Biophysical analysis of diffusion controlled processes in the budding yeast nucleus

Inauguraldissertation

zur Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von

Lutz Robert Gehlen
aus Deutschland

Basel, im Jahre 2009

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von

    Prof. Dr. Susan Gasser
    Prof. Dr. Jörg Langowski
    Dr. Dirk Schübeler
    Dr. Konstantin Klenin

Basel, den 28.04.2008   _____

                                  Prof. Dr. Eberhard Parlow (Dekan)

Abstract

The spatial organization of chromosomes and the dynamics of chromosome reorganization have been shown to be crucial for the regulation of gene expression and various aspects of DNA metabolism. However, the mechanisms that establish and maintain nuclear organization and coordinate changes to this organization are poorly understood.

I used computer simulations based on random walk and polymer chain models to investigate the diffusion controlled behavior of chromosomes and extrachromosomal elements in the yeast nucleus. I also investigated the influence of fundamental geometrical and physical parameters on this behavior. Concretely, I analyzed the following systems:

- The distribution of intrachromosomal telomere-telomere distances in yeast and the effects of telomere anchoring: I could show that the intrachromosomal telomere-telomere distances of chromosomes 3 and 6 obtained using fluorescence microscopy measurements are shorter on average than predicted by the model for the respective chromosomes in free solution in the nucleus, suggesting additional constraints. Furthermore, telomeric anchoring leads to a further increase in the average distance and can therefore not be directly responsible for the close juxtaposition.

- The influence of nuclear geometry on the diffusion of a plasmid during nuclear division: In budding yeast, autonomously replicating sequence (ARS) plasmids show a strong tendency to segregate to the mother cell at mitosis. I showed that the geometric shape of the dividing nucleus and the limited length of mitosis impose a severe barrier on passive diffusion into the daughter nucleus, explaining the asymmetry in plasmid distribution. In collaboration with a colleague, I could show experimentally and theoretically that tethering of ARS plasmids to the inner nuclear membrane can increase the efficiency of plasmid partitioning. Our results suggest that the asymmetric morphology of mitosis could potentially contribute to rejuvenation by imposing physical constraints on the diffusion of damaged material into the daughter.

- The influence of nuclear organization and specifically telomeric anchoring on the search for a template during homologous recombination: DNA double-strand breaks are the most deleterious DNA lesions. Homologous recombination uses a homologous template to repair a double-strand break accurately and is very efficient in yeast. However, the process by which the break and template site find each other within the vast quantity of non-homologous DNA is not well understood. We have developed a combined experimental and theoretical approach to study homology search and its relationship to nuclear organization in yeast. I introduce our strategy and present some first theoretical results that prove the concept of the approach. Within the ongoing work on homology search in yeast, I am going to investigate the important question of the influence of locus mobility on nuclear processes in yeast.

Contents

List of Figures

List of Tables

Abbreviations

Abbreviations

ACF	autocorrelation function
BD	BROWNian dynamics
BER	base excision repair
DNA	deoxyribonucleic acid
DSB	double-strand break
FISH	fluorescence *in situ* hybridization
HR	homologous recombination
IF	immunofluorescence
LSM	laser scanning microscope
MC	Monte Carlo
MEC	method of excess collisions
MSD	mean squared displacement
NA	numerical aperture
NE	nuclear envelope
NER	nucleotide excision repair
NHEJ	non-homologous end joining
SPB	spindle pole body
wt	wild type

Units

bp	base pairs
kb	kilobases = 1000 base pairs
h	hour
min	minute
s	second
µs	microsecond
µm	micrometer
nm	nanometer

Chapter 1

Introduction

1.1 Genetics and epigenetics

1.1.1 Historical overview of genetics

The observation that children resemble their parents is both striking and fundamental, and it is likely that people have been thinking about the origins of heredity for a very long time. Moreover, plants and animals have been domesticated for at least 10000 and probably for 20000 years [Roberts, 1965]. This process involved the selection of the best specimens for the next generation, and it is therefore plausible that at least those who carried out the breeding had a basic concept of heredity.

The first scientific breeding studies were carried out in the seventeenth century and led to the sexual theory of flowering plants [Carlson, 2004]. In the eighteenth century, many breeders studied heredity in peas, but it was GREGOR JOHANN MENDEL's achievement to derive a consistent theory from his results. An important difference is that earlier studies looked at all traits of the species in a single specimen whereas MENDEL monitored a single trait at a time in a single specimen [Carlson, 2004]. He observed seven individual traits over several generations, which allowed him to derive two combinatorial laws about heredity. He postulated that a trait is controlled by a pair of factors (now called genes), one of which is inherited from the mother, the other from the father. If the two versions (or alleles) of a gene are different, often one of them determines the visible outcome (phenotype). This allele is called dominant, the other one is called recessive. MENDEL's laws of inheritance explain the distribution of phenotypes based on the assumptions above [Watson et al., 2003]. These findings are considered as the starting point for what is called "classical genetics".

MENDEL's work was not accepted by the scientific community and largely ignored until it was rediscovered in the early twentieth century [Carlson, 2004]. By that time, the cell nucleus had been established as the body that transmits the genetic information. The concept of chromosomes as the basis of heredity arose with improved microscopes which allowed WALTHER FLEMMING to identify thread-like objects that were duplicated and

segregated to daughter cells during cell division [Flemming, 1879]. In 1871, FRIEDRICH MIESCHER published his isolation and characterization of "Nukleinsäure" or DNA, which he also proposed to might be the bearer of inheritance [Miescher, 1871]. WALTER SUTTON was the first to realize that chromosomes obey MENDELian rules, and he hypothesized that genes are parts of the chromosomes [Sutton, 1903].

The chromosomal theory of heredity was confirmed by the fact that there are groups of traits that are always inherited together. With more and more mutants available, it became apparent that the number of independent groups is equal to the number of chromosomes. This result can be explained by SUTTON's hypothesis that a gene is a part of a chromosome and that therefore, genes on the same chromosome are inherited together.

In the following decades, two questions dominated genetic research, namely which material carries the genetic information and how genes carry out their function in the cell. It was an accepted hypothesis that genes control the synthesis of enzymes, and it was assumed that they exert this control by determining the amino acid sequence of proteins [Watson et al., 2003]. However, there was no convincing model how genes can encode protein sequences, nor how this information could be copied before cell division.

In 1928, FREDERICK GRIFFITH discovered that genetic material from dead bacteria could be integrated and functional in living cells. In 1944, OSWALD AVERY showed using specific enzymes to degrade DNA, RNA, or proteins, that it is DNA that carries the genetic information [Avery et al., 1944]. This was confirmed by experiments of ALFRED HERSHEY and MARTHA CHASE, based on work of SALVADOR LURIA and MAX DELBRÜCK, who showed that DNA is injected by bacteriophages to transform bacteria [Hershey and Chase, 1952].

Meanwhile, X-ray diffraction images of DNA were being taken with increasing quality [Franklin and Gosling, 1953]. These images suggested that the structure of DNA was helical and composed of at least two strands. Finally, in 1953, the correct structure was discovered by FRANCIS CRICK and JAMES WATSON [Watson and Crick, 1953]. It is composed of two strands each consisting of a sequence of building blocks (nucleotides). Only four types of nucleotides occur in DNA, and each of them found along one strand is bound to a specific partner on the other strand: adenine is always paired with thymine and cytosine with guanine, consistent with CHARGAFF's rules, which had been formulated by ERWIN CHARGAFF based on the relative ratios on the four bases in genetic material [Chargaff and Magasanik, 1949]. This complementary structure conceptually solved the question of how the genetic information can be copied before cell division. Since both strands carry the same information, either one of them can be used as a template for the synthesis of a new strand.

The discovery of the DNA structure started the field of molecular genetics. Soon afterwards, the so-called central dogma was established. It states that the DNA functions as the template for messenger RNA molecules, which leave the nucleus and act themselves as templates for protein synthesis. Importantly, according to the central dogma, this flow of information is unidirectional: proteins do not act as templates for RNA, and RNA does not

determine DNA sequence. It is now known that the latter statement is not strictly true, but the dogma is still essentially valid.

For the work presented here, it is important that the genetic information resides in the nucleus in the form of long linear DNA molecules. A gene is a stretch of DNA, which — in many cases — encodes the amino acid sequence of a protein. The encoded proteins then act as enzymes, which catalyze chemical reactions in the cell, or as structural components in- or outside the nucleus.

1.1.2   Epigenetics

The genome of a cell can be envisioned as a library of construction plans for all the tools a cell might ever need. Obviously, not all of the tools are needed all of the time. Many of them are only used in special situations (e.g. when the cell or organism encounters nutritional changes or in case of stress). When these special circumstances are not present, the respective tools should not be produced. Under certain circumstances, changes in gene expression can become stable and are remembered through mitotic division. The field of science that studies the complex mechanisms that control the expression of individual genes in a heritable manner is called epigenetics. An epigenetic phenomenon can be defined as a "change in phenotype that is heritable but does not involve DNA mutation" [Allis et al., 2006].

Apart from the reaction to environmental pressure, the epigenetic silencing of genes plays an important roll in cell differentiation. A multicellular organism consists of cells of many different types. Despite their highly specific morphology and functions, it is remarkable that all cells contain the same genome. Thus, although the genome holds construction plans for all proteins[1] that any cell needs, a certain cell type requires only a fraction of them. Therefore, in the process of cell differentiation, during which specialized cells emerge from pluripotent stem cells, many genes are made permanently unreadable [Delcuve et al., 2009]. Once the fate of a cell is determined, the expression pattern is stably inherited through mitotic division.

Furthermore, in diploid organisms there are two sets of nearly identical chromosomes, paternal and maternal, which often undergo selective repression, allowing information from only one parent to be expressed.

Epigenetic regulation of gene expression is obtained by an interplay of chemical modifications, local structure, and long range spatial organization of the genome. I give a brief summary of these phenomena in the following sections.

---

[1]As mentioned before, proteins are not the only regulatory elements that are encoded in the genome. However, the field of non-coding RNAs is beyond the scope of this introduction, see e.g. [Hüttenhofer et al., 2005].

1.1.2.1   Chromatin

The DNA is not present as a naked molecule in the nucleus. Instead, a chromosome is organized in several layers of compaction. This compaction is accomplished by a set of DNA binding proteins. The condensed structure of DNA and proteins that form a chromosome is called chromatin. The compaction does not only serve the purpose of packaging the large amount of DNA into the small volume of the nucleus, but the proteins that are part of the condensed structure also serve as platform for epigenetic marks (see below).

The different levels of compaction are shown in Figure 1.1. The first stage of compaction is the nucleosome where 147 base pairs of DNA are wrapped almost two times around an octamer of histone proteins [Luger, 2003]. A nucleosome has a diameter of about 10 nm. The histones are small proteins that consist of a central, highly structured region and unstructured C- and N-terminal tails. This repeated structure of DNA and histones results in a compaction of the DNA into a "beads-on-a-string fiber", called the 10 nm fiber.
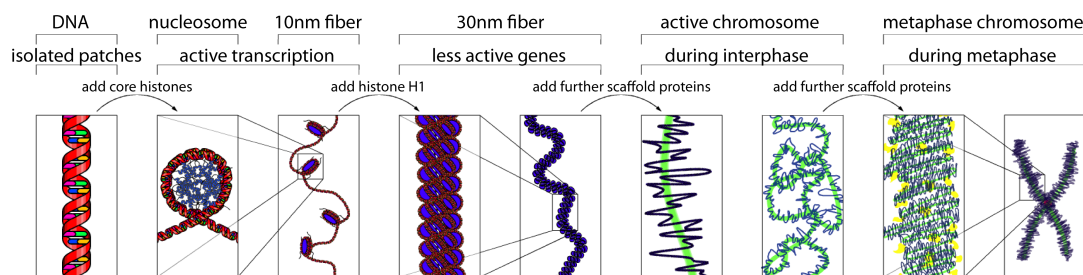
Figure 1.1: Levels of chromosome compaction. This illustration is copied from wikimedia commons (`http://en.wikipedia.org/wiki/File:Chromatin_Structures.png`) with permission under the GNU Free Documentation License (`http://www.gnu.org/copyleft/fdl.html`).

In the next level of compaction, nucleosomes are packaged into the so-called 30 nm fiber. The exact structure of the 30 nm fiber is still under debate [van Holde and Zlatanova, 2007]. Moreover, the vast majority of studies investigating the 30 nm fiber have been carried out *in vitro* where linker histones, such as histone H1, stabilize the compact 30 nm structure [Robinson et al., 2006]. It is not clear if chromatin actually forms a regular 30 nm fiber in the nucleus [Campbell et al., 1978, Widom, 1998, van Holde and Zlatanova, 2007, Dekker, 2008].

Even less is known about higher levels of compaction. The chromosome fiber is assumed to form loops of varying levels of compaction according to the stage of the cell cycle [Bohn et al., 2007, Mateos-Langerak et al., 2009].

1.1.2.2 Heterochromatin and euchromatin

Although the structure of the 30 nm chromatin fiber *in vivo* is not fully elucidated, chromatin structure is believed to contribute to the regulation of gene expression [Allis et al., 2006]. In general, inactive regions are supposed to form a more compacted, inaccessible structure called heterochromatin. In contrast, euchromatin, or active chromatin, is present in a more open structure, more accessible to the transcription machinery. This has been demonstrated by monitoring the accessibility of DNA to endonuclease digestion *in vivo* or in isolated nuclei. Active chromatin does not necessarily imply that the genes in this region are ubiquitously expressed. It only means that this is possible.

A large number of chemical modifications of DNA and primarily of histones have been identified that contribute to the formation of eu- or heterochromatin. These include acetylation, methylation, ubiquitination, and SUMOylation of lysine residues that are generally enriched in the tails of the four core histones. Phosphorylation of serines and threonines are also occasionally of major importance. In addition to targeting the histone tails, there are several important modifications that are found in the body or on the "face" of the assembled nucleosome. The modifications have been shown to alter chromatin structure by changing the charge of the nucleosomes or by recruiting structural proteins and nucleosome remodelers. In general, histone modifications are at the base of epigenetic changes, although not all histone modifications are heritable.

1.2 Nuclear organization

The nucleus is the cellular compartment that incloses the genome. It contains several structural elements that can serve as scaffold for the spatial organization of the genome. First, there is the nuclear envelope (NE), a double lipid bilayer contiguous with the endoplasmatic reticulum that separates the nuclear content from the rest of the cell. The NE is studded with nuclear pores, which are used for transport of proteins and RNAs into and out of the nucleus, but also provide a compartmentalization of the NE. A second structural element of the nucleus is the nucleolus which enables the transcription, processing, and assembly of ribosomal RNA into pre-ribosomes. The nucleolus is not enclosed in a membrane, but usually occupies a distinct region of the nucleus. The nuclei of higher eukaryotes contain a fibrous network of intermediate filament proteins called the nuclear lamina. These are found both underlying the nuclear envelope and at sites throughout the nucleoplasm. The internal lamin might be part of a more extensive nucleoskeleton, although the components of this structure have never been elucidated.

The spatial localization of chromosome domains to these subcompartments of the nucleus has been shown to play a role in the activation or repression of genes (reviewed in [Sexton et al., 2007, Towbin et al., 2009]). Silent chromatin is often associated with the NE or the nucleolus. However, it is also known that peripheral localization alone is not sufficient to

ensure gene silencing. The additional presence of binding sites for silencing factors is required. More recently, it was shown that the relocalization of certain genes to nuclear pores in yeast even has an activating effect (reviewed in [Akhtar and Gasser, 2007]).

Many studies about nuclear organization and gene expression have been carried out in the budding yeast *Saccharomyces cerevisiae*. My thesis work focuses on processes within the nucleus of this organism. Therefore, I give an introduction to yeast and its nuclear organization in the next section.

1.2.1 Budding yeast

The budding yeast *Saccharomyces cerevisiae*, also called baker's or brewer's yeast because of its use as a fermenting agent, has been used as a model system since the 1930's, when it was used to identify metabolic pathways. Yeast unites the practical advantages of bacteria with the central characteristics of all eukaryotic cells. On one hand, it has a relatively small genome, can be grown rapidly (the doubling time is about 90 min under optimal conditions), and is easy to manipulate genetically. On the other hand, being eukaryotic, it has linear chromosomes packaged into chromatin and contained in a discrete nucleus.

A further advantage of yeast is that it grows in either haploid (one copy of each chromosome) or diploid (two copies) state. Since the experimental analysis of gene function is much simpler in haploid cells, laboratory strains are usually kept in haploid state and prevented from forming diploids. In contrast, yeast cells are usually diploid in the wild. Haploid cells exist in two different mating types, termed **a** and α, reminiscent of the two sexes in mammals [Madhani, 2006]. If two haploid cells of different mating types encounter one another, mating type specific pheromones (called **a** and α factors) bind to the opposite cell and trigger cell cycle arrest and the production of proteins required for mating. The two cells then fuse and produce a diploid cell. Importantly, the fact that **a** and α factors stop cell cycle progression of the other mating type can be used to synchronize cell populations for cell cycle stage analyses.

In response to starvation conditions, diploid cells undergo sporulation resulting in a tetrad of four spores in a protective container called ascus. Two of these spores are of mating type **a**, the other two of mating type α. When conditions improve, the cells return to the proliferate state. Using this natural mechanism, haploid strains can be generated for laboratory work.

Cells divide by budding in both haploid and diploid state. Haploid wild type cells can switch their mating type at each cell division [Madhani, 2006]. Mating type switching is provoked by a site specific endonuclease (HO) that induces a DNA double-strand break at the mating type locus *MAT*. The break is repaired by homologous recombination (see section 1.3). The recombination template can be one out of two silent mating type loci, each representing one mating type. Since mating type switching occurs very frequently in wild type cells, they usually form diploids immediately after spores start growing. To prevent this in laboratory strains, mating type switching is inhibited by deletion of the *HO* gene.

6

Before I introduce the important aspects of nuclear organization specific to budding yeast, I give an introduction to the yeast cell cycle.

1.2.2   The cell cycle of budding yeast

Yeast cells divide in both haploid and diploid state. The term cell cycle subsumes the cellular events that lead to the birth of a new cell and the return of the mother cell to its initial state. This process comprises a sequence of tightly controlled steps including the duplication of each chromosome and the accurate segregation of the chromosome pairs into the two daughter cells.

The cell cycle can be divided into four phases (Figure 1.2). When cells are quiescent or not dividing, they generally accumulate in a phase called G0. Recovery from this quiescent state starts by entry into G1, a gap phase, in which cells accumulate the resources required for successful reproduction. The duplication of the genome by DNA replication takes place in the synthesis phase (S). S phase is followed by a second gap phase (G2). The separation of the chromosomes and the actual cell division take place during mitosis (M). Afterwards, both cells enter a new round of G1, from which they can either start a new division cycle or switch to G0.

Mitosis can be divided further into five subphases, which are defined on the basis of chromosome morphology in mammalian cells [Morgan, 2006]. In prophase, chromosomes start to condense and become visible under a light microscope. The nuclear envelope breaks down (not in yeast, see below), and chromosomes start to attach to the mitotic spindle apparatus, which will pull them later into the correct daughter cell. This intermediate stage is called prometaphase. In metaphase, the chromosomes are highly condensed and aligned at the center of the spindle apparatus.

The separation of sister chromatids (the two copies of one chromosome) takes place in anaphase. Telophase is the final subphase of mitosis, in which the nuclear membrane is reformed and the separation of the two genomes is finished.

The most important deviation of the budding yeast cell cycle from the progress described above is that yeast has a so-called closed mitosis. This means that there is no nuclear envelope breakdown in prometaphase. Instead, the nucleus itself divides (Figure 1.2). The spindle apparatus, which separates the sister chromatids, is connected to one of the two spindle pole bodies (SPB), which are embedded into the nuclear membrane [Jaspersen and Winey, 2004].

Additionally, the division of the cell starts much earlier in budding yeast. *S. cerevisiae* reproduces by budding which gave the species its common name. Already during S phase when the genome is replicated, a bud starts to bulge out of the mother cell. The bud enlarges during the cell cycle and separates from the mother at the end of mitosis. The daughter cell is considerably smaller than the mother [Porro et al., 2009]. Importantly, the continuous growth of the bud during the cell cycle provides a simple morphological indicator for the cell cycle stage.
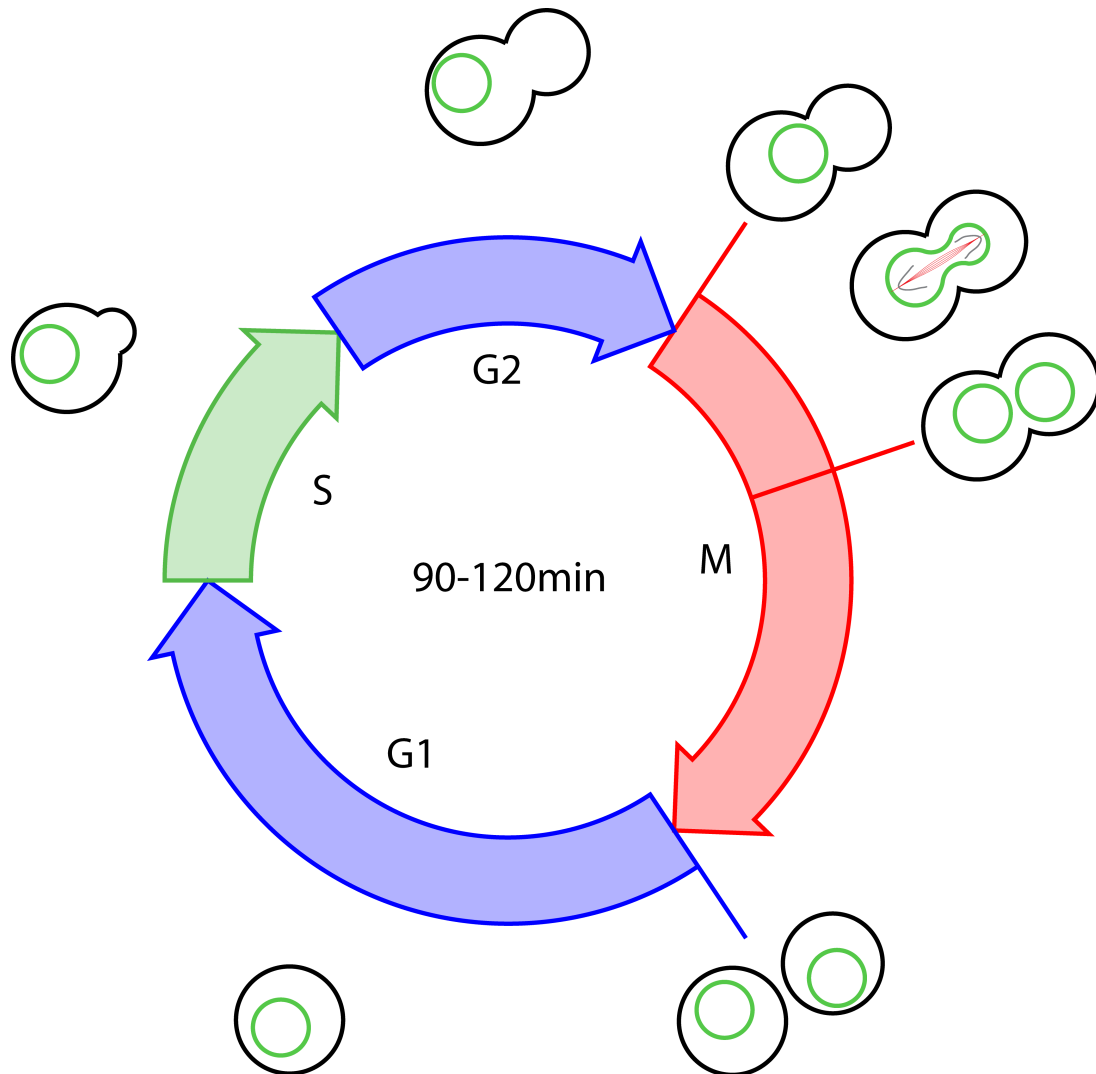
Figure 1.2: The cell cycle of budding yeast. The stage of the cell cycle can be determined by looking at the size of the bud and location and shape of the nucleus. The duration of the whole cell cycle lies between 90 min and 120 min, depending on genomic background and growth conditions. The duration of the different phases is drawn on scale according to [Lord and Wheals, 1981].

1.2.3 Nuclear organization in yeast

Much information about the relationship between nuclear organization and nuclear functions has come from studies in yeast. The chromosomes in budding yeast show a distinct spatial organization. The telomeres (chromosome ends), ribosomal DNA (located in the nucleolus), and the silent mating type loci (*HML* and *HMR*) assume a heterochromatin-like structure whose integrity is important for processes such as epigenetic silencing, chromosome cohesion, perinuclear anchoring, and suppression of recombination. The 32 telomeres of a haploid yeast cell form 2-8 foci, which are preferentially located close to the nuclear periphery [Palladino et al., 1993, Gotta et al., 1996]. These telomere clusters sequester the silent information regulator proteins Sir2, Sir3, and Sir4, which nucleate telomeric silencing (repression of subtelomeric genes), away from internal loci [Palladino et al., 1993, Gotta et al., 1996, Taddei et al., 2009]. The centromeres are located close to the spindle pole body and opposite to the nucleolus throughout the cell cycle [Bystricky et al., 2005].

1.2.3.1 Telomere anchoring pathways

Yeast telomeres are anchored to the nuclear envelope via at least two partially redundant pathways [Hediger et al., 2002, Taddei et al., 2004]. One involves the Sir4 protein, which binds to the membrane-associated protein Esc1, the other one involves the yKu heterodimer. Different telomeres vary with respect to the preferred anchoring pathway and to the cell cycle dependency of telomeric anchoring.

Recently, an additional pathway for telomeric anchoring has been described, which relies on yKu and the telomeric proteins Est2 and Est1, as well as the integral membrane protein Mps3 [Schober et al., 2009].

1.2.3.2 Nuclear pores

There is growing evidence that the nuclear envelope in yeast is not homogeneous with respect to transcriptional control. In contrast to the transcriptionally silent telomere pools, nuclear pores seem to be centers for activation of certain genes (reviewed in [Akhtar and Gasser, 2007]). Additionally, the nuclear pore complex seems to be involved in DNA double-strand break repair (see also next section). [Therizols et al., 2006] reported that a part of the nuclear pore, the Nup84 complex, is required for the efficient repair of double-strand breaks (DSBs) in subtelomeric regions. Furthermore, it was shown that irreparable DSBs relocate to nuclear pores [Nagai et al., 2008, Kalocsay et al., 2009].

1.3 DNA damage

The survival of a species depends on the reliable propagation of the genetic material from generation to generation. High mutation rates in the germ line would imperil the species,

and high mutation rates in the soma could endanger survival of the individual. Mutations arise not only from erroneous DNA replication but also from DNA damage, which can occur spontaneously or can be caused by chemicals or radiation. Therefore, the stable maintenance of the genetic material requires not only accurate DNA replication, but also mechanisms for DNA damage repair. For the latter, cells have developed a complex control system to recognize DNA lesions and to prevent permanent mutations by a strict synchronization of damage repair with the cell cycle. If DNA damage is detected, cell cycle progression is stopped by checkpoint-signaling networks to allow repair before an alteration becomes permanent [Zhou and Elledge, 2000]. If the damage cannot be repaired, controlled cell death by apoptosis is induced.

1.3.1 Sources and types of DNA damage

DNA damage can be caused by external damaging agents like ultraviolet or ionizing radiation or chemical reagents. However, the source of the damage can also lie within the cell. Byproducts of cellular metabolism can lead to oxidation or alkylation of nucleotides [Hoeijmakers, 2001, Watson et al., 2003]. Over 100 oxidative modifications of DNA have been identified [Cadet et al., 1997]. Moreover, even without any direct source of damage, DNA is inherently unstable because spontaneous disintegration of chemical bonds can lead to deamination or depurination of nucleotide residues [Lindahl, 1993, Watson et al., 2003].

Apart from chemical modifications, errors in DNA replication can result in the insertion or deletion of bases leading to base pair mismatches. Furthermore, the topological structure of the DNA can be damaged by single-strand breaks, intra- or interstrand crosslinks, or double-strand breaks [Hoeijmakers, 2001]. DNA double-strand breaks (DSBs) are the most deleterious DNA lesions because they disrupt the connectivity of the DNA and can lead to loss of genetic information or deleterious gene fusions if inaccurately repaired. They can be generated by ionizing radiation, free radicals, and certain anti-tumor agents. Importantly, less severe types of DNA damage can indirectly cause DSBs during replication. For example, a replication fork can collapse when it encounters a nick in the DNA, and this collapse gives rise to a DSB [Watson et al., 2003]. Other lesions can also bring a replication fork to stall and eventually to collapse.

An immediate consequence of DNA damage is the obstruction or inhibition of transcription and replication. Long-term consequences of error-prone repair are permanent mutations in the genomic sequence and aberrant chromosomal translocations, which often correlate with cancer.

1.3.2 DNA damage repair

Considering the diversity of DNA damage, it is not surprising that there is no single mechanism that can repair all sorts of damage. There are at least four, partly overlapping pathways [Hoeijmakers, 2001], which are conserved from yeast to mammals: base excision

repair (BER), nucleotide excision repair (NER), non-homologous end joining (NHEJ), and homologous recombination (HR). Here, I focus on the repair of DSBs which can be achieved by NHEJ and by HR.

NHEJ simply religates the two ends of the break (Figure 1.3B). Depending on the origin of the lesion, this can reproduce the original sequence exactly or cause the loss of base pairs. Breaks created by a nuclease often retain complementary bases in an overhang and can be repaired by NHEJ without loss of information. In contrast, ionizing radiation usually leads to the damage or removal of base pairs at the break site. These lesions can only be repaired by imprecise NHEJ and lead to a change in the sequence [Shrivastav et al., 2008]. Thus, NHEJ is often referred to as "error-prone" DSB repair.

In contrast to NHEJ, HR is generally error-free. During HR, the ends of the DSB get partially resected leading to single-stranded 3' overhangs, and the missing sequence is resynthesized using a homologous sequence elsewhere in the genome (sister chromatids, homologous chromosomes, or homologous regions on the same or a different chromosome), see Figure 1.3A. After alignment of the homologous sequences, the single-stranded ends invade the homologous DNA template sequence and anneal to the respective complementary strand. This enables the repair polymerase to restore the destroyed and resected DNA sequences. The resulting intermediate structure (HOLLIDAY junction) can be resolved in two ways leading either to a gene conversion or a crossing over event [Watson et al., 2003].

NHEJ and HR compete for DSB repair, and the preference for one or the other depends on various conditions including species, cell type, and cell cycle stage. In mammals, both precise and imprecise NHEJ are very efficient. HR is only upregulated in S and G2 phase of the cell cycle when sister chromatids are available for HR. In mammals, HR may be particularly important for the restart of stalled or collapsed replication forks [Shrivastav et al., 2008].

In yeast, the fraction of "clean" breaks that are repaired by precise NHEJ has been estimated to be at least 25% [Clikeman et al., 2001]. However, imprecise NHEJ is very inefficient (0.01% to 0.2% cell survival if DSBs can only be repaired by imprecise NHEJ [Shrivastav et al., 2008]). Therefore, HR is the predominant DSB repair pathway, at least for "dirty" breaks such as those caused by ionizing radiation. Even in yeast, sister chromatids are the preferred template for repair if they are available [Shrivastav et al., 2008]. Intriguingly, HR in yeast is very efficient even for non-allelic (ectopic) homologous sequences [Aylon et al., 2003].

1.4   Chromatin movement and diffusion

Chromatin in the interphase nucleus is not static. Chromosomal loci have been shown to move up to 1 µm in a few seconds [Gasser, 2002, Soutoglou and Misteli, 2007]. The movement of genomic loci is involved in many types of DNA metabolism including transcription, replication and DNA damage repair [Akhtar and Gasser, 2007]. Although little is known

A   Homologous recombination   B   Non-homologous end joining

Double-strand break   Double-strand break

Resection   Ligation

3'   5'

5'   3'

Stand invasion,
D-loop formation,
DNA synthesis

Second end capture
DNA synthesis,
Ligation

Branch migration,
Resolution

Non-crossover

or

Crossover
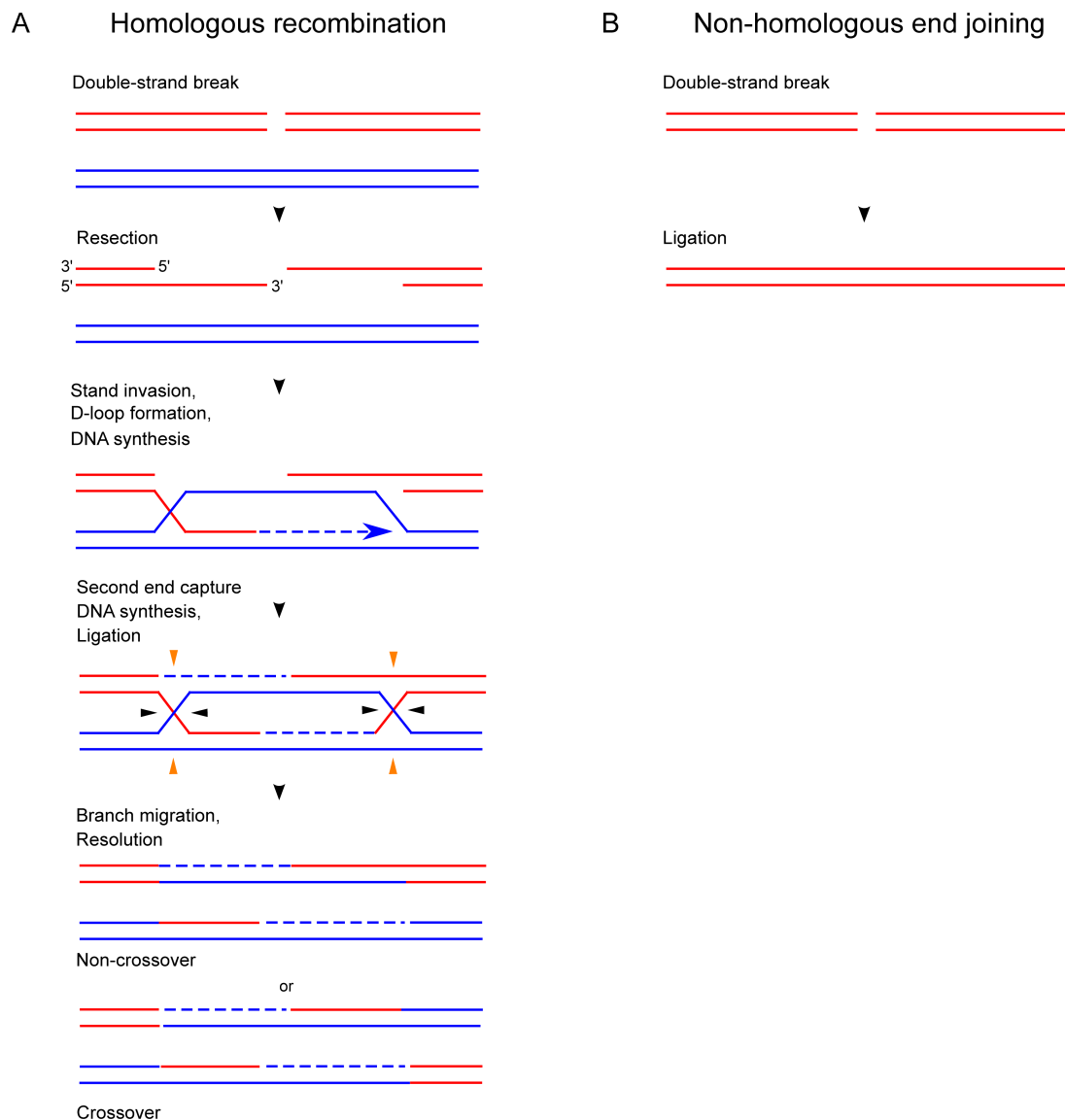
Figure 1.3: Schematic overview of the pathways for DNA double-strand break repair. HR uses a homologous template to repair the break accurately. This process involves the resection of the break ends and the invasion of the homologous sequence by the created 3' overhangs. The resolution of the resulting HOLLIDAY junction leads to either a gene conversion or a crossing over event. NHEJ simply religates the broken ends, possibly leading to loss or change of genetic information. See e.g. [Watson et al., 2003] for a detailed description of both pathways. This illustration is based on work on wikimedia commons (`http://en.wikipedia.org/wiki/File:HR_schematic_diagram.svg`, licensed under the Creative Commons Attribution ShareAlike 3.0 License, `http://creativecommons.org/licenses/by-sa/3.0/`).

about the source of the movement, it has been shown that a depletion of energy by lowering ATP levels leads to a dramatic decrease of chromatin movement [Gartenberg et al., 2004, Heun et al., 2001]. We could further show that chromatin remodeling complexes are able to increase the mobility of a chromosomal locus, possibly by increasing chromatin flexibility [Neumann et al., *in preparation*].

Despite these indications that the movement of chromosomes is not simply thermal motion, diffusive movement is nonetheless an important model. First of all, it is inescapable that there is always a thermal component in the movement of a chromosomal locus. The second and possibly more important reason is that it can be very insightful to compare experimentally derived data to the results of a diffusion model. This analysis can help to reveal the mechanisms that work in the cell in order to deviate from random motion.

1.4.1   Diffusion

In a physical system consisting of many particles each object possesses a certain amount of energy which is — on average over time — proportional to the absolute temperature. Therefore all particles are in constant motion and (at sufficient particle density, e.g. in a liquid) frequently collide with each other and change their direction of movement. This random movement of particles is called diffusion.

This process has two important consequences. First, concentration gradients are not stable. Diffusion leads to a transport of particles from regions with higher concentrations to those with lower concentrations until in equilibrium, the concentration of each species of particles is constant. Concentration gradients can only be actively maintained by the presence of sources and drains of concentration (e.g. regions of production and destruction of particles, respectively). This process of net particle transport along concentration gradients is called diffusion (in narrower sense) and is discussed in the following section.

Second, the position of a single object of interest is not stable. Due to the collisions with other particles it exerts a random walk or so-called BROWNian motion. I analyze random walks in more detail in chapter 2.

1.4.2   Diffusion as a transport process

Molecular diffusion is the process of particle transport from regions of higher concentration to those of lower concentration. A classical demonstration experiment is the mixture of a colored solution (e.g. an iodine solution) with water. The water is carefully poured on top of the iodine solution in a cylinder so that no convection occurs. At the beginning, the two parts are well separated by a sharp border, but with time, the upper part gets more and more colored while the color fades in the lower part until finally, the whole solution is uniformly colored.

The assumption that the diffusing particles move independently justifies the reasoning that the flux of the diffusing substance through unit area of a section is proportional to the

concentration gradient normal to that section:

$$\vec{F} = -D\nabla C \tag{1.1}$$

where $\vec{F}$ is the flux and C is the concentration of the substance. This relation is known as FICK's first law. The proportionality constant D is called the diffusion coefficient of the substance. In dilute solution, it can usually be taken as independent of the concentration.

Under the assumption that particles cannot be created or destroyed, a combination of the divergence theorem[2] and the continuity equation[3] yields FICK's second law:

$$\frac{\partial \, \mathrm{C}(\vec{x}, t)}{\partial t} = \nabla(D\nabla C) \tag{1.2}$$

and with constant D:

$$\frac{\partial \, \mathrm{C}(\vec{x}, t)}{\partial t} = D \sum_{i=1}^{3} \frac{\partial^2 \, \mathrm{C}(\vec{x}, t)}{\partial x_i^2}$$
$$\frac{\partial \, \mathrm{C}(\vec{x}, t)}{\partial t} = D\Delta \mathrm{C}(\vec{x}, t) \tag{1.3}$$

where Δ is the LAPLACE operator [Boas, 1983]. A more detailed derivation can be found in [Crank, 1975].

1.5 Aim of this study

The spatial organization of chromosomes and the dynamics of chromatin reorganization have been shown to be crucial for gene regulation and various aspects of DNA metabolism. However, the underlying mechanisms of the establishment and maintenance and the coordinated change of nuclear organization are poorly understood.

Mathematical and computational modeling are well-suited to address these questions. In contrast to a biological experiment, a theoretical model is under full control of the researcher, both actively (parameter manipulation) and passively (parameter readout). Single parameters, such as the position or mobility of a specific genomic locus, can be manipulated exactly without affecting the rest of the system. Moreover, even situations that are physiological, but difficult or impossible to induce in the laboratory (e.g. extreme gradients in temperature or viscosity, immobilization of an object at a specific position etc.) can be easily set up. The study of these situations can help substantially to elucidate the influence of a given parameter on the system. On the other hand, the values of variables of

---

[2]`http://mathworld.wolfram.com/DivergenceTheorem.html`
[3]`http://en.wikipedia.org/wiki/Continuity_equation`

interest can be measured with arbitrary precision that is only limited by the speed of the computer and the size of its memory.

In addition to the complete control over a theoretical model, the design of the model itself can be used to elucidate the biological situation. In most cases, it is necessary to simplify the model compared to the real situation. However, this is not necessarily a disadvantage, but instead can be a very powerful feature of theoretical modeling, because it allows the identification of the key components of the studied system that are essential to produce the observed behavior. One possible approach is therefore to start out with a minimal model and to investigate which components have to be added in order to reproduce the general behavior of the biological system.

The aim of my thesis work was to apply this approach to different aspects of the spatial organization of chromosomes and extrachromosomal elements in yeast. My special interest lay in the diffusion controlled behavior of these systems and in the influence of fundamental geometrical and physical parameters on this behavior. Concretely, I analyzed the following systems:

- the distribution of intrachromosomal telomere-telomere distances in yeast and the effects of the anchoring of telomeres (chapter 3)

- the influence of nuclear geometry on the diffusion of a plasmid during nuclear division (chapter 4)

- the influence of nuclear organization and specifically telomeric anchoring on the search for a template during homologous recombination in DNA damage repair (chapter 5).

Theoretical modeling of biological systems is most powerful in combination with wet lab experiments. In the GASSER lab, fluorescence microscopy is used extensively to study the spatial organization of the nucleus. In addition to the simulations introduced above that I have used to address important unresolved questions of nuclear organization, I have analyzed common methods that are used routinely to extract quantitative parameters from microscopic images in order to expand their applicability. I present this work in section 2.3.

Chapter 2

Methods

2.1 Fluorescence microscopy

A large part of my thesis work deals with fluorescence microscopy images. I analyzed methods that are frequently used to quantify fluorescent images (section 2.3), and I used fluorescence microscopy data obtained in budding yeast as a starting point as well as for experimental validation in various modeling projects (chapters 3, 4, 5). Therefore, I give an introduction into the methods we use for acquisition of fluorescence microscopy images in yeast. This section is based on [Meister et al., 2010].

Quantitative imaging techniques have improved dramatically in the last 15 years, both thanks to the rapid adaptation of naturally fluorescent proteins to cell biology and improvements in fluorescence microscopy. Methods are being continually perfected to enable the analysis and localization of endogenous yeast proteins and chromosomal loci in living cells [Michaelis et al., 1997, Rohner et al., 2008, Straight et al., 1996, Straight et al., 1997]. In addition to new microscopes, novel computational approaches for image analysis have become available, which facilitate the high resolution analysis of digital image stacks. These are generally captured on deconvolution widefield microscopes or with point scan or spinning disk confocal instruments [Hom et al., 2007]. While techniques of live microscopy are powerful, they are also not trivial to perform correctly. It is often hard to visualize more than two fluorophores at the same time, and the exposure to laser or high intensity light sources can provoke damage of both DNA and proteins. Particular care must be taken when dealing with mutants that can be damaged by conditions of fluorescence imaging. Maintenance of unperturbed growth conditions is essential for meaningful results, particularly for live imaging. Moreover, high resolution time-lapse microscopy often captures only one or a few cells per 3D stack, which means that the imaging itself takes considerable time. This makes it difficult to perform time-course experiments when several strains need to be analyzed by live microscopy in parallel. In these cases, cells can be fixed rapidly by formaldehyde, and the localization of proteins or DNA can be achieved by immunofluorescence (IF) or

fluorescence *in situ* hybridization (FISH) or a combination of the two [Gotta et al., 1999]. In this overview, I focus on live cell imaging.

2.1.1   Tagging chromatin *in vivo*

The study of chromatin organization in living cells often exploits the recognition of integrated arrays by fluorescently labeled bacterial DNA binding factors (reviewed in [Hediger et al., 2004, Neumann et al., 2006]). The arrays consist of 128 to 256 copies of the recognition consensus. As few as 24 binding sites are probably sufficient to allow the formation of a visible spot in yeast, depending on the expression level of the fluorescently tagged binding protein.

Tagging yeast chromatin *in vivo* is a two step process. The first step involves the expression of a fusion between a DNA binding protein, a fluorescent protein and a nuclear localization signal. The most commonly used DNA binding proteins are the lac repressor fused to green fluorescent protein (GFP) as well as its cyan and yellow variants (CFP, YFP), and the tet repressor with GFP, CFP, YFP, and the monomeric variant of the red fluorescent protein (mRFP) [Lisby et al., 2003]. To improve the fluorescence signal, the fluorescent protein can be used in a tandem array (3x CFP), although this can also induce unwanted CFP-CFP interaction [Bressan et al., 2004]). Expression levels of these proteins have to be kept low, as overexpression leads to non-specific binding and slow-growing colonies.

It is often useful to insert a low number of binding sites for another DNA-binding protein next to the repeats integrated at specific loci. This allows one to target another protein to the site of interest and to monitor its effect on the location of the locus. For example, often near an array of lacO sites, we have integrated lexA sites to allow targeting of the labeled locus to specific subnuclear compartments using lexA fusion proteins. The expression of a fusion between lexA and Yif1, an integral membrane protein, for example, anchors the chromatin at the nuclear envelope.

2.1.2   Determining the position of the nucleus

For precise chromatin localization studies as well as for dynamics, the nuclear volume has to be outlined. This can be achieved either by the expression of a nucleoporin fused to a fluorescent protein (commonly Nup49-GFP) or by using the nuclear background fluorescence created by a GFP fusion to an unbound DNA binding protein. However, diffuse fluorescence is only observed in cells expressing a tetR fusion, as lacI fusions tend to give very little background, probably due to a difference in the expression level of the proteins.

2.1.3   Image acquisition

Depending on the aim of the experiment, different setups can be used to image tagged chromatin *in vivo*. Whichever system is used, it is essential to check that the cells are able to

survive the high intensity light used for fluorescence illumination. It is therefore important to confirm for each experiment that illumination alone has no influence on cell physiology by testing cell cycle length in illuminated versus non-illuminated cells. Every microscopic system is a compromise between speed of acquisition (the higher the speed, the lower the amount of light that can be recorded), field of acquisition (in general, the bigger the field, the lower the speed) and resolution (higher resolutions decrease speed and signal, since each pixel on the image corresponds to a smaller part of the sample and more pixels take more time to acquire). A common rule for all systems is that since the yeast nucleus is small, the objective's magnification should be at least 63x or ideally 100x. The numerical aperture (NA) should be as high as possible (between 1.3 and 1.45) to obtain as many details as possible from the sample (resolution power is inversely proportional to the NA).

The first image acquisition setup described here is based on an improved widefield microscope setup, with a regulated light source, rapid and precise z motion and rapid camera for image capture. Since there is no pinhole, light from out-of-focus planes will be recorded, which can be later used in deconvolution (image restoration algorithms which recalculate position of the emitted light based on an ideal or measured light spreading function). The main drawback of this system is phototoxicity from illumination, which prevents rapid frequent time lapse acquisitions.

The second, widely available system is a laser scanning microscope (LSM). These systems have been proven very useful in acquiring very fast time lapse recordings. Their limitation is again their phototoxicity, as the laser beam is focused on a single point in the sample, as well as the scan speed. Imaging is therefore always a compromise between laser power (which increases phototoxicity, while allows speed increase) and scan speed (essential for the identification of rapid movements observed for chromatin *in vivo*).

The third emerging system is based on a widefield high precision rapid microscope, but the light source is a laser, whose beam is focused on a disk with pinholes rotating at high speed. Similar speed as for LSM systems can be achieved, the out-of-focus light is filtered through the pinholes and phototoxicity is not critical, as the whole laser power is never focused on a single point in the sample.

Below, I discuss the individual setups and the critical points of each of these.

2.1.3.1   Widefield high precision rapid microscopy

For the imaging of a large number of cells, best results are obtained with a modified widefield high precision microscope. These microscopes are equipped with a piezo-electric focus either with the objective mounted on it directly or a piezo-electric table, which enables rapid scanning in z and capture of stacks of focal planes. The z distance between planes is carefully controlled and highly reproducible, while movement from one plane to the next is quasi instantaneous. To reduce phototoxicity from mercury lamp illumination or out of focus xenon illumination, generally a xenon lamp is coupled with a monochromator, which restricts the wavelength for fluorescence excitation by nanometer steps (320-680 nm continuous spectrum,

using either a 20 nm or 2-20 nm variable windows). Acquisition is achieved with a high resolution and highly sensitive CCD camera. Since yeast nuclei are small, it is essential if one wants to see fine details that the final pixel size with a 100x objective is between 60 and 80 nm. The readout of the camera by the computer is often the rate limiting factor of the system. Typically, high speed CCD cameras achieve about 30 frames per second, which prevents exposure times shorter than 30 ms.

This modified widefield microscopy is well adapted for measurements of the position of a locus relative to another locus or to a fixed structure (spindle pole body, nuclear periphery, and nucleolus) in a large number of cells. If two loci are to be observed, either two different excitation colors have to be used or the spots have to be of significantly different sizes (achieved by a large difference in the number of binding site repeats between the two tagged loci). 3D stacks of images are necessary to assess 3D localization of the locus, especially when measuring either spot-to-spot distances or evaluating the distance to the periphery (see section 2.3).

2.1.3.2   Laser scanning microscopy

Laser scanning systems are based on the rapid scanning of the sample by an excitation laser and recording of the emitted signal by photomultipliers. The out-of-focus light is blocked by a pinhole. While these systems are extremely well suited to discriminate wavelengths, the scanning speed is often the limiting factor for image acquisition. To track chromatin in individual cells, commercially available systems are well adapted. These systems, although slower than the newer spinning disks (see below), are fast enough to track very rapid chromatin movements (more 0.5 µm in 10 s, measured at 1.5 s intervals) [Heun et al., 2001]. Since the pixel size is set by the user, we recommend using pixel sizes no bigger than 100 nm to track chromatin *in vivo*. Similarly to the z focusing devices described above, a piezo table is essential to achieve speed and reproducibility in z position.

2.1.3.3   Spinning-disk systems

An important alternative to widefield and laser scanning microscopes are spinning disk (SD) confocals. SD microscopes are very similar to widefield systems in their shape. However, the excitation wavelengths are provided by lasers, which focus a beam on pinholes located on a high speed rotating disk. Every point of the focal plane is therefore illuminated several thousand times per second, but only for a fraction of a µs. The emitted light is filtered by passing through the pinholes to remove out-of-focus photons. Acquisition is achieved with a CCD camera, as for widefield systems. The overall quality of the picture is improved since due to the confocality of the system there is not the haze observed in widefield images. Moreover, due to the intermittent excitation of fluorophores by the laser created by the SD, these systems show little bleaching and create less damage to cells. They also allow high frequency imaging (limited again by the acquisition rate of the camera).

2.2 Random walks

In section 2.1, I have described how microscopy images can be acquired. Before I continue with an explanation of how we can extract quantitative kinetic data from these images, I give an introduction to random walks, which can be used as a theoretical model for the movement of a cellular object. These theoretical considerations are the basis both for the analysis (section 2.3) and for the prediction (e.g. chapter 4) of the movement of fluorescently labeled objects.

2.2.1 Free random walks

The most general random walk is a series of steps $\vec{r}_i$ where each step is drawn from a random distribution.

$$\vec{r}_N = \sum_{i=1}^{N} \vec{r}_i \quad \text{with } \vec{r}_i = \begin{pmatrix} a_i \\ b_i \\ c_i \end{pmatrix}. \tag{2.1}$$

I make the usual assumption that each step is independent of all previous steps which makes the random walk a Markov chain [Feller, 1968]. However, I do not assume generally that each step direction is drawn with the same probability, but only that each direction is equally probable as the exactly opposite direction[1]. This set of requirements includes, for example, random walks on a cartesian lattice where every step is parallel to one of the coordinate axes[2]. Importantly, the results presented here do also not require constant step length. They are therefore directly applicable to the analysis of time lapse microscopy movies (see section 2.3.3).

An obvious quantity to ask for is the mean displacement of the walker after a given number of steps. However, from the independency of the steps and the fact that a step in any direction is equally probable as the exactly opposite step it follows that the mean displacement vanishes:

$$\left| \left\langle \sum_{i=1}^{N} \vec{r}_i \right\rangle \right| = \left| \sum_{i=1}^{N} \langle \vec{r}_i \rangle \right| = 0.$$

---

[1]Random walks that do not fulfill this requirement are called random walks with drift, see e.g. [Berg, 1993].

[2]Although for obvious reasons, lattice walks strongly deviate from the real movement of an object or the real conformation of a polymer, they can be used to study so-called universal quantities, which do not depend on the choice of the representative of the respective universality class [Binder, 1996].

However, the mean **squared** displacement can be calculated as

$$\langle \vec{r}_N^2 \rangle = \left\langle \left(\sum_{i=1}^{N} \vec{r}_i \right) \cdot \left(\sum_{j=1}^{N} \vec{r}_j \right) \right\rangle \tag{2.2}$$

$$= \left\langle \sum_{i=1}^{N} (a_i^2 + b_i^2 + c_i^2) + \sum_{i \neq j} (a_i a_j + b_i b_j + c_i c_j) \right\rangle$$

$$= \sum_{i=1}^{N} \left\langle a_i^2 + b_i^2 + c_i^2 \right\rangle + \sum_{i \neq j} \left(\langle a_i a_j \rangle + \langle b_i b_j \rangle + \langle c_i c_j \rangle \right).$$

In the left sum, every summand is the square of the step length l_i. In the right sum, a_i is independent of a_j for $i \neq j$ and therefore $\langle a_i a_j \rangle = \langle a_i \rangle \langle a_j \rangle$. Due to the isotropy of the movement, $\langle a_i \rangle$ vanishes because each positive value of a_i is equally probable as the corresponding negative value. The same is, of course, true for b and c. Thus, we get:

$$\langle \vec{r}_N^2 \rangle = \sum_{i=1}^{N} l_i^2 + 0$$

$$\langle \vec{r}_N^2 \rangle = N l^2 \tag{2.3}$$

where $l = \sqrt{\frac{1}{N} \sum l_i^2}$ is the quadratically averaged step length. This means that for a free random walk, the mean squared displacement is proportional to the number of steps, i.e. proportional to the time between two measurements (see also section 2.2.2).

This is not the case in the presence of obstacles (subdiffusion[3]) or for directed motion (superdiffusion). I did not touch the effects of obstacles and directed motion in this work (see e.g. [Bouchaud and Georges, 1990]). However, I have studied extensively the effects of spatial confinement of the walker. I present the mean squared displacement (MSD) of this kind of movement in section 2.2.3. The MSD analysis is an important tool for quantifying the movement of fluorescently labeled objects in the cell (section 2.3.3).

2.2.2 Random walks and diffusion

In section 1.4, I have introduced diffusion as the random movement of particles. The effects of this process on concentration gradients are described by FICK's first and second law. However, since each of the diffusing particles performs a random walk, there must be a connection between the quantities describing a random walk and those describing diffusion, namely the diffusion coefficient. In this section, I derive FICK's second law based on the properties of a random walk. A comparison of coefficients allows the diffusion coefficient to be expressed using the step length and the time step of the random walk. These are the data we obtain from live cell imaging.

---

[3]also called anomalous or obstructed diffusion

The mean squared displacement is independent of the time between observations.
In section 2.2.1, I have shown that the mean squared displacement of a random walker is
proportional to the number N of steps between the two points of measurement (equation
(2.3)). If one step takes time τ then one can write instead of equation (2.3):

$$\langle \vec{r}^2(t) \rangle = \frac{t}{\tau} l^2$$
$$\langle \vec{r}^2(t) \rangle = \frac{l^2}{\tau} t \tag{2.4}$$

with $t = N\tau$.

If the position is measured only every n steps (e.g. when doing time lapse microscopy
where it is impossible to resolve each step of a molecule), each set of n steps is summed up
to one larger step of quadratically averaged length l_n:

$$l_n^2 = n l^2 \qquad \text{(equation (2.3))}. \tag{2.5}$$

The time between two measurements is $\tau_n = n\tau$ and the mean squared displacement
calculated from what I will call a "coarsened" walk is:

$$\langle \vec{r}^2(t) \rangle = \frac{l_n^2}{\tau_n} t = \frac{n l^2}{n\tau} t = \frac{l^2}{\tau} t,$$

giving the same result as before. This means that the MSD of a free random walk can
be calculated from a simplified walk where each n steps are combined into one larger step
without changing the result. To emphasize that this result is not trivial, one can compare it
to the length of the measured track. Even without an exact calculation (which is difficult
because the mean absolute displacement is much harder to calculate than the mean squared
displacement) it is obvious that the measured track length decreases dramatically when
coarse-graining the track.

In a calculation or simulation, the reverse process is also possible, which is the replacement
of each step by a series of n smaller steps of length $l/\sqrt{n}$. In general, the time step and the
step length can be changed without changing $\langle \vec{r}^2(t) \rangle$ as long as $\frac{l^2}{\tau}$ stays constant:

$$\frac{l^2}{\tau} = c. \tag{2.6}$$

Derivation of Fick's second law I mainly follow the derivation of Einstein [Coffey
et al., 2004, Fürth, 1926]. $C(\vec{x}, t)$ denotes the concentration of particles in a small volume
element at position $\vec{x}$ and at time t. After a short time τ, the concentration is the sum of
the influx of concentration from all volume elements (including the one at position $\vec{x}$):

$$C(\vec{x}, t + \tau) = \int_V C(\vec{x} + \vec{s}, t)\, \Phi(\vec{s})\, d\vec{s}. \tag{2.7}$$

$\Phi(\vec{s})$ denotes the probability for a random walker to make a step of $\vec{s}$. For a random walk with constant step length l, Φ has the form $\Phi(\vec{s}) = \delta(|\vec{s}| - l)$ where δ denotes the delta distribution [Boas, 1983]. However, this is not required for the following derivation.

Under the assumption that τ is small, $C(\vec{x}, t + \tau)$ can be expanded in powers of τ:

$$C(\vec{x}, t + \tau) = C(\vec{x}, t) + \tau \frac{\partial C(\vec{x}, t)}{\partial t} + \mathcal{O}(\tau^2). \tag{2.8}$$

Equally, $C(\vec{x} + \vec{s}, t)$ can be developed in powers of $\vec{s}$:

$$C(\vec{x} + \vec{s}, t) = C(\vec{x}, t) + \sum_{i=1}^{3} s_i \frac{\partial C(\vec{x}, t)}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^{3} s_i s_j \frac{\partial^2 C(\vec{x}, t)}{\partial x_i \partial x_j} + \dots \tag{2.9}$$

With this, equation (2.7) turns into:

$$C(\vec{x}, t) + \tau \frac{\partial C(\vec{x}, t)}{\partial t} + \mathcal{O}(\tau^2) = C(\vec{x}, t) \int_V \Phi(\vec{s}) \, d\vec{s} + \sum_{i=1}^{3} \frac{\partial C(\vec{x}, t)}{\partial x_i} \int_V s_i \, \Phi(\vec{s}) \, d\vec{s}$$

$$+ \frac{1}{2} \sum_{i,j=1}^{3} \frac{\partial^2 C(\vec{x}, t)}{\partial x_i \partial x_j} \int_V s_i s_j \, \Phi(\vec{s}) \, d\vec{s} \tag{2.10}$$

$$+ \frac{1}{6} \sum_{i,j,k=1}^{3} \frac{\partial^2 C(\vec{x}, t)}{\partial x_i \partial x_j \partial x_k} \int_V s_i s_j s_k \, \Phi(\vec{s}) \, d\vec{s} + \dots$$

$\Phi(\vec{s})$ is a probability density function with $\Phi(\vec{s}) = \Phi(-\vec{s})$ (because the walk is unbiased). From this, it follows (see appendix A.1):

$$\int_V s_i^n \, \Phi(\vec{s}) \, d\vec{s} = 0 \text{ for } n \in \{1, 3, 5, \dots\}.$$

Moreover, since s_i and s_j are independent for $i \neq j$, also the corresponding integrals over products of components of $\vec{s}$ vanish if at least one of them occurs in an odd power. Therefore it follows:

$$\int_V \Phi(\vec{s}) \, d\vec{s} = 1$$

$$\int_V s_i \, \Phi(\vec{s}) \, d\vec{s} = 0$$

$$\int_V s_i s_j \, \Phi(\vec{s}) \, d\vec{s} = 0 \text{ for } i \neq j$$

$$\int_V s_i^2 \, \Phi(\vec{s}) \, d\vec{s} = \langle s_i^2 \rangle$$

$$\int_V s_i s_j s_k \, \Phi(\vec{s}) \, d\vec{s} = 0$$

The assumption that all higher terms in equation (2.9) are at least of order τ^2 is justified by equation (2.6). Thus, equation (2.10) can be simplified to:

$$\tau \frac{\partial \, \mathrm{C}(\vec{x},t)}{\partial t} = \tfrac{1}{2} \sum_{i=1}^{3} \langle s_i^2 \rangle \frac{\partial^2 \, \mathrm{C}(\vec{x},t)}{\partial x_i^2}$$

and with $\langle s_1^2 \rangle = \langle s_2^2 \rangle = \langle s_3^2 \rangle = \frac{l^2}{3}$

$$\frac{\partial \, \mathrm{C}(\vec{x},t)}{\partial t} = \frac{l^2}{6\tau} \sum_{i=1}^{3} \frac{\partial^2 \, \mathrm{C}(\vec{x},t)}{\partial x_i^2}$$

$$\frac{\partial \, \mathrm{C}(\vec{x},t)}{\partial t} = \frac{l^2}{6\tau} \Delta \mathrm{C}(\vec{x},t) \tag{2.11}$$

where Δ is the LAPLACE operator [Boas, 1983]. This equation is equivalent to FICK's second law (1.3) and a comparison of coefficients yields:

$$D = \frac{\langle s_i^2 \rangle}{2\tau} = \frac{l^2}{6\tau}. \tag{2.12}$$

With this result, equation (2.3) can be written as:

$$\langle \vec{r}^2(t) \rangle = 6Dt. \tag{2.13}$$

It is important to note that this formulation depends on the number of dimensions within which the walk takes place. The first equality in (2.12) always holds, but $\langle s_i^2 \rangle$ is not necessarily equal to $\frac{l^2}{3}$. It is equal to $\frac{l^2}{d}$ where d is the number of dimensions of the walk. The general form of equations (2.12) and (2.13) is therefore:

$$D = \frac{l^2}{2d\tau} \tag{2.14}$$

$$\langle \vec{r}^2(t) \rangle = 2dDt. \tag{2.15}$$

This result shows that the diffusion coefficient can be derived directly from the mean squared displacement of the random walk.

2.2.3   Confined random walks

In many biological situations where the movement of an object can be modeled as a random walk, the walk is not entirely free. If nothing else, the envelope of the cell or a cellular compartment puts a constraint on the object's movement: the random walk is confined. As a consequence of this confinement, the MSD of the random walk cannot be strictly linear any more because two points of the trajectory cannot be further apart from each other than the

greatest extension of the confining volume. Thus, the MSD curve is bounded. Moreover, with an increase of the time window, the correlation of two points on the trajectory will decrease and finally vanish in the limit of infinite time windows. From these two properties, it follows that the MSD curve must reach a plateau at large time windows. One can expect that the value of this plateau contains information about the size of the confinement. Furthermore, the confinement should have only a small impact on the displacement of the walker on small time scale. Therefore the slope of the curve for very small time windows should reveal information about the diffusion coefficient of the walker.

In this section, I present my results about these properties for different types of confinement.

2.2.3.1 Random walks inside a ball

The most simple type of confinement is a ball. This could be, for example, the whole yeast nucleus, which 3D confocal reconstruction confirms can be approximated as a sphere in G1 and early S phase cells. The simulation of random walks inside such a spherical confinement reveals that the MSD curve is very well described by an exponential approximation to the plateau (see Figure 2.1):

$$\langle r_N^2 \rangle \approx a \left(1 - e^{-N/b} \right) \tag{2.16}$$
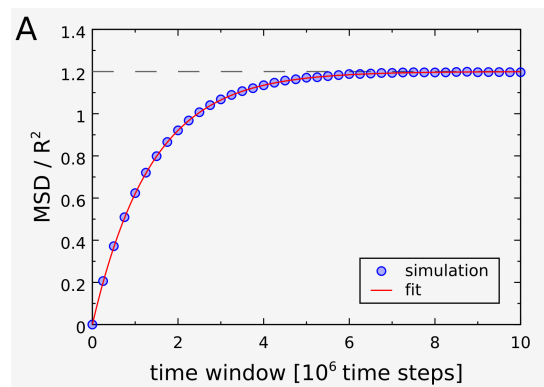
Figure 2.1: MSD of a random walk in a spherical confinement. The curve is very well described by equation (2.16). The data set was derived from 100000 random walk simulations with a step length of 0.001 inside a ball of radius 1.

The agreement is so good that one might speculate that this is indeed the exact relationship. However, I do not know of a proof for this conjecture. Although the entire curve can therefore not be calculated, the value of the plateau can.

Calculation of the plateau value As mentioned in the introductory paragraph the positions of the walker lose their correlation with increasing time window. Therefore the

value of the MSD plateau is the expectation value of the squared distance of two independent points inside the confining ball. I present two different ways to calculate the plateau value.

Calculation using the distance distribution   It is more work to calculate the distribution function of the distance of two uniformly distributed spots than just the average squared distance. However, if the former is available, the latter is derived easily. Since I calculated the distribution function within the scope of one of my projects (section 3.4.2), I show this calculation here.

$P(\rho)d\rho$ is the probability that the distance of two uniformly distributed spots inside a sphere of radius 1 falls into the interval $[\rho, \rho + d\rho]$. I use equation (3.6):

$$P(\rho)d\rho = 3\rho^2 - \tfrac{9}{4}\rho^3 + \tfrac{3}{16}\rho^5 \tag{3.6}$$

With this we get the expected squared distance as

$$\langle\rho^2\rangle = \int\limits_0^2 \rho^2 \cdot P(\rho)\, d\rho$$

$$= \int\limits_0^2 \left(3\rho^4 - \tfrac{9}{4}\rho^5 + \tfrac{3}{16}\rho^7\right) d\rho$$

$$= \tfrac{3}{5}\rho^5 - \tfrac{3}{8}\rho^6 + \tfrac{3}{128}\rho^8 \Big|_0^2$$

$$\langle\rho^2\rangle = \tfrac{6}{5} \tag{2.17}$$

Remark: In a sphere of radius R, we thus have $\langle r^2\rangle = R^2\langle\rho^2\rangle = \tfrac{6}{5}R^2$.

Direct derivation   In the following, I present a more direct way to calculate the radius of constraint. Using this method, I show how the plateau of the MSD calculated from a projected trajectory relates to the full 3D MSD plateau. This is relevant for 3D imaging over time because the number of z-stacks that can be taken rapidly is too small to maximize resolution in z. Therefore a 2D projection can increase the accuracy of analysis (see also section 2.3.3.4).

The idea of the direct derivation is to multiply the probability for spot 1 being at position $\vec{r}_1$ and spot 2 being at position $\vec{r}_2$ with the respective squared distance $(\vec{r}_2 - \vec{r}_1)^2$ and integrate this term over the volumes V_1 and V_2 (which are both the whole ball). Since the spots are uniformly distributed, the probability for residing in a small volume dV is — independently of the position — $\frac{dV}{V}$, where V ($= \frac{4}{3}\pi R^3$) is the volume of the ball. The desired mean squared distance can now be calculated as follows:

$$\langle r^2\rangle = \frac{1}{V^2} \int_V \int_V \left[(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2\right] dV_1\, dV_2$$

$$= \frac{1}{V^2} \int_V \int_V (x_2 - x_1)^2 \, dV_1 \, dV_2$$

$$+ \frac{1}{V^2} \int_V \int_V (y_2 - y_1)^2 \, dV_1 \, dV_2$$

$$+ \frac{1}{V^2} \int_V \int_V (z_2 - z_1)^2 \, dV_1 \, dV_2$$

The problem is symmetric in x, y, and z, and therefore the three integrals above are equal:

$$\langle r^2 \rangle = \frac{3}{V^2} \int_V \int_V (x_2 - x_1)^2 \, dV_1 \, dV_2 \tag{2.18}$$

Moreover, a projection of the two points onto the xy plane just means that z_1 and z_2 are equal to 0, and the third summand vanishes. In general, if we project onto n dimensions we obtain:

$$\langle r^2 \rangle = \frac{n}{V^2} \int_V \int_V (x_2 - x_1)^2 \, dV_1 \, dV_2 \tag{2.19}$$

This integral can be further simplified:

$$\langle r^2 \rangle = \frac{n}{V^2} \int_V \int_V (x_2 - x_1)^2 \, dV_1 \, dV_2$$

$$= \frac{n}{V^2} \int_V \int_V \left(x_2^2 - 2x_1 x_2 + x_1^2 \right) \, dV_1 \, dV_2$$

$$= \frac{n}{V^2} \left[V \int_V x_2^2 \, dV_2 + V \int_V x_1^2 \, dV_1 - 2 \int_V \int_V x_1 x_2 \, dV_1 \, dV_2 \right].$$

In the first two summands, one trivial integration has been executed. The resulting integrals are again equal due to symmetry between x_1 and x_2. The third integral vanishes, because the integration interval for x_1 (the same argument holds for x_2) is symmetric with respect to 0 and the function $x_1 x_2$ is antisymmetric in x_1 (see appendix A.1). Therefore we get:

$$\langle r^2 \rangle = \frac{2n}{V} \int_V x^2 \, dV. \tag{2.20}$$

This integral can be calculated in various ways, I use spherical coordinates[4] here:

$$\langle r^2 \rangle = \frac{2n}{V} \int_V (r \sin\theta \cos\phi)^2 \cdot r^2 \sin\theta \, dr d\theta d\phi$$

$$= \frac{2n}{V} \int_0^R r^4 \, dr \cdot \int_0^\pi \sin^3\theta \, d\theta \cdot \int_0^{2\pi} \cos^2\phi \, d\phi$$

---

[4] `http://mathworld.wolfram.com/SphericalCoordinates.html` (different usage of θ and ϕ)

These are standard integrals, see appendix A.2. We finally get:

$$\langle r^2 \rangle = \frac{2n}{\frac{4}{3}\pi R^3} \cdot \frac{R^5}{5} \cdot \frac{4}{3} \cdot \pi$$

$$\langle r^2 \rangle = \frac{2n}{5} R^2 \tag{2.21}$$

Without projection, we get $\langle r^2 \rangle = \frac{6}{5}R^2$, the same result as with the first approach. If the points are projected onto the xy plane, the MSD plateau has the value $\langle r_p^2 \rangle = \frac{4}{5}R^2$. From this, we can see that at least the plateau of the MSD curve of a confined random walk shows the same relationship between a full 3D trajectory and a trajectory projected to the xy plane as a free random walk (see equation 2.73): $\langle r_p^2 \rangle = \frac{2}{3}\langle r^2 \rangle$. These results will be published in [Neumann et al., *in preparation*].

Importantly, this analysis shows that the radius of the confining ball can be reconstructed from the plateau of the MSD curve:

$$R = \sqrt{\frac{5}{6}\langle r^2 \rangle} \text{ in 3D}$$

$$R = \sqrt{\frac{5}{4}\langle r_p^2 \rangle} \text{ in a 2D projection} \tag{2.22}$$

The slope of the MSD curve As pointed out before, the slope of the MSD curve for small time windows should contain information about the diffusion coefficient of the walker. In fact, it would be plausible if the slope at a time window of 0 would match the slope of the corresponding unconfined walk. The derivative of equation (2.16) is (with $a = \frac{6}{5}R^2$, equation (2.21)):

$$\frac{d}{dN}\frac{6}{5}R^2\left(1 - e^{-N/b}\right) = \frac{\frac{6}{5}R^2}{b}e^{-N/b}$$

$$\left.\frac{d}{dN}\frac{6}{5}R^2\left(1 - e^{-N/b}\right)\right|_{N=0} = \frac{\frac{6}{5}R^2}{b}$$

The slope of the free walk MSD is l^2 (equation (2.3)). If this is equal to the initial slope of the confined walk MSD, then b can be calculated:

$$\frac{\frac{6}{5}R^2}{b} = l^2$$

$$b = \frac{\frac{6}{5}R^2}{l^2} = \frac{6R^2}{5l^2} \tag{2.23}$$

and

$$\langle r_N^2 \rangle \approx \frac{6}{5}R^2\left(1 - e^{-5Nl^2/6R^2}\right) \tag{2.24}$$

$$\langle r^2(t) \rangle \approx \frac{6}{5}R^2\left(1 - e^{-5Dt/R^2}\right). \tag{2.25}$$

In the last line, I have used equation (2.12).

In the case of a simulation, equation (2.24) does not contain any free parameters any more. Therefore it can be compared directly to the MSD calculated from a simulation. If there is a good agreement, then — under the assumption of confinement to a ball — a least squares fit of equation (2.25) to an experimental data set can be used to acquire both the radius of constraint and the unconstrained diffusion coefficient simultaneously (see section 2.3.3).

Figure 2.2 shows the MSD of a simulated random walk confined to a ball in comparison to the theoretical prediction. Equation (2.24) reproduces the simulation result very well, but not as accurately as a fit of function (2.16) with two free parameters (compare Figure 2.1). This does not seem to be a step length effect since the simulation results are indistinguishable for step sizes between 0.001 and 0.1 (Figure 2.2B). Nonetheless, this shows that equation (2.25) is a useful model for the diffusion of a cellular object in a spherical volume.
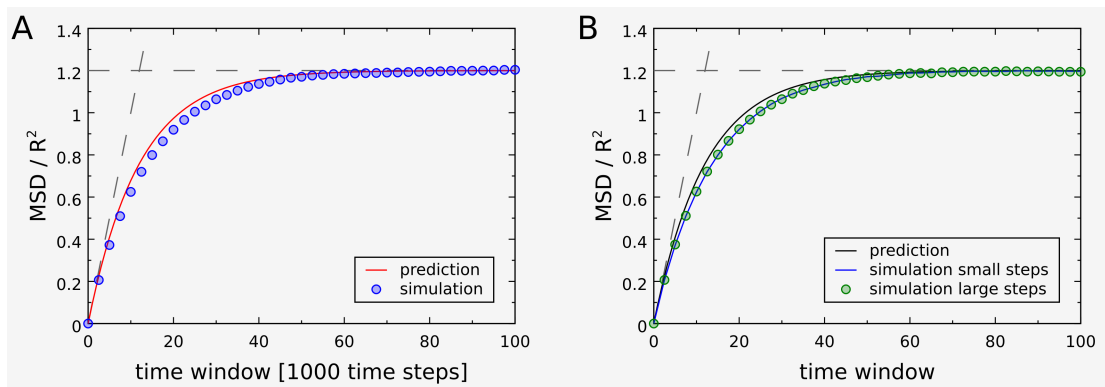
Figure 2.2: Theoretical prediction for the MSD of a random walk inside a ball. (A) MSD predicted by equation (2.24) and calculated from 100000 simulations with $R = 1$ and $l = 0.01$. The dashed lines show the plateau value and the MSD of the corresponding free random walk. (B) MSDs from each 100000 simulations with $l = 0.001$ and $l = 0.1$. The x values had to be rescaled in order to compare the results. A value of 100 on the abscissa corresponds to 10000 time steps in the small step simulation and to 1 time step in the large step simulation (see section 2.2.2).

Importantly, the decay constant b depends only on the quotient $\frac{R}{l}$ and not on R or l explicitly (equation (2.23)). In a simulation, it is obvious that the time required for the MSD to reach the plateau does not change if the same factor is applied to the radius of constraint and the step length. A model function that would not have this property could not be correct. A consequence is that the time to reach the plateau depends quadratically on the radius of constraint (see also equation (2.25)). If the radius of constraint is doubled, both the value of the MSD plateau and the time required to reach it are multiplied by 4.

2.2.3.2 Confinement to the surface of a sphere

An important scaffold for nuclear organization in yeast is the nuclear periphery (see section 1.2.3). Although the interaction of nuclear objects with the periphery is typically transient, certain protein-bound sequences have high affinity for perinuclear structures leading to an enrichment of objects at the nuclear periphery. Therefore, the MSD of movement restricted to the surface of a sphere is worth investigating.

The calculation of the plateau value is analogous to that performed for movement inside the whole ball. One can use the distance distribution from section 3.4.2 (see also page 27):

$$\mathrm{P}(\rho)d\rho = \tfrac{\rho}{2} \tag{3.13}$$

and

$$\langle \rho^2 \rangle = \int\limits_0^2 \rho^2 \cdot \mathrm{P}(\rho)\, d\rho$$

$$= \int\limits_0^2 \tfrac{\rho^3}{2}\, d\rho$$

$$= \tfrac{\rho^4}{8}\Big|_0^2$$

$$\langle \rho^2 \rangle = 2 \tag{2.26}$$

$$\langle r^2 \rangle = 2R^2. \tag{2.27}$$

Alternatively, one can perform the direct calculation (see page 27):

$$\langle r^2 \rangle = \frac{1}{S^2} \int_S \int_S \left[(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2 \right] dS_1\, dS_2$$

and with the same simplifications as before:

$$\langle r^2 \rangle = \frac{2n}{S} \int_S x^2\, dS. \tag{2.28}$$

$S = 4\pi R^2$ is here the surface of the sphere. The integral can be calculated as:

$$\langle r^2 \rangle = \frac{2n}{S} \int_S (R \sin\theta \cos\phi)^2 \cdot R^2 \sin\theta\, d\theta d\phi$$

$$= \frac{2nR^4}{S} \int\limits_0^\pi \sin^3\theta\, d\theta \cdot \int\limits_0^{2\pi} \cos^2\phi\, d\phi$$

$$\langle r^2 \rangle = \frac{2nR^4}{4\pi R^2} \cdot \frac{4}{3} \cdot \pi$$
$$\langle r^2 \rangle = \frac{2n}{3} R^2. \tag{2.29}$$

For $n = 3$ (i.e. without projection), this agrees with the previous result of $\langle r^2 \rangle = 2R^2$.

Computer simulations show that the MSD of a random walk on a spherical surface is also well modeled by an exponential approach towards the plateau, and analogous calculations as before lead to the theoretical prediction:

$$b = \frac{2R^2}{l^2} \tag{2.30}$$

$$\langle r_N^2 \rangle \approx 2R^2 \left(1 - e^{-Nl^2/2R^2} \right) \tag{2.31}$$

$$\langle r^2(t) \rangle \approx 2R^2 \left(1 - e^{-3Dt/R^2} \right). \tag{2.32}$$
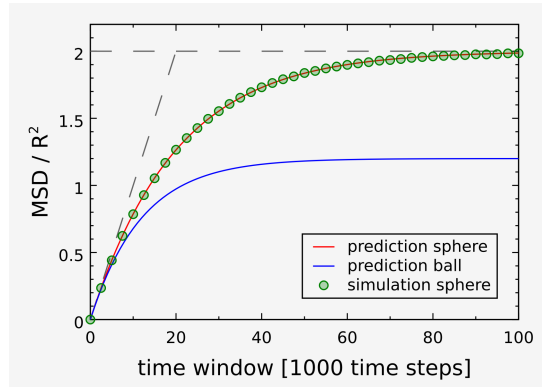
Figure 2.3: Theoretical prediction for the MSD of a random walk on the surface of a sphere. (A) MSD predicted by equation (2.31) and calculated from 100000 simulations with $R = 1$ and $l = 0.01$. The dashed lines show the plateau value and the MSD of the corresponding free random walk. The prediction for the walk inside the sphere is shown in comparison.

I did not perform a statistical analysis but there is no visible deviation between prediction and simulation (Figure 2.3). It might look surprising that the initial slopes of both a walk inside the ball and on its surface are well described by the slope of the free walk MSD. It becomes plausible, however, if one considers the very broad applicability of equation (2.3). Indeed, relevant for cases like this one, section 2.2.1 mentions explicitly that a uniform distribution of step directions is not required for the derivation of this relation.

It has to be noted, however, that the above comparison of walks inside and on the surface of a ball require that the (quadratically averaged) step lengths are equal. This is not the

case, for example, if the surface walk is enforced by just eliminating the vector component of the step that is perpendicular to the surface. This results in a walk with a reduced step length $\hat{l}$ with $\hat{l}^2 = \frac{2}{3}l^2$ [5].

2.3 Extraction of position and mobility of an object from microscopic data

In section 2.1, I have introduced how fluorescent labeling of nuclear objects can be used to monitor their position under the microscope. However, image acquisition is only the first step in determining the position and mobility of a tagged object. In this section, I describe how single images and time lapse movies can be analyzed in order to obtain quantitative parameters in consideration of the limitations of fluorescence microscopy (e.g. with respect to resolution in time and space).

The first part of this section deals with extracting the position of a labeled object with respect to the nuclear envelope or other labeled regions of the nucleus. The second part introduces methods to determine the diffusion coefficient and the size of a possible confinement of the movement. This part makes substantial use of the previous section on random walks.

2.3.1 The 3D position of a tagged locus relative to the nuclear envelope

2.3.1.1 The three zone measurement

In order to determine the position of a tagged locus inside the nucleus, both the position of the center of the nucleus and of the locus have to be reconstructed from the microscopic images. As described in section 2.1, the locus is usually labeled using lac or tet repressors fused to a fluorescent protein. The outline of the nucleus can be determined either by labeling a component of the nuclear pore complex or by using the background fluorescence given by unbound repressor proteins filling the nuclear volume. While the latter method allows reliable identification of the center of the nucleus, it is difficult to measure its exact size since background fluorescence fades at the boundary. Whenever the size of the nucleus or the exact location of the nuclear envelope (NE) is required, nuclear pore staining is recommended, as the boundaries of the nucleus are sharper.

The extraction of the shape of the NE and the position of a fluorescent spot from a stack of microscopic images has to deal with the anisotropy of the data, i.e. the difference in optical resolution along the optical axis of the microscope (z-axis) and perpendicular to

---

[5]I discuss the projection of walks in detail in section 2.3.3.4. Here I use this result, which is true for $l \ll R$, without derivation.

it (x/y axes). One image (x/y direction) has a typical optical resolution of 200 nm (with a 100x objective) and is sampled with a pixel size of 50-100 nm. In contrast, the resolution in z is not better than 300 nm even for a confocal microscope, and the images of a stack are typically taken at 200 nm steps. In addition, the fluorescent staining of the nuclear pores close to the top and bottom of the nucleus becomes diffuse and fuzzy which impairs the reconstruction of the NE.

Ideally, one would want to directly measure the 3D distance between the nuclear rim and the tagged locus. A budding yeast nucleus can be accurately represented by an ellipsoid or even a sphere. One possibility is therefore to fit an ellipsoid to the nuclear pore staining and use it as a model for the NE. Equally, a 3D Gaussian distribution can be fitted to the staining of the locus in order to determine its position with high accuracy. The distance between the locus and the NE (or the center of the nucleus) can then be calculated using the ellipsoid and the position of the spot. However, due to the limited microscopic resolution in the z direction and the small size of the yeast nucleus, the localization of the NE is difficult. Attempts to solve this problem with high accuracy require custom-tailored multi-step processing [Berger et al., 2008], and to date no standard software has been established. On the other hand, if the position of the locus and of nuclear structures like the nucleolus have been determined accurately, a more detailed analysis of nuclear organization can be performed. If only the distance of a locus to the nuclear center is measured, the nucleus is treated as spherically symmetric, which is not the case. For example, the nucleolus and the spindle pole body are located at opposite ends of the nucleus and thereby define a distinct axial pattern. It has been proposed to use this axis as an additional landmark for locus position [Berger et al., 2008]. This allows to detect deviations of the distribution of spot positions from spherical symmetry.

To deal with the poor z resolution of microscopic stacks, an alternative method does not calculate the 3D distance between the spot and the NE directly, but exploits the fact that resolution is better in xy and a spot can be assigned to a specific plane of an image stack. One chooses the plane where the spot is brightest. In this plane, the nucleus is a circle, which can be partitioned into three concentric zones of equal area (Figure 2.4B). The spot position is then sorted into the outermost (zone 1), the intermediate (zone 2), or the innermost zone (zone 3). In order to obtain equal areas for the three zones, the circles separating zones 1 and 2 and 2 and 3, respectively, have to have radii of $\sqrt{2/3}R$ and $\sqrt{1/3}R$, where R is the radius of the nucleus in the chosen plane.

For practical applications, we use the following procedure:

1. Measure the distance between the spot and the periphery along a nuclear diameter as well as the diameter itself. Several programs can be used to extract the coordinates of points of interest from an image. For this task, the freely available pointpicker plug-in[6] for ImageJ is particularly useful.

---

[6]http://bigwww.epfl.ch/thevenaz/pointpicker/

A   B
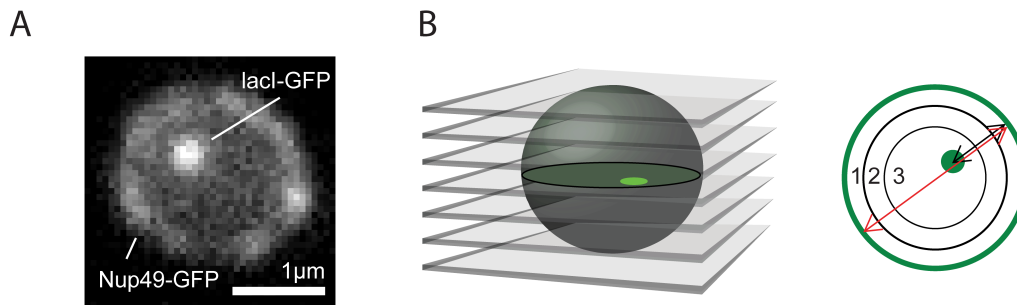
Figure 2.4: (A) Fluorescence microscopy image of a yeast nucleus (1 plane of a stack). Nup49, a component of the nuclear pore complex, is fused to GFP to visualize the nuclear envelope. An array of lac operators is integrated into the genome and gets bound by a lacI-GFP fusion leading to a fluorescent spot. (B) For quantification, the ring representing the nuclear envelope in the plane where the spot is brightest is partitioned into three zones of equal area. The nuclear diameter in this plane (red arrow) and the distance of the spot to the periphery (black arrow) are measured and the spot is sorted into zone 1, 2, or 3.

2. Normalize the spot pore distance to the radius[7] of the circle.

3. Sort the spot into zone 1 (if the normalized distance is less than $1 - \sqrt{2/3}$), zone 2 (if it is between $1 - \sqrt{2/3}$ and $1 - \sqrt{1/3}$), or zone 3 (greater than $1 - \sqrt{1/3}$).

4. Compare the measured distribution to another one (different strain, condition etc.) or to a uniform distribution using e.g. a χ^2- test. If only percentages of one zone (e.g. the outermost zone) are compared, a proportional test should be used.

By this method, the whole nucleus is divided into three equally sized volumes (this follows from the equality of the three cross sections at each z according to CAVALIERI's principle). Therefore a uniformly distributed locus will be found with an equal probability of $\frac{1}{3}$ in each of the three zones. However, the three zones do not coincide exactly with three concentric shells of equal volume, which is the desired partition of the nucleus for assessing if a locus is enriched at the nuclear periphery (Figure 2.5). Because of this deviation, the zone measurement is not an exact method to measure whether a labeled locus is enriched in one of the three shells, but importantly, the uniform distribution of a spot is detected without error (this is a direct consequence of the volume equality).

---

[7]The distance is normalized to the radius according to mathematical convention. In principle, it is equally possible to normalize to the diameter, but then the following numbers have to be adapted.
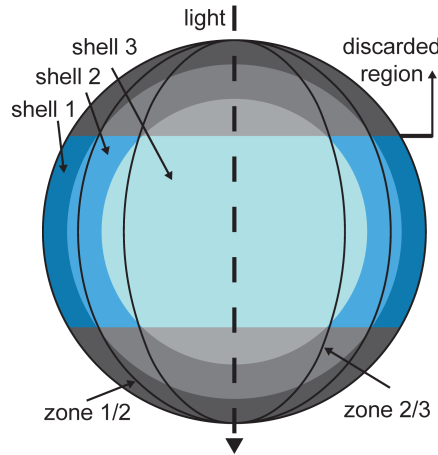
Figure 2.5: Vertical cut through the nucleus. Three shells of equal volume are shown in shades of blue. The three zones also add up to three equal volumes (the boundaries are shown as black lines), but these do not coincide exactly with the shells. For image quality reasons (see text), measurements in the pole regions of the nucleus have to be removed from the sample. These regions are displayed in grey. The removal affects neither each zone nor each shell to equal extent.

As mentioned before, the measurement of spot position with respect to the NE is particularly difficult close to the poles of the nucleus. To avoid severe measurement errors, cells in which the tagged locus resides close to the top or bottom of the nucleus have to be discarded. This decapping can include up to 25% of the planes on each side. This affects both shell (ideal 3D distance measurement) and zone measurements (as described above), because peripheral spots are more likely to be discarded than interior ones.

Below, I calculate the error that is introduced by the zone measurement and the impact of decapping on both shell and zone measurement results.

2.3.1.2   Calculation of the results of shell and zone measurement

The approach is to calculate the volumes of the nine intersections $Z_i \cap S_j, 1 \leq i, j \leq 3$ where Z denotes a zone and S a shell. If a distribution of spots in the shells is given, one can calculate how many spots will be measured in each zone or shell. The decapping of the nucleus can be expressed by only taking the volumes for $-c \leq z \leq c$ into account.

I normalize all lengths to R (the radius of the nucleus), all areas to πR^2, and all volumes to $\frac{4}{3}\pi R^3$. Usually, Greek letters denote the dimensionless partners of their Latin counterparts, but this Greek/Latin correspondence cannot always be held up. I use the following identifiers:

- $\gamma = \frac{c}{R}$: decapping parameter (see above)

- $\zeta = \frac{z}{R}$: normalized z coordinate

- $\Omega(\mathfrak{S}, \gamma)$: normalized volume of a subset $\mathfrak{S}$ of the nucleus with the restriction $-\gamma \leq \zeta < \gamma$

- N: total number of cells including the ones where the measurement is discarded

- ϵ_j: fraction of spots in shell S_j (the quantity one wants to measure)

Then the measurement results can be expressed as follows:

- $\epsilon_j N \cdot \dfrac{\Omega(S_j, \gamma)}{\Omega(S_j, 1)}$: spots in shell S_j that are not discarded.

- $\epsilon_j N \cdot \dfrac{\Omega(S_j, \gamma)}{\Omega(S_j, 1)} \cdot \dfrac{\Omega(Z_i \cap S_j, \gamma)}{\Omega(S_j, \gamma)} = \epsilon_j N \cdot \dfrac{\Omega(Z_i \cap S_j, \gamma)}{\Omega(S_j, 1)}$: spots in shell S_j that are measured in zone Z_i

- $\alpha_i = \dfrac{\sum_j \epsilon_j N \cdot \dfrac{\Omega(Z_i \cap S_j, \gamma)}{\Omega(S_j, 1)}}{\sum_j \epsilon_j N \cdot \dfrac{\Omega(S_j, \gamma)}{\Omega(S_j, 1)}} = \dfrac{\sum_j \epsilon_j \cdot \dfrac{\Omega(Z_i \cap S_j, \gamma)}{\Omega(S_j, 1)}}{\sum_j \epsilon_j \cdot \dfrac{\Omega(S_j, \gamma)}{\Omega(S_j, 1)}}$:

  fraction of spots measured in zone Z_i, the readout of the zone measurement. It should be noted that the number of the spots measured in zone i has to be normalized to the total number of measured spots, which is not N. This is done in the analysis of experiments, and otherwise the fractions of zones 1, 2, and 3 would not sum up to 1. The number of measured spots depends on the three epsilon values, and this is the reason why α_i can depend non-linearly on ϵ_i (see e.g. Figure 2.7B).

- $\beta_i = \dfrac{\epsilon_i \cdot \dfrac{\Omega(S_i, \gamma)}{\Omega(S_i, 1)}}{\sum_j \epsilon_j \cdot \dfrac{\Omega(S_j, \gamma)}{\Omega(S_j, 1)}}$: fraction of spots measured in shell S_i.

Given these considerations, the problem has been reduced to calculating the volumes $\Omega(S_j, \gamma)$ and $\Omega(Z_i \cap S_j, \gamma)$ for $1 \leq i, j \leq 3$ and $0 \leq \gamma \leq 1$. I decided to use cartesian coordinates. The main integration is the one along the z axis. Due to the symmetry of the problem with respect to this axis all xy cross sections of the volumes of interest are rings (possibly degenerated to discs). Furthermore, the problem is symmetric with respect to the xy plane. I use the following identifiers:

- $A(\zeta)$: a general normalized area at height ζ

- $\Gamma(\mu_1, \mu_2)$: a general normalized volume between heights $0 \leq \mu_1 \leq \mu_2 \leq 1$. $\Gamma(\mu_1, \mu_2)$ already takes the symmetry with respect to the xy plane into account. This means that the volume comprises both the regions with $\mu_1 \leq \zeta \leq \mu_2$ and with $-\mu_2 \leq \zeta \leq -\mu_1$.

The relation between Γ and an integral over A has to respect the different types of normalization (the factor 2 reflects the symmetry that I just discussed):

$$\overbrace{\tfrac{4}{3}\pi R^3\, \Gamma(\mu_1, \mu_2)}^{\text{unnormalized volume}} = 2 \int_{\mu_1}^{\mu_2} \overbrace{\pi R^2\, \mathrm{A}(\zeta)}^{\text{unn. area}} \cdot \overbrace{R d\zeta}^{\text{unn. height}}$$

$$\Gamma(\mu_1, \mu_2) = \tfrac{3}{2} \int_{\mu_1}^{\mu_2} \mathrm{A}(\zeta)\, d\zeta$$

As mentioned, the xy cross sections of the volumes of interest are rings. Therefore A and Γ can be rewritten as:

$$\pi R^2\, \mathrm{A}(\zeta) = \pi(R\, \Psi_2(\zeta))^2 - \pi(R\, \Psi_1(\zeta))^2$$
$$\mathrm{A}(\zeta) = \Psi_2^2(\zeta) - \Psi_1^2(\zeta),$$

where Ψ_1 and Ψ_2 are the normalized outer and inner radius of the ring, and

$$\Gamma(\mu_1, \mu_2) = \tfrac{3}{2} \int_{\mu_1}^{\mu_2} (\Psi_2^2(\zeta) - \Psi_1^2(\zeta))\, d\zeta.$$

The boundaries of the rings are always either the boundary of a zone or the boundary of a shell, which I denote with ρ and σ, respectively. I use the following constants:

- $\rho_0 = 1$, $\sigma_0 = 1$, for convenience in notation

- $\rho_1 = \sqrt{\tfrac{2}{3}}$, $\sigma_1 = \sqrt[3]{\tfrac{2}{3}}$, separating zones 1/2 and shells 1/2, respectively

- $\rho_2 = \sqrt{\tfrac{1}{3}}$, $\sigma_2 = \sqrt[3]{\tfrac{1}{3}}$, separating zones 2/3 and shells 2/3, respectively

- $\rho_3 = 0$, $\sigma_3 = 0$, for convenience in notation

The radius (i.e. the distance from the z axis) that separates two zones at height ζ is always of the form $\rho_i \sqrt{1 - \zeta^2}$, and the radius that separates two shells is of the form $\sqrt{\sigma_i^2 - \zeta^2}$. This follows directly from the definition of zones and spheres and the theorem of PYTHAGORAS.

Four types of integrals will occur during the calculation. Always ρ and ρ' are elements of $\{\rho_0, ..., \rho_3\}$, and σ and σ' are elements of $\{\sigma_0, ..., \sigma_3\}$.

1. Both boundaries are zone boundaries:

$$
\Gamma^{(\rho\rho)}(\mu_1, \mu_2, \rho, \rho') = \tfrac{3}{2} \int\limits_{\mu_1}^{\mu_2} \left[\rho^2(1 - \zeta^2) - \rho'^2(1 - \zeta^2) \right] d\zeta
$$

$$
= \tfrac{3}{2} \int\limits_{\mu_1}^{\mu_2} \left[(\rho^2 - \rho'^2) - (\rho^2 - \rho'^2)\zeta^2 \right] d\zeta
$$

$$
\Gamma^{(\rho\rho)}(\mu_1, \mu_2, \rho, \rho') = \tfrac{3}{2}(\rho^2 - \rho'^2)(\mu_2 - \mu_1) - \tfrac{1}{2}(\rho^2 - \rho'^2)(\mu_2^3 - \mu_1^3). \tag{2.33}
$$

2. The outer boundary is a zone boundary, the inner a shell boundary:

$$
\Gamma^{(\rho\sigma)}(\mu_1, \mu_2, \rho, \sigma) = \tfrac{3}{2} \int\limits_{\mu_1}^{\mu_2} \left[\rho^2(1 - \zeta^2) - (\sigma^2 - \zeta^2) \right] d\zeta
$$

$$
= \tfrac{3}{2} \int\limits_{\mu_1}^{\mu_2} \left[(1 - \rho^2)\zeta^2 - (\sigma^2 - \rho^2) \right] d\zeta
$$

$$
\Gamma^{(\rho\sigma)}(\mu_1, \mu_2, \rho, \sigma) = \tfrac{1}{2}(1 - \rho^2)(\mu_2^3 - \mu_1^3) - \tfrac{3}{2}(\sigma^2 - \rho^2)(\mu_2 - \mu_1). \tag{2.34}
$$

3. The outer boundary is a shell boundary, the inner a zone boundary:

$$
\Gamma^{(\sigma\rho)}(\mu_1, \mu_2, \sigma, \rho) = -\,\Gamma^{(\rho\sigma)}(\mu_1, \mu_2, \rho, \sigma) \tag{2.35}
$$

4. Both boundaries are shell boundaries:

$$
\Gamma^{(\sigma\sigma)}(\mu_1, \mu_2, \sigma, \sigma') = \tfrac{3}{2} \int\limits_{\mu_1}^{\mu_2} \left[(\sigma^2 - \zeta^2) - (\sigma'^2 - \zeta^2) \right] d\zeta
$$

$$
= \tfrac{3}{2} \int\limits_{\mu_1}^{\mu_2} \left[(\sigma^2 - \sigma'^2) \right] d\zeta
$$

$$
\Gamma^{(\sigma\sigma)}(\mu_1, \mu_2, \sigma, \sigma') = \tfrac{3}{2}(\sigma^2 - \sigma'^2)(\mu_2 - \mu_1). \tag{2.36}
$$

The surface of the whole sphere can be used either as a zone or as a shell boundary. In the following, I consider it as a shell boundary. If the inner boundary is degenerated to the z axis, it can be considered as a zone boundary with $\rho = \rho_3 = 0$.

Depending on the size of the removed caps (i.e. the value of γ), some case differentiations have to be done. The boundary values of the case differentiations are the values of ζ at which zone and shell boundaries intersect (see Figure 2.6):
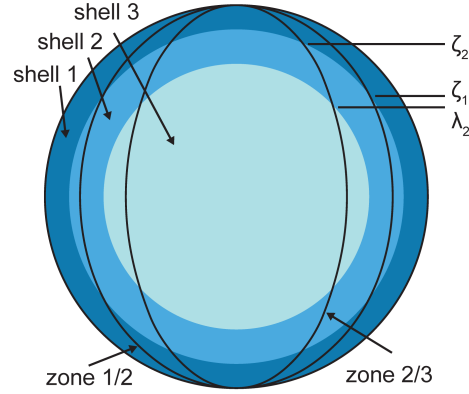
Figure 2.6: Vertical cut through the nucleus. Three shells of equal volume are shown in shades of blue. The zone boundaries are shown as black lines. Important for the calculations are the values of ζ for the intersections of zone and shell boundaries. $\zeta = \zeta_1$ marks the intersection of the boundaries between zones 1/2 and shells 1/2, $\zeta = \zeta_2$ the intersection of the boundaries between zones 2/3 and shells 1/2, and $\zeta = \lambda_2$ the intersection of the boundaries between zones 2/3 and shells 2/3.

$$\rho_1(1 - \zeta_1^2) = \sigma_1^2 - \zeta_1^2 \quad \Longleftrightarrow \quad \zeta_1 = \sqrt{\frac{\sigma_1^2 - \rho_1^2}{1 - \rho_1^2}}$$

$$\rho_2(1 - \zeta_2^2) = \sigma_1^2 - \zeta_2^2 \quad \Longleftrightarrow \quad \zeta_2 = \sqrt{\frac{\sigma_1^2 - \rho_2^2}{1 - \rho_2^2}}$$

$$\rho_2(1 - \lambda_2^2) = \sigma_2^2 - \lambda_2^2 \quad \Longleftrightarrow \quad \lambda_2 = \sqrt{\frac{\sigma_2^2 - \rho_2^2}{1 - \rho_2^2}}$$

λ_1 would denote the height at which the boundaries between zones 1/2 and shells 2/3 intersect, but this intersection does not exist.

With these preparatory results, the volumes of the shells and their intersections with the zones can be written as:

$$\Omega(S_1, \gamma) = \begin{cases} \Gamma^{(\sigma\sigma)}(0, \gamma, \sigma_0, \sigma_1), & \text{for } 0 \leq \gamma < \sigma_1 \\ \Gamma^{(\sigma\sigma)}(0, \sigma_1, \sigma_0, \sigma_1) + \Gamma^{(\sigma\rho)}(\sigma_1, \gamma, \sigma_0, \rho_3), & \text{for } \sigma_1 \leq \gamma \leq 1 \end{cases} \tag{2.37}$$

$$\Omega(S_2, \gamma) = \begin{cases} \Gamma^{(\sigma\sigma)}(0, \gamma, \sigma_1, \sigma_2), & \text{for } 0 \le \gamma < \sigma_2 \\ \Gamma^{(\sigma\sigma)}(0, \sigma_2, \sigma_1, \sigma_2) + \Gamma^{(\sigma\rho)}(\sigma_2, \gamma, \sigma_1, \rho_3), & \text{for } \sigma_2 \le \gamma < \sigma_1 \\ \Gamma^{(\sigma\sigma)}(0, \sigma_2, \sigma_1, \sigma_2) + \Gamma^{(\sigma\rho)}(\sigma_2, \sigma_1, \sigma_1, \rho_3), & \text{for } \sigma_1 \le \gamma \le 1 \end{cases} \quad (2.38)$$

$$\Omega(S_3, \gamma) = \begin{cases} \Gamma^{(\sigma\rho)}(0, \gamma, \sigma_2, \rho_3), & \text{for } 0 \le \gamma < \sigma_2 \\ \Gamma^{(\sigma\rho)}(0, \sigma_2, \sigma_2, \rho_3), & \text{for } \sigma_2 \le \gamma \le 1 \end{cases} \quad (2.39)$$

$$\Omega(Z_1 \cap S_1, \gamma) = \begin{cases} \Gamma^{(\sigma\sigma)}(0, \gamma, \sigma_0, \sigma_1), & \text{for } 0 \le \gamma < \zeta_1 \\ \Gamma^{(\sigma\sigma)}(0, \zeta_1, \sigma_0, \sigma_1) + \Gamma^{(\sigma\rho)}(\zeta_1, \gamma, \sigma_0, \rho_1), & \text{for } \zeta_1 \le \gamma \le 1 \end{cases} \quad (2.40)$$

$$\Omega(Z_1 \cap S_2, \gamma) = \begin{cases} \Gamma^{(\sigma\rho)}(0, \gamma, \sigma_1, \rho_1), & \text{for } 0 \le \gamma < \zeta_1 \\ \Gamma^{(\sigma\rho)}(0, \zeta_1, \sigma_1, \rho_1), & \text{for } \zeta_1 \le \gamma \le 1 \end{cases} \quad (2.41)$$

$$\Omega(Z_1 \cap S_3, \gamma) = 0 \quad (2.42)$$

$$\Omega(Z_2 \cap S_1, \gamma) = \begin{cases} 0, & \text{for } 0 \le \gamma < \zeta_1 \\ \Gamma^{(\rho\sigma)}(\zeta_1, \gamma, \rho_1, \sigma_1), & \text{for } \zeta_1 \le \gamma < \zeta_2 \\ \Gamma^{(\rho\sigma)}(\zeta_1, \zeta_2, \rho_1, \sigma_1) + \Gamma^{(\rho\rho)}(\zeta_2, \gamma, \rho_1, \rho_2), & \text{for } \zeta_2 \le \gamma \le 1 \end{cases} \quad (2.43)$$

$$\Omega(Z_2 \cap S_2, \gamma) = \begin{cases} \Gamma^{(\rho\sigma)}(0, \gamma, \rho_1, \sigma_1), & \text{for } 0 \le \gamma < \lambda_2 \\ \Gamma^{(\rho\sigma)}(0, \lambda_2, \rho_1, \sigma_1) + \Gamma^{(\rho\rho)}(\lambda_2, \gamma, \rho_1, \rho_2), & \text{for } \lambda_2 \le \gamma < \zeta_1 \\ \Gamma^{(\rho\sigma)}(0, \lambda_2, \rho_1, \sigma_1) + \Gamma^{(\rho\rho)}(\lambda_2, \zeta_1, \rho_1, \rho_2) \\ \quad + \Gamma^{(\sigma\rho)}(\zeta_1, \gamma, \sigma_1, \rho_2), & \text{for } \zeta_1 \le \gamma < \zeta_2 \\ \Gamma^{(\rho\sigma)}(0, \lambda_2, \rho_1, \sigma_1) + \Gamma^{(\rho\rho)}(\lambda_2, \zeta_1, \rho_1, \rho_2) \\ \quad + \Gamma^{(\sigma\rho)}(\zeta_1, \zeta_2, \sigma_1, \rho_2), & \text{for } \zeta_2 \le \gamma \le 1 \end{cases} \quad (2.44)$$

$$\Omega(Z_2 \cap S_3, \gamma) = \begin{cases} \Gamma^{(\sigma\rho)}(0, \gamma, \sigma_2, \rho_2), & \text{for } 0 \le \gamma < \lambda_2 \\ \Gamma^{(\sigma\rho)}(0, \lambda_2, \sigma_2, \rho_2), & \text{for } \lambda_2 \le \gamma \le 1 \end{cases} \quad (2.45)$$

$$\Omega(Z_3 \cap S_1, \gamma) = \begin{cases} 0, & \text{for } 0 \le \gamma < \zeta_2 \\ \Gamma^{(\rho\sigma)}(\zeta_2, \gamma, \rho_2, \sigma_1), & \text{for } \zeta_2 \le \gamma < \sigma_1 \\ \Gamma^{(\rho\sigma)}(\zeta_2, \sigma_1, \rho_2, \sigma_1) + \Gamma^{(\rho\rho)}(\sigma_1, \gamma, \rho_2, \rho_3), & \text{for } \sigma_1 \le \gamma \le 1 \end{cases} \quad (2.46)$$

$$\Omega(Z_3 \cap S_2, \gamma) = \begin{cases} 0, & \text{for } 0 \le \gamma < \lambda_2 \\ \Gamma^{(\rho\sigma)}(\lambda_2, \gamma, \rho_2, \sigma_2), & \text{for } \lambda_2 \le \gamma < \sigma_2 \\ \Gamma^{(\rho\sigma)}(\lambda_2, \sigma_2, \rho_2, \sigma_2) + \Gamma^{(\rho\rho)}(\sigma_2, \gamma, \rho_2, \rho_3), & \text{for } \sigma_2 \le \gamma < \zeta_2 \\ \Gamma^{(\rho\sigma)}(\lambda_2, \sigma_2, \rho_2, \sigma_2) + \Gamma^{(\rho\rho)}(\sigma_2, \zeta_2, \rho_2, \rho_3) \\ \quad + \Gamma^{(\sigma\rho)}(\zeta_2, \gamma, \sigma_1, \rho_3), & \text{for } \zeta_2 \le \gamma < \sigma_1 \\ \Gamma^{(\rho\sigma)}(\lambda_2, \sigma_2, \rho_2, \sigma_2) + \Gamma^{(\rho\rho)}(\sigma_2, \zeta_2, \rho_2, \rho_3) \\ \quad + \Gamma^{(\sigma\rho)}(\zeta_2, \sigma_1, \sigma_1, \rho_3), & \text{for } \sigma_1 \le \gamma \le 1 \end{cases} \quad (2.47)$$

$$\Omega(Z_3 \cap S_3, \gamma) = \begin{cases} \Gamma^{(\rho\rho)}(0, \gamma, \rho_2, \rho_3), & \text{for } 0 \le \gamma < \lambda_2 \\ \Gamma^{(\rho\rho)}(0, \lambda_2, \rho_2, \rho_3) + \Gamma^{(\sigma\rho)}(\lambda_2, \gamma, \sigma_2, \rho_3), & \text{for } \lambda_2 \le \gamma < \sigma_2 \\ \Gamma^{(\rho\rho)}(0, \lambda_2, \rho_2, \rho_3) + \Gamma^{(\sigma\rho)}(\lambda_2, \sigma_2, \sigma_2, \rho_3), & \text{for } \sigma_2 \le \gamma \le 1 \end{cases} \qquad (2.48)$$

With these results, the given equations for α_i and β_i (see page 37) can be used to calculate the zone and shell measurement results for any given γ and $\epsilon_{1,2,3}$. I have implemented the calculation as a Perl[8] class.

2.3.1.3 Results and discussion

In order to determine the most accurate method for quantifying object distributions (2D-zone or 3D-shell), I calculated their deviations from the respective true distribution. Figure 2.7 shows how well enrichments in the three different zones are measured by either shell or zone measurements with no or maximal decapping ($c = 0.5R$) of the nucleus. Without decapping the shell measurement gives exactly what is supposed to be measured and is therefore without error (Figure 2.7, panels A, C, E). The zone measurement reflects a uniform distribution of spots correctly, but without decapping it shows a considerable deviation from the true enrichment, especially in the intermediate zone. Importantly, this deviation is always conservative, i.e. both depletion and enrichment are underestimated. Therefore a measured depletion or enrichment cannot be an artifact of the method.

With a decapping of $0.5R$ at the top and bottom of the nucleus, both zone and shell measurement results are changed (Figure 2.7, panels B, D, F). The shell measurement stays accurate only for extreme depletions or enrichments. An enrichment in the outermost shell is now more precisely reported by the zone measurement than by the shell measurement. However, this is not the case for the intermediate shell, where the shell measurement performs much better. For the innermost shell, the shell measurement is best at extreme enrichment values, but shows a moderate enrichment even when there is none. The zone measurement performs very well for depletion and small enrichments, but quite badly for strong enrichments. However, the zone measurement — in contrast to the shell measurement — is always conservative, i.e. closer to a uniform distribution than is the true distribution of spots.

Figure 2.8 compares the measured enrichments as a function of the amount of decapping. As a consequence of the design of the zone measurement, it always reflects the uniform distribution of spots exactly, independently of the amount of decapping. The data for an enrichment of 60% in the respective zone show again that the zone measurement is best for the outermost shell, where it performs better than the shell measurement in the practical range of decapping, the shell measurement is best for the intermediate shell, and they are both rather inaccurate for the innermost shell. The shell measurement consistently overestimates an enrichment in the innermost shell. Both methods perform better at $0.4R$ than at $0.5R$ decapping.

---

[8]see `http://www.perl.org`
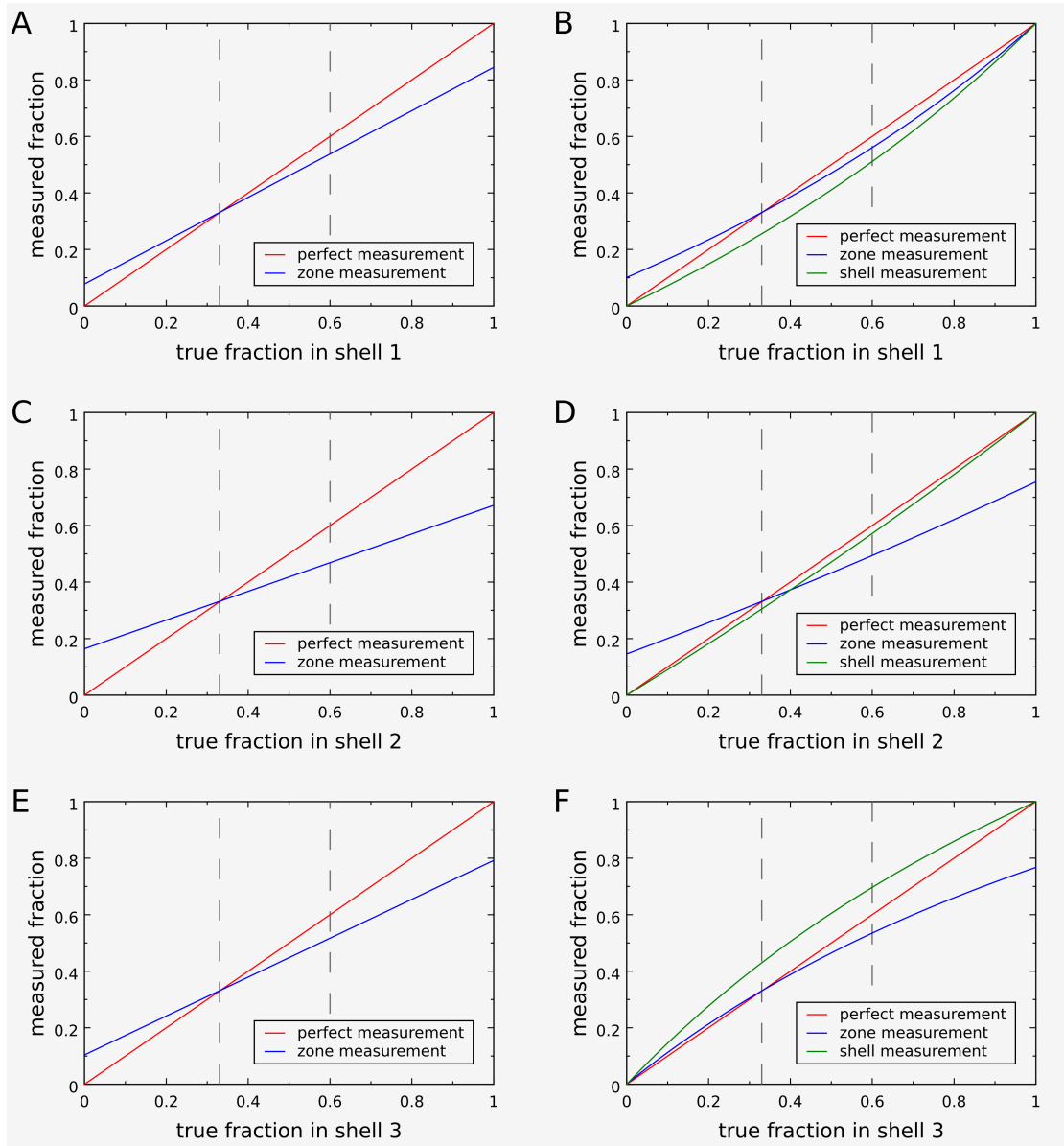
Figure 2.7: Measured fractions in zones and shells depending on the true fraction in the respective shell. Without decapping of the nucleus (A, C, E) the shell measurement is always exact. If the measurements in the pole caps of the nucleus (height $0.5R$, compare Figure 2.5) are discarded (B, D, F) both types of measurements deviate from the true enrichment. A fraction of $\frac{1}{3}$ corresponds to a uniform distribution, 0.6 is a typical fraction e.g. for an anchored yeast telomere [Hediger et al., 2002].
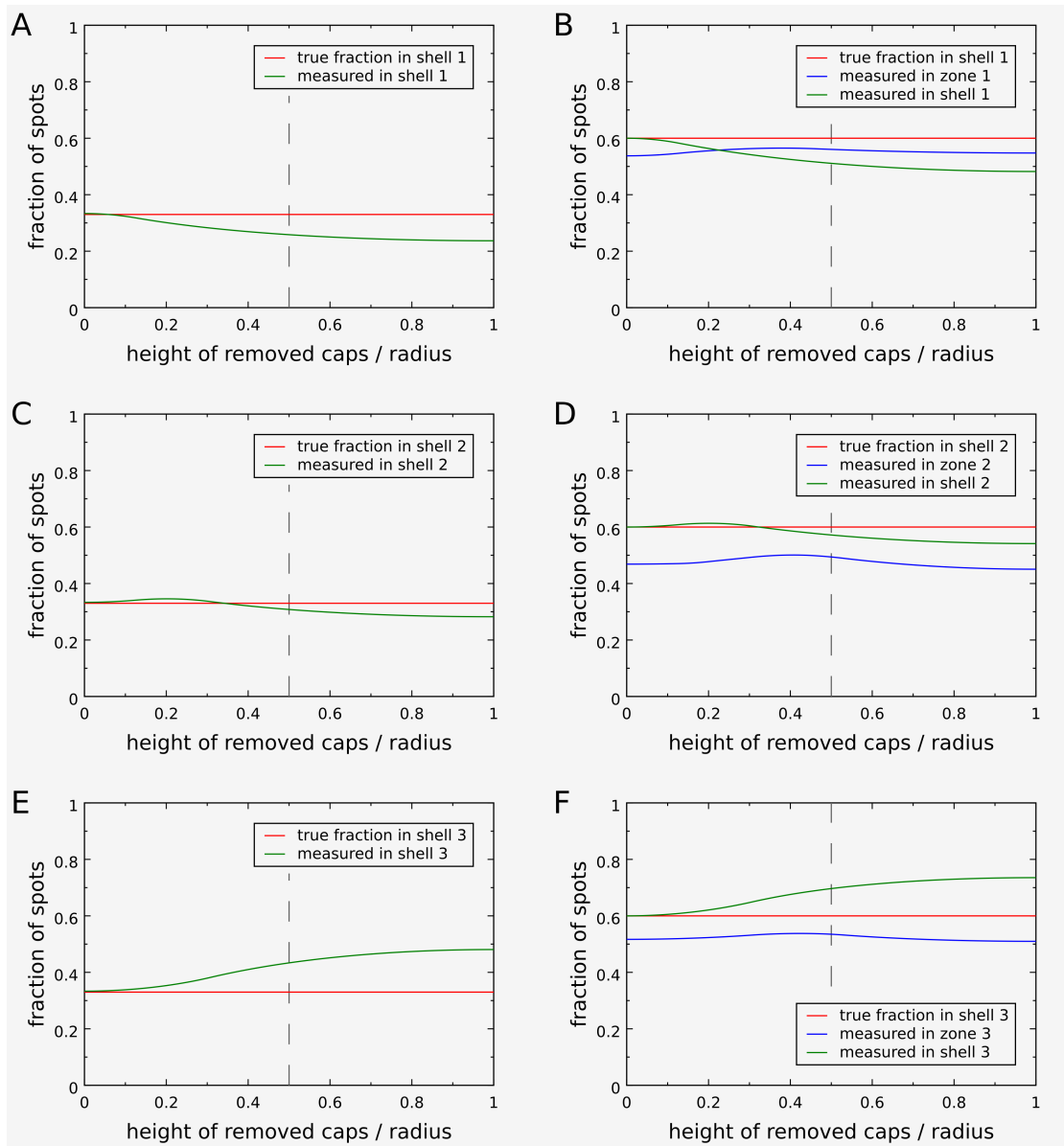
Figure 2.8: Measured fractions in zones and shells depending on the amount of decapping. For a uniform distribution of spots (A, C, E) the zone measurement is exact, independently of the amount of decapping. The shell measurement is best in shell 2 but suffers severely from the decapping in shells 1 and 3. Panel B, D, F show the measured fractions of spots for an fraction of 60% in the respective shell.

Although Figures 2.7 and 2.8 are the natural way to present the results of my calculations, it remained difficult to assess what they imply for the three zone analysis of experimental data. Therefore, I plotted the results for some scenarios that can arise in experimental data sets (Figure 2.9). As one would expect based on the results shown above, a high fraction (80%) of spots in the outermost shell is well reproduced by the zone measurement (panel A). This is also true for depletion in this shell, although it is slightly underestimated (panel B).

In contrast, an enrichment in the intermediate shell is severely underestimated by the zone measurement (panels C and D). Nonetheless, a fraction of 50% (panel D) can still be distinguished significantly from a random distribution at a sample size of 90 or more. For the innermost shell (panels E and F), even a sample size of 50 is sufficient.

I have already pointed out that the zone measurement is always conservative and that therefore the depletion or enrichment of a locus in one of the zones cannot be an artifact of the method. In contrast to this reliability of the method when comparing the fraction in one zone to that of a uniform distribution, it can be misleading to compare the values for different zones to each other without correction. For example, it cannot be concluded from Figure 2.9C that the spot resides more often in the outermost than in the innermost shell (the true fractions are both 10%). The same is true for the shell measurement in the presence of decapping (e.g. Figure 2.9D).

The analytical calculation of the errors of zone and shell measurement raises the question if the results can be used to compensate for these errors. However, a critical requirement for such a correction is that the amount of decapping is known and constant throughout the measured sample. In practical situations, this is often not the case, as it depends on the quality of the images. Furthermore, the above analysis is based on the assumption that the spot distribution is spherically symmetric. In a cell, this assumption is not always fulfilled exactly. For example, the nucleolus occupies a certain region close to the periphery and thereby blocks it for many objects. These restrictions do not invalidate the above analysis but they make it questionable if an error compensation based on the analysis would make the results more reliable.

In summary, the zone measurement is a useful tool to determine the 3D position of a fluorescent spot with respect to the nuclear envelope, specifically if spots in the pole regions of the nucleus are discarded. It performs best in the outermost zone (zone 1), that is for identifying perinuclear or NE association. The deviations of the measured fractions of spots in zone 1, 2, and 3 from the true fractions in the respective shells are always conservative for both depletions and enrichments, which implies that a measured deviation from the uniform distribution cannot be an artifact. In contrast, the comparison of two zones to each other can be misleading. One practical outcome of this analysis is that a decapping of $0.4R$ is favorable compared to $0.5R$.
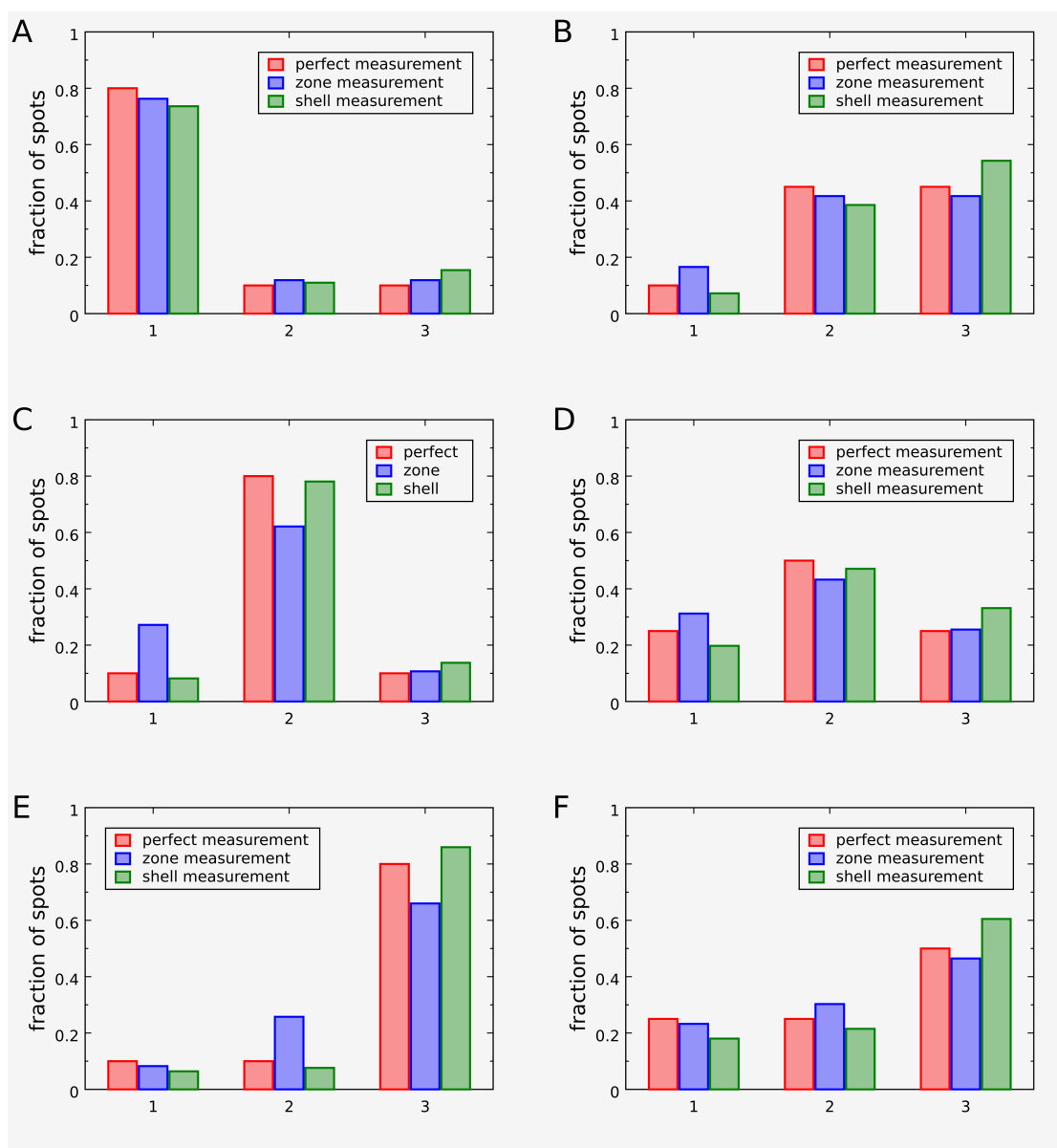
Figure 2.9: Measured fractions in zones and shells in typical scenarios: (A) High fraction (80%) in the outermost shell (shell 1). (B) Depletion (10% of the spots) in shell 1. (C) High fraction (80%) in the intermediate shell (shell 2). (D) Moderate enrichment (50% of the spots) in shell 2. (E) High fraction (80%) in the innermost shell (shell 3). (F) Moderate enrichment (50% of the spots) in shell 3.

2.3.2   The position relative to other structures

In order to further investigate the function of locus position, it is interesting to know if a locus binds to a structural component of the nucleus. This can be investigated by staining the locus in one color and the structure of interest in another color and observing their colocalization. However, unless the locus is excluded from the structure, even in the absence of specific binding, a certain level of random colocalization will be detected, and its degree depends on size and form of the structure. In order to assess whether experimentally obtained colocalization values are significant or not, it is helpful to know the expected degree of colocalization for a uniformly distributed locus. This can be calculated as the ratio between the volume of the region in which the spot is considered colocalizing with the structure and the total volume available to the spot.

An application of this approach is the binding of a locus to nuclear pores [Schober et al., 2009]. Binding of chromatin to pores has been shown to be involved in important nuclear processes such as transcriptional activation (reviewed in [Akhtar and Gasser, 2007]) and processing of irreparable double-strand breaks [Nagai et al., 2008]. The diffraction limited resolution of a light microscope is not sufficient to distinguish the binding of a locus to nuclear pores from its binding to other components at the nuclear envelope (NE). A genetic trick to circumvent this problem is to examine a yeast strain with an N-terminal deletion of the nuclear pore component *NUP133*. In this mutant the pores are not distributed all over the NE but clustered on one side of the nucleus (Figure 2.10). A high degree of colocalization of a locus with this pore cluster can be an indication for binding to nuclear pores.
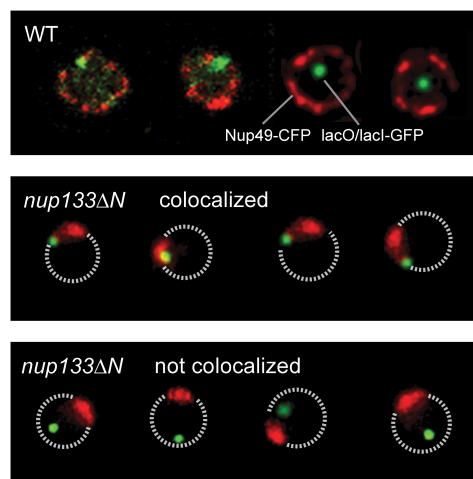
Figure 2.10: Nuclear pores (red) and tagged locus (green). The two left images in the upper panel are not deconvolved, the other images are deconvolved. In the *nup133ΔN* mutant the nuclear pores form a cluster [Schober et al., 2009].

2.3.2.1   Calculation of the spot pore cluster colocalization

In our model, the pore cluster is represented by a conical disc at the periphery of the nucleus. The spot is considered to colocalize with the cluster if it at least touches it (Figure 2.11). This definition results in a certain region for the center of the spot which represents colocalization. The expected degree of colocalization is the ratio between this colocalization volume and the total volume that is available to the spot. This ratio can be calculated analytically.
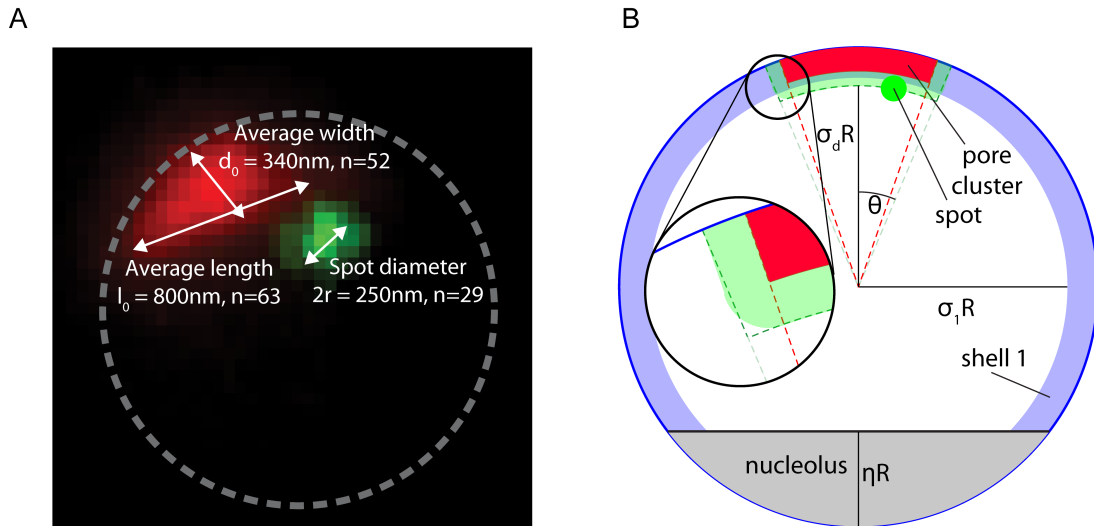
Figure 2.11: Analytical model for spot cluster colocalization [Schober et al., 2009]. (A) The pore cluster is modeled as a conical disk at the periphery of the nucleus. The thickness d_0, the diameter l_0, and the radius of the spot r were measured in 3D reconstructions of microscopic images. A possible outline of the nucleus is shown for illustration. (B) The expected colocalization for a random spot was calculated analytically as a fraction of volumes (see text). The figure shows a cut through the model. The pore cluster is a conical layer shown in red. The spot is considered as colocalizing if it at least touches the pore cluster, which results in the colocalization region shown in green. For the calculation this region is approximated again by a conical layer outlined in dark green (see also the blow up in the left part of the figure). In the parameter range used here this causes an error in the volume calculation of less than 5%. The nucleolus is shown in grey. It is assumed not to overlap with the pore cluster, and under this condition, its exact position is irrelevant to the calculation. Shell 1 (see text) is shown in light blue.

First I calculate the expected colocalization for a spot whose position is uniformly distributed throughout the nucleus. The key parameters are the diameter and thickness of the cluster and the radius of the spot. A volume fraction which is inaccessible for the spot (e.g. the nucleolus) can easily be included. A non-pore-associated enrichment in

the outermost shell of the nucleus is relevant for important chromosomal interactions and therefore calculated subsequently.

I use the following identifiers:

- R: radius of the nucleus

- $V_{\text{tot}} = \frac{4}{3}\pi R^3$: total volume of the nucleus

- C: volume of the colocalization region

- $\gamma = \frac{C}{V_{\text{tot}}}$: dimensionless variant

- V: accessible volume

- $\nu = \frac{V}{V_{\text{tot}}}$: dimensionless variant

- d: thickness of the colocalization volume (not the cluster)

- $\sigma_d = \frac{R-d}{R}$: see Figure 2.11

- Θ: measure for the radius of the colocalization volume (not the cluster), see Figure 2.11

With this, the expected colocalization ζ can be written as:

$$\zeta = \frac{C}{V} = \frac{\gamma}{\nu} \tag{2.49}$$

Using spherical coordinates[9], the colocalization volume can be calculated as the following integral:

$$C = \int_{r_1}^{r_2} \int_0^{\pi} \int_0^{\Theta} r^2 \sin\theta \, dr d\theta d\phi \tag{2.50}$$
$$= 2\pi \cdot \tfrac{1}{3} \left(r_2^3 - r_1^3\right) (1 - \cos\Theta)$$
$$= 2\pi \cdot \tfrac{1}{3} \left(R^3 - (\sigma_d R)^3\right) (1 - \cos\Theta)$$
$$C = \tfrac{2}{3}\pi R^3 \left(1 - \sigma_d^3\right) (1 - \cos\Theta) \tag{2.51}$$
$$\gamma = \tfrac{1}{2} \left(1 - \sigma_d^3\right) (1 - \cos\Theta). \tag{2.52}$$

For $\nu = 1$, ζ is equal to γ:

$$\zeta = \tfrac{1}{2} \left(1 - \sigma_d^3\right) (1 - \cos\Theta). \tag{2.53}$$

---

[9]`http://mathworld.wolfram.com/SphericalCoordinates.html` (different usage of θ and ϕ)

The dependency on thickness and diameter of the cluster is discussed in the following section (see Figure 2.12). The prediction is in general much lower than the experimental values we measured.

This big deviation from experimental colocalization values made me investigate possible influences that may increase the expected colocalization. An obvious possibility is a reduction of the available volume ν. The illustration in Figure 2.11 shows the nucleolus, but for a uniform distribution of spots the source of the reduction does not matter. As long as the occupying structure does not overlap with the colocalization volume it just reduces ν in equation (2.49). If a fraction κ of the nucleus is occupied and unavailable for the locus of interest, equation (2.53) turns into:

$$\zeta = \tfrac{1}{2\nu}\left(1 - \sigma_d^3\right)(1 - \cos\Theta)$$
$$\zeta = \tfrac{1}{2(1-\kappa)}\left(1 - \sigma_d^3\right)(1 - \cos\Theta). \tag{2.54}$$

In case of a nucleolus shaped as in Figure 2.11, κ can be written as[10]:

$$V_{\text{nucleolus}} = \tfrac{1}{3}\pi(\eta R)^2(3R - \eta R)$$
$$= \tfrac{4}{3}\pi R^3 \cdot \tfrac{1}{4}\eta^2(3 - \eta)$$
$$\kappa = \tfrac{1}{4}\eta^2(3 - \eta). \tag{2.55}$$

The dependency on κ is shown in Figure 2.13 in the next section. As expected, the random colocalization increases with growing κ. However, the lowest experimental value is only reached for very high κ values of almost 70%.

Furthermore, the expected colocalization is affected by the position of the spot relative to the nuclear periphery. I calculated this effect by dividing the nucleus into the outermost shell (shell 1) and the interior (shells 2 and 3, see section 2.3.1). If out of N spots ϵN are in shell 1, then the expected colocalization ζ can be calculated as

$$\zeta = \epsilon\zeta_1 + (1 - \epsilon)\zeta_{23} \tag{2.56}$$

with

$$\zeta_1 = \frac{C_1}{V_1} = \frac{\gamma_1}{\nu_1} \tag{2.57}$$

$$\zeta_{23} = \frac{C_{23}}{V_{23}} = \frac{\gamma_{23}}{\nu_{23}}. \tag{2.58}$$

Again C stands for the colocalization volume and V for the volume available to the spot (nucleolus already subtracted). γ and ν denote their dimensionless counterparts, as above. If ϵN spots are in shell 1 then $\epsilon N \frac{C_1}{V_1}$ spots will be in shell 1 **and** colocalizing with the pore

---

[10]See e.g. `http://mathworld.wolfram.com/SphericalCap.html`.

cluster. Accordingly, $(1 - \epsilon)N\frac{C_{23}}{V_{23}}$ spots will reside in shells 2/3 and colocalize. To be precise, C_1 is the intersection of shell 1 with the colocalization region, C_{23}, V_1, and V_{23} accordingly.

It is noteworthy that C_{23} can be empty depending on the set of parameters. Similarly, the nucleolus can lie entirely in shell 1 or cut into shells 2/3. This leads to case differentiations during the derivations. I will present only the case of both colocalization region and nucleolus extending into shells 2/3 because this is the parameter range we wanted to investigate.

The colocalization volumes C_1 and C_{23} can be calculated using equation (2.50):

$$C_1 = \tfrac{2}{3}\pi R^3 \left(1 - \sigma_1^3\right)\left(1 - \cos\Theta\right) \tag{2.59}$$

$$\gamma_1 = \tfrac{1}{2}\left(1 - \sigma_1^3\right)\left(1 - \cos\Theta\right) \overset{\sigma_1^3 = \frac{2}{3}}{=} \tfrac{1}{6}(1 - \cos\Theta) \tag{2.60}$$

$$C_{23} = \tfrac{2}{3}\pi R^3 \left(\sigma_1^3 - \sigma_d^3\right)\left(1 - \cos\Theta\right) \tag{2.61}$$

$$\gamma_{23} = \tfrac{1}{2}\left(\sigma_1^3 - \sigma_d^3\right)\left(1 - \cos\Theta\right) \overset{\sigma_1^3 = \frac{2}{3}}{=} \tfrac{1}{2}\left(\tfrac{2}{3} - \sigma_d^3\right)\left(1 - \cos\Theta\right). \tag{2.62}$$

Here, $\sigma_1 = \sqrt[3]{\frac{2}{3}}$ is the radius (normalized to R) that separates shells 1 and 2 (see also section 2.3.1).

An important intermediate result for the calculation of V_1 and V_{23} is the fraction of the nucleolus that lies in shells 2/3. Evidently, the next part makes use of the assumption that the nucleolus indeed extends into shells 2/3, i.e. $\eta > 1 - \sigma_1$:

$$V_{\text{cap23}} = \tfrac{1}{3}\pi h_{\text{cap23}}(3R\sigma_1 - h_{\text{cap23}})$$
$$h_{\text{cap23}} = \eta R - (R - \sigma_1 R)$$
$$= (\eta + \sigma_1 - 1)R$$
$$V_{\text{cap23}} = \tfrac{1}{4}V_{\text{tot}}(\eta + \sigma_1 - 1)^2(2\sigma_1 - \eta + 1) \tag{2.63}$$
$$\nu_{\text{cap23}} = \tfrac{1}{4}(\eta + \sigma_1 - 1)^2(2\sigma_1 - \eta + 1). \tag{2.64}$$

One way to derive a formula for V_1 is to subtract the outer part of the nucleolus from the whole shell volume:

$$V_1 = (1 - \sigma_1^3)V_{\text{tot}} - (\kappa V_{\text{tot}} - V_{\text{cap23}}) \tag{2.65}$$
$$\nu_1 = 1 - \sigma_1^3 - (\kappa - V_{\text{cap23}})$$
$$= 1 - \sigma_1^3 - \kappa + \tfrac{1}{4}(\eta + \sigma_1 - 1)^2(2\sigma_1 - \eta + 1)$$
$$\nu_1 = \tfrac{1}{3} - \kappa + \tfrac{1}{4}(\eta + \sigma_1 - 1)^2(2\sigma_1 - \eta + 1). \tag{2.66}$$

V_{23} can also be expressed using V_{cap23}:

$$V_{23} = \tfrac{4}{3}\pi(\sigma_1 R)^3 - V_{\text{cap23}} \tag{2.67}$$
$$\nu_{23} = \sigma_1^3 - \tfrac{1}{4}(\eta + \sigma_1 - 1)^2(2\sigma_1 - \eta + 1)$$
$$\nu_{23} = \tfrac{2}{3} - \tfrac{1}{4}(\eta + \sigma_1 - 1)^2(2\sigma_1 - \eta + 1). \tag{2.68}$$

With these results ζ can be written as:

$$\zeta = \epsilon \frac{\frac{1}{2}\left(1 - \sigma_1^3\right)\left(1 - \cos\Theta\right)}{1 - \sigma_1^3 - \kappa + \frac{1}{4}(\eta + \sigma_1 - 1)^2(2\sigma_1 - \eta + 1)}$$
$$+ (1 - \epsilon)\frac{\frac{1}{2}\left(\sigma_1^3 - \sigma_d^3\right)\left(1 - \cos\Theta\right)}{\sigma_1^3 - \frac{1}{4}(\eta + \sigma_1 - 1)^2(2\sigma_1 - \eta + 1)} \tag{2.69}$$

and with $\sigma_1^3 = \frac{2}{3}$

$$\zeta = \epsilon \frac{\frac{1}{6}(1 - \cos\Theta)}{\frac{1}{3} - \kappa + \frac{1}{4}(\eta + \sigma_1 - 1)^2(2\sigma_1 - \eta + 1)}$$
$$+ (1 - \epsilon)\frac{\frac{1}{2}\left(\frac{2}{3} - \sigma_d^3\right)\left(1 - \cos\Theta\right)}{\frac{2}{3} - \frac{1}{4}(\eta + \sigma_1 - 1)^2(2\sigma_1 - \eta + 1)}. \tag{2.70}$$

2.3.2.2   Results and discussion

HEIKO SCHOBER measured the following dimensions of the pore cluster and the spot:

- $d_0 = 340\,\text{nm}$: thickness of the cluster

- $l_0 = 800\,\text{nm}$: extension of the cluster

- $r = 125\,\text{nm}$: radius of the spot.

With these parameters, I calculated the following quantities, using a nuclear radius $R = 1000\,\text{nm}$ and a nucleolar volume fraction of $\kappa = 0.3$:

- $d = d_0 + r = 465\,\text{nm}$: thickness of the colocalization volume

- $\Theta_0 = \sin^{-1}(\frac{l_0}{2R}) = 0.4115$: angle measuring the extension of the pore cluster

- $\Theta = \Theta_0 + \sin^{-1}(\frac{r}{2R}) = 0.5366$: angle representing the extension of the colocalization volume, see Figure 2.11B

- $\eta = 0.7265$: height of the nucleolus, derived from κ using equation (2.55)

- $\rho_d = \frac{R-d}{R} = 0.535$: see Figure 2.11B.

The dependency of the expected colocalization on the thickness and the extension of the pore cluster is shown in Figure 2.12. Because a spot touching the cluster is considered colocalizing, the colocalization volume is finite even for a very thin cluster. When the cluster thickness approaches the nuclear radius, the colocalization volume increases very slowly leading to a very flat curve. The overall dependency on the cluster thickness is smaller than
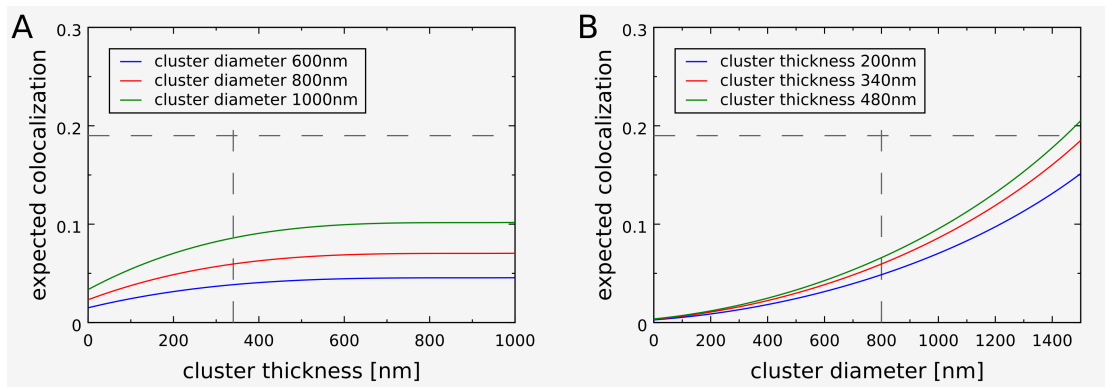
Figure 2.12: (A) Expected colocalization versus cluster thickness for three different cluster diameters. The dashed lines show the measured cluster thickness and the smallest measured colocalization. (B) Expected colocalization versus cluster diameter for three different cluster thicknesses. The vertical dashed line shows the measured cluster diameter.

one might expect from looking at Figure 2.11B. The likely reason is that the figure shows a cut through the nucleus. In 3D, the volume distribution is even more slanted towards the periphery of the nucleus.

The most striking observation is that the expected colocalization within a realistic parameter range is at least two fold lower than all experimental colocalization values that we measured, including control loci (e.g. *ARS607* bound by lexA) that are not expected to colocalize with pores [Schober et al., 2009]. The smallest of these values is shown in Figure 2.12 for comparison.

I investigated two possible explanations for this observation. Obviously, a reduction of the space apart from the pore cluster which is accessible for the locus increases the chance of random colocalization with the cluster. Such a reduction is caused by other objects or structures in the nucleus that occupy space, e.g. the nucleolus. The result is shown in Figure 2.13A. Although the effect is as expected, it is not able to explain the observed discrepancy completely.

Another factor that can increase the random colocalization of a spot with the pore cluster is a non-pore-associated enrichment of the spot at the nuclear periphery. If the spot is more often located close to the nuclear envelope it is more likely to colocalize with the cluster even if it does not bind to pores. However, the effect is rather minor (see Figure 2.13B). An increase of the fraction of spots in shell 1 from 33% to 60% raises the random colocalization from 8.6% to only 9.5%.

One possible conclusion from these data is, of course, that the observed loci in fact do have an association with pores. This association does not have to be specific but can be due to unspecific binding of DNA to nuclear pores. This would be consistent with the high
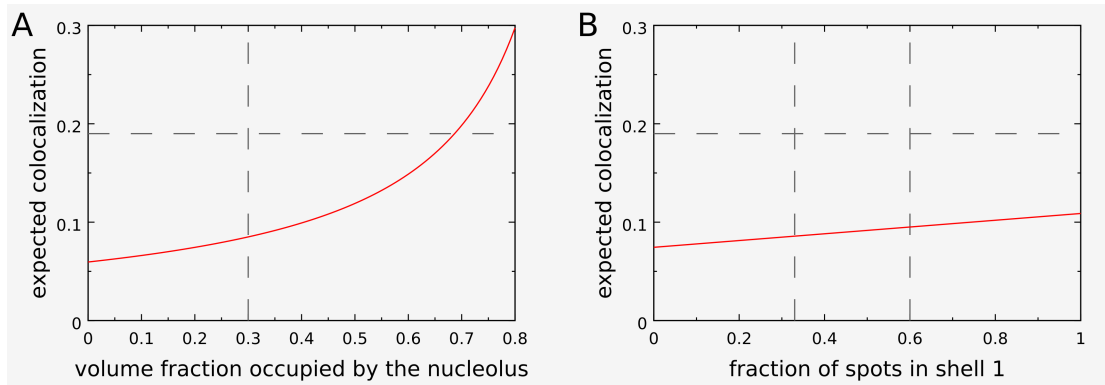
Figure 2.13: (A) Expected colocalization versus the volume fraction occupied by the nucleolus or another "excluding" structure. The expected colocalization increases when the total accessible volume decreases. This can be the case if a certain volume fraction is occupied e.g. by the nucleolus. The vertical dashed line indicates a volume fraction of 30%. Cluster thickness and diameter have their default values of 340 nm and 800 nm, respectively. (B) Expected colocalization versus enrichment in shell 1. The vertical dashed lines indicate the uniform distribution (fraction $\frac{1}{3}$) and a typical enriched value of 0.6. Cluster thickness and diameter have their default values of 340 nm and 800 nm. The nucleolus in this case occupies 30% of the nucleus.

frequency of pore-contact detected by [Schmid et al., 2006]. Alternatively, there could be an association of commonly used protein binding sequences like lexA with pores. Another possibility is that the volume inaccessible to the studied loci is bigger than assumed above. For very high occupation values the observed colocalization values can be reached (see Figure 2.13A). As mentioned earlier, nuclear structures other than the nucleolus can contribute to this occupied volume: chromosomes, proteins etc.. However, they can also block access to the pore cluster itself, thereby reducing the expected colocalization. It is hard to estimate which effect dominates over the other. This will depend on the preferred location of the chromosomes and of the pore cluster.

The pore cluster in a *nup133ΔN* mutant does not assume a strict location with respect to the bud neck or the spindle pole body [Doye et al., 1994]. If it displays a specific location with some other structure, then a third explanation for high colocalization is possible. The calculations that I have presented here are based on the assumption that the locus of interest explores the whole accessible volume. This assumption might not be fulfilled. If, in contrast, both the pore cluster and the locus reside in specific regions of the nucleus and if these regions overlap, this can lead to higher colocalization than expected for a cluster and a spot with random locations. A direct interaction is not required.

2.3.3   Mobility of a tagged locus

A stretch of chromatin (or any other object) inside the nucleus is exposed to numerous hits of water or other small molecules, proteins and other macromolecules, as well as other chromatin fibers. Due to these interactions, it inevitably performs a seemingly random movement called BROWNian motion. This motion is limited by the nuclear envelope, but in many cases locus diffusion is even more constrained, either confined to a certain area or obstructed by obstacles. The random movement can also be temporarily or continuously superimposed by active displacement which possibly expresses itself as increased speed and/or directionality of movement.

2.3.3.1   Tracking of a fluorescent spot

The first step of the quantitative analysis of chromatin movement is the determination of the position of the locus and the nuclear center for each time point of the movie. Indeed, since the nucleus itself is moving inside the cytoplasm, one must compensate for its displacement to measure the movement of a locus relative to the nucleus. Several general purpose software packages like Imaris[11] offer object tracking functionality but usually require uniformly high contrast images. The algorithms are mostly based on threshold principles, and it is difficult to correct insufficient results by hand. In collaboration with DANIEL SAGE and MICHAEL UNSER, a dynamic programming algorithm was developed which is dedicated to the tracking of single spots in noisy images and can be applied to 2D or 3D time-lapse movies [Sage et al., 2005]. The algorithm is implemented as a publicly available plug-in for the free software ImageJ[12]. The tracking works in two steps: First the images are aligned with respect to the center of the nucleus in order to compensate for the movement of the entire nucleus throughout the movie.

1. The fluorescence of the spot is more intense than the background fluorescence surrounding it.

2. Within one time step, the spot can only travel a limited distance.

3. In contrast to nuclear pores, the spot can be located in the nuclear interior.

To reflect these properties the tracking algorithm uses four different criteria to determine the spot position at a given time point:

1. the pixel intensity

2. the displacement from the location at the previous time point

3. the displacement from the last user-defined position (see below)

---

[11]http://www.bitplane.com/go/products/imaris
[12]http://bigwww.epfl.ch/sage/soft/spottracker/

4. the distance from the nuclear center.

The user can give different weights to these criteria in order to optimize the performance of the algorithm for different situations or image qualities. Most importantly, the plug-in offers the possibility to correct the trajectory manually by forcing it to pass through a given pixel at a certain time point. The output of the plug-in is the position of spot and nuclear center for each time point.

2.3.3.2 Basic quantification parameters for locus mobility

Track length A very simple and robust parameter is the track length over the whole movie. Although it is a measure for the average mobility of a locus and can be used for comparison, it has to be kept in mind that it is a very artificial parameter because the true trajectory of the spot is inaccessible due to the lack of temporal and spatial resolution and is much longer than the measured track length. The measured length is indeed highly dependent on the time step of the movie, and only movies with the same time step can be compared to each other (see also section 2.2.2).

Step size and large steps Another readout is the average step size of the movie. As the track length, it depends on the time step used during image acquisition. Moreover, directed movement does not necessarily reveal itself in large single steps but rather in several successive correlated steps. Therefore it is useful to look for exceptionally high displacements ("large steps") within a certain time window. It turned out empirically that a useful parameter for distinguishing patterns of mobility is scoring steps larger than 500 nm during 10.5 s [Heun et al., 2001, Gartenberg et al., 2004].

2.3.3.3 Mean squared displacement (MSD) analysis

A robust method to analyze the global properties of an object's movement is the mean squared displacement (MSD) analysis. An object in solution changes its direction when it bumps into solvent molecules and moves linearly in between: It performs a random walk. If a number of objects would be initially confined in a small volume and then released they would spread over time. It can be derived mathematically that for free diffusion the mean of the squared distance from one point on the trajectory to another (see Figure 2.14) is proportional to the time difference t (see section 2.2):

$$\langle \vec{r}^2(t) \rangle = 6Dt \ . \tag{2.13}$$

The resolution of fluorescence microscopy is far too low both in time and space to resolve each step of the object's trajectory. However, an important property of the MSD is that it can also be calculated from a coarsened trajectory (see section 2.2.2).
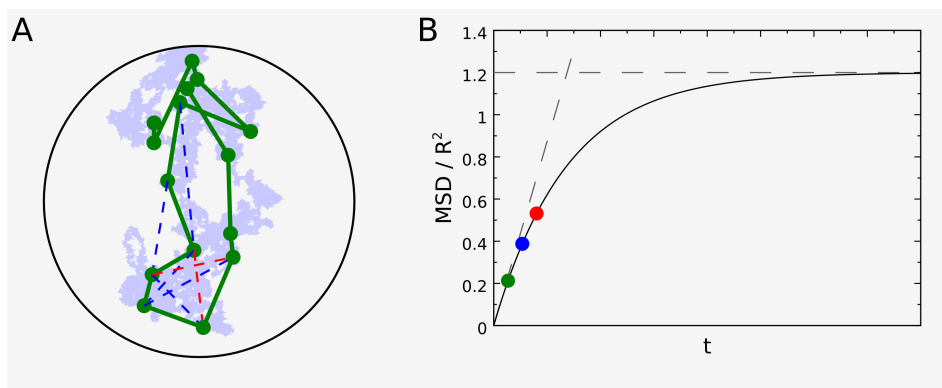
Figure 2.14: Mean squared displacement (MSD) analysis. (A) The full trajectory of
microscopic movement (light blue) cannot be detected by fluorescence microscopy due
to limited resolution in time and space. A coarsened trajectory (green) is recorded
instead. (B) The mean of all squared spatial distances between each two points at
a given time difference results in one point on the MSD graph. The mean squared
distance between a point and its successor on the trajectory is the first point on the
MSD graph (green). The mean squared distance between a point and its second
successor yields the second point (blue) and so on. A few examples of the averaged
distances are also shown in (A) in the respective color.

In a cellular environment there is no free diffusion. The free movement of an object
can be impaired by confinement, obstacles, and the binding to immobile or actively moved
structures. The most inevitable restriction is the confinement of the object's movement to a
nuclear or cellular compartment or at least the cell. This implies that the distance of any
two points of the trajectory cannot exceed the maximal extension of the confining volume.
Therefore the MSD curve has to reach a plateau for large time windows (Figure 2.14B). In
the case of a spherical confinement, the value of the plateau can be calculated as $\frac{6}{5}R^2$ where
R is the radius of the sphere (see section 2.2.3). Thus, the so-called radius of constraint or
the plateau value directly can be used as a measure for the region explored by the object.

For free diffusion, the slope of the MSD line is a measure for the diffusion coefficient of
the object, as mentioned above. In the case of confined diffusion, the slope of the MSD curve
is not constant. The curve is steepest at t = 0, and then the slope decreases monotonously
(Figure 2.14B). However, one can still use the initial slope of the curve to compare the
intrinsic mobility of different objects or one object under different conditions. See section
2.2.3 for a detailed analysis.

Due to the difficulties in reconstructing the 3D position of fluorescent spots (see section
2.3.1.1), the movement is often observed in a 2D projection of the microscopic stacks.
Therefore, I discuss the effects of projection on the MSD curve in the following section.

2.3.3.4 MSD analysis of projected trajectories

As discussed in section 2.1, it is difficult to acquire sufficiently robust microscopic data to allow deconvolution and correction of the z-stretch fast and accurately in 3D. Therefore the 3D data sets are often projected onto the xy plane, either already during the image acquisition or later during the analysis.

If the movement is isotropic, one can hope to reconstruct the 3D properties of the movement even from the projected data set. For the derivation of equation (2.3), I have only assumed that each step is equally probable as the opposite step. This is still true for the projected steps. However, the quadratically averaged projected step length is shorter than the 3D step length l.

Due to the symmetry of the problem, the attitude of the projection plane does not matter. By convention, the optical axis of the microscope defines the z axis which means in this case that the trajectory is projected onto the xy plane. Mathematically, this means that all c_i are set to 0 in equation (2.1):

$$\vec{r}_{N,p} = \sum_{i=1}^{N} \vec{r}_{i,p} \ \text{ with } \vec{r}_{i,p} = \begin{pmatrix} a_i \\ b_i \\ 0 \end{pmatrix}. \tag{2.71}$$

The mean squared displacement of a free walk can be calculated as in section 2.2.1:

$$\begin{aligned}
\langle \vec{r}_{N,p}^2 \rangle &= \left\langle \left(\sum_{i=1}^{N} \vec{r}_{i,p} \right) \cdot \left(\sum_{j=1}^{N} \vec{r}_{j,p} \right) \right\rangle \\
&= \left\langle \sum_{i=1}^{N} (a_i^2 + b_i^2 + 0) + \sum_{i \neq j} (a_i a_j + b_i b_j + 0) \right\rangle \\
&= \sum_{i=1}^{N} \left\langle a_i^2 + b_i^2 \right\rangle + \sum_{i \neq j} \left(\langle a_i a_j \rangle + \langle b_i b_j \rangle \right) \\
&= N \left(\langle a^2 \rangle + \langle b^2 \rangle \right).
\end{aligned}$$

It follows from symmetry that $\langle a^2 \rangle = \langle b^2 \rangle = \langle c^2 \rangle = \frac{l^2}{3}$ and therefore:

$$\langle \vec{r}_{N,p}^2 \rangle = \tfrac{2}{3} N l^2 \tag{2.72}$$

$$\langle \vec{r}_{N,p}^2 \rangle = 4Dt. \tag{2.73}$$

If the movement is confined, one can only expect to reconstruct the 3D MSD if the confinement has similar extension in each coordinate direction. During image acquisition, the yeast nucleus is randomly oriented. Therefore, only a spherical confinement strictly fulfills the above requirement.

Section 2.2.3.1 discusses random walks in a spherical confinement. The MSD of such a walk can be approximated accurately by the following function:

$$\langle r(t)^2 \rangle \approx a \left(1 - e^{-mt/a} \right) \tag{2.74}$$

where a is the height of the MSD plateau and m is the slope of the MSD curve at $t = 0$. Furthermore, m is very close to $6Dt$, the MSD slope of the corresponding unconfined walk, leading to:

$$\langle r^2(t) \rangle \approx \tfrac{6}{5} R^2 \left(1 - e^{-5Dt/R^2} \right). \tag{2.25}$$

If the trajectory is projected onto two dimensions, the plateau value is $\tfrac{4}{5} R^2$ instead of $\tfrac{6}{5} R^2$, and the slope of the MSD for free diffusion is $4Dt$ instead of $6Dt$, as shown above. This results in:

$$\langle r_p^2(t) \rangle \approx \tfrac{4}{5} R^2 \left(1 - e^{-5Dt/R^2} \right). \tag{2.75}$$

Notably, the decay constant of the exponential is the same as in 3D, and it follows:

$$\langle r_p^2(t) \rangle = \tfrac{2}{3} \langle r^2(t) \rangle. \tag{2.76}$$

Under the assumption of (2.74), the 3D MSD curve can be fully recovered from the MSD calculated from the projected trajectory. I confirmed this result with a simulation (Figure 2.15). It has to be emphasized, however, that this derivation requires that the movement of the labeled object is isotropic and the confinement is spherical.

2.3.3.5   The mean squared change of spot-spot distance

The movement of a locus relative to the nucleus is superimposed by the movement of the nucleus itself. Translational movement of the nucleus can be subtracted by aligning the nuclear center throughout the time course (see section 2.3.3.1). If two spots are observed, there is the alternative possibility to align one of the spots throughout the movie and analyze the movement of the other spot relative to the first one. This procedure also eliminates the global movement of the nucleus. If $\vec{u}$ and $\vec{v}$ denote the positions of the two spots, the mean squared change of the connecting vector $\vec{w} = \vec{v} - \vec{u}$ can be calculated analogously to the derivation of the classical MSD given in section 2.2.1:

$$\langle \vec{w}_N^2 \rangle = \left\langle \left(\sum_{i=1}^{N} (\vec{v}_i - \vec{u}_i) \right) \cdot \left(\sum_{j=1}^{N} (\vec{v}_j - \vec{u}_j) \right) \right\rangle$$

$$= \left\langle \sum_{i=1}^{N} (\vec{v}_i^2) \right\rangle + \left\langle \sum_{i=1}^{N} (\vec{u}_i^2) \right\rangle.$$
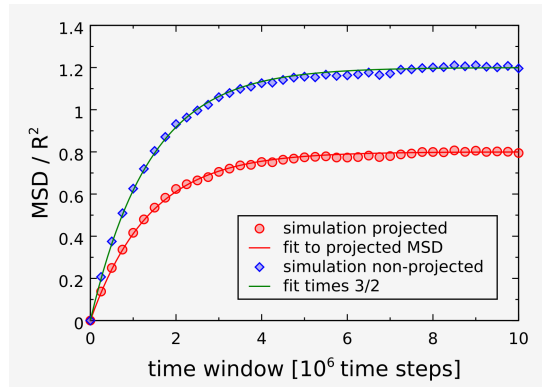
Figure 2.15: MSD of a random walk inside a ball, calculated from the full trajectory
and from the trajectory projected onto the xy plane. A fit of equation (2.74) to
the projected trajectory, multiplied by $\frac{3}{2}$, reproduces the true MSD very well. The
curves were calculated from 100000 simulations with step length $1\,\text{nm}$ and a radius of
confinement of $1000\,\text{nm}$.

The last step makes use of the assumption that the movement of one spot is independent
of the movement of the other spot and that a step of one spot is independent of the steps
before. The result is that the MSD calculated from the connection vector of the two spots
is the sum of the individual MSDs. If the assumption is justified that the two spots move
similarly (e.g. in the case of two alleles of the same gene in a diploid cell), then the MSD of
one locus is just the half of the measured MSD.

However, neither the alignment of the nuclear center nor the alignment of one spot
eliminates the rotational movement of the nucleus. A possibility to obtain a quantification of
locus mobility that is independent of nuclear rotation is to observe the distance between the
two loci and calculate the mean squared change of this distance (see Figure 2.16). Since the
distance between the two spots is unaffected by both translation and rotation of the nucleus,
this "distance MSD" is only influenced by the individual movement of the two spots.

The distance MSD curve shows similar behavior to a classical MSD curve (Figure 2.17A)
and has been used to derive diffusion coefficients and radii of constraints [Marshall et al.,
1997]. However, it is important to note that the authors assumed that both loci are confined
to the same region. If this is not the case, the height of the plateau as well as the initial
slope of the curve do not only depend on the mobility of the loci but also on the distance of
the regions of constraint for the two spots (Figure 2.17C+D). I calculated the dependency
of the plateau value on this distance analytically in the case of non-overlapping regions of
constraints.
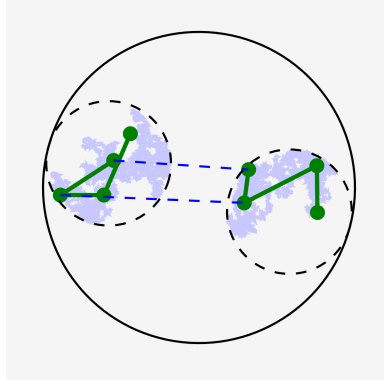
Figure 2.16: Distance MSD analysis. In contrast to a classical MSD analysis, the mean squared change of the distance between two spots (blue) instead of the mean squared change of the position of one spot is analyzed.

2.3.3.6   Calculation of the distance MSD plateau

As mentioned before, two positions on the trajectory of a spot lose their correlation when the time window between them gets sufficiently large. Equally, the distance between two spots becomes uncorrelated, and the distance MSD plateau can in principle be calculated as the squared difference between the distances of two pairs of points, which are all randomly located in their respective region of constraint:

$$
\begin{aligned}
\langle s^2 \rangle &= \frac{1}{V_1^2 V_2^2} \int_{V_1} \int_{V_2} \int_{V_1} \int_{V_2} \left[|\vec{r}_2 - \vec{r}_1| - |\vec{r}_4 - \vec{r}_3| \right]^2 \, dV^4 \\
&= \frac{1}{V_1^2 V_2^2} \int_{V_1} \int_{V_2} \int_{V_1} \int_{V_2} \left[\sqrt{(\vec{r}_2 - \vec{r}_1)^2} - \sqrt{(\vec{r}_4 - \vec{r}_3)^2} \right]^2 \, dV^4 .
\end{aligned}
\tag{2.77}
$$

Here, spots 1 and 3 are confined to volume V_1, and spots 2 and 4 are confined to volume V_2. Unfortunately, this integral is very difficult to solve. However, I solved the underlying problem in two special cases in which V_1 and V_2 are both balls and are either identical or do not overlap.

Confinement to the same ball   If V_1 and V_2 are identical (i.e. both spots are confined to the same region), the distance distribution (3.6) can be used to calculate the plateau value. The plateau is the product of the squared difference of distances $(\sigma_2 - \sigma_1)^2$ with the
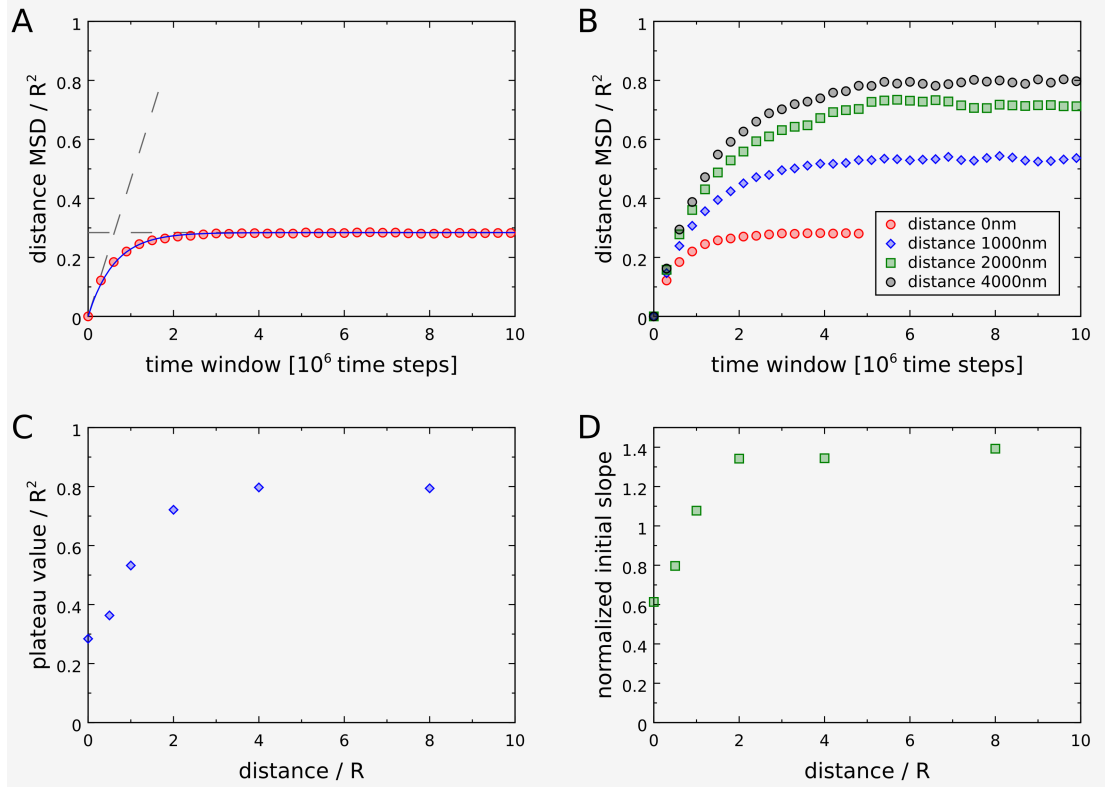
Figure 2.17: (A) Similarly to a classical MSD, the distance MSD curve is well described by equation (2.16) (blue). The MSD is normalized to the square of the radius of confinement R. The data set shown in red was derived from 1000000 simulations of two random walks confined to the same ball with radius 1000 nm. The step length was 1 nm. (B) The plateau value of the distance MSD depends on the distance between the confinement regions for the two walks. In all cases, the two walkers were each confined to a ball with a radius of 1000 nm. The distance between the centers of the balls is indicated in the figure. Each curve was calculated from 100000 simulations. The step size was always 1 nm. (C) Plateau values depending on the distance of the confining regions. The values were derived by fitting equation (2.16) to the data sets shown in (B). (D) Initial slopes depending on the distance of the confining regions. The values were derived by fitting equation (2.16) to the data sets shown in (B). They are therefore normalized to R^2 and 10^6 time steps.

probabilities for the distances σ_1 and σ_2, integrated over all possible distances:

$$\langle q^2 \rangle = R^2 \int_0^2 \int_0^2 (\sigma_2 - \sigma_1)^2 \, \mathrm{P}(\sigma_1) \, \mathrm{P}(\sigma_2) \, d\sigma_1 d\sigma_2 \tag{2.78}$$

$$= R^2 \int_0^2 \int_0^2 (\sigma_2 - \sigma_1)^2 \, (3\sigma_1^2 - \tfrac{9}{4}\sigma_1^3 + \tfrac{3}{16}\sigma_1^5)(3\sigma_2^2 - \tfrac{9}{4}\sigma_2^3 + \tfrac{3}{16}\sigma_2^5) \, d\sigma_1 d\sigma_2$$

$$= R^2 \int_0^2 \tfrac{3}{560}\sigma_1^2(\sigma_1 - 2)^2(\sigma_1 + 4)(35\sigma_1^2 - 72\sigma_1 + 42) \, d\sigma_1$$

$$\langle q^2 \rangle = \tfrac{348}{1225} R^2. \tag{2.79}$$

Confinement to non-overlapping balls I exploit the symmetry of the problem to calculate the distance distribution in the case of non-overlapping balls. The larger ball is centered at the origin, the smaller one at $(d, 0, 0)$ (see Figure 2.18). I use the following identifiers:

- R: radius of the larger ball (ball 1)

- ρR: radius of the smaller ball (ball 2)

- d: center-to-center distance of the balls

- $C_2 = (d, 0, 0)$: center of ball 2

- $\delta = \frac{d}{R}$

- s: distance between the two spots

- $\sigma = \frac{s}{R}$

- a_1: see Figure 2.18

- h_1: see Figure 2.18.

- a_2: see Figure 2.18

- h_2: see Figure 2.18.

- x: see Figure 2.18

- y: see Figure 2.18.

Since the spots are assumed to be uniformly distributed in their respective confinements, most probability calculations are in fact volume calculations. However, the probability that the distance between the spots is between s and $s + ds$ can be written as:

$$\mathrm{P}(s)ds = \int_{-R}^{R} \tilde{\mathrm{P}}(x, s) \, dx ds.$$

If spot 1 is at position $(x, 0, 0)$, then it could also be anywhere else on the corresponding shell cap with radius $d - x$ around C_2 without changing the probability for spot 2 being at distance s (see Figure 2.18). Therefore, $\tilde{\mathrm{P}}(x, s)$ denotes the probability that spot 1 is at distance $d - x$ from C_2 **and** that spot 2 is at distance s from spot 1. This probability can be split up further:

$$\tilde{\mathrm{P}}(x, s) = \hat{\mathrm{P}}(x) \, \hat{\tilde{\mathrm{P}}}(x, s) \tag{2.80}$$

where $\hat{\mathrm{P}}(x)$ is the probability for spot 1 being at distance $d - x$ from C_2 and $\hat{\tilde{\mathrm{P}}}(x, s)$ is the probability for spot 2 being at distance s from spot 1, **given** that spot 1 is at distance $d - x$ from C_2.
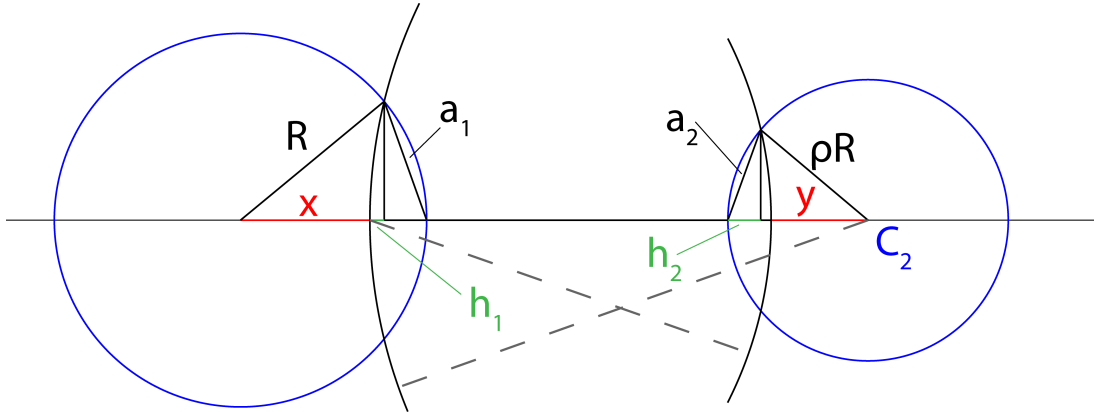
Figure 2.18: Identifiers for the distance MSD calculation.

In order to calculate $\hat{\mathrm{P}}(x)$, the height h_1 of the shell cap is required (see Figure 2.18):

$$\begin{aligned}
a_1^2 &= (d - x)^2 - (d - x - h_1^2) \\
&= 2h_1(d - x) - h_1^2 \\
a_1^2 &= (R + (x + h_1))(R - (x + h_1)) \\
&= R^2 - x^2 - 2xh_1 - h_1^2.
\end{aligned}$$

From these two relations, it follows:

$$2h_1(d-x) = R^2 - x^2 - 2xh_1$$

$$h_1 = \frac{R^2 - x^2}{2d}.$$

For the probability $\hat{P}(x)$ follows[13]:

$$\hat{P}(x)dx = 2\pi(d-x) \cdot \frac{R^2 - x^2}{2d} \cdot \frac{1}{\frac{4}{3}\pi R^3}dx$$

$$= \frac{3(d-x)(R^2 - x^2)}{4d}dx$$

$$\hat{P}(\xi)d\xi = \frac{3(\delta - \xi)(1 - \xi^2)}{4\delta}d\xi \tag{2.81}$$

where $\xi = \frac{x}{R}$.

If spot 1 is residing on the shell cap with distance $d - x$ from C_2, then the probability for spot 2 being at distance s from spot 1 is again proportional to the surface of a spherical cap (see Figure 2.18):

$$a_2^2 = s^2 - (s - h_2)^2$$

$$= 2h_2 s - h_2^2$$

$$a_2^2 = (\rho R + (y + h_2))(\rho R - (y + h_2))$$

$$= (\rho R)^2 - y^2 - 2yh_2 - h_2^2$$

$$h_2 = \frac{(\rho R)^2 - (d - x - s)^2}{2(d-x)}$$

$$\hat{\hat{P}}(x,s)ds = 2\pi s \cdot \frac{(\rho R)^2 - (d - x - s)^2}{2(d-x)} \cdot \frac{1}{\frac{4}{3}\pi(\rho R)^3}dx$$

$$\hat{\hat{P}}(\xi,\sigma)d\sigma = \frac{3\sigma(\rho^2 - (\delta - \xi - \sigma)^2)}{4\rho^3(\delta - \xi)}d\sigma \tag{2.82}$$

$$\hat{\hat{P}}(\xi,\gamma)d\gamma = \frac{3(\gamma + \delta)(\rho^2 - (\gamma + \xi)^2)}{4\rho^3(\delta - \xi)}d\gamma \tag{2.83}$$

with $\gamma = \sigma - \delta$.

With these results, the probability for spot 1 being at distance $\delta - \xi$ from C_2 **and** spot 2 being at distance $\sigma = \gamma + \delta$ turns into:

$$\tilde{P}(\xi,\gamma) = \hat{P}(\xi)\hat{\hat{P}}(\xi,\gamma)$$

$$\tilde{P}(\xi,\gamma) = \frac{9(\gamma + \delta)(1 - \xi^2)(\rho^2 - (\gamma + \xi)^2)}{16\delta\rho^3}. \tag{2.84}$$

---

[13]http://mathworld.wolfram.com/SphericalCap.html

Integration over ξ finally yields the probability for spots 1 and 2 being at distance σ:

$$
\mathrm{P}(\gamma)d\gamma = \int_{\xi_1}^{\xi_2} \tilde{\mathrm{P}}(\xi, \gamma)\, d\xi d\gamma
$$

$$
= \frac{9(\gamma + \delta)}{16\delta\rho^3} \left(\tfrac{1}{5}(\xi_2^5 - \xi_1^5) + \tfrac{1}{2}\gamma(\xi_2^4 - \xi_1^4) + \tfrac{1}{3}(\gamma^2 - \rho^2 - 1)(\xi_2^3 - \xi_1^3) \right.
$$
$$
\left. + \gamma(\xi_2^2 - \xi_1^2) + (\rho^2 - \gamma^2)(\xi_2 - \xi_1) \right). \tag{2.85}
$$

It is important to note that depending on σ, ξ_1 and ξ_2 are not necessarily -1 and 1, respectively. The reason is that $\hat{\mathrm{P}}(\xi, \sigma)$ can vanish if e.g. $x + \sigma > \delta + \rho$. This is not reflected by the given formulas; they assume $x + \sigma \in [\delta - \rho, \delta + \rho]$. Therefore, the integration intervals have to be chosen properly:

$$
\begin{array}{llll}
\xi_1 = -\gamma - \rho, & \xi_2 = 1 & \text{for} & -\rho - 1 \leq \gamma < \rho - 1 \\
\xi_1 = -\gamma - \rho, & \xi_2 = -\gamma + \rho & \text{for} & \rho - 1 \leq \gamma < -\rho + 1 \\
\xi_1 = -1, & \xi_2 = -\gamma + \rho & \text{for} & -\rho + 1 \leq \gamma < \rho + 1.
\end{array} \tag{2.86}
$$

This results in an average distance of:

$$
\langle \sigma \rangle = \int_{-\rho-1}^{\rho-1} (\gamma + \delta)P(\gamma)d\gamma + \int_{\rho-1}^{-\rho+1} (\gamma + \delta)P(\gamma)d\gamma + \int_{-\rho+1}^{\rho+1} (\gamma + \delta)P(\gamma)d\gamma
$$

$$
\langle \sigma \rangle = \frac{1 + 5\delta^2 + \rho^2}{5\delta}
$$

$$
\langle s \rangle = \frac{1 + 5\delta^2 + \rho^2}{5\delta}R. \tag{2.87}
$$

Furthermore, the plateau of the distance MSD can be calculated as:

$$
\langle q^2 \rangle = R^2 \int_{-\rho-1}^{\rho+1} \int_{-\rho-1}^{\rho+1} (\gamma_2 - \gamma_1)^2 \, \mathrm{P}(\gamma_1)\,\mathrm{P}(\gamma_2)\, d\gamma_1 d\gamma_2.
$$

Both integrals have to be split up as above, resulting in a sum of 9 integrals which add up to:

$$
\langle q^2 \rangle = \frac{2(1 + \rho^2)(5\delta^2 - \rho^2 - 1)}{25\delta^2}R^2 \tag{2.88}
$$

$$
\langle q^2 \rangle = \frac{4(5\delta^2 - 2)}{25\delta^2}R^2 \text{ for } \rho = 1. \tag{2.89}
$$

The limit for large distances of the confining regions is:

$$\langle q^2 \rangle = \frac{2(1+\rho^2)}{5} R^2 \tag{2.90}$$

$$\langle q^2 \rangle = \tfrac{4}{5} R^2 \text{ for } \rho = 1. \tag{2.91}$$

2.3.3.7 Results and discussion

Figure 2.19 shows the dependency of the distance MSD plateau on the distance d of two equally sized confining balls with radius R. The radius of constraint can be reliably derived from the plateau value if the confining regions are identical ($d = 0$) or are sufficiently far from each other ($d \gtrsim 3R$).
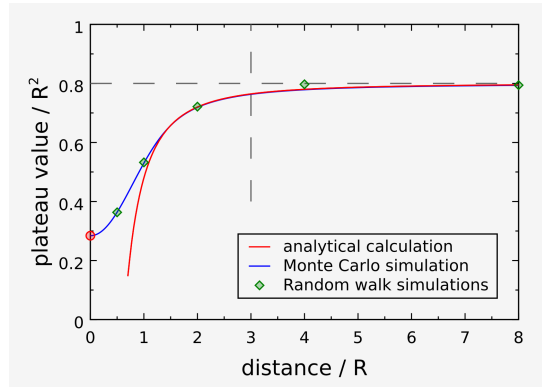
Figure 2.19: Dependency of the distance MSD plateau $\langle q^2 \rangle$ on the distance d between the two regions of confinement. The analytical result for $d = 0$ (2.79) and $d \geq 2R$ (2.88) is shown in red. The distance MSD plateau value can be calculated much faster by a Monte Carlo simulation of uniformly distributed populations of spots (blue) than by a random walk simulation (green, see also Figure 2.17). For $d \gtrsim 3R$, the approximation $\langle q^2 \rangle \approx \tfrac{4}{5} R^2$ causes only small errors (see text).

If the latter is the case, then the distance MSD plateau $\langle q^2 \rangle$ is equal to $\tfrac{4}{5} R^2$ and therefore

$$R = \sqrt{\tfrac{5}{4} \langle q^2 \rangle}. \tag{2.92}$$

However, it can be tricky to determine if $\frac{d}{R}$ is large enough because this condition is reflexive. If d gets too small, then also the plateau decreases and with it the radius of constraint calculated under the assumption of (2.92). This can then lead to the wrong conclusion that the assumption of $\frac{d}{R}$ being sufficiently large was correct. The assumption can therefore be self-fulfilling.

To investigate how severe this effect is, I plotted the measured versus the true $\frac{d}{R}$ (Figure 2.20A). It turned out that the error is small. If one calculates R under the assumption that (2.92) is correct, then one will conclude that $d > 3R$ already when the true value is only $2.93R$, which means that the error is smaller than 3%. Likewise, the error in the measured radius of constraint under the assumption of (2.92) is small if $d > 3R$ is fulfilled (Figure 2.20B). Under this premise, the error is always smaller than 2.5%.
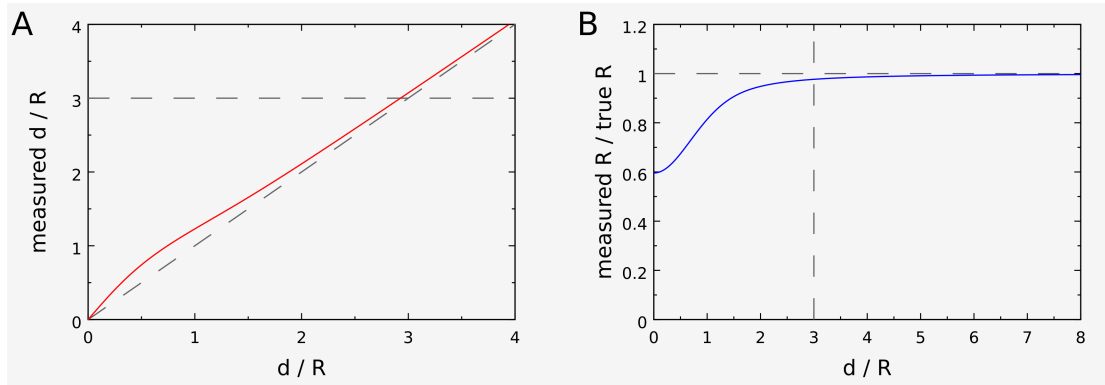
Figure 2.20: Errors in deriving the radius of constraint from the distance MSD plateau. (A) $\frac{d}{R}$, derived under the assumption (2.92), versus the exact value for $\frac{d}{R}$. The error in the assessment if $\frac{d}{R} > 3$ is in most cases neglectable compared to other errors. (B) Quotient of the radius of constraint, calculated under the assumption (2.92), and the true radius of constraint versus the exact value for $\frac{d}{R}$. For $\frac{d}{R} \gtrsim 3$ the approximation is sufficient for most practical applications.

In principle, it would be possible to use the simulation data shown in Figure 2.19 to derive the radius of constraint from the distance MSD plateau also for $0 < d < 3R$. Maybe, even an analytical calculation is possible in this range as well. However, other errors will have more severe effects in this case, especially the uncertainty in the measurement of the distance d. Since the curve is very steep for $d < 3R$, an error in d has a strong effect. Additionally, one has to keep in mind that the derivations presented above are based on the assumptions that the confining regions are balls. As for classical MSDs, the calculated radius of constraint can still be used as a general measure for the size of the confinement region if this condition is not fulfilled. However, this concept becomes questionable if the results depend very sensitively on other parameters like d in this case. Therefore, I conclude that the distance MSD analysis is a valid technique to determine a radius of constraint for two diffusing spots if one of the two following conditions is fulfilled. Either a) one can assume that the confining regions are identical (e.g. the whole nucleus) or b) they are sufficiently far from each other ($d \gtrsim 3R$).

My calculations presented in the previous section include the possibility of two confinement regions with different sizes. Obviously, this also influences the value of the distance MSD plateau. In the discussion above, I assumed that both regions are equal in size. In cases

where this assumption is not justified, the general behavior for large distances will still be the same (see equations (2.88) and (2.90)). However, it will be very difficult to dissect the contributions of the two radii of constraint and the distance to the plateau value from each other. Probably, a comparison of the same system in different situations is the best one can achieve. The derivation of absolute values is likely to be impossible.

The second parameter that can be derived from classical MSD analysis is the diffusion coefficient of the monitored object. I have not tried to calculate the diffusion coefficients of the two spots from the distance MSD curve, but it should be possible for the following reason. If the radii of constraint and the step sizes of the random walkers are all multiplied by a factor $c > 0$ then the curve in Figure 2.17A stays exactly the same. Since the MSD is normalized to R^2, the slope of the curve in nm/time step increases by a factor of c^2. It is plausible that the initial slope of the distance MSD curve is independent of R because — as for the classical MSD — the walkers do not feel the confinement at very small time intervals. Therefore, the factor c^2 is caused by the increase in step size, and thus the initial slope of the MSD curve shows the same step size dependency as the diffusion coefficient (equation (2.12)).

This consideration suggests that a common diffusion coefficient of the two spots is proportional to the initial slope of the distance MSD curve. Furthermore, the slope also seems to lose its dependency on d for $d \gtrsim 3R$ (Figure 2.17D). If the diffusion coefficients are different, one can expect that the initial slope reflects a value between the two.

2.4 Polymer modeling

In many situations, a random walk model can be used as a first approximation for the movement of a cellular object. However, this model cannot be very accurate in the case of a chromosomal locus because the fiber restricts the movement of the locus. The entanglement of a chromosome with itself or other chromosomes influences the mobility of the locus, too. Especially if one is interested in non-equilibrium dynamics of a chromosomal site, it is inevitable to use a more accurate model.

Depending on the size of the simulated system and on the parameters of interest, a chromosome can be modeled at different levels of abstraction. In my simulations, I considered the chromatin fiber as a thread. All details below chromatin were simplified to homogeneous mechanical properties of the fiber. In the following sections, I introduce how the behavior of such a thread can be modeled in a computer simulation.

2.4.1 The ideal polymer chain

The ideal chain model describes a chain of immaterial segments which — apart from being connected with their immediate neighbors — do not interact with each other. In particular, there is no excluded volume interaction, i.e. the segments do not occupy any space. Biological

macromolecules only rarely behave like ideal chains, but the model is nonetheless useful to study the statistical properties of their thermal motion. Different realizations of the ideal chain model have the properties that are mentioned above in common, but they differ in how the individual segments are connected to each other.

2.4.1.1   The freely jointed chain

One of the simplest representations of the ideal chain model is the freely jointed chain. It is composed of N stiff segments, which can rotate freely around each other. This implies that no energy is stored in elastic stretching, bending, or twisting of the chain, and thus all conformations have the same energy. Therefore, in thermal equilibrium, the chain will be driven towards the conformation with the highest entropy.

The mean squared end-to-end distance of a chain can be used for a simple characterization of the extension of the chain:

$$\langle r^2 \rangle = \left\langle \left(\sum_{i=1}^{N} \vec{u}_i \right)^2 \right\rangle \tag{2.93}$$

where $\vec{u}_i$ is the vector along the ith segment of the chain. Equation (2.93) is exactly identical to equation (2.2), which describes the mean squared displacement of a random walk. In fact, from the mathematical point of view, a free random walk and a freely jointed chain are exactly the same thing.

Therefore, the same derivation as in section 2.2.1 can be used to derive the following identity:

$$\langle r^2 \rangle = N s^2 = L_C s \tag{2.94}$$

where s is the length of one segment and L_C ($= Ns$) is the contour length of the chain. $\sqrt{\langle r^2 \rangle}$ as a measure of the extension of a long freely jointed chain is therefore proportional to the square root of the number of segments.

2.4.1.2   The persistent or worm-like chain

The freely jointed chain can be treated well theoretically, but it is not a very realistic model for a macromolecule, because chemical bonds usually give the molecule a general stiffness. A more realistic model is the persistent or worm-like chain. In this model, the polymer is approximated by a homogeneous, elastic thread with constant flexibility.

Although the directions of two segments in close proximity are correlated now, this correlation is lost at sufficiently large contour distance between the segments. A directional correlation of two segments means that the average cosine of the angle between the segments does not vanish:

$$\langle \vec{u}_i \vec{u}_j \rangle = \langle \cos \theta_{ij} \rangle \neq 0.$$

If the flexibility of the chain is constant, then this quantity depends only on the contour distance z between the two segments, and one can define:

$$\zeta(z) = \langle \cos \theta(z) \rangle.$$

For many models, $\zeta(z)$ is multiplicative, i.e. $\zeta(z + z') = \zeta(z) \cdot \zeta(z')$ [Grosberg and Khokhlov, 1994]. This relation is the characteristic property of the exponential function, and because of $\zeta(0) = 1$, it follows:

$$\zeta(z) = e^{-\frac{z}{L_p}}. \tag{2.95}$$

The decay constant is called the persistence length and is characteristic for the flexibility of the chain. It can be pictured as the length scale on which the chain can be considered rod-like since from $z \ll L_p$ it follows that $\zeta(z) = \langle \cos \theta(z) \rangle \approx 1$. This means that the chain has nearly identical tangent vectors at the endpoints of a section of length z (see Figure 2.21).
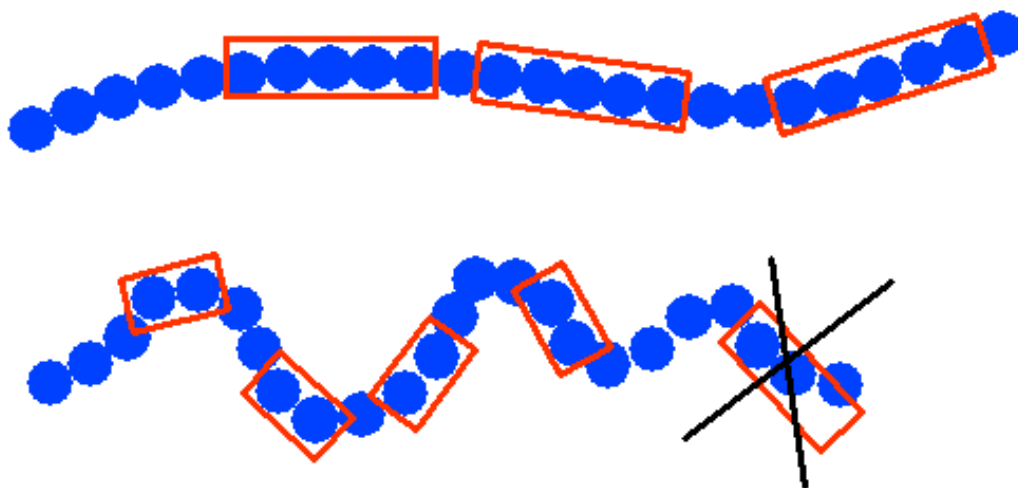
Figure 2.21: Illustration of the concept of the persistence length of a polymer chain. The persistence length defines the length scale on which the chain can be considered as a stiff rod. The illustration is based on work by FRANK AUMANN.

The notion of the persistence length suggests to approximate a long polymer with the freely jointed chain model with a segment length in the order of the persistence length of the polymer. The proportionality of the mean squared end-to-end distance to the contour length L_C of the chain is a characteristic property of the ideal chain, and one can use equation (2.94) to find the segment length L_K of the freely jointed chain with the same end-to-end

distance. L_K is called the effective segment length or KUHN length. For simple models as the worm-like chain model, the KUHN length can be calculated analytically. For many models and real macromolecules, it can be approximated as [Grosberg and Khokhlov, 1994]:

$$L_K \approx 2L_p. \tag{2.96}$$

A more thorough analysis of the relationship of the mean squared end-to-end distance and the persistence length leads to an improved formula, which yields better results specifically for short chains or chain parts [Rivetti et al., 1996]:

$$\langle r^2 \rangle = 2L_p L_c \left(1 - \frac{L_p}{L_c} \left(1 - e^{-\frac{L_p}{L_c}} \right) \right) \overset{L_c \gg L_p}{\approx} 2L_p L_c. \tag{2.97}$$

2.4.1.3   The distribution of the end-to-end distance

A more detailed description than $\langle r^2 \rangle$ is given by the distribution $\mathrm{P}(\vec{r})$ of the end-to-end vector. In the freely jointed chain model, $\vec{r}$ is the sum of N independent vectors. It follows from the central limit theorem [Feller, 1968] that $\vec{r}$ is normally distributed [Grosberg and Khokhlov, 1994]:

$$\mathrm{P}(\vec{r}) = \left(\frac{2\pi N s^2}{3} \right)^{-\frac{3}{2}} e^{-3\frac{\vec{r}^2}{2Ns^2}}.$$

For other models, the derivation is more difficult because the segment vectors are not independent of each other. However, it can be shown that the central limit theorem also holds in the case of exponentially decaying correlation [Grosberg and Khokhlov, 1994].

The probability density $\mathrm{P}(\vec{r})$ depends only on the modulus of $\vec{r}$. Therefore, the distribution of the end-to-end distance is:

$$\mathrm{P}(r) = \int \mathrm{P}(\vec{r}) \, d\Omega = 4\pi \left(\frac{2\pi N s^2}{3} \right)^{-\frac{3}{2}} e^{-3\frac{r^2}{2Ns^2}}. \tag{2.98}$$

Thus, the end-to-end distance is also normally distributed.

2.4.2   Deviations from the ideal chain model

2.4.2.1   Excluded volume interaction

The most important deviation of a real macromolecule from the model of the ideal chain is the so-called excluded volume. A real polymer has a finite thickness, and the segments cannot overlap with or cross each other. Furthermore, the segments can attract or repel each other electrostatically.

The immediate effect of the excluded volume interaction is a change in the size of the coil that is formed by the chain. It has turned out that at low concentrations, this effect can

be calculated independently of the precise realization of the excluded volume interaction [Grosberg and Khokhlov, 1994]. The coil swells if the interaction is predominantly repulsive. In this case, the size of the coil is proportional to $N^{\frac{3}{5}}$. If the interaction is dominated by an attractive part, the coil shrinks, and its size is proportional to $N^{\frac{1}{3}}$.

2.4.2.2 Interaction with the solvent

If a macromolecule has charged subgroups, they interact with molecules of the solvent and with dissolved ions. In many cases, these effects can be taken into account by introducing an effective potential for the excluded volume interaction. Then an explicit treatment of the solvent interaction is not necessary.

2.4.3 Modeling chromosome dynamics over time

The simulation of the conformation of a macromolecule over time is identical to the solution of the equations of motion of this molecule and of the surrounding solvent molecules. To do this for an entire chromosome at atomic resolution and on time scales of any interest, however, is impossible on current computers. Therefore it is necessary to simplify the model that is used in a simulation. In many cases, the full resolution is not even required to solve the given problem, and therefore certain simplifications are well justified:

- The molecule is not simulated on atomic level but as a polymer chain consisting of stiff segments. This coarse-graining process is specifically justified if one is mainly interested in global properties of the chain, e.g. the end-to-end distance.

- The solvent is modeled as a continuous medium. This is justified by the fact that the solvent molecules move much faster on average than the macromolecules and that in a time step that is relevant for the chain each segment is hit by many solvent molecules. The thermal motion of the solvent is represented in the model as a random force on the segment joints of the chain.

The simulation of a microscopic system in which the influence of the solvent is represented as stochastic force on the elements of the system is called a BROWNian dynamics (BD) simulation.

2.4.4 Sampling conformations

Although BROWNian dynamics simulations make use of important simplifications, they still take a lot of time. One important reason for this is that one needs to use very small time steps because the forces that different parts of the chain exert on each other are taken to be constant during one time step. A consequence is that the conformations of the chain still

correlate with the initial conformation after many time steps. Therefore, a BD simulation is not an efficient way to sample the conformational space of a polymer.

A better way to obtain a representative ensemble of conformations is a so-called Monte Carlo (MC) simulation. In an MC simulation, the conformation of the chain is not necessarily changed according to the forces that act on the segment joints. In fact, the changes can be made quite arbitrarily as long as they allow to explore the whole conformational space. A commonly used step is to randomly choose one segment joint and rotate one of the chain arms around a random axis through this joint. The important part of an MC step is that this conformational change is only a suggestion. Based on the temperature T used in the simulation and on the energy change ΔE caused by the step, the step is accepted if the BOLTZMANN factor $\exp(-\frac{\Delta E}{k_B T})$ is larger than a random number drawn from a uniform distribution between 0 and 1 [Binder, 1996].

This approach allows much faster sampling of the conformational space because much larger conformational deviations are possible within one step. It has to be noted, however, that there is no time information in a Monte Carlo simulation. Two conformations that follow each other in the course of a simulation would not usually follow each other in real time. They are just two possible conformations within the generated ensemble. Therefore, a BD simulation has to be carried out if the a full trajectory of the chain in space and time is required.

2.4.5   The program `corchy++`

The program `corchy++` was developed by KONSTANTIN KLENIN for the simulation of linear and circular DNA. However, the broad applicability to persistent polymer chains in general was a development principle from the start. A detailed description of the approach can be found in [Klenin et al., 1998]. In the following sections, I introduce some aspects that are important for my thesis work.

2.4.5.1   Brownian dynamics simulations

`corchy++` can be used both for BD and MC simulations. The former is implemented as a second-order BROWNian dynamics algorithm using two half-steps [Klenin et al., 1998]. It takes stretching, bending, and twisting forces and bending and twisting torques into account. The excluded volume interaction is calculated as the electrostatic repulsion of the charged DNA backbone using the DEBYE-HÜCKEL approximation [Debye and Hückel, 1923]. The hydrodynamic interactions, which occur when the motion of one part of the chain disturbs the solvent around another, are described by the ROTNE-PRAGER tensor [Rotne and Prager, 1969].

2.4.5.2   Monte Carlo simulations

The Monte Carlo simulations use segments of fixed length. Therefore, there is no stretching energy. The total energy of a chain conformation is the sum of bending, twisting, and excluded volume energy. In the case of linear chains, the only MC step is the rotation of a chain arm around a segment joint. My adaptations to the code are described in section 3.3.2.

Chapter 3

Telomere-Telomere Interaction

3.1 Summary

Nuclear organization into subcompartments (such as the nucleolus, the nuclear periphery, and the interior of the nucleus) is important for many nuclear functions, including gene expression, recombination, and DNA repair. The telomeres of budding yeast form clusters close to the nuclear periphery, but the composition and dynamics of the clusters are unknown. A way to gain insight into these phenomena is to observe the distance distributions of pairs of telomeres. It seems that the intrachromosomal distances between some pairs of telomeres are smaller than interchromosomal telomere-telomere distances. The peripheral telomeres of the short chromosomes 3 and 6 are particularly close to their respective intrachromosomal partners. I used a polymer chain model and Monte Carlo simulations to determine the expected telomere-telomere distances of these chromosomes and the effect of peripheral anchoring of telomeres on these distances. I found that the experimentally observed telomere-telomere distances are indeed smaller on average than predicted by my simulations. This juxtaposition cannot be explained by peripheral anchoring alone because the restriction of telomere movement to the nuclear periphery leads to an increase of the average telomere-telomere distance.

3.2 Introduction

As discussed in section 1.2.3, the telomeres of budding yeast form clusters close to the nuclear periphery, which is important for the maintenance of silent chromatin [Palladino et al., 1993, Gotta et al., 1996]. These early studies relied on fluorescent labeling of the telomere binding protein Rap1 and fluorescence *in situ* hybridization (FISH) of subtelomeric Y' elements. These methods visualize all telomeres at once and therefore do not allow a determination of the composition of the pools. A later *in vivo* imaging study observing single telomeres [Hediger et al., 2002] revealed that telomere anchoring is dynamic and allows the

telomeres to slide along the nuclear envelope within a certain perimeter[1], but it remained unclear if the observed movement reflects the movement of an entire telomere cluster or of a single telomere independently of its cluster, possibly switching from one pool to another.
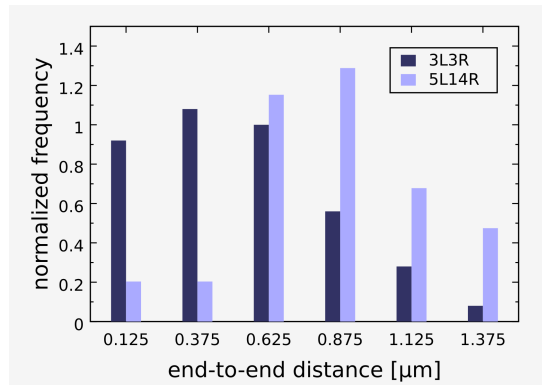
Figure 3.1: Telomeres of the same chromosome are often closer together than telomeres of distinct chromosomes: Shown are the distance distributions of 3L3R and 5L14R [Bystricky et al., 2005].

One way to shed light on the composition and dynamics of telomere clusters is to look at the distance distribution of telomere pairs. In a haploid yeast cell with 16 chromosomes and 32 telomeres, there are $\frac{32!}{30! \cdot 2!} = 496$ combinations of two telomeres. Of course, it is very time consuming to measure all of them. Nonetheless, the comparison of selected examples might still help to elucidate the dynamics of cluster composition. At the starting time of this project, it was known from [Bystricky et al., 2005] that the telomeres of the same chromosome are noticeably close together compared to the telomeres of distinct chromosomes (Figure 3.1). The telomeres of the small chromosomes 3 and 6 are particularly close to their respective partners (Figure 3.2).

The fact that these telomeres are also particularly well anchored at the nuclear periphery raised the question of whether telomeric anchoring could contribute to their close proximity. The investigation of *sir4Δ* and *yku70Δ* mutant strains, in which telomere anchoring is impaired (see section 1.2.3), does not give a clear result: the consequences of deleting *SIR4* or *YKU70* on anchoring are quite different, yet the effects on the telomere-telomere distance are very similar (Figure 3.3). Therefore, I used a theoretical model to elucidate the effects of telomeric anchoring on the telomere-telomere distance of chromosomes. In contrast to an experimental approach, analytic calculations and computer simulations allow selected parameters to be manipulated without effecting any other properties of the system. Since the chromosome is constrained to the nucleus, I could not apply known analytical formulas for

---

[1]Here and in the following, "anchored" means that the movement of a telomere is restricted to the nuclear periphery but free to move along it.
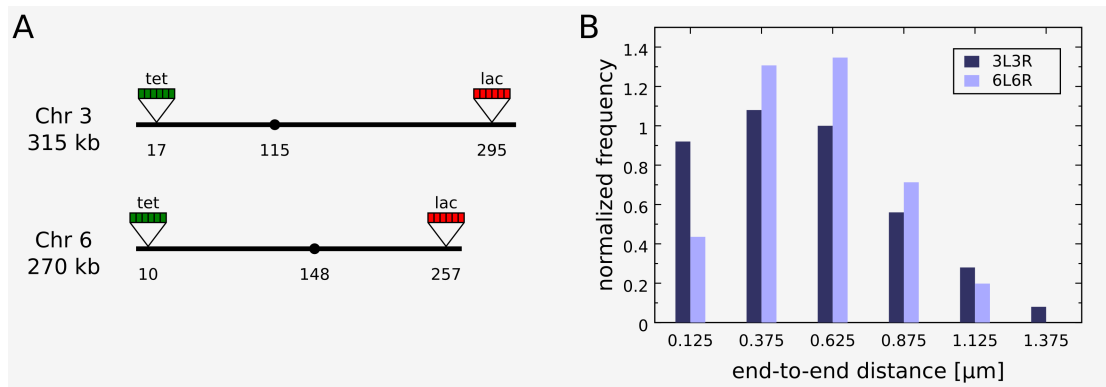
Figure 3.2: (A) The telomere-telomere distance distributions of chromosomes 3 and 6 were measured microscopically by introducing arrays of lac and tet operators close to the telomeres. (B) Telomere-telomere distance distribution for chromosomes 3 and 6 [Bystricky et al., 2005].

the end-to-end distance distribution of a unconfined polymer. I developed a computational model to investigate the following questions:

1. Are the two telomeres of chromosomes 3 and 6 indeed closer together to each other than expected, or is their close proximity rather a consequence of their small length?

2. What effect does the restriction of telomeric anchoring have on the distance between two telomeres?

3.3 A polymer model for telomere-telomere interaction

Numerous studies on the flexibility of the chromatin fiber have been carried out, but the influence of spatial confinement on the conformation of chromatin has only been investigated in a few cases. I modeled a yeast chromosome as a single flexible polymer chain confined to a spherical volume with a radius of 900 nm (Figure 3.4). End-to-end distance distributions were derived from conformation ensembles generated by Monte Carlo simulations.

In the model, the chain is composed of a succession of beads and segments that are represented by balls and cylinders of 30 nm diameter. The excluded volume interaction is realized as a hard core interaction (see section 2.4). The centromere bead can be fixed completely or attached elastically to the spindle pole body. Both the excluded volume interaction and the movement of the centromere ($\pm 0.3\,\mu m$) had only minor effects on the end-to-end distance distribution (data not shown).
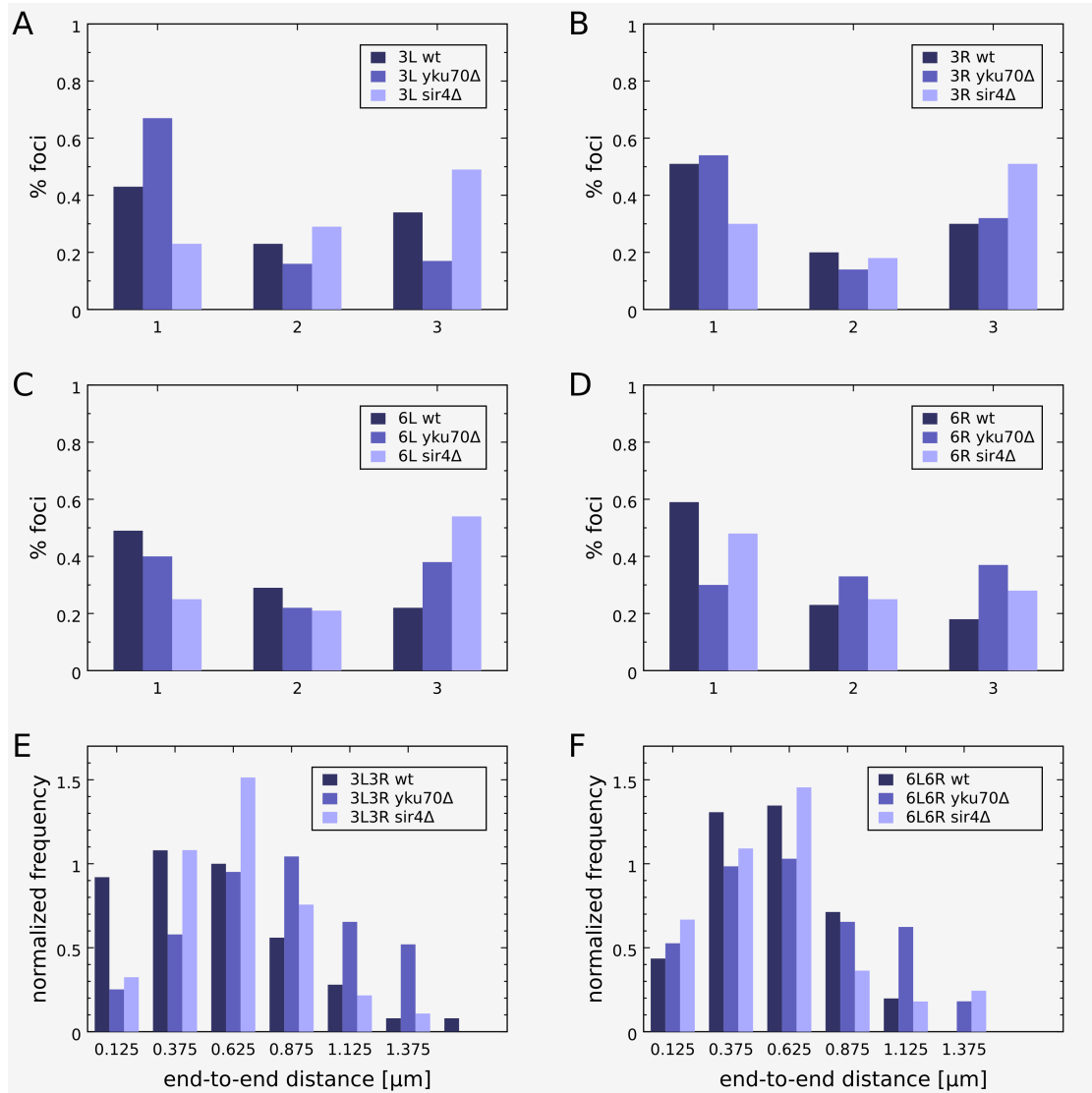
Figure 3.3: The deletion of *YKU70* or *SIR4* has different effects on the anchoring of the telomeres 3L/3R and 6L/6R but similar effects on the intrachromosomal telomere-telomere distances [Hediger et al., 2002, Bystricky et al., 2004, Bystricky et al., 2005]. (A)+(B) Position (see section 2.3.1) of telomeres 3L and 3R in wt, *yku70*Δ, and *sir4*Δ. The deletion of *KU70* leads to an even tighter anchoring of 3L, while *sir4*Δ releases both telomeres. (C)+(D) Position of telomeres 6L and 6R. The deletion of *KU70* releases 6R, the deletion of *SIR4* releases 6L. (E)+(F) The absence of yKu70 leads to a significant increase in telomere-telomere distances while the absence of Sir4 does not.
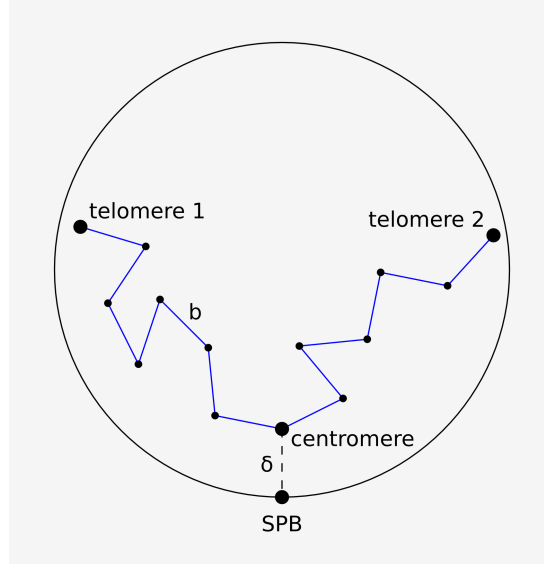
Figure 3.4: The chromosomes were modeled as flexible chains with segment length b. One bead representing the centromere was fixed or elastically attached to the spindle pole body (SPB) at a distance $\delta \approx 300\,\text{nm}$ [Bystricky et al., 2005].

3.3.1 Contour length and persistence length

The most important parameters for the model are the contour length L_C and the persistence length L_p, which defines the bending rigidity of the polymer (see section 2.4.1).

The length of a chromosome is known as the number of base pairs, but to get its contour length in space the linear mass density has to be determined. We have shown in a previous study [Bystricky et al., 2004] that yeast interphase chromatin exists in a rather compact structure with a persistence length $L_p = 170\text{-}220\,\text{nm}$ and a mass density $c = 110\text{-}150\,^{\text{bp}}/_{\text{nm}}$. These results were obtained by fitting the KRATKY-POROD equation (see section 2.4.1, [Doi and Edwards, 1986, Rivetti et al., 1996]) to experimentally determined point-to-point distances r on a chromosome:

$$\langle r^2 \rangle = 2L_p^2 \cdot \left(e^{-\frac{L_C}{L_p}} + \frac{L_C}{L_p} - 1 \right). \tag{3.1}$$

$$\langle r^2 \rangle = 2L_p^2 \cdot \left(e^{-\frac{L_g}{cL_p}} + \frac{L_g}{cL_p} - 1 \right). \tag{3.2}$$

Here, L_g is the genomic distance between the two tagged loci on the chromosome. Then $\frac{L_g}{c}$ gives the contour length L_C.

For my simulations, I used a linear mass density of $130\,^{\text{bp}}/_{\text{nm}}$, which results in contour lengths of $2438\,\text{nm}$ and $2077\,\text{nm}$ for chromosomes 3 and 6, respectively.

The persistence length L_p of a polymer chain is defined as the decay constant of the directional correlation of the tangent vectors at two points with increasing chain length between them (see section 2.4.1). A greater persistence length is equal to a stiffer chain. Figure 3.5 illustrates the impact of the bending flexibility of a polymer on its end-to-end distance using the freely jointed chain model. These simple simulations show that increasing the flexibility of a chromosome has dramatic consequences on the distribution of its end-to-end distance, allowing the two ends to get closer to each other.
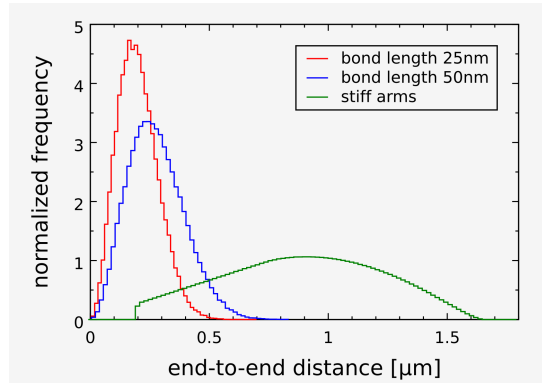
Figure 3.5: End-to-end distance distribution for a freely jointed chain with segment lengths b of 25 nm (higher flexibility) and 50 nm (lower flexibility). The distributions are normalized such that the integral gives 1 µm. I used arm lengths of 1139 nm and 938 nm, corresponding to the arms of chromosome 6. The distribution of a model where arms are represented as stiff rods is shown for comparison.

The freely jointed chain model gives a good first approximation of the effect of increasing polymer flexibility. However, its validity is restricted to chains that are longer than ten KUHN segments (see section 2.4.1), which is not the case for yeast chromosomes 3 and 6. For a quantitative comparison of the model with experimental data, I used the more detailed worm-like chain model (section 2.4.1).

The code of my simulation program is based upon the previously described Monte Carlo and BROWNian dynamics simulation package `corchy++` (see section 2.4.5, [Klenin et al., 1998]). A chain is modeled as a linear sequence of segments connected by flexible joints; the bending energy E_i at the ith joint is given by

$$\frac{E_i}{k_B T} = \alpha \beta_i^2 \tag{3.3}$$

where β_i is the joint angle.

The bending rigidity parameter α is directly related to the persistence length L_p [Klenin et al., 1998]. For a free polymer, L_p can be obtained from the mean squared end-to-end distance of the chain using equation (3.2), but for a confined chain this relation is not valid.

As mentioned before, we had obtained c and L_p from our experimental data by fitting equation (3.2) to the measured distances. Now, I used the same approach to determine the effective persistence length of the simulated chains (see Figure 3.6).
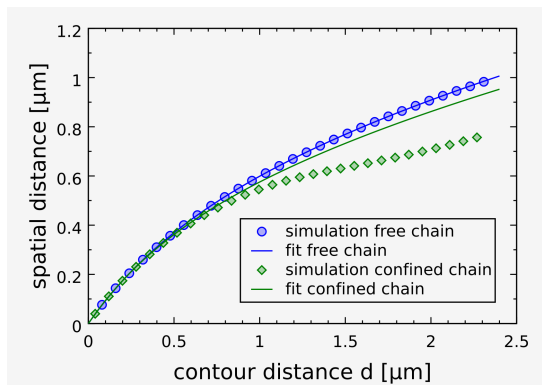
Figure 3.6: Fit of the square root of the KRATKY-POROD formula (3.2) to $\sqrt{\langle r^2 \rangle}$ for subchains of contour length d, where r is the end-to-end distance of the subchain. For a free chain the data correspond very well to the analytical formula, but for a confined chain there is an agreement only for small d.

For a free chain, the data are perfectly consistent with equation (3.2). When confining the chain to the nuclear volume, however, there is only an agreement for short subchains (Figure 3.6). A fit of the KRATKY-POROD formula to the data with $d < 400\,\text{nm}$ shows that the contour length given as a parameter for the simulation is well reproduced (with an error of less than 3%), but that the persistence length obtained from the fit is reduced by about 15% as a result of the confinement. For the following simulations, I chose the rigidity parameter such that the measured effective persistence length was $200\,\text{nm}$, to match the experimentally determined value [Bystricky et al., 2004]. In the absence of confinement, one would measure a persistence length of $225\,\text{nm}$ for this chain.

3.3.2 Main adaptations to `corchy++`

The program `corchy++` was developed for the simulation of DNA in free solution (see section 2.4.5). The physical properties of the polymer chain are controlled via an input file and do not require changes of the code in most cases. However, the confinement of the chain and the peripheral anchoring of telomeres required several adaptations.

- The original version of `corchy++` implements the electrostatic repulsion between different segments of the DNA backbone. This leads to an excluded volume interaction of the polymer chain. However, this is not a realistic model for the chromatin fiber, which does not carry an effective electric charge. Therefore, I changed the excluded

volume interaction into a hard-core interaction. The implementation for a Monte Carlo simulation is straight-forward: A conformation with overlapping segments is punished with a high energy penalty, which effectively prohibits this conformation.

Even if the effect of the exchange of the excluded volume interaction on the results might be small, the hard-core interaction is supposed to be computationally much less expensive than the electrostatic interaction.

- In the case of non-anchored telomeres, the confinement to the spherical volume of the nucleus is realized in a similar way. A conformation that comprises a bead outside of the nucleus is forbidden by an energy penalty.

  In the presence of a confinement, the absolute position of the chain matters. Therefore, the rotation of chain arms as the only Monte Carlo step is not sufficient. I added a translation step which moves the whole chain in a random direction.

- In the case of an anchored telomere, the Monte Carlo steps have to be modified in order to preserve the distance of the anchored telomere from the center of the nucleus. After generation of a start conformation with peripheral telomeres, two kinds of steps were used.

  1. Internal rotation: A subchain is rotated around the axis connecting the two terminal beads. This does not affect telomere position.

  2. Tail rotation: One part of a chromosome arm beginning at a randomly chosen bead is rotated around the nuclear diameter through this bead. Such a rotation preserves distances from the nuclear center and therefore moves the chain end along the periphery.

- The attachment of the centromere to the spindle pole body was realized as a harmonic interaction which was added to the total energy of a chain conformation.

3.4 Results

3.4.1 The telomeres of chromosomes 3 and 6 are closer to each other than predicted for free telomeres

I used my modified version of `corchy++`, which I introduced in section 3.3, to compute the telomere-telomere distance distributions of chromosomes 3 and 6. Figure 3.7 shows the results for chromosome 6 with free telomeres in comparison to the experimental data. The results for chromosome 3 are very similar [Gehlen et al., 2006]. On average, the predicted telomere-telomere distance is significantly higher than measured in wild type ($p < 10^{-9}$ in a t-test). Although the preferred peripheral location of telomere 6L is compromised in a *sir4*Δ

mutant, the absence of Sir4 did not result in a significant change of the telomere-telomere distance (Figure 3.3C). In contrast, the deletion of *YKU70*, which leads to a random position of telomere 6R in the nucleus, significantly increased the distance between the two telomeres, although this distance is still slightly smaller than in the simulations. This result reinforces the question if telomeric anchoring is responsible for small distances observed *in vivo*.
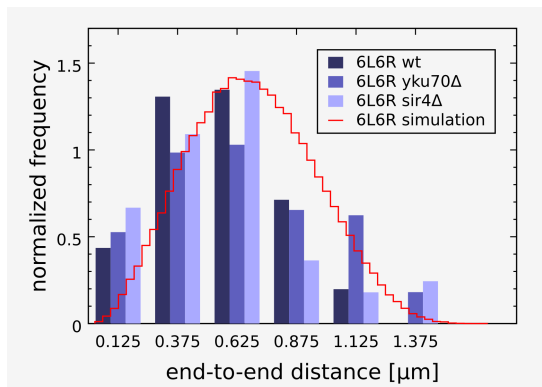
Figure 3.7: Monte Carlo simulation of chromosome 6. The telomere-telomere distance distribution of chromosome 6 in wt and anchoring mutants (see Figure 3.3) compared to the theoretical prediction for free telomeres. The difference between the wt and the simulated distributions is highly significant ($p < 10^{-9}$ in a t-test). The distances in *yku70*Δ are significantly larger than in wt ($p < 0.004$) and close to the prediction ($p < 0.03$) while the distribution for *sir4*Δ does not differ significantly from wt. The histograms for the simulation data were each calculated from 150000 independent conformations.

3.4.2 Peripheral anchoring increases the distance between two uniformly distributed spots, but also increases the probability to be in very close proximity

With increasing length of a chromosome, the movement of the two telomeres gets less and less correlated. In the absence of additional regulatory elements, the telomeres of a sufficiently long chromosome should behave like two independent spots. The distance distribution of two such spots inside a spherical volume and the impact of restriction to the surface of the sphere can be calculated analytically. Before I present the simulation results for the anchoring of the telomeres of chromosomes 3 and 6, I therefore show these analytical calculations. They already give a first indication of the influence of telomeric anchoring.

I calculated the probability density functions $P(r)$ [2] for two spots inside or on the surface

---

[2]I use the same letters P, $\tilde{P}$, $\hat{P}$, and $\hat{\tilde{P}}$ in all the three cases (both spots internal, one peripheral, both

of a sphere being at a certain distance r. I made use of two properties to simplify the calculations:

- The whole problem scales with the radius R of the sphere. Therefore, one can safely assume $R = 1$.

- The sphere can always be rotated such that the first spot lies on the z axis. Therefore, only its distance ρ from the sphere center has to be considered. $0 \le \rho \le 1$ always holds.

$P(r)dr$ is the probability that the two spots have a distance between r and $r + dr$. This term can be split into

$$P(r)dr = \int_0^1 \tilde{P}(\rho, r)d\rho dr \tag{3.4}$$

where $\tilde{P}(\rho, r)d\rho dr$ is the probability that the first spot is between ρ and $\rho + d\rho$ from the sphere center and the second spot is at a distance between r and $r + dr$ from the first spot. We can further write:

$$\tilde{P}(\rho, r) = \hat{P}(\rho) \cdot \hat{\tilde{P}}(\rho, r) \tag{3.5}$$

where $\hat{P}(\rho)d\rho$ is the probability for spot 1 being at a distance ρ' with $\rho \le \rho' \le \rho + d\rho$ and $\hat{\tilde{P}}(\rho, r)dr$ is the probability for spot 2 being at distance r', $r \le r' \le r + dr$ **under the condition** $\rho' = \rho$.

3.4.2.1   Both spots inside the sphere

In this case, the probability for one spot to be in a volume V inside the sphere is always $\frac{V}{4/3\pi}$, independently of the location of the volume. If spot 1 is at a distance between ρ and $\rho + d\rho$ from the sphere center, the representing region is a shell with radius ρ and thickness $d\rho$ and has a volume of $4\pi\rho^2 d\rho$. It follows

$$\hat{P}(\rho) = \frac{4\pi\rho^2}{4/3\pi} = 3\rho^2.$$

The calculation of $\hat{\tilde{P}}(\rho, r)$ consists also of surface calculations of spheres and spherical caps and is not difficult *per se*, but complicated because several case differentiations have to be made (see Figure 3.8 for illustration). I give the full derivation in appendix B.1 and here only the result:

$$P(r)dr = 3r^2 - \tfrac{9}{4}r^3 + \tfrac{3}{16}r^5. \tag{3.6}$$

---

peripheral). This should not lead to any confusion.

From that, we get the average distance:

$$\langle r \rangle = \int_0^2 r \cdot \mathrm{P}(r) dr$$

$$= \frac{3}{4}r^4 - \frac{9}{20}r^5 + \frac{3}{112}r^7 \Big|_0^2$$

$$\langle r \rangle = \frac{36}{35}. \tag{3.7}$$
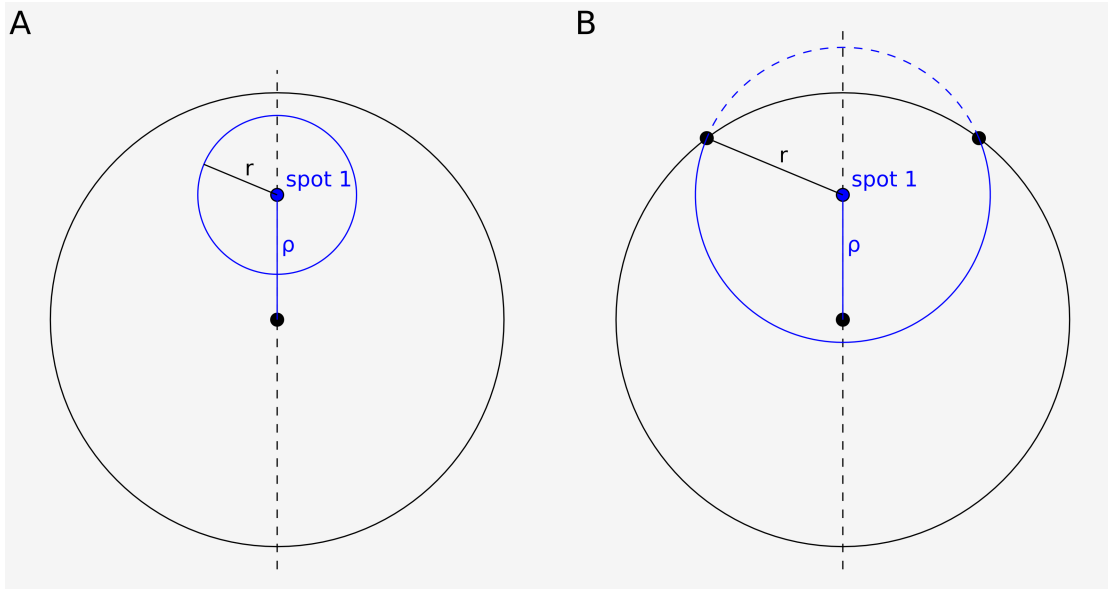
Figure 3.8: If spot 1 is at a distance ρ from the nuclear center, then the probability for spot 2 being at a distance r from spot 1 is proportional to the surface of a spherical cap (blue). Depending on the values of ρ and r, different cases have to be considered, two of which are shown in panel A and B. For a full analysis of the problem see appendix B.1.

3.4.2.2 One spot inside the sphere, one on the surface

The result from the last section can be applied to this case:

$$\hat{\mathrm{P}}(\rho) = \begin{cases} 0 \text{ for } \rho = 1 \\ 1 \text{ otherwise} \end{cases}$$

or more elegantly with the δ distribution [Boas, 1983]

$$\hat{\mathrm{P}}(\rho) = \delta(1 - \rho).$$

With the results from the previous section (see appendix B.1) we get:

$$\mathrm{P}(r) = \int_0^1 \tilde{\mathrm{P}}(\rho, r) d\rho = \int_0^1 \hat{\mathrm{P}}(\rho) \, \hat{\tilde{\mathrm{P}}}(\rho, r) d\rho = \int_0^1 \hat{\mathrm{P}}(\rho) \, \delta(1 - \rho) d\rho$$

$$= \hat{\tilde{\mathrm{P}}}(1, r)$$

$$= \tfrac{3}{4} r \left(1 - (r - 1)^2\right)$$

$$\mathrm{P}(r) = \tfrac{3}{4} r^2 (2 - r) \tag{3.8}$$

and

$$\langle r \rangle = \int_0^2 r \cdot \mathrm{P}(r) dr$$

$$= \left. \tfrac{3}{8} r^4 - \tfrac{3}{20} r^5 \right|_0^2$$

$$\langle r \rangle = \tfrac{6}{5}. \tag{3.9}$$

3.4.2.3 Both spots on the surface of the sphere

Here I used a different approach. If spot 1 resides at $(0, 0, 1)$ then the following relationship between the spot-spot distance r and the angle θ (see Figure 3.9) follows from the law of cosines:

$$r = \sqrt{2(1 - \cos \theta)}. \tag{3.10}$$

The probability $\mathrm{P}(\theta)$[3] for spot 2 being at a position with angle θ (and thus at distance r since equation (3.10) is a bijection[4]) is proportional to the circumference circle with radius $\sin \theta$:

$$\mathrm{P}(\theta) d\theta = \frac{2\pi \sin \theta d\theta}{4\pi}$$

$$\mathrm{P}(\theta) d\theta = \tfrac{1}{2} \sin \theta d\theta \tag{3.11}$$

---

[3]For simplicity, I reuse the identifier P for this probability although strictly speaking, $\mathrm{P}(\theta)$ and $\mathrm{P}(r)$ are different entities.

[4](3.10) establishes a one-to-one relationship between θ and r. Therefore, it is effectively the same thing to talk about θ or r.
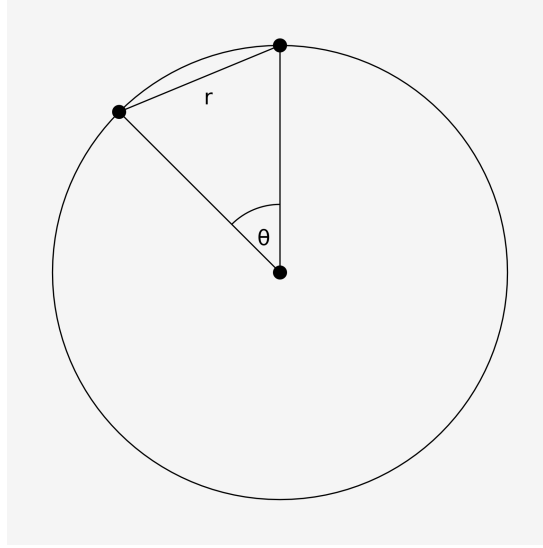
Figure 3.9: Spot 1 is at the top. Then there is a bijective map from the angle between
the z axis and the vector to spot 2 to the distance between the spots.

This probability density can now be transformed as follows [Brandt, 1998]:

$$\mathrm{P}(r)\left|\frac{dr}{d\theta}\right| = \mathrm{P}(\theta) \tag{3.12}$$

$$\mathrm{P}(r)\left|\frac{\sin\theta}{r}\right| = \tfrac{1}{2}\sin\theta$$

$$\mathrm{P}(r)\frac{\sin\theta}{r} = \tfrac{1}{2}\sin\theta$$

$$\mathrm{P}(r) = \tfrac{r}{2} \tag{3.13}$$

and

$$\langle r \rangle = \int\limits_0^2 r \cdot \mathrm{P}(r)dr$$

$$= \frac{r^3}{6}\bigg|_0^2$$

$$\langle r \rangle = \tfrac{4}{3}. \tag{3.14}$$

3.4.2.4   Analysis

Figure 3.10 shows the results together with Monte Carlo simulations as a control. The
restriction of first one and then both spots to the periphery of the sphere shifts the average

distance between the spots from $\frac{36}{35}R \approx 1.03R$ to $\frac{6}{5}R = 1.2R$ and then to $\frac{4}{3}R \approx 1.33R$. This means that on average the restriction of one or two spots to the periphery does not bring the spots closer together. On the contrary, it increases their distance.
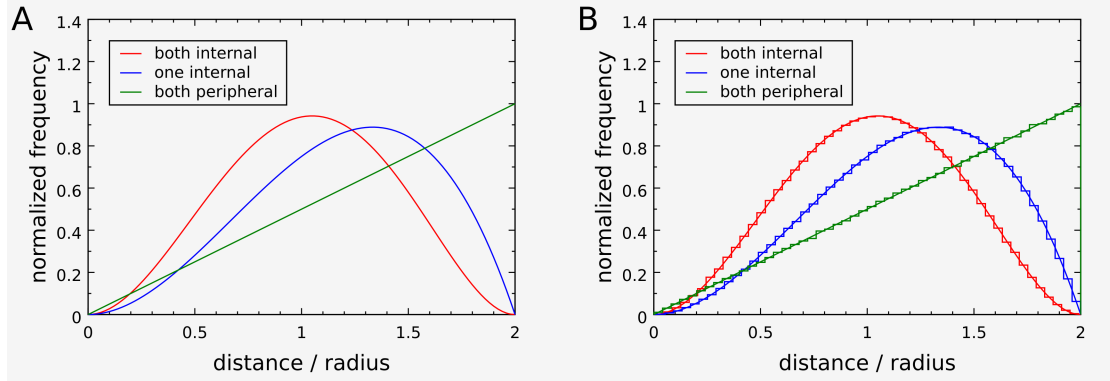
Figure 3.10: (A) Distance distribution for two spots inside a sphere or on its periphery. (B) I confirmed my calculations by computing the distance histogram from 1000000 pairs of points for each condition.
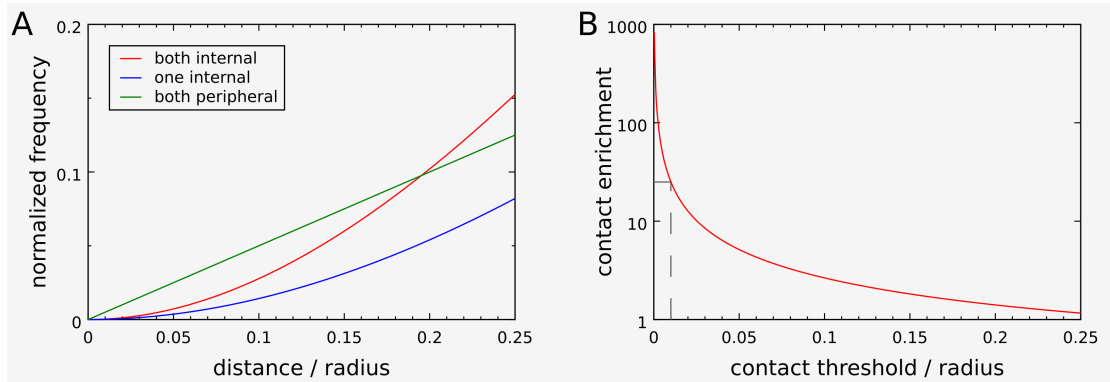
Figure 3.11: (A) Low distance part of Figure 3.10. Two peripheral spots are more likely to be very close together than two free ones. (B) Integrated probability for two peripheral spots being closer together than the contact threshold divided by the integrated probability for two free spots. For example, the probability for being closer together than $10\,\text{nm}$ $(0.01R)$ is about 25 times higher for peripheral spots than for free ones.

On the other hand, a closer look at the frequency of very small distances reveals that, while the restriction of one spot decreases the probability density, it is highest for two

peripheral points (Figure 3.11A). The integrated probability for the two spots being closer together than $0.01R$ (10 nm in a nucleus with a diameter of $2\,\mu m$) is less than 10^{-6} for two free spots but $2.5 \cdot 10^{-5}$ for two peripheral spots. This means that — despite the higher average distance — the chance of touching each other is more than 25 times higher for peripheral spots than for internal ones. It should be mentioned though that this value is highly dependent on the threshold distance for touching, which I have arbitrarily set to 10 nm here (Figure 3.11B). One has to keep in mind, however, that the contact probabilities are in general very low. It is also important to note that one cannot make direct conclusions from this data about the time two spots need to get into close contact (see chapter 5).

3.4.3 Telomeric anchoring causes an increase, not a decrease of telomere-telomere distances
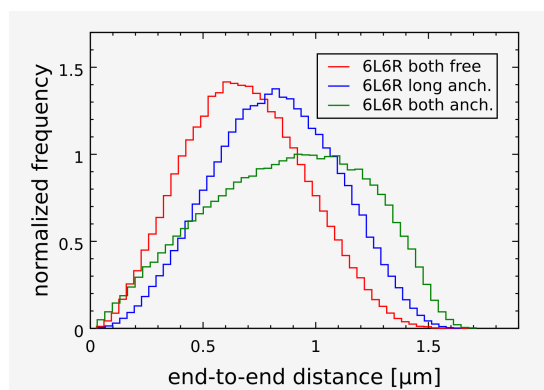
Figure 3.12: Modeled telomere-telomere distance distribution of chromosome 6 for two free telomeres, anchored telomere 6L (long arm), and two anchored telomeres. The result for anchoring of telomere 6R is almost identical to the result for anchoring of 6L and is omitted for clarity. The histograms were each calculated from 150000 independent conformations.

The analytical calculations for two independent spots have shown that restriction to the periphery of the sphere leads to an increase in the spot-spot distance. To investigate if this is the same for the telomeres of chromosomes 3 and 6, I included the anchoring of telomeres into the simulation (see section 3.3.2). Figure 3.12 shows the results for chromosome 6. Again, the observations for chromosome 3 are very similar. In agreement with the results for two independent spots, restriction of the telomere movement to the nuclear periphery leads to further increase in the telomere-telomere distance. The distribution characteristics of the experimental and the simulated data are summarized in Table 3.1. From these results, it follows that peripheral anchoring *per se* cannot be the reason for the small telomere-telomere

distances observed for chromosomes 3 and 6. However, the chromosome simulations show in agreement with the calculations for two spots that the probability whereby the telomeres come extremely close together is higher for anchored telomeres than for free ones. The probability for the telomeres being closer together than $10\,\mathrm{nm}$ [5] is 5 ± 2 times higher for anchored telomeres.

Telomeres	experiment/simulation	n	$r\,[\mathrm{µm}]$	$\sigma\,[\mathrm{µm}]$
6R 6L wt	experiment	153	0.558	0.259
6R 6L *yku70Δ*	experiment	192	0.650	0.351
6R 6L *sir4Δ*	experiment	66	0.564	0.328
6R free 6L free	simulation	150000	0.688	0.263
6R anchored 6L free	simulation	150000	0.835	0.275
6R free 6L anchored	simulation	150000	0.821	0.275
6R anchored 6L anchored	simulation	150000	0.871	0.346
3R 3L wt	experiment	56	0.548	0.358
3R 3L *yku70Δ*	experiment	108	0.801	0.354
3R 3L *sir4Δ*	experiment	37	0.611	0.282
3R free 3L free	simulation	150000	0.736	0.281
3R anchored 3L free	simulation	150000	0.874	0.298
3R free 3L anchored	simulation	150000	0.900	0.297
3R anchored 3L anchored	simulation	150000	0.946	0.378
5L 14R wt	experiment	58	0.838	0.311
5L free 14R free	simulation	150000	0.697	0.267
5L anchored 14R anchored	simulation	150000	0.875	0.360

Table 3.1: Sample size n, mean r and standard deviation σ for the experimental and simulated distributions.

The simulations of chromosomes 3 and 6 showed that their telomeres are closer to each other than predicted for free telomeres and that anchoring leads to a further increase of this distance. However, the starting point of this investigation was that the distances of telomeres on distinct chromosomes are much higher than those on the same chromosomes. Therefore, I also determined the predicted distance distribution of telomeres 5L and 14R [6], which are also well anchored [Bystricky et al., 2005]. Unlike the intrachromosomal distances of telomeres 3L3R and 6L6R, the average distance of telomeres 5L and 14R in wild type

---

[5]Due to the excluded volume interaction, the center-to-center distance of two beads is always greater than or equal to $30\,\mathrm{nm}$. Closer than $10\,\mathrm{nm}$ means that the center-to-center distance is less than or equal to $40\,\mathrm{nm}$.

[6]The version of `corchy++` that I used for this project always simulates one chromosome. Therefore I mimicked the situation of two unconnected chromosome arms by making the centromere bead completely flexible. I expect that the simulation of two entirely unconnected chromosome arms would lead to slightly larger distances than shown here.

cells is significantly higher than predicted by the simulation for free telomeres ($p < 10^{-5}$, see Figure 3.13). Moreover, the mean of the simulated distribution for anchored telomeres agrees well with the mean measured distance ($p > 0.2$), although the distribution is considerably broader than in the experiments (Figure 3.13).
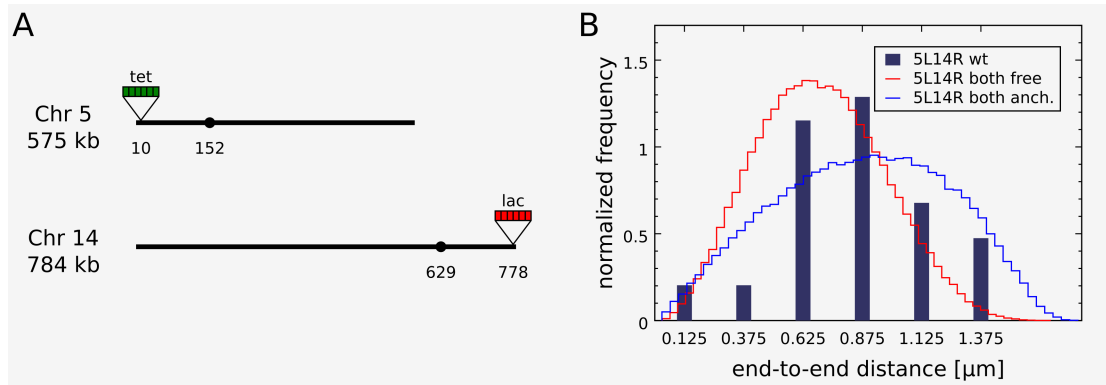
Figure 3.13: Monte Carlo simulation of chromosome arms 5L and 14R: The histograms for the simulation data were each calculated from 150000 independent conformations. (A) As before, the insertion of lac and tet operators close to the telomeres were used measure the telomere-telomere distance microscopically. (B) The distance distribution of telomeres 5L and 14R obtained experimentally and predicted theoretically. The measured average distance is significantly larger than predicted for free telomeres ($p < 10^{-5}$ in a t-test), but agrees well with the prediction for anchored telomeres ($p > 0.2$).

3.5 Discussion

The telomeres of chromosomes 3 and 6 appear to be noticeably closer together *in vivo* than the telomeres of distinct chromosomes. Additionally, they are particularly well anchored at the nuclear periphery. I used analytical calculations and computer simulations to elucidate the unperturbed distance distribution of these telomere pairs and the relationship between telomere anchoring and telomere-telomere distance.

I found that the telomeres of chromosome 3 and 6 are indeed significantly closer together than predicted by the model for free telomeres. In the model I used, chromatin is represented as a homogeneous flexible polymer with a linear mass density of $130\,^{bp}/_{nm}$ and a persistence length of $200\,nm$. A recent study of [Dekker, 2008] resulted in widely scattered, but substantially lower values for both parameters. The results are centered around a mass density of about $25\,^{bp}/_{nm}$ and a persistence length of about $100\,nm$. The contour length of a chromosome is inversely proportional to the linear mass density, and the mean squared

end-to-end distance is approximately proportional to the product of the contour length and the persistence length. These relationships lead to the assumption that the modeled telomere-telomere distances would be even greater based on the parameter values of [Dekker, 2008] than for the values used for this study. Furthermore, telomeric anchoring leads to a further increase in the average distance distance both for two independent spots and for the investigated telomeres. Despite this effect on the mean distance, the frequency of close contact is increased by peripheral anchoring.

I started with the simplest model possible: A single chromosome in the nucleus without any restrictions. This model does not reproduce the small telomere-telomere distances observed *in vivo* and the question is how the model can be modified to improve the prediction. One important simplification of the model is that it assumes that the chromosome samples the whole conformation space, i.e. that it has enough time and is not hampered to assume every possible conformation. This is not necessarily the case in a cell. Firstly, time is not unlimited. If the chromosome ends up in a certain conformation after mitosis and moves very slowly then the conformation at the beginning of the next mitosis will still be correlated with the start conformation. The question how fast the chromosome samples the conformation space cannot be answered by a Monte Carlo simulation (see section 2.4.4). Secondly, especially the movement at the nuclear periphery can be obstructed by obstacles or the binding to other telomeres, which restricts the telomere to a certain region of the nuclear periphery.
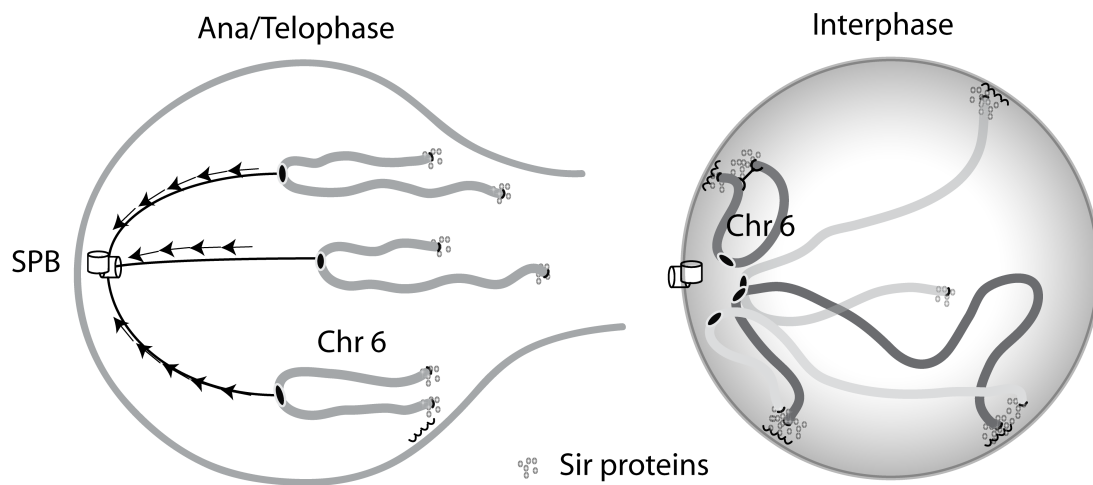
Figure 3.14: The Rabl conformation of chromosomes in late anaphase promotes the juxtaposition of telomeres on equally long arms and on the same chromosome. This configuration could then be preserved by telomeric anchoring. The illustration is adapted from [Schober et al., 2008].

Previous studies have shown that the movement of anchored telomeres is restricted at least on the time scale of minutes [Hediger et al., 2002] and have quantified the region of

constraint [Rosa et al., 2006]. It is possible to include this restriction into my model but the value of the results would be quite limited. One can expect that in this case, the simulated distance distributions would be trivially quite close to the measured ones. However, the restriction of movement explains that the measured distributions in general tend to be narrower than the simulated ones.

I have laid out that the restriction of telomeric movement on the time scale of the yeast cell cycle would conserve the telomere-telomere distance of the start conformation. This could explain the difference between intra- and interchromosomal telomere-telomere distances if the telomeres of the same chromosome start particularly close together. Possible reasons for this have been further investigated by [Schober et al., 2008]. Apart from a unique sequence element on telomere 6L, close proximity is favored between telomeres of chromosomes with equal arm lengths. This can be explained by the polarized RABL conformation of yeast chromosomes in anaphase which is then conserved by a restriction of movement during interphase (see Figure 3.14).

Chapter 4

Asymmetric Segregation of ARS Plasmids

4.1   Summary

Asymmetric cell division in single-celled organisms is thought to have arisen from the need for rejuvenation. In multicellular organisms, asymmetric division drives the generation of differentiated cells and maintenance of a stem cell pool. In budding yeast, autonomously replicating sequence (ARS) plasmids show a strong tendency to segregate to the mother cell at mitosis. The accumulation of these plasmids is a cause of aging in yeast. I used a computational model to show that the geometric shape of the dividing nucleus and the limited length of mitosis impose a severe barrier on passive diffusion into the daughter nucleus, resulting in a strong asymmetry in plasmid distribution. Consistent with the simulation, plasmid partitioning is improved in a mutant with an extended mitotic duration. Previous work identified subtelomeric sequences that overcome the physical barrier to segregation. In this project, we have shown experimentally and theoretically that tethering of ARS plasmids to the inner nuclear membrane can increase the efficiency of plasmid partitioning. Similarly, association with a chromosome also enhances partitioning efficiency.

4.2   Introduction

Asymmetric cell division is a fundamental process in the life cycle of an organism. During cell differentiation, precursor cells give rise to the variety of different cell types that make up a multicellular organism, yet retain their properties as pluripotent stem cells. This asymmetry is induced by the unequal distribution of cellular content during cell division [Gönczy and Rose, 2005]. Additionally, cells accumulate various forms of damage to DNA, proteins, and subcellular organelles, which are caused by vital cellular activities. This has been proposed to be a cause for cellular aging [Partridge and Barton, 1993]. Since the germ cell lineage can

be passed on from generation to generation indefinitely, germ cells must be protected from this process. Indeed, differential segregation of damaged material during cell division has been shown to play an important role in restricting senescence to one daughter from bacteria to mammals [Sinclair and Guarente, 1997, Ackermann et al., 2003, Hernebring et al., 2006].

Aging in budding yeast is characterized by increased generation time, increase in size, decline in mating ability [Müller, 1985, Kennedy et al., 1997], and switch to a hyperrecombinational state [McMurray and Gottschling, 2003]. Wild type yeast cells enter terminal senescence after 20-30 divisions [Jazwinski, 1993]. Nonetheless, the replicative life span of daughter cells depends only weakly on the age of the mother [Kennedy et al., 1994] thereby making a cell population immortal. Furthermore, when old cells are mated to young cells, the resultant diploids live for a number of generations most similar to the remaining life span of the older cells [Müller, 1985]. These observations led to the proposal that mother cells accumulate a "senescence factor" with age, which is prevented from transmission to the offspring [Kennedy et al., 1994].

Several factors, such as damaged proteins [Laun et al., 2001, McMurray and Gottschling, 2003], dysfunctional mitochondria [Lai et al., 2002], and extrachromosomal DNA [Sinclair and Guarente, 1997, Falcón and Aris, 2003], have been shown to contribute to aging in yeast. In the latter case, this was first demonstrated for extrachromosomal rDNA circles (ERCs) [Sinclair and Guarente, 1997]. Consistently, genetic manipulations that either increased or decreased the amount of ERCs shortened or lengthened life span [Sinclair et al., 1997, Defossez et al., 1999, Kaeberlein et al., 1999, Kim et al., 1999]. Later it was shown that the acceleration of aging is not limited to ERCs but is caused by the accumulation of plasmids without centromeres in general, especially by plasmids carrying an autonomously replicating sequence (ARS) [Falcón and Aris, 2003]. Importantly, both damaged proteins and ARS plasmids (without active partitioning mechanism) are retained in mother cells [Aguilaniu et al., 2003, Murray and Szostak, 1983].

Despite these insights into the physiological aspects of asymmetric plasmid segregation, it remained unresolved what causes the segregation bias. The volume ratio of the nucleus between mother and daughter has been shown to be approximately 3 to 2 [Gordon, 1977, Heun et al., 2001] and is therefore not able to account for the strong asymmetry in plasmid distribution. Two possibilities can be envisioned to explain this asymmetry. First, plasmids could associate with some site or structure that predominantly stays in the mother cell during mitotic division or are retained by a diffusion barrier at the neck. A recent study [Shcheprova et al., 2008] supports this model. The authors showed that a "non-centromeric" plasmid created by the excision of the centromeric sequence from a CEN-ARS plasmid is associated with nuclear pores. Furthermore, nuclear pores were shown to be retained in the mother cells due to a diffusion barrier at the bud neck thereby also retaining the bound plasmid.

The other possibility is that the mother-daughter segregation bias of ARS plasmids is intrinsic to the geometry of the dividing yeast cells. This mechanism relies on the asymmetric

morphology of the nucleus during division in budding yeast. The daughter nucleus bulges out of the mother driven by microtubule extension and stays connected with it through a thin tunnel until the two nuclei finally separate (Figure 4.1). At the beginning of nuclear division the plasmid naturally resides in the mother nucleus. If the plasmid diffuses freely and there is enough time for the system to equilibrate, then the plasmid will be located in the mother or in the daughter with frequencies proportional to their volumes. In contrast, if the division of the nucleus is too fast, the intrinsic asymmetry of the division will be transferred onto the plasmid distribution leading to an enrichment in the mother nucleus. In this case, the diffusive barrier that results from the shape of the dividing nucleus itself gives rise to the asymmetric plasmid segregation. The observation that 1.45 kb circles, which should diffuse faster due to their small size, do not show a strong segregation bias supports this view [Zakian and Kupfer, 1982].
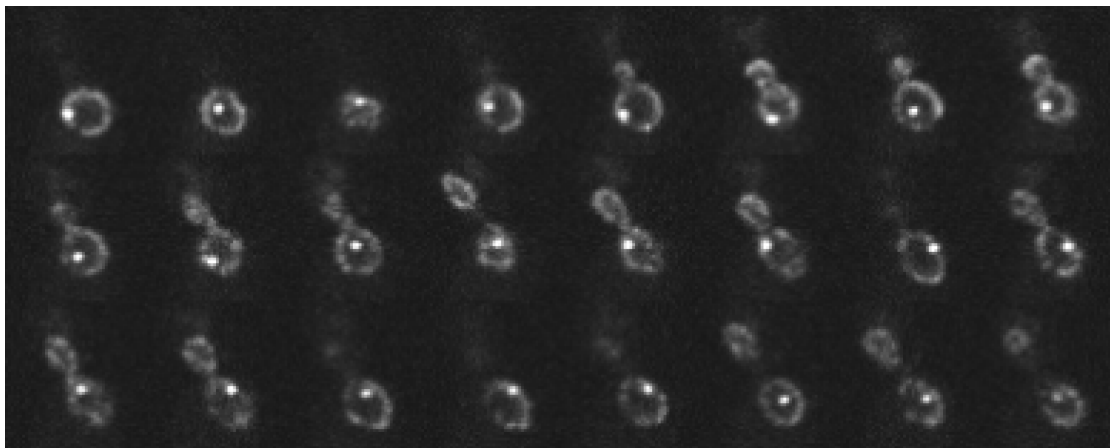
Figure 4.1: Representative time lapse image series of the excised *LYS2* ring during mitotic division. Image stacks were collected at 30 s intervals and projected to the *xy* plane.

In collaboration with SHIGEKI NAGAI, I have investigated the role of these two conflicting models in the asymmetric distribution of plasmids during cell division in budding yeast. All experimental work shown here was carried out by SHIGEKI, all theoretical modeling was developed and performed by me. Using live cell imaging, we could show that an extrachromosomal ring and an *ARS* plasmid move freely inside the nucleus, but nonetheless show a strong segregation bias. Computational simulations revealed that the shape of the dividing nucleus imposes a substantial constraint on the ring's mother-daughter transition by passive diffusion. This result is consistent with our experimental finding that the stability of *ARS* plasmids can be increased by extending the duration of mitosis. These data indicate that the mother-daughter segregation bias of *ARS* plasmids is, at least in part, mediated by the structural burden at the bud neck, which inhibits the efficient diffusion of *ARS* plasmids

into the daughter nucleus.

Our finding that the asymmetric distribution of *ARS* plasmids is caused by the geometry of the dividing nucleus raises the question of how symmetric segregation can be achieved mechanistically. Several ways to overcome the bias have been discovered including the targeting of the telomere anchoring factor Sir4PAD [Ansari and Gartenberg, 1997]. This opened the question of whether either the binding of Sir4PAD to the nuclear envelope or to telomeres is responsible for the stabilization of the plasmid. We have shown that tethering of an *ARS* plasmid either to the nuclear envelope or to telomeres overcomes the diffusional barrier. Here, I summarize these results and discuss the theoretical aspects of how peripheral anchoring can facilitate the escape from a confining volume.

4.3   Measuring the segregation bias

Before I present our results on the underlying mechanism for asymmetric plasmid segregation, I explain how the segregation bias can be measured in living cells. The two classical parameters are the fraction of plasmid-bearing cells under conditions selecting for the plasmid [Stinchcomb et al., 1979, Hsiao and Carbon, 1979] and the rate at which plasmid-free cells are generated under non-selective conditions [Fitzgerald-Hayes et al., 1982].

The first parameter, termed mitotic stability, can be determined by culturing cells under selective conditions followed by plating them on selective and non-selective conditions. It is important to know that cells which have lost the plasmid usually can still undergo several rounds of division under conditions selecting for the presence of the plasmid due to continued presence of the plasmid-encoded gene product [Murray and Szostak, 1983]. However, this is not sufficient to form a colony on plate.

The second parameter, the plasmid loss rate, can be derived from mitotic stability measurements. The fraction of plasmid-bearing cells is measured before and after a fixed number of cell cycles under non-selective conditions. At the beginning there are N cells. Let S_i be the number of plasmid-bearing cells after i divisions. Then $S_0 = s_0 N$ where s_0 is the measured fraction of plasmid-bearing cells at the start of the experiment. In each round of division each plasmid-bearing cell produces a plasmid-free daughter with a probability p (the plasmid loss rate). Thus, it produces two plasmid-bearing daughters with a probability of $1 - p$ and one plasmid-bearing daughter otherwise:

$$\begin{aligned} S_i &= S_{i-1} + S_{i-1} \cdot (1 - p) \\ &= S_{i-1} \cdot (2 - p) \\ &= s_0 N \cdot (2 - p)^i \end{aligned}$$

At the same time, the total number of cells doubles in each round of division. Therefore the

fraction s_i of plasmid-bearing cells can be calculated as

$$s_i = \frac{S_i}{2^i N}$$
$$= s_0 \left(\frac{2-p}{2} \right)^i$$
$$= s_0 \left(1 - \frac{p}{2} \right)^i$$

and the plasmid loss rate calculated after n divisions is:

$$p = 2 \left(1 - \sqrt[n]{\frac{s_n}{s_0}} \right). \tag{4.1}$$

4.4 Results

4.4.1 An extrachromosomal *LYS2* ring moves freely in the nucleus at mitosis, yet exhibits a strong bias to segregate to the mother cell

The mitotic stability and plasmid loss rate are population parameters of an entire yeast culture and therefore highly averaged quantities. Moreover, they depend not only on the segregation efficiency but also on the average copy number of the plasmid, the degradation of plasmids, and the life span of plasmid-free cells under selective conditions. The fundamental parameter is the probability that one particular plasmid reaches the daughter nucleus during nuclear division.

In order to measure this transfer probability directly, we examined the dynamics of an extrachromosomal ring in the nucleus at mitosis using live cell imaging. In this system, an inducible site-specific recombination event gives rise to a 16.5 kb ring bearing a fragment of the *LYS2* gene and an array of lac operators which bind a coexpressed lac repressor-GFP fusion [Gartenberg et al., 2004, Gehlen et al., *in preparation*]. This allowed us to observe the ring using fluorescence microscopy. Importantly, the ring does not have functional origins of DNA replication and is therefore rarely replicated. As a consequence, most cells contain not more than one ring, which allowed us to follow it under the microscope to determine whether it crossed over to the daughter nucleus or not.

In order to examine if the excised ring shows a mother-daughter segregation bias at mitosis, we scored the presence of the ring in the mother and daughter nuclei in G1 phase of the cell cycle. The mother nucleus can be distinguished from its daughter by its larger size. We found that 90% of the excised rings stay in the mother, whereas only 10% of them segregate to the daughter cell (Figure 4.2), showing that the excised ring exhibits a strong segregation bias towards the mother nucleus at mitosis.
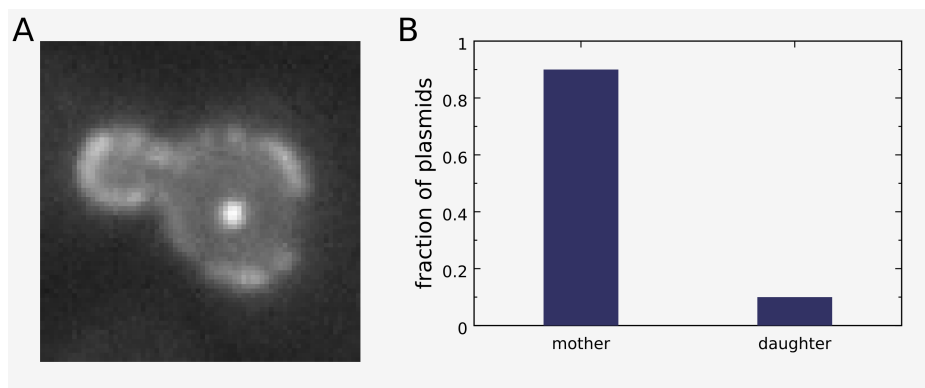
Figure 4.2: The excised *LYS2* ring shows a strong segregation bias. (A) Representative image of a dividing nucleus with an excised ring. (B) In 90% of the cases the ring is found in the mother nucleus after mitosis.

To monitor the dynamics of the ring, three dimensional (3D) confocal stacks were collected on living cells during mitosis at an interval of 30 s. Visual inspection of the movies showed that the excised ring is highly mobile at mitosis (Figure 4.1). Moreover, a mean squared displacement (MSD) analysis of the movement in G1 phase confirmed that the dynamics of the ring show the characteristics of a freely diffusing particle in a spherical volume with a radius of constraint of $\sim 0.8\,\mu\text{m}$ (Figure 4.3, see also section 2.2.3). This value is close to the average nuclear radius of haploid yeast cells ($\sim 0.98\,\mu\text{m}$).
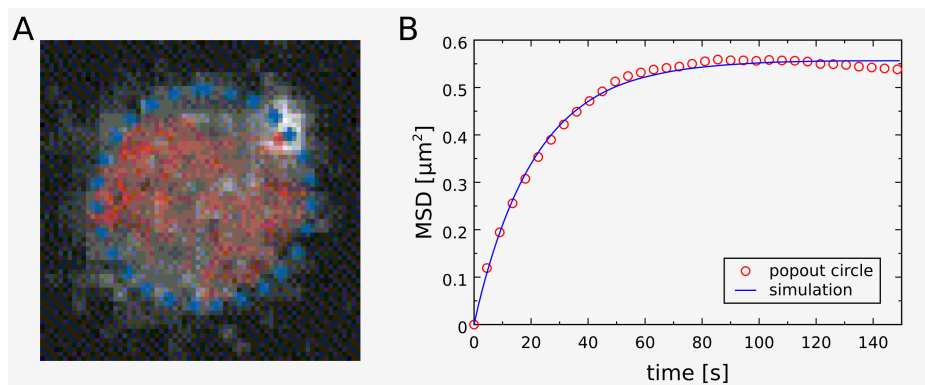
Figure 4.3: The excised *LYS2* ring is highly mobile [Gartenberg et al., 2004, Gehlen et al., *in preparation*]. (A) Representative trajectory of a ring in G1. (B) The movement of the ring shows the same properties as a random walk inside a spherical volume.

Considering the reported observation that an *ARS* plasmid is associated with nuclear

pores [Shcheprova et al., 2008], we also measured the distance between the nuclear envelope and the excised ring at mitosis (Figure 4.4). A comparison to the expected distribution of a uniformly distributed particle, which I calculated using the computational model described below, indicates that the ring is not associated with the nuclear envelope.
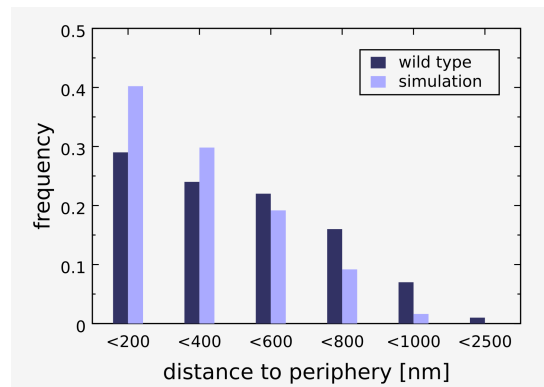
Figure 4.4: Distribution of the distance between the excised *LYS2* ring and the NE. The ring is even further from the periphery than expected for a uniformly distributed spot.

In order to rule out the possibility that the excised ring behaves differently from an autonomously replicating plasmid, we corroborated our findings with the inspection of an *ARS* plasmid[1] carrying a lac operator array. Although the measurement of cells with multiple plasmids is difficult and it was hard to collect enough mitotic cells with few plasmids to perform a statistical analysis, our results suggest that the dynamic movement is not a special property of the excised ring, but a more general feature of extrachromosomal circles [Gehlen et al., *in preparation*].

Taken together, our results show that plasmids can move freely in the yeast nucleus at mitosis and nevertheless show a strong segregation bias. This implies that the binding to a structure in the mother is not required for an asymmetric distribution of an extrachromosomal DNA ring.

4.4.2 Passive diffusion and the geometric constraint of the dividing nucleus are sufficient to induce a segregation bias

Our data argue that the binding of an excised ring to an immobile structure in the mother nucleus is not required for a strong segregation bias. This led us to investigate the possibility

---

[1]In this study, we used the pAA4 plasmid which contains the *URA3* gene, the 2 μm *ARS*, and an array of lexA operators [Ansari and Gartenberg, 1997]. For this experiment, the plasmid was modified by the insertion of lac operators, which allowed us to visualize the position of the plasmid using a GFP-lac repressor fusion.

that the geometric constraint which the shape of the dividing nucleus imposes on the plasmid diffusion effectively obstructs the transition to the daughter nucleus (see introduction).

I developed a computational model to investigate the effects of geometry and duration of the nuclear division on the diffusion-based plasmid distribution. The plasmid is modeled as a particle diffusing freely inside the nucleus, which has a spherical shape at the beginning. The daughter nucleus then bulges out of the mother with constant volume rate (Figure 4.5).
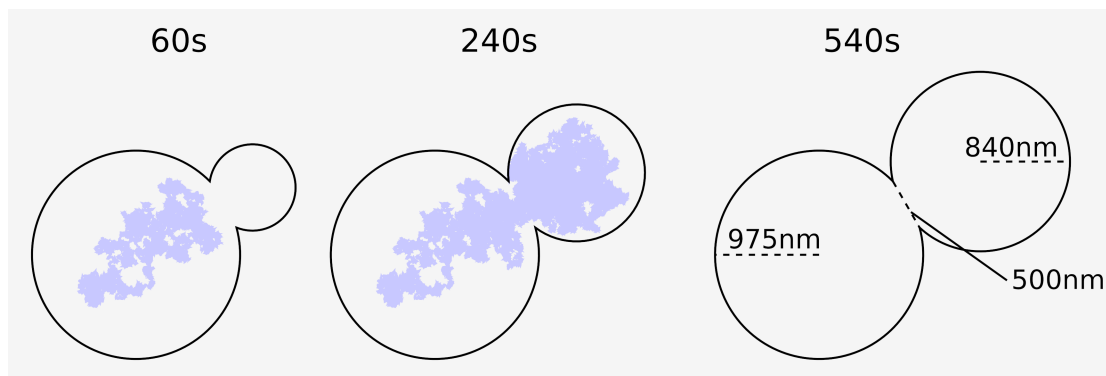
Figure 4.5: Model of the dividing budding yeast nucleus. The daughter bulges out of the mother during 9 min with constant volume rate.

The model uses the following parameters:

- Step size: The step size has to be small compared to the typical length scale of the confinement which is in the order of 100 nm. Between step size values of 10 nm and 5 nm there is still a detectable difference in the results but not for a further decrease beyond 5 nm. I used a value of 1 nm.

- Time step: The number of steps per second depends on the step size. I applied our tracking data shown in Figure 4.3B to calibrate the simulation against experimental data. In order to do that, I fitted function (2.16) to the mean squared displacement (MSD) of the excised ring:

$$\langle d_1^2(\tau) \rangle = a \left(1 - e^{-\tau/b_1} \right).$$

I then simulated a random walk inside a spherical volume with the radius derived from a (see equation (2.22)) and fitted

$$\langle d_2^2(n) \rangle = a \left(1 - e^{-n/b_2} \right).$$

to the MSD of the walk. $\frac{b_2}{b_1}$ gives the desired number of steps per second. For the step size of 1 nm, I used 44602 s^{-1}.

- Duration of mitosis: The duration of different sections of the budding yeast cell cycle has been measured by [Lord and Wheals, 1981]. The time between the relocalization of the nucleus to the bud isthmus and the separation into two nuclei was measured to (8.8 ± 0.4) min. I used a value of 9 min.

- Radii of mother and daughter: The radii of mother and daughter nuclei directly after mitosis were measured in [Heun et al., 2001] to $R_{\text{mother}} = 975$ nm and $R_{\text{daughter}} = 840$ nm. In the model the daughter grows with constant volume rate. Although this assumption was not verified experimentally, it is unlikely to radically impact the simulation results.

- Final mother-daughter distance: The center-to-center distance of mother and daughter nucleus grows with constant speed. As for the volume growth rate, this was the simplest assumption I could make. I measured the final distance[2] in movies taken by SHIGEKI NAGAI. However, it is difficult to define the end of mitosis precisely and the final distances varied between 2200 nm and 4200 nm. This problem is alleviated by my finding that the results depend much less on the final mother-to-daughter distance than on the neck diameter (data not shown). I used a value of 2500 nm.

- The neck diameter: Obviously, the diameter of the neck between mother and daughter is crucial for the model. Moreover, it is very hard to determine. Because we use fluorescent labeling of nuclear pores to visualize the periphery and because there are almost no pores in the neck, it is impossible to measure the neck diameter from our fluorescent images. To my knowledge, no systematic study of this parameter using different staining or imaging techniques has been carried out so far. As part of this study, I have performed a screen over varying neck diameters to corroborate the results (see below). However, I estimated a default value for the diameter from electron micrographs published in [Gordon, 1977]. I measured a value of 450 nm and used 500 nm as default value for my simulations.

After equilibration, the plasmid would be found in the daughter nucleus in 39% of the cases, reflecting the volume difference of daughter and mother. In my simulation, on the other hand, I found that it reaches the daughter only with a frequency of 24% by 9 min. The system reaches equilibrium after about 25-30 min (Figure 4.6). This result implies that the mother-daughter transition of the plasmid is detectably obstructed solely by the geometric constraint of the dividing nucleus.

---

[2]To be precise: I measured the total extension of the dividing nucleus because this avoids the error-prone determination of the centers of mother and daughter. From this, I calculated the center-to-center distance by subtracting the two radii.
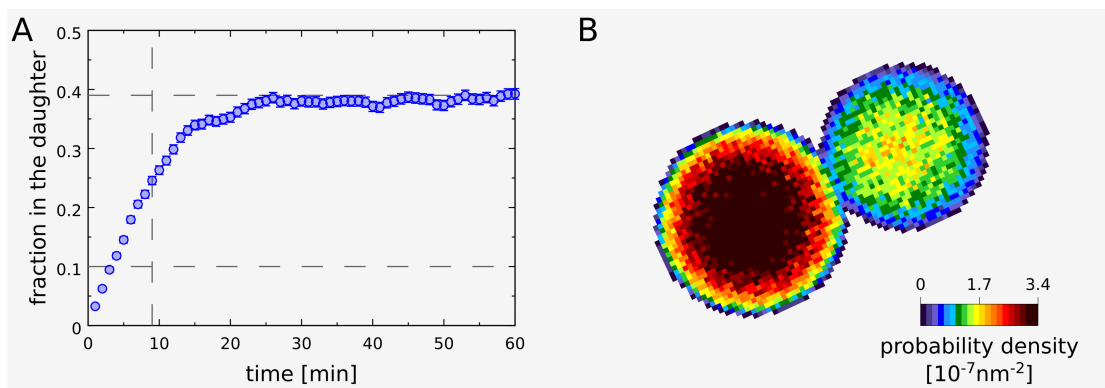
Figure 4.6: (A) The fraction of plasmids in the daughter nucleus over time, derived from 5000 simulations. The system does not equilibrate in 9 min but only reaches a fraction of 24% in the daughter nucleus. The fraction in the daughter at equilibrium (39%) and in the experiment (10%) are shown in comparison. (B) Probability density of the plasmid's position at the end of mitosis, calculated from 200000 simulations and projected onto the xy plane.

4.4.3 A refined model together with molecular crowding in the neck can, at least in part, explain the segregation bias

Despite the result that the geometry of the nucleus imposes a notable constraint onto diffusion into the daughter nucleus, the transition probability of 24% is still substantially higher than the frequency of 10% observed experimentally (Figure 4.2B). But the result that the diffusive system does not reach equilibrium implies that the exact shape of the dividing nucleus plays an important role in the extent of the segregation bias. I therefore refined our model by introducing a tunnel between mother and daughter nucleus to better mimic the shape of the nucleus during division (Figure 4.7). As expected, the transition probability drops further to 15% (equilibrium: 39%) and the system needs much longer to equilibrate (Figure 4.8).

A high concentration of proteins and other molecules in the tunnel can increase the viscosity of the nucleoplasm and thereby slow down plasmid diffusion. I therefore investigated to which extent an increased viscosity in the neck would contribute to the segregation bias. In general, the effect is very little. Only a dramatic decrease of the plasmid's diffusion coefficient in the neck exerts a visible change on the segregation bias (Figure 4.9A). Moreover, the effect depends radically on the movement of the solvent in the tunnel. I investigated three possible scenarios. If the solvent does not move, then an increased viscosity eventually leads to a retention of the plasmid, as expected (Figure 4.9A, blue curve). However, if the solvent in the tunnel moves as fast as the daughter nucleus, a high viscosity leads to a stabilization of the plasmid (Figure 4.9A, red curve). This is plausible because the plasmid gets stuck

in the high viscosity medium if it enters the tunnel. If this happens early in mitosis, the plasmid is dragged to the other side by the medium. In an intermediate scenario, in which the velocity of the medium is interpolated continuously between mother and daughter, the result is essentially the same as in the no-drift scenario.
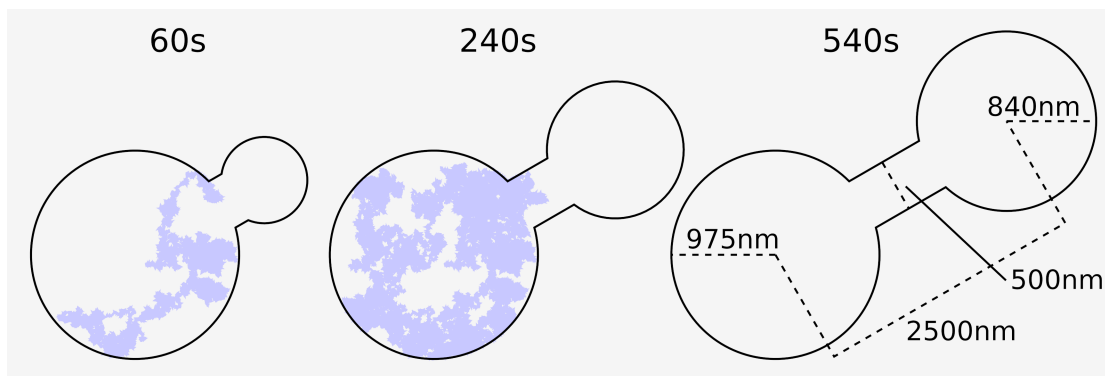
Figure 4.7: Refined model of the dividing budding yeast nucleus. The extended thin section between mother and daughter nucleus is modeled as a tunnel.
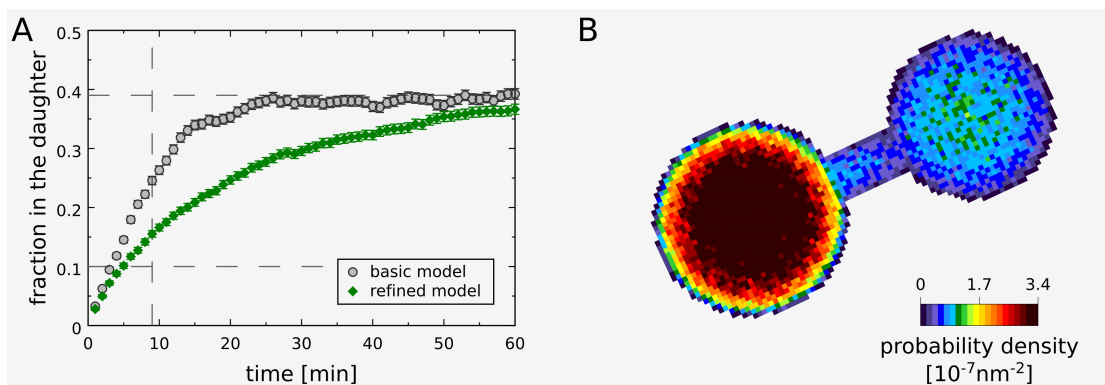
Figure 4.8: (A) The fraction of plasmids in the daughter nucleus over time for both models, derived from 5000 simulations each. In the refined model, the system does not even equilibrate fully in 60 min. In 9 min, it only reaches a fraction of 15% in the daughter nucleus. (B) Probability density of the plasmid's position at the end of mitosis, calculated from 200000 simulations and projected onto the xy plane.

Since the velocity profile of the medium in the tunnel is unknown, it is difficult to conclude how high viscosity in the tunnel would influence the segregation bias. Importantly, the flow of the medium has no impact on the plasmid distribution for all realistic values of the solvent viscosity.

Not only increased viscosity could impair the plasmid transition from mother to daughter. For example, the presence of chromosomes and microtubules in the tunnel reduces the cross section which is accessible to the plasmid. Therefore, I ran a set of simulations to find out how far the effective tunnel diameter has to be reduced in order to retain the plasmid as efficiently as observed in the experiment. This analysis yielded the refined model with a tunnel diameter reduced to 160 nm, corresponding to 40% of the total cross section (Figure 4.9B).
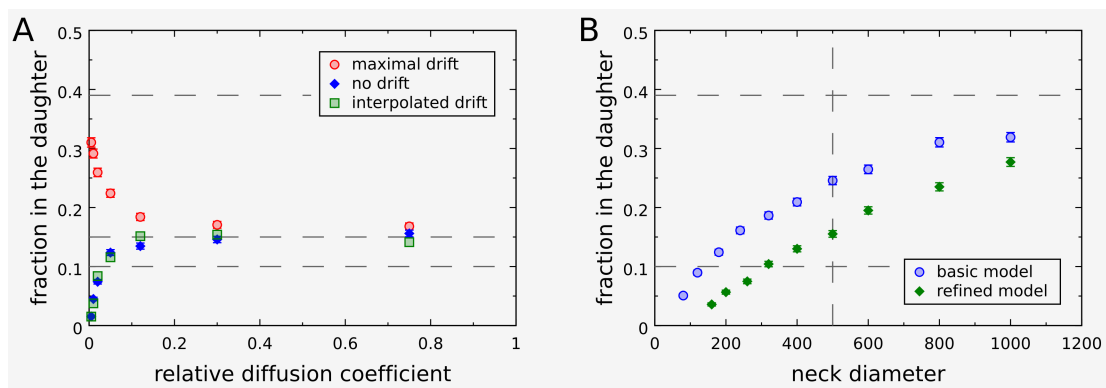
Figure 4.9: (A) Effect of increased solvent viscosity inside the tunnel. The fraction in the daughter at the end of mitosis is shown in dependency of the plasmid's diffusion coefficient in the tunnel normalized to the default diffusion coefficient. The fraction in the daughter at equilibrium (0.39), for the default viscosity (0.15), and in the experiment (0.10) are shown in comparison. The viscosity in the neck shows only an effect for extremely high viscosity values. In this case, the outcome depends crucially on the chosen model for the flow of the solvent in the tunnel (see text). (B) Effect of the tunnel diameter on the segregation bias for both models. The fraction in the daughter at equilibrium (0.39) and in the experiment (0.10) are shown in comparison. In the refined model, a reduction of the diameter to 320 nm (corresponding to 40% of the default cross section) reproduces the experimental segregation bias.

4.4.4 Extension of mitotic duration improves plasmid partitioning

The result that the diffusive system of the dividing nucleus does not reach its equilibrium before the nuclei separate predicts that a prolongation of the division would improve plasmid segregation into the daughter cell. To test this hypothesis, we measured the stability of the *ARS* plasmid in a mutant with an extended mitotic duration and at a lower temperature which also prolongs cell division.

We focused on the *yku70Δ* mutant, in which the duration of the G2/M phase is extended due to the checkpoint activation caused by damaged telomeres [Maringele and Lydall, 2002].

The G2/M delay in the *yku70Δ* mutant is largely suppressed by the deletion of the *CHK1* gene [Maringele and Lydall, 2002], which plays an important role in the G2/M checkpoint in budding yeast [Heideker et al., 2007]. We measured mitotic stability and plasmid loss rate of the pAA4 plasmid (see footnote on page 103) in wild type, a *yku70Δ* mutant, and a *yku70Δchk1Δ* double mutant. As predicted, the deletion of *YKU70* improves the persistence of the plasmid significantly: the mitotic stability was improved from 21% to 31.5% (Figure 4.10A), and the plasmid loss rate dropped from 0.3 to 0.15 [Gehlen et al., *in preparation*]. Moreover, simultaneous deletion of *YKU70* and *CHK1* reverses this effect: both the mitotic stability (18.8%) and the plasmid loss rate (0.27) show values comparable to those of wild type cells. Therefore, the effect does not result from the loss of yKu70 *per se*. Although we cannot completely exclude the possibility that this behavior results from effects other than G2/M delay, these data strongly suggest that the extension of G2/M phase improves the diffusion of the *ARS* plasmids into the daughter nucleus, confirming the theoretical prediction.

To further support the observations in the *yku70Δ* mutant, we measured the mitotic stability of the plasmid at different temperatures. It is known that the cell cycle is extended by shifting yeast cells to lower temperature. The duration of the combined G2/M phase doubles under a shift from 30 °C to 20 °C [Vanoni et al., 1984]. However, at the same time also diffusion slows down at decreased temperature which should destabilize the plasmid. Under the assumption of the STOKES-EINSTEIN-relation [Coffey et al., 2004], the diffusion coefficient D is proportional to T/η where T is the absolute temperature and η the dynamic viscosity of the solvent. From this follows: $\frac{D(293\,K)}{D(303\,K)} = \frac{293\,K\cdot\eta(303\,K)}{303\,K\cdot\eta(293\,K)}$. Assuming that the temperature dependence of the diffusion coefficient of the nucleoplasm is similar to that of water, this results in $D(293\,K) = 0.77 D(303\,K)$ [Aleksandrov and Trakhtengerts, 1974]. In my simulations, this decreases the fraction of plasmids in the daughter nucleus only from 15.3% to 14.2% (Figure 4.10C). An accompanying increase of mitotic duration to 18 min leads to a transfer probability of the plasmid of 20.1% (Figure 4.10D). Despite this rather moderate increase, these results suggest that the prolongation of mitosis is the predominant consequence of the temperature shift and that a decrease in temperature might stabilize a plasmid.

We measured the mitotic stability of pAA4 in wild type at 30 °C and 20 °C. Although the mitotic stability is lower in general in this experiment, we found that, consistent with the prediction, it is higher at 20 °C than at 30 °C (20.7% versus 12.3%, Figure 4.10B).

4.4.5 Anchoring to the nuclear envelope improves plasmid partitioning

Our results revealed that no active molecular mechanism is required to explain the asymmetric distribution of *ARS* plasmids. This raises the question of which mechanisms can overcome the bias and lead to a symmetric distribution. It was shown in [Ansari and Gartenberg, 1997]

that targeting a specific domain of Sir4, the partitioning and anchoring domain (Sir4PAD), directly to *ARS* plasmids improves their mitotic segregation. Sir4 is involved in anchoring telomeres to the nuclear envelope (NE), and it could therefore link the plasmid either to telomeres or to the NE. The Sir4PAD-mediated enhancement of *ARS* plasmid segregation requires the nuclear periphery protein Esc1 [Andrulis et al., 2002], which is also bound by Sir4PAD. This observation led us to hypothesize that anchoring to the NE is responsible for the Sir4PAD-mediated improvement of mitotic segregation. We targeted a fusion of lexA with the integral membrane protein Yif1 [Andrulis et al., 1998] to the excised ring in order to test this hypothesis.
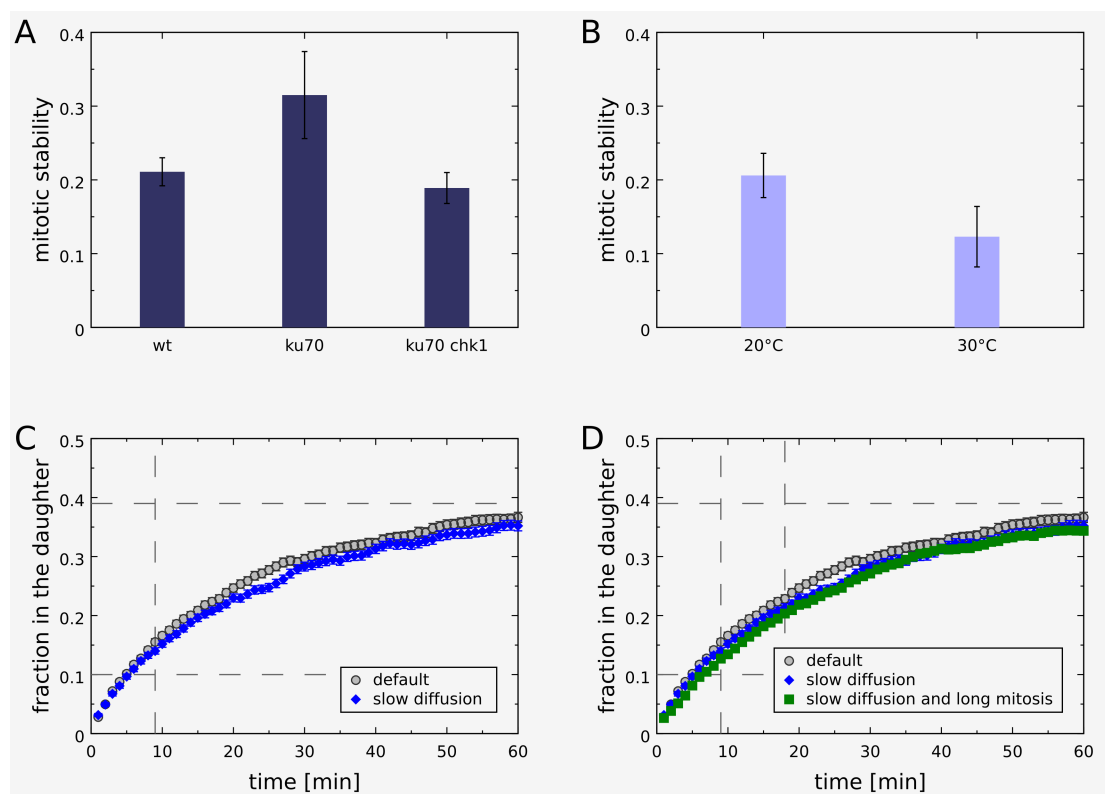
Figure 4.10: Effects of prolonged mitosis on the segregation bias. (A) Effect of *yku70Δ* and *yku70Δchk1Δ* deletions on the mitotic stability of the pAA4 plasmid. The values were calculated as averages of six experiments with independent transformants. (B) Effect of different temperatures on the mitotic stability of the pAA4 plasmid in wild type. (C) Effect of slower diffusion at 20 °C. The fraction of plasmids in the daughter nucleus after 9 min drops from 15.3% to 14.2%. (D) Combined effect of slower diffusion and slower mitosis at 20 °C. All three curves are very similar. The important impact of lowered temperature is that mitosis takes longer. After 18 min, 20.1% of the plasmids are in the daughter nucleus.

We have shown in a previous study that targeting of a lexA-Yif1 fusion relocalizes an internal GFP-tagged chromosomal locus to the nuclear periphery in interphase [Taddei et al., 2004]. We showed that targeting of lexA-Yif1 results in a relocalization of the excised ring to the periphery at mitosis, compared to targeting of lexA alone (Figure 4.11A).
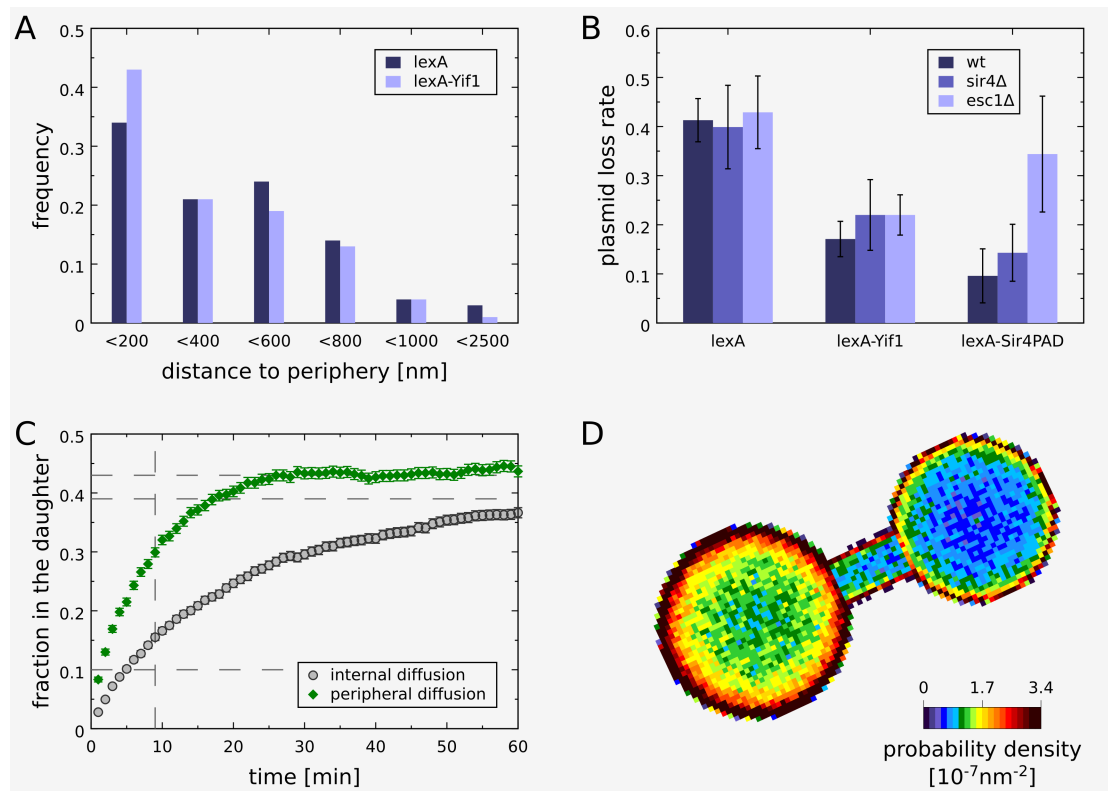
Figure 4.11: Effects of anchoring to the NE on the segregation bias. (A) The lexA-Yif1-targeted ring is more frequently close to the periphery than a ring targeted with lexA alone. (B) Effect of lexA-Yif1-targeting on the plasmid loss rate of pAA4. The values were calculated as averages of six experiments with independent transformants. (C) Effect of restriction of plasmid movement to the surface of the confining volume. The fraction of plasmids in the daughter nucleus after 9 min increases from 15.3% to 29.9%. The fraction in equilibrium is also increased (43% instead of 39%) because the surface ratio of daughter and mother is larger than their volume ratio. (D) Projected probability density of the plasmid's position at the end of mitosis.

To test whether anchoring to the NE improves plasmid partitioning, lexA-Yif1 was also targeted to the pAA4 plasmid, which contains four adjacent lexA sites. As shown before [Ansari and Gartenberg, 1997], targeting of Sir4PAD stabilizes pAA4 in wild type cells in a Esc1 dependent manner. We found that lexA-Yif1 also stabilizes the plasmid (Figure 4.11B).

To confirm that this stabilization is due to the binding of Yif1 to the plasmid DNA and not to its mere presence in the nucleus, the stability of a plasmid with no lexA binding sites (pAA0) was measured in the presence of lexA-Yif1. Expression of lexA-Yif1 did not stabilize pAA0 compared to lexA alone [Gehlen et al., *in preparation*].

An improvement in plasmid stability could result not only from improved plasmid segregation, but also from enhanced replication efficiency. To rule out this second possibility, we determined the average copy number of the pAA4 plasmid per plasmid-bearing cell by southern hybridization. An increase in replication efficiency would lead to an increase in plasmid copy number per plasmid-bearing cell, whereas an improvement of plasmid segregation would be accompanied by a decrease in plasmid copy number. We found that cells expressing lexA alone have 20.6 plasmids per plasmid-bearing cell, while cells expressing lexA-Yif1 have only 5.4. Thus, we conclude that the improvement of plasmid stability by Yif1 results from an improvement of plasmid segregation. The fact that this improvement is independent of Esc1 and Sir4 (Figure 4.11B) reinforces our conclusion that physical anchoring of the plasmid to the NE is sufficient to improve plasmid segregation. This conclusion is in agreement with the observation that inner nuclear membrane proteins are able to diffuse into the daughter nucleus during mitosis [Shcheprova et al., 2008].

Our findings raise the question of how binding to the NE can stabilize a plasmid. The mathematical theory of narrow escape [Singer et al., 2006] indeed predicts that the time a diffusing particle needs to escape from a spherical volume through a small circular window is shorter for diffusion along the periphery than for internal diffusion. This result requires that the window is very small compared to the size of the sphere which is fulfilled only approximately in our case. I simulated the plasmid segregation with the same model as before but with the restriction that the plasmid moved on the periphery of the dividing nucleus. I found that the fraction of plasmids in the daughter after mitosis is indeed considerably bigger (30% compared to 15%) than for internal diffusion (Figure 4.11).

4.5   Discussion

In this study, we found that the segregation bias of *ARS* plasmids can be explained by the geometric constraint that the dividing nucleus imposes on plasmid diffusion, together with the limited duration of mitosis. Furthermore, we showed that the binding of the plasmid to the nuclear envelope improves its partitioning. This can be explained by the mathematical result that movement on the periphery of a confining volume reduces the escape time from this volume.

We used time lapse microscopy to show that an excised chromatin ring moves freely inside the nucleus and nevertheless shows a strong segregation bias. This finding supports the hypothesis that binding to a nuclear structure is not required for asymmetric plasmid distribution. I developed a computational model to study the transition of the plasmid from mother to daughter based on unobstructed diffusion in the dividing nucleus. With

this model, I could show that the diffusional barrier created by the shape of the dividing nucleus effectively inhibits the mother-daughter transition of a plasmid within the limited time of mitosis. The model predicts that both a widening of the neck and a prolongation of mitosis would improve plasmid partitioning. We confirmed the case of prolonged mitosis by using a *yku70Δ* mutant and by lowering the temperature. Additionally, we confirmed the recent result of [Shcheprova et al., 2008] that the deletion of *BUD6* increases plasmid stability (data not shown). Bud6 is located at the cellular bud neck during late mitosis, and the deletion leads to a thicker neck [Amberg et al., 1997]. Although we do not know if this also causes a widening of the nuclear neck during the division of the nucleus, this effect would indeed cause an increase in plasmid stability according to our model. Furthermore, the deletion of *BUD6* causes a decrease in growth rate and an accumulation of cells in late mitosis [Amberg et al., 1997]. We propose that the stabilizing effect of *BUD6* deletion is a result of prolonged mitosis and possibly a widening of the neck of the dividing nucleus.

[Shcheprova et al., 2008] reported that an *ARS* plasmid is associated with nuclear pores. They further showed that nuclear pores stay in the mother due to a diffusion barrier at the bud neck and proposed that this barrier prevents the transition of an *ARS* plasmid to the daughter nucleus. We found, however, that an excised chromatin ring moves freely inside the nucleus and nonetheless shows a strongly asymmetric distribution. The contradiction might stem from the different systems used in [Shcheprova et al., 2008] and our study. The authors used a system in which excision of the centromeric sequence from a *CEN-ARS* plasmid produces a "non-centromeric" plasmid. Since centromeres in budding yeast have been shown to interact with the spindle pole body, which is embedded in the nuclear envelope [Jaspersen and Winey, 2004], it is possible that an epigenetic mechanism retains the "non-centromeric plasmid" at the nuclear envelope even after the *CEN* element is excised. Accordingly, in budding yeast, the centromere-specific histone H3 variant Cse4 is present in a 20 kb domain surrounding the 120 bp *CEN* sequence in G2/M phase of the cell cycle [Riedel et al., 2006]. It is possible that Cse4 is still present after excision of *CEN* sequences from the plasmid, keeping it at the periphery.

In spite of other mechanisms that could act on top, we have shown that the retention of *ARS* plasmids in the mother nucleus already occurs passively due to the geometric constraints that the mitotic nucleus imposes on plasmid diffusion. The model predicts a transition probability of 15% for the excised DNA ring while we observed an experimental value of 10%. Two factors could explain this difference. First, there could be an active gate keeping mechanism at the neck between mother and daughter nucleus that specifically retains the plasmid. Although the model by [Shcheprova et al., 2008] supports this possibility, this model relies on a diffusion barrier at the nuclear envelope. This mechanism is less plausible for plasmids that are not associated with the nuclear envelope, such as our *LYS2* ring.

Alternatively, the difference could be the result of a reduced effective cross section of the neck due to the presence of chromosomes and microtubules. Furthermore, the tunnel between mother and daughter thins at the end of mitosis until the two nuclei separate. This

process is not included in our model. My simulations show that with a default neck diameter of 500 nm the available cross section has to be reduced to 40%. If 32 chromatid arms are present in the neck this means that each chromatid arm together with its microtubule would have an effective diameter (i.e. the diameter of a cylinder with the same cross section) of about 70 nm. This highly simplified calculation suggests that the crowding of chromatid arms and microtubules in the neck can explain the discrepancy between our theoretical modeling and our experimental results. A more detailed analysis would have to make assumptions about the exact position of the chromosomes and microtubules in the neck. Our results are supported by the fact that a small 1.45 kb *TRP1* circle does not show a segregation bias [Zakian and Scott, 1982], suggesting that the diffusion coefficient and the available space at the bud neck are important for plasmid transition.

A further prediction of our model is that, since there is no specific retention of the plasmid, it should be stabilized by association with a structure that partitions well between mother and daughter. Indeed, plasmids bearing a centromeric sequence are present in 90% of the cells [Clarke and Carbon, 1980]. It has been speculated that also the binding to telomeres might stabilize an *ARS* plasmid [Ansari and Gartenberg, 1997]. Within this project, we could show that targeting of the Ku80 protein associates a plasmid with telomeres and indeed increases plasmid stability, confirming this hypothesis.

It has been proposed that cell polarity might have arisen originally as a solution to cope with the accumulation of damaged proteins and DNA, which are caused by vital cellular activities [Partridge and Barton, 1993, Ackermann et al., 2007, Macara and Mili, 2008]. Polarized cell division enables the rejuvenation of one daughter cell, at the expense of the other, which takes over accumulated material. While many systems appear to rely on active molecular mechanisms to ensure that damaged material is prevented from being transmitted to offspring [Aguilaniu et al., 2003, Macara and Mili, 2008], our results suggest that asymmetric morphology of mitosis could potentially contribute to rejuvenation by imposing physical constraints on the diffusion of damaged material into the daughter. Thus, polarized mitotic morphology of budding yeast might have arisen from the need for rejuvenation during evolution.

Our results raise the possibility that the neck geometry could also contribute to the asymmetric segregation of other damaged material. Interestingly, [McMurray and Gottschling, 2003] suggested the possibility that a centromere-less fragment (acentric fragment) that can be created in the presence of unrepaired double-strand breaks might be preferentially retained in the mother cell when it is completely separated from the rest of the chromosome. It was proposed that this mechanism could account for the daughter-biased occurrence of age-related loss of heterozygosity (LOH) in yeast. If the acentric fragment is preferentially retained in the mother, the daughter will repair the broken fragment by break-induced replication, whereas the mother cell could repair the broken fragment via non-homologous end-joining or homologous recombination. This would result in daughter-specific LOH.

Another possibility is that the physical constraint at the neck might also impose a barrier

on passive diffusion of proteins into the daughter nucleus. This is an attractive option since it could ensure that most of these proteins are synthesized *de novo* in the daughter, while old and damaged proteins are predominantly retained in the mother nucleus. However, it has to be noted that proteins diffuse much faster in the nucleus than a plasmid. Computational simulations using our model, in combination with live imaging of single molecule proteins, could be used to shed light on this possibility.

Our results suggest that the asymmetric division of the budding yeast nucleus together with the limited duration of mitosis provides the cell with a powerful mechanism to retain detrimental material in the mother nucleus. They demonstrate on the other hand that passive diffusion is already on the length scale of the yeast nucleus not an efficient way to segregate objects to a distant compartment. We showed that binding to the nuclear envelope is one way to alleviate this effect. However, in larger systems and especially if they bear inherent asymmetry, active transport is indispensable to ensure the proper distribution of cellular objects.

Chapter 5

Homology Search

5.1 Summary

DNA double-strand breaks are the most deleterious DNA lesions (see section 1.3). Homologous recombination uses a homologous template to repair a double-strand break accurately and is very efficient in yeast. However, the process by which the break and template site find each other within the vast quantity of non-homologous DNA is not well understood. We have developed a combined experimental and theoretical approach to study homology search and its relationship to nuclear organization in yeast. In this chapter, I introduce our strategy and present some first theoretical results that proof the concept of the approach.

5.2 Introduction

In order to repair a DNA double-strand break (DSB) via homologous recombination (HR), the break and template site have to colocalize and form a HOLLIDAY junction. The formation and resolution of HOLLIDAY junctions during HR have been studied extensively [Nassif et al., 1994, Szostak et al., 1983]. In contrast, the preceding process through which the two homologous loci get together is ignored in most models. Conceivable mechanisms that can lead to the colocalization of a DSB site with its template for HR include a linear scan of the genome, 3D diffusion with repeated unspecific binding of randomly selected loci, and a combination of these two. It has been shown that such a combination of 1D and 3D diffusion can significantly reduce the time needed by a DNA binding protein to find its binding site [Merlitz et al., 2006].

In yeast, DSB repair by interchromosomal ectopic recombination occurs with high efficiency (>90%) within 2 hours after DSB formation [Aylon et al., 2003]. A simplified calculation suggests that a linear scan of the genome within that time would have to be exerted at a unrealistically high speed compared to the processivity of DNA polymerase [Barzel and Kupiec, 2008]. This observation has led to the hypothesis that homologous

sequences might be aligned to each other even before a break occurs.

However, quantitative predictions for the duration of homology search and the influence of the spatial organization of the genome are difficult. To assess different models for homology search in the light of experimental results, it would be very helpful to know how long the search process would take based on different diffusion models (1D, 3D, a combination thereof) and how this time is influenced e.g. by anchoring break or template site to the nuclear periphery or by changing their mobility.

In collaboration with Vincent Dion, I am addressing these questions using a combined experimental and theoretical approach. In the following two sections, I introduce the experimental system designed by Haico van Attikum, Benjamin Towbin, and Vincent Dion and my computational model.

5.3   Experimental approach to study homology search

DNA double-strand breaks can be induced artificially by exposing cells to ionizing radiation, DNA damaging agents, or specific endonucleases [Feuerhahn and Egly, 2008, Lundin et al., 2005, Koç et al., 2004]. In the first two cases, one cannot predict where in the genome the damage will occur. Using lasers, the damage can be restricted to certain regions of the nucleus, but still it is not possible to induce damage at a given genomic location. This specificity can be achieved, however, by the use of site specific endonucleases, which recognize a specific sequence in the genome and induce a double-strand break (DSB) [Feuerhahn and Egly, 2008].

The common way to induce site-specific DSBs in yeast is to use the HO endonuclease, which plays an essential role in the mating type switching system of budding yeast [Haber, 1998]. While this endogenous enzyme is highly specific and efficient in its cleavage pattern, one risks that this system shows specific repair characteristics due to its endogenous function in mating type switching. Therefore, we use a different endonuclease, I-SceI whose recognition site is not present in the wild type yeast genome, but which has also been successfully used to induce DSBs in yeast [Haber, 1995].

Our system consists of two cassettes, which can be integrated at different positions in the genome. One cassette contains an I-SceI cut site and an array of lac operator sequences. The other one contains a template sequence for HR, a stretch of lexA operators, and an array of tet operators (see Figure 5.1). The I-SceI protein is expressed under control of a *GAL1-10* galactose-induced promoter[1]. Therefore, the expression can be triggered by shifting the cells to media containing galactose. Break and template site can be visualized microscopically in different ways (see section 2.1). A CFP-lacI fusion binds the lac operator sequences. Additionally, the break site can be visualized by the presence of a Rad52-GFP fusion. Importantly, Rad52 only forms a spot when the break has been induced [Lisby et al.,

---

[1]http://www.yeastgenome.org/cgi-bin/locus.fpl?locus=GAL1

2001]. Therefore, a colocalization of the CFP and GFP signals indicate the induction of a break. The tet operator sequences are bound by an mCherry-TetR fusion.
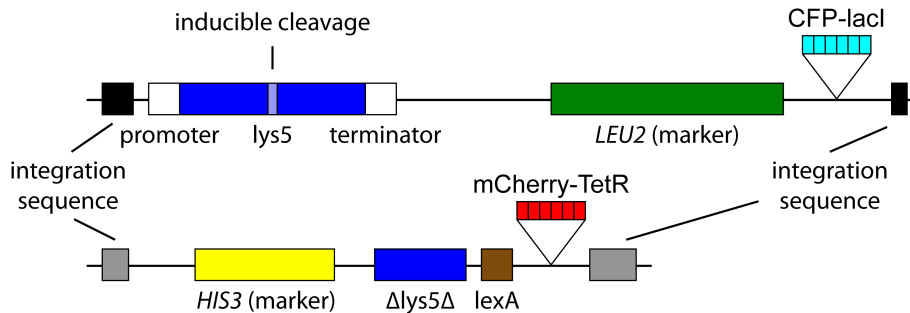
Figure 5.1: Integration cassettes to study homology search *in vivo*. One cassette contains a copy of the *LYS5* gene with a recognition site for the I-SceI endonuclease. Next to it is a sequence of lac operators, which can be bound by a CFP-lacI fusion. A *LEU2* marker is used to check for the presence of the cassette. The other cassette contains a truncated copy of the *LYS5* gene and arrays of lexA and tet operators, which also allow the binding of the respective repressor proteins. The cassettes can be integrated at different sites of the genome. Both *LYS5* copies are dysfunctional. Only the repair of the break via gene conversion leads to a functional copy.

To study the duration of homology search, a break is induced by shifting the cells to galactose. The colocalization of CFP and GFP indicates the successful induction of a break. The colocalization of CFP/GFP with mCherry signals the end of homology search.

Apart from this assay, the system allows to measure the efficiency of repair genetically. The break site is located in a copy of the *LYS5* gene, which encodes an enzyme of the lysine synthesis pathway[2]. Due to the the insertion of the I-SceI recognition sequence, the gene is not functional. The template site also contains a copy of *LYS5*, but only the middle part. If the endogenous copy of *LYS5* is deleted, there is no functional copy, and the cells cannot grow on media without lysine. However, if a break is induced and repaired via gene conversion[3], the break site is removed, and the cell gains a functional *LYS5* copy. If the damage is repaired by error-free NHEJ, then the cut site remains and the cut will be induced again. However, repair via erroneous NHEJ stably repairs the damage, but leads to a cell which cannot synthesize lysine. This fact can be used to measure the efficiency of repair via HR versus erroneous NHEJ. The presence of the lexA sequences allow the targeting of additional proteins. This can be used to influence position and mobility of the template site.

---

[2]`http://www.yeastgenome.org/cgi-bin/locus.fpl?locus=LYS5`

[3]The cassettes are designed in a way that a crossing over event leads to a dicentric chromosome and therefore to an inviable cell [Watson et al., 2003].

The cassettes described here are under construction. Therefore I cannot present any empirical results. Nonetheless, I present the initial studies performed for computer simulations of the homology search experiments, which will serve as a basis for interpretation and modeling of experimental results.

5.4 Theoretical approach to study homology search

To date, not only the biochemical details of homology search are unclear, but also the general mechanism has remained enigmatic. We will use the experimental system described above to investigate how the homology search system of the nucleus reacts to interference. In addition, to elucidate the underlying homology search mechanism, we have modeled some simple mechanisms and will compare their performance to that which occurs in yeast.

The most basic possibility is that the two affected chromosomes diffuse freely in the nucleus until break and template site get in contact by chance. The knowledge of how long the search takes based on this mechanism would already allow us to assess if the homology search could be the rate limiting step of homologous recombination and how much more efficient the true mechanism has to be.

In the model described above, the two chromosomes diffuse freely in 3D until the distance between the two loci of interest becomes lower than a given threshold. This event is called a collision, and the quantity we are interested in is the time of the first collision. The straightforward way to calculate the mean first collision time τ_F is the following. The chromosomes start from random initial conformations and evolve according to the BROWNian dynamics algorithm (see section 2.4.3). The simulation cycle is finished when the first collision takes place. τ_F is calculated as the average of many simulation cycles.

This approach is inefficient if τ_F is large compared to the relaxation time of the distance between the loci. Since the system "forgets" its initial distance after the relaxation time large parts of a simulation cycle are essentially redundant. In the following section, I introduce a method that avoids this redundancy and can thereby greatly reduce the computation time. This method has been proposed in [Klenin and Langowski, 2001, Klenin and Langowski, 2004] and has been successfully applied to the search of a DNA binding protein for its binding site [Merlitz et al., 2006].

5.4.1 The method of excess collisions

The method of excess collisions (MEC) can be applied in general to a system that can assume certain states A and B (but can also assume states that are neither A nor B). In our case, A is the state in which the two loci are in contact (i.e. closer to each other than a threshold distance d_A), and B is the state where the loci are far apart (i.e. the distance is larger than a second threshold d_B). We are interested in the time τ_{BA}, the average time that the system needs to switch from state B to state A. In our system, this process is usually much slower

than the opposite transition from A to B.

When the system enters state A, the event is counted as a collision. Each collision that occurs for the first time after the system has visited state B is called a prime collision. If the system is observed for a sufficiently long[4] time interval T, one can define the recurrence times

$$\tau_R = \frac{T}{n(T)}$$
$$\tau'_R = \frac{T}{n'(T)}$$

(5.1)

where $n(T)$ and $n'(T)$ are the numbers of collisions and prime collisions, respectively, during the time interval T. Between two prime collisions, the system switches from A to B and back to A. τ'_R is therefore equal to $\tau_{AB} + \tau_{BA}$ which reduces the measurement of τ_{BA} to the measurement of τ_{AB} and τ'_R:

$$\tau_{BA} = \tau'_R - \tau_{AB}.$$

(5.2)

From equation (5.1) follows

$$\tau'_R = N\tau_R \text{ with } N = \frac{n(T)}{n'(T)}.$$

(5.3)

N is the average number of collisions between two visits to state B, which is independent of T. Thus, equation (5.2) turns into:

$$\tau_{BA} = N\tau_R - \tau_{AB}.$$

(5.4)

Both τ_{AB} and N can be determined by a simulation of the transition from A to B. This is useful if $\tau_{AB} \ll \tau_{BA}$. In this case, the problem of an efficient determination of τ_{BA} has been reduced to the calculation of the recurrence time τ_R (see next section).

Using the method of excess collisions, a simulation cycle looks like the following: The system starts in state A and ends when the system reaches state B. From a series of simulation cycles, one can calculate the average transition time τ_{AB} and the average number of collisions N_C. The first collision is not part of the simulation since the system starts in state A. Therefore, it follows:

$$N = N_C + 1$$
$$\tau_{BA} = (N_C + 1)\tau_R - \tau_{AB}$$

(5.5)

$$\tau_{BA} = (N_C - \frac{\tau_{AB}}{\tau_R} + 1)\tau_R$$

(5.6)

$$\tau_{BA} = (N_E + 1)\tau_R.$$

(5.7)

---

[4] "Sufficiently long" means that a large number of collisions and prime collisions occur within T. I exclude the pathological case of a system that never enters state A or B.

Equation (5.6) is the formula that is used in practical applications to calculate τ_{BA}. In equation (5.7), $N_C - \frac{\tau_{AB}}{\tau_R}$ has been merged to N_E, the mean number of excess collisions, from which the method got its name. $\frac{\tau_{AB}}{\tau_R}$ is the mean number of collisions that would occur within a simulation cycle of length τ_{AB} starting from a random state. N_E is therefore the mean number of excess collisions, which only occur because the system starts in state A.

5.4.2 An efficient way to calculate the mean recurrence time

In section 5.4.1, the problem of determining the mean transition time τ_{BA} has been reduced to finding the mean recurrence time τ_R of the system. In this section, I demonstrate how τ_R can be obtained from faster computations in a small test system.

If p is the average probability for a collision during the next short time interval $\Delta\tau$, then the expected number of collisions during k intervals is kp. It follows from the definition of τ_R that the expected number of collisions during τ_R is 1. Thus, p must be equal to $\frac{\Delta\tau}{\tau_R}$.

On the other hand, one can define $\mathfrak{S}$ as the set (or a superset thereof) of all system states X in which the system has a finite probability to reach A within the time interval $\Delta\tau$. Then, p can also be written as:

$$p = \mathrm{P}(X \in \mathfrak{S}) \cdot \hat{p}$$

where $\hat{p}$ is the average transition probability from X to A within $\Delta\tau$, given $X \in \mathfrak{S}$. Importantly, $\hat{p}$ does not depend on the system size, and one can write:

$$\tau_R = \frac{\Delta\tau}{\hat{p} \cdot \mathrm{P}(X \in \mathfrak{S})}$$
$$\tau_R \sim \frac{1}{\mathrm{P}(X \in \mathfrak{S})} \text{ for varying system sizes.} \tag{5.8}$$

This result opens the possibility to calculate τ_R in a small test system (for example by confining the two loci of interest to a small ball), in which it is much lower, and extrapolate the result using (5.8). $\mathrm{P}(X \in \mathfrak{S})$ can be computed efficiently for both systems using a Monte Carlo simulation (see section 2.4.4) or — in simple cases — even an analytical calculation.

5.5 Proof of concept and discussion

5.5.1 Test of the MEC approach

I used a random walk model to test the method of excess collisions in the case of homology search. In the model, the two loci of interest are represented by two randomly moving points, which are confined to the same spherical volume. This highly simplified model allows a straightforward simulation of the homology search, which can be used to verify the results obtained by the MEC approach and to estimate the gain in computation time.

As a first test, I ran simulations of the full cycle starting from state A, reaching state B, and eventually returning to A. The first part of the cycle can be used for the MEC approach, and the second part is the direct simulation of the homology search. Whenever the system reenters state A, a collision is counted. The final arrival in state A, which terminates the cycle, is counted as a collision, but not the start. With these settings, all simulated cycles can be virtually concatenated to one long simulation giving direct access to the recurrence time τ_R. I used a ball radius of 1000 and the average distance $\frac{36}{35} \cdot 1000 = 1028.6$ (see equation (3.7)) between the spots as the threshold distance d_B. The step length of the random walks was 1. Table 5.1 summarizes the results for different threshold distances d_A.

d_A	n	$\tau_{AB}[10^6]$	N	$\tau_R[10^6]$	$\tau_{BA}^{(\text{MEC})}[10^6]$	$\tau_{BA}^{(\text{direct})}[10^6]$
5	1000	0.722 ± 0.016	5.42 ± 0.16	40.09 ± 0.54	216.6 ± 7.0	216.6 ± 6.8
10	1000	0.739 ± 0.018	10.33 ± 0.33	10.94 ± 0.11	112.2 ± 3.8	112.2 ± 3.4
20	1000	0.743 ± 0.017	21.69 ± 0.68	2.47 ± 0.17	52.9 ± 1.7	52.9 ± 1.6
40	1000	0.725 ± 0.017	43.0 ± 1.4	0.6586 ± 0.0032	27.57 ± 0.91	27.57 ± 0.85
80	1000	0.716 ± 0.016	89.4 ± 2.9	0.1697 ± 0.0006	14.46 ± 0.49	14.46 ± 0.46

Table 5.1: Parameters and results for the simulation of the whole A-B-A cycle: threshold distance for entering A, number of simulation cycles, average transition time from A to B, number of collisions during the A to B transition (including the prime collision at the start), mean recurrence time, average transition time from B to A (calculated using the MEC approach), average transition time from B to A (measured directly).

It has to be noted that these simulations cannot be used as a solid verification of the MEC approach. Since all data were calculated from the same simulations, the agreement of $\tau_{BA}^{(\text{MEC})}$ and $\tau_{BA}^{(\text{direct})}$ only reflects the arithmetic derivations presented above. Additionally, at the start of the B-A simulation cycle the distance between the two spots is exactly d_B, which is artificial. It would be more realistic to draw the start positions of the two spots from a uniform distribution across the confining ball. However, if the relaxation time of the distance is much shorter than τ_{BA}, then the correlation with the starting distance is lost very fast and it is safe to always start with the mean distance. The relaxation time can be calculated via an autocorrelation analysis [Priestley, 1982].

I ran a simulation of two random walks with $2 \cdot 10^9$ steps inside a ball of radius 1000 and a step length of 1. I calculated the autocorrelation function (ACF) $A(\Delta t)$ as

$$A(\Delta t) = \frac{\langle d(t)d(t+\Delta t) \rangle - \langle d(t) \rangle^2}{\langle d(t)^2 \rangle - \langle d(t) \rangle^2}.$$

The result is shown in Figure 5.2. A fit of an exponential function to the ACF yielded a decay constant of 553600 time steps. Thus, the distance loses its correlation after a few million time steps, which is acceptable for the simulations shown above. The result of this

analysis is confirmed by a set of simulations of the *B-A* transition starting from randomly located spots (Table 5.2). These data also corroborate the validity of the MEC approach.
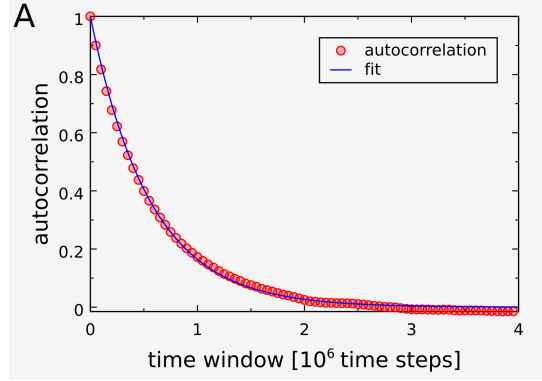
Figure 5.2: Spot-spot distance autocorrelation, computed from a random walk simulation of two spots inside a spherical volume of radius 1000. The simulation comprised $2 \cdot 10^9$ time steps, and the step length was 1. The function $\exp(-\frac{\Delta t}{a})$ was fitted to the ACF and yielded a decay constant of 553600 time steps.

d_A	n	$\tau_{BA}^{(\mathrm{MEC})}[10^6]$	$\tau_{BA}^{(d_B)}[10^6]$	$\tau_{BA}^{(\mathrm{random})}[10^6]$
5	1000	216.6 ± 7.0	224.0 ± 7.4	230.9 ± 7.4
10	1000	112.2 ± 3.8	113.4 ± 3.6	115.3 ± 3.6
20	1000	52.9 ± 1.7	55.0 ± 1.7	56.7 ± 1.9
40	1000	27.57 ± 0.91	28.43 ± 0.89	28.38 ± 0.88
80	1000	14.46 ± 0.49	14.91 ± 0.46	14.57 ± 0.46

Table 5.2: Parameters and results for the simulation of the *B-A* cycle: threshold distance for entering *A*, number of simulation cycles, average transition time from *B* to *A* (calculated using the MEC approach, see Table 5.1), average transition time from *B* to *A* (measured directly), average transition time from *B* to *A* (measured directly with random start positions).

The data shown so far confirm that τ_{BA} can be recovered accurately from a set of *A-B* simulations if τ_R can be obtained from elsewhere. Additionally, the results can be used to get an impression of the properties of the search and of the impact of the MEC approach. As expected, τ_{AB} is independent of d_A as long as $d_A \ll d_B$. Moreover, in the observed range, τ_{AB} is consistently 1 to 2 orders of magnitude smaller than $N\tau_R$. Therefore, the calculation of τ_{BA} via equation (5.5) is dominated by the first term. As a consequence, the number of collisions in a simulation cycle of the inverse homology search is the important quantity. The length of the cycle is almost insignificant.

The results suggest that the number N of collisions during one *A-B* cycle is proportional to d_A and that the recurrence time τ_R is inversely proportional to d_A^2 (see Table 5.1 and Figure 5.3A). As mentioned above, equation (5.5) for the calculation of τ_{BA} is dominated by $N\tau_R$. Therefore, in the parameter range used here, τ_{BA} should be inversely proportional to d_A. The simulation data confirm this relationship (Figure 5.3B).
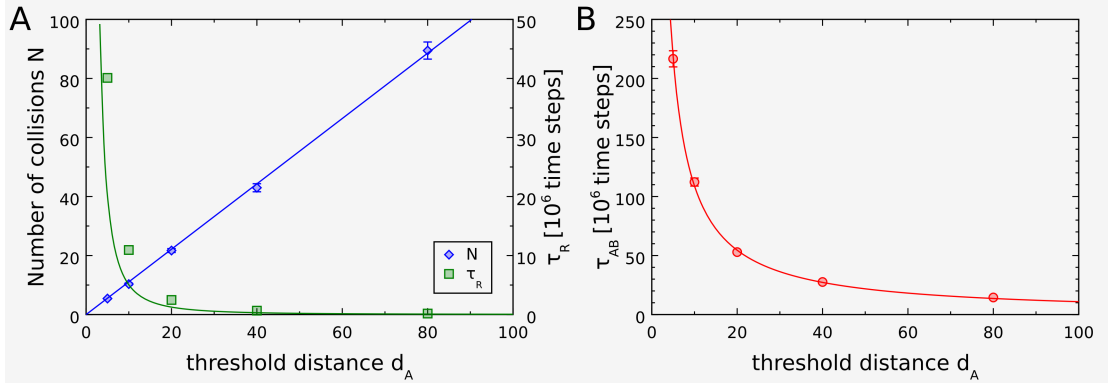
Figure 5.3: (A) Dependency of the number of collisions N and the recurrence time τ_R on the threshold distance d_A. N is well described by a linear relationship to d_A, τ_R appears to be inversely proportional to d_A^2. (B) τ_{BA} (the direct measurement is shown here) can be modeled as $\tau_{BA} \sim d_A^{-1}$.

The gain in computation time that is accomplished by the MEC approach is the ratio between τ_{BA} and τ_{AB}. Since τ_{AB} depends only very weakly on d_A, the gain is also proportional to d_A^{-1}. In the parameter range covered here, it reaches a factor of 300 in the case of $d_A = 5$.

5.5.2   Computation of the recurrence time

As laid out in section 5.4.2, the recurrence time τ_R can be calculated in a smaller confinement and extrapolated using the relationship (5.8).

In this case, $P(X \in \mathfrak{S})$ can be calculated analytically. I have defined $\mathfrak{S}$ as the set of system states from which the state A can be reached within a given small time interval $\Delta\tau$. It is convenient to chose $\Delta\tau$ as one time step of the simulation.

It is important to note that we are not interested in what happens while the two loci are in contact. Therefore, the clock is halted as long as the system is in state A. This is equivalent to a setting in which the volume in the state space that is occupied by state A is inaccessible (in this case, the attempt to enter is counted as a collision). Therefore, the probability that the system state lies in $\mathfrak{S}$ is the probability that the distance between the spot lies between d_A and $d_A + 2l$ where l is the step length of the walks. This probability

can be derived from equation (3.6):

$$\mathrm{P}(\delta)d\delta = 3\delta^2 - \tfrac{9}{4}\delta^3 + \tfrac{3}{16}\delta^5 \text{ with } \delta = \tfrac{d}{R} \tag{3.6}$$

$$\mathrm{P}_0(d_A \leq d \leq d_A + 2l) = \int_{\frac{d_A}{R}}^{\frac{d_A+2l}{R}} \mathrm{P}(\delta)d\delta \tag{5.9}$$

$$\tag{5.10}$$

However, this probability has to be renormalized to account for inaccessibility of distances below d_A:

$$\mathrm{P}(X \in \mathfrak{S}) = \mathrm{P}(d_A \leq d \leq d_A + 2l) \tag{5.11}$$

$$= \frac{\displaystyle\int_{\frac{d_A}{R}}^{\frac{d_A+2l}{R}} \mathrm{P}(\delta)d\delta}{1 - \displaystyle\int_{0}^{\frac{d_A}{R}} \mathrm{P}(\delta)d\delta}$$

$$\mathrm{P}(X \in \mathfrak{S}) = \frac{6\frac{l}{R}\left(\frac{d_A}{R}\right)^2 + 12\left(\frac{l}{R}\right)^2\frac{d_A}{R} + 8\left(\frac{l}{R}\right)^3}{1 - \left(\frac{d_A}{R}\right)^3 + \frac{9}{16}\left(\frac{d_A}{R}\right)^4 - \frac{1}{32}\left(\frac{d_A}{R}\right)^6}$$

$$+ \frac{\frac{9}{16}\left(\left(\frac{d_A}{R}+2\frac{l}{R}\right)^4 - \left(\frac{d_A}{R}\right)^4\right) + \frac{1}{32}\left(\left(\frac{d_A}{R}+2\frac{l}{R}\right)^6 - \left(\frac{d_A}{R}\right)^6\right)}{1 - \left(\frac{d_A}{R}\right)^3 + \frac{9}{16}\left(\frac{d_A}{R}\right)^4 - \frac{1}{32}\left(\frac{d_A}{R}\right)^6} \tag{5.12}$$

$$\mathrm{P}(X \in \mathfrak{S}) = \frac{6\frac{l}{R}\left(\frac{d_A}{R}\right)^2 + 12\left(\frac{l}{R}\right)^2\frac{d_A}{R} + 8\left(\frac{l}{R}\right)^3 + \mathcal{R}(\frac{d_A}{R},\frac{l}{R})}{1 - \left(\frac{d_A}{R}\right)^3 + \frac{9}{16}\left(\frac{d_A}{R}\right)^4 - \frac{1}{32}\left(\frac{d_A}{R}\right)^6}. \tag{5.13}$$

In the observed range ($\frac{d_A}{R} < 0.1$), the denominator is larger than 0.999 and therefore well approximated by 1. In the rest term $\mathcal{R}(\frac{d_A}{R}, \frac{l}{R})$, the sum of the exponents of $\frac{d_A}{R}$ and $\frac{l}{R}$ is at least 4, suggesting that they can be neglected for small values of d_A and l. Although the constant and the linear term become more important for very small values of d_A, the sum is governed by the first term in the parameter range investigated here (Figure 5.4). This behavior explains the proportionality of τ_R to d_A^{-2} observed in the simulations above.

In order to test the proportionality of τ_R to $\mathrm{P}(X \in \mathfrak{S})^{-1}$, I ran simulations with $d_A = 10$ and different values for the confinement radius R. At first glance, it seems to be the same to vary d_A or R since $\mathrm{P}(X \in \mathfrak{S})$ depends only on the ratio $\frac{d_A}{R}$. However, $\mathrm{P}(X \in \mathfrak{S})$ also depends on the step length l, and it does matter if $\frac{l}{R}$ or $\frac{l}{d_A}$ stays constant. Therefore, I did not use the simulations with varying threshold distance above, but ran a separate set of simulations with varying confinement radius.
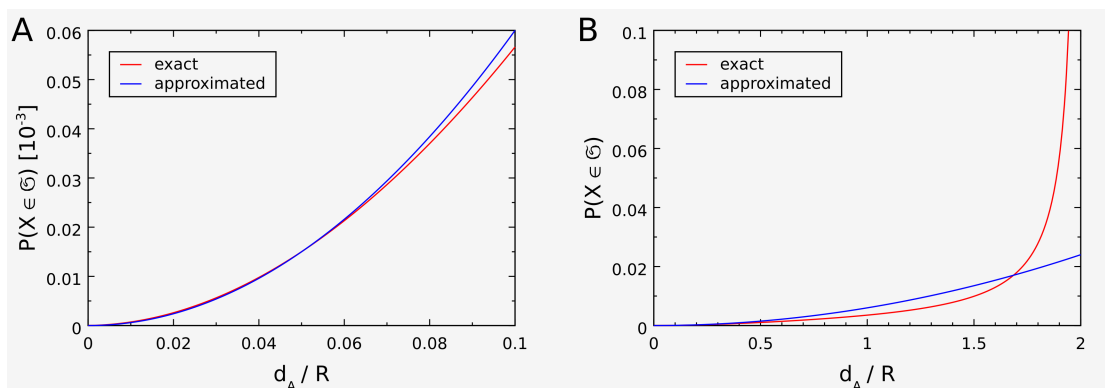
Figure 5.4: Probability for being close to a collision versus the collision distance d_A. (A) $P(d_A \leq d \leq d_A + 2l)$ in the observed range of d_A. The probability is well approximated by $6\frac{l}{R}(\frac{d_A}{R})^2$. (B) $P(d_A \leq d \leq d_A + 2l)$ in the full range of d_A.

The results are shown in Figure 5.5. By reducing the confinement radius from 1000 to 20, τ_R (and thereby the computation time) can be reduced by five orders of magnitude (panel A). The estimated true recurrence time is largely independent of the the confinement radius in the observed range. Step length effects are not detectable. However, the directly measured recurrence time is consistently underestimated. Further investigation is required to elucidate this effect.
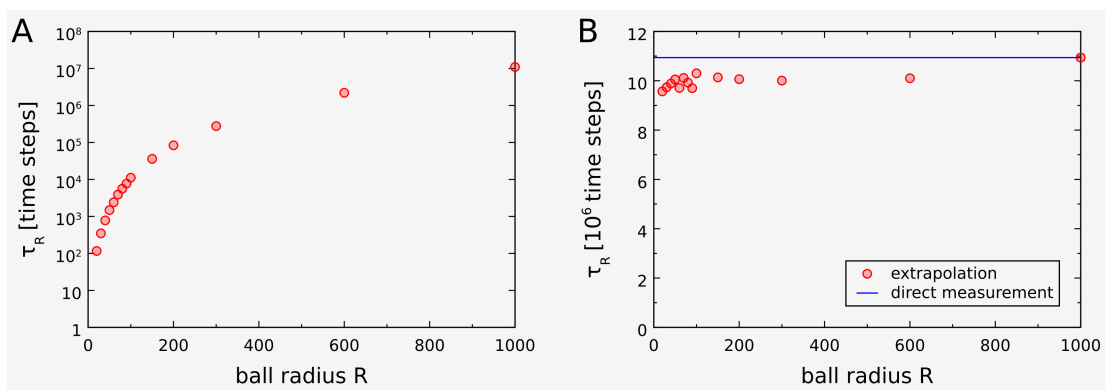
Figure 5.5: (A) Recurrence time τ_R for different confinement radii. Each value was calculated from 10000 simulations with collision threshold $d_A = 10$ and step length $l = 1$. (B) Estimated full recurrence times were obtained by rescaling the value for τ_R measured in the small system with the factor $P_R(X \in \mathfrak{S})/P_{1000}(X \in \mathfrak{S})$, according to equation (5.8). The value that was measured directly (see Table 5.1) is consistently underestimated.

5.5.3 The effect of peripheral anchoring

As discussed before, the nuclear envelope is an important scaffold for nuclear organization in budding yeast. There is evidence that it also plays a role in DNA damage repair [Therizols et al., 2006, Nagai et al., 2008]. One question within this project is how nuclear organization influences specifically the duration of homology search. I used the simple random walk model that I introduced above, to get an impression of the impact that the peripheral anchoring of one or both spots could have on homology search.

   I looked at five scenarios:

1. Both spots are moving inside the confinement.

2. One spot is moving internally and the other on the surface of the confinement.

3. One spot is moving internally, the other is fixed to one position at the periphery.

4. Both spots are moving on the surface.

5. One spot is moving on the surface of the confinement, the other one is fixed to one position.

   In this study, I focused on a collision threshold of $d_A = 10$ and a confinement radius of 1000. The threshold distance d_B for entering state B was set to the average distance in the respective scenario. It was set to $\frac{36}{35} \cdot 1000 = 1028.6$ in scenario 1, $\frac{6}{5} \cdot 1000 = 1200$ in scenarios 2 and 3, and to $\frac{4}{3} \cdot 1000 = 1333.3$ in scenarios 4 and 5 (see section 3.4.2). I did not make use of the fast calculation of τ_R, but carried out simulations of the whole *A-B-A* cycle. However, τ_R could be computed in the same way as presented in the previous section. The only modification that would be required is the replacement of the probability density function $P(\delta)$ by the correct one for the respective scenario (see also section 3.4.2).

   The results of my simulations are shown in Table 5.3. The transition time τ_{BA} is the smallest by far in scenario 4, in which both loci move along the periphery. It is more than 5 times smaller than for two internal loci (scenario 1) and more than 20 times smaller than in the slowest scenario (scenario 3, one spot internal, one fixed at the periphery).

   It is important to note that the two fastest scenarios (4 and 5) both require that the movement (or fixed position) of both loci is restricted to the surface of the confinement. This requirement is difficult to realize for a cell in response to the detection of a DSB. Since the homology search itself results in the identification of the homologous template, this template site cannot be specifically relocated to the periphery at the beginning of the search. Moreover, the relocation of only the break site to the periphery (scenarios 2 and 3) would be most detrimental. At least in this simple diffusive model, it is the best reaction with respect to the speed of homology search not to change the position of an internal site of damage. If the damaged locus is already at the periphery, the recommended reaction depends on assumptions about the location of the template locus. If the template can be expected to be

Scenario	n	$\tau_{AB}[10^6]$	N	$\tau_R[10^6]$	$\tau_{BA}^{(\text{direct})}[10^6]$
1	1000	0.739 ± 0.018	10.33 ± 0.33	10.94 ± 0.11	112.2 ± 3.4
2	1000	1.062 ± 0.024	11.31 ± 0.34	20.93 ± 0.20	235.7 ± 7.3
3	1000	1.808 ± 0.038	15.58 ± 0.48	28.36 ± 0.23	439 ± 13
4	1000	1.152 ± 0.027	42.8 ± 1.3	0.5055 ± 0.0024	20.50 ± 0.68
5	1000	2.395 ± 0.056	67.9 ± 2.1	0.6505 ± 0.0025	41.8 ± 1.4

Table 5.3: Parameters and results for the simulation of the whole *A-B-A* cycle in 5 different locus position scenarios (see text): number of simulation cycles, average transition time from *A* to *B*, number of collisions during the *A* to *B* transition (including the prime collision at the start), mean recurrence time, average transition time from *B* to *A* (measured directly). The threshold distance for entering *A* was 10 nm.

attached to the periphery as well (e.g. both sites are telomeres), then the break site should stay attached as well. If in contrast, the template must be expected to be internal, then the site of damaged should be released.

At the current point of this study, I have not investigated the effects of increased mobility on the duration of homology search. It is plausible, however, that an increased speed of either one or both loci would speed up the search. It has been shown that certain chromatin remodeling complexes are able to increase the speed of the movement of a chromosomal locus, possibly by increasing the flexibility of the chromatin fiber [Neumann et al., *in preparation*]. On the other hand, fast movement of the loci might impair the recognition of the homology. This is a question that cannot be addressed within this simple model.

In the preceding discussion, I have only compared the simulation results to each other internally. This has also been the main purpose of my simulations carried out so far: a confirmation of the applicability and usefulness of the MEC approach and a first comparison of different positioning scenarios of loci in the nucleus. In order to relate the results to biological data, one has to define the time step of the simulation. In any case, the applicability of the results presented here to the homology search of chromosomal loci is limited because they cannot be expected to move in a random-walk-like fashion.

However, to get a first idea of the duration of homology search, one can use the time calibration for the movement of an excised chromatin ring (see chapter 4). In this project, I found that the movement of such a ring is well described by a random walk with step size 1 nm and 44602 steps per second. Using this value, one can calculate that it would take two excised rings about 8 min on average to get in contact (defined by a collision distance of 10 nm) if moving along the periphery and about 40 min if moving internally.

Chromosomal loci have been shown to move more slowly due to the drag of the chromatin fiber [Gartenberg et al., 2004]. The possible entanglement of the chromosomes may lead to a further increase of the duration of homology search. Moreover, it has to be investigated

how this entanglement might impair the MEC approach. Additionally, an important, but not well known parameter is the collision threshold d_A. This parameter does not have to be the physical distance that break and template site have to reach. Depending on the exact question one wants to answer, it can also comprise various reaction parameters, e.g. the probability that strand invasion actually takes place when the two sites come close to each other [Merlitz et al., 2006]. These questions are subject to further investigations, which will be based on polymer modeling with `corchy++` (see section 2.4).

Chapter 6

Discussion

The aim of my thesis was to develop and apply mathematical and computational models to elucidate fundamental mechanisms of chromosomal organization in budding yeast. For a long time it has been clear that the spatial organization of chromosomes within the nucleus is not random, and the relationship between chromatin reorganization and nuclear processes such as transcriptional regulation and DNA damage repair is still a much studied and poorly understood field. A key component of my approach was to start out with a minimal system and apply successive refinements in order to approximate the situation in a living cell. For each aspect of nuclear organization I investigated, a constant evaluation of the models with respect to experimental results allowed the identification of important determinants that govern the behavior of the investigated systems.

I started by analyzing the previously observed looping of telomeres, focusing on the relationship between telomere anchoring and telomere-telomere interaction. My initial model described a single chromosome in the yeast nucleus by a polymer model. Apart from confinement to the nucleus and the elastic attachment of the centromere to the spindle pole body, the conformation of the chromosome was unconstrained. The results of the computer simulations showed that this minimal model did not reproduce our experimental results. However, this is not a negative result *per se*, but is part of the approach and useful information could still be extracted.

The results showed that the initial model was oversimplified. The chromosomes showed on average a smaller telomere-telomere distance in laboratory microscopy experiments than modeled in free solution, suggesting that the telomeres are brought together by some influence not included in the model. Importantly, the modeling showed that the anchoring of the telomeres alone has the opposite effect and increased the average telomere-telomere distance. It would be very difficult to show this direct effect experimentally since one can usually not exclude that a manipulation of anchoring would affect telomere-telomere interaction in other ways.

However, the model with anchored telomeres allowed free movement along the periphery. Therefore, the anchoring of the telomeres could decrease the distance more indirectly. First,

the telomeres could be fixed to the nuclear envelope in close proximity, which would trivially lead to a small average distance. Alternatively, the attachment could only slow down the movement of telomeres along the periphery. This would allow the exploration of the whole nuclear periphery in principle but maybe not during the limited time of interphase.

Again, different explanations for slow movement at the periphery are possible. Telomeres are attached to an anchor in the nuclear membrane. This anchor has to diffuse inside the membrane, which must be expected to be slower than diffusion in free solution. Also, the association with a telomere cluster could slow down the movement of a single telomere.

In any case, fixing or reducing the movement of telomeres at the nuclear periphery can only explain the perpetuation of a small telomere-telomere distance. In this case, the question is how a small distance between telomeres is established. This is a biological question that cannot be answered by the initial model. This result provides us with an example of how theoretical modeling can yield a feedback to the experimental investigation. In this case, the modeling showed that there must be a mechanism that leads to the observed juxtaposition of telomeres of the same chromosome. [Schober et al., 2008] have proposed a model in which the RABL conformation of chromosomes gets fixed by attachment of the telomeres at the end of mitosis. This model could be the basis for further computational modeling investigating if the mitotic movement of a chromosome can result in a telomere-telomere distance as small as determined experimentally.

The simulation of intrachromosomal distances revealed that telomeric anchoring increases the average telomere-telomere distance. However, it also showed that in contrast, the frequency of very close juxtaposition is increased. This result makes an interesting link to my investigations of the homology search process during DNA double-strand break repair. The random walk simulations in the context of this project showed that the restriction of both loci to movement only on the surface of the confinement leads to a dramatic decrease of the duration of homology search. This effect also remains – although to smaller extent – if one of the loci is fixed to one position on the nuclear periphery.

This result is in turn connected to my investigations of the asymmetric distribution of plasmids after mitosis in budding yeast. In this project, I found that the segregation bias of *ARS* plasmids in yeast can be explained by the geometric barrier that the dividing nucleus imposes onto the plasmid diffusion. In agreement with the results of the homology search simulations, the transfer probability of the plasmid into the daughter nucleus was greatly increased if the plasmid movement was restricted to the nuclear periphery. This behavior agrees with the mathematical prediction that the time to escape from a confining volume through a small window is decreased if the object moves along the surface of the confinement. This analytical result is restricted to the case of very small escape windows and is therefore not directly applicable to the plasmid segregation problem. However, it suggests that the plasmid shows the same behavior, which was confirmed by my simulations. Interestingly, the mathematical result is even more applicable to the homology search problem because the collision threshold is usually smaller than the tunnel of the dividing nucleus. The escape

of the plasmid from the mother nucleus and the collision of the randomly moving locus with its fixed homologous partner are — within the scope of the respective models — the same thing. It would be interesting to find out if the theory of narrow escape can be extended to the case of two moving spots instead of one moving spot and one fixed escape window.

The investigations of the spatial organization of yeast chromosomes within my thesis work focused on the effects of change in position, especially anchoring to the nuclear periphery. However, it has been shown that also the mobility of chromosomal loci is regulated [Neumann et al., *in preparation*]. We showed in this work that chromatin remodelers are able to increase the mobility of a locus. We propose that the chromatin remodeling increases the flexibility of the chromatin fiber which then weakens the restraint that the fiber imposes on the movement of a locus. Within the ongoing investigation of homology search in yeast, I am going to investigate this important question of the influence of locus mobility on the duration of homology search.

Appendix A

Mathematical Tools

A.1 Antisymmetric functions

Let $f : \mathbb{R} \to \mathbb{R}, x \mapsto f(x)$ be a function. f is called antisymmetric with respect to a if

$$f(a - x) = -f(a + x) \; \forall \, x \in \mathbb{R}.$$

If a is not mentioned, it is assumed to be 0. If a function f is antisymmetric with respect to a, then any integral over an interval symmetric to a vanishes:

$$
\begin{aligned}
\int_{a-b}^{a+b} f(x) \, dx &= \int_{a-b}^{a} f(x) \, dx + \int_{a}^{a+b} f(x) \, dx \\
&= \int_{-b}^{0} f(a+x) \, dx + \int_{0}^{b} f(a+x) \, dx \\
&= -\int_{-b}^{0} f(a-x) \, dx + \int_{0}^{b} f(a+x) \, dx \quad \text{(antisymmetry)} \\
&= \int_{b}^{0} f(a+x) \, dx + \int_{0}^{b} f(a+x) \, dx \quad \text{(substitution } x \mapsto -x) \\
&= -\int_{0}^{b} f(a+x) \, dx + \int_{0}^{b} f(a+x) \, dx \\
&= 0
\end{aligned}
$$

A.2   Integrals

I assume standard knowledge about integration including the primitives of elementary functions like power functions, exponential functions etc. and integration methods like integration by parts and substitution. All these things can be found in any elementary calculus text book. The following primitive functions are also not very advanced, but I include them for completeness.

A.2.1   Integrals involving trigonometric functions

$$\int \sin^2 x \, dx = \int \left(\tfrac{1}{2}\sin^2 x + \tfrac{1}{2}\cos^2 x + \tfrac{1}{2}\sin^2 x - \tfrac{1}{2}\cos^2 x \right) dx$$

$$= \int \left(\tfrac{1}{2} - \tfrac{1}{2}\cos(2x) \right) dx$$

$$= \tfrac{1}{2}x - \tfrac{1}{4}\sin(2x) + C$$

$$\int \cos^2 x \, dx = \int \left(1 - \sin^2 x \right) dx$$

$$= x - \left(\tfrac{1}{2}x - \tfrac{1}{4}\sin(2x) \right) + C$$

$$= \tfrac{1}{2}x + \tfrac{1}{4}\sin(2x) + C$$

$$\int \sin^3 x \, dx = \int \left(1 - \cos^2\theta \right) \sin\theta \, d\theta$$

$$= -\cos\theta + \tfrac{1}{3}\cos^3\theta + C$$

Appendix B

Detailed Derivations

B.1 Two spots inside a sphere
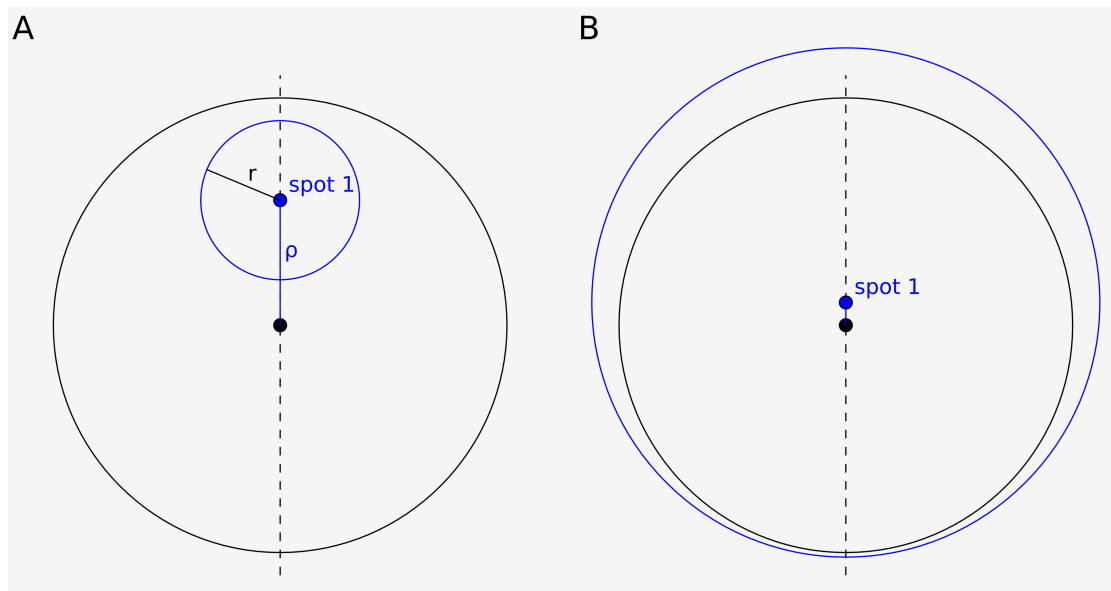
Figure B.1: (A) Case 1.1. (B) Case 2.1.

Case 1: $0 \leq r \leq 1$ (the distance is smaller than the radius of the sphere)
Case 1.1: $0 \leq \rho \leq 1 - r$ (no intersection, Figure B.1A)

$$\hat{\hat{P}}(\rho, r) = \frac{4\pi r^2}{\frac{4}{3}\pi}$$

$$= 3r^2$$

$$\begin{aligned}
\tilde{\mathrm{P}}(\rho, r) &= \hat{\mathrm{P}}(\rho) \cdot \hat{\tilde{\mathrm{P}}}(\rho, r) \\
&= 3\rho^2 \cdot 3r^2 \\
&= 9\rho^2 r^2
\end{aligned}$$
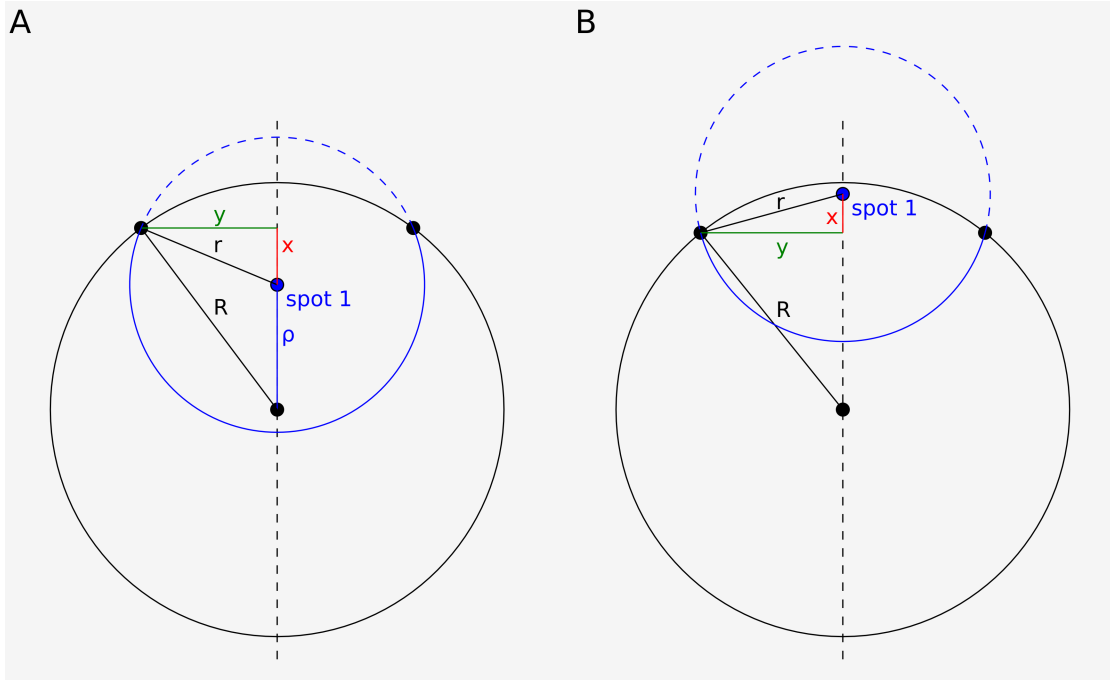
Figure B.2: (A) Case 1.2.1. (B) Case 1.2.2.

Case 1.2: $1 - r < \rho \le 1$ (intersection, Figure B.2)

Case 1.2.1: $\rho^2 \le 1 - r^2$

Let h be the height of the spherical cap whose surface is needed: $h = r + x$. The theorem of PYTHAGORAS yields:

$$\begin{aligned}
1 &= (\rho + x)^2 + y^2 \\
r^2 &= x^2 + y^2 \\
h &= \frac{1 - (r - \rho)^2}{2\rho}.
\end{aligned}$$

Case 1.2.2: $\rho^2 > 1 - r^2$

$$\begin{aligned}
h &= r - x \\
1 &= (\rho - x)^2 + y^2 \\
r^2 &= x^2 + y^2
\end{aligned}$$

$$h = \frac{1 - (r - \rho)^2}{2\rho}.$$

Let S be the surface of the spherical cap (which is also correct if $h > R$):

$$S = 2\pi R h.$$

From this, it follows:

$$\hat{\tilde{P}}(\rho, r) = \frac{3r}{4\rho}(1 - (r - \rho)^2)$$
$$\tilde{P}(\rho, r) = \tfrac{9}{4}\rho r (1 - (r - \rho)^2).$$

This yields the following result for case 1:

$$P(r) = \int\limits_0^1 \tilde{P}(\rho, r)\, d\rho$$
$$= 3r^2 - \tfrac{9}{4}r^3 + \tfrac{3}{16}r^5.$$

Case 2: $1 < r \leq 2$ (the distance is larger than the radius of the sphere)
Case 2.1: $\rho \leq r - 1$ (no intersection, Figure B.1B)

$$\hat{\tilde{P}}(\rho, r) = 0$$
$$\tilde{P}(\rho, r) = 0.$$

Case 2.2: $r - 1 < \rho \leq 1$ (intersection, Figure B.3)
Case 2.2.1: $\rho^2 \leq r^2 - 1$

$$h = r - x$$
$$1 = (x - \rho)^2 + y^2$$
$$r^2 = x^2 + y^2$$
$$h = \frac{1 - (r - \rho)^2}{2\rho}.$$

Case 2.2.2: $\rho^2 > r^2 - 1$

$$h = r - x$$
$$1 = (\rho - x)^2 + y^2$$
$$r^2 = x^2 + y^2$$

Figure B.3: (A) Case 2.2.1. (B) Case 2.2.2.

$$h = \frac{1 - (r - \rho)^2}{2\rho}.$$

$$\hat{\tilde{P}}(\rho, r) = \frac{3r}{4\rho}(1 - (r - \rho)^2)$$
$$\tilde{P}(\rho, r) = \tfrac{9}{4}\rho r(1 - (r - \rho)^2).$$

This yields the same result as for case 1:

$$P(r) = \int_0^1 \tilde{P}(\rho, r)\, d\rho$$
$$= 3r^2 - \tfrac{9}{4}r^3 + \tfrac{3}{16}r^5.$$

Appendix C

Driver's License

Once you have written your PhD thesis it will accompany you for the rest of your life. You have poured many thoughts into it which you can always refer back to. Wouldn't it be nice if your thesis could also help you in situations of everyday life? Just imagine you get stopped by the police and asked for your driver's license and you can just answer: "It's in my thesis." Well, I can...

Figure C.1: Driver's license outside.
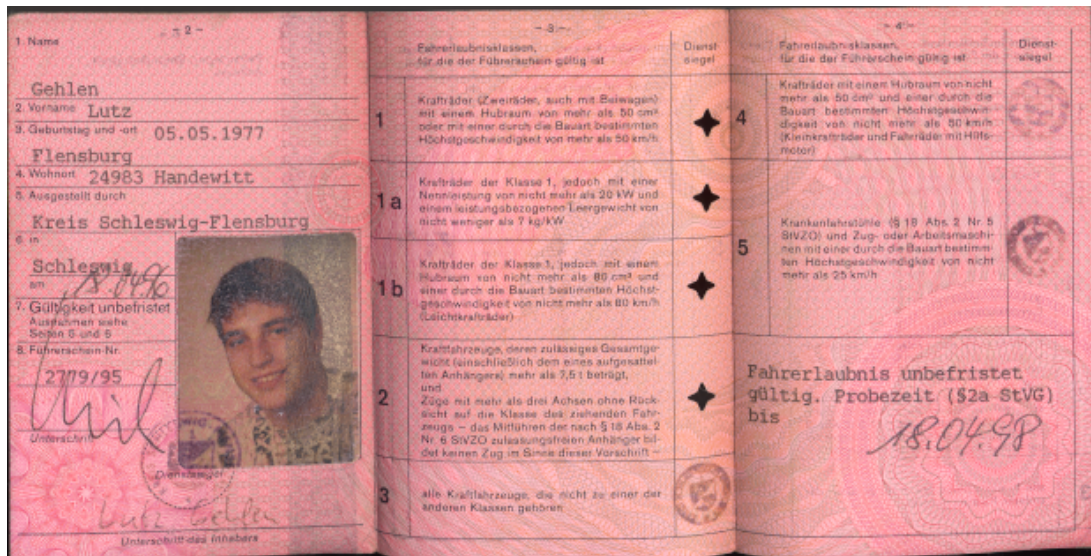
Figure C.2: Driver's license inside.

References

[Ackermann et al., 2007] Ackermann, M., Chao, L., Bergstrom, C. T., and Doebeli, M. (2007). On the evolutionary origin of aging. *Aging Cell*, 6:235–244.

[Ackermann et al., 2003] Ackermann, M., Stearns, S. C., and Jenal, U. (2003). Senescence in a bacterium with asymmetric division. *Science*, 300:1920.

[Aguilaniu et al., 2003] Aguilaniu, H., Gustafsson, L., Rigoulet, M., and Nyström, T. (2003). Asymmetric inheritance of oxidatively damaged proteins during cytokinesis. *Science*, 299:1751–1753.

[Akhtar and Gasser, 2007] Akhtar, A. and Gasser, S. M. (2007). The nuclear envelope and transcriptional control. *Nat. Rev. Genet.*, 8:507–517.

[Aleksandrov and Trakhtengerts, 1974] Aleksandrov, A. A. and Trakhtengerts, M. S. (1974). Viscosity of water at temperatures of -20 to 150°C. *J. Eng. Phys. Thermophys.*, 27:1235–1239.

[Allis et al., 2006] Allis, C. D., Jenuwein, T., and Reinberg, D., editors (2006). *Epigenetics*. Cold Spring Harbor Laboratory Press, 1. edition.

[Amberg et al., 1997] Amberg, D. C., Zahner, J. E., Mulholland, J. W., Pringle, J. R., and Botstein, D. (1997). Aip3p/Bud6p, a yeast actin-interacting protein that is involved in morphogenesis and the selection of bipolar budding sites. *Mol. Biol. Cell*, 8:729–753.

[Andrulis et al., 1998] Andrulis, E. D., Neiman, A. M., Zappulla, D. C., and Sternglanz, R. (1998). Perinuclear localization of chromatin facilitates transcriptional silencing. *Nature*, 394:592–595.

[Andrulis et al., 2002] Andrulis, E. D., Zappulla, D. C., Ansari, A., Perrod, S., Laiosa, C. V., Gartenberg, M. R., and Sternglanz, R. (2002). Esc1, a nuclear periphery protein required for Sir4-based plasmid anchoring and partitioning. *Mol. Cell. Biol.*, 22:8292–8301.

[Ansari and Gartenberg, 1997] Ansari, A. and Gartenberg, M. R. (1997). The yeast silent information regulator Sir4p anchors and partitions plasmids. *Mol. Cell. Biol.*, 17:7061–7068.

[Avery et al., 1944] Avery, O. T., McLeod, C. M., and McCarty, M. (1944). Studies on the chemical nature of the substance inducing transformation of pneumococcal types. Induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus Type III. *J. Exptl. Med.*, 79:137–158.

[Aylon et al., 2003] Aylon, Y., Liefshitz, B., Bitan-Banin, G., and Kupiec, M. (2003). Molecular dissection of mitotic recombination in the yeast *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, 23:1403–1417.

[Barzel and Kupiec, 2008] Barzel, A. and Kupiec, M. (2008). Finding a match: how do homologous sequences get together for recombination? *Nat. Rev. Genet.*, 9:27–37.

[Berg, 1993] Berg, H. C. (1993). *Random Walks in Biology*. Princeton University Press, expanded edition.

[Berger et al., 2008] Berger, A. B., Cabal, G. G., Fabre, E., Duong, T., Buc, H., Nehrbass, U., Olivo-Marin, J. C., Gadal, O., and Zimmer, C. (2008). High-resolution statistical mapping reveals gene territories in live yeast. *Nat. Methods*, 5:1031–1037.

[Binder, 1996] Binder, K., editor (1996). *Monte Carlo and Molecular Dynamics Simulations in Polymer Science*. Oxford University Press.

[Boas, 1983] Boas, M. L. (1983). *Mathematical Methods in the Physical Sciences*. Wiley & Sons, 2. edition.

[Bohn et al., 2007] 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.

[Bouchaud and Georges, 1990] Bouchaud, J.-P.. and Georges, A. (1990). Anomalous diffusion in disordered media: statistical mechanisms, models and physical applications. *Physics Reports*, 195:127–293.

[Brandt, 1998] Brandt, S. (1998). *Statistical and Computational Methods for Scientists and Engineers*. Springer, 3. edition.

[Bressan et al., 2004] Bressan, D. A., Vazquez, J., and Haber, J. E. (2004). Mating type-dependent constraints on the mobility of the left arm of yeast chromosome III. *J. Cell Biol.*, 164:361–371.

[Bystricky et al., 2004] 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. U.S.A.*, 101:16495–16500.

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

[Cadet et al., 1997] Cadet, J., Berger, M., Douki, T., and Ravanat, J. L. (1997). Oxidative damage to DNA: formation, measurement, and biological significance. *Rev. Physiol. Biochem. Pharmacol.*, 131:1–87.

[Campbell et al., 1978] Campbell, A. M., Cotter, R. I., and Pardon, J. F. (1978). Light scattering measurements supporting helical structures for chromatin in solution. *Nucleic Acids Res.*, 5:1571–1580.

[Carlson, 2004] Carlson, E. A. (2004). *Mendel's Legacy: The Origin of Classical Genetics.* Cold Spring Harbor Laboratory Press.

[Chargaff and Magasanik, 1949] Chargaff, E. and Magasanik, B. (1949). The nucleotide composition of ribonucleic acids. *J Am Chem Soc*, 71:1513.

[Clarke and Carbon, 1980] Clarke, L. and Carbon, J. (1980). Isolation of a yeast centromere and construction of functional small circular chromosomes. *Nature*, 287:504–509.

[Clikeman et al., 2001] Clikeman, J. A., Khalsa, G. J., Barton, S. L., and Nickoloff, J. A. (2001). Homologous recombinational repair of double-strand breaks in yeast is enhanced by MAT heterozygosity through yKU-dependent and -independent mechanisms. *Genetics*, 157:579–589.

[Coffey et al., 2004] Coffey, W. T., Kalmykokov, Yu.. P., and Waldron, J. T. (2004). *The Langevin Equation.* World Scientific, 2. edition.

[Crank, 1975] Crank, J. (1975). *The Mathematics of Diffusion.* Oxford University Press, 2. edition.

[Debye and Hückel, 1923] Debye, P. and Hückel, E. (1923). The theory of electrolytes. i. Lowering of freezing point and related phenomena. *Z. Phys.*, 24:185.

[Defossez et al., 1999] Defossez, P. A., Prusty, R., Kaeberlein, M., Lin, S. J., Ferrigno, P., Silver, P. A., Keil, R. L., and Guarente, L. (1999). Elimination of replication block protein Fob1 extends the life span of yeast mother cells. *Mol. Cell*, 3:447–455.

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

[Delcuve et al., 2009] Delcuve, G. P., Rastegar, M., and Davie, J. R. (2009). Epigenetic control. *J. Cell. Physiol.*, 219:243–250.

[Doi and Edwards, 1986] Doi, M. and Edwards, S. F. (1986). *The Theory of Polymer Dynamics*. Oxford University Press.

[Doye et al., 1994] Doye, V., Wepf, R., and Hurt, E. C. (1994). A novel nuclear pore protein Nup133p with distinct roles in poly(A)+ RNA transport and nuclear pore distribution. *EMBO J.*, 13:6062–6075.

[Falcón and Aris, 2003] Falcón, A. A. and Aris, J. P. (2003). Plasmid accumulation reduces life span in *Saccharomyces cerevisiae*. *J. Biol. Chem.*, 278:41607–41617.

[Feller, 1968] Feller, W. (1968). *An Introduction to Probability Theory and Its Applications: 1*. Wiley & Sons.

[Feuerhahn and Egly, 2008] Feuerhahn, S. and Egly, J. M. (2008). Tools to study DNA repair: what's in the box? *Trends Genet.*, 24:467–474.

[Fitzgerald-Hayes et al., 1982] Fitzgerald-Hayes, M., Clarke, L., and Carbon, J. (1982). Nucleotide sequence comparisons and functional analysis of yeast centromere DNAs. *Cell*, 29:235–244.

[Flemming, 1879] Flemming, W. (1879). Ueber das Verhalten des Kerns bei der Zelltheilung und über die Bedeutung mehrkerniger Zellen. *Arch. Pathol. Anat. Physiol.*, 77:1–29.

[Franklin and Gosling, 1953] Franklin, R. E. and Gosling, R. G. (1953). Molecular configuration in sodium thymonucleate. *Nature*, 171:740–741.

[Fürth, 1926] Fürth, R. H., editor (1926). *Investigations on the Theory of Brownian Movement*. Methuen.

[Gartenberg et al., 2004] Gartenberg, M. R., Neumann, F. R., Laroche, T., Blaszczyk, M., and Gasser, S. M. (2004). Sir-mediated repression can occur independently of chromosomal and subnuclear contexts. *Cell*, 119:955–967.

[Gasser, 2002] Gasser, S. M. (2002). Visualizing chromatin dynamics in interphase nuclei. *Science*, 296:1412–1416.

[Gehlen et al., *in preparation*] Gehlen, L. R., Nagai, S., Taddei, A., Gartenberg, M. R., and Gasser, S. M. The asymmetric segregation of *ARS* plasmids is a consequence of nuclear geometry and can be overcome by attachment either to telomeres or to the nuclear envelope.

[Gehlen et al., 2006] Gehlen, L. R., Rosa, A., Klenin, K., Langowski, J., Gasser, S. M., and Bystricky, K. (2006). Spatially confined polymer chains: implications of chromatin fibre flexibility and peripheral anchoring on telomere-telomere interaction. *Journal of Physics: Condensed Matter*, 18(14):S245–S252.

146

[Gordon, 1977] Gordon, C. N. (1977). Chromatin behaviour during the mitotic cell cycle of *Saccharomyces cerevisiae*. *J. Cell. Sci.*, 24:81–93.

[Gotta et al., 1996] Gotta, M., Laroche, T., Formenton, A., Maillet, L., Scherthan, H., and Gasser, S. M. (1996). The clustering of telomeres and colocalization with Rap1, Sir3, and Sir4 proteins in wild-type *Saccharomyces cerevisiae*. *J. Cell Biol.*, 134:1349–1363.

[Gotta et al., 1999] Gotta, M., Laroche, T., and Gasser, S. M. (1999). Analysis of nuclear organization in *Saccharomyces cerevisiae*. *Meth. Enzymol.*, 304:663–672.

[Grosberg and Khokhlov, 1994] Grosberg, A. Yu.. and Khokhlov, A. R. (1994). *Statistical Physics of Macromolecules*. AIP Press.

[Gönczy and Rose, 2005] Gönczy, P. and Rose, L. S. (2005). Asymmetric cell division and axis formation in the embryo. *WormBook*, pages 1–20.

[Haber, 1995] Haber, J. E. (1995). *In vivo* biochemistry: physical monitoring of recombination induced by site-specific endonucleases. *Bioessays*, 17:609–620.

[Haber, 1998] Haber, J. E. (1998). Mating-type gene switching in *Saccharomyces cerevisiae*. *Annu. Rev. Genet.*, 32:561–599.

[Hediger et al., 2002] Hediger, F., Neumann, F. R., Van Houwe, G., Dubrana, K., and Gasser, S. M. (2002). Live imaging of telomeres: yKu and Sir proteins define redundant telomere-anchoring pathways in yeast. *Curr. Biol.*, 12:2076–2089.

[Hediger et al., 2004] Hediger, F., Taddei, A., Neumann, F. R., and Gasser, S. M. (2004). Methods for visualizing chromatin dynamics in living yeast. *Meth. Enzymol.*, 375:345–365.

[Heideker et al., 2007] Heideker, J., Lis, E. T., and Romesberg, F. E. (2007). Phosphatases, DNA damage checkpoints and checkpoint deactivation. *Cell Cycle*, 6:3058–3064.

[Hernebring et al., 2006] Hernebring, M., Brolén, G., Aguilaniu, H., Semb, H., and Nyström, T. (2006). Elimination of damaged proteins during differentiation of embryonic stem cells. *Proc. Natl. Acad. Sci. U.S.A.*, 103:7700–7705.

[Hershey and Chase, 1952] Hershey, A. D. and Chase, M. (1952). Independent function of viral protein and nucleic acid on growth of bacteriophage. *J. Gen. Physiol.*, 36:39–56.

[Heun et al., 2001] Heun, P., Laroche, T., Shimada, K., Furrer, P., and Gasser, S. M. (2001). Chromosome dynamics in the yeast interphase nucleus. *Science*, 294:2181–2186.

[Hoeijmakers, 2001] Hoeijmakers, J. H. (2001). Genome maintenance mechanisms for preventing cancer. *Nature*, 411:366–374.

[Hom et al., 2007] Hom, E. F., Marchis, F., Lee, T. K., Haase, S., Agard, D. A., and Sedat, J. W. (2007). AIDA: an adaptive image deconvolution algorithm with application to multi-frame and three-dimensional data. *J Opt Soc Am A Opt Image Sci Vis*, 24:1580–1600.

[Hsiao and Carbon, 1979] Hsiao, C. L. and Carbon, J. (1979). High-frequency transformation of yeast by plasmids containing the cloned yeast ARG4 gene. *Proc. Natl. Acad. Sci. U.S.A.*, 76:3829–3833.

[Hüttenhofer et al., 2005] Hüttenhofer, A., Schattner, P., and Polacek, N. (2005). Non-coding RNAs: hope or hype? *Trends Genet.*, 21:289–297.

[Jaspersen and Winey, 2004] Jaspersen, S. L. and Winey, M. (2004). The budding yeast spindle pole body: structure, duplication, and function. *Annu. Rev. Cell Dev. Biol.*, 20:1–28.

[Jazwinski, 1993] Jazwinski, S. M. (1993). The genetics of aging in the yeast *Saccharomyces cerevisiae*. *Genetica*, 91:35–51.

[Kaeberlein et al., 1999] Kaeberlein, M., McVey, M., and Guarente, L. (1999). The SIR2/3/4 complex and SIR2 alone promote longevity in *Saccharomyces cerevisiae* by two different mechanisms. *Genes Dev.*, 13:2570–2580.

[Kalocsay et al., 2009] Kalocsay, M., Hiller, N. J., and Jentsch, S. (2009). Chromosome-wide Rad51 spreading and SUMO-H2A.Z-dependent chromosome fixation in response to a persistent DNA double-strand break. *Mol. Cell*, 33:335–343.

[Kennedy et al., 1994] Kennedy, B. K., Austriaco, N. R., and Guarente, L. (1994). Daughter cells of *Saccharomyces cerevisiae* from old mothers display a reduced life span. *J. Cell Biol.*, 127:1985–1993.

[Kennedy et al., 1997] Kennedy, B. K., Gotta, M., Sinclair, D. A., Mills, K., McNabb, D. S., Murthy, M., Pak, S. M., Laroche, T., Gasser, S. M., and Guarente, L. (1997). Redistribution of silencing proteins from telomeres to the nucleolus is associated with extension of life span in *S. cerevisiae*. *Cell*, 89:381–391.

[Kim et al., 1999] Kim, S., Benguria, A., Lai, C. Y., and Jazwinski, S. M. (1999). Modulation of life-span by histone deacetylase genes in *Saccharomyces cerevisiae*. *Mol. Biol. Cell*, 10:3125–3136.

[Klenin et al., 1998] Klenin, K., Merlitz, H., and Langowski, J. (1998). A Brownian dynamics program for the simulation of linear and circular DNA and other wormlike chain polyelectrolytes. *Biophys. J.*, 74:780–788.

[Klenin and Langowski, 2001] Klenin, K. V. and Langowski, J. (2001). Kinetics of intrachain reactions of supercoiled DNA: Theory and numerical modeling. *The Journal of Chemical Physics*, 114(11):5049–5060.

[Klenin and Langowski, 2004] Klenin, K. V. and Langowski, J. (2004). Modeling of intramolecular reactions of polymers: an efficient method based on Brownian dynamics simulations. *J Chem Phys*, 121:4951–4960.

[Koç et al., 2004] Koç, A., Wheeler, L. J., Mathews, C. K., and Merrill, G. F. (2004). Hydroxyurea arrests DNA replication by a mechanism that preserves basal dNTP pools. *J. Biol. Chem.*, 279:223–230.

[Lai et al., 2002] Lai, C. Y., Jaruga, E., Borghouts, C., and Jazwinski, S. M. (2002). A mutation in the ATP2 gene abrogates the age asymmetry between mother and daughter cells of the yeast *Saccharomyces cerevisiae*. *Genetics*, 162:73–87.

[Laun et al., 2001] Laun, P., Pichova, A., Madeo, F., Fuchs, J., Ellinger, A., Kohlwein, S., Dawes, I., Fröhlich, K. U., and Breitenbach, M. (2001). Aged mother cells of *Saccharomyces cerevisiae* show markers of oxidative stress and apoptosis. *Mol. Microbiol.*, 39:1166–1173.

[Lindahl, 1993] Lindahl, T. (1993). Instability and decay of the primary structure of DNA. *Nature*, 362:709–715.

[Lisby et al., 2003] Lisby, M., Mortensen, U. H., and Rothstein, R. (2003). Colocalization of multiple DNA double-strand breaks at a single Rad52 repair centre. *Nat. Cell Biol.*, 5:572–577.

[Lisby et al., 2001] Lisby, M., Rothstein, R., and Mortensen, U. H. (2001). Rad52 forms DNA repair and recombination centers during S phase. *Proc. Natl. Acad. Sci. U.S.A.*, 98:8276–8282.

[Lord and Wheals, 1981] Lord, P. G. and Wheals, A. E. (1981). Variability in individual cell cycles of *Saccharomyces cerevisiae*. *J. Cell. Sci.*, 50:361–376.

[Luger, 2003] Luger, K. (2003). Structure and dynamic behavior of nucleosomes. *Curr. Opin. Genet. Dev.*, 13:127–135.

[Lundin et al., 2005] Lundin, C., North, M., Erixon, K., Walters, K., Jenssen, D., Goldman, A. S., and Helleday, T. (2005). Methyl methanesulfonate (MMS) produces heat-labile DNA damage but no detectable in vivo DNA double-strand breaks. *Nucleic Acids Res.*, 33:3799–3811.

[Macara and Mili, 2008] Macara, I. G. and Mili, S. (2008). Polarity and differential inheritance – universal attributes of life? *Cell*, 135:801–812.

[Madhani, 2006] Madhani, H. (2006). *From a to α: Yeast as a Model for Cellular Differentiation.* Cold Spring Harbor Laboratory Press.

[Maringele and Lydall, 2002] Maringele, L. and Lydall, D. (2002). EXO1-dependent single-stranded DNA at telomeres activates subsets of DNA damage and spindle checkpoint pathways in budding yeast yku70Δ mutants. *Genes Dev.*, 16:1919–1933.

[Marshall et al., 1997] Marshall, W. F., Straight, A., Marko, J. F., Swedlow, J., Dernburg, A., Belmont, A., Murray, A. W., Agard, D. A., and Sedat, J. W. (1997). Interphase chromosomes undergo constrained diffusional motion in living cells. *Curr. Biol.*, 7:930–939.

[Mateos-Langerak et al., 2009] 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.

[McMurray and Gottschling, 2003] McMurray, M. A. and Gottschling, D. E. (2003). An age-induced switch to a hyper-recombinational state. *Science*, 301:1908–1911.

[Meister et al., 2010] Meister, P., Gehlen, L. R., Varela-Sanz, E., Kalck, V., and Gasser, S. M. (2010). *Visualizing yeast chromosomes and nuclear architecture* in Methods in Enzymology.

[Merlitz et al., 2006] Merlitz, H., Klenin, K. V., Wu, C. X., and Langowski, J. (2006). Facilitated diffusion of DNA-binding proteins: efficient simulation with the method of excess collisions. *J Chem Phys*, 124:134908.

[Michaelis et al., 1997] Michaelis, C., Ciosk, R., and Nasmyth, K. (1997). Cohesins: chromosomal proteins that prevent premature separation of sister chromatids. *Cell*, 91:35–45.

[Miescher, 1871] Miescher, F. (1871). Ueber die chemische Zusammensetzung des Eiters. *Med.-Chem. Unters.*, 4:441–460.

[Morgan, 2006] Morgan, D. O. (2006). *The Cell Cycle: Principles of Control.* Oxford University Press.

[Müller, 1985] Müller, I. (1985). Parental age and the life-span of zygotes of *Saccharomyces cerevisiae*. *Antonie Van Leeuwenhoek*, 51:1–10.

[Murray and Szostak, 1983] Murray, A. W. and Szostak, J. W. (1983). Pedigree analysis of plasmid segregation in yeast. *Cell*, 34:961–970.

[Nagai et al., 2008] Nagai, S., Dubrana, K., Tsai-Pflugfelder, M., Davidson, M. B., Roberts, T. M., Brown, G. W., Varela, E., Hediger, F., Gasser, S. M., and Krogan, N. J. (2008). Functional targeting of DNA damage to a nuclear pore-associated SUMO-dependent ubiquitin ligase. *Science*, 322:597–602.

[Nassif et al., 1994] Nassif, N., Penney, J., Pal, S., Engels, W. R., and Gloor, G. B. (1994). Efficient copying of nonhomologous sequences from ectopic sites via P-element-induced gap repair. *Mol. Cell. Biol.*, 14:1613–1625.

[Neumann et al., *in preparation*] Neumann, F. R., Gehlen, L. R., Tsai-Pflugfelder, M., Dion, V., Taddei, A., and Gasser, S. M. A targeted INO80 remodeler increases the dynamic range of chromatin and enhances homologous recombination.

[Neumann et al., 2006] Neumann, F. R., Hediger, F., Taddei, A., and Gasser, S. M. (2006). *Tracking Individual Chromosomes with Integrated Arrays of lacop sites and GFP-lacI repressor: Analysing Position and Dynamics of Chromosomal Loci in Saccharomyces cerevisae* in Cell Biology A Laboratory Handbook, Vol. 2. Elsevier.

[Palladino et al., 1993] Palladino, F., Laroche, T., Gilson, E., Axelrod, A., Pillus, L., and Gasser, S. M. (1993). SIR3 and SIR4 proteins are required for the positioning and integrity of yeast telomeres. *Cell*, 75:543–555.

[Partridge and Barton, 1993] Partridge, L. and Barton, N. H. (1993). Optimality, mutation and the evolution of ageing. *Nature*, 362:305–311.

[Porro et al., 2009] Porro, D., Vai, M., Vanoni, M., Alberghina, L., and Hatzis, C. (2009). Analysis and modeling of growing budding yeast populations at the single cell level. *Cytometry A*, 75:114–120.

[Priestley, 1982] Priestley, M. B., editor (1982). *Spectral Analysis and Time Series*. Academic Press.

[Riedel et al., 2006] Riedel, C. G., Katis, V. L., Katou, Y., Mori, S., Itoh, T., Helmhart, W., Gálová, M., Petronczki, M., Gregan, J., Cetin, B., Mudrak, I., Ogris, E., Mechtler, K., Pelletier, L., Buchholz, F., Shirahige, K., and Nasmyth, K. (2006). Protein phosphatase 2A protects centromeric sister chromatid cohesion during meiosis I. *Nature*, 441:53–61.

[Rivetti et al., 1996] Rivetti, C., Guthold, M., and Bustamante, C. (1996). Scanning force microscopy of DNA deposited onto mica: equilibration versus kinetic trapping studied by statistical polymer chain analysis. *J. Mol. Biol.*, 264:919–932.

[Roberts, 1965] Roberts, H. F. (1965). *Plant Hybridization Before Mendel*. Hafner.

[Robinson et al., 2006] Robinson, P. J., Fairall, L., Huynh, V. A., and Rhodes, D. (2006). EM measurements define the dimensions of the "30-nm" chromatin fiber: evidence for a compact, interdigitated structure. *Proc. Natl. Acad. Sci. U.S.A.*, 103:6506–6511.

[Rohner et al., 2008] Rohner, S., Gasser, S. M., and Meister, P. (2008). Modules for cloning-free chromatin tagging in *Saccharomyces cerevisiae*. *Yeast*, 25:235–239.

[Rosa et al., 2006] Rosa, A., Maddocks, J. H., Neumann, F. R., Gasser, S. M., and Stasiak, A. (2006). Measuring limits of telomere movement on nuclear envelope. *Biophys. J.*, 90:L24–26.

[Rotne and Prager, 1969] Rotne, J. and Prager, S. (1969). Variational treatment of hydrodynamic interaction in polymers. *J. Chem. Phys.*, 50(11):4831–4837.

[Sage et al., 2005] Sage, D., Neumann, F. R., Hediger, F., Gasser, S. M., and Unser, M. (2005). Automatic tracking of individual fluorescence particles: application to the study of chromosome dynamics. *IEEE Trans Image Process*, 14:1372–1383.

[Schmid et al., 2006] Schmid, M., Arib, G., Laemmli, C., Nishikawa, J., Durussel, T., and Laemmli, U. K. (2006). Nup-PI: the nucleopore-promoter interaction of genes in yeast. *Mol. Cell*, 21:379–391.

[Schober et al., 2009] Schober, H., Ferreira, H., Kalck, V., Gehlen, L. R., and Gasser, S. M. (2009). Yeast telomerase and the SUN domain protein Mps3 anchor telomeres and repress subtelomeric recombination. *Genes Dev.*, 23:928–938.

[Schober et al., 2008] Schober, H., Kalck, V., Vega-Palas, M. A., Van Houwe, G., Sage, D., Unser, M., Gartenberg, M. R., and Gasser, S. M. (2008). Controlled exchange of chromosomal arms reveals principles driving telomere interactions in yeast. *Genome Res.*, 18:261–271.

[Sexton et al., 2007] Sexton, T., Schober, H., Fraser, P., and Gasser, S. M. (2007). Gene regulation through nuclear organization. *Nat. Struct. Mol. Biol.*, 14:1049–1055.

[Shcheprova et al., 2008] Shcheprova, Z., Baldi, S., Frei, S. B., Gonnet, G., and Barral, Y. (2008). A mechanism for asymmetric segregation of age during yeast budding. *Nature*, 454:728–734.

[Shrivastav et al., 2008] Shrivastav, M., De Haro, L. P., and Nickoloff, J. A. (2008). Regulation of DNA double-strand break repair pathway choice. *Cell Res.*, 18:134–147.

[Sinclair and Guarente, 1997] Sinclair, D. A. and Guarente, L. (1997). Extrachromosomal rDNA circles – a cause of aging in yeast. *Cell*, 91:1033–1042.

[Sinclair et al., 1997] Sinclair, D. A., Mills, K., and Guarente, L. (1997). Accelerated aging and nucleolar fragmentation in yeast sgs1 mutants. *Science*, 277:1313–1316.

[Singer et al., 2006] Singer, A., Holcman, D., and Eisenberg, R. S. (2006). Narrow Escape, Part I. *J. Stat. Phys.*, 122:437–463.

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

[Stinchcomb et al., 1979] Stinchcomb, D. T., Struhl, K., and Davis, R. W. (1979). Isolation and characterisation of a yeast chromosomal replicator. *Nature*, 282:39–43.

[Straight et al., 1996] Straight, A. F., Belmont, A. S., Robinett, C. C., and Murray, A. W. (1996). GFP tagging of budding yeast chromosomes reveals that protein-protein interactions can mediate sister chromatid cohesion. *Curr. Biol.*, 6:1599–1608.

[Straight et al., 1997] Straight, A. F., Marshall, W. F., Sedat, J. W., and Murray, A. W. (1997). Mitosis in living budding yeast: anaphase A but no metaphase plate. *Science*, 277:574–578.

[Sutton, 1903] Sutton, W. S. (1903). The chromosomes in heredity. *Biol. Bull.*, 4:231–251.

[Szostak et al., 1983] Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J., and Stahl, F. W. (1983). The double-strand-break repair model for recombination. *Cell*, 33:25–35.

[Taddei et al., 2004] Taddei, A., Hediger, F., Neumann, F. R., Bauer, C., and Gasser, S. M. (2004). Separation of silencing from perinuclear anchoring functions in yeast Ku80, Sir4 and Esc1 proteins. *EMBO J.*, 23:1301–1312.

[Taddei et al., 2009] Taddei, A., Van Houwe, G., Nagai, S., Erb, I., van Nimwegen, E., and Gasser, S. M. (2009). The functional importance of telomere clustering: Global changes in gene expression result from SIR factor dispersion. *Genome Res.*, 19:611–625.

[Therizols et al., 2006] Therizols, P., Fairhead, C., Cabal, G. G., Genovesio, A., Olivo-Marin, J. C., Dujon, B., and Fabre, E. (2006). Telomere tethering at the nuclear periphery is essential for efficient DNA double strand break repair in subtelomeric region. *J. Cell Biol.*, 172:189–199.

[Towbin et al., 2009] Towbin, B. D., Meister, P., and Gasser, S. M. (2009). The nuclear envelope – a scaffold for silencing? *Curr. Opin. Genet. Dev.*, 19:180–186.

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

[Vanoni et al., 1984] Vanoni, M., Vai, M., and Frascotti, G. (1984). Effects of temperature on the yeast cell cycle analyzed by flow cytometry. *Cytometry*, 5:530–533.

[Watson et al., 2003] Watson, J. D., Baker, T. A., Bell, S. P., Gann, A., Levine, M., and Losick, R. (2003). *Molecular Biology of the Gene.* Addison-Wesley Longman, 5. edition.

[Watson and Crick, 1953] Watson, J. D. and Crick, F. H. C. (1953). Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature*, 171:737–738.

[Widom, 1998] Widom, J. (1998). Structure, dynamics, and function of chromatin in vitro. *Annu Rev Biophys Biomol Struct*, 27:285–327.

[Zakian and Kupfer, 1982] Zakian, V. A. and Kupfer, D. M. (1982). Replication and segregation of an unstable plasmid in yeast. *Plasmid*, 8:15–28.

[Zakian and Scott, 1982] Zakian, V. A. and Scott, J. F. (1982). Construction, replication, and chromatin structure of TRP1 RI circle, a multiple-copy synthetic plasmid derived from *Saccharomyces cerevisiae* chromosomal DNA. *Mol. Cell. Biol.*, 2:221–232.

[Zhou and Elledge, 2000] Zhou, B. B. and Elledge, S. J. (2000). The DNA damage response: putting checkpoints in perspective. *Nature*, 408:433–439.

Acknowledgments

I would like to thank my supervisor, Susan Gasser, for inviting me into her research group for my PhD work. She once said that me, a physicist, doing my PhD in her lab was an experiment for both of us, and I think she was right. I hope that I am already allowed to say that it was a successful experiment. I would like to thank Susan for getting into this venture, for giving me the freedom to find my own way through it, and for the trust into what I was doing to large extent on my own.

I thank my thesis committee consisting — in addition to Susan — of Jörg Langowski, Dirk Schübeler, and Konstantin Klenin for accompanying me through the time of my PhD. I like to work on my own and tend to only ask questions when I am really stuck, but I want to thank my committee members for helping me in these situations and for giving me the feeling that they would always try to help if I would ask for it. Furthermore, I thank Erik van Nimwegen for chairing my thesis defense.

I would like to thank all Gasser lab members and especially those who share or shared my office, in which I have spent most of my working time. I find it really amazing what a good overall atmosphere we have in the group, which contributed a lot to making my time in Basel very enjoyable.

A big part of my thesis work consisted of collaborations within the group. I would like to thank Frank Neumann, who guided me at the beginning of my PhD and provided me with an initial small project, which made my start much easier. I thank Haico van Attikum and Monika Tsai for teaching me during my short "lab sabbaticals" with great competence and patience. I am thankful to Shigeki Nagai and Angela Taddei for their collaboration within the plasmid segregation project, probably my favorite project of my PhD. I thank Benjamin Towbin not only for reinspiring me to the zoning assay analysis, but especially for his critical and insightful interest also in the details of my work, which I always found very motivating. I wish to thank Kerstin Bystricky for thinking along the lines of computational modeling in biology and suggesting interdisciplinary projects. Finally, I would like to thank Haico, Ben, and Vincent Dion for their collaboration within the homology search project.

I thank the members of the FMI IT facility and specifically Dean Flanders, who always had an open ear for my special requests.

Special thanks go to Michael Stadler and Ed Oakeley. The work of the members of our computational biology progress report is related by the use of computers, but not by much more. The insight and motivation with which Michael and Ed attend the presentations of people who have completely different projects than themselves were always a great source of encouragement and motivation for me. I am very thankful. In this context, I would also like to thank Samuel Straumann. During all these years and surrounded by biologists, Sämi was my Basel connection to mathematics and — maybe more to my pleasure than to his — to physics.

I would like to thank Stephanie Küng, Vincent Dion, Anna Maria Friedel, and Juergen Reiter for carefully reading my thesis and giving me a lot of useful comments and Michael Doyle for a very useful discussion. I also thank the community members of Matroids Matheplanet[1], especially Stefan_K and Marco_D. They have a really amazing knowledge of LaTeX and helped me competently and tremendously fast with all my LaTeX problems. I am very thankful to all of these people, who helped me a lot with writing this thesis, but I really want to give special thanks to Brietta Pike, who tirelessly accompanied me through the last days of thesis writing and provided a lot of help and encouragement.

---

[1] `http://www.matheplanet.com`

Curriculum Vitae
Lutz Robert Gehlen

Date of birth: 05 May 1977
Nationality: German
Address: Beim Letziturm 10
4052 Basel, Switzerland
+41 61 697 6975
lutz.gehlen@fmi.ch

Education

PhD in biophysics (magna cum laude) — **May 2009**
University of Basel — Basel, Switzerland
research performed at **Friedrich Miescher Institute for Biomedical Research** under guidance of Prof. Susan Gasser
Project title: "Biophysical analysis of diffusion controlled processes in the budding yeast nucleus"

Diploma in physics (very good with distinction) — **February 2004**
Ruprecht-Karls-Universität — Heidelberg, Germany
Diploma thesis with Prof. Jörg Langowski at the **German Cancer Research Center**, Heidelberg: "Study of the mobility of chromosomes in *Saccharomyces cerevisiae*: Fluorescence correlation spectroscopy and Brownian dynamics simulations"

ERASMUS Internship — **August 2001 – March 2002**
University of Uppsala — Uppsala, Sweden

Intermediate Diploma (Vordiplom) (very good) — **June 1999**
Ruprecht-Karls-Universität — Heidelberg, Germany

Experience

Post-doctoral fellow at — **since June 2009**
Friedrich Miescher Institute for Biomedical Research in the group of Prof. Susan Gasser
Computational Biology: Spatial organization of the nucleus and its dynamics, diffusion controlled processes
Computational modeling and quantification of imaging data

Graduate studies in biophysics — **October 2004 - May 2009**
Friedrich Miescher Institute for Biomedical Research under guidance of Prof. Susan Gasser (molecular biology), Prof. Jörg Langowski (German Cancer Research Center, Heidelberg, biophysics), and Dr. Konstantin Klenin (Research Center Karlsruhe, theoretical physics) — Basel, Switzerland
Project title: "Biophysical analysis of diffusion controlled processes in the budding yeast nucleus"
Interdisciplinary research on nuclear organization in yeast using computational modeling (focus on polymer chain and random walk models), fluorescence microscopy on living cells, and mathematical and computational analysis of fluorescence imaging data

Software developer at — **August 2004 – December 2004**
TrekStor GmbH & Co. KG — Lorsch, Germany
Project: Implementation of a return merchandise authorization (RMA) system using Perl/MySQL
Design of the underlying relational database, design and implementation of the browser based frontend, object oriented design and implementation of the processing, storage, and display engine in Perl

Linux System Administrator (voluntary work) at — **since August 2002**
Nightline Heidelberg e.V. — Heidelberg, Germany
Linux server administration including
Installation and maintenance of a Linux server including a web server (Apache), mail server (Postfix), mailing list server (Mailman), wiki based content management including own plugins (Foswiki), database server (MySQL)

Publications

1. **Gehlen, L.R.**, Gasser, S.M.: "Techniques and errors for the determination of locus position within a spherical nucleus"
   in preparation for Nature Methods
2. Neumann, F.R., **Gehlen, L.R.**, Tsai-Pflugfelder, M., Dion, V., Taddei, A., Gasser, S.M.: "A targeted INO80 remodeler increases the dynamic range of chromatin and enhances homologous recombination"
   in preparation
3. **Gehlen, L.R.**[*], Nagai, S.[*], Shimada, K., Taddei, A., Gasser, S.M.: "Nuclear geometry and rapid mitosis ensure asymmetric episome segregation and rejuvenation in yeast."
   under review at Science

4. Meister, P., Gehlen, L.R. Varela, E., Kalck, V. and S. M. Gasser (2009): "Visualizing yeast chromosomes and nuclear architecture", in Methods in Enzymology, Guide to Yeast Genetics, Guthrie, C.and Fink, G., Academic Press
    in press
5. Schober, H., Kalck, V., Ferreira, H., **Gehlen, L.R.**, Gasser, S.M.: "A role for telomerase in telomere anchoring in budding yeast."
    Genes Dev. 2009 Apr; 23(8):928-38.
6. Bystricky, K., van Attikum, H., Montiel, M.D., Dion, V., **Gehlen, L.**, Gasser, S.M..: "Regulation of nuclear positioning and dynamics of the silent mating type loci by the yeast Ku70/Ku80 complex."
    Mol Cell Biol. 2009 Feb;29(3):835-48.
7. **Gehlen, L.R.**, Rosa, A., Klenin, K., Langowski, J., Gasser, S.M., Bystricky, K.: "Spatially confined polymer chains: implications of chromatin fibre flexibility and peripheral anchoring on telomere–telomere interaction."
    J. Phys.: Condens. Matter 2006, 18 S245-S252
8. Bystricky, K., Heun, P., **Gehlen, L.**, Langowski, J., Gasser S.M.: "Long-range compaction and flexibility of interphase chromatin in budding yeast analyzed by high-resolution imaging techniques."
    Proc Natl Acad Sci U S A. 2004 Nov 23;101(47):16495-500

[*] these authors contributed equally to the work

International Conferences

1. Gordon Research Conference: Chromosome Dynamics (**2009**, Lucca, Italy) – **oral presentation**
2. Lorentz Workshop: The physics of genome folding and function (**2008**, Leiden, Netherlands)
3. Biology in Motion (**2007**, Evian, France) – **oral presentation**
4. EMBO Workshop: Systems Biology Meets Chromatin Function (**2006**, Gosau, Austria) – **invited speaker**
5. Summer School: DNA and Chromosomes: Physical and Biological Approaches (**2006**, Cargèse, France) – **poster**
6. Elucidating Biomolecular Networks by Single-Molecule Technologies (**2003**, Ascona, Switzerland) – **poster**

Teaching Experience

1. Tutor in a practical course "Physics for medical and dental students" – May - August 2002
2. Tutor in "Linear Algebra" – May - August 2001

Professional Affiliations

1. Swiss Society for Biochemistry (SSB)
2. German Physical Society (DPG)

Computational Skills

Main fields of research
*   Brownian dynamics and Monte Carlo simulations of polymer chains
*   Random walks as a model for diffusion controlled processes

Programming skills
1. Perl – excellent
2. C++ – very good
3. Visual Basic – good
4. PHP – good

Spoken Languages

1. German – mother tongue
2. English – fluent
3. Swedish – good
4. French – poor