherical chromocenters results in a chromocenter radius of 19% of the nuclear radius (calculation in Methods). The total amount of monomers was kept at 4000, implying that the euchromatin resolution slightly increased to 65Kbp/monomer).

We expanded the LCC model by including chromatin loops in two different ways (figure 2.1), based on previously described models of interphase chromatin organization. We

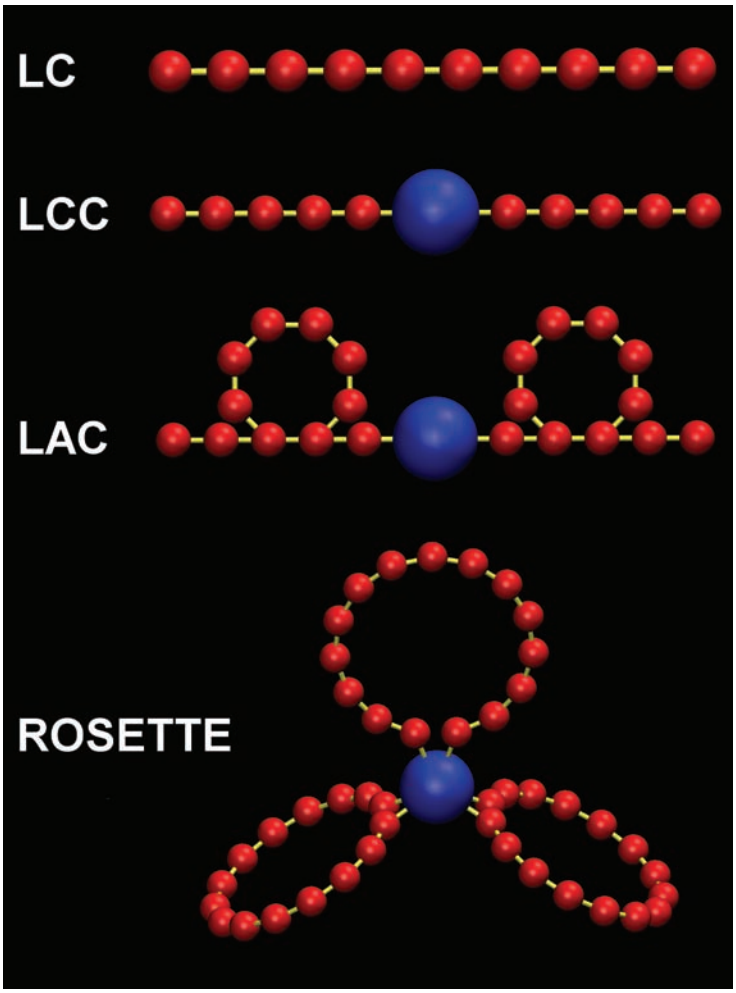



Figure 2.1: graphical depiction of the various models. The most simple model is the linear chain (LC) model, in which chromosomes are modelled as consisting of identical monomers (red) arranged in linear chains with harmonic spring potentials (yellow) connecting the monomers. The linear chains with chromocenters (LCC) model is almost identical to the LC model, but models the centromeric area of the chromosome as a large chromocenter (blue). An expansion of the LCC model is the looped arms with chromocenters model (LAC), in which the chains contain loops. In the rosette model (after Fransz *et al*. 2002) the chromosome arms loop out from a chromocenter in several loops. Chromocenters, monomers and bonds are not drawn according to scale.

thereby assume that loops are formed by a so far unknown mechanism that favours intra-chromosomal loops over inter-chromosomal loops. In Arabidopsis FISH data suggest that chromosomes contain euchromatic loops of ~0.1-1.5Mbp emanating from chromocenters (Fransz *et al*., 2002), which we implemented in simulations (the rosette model). In our model we initially set the loop size to ~1.5-2Mbp (in order to restrict the number of monomers needed to provide sufficient resolution in the loops) and included 15 loops per chromosome (figure 2.1). The terminal monomers of each loop are attached to the centre of the chromocenter by a harmonic spring potential, allowing these loop attachment points to slide over the chromocenter surface freely.

The simulations of the rosette model were compared with simulations of a model in which loops exist, but do not attach to chromocenters. Loops were introduced in the arms of chromosomes by adding additional harmonic spring potentials between monomers not in sequence, resulting in the loops and chromocenter (LAC) model (figure 2.1, Methods). Several variations have been tested: a model where every monomer is in a loop (LAC 100%), one where 50% of the monomers are incorporated into loops, and one in which only 10% of the monomers are incorporated into loops (see Methods). The loop size was again set to 1.5-2Mbp.

Time traces of the total energy, end-to end distance, contour length and mixing in a LC model simulation show that equilibration takes about $1.6 \cdot 10^4$ MD time units ($10^7$ integration cycles, supplementary data figure 2.1). The bell-shaped curves of fluctuations around mean values of these parameters after equilibration (supplementary data figure 2.1) and autocorrelation curves of these fluctuations (supplementary data figure 2.1) show that the simulations are properly equilibrated and that correlations in the fluctuations decay over a time span on the order of $10^2$ MD time units, short compared to total simulation time ($4.2 \cdot 10^4$ MD time units), thus allowing sufficient sampling for equilibration statistics.

We first examined how the various models behaved with respect to the formation of stable chromosome territories. Therefore we quantified mixing from a sample of equilibrated simulation states from each model by calculating a measure for mixing (mixing parameter, MP) from the frequency of interaction between monomers of different chains (see Methods), which are normalized on a scale from 0 (no mixing) to 1 (full mixing, supplementary data figure 2.2). In a simulation with linear chromosomes (LC model) at a 15% volume density, mixing occurs to a MP value of 0.7. Because of uncertainty in the actual *in vivo* chromatin density (Methods), simulations at different densities were also performed. These show that there is a positive relationship between chromatin density and mixing (supplementary data figure 2.3). Mixing already occurs to a MP value of 0.35 at 5% density, which represents an absolute minimal value for chromatin density *in vivo* (calculation in Methods) and MP reaches 0.9 at a density of 30% . Simulations of the LCC model showed that the addition of chromocenters has no significant effect on the mixing behaviour of the polymers (at 15%

density MP values are 0.67 and 0.68 respectively). However, the level of mixing of looped chromosomes is theoretically expected to be reduced (Cremer *et al*., 2000), because polymers that contain internal branches or loops are expected to mix less than linear polymers (Jun & Mulder, 2006). In our simulations, the introduction of loops reduced mixing dramatically: in simulations of the rosette model (15 loops) the chromosomes mix to a MP value of only 0.07. Variation of the number of loops in the rosette model showed that the level of mixing is dependent on the number of loops, but that even in simulations in which chromosomes had just 3 loops, the level of mixing was already reduced from a MP value of 0.7 to a value of 0.35. Simulations of the LAC model lead to similar results: an increasing percentage of monomers in loops leads to progressive decrease in the amount of mixing. The rosette and LAC 100% models, which have identical loop sizes and loop numbers, both almost completely prevent mixing (figure 2.2 first column, supplementary data figure 2.3).

**Chromocenter positions in simulations of the various models**

In Arabidopsis chromocenters preferentially are found within the nuclear periphery, and 2 or more chromocenters can be fused as between 6 to 10 spatially separated chromocenters are usually observed (Fang & Spector, 2005; Fransz *et al*., 2002; Berr & Schubert, 2007). The localization and fusion of chromocenters was analysed in simulations of models of the previous section (LCC, LAC, rosette models). Average radial positions of chromocenters were determined for each model, and distances between chromocenters were measured to determine the frequency of fusion (Methods).

In all 3 models the chromocenters preferentially localize to the periphery. However, the models differ in chromocenter distribution over the periphery. For the LCC model, this results in peripheral localisation of all chromocenters (figure 2.3). Clustering analysis (supplementary data figure 2.4) shows that most chromocenters cluster together in large groups of 3-6 chromocenters. On average about 4 clusters are present (figure 2.3).

A similar localisation and clustering is observed in simulations of the LAC models. Chromocenters in the LAC model mostly localise into one or two big peripheral clusters (figure 2.3, supplementary data figure 2.4).

In simulations of the rosette model, the radial chromocenter distribution is bimodal, with one large peak representing peripheral chromocenters and the other smaller peak an inner shell of chromocenters (figure 2.3). The preferred localization of chromocenters is on the periphery of the nucleus and chromocenters only localize more internally when the outer shell is filled with chromocenters and their associated loops (supplementary data, supplementary data table 1). Clustering analysis of simulations of the rosette models revealed that chromocenters with 5 or more loops do not associate with other chromocenters (figure 2.3), but chromocenters with 3 loops can occasionally form clusters of 2 chromocenters (data not shown).

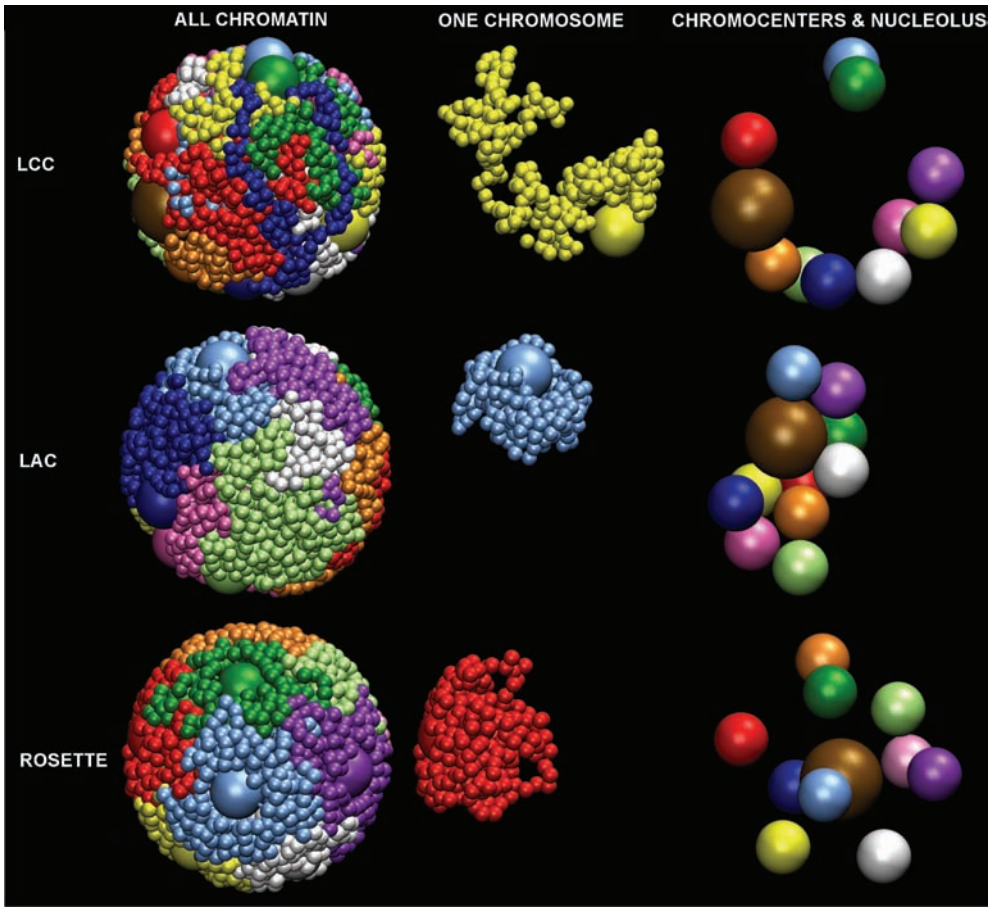



Figure 2.2: overview showing orthographically rendered snapshots of single configurations from LCC (top row), LAC (middle row) and rosette (bottom row) model simulations containing 10 chromosomes, with chromocenters, in different colours. Of each configuration 3 images are provided: left column shows all monomers in the simulation, middle column shows one chromosome, and the right column shows the localisation of chromocenters and nucleoli (brown) only.

While all models predict the radial localization of chromocenters correctly, the LCC and LAC models predict association of many or even all chromocenters (figure 2.2), thus reducing the number of chromocenter clusters to 4 per nucleus or even less. This is lower than the 6 to 10 chromocenters that are normally observed in Arabidopsis nuclei. The rosette model never shows any clustering (figure 2.2), while in most Arabidopsis nuclei some

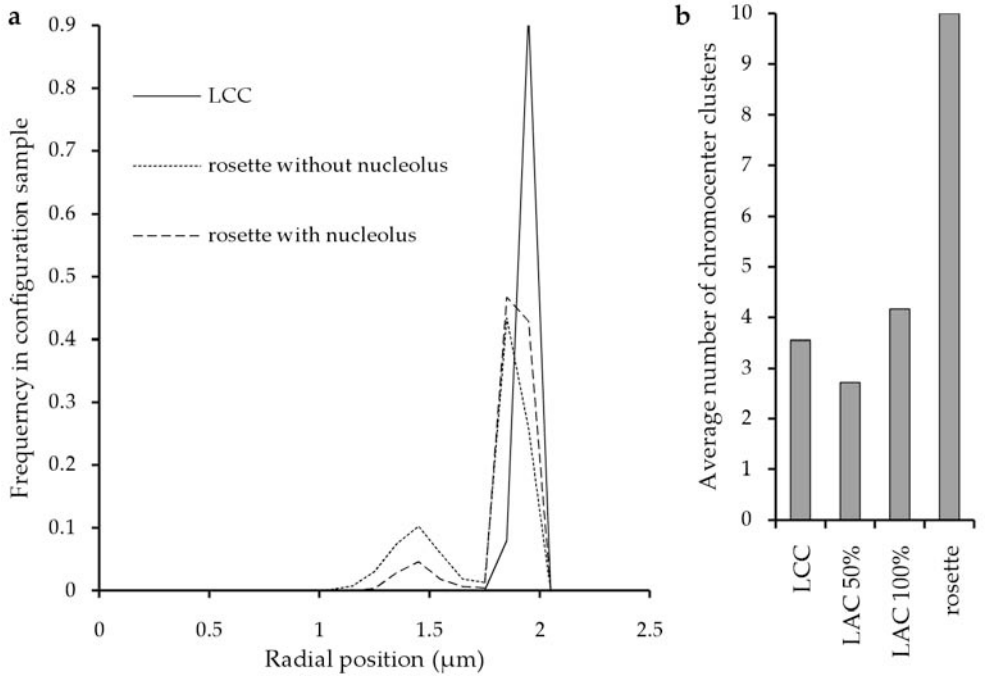



Figure 2.3: chromocenter localisation and clustering. a: histogram showing the fraction of chromocenters in each radial position bin in simulations implementing the LCC model and the rosette model with and without a 1.5μm nucleolus. b: Chromocenter clustering analysis on the simulations of figure 2.3a, and on simulations implementing LAC models.

clustering occurs. Furthermore, the inner shell of chromocenters that was found in simulations of the rosette model does not occur in Arabidopsis nuclei. So additional parameters must affect the behaviour of chromocenters. Since the nucleolus forms a large excluded volume in nuclei and therefore could affect the positioning of chromocenters, we included it in the simulations.

**Effect of nucleolus on chromocenter positions**

The nucleolus is the most conspicuous sub-nuclear structure. Due to its size, the nucleolus represents a significant excluded volume within the nucleus. Therefore it is to be expected that the nucleolar position is determined by non-specific interactions and that the presence of a nucleolus influences the localisation of chromocenters. To test this, we measured nucleolar positions and sizes in Arabidopsis nuclei and compared the results with those obtained in simulations.

To visualise the nucleolus *in vivo*, Arabidopsis plants expressing a H2B-YFP construct were used (Willemse, 2006). Whole, living seedlings were observed using a confocal microscope. In z-stacks made of nuclei of leaf mesophyll cells the nucleolus can readily be observed as a spherical region of low fluorescence within the nucleus. Nuclei with more than one nucleolus (which occurred in ~10-20% of nuclei) were omitted from the data analysis.

To determine the nuclear/nucleolar radius and nucleolar radial position, a threshold-based automated approach was adopted (see Methods). This analysis revealed that in Arabidopsis leaf mesophyll nuclei the nucleolus/nucleus radius is 0.30 (+/- 0.05, n=39). The nucleolus was found to be localised in most cases at or near the centre of the nucleus (figure 2.4).

Models were designed to simulate the behaviour of the nucleolus *in silico*. A nucleolus was added to the rosette (15 loops) model as a sphere of 0.30 times the nuclear diameter. As a first approach the nucleolus was not attached to the chromatin and this resulted in a central position of the nucleolus in the nucleus (figure 2.2). In addition, in the presence of a nucleolus more chromocenters localise to the periphery compared to simulations of the rosette model without a nucleolus (figure 2.3). The small amount of internally localised chromocenters when a nucleolus is present, is due to chromocenters that remain there only for short periods of time during the simulation. Clustering of chromocenters does not take place (figure 2.3).

In the same way as for the rosette model, a nucleolus was introduced into the LCC and LAC 100% models. In these simulations the nucleolus localises to the nuclear periphery (figure 2.2, figure 2.4). So only the rosette model is consistent with the central nucleolar position observed *in vivo*.

In Arabidopsis, the nucleolus organizing regions (NORs) are located at the ends of the short arms of chromosomes 2 and 4. Since usually 2 to 3 of the NORs are associated with the nucleolus (Fransz *et al*., 2002), we tested the effect of attaching the monomers at the ends of all NOR containing arms to the nucleolus on the localization of nucleolus and chromocenters in the rosette model. This did neither influence the localization of chromocenters nor that of the nucleolus (data not shown).

In FISH studies telomeres were found exclusively localized around the Arabidopsis nucleolus (Fransz *et al*., 2002). In the LCC/LAC and rosette models, the terminal monomers do not have a preferred localization (data not shown). Hypothesizing that the *in vivo* localization is caused by physical interaction between telomeres and the nucleolus, we attached all chromosome ends to the nucleolus in the rosette model. This again did neither affect the localization of the nucleolus nor the chromocenters.

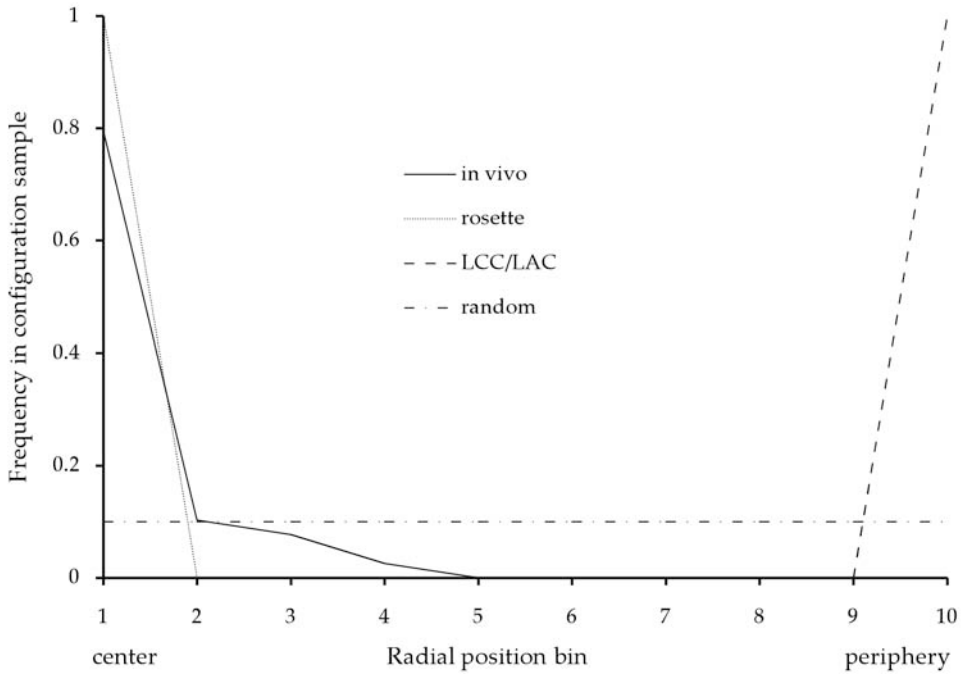



Figure 2.4: nucleolus positions in Arabidopsis. On the horizontal axis nucleolus position bins are shown each representing a shell of 0.1 times the nuclear volume available to the nucleolus. On the vertical axis the fraction of nucleoli in each bin is shown. The solid line shows measured positions in Arabidopsis mesophyl nuclei, the dotted line shows the prediction derived from the rosette model, the dashed line shows predictions from the LAC model, and the line with alternating dashes and dots shows random localisation (assuming the nucleolus to have an equal chance to localise to every available position in the nucleus).

## Discussion

By implementing models for nuclear organisation in molecular dynamics simulations, we show that non-specific interactions are sufficient to explain the peripheral localisation of heterochromatic chromocenters and central localisation of nucleoli in interphase Arabidopsis nuclei. It is therefore not necessary to explain this localisation through specific interactions such as between heterochromatin proteins and the nuclear envelope or lamina (Baricheva *et al*., 1996; Taddei *et al*., 2004) and the involvement of nucleolar proteins in a nuclear matrix (Calikovski *et al*., 2003). Interestingly, chromocenters localise peripherally regardless of the way euchromatin arms are represented: as linear chains attached to the

chromocenter (LCC model), chains containing loops (LAC model), or as loops emanating from the chromocenter (rosette model). This implies that the peripheral chromocenter localisation of other species than Arabidopsis could be explained through depletion interactions even if their chromosomes do not form rosette structures. Also, even in species or cell types where heterochromatin does not form chromocenters, heterochromatin is still more condensed and less dynamic than euchromatin. Such a difference in structure could lead to depletion attraction between the less dynamic heterochromatin and the nuclear periphery or the nucleolus, leading to spatial separation of eu- and heterochromatin such as observed in for instance human nuclei.

In animals many interactions between lamina proteins and heterochromatin proteins have been discovered (Taddei *et al*., 2004). However, the existence of these interactions does not imply that they are the direct or the sole cause of peripheral heterochromatin localisation. Here, we show that peripheral heterochromatin localisation is expected to occur even without the presence of any lamina proteins. Interactions between lamina proteins and heterochromatin may further stabilise this localisation in animals and may be important to keep heterochromatin away from nucleopores, where active transcription occurs (Taddei, 2007).

Predictions of the nucleolar localisation in simulations based on the rosette model fit more closely to those observed *in vivo* than in simulations of the LAC model. So an Arabidopsis model of chromatin organization in which euchromatic loops extend from heterochromatic chromocenters (rosette model) has sufficient self organizing potential, whereas an organization in which loops do not attach to chromocenters (LAC) would require specific interactions for nucleolus positioning.

However, neither model can fully explain the observed clustering of chromocenters. Clustering of chromocenters could be expected based on depletion interactions, because it would result in an entropic gain. In the LAC model, that is exactly what happens. However, in the rosette model clustering of chromocenters does not occur because the emanating loops prevent depletion interactions between chromocenters. The *in vivo* situation, in which chromocenters cluster to a limited extent and form approximately 6 to 10 spatially separate structures, is somewhere in between the extremes predicted by our models. Limited clustering is observed in rosette models that include few chromocenter-emanating loops (supplementary data figure 2.2). While the loop sizes in these simulations (~5Mbp) are larger than observed *in vivo* (1.5Mbp; Fransz *et al*., 2002), DNA is more flexible than our simulated polymers and therefore might permit some limited association between chromocenters surrounded by loops of sizes as observed *in vivo*.

Alternatively, the *in vivo* clustering could be explained by hypothesizing that limited chromocenter clustering occurs when the rosette structure is not fully present, for instance

during chromosome decondensation after mitosis and during major developmental cell fate changes. Chromocenter clustering in many species is a dynamic process associated with cell differentiation and other major developmental cell fate changes (Ceccarelli *et al*., 1997). In Arabidopsis, chromocenters partially decondense during floral transition (Tessadori *et al*., 2007a) and upon dedifferentiation after protoplastation (Tessadori *et al*., 2007b). Presumably rosette loops also dissociate when chromocenters decondense. Our results on the clustering of rosette structures with few loops show that shielding becomes less efficient when fewer loops are present and that in the case of 3 loops per chromocenter, chromocenters can cluster. Therefore rosette structure disruption (loop release) during major cell fate switches might allow chromocenters to cluster.

Here we show that chromatin loops also have a major effect on formation of separated chromosome territories, as was predicted in (Münkel *et al*., 1999). We have provided quantitative proof that in the absence of chromatin loops, chromosomes mix. The introduction of loops in the mega-base size range proved to be sufficient for chromosome territory stability in our simulations. Two parameters determine the amount of mixing in looped chromatin: average loop size and the percentage of monomers in loops. Both the rosette and LAC models achieve similar MP values when loop amounts and sizes are the same (15 loops per chromosome / loop size ~1.5-2Mbp), indicating that the MP value is not dependent on a specific loop model. This indicates that preventing mixing by loop formation may be a universal mechanism irrespective of how loops are formed in a certain species. Recent work has provided computational evidence that the inability of human chromosomes (represented as a 30nm fibre) to fully mix is caused by the slow relaxation kinetics of very long entangled polymers involved (Rosa & Everaers, 2008). This would imply that chromosome territory formation could in fact also be explained by kinetic arguments, rather than being a consequence of equilibrium statistics as we argue is the case here. In that respect it should be pointed out that not only are the Arabidopsis chromosomes significantly shorter than human ones, they also have, due to the presence of the chromocenters, a markedly different topology. In this case one expects the relaxation time to be dominated by the size of the longest loop present, which is at least an order of magnitude smaller than the total chromosome length. Due to the very strong length dependence of the relaxation time (~ L3) this would already imply a reduction in time-scales of three orders of magnitude. Moreover, our chromosomes are confined to a nuclear volume with a radius comparable in size to the radii of gyration of the chromosome arms/loops involved. There is strong evidence that under these conditions polymer dynamics is significantly enhanced with respect to bulk behaviour (Shin *et al*., 2007), which was not taken into account in (Münkel *et al*., 1999).

Some recent data indicates that territory segregation is not complete (Branco & Pombo, 2006). Our models predict different amounts of mixing depending on factors such as density and loop architecture, and quantitative data on the level of mixing *in vivo* could be used to

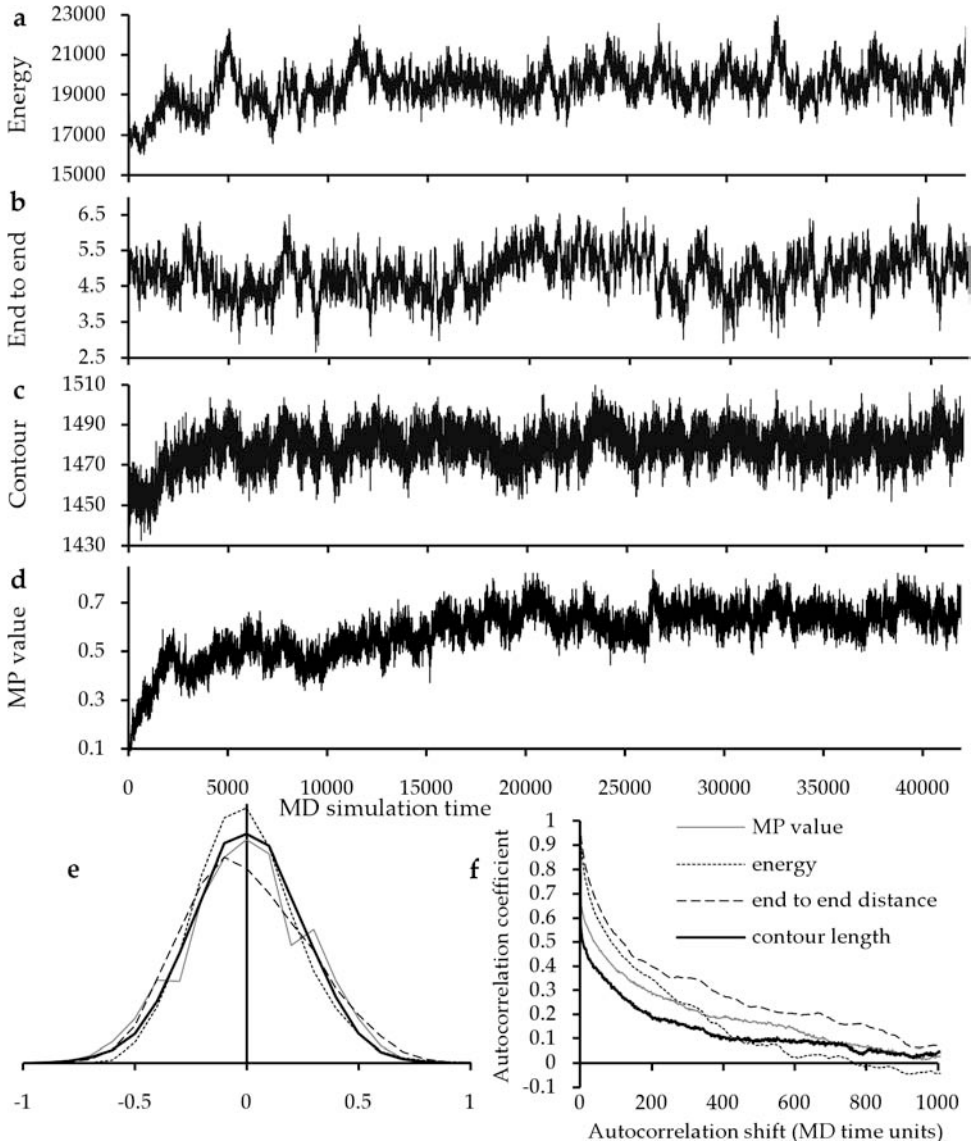
test them. However, unfortunately such data are not yet available for Arabidopsis. Still, our simulation methods provide a new way to test models of large-scale chromatin organisation against biological data and lead to new insights about the effects of large-scale chromatin looping. For instance, it has to our knowledge not been suggested before that chromocenter-emanating loops act as barriers between chromocenters, preventing their coagulation. Our methods can be applied to the chromatin organisation of other species, although resolution problems may occur with genomes of larger size than Arabidopsis. To overcome this limitation and to improve the speed of simulations of Arabidopsis nuclei, a more coarse-grained potential to represent larger chromatin chain sub-domains could be developed (see e.g. Hansen *et al*., 2005).

## Supplementary materials

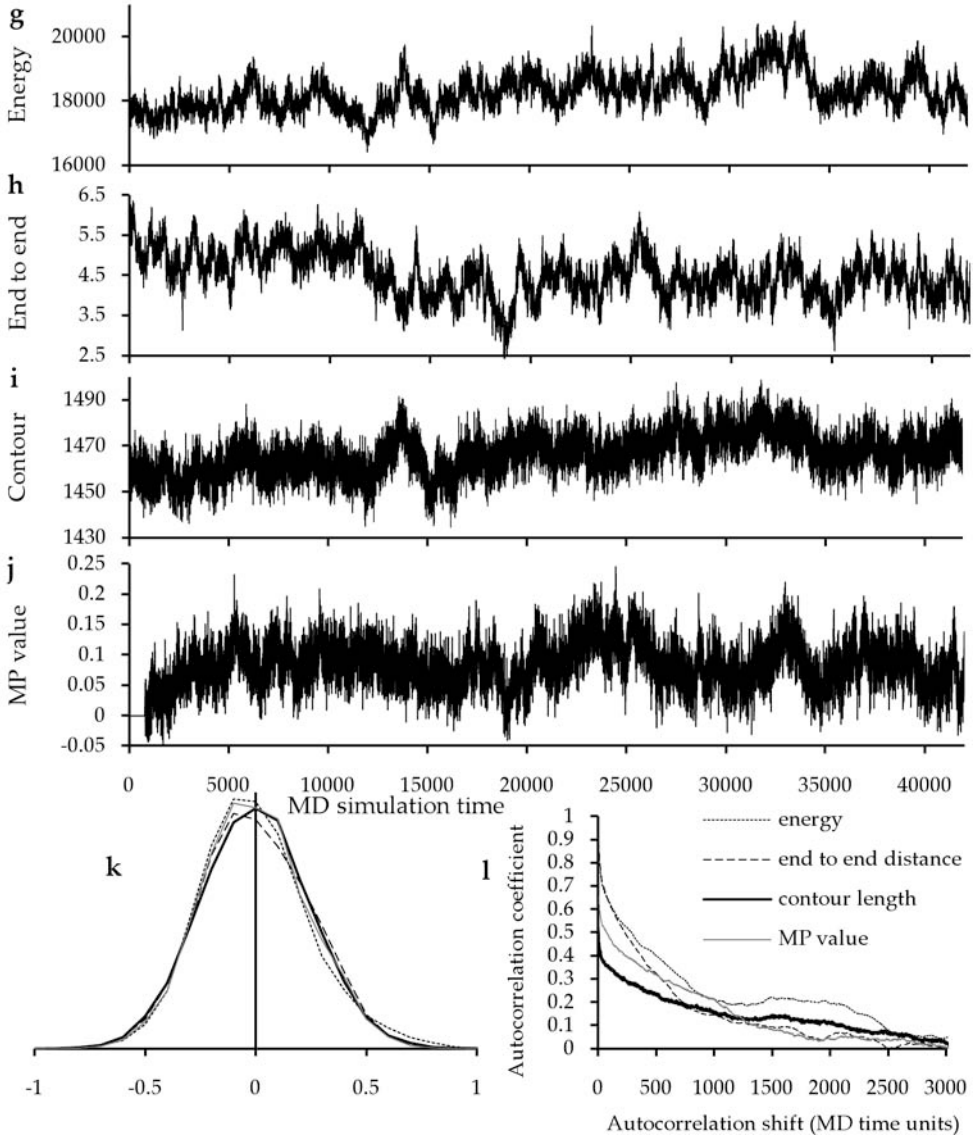
**Mixing algorithm**

An algorithm was developed to quantify the amount of mixing in a simulation state. While methods based on measurable polymer properties such as the radius of gyration of the polymers or end-to-end distances can be used for determining to what extent polymers mix, statistical noise in these measurements dominated the mixing signal. To reduce statistical noise a more thorough sampling of the simulation was obtained by monomer-based methods. Initially the nucleus was divided in grid cells and for each grid cell the amount of chains present inside was determined. However, in this method the amount of monomers per grid cell can vary when density differences are present, leading to sample bias. Since density variations in the nucleus occur due to monomer depletion at the periphery of the nucleus, we applied the method used in this paper. By sampling a fixed number of closest monomers to each monomer, we automatically compensate for density differences. Sampling bias still occurs for monomers at the nuclear periphery or next to large structures such as the nucleolus, since there the presence of excluded volume leads to longer-range samples and therefore potentially to a higher mixing value. To minimise this potential bias the amount of neighbour monomers included in the calculation must be kept as low as possible. However, a lower limit on the amount of sampled monomers follows from the consideration that the closest monomers to any monomer are almost always the ones in the same chain. Therefore it is essential to include more than at least 4 monomers to see monomers belonging to different chains at all. In practice, including the 10 neighbouring monomers resulted in a good compromise between these conflicting demands.

The exact value of the mixing parameter is dependent on the amount of monomers, since in the case where polymers occur in territories an increase in chain length leads to a larger fraction of monomers inside a territory and therefore to a lower mixing parameter value. Therefore it is essential to only compare configurations that have the same amount of monomers.




Supplementary figure 2.1: equilibration of a simulation with linear chains (LC model). a-d: development of parameters in time, including equilibration. a: total energy. b: average displacement length. c: average contour length. d: mixing parameter. e: distributions of the normalized fluctuation of parameters around their average after equilibration (equilibration cut-off at 16000 MD time units). f: autocorrelation of these fluctuations at different distribution shifts.



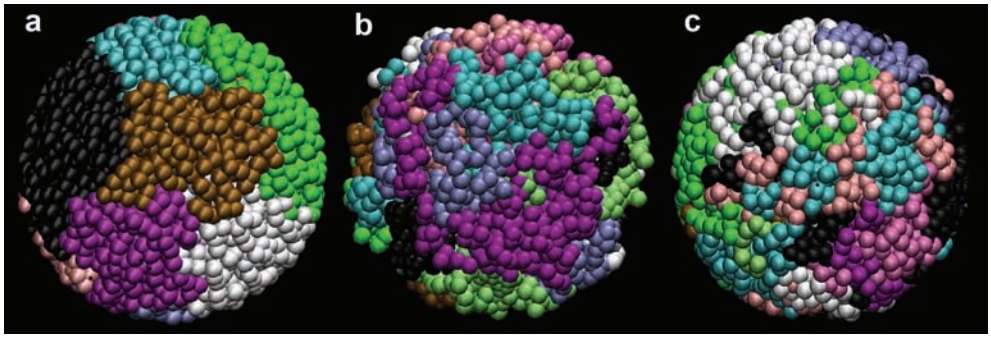

g-l: same as a-f, but equilibration of simulations implementing LAC 100% model.

**Chromocenter shell distribution in rosette model**

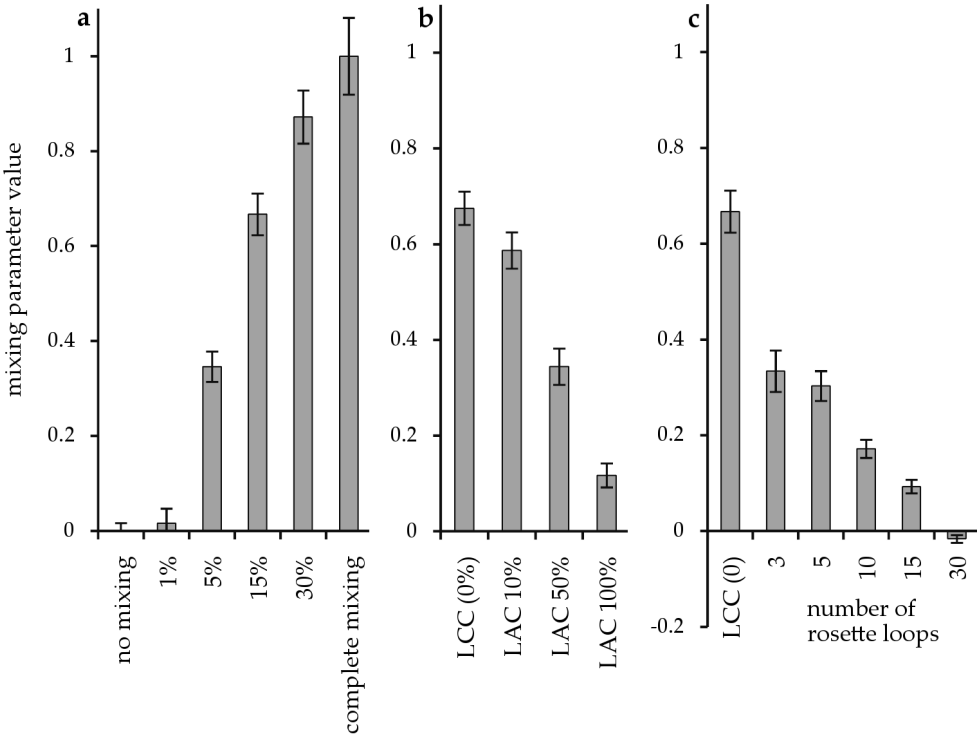
To study the preferred localization of chromocenters in the rosette model, we examined the effect the number of chromocenters has on the distribution of chromocenters over these shells. The amount of chromosomes in the nucleus was varied by randomly adding or removing chromosomes without changing the density of the monomers in the simulations. The results in supplementary table 1 show that in simulations with less than 8 chromosomes, all chromocenters localize to the periphery of the nucleus most of the time, although occasionally individual chromocenters move to the inside for short periods of time. When more chromocenters are included, the outer shell fills up completely due to steric hindrance between loops on different chromocenters and an inner shell directly underneath the outer chromocenters fills up with the remaining chromocenters. Therefore we conclude that the preferred localisation of chromocenters is peripheral and only when the periphery is filled up completely, chromocenters start to localise to more internal positions.

| Number of Chromosomes | Peripheral shell | Inner shell |
|---|---|---|
| 4 | 0.90 | 0.10 |
| 6 | 0.92 | 0.08 |
| 8 | 0.81 | 0.19 |
| 10 | 0.69 | 0.31 |
| 12 | 0.66 | 0.34 |
| 16 | 0.59 | 0.41 |

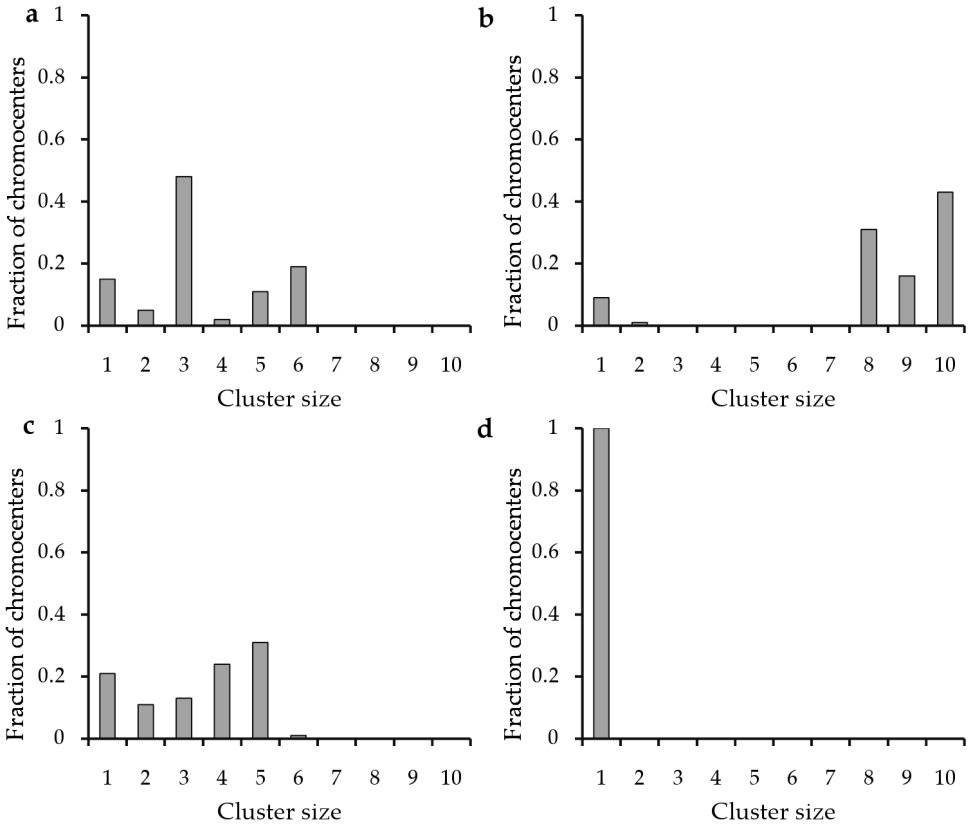
Supplementary table 1: distribution of chromocenters over shells in rosette models with varying chromosome numbers (without nucleolus). Note: numbers in second and third column refer to the fraction of chromocenters in each shell.




Supplementary figure 2.2: snapshots of chromosome distributions at different MP values. All images are front view orthographic projections of the spherical nucleus and contain 4000 monomers divided over 10 linear chains (each a different colour) with relative lengths as in Methods. a: no mixing enforced MP value of 0. b: 5% density MP value of 0.34. c: MP value of 0.65. Images were generated using the VMD software.




Supplementary figure 2.3: Dependence of MP value on chromatin structure. a shows the average MP value of confined LCC polymer chains at several densities ranging from 1 volume percent to 30 volume percent. Also shown is a control in which polymer territory formation was enforced by a strong repulsive potential between monomers belonging to different chains, representing no mixing, and a control in which repulsive non-bonded monomer interactions were completely turned off to simulate perfect mixing. In both controls the density was set to 15 percent, and in all simulations the parameters were set according to Methods section. b shows influence of loop formation on MP value in LCC model and LAC models with 10, 50 and 100% of monomers in loops. c shows the effect of various amounts of rosette loops per chromocenter on MP value.




Supplementary figure 2.4: detailed clustering analysis graphs showing the average fractions of chromocenters in clusters of different sizes. a. LCC model. b. LAC 50% model. c. LAC 100% model. d. rosette model (15 loops per chromocenter)

# 3. Effective interaction potentials for Arabidopsis chromosomes

S. de Nooijer[1], A. Bailey[2], J. Wellink[1], T. Bisseling[1], B. Mulder[2,3]

1: Laboratory for Molecular Biology, Wageningen University, Droevendaalsesteeg 1, Wageningen, The Netherlands
2: FOM institute for atomic and molecular physics (AMOLF), Science Park 104, Amsterdam, The Netherlands
3: Laboratory for Plant Cell Biology, Wageningen University, Droevendaalsesteeg 1, Wageningen, The Netherlands

## Abstract

Unravelling the role of entropic forces in shaping the chromosomal architecture within the nucleus of eukaryotic cells requires large-scale computer simulations. To date these simulations are typically based on polymer representations of the chromosomes, which, although strongly coarse grained with respect to the atomistic level, still involve $10^2$ −$10^3$ beads per chromosome. The associated computational effort severely limits the possibilities of exploring the multi-parameter phase space of nuclear organization, involving e.g. changes in chromosome density, chromosome number, and chromosome topology. Here we explore the use of simulation-derived effective pair potentials in addressing these limitations. We derive effective intra-chromosomal potentials including the effect of confinement to a nucleus with a radius comparable to the radius of gyration of the model chromosomes. We apply our approach to specific case of the nucleus of the model plant species Arabidopsis, which possesses well-characterized chromosomes, consisting of a dense heterochromatic core, a so-called 'chromocenter', surrounded by euchromatin loops. We specifically investigate the forces involved in chromocenter localization, nucleolus localization and the influence of the number of euchromatic loops, comparing where possible to extant biological data.

## Introduction

*Arabidopsis thaliana* is an interesting system for the study of nuclear organization because of its small (150Mbp), fully sequenced genome and low repetitive sequence content of ~15% (Arabidopsis genome initiative, 2000). During interphase these repetitive sequences associate into 6 to 10 chromocenters, which are localized at the nuclear periphery (Fang and Spector, 2005). Fluorescent in situ hybridisation studies have shown that euchromatin emanates from the chromocenters in loops of 0.1-1.5Mbp (Fransz *et al*. 2002). Based on these studies a model has been proposed for interphase organization where sequences from the euchromatic arms associate with the chromocenter, creating a rosette-like structure with approximately 15-50 loops emanating from each chromocenter (Fransz *et al*. 2002). In chapter 2 (de Nooijer *et al*., 2009), we performed MD simulations implementing chromosomes folded into rosettes, to explore the consequences of the rosette structure for nuclear organization. In these simulations the dispersed localization of chromocenters and central localization of nucleoli, which occurs in Arabidopsis nuclei, was reproduced. We hypothesized that chromocenters repel each other through interactions between their associated rosette loops, but the simulations of chapter 2 did not provide further understanding of these forces. Therefore, we here derive effective potentials for the interactions between rosettes, nucleoli and the nuclear periphery. The shapes of these potentials then provide insight into the forces driving nuclear organization.

The most important structures in interphase nuclei are the chromosomes, which occupy the bulk of the nucleus. Due to their considerable length, approximately $3*10^7$ bp in Arabidopsis, chromosomes could above a certain length scale behave as classical excluded volume polymers. The magnitude of this appropriate coarse graining length-scale is as yet unknown, but a probably safe lower bound can be given in the commonly accepted persistence length of the 30 nm fibre being 100-200 nm and comprising 3-20Kbp (Langowski, 2006). This uncertainty in the size of the coarse-grained "monomers" is not a major obstacle in elucidating large-scale polymer behaviour of chromosomes, as statistical physics tells us that polymers of sufficient length relative to the monomer size possess universal properties that obey well-understood scaling laws and are to a large extent independent on the precise physical and chemical composition of the chain. Accepting that a chromosome behaves as a classical polymer implies that entropic effects may be important in shaping the folding of chromosomes both individually and collectively within the confines of a bacterium or a eukaryotic cell nucleus. Over the past few years this idea, which hitherto did not play a role in the relevant biological literature, has led to new insights in human chromosome folding (Munkel *et al*., 1999), bacterial chromosome segregation (Jun & Mulder, 2006), the sub-nuclear localization of more and less-expressed human chromosomes (Cook & Marenduzzo, 2009) and the chromosomal arrangements in Arabidopsis (chapter 2). In all these examples,

computer simulations, either of the Monte Carlo (MC) or (hybrid) Molecular Dynamics (MD) type, were used as the primary research tool, because the combination of multiple self-avoiding chains and spatial confinement is beyond the reach of present day analytical theories.

However, given the need for sufficient resolution of the polymeric degrees of freedom and the numbers of chromosomes concerned, these simulations require significant amounts of computer time. Our earlier work (chapter 2) on the nuclear organization of Arabidopsis involved 10 chromosomes each represented by approximately $10^3$ monomers. A typical simulation encompassing both an initial equilibration phase and a sufficient data collection phase, would take up to two weeks wall-clock time on current PC hardware. Ideally, in order to understand the dependence of nuclear organization on chromosome folding, one would like to vary the nuclear dimensions, the number of chromosomes, add other nuclear constituents of varying numbers and sizes such as the nucleolus and Cajal bodies, and change structural parameters of the chromosomes. But computational time requirements effectively preclude such a wide exploration of the systems in question. And even if extensive series of simulations were computationally feasible, they would not directly provide quantitative insight into the entropic forces driving nuclear organization.

A solution to both these problems can be provided by constructing effective pair potentials between the chromosomes as a whole. Such effective potentials reduce the complexities of the inter-polymer interactions to a single distance dependent potential between the polymer centres of mass, and have been used extensively in the study of polymeric systems (see e.g. Bolhuis and Louis, 2002). Here we show how effective potentials can be constructed that take into account both chromosome structure and spatial confinement. In view of the latter aim we have introduced additional effective interactions with a curved hard wall, representing the nuclear periphery. Side-by-side comparisons between simulations implementing effective potentials and simulations using full polymeric simulations, allow us to gauge the validity of the approach. Apart from speeding up the simulations by four orders of magnitude, the shape of the effective potentials themselves also provides a direct higher level insight into the effects of chromosome folding. The dependencies of this shape e.g. on the sizes of nuclear constituents has predictive value in terms of the expected nuclear organization. In this way, we quantify the previously discovered (chapter 2) interactions between rosettes, nucleoli and the nuclear periphery, and explain how the interplay between these forces leads to the observed nuclear organization.

## Methods

**The rosette model for chromosome folding**

Chromosomes are modelled as loops of bead chains attached on both ends to a central larger bead representing the chromocenter. Each bead represents ~70Kbp of chromatin. Both beads and chromocenters interact with each other and with the spherical confinement, representing the nuclear periphery, using pairwise WCA potentials (Weeks *et al*., 1971), of which the interaction ranges are set such that beads have an effective radius (s) of 84nm and chromocenters a radius (S) of 500nm, unless otherwise indicated. Beads are assembled into chains using harmonic spring interactions between their centres of mass, with an equilibrium range of 2s. Chromocenters then have 15 associated loops, each consisting of 26 beads. The ends of these chains are attached to the chromocenter using a harmonic spring interaction with an equilibrium range of s+S.

**Setup and equilibration of simulations involving chromosome-representing polymers**

10 identical rosettes were placed together in a 2500nm radius spherical confinement, using the rosette construction and equilibration methods described in chapter 2, and the ESPResSo simulation software package (Limbach *et al*., 2006). Some simulations include a single bead that only interacts with other beads through WCA interactions, representing the nucleolus or other smaller nuclear sub-compartments. After equilibration for $5*10^3$ MD time units (chapter 2, chapter 2 supplementary data), simulations were performed for at least $10^4$ MD time units (see chapter 2). Samples were taken every 4 time units, based on which parameters of interest were calculated.

**Setup and equilibration of simulations used to generate configurations upon which effective interaction potentials were calculated.**

Using ESPResSo, a single rosette such as described above is placed in periodic boundary conditions, with sufficient space to prevent the polymer from encountering itself in another period (dilute conditions). Then, the rosette is equilibrated for at least $1.5*10^4$ MD time units. The resulting equilibrium configurations are used for effective potential calculations (see results).

**Setup and equilibration of simulations implementing effective potentials**

Using ESPResSo, 10 point particles with associated tabulated effective interaction potentials were placed at random positions in a R=2500nm confinement. In simulations implementing a nucleolus, a single point particle with appropriate potentials was added. Equilibration occurred almost instantaneously, after which the simulation was performed for at least $2*10^5$ MD time units. Sample configurations were stored every 40 MD time units.

## Results

**Construction of effective inter-chromosomal potentials**

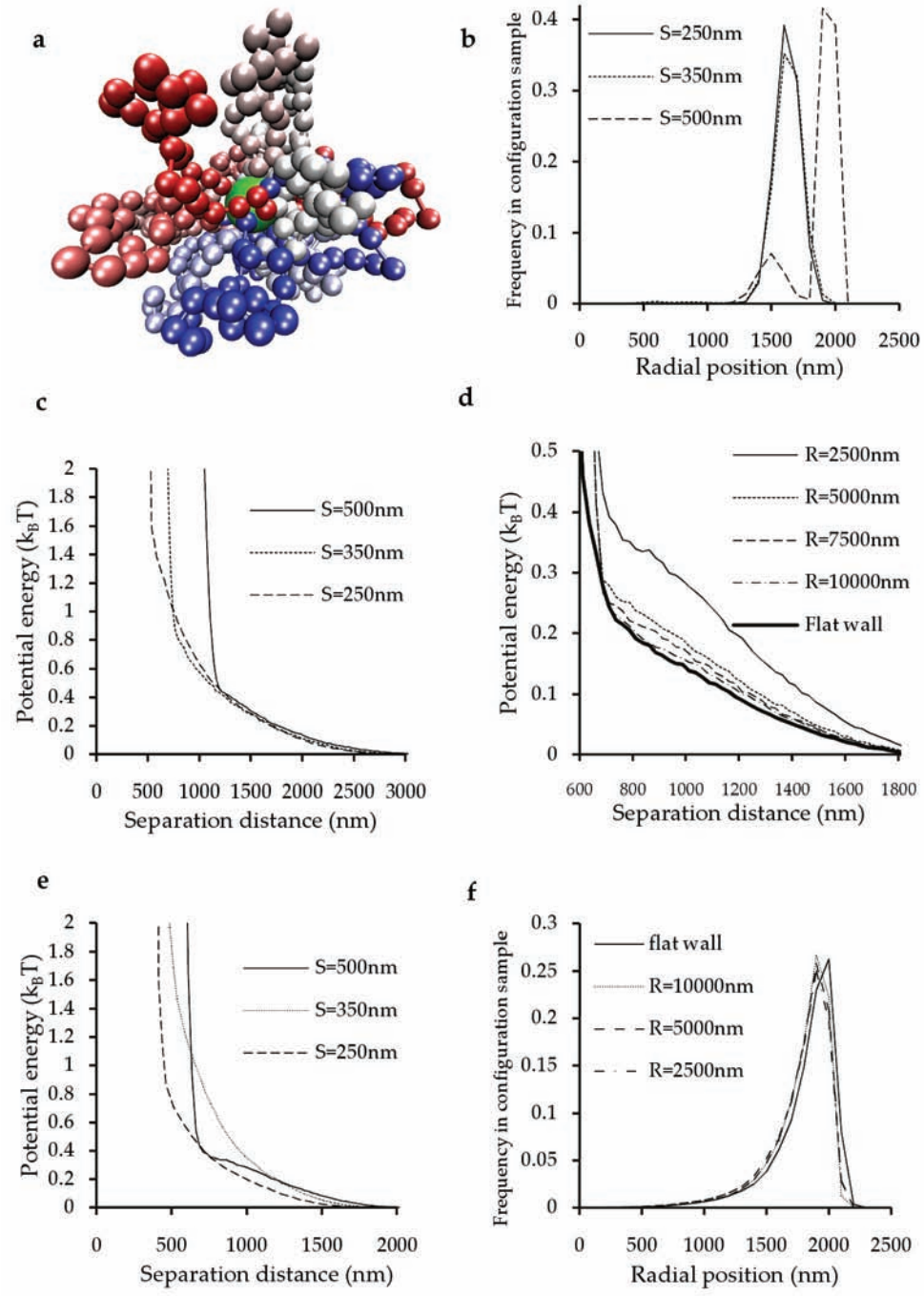
To get more insight in the effects of parameter variations such as the sizes of chromocenters and nucleoli on the entropic forces driving nuclear organization, we describe the Arabidopsis rosette model with the help of effective potentials. An effective potential, or potential of mean force (PMF), is the average force between two entities – such as two model chromosomes – sampled over all possible configurations. Effective potentials are given by

$$U(r) = -k_B T \ln g(r) + \text{const} \qquad \text{(formula 3.1)}$$

where U(r) is the PMF; r the reaction coordinate; $k_B$ is Boltzmann's constant; T the temperature; and g(r) the radial distribution function (Frenkel & Smit, 1995). The PMF can of course only be determined up to a constant value (const).

To describe the full rosette model in terms of effective interactions, we require three different effective potentials: chromosome-chromosome, chromosome-nuclear membrane, and chromosome-nucleolus. The reaction coordinate for each, r, is chosen to simply be the separation distance. For all cases r is measured from the centre of the chromocenter, and extends to the centre of a second chromocenter, to the edge of the confinement membrane, or to the centre of the nucleolus, respectively.

We first make the simplifying assumption that the intra-chromosomal potentials calculated in the zero-density limit can be used for a simulation set in a nucleus, i.e. a crowded environment (see discussion). Using Espresso, we equilibrate a solitary rosette such as described above in an infinite space. After equilibration, repeatedly sampled configurations are then used for calculation of the PMF. Formula 3.1 shows that to arrive at U(r), we must calculate g(r). Calculating g(r) is problematic for small r, because the PMF diverges strongly when chromocenters approach each other. Thus the energy landscape is difficult to sample efficiently because rosettes rarely come into contact, let alone ascend the diverging PMF to achieve short separation distances. To efficiently sample the free energy landscape in this regime, we therefore used umbrella sampling (Torrie & Valeau, 1977), which adds a "biasing potential" to the system in order to bias a reaction coordinate to be within a certain range, specifically one that is otherwise rarely sampled. In this way, one is forcing the system to reside in a specific area of the free energy landscape. The biasing effect of the additional potential is afterwards removed to calculate the free energy differences in its absence (Frenkel & Smit, 1995).


a
b
Frequency in configuration sample
S=250nm
S=350nm
S=500nm
Radial position (nm)
c
Potential energy ($k_BT$)
S=500nm
S=350nm
S=250nm
Separation distance (nm)
d
Potential energy ($k_BT$)
R=2500nm
R=5000nm
R=7500nm
R=10000nm
Flat wall
Separation distance (nm)
e
Potential energy ($k_BT$)
S=500nm
S=350nm
S=250nm
Separation distance (nm)
f
Frequency in configuration sample
flat wall
R=10000nm
R=5000nm
R=2500nm
Radial position (nm)

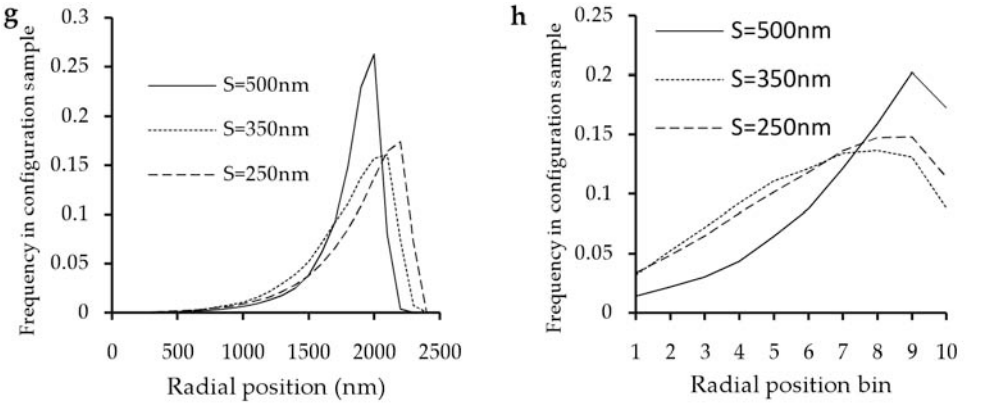



Figure 3.1: nuclear localization of chromocenters. a: Perspective rendering of an equilibrated rosette used for effective potential calculations. Here, chromocenter (large green sphere) radius is 500nm, and the monomers have a radius of 84nm. Monomers are coloured based on their position in the chromosome, ranging from red for the first monomer in the first loop through white to blue for the last monomer in the last loop. Thus the rosette loops can be easily discerned. b: Radial positions of chromocenters in sample configurations from MD simulations, implementing 10 rosettes as depicted in a, confined in a sphere of 2500nm radius. c: effective potentials for the forces operating between the chromocenters of two rosettes as a function of separation distance. Shown are potentials for rosettes containing chromocenters of three different sizes. d: effective potentials for the forces operating between a rosette of size S=500nm and various confinements, both flat and curved with different radii. e: effective potentials for the forces operating between rosettes containing chromocenters of different sizes, and a curved confinement with R=2500nm. f-h: results of the implementation of the derived effective potentials in simulations. f: the effect of wall curvature on average radial chromocenter positions. g: the radial distributions of chromocenters of different sizes. h: same as g, but with chromocenter positions binned to ten concentric bins of equal volume.

Here, we chose the biasing potential to be an harmonic spring with a spring constant k and equilibrium separation $r_0$. For each potential a set of simulations are carried out such that the biasing potential of each simulation is tuned to sample a specific range of the reaction coordinate. One can achieve this by varying $r_0$ and k. The output from each simulation is a free energy curve that extends along only a fraction of the full range of the reaction coordinate. The free energy curves for the separate simulations together cover the entire range of interest. They can then be combined to form the full free energy curve using a method such as self-consistent histogram method (Bennett, 1976). The resulting curves can be implemented as tabulated potentials in Espresso MD simulations.

**Effective potentials for rosette interactions**

Chromocenters preferentially localize to the periphery in real nuclei (Fang & Spector, 2005) as well as in simulations implementing chromosomes as rosettes, consisting of chromocenters with a radius of 500nm from which euchromatin emanates in 15 loops (chapter 2, figure 3.1a). To accommodate the 30 beads forming the attachment points between the loops and the chromocenter, the surface area of the sphere surrounding the chromocenter at a radius of S+s needs to be bigger than approximately $30\pi s2$ . Solving for S results in a minimum chromocenter radius of 156nm when an optimal hexagonal packing of spheres is assumed. However, the density of chromatin around the chromocenter should not be much higher than the chromatin density in bulk. To achieve a chromatin density of 15% around the chromocenter (as in chapter 2), and ignoring any density contributions from beads that are further in the chain, the spherical shell around the chromocenter between the radial coordinates S and S+2s should be filled for <= 15% by the 30 terminal beads of the loops. This requirement leads to S>400nm. For smaller chromocenters the chromatin surrounding the chromocenter will be more dense than the bulk and thus less easily moved away, preventing close association with other nuclear constituents such as the nuclear periphery.

We tested this hypothesis by implementing chromocenters with S=500nm, S=350nm and S=250nm in MD simulations which use polymers to represent rosette loops (see Methods). In these simulations, chromocenters with R=500nm localize primarily to the nuclear periphery, with a small interior shell (see chapter 1). Chromocenters of size R<=350nm instead localize to a slightly more interior radial shell between 1500 and 2000nm, or 0.6R-0.8R (figure 3.1b), as expected. In all simulations, chromocenters occur equally spaced over the nuclear periphery as described in chapter 2.

To quantify the forces positioning chromocenters, effective potentials were derived for rosette-wall and rosette-rosette interactions, using rosettes which contain chromocenters with S = 500nm, 350nm and 250nm. Rosette-rosette potentials (figure 3.1c) consist of two parts: a steeply diverging part where chromocenters touch, which directly derives from the Weeks-Chandler-Andersen potential operating between them in the simulations used to derive the potentials, and a long tail, falling off from less than $1k_BT$ to 0 over a distance of up to 3μm. All rosettes have a very similar tail region, but the position where divergence occurs is of course dependent on chromocenter size.

In the derivation of an effective potential for rosette-periphery interactions, wall curvature plays a role. The nuclear periphery is significantly curved at the size scale of a typical rosette, and this curvature is likely to increase the repulsion between chromocenters and the periphery. To quantify this effect, we derived the rosette-wall effective potential from a rosette with a S=500nm chromocenter, interacting with a flat wall and with curved walls

with R=10000nm, 5000nm and 2500nm, the latter of which corresponds to the curvature in actual nuclei. Results (figure 3.1d) show that all rosette-periphery potentials are similar to rosette-rosette potentials, although the range over which they operate is halved (there being only one chromocenter) and tails start higher, but drop off faster with distance. Therefore repulsion between rosettes and the periphery is stronger at short distances than between two chromocenters. Increasing wall curvature increases the height of the tails. There is also a slight but marked difference in shape between potentials derived using curved walls and using a flat wall, which we cannot explain. Using the R=2500nm curved wall, we also derived potentials for S=350nm and S=250nm chromocenter containing rosettes (figure 3.1e).

To check to what extent the derived effective interaction potentials can explain the nuclear localization of chromocenters, MD simulations were performed in which 10 chromocenters interacted with each other through implemented effective potentials (Methods). Where integration progresses at a speed of ~$2*10^2$ MD time units per hour for simulations implementing rosettes as polymers, in simulations implementing effective interactions, integration progressed at ~$1*10^6$ MD time units per hour. Therefore, derivation of effective interactions speeds up our simulations by ~4 orders of magnitude.

In simulations implementing effective potentials, S=500nm chromocenters again localize at the nuclear periphery. Wall curvature drives chromocenters ~100nm closer to the nuclear centre, but they remain peripheral (figure 3.1f). We selected the S=2500nm curved wall for analysis of the effect of chromocenter size on localization, and found that S=350nm and S=250nm chromocenters have a slightly broader radial distribution which is relatively more central (figure 3.1g,h), showing that the increased density of monomers around smaller chromocenters drives rosettes slightly away from the wall.

**Effective potentials for interactions involving the nucleolus**

To investigate the effect of nucleolar size on its location, we performed simulations implementing chromosomes as rosette-shaped polymers, and nucleoli as single monomers not attached to rosettes. Nucleoli with effective sizes of up to that of a euchromatic loop monomer, S=84nm, localize dispersed throughout the nucleus (figure 3.2a,b). Larger nucleoli localize increasingly to the periphery, resulting in nearly exclusive peripheral localization for S>200nm. However, nucleoli of size S>400nm start to occupy the centre of the nucleus as well, and nucleoli of size S>625nm are exclusively centrally localized.

Effective potentials were derived for interactions between nucleoli and rosettes (figure 3.2c). Results show that potentials are again composed of a steep part, where the nucleolus touches the chromocenter, and a long tail where the nucleolus interacts with rosette loops. Small nucleoli (S=50nm) move through this tail virtually without being repelled, but the repulsion becomes ever higher for larger nucleoli. This explains why only small nucleoli localize throughout the nucleus.

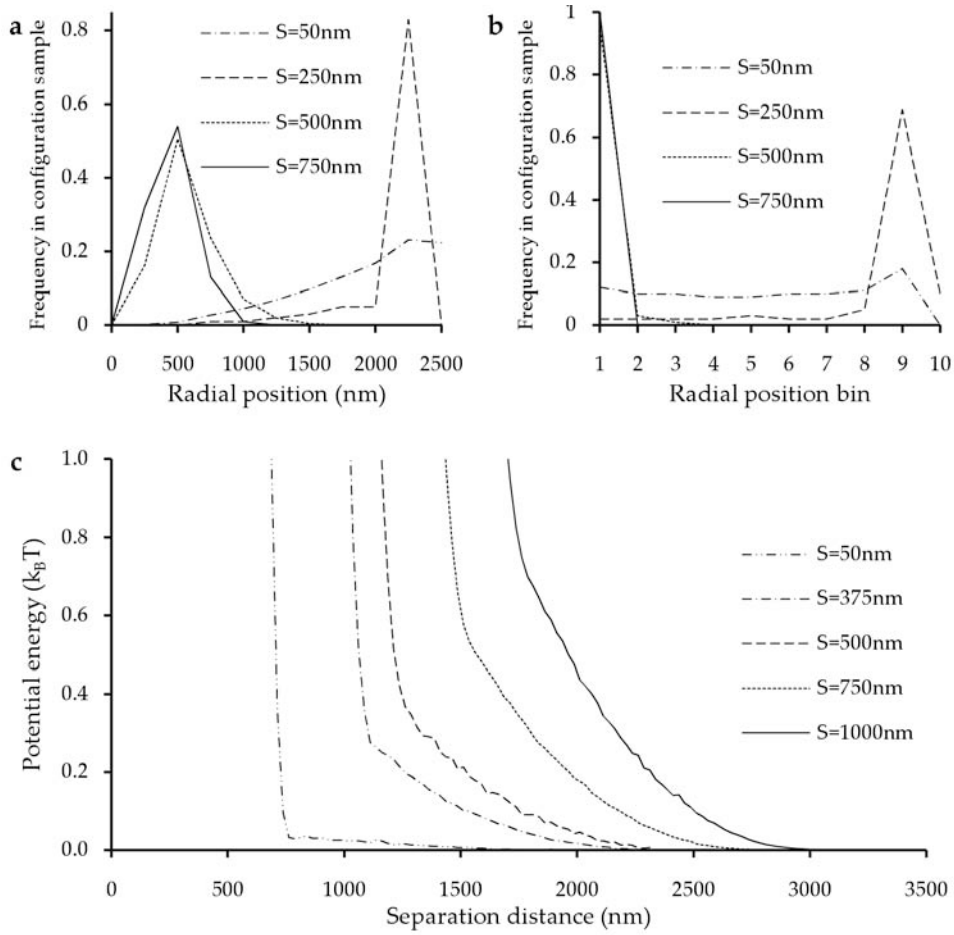
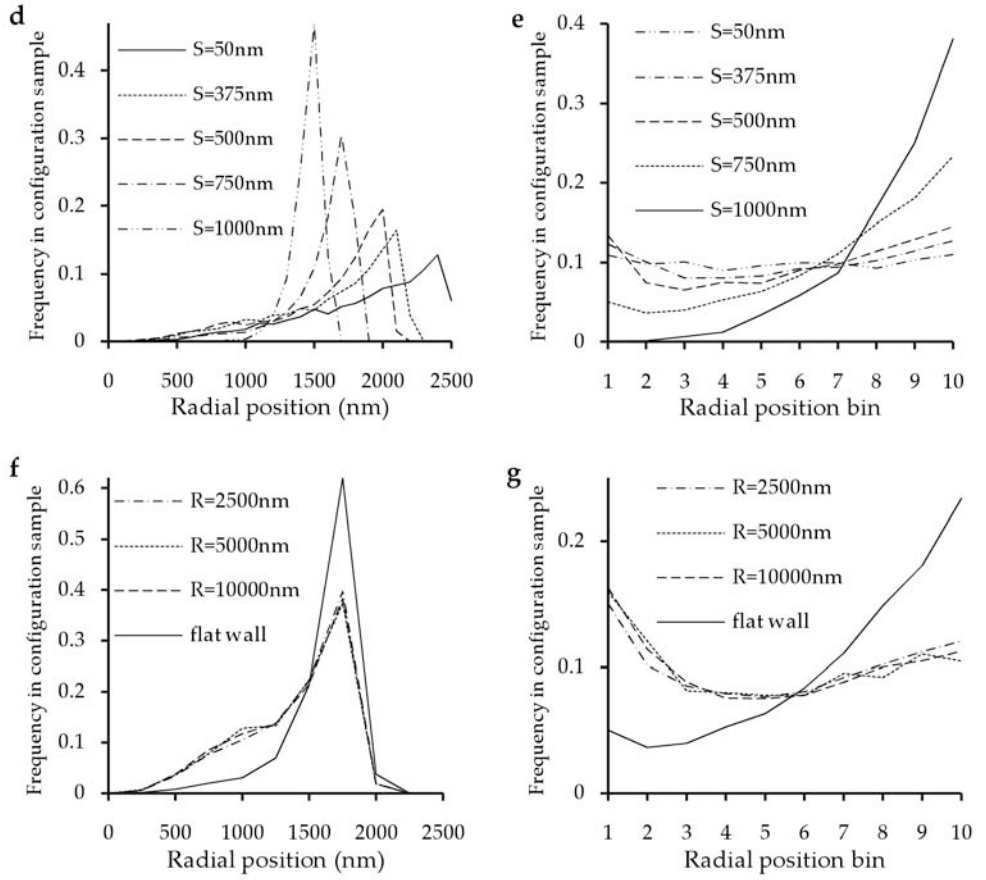



Figure 3.2: nuclear localization of nucleoli. a: Radial positions of nucleoli of various sizes in sample configurations from MD simulations, implementing 10 rosettes as depicted in figure 3.1a together with a single nucleolus, confined in a sphere of 2500nm radius. b: same as a, but with nucleoli binned to ten concentric shells of equal volume. c: effective potentials for the forces operating between a chromocenter and a nucleolus as a function of separation distance. d-g: results of the implementation of the derived effective potentials (including the S=500nm rosette-rosette and rosette-flat wall potential) in simulations. d: nuclear localization of nucleoli of various sizes. e: same as d, but nucleoli binned to ten concentric shells of equal volume. f: nuclear localization in simulations implementing various rosette-wall potentials. g: same as f, but nucleoli binned to ten concentric shells of equal volume.




Again, to determine to what extent effective interactions can explain the nuclear localization of nucleoli, these potentials were implemented in simulations, together with the previously derived effective potentials for rosette-rosette and rosette-flat wall interactions. For interactions of nucleoli with the nuclear periphery a WCA potential was used. Despite the absence of any depletion attraction between nucleoli and the periphery, results show that nucleoli with radius S<500nm localize dispersed over the nucleus (figure 3.2d,e), whereas larger nucleoli localize increasingly to the periphery. Central localization of nucleoli was not observed in simulations implementing a rosette-flat wall potential. When the rosette-curved wall potentials are implemented into simulations, nucleoli localize more centrally (figure 3.2f,g), but nevertheless remain predominantly localized at the periphery.

## Discussion

In chapter 2 we found that in MD simulations implementing the rosette model for Arabidopsis interphase chromosome organization, the localization of many nuclear constituents was correctly reproduced. These rosettes consist of two fractions of chromatin: a relatively compact core called the chromocenter, out of which less condensed loops emanate. Simulations correctly reproduced the peripheral localization of chromocenters (figure 3.1b), the otherwise spatially dispersed distribution of chromocenters over the nuclear periphery, and the central localization of nucleoli (figure 3.2a). We hypothesized that the dispersion of chromocenters over the nuclear periphery is due to repulsion between chromocenters caused by their associated loops. To further investigate this interaction between rosettes, we here have derived effective potentials that describe it. These potentials (figure 3.1c) consist of two parts: a steeply diverging part where chromocenters touch and a long tail, falling off from less than $1k_BT$ to 0 over a distance of up to 3μm. This at first glance explains why chromocenters do not cluster. However, in chapter 2 we identified a depletion attraction between chromocenters due to the exclusion of monomers around them. Using the AO theory (Asakura & Oosawa, 1958; formula 3.2, where ΔF is the free energy gain when two spheres of size D come into contact in a medium filled to volume fraction n with small spheres of size d) we can calculate that the free energy gain when two chromocenters

$$\Delta F = (1+\frac{3}{2}\frac{D}{d})nk_BT \qquad \text{(formula 3.2)}$$

come into contact is $1.48k_BT$. When the chromocenters touch each other, the rosette-rosette repulsive potential results in a free energy loss of $0.68k_BT$. Therefore, the repulsive potential seems inadequate to overcome the depletion attraction. However, it should be noted that the effective potentials used here are calculated in dilute conditions. Within the nucleus, rosette loops are confined into a much smaller volume, resulting in a shorter-ranged but steeper repulsive potential which could overcome depletion attraction.

Rosettes are also repelled by the nuclear periphery. We measured the effective interaction between a flat wall and a rosette (figure 3.1e), and found that it is similar in shape to the repulsion between two rosettes: again a steeply diverging potential when the chromocenter touches the nuclear periphery, and a low tail for the interactions of loops with the periphery, now extending up to 1.8μm. Because the nuclear periphery is curved, we also measured the potential between rosettes and the interior wall of spheres of various diameter. These potentials are, as expected, slightly steeper than the potential with a flat wall (figure 3.1d). Because the nuclear diameter is 5μm and, therefore, the maximum distance between the centres of chromocenters (diameter 1μm) confined therein always less than 4μm, the 10 chromocenters confined within a nucleus never fully escape repulsion by most of their

colleagues, which as described above extends up to 3μm, chromocenters collectively push each other towards the periphery. Since each chromocenter only has 1 interaction with the wall while being repelled by most of the other rosettes, the rosette-periphery interaction is overcome by the combined repulsion by all other rosettes. Therefore we expect chromocenters to localize at the periphery, which is also seen in simulations implementing the effective interactions (figure 3.1f). Besides the depletion attraction operating between chromocenters and the wall, the repulsion between rosettes is thus a second force driving chromocenters to the periphery.

Interestingly, chromocenters smaller than 350nm localized slightly towards the nuclear interior in simulations implementing the rosette loops as chains of monomers. Decreasing the chromocenter radius results in more confined loops around the chromocenter, which are more difficult to displace when the chromocenter approaches the wall. This is reflected in the repulsive effective potentials with the nuclear periphery (figure 3.1d). Implementing these potentials resulted in a broader radial distribution of chromocenters and a stronger preference for a slightly more internal location (figure 3.1g). However, *in vivo* the repulsion between rosettes and the nuclear periphery may be smaller than in simulations, because the persistence length of real chromatin is smaller than our monomer size. Therefore, real chromatin might be more easily deformed when chromocenters approach the nuclear confinement.

Rosettes also repel nucleoli and other nuclear sub-compartments. Nucleoli were modelled as single spheres that interact with the nuclear confinement based on excluded volume. We measured effective interactions between these and rosettes, and found that, as expected, the repulsion is dependent on nucleolar size (figure 3.2c). In chapter 2, no explanation was given for the central localization of nucleoli with radii of 750nm. Again, depletion attraction would result in peripheral localization of nucleoli. We hypothesized that the combined effects of rosette repulsion on each other, on the confinement and on the nucleolus could drive the nucleolus towards the nuclear centre. To investigate this, we implemented the derived rosette-nucleolus effective potentials into a simulation containing the 10 chromocenters and a single nucleolus, and found that the nucleolus has a slight preference for the nuclear interior but a stronger preference for the nuclear periphery. Both preferences initially increase with nucleolar size, but for bigger nucleoli the peripheral localization becomes predominant (figure 3.2d,e). Increased confinement of rosettes forces the nucleolus towards the nuclear centre, because there is less space for it at the nuclear periphery. Because the potentials are calculated in dilute conditions, the real rosette-rosette and rosette-periphery potentials might be much stronger, albeit shorter in range, thus further increasing the propensity of the nucleolus towards central localization.

In conclusion, deriving effective potentials for interactions between nuclear constituents has provided new insights in the forces shaping the Arabidopsis nucleus. We have shown that

implementation of these effective interactions speeds up integration by 4 orders of magnitude, yet reproduces the localization of chromocenters well, while only partly reproducing the localization of nucleoli. Considering that many effects, such as depletion attraction and other many-body interactions, are missing in these simulations, our coarse-grained description of Arabidopsis nuclear organization nevertheless models the nuclear organization of Arabidopsis surprisingly well. The derivation of effective interactions is therefore a powerful tool to ease the exploration of parameter variation in models for chromosome folding.

# 4. Intermingling of sub-chromosomal genomic domains

S. de Nooijer[1], J. Wellink[1], R. van Driel[2], B. Mulder[3,4], T. Bisseling[1]

1: Laboratory of Molecular Biology, Wageningen university,
Droevendaalsesteeg 1, Wageningen, The Netherlands
2: Swammerdam Institute for Life Sciences, Faculty of Science, University of Amsterdam,
Kruislaan 318, Amsterdam , The Netherlands
3: Laboratory of Plant Cell Biology, Wageningen university,
Droevendaalsesteeg 1, Wageningen, The Netherlands
4: FOM institute for atomic and molecular physics (AMOLF),
Science Park 104, Amsterdam, The Netherlands

## Abstract

Understanding the three-dimensional folding of human chromosomes into chromosome territories is of great importance because chromatin folding is involved in regulation of gene expression . Previous fluorescent in situ hybridization experiments using pooled sequence probes spanning Mbps of DNA have revealed that chromosomes do not only fold into chromosome territories, but also that chromosome arms and even domains of approximately 10Mbp appear to fold into their own sub-territories with little visible intermingling. However, what implications these results have for models that describe folding of interphase chromosomes, is unclear. In this paper, we simulate linear chain and random loop models for interphase folding of a single human chromosome 11, and analyse these using a virtual confocal microscope that emulates the FISH technique used in biological experiments. We then compare the results to those of *in vivo* experiments and thus show that typical FISH experiments severely underestimate chromosome sub-domain intermingling. Nevertheless we conclude that extensive chromatin loop formation between neighbouring 10Mbp domains can be excluded because it would create more overlap than observed in experiments.

## Introduction

Chromatin fibres are highly folded inside the interphase nucleus. Each chromosome folds into its own spatial territory within the nucleus (Cremer & Cremer, 2001), and within these territories further sub-compartmentalization is observed (Dietzel *et al*., 1998; Goetze *et al*., 2007). Territory formation can be explained through the formation of intra-chromosomal loops (chapter 2), yet at first sight such loops would reduce territory sub-compartmentalization because the loops would connect and thus co-localize distant parts of the chromosome. To resolve this apparent contradiction, we here investigate what the effect is of chromatin loops on chromosome territory sub-compartmentalization.

Chromatin loops are formed due to the interaction between genomic regulatory elements, including promoters, enhancers and insulators (Simonis & de Laat, 2008). Also, groups of co-regulated genes have been shown to congress in transcription factories (Schoenfelder *et al*., 2010; Sutherland and Bickmore, 2009), resulting in further looping. Most looping is intra-chromosomal and only a small percentage is between chromosomes (Lieberman-Aiden *et al*., 2009). Single cell fluorescent in situ hybridisation (FISH) data sets and biochemical methods, i.e. 3C-related techniques (Simonis *et al*., 2007), reveal a considerable cell-to-cell heterogeneity in chromatin folding in general and looping in particular (Mateos-Langerak *et al*., 2009; Simonis *et al*., 2006). A likely explanation is that chromatin looping is dynamic. In agreement with this, interphase chromatin shows rapid constrained diffusion, allowing individual sequence elements to explore a nuclear volume with a diameter of a few tenths of a μm at a sub-minutes time scale (Heun *et al*., 2001). Despite the freedom of movement of chromatin, each interphase chromosome occupies its own territory. Within individual chromosome territories chromatin from different parts of a chromosome also show remarkably little mixing. For instance, for human chromosomes 2 and 6 it has been shown that the two arms (each tens of Mbp) show little mixing at the light microscopy level (Dietzel *et al*., 1998). Similarly, (Goetze *et al*., 2007) showed, using fluorescent in situ hybridisation (FISH), that 2 to 10 Mbp domains on the same chromosome arm hardly intermingle. Chromatin is thus compartmentalized at many different length-scales, and any model for chromatin folding must be compatible with this observation. Therefore the fractal globule model was proposed (Lieberman-Aiden *et al*., 2009), in which a chromosome fills its territory through a fractal-like path that results in sub-compartmentalisation at all length-scales, but that model does not take into account chromatin looping. Also, it does not explain the presence of distinct chromosome territories because its folding is thermodynamically unstable. Therefore, a better model is required.

To describe the folding of a chromatin chain into a three-dimensional territory through loop formation, the random loop model (Bohn *et al*., 2007; Mateos-Langerak *et al*., 2009) was proposed, in which intra-chromosomal loops are formed by association of two loci

randomly distributed on the chromosome, creating a network polymer with loops over a broad length scale. The random loop model successfully explains data that relate genomic distance to physical distance in interphase nuclei (Bohn *et al*., 2007) and chromosome territory formation (chapter 5). We here investigate whether the random loop model is also able to explain chromosome territory sub-compartmentalization. Long-range loops would bring distant domains in close proximity, which would increase contacts between these distant domains. The random loop model, which contains 15 to 150 long-range loops per 100 Mbp, seems incompatible with sub-compartmentalisation (Mateos-Langerak *et al*., 2009). However, a quantitative approach is required to study how much compartmentalization is expected in an unbranched polymer and to what extent loop-containing chromosomes can sub-compartmentalise the folded polymer. In this paper, we provide such quantitative analysis by simulating the folding of a chromosome under different looping constraints. We then analyse configurations for overlap between chromosomal sub-domains, using a virtual confocal microscope. This enables us to make a direct comparison between the results of our simulations and data from real FISH experiments. Based on this analysis we are able to explain how loop formation affects sub-chromosomal compartmentalisation, and also to show that the level of intermingling is often underestimated in FISH experiments.

# Methods

**Simulation of chromatin folding**

Simulations were carried out using the ESPResSo software package (Limbach *et al*., 2006) using the method described in chapter 2. Human chromosome 11 (135 Mbp) was modelled as a homogeneous, unbranched polymer, in order to facilitate comparison to the experimental analysis of (Goetze *et al*., 2007). The polymer consists of $10^4$ soft monomers that are connected by harmonic springs without rotational constraints, each monomer representing 13.5Kbp. This monomer size is in the upper range of the experimentally observed persistence length for nuclear chromatin: 3-20Kbp (Langowski, 2006). The polymer was confined inside a spherical volume mimicking the territory of chromosome 11. Assuming a spherical nucleus with a diameter of 10μm (523$\mu m^3$ volume) and equal density of all chromosomes, the territory of chromosome 11 was calculated to occupy a volume of 23$\mu m^3$, corresponding to a sphere with a diameter of 3.5μm. It should be noted that in the simulations this size is arbitrary, as monomers are scaled to occupy 10% of the territory volume (as in Cook & Marenduzzo, 2009). All other relevant parameters are scaled based on the monomer size. Values in micro- and nanometres are given only to allow direct comparison with biological data. Thus each monomer should occupy a volume of $2.3*10^5$ $nm^3$, which corresponds to a monomer radius ($\sigma$) of 38nm. In the ESPResSo software monomers are represented by point centres of force, with an interaction range between particles establishing an excluded volume around each point particle. Therefore, the range of interaction potentials was set based on $\sigma$ to result in the appropriate excluded volume.

For interactions between monomers we used the Weeks-Chandler-Andersen potential (Weeks *et al*., 1971) in the same way as described in chapter 2. For bonded interactions between monomers, harmonic spring potentials were used with an equilibrium position at $2\sigma$ and a spring constant of 50. This is sufficiently high to prevent extensions through thermal fluctuations that would allow two chains to pass through each other.

**Starting configuration**
The monomers were positioned in a cylindrical spiral shape with a period of 125 monomers, a distance of $2\sigma$ between successive monomers and a displacement of $2\sigma$ per winding, mimicking to some extent the metaphase chromosome configuration. This structure was placed at the centre of the confining sphere. To create a linear chain, monomers were connected sequentially (following the spiral). In the random loop model, additional (initially very long) bonds were created between randomly selected monomers (see figure 4.1a for a perspective rendering of the starting configuration) spanning the interior of the cylinder.

**Equilibration and simulation**
Molecular dynamics simulations were carried out through velocity-Verlet integrations in the presence of a Langevin thermostat. The time step was 0.01 (see chapter 2 for units). Both linear and random loop models were subjected to a warm up phase with a capped Weeks-Chandler-Andersen potential (Weeks *et al*., 1971) to prevent excessive energies due to the starting configuration. The cap was implemented as a maximum force limit for a single interaction, starting at $3k_BT$ ($k_B$ = the Boltzmann constant) and was raised every $10^3$ integration cycles by another $3k_BT$, until it was removed altogether after reaching $120k_BT$. Since the bonds in the random loop model were extremely stretched in the initial configuration, warm up integrations were performed until all harmonic bonds, including the very stretched bonds that implemented the random loops, reached a state close to equilibrium (extensions maximum 2 times harmonic bond equilibrium extension). During this phase, the confining sphere was slowly shrunk to its final size of 3.5μm. After warm up, equilibration continued with uncapped forces until equilibrium was achieved (see Results section) and sample configurations were stored every $10^3$ integration cycles.

To evaluate equilibration, simulations of polymers containing a non-looped polymer were integrated for $2*10^5$ MD time units. We verified that the simulation reached equilibrium by plotting the distribution around average of general polymer parameters, including end-to-end distance, contour length, radius of gyration and energy, as well as parameters related to folding into domains, i.e. the number of confocal volumes (see below) containing at least one monomers of a reference region of 743 monomers (equivalent to 10Mbp). Supplementary figure 1 shows that equilibration was achieved for all parameters after $5*10^4$ MD time units. Autocorrelation analysis of fluctuations around average of these parameters showed that autocorrelation half-times for the slowest parameters, end to end distance and

total signal volume, were approximately 200 MD time units. We therefore concluded that sufficient sampling could be achieved by analysing configurations sampled from a period 500 times longer, or $10^5$ MD time units following equilibration.

Since random loop model polymers are sterically confined due to their network geometry, they equilibrate much faster than non-looping polymers. Equilibration of all parameters occurs within $5*10^3$ MD time units. Autocorrelations likewise occurred considerably faster, with the slowest parameter, here the end-to-end distance of the two linear fragments emanating from the network, having an autocorrelation half-time of 20 MD time units. Again, we simulated 500 times longer than that, or $10^4$ MD time units after equilibration.

**Virtual confocal microscopy**

Experimental data on the sub-compartmentalization of chromosome territories are based on FISH studies. Imaging of FISH labelled nuclei by confocal microscopy is limited by the point spread function, resulting in a confocal volume. This has an approximately ellipsoidal shape of about 250x250x500 nm for excitation light of 500nm. In practice, the confocal volume is often larger due to suboptimal microscope alignment and sample quality, resulting in a larger confocal volume and therefore lower spatial resolution. A second important parameter is the detection threshold of the FISH labelled object, which is determined by (i) the labelling density, i.e. the number of fluorophores per unit of genomic length, (ii) the signal-to-noise ratio, (iii) the sensitivity of the detector, and (iv) the scanning rate and fluorophore bleaching. In the study of Goetze *et al*. (Goetze *et al*., 2007) and many others that use bacterial artificial chromosomes (BAC) probes, the typical detection limit is in the range of 30 to 100Kbp. To compare the behaviour of the polymer model with microscopic observations a "virtual confocal microscope" algorithm was applied (similar to (Munkel *et al*., 1999), but simplified to allow high-throughput analysis of many configurations). The algorithm produces a three-dimensional spatial binning of monomers to rectangular voxels with dimensions comparable to the confocal volume of microscopes used for biological experiments. As in a real microscope, a signal in a voxel is detected only if sufficient labelled DNA is present in that voxel. In our algorithm, this detection threshold is implemented as a minimum amount of monomers required for signal detection (figure 4.1c-d for a graphic illustration of the algorithm).

**Calculating overlap between adjacent polymer domains**

To calculate the overlap of the signal from two adjacent polymer regions of 743 monomers (equivalent to 10Mbp chromatin of human chromosome 11), a reference region and a target region are defined, similar to the experiments of Goetze *et al*. (Goetze *et al*., 2007). For each voxel, the number of monomers belonging to either region is scored. If the number of monomers of the reference region within a voxel equals or exceeds the detection threshold, it is scored as resulting in a detectable signal. If that voxel also contains a number of

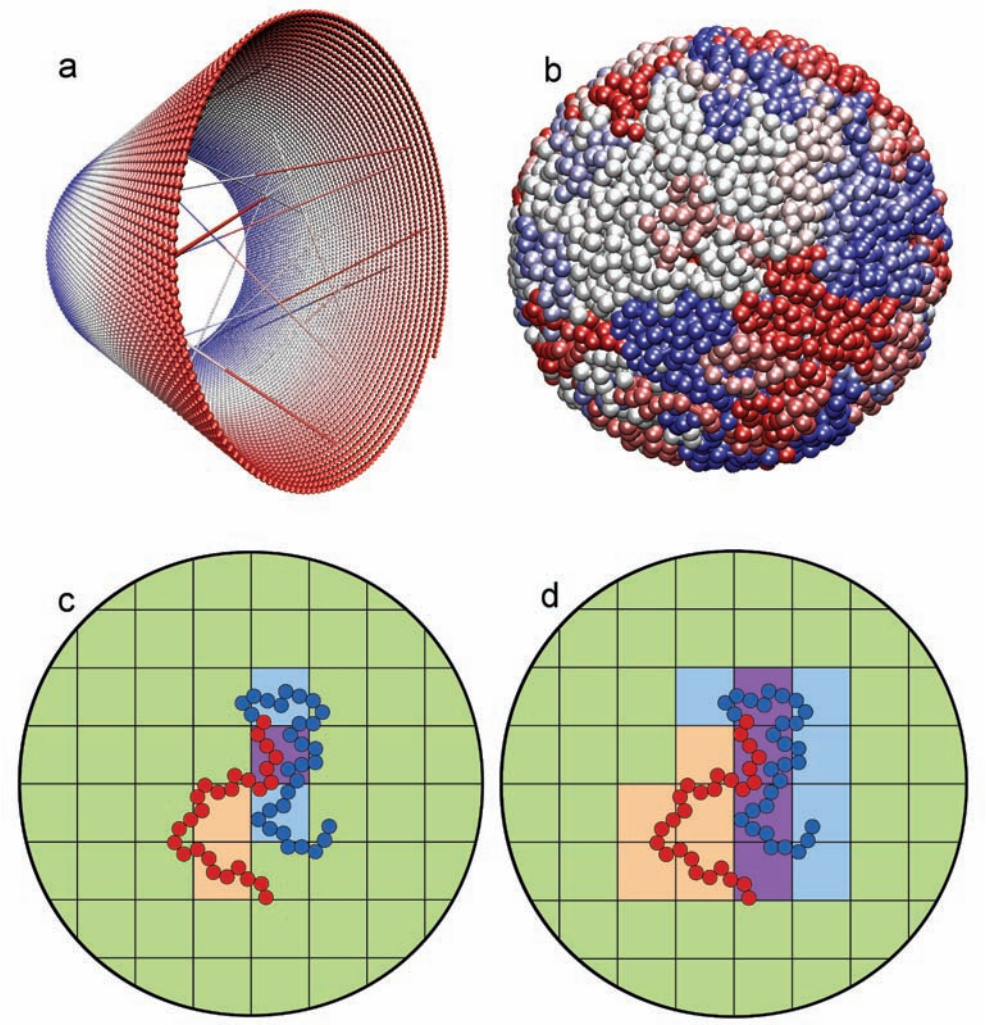



Figure 4.1: molecular dynamics simulations and virtual confocal microscope. a: Perspective projection of spiral-shaped polymer starting configuration. Colours indicate monomer position in the chromosome and range from red (position 0Mbp) through white (67Mbp) to blue (135Mbp). Subsequent monomers are connected through bonds along the spiral. Random loops, when present, are implemented through very long bonds between random monomers, which span the interior of the spiral. Each half of these long bonds has a different colour, matching that of the monomers at either end. To display the bonds, monomers are scaled down to half diameter in this image.

b: Orthogonal projection of equilibrated chromosome (without random loops) in confinement. Monomers are here shown at real size. Figures a and b were created using VMD software. c and d: Schematic explanation of the virtual confocal microscope algorithm. The nucleus is divided into confocal volumes (here represented in 2D with squares). Two regions, a reference region (blue) and a target region (red) of a monomer chain are shown. The algorithm determines for each confocal volume if the monomer number belonging to either region exceeds the signal threshold (>=5 monomers). In c, signal from the reference region exceeds the threshold in 3 volumes (blue squares) and signal from the targ*et al*so (red squares), and one of these volumes is shared (purple square). Therefore, the overlap percentage in this image is 33%. In d, the same configuration is re-analysed at a detection threshold of 1 monomer. Here signal from the reference region (blue) exceeds the threshold in 9 squares, of which 4 are shared with the target region. Therefore, the overlap percentage at a signal threshold of 1 is 44%. Note that the real algorithm is applied in three dimensions and that nucleus, monomers and confocal volumes are not drawn to scale.

monomers above the detection limit of the target region, it is scored as containing overlap between the two adjacent regions. The total overlap between the two domains is then calculated as the ratio of the number of voxels containing both signals to total number of voxels containing a detectable signal from the reference domain. To plot the distribution of overlap values in the configuration sample, values were binned to 30 bins, each with a width of 0.03. Then the fractional contribution of each bin to the total dataset was plotted.

## Results

**The spatial distribution of sub-domains in polymers with and without loops**

To investigate the effect of intra-chromosomal loop formation on chromosome territory sub-compartmentalization, we simulate the folding of a $10^4$ monomers polymer (see Methods), representing human chromosome 11, confined within a spherical volume of 3.5μm diameter, mimicking the territory of this chromosome (for calculation see Methods). We simulated both unlooped and looped polymers; for looped polymers, we selected a looping frequency in the random loop model which can explain chromosome territory formation (Mateos-Langerak *et al*., 2009). This resulted in 100 loops, which were randomly introduced in the polymer. Ten independent simulations were performed with different loop configurations. Sample configurations taken from simulations after equilibration (see Methods) were convoluted by virtual 3D confocal microscopy (see Methods).

First, the distribution of monomers over confocal volumes in simulations of both a looped and an unlooped polymer confined into a territory was analysed. We defined a 743 monomer (representing 10Mbp) sub-domain and then counted how many of the sub-domain monomers are present in each of the 658 confocal volumes within the territory. We re-analysed each configuration sample multiple times by shifting the 743 monomers long

sub-domain along the chromosome in steps of 500 monomers. The numbers of confocal volumes occupied by sub-domain monomers were averaged for all shifted sub-domains. In simulations of polymers without loops, monomers from the sub-domain occurred in 118 +/- 1.4 (=standard deviation) confocal volumes, which is 18% of the territory. In figure 4.2, for the subset of confocal volumes that contains at least one monomer of the sub-domain, the distribution of the number of the sub-domain monomers over this subset was plotted. This shows that, of the confocal volumes containing at least 1 monomer from the sub-domain, the majority contains less than 4 monomers of the sub-domain, while less than 10% of confocal volumes contains at least 10 monomers. This means that the space occupied by the sub-domain consists of a core in which sub-domain monomers are highly concentrated, surrounded by an increasingly more diffuse cloud where few monomers belonging to the sub-domain occur. Repeating this analysis on simulations of a randomly looped polymer showed that here, sub-domain monomers occur in 149 +/- 1.5 confocal volumes, which is 23% of the total volume of the territory. This shows that under confinement, sub-domains are more expanded in a randomly looped polymer than in an unlooped polymer. Nevertheless, figure 4.2 shows that here, too, the sub-domain consists of a core of small number of confocal volumes containing a large number of sub-domain monomers, and a larger cloud of confocal volumes containing only few sub-domain monomers.

These results pose a problem for the definition of overlap between two sub-domains. Should we only consider overlap of the concentrated cores of each sub-domain, or also of (part of) the diffuse clouds surrounding them? An argument for either could be made, but here we are interested in a comparison of simulation data with experimental data. We therefore have equipped our virtual confocal microscope with the same sensitivity as a real microscope through the introduction of a detection threshold: a minimum amount of monomers per confocal volume that is required to observe a signal. Then the values on the x-axis of figure 4.2 can be reinterpreted as the minimum number of monomers required to see a signal in the microscope (detection limit), and the solid lines now show the fraction of confocal volumes containing sub-domain monomers that can be detected by the microscope at various thresholds.

In practice, detection limits of 30-100Kbp are commonly encountered in FISH studies. Our results show that, assuming unlooped chromosomes, in FISH experiments with a detection threshold of ~30-40Kbp (2-3 monomers per voxel in our simulations), only 60-75% of the volume occupied by the labelled DNA (sub-domain monomers in our simulation) is actually observed. In figure 4.2, the dotted lines show the fraction of sub-domain monomers detected by the microscope. At a 2-3 monomer detection threshold, 95-85% of the sub-domain monomers (and thus labelled DNA) is observed, respectively. The remaining 5-15% of the DNA is present in invisible loops emanating from the visible sub-domain. With a 100Kbp detection threshold (7-8 monomers), only 15-20% of the volume occupied by sub-domain monomers is still observed, and over half of the sub-domain monomers then occurs in

voxels that do not exceed the detection threshold. The detection threshold thus greatly influences the size of the space which a 10Mbp sub-domain would seem to occupy. The presence of random loops in chromosomes would change all these numbers slightly downwards (figure 4.2), but the trend remains the same.

**Effects of loop formation on overlap of adjacent sub-domains**

To investigate the effect of loop formation on overlap between two adjacent sub domains within a polymer, we calculated the amount of overlap between two adjacent labelled regions. Again, we analysed each configuration (total 1018 configurations) multiple times by shifting the 743 monomers reference region plus an adjacent target region (each equivalent to 10Mbp) along the chromosome in steps of 500 monomers (equivalent to 6.7Mbp) and recalculating the overlap after each shift. The degree of overlap as detected by the virtual microscope was calculated for different detection thresholds. Without detection limit, the overlap in an polymer without loops is 17% +/- 8.9%. This means that on average 17% of the voxels that contain at least one monomer of the reference region also contain at least one monomer of the target region. There is no significant difference between overlap percentages calculated at different positions of the reference and target region pairs along the polymer, indicating that their position in the polymer does not affect their intermingling.

The detectable overlap rapidly decreases with increasing detection threshold. A threshold of 4 monomers (54Kbp) per voxel results in an apparent overlap of only 5.5 +/- 4.6%, whereas at 7 monomers (94Kbp) per voxel the apparent overlap is reduced to 1.1 +/- 2.7% (for distributions of overlap over configurations, see figure 4.3a). These latter overlap values are similar to what Goetze *et al*. (2007) found experimentally (less than 5%). To eliminate effects of varying the detection threshold, all simulations hereafter are analysed at a detection threshold of 5 monomers (67Kbp). For unlooped polymers, sub-domain overlap is then 4.8 +/- 5.7%.

Next, we repeated the analysis on simulations containing a looped polymer. Here, 25% of the confocal volumes containing sub-domain monomers, exceeded the detection threshold of 5 monomers per voxel. The average detectable overlap between two adjacent 743 monomer domains was 1.9% +/- 3.0%, which is lower than found for a unlooped polymer. Because the overlap percentage is not distributed normally over configurations, the overlap distributions of both looped and non-looped polymers are plotted in figure 4.3c, which shows that while both distributions are centred at 0 overlap, the introduction of loops in the polymer leads to a much smaller number of configurations in which overlap takes place and thus an overall lower overlap percentage.

To understand why adjacent sub-domains of random loop model polymers overlap less than non-looped polymers we consider two opposing effects. First, loops physically connect distant chromosomal regions and thus increase overlap between different regions of the

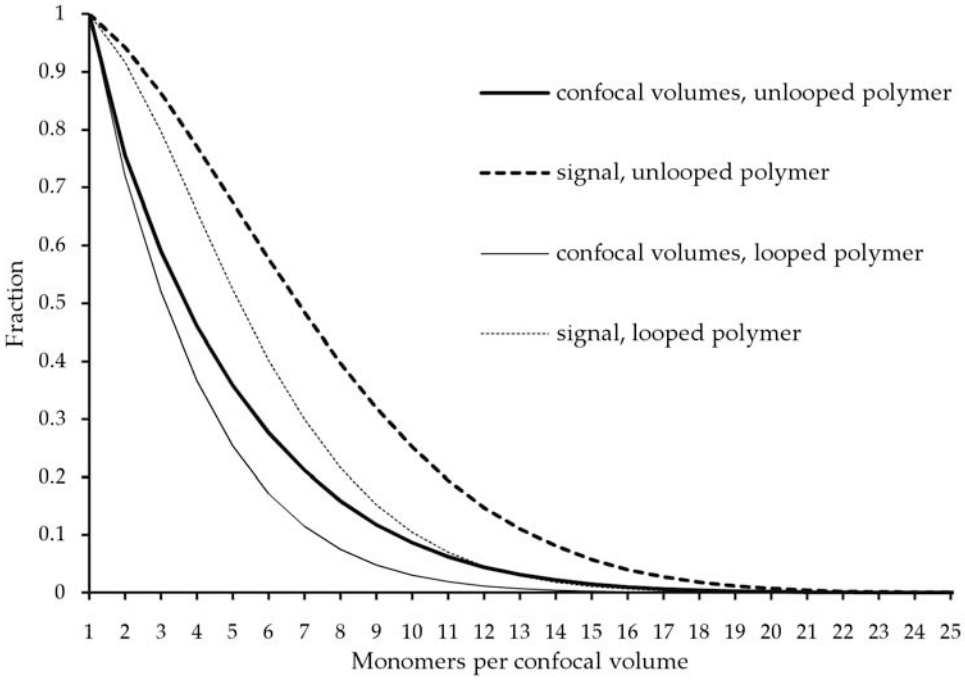



Figure 4.2: the distribution of monomers from a 743 monomer sub-domain over confocal volumes. Solid lines show, for the subset of confocal volumes that contains at least one monomer of the sub-domain, the fraction of confocal volumes containing a certain number of sub-domain monomers. thick lines are derived from simulations of confined unlooped polymers, thin lines of confined looped polymers. The solid lines can also be interpreted as the fraction of confocal volumes visible at different signal thresholds (which is the number of monomers that have to be present in a confocal volume to give a visible signal) of the virtual confocal microscope. The dotted lines show the fraction of sub-domain monomers visible under each signal threshold.

polymer. Second, loop formation contributes to spatial partitioning of the polymer, because loops intermingle less readily than unlooped chains (Jun & Mulder, 2006). Thus it is important to unravel the contribution of each of the two effects on intermingling. To analyse the effect of loops that connect the reference and the target region on overlap of the two sub-domains, we performed simulations of polymers with different numbers of loops exclusively connecting the two 743 monomer regions (in the centre of the polymer). Results show that 2 loops increased the overlap from 4.8 +/- 5.7% to 6.9 +/- 5.0%. Overlap increases to 18 +/-6.8% if 5 loops are introduced and rapidly increases, up to 50% +/- 7.6% for 50 loops (figure 4.3d shows distributions of overlap values over sampled configurations). To calculate a maximum value for overlap we zipped the two regions together by connecting each monomer in the reference region to its counterpart at the same position in the target

region. This resulted in an overlap of 59 +/- 8.0%. Note that 100% overlap does not occur in these analyses because adjacent labelled chains may reside in different voxels and thus will not exceed the detection threshold in the same voxels.

To quantify the effect on overlap of two sub-domains of loops that do not connect the two regions for which overlap is measured, we first created models with 25, 50 or 100 random loops. Subsequently, loops that connect the two adjacent domains that are analysed for overlap are removed (which usually implies the removal of 1 loop). Results of these simulations reveal that the two regions show an overlap in the three models of 0.6% +/- 2%, 0.6% +/- 2% and 0.8% +/- 2% respectively (distributions of overlap over configurations in figure 4.3c). This is lower than what is found for the random loop polymer with 100 loops in which the loops between the regions were not omitted (1.9 +/- 3.0%). These results are compatible with the idea that loops that do not directly physically link the analysed regions decrease overlap percentages through increased polymer partitioning. In contrast, as expected, loops between the two domains increase overlap (figure 4.3d).

**Effect of confinement on overlap between sub-domains**

Apart from loop formation, volume confinement too is expected to influence the spatial distribution of sub-domains. A polymer in dilute conditions expands to a much larger size than is available to chromosomes in the nucleus. Confinement thus brings distant parts of the polymer closer together and therefore is expected to increase sub-domain overlap. To quantify the effect of volume confinement of a self-avoiding random walk polymer on overlap between two adjacent sub-domains, polymer configurations were simulated without confinement, allowing the polymer to expand to an ideal self-avoiding configuration. Under these conditions monomers from a labelled 743 monomer region were present in 184 +/-7.5 voxels, showing an increase in volume of 55% compared to a spatially confined polymer. With a detection limit of 5 monomers per voxel, only 16.9 +/- 3.8 voxels (9% of the total volume occupied by the 743 monomer domain) contained a detectable signal from the labelled domain. To measure overlap between two sub-domains, adjacent 743 monomer domains were labelled and analysed as described above. The number of voxels with a signal from both adjacent domains above the detection threshold of 5 monomers is very low under these conditions: only 0.01%+/- 0.17% of the voxels, i.e. no overlap is detectable and the domains appear fully separated in the virtual microscope.
The same analysis was applied to a randomly looped polymer without confinement to a pre-set volume. This resulted in a 18% increase in sub-domain volume, less than in the case of linear polymers because the loops restrict the looped polymer in its expansion. Nevertheless, the overlap percentage between sub-domains likewise dropped to 0.05%+/- 0.2%.

**Effect of spatial resolution on overlap between sub-domains**

The size of confocal volumes into which monomers are binned affects the observed overlap. In general, for smaller spatial bins less overlap is found because the chance that monomers

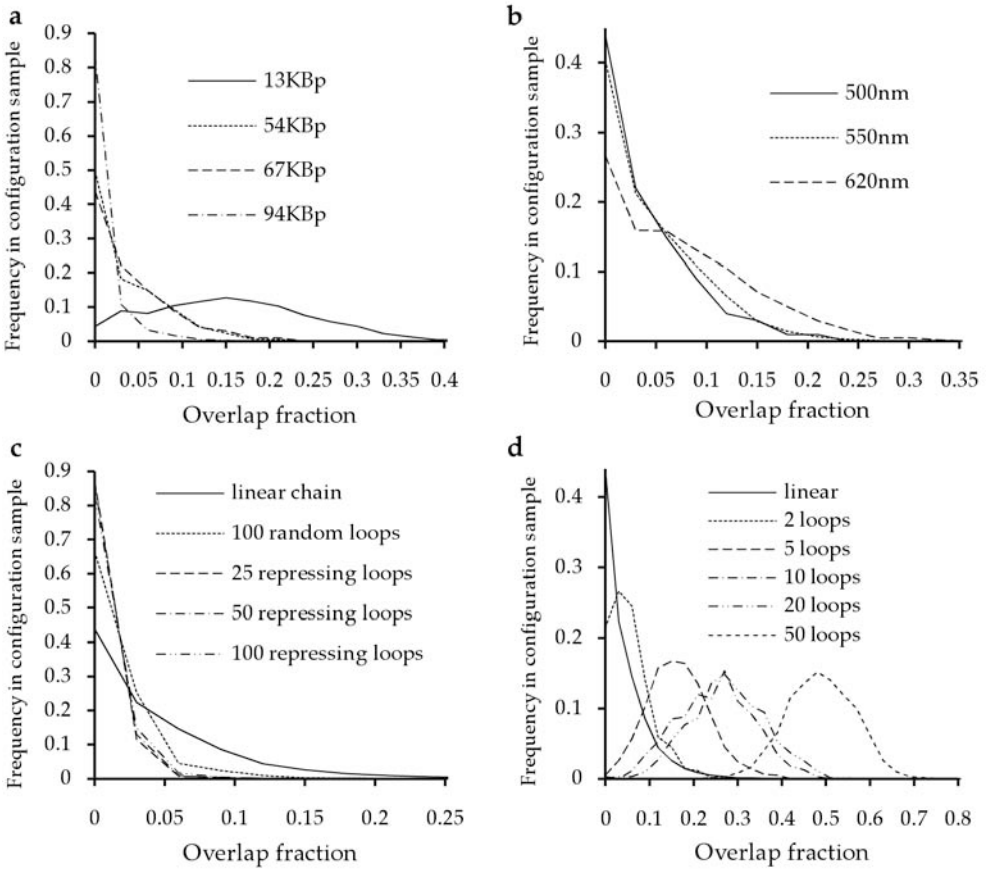



Figure 4.3: distribution over sample configurations of the domain overlap between two labelled adjacent 10Mbp (743 monomers) domains, as analysed using a virtual confocal microscope algorithm with a threshold of 5 monomers per voxel. a shows the dependence of measured overlap on the detection threshold of the virtual confocal microscope. b shows the dependence of measured overlap on confocal volume. c and d show how overlap changes in looped polymers. In c, the dotted line shows the effect of 100 random loops compared to a linear chain. The dashed lines show the partitioning (overlap decreasing) effect of 25, 50 and 100 random loops, respectively, when any loops connecting the reference and target regions are removed (therefore the remaining loops are called "overlap repressing loops"). In d, the overlap enhancing effect of various amounts of loops connecting the reference and target regions, in the absence of other loops, is shown.

from different regions come together in a spatial bin decreases. For optimal sampling, the spatial bins should be small compared to the total volume occupied by the sub-domains of interest but contain sufficient monomers to avoid sampling errors. The size used in this paper, 250x250x500nm, satisfied both criteria mentioned above well: each confocal volume contains approximately 15-30 monomers, and a 10Mbp-subdomain is spread over 118 confocal volumes. This volume size has the additional advantage of being comparable to the size of confocal volumes of real microscopes, facilitating comparison of simulation data with experiments. However, the size of the confocal volume can vary between experiments, and therefore it is interesting to see the effects of changes in spatial resolution on measured overlap. In a real well-aligned confocal microscope combined with optimal sample preparation, the size of the confocal volume depends on the wavelength of the excitation light. While we assumed a wavelength of ~500nm for the excitation light so far, other commonly used wavelengths include ~550 and ~630nm. We therefore scaled the confocal volume to 275x275x550nm (volume increase 33%) and 315x315x630nm (volume increase 100%), respectively, and recalculated the overlap in a simulation of unlooped polymers, using a detection limit of 5 monomers per voxel (67Kbp). Results show an increase of apparent overlap to 5.6% +/- 5.3% and 8.7 +/- 6.8%, respectively, compared to 4.8 +/- 5.7% at 500 nm (for distributions of overlap over configurations, see figure 4.3b). As expected, the overlap scales with the size of the confocal volume.

## Discussion

In this work, we studied whether intra-chromosomal random loop formation, which explains why chromosomes form territories, can also explain sub-compartmentalisation of chromosome territories. At first sight, randomly looped chromatin contains many long-distance loops which interconnect different chromosomal sub-domains. Indeed, our results show that sub-domains are spread out over more volume in looped polymers, because the loops interconnect sub-domains with other areas of the polymer. However, our results also show that overlap between adjacent sub-domains in randomly looped polymers is lower than in unlooped polymers. How then to resolve this apparent contradiction? Looking closer, we demonstrate that loop formation has two opposing effects on overlap between chromosomal sub-domains. Loops that connect the two domains bring these domains closer together and increase their overlap. In contrast, loops that connect parts of the chromatin fibre outside the two sub-domains contribute to partitioning of the chromosome territory and thus decrease sub-domain overlap. In the random loop model for interphase chromosome folding, the chance that two random 10Mbp sub-domains are connected by a loop is 67%, while more than 2 loops between sub-domains are required to significantly raise the amount of overlap between them (figure 4.3d). Therefore the repressive effect on overlap of loop formation is dominant over the enhancing effect and, counter-intuitively, random loops decrease overlap between adjacent sub-domains.

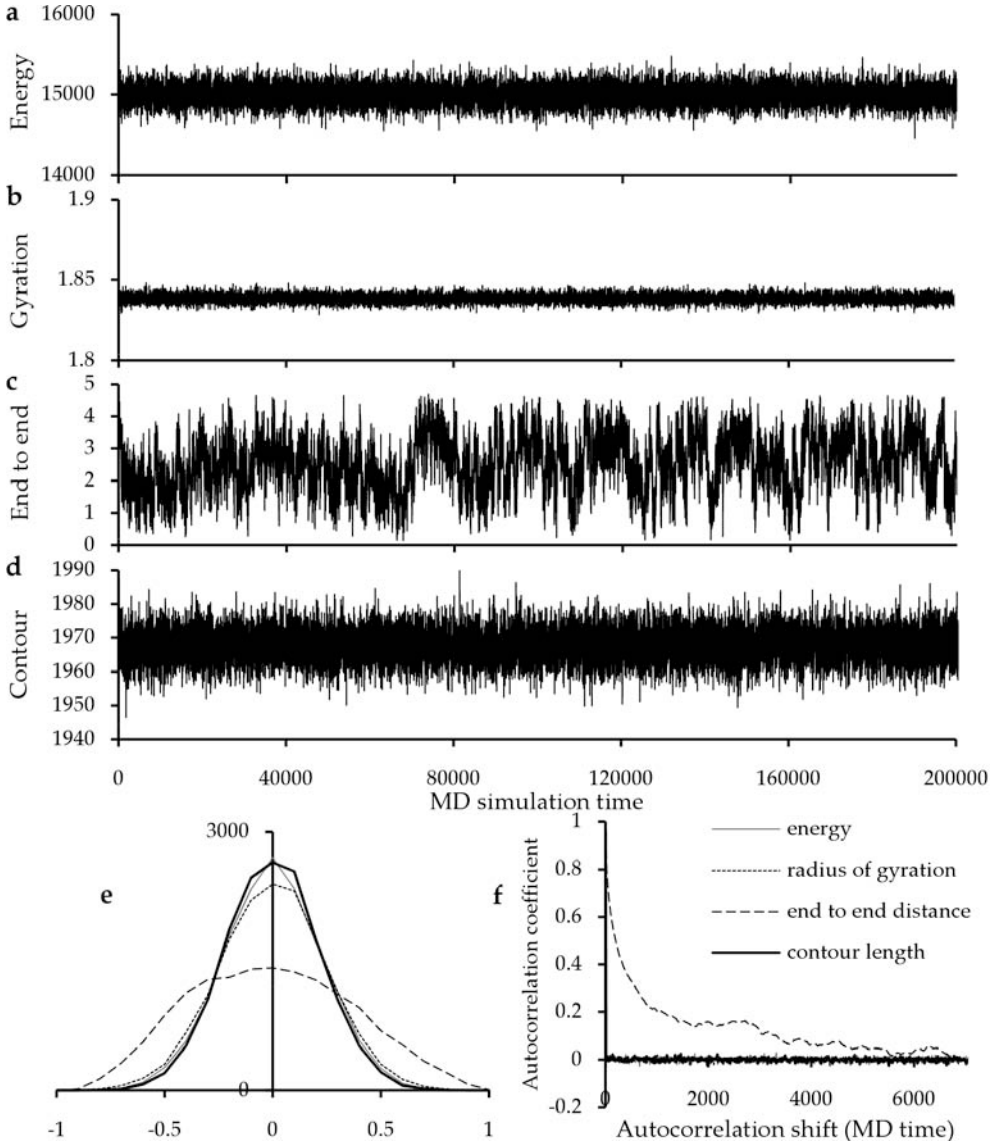
In the absence of confinement, linear polymers fold into a spatially extended configuration in which different parts of the chain become unlikely to encounter each other, resulting in very low overlap between sub-domains. Looped polymers de-intermingle easily (Jun *et al*. 2006), and in randomly looped polymers each branch of the network tries to form its own sub-territory within the spatial constraints imposed by the network structure. This, too, results in very low sub-domain overlap. Therefore, both in linear and in looped polymers, confinement is the ultimate cause of sub-domain overlap.

We also show that in FISH experiments only a relatively small fraction of the chromatin fibre will be observed using a confocal microscope. Depending on the sensitivity of the microscope, only around 50-85% of the labelled fibre is detectable under typical experimental conditions, because only those voxels that contain sufficient fluorophores associated with the FISH labelled chromatin fibre give rise to a signal. Not less than 25-60% of the voxels in which sub-domain chromatin occurs, has a signal strength below the detection threshold and is not visible as a part of the spatial sub-compartment. This predicted limitation in the visualisation of the chromatin fibre under typical FISH labelling and microscopic conditions offers a simple explanation for the lack of intermingling between sub-chromosomal domains that is observed in experiments (Goetze *et al*., 2007).

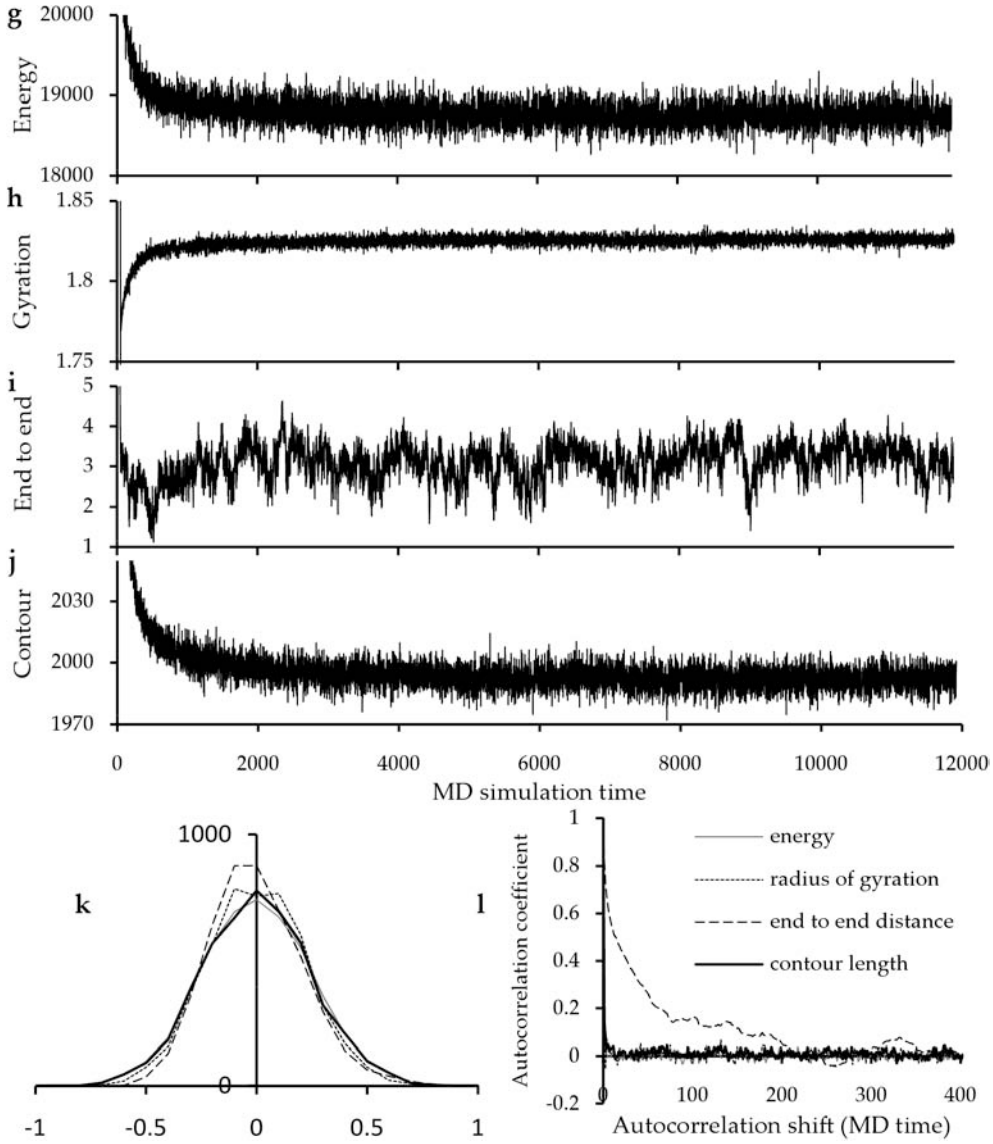
Our analysis of the overlap between sub-chromosomal domains may be extended to the issue of intermingling between individual chromosomes. Although it is generally accepted that there is little mixing between interphase chromosomes (Cremer & Cremer, 2001), each occupying a separate territory, a study of Branco and Pombo (Branco & Pombo, 2006) shows overlap in the 30% range, based on FISH labelling on thin sections of nuclei, rather than complete nuclei. Similarly, they showed significant intermingling of chromatin from the two arms of chromosome 3. Their experimental approach is likely to considerably increase the signal-to-noise ratio, thereby lowering the detection threshold. Based on our model calculations, this may well explain their high intermingling values, suggesting that classic whole nucleus measurements considerably underestimate mixing of chromosomes in interphase.

Based on the work presented here, recommendations for future FISH experiments can be made. In order to compare overlap between experiments, it is important to standardize the confocal volume of the microscope, and to use labelling and observations methods that result in high but standardized signal intensity. Simulation techniques such as presented in this paper can then be used to help interpret experimental observations and to use such data to develop and test models for chromosome folding. The currently available data on chromosome sub-compartmentalisation are consistent with the random loop model, which therefore remains the best available model for chromosome folding.

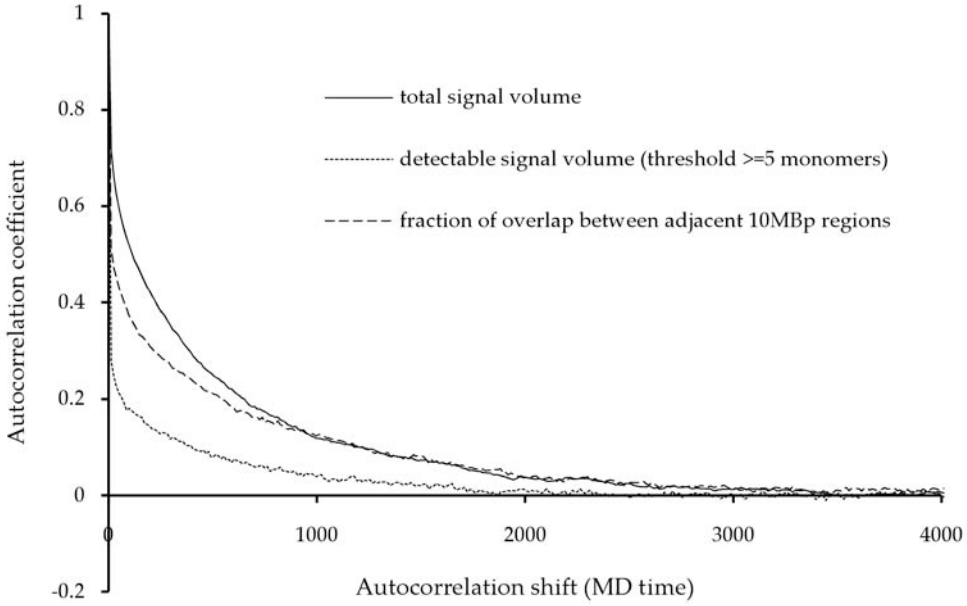
## Supplementary materials




Supplementary figure 1: equilibration of simulations of linear (a-f) and looped (g-l) polymers. Traces are shown of the development in time of total (kinetic and potential) energy of all particles in the simulation (a and g), radius of gyration of the (confined) polymers (b and h), end to end distance (c and i)




and contour length (the total length of all bonds within the polymer, including those forming loops in looped polymers, d and j). These traces show that parameters stabilize around equilibrium after approximately 50,000 time units for linear polymers and 5,000 time units for looped polymers. Figures e and k prove that after equilibration, parameters fluctuate around equilibrium. Autocorrelation of these fluctuations (figures f and l) shows that the slowest parameter, end to end distance, has a halftime of ~200 time units for linear polymers and ~20 time units for looped polymers.




Supplementary figure 2: autocorrelation for fluctuations around equilibrium of three parameters related to domain overlap between 2 labelled 10Mbp regions analysed by our virtual confocal microscope algorithm. These are the number of confocal volumes occupied by at least one monomer from the labelled reference region, the number of confocal volumes with enough reference region monomers to reach a 5 monomer detection threshold, and the amount of domain overlap measured between reference and target regions using the same detection threshold. Results show that autocorrelation half-times for the slowest parameter are approximately 200 MD time units.

# 5. Weak interactions can act cooperatively to organise interphase human nuclei

S. de Nooijer[1], J. Wellink[1], T. Bisseling[1], P.R. Cook[2], D. Marenduzzo[3], B. Mulder[4,5]

1: Laboratory for Molecular Biology, Wageningen University,
Droevendaalsesteeg 1, Wageningen, The Netherlands
2: Sir William Dunn School of Pathology, University of Oxford,
South Parks Road, Oxford, OX1 3RE, United Kingdom
3: SUPA, School of Physics, University of Edinburgh,
Mayfield Road, Edinburgh, EH9 3JZ, United Kingdom
4: FOM Institute AMOLF, Science Park 104, Amsterdam, The Netherlands
5: Laboratory for Plant Cell Biology, Wageningen University,
Droevendaalsesteeg 1, Wageningen, The Netherlands

## Abstract

We present large scale computer simulations of human interphase nuclei in which chromosomes interact with each other and with the lamina via a coarse grained force field which depends on local GC content. The interaction between GC poor segments and either the lamina or other GC poor segments is assumed to be attractive, representing the fact that these regions are heterochromatic, and the magnitude of these interactions are comparable and in the Kcal/mol range. As the coarse grained units represent 200-1,000Kbp each, these are remarkably weak forces: our simulations surprisingly show that they are nevertheless sufficient to explain published data on chromosome positioning in human lymphocytes and to reproduce in simulations the distributions of eu- and heterochromatin observed *in vivo*.

## Introduction

During interphase, human chromosomes each form separate three-dimensional territories. The spatial organisation of chromosomes during interphase plays an important role in determining gene expression and thus cell fate in eukaryotes. Gene-rich chromosomes, on which relatively more transcription takes place, are typically in the centre of the nucleus, while the gene-poor or less active ones are more peripheral (Cremer *et al*., 2001; Branco *et al*., 2008). Similarly, some genes are localized in the nuclear centre when they are transcriptionally active and at the periphery when inactive (Brown *et al*., 1999; Skok *et al*., 2001). Thus there is a relationship between nuclear localization and transcriptional activity, both on the scale of individual genes and of whole chromosomes. However, the mechanisms underlying nuclear organization are still poorly understood.

Within actively transcribed loci, different proteins and different chemical DNA and histone modifications occur than in sequences that are not transcribed (Venters & Puch, 2009). On average, gene dense regions also have different chromatin composition than gene poor regions. These variations in chromatin composition are often correlated with each other and with the amount of transcription taking place, resulting in a predominant distinction of two chromatin states: euchromatin and heterochromatin. Compared to euchromatin, heterochromatin has a relatively high volumetric DNA content, contains relatively less actively transcribed sequences and more repetitive sequences, less histone modifications that are associated with active transcription and more that are associated with transcriptional repression, a different chromatin protein content and a different GC/AT base-pair composition. These observations have led to a model in which heterochromatin is a compact, relatively inaccessible chromatin state that contributes to transcriptional silencing, whereas euchromatin is more open and thus can be more easily transcribed (Zhao *et al*., 2009). These two fractions occur as interspersed spatial domains within most human nuclei, but in some cell types (e.g. Bartova *et al*., 2002) heterochromatin occurs in chromocenters, distinct spatial domains that can be easily visualized by light microscopy after staining of DNA. In many other species, for instance in mouse and Arabidopsis, the majority of the nuclei have chromocenters. Regardless of the presence of chromocenters and in correspondence with the aforementioned spatial distribution of active genes, the interior of the nucleus is generally enriched in euchromatin, while heterochromatin is preferentially located at the nuclear periphery and near the nucleolus. One notable exception to this rule is the organization found in retina nuclei of nocturnal mammals, where heterochromatin is localized in a central sphere with peripheral euchromatin (Solovei *et al*., 2009). This special organization is thought to be an evolutionary adaptation to aid in low-light vision because the nucleus in these cells helps to focus light towards the photoreceptors underneath.

MeCP2 (Sharma *et al*., 2005) and HP1 (James & Elgin, 1986) are examples of proteins thought to be involved in the condensation of chromatin into heterochromatin, but their binding to methylated DNA and modified histones, respectively, could also cause cross-linking and thus coagulation of genomically distant heterochromatic regions. Such coagulation would then lead to the formation of heterochromatin domains and, through exclusion, euchromatin domains at other places within the nucleus. Heterochromatin has also been found to interact with the nuclear lamina (Yuan *et al*., 1991), a network of intermediate filaments that is localized at the inner surface of the nuclear membrane. These interactions, for instance mediated by the Lamin B receptor (LBR, Worman *et al*., 1988), would tether heterochromatin to the nuclear lamina. However, these interactions have only been described in qualitative terms and their effect on nuclear organization is unclear. In this chapter, we aim to study how these interactions influence nuclear organization through a modelling approach.

Modelling approaches have been employed to study several aspects of interphase chromosome folding. (Münkel *et al*., 1999) modelled human chromosomes as polymers consisting of monomers representing 200Kbp each, and then used Monte Carlo simulation methods to evaluate the folding of chromosomes into territories. In chapter 3, we have applied a similar model in Molecular Dynamics simulations to study the intermingling of different domains within a single human chromosome 11. In both these studies the chromosome was modelled as a fibre built from identical subunits, with no distinction between eu- and heterochromatin. Such a distinction was made in chapter 2, where heterochromatic chromocenters were modelled as single spheres. This method cannot be applied to human nuclei which mostly lack chromocenters, and therefore we model human chromosomes as block polymers containing monomers of two types, representing eu- and heterochromatin. All monomers interact with each other using an excluded volume interaction, and furthermore we represent the action of proteins such as MeCP2, HP1 and LBR by a short-ranged attraction of heterochromatic monomers to each other and to the nuclear periphery. Thus we show that these interactions can drive an organisation of eu- and heterochromatin that looks similar to the distribution observed in a typical human cell, and that subtle changes in relative interaction strengths can lead to switches between organizations actually found in nature. This indicates that cells can easily change their nuclear organisation by minor alterations of interaction strengths through activation or repression of the genes involved in these interactions, and therefore it is feasible for cells to control the expression of many genes through nuclear reorganisation.

## Methods

**Self-avoiding semi-flexible self-interacting chromosomes**

We model human chromosomes as flexible block polymers consisting of chains of beads connected by harmonic springs (chapter 2). Each bead represents 1Mbp in coarse-grained

simulations and 200Kbp in finer-grained simulations, and the number of beads in simulated chromosomes is chosen such that their total chain lengths correspond to their *in vivo* counterparts. All chromosomes are together confined within a 8μm diameter sphere representing the nuclear membrane and associated lamina structures. Beads are modelled as points with associated potentials creating an excluded volume around them and can have either of two types: euchromatic or heterochromatic. A good predictor for heterochromatin is the average GC to AT base-pair ratio of DNA, and to assign identity to monomers we have used this measure. All human chromosomes were divided in regions of either 200Kbp or 1Mbp, for which average GC content was determined using the UCSC human genome database. Monomers, which each represented one such region, of which the G + C content exceeded a 41% cut-off ratio were assigned euchromatic identity, whereas AT-rich monomers became heterochromatic.

Monomers interact with each other using a shifted and truncated Lennard-Jones potential (formula 1) operating between their centre points with a $\sigma$ of 2r. The same potential, but with a range of r, is used for interactions with the wall. For excluded volume interactions, we shifted the LJ potential up by 1 $k_BT$ and truncated it at 2.24r, creating a Weeks-Chandler-Andersen potential (Weeks *et al*., 1971). For partially attractive Heterochromatin-Heterochromatin and Heterochromatin-Lamina interactions, we did not shift the LJ potential up but truncated the LJ potential at 3r, creating a short-ranged attractive potential between 2r and 3r. We used $\varepsilon$ to control the strength of this attraction. The bead size parameter was used to scale the polymers such that they together occupied 10% of the nuclear volume (analogous to Cook & Marenduzzo, 2009). In 1Mbp simulations this resulted in a particle radius of 100nm and in 200Kbp simulations a radius of 58nm. Beads were connected through a harmonic spring potential with an equilibrium point of 2r and a spring constant large enough to render extensions that would allow two chains to pass through each other unlikely, empirically set to a value of $50k_BT/r^2$. These harmonic springs were used to assemble beads into linear chromosomes in all simulations, but for simulations implementing a random loop model chromosome extra springs were added between random monomers within each of the chromosomes. On average 1 loop per 20 beads was thus created.

**Molecular dynamics simulations of nuclear chromosomes**

Simulations were carried out using the ESPResSo software package (Limbach *et al*., 2006). We created initial configurations using the pseudo self avoiding walk algorithm provided in ESPResSo to create each chromosome, with a step size of 2r and a random first bead position within 5.6μm of the centre of the confining sphere, which had a diameter of 16μm at the start of simulations.

Integrations were performed using a velocity-Verlet algorithm while implementing a Langevin thermostat. The integration time step was set to 0.002 MD time units (chapter 2 for

a discussion of units). The simulations started with an equilibration phase during which the confining sphere was slowly shrunk by 0.01μm/MD time unit, until reaching its eventual size of 8μm diameter. After the confining sphere reached its eventual size, equilibration continued until parameters such as energy and end-to-end distance reached equilibrium. After equilibration samples were stored every 4 MD time units. Autocorrelation of the fluctuations around equilibrium were calculated to determine that integration should continue for at least 5000 MD time units after equilibration to obtain sufficient sampling.

## Results

**Model of the human chromosomes**

In order to investigate the effects of heterochromatic interactions on nuclear organization, we modelled the nucleus as a sphere of 8μm in diameter, with the border representing the nuclear lamina and membrane. Within that sphere, each of the 46 human chromosomes is represented as a polymer composed of a string of spherical monomers of similar diameter, with polymer length reflecting chromosomal length. Each monomer (diameter 117 or 200 nm) represents 0.2 or 1 Mbp (1,000 or 5,000 nucleosomes) and no two monomers are allowed to occupy the same volume. Monomers are not identical; the varying eu-/heterochromatic identity of monomers along each polymer reflects the variation in GC content along the corresponding chromosome. Based on GC/AT ratio a distinction into 5 classes of relatively uniform ~300Kbp regions (named isochores) was proposed, with the most important distinction between low and high GC/AT content at a 41% GC/AT ratio threshold (Costantini *et al*. 2006). We use this threshold to determine whether a monomer is eu- or heterochromatic.

Each heterochromatic monomer is given a specified attraction for the lamina and/or another heterochromatic monomer (which might be on the same or a different chromosome). We denote these two interactions as HL (heterochromatin-lamina) and HH (heterochromatin-heterochromatin) respectively. They could arise from specific interactions (e.g., hydrogen bonds, charge interactions) mediated by chromatin proteins, or be purely non-specific (entropic) in origin (e.g., because heterochromatin is more compact, and so more likely than euchromatin to contact other heterochromatin or the lamina; Cook & Marenduzzo, 2009). Finally, polymers are allowed to "diffuse" in the virtual nucleus until they reach equilibrium.

**Weak HH and HL interactions yield the organization found in nature**

Figure 5.1a illustrates the results of a typical simulation using 0.2-Mbp beads with HH and HL interactions of 1 and 1.5 $k_BT$. Heterochromatic monomers (blue) clump under the lamina – a pattern strikingly similar to the organization of heterochromatin in human nuclei. Using 1Mbp beads to speed up simulations yielded the same organization. We analysed how interaction strength affects the organization (figure 5.1b). Without any HH or HL

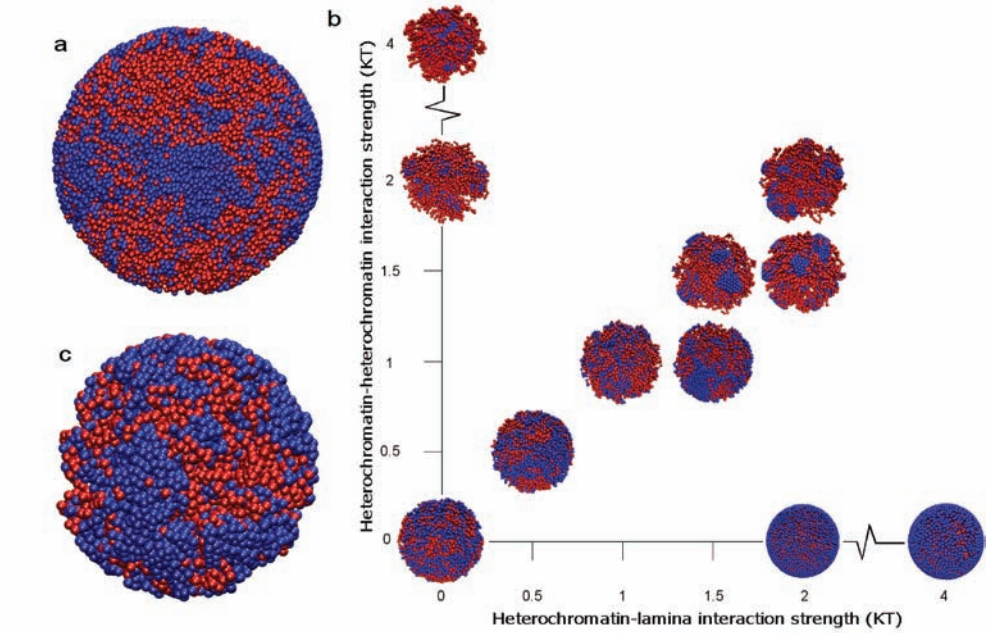



Figure 5.1: weak heterochromatin-heterochromatin (HH) and heterochromatin-lamina (HL) interactions profoundly affect nuclear organization. Blue and red beads represent hetero- and euchromatic regions. a. A snapshot of a configuration sampled from an equilibrated (MD) simulation (HH = 1 $k_BT$, HL = 1.5 $k_BT$) of unlooped polymers, with monomers representing 200Kbp each. b. "Phase" diagram showing how strengths of HH and HL interactions control nuclear organization. Polymers are unlooped and monomers represent 1Mbp each. c. A snapshot of a configuration sampled from an equilibrated (MD) simulation (HH = 1 $k_BT$, HL = 1.5 $k_BT$) with looped polymers, monomers representing 1Mbp each.

interactions, eu- and heterochromatic regions intermingle (figure 5.1b, origin). With HH interactions of 2-4 $k_BT$ (but no HL interactions), heterochromatin clumps centrally to give the "inverted" organization found only in the rod cells of diurnal mammals (Sovolei *et al*., 2009). Conversely, with HL interactions of 2-4 $k_BT$ (but no HH interactions), heterochromatin coats the periphery; although heterochromatin in nature is often peripheral, it is usually not distributed so uniformly. Only the combination of weak HH and HL interactions (i.e., on the diagonal) yields peripheral clumps of heterochromatin which, under interaction strengths of up to 1.5 $k_BT$, have a relatively loose structure reminiscent of most human nuclei, and at interaction strengths above 1.5 $k_BT$ appear as distinct chromocenters, as found in mouse nuclei (chapter 2; Sovolei *et al*., 2009).

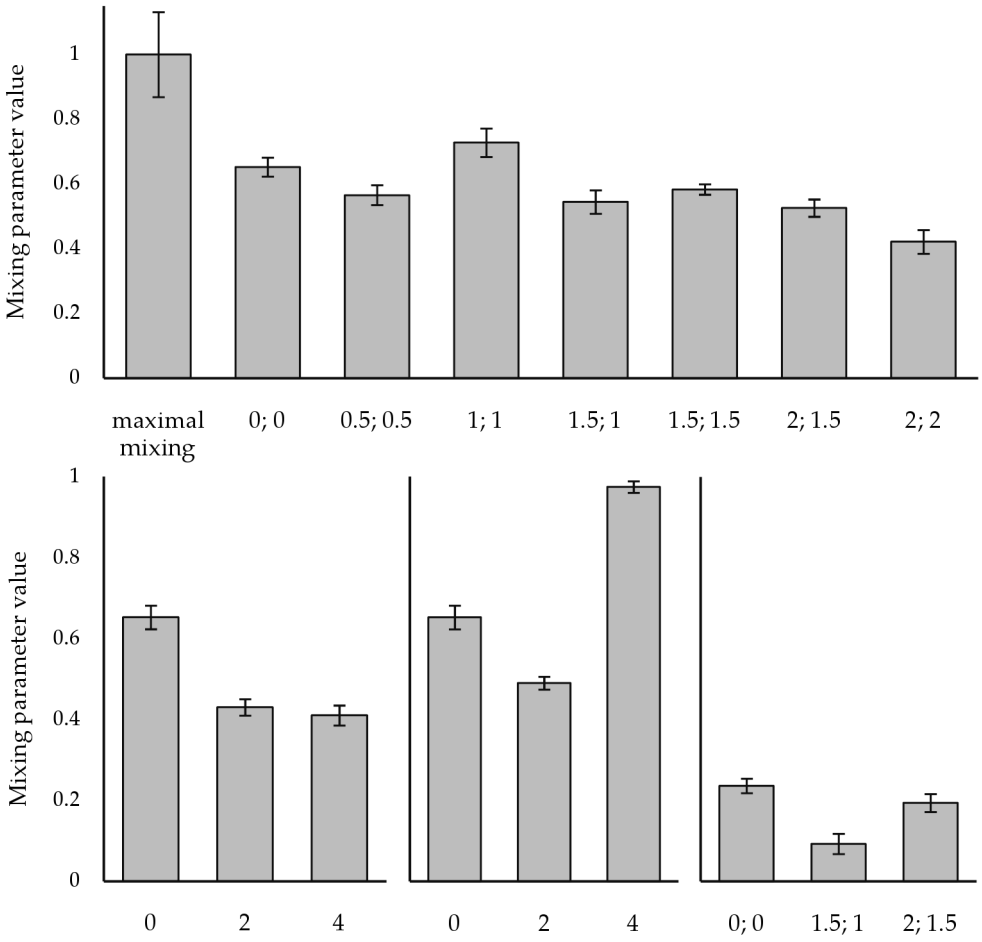



Figure 5.2: effect of varying the strength of heterochromatin-heterochromatin (HH) and heterochromatin-lamina (HL) interactions on territory formation. Each measurement was taken post-equilibration and is based on >2,000 sample configurations; each monomer represents 1Mbp. The amount of mixing is the average number of different chromosomes each monomer from each chromosome is surrounded by, normalized to a scale where 0 represents no mixing and 1 the mixing that occurs when all chromosomes are infinitesimally thin and non-interacting. Error bars denote standard deviations. Numbers on X-axes denote interaction strengths in $k_BT$, in format "HL; HH", except for B and C where only one interaction is varied and the strength of the other is 0. a. Varying the strength of both HL and HH interactions with unlooped polymers. b. Varying the strength of HL interactions only, unlooped polymers. c. Varying the strength of HH interactions only, unlooped polymers. d. Varying strengths of both interactions, with looped polymers.

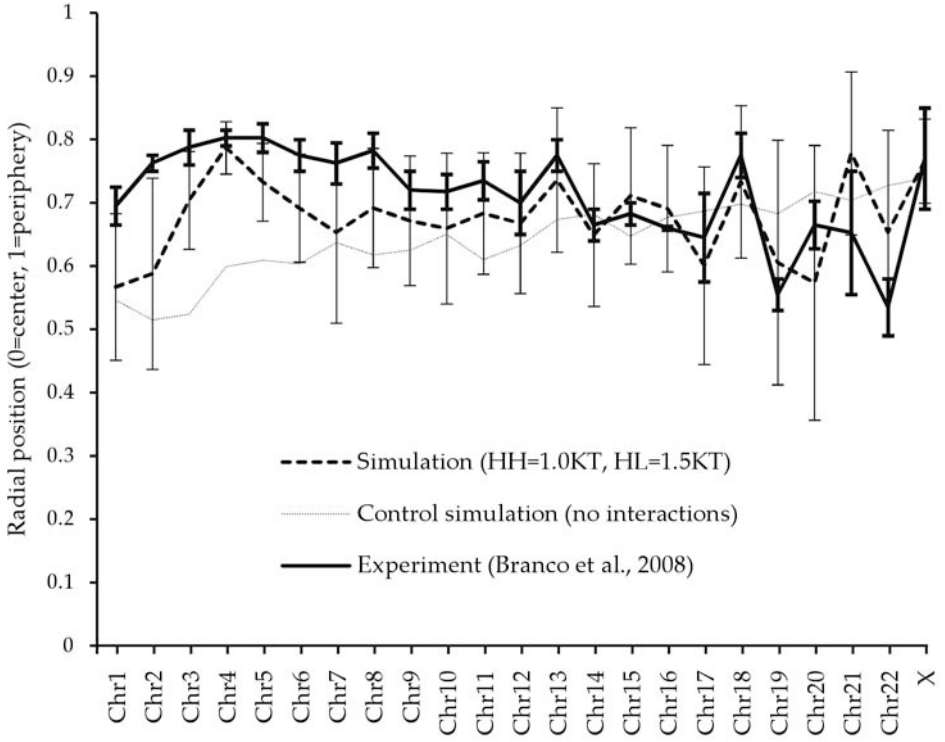



Figure 5.3: average radial position of the centers of mass of chromosomes in human lymphocytes given by simulations with HH = 1 $k_BT$ and HL = 1.5 $k_BT$ and RNA FISH (data from Branco *et al*., 2008). Grey error bars represent standard deviation in simulations and black error bars in experiments. Pearson correlation coefficient between these two curves is 0.49. Also shown is a control simulation in which there were no attractive HH and HL interactions.

**HL interactions and looping favour territory formation**

Simulations of chromosomes modelled as linear self-avoiding polymers give intermingled structures with many inter-fibre contacts (as in cooked spaghetti) and not compact territories with many intra-fibre ones (Münkel *et al*., 1999; Bohn *et al*., 2007; Jhunjhunwala *et al*., 2008; Cook and Marenduzzo, 2009; chapter 2); therefore, we examined the influence of HH and HL interactions on mixing of the polymers. Mixing was quantified by computing for each monomer the number of different chromosomes the 10 most proximal monomers belong to using the algorithm of chapter 2, taking the average over all monomers and normalizing results on a scale from 0 (discrete territories) to 1 (complete mixing given by

infinitesimally thin and non-interacting polymers). In the absence of HH or HL interactions, simulations yield a value of 0.65 (figure 5.2). HL interactions promote territory formation (figure 5.2b), while low-strength HH interactions also promote territory formation, but high-strength HH interactions increase mixing (figure 5.2c); if both are present, effects are not additive (figure 5.2a).

While HH interactions create looped polymers, loops may also be created by other mechanisms (see thesis introduction). We therefore investigate the effect of additional loops on the (nuclear) distribution of eu- and heterochromatin in the presence of HH and HL interactions. On average 30 loops per chromosome were introduced by randomly connecting monomers within each polymer (Bohn *et al*., 2007; detailed algorithm in chapter 3). Results (figure 5.1c) show no differences in eu- and heterochromatin distribution. Because simulations in previous work indicate that looping promotes territory formation (Münkel *et al*., 1999; Jhunjhunwala *et al*., 2008; Cook and Marenduzzo, 2009; chapter 2) we investigated how looped polymers mix in the presence of HH and HL interactions, and found that these further promote territory formation (figure 5.2d).

**Weak HH and HL interactions could position human chromosomes in lymphocytes**

Finally, we compared radial distributions of the centers-of-mass of chromosomes in simulations (using unlooped polymers and 1-Mbp monomers) with those found experimentally by "chromosome painting" in lymphocytes (Branco and Pombo, 2008). In the absence of HH and HL interactions, short chromosomes tend (incorrectly) to lie close to the nuclear periphery (figure 5.3), resulting in a Pearson correlation coefficient with experiments of -0.71. The localization of small chromosomes near the nuclear periphery is probably the consequence of the entropic forces described in Cook & Marenduzzo (Cook & Marenduzzo, 2009). Strikingly, adding HH and HL interactions of 1.5 and 1 $k_BT$ reverses this trend to yield a correlation coefficient of 0.45 with experiments (figure 5.3). Adding just the HL interactions yielded a correlation coefficient of 0.44, indicating that HL interactions are the main cause of the experimentally observed radial chromosome distribution. Using looped instead of unlooped polymers resulted in a similar correlation coefficient of 0.50.

## Discussion

We describe a simple model of the human genome in an interphase nucleus in which the 46 chromosomes are represented by polymers composed of strings of monomers that reflect the local GC content. Each monomer represents 0.2 or 1 Mbp of either eu- or heterochromatin. Short range attractive forces operate between heterochromatic monomers, and between heterochromatic monomers and the lamina. Whether a monomer represents eu- or heterochromatin is determined based on GC content of the sequence it represents, with a threshold value of 41%. We then find remarkably weak forces of only a few kcal/mol

can position chromosomes in the way found in the nuclei of human lymphocytes (figures 5.1-5.3). Abolishing the attraction of heterochromatin for the lamina generates the "inverted" organization found exceptionally in the rod cells of diurnal mammals (Sovolei *et al*., 2009). Under interactions that lead to a nuclear distribution of eu- and heterochromatin most resembling that of typical human cells, GC-poor chromosomes localize to the periphery in agreement with experimental data (figure 5.3, Cremer *et al*., 2001; Branco & Pombo, 2008). Simulations with HL interactions only lead to a correlation between radial chromosome positions in simulations and experiment of 0.44, whereas in the absence of interactions there is a strong anti-correlation of -0.71. Therefore the HL interaction is the strongest determinant of the radial chromosome distribution. However, standard deviations of chromosome positions in simulations are much larger than in experiments (figure 5.3). We hypothesize that in the experimental data (which are derived from a single cell type), the specific gene expression program causes several additional interactions which explain both the smaller variance in chromosomal positions in the experimental data and the differences in position of some of the chromosomes (figure 5.3). The large errors in the prediction of the localization of the smallest autosomal chromosomes, 21 and 22, might be caused by the nucleolar organizing regions they contain, which leads to association of these chromosomes with nucleoli, changing their localization.

Because the distribution of eu- and heterochromatin over chromosomes in our simulations is determined by GC content, these results further support the result of (Cremer *et al*., 2001) that the underlying DNA sequence is a major determinant of chromosomal position. Not surprisingly, a very similar eu/heterochromatin distribution was obtained when C-banding, in which the cytosine nucleotides are stained, was used as determinant for heterochromatic identity instead of GC content. Maps based on other marks associated with heterochromatin, such as H3K27 tri-methylation, mark a smaller chromosomal fraction as heterochromatin than GC ratio based maps, and were thus not used.

Our data shows that chromosome territory formation is aided by both HH and HL interactions. This result is expected because the interactions create loops, and looped polymers are less likely to mix. HH and HL interactions are less efficient at preventing intermingling of chromosomes than chromosome-specific loops because they do not discriminate between monomers of different chromosomes and/or the periphery. Also, both HH and HL interactions are in competition for space on the surface of monomers, which explains why their effect on territory formation is not cumulative (figure 5.2d). HL interactions are more efficient at stimulating territory formation because they attract chromosomes to the nuclear periphery, away from the interior where most mixing occurs. Interestingly, very strong HH interactions increase mixing, because they compete with the bonded interactions that connect monomers. This leads to the formation of stretched bonds in and especially at the surface of the central heterochromatin compartment, which allows chains to pass through each other and thus greatly increases mixing.

Our results indicate that the forces required to reorganize the nucleus are weak, even though in combination they can relocate a structure as large as a chromosome. In simulations, changes of ~0.5Kcal/mol in pairwise potentials between monomers representing 1Mbp of DNA are strong enough to form or dissociate chromocenters. However, since the attractive interactions in our simulations span up to 40 nm whereas a hydrogen bond spans less than 1 nm, stronger interactions would be required to keep two monomers attached to each other. Also, the monomers of our simulations each represent very long polymer tangles, which would need to be tethered at several different places along their length to remain near each other (chapter 3). Nevertheless, the forces required to reorganize the nucleus remain weak, and tuning such weak forces is clearly an efficient way of controlling nuclear architecture. This begs the question: which forces might be involved? The obvious candidates are the hydrogen bonds and other interactions that act between histones and other chromatin proteins to interconvert eu- and heterochromatin (Ho & Crabtree, 2010). A single hydrogen bond has a binding energy in the order of magnitude of the aforementioned 0.5Kcal/mol, and thus of the millions of potential H-bond donors and acceptors present on the surfaces of the approximately 5000 nucleosomes in a 1Mbp stretch of chromatin, only a tiny fraction would need to be involved in HH or HL interactions, probably far below 1 interaction per nucleosome. This means that it is feasible to tune these interactions through epigenetic modifications and/or through expression of proteins that mediate these interactions.

Other candidates are the weak non-specific (entropic) forces that act in crowded environments (Cook & Marenduzzo, 2009; chapter 2). Such forces are influenced by the flexibility and compaction of the chromatin fibre, which – in turn – probably depend on the underlying DNA sequence (Travers, 2004). For example, compact (heterochromatic) fibres tend to have a smaller radius of gyration and so are able to approach closer to the surrounding lamina (Cook & Marenduzzo, 2009). A third possibility involves chemical modification of many nucleosomes in a single 1-Mbp segment, and this would subtly alter some critical property of the fibre (e.g., its compaction); in turn, this would affect the scale of the non-specific (entropic) forces positioning that fibre. Whichever proves to be the case, our results suggest that modifications on only few histones could already cause the weak forces required to remodel nuclear organization, and that such modifications might subtly and indirectly affect the underlying forces. Therefore, it will be difficult to identify the molecular mechanisms reorganizing the nucleus from epigenetic data.

# 6. General discussion

## Chromatin loops

Chromosomes, with a linear length in the order of magnitude of millimetres to centimetres, fold into a three-dimensional territory during interphase. Chromosome folding determines the position of genes in the nucleus, which for many genes plays a role in the regulation of their expression (Oliver & Misteli, 2005). Chromosome folding furthermore determines through non-specific interactions the localization of nuclear sub-compartments, and thus the general spatial organization of the nucleus. An important factor in determining the way chromosomes fold is chromatin loop formation. Chromatin loops are formed when two distant regions of a chromatin fibre are physically tethered. The chromatin between these anchor points then forms a loop. Depending on the strength and dynamics of the tether, loops may be either dynamic or static. There are many biological processes which may lead to loop formation, and the size of the loops formed may vary between less than 1Kbp to a whole chromosome arm. Small loops are for example formed between genes and their regulatory elements, including enhancers/insulators and boundary elements (Kadauke & Blobel, 2009). Since these loops are typically smaller than 100Kbp, they are not relevant for the effect of looping on large-scale chromosome folding because these loops encompass only a small part of the chromosome. Large transcription-associated loops can be formed when distant genes co-localize into a transcription factory (Mitchell & Fraser, 2008). Other long-distance loops are formed when heterochromatin, consisting mostly of transcriptionally inactive sequences, co-localizes into condensed domains.

Regardless of how long-distance loops are formed, they profoundly affect folding, because loops form their own sub-territories within the nucleus. Since the loop forms a ring polymer, its chromatin is less likely to mix with other chromatin domains surrounding it (Jun & Mulder, 2006; chapter 4). Sequences within a loop will thus mainly interact with other sequences inside the loop, and less with sequences outside the loop. On the scale of whole chromosomes, this also means that looped chromosomes mix less readily with each other than unlooped chromosomes, contributing to the formation of chromosome territories (chapters 2 and 5). Our results show that the amount and size of loops, but not the specific loop architecture (the loop model), determined the amount of mixing (chapter 2). The subterritories formed by loops can also act as barriers within the nucleus, which for instance keep heterochromatic domains separate (chapters 2, 3 and 5) and restrict the spatial distribution of nuclear sub-compartments such as nucleoli (chapters 2 and 3).

While for territory formation, the specific loop architecture and the biological mechanisms of loop formation are less relevant than generic parameters such as the size and number of loops, these are relevant when the position of genes inside the nucleus is studied, or the effects of loop formation on the spatial distribution of nuclear sub-compartments. In this thesis, we performed a theoretical analysis of chromatin looping for which we used simulations implementing various polymer based models of both human and Arabidopsis chromosomes. Here, we first compare the genomes of these species to explore how the differences between them affect chromosome loop formation, and then discuss for each species the consequences of these loop models for nuclear organization.

## Comparison of human and Arabidopsis genome organization

Differences in the organization of the human and Arabidopsis genome result in differences in chromosome loop formation. The human genome contains approximately 20 times more DNA, but contains slightly less genes (23000; international human genome sequencing consortium, 2001) than Arabidopsis (26500; arabidopsis genome initiative, 2001) and more repetitive sequences. In Arabidopsis, most genes are located in euchromatin, which has an average gene density of more than 1 gene per ~5Kbp. This implies that there is very limited room for regulatory elements, which are mostly localized in a small 2Kbp region directly upstream of the transcriptional start sites of genes. Consequently, regulatory element associated looping mostly forms tiny loops of at most several Kbp, which do not influence chromosome folding. Formation of larger transcription associated loops may occur if genes localize to transcription factories, but these have not been demonstrated in Arabidopsis. Therefore, the dominant loop formation mechanism in Arabidopsis seems to be heterochromatin mediated looping. In Arabidopsis, most repetitive sequences are located around the centromeres of each chromosome, resulting in a clear separation of hetero- and euchromatin. Heterochromatin condenses into compact chromocenters during interphase (Tessadori *et al*., 2007b). This condensation is thought to be mediated by proteins which recognize specific heterochromatic sequences, chromatin modifications or heterochromatin proteins. While the identities of these proteins are not known, the result is that chromocenters are relatively compact and rigid structures, as demonstrated by their ability to resist deformation (as in Soppe *et al*., 2002). While most heterochromatin occurs around the centromeres and the nucleolar organizing regions, within the arms of chromosomes there are heterochromatic sequences. The most obvious of these is the heterochromatic knob on chromosome 4 (Fransz *et al*., 2000), but also in other regions heterochromatic repeat sequences occur interspersed in the euchromatin. If proteins occur in the nucleus that coagulate centromeric heterochromatin into chromocenters, these same proteins might recognize heterochromatic chromatin modifications on sequences within chromosome arms and coagulate these into chromocenters as well. Such a process would result in a rosette structure for Arabidopsis chromosome organization.

In contrast, the human genome consists for 45% of repetitive sequences and these occur much more dispersed throughout the chromosome arms. Human gene density is much lower than in Arabidopsis, within euchromatin 1 gene per 50Kbp. Chromocenters are not commonly found in human nuclei; instead, small heterochromatin domains occur dispersed throughout the nucleus. Here human nuclear organization differs from that of mouse and many other mammals which have chromocenters (Solovei *et al*., 2009), although the genome of mouse resembles the human genome. In chapter 5 we showed that human nuclei can form chromocenters if the strength of heterochromatic interactions is increased. Therefore, the strength of heterochromatic interactions may explain the existence of chromocenters in mouse. In human however, loop formation through associations between genomically distant stretches of heterochromatin leads to a network structure rather than a rosette structure. Although regulatory elements at a distance of hundreds of kilobases have been identified for some human genes (Palstra *et al*., 2008), resulting in the formation of some loops large enough to influence nuclear organization, still most of these loops connect sites that are genetic neighbours. Because genes regulated by the same transcription factors are located at large genomic distances, co-localization of these genes into transcription factories would create large-scale loops in a chaotic manner, leading to further network-like loop formation. Thus the resulting loop structure in human chromosomes is a network in which loops are formed between sites dispersed throughout the genome (Lieberman-Aiden *et al*., 2009), rather than a rosette structure consisting of a heterochromatic core surrounded by euchromatic loops, such as in Arabidopsis.

## A rosette model for Arabidopsis chromosome organization

In chapter 1, we show that the distribution of chromocenters over the nucleus can be explained when chromosomes are assumed to form rosette-like structures in which euchromatic loops emanate from the chromocenters. Chromocenter emanating loops of approximately 0.1-1.5Mbp have been observed in in situ hybridisation studies (Fransz *et al*., 2002), but unclear is to what extent chromosomes form such loops. Nevertheless, our simulations show that the rosette model could explain several aspects of nuclear organization without the need for introducing further specific interactions as would be required for other chromosome models.

The first of these aspects is the apparent repulsion of chromocenters. Chromocenters do not cluster, even though our results show that depletion attraction would occur between them. Therefore, a stronger force must exist which keeps chromocenters apart. In chapter 1, we hypothesized that the loops associated with each chromocenter would form a barrier that would keep chromocenters separated. In chapter 3 we characterized this force by calculating an effective potential operating between pairs of rosettes. We found that chromocenters indeed repel each other, with an interaction strength in the same order of magnitude as the depletion attraction that would keep chromocenters together. However, depletion attraction

requires prospective partners to approach each other to a distance of at most some nanometres, while the loop-mediated repulsion between chromocenters operates on a scale of micrometers. Therefore there is a significant potential barrier to overcome before chromocenters can associate.

In Arabidopsis nuclei, usually 6 to 10 separate chromocenters are observed (Soppe *et al.*, 2002), while in theory every chromosome, of which there are 10, would form its own chromocenter. Therefore, a small number of chromocenters contain heterochromatin from two different chromosomes. How is this possible when rosettes repel each other? We hypothesize that if chromocenters approach each other closely, strong short-ranged attractive interactions, both specific and non-specific, will coagulate them and that these short-ranged attractive interactions are stronger at short range than the long-ranged rosette repulsion. Coagulation is then a rare event because of the barrier potential that exists between rosettes. However, in the case that chromocenter-associated loops are not present or not formed yet, the barrier potential would be lower and coagulation would readily occur when heterochromatin from two chromosomes meets. This indeed occurs in simulations presented in chapter 2, where chromocenters of rosettes with only 3 loops coagulate.

Related to the question of why some chromocenters coagulate, is thus the question how chromosomes form rosette structures. Loops are presumably formed when certain sequences interspersed in the euchromatic arms relocate into the chromocenter (Fransz *et al.*, 2002). During mitosis, chromosomes become very condensed and have a linear structure. Chromocenters therefore might coagulate during decondensation, and later associate with the euchromatic arms of both chromosomes to form a rosette. Also, rosette loops may be transient, forming and dissociating during interphase. However, this would not change the rosette structure, because a dissociating anchor point would lead to the formation of a bigger loop out of the two flanking loops, whereas a new association of a locus with the chromocenter would split the loop it was part of in two smaller loops. As long as the majority of loops are present, rosettes would still repel each other and thus prevent association of heterochromatic sequences from one chromosome with a chromocenter belonging to another chromosome. Similarly, loops may be cell-type specific when the heterochromatic identity of anchor points differs based on transcriptional status of genes in their vicinity. This too would not affect rosette functioning as long as the average amount of loops is similar in different cell-types.

Another aspect of nuclear organization that is only correctly reproduced in simulations of models based on a rosette structure, is the central localization of nucleoli in nuclei of Arabidopsis (chapter 2). This is counter-intuitive because, being large structures, nucleoli should cluster with chromocenters at the nuclear periphery due to depletion attraction. While nucleoli tend to associate with NOR-containing chromocenters, which is caused by the specific interactions of 45S-rDNA with nucleoli, they do not associate with the other

chromocenters. In chapter 3, we show that rosette loops again provide a barrier, now between nucleoli and chromocenters, preventing their coagulation. Between the nucleolus and nuclear periphery no such barriers exist in our models, yet the simulations of chapter 2 show that in simulations implementing the rosette model, the nucleolus has a preference for the nuclear centre. We cannot fully explain this central localization from simulations implementing effective potentials (chapter 3), because the nucleolus localized mostly to the periphery in these simulations. Apparently, some other effect, possibly due to many-body interactions that are not represented in pairwise effective potentials, keeps the nucleolus in the nuclear centre in the simulations of chapter 2. In actual nuclei, there might also be additional reasons for this central localization, such as interactions with one or more NOR-containing chromosomes, which might position nucleoli in the nuclear interior.

## Loop models of human chromosomes

For human chromosomes, the model that best fits experimental data on the relationship between genomic and physical distance in interphase nuclei, is a model in which loop formation between loci within a chromosome is assumed to be random (RL model; Bohn *et al*., 2007; chapter 4 and 5). This model explains the formation of chromosome territories (see below, chapter 4) and leads to a network-like loop structure, which, as described before, is the expected structure of human chromatin based on the interactions which cause loop formation. The random formation of loops in the RL model is based on the assumption that within chromosomes, due to various loop-forming mechanisms such as heterochromatic interactions and transcription-factory associated looping, loops of all lengths occur. Then the effects of these loops on nuclear organization would be similar to loops sampled from a random distribution, especially when an ensemble of many nuclei is considered, in each of which different loops occur.

Recently, the chromatin conformation capture technique has made it possible to measure physical proximity between markers during interphase in an ensemble of nuclei. Lieberman-Aiden *et al*. (2009) studied marker proximity for several human chromosomes. The chance that a locus is in physical proximity with another locus decreases with genomic distance, and based on this observation the authors proposed a fractal globule (FG) model, in which each locus is mainly surrounded by other loci that are in close proximity. However, the FG model is not thermodynamically stable and would evolve into an equilibrium globule (chapter 3), which is less partitioned than the FG model. The FG model does not take chromatin loops into account. In chapter 3 we showed that in the RL model, the chromosome territory is more partitioned internally than in models of unlooped chromosomes that fold into an equilibrium globule. The average folding of an ensemble of different randomly looped polymers therefore resembles that of an ensemble of fractal globules, while being thermodynamically stable.

Thus the RL model best explains the data available on human chromosome folding. However, there is still no satisfying biological model for chromatin looping in human. The RL model is based on the very crude estimate that looping in real chromosomes is approximately random, but exclusively intra-chromosomal. We have made a first attempt to improve upon this estimate by introducing interactions, and thus loops, that are dynamically and randomly formed between heterochromatic domains (chapter 5). We showed that even interactions that can cross-link chromosomes can serve to separate chromosomes into territories, so the assumption of loops only being formed intra-chromosomally is not necessary. Furthermore, we showed that in simulations of model human chromosomes in which heterochromatic regions attract each other and the nuclear lamina, the position of the chromosomes is similar as observed in experiments (Branco *et al.*, 2008), which is not the case in simulations of chromosomes in which there is no attraction between heterochromatic regions and the lamina. Thus we have shown that simulation approaches which implement interactions discovered in experiments can contribute to our understanding of the forces shaping nuclear organization.

## Perspectives

The RL model can be refined based on data from genome-wide chromatin conformation capture experiments, which measure contacts between distant loci. Whenever a chromatin loop is formed, such contacts increase in frequency. Thus chromatin loop formation can be studied in an ensemble of cells, resulting in an interaction frequency map. Already the limited datasets generated so far (Liebermann-Aiden *et al.*, 2009) show clear interaction hotspots which indicate a propensity for the formation of chromatin loops at these locations. When a more elaborate genome-wide dataset is available, these data can be used to bias the loop-forming stochastic component of the RL model. Instead of implementing an equal chance for each monomer to form a bond with any other monomer, the chance to form each bond would be based on the frequency of the corresponding measured interaction in the genome-wide conformation capture dataset. Thus a new stochastic loop (SL) model would be created that, when used to simulate an ensemble of nuclei, accurately models folding as it occurs in real nuclei. Such a model can then be used as a reference to compare with experimental results on localization of individual loci, for instance genes of interest. If a deviation is then found between model and experimental data, this may point to a specific interaction operating on the gene of interest. Thus an SL model based on actual interaction data will contribute to the understanding of the relationship between nuclear structure and function.

For Arabidopsis, experimental data is now required to test the rosette model for genome organization. More specifically, the sequences or modifications underlying attachment of parts of the euchromatic chromosome arms to the centromere need to be found to verify the

model. The existing in situ hybridisation data is not suitable to verify the rosette model, and a more systematic study will be necessary. However, if loops vary between cell types and especially if loops are transiently formed, the amount of FISH experiments needed to accurately map loops in a cell population would be prohibitive. Genome-wide chromatin conformation capture studies would be able to find loop attachment points even if they are dynamic and occur only in a fraction of cells, if in the data analysis of such experiments all interactions of sites within the chromocenter are pooled and treated as one locus. Like for the human SL model proposed above, results could be used to further refine the rosette model to accurately map folding of chromosomes, and then be used as a reference for finding specific interactions operating on loci of interest.

Ultimately however, the most important contribution of models for chromosome folding lies not in the discovery of new interactions in nuclear organization, for that can only be done using biological experiments. Modelling studies are essential to understand how these interactions, both specific and non-specific, influence the organization of the nucleus. However, the goal should not be to create a big model that most accurately represents the diversity of nuclear structures that exist in nature, but should be to design a variety of small models which each include only those interactions of which the effects are under investigation and are used only to give answers to specific research questions. Using this method, we have in this thesis answered several questions regarding the mechanisms behind nuclear organization. Yet we have only scratched the surface of the amount of known interactions occurring within nuclei. All of these may potentially influence nuclear structure, and more of these are discovered every day. Therefore, modelling of nuclear organization will remain an important technique in the field for years to come.

# Samenvatting

Het belangrijkste organel in eukaryote cellen is de celkern. In de celkern vindt gentranscriptie plaats, de eerste stap in het produceren van eiwitten. Door regulatie van de genexpressie wordt bepaald welke eiwitten er in een cel beschikbaar zijn en dus wat de functie van de cel is. Fouten hierbij veroorzaken vele ziekten, en ook voor vele andere biologische en biotechnologische toepassingen is het van belang de mechanismen van regulatie van genexpressie goed te begrijpen. Regulatie van genexpressie is een complex proces waarbij zeer vele factoren een rol spelen. Een daarvan is de ruimtelijke positie en context waarbinnen expressie plaatsvindt. Om die reden is het van belang te begrijpen hoe de ruimtelijke organisatie van de celkern tot stand komt. De ruimtelijke organisatie van de celkern wordt bepaald door interacties tussen de moleculen en andere (grotere) bestanddelen waaruit een kern is opgebouwd. Veel van deze interacties zijn specifiek, dat wil zeggen dat ze optreden tussen moleculen die een specifieke bindingscapaciteit voor elkaar hebben. Voorbeelden hiervan zijn een eiwit dat alleen DNA bindt wanneer dat een bepaalde basenvolgorde heeft, of eiwitten die uitsluitend elkaar binden omdat ze precies op elkaar passen. Zulke interacties leiden tot een clustering van kernbestanddelen tot ruimtelijke sub-domeinen die, daar ze verschillende bestanddelen bevatten, elk verschillende functies uitvoeren. Maar naast deze specifieke interacties zijn er ook niet-specifieke interacties die helpen de kern te organiseren. Dit zijn interacties die altijd en overal plaatsvinden wanneer moleculen elkaar tegenkomen. In dit proefschrift wordt de bijdrage van niet-specifieke interacties aan het ontstaan van organisatie in de celkern bestudeerd.

Een belangrijke groep van niet-specifieke interacties die plaatsvinden in de celkern, zijn de zogenoemde 'uitgesloten-volume' interacties. Deze zijn belangrijk in de celkern door de hoge concentratie van macromoleculen (voornamelijk eiwitten en nucleïnezuren), die 10-40% (v/v) bedraagt. Omdat deze macromoleculen ook clusteren tot kleinere en grotere complexen en zelfs tot deeltjes die de colloïdale schaal kunnen bereiken, is vooral depletie-attractie een sterke kracht in de celkern. Depletie-attractie komt voort uit de toevallige Browniaanse bewegingen van deeltjes en leidt ertoe de grote en kleine deeltjes ruimtelijk gescheiden raken.

Een andere belangrijke groep van niet-specifieke interacties die optreden in de celkern zijn de effecten die het gevolg zijn van het opsluiten van de relatief erg lange chromosomen in een klein volume. Het chromosomale DNA zou een veel groter volume willen innemen, en daarom is het in de celkern gebonden aan eiwitten, waarmee het een complex vormt, chromatine genaamd. Ook het chromatine heeft echter een zeer hoge lengte-dikte

verhouding, en gedraagt zich daarom als een polymeer, wat gevolgen heeft voor de manier waarop het zich opvouwt. In de celkern vinden verder diverse processen plaats, zoals transcriptie en vorming van heterochromatine, die ertoe leiden dat er lussen ontstaan in het chromatine. De precieze manier waarop deze lussen gevormd worden, is van groot belang voor de manier waarop het chromatinepolymeer zich op kan vouwen in de celkern, en daarmee dus ook voor de organisatie van de celkern.

Om de effecten van niet-specifieke interacties te bestuderen, worden vaak technieken uit de statistische mechanica gebruikt. Een van deze technieken is moleculaire dynamica (MD) simulatie, waarbij de Browniaanse beweging van moleculen en deeltjes gesimuleerd wordt. Uit deze simulaties kunnen vervolgens steekproeven genomen worden van de organisatie van het gesimuleerde systeem. Met behulp van deze steekproeven kan dan afgeleid worden hoe niet-specifieke interacties de organisatie beïnvloeden. In dit proefschrift worden MD-simulaties gebruikt om het effect van chromatinelusvorming op de kernorganisatie te bestuderen. Daartoe worden modellen voor chromatinelusvorming in simulaties geïmplementeerd als ketens van monomeren. Afhankelijk van de precieze mechanismen van lusvorming kan dit simpele lussen, maar ook vertakkingen en netwerken opleveren. Door simulaties uit te voeren met elk van deze modellen, en door de hieruit genomen steekproeven te vergelijken met de organisatie van echte kernen, kunnen modellen voor chromatinelusvorming goed- of afgekeurd worden.

In hoofdstuk 2 van dit proefschrift worden de organiserende effecten van niet-specifieke interacties bestudeerd in *Arabidopsis thaliana*, de Zandraket. Deze soort is het belangrijkste modelsysteem voor de studie van planten. Eerder gepubliceerde modellen voor de organisatie van chromatine in Arabidopsis worden geïmplementeerd in MD-simulaties.. Hieruit blijkt dat niet-specifieke interacties de in celkernen gemeten posities van nucleoli en chromocenters, twee van de bestanddelen van de celkern, kunnen verklaren. Verder wordt bewezen dat chromatinelusvorming bijdraagt aan het ruimtelijk gescheiden houden van chromosomen (chromosoomterritoria). Alleen het eerder gepubliceerde rosette model blijkt alle geteste aspecten van kernorganisatie goed te verklaren. Daarom wordt in hoofdstuk 3 van dit proefschrift het rosette model nader onderzocht. De grootte van nucleoli en chromocenters wordt gevarieerd, en daarbij worden de krachten gemeten die optreden tussen deze structuren onderling, en met het omsluitende kernmembraan. Hoewel deze krachten zwak zijn, kunnen ze de chromocenters toch op hun plaats krijgen. Met name de verspreiding van rosettes over de kern kan verklaard worden vanuit de zwakke afstoting tussen deze rosettes onderling. Zo leidt dus dit proefschrift tot een beter begrip van de krachten die spelen in de celkern. Tot slot kunnen de gemeten krachten gebruikt worden voor nieuwe simulaties, die ongeveer 10.000 keer minder rekenkracht vergen.

In hoofdstuk 4 worden MD simulaties gebruikt om de vouwing van het menselijk chromosoom 11 te bestuderen. De steekproeven genomen uit deze simulaties worden

vervolgens geanalyseerd met behulp van een virtuele confocale microscoop, een algoritme dat dezelfde beperkingen heeft als een echte confocale microscoop en daardoor een directe vergelijking mogelijk maakt tussen data uit de simulatie en experimentele fluorescent *in situ* hybridisatie (FISH) data. Hieruit blijkt dat chromosoomlusvorming het volume ingenomen door een 10Mbp groot sub-domein van het chromosoom vergroot, maar dat tevens de waarneembare overlap tussen twee naburige sub-domeinen kleiner wordt. Verder laten de resultaten zien dat de hoeveelheid overlap, gemeten met een confocale microscoop, erg gevoelig is voor veranderingen in de detectielimiet en de grootte van het confocaal volume van de microscoop. Dit zou betekenen dat in veel FISH experimenten de hoeveelheid gemeten overlap tussen twee naburige domeinen veel lager is dan in werkelijkheid. Op basis van de resultaten gepresenteerd in dit proefschrift kunnen de experimentele protocollen voor dergelijke proeven verbeterd en gestandaardiseerd worden.

Naast de vouwing van een enkel chromosoom, is ook het samenspel van specifieke en niet-specifieke interacties bij de vouwing van alle menselijke chromosomen samen onderzocht (hoofdstuk 5). Het menselijk genoom kan opgedeeld worden in twee gebieden, euchromatine, waar de meeste genen liggen, en heterochromatine, waar minder genen liggen. Er zijn specifieke interacties bekend tussen heterochromatine en de lamina, een netwerk van eiwitten bij het kernmembraan. Ook zijn er interacties bekend tussen heterochromatische gebieden onderling. In hoofdstuk 5 is onderzocht of deze twee specifieke interacties, samen met niet-specifieke interacties, de organisatie van menselijke kernen kunnen verklaren. Daartoe wordt een model van 23 paar menselijke chromosomen gesimuleerd, die elk bestaan uit eu- en heterochromatische gebieden. De heterochromatische gebieden ondergaan in deze simulatie een attractieve interactie met elkaar en met het kernmembraan. Afhankelijk van de sterkte van deze interacties, blijkt een breed spectrum aan kernorganisatie te ontstaan, en diverse van deze organisaties zijn ook in echte kernen waargenomen. In de simulaties die het meest overeenkomen met de organisatie van menselijke kernen blijkt bovendien de in echte kernen gemeten radiale verdeling van chromosomen over de kern gereproduceerd te worden. Deze resultaten laten zien dat een cel zijn kernorganisatie eenvoudig kan aanpassen door de sterktes van deze interacties aan te passen, bijvoorbeeld door middel van het wijzigen van de expressie van een enkel eiwit of door eiwitmodificatie.

Tot slot worden in hoofdstuk 6 de implicaties van de in dit proefschrift gepresenteerde resultaten voor de mechanismen van kernorganisatie en modellen voor chromosoomvouwing bediscussieerd. De verschillen tussen chromosoomvouwing in de mens en in Arabidopsis blijken terug te voeren op verschillen in genoomstructuur, maar de krachten en organiserende principes die een rol spelen in de kernorganisatie, blijken niet wezenlijk te verschillen tussen de beide soorten. Het inzicht hoe niet-specifieke en specifieke krachten samen chromosomen vouwen en sub-compartimenten positioneren, is daarmee de belangrijkste bijdrage van dit proefschrift aan de wetenschap.

# Resume

Silvester de Nooijer was born in Goes, the Netherlands, on 3 November 1982. He has attended high school (Scholengemeenschap Nehalennia) in Middelburg from 1994, graduating with honours (*cum laude*) in 2000. Between 2000 and 2006, he studied at Wageningen University, successfully completing the BSc curriculum 'Moleculaire Wetenschappen' and the MSc curriculum 'Molecular Sciences', specialisation 'Biological Chemistry' (*cum laude*). These studies resulted in four theses and internships: at the cytogenetics group of prof. dr. Hans de Jong, Wageningen UR, at the laboratory for molecular biology in the group of dr. R. Geurts and the laboratory for plant cell biology in the group of prof. dr. A. Emons, at the FOM institute AMOLF in the group of prof. dr. B. Mulder, and at the Allan Wilson centre, Massey university, New Zealand, in the group of prof. dr. D. Penny. Three of these have resulted in publications in international scientific journals.

In 2006, Silvester returned to the laboratory of molecular biology in Wageningen as a PhD student, which resulted in this thesis. His project, in the writing of which he was intimately involved, was to study the role of non-specific interactions in the organization of Arabidopsis nuclei. Later, the project was extended to include the influence of non-specific interactions on the organization of human nuclei. As part of this work, Silvester spent 3 weeks as a guest at the EPCC, university of Edinburgh.

# Publications

1. Silvester de Nooijer, Joan Wellink, Bela Mulder, Ton Bisseling (2009) Non-specific interactions are sufficient to explain the position of heterochromatic chromocenters and nucleoli in interphase nuclei. Nucleic Acids Res. 37(11):3558-68.

2. Silvester de Nooijer, Barbara Holland, David Penny (2009) The emergence of predators in early life: there was no Garden of Eden. PLoS One. 4(6):e5507.

3. Silvester de Nooijer, Tijs Ketelaar, Anne Mie Emons and Bela Mulder (2008) Sphere size distributions from finite thickness sections: A forward approach using a genetic algorithm. Journal of Microscopy 231, 257-264.

4. Joke van Vugt, Silvester de Nooijer, Richard Stouthamer and Hans de Jong (2005) NOR activity and repeat sequences of the paternal sex ratio chromosome of the parasitoid wasp Trichogramma kaykai. Chromosoma 114, 410-419.

# Acknowledgements

# Bibliography

Arnold, A., and Jun, S. (2007). Time scale of entropic segregation of flexible polymers in confinement: implications for chromosome segregation in filamentous bacteria. Phys Rev E Stat Nonlin Soft Matter Phys 76, 031901.

Bancaud, A., Huet, S., Daigle, N., Mozziconacci, J., Beaudouin, J., and Ellenberg, J. (2009). Molecular crowding affects diffusion and binding of nuclear proteins in heterochromatin and reveals the fractal organization of chromatin. Embo J 28, 3785-3798.

Baricheva, E.A., Berrios, M., Bogachev, S.S., Borisevich, I.V., Lapik, E.R., Sharakhov, I.V., Stuurman, N., and Fisher, P.A. (1996). DNA from Drosophila melanogaster beta-heterochromatin binds specifically to nuclear lamins *in vitro* and the nuclear envelope in situ. Gene 171, 171-176.

Bartova, E., Krejci, J., Harnicarova, A., Galiova, G., and Kozubek, S. (2008). Histone modifications and nuclear architecture: a review. J Histochem Cytochem 56, 711-721.

Bartova, E., Kozubek, S., Jirsova, P., Kozubek, M., Gajova, H., Lukasova, E., Skalnikova, M., Ganova, A., Koutna, I., and Hausmann, M. (2002). Nuclear structure and gene activity in human differentiated cells. J Struct Biol 139, 76-89.

Bennett, C. (1976). Efficient estimation of free energy differences from Monte Carlo data. J. Comp. Phys. 22, 245-268.

Berr, A., and Schubert, I. (2007). Interphase chromosome arrangement in *Arabidopsis thaliana* is similar in differentiated and meristematic tissues and shows a transient mirror symmetry after nuclear division. Genetics 176, 853-863.

Bode, J., Benham, C., Ernst, E., Knopp, A., Marschalek, R., Strick, R., and Strissel, P. (2000). Fatal connections: when DNA ends meet on the nuclear matrix. J Cell Biochem Suppl Suppl 35, 3-22.

Bohn, M., Heermann, D.W., and van Driel, R. (2007). Random loop model for long polymers. Phys Rev E Stat Nonlin Soft Matter Phys 76, 051805.

Bohrmann, B., Haider, M., and Kellenberger, E. (1993). Concentration evaluation of chromatin in unstained resin-embedded sections by means of low-dose ratio-contrast imaging in STEM. Ultramicroscopy 49, 235-251.

Bolhuis, P., and Louis, A. (2002). How To Derive and Parameterize Effective Potentials in Colloid−Polymer Mixtures. Macromolecules 35, 1860-1869.

Branco, M.R., and Pombo, A. (2006). Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. PLoS Biol 4, e138.

Branco, M.R., and Pombo, A. (2007). Chromosome organization: new facts, new models. Trends Cell Biol 17, 127-134.

Branco, M.R., Branco, T., Ramirez, F., and Pombo, A. (2008). Changes in chromosome organization during PHA-activation of resting human lymphocytes measured by cryo-FISH. Chromosome Res 16, 413-426.

Brown, K.E., Baxter, J., Graf, D., Merkenschlager, M., and Fisher, A.G. (1999). Dynamic repositioning of genes in the nucleus of lymphocytes preparing for cell division. Mol Cell 3, 207-217.

Bystricky, K., Heun, P., Gehlen, L., Langowski, J., and Gasser, S.M. (2004). Long-range compaction and flexibility of interphase chromatin in budding yeast analyzed by high-resolution imaging techniques. Proc Natl Acad Sci USA 101, 16495-16500.

Cacciuto, A., and Luijten, E. (2006). Self-avoiding flexible polymers under spherical confinement. Nano Lett 6, 901-905.

Calikowski, T.T., Meulia, T., and Meier, I. (2003). A proteomic study of the arabidopsis nuclear matrix. J Cell Biochem 90, 361-378.

Castano, E., Philimonenko, V.V., Kahle, M., Fukalova, J., Kalendova, A., Yildirim, S., Dzijak, R., Dingova-Krasna, H., and Hozak, P. (2010). Actin complexes in the cell nucleus: new stones in an old field. Histochem Cell Biol 133, 607-626.

Ceccarelli, M., Morosi, L., and Cionini, P. (1997). Chromocenter association in plant cell nuclei: determinants, functional significance, and evolutionary implications. Genome 41, 96-103.

Chambeyron, S., and Bickmore, W.A. (2004). Does looping and clustering in the nucleus regulate gene expression? Curr Opin Cell Biol 16, 256-262.

Cheutin, T., McNairn, A.J., Jenuwein, T., Gilbert, D.M., Singh, P.B., and Misteli, T. (2003). Maintenance of stable heterochromatin domains by dynamic HP1 binding. Science 299, 721-725.

Chubb, J.R., and Bickmore, W.A. (2003). Considering nuclear compartmentalization in the light of nuclear dynamics. Cell 112, 403-406.

Cook, P.R. (2002). Predicting three-dimensional genome structure from transcriptional activity. Nat Genet 32, 347-352.

Cook, P.R., and Marenduzzo, D. (2009). Entropic organization of interphase chromosomes. J Cell Biol 186, 825-834.

Costantini, M., Clay, O., Auletta, F., and Bernardi, G. (2006). An isochore map of human chromosomes. Genome Res 16, 536-541.

Cremer, M., von Hase, J., Volm, T., Brero, A., Kreth, G., Walter, J., Fischer, C., Solovei, I., Cremer, C., and Cremer, T. (2001). Non-random radial higher-order chromatin arrangements in nuclei of diploid human cells. Chromosome Res 9, 541-567.

Cremer, T., and Cremer, C. (2001). Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nat Rev Genet 2, 292-301.

Cremer, T., Cremer, C., Schneider, T., Baumann, H., Hens, L., and Kirsch-Volders, M. (1982). Analysis of chromosome positions in the interphase nucleus of Chinese hamster cells by laser-UV-microirradiation experiments. Hum Genet 62, 201-209.

Cremer, T., Kreth, G., Koester, H., Fink, R.H., Heintzmann, R., Cremer, M., Solovei, I., Zink, D., and Cremer, C. (2000). Chromosome territories, interchromatin domain compartment, and nuclear matrix: an integrated view of the functional nuclear architecture. Crit Rev Eukaryot Gene Expr 10, 179-212.

Davie, J.R. (1995). The nuclear matrix and the regulation of chromatin organization and function. Int Rev Cytol 162A, 191-250.

de Gennes, P. (1979). Scaling concepts in polymer physics. (Ithaca, New York: Cornell University Press).

de Nooijer, S., Wellink, J., Mulder, B., and Bisseling, T. (2009). Non-specific interactions are sufficient to explain the position of heterochromatic chromocenters and nucleoli in interphase nuclei. Nucleic Acids Res 37, 3558-3568.

de Wit, E., and van Steensel, B. (2009). Chromatin domains in higher eukaryotes: insights from genome-wide mapping studies. Chromosoma 118, 25-36.

Dernburg, A.F., Broman, K.W., Fung, J.C., Marshall, W.F., Philips, J., Agard, D.A., and Sedat, J.W. (1996). Perturbation of nuclear architecture by long-distance chromosome interactions. Cell 85, 745-759.

Dietzel, S., Jauch, A., Kienle, D., Qu, G., Holtgreve-Grez, H., Eils, R., Munkel, C., Bittner, M., Meltzer, P.S., Trent, J.M., and Cremer, T. (1998). Separate and variably shaped chromosome arm domains are disclosed by chromosome arm painting in human cell nuclei. Chromosome Res 6, 25-33.

Dorier, J., and Stasiak, A. (2009). Topological origins of chromosomal territories. Nucleic Acids Res 37, 6316-6322.

Dorigo, B., Schalch, T., Kulangara, A., Duda, S., Schroeder, R.R., and Richmond, T.J. (2004). Nucleosome arrays reveal the two-start organization of the chromatin fiber. Science 306, 1571-1573.

Eltsov, M., Maclellan, K.M., Maeshima, K., Frangakis, A.S., and Dubochet, J. (2008). Analysis of cryo-electron microscopy images does not support the existence of 30-nm chromatin fibers in mitotic chromosomes in situ. Proc Natl Acad Sci U S A 105, 19732-19737.

Engel, J.D., and Tanimoto, K. (2000). Looping, linking, and chromatin activity: new insights into beta-globin locus regulation. Cell 100, 499-502.

Fang, Y., and Spector, D.L. (2005). Centromere positioning and dynamics in living Arabidopsis plants. Mol Biol Cell 16, 5710-5718.

Fedorova, E., and Zink, D. (2008). Nuclear architecture and gene regulation. Biochim Biophys Acta 1783, 2174-2184.

Finch, J.T., and Klug, A. (1976). Solenoidal model for superstructure in chromatin. Proc Natl Acad Sci U S A 73, 1897-1901.

Finlan, L.E., Sproul, D., Thomson, I., Boyle, S., Kerr, E., Perry, P., Ylstra, B., Chubb, J.R., and Bickmore, W.A. (2008). Recruitment to the nuclear periphery can alter expression of genes in human cells. PLoS Genet 4, e1000039.

Fischer, C., Bouneau, L., Coutanceau, J.P., Weissenbach, J., Volff, J.N., and Ozouf-Costaz, C. (2004). Global heterochromatic colocalization of transposable elements with minisatellites in the compact genome of the pufferfish Tetraodon nigroviridis. Gene 336, 175-183.

Fiserova, J., and Goldberg, M.W. (2010). Relationships at the nuclear envelope: lamins and nuclear pore complexes in animals and plants. Biochem Soc Trans 38, 829-831.

Flemming, W. (1880). Beitraege zur kenntniss der Zelle und ihrer Lebenserscheinungen, Theil II. . Archiv fuer Mikroskopische Anatomie 18, 151-259.

Fox, A.H., Lam, Y.W., Leung, A.K., Lyon, C.E., Andersen, J., Mann, M., and Lamond, A.I. (2002). Paraspeckles: a novel nuclear domain. Curr Biol 12, 13-25.

Fransz, P., De Jong, J.H., Lysak, M., Castiglione, M.R., and Schubert, I. (2002). Interphase chromosomes in Arabidopsis are organized as well defined chromocenters from which euchromatin loops emanate. Proc Natl Acad Sci USA 99, 14584-14589.

Fransz, P.F., Armstrong, S., de Jong, J.H., Parnell, L.D., van Drunen, C., Dean, C., Zabel, P., Bisseling, T., and Jones, G.H. (2000). Integrated cytogenetic map of chromosome arm 4S of *A. thaliana*: structural organization of heterochromatic knob and centromere region. Cell 100, 367-376.

Frenkel, D., and Smit, B. (1995). Understanding Molecular Simulations: From Algorithms to Applications. (San Diego: Academic Press).

Furukawa, K., Pante, N., Aebi, U., and Gerace, L. (1995). Cloning of a cDNA for lamina-associated polypeptide 2 (LAP2) and identification of regions that specify targeting to the nuclear envelope. Embo J 14, 1626-1636.

Geiman, T.M., and Robertson, K.D. (2002). Chromatin remodeling, histone modifications, and DNA methylation-how does it all fit together? J Cell Biochem 87, 117-125.

Goetze, S., Mateos-Langerak, J., Gierman, H.J., de Leeuw, W., Giromus, O., Indemans, M.H., Koster, J., Ondrej, V., Versteeg, R., and van Driel, R. (2007). The three-dimensional structure of human interphase chromosomes is related to the transcriptome map. Mol Cell Biol 27, 4475-4487.

Guarda, A., Bolognese, F., Bonapace, I.M., and Badaracco, G. (2009). Interaction between the inner nuclear membrane lamin B receptor and the heterochromatic methyl binding protein, MeCP2. Exp Cell Res 315, 1895-1903.

Hancock, R. (2004). Internal organisation of the nucleus: assembly of compartments by macromolecular crowding and the nuclear matrix model. Biol Cell 96, 595-601.

Handwerger, K.E., and Gall, J.G. (2006). Subnuclear organelles: new insights into form and function. Trends Cell Biol 16, 19-26.

Hansen, J., Addison, C., and Louis, A. (2005). Polymer solutions: from hard monomers to soft polymers. J. Phys.: Condens. Matter 17, S3185-S3193.

Hebert, M.D., and Matera, A.G. (2000). Self-association of coilin reveals a common theme in nuclear body localization. Mol Biol Cell 11, 4159-4171.

Heitz, E. (1928). Das Heterochromatin der Moose. Jahrb. Wiss. Botanik. 69, 762-818.

Heun, P., Laroche, T., Shimada, K., Furrer, P., and Gasser, S.M. (2001). Chromosome dynamics in the yeast interphase nucleus. Science 294, 2181-2186.

Ho, L., and Crabtree, G.R. (2010). Chromatin remodelling during development. Nature 463, 474-484.

Ishov, A.M., Sotnikov, A.G., Negorev, D., Vladimirova, O.V., Neff, N., Kamitani, T., Yeh, E.T., Strauss, J.F., 3rd, and Maul, G.G. (1999). PML is critical for ND10 formation and recruits the PML-interacting protein daxx to this nuclear structure when modified by SUMO-1. J Cell Biol 147, 221-234.

Jackson, D.A., and Cook, P.R. (1995). The structural basis of nuclear function. Int Rev Cytol 162A, 125-149.

Jackson, D.A., Hassan, A.B., Errington, R.J., and Cook, P.R. (1993). Visualization of focal sites of transcription within human nuclei. Embo J 12, 1059-1065.

Jackson, D.A., Iborra, F.J., Manders, E.M., and Cook, P.R. (1998). Numbers and organization of RNA polymerases, nascent transcripts, and transcription units in HeLa nuclei. Mol Biol Cell 9, 1523-1536.

James, T.C., and Elgin, S.C. (1986). Identification of a nonhistone chromosomal protein associated with heterochromatin in Drosophila melanogaster and its gene. Mol Cell Biol 6, 3862-3872.

Jhunjhunwala, S., van Zelm, M.C., Peak, M.M., Cutchin, S., Riblet, R., van Dongen, J.J., Grosveld, F.G., Knoch, T.A., and Murre, C. (2008). The 3D structure of the immunoglobulin heavy-chain locus: implications for long-range genomic interactions. Cell 133, 265-279.

Jun, S., and Mulder, B. (2006). Entropy-driven spatial organization of highly confined polymers: lessons for the bacterial chromosome. Proc Natl Acad Sci USA 103, 12388-12393.

Kadauke, S., and Blobel, G.A. (2009). Chromatin loops in gene regulation. Biochim Biophys Acta 1789, 17-25.

Kohler, A., and Hurt, E. (2010). Gene regulation by nucleoporins and links to cancer. Mol Cell 38, 6-15.

Kumaran, R.I., and Spector, D.L. (2008). A genetic locus targeted to the nuclear periphery in living cells maintains its transcriptional competence. J Cell Biol 180, 51-65.

Kurukuti, S., Tiwari, V.K., Tavoosidana, G., Pugacheva, E., Murrell, A., Zhao, Z., Lobanenkov, V., Reik, W., and Ohlsson, R. (2006). CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. Proc Natl Acad Sci USA 103, 10684-10689.

Lallemand-Breitenbach, V., and de The, H. (2010). PML nuclear bodies. Cold Spring Harb Perspect Biol 2, a000661.

Langowski, J. (2006). Polymer chain models of DNA and chromatin. Eur Phys J E Soft Matter 19, 241-249.

Langowski, J., and Heermann, D.W. (2007). Computational modeling of the chromatin fiber. Semin Cell Dev Biol 18, 659-667.

Leimgruber, E., Seguin-Estevez, Q., Dunand-Sauthier, I., Rybtsova, N., Schmid, C.D., Ambrosini, G., Bucher, P., and Reith, W. (2009). Nucleosome eviction from MHC class II promoters controls positioning of the transcription start site. Nucleic Acids Res 37, 2514-2528.

Li, G., Margueron, R., Hu, G., Stokes, D., Wang, Y.H., and Reinberg, D. (2010). Highly compacted chromatin formed *in vitro* reflects the dynamics of transcription activation *in vivo*. Mol Cell 38, 41-53.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., Sandstrom, R., Bernstein, B., Bender, M.A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L.A., Lander, E.S., and Dekker, J. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326, 289-293.

Limbach, H., Arnold, A., Mann, B., and Holm, C. (2006). ESPResSo - An Extensible Simulation Package for Research on Soft Matter Systems. Comput. Phys. Commun. 174, 704-727.

Lorson, C.L., Hahnen, E., Androphy, E.J., and Wirth, B. (1999). A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. Proc Natl Acad Sci USA 96, 6307-6311.

Marella, N.V., Bhattacharya, S., Mukherjee, L., Xu, J., and Berezney, R. (2009). Cell type specific chromosome territory organization in the interphase nucleus of normal and cancer cells. J Cell Physiol 221, 130-138.

Marenduzzo, D., Finan, K., and Cook, P.R. (2006a). The depletion attraction: an underappreciated force driving cellular organization. J Cell Biol 175, 681-686.

Marenduzzo, D., Micheletti, C., and Cook, P.R. (2006b). Entropy-driven genome organization. Biophys J 90, 3712-3721.

Mateos-Langerak, J., Bohn, M., de Leeuw, W., Giromus, O., Manders, E.M., Verschure, P.J., Indemans, M.H., Gierman, H.J., Heermann, D.W., van Driel, R., and Goetze, S. (2009). Spatially confined folding of chromatin in the interphase nucleus. Proc Natl Acad Sci U S A 106, 3812-3817.

Mattioli, E., Columbaro, M., Capanni, C., Santi, S., Maraldi, N.M., D'Apice, M.R., Novelli, G., Riccio, M., Squarzoni, S., Foisner, R., and Lattanzi, G. (2008). Drugs affecting prelamin A processing: effects on heterochromatin organization. Exp Cell Res 314, 453-462.

Misteli, T. (2001). The concept of self-organization in cellular architecture. J Cell Biol 155, 181-185.

Misteli, T. (2005). Concepts in nuclear architecture. Bioessays 27, 477-487.

Mitchell, J.A., and Fraser, P. (2008). Transcription factories are nuclear subcompartments that remain in the absence of transcription. Genes Dev 22, 20-25.

Morris, G.E. (2008). The Cajal body. Biochim Biophys Acta 1783, 2108-2115.

Munkel, C., and Langowski, J. (1998). Chromosome structure predicted by a polymer model. Physical reviews E. 57, 5888-5896.

Munkel, C., Eils, R., Dietzel, S., Zink, D., Mehring, C., Wedemann, G., Cremer, T., and Langowski, J. (1999). Compartmentalization of interphase chromosomes observed in simulation and experiment. J Mol Biol 285, 1053-1065.

Oliver, B., and Misteli, T. (2005). A non-random walk through the genome. Genome Biol 6, 214.

Osborne, C.S., Chakalova, L., Brown, K.E., Carter, D., Horton, A., Debrand, E., Goyenechea, B., Mitchell, J.A., Lopes, S., Reik, W., and Fraser, P. (2004). Active genes dynamically colocalize to shared sites of ongoing transcription. Nat Genet 36, 1065-1071.

Palstra, R.J. (2009). Close encounters of the 3C kind: long-range chromatin interactions and transcriptional regulation. Brief Funct Genomic Proteomic 8, 297-309.

Palstra, R.J., de Laat, W., and Grosveld, F. (2008). Beta-globin regulation and long-range interactions. Adv Genet 61, 107-142.

Pecinka, A., Schubert, V., Meister, A., Kreth, G., Klatte, M., Lysak, M.A., Fuchs, J., and Schubert, I. (2004). Chromosome territory arrangement and homologous pairing in nuclei of *Arabidopsis thaliana* are predominantly random except for NOR-bearing chromosomes. Chromosoma 113, 258-269.

Rajapakse, I., Perlman, M.D., Scalzo, D., Kooperberg, C., Groudine, M., and Kosak, S.T. (2009). The emergence of lineage-specific chromosomal topologies from coordinate gene regulation. Proc Natl Acad Sci U S A 106, 6679-6684.

Reddy, K.L., Zullo, J.M., Bertolino, E., and Singh, H. (2008). Transcriptional repression mediated by repositioning of genes to the nuclear lamina. Nature 452, 243-247.

Rivas, G., Ferrone, F., and Herzfeld, J. (2004). Life in a crowded world. EMBO Rep 5, 23-27.

Rosa, A., and Everaers, R. (2008). Structure and dynamics of interphase chromosomes. PLoS Comput Biol 4, e1000153.

Rosenfeld, J.A., Wang, Z., Schones, D.E., Zhao, K., DeSalle, R., and Zhang, M.Q. (2009). Determination of enriched histone modifications in non-genic portions of the human genome. BMC Genomics 10, 143.

Sachs, R.K., van den Engh, G., Trask, B., Yokota, H., and Hearst, J.E. (1995). A random-walk/giant-loop model for interphase chromosomes. Proc Natl Acad Sci U S A 92, 2710-2714.

Schalch, T., Duda, S., Sargent, D.F., and Richmond, T.J. (2005). X-ray structure of a tetranucleosome and its implications for the chromatin fibre. Nature 436, 138-141.

Schoenfelder, S., Sexton, T., Chakalova, L., Cope, N.F., Horton, A., Andrews, S., Kurukuti, S., Mitchell, J.A., Umlauf, D., Dimitrova, D.S., Eskiw, C.H., Luo, Y., Wei, C.L., Ruan, Y., Bieker, J.J., and Fraser, P. (2010). Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. Nat Genet 42, 53-61.

Sereda, Y.V., and Bishop, T.C. (2010). Evaluation of elastic rod models with long range interactions for predicting nucleosome stability. J Biomol Struct Dyn 27, 867-887.

Sharma, R.P., Grayson, D.R., Guidotti, A., and Costa, E. (2005). Chromatin, DNA methylation and neuron gene regulation--the purpose of the package. J Psychiatry Neurosci 30, 257-263.

Shin, K., Obukhov, S., Chen, J.T., Huh, J., Hwang, Y., Mok, S., Dobriyal, P., Thiyagarajan, P., and Russell, T.P. (2007). Enhanced mobility of confined polymers. Nat Mater 6, 961-965.

Simonis, M., and de Laat, W. (2008). FISH-eyed and genome-wide views on the spatial organisation of gene expression. Biochim Biophys Acta 1783, 2052-2060.

Simonis, M., Kooren, J., and de Laat, W. (2007). An evaluation of 3C-based methods to capture DNA interactions. Nat Methods 4, 895-901.

Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B., and de Laat, W. (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat Genet 38, 1348-1354.

Skok, J.A., Brown, K.E., Azuara, V., Caparros, M.L., Baxter, J., Takacs, K., Dillon, N., Gray, D., Perry, R.P., Merkenschlager, M., and Fisher, A.G. (2001). Nonequivalent nuclear location of immunoglobulin alleles in B lymphocytes. Nat Immunol 2, 848-854.

Solovei, I., Grandi, N., Knoth, R., Volk, B., and Cremer, T. (2004). Positional changes of pericentromeric heterochromatin and nucleoli in postmitotic Purkinje cells during murine cerebellum development. Cytogenet Genome Res. 105, 302-310.

Solovei, I., Kreysing, M., Lanctot, C., Kosem, S., Peichl, L., Cremer, T., Guck, J., and Joffe, B. (2009). Nuclear architecture of rod photoreceptor cells adapts to vision in mammalian evolution. Cell 137, 356-368.

Soppe, W.J., Jasencakova, Z., Houben, A., Kakutani, T., Meister, A., Huang, M.S., Jacobsen, S.E., Schubert, I., and Fransz, P.F. (2002). DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in Arabidopsis. Embo J 21, 6549-6559.

Soutoglou, E., and Misteli, T. (2007). Mobility and immobility of chromatin in transcription and genome stability. Curr Opin Genet Dev 17, 435-442.

Spector, D.L. (2001). Nuclear domains. J Cell Sci 114, 2891-2893.

Spector, D.L. (2003). The dynamics of chromosome organization and gene regulation. Annu Rev Biochem 72, 573-608.

Sutherland, H., and Bickmore, W.A. (2009). Transcription factories: gene expression in unions? Nat Rev Genet 10, 457-466.

Taddei, A. (2007). Active genes at the nuclear pore complex. Curr Opin Cell Biol 19, 305-310.

Taddei, A., Hediger, F., Neumann, F.R., and Gasser, S.M. (2004). The function of nuclear architecture: a genetic approach. Annu Rev Genet 38, 305-345.

Teixeira, F.K., and Colot, V. (2010). Repeat elements and the Arabidopsis DNA methylation landscape. Heredity 105, 14-23.

Tessadori, F., Schulkes, R.K., van Driel, R., and Fransz, P. (2007a). Light-regulated large-scale reorganization of chromatin during the floral transition in Arabidopsis. Plant J 50, 848-857.

Tessadori, F., Chupeau, M.C., Chupeau, Y., Knip, M., Germann, S., van Driel, R., Fransz, P., and Gaudin, V. (2007b). Large-scale dissociation and sequential reassembly of pericentric heterochromatin in dedifferentiated Arabidopsis cells. J Cell Sci 120, 1200-1208.

The Arabidopsis Genome Sequencing Initiative. (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408, 796-815.

The Human Genome Sequencing Initiative. (2001). Initial sequencing and analysis of the human genome. Nature 409, 860-921.

Toan, N.M., Marenduzzo, D., Cook, P.R., and Micheletti, C. (2006). Depletion effects and loop formation in self-avoiding polymers. Phys Rev Lett 97, 178302.

Torrie, G., and Valleau, J. (1977). Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. J. Comp. Phys. 23, 187-199.

Tremethick, D.J. (2007). Higher-order structures of chromatin: the elusive 30 nm fiber. Cell 128, 651-654.

van Holde, K., and Zlatanova, J. (2007). Chromatin fiber structure: Where is the problem now? Semin Cell Dev Biol 18, 651-658.

Venters, B.J., and Pugh, B.F. (2009). How eukaryotic genes are transcribed. Crit Rev Biochem Mol Biol 44, 117-141.

Virstedt, J., Berge, T., Henderson, R.M., Waring, M.J., and Travers, A.A. (2004). The influence of DNA stiffness upon nucleosome formation. J Struct Biol 148, 66-85.

Vyakarnam, A., Lenneman, A.J., Lakkides, K.M., Patterson, R.J., and Wang, J.L. (1998). A comparative nuclear localization study of galectin-1 with other splicing components. Exp Cell Res 242, 419-428.

Wallace, J.A., and Felsenfeld, G. (2007). We gather together: insulators and genome organization. Curr Opin Genet Dev 17, 400-407.

Weeks, J., Chandler, D., and Andersen, H. (1971). Role of Repulsive Forces in Determining the Equilibrium Structure of Simple Liquids. J. Chem. Phys. 54, 5237-5247.

Willemse, J. (2006). Histone 2B exchange in Arabidopsis. In: a microscopic analysis of Arabidopsis chromatin (Wageningen: Wageningen University), 45-60.

Woodcock, C.L., Frado, L.L., and Rattner, J.B. (1984). The higher-order structure of chromatin: evidence for a helical ribbon arrangement. J Cell Biol 99, 42-52.

Worman, H.J., Yuan, J., Blobel, G., and Georgatos, S.D. (1988). A lamin B receptor in the nuclear envelope. Proc Natl Acad Sci USA 85, 8531-8534.

Yan, H., Kikuchi, S., Neumann, P., Zhang, W., Wu, Y., Chen, F., and Jiang, J. (2010). Genome-wide mapping of cytosine methylation revealed dynamic DNA methylation patterns associated with genes and centromeres in rice. Plant J 63, 353-365.

Ye, Q., and Worman, H.J. (1996). Interaction between an integral protein of the nuclear envelope inner membrane and human chromodomain proteins homologous to Drosophila HP1. J Biol Chem 271, 14653-14656.

Ye, Q., Callebaut, I., Pezhman, A., Courvalin, J.C., and Worman, H.J. (1997). Domain-specific interactions of human HP1-type chromodomain proteins and inner nuclear membrane protein LBR. J Biol Chem 272, 14983-14989.

Yuan, J., Simos, G., Blobel, G., and Georgatos, S.D. (1991). Binding of lamin A to polynucleosomes. J Biol Chem 266, 9211-9215.

Zhao, R., Bodnar, M.S., and Spector, D.L. (2009). Nuclear neighborhoods and gene expression. Curr Opin Genet Dev 19, 172-179.

Zorn, C., Cremer, C., Cremer, T., and Zimmer, J. (1979). Unscheduled DNA synthesis after partial UV irradiation of the cell nucleus. Distribution in interphase and metaphase. Exp Cell Res 124, 111-119.

**Education Statement of the Graduate School**

**Experimental Plant Sciences**

The Graduate School EXPERIMENTAL PLANT SCIENCES

**Issued to:** **Silvester de Nooijer**
**Date:** **26 October 2010**
**Group:** **Laboratory of Molecular Biology, Wageningen University**

| **1) Start-up phase** | *date* |
|---|---|
| ▸ **First presentation of your project** | |
| Physical modelling of chromatin in Arabidopsis nuclei | Dec 01, 2006 |
| ▸ **Writing or rewriting a project proposal** | |
| Modelling chromatin organization in Arabidopsis root cell nuclei | Dec 2006 |
| ▸ **Writing a review or book chapter** | |
| ▸ **MSc courses** | |
| ▸ **Laboratory use of isotopes** | |
| *Subtotal Start-up Phase* | *7.5 credits** |

| **2) Scientific Exposure** | *date* |
|---|---|
| ▸ **EPS PhD Student Days** | |
| EPS PhD student day, Wageningen University | Sep 19, 2006 |
| EPS PhD student day, Wageningen University | Sep 13, 2007 |
| 1st joint international PhD retreat, Wageningen University | Oct 02-03, 2008 |
| 2nd joint international PhD retreat, Max Planck institute for plant sciences, Cologne, Germany | Apr 14-17, 2010 |
| ▸ **EPS Theme Symposia** | |
| EPS Theme 4 Symposium 'Genome Plasticity', Radboud University Nijmegen | Dec 08, 2006 |
| EPS Theme 1 Symposium 'Developmental Biology of Plants', Wageningen University | Oct 11, 2007 |
| EPS Theme 4 Symposium 'Genome Plasticity', Leiden University | Dec 07, 2007 |
| EPS Theme 4 Symposium 'Genome Plasticity', Wageningen University | Dec 12, 2008 |
| EPS Theme 4 Symposium 'Genome Plasticity', Radboud University Nijmegen | Dec 11, 2009 |
| ▸ **NWO Lunteren days and other National Platforms** | |
| NWO-ALW lunteren day 2006 Nucleic acids | Dec 04, 2006 |
| NWO-ALW Lunteren days 2007 Experimental Plant Sciences (EPS) | Apr 02-03, 2007 |
| NWO-ALW Lunteren day 2007 Nucleic Acids & Protein | Dec 2007 |
| NWO-ALW Lunteren days 2008 EPS | Apr 07-08, 2008 |
| NWO-ALW Lunteren days 2009 EPS | Apr 06-07, 2009 |
| NWO-ALW Lunteren days 2010 EPS | Apr 19-20, 2010 |
| ▸ **Seminars (series), workshops and symposia** | |
| flying seminar Jim Carrington | Mar 26, 2007 |
| flying seminar Rob Martienssen | Oct 23,2006 |
| flying seminar Scott Poethig | Sep 24, 2007 |
| flying seminar Greg Amoutzias | Oct 22, 2007 |
| seminar Ineke Braakman | Nov 09, 2007 |
| seminar Sander Tans | Nov 09, 2007 |
| flying seminar Hiroo Fukuda | Nov 26, 2007 |
| flying seminar Richard Vierstra | Apr 14, 2008 |
| flying seminar Simon Gilroy | May 18, 2008 |
| flying seminar Zhenbiao Yang | Jun 23, 2008 |
| seminar Enrico Scarpella | Oct 16, 2008 |
| Workshop the physics of genome folding and function | Oct 20-23, 2008 |
| Farewell symposium prof dr. A. J. W. Visser | Sep 24, 2008 |
| lecture dr. Pamela Hines (Science senior editor) | Nov 06, 2008 |

| | | |
|---|---|---|
| | IP/OP Systems biology Symposium | Jun 10, 2009 |
| | Workshop systems biology for plant design | Jul 09-11, 2009 |
| | NCSB kickoff Symposium | Oct 16, 2009 |
| ▸ | **Seminar plus** | |
| | Seminar + Rob Martienssen | Oct 23,2006 |
| | Seminar + Jim Carrington | Mar 26, 2007 |
| ▸ | **International symposia and congresses** | |
| | International chromosome conference 2007 (Amsterdam) | Aug 2007 |
| ▸ | **Presentations** | |
| | Poster: PhD retreat, Wageningen: Non-specific interactions are sufficient to explain the localisation of chromocenters and nucleoli in interphase nuclei | Oct 02, 2008 |
| | Poster: NWO-ALW Lunteren days 2010 EPS: Modelling chromatin organization in human interphase nuclei | Apr 19, 2010 |
| | Oral: EPS theme 4 symposium, Nijmegen | Dec 08, 2006 |
| | Oral: CBSG summit, Wageningen | Feb 06, 2007 |
| | Oral: International chromosome conference 2007, Amsterdam | Aug 25, 2007 |
| | Oral: IPK, Gatersleben | Nov 27, 2007 |
| | Oral: Biomeeting AMOLF | Mar 25, 2008 |
| | Oral: Physical and Colloid Chemistry group, Wageningen | Jan 18, 2009 |
| | Oral: Rijk Zwaan, Fijnaart | Feb 04, 2009 |
| | Oral: Biophysics group, Edinburgh University | Mar 31, 2009 |
| | Oral: EPS/ALW meeting, Lunteren | Apr 07, 2009 |
| | Oral: IP/OP systems biology symposium | Jun 10, 2009 |
| | Oral: NCSB kickoff symposium | Oct 16, 2009 |
| | Oral: PhD retreat, Cologne | Apr 15, 2010 |
| ▸ | **IAB interview** | Dec 05, 2008 |
| ▸ | **Excursions** | |
| | *Subtotal Scientific Exposure* | *23.0 credits** |

| **3) In-Depth Studies** | *date* |
|---|---|
| ▸ **EPS courses or other PhD courses** | |
| Statistical analysis of omics data | Dec 2008 |
| Statistics: Linear models | Jun 20-22, 2010 |
| ▸ **Journal club** | |
| Literature discussion at Mol. Biology (once every two weeks) | 2006-2010 |
| ▸ **Individual research training** | |
| MPI computing (HPC/EPCC, Edinburgh University) | Feb 2009 |
| EsPResSo/TCL programming instructions (Axel Arnold) | Oct 2006 |
| *Subtotal In-Depth Studies* | *7.2 credits** |

| **4) Personal development** | *date* |
|---|---|
| ▸ **Skill training courses** | |
| Mobilising your scientific network | May 23 & Jun 06, 2008 |
| Career Perspectives | Oct-Nov 2009 |
| Nationale carrierebeurs | Mar 12, 2010 |
| BCF career event | May 12, 2010 |
| ▸ **Organisation of PhD students day, course or conference** | |
| ▸ **Membership of Board, Committee or PhD council** | |
| *Subtotal Personal Development* | *3.4 credits** |

| | |
|---|---|
| **TOTAL NUMBER OF CREDIT POINTS*** | **41.1** |

Herewith the Graduate School declares that the PhD candidate has complied with the educational requirements set by the Educational Committee of EPS which comprise a minimum total of 30 ECTS credits

** A credit represents a normative study load of 28 hours of study*

The research described in this thesis was co-financed by the Centre for BioSystems Genomics (CBSG) under the auspices of the Netherlands Genomics Initiative (NGI), and the IP/OP systems biology programme of Wageningen University. The research described in chapter 5 was partly carried out under the HPC-EUROPA2 project (project number: 228398) with the support of the European Commission - Capacities Area - Research Infrastructures.