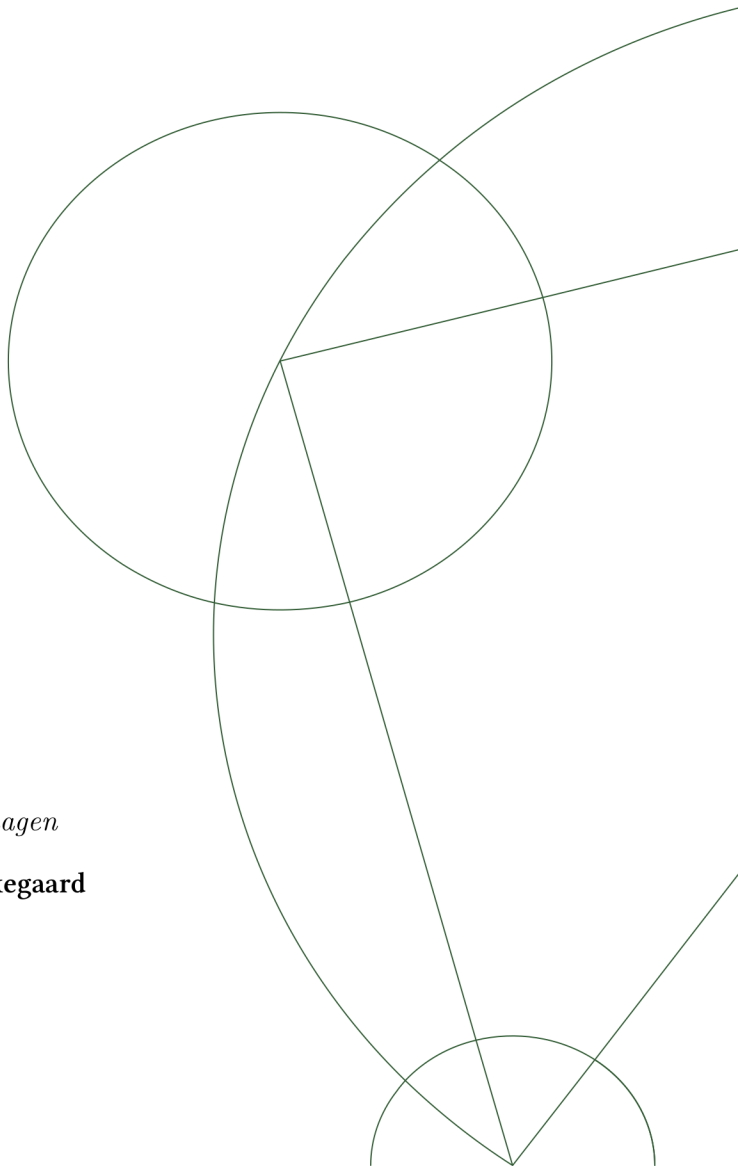
UNIVERSITY OF COPENHAGEN
FACULTY OF SCIENCE

Computational Nucleosome Dynamics in 3D

Epigenetic Switching and Pressure Potential Effects

Master's Thesis in Physics



Lars Erik Johnsen Skjegstad
zfj803@alumni.ku.dk
Niels Bohr Institute, University of Copenhagen

Advisors: Kim Sneppen and Julius Bier Kirkegaard

Submitted: May 20, 2022

Abstract

Nucleosome dynamics and a computational 3D polymer were coupled to model heterochromatic establishment in a genetic region. It was investigated whether qualitative features, such as bistability, hysteresis, and epigenetic memory and switching could be produced. Simulation results showed that all of these features could be modelled, and a parameter regime could be found which additionally yielded a monostable system following the introduction of an epigenetically silent section in the genetic region, where heterochromatin nucleation, spreading, and subsequent gene silencing were direct consequences.

In addition, the effects of implementing a pressure potential for the Brownian dynamics used in the model were investigated. Differences in the number of global nucleosome-nucleosome interactions, the duration of system switching times, and the distributions of establishment times were used as evidence to support the need for such a potential in this model. Additionally, qualitative and quantitative comparisons with real data were also used as indicators. All results were found to provide support for the argument that the pressure potential is an essential component in the kind of model presented in this thesis.

Acknowledgements

I would like to thank the Center for Models of Life (CMOL) at the Niels Bohr Institute (NBI) for being such an engaging and fun place to work at. Your open, inviting and passionate spirit is truly unique, and is something that has given me joy and inspiration during my time as a master's student.

An especially big thank you to my supervisors Kim Sneppen and Julius B. Kirkegaard, as well as Jan F. Nickels, for our great collaboration. It has been a most interesting time, in which I have developed enormously as a computational physicist. Thank you for always keeping the door open, and being available for questions and discussion, more or less whenever it suited me. Thank you for your big enthusiasm, sound advice, and for sharing the wealth of knowledge that you possess. And last but not least, thank you for accepting me as a master's student in your group on a rather short notice - it really made a world of difference.

Finally, to my wonderful wife, Anna. Thank you for all your unconditional support and love, for all the meals you made me, and for making me a better person. I am forever grateful for having you in my life.

Contents

1 Introduction

Deoxyribonucleic acid, or *DNA* for short, is an essential part of life, in that it constitutes "the primary unit of heredity in organisms of all types" [1]. It has been known to exist ever since it was discovered in 1869 as an important cellular component separate from proteins, by Swiss psychological chemist Friedrich Miescher [2]. The knowledge of its chemical composition was enhanced over the next many decades, until its spatial structure was finally discovered in the middle of the 20th century. The paper describing the double-helix structure of DNA was famously published in *Nature* in 1953 and nine years later, the authors James Watson and Francis Crick, along with Maurice Wilkins (who had contributed with X-ray diffraction patterns used by Watson and Crick), were awarded the Nobel Prize in Physiology or Medicine "for their discoveries concerning the molecular structure of nucleic acids and its significance for information transfer in living material" [3]. Ever since then, the knowledge of DNA has only been strengthened, and it now plays a big role in a wide variety of fields, such as genetics, epidemiology, biotechnology, and archaeology.

Epigenetics constitutes another such field, the term of which has seen a large increase in popularity in scientific literature after the year 2000, even to the degree of being labelled as "fashionable" [4]. It is generally accepted as "the study of changes in gene function that are mitotically and/or meiotically heritable and that do not entail a change in DNA sequence" [5], and it bears a close relation to gene expression by means of the dynamics of these very changes or modifications. A prominent example is the inactivation of one of the two X-chromosomes in female mammalian cells due in part to the presence of methyl groups on the associated proteins of the DNA [6]. The modifications lead to the creation and spreading of *heterochromatin*, or a compact form of DNA and the associated proteins. This is an example of *epigenetic switching*, in that the genetic region is "switched off" due to epigenetic modifications.

1.1 Objectives

In this thesis we study and model the dynamics of epigenetic switching within a genetic region found in the fission yeast *Schizosaccharomyces pombe*. This is an important *model organism*, partly because it is unicellular, which makes it easy to study the genes of a large number of individuals. Furthermore, it is a eukaryote, which makes it possible to relate findings for this organism to other eukaryotic species, such as humans [7].

We here present a computational-physical 3D polymer model of DNA, coupled with epigenetic state dynamics. The study has two main objectives:

1. To develop a computational-physical model of *heterochromatin establish-*

ment on a DNA chain. The model should extend previous knowledge of nucleosome dynamics in 0D to a 3D polymer, taking spatial interactions into account, as well as the biological and physical surroundings of the system in question. The model should be able to reproduce qualitative features, such as *bistability*, *hysteresis*, *epigenetic switching* and *epigenetic memory*.

2. To investigate the conditions under which a 3D model is capable of reproducing real data. More concretely, the effects of adding a *pressure potential* to the Brownian dynamics of the 3D polymer will be studied in greater depth.

1.2 Outline

Following the introduction in chapter 1, an introduction to the relevant biological topics will be provided in chapter 2. Here we will look at the connections between chromatin structure and gene silencing, as well as diving deeper into the mechanisms behind the dynamics of epigenetics.

In chapter 3, we study a 0D model for nucleosome dynamics, a model which defines some powerful basic rules and assumptions, yielding interesting results on epigenetic switches. We will aim to produce similar results in our 3D model, using some of the rules and assumptions described in the simpler model.

Chapter 4 concerns the construction of the 3D model, built for the purpose of this thesis. We will look at the theoretical foundations for the Brownian dynamics and the nucleosome dynamics, as well as the motivations and assumptions made in order to achieve this. This will be followed by a brief discussion about the differences between this model and other 3D polymer models from the literature.

Then follows chapter 5, in which the results from simulations using the 3D model will be presented. Among other things, we look at the resulting spatial structures of the polymers, qualitative traits for epigenetic switches, such as bistability and hysteresis, and the effect of introducing a pressure potential in the implementation of Brownian dynamics. Discussions will follow the presentation of the results.

In chapter 6, conclusions will be made on the findings, and finally, chapter 7 looks ahead at the possible near-future developments and uses for the model.

1.2.1 Source code

The implementation was written in the Python programming language, and the source code can be found on GitHub in the following repository:

`https://github.com/larserik-js/heterochromatin-establishment`

1.2.2 Note on figures

- Figure 2.1 was made using the BioRender platform: `www.biorender.com`.
- Figures 2.2-2.3 were taken from *Molecular Biology of the Cell*, 6th ed. (2015) [8].
- All other figures were created using the MATPLOTLIB library in Python.

2 Background

2.1 DNA, genes, and chromosomes

DNA is found in all prokaryotic and eukaryotic cells, as well as in many viruses, and is what codes genetic information for the transmission of inherited traits [9]. It is sectioned into *genes*, or segments of code in the form of different molecules or *bases*, evenly distributed on double-stranded macromolecules, where each strand forms the template for replication and transcription of mRNA corresponding to the opposite strand. There are four different bases, namely adenine (A), cytosine (C), guanine (G), and thymine (T). Each base forms a *base pair* (bp) with exactly one of the other bases on the opposite strand through hydrogen bonds; A pairs with T, and C pairs with G. All genes contain code on this form, and the collection of genes constitutes "the information that specifies all the RNA molecules and proteins that make up an organism — including information about when, in what types of cells, and in what quantity each RNA molecule and protein is to be made" [8].

The DNA is distributed on *chromosomes*. Chromosomes are big macromolecules, located in the cell nucleus, and consist of a very long section of DNA, as well as different proteins which contribute to holding the structure together. There is a huge variety between organisms in the number of genes, but also in the number of chromosomes. Furthermore, some organisms, like humans, have mostly *diploid* cells, i.e. cells that contain two pairs of each chromosome. Others, like the fission yeast *Schizosaccharomyces pombe*, have *haploid* cells, or only one pair of each chromosome. Humans are multicellular organisms that have 23 pairs of chromosomes, or 46 in total, in each cell, whereas *S. pombe* is a unicellular organism, the genetic material of which is distributed on 3 chromosomes only [10].

2.1.1 Chromatin

The material that makes up the chromosomes is called *chromatin*, and consists of the DNA molecules, together with associated proteins, including *histones*, which are packing proteins, and *non-histone chromosomal proteins.* In total, the chromatin is roughly one-third DNA and two-thirds protein by mass [8]. The structure can generally be divided into two types: *Heterochromatin*, which is a highly condensed form of chromatin, and *euchromatin*, which is less condensed. Heterochromatin is generally found in some specialized regions of the chromosomes, but can also be found in areas that depend on the physiological state of the cell [8]. This is intimately connected to the dynamics of *gene silencing*, as we will see.

Control of gene expression is what makes cell differentiation possible. The

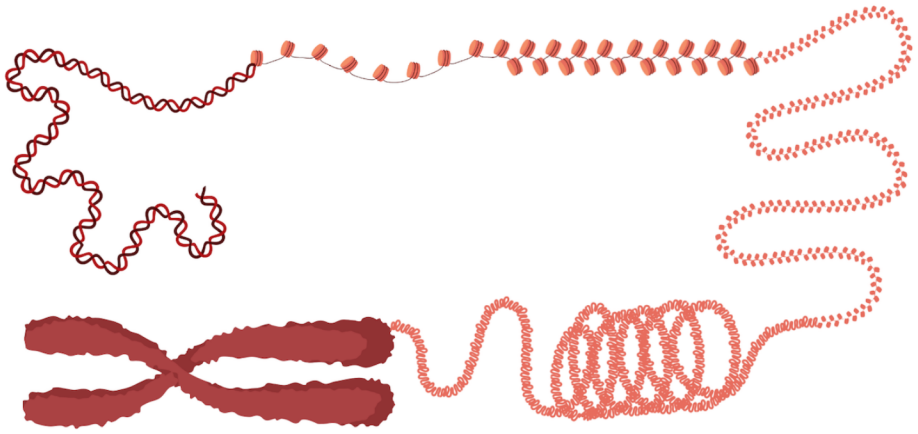


Figure 2.1: The different levels of chromatin structure. The DNA winds around individual nucleosomes, which condense to form heterochromatin. This further folds into coils and loops, and eventually becomes a full chromosome at the top level. (Created on the BioRender platform.)

transcription of genes can be switched on and off by so-called *chromatin modifiers*, of which there exists a large variety. In general, these chromatin modifiers alter the chromatin structure to control the availability of the DNA for transcription - euchromatin is associated with genes that are available for transcription, while heterochromatin is associated with genes that are resistant to transcription, due to its highly condensed form. Indeed, genes are generally switched off in the heterochromatic state [8]. This is a dynamic process: heterochromatin, once initially formed through nucleation within a certain gene, can propagate through the gene in a self-amplifying manner, where the same enzymes which recognize certain modifications associated with heterochromatin also catalyze the addition of the very same modifications in the surroundings. This is thus the basis for a *positive feedback* mechanism, which will be described below.

Both euchromatin and heterochromatin can additionally be *inherited* through cell generations. This means that the compact chromatin can survive through processes such as DNA replication and cell division, both processes which cause significant disruptions to the organization of the chromosomes. This is an example of *epigenetic memory*, and will be elaborated upon in section 2.2.

2.1.2 DNA packing

The two types of chromatin, and thus chromosomes in general are essentially DNA which have been packed together by proteins. It turns out that the DNA has to fold and compact itself to a very high degree to be able to fit inside the cell nucleus (see figure 2.1). For example, the complete human genome from one single cell measures approximately 2 m (if hypothetically stretched out),

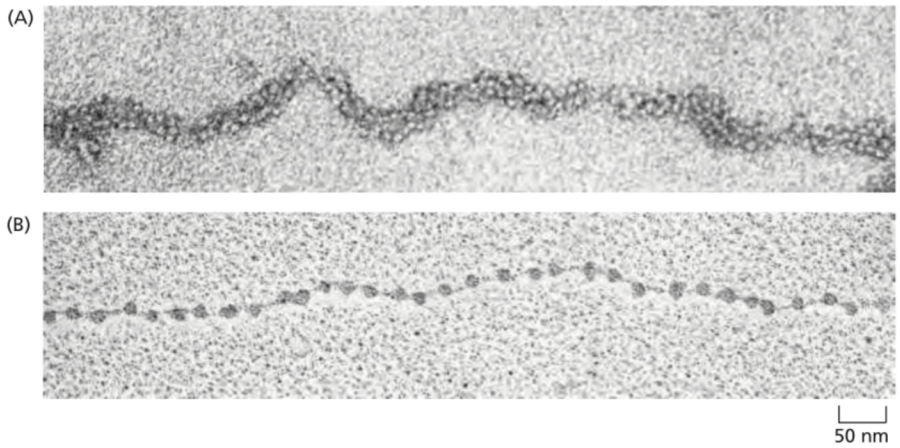


Figure 2.2: Electron microscopy images of **(A)** the heterochromatic fiber, and **(B)** euchromatin, achieved by experimentally unpacking heterochromatin. Note the beads-on-a-string structure of nucleosomes. (From *Molecular Biology of the Cell*, 6th ed. (2015), p. 188 [8]).

whereas the cell nucleus has a diameter of about 6 µm [8]. This packaging is accomplished on several different structural levels, as well as by several different mechanisms.

At the very lowest level the chromatin adapts a beads-on-a-string-structure, which consists of protein-DNA complexes known as *nucleosomes*. Here, a section of DNA (147 bp in humans) is wound around *nucleosome core particles*, which are protein complexes that consist of four different pairs of two equal histones, i.e. eight histones in total. Each histone in turn consists of one core complex with an N-terminal amino acid *tail* [8].

Counting both the DNA which is wound around the nucleosome core particle, as well as the *linker DNA* which runs between these particles, one nucleosome in a human cell on average contains DNA the length of about 200 bp. The nucleosome should thus be thought of as a basic building block for the lowest-level organization and compaction of DNA. Considering the packing due to the nucleosome core particles only, the DNA molecule is reduced to about one-third of its length [8]. In the open beads-on-a-string configuration, this is what can be recognized as euchromatin. However, the nucleosomes can condense even further via nucleosome-nucleosome interactions, and these interactions constitute an essential ingredient in the establishment and spreading of heterochromatin. The fundamental mechanism behind this are biochemical modifications of the nucleosomes, which is a central topic in the field of *epigenetics*. This will be explored in the following.

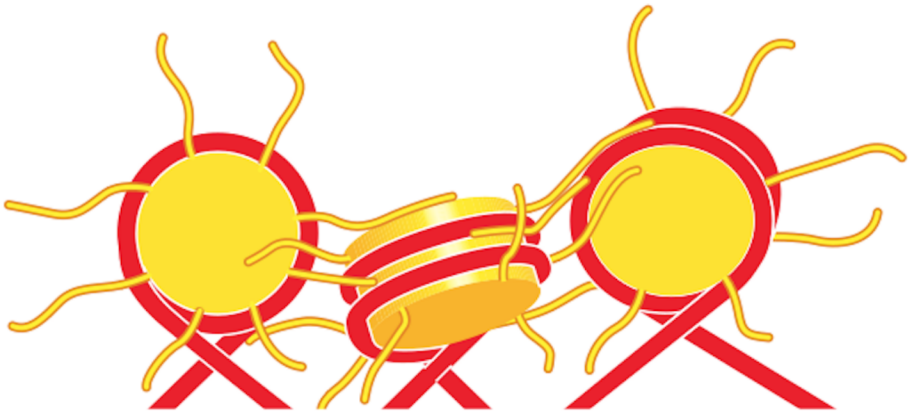


Figure 2.3: Nucleosome core particles, which consist of eight histones, each of which has an amino acid tail. This also illustrates the interactions that lead to the packing of the nucleosomes due to modifications on the tails. (From *Molecular Biology of the Cell*, 6th ed. (2015), p. 193 [8]).

2.2 Epigenetics

Epigenetics is concerned with the changes and modifications that occur on the DNA molecule itself (or on the associated proteins), but crucially does not change the genetic code. These alterations generally comprise "histone variants, posttranslational modifications of amino acids on the amino-terminal tail of histones, and covalent modifications of DNA bases" [11].

Here we look at the covalent modifications of nucleosomes, as this is what induces the interactions that ultimately lead to heterochromatin creation and spreading. As mentioned, nucleosomes can pack together to form more condensed structures; it is believed that the nucleosomes form a semi-regular *zigzag* pattern at this lowest level of organization [8], where nucleosomes located closely together on the beads-on-a-string structure associate via interactions with *bridging proteins*. This makes out a 30 nm fiber, distinguishable from the open euchromatic state [8] (see figure 2.2). One important protein is the *H1 linker histone*, which is *not* one of the histones that are part of the nucleosome core particles, but rather binds directly to the DNA at the sites where the DNA enters and exits the nucleosome core particle, and stabilizes the fiber by controlling the angle between the DNA string on each side of the core particle [12]. Important mechanisms include reversible processes where highly specific enzymes add or remove chemical groups on specific amino acid locations on the aforementioned histone tails, which can lead to bridging interactions between the nucleosomes (see figure 2.3). As an example, acetylation of lysines on the histone tails contributes to less condensed chromatin, partly because adding an acetyl group to lysine removes its positive charge, which further reduces the affinity of the tails for adjacent nucleosomes [8]. The modifications described

in the following section are essential to the aforementioned positive feedback mechanism and self-propagation of heterochromatin, because they can recruit enzymes which are also capable of adding the same modification to other nucleosomes in the environment, in addition to associating with bridging proteins to bind nucleosomes together. This is the basic process of heterochromatin formation and spreading.

2.2.1 Selected modification reactions

The number of epigenetic modifications, reactions, and corresponding enzymes is vast, so for the purpose of this thesis we will look at some select few that are particularly important. The organism of interest in this thesis is the model organism *S. pombe*, as the physical model developed here is largely based on results from studies involving this organism. The reactions in question concern modifications that are involved in promoting heterochromatin, and those that are involved with repressing it. Both of these reactions are reversible, so overall, we can divide the modifications and their corresponding enzymes into four broad groups:

- **Methylation:** This modification is carried out by *histone methyltransferases* (HMTs), and can be identified specifically as the trimethylation of lysine at histone H3K9 (H3K9me3). It is believed that this process can happen in two different places on each nucleosome, due to the fact that the histones come in pairs, as mentioned. HMTs are capable of spreading heterochromatin, as the proteins both recognize H3K9me3 and methylate adjacent nucleosomes [13]. In *S. pombe* the specific enzyme is Clr4, and the methylation of H3K9 additionally attracts the bridging protein Swi6 (an ortholog of HP1) [14].

- **Demethylation:** Demethylation is performed by *histone demethylases* (HDMs). The modification has been attributed to so-called JmjC-domain-containing proteins [15].

- **Acetylation:** The acetylation of specific lysines is carried out by *histone acetyl transferases* (HATs). As mentioned, acetylation contributes to the euchromatic state, as it removes the positive charge of the N-terminal histone tail [8]. Furthermore, some HATs have been shown to bind to acetylated histones, which is evidence for a self-propagating euchromatic state [16].

- **Deacetylation:** The deacetylation of lysines is performed by *histone deacetylase complexes* (HDACs). The absence of specific acetylated lysines is associated with heterochromatin formation, and thus transcriptional silencing. In *S. pombe*, the enzymes Clr3 and Clr6 are associated with heterochromatin formation, in cooperation with HMTs [17].

2.2.2 Heterochromatin nucleation

In this study, we are mainly concerned with a specific region of DNA, namely the *mating-type region* of *S. pombe*. The density of nucleosomes in this organism is on average 6.5 kb^{-1} [18], or approximately 153.8 bp per nucleosome, and the region within which heterochromatin can establish has the size of about 23 kb [19][20], which yields approximately 150 nucleosomes in total. The nucleation occurs at a $\sim$ 4.3 kb region known as the *cenH* region, and heterochromatin subsequently spreads from this element to the rest of the mating-type region [21]. This gives a ratio of the length of the *cenH* region to the entire system of 0.187. Furthermore, the nucleation is partly driven by RNA interference (RNAi) [22], and occurs stochastically, as well as at a high rate [19]. Notably, it has been found that strains that have been reduced to approximately half the original size, as well as lacking the *cenH* region (ΔK *mutants*), display *bistable* behavior (explained below), switching between a euchromatic and a heterochromatic state approximately every 2 000 cell generations on average [23].

2.2.3 Epigenetic switching

Since the nucleation of *cenH* eventually leads to heterochromatin establishment and subsequent silencing of the chromosomal region, and since this event is regulated in the cell, it is an example of an *epigenetic switch*. Two features are associated with epigenetic switching: bistability and hysteresis [24]. A bistable system is a system in which there are two different metastable system states, each of which can exist over a considerable amount of time. It is worth noting that even though a system can be bistable, this does not necessarily imply anything about the ability of the system to *switch* between the states. In fact, even a system that can stay in both its initial states infinitely long would still be a bistable system by this definition. This is important for epigenetic inheritance and memory; the system states should remain stable through disruptive perturbations such as cell division. And when modelling a system with an epigenetic switch, it is desirable to be fully able to control the system states.

Another implication of bistability is that once the system is in one of the states, this state can be sustained by the system itself; switching to the other state requires a change in one of the controlling parameters, e.g. the enzymatic rates of the epigenetic modification reactions. This applies to both states, which means that the transition points for the switches are not equal. This entails a system memory, in that the state of the system depends on the history of the system itself. This effect is referred to as *hysteresis* [8].

Bistability and hysteresis are important traits of epigenetic switches, both of which we will try to produce using our model. We will now take the first steps towards constructing our model, by looking at a physical model for nucleosome dynamics.

3 The 0D model

As should be evident from chapter 2, is that biological systems are complex. We have already seen that there exists a myriad of enzymatic processes involving epigenetic modifications, and that there are modifications that lead to actual physical interactions between the nucleosomes, which in turn leads to the packing structures that constitute heterochromatin. Furthermore, we have seen that these structures are dynamic, and that there is a complex interplay between the individual nucleosome states, which determines the state of the system overall.

It is desirable to build a physical model that can explain these effects. Such a model could take into account a range of different important factors, such as:

- The number of different epigenetic states.
- The epigenetic state of an individual nucleosome at a given point in time.
- The ability for a specific epigenetic state to affect, and possibly change, other specific epigenetic states.
- The physical behavior of a nucleosome while in a specific epigenetic state.
- The enzymatic rates for different modification reactions, and the effect that these have on epigenetic state dynamics.
- Qualitative system traits, such as bistability and hysteresis.

These are just some of the system characteristics that could be interesting to include in our model. Already at this point, however, we see that a significant amount of simplification needs to be undertaken. For example, we know that the number of epigenetic modifications possible is vast. For a sensible physical model to work, we need to reduce the assumed number of modifications, and thus the number of possible epigenetic states. In the following we will look at a model which has used this assumption among others, and that has successfully emulated a system of interacting nucleosomes for the fission yeast *S. pombe*. The model was developed by I. B. Dodd et al. [25], and will constitute an important foundation for the 3D model developed for the purpose of this thesis. The model makes use of some important assumptions, and sets up some useful rules for how the system develops. These assumptions and rules will be outlined in the following.

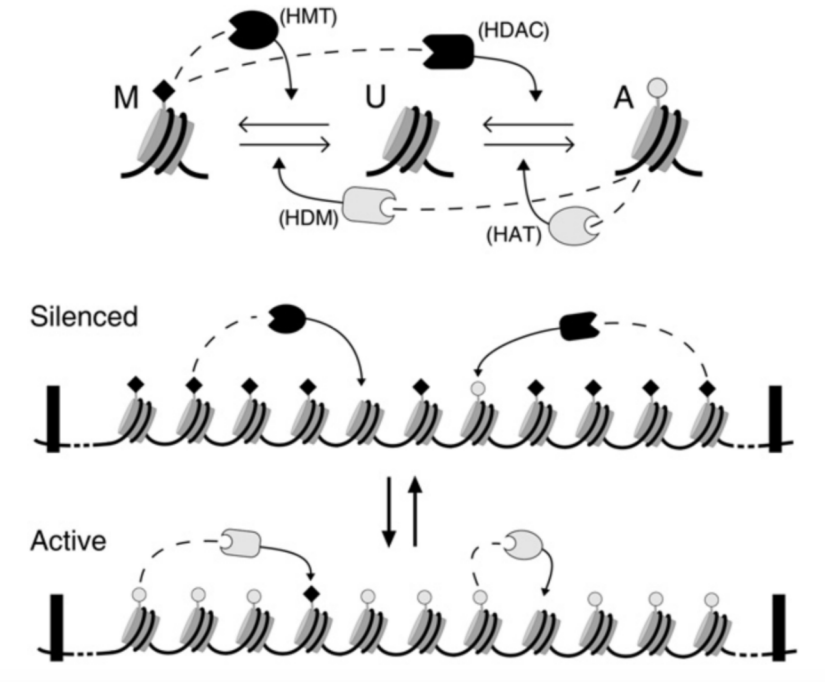


Figure 3.1: Top: Reaction scheme involving the three different states. The four different enzymes that each catalyze one of the reactions are indicated. **Bottom:** The system is bistable, with one overall silenced state, and one overall active state. (Figure from Dodd et al. [25])

3.1 Model assumptions and rules

3.1.1 Assumptions

- The model system is a set of a constant number of monomers. Each monomer represents a nucleosome, and the system in total represents an isolated section of the DNA polymer.
- The model does not include the concept of space, so the monomers have no spatial position, nor a defined distance between them. This is of course a simplification of reality, as we know that DNA polymers are 3D structures. The motivation for this design is the fact that nucleosomes that are not necessarily nearest neighbors on the chain also interact and affect each other, and the 0D approach eliminates distance being a factor.
- Each monomer can be in one of three states: M, U, and A (for *methylated, unmodified,* and *active*, respectively). The states each represent one or more epigenetic modifications, the number of which in reality is much

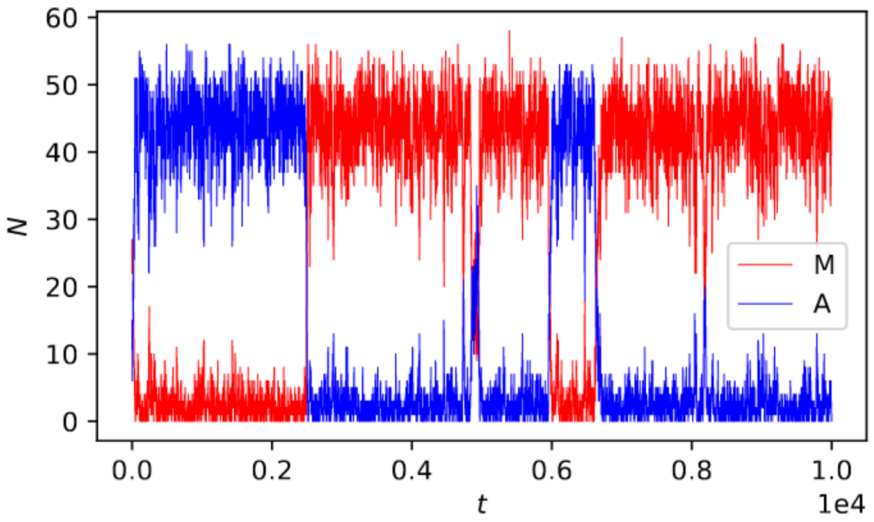


Figure 3.2: The number of nucleosomes in the M and A state for $\alpha = 0.59$ in a system of a total of 60 nucleosomes. Notice the bistability of the system.

larger than three. However, as we will see, only three different behaviors are required for the purpose of this model, so this is also a useful simplification.

3.1.2 Rules

A stochastic simulation is now performed, and the system can develop according to some simple rules:

- For each iteration step, one of two types of conversions is attempted: a *recruited* or a *noisy* conversion.
- A nucleosome n_1 is chosen at random. Then, with probability α, a recruited conversion is attempted. Else (with probability $1 - \alpha$) a noisy conversion is attempted.
- In case of a recruited conversion, another nucleosome n_2 is also chosen at random. If the state of n_2 is M or A, the state of nucleosome n_1 is changed one step towards the state of nucleosome n_2, according to figure 3.1 (top) Thus, if the state of n_1 is U, and the state of n_2 is M, the state of n_1 is changed from U to M. Biologically, this corresponds to the methylation of the nucleosome n_1 by the Clr4 enzyme bound to nucleosome n_2 (see section 2.2.1). Unmodified nucleosomes do not have the capability to change the

states of other nucleosomes, which reflects the assumption that modifying enzymes must first bind to an already modified nucleosome.

- If a noisy conversion is to be attempted, nucleosome n_1 is changed one step in either direction with the probability 1/3.

Implementation and discussion of features

Implementing the model using a system size of 60 nucleosomes with a probability parameter of $\alpha = 0.59$ yields the result seen in figure 3.2. It can be observed that both the overall M state and the A state are stable over longer periods of time. The system state switches back and forth between having most nucleosomes in the M state to having most nucleosomes in the A state, and when the switches occur, they occur fast, and there is no third stable overall state (e.g. the U state never dominates, and the system virtually never stays in a state where half the monomers are in the M and the A state, respectively). This mimics the inherent bistability of the system, and this very simple model is in other words able to reproduce this important trait of epigenetic switching. The observed switches reflect the behavior of the ΔK mutants, as discussed in section 2.2.2.

As the authors note, the key mechanism to model a bistable system is the ability for the nucleosomes to "stimulate modification beyond their neighbor nucleosomes, arguing against a simple continuous spreading of nucleosome modification" [25]. The need for longer-range interactions is an important point, and something that thus has to be included in the 3D model. In the 0D model, it is the very absence of space that enables the monomers to recruit each other over such long distances. With a 3D model, interacting nucleosomes have to be *spatially close*, even though they are located far apart on the *chain*. This is one of the challenges that lie ahead when we now start building the model for the purpose of this thesis.

4 The 3D model

As we have seen, it is possible to reproduce some desirable qualitative attributes of the biological system in question using a relatively simple physical model. Notably, it is possible to achieve a *bistable* system, assuming that nucleosomes interact on a longer range than between nearest neighbors on the chain. It is, however, also desirable to create models that take space into account. After all, chromatin is a 3D structure, so a 3D model should be at least as accurate as a model which does not use space - indeed, it is believed that there exists a correlation between repressive epigenetic modifications and chromatin compaction [8][26]. There are several possible advantages of doing this:

- Nucleosomes have different spatial distances between them, and this makes it likely that the interactions that occur between them have a distance-dependence which is non-linear. It is highly interesting to investigate this effect, and this can be captured most accurately via a 3D model.
- Heterochromatin consists of nucleosomes which are bound together by proteins. Since nucleosome states are highly dynamic, a DNA chain is expected to be a highly complex structure, which may contain areas of both heterochromatin and euchromatin. These structures contribute steric effects to the system, which in turn leads to epigenetic repression due to the correlation with chromatic compaction [24]. In addition, different nucleosome states produce different physical interactions, e.g. the bridging interactions that occur due to methylation (see section 2.2.1), and this would also be desirable to model.
- A 3D model opens the possibility of including knowledge of the actual three-dimensional structure of heterochromatin in its different levels of organization. This would in other words add a biological dimension to the model.

These are some of the motivational questions behind developing a 3D model. Notably, the introduction of space requires, among other things, that the Brownian dynamics of the DNA chain be taken into account. We are thus now ready to develop our own 3D model, in which the rules and assumptions from the 0D model will be combined with the Brownian dynamics of the DNA chain to create a computational-physical polymer which can be used as a model for epigenetic switching, as explained in section 2.2.

4.1 Brownian dynamics

The isolated section of the DNA chain is represented by a polymer consisting of N monomers. In this model these monomers each represent one nucleosome

(although the possibility for considering one monomer as a coarse-grained representation of several nucleosomes can also be conceived, and this has indeed been done in other models [24][26][27]). Each monomer can be in one of three possible states at a given time, each of which represents an epigenetic state (see section 2.2.1):

- **S:** This state represents the 'silent' nucleosome, or nucleosomes that have been epigenetically modified via the methylation of H3K9, such that they condense into a heterochromatic state. Here we will assume that S state monomers are attractive to other S state monomers, and that they are capable of spreading their state to other monomers. Additionally, whenever the *cenH* region is considered activated during simulations, all monomers within this region will be set constantly to the S state.
- **U:** This state represents 'unmodified' nucleosomes, which corresponds to no epigenetic modification. Unmodified nucleosomes are assumed not to condense into heterochromatin, and are thus modelled as non-attractive, sterically repulsive monomers.
- **A:** This state represents 'active', or euchromatic nucleosomes. These nucleosomes are assumed to be epigenetically modified in such a way that they attract the same enzymes that perform the modification, i.e. the HATs. For simplicity it will be assumed that the A state monomers represent acetylated nucleosomes, and that the HATs associate with this chemical group and self-propagates in the same manner as the HMTs do with regards to the nucleosomes modified by methyl-groups. Since acetylated nucleosomes are associated with the euchromatic state [8], the A state monomers will be modelled as non-attractive, but they will still be capable of spreading their state to other monomers.

The movement of the monomers is purely determined by the aggregate potential U, which serves to introduce repulsion and attraction to the system. (These will be described more fully in section 4.1.2.) All positions and distances are measured in units of the preset distance between neighboring monomers, l_0. (The specifics on different parameters and their values will be laid out in section 5.1.1).

4.1.1 Diffusion

Chromosomes exist in the nucleoplasm of the cell nucleus. This is a viscous liquid, which means that a drag force is induced on the moving macromolecules. On the time scale of interest, the inertia-relaxation time of the moving polymer can be neglected, i.e. the movement of the polymer can be assumed to be *non-inertial*, which means that it can be described using Brownian dynamics. This is achieved using the over-damped Langevin equation:

$$\eta \frac{d}{dt}\mathbf{X}_i(t) = -\nabla_{\mathbf{X}_i} U(\mathbf{X}(t)) + \mathbf{R}_i(t) \tag{4.1}$$

Here η is the viscosity, set to unity without loss of generality, $\mathbf{X}_i(t)$ is the position of monomer i at time t, and $\nabla_{\mathbf{X}_i} U$ is the gradient of the aforementioned potential with regards to the position of the i'th monomer. Notably, the potential depends on the position of *all* monomers. Computationally, $\mathbf{X}(t)$ is expressed as an $N \times 3$ matrix $\mathbf{X}(t)$, where N is the number of monomers, and where the i'th row contains the position of the i'th monomer. Lastly, $\mathbf{R}_i(t)$ is the random Gaussian noise on the i'th monomer, where every component R_{ij} satisfies:

$$\langle R_{ij} \rangle = 0, \tag{4.2}$$

and

$$\langle R_{ik}(t) R_{jl}(t') \rangle = 2D\delta(t - t')\delta_{ij}\delta_{kl}, \tag{4.3}$$

where $\delta(t)$ is the Dirac delta function, and D is the diffusion constant. We set $D = 1$, and thus let this define our units.

The system can now evolve through numerical simulations, using the Euler-Maruyama method (for the full derivation, see Appendix A):

$$\mathbf{X}_i(t + \Delta t) = \mathbf{X}_i(t) - \nabla_{\mathbf{X}_i} U(\mathbf{X}(t))\Delta t + \Delta \mathbf{W}_i(t), \tag{4.4}$$

where $\Delta \mathbf{W}$ is Gaussian noise, of which the components dW_{ij} satisfy:

$$\langle dW_{ij} \rangle = 0, \tag{4.5}$$

and:

$$\langle dW_{ij}^2 \rangle = 2\Delta t \tag{4.6}$$

The most involved term in eq. 4.4 is the potential term, which will be explained in the following.

4.1.2 Potentials

The aggregate potential is a sum of two terms: $U = U_{\text{interaction}} + U_{\text{pressure}}$. Notably, the potential experienced by each monomer depends on all of the other monomers and their states. As we have seen in eq. 4.4, it is necessary to calculate the *gradient* of the potential function. In practice, this would take infeasibly long: simulation time is $\mathcal{O}(N^2)$, due to the need to calculate the distances between all monomers at each time step. As it turns out, however, it can be accomplished using *differentiable programming* (see section 4.1.4). First we look at the calculation of the individual potential terms, and then we see exactly how the potential gradient can be found.

Spring rigidity

As we have seen, chromatin at its lowest level of organization is a 'beads-on-a-string' structure. The nucleosomes are connected to each other by the DNA itself, and the position of each nucleosome relative to its nearest neighbors is

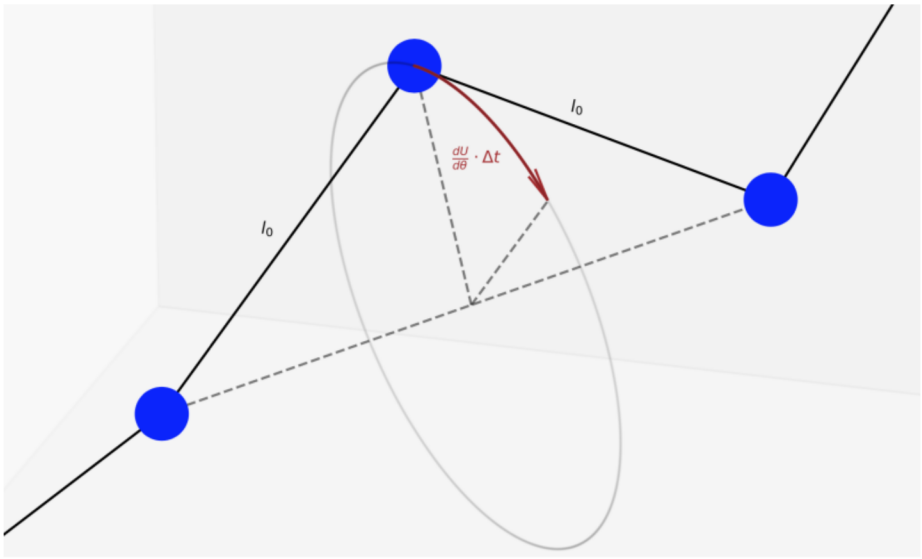


Figure 4.1: Illustration of the incremental movement of a monomer. The monomer always has a fixed distance, l_0, to both of its neighbors, and can only rotate around an axis which goes between its neighboring monomers, as shown.

assumed to be fairly rigid. A standard way to model this would be via a harmonic potential, using a proper spring constant to ensure a stiff spring. While this seems fine in principle, it turned out to be a weakness for the simulations: to obtain the necessary stiffness, the spring constant needed to be set to a relatively high value, which in turn led to numerical blowups, and which further required the finite iteration time step, Δt, to be set to a correspondingly low value. This naturally slowed simulation time drastically, so a choice was made *not* to implement a spring potential at all, but to emulate the beads-on-a-string structure in a completely different way.

The distance between neighboring monomers was thus set to a fixed value, l_0, as mentioned. To maintain this distance at all times, the monomers (except the monomers at the ends of the polymer) were restricted to *rotate* along a circle, with a rotation axis going between the neighboring monomers (see figure 4.1). The monomers at the ends of the polymer could move freely, although still with a fixed distance l_0 to their one neighbor. The result was an infinitely rigid spring, with the added benefit of numerically stable simulations.

Interaction potential

This potential arises due to the different physical interactions between the monomers. Even though there is no defined monomer size in this model, it was assumed that the monomers exhibit repulsive behavior when they get close to each other. In addition, to represent the bridging of two methylated nu-

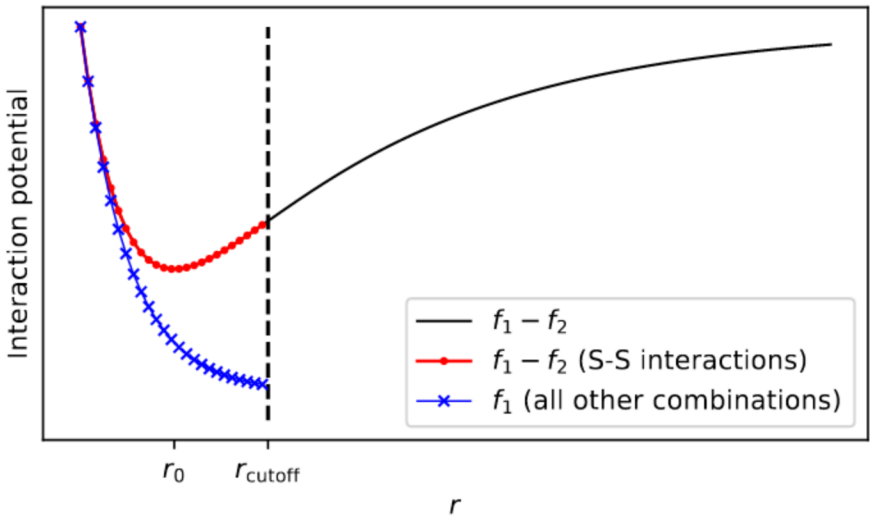


Figure 4.2: A qualitative, one-dimensional representation of the interaction potential in its different forms, shown for the interaction between two monomers. The entire potential function, $U_{\text{interaction}}$, is shown in black. Monomers which both are in the S state experience this potential below the cutoff distance (shown in red), while monomers in all other combinations of states experience the repulsion-only potential (shown in blue).

cleosomes, monomers of state S were assumed to be mutually attractive when within a certain distance of each other. This was achieved by constructing the interaction potential function, $U_{\text{interaction}}$ using two decaying exponential functions,

$$f_1 = \exp\left[\frac{-2r}{r_0}\right] \tag{4.7}$$

and

$$f_2 = \exp\left[\frac{-2r}{br_0}\right], \tag{4.8}$$

where r is the distance between two monomers, r_0 is the equilibrium distance between two monomers (which was preset as a model parameter), and $b = \text{Re}\left[\frac{-2}{\text{LambertW}(-2e^{-2})}\right]$ is a constant which ensures that r_0 becomes the global extremum for the potential function. In total, this gives us:

$$U_{\text{interaction}} = c_{\text{interaction}}\left[\sum_{r\notin A_S} f_1(r) + \sum_{r\in A_S}\left(f_1(r) - f_2(r)\right)\right] \tag{4.9}$$

Here, A_S is the set of all distances between pairs of S state monomers (excluding pairs where the monomers are the same monomer). The constant $c_{\text{interaction}}$ is a model parameter which controls the size of the potential.

Subtracting the exponential functions yields a potential well (see figure 4.2), while the function f_1 by itself is a purely exponentially decaying function, which was used for repulsion. In addition, a cutoff distance was added; monomers within the cutoff distance are affected by the potential, while monomers outside of this distance are not. This was implemented to prevent physical interactions of very long range; in reality these interactions are not physical forces, but rather occur on short distances, as they are the result of enzymes mediating the binding together of nucleosomes.

In summary, the monomers experience two different interaction potentials:

- S monomers interact with each other via a potential well with the equilibrium distance r_0.
- All other combinations of monomers interact with each other repulsively via a decaying exponential function.

As a particular feature in this model, a limit was introduced on how many S state monomers could attract each other at a given time. Based on the assumption that methylation can occur on two different locations for each nucleosome (see section 2.2.1), each S state monomer was only allowed attract a maximum of two other S state monomers at a given time. This had some significant implications on the resulting 3D folding structure of the polymer (see section 5.2). To achieve this, an algorithm was constructed, with the following requirements:

- A set of pairs of interacting S state monomers should be produced.
- No monomer should appear in more than two interaction pairs.
- The algorithm should minimize the total interaction distance for all S state monomers.

The algorithm was defined as follows:

1. Define an initially empty set of pairs of interacting monomers.
2. List all distances between all monomers (excluding distances to self, i.e. distance 0) from smallest to largest.
3. For each distance in the list, starting at the smallest distance, identify the monomers between which the distance in question applies. If both monomers are in state S, and if none of the monomers is already interacting with two other monomers, add the monomer pair to the set of interacting monomers.

Given this algorithm, if there are N_{S} S-state monomers in the system at a given time, the theoretical maximum number of possible S-S interactions at this time is equal to N_{S}, while using our implementation, the number of distances to iterate through to establish the interactions is equal to $\frac{1}{2}(N^2 - N)$, where N is the total number of monomers in the system. Since $\frac{1}{2}(N^2 - N) >> N_{\mathrm{S}}$, several stopping mechanisms were therefore added in the computational implementation of the algorithm to ensure as early an exit as possible (see Appendix C for the full implementation).

Pressure potential

The second contribution to the potential function is the pressure potential. This is a generalized potential which was conjectured to be the result of many different forces in the environment of the DNA chain. The DNA chain exists in the cell nucleus, which is surrounded by the nuclear envelope. This exerts a force inwards toward the center of the nucleus. The nucleoplasm is a viscous liquid, in which a myriad of different structures are suspended, including the chromosomes themselves. In addition to these viscous effects, steric effects on the DNA chain are expected from the immediate neighborhood on the chain itself, i.e. from other sections of the chromosome. All of these factors contribute to keeping the polymer more or less centered in a restricted location in the nucleus.

The pressure potential function was thus constructed:

$$U_{\mathrm{pressure}} = c_{\mathrm{pressure}} \sum_i \|\mathbf{X}_i - \mathbf{m}_{\mathrm{init}}\|^2, \tag{4.10}$$

where c_{pressure} is a model parameter, controlling the size of the potential, and $\mathbf{m}_{\mathrm{init}}$ is the center of mass of the polymer at the initialization of simulation. This created an assumed fixed location of the polymer, around which it could move.

It is worth noting that this potential term has significant effects on the root-mean-square (RMS) value of the distances from the monomers to the center of mass of the polymer, and that this in turn has profound effects on the ability of the monomers to perform recruited conversions of states. This will be explained in later sections.

4.1.3 Random noise

The last term of eq. 4.4 concerns the movement due to random noise. This noise is assumed to arise from random collisions with other molecules in the nucleoplasm, and is the sum of many such collisions occurring within the simulation time step Δt. The components of $\Delta\mathbf{W}$ are assumed to be Gaussian variables (see Appendix A), whose distributions are defined as follows:

$$p(\Delta W_{ij}) = \frac{1}{\sqrt{4\pi c_{\mathrm{noise}}^2 \Delta t}} \exp\left[\frac{-\Delta W_{ij}^2}{4c_{\mathrm{noise}}^2 \Delta t}\right] \tag{4.11}$$

In other words, the distribution is a Gaussian with mean zero and standard deviation $c_{\text{noise}}\sqrt{2\Delta t}$. The constant c_{noise} is a model parameter which controls the amount of noise.

4.1.4 Computing the gradients

The potential function terms given by eq. 4.9 and 4.10 are non-linear, and as mentioned, it was necessary to compute the gradient of this function for each monomer. This would take infeasibly long in practice, but could be achieved using automatic numerical differentiation via the `torch.autograd` functionality from the PyTorch library [28]. This tool is used heavily in machine learning, as one of the essential steps in this type of machine learning is *backpropagation*, in which model parameters are optimized based on gradient descent on loss functions, and this functionality is exploited using the potential functions. Every mathematical operation on the relevant model parameters (the monomer positions in our case) can thus be tracked, and using the functions involved in the iteration process, the numerically computed gradients are found and stored, after which they can easily be retrieved and used automatically for the purpose of our 3D simulation.

4.2 Nucleosome dynamics

We now have a three-dimensional polymer which evolves in time and space. We have seen how the potential experienced by each monomer depends on the state of the monomer: S state monomers are mutually attractive (up to two other monomers), while all other combinations of states yield repulsive monomers. Furthermore, these monomer states represent nucleosome modifications, which arise due to enzymatic reactions, as discussed in chapter 2. These states change due to interactions between the nucleosomes, so to represent these dynamics, it is thus now desirable to implement the 0D model discussed in chapter 3, with a few changes and modifications for the 3D scenario. The details are laid out in the following:

4.2.1 Implementation

The main difference between the implementation of nucleosome dynamics in the 0D model and in the 3D model is the way nucleosomes/monomers are chosen for recruited conversions. In the 0D model, given that a nucleosome has been chosen for recruitment, the probability of choosing any of the remaining nucleosomes is constant. In the 3D model, there is a constant preset maximum spatial distance that the monomers can be apart from each other for a recruitment to occur. This leads to a *distance-dependence* to arise naturally for the polymer; monomers closer to each other in the chain are expected to interact more often, while monomers further away will have fewer interactions.

For clarity, the algorithm for the nucleosome dynamics step will be fully described below. These steps are performed for every time step in the Euler-Maruyama iteration for the polymer dynamics.

Recruited conversion steps

1. Choose a *random* monomer n_1 on which to attempt a recruited conversion.
2. If the monomer n_1 is part of the *cenH* region, go directly to the noisy conversion steps.
3. Find the set of all other monomers within the recruitment distance r_{recruit}. If this set is empty, go directly to the noisy conversion steps.
4. Choose a *random* recruiting monomer n_2 from the set of monomers within distance r_{recruit}.
5. If the states of the monomers n_1 and n_2 are equal, or if the state of the recruiting monomer n_2 is U, go directly to the noisy conversion steps.
6. Perform the recruited conversion step according to the state of the recruiting monomer n_2:
   (a) State S: With probability α_1 (which is a model parameter), change the state of n_1 one step towards the state of n_2, i.e. from A to U or U to S.
   (b) State A: With probability α_2 (which is a model parameter), change the state of n_1 one step towards the state of n_2, i.e. from S to U or U to A.
7. Go directly to the noisy conversion steps.

Noisy conversion steps

1. Choose a *random* monomer n_3 on which to attempt a noisy conversion.
2. If the monomer n_3 is part of the *cenH* region, exit the noisy conversion steps.
3. With probability β (which is a model parameter), perform the noisy conversion step according to the state of monomer n_3:
   (a) State S: Change the state of n_3 to U.
   (b) State A: Change the state of n_3 to U.
   (c) State U: Randomly choose the state towards which a noisy conversion should be attempted.
      i. If the state S is chosen, with probability α_1, change the state of n_3 to S.

ii. If the state A is chosen, with probability α_2, change the state of n_3 to A.

At this point it is worth noting an important thing: In contrast to the 0D model by Dodd et al., there are two different conversion probabilities, one for each direction. The reason for introducing an extra model parameter is that the different physical and spatial interactions for monomers of different states also create unequal opportunities for the monomers to recruit each other. As we have seen, the monomers have to be within a certain distance to interact, and since physical attraction (as occurs for S state monomers) force more monomers of the same state to be closer together, more reactions and therefore a higher reaction probability is expected. As we will see, setting $\alpha_1 = \alpha_2$ does *not* create a symmetric behavior for the two competing states.

4.3 Comparison with existing models

Here we look at the most important differences between the model developed here and models already developed in the literature. This model is largely developed from scratch, and as such, there are many differences in e.g. the physical parameters used, and in the implementation of the Brownian dynamics. We therefore highlight the points that are considered to differ the most from established ideas.

4.3.1 Potentials

As noted in section 4.1.2, we here introduced rotation as the only allowed movement for monomers not located at the end of the polymer. This ensured a higher numerical stability, in addition to enforcing an infinitely stiff spring. This is markedly different from established 3D polymer models, where *harmonic* potentials are typical [24][27][29].

Another common type of potential is the *bending potential* [24][26][29][30]. This potential is not directly implemented in this model; however, as the rotating-monomer approach does restrict movement in some directions, it is expected that this results in some stiffness, the degree to which remains unclear.

Lastly, the pressure potential, which *is* implemented in this model has, to our best knowledge, not been used in other models to this point in time. As we will see, one of the main consequences of this potential is that monomers spend more time in closer proximity to each other. This is achieved in other models by other means, for example by letting a large number of monomers attract one another - this can lead to what is referred to in some studies as "compact-ordered states" [24][29][31][32]. This leads us to the next point, which concerns the physical monomer-monomer interactions.

4.3.2 Physcial interactions

In our model, we let U state monomers correspond to unmodified nucleosomes, whereas A state monomers represent acetylated, euchromatic nucleosomes, and S state monomers represent methylated, compacting nucleosomes. In other words, there is only one type of monomer which can attract other monomers. This differs from many estblished models, which usually implement attraction for *both* types of modified or 'marked' monomers (see e.g. [24][27][29][31][32]). In these models, one of the states usually corresponds to an 'unmarked' nucleosome, which induces steric repulsion towards all other monomers, whereas both of the other states induce attraction toward other monomers of the same state. Depending on the recruitment rates for the different states, this usually leads to different *phases*, where the polymer becomes 'compact-ordered' above some transition point due to clustering of the different monomers types, and 'swollen-disordered' below the transition point. As we will see in section 5.2, the resulting 3D structures from our model also vary in compactness, but only due to the presence of the S state. We view the A state as antagonistic towards compaction, and this is a significant difference between our model and the established ones.

Limit on interaction pairs

The upper limit on the number of interaction pairs one S state monomer can make with other S state monomers is another feature which, to our best knowledge, has not been seen in other models. The motivation behind limiting this number to 2 is a biological one, as it is assumed that the number of nucleosomes that can bind together is limited by the number of possible methylation locations on the histones (see section 2.2.1). This naturally has further implications on the emerging 3D structures, and it is a reminder that using physical potential functions to represent interactions between proteins has some unrealistic consequences, as the binding interactions in question are actually driven by enzymatic reactions, and not by force fields.

4.3.3 Local vs. global conversion rates

The last important difference between our model and some established models lies in the modelling of the recruitment process between monomers, and the dependence on spatial distances. In our model, only monomers within a preset spatial distance can recruit each other, and there is no discrimination between monomers closer or further away from each other on the polymer. We highlight that the resulting difference in recruited conversion reactions emerge as a result of spatial distance only, and that differences in distance on the polymer itself is a consequence of this spatial distance, and thus a lower probability of interactions between the monomers (see section 5.4.1). The selection of spatially proximate monomers for these recruitment reactions is also seen in other models, e.g. [24][31]. However, some other studies discriminate between monomer state conversion rates for nearest-neighbor reactions (sometimes referred to as

local or *cis* reactions) and non-nearest neighbors reactions (*global* or *trans* reactions) [26]. Studies by Obersriebnig et al. [19], as well as Nickels et al. [20] (the latter of which will be studied in more detail in sections 5.4.2 and 5.4.3), both highlight the *need* for local and global rates as distinctively different: one rate applies only to nearest-neighbor recruitments and the other applies to all other recruitments. It should be noted that no 3D substrate is implemented for the nucleosomes in the models from these studies; rather, a 1D spatial distance dependence is introduced for the global reaction rates. Furthermore, the authors of both studies only allow global reactions for some of the reaction types, i.e. Obersriebnig et al. only allow global reactions for the A $\rightarrow$ U reaction. Again, in our model, the only difference in behavior between reaction types is seen in the reaction *rates*. As such, this is a reduction in the number of necessary model parameters.

4.4 Final words on the model

We now have a working 3D model to represent both the Brownian dynamics as well as the nucleosome dynamics of a DNA chain. In the next chapter, we will see how this model was used for simulations to generate data and results to meet our main objectives described in chapter 1.

5 Results and discussion

5.1 Model parameters

Before looking at the results from the simulations, it is prudent to keep in mind the different model parameters. Some values were fixed for all simulations, whereas others were varied to yield the different results seen in this chapter. The fixed and the varied parameters can be found in tables 5.1 and 5.2 respectively, and include names, descriptions, and values used for each parameter.

Generally, the values of these parameters were found after performing a large number of simulations, and investigating qualitative and quantitative behavior of the polymer. Here the values do not have units, but knowledge on the actual scales involved makes it possible to assign correct units, since the relationship between the parameters is established in the model. For example, the parameter l_0 represents the distance between nucleosome core particles, the value of which can possibly be experimentally determined. The time steps t are unitless integers, which also could potentially be rescaled using knowledge of the time scales involved. It is also crucial to note that the relation between the time scales of the Brownian dynamics and the nucleosome dynamics is not established for this model.

All in all, the parameter values arrived at here give us a model polymer which has the desired behavior, and additionally allows us to perform parameter optimization to a certain degree. All of this will be explored in the following.

5.1.1 On the specific parameters

Fixed parameters

- N: The system size is a compromise between modelling a realistic system and being able to run a sufficient number of simulations. As we saw in chapter 2, the mating-type region of *S. pombe* includes approximately 150 nucleosomes. The simulation time however becomes infeasibly long with such a large system, so 40 monomers were found suitable for this purpose.

- l_0: As noted above, this parameter can in principle be linked to the real length between nucleosomes in the model system. Here it was set to unity without loss of generality.

- c_{noise}: The value of the parameter was arrived at through inspection of the results of various simulations, and was found to yield realistcally looking dynamics.

- Δt: This value was sought maximized while maintaining realistic dynamics. Notably, the Gaussian noise scales with the square root of this value.

Parameter	Value	Description
N	40	The number of monomers.
l_0	1.0	The fixed distance between neighboring monomers.
c_{noise}	0.5	The scaling parameter for the random Gaussian noise for the Brownian dynamics.
Δt	0.02	The finite time step value.
interaction size	50	The scaling parameter for the interaction potential.
interaction potential cutoff	$1.0 \cdot l_0$	The maximum distance for the interaction potential.
r_0	$l_0/2$	The equilibrium distance for the interaction potential.
r_{recruit}	$4.0 \cdot r_0$	The distance within which two monomers must be for a recruited conversion to happen.
cenH initial index	16	The initial monomer index of the *cenH* region (using one-based indexing).
α_2	0.1	Proportionality constant for the reaction probability of recruited conversions from S $\rightarrow$ U and U $\rightarrow$ A.
β	$0.04\alpha_2$	Proportionality constant for noisy conversion.

Table 5.1: The fixed model parameters used, along with their values and descriptions.

- interaction size: The value of the parameter was arrived at through inspection of the results of various simulations, and was found to yield realistically looking dynamics.
- interaction potential cutoff: The choice to implement a potential cutoff was based on the assumption that nucleosomes condense together by binding to proteins, which constitutes a short-range interaction. The value itself was conjectured using the same assumption, and set in relation to l_0.
- r_0: The motivation behind this parameter is based on the assumption that nucleosomes pack together in a zigzag pattern to form a heterochromatic fibre [8], and the actual value was thus set to best facilitate the emergence of such a structure.
- r_{recruit}: The value of this parameter again uses the assumption that the

Parameter	Value	Description
total time	Any	The number of time steps for a simulation. This value is not restricted, and can be set to anything depending on the purpose of the simulation.
RMS	[1.677, 4.130]	The interval of the allowed root-mean-square (RMS) values of the distances from the monomers to the center of mass of the initial polymer configurations. This value directly determines the value of the pressure potential; see section 5.1.2.
initial state	{S, U, A}	Refers to the initial state of all monomers except those within the *cenH* region.
cenH size	{6, 7, 8}	The possible number of monomers in the *cenH* region.
α_1	(0, 1.0]	Proportionality constant for the reaction probability of recruited conversions from A $\rightarrow$ U, and U $\rightarrow$ S.

Table 5.2: The model parameters that were varied for different simulations, along with their values and descriptions.

states of nucleosomes change due to interactions with enzymes, which are short-range interactions. Furthermore, there is an important tradeoff: the higher the value, the more monomers will be candidates for a recruited conversion, but the behavior arising from the different spatial distances become less and less important. The actual value was conjectured to be a factor 4 times bigger than the equilibrium distance r_0.

- *cenH* initial index: The value was set based on data from Nickels et al. [20].
- α_2: The value was set as low as possible, but still such that results could be achieved within reasonable simulation time. The actual relation to real time is not established here.
- β: This parameter represents the rate of nucleosome state conversions that arise due to the presence of histone-modifying enzymes in the nucleoplasm. These reactions thus do *not* arise from recruitments by other nucleosomes. The value was set based on the assumption that noisy conversions happen much more rarely than recruited conversions, and was set several orders of magnitude lower than the value of α_2.

Varied parameters

- total time: This value was varied based on the needs for the particular simulation. Note that the time steps t are without units, and thus simply set to increasing integers. The value can be rescaled using knowledge on real time scales.

- RMS: The parameter should be viewed as an indirect parameter; setting the value directly determines the value of the pressure parameter c_{pressure} (see eq. 4.10) via a function constructed for the purpose. (This will be explained in section 5.1.2). As such, the given RMS values used to produce the results in this chapter should be considered approximate values for the initial condition of the polymer. The parameter was used as input to achieve a desired average RMS for the initial polymer configuration, and the value was varied to demonstrate the effects of the pressure potential on many of the results.

- initial state: This parameter was varied according to the purpose of the simulation. For the simulations involving the *cenH* region, the initial state for the whole system (except for the *cenH* monomers) was always set to A to demonstrate the establishment of the S state.

- *cenH* size: The value was set to 8 for all simulations representing the wild-type mating-type region of *S. pombe*, as this achieved the real ratio of *cenH* size to system size. As mentioned in section 2.2.2, the nucleation of the *cenH* region actually is a reversible reaction occurring at a high rate, so for modelling purposes, it was assumed that this rate was significantly higher than the rate of monomer conversions. As such, it was decided to set *cenH size* independent of time, and thus implement it as a fixed-length region throughout simulation. These monomers could thus neither undergo recruited, nor noisy conversions. Crucially however, they could participate in recruited conversions as the recruiting monomer. For different simulations, the value was varied to demonstrate different mean S state establishment times for different ratios of the aforementioned *cenH* size to system size.

- α_1: This value was only varied to demonstrate the hysteresis effect, and during optimization (see section 5.4.3). For all other simulations, the value was held fixed at $\alpha_1 = 0.07$. This value was used, as it produced a regime where both the overall S and A states were stable throughout simulation time (see section 5.3.1), while at the same time producing a monostable system for the S state when the *cenH* region was present (see section 5.3.3).

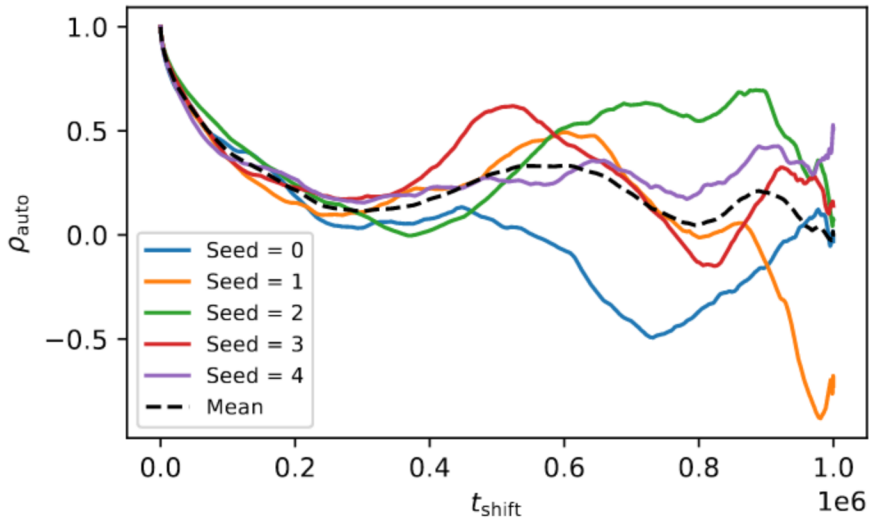


Figure 5.1: Auto-correlation of the distance vectors from each monomer to the center of mass of the polymer, averaged over the number of monomers and time. All monomers were initially in the A state, and were not allowed to change during simulation to maintain the behavior of a free polymer. Parameters: total time = 10^6, RMS ≈ 4.13.

5.1.2 Initial conditions

Quasi-random polymers

It was desirable to start each simulation with the polymer taking a *quasi-random* initial position. By quasi-random, it is implied that the configuration of the polymer in space should appear random, since by the nature of computational simulation, the initial position will always be deterministic. In this case, all simulations were additionally run using preset random seeds, such that experimental results could be reproduced, hence the initial positions are not truly random.

To arrive at a quasi-random initial position, the dynamics of a free polymer (i.e. a polymer without pressure implemented) were investigated. In this way it was possible to get an idea of the time scale of the Brownian dynamics. The measure used for the movement of the polymer was the auto-correlation of the distance vectors from the monomers to the center of mass of the polymer, averaged over the number of monomers and time, which has the following expression:

$$\rho_{\text{auto}}(t_{\text{shift}}) = \frac{\langle \frac{1}{N} \sum_{i=1}^{N} \mathbf{d}_{i,t} \cdot \mathbf{d}_{i,t+t_{\text{shift}}} \rangle_t}{\langle \frac{1}{N} \sum_{i=1}^{N} ||\mathbf{d}_{i,t}||^2 \rangle_t}, \tag{5.1}$$

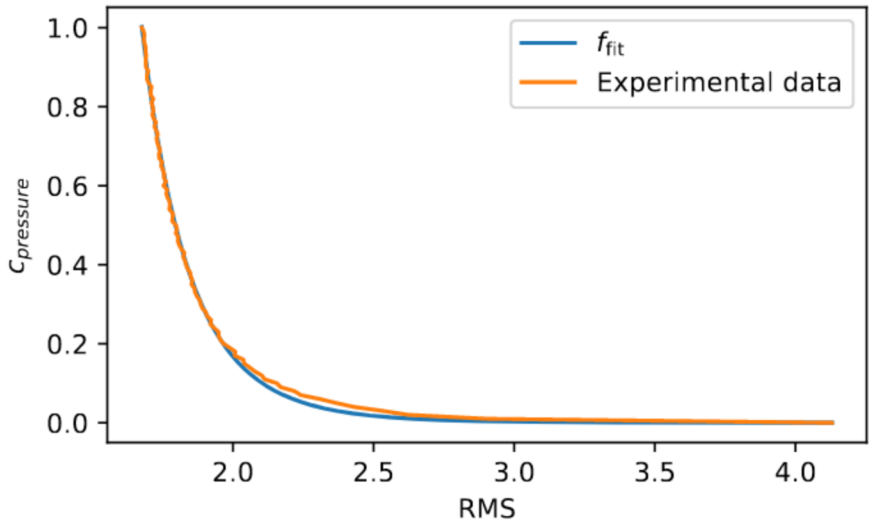


Figure 5.2: The relationship between the pressure constant and the RMS of the polymer.

where $\mathbf{d}_{i,t}$ is the distance vector from the i'th monomer to the center of mass at time t. The results for five different free polymers developing through 10^6 time steps can be seen in figure 5.1. Here the auto-correlation shows exponential decay from the beginning of the simulation, as expected. Around approximately $3 \cdot 10^5$ time steps, the polymers start to exhibit widely different behavior, which is expected to arise from two sources:

- Less statistics, which occurs due to the fact that the higher the value of t_{shift}, the less data to average over.
- Configurations can be correlated over longer time intervals if the polymer stays repeatedly in similar configurations for certain periods of time.

The mean correlation, however, does not deviate far from 0. Using these results, it will be assumed that the configuration of most polymers are quasi-random at least after $t = 10^6$. *To generate initial configurations for the simulations, 100 polymers were thus allowed to move through a total of* 10^6 *time steps.* The positions were then used as a basis for generating initial polymer positions with different RMS values.

Adding RMS for initial positions

Using these initial configurations of free polymers, different RMS values were then input to get quasi-random polymers for each different value. *After inputting*

these RMS values, the polymers were allowed to move for another 10^6 *time steps.* The result was 100 different RMS values for which 100 different quasi-random configurations could be initialized for simulation.

RMS and pressure

In eq. 4.10, the constant c_{pressure} appears as a scaling constant for the pressure potential. As we will see, the higher this constant is, the more compact the polymer is, and the lower the RMS, whereas a low value yields a polymer with a relatively high RMS. The relationship between c_{pressure} and the RMS (resulting from experimental data) can be seen in figure 5.2.

The experimental data was collected from the production of initial quasi-random polymer positions, as explained above. A function was then constructed from fitting to the data:

$$f_{\text{fit}}(x) = 186.8 \cdot \left[\frac{1}{1+\exp(x-38.85)}\right] \cdot x^{-10.12}, \tag{5.2}$$

which essentially is a scaled power-law function with an exponential cutoff. In this way, an analytical value for the constant c_{pressure} could be retrieved by inputting the desired RMS value. As noted in table 5.2, there is both a lower and an upper bound for the possible RMS values. It is important to note that these values are based on the output from the *data*; values outside of this interval are possible in theory. For comparison, the theoretically possible interval of RMS values, i.e. for a fully stretched polymer, would be $\left[0, \sqrt{\frac{1}{N}\sum_i^N (\frac{N+1}{2} - i)^2}\right]$, which gives an interval of [0, 11.54] for $N = 40$.

5.2 3D structure

In figures 5.3-5.4 some examples of different polymers can be seen for RMS ≈ 2.0 and RMS ≈ 4.0, respectively. The volume the different polymers take up due to the difference in pressure potential is noticeable. The polymers are also visibly more compacted in regions where the monomers are in the S state, as these monomers are mutually attractive, whereas the sections of A state monomers have a more open structure. In figure 5.4 (bottom), the structure emerging from the limitation on two interactions per S state monomer becomes particularly noticeable: the polymer has folded on a macroscopic level, and the monomers are intertwined over long stretches of the chain. As such, this is not an example of a zigzag pattern of close-neighbor interactions, as mentioned in section 2.2, but it does illustrate the desired close affinity between the S state monomers. It also appears that this polymer is locked in a configuration where S state monomers far apart on the chain are interacting attractively. The result is a higher-level packing, which indicates that the simple monomer interactions could account for more macroscopic structures.

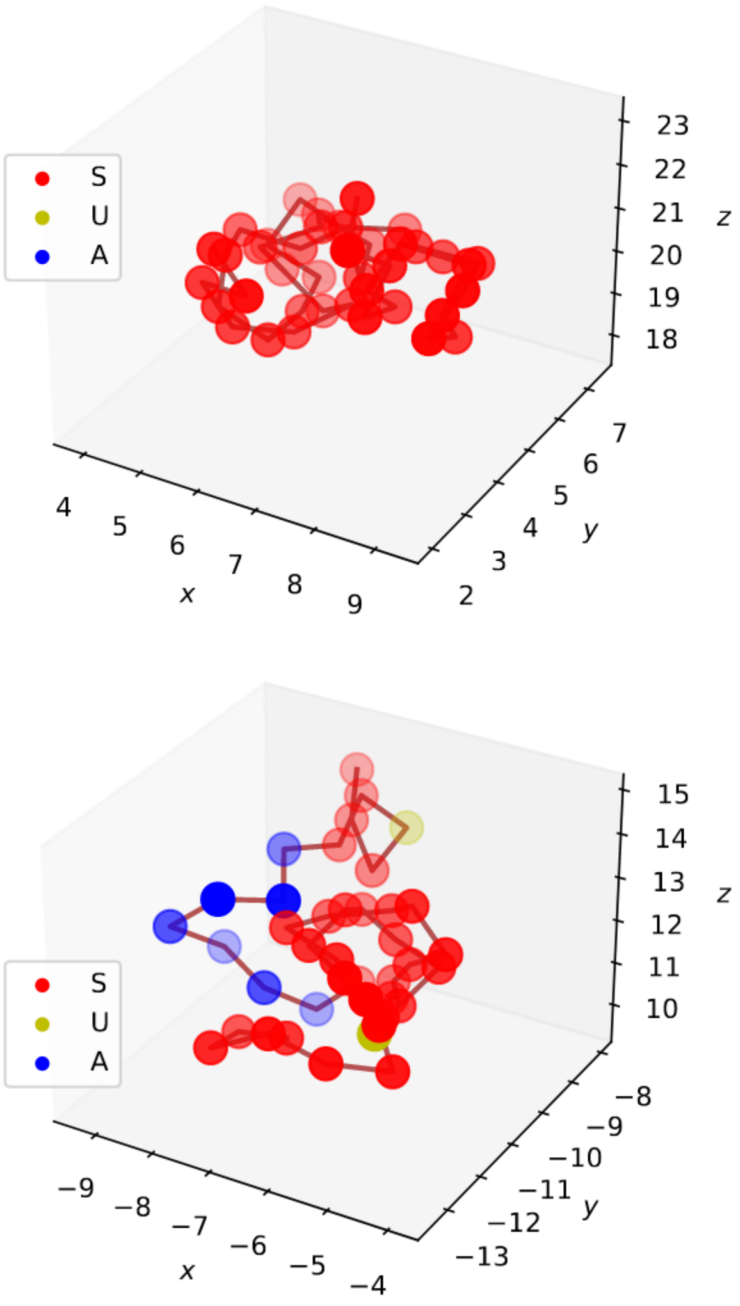


Figure 5.3: Two examples of polymers after $5 \cdot 10^4$ time steps. Parameters: RMS ≈ 2.0, initial state = A, *cenH* size = 8, $\alpha_1 = 0.07$.

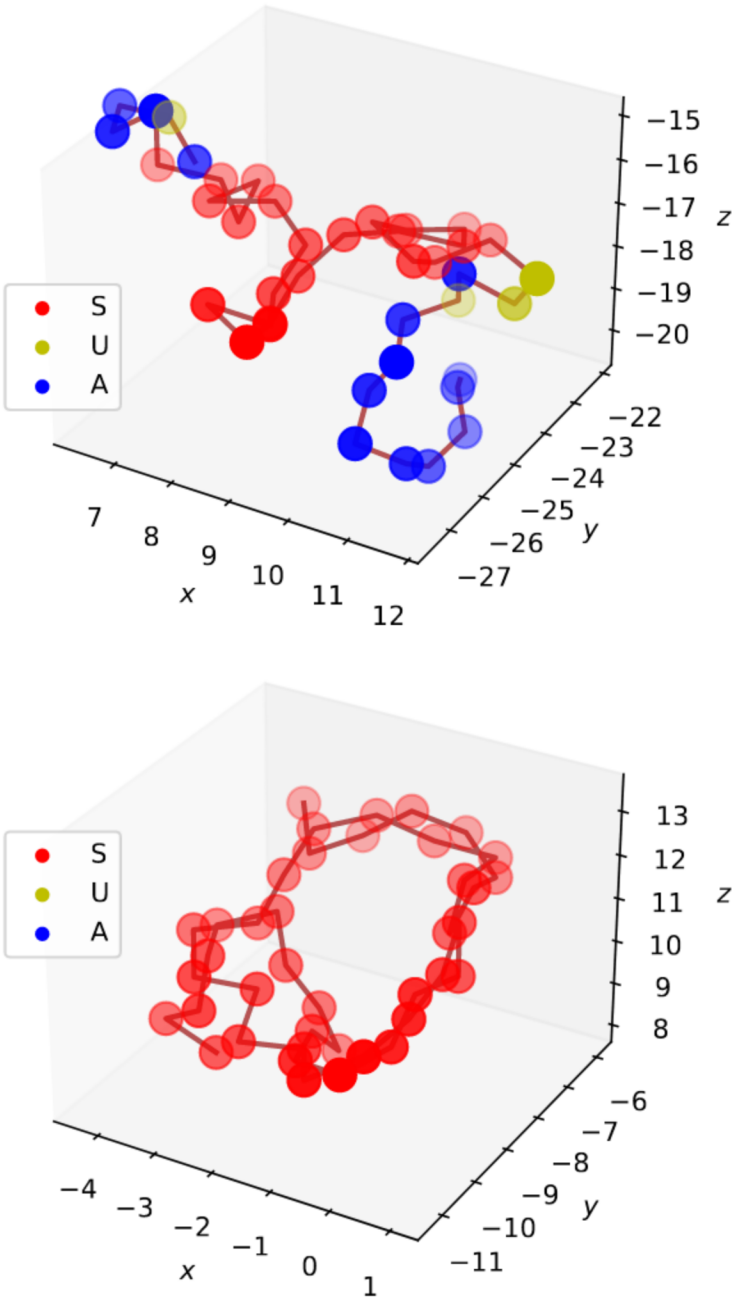


Figure 5.4: Two examples of polymers after $5 \cdot 10^4$ time steps. Parameters: RMS ≈ 4.0, initial state = A, *cenH* size = 8, $\alpha_1 = 0.07$.

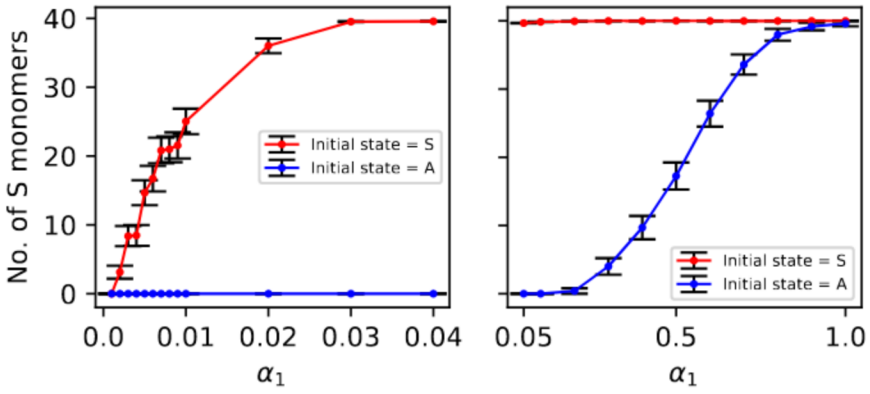


Figure 5.5: The mean number of S state monomers at the end of 100 simulations, each with a total time = 10^5, and keeping $\alpha_2 = 0.1$ constant. Note the different scales on the abscissae. Parameters: RMS ≈ 2.0, *cenH* size = 0.

5.3 Epigenetic switching

5.3.1 Bistability and hysteresis

The 3D polymer developed here exhibits both *bistable* behavior, following the definition in chapter 2, and *hysteresis*. As can be seen in figure 5.5, depending on the overall initial state of the polymer, there are two different transition points for when the polymer switches between the overall S and A states. Keeping $\alpha_2 = 0.1$ constant, we see that a switch from A to S for all monomers after 10^5 simulations requires a value of α_1 roughly a factor 10 higher. Starting in the overall S state, the polymer only switches to an overall A state for values of α_1 about 100 times lower. Note also that the transition is more rapid for the overall initial state S. This is expected to be a consequence of the different physical interaction rules for the S state and the A state, and indicates that the presence of physical interactions between the monomers *increases* a higher overall stability for those monomers.

It is also worth noting that in the approximate range for α_1 of [0.03, 0.2], both initial states are stable, and do not switch to the other state during simulation time. A concrete example can be seen in figure 5.6, where the number of monomers in each different state given two different initial conditions are shown as a function of time, for $\alpha_1 = 0.07$. Initializing the polymer with all monomers in either the S state or the A state, the polymer never switches to the other state during a simulation time of 10^6 time steps. The level of noise is low, and the two states are never truly competing. The figure shows representative behavior for the polymer in general - running many simulations with different preset seeds yields the same results.

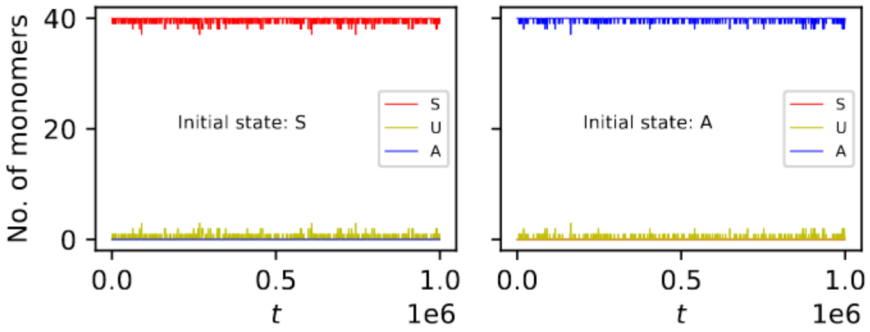


Figure 5.6: The number of monomers in each state given two different initial conditions. Both states are clearly stable over long periods of time. Parameters: total time = 10^6, RMS ≈ 2.0, *cenH* size = 0, $\alpha_1 = 0.07$.

The simulation time is fairly long, and a switch between polymer states is not observed. It is important to remember that there is a finite probability that such a switch can occur, but that the definition of bistability given in section 2.2.3 does take this into account. For an epigenetic switch, stability over very long periods of time is crucial, but it is of course impossible to conclude anything about the behavior over infinite time.

The system as seen here corresponds to the ΔK mutants, i.e. small strains that lack the *cenH* region (see section 2.2.2), as no *cenH* region is present. These mutants show bistable behavior, but additionally, *switches* between both states are observed every $\sim 2\,000$ cell generations. Here these switches are not observed during simulation time. This is an indication that the number of noisy conversions were not sufficient to overcome the stability of the two states, and thus that the proportionality constant for noisy conversions, β, needs to be set to a higher value to enable spontaneous switches between the states without a *cenH* region implemented. On the other hand, achieving stability throughout simulation time for both the overall S and A state was prioritized in this model, as it is a prerequisite for a working epigenetic switch.

5.3.2 Epigenetic memory

Next, it was investigated whether the model could be used to represent epigenetic memory. This was done by introducing *cell division events*, which are disruptive events to the system, at regular time intervals, and then testing the stability of the two overall states. Two assumptions were made:

- In real cells, the spatial location of the DNA chain is altered as the chromosome undergoes compaction and subsequent division. This was simulated as a re-initialization of the polymer position at the cell division event, i.e.

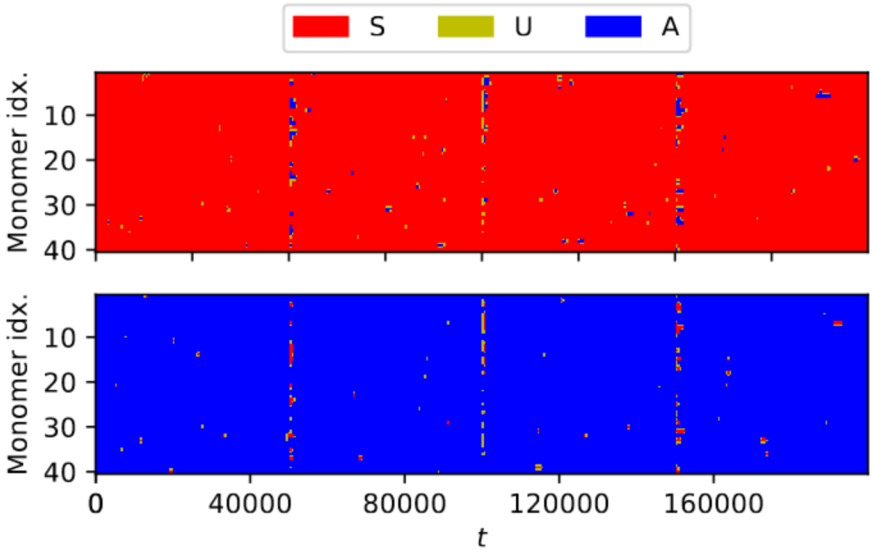


Figure 5.7: Time-space plots of monomer states with regular cell-division events, for polymers of initial state S (top) and initial state A (bottom). Both polymers remain in their overall initial states. Parameters: total time $= 2 \cdot 10^5$, RMS ≈ 2.0, *cenH* size $= 0$, $\alpha_1 = 0.07$.

a random initial position was chosen (as described in section 5.1.2), at which the polymer was placed.

- After DNA replication and the subsequent cell division, the epigenetic modifications are distributed randomly on the two daughter chromosomes. In this case, the polymer on average retains 50% of the original modifications. This was implemented by randomly changing the monomer states (excluding the *cenH* monomers, following the assumption that *cenH* nucleosomes immediately change back to the S state due to the activity of RNAi) to the U state at the cell division event.

The results from simulations with cell division events can be seen in figure 5.7, where a cell division was simulated every $5 \cdot 10^4$ time steps. The highly noisy events clearly have little effect on the overall stability of either of the initial states: after each event, the individual monomer states quickly change back to the initial state. This mimics epigenetic memory, as it shows that the overall state can be inherited by daughter chromosomes, and corroborates our claim of the stability of both of the overall initial states.

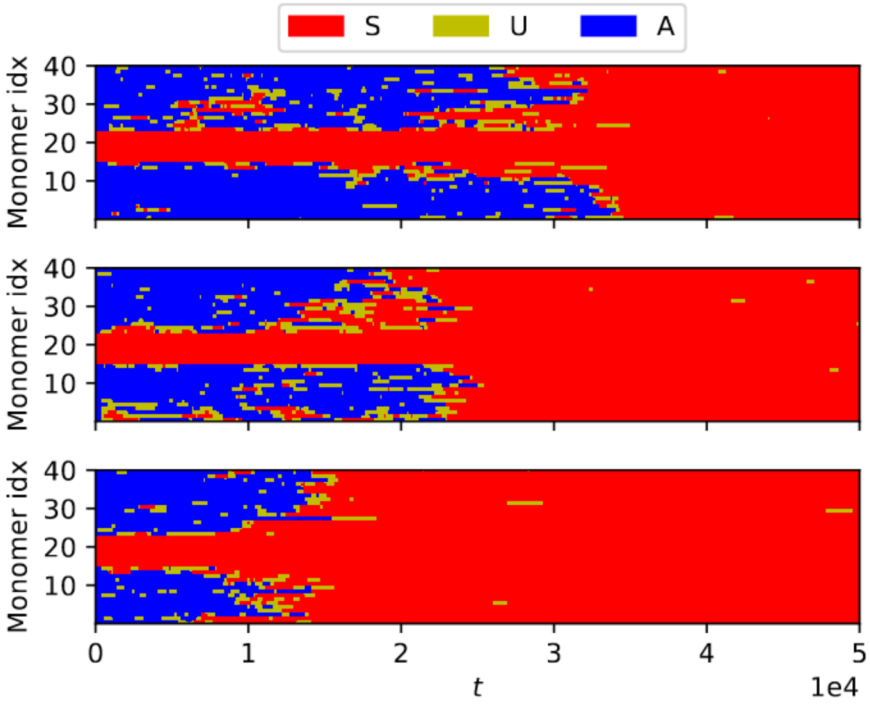


Figure 5.8: Time-space plots of the individual monomer states for three different simulations. The plots give a representative view of the time scale for the epigenetic switching time, the mechanism for which can clearly be seen in all three plots. Parameters: total time $= 5 \cdot 10^4$, RMS ≈ 2.0, initial state = A, *cenH* size = 8, $\alpha_1 = 0.07$.

5.3.3 Activating the switch

Now we look at results from simulations where the *cenH* region was present. The results can be seen in figure 5.8. Here, a number of different things can be observed:

- In each case, a *switch* from the A state to the S state can clearly be seen. The presence of the *cenH* region adds a bias to the system that is enough both to induce the actual switch, and to make the system stay in the S state.
- The switch occurs *stochastically* in time. This is illustrated by the three different time-space plots in the figure.
- When the switch occurs, it occurs in a sudden *burst*. From the start of each simulation, noise is observed as small local S-state patches before the

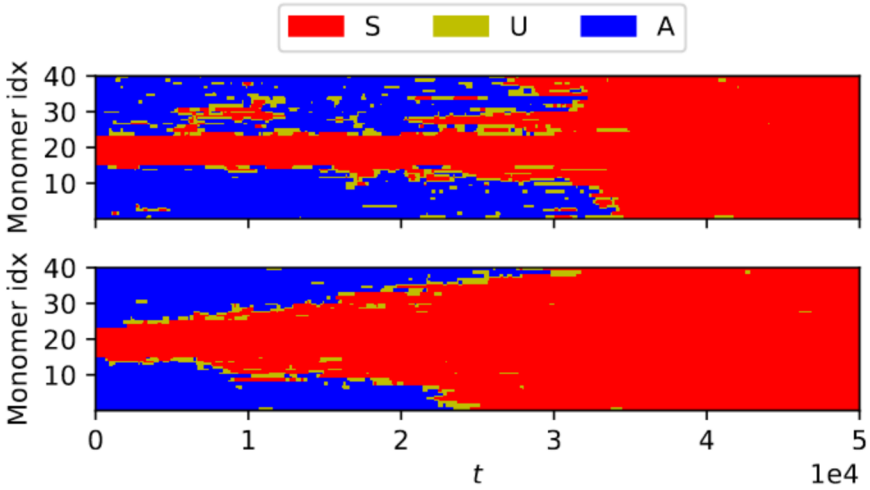


Figure 5.9: Time-space plots of polymer showing heterochromatin establishment and state switching. **Top:** Polymer with pressure implemented (RMS $\approx$ 2.0). The switch occurs within a short time interval, and S-state patches are observed. **Bottom:** Polymer without pressure implemented (RMS $\approx$ 4.13). The switch occurs more gradually, and there are almost no S-state patches. Parameters: total time $= 5 \cdot 10^4$, initial state = A, *cenH* size = 8, $\alpha_1 = 0.07$.

whole system suddenly switches within a fairly short time interval.

- All of the observations made above are in line with the properties of a biological epigenetic switch, as described in section 2.2.3.

Even though figure 5.8 only shows the behavior for a small time interval, it is important to highlight that after the switch occurs, the polymer stays in an overall S state for the entirety of simulation time. More concretely, *with the cenH region present, the polymer always switches from the A state to the S state, but a switch back from the S state to the A state is never observed.* This exactly mimics the behavior of an epigenetic switch.

5.4 The effects of external pressure

In this section we will see the profound effects the pressure potential has on the simulation results. We will see that the pressure potential is essential for enabling sufficient long-range monomer interactions, and thus recruited conversions. This in turn affects the *globalness* of the interaction activity, and the

switching time for the polymer. We will look at three main topics:

1. First we will look at some general comparisons between results from simulations where pressure was implemented, and when it was not. Here we will find support for the argument that pressure is an essential ingredient in reproducing realistic results for the epigenetic switch.

2. Second, we will investigate the properties of the S state establishment time of the polymer, and compare the results with a study by Nickels et al. [20]. Here we will find further support for the aforementioned argument.

3. Third, having shown that the establishment times from real experiments can be reproduced using this 3D model, we will optimize the model with regards to the recruitment probability constant parameter α_1, using the results from the study as a benchmark. This will further corroborate the argument that the pressure potential is an essential component in the model.

5.4.1 General results

A comparison of time-space plots of simulations with and without pressure effects implemented can be seen in figure 5.9. The time-space plots look markedly different, and the first important observation is that the S-state establishment happens over a significantly longer time interval for the polymer where pressure is not implemented (bottom figure). Here, establishment starts more or less at the start of the simulation, and is completed after around 30 000 time steps. The polymer where pressure effects are implemented (top figure) undergoes a much more sudden switch; it remains more or less in the same state from the start of the simulation until about 30 000 time steps, after which the switching starts, and then finishes around 4 000 time steps later. There is also visibly more noise in this plot, and it appears that many more monomers undergo state changes during the simulations, and that conversion interactions happen between monomers far apart on the chain. For the polymer without pressure effects, there is very little noise, and the S state appears to spread mainly between neighboring monomers. In other words, *pressure seems to induce a higher number of state conversions, both occurring between non-neighboring monomers, and in absolute terms.* Figure 5.9 shows a qualitative comparison, but these effects can also be quantified.

Global interactions, switching times, and S state patches

The number of successful recruited conversions from state U to A as a function of the monomer index difference can be seen in figure 5.10. Here, a polymer with no *cenH* was developing for 10^7 time steps, and in this case recruited conversions occurred almost exclusively between these two states. This was expected, as the polymer started in the overall A state, and with no *cenH* present, the initial state is stable through simulation time (as we have seen in

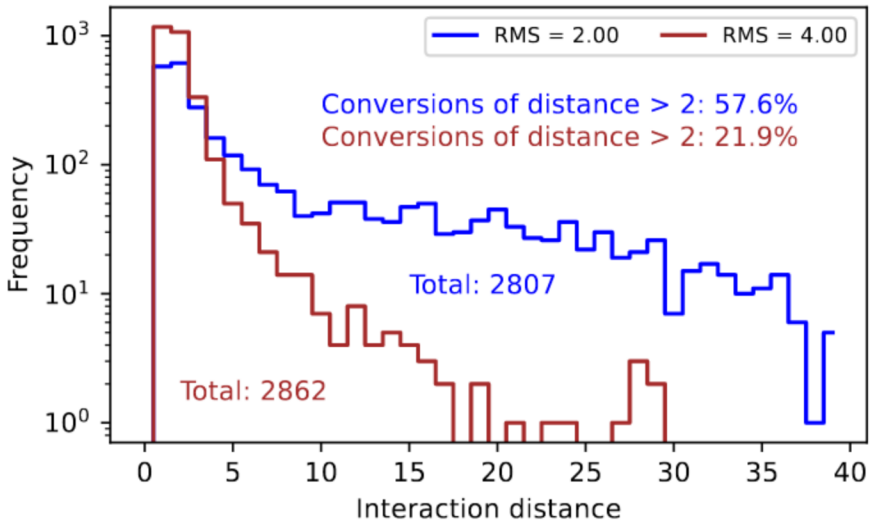


Figure 5.10: The number of successful recruited conversions from state U to A as a function of the index difference of monomers for different RMS values. Parameters: total time = 10^7, initial state = A, *cenH* size = 0, $\alpha_1 = 0.07$.

figure 5.6). The total number of conversions is comparable for both RMS values, but there is a significant difference: for RMS ≈ 2.0, the number of successful recruited conversions occurring between monomers further away from each other than the next-nearest neighbor is almost a factor 3 higher than for RMS ≈ 4.0 (57.6% vs. 21.9%). Furthermore, the number of conversions appears to decay very slowly all the way up to the maximum distance of 39. For RMS ≈ 4.0, the values decay very fast, and very few if any recruited conversions happen across a distance corresponding to half the chain length.

Furthermore, this figure illustrates a *distance-dependence* in the system, as the number of successful conversions clearly decays as a function of the distance between monomers on the chain. It is important to note that *this behavior arises naturally from the rules and parameters for this system*; it does neither require an assumed function for distance-dependence, nor different rate parameters for interactions occurring locally and globally, as was described in section 4.3.3.

Finally, the ratio of the total number of successful recruited conversions to the total number of time steps is $\sim 3 \cdot 10^{-4}$. This is an indication of the actual rate of monomer state conversions occurring. The rates are expected to differ for different initial polymer states, as S state monomers are attractive towards other S state monomers, and since two S state monomers, once interacting within the equilibrium distance r_0, are expected to stay in that position for extended periods of time, preventing further recruited conversions from taking place.

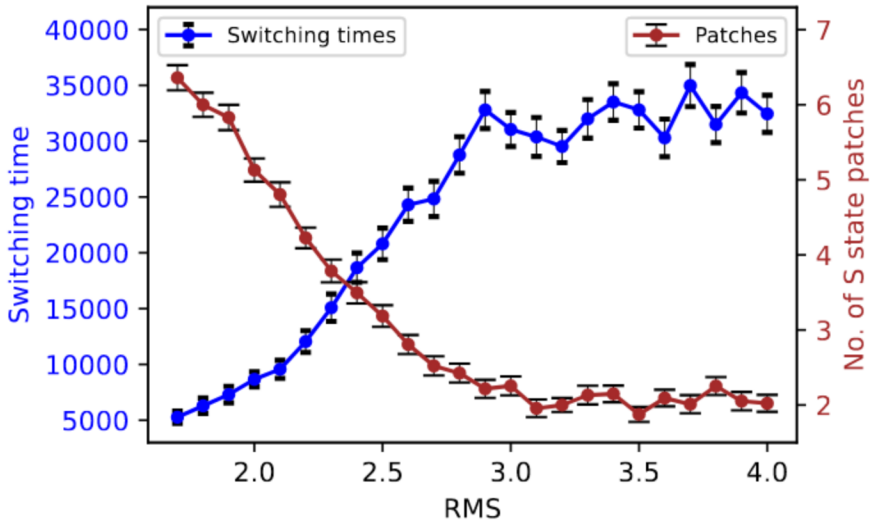


Figure 5.11: The mean switching time and number of S state patches at the time where 50% of the monomers were in the S state, resulting from 100 simulations for each RMS value. Parameters: initial state = A, *cenH* size = 8, $\alpha_1 = 0.07$.

Referring again to figure 5.9, we see that when the pressure potential is implemented, there is significantly more noise in the form of isolated S state areas, or patches. Since the number of random conversions should be the same for both RMS values, these patches are certain to arise from longer-distance recruited conversion interactions. This was quantified as the number of isolated, contiguous S-state patches, (or regions), of any size in the polymer, at the time where 50% of the polymer was in the S state. Furthermore, we see that the switching time for a polymer under pressure potential is significantly shorter. The switching was defined as the time interval between the time at which 50% of the polymer was in the S state, and the time at which 90% of the polymer was in the S state. In figure 5.11, these results can be seen for polymers with different RMS values. The results were taken as an average over 100 simulations. The number of S state patches is markedly higher for lower RMS values, averaging more than 6 patches, whereas the number falls as the RMS grows; when the RMS is above 3.0, the average number of patches falls to about 2. Taking into account that any random state conversion to an S state could result in a patch of size 1, which still was counted, the average value of about 2 patches indicates that the S state spreads almost entirely between nearest neighbors. There is very few, if any, recruited conversions that occur on longer distances. This corroborates the results from figure 5.10.

The switching time clearly trends upwards for higher RMS values, and thus

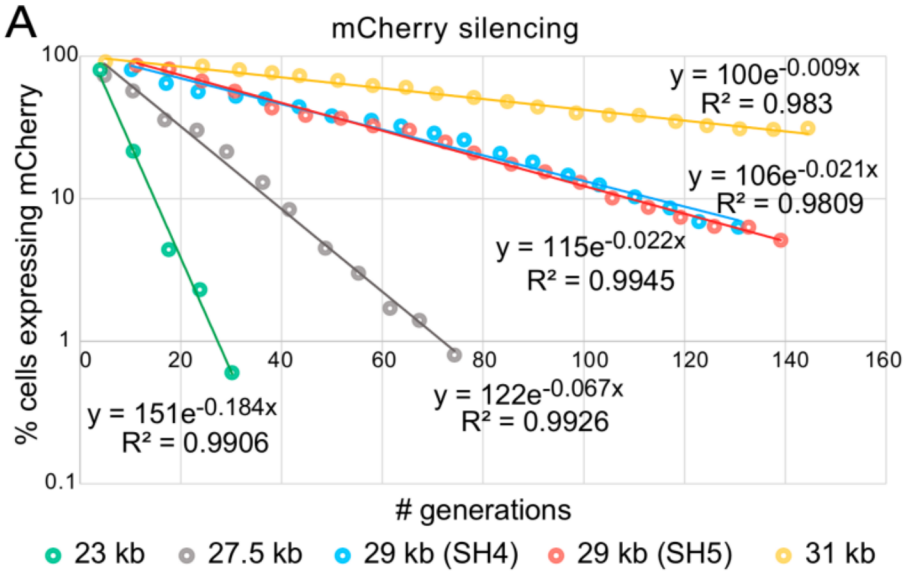


Figure 5.12: Heterochromatin establishment times in *S. pombe* mating-type regions of different lengths. (From *Establishment of heterochromatin in domain-size-dependent bursts* (2021), Nickels et al. [20].)

lower pressure. When a significant amount of pressure is implemented, the switching time becomes fast, and mimics the behavior seen in figure 5.9 (top), whereas with a higher RMS value, and thus lower pressure, the switching time becomes much longer; the difference between the most extreme values are about a factor 6. This in turn mimics what it seen in figure 5.9 (bottom). As discussed in section 2.2.2, during heterochromatin establishment, the overall state switching of the DNA chain occurs stochastically and in a burst-like manner. The results seen here thus constitute further evidence that pressure is needed to induce this fast state-switching nature.

5.4.2 Investigating the S state establishment time

In the following, we investigate the *establishment time* of the S state, i.e. the time it takes from the start of simulation until state switching has taken place (defined here as the time at which 90% of the polymer was in the S state). This was inspired by the study done by Nickels et al. [20], in which the main findings for this purpose were:

- Heterochromatin establishment times follow an *exponentially decaying* probability distribution.
- The mean establishment time depends on system size, i.e. a bigger system yields longer mean establishment times.

The authors investigated heterochromatin establishment by creating strains of *S. pombe*, with two fluorescent reporters embedded strategically within the mating-type region: the *yellow fluorescent protein* (YFP) was placed in the *cenH* region, while the *mCherry* (which is a red fluorescent protein) was placed at a "peripheral location". In this way nucleation at the *cenH* region could be captured by the absence of yellow light, while an absence of red light could be used as an indicator that heterochromatin had spread through the whole region, based on the peripheral location of the mCherry reporter. The authors highlight that this reporter has no intermediate states between 'ON' and 'OFF', so that an 'OFF' state mCherry reporter could be counted as an 'OFF' state, or 'silenced' cell [20].

The authors also investigated the effect of system size on the heterochromatin establishment time by creating strains with different lengths: 23 kb (wild-type), 27.5 kb, 29 kb, and 31 kb. Crucially, these strains were obtained by inserting DNA-sections between the two fluorescent reporters, so that the system sizes could be reflected in the difference in time between the silencing of the reporters.

The authors found that the silencing of the YFP happened "fast in all cases", due to its location within the nucleation region itself, but that the silencing times of mCherry "was slower and differed greatly between strains". In other words, there was huge variation in time duration for heterochromatin establishment within the strains. In figure 5.12, these results can be seen. The fraction of cells expressing mCherry ('ON' cells) is plotted as a function of time. These fractions are distributed by exponentially decaying functions. It follows that the establishment times themselves are also exponentially distributed: given an exponentially decaying, normalized, probability distribution $p(t') = \frac{1}{\tau}e^{t'/\tau}$, where t' is the time at which the polymer switches due to heterochromatin establishment, the fraction f_{ON} of 'ON' cells is given as:

$$f_{\mathrm{ON}}(t) = 1 - \int_0^t p(t')dt' = e^{-t/\tau} \tag{5.3}$$

The other important finding is that the mean values of the distribution functions depend on system size - the bigger the system, the higher the mean value of heterochromatin establishment times.

The authors also created a physical model to reproduce the data, a model which is markedly different from our model (see section 4.3). In the following, the results obtained using our 3D model will be presented.

Implementation

It was desirable to reproduce the data by simulating different system sizes. Due to computational constraints - simulation time becomes infeasibly long with longer polymers - it was instead decided to vary the *cenH* size and keep the system size constant to achieve the same *ratio* of *cenH* size to system size as in the study by Nickels et al. It is also worth noticing that the system size used here (40 monomers) is smaller than what would have been needed to represent a 23 kb DNA chain; in their model implementation, the authors used 153 monomers

Strain size (Nickels et al.)	No. of nucleosomes (Nickels et al.)	Ratio of *cenH* size to system size	*cenH* size in 3D model (system size 40)
23 kb	153	0.20	8
27.5 kb	182	0.17	7
29 kb	191	0.16	7
31 kb	203	0.15	6

Table 5.3: Comparison of the models by Nickels et al. and the 3D model presented in this study. A constant *cenH* size of 31 monomers was used by the authors, and the fractions arise by varying the system size. The corresponding *cenH* size in our 3D model can be seen in the rightmost column.

to represent the same chain, and a constant *cenH* size of 31 monomers while varying the system size.

A summary of the different models and their parameters can be seen in table 5.3. As is shown here, as well as in tables 5.1 and 5.2, it was decided to use the *cenH* values of 6, 7, and 8, while keeping the system size constant at 40 monomers. Simulations could now be run, and the S state establishment times could be measured. *Establishment time was defined as the time between the start of the simulation until 90% of the polymer was in the S state.* To estimate the mean parameter τ, and thus define the distribution for each *cenH* size, the data was then collected from 100 simulations each.

It is worth noting that the since the sampled values are assumed to be distributed according to a decaying exponential function, which has a domain of $[0, \infty]$, some values would inevitably never be sampled due to finite simulation time. The data from such 'shortened' simulations would in turn be biased towards lower values. To avoid running very long simulations, it was decided to introduce a cutoff time $t_{\max}$ at some value of t within which a high number of samples was expected. Using the knowledge of the number of simulations performed, the number of data points sampled, and the values of these data points, an estimate of the real mean parameter could then be achieved using *maximum likelihood estimation*. The values generated from this estimator are also biased, but since this bias becomes insignificant for high simulation numbers (i.e. 100 simulations was sufficient), the raw estimates were used as is. The details on the derivation and construction of the estimator can be found in Appendix B.

Results and discussion

The results from these simulations can be seen in figure 5.13. Here, eq. 5.3 was used to arrive at the values seen in the main plot. We observe that the smaller the *cenH* length, which corresponds to a larger system size in the study by Nickels et al., the larger the mean S state establishment time for the given system. This is exactly the same trend seen in the real data (see figure 5.12). Furthermore, there are some clear differences in the behavior for different values

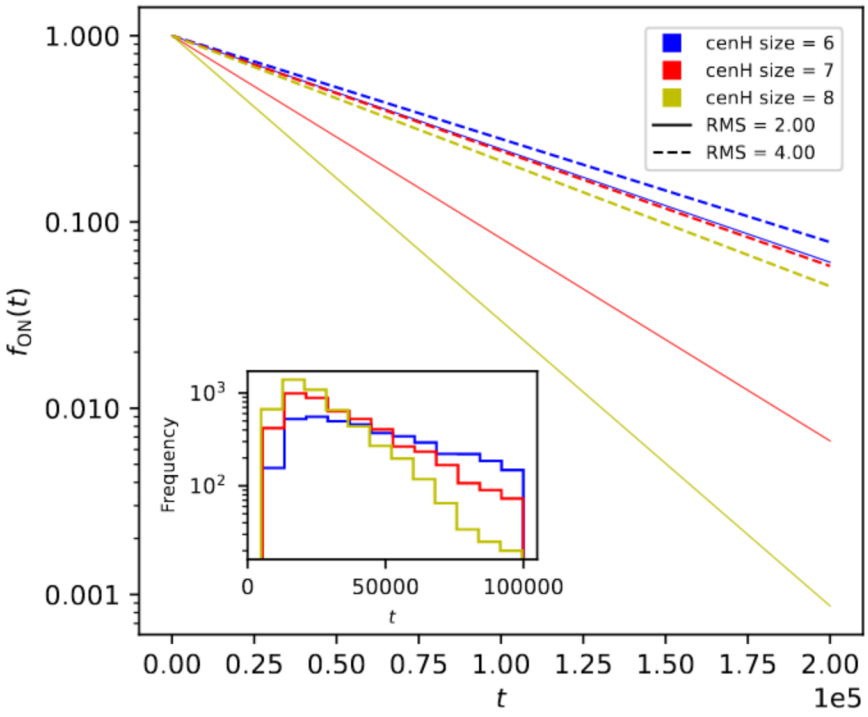


Figure 5.13: Main: The estimated exponentially decaying fractions of 'ON' polymers with different *cenH* sizes, and for different RMS values, after a total of 5 000 simulations. The bigger the fraction of *cenH* to the total polymer length, the lower the average switching time. **Inset:** Sampled establishment times from 5 000 simulations for RMS $\approx$ 2.0. Parameters: total time = 10^5, initial state = A, $\alpha_1 = 0.07$.

of RMS. First, for RMS $\approx$ 4.0, all mean establishment times are significantly longer than for their corresponding *cenH* sizes with RMS $\approx$ 2.0. Second, for RMS $\approx$ 2.0, the slopes are markedly different, whereas for RMS $\approx$ 4.0, the slopes are almost identical, and there is only a slight effect of different *cenH* sizes. This may be attributed to what was remarked earlier, and what can be seen in figure 5.9, namely that without a significant amount of pressure implemented, the spreading of the S state on the polymer happens almost exclusively from neighbor to neighbor, and with the same frequency. Thus the effect on the establishment times should only be attributed to the mere fact that the *cenH* sizes differ by one or two monomers. With pressure effects, and thus a smaller RMS, there are more global interactions and recruitment conversions, and the data indicates that this is a significant effect. This once again gives support to the argument for implementing the pressure potential.

In the inset plot of the same figure, the actual sampled establishment times can also be seen. After an initial period of $\sim 10\,000$ time steps, a decay is observed which appears to be exponential. Furthermore, we see that with a bigger *cenH* region, the faster the decay. As a caveat, we note that the exponential decay is only observed from $\sim 10\,000$ until the cutoff time. This most likely stems from the fact that some time interval is *required* from the beginning of the simulation until state switching can occur, as the monomers states require finite time to spread. This could have implications on the assumed probability distribution of the establishment times, which is something that should be investigated in future work.

5.4.3 Optimizing the parameter α_1

Lastly, it was attempted to optimize the probability constant α_1 to investigate if the slopes in figure 5.13 could be tuned to match with those in figure 5.12. It is worth repeating that in this study, there is no utilization of the real time scales in play, whereas in the data with which the results from this model was compared, the real time scales are known. The following should therefore be viewed as an initial attempt of parameter optimization, and as evidence that the S-state establishment times could be tuned in our model.

Because of the above, it was decided to optimize using the *fractions* of the slopes of the data in figure 5.12. Polymers with a *cenH* size of 6 and 8 were chosen, and using the same maximum likelihood estimation procedure as described in Appendix B, the fractions of the (linear) slopes of the estimated distributions of S state establishment times for these two systems were used as the value to compare with the data. In the data seen in figure 5.12, the actual value of the fraction of the slopes (representing the system sizes of 31 kb and 23 kb, respectively) was:

$$f_{\text{slopes}} = \frac{-0.009}{-0.184} = 0.04891 \tag{5.4}$$

Using this, the loss function could be constructed:

$$f_{\text{loss}}(\alpha_1) = [f_{\text{slopes}} - f_{\text{res}}(\alpha_1)]^2, \tag{5.5}$$

where $f_{\text{res}}(\alpha_1)$ is the estimated slopes from using the optimization procedure with our model.

Some proposed values of α_1 were too low for any of the simulations to yield a single establishment time (i.e. none of the polymers switched from the overall initial A state to the S state). In this case, the constructed loss function was:

$$f_{\text{loss}}(\alpha_1) = \frac{1}{10\alpha_1} \tag{5.6}$$

This function ensured high values for these low values of α_1, while still yielding values within the same order of magnitude as the loss function in eq. 5.5.

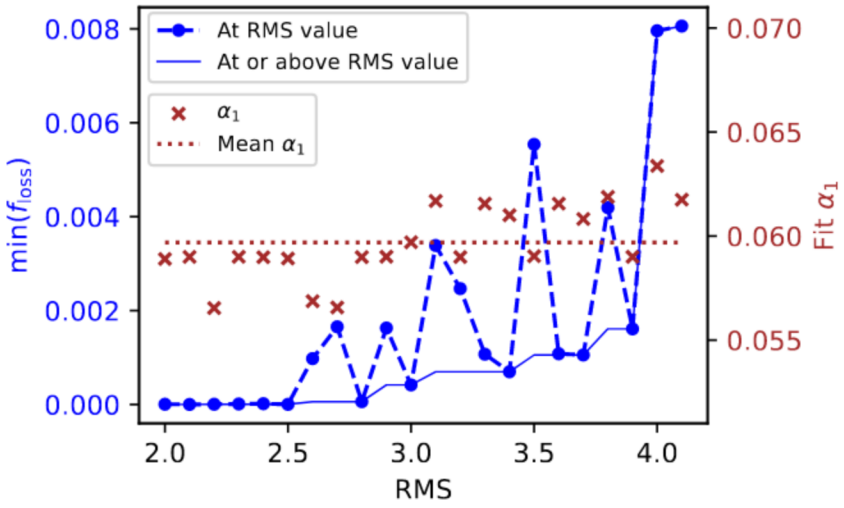


Figure 5.14: The fit values of α_1 along with the corresponding minimum values of the loss function, f_{loss}. These values are shown both for the actual RMS value used, and for the RMS value or higher. Parameters: total time $= 5 \cdot 10^4$, initial state = A.

Implementation

The optimizer algorithm used was the `forest_minimize` tree-based regressor from the SCIKIT-OPTIMIZE Python library. The model minimizes a value by evaluating the parameter that yielded the next lowest fit value up until the current iteration. For each proposed value of α_1, 100 evaluations were done for each of the two values of *cenH* size, including 10 initial, random evaluations. The search range for the value of α_1 was set to $[0.01, 0.2]$, based on our knowledge of the behavior of the polymer from earlier results. Finally, optimization was done for the RMS values in the range of $[2.0, 4.1]$. (For a visualization of the procedure, see Appendix D).

Results and discussion

The results from the optimization can be seen in figure 5.14. We first note that the values of α_1 resulting from the optimization all lie close to the value of $\alpha_1 \approx 0.06$. This value is within the same regime as the value used for generating most results in this thesis, namely $\alpha_1 = 0.07$ (see figure 5.5). A slight possible trend towards higher values can be observed with higher RMS values. As seen in figure 5.13, the S state takes relatively long to spread at high RMS values, so higher fit values for α_1 could be an indication that in this regime, a higher probability of the S state spreading is preferred. On the other hand, it is too

early to conclude that this could be the case, due to insufficient data, and possibly noise from the simulations.

Generally, we observe the trend that the minimum values of the loss function are stably low for RMS values of 2.0-2.5, all values lying around 0. For higher values, the values begin to fluctuate, in addition to trending upwards. The fluctuation could indicate that the optimization functions did not fully converge during the simulations - running more simulations might have reduced this noise. However, the upward trend tells us that there might be a *lower limit* as to how much pressure has to be implemented for the model to be able to represent the real data, as we saw in section 5.4.2, and more specifically in figure 5.13. Here, a pressure corresponding to an initial RMS value of 2.5 seems to be necessary for convergence to be achieved, and a lower loss value is never observed for higher RMS values. This adds further support for the argument that pressure is a *necessary* component in this 3D polymer model, and possibly that there exists a definable limit as to how much of an effect is required.

6 Conclusion

In this study, it was investigated whether nucleosome dynamics from a 0D model (see chapter 3) could be coupled with 3D polymer dynamics to yield a model for heterochromatin establishment on a DNA chain (chapter 4). The ability of the model to produce a system with one or more of the essential features of an epigenetic switch was gauged (section 5.3). Furthermore, the effects from adding a pressure potential to the Brownian dynamics were investigated (section 5.4).

6.1 Epigenetic switching

The model was found to be *suitable* for modelling an epigenetic switch. Both *bistability* and *hysteresis* were observed in simulations, with high stability and system memory for both system states (section 5.3.1). Switches back and forth between the system states were not observed, possibly due to low noise, but was not considered crucial to the overall result, as ensuring state stability and system memory was prioritized. *Epigenetic memory* was observed by introducing cell division events (section 5.3.2). Finally *heterochromatin establishment* with subsequent *epigenetic switching* were observed throughout. The switches occurred stochastically, and within relatively short time intervals (section 5.3.3). Additionally, it was possible to model the nucleation at the *cenH* region, and to find a parameter regime where this event biased the system towards a monostable S-state system, representing heterochromatin establishment, and subsequent gene silencing.

6.2 The effects of external pressure

Furthermore, it can be concluded that the introduction of a pressure potential was *crucial* in the model presented in this study. This is based on evidence from several of the simulations performed here. S-state spreading happened only *gradually, and mostly via nearest-neighbor interactions* for low pressure values, whereas *stochastic, sudden switches* were observed with significant pressure (section 5.4.1). A clear difference in the number of 'global' interactions was also observed, where the pressure appeared to ensure that monomers far apart on the chain were able to interact with a much higher frequency. This finding was supported by the differences in the number of isolated S-state patches, where higher pressure values yielded more patches, and by differences in the switching times, which were significantly shorter with higher pressure, in line with real data (section 5.4.1). In the comparison with the study by Nickels et al. [20], a qualitative match between the model data and the data from the

study, i.e. exponentially decaying behavior for establishment times, along with a connection between mean heterochromatin establishment times and system size, could only be achieved when pressure was implemented (section 5.4.2). Furthermore, the quantitative comparisons via parameter optimization yielded the lowest errors for higher pressure values (section 5.4.3). All of these findings support the argument for the use of a pressure potential in this model.

7 Outlook

There is a wealth of possibilities for further improvements and developments of the model presented in this thesis. First it is prudent to look at what the model was not able to accomplish. Looking at the three-dimensional structures emerging from the simulations (figures 5.3 and 5.4), we see a plethora of microstructures for both RMS values, including loops and coils. However, a zigzag pattern at the lowest level of organization was not observed (see section 2.2), so this could be a starting point in working towards a model which is better able to represent real three-dimensional heterochromatic structures. On the other hand, the observed structures indicate that it is possible to produce 3D structures at higher levels of organization than the zigzag structure. This could also be an interesting point of departure for further investigations.

Furthermore, the model was not able to mimic the behavior of ΔK mutants, in that switches back and forth between system states were not observed (section 2.2.2). This points to a mismatch between recruited and noisy nucleosome conversion rates, something which could easily be investigated, although there obviously is a trade-off between inducing system state switches and maintaining system state stability. In general, knowledge on the relationship between the dynamic time scale and the time scale for nucleosome conversions should be implemented to get realistic system dynamics.

Lastly, the incorporation of further biological features is another low-hanging fruit, and some can be implemented relatively easily in our model. For example, the implementation of inert borders in the genetic region to maintain heterochromatic areas within the region could be an interesting next step in future work.

Appendices

A Derivation of the expression for the numerical Euler iteration of the non-linear Langevin equation

In the following we will show the derivation [33] for eq. 4.1. Let $\mathbf{X}(t)$ be the $N \times 3$ monomer position matrix, where $\mathbf{X}_i(t)$ is the position of the i'th monomer at time t. Then, by the over-damped Langevin equation:

$$\frac{d}{dt}\mathbf{X}_i(t) = -\nabla_{\mathbf{X}_i(t)} U(\mathbf{X}(t)) + \mathbf{R}_i(t), \tag{A.1}$$

where U is the aggregate potential (which depends on the total position matrix), and $\mathbf{R}_i(t)$ is the random Gaussian noise on the i'th monomer at time t. This equation can be integrated from t to Δt as:

$$\int_t^{\Delta t} \frac{d}{dt'} X_{ij}(t')dt' = -\int_t^{\Delta t} \nabla_{X_{ij}(t')} U dt' + \int_t^{\Delta t} R_{ij}(t')dt', \tag{A.2}$$

where X_{ij} is a component of the position of the i'th monomer. The LHS can be expanded as:

$$\int_t^{\Delta t} \frac{d}{dt'} X_{ij}(t')dt' = X_{ij}(t+\Delta t) - X_{ij}(t) \equiv \Delta X_{ij}(t) \tag{A.3}$$

The first term on the RHS requires a bit more work. Expanding around Δt, we get:

$$\begin{aligned} \int_t^{\Delta t} \nabla_{X_{ij}(t')} U dt' &= \nabla_{X_{ij}(t')} U \Delta t + \frac{1}{2}\frac{d}{dt}\nabla_{X_{ij}(t')} U \Delta t + \dots \\ &\approx \nabla_{X_{ij}(t')} U \Delta t + \frac{1}{2}[\nabla_{X_{ij}(t')} U]' \Delta X_{ij}(t) \Delta t \end{aligned} \tag{A.4}$$

Now for the second term on the RHS of eq. A.2:

$$\int_t^{\Delta t} R_{ij}(t')dt' \equiv \Delta W_{ij}(t) \tag{A.5}$$

This is a stochastic integral, as R_{ij} is random Guassian noise, with the following properties: $\langle R_{ij}(t) \rangle = 0$, and $\langle R_{ik}(t) R_{jl}(t') \rangle = 2\delta(t-t')\delta_{ij}\delta_{kl}$, where $\delta(t)$ is the Dirac delta function. By the central limit theorem [34], the sum S of n independent variables becomes Gaussian as $n \to \infty$, as long as the mean μ and variance V of the distribution(s) these variables are taken from is/are defined.

In addition, $\langle S \rangle = \sum_k \mu_k$ and $V(S) = \sum_k \sigma_k^2$. This means that $\Delta W_{ij}(t)$ is also a Gaussian variable, and its properties can be derived from what we know about R_{ij}:

$$\langle \Delta W_{ij}(t) \rangle = \int_t^{\Delta t} \langle R_{ij}(t') \rangle dt' = 0, \tag{A.6}$$

and:

$$\begin{aligned} \langle \Delta W_{ij}(t)^2 \rangle &= \langle \int_t^{\Delta t} R_{ij}(t'')dt'' \int_{t'}^{\Delta t'} R_{ij}(t''')dt''' \rangle \\ &= \int_t^{\Delta t} dt'' \int_{t'}^{\Delta t'} dt''' \langle R_{ij}(t'')R_{ij}(t''') \rangle \\ &= \int_t^{\Delta t} dt'' \int_{t'}^{\Delta t'} dt''' 2\delta(t'' - t''') \\ &= 2\int_t^{\Delta t} dt'' = 2\Delta t \end{aligned} \tag{A.7}$$

Using these properties, we can construct the probability distribution for $\Delta \mathbf{W}(t)$:

$$p(\Delta W_{ij}(t)) = \frac{1}{\sqrt{4\pi\Delta t}} \exp\left[\frac{-\Delta W_{ij}^2}{4\Delta t}\right] \tag{A.8}$$

Since $\Delta W_{ij}(t)^2$ is $O(\Delta t)$, then $O(\Delta X_{ij}(t)) = O(\sqrt{\Delta t})$. Now, putting everything together, and neglecting all terms higher than $O(\Delta t)$, eq. A.2 can be written out as:

$$\mathbf{X}_i(t + \Delta t) = \mathbf{X}_i(t) - \nabla_{\mathbf{X}_i(t)} U(\mathbf{X}(t))\Delta t + \Delta \mathbf{W}_i(t), \tag{A.9}$$

which is the final result.

B Maximum likelihood estimation

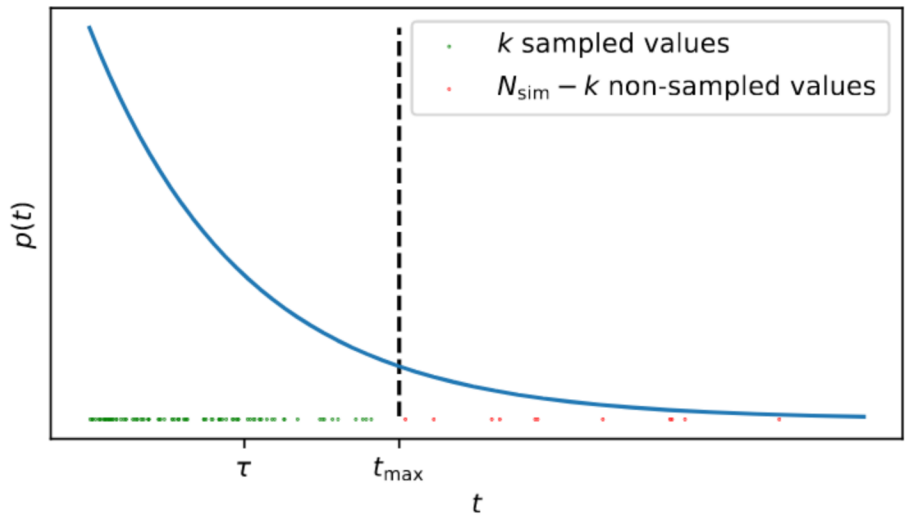


Figure B.1: Sampling from an exponential distribution, using a simulation time cutoff, $t_{\max}$. The values in green are the actual sampled values, while the red values are hypothetical values that are never sampled because of the cutoff time.

In the following we will see how to obtain an estimate of the parameter τ using simulations with a cutoff time. The decay times of a polymer is assumed to be distributed according to an exponential distribution, which can be seen in figure B.1. The distribution has the following expression:

$$p(t) = \tfrac{1}{\tau} e^{-t/\tau} \tag{B.1}$$

where τ is the characteristic decay time for a polymer with a given set of parameters. Simulations are stopped at the cutoff-time $t_{\max}$, so only values below this value can be sampled. Let the set of sampled values be given as $\{t_i | 1 \leq i \leq k\}$, and the total number of simulations as N_{sim}. In total, k values of t will be at or below $t_{\max}$. Using *maximum likelihood* estimation, the following expression is obtained:

$$\mathcal{L}(\tau) = \prod_{i=1}^{k} \tfrac{1}{\tau} e^{\frac{-t_i}{\tau}} \cdot \prod_{k+1}^{N_{\text{sim}}} e^{\frac{-t_{\max}}{\tau}} \tag{B.2}$$

Expanding and setting the derivative of the natural logarithm of eq. B.2 to 0, we get:

$$\begin{aligned}\mathcal{L}(\tau) &= \prod_i^k \tfrac{1}{\tau} e^{\frac{-t_i}{\tau}} \cdot \prod_{k+1}^{N_{\text{sim}}} e^{\frac{-t_{\max}}{\tau}} \\ &= \prod_i^k \tfrac{1}{\tau} e^{\frac{-t_i}{\tau}} \cdot e^{\frac{-(N_{\text{sim}}-k)t_{\max}}{\tau}}\end{aligned} \tag{B.3}$$

We then take the logarithm of the above:

$$\ln \mathcal{L}(\tau) = \sum_i^k (-\ln \tau - \tfrac{t_i}{\tau}) - \tfrac{(N_{\text{sim}}-k)t_{\max}}{\tau} \tag{B.4}$$

Next, we set the derivative to 0:

$$\frac{\partial}{\partial \tau} \ln \mathcal{L}(\tau) = \sum_i^k (-\tfrac{1}{\tau} + \tfrac{t_i}{\tau^2}) + \tfrac{(N_{\text{sim}}-k)t_{\max}}{\tau^2} = 0, \tag{B.5}$$

and solve for the estimate of τ:

$$\hat{\tau} = \tfrac{1}{k} \sum_i^k t_i + (\tfrac{N_{\text{sim}}}{k} - 1) t_{\max} \tag{B.6}$$

This is the main result, which gives us an estimation for the parameter τ. The second derivative can be used to find the error:

$$\text{Var}(\hat{\tau}) = -\left[\frac{\partial^2}{\partial \tau^2} \ln \mathcal{L}(\tau)\right]^{-1} = \left[2 \sum_{i=1}^k \tfrac{t_i}{\tau^3} - \tfrac{k}{\tau^2} + 2(N_{\text{sim}} - k)\tfrac{t_{\max}}{\tau^3}\right]^{-1} \tag{B.7}$$

The estimator is biased for small values of N_{sim}. This can be seen by calculating the average of the computed estimators:

$$\langle \hat{\tau} \rangle = \langle \tfrac{1}{k} \sum_i^k t_i \rangle + (N_{\text{sim}} \langle \tfrac{1}{k} \rangle - 1) t_{\max} \tag{B.8}$$

Here the first term is the expectation value of the values $t_i \leq t_{\max}$. This can be found analytically by integrating the (normalized) probability distribution over the interval $[0, t_{\max}]$:

$$\langle \tfrac{1}{k} \sum_i^k t_i \rangle = \int_0^{t_{\max}} (1 - e^{\frac{-t_{\max}}{\tau}})^{-1} \tfrac{t}{\tau} e^{\frac{-t_i}{\tau}} dt = \tau - \frac{t_{\max}}{e^{t_{\max}/\tau} - 1} \tag{B.9}$$

The second term of eq. B.8 requires a bit more work. First, the expectation of $\frac{1}{k}$ needs to be found, using the distribution for k. For a given simulation, k is

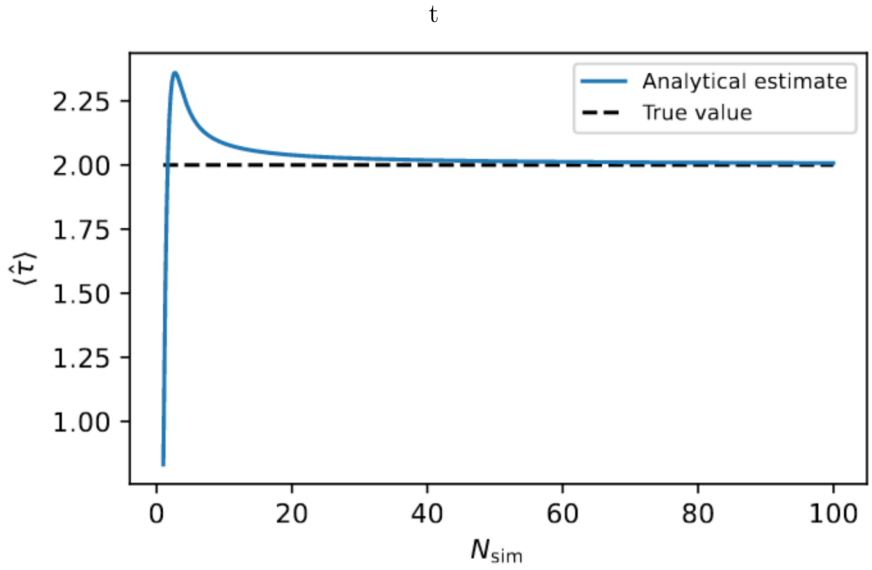


Figure B.2: The average estimate for τ becomes better with a higher number of simulatons. This example is shown assuming a true value of $\tau = 2$.

the number of data points that fall below the cutoff, while the rest fall above. k is thus binomially distributed, according to:

$$P(k) = \binom{N_{\text{sim}}}{k} p^k (1-p)^{N_{\text{sim}}-k} \tag{B.10}$$

where the probability parameter $p = 1 - e^{\frac{-t_{\max}}{\tau}}$. Using this, we can compute the expectation of $\frac{1}{k}$:

$$\begin{aligned} \langle \tfrac{1}{k} \rangle &= \sum_{k=1}^{N_{\text{sim}}} \tfrac{1}{k} P(k) \\ &= \sum_{k=1}^{N_{\text{sim}}} \tfrac{1}{k} \binom{N_{\text{sim}}}{k} p^k (1-p)^{N_{\text{sim}}-k} \\ &= N_{\text{sim}} p (1-p)^{N_{\text{sim}}-1} \cdot {}_3F_2(1, 1, 1 - N_{\text{sim}};\ 2, 2;\ \tfrac{p}{p-1}) \end{aligned} \tag{B.11}$$

where ${}_3F_2$ is the generalized hypergeometric function. Combining the results from equations B.8-B.11, we get:

$$\langle \hat{\tau} \rangle = \tau - \left[\frac{1}{e^{t_{\max}/\tau} - 1} - N_{\text{sim}}^2 p (1-p)^{N_{\text{sim}}-1} \cdot {}_3F_2(1, 1, 1 - N_{\text{sim}};\ 2, 2;\ \tfrac{p}{p-1}) + 1 \right] t_{\max} \tag{B.12}$$

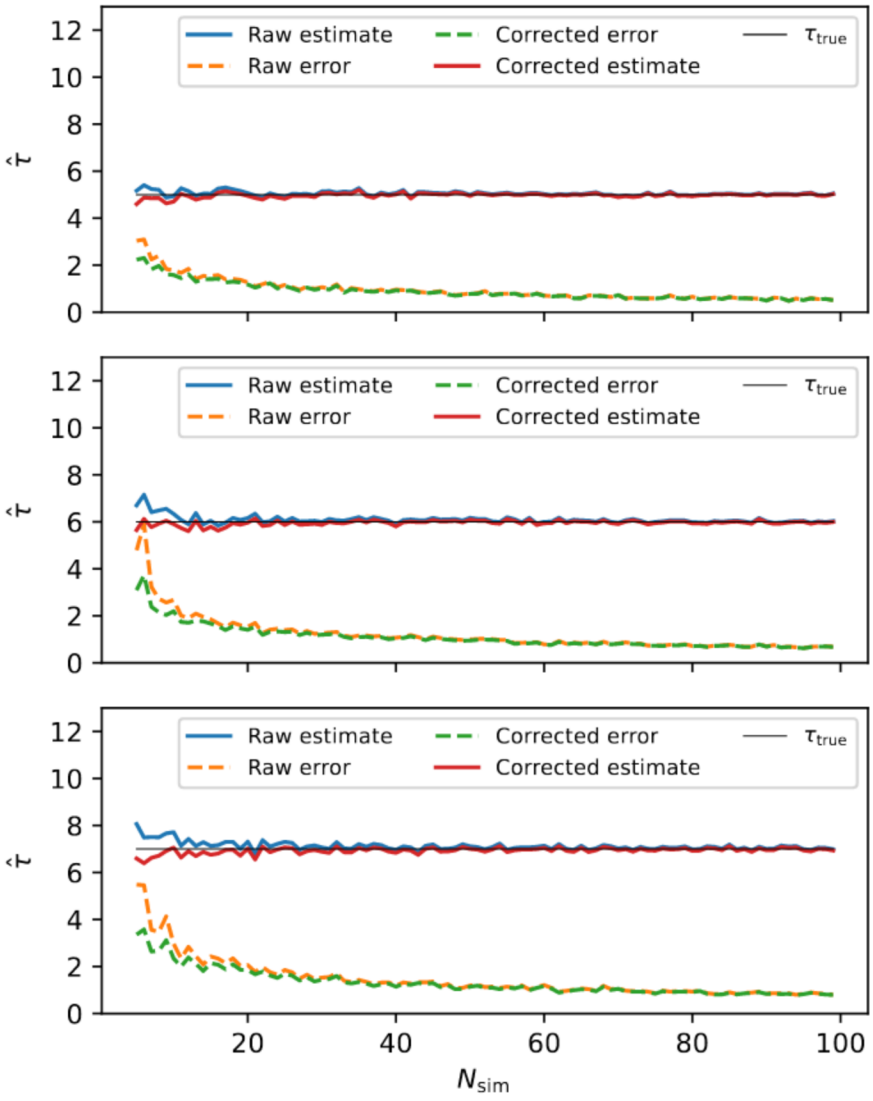


Figure B.3: Average values of estimates and errors based on 150 runs for each value of N_{sim}.

This function can be seen in figure B.2, and shows that the average estimate of τ becomes more and more accurate with a higher number of simulations. By running some simulations, this could be tested empirically. For different values of N_{sim}, the estimates of τ and its error were generated, taking the averages of 150 independent runs (see figure B.3). The estimates and errors were also corrected using the following constructed function:

$$\hat{\tau} = \frac{N\tau}{N - \pi \cdot \exp[-t_{\max}/\tau]} \tag{B.13}$$

As can be seen in the figure, however, the difference between the biased, raw estimate and the values from the constructed function vanish for high simulation numbers. Using a sufficient number of simulations, the bias can thus be neglected.

C Implementation of the algorithm for selecting interaction pairs for the S state monomers

```
import numpy as np
from numba import njit


@njit
def get_two_interaction_mask(norms_all, state_two_interaction,
                             i_idx, j_idx, N_ALLOWED_INTERACTIONS=2):

    # Total number of monomers
    N = len(norms_all)

    # Shows which monomers interact with which
    two_interaction_mask = np.zeros(norms_all.shape, dtype=np.bool_)

    # Sort distances
    norms_all_flattened = norms_all.flatten()
    # The indices that sort norms_all_flattened
    sorted_indices = np.argsort(norms_all_flattened)
    # The indices that sort norms_all
    i_idx = i_idx.flatten()[sorted_indices]
    j_idx = j_idx.flatten()[sorted_indices]

    # Counts no. of interactions per monomer
    n_interactions = np.zeros(N, dtype=np.uint8)
    # For stopping criterion
    has_counted = np.zeros(N, dtype=np.uint8)
    total_2_interactions = 0

    # Loop over combinations of indices
    for k in range(len(i_idx)):
        i = i_idx[k]
        j = j_idx[k]

        # Checks if both monomers are of the same
        # (interacting) state
        two_interaction = (state_two_interaction[i]
                           and state_two_interaction[j])

        # Only S states can interact
        if not two_interaction:
            continue

        # If the monomers are the same
        # or nearest neighbors
        if i == j or i == j+1 or i == j-1:
```

```
            continue

        # If there already exists an interaction
        # between the two monomers
        if two_interaction_mask[i,j] and two_interaction_mask[j,i]:
            continue

        # Two-interaction state monomers can only interact with
        # max. 2 other monomers
        if (n_interactions[i] >= N_ALLOWED_INTERACTIONS
                or n_interactions[j] >= N_ALLOWED_INTERACTIONS):
            continue

        # Create interaction
        two_interaction_mask[i, j] = 1
        two_interaction_mask[j, i] = 1

        n_interactions[i] += 1
        n_interactions[j] += 1

        # For stopping criterion
        if (not has_counted[i]
                and n_interactions[i] == N_ALLOWED_INTERACTIONS):
            total_2_interactions += 1
            has_counted[i] = True
        if (not has_counted[j]
                and n_interactions[j] == N_ALLOWED_INTERACTIONS):
            total_2_interactions += 1
            has_counted[j] = True

        if total_2_interactions >= N - 1:
            break

    return two_interaction_mask
```

D Optimization process visualization

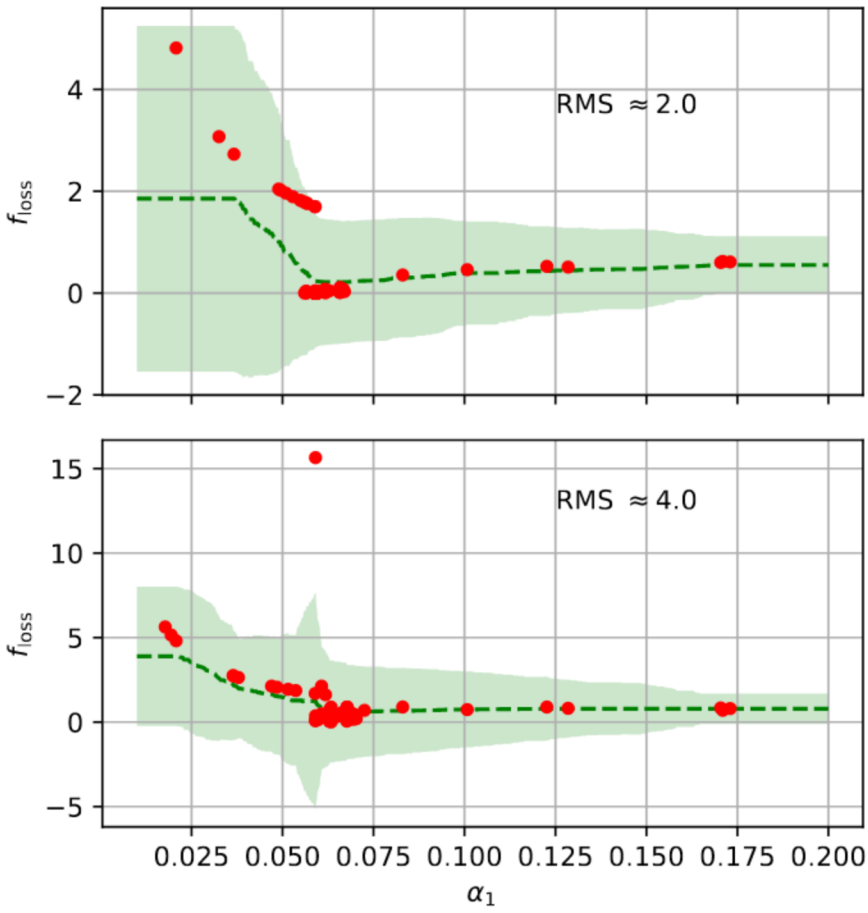


Figure D.1: Examples of the fitting procedure from the optimization of the parameter α_1. The red dots represent the proposed values of α_1, and the green area with the dotted green line shows the fitting landscape, with loss values and their uncertainty indicated.

Bibliography

[1] Scitable by Nature Education. A brief history of genetics: Defining experiments in genetics. unit 5: What is dna? what does dna do? https://www.nature.com/scitable/ebooks/a-brief-history-of-genetics-defining-experiments-16570302/126132310/, 2022. Accessed: May 19, 2022.

[2] Scitable by Nature Education. Discovery of dna structure and function: Watson and crick. https://www.nature.com/scitable/topicpage/discovery-of-dna-structure-and-function-watson-397/, 2022. Accessed: May 19, 2022.

[3] NobelPrize.org. Nobel Prize Outreach AB 2022. The nobel prize in physiology or medicine 1962. https://www.nobelprize.org/prizes/medicine/1962/summary/, 2022. Accessed: May 19, 2022.

[4] Ute Deichmann. Epigenetics: The origins and evolution of a fashionable topic. *Developmental biology*, 416(1):249–254, 2016.

[5] C-T Wu and James R Morris. Genes, genetics, and epigenetics: a correspondence. *Science*, 293(5532):1103–1105, 2001.

[6] Philip Avner and Edith Heard. X-chromosome inactivation: counting, choice and initiation. *Nature Reviews Genetics*, 2(1):59–67, 2001.

[7] Charles S Hoffman, Valerie Wood, and Peter A Fantes. An ancient yeast for young geneticists: a primer on the schizosaccharomyces pombe model system. *Genetics*, 201(2):403–423, 2015.

[8] Bruce Alberts, Alexander Johnson, Julian Lewis, David Morgan, Martin Raff, Keith Roberts, and Peter Walter. *Molecular Biology of the Cell (6th ed.)*. Garland Science, 2015.

[9] Encyclopaedia Britannica. Dna. https://www.britannica.com/science/DNA, 2022. Accessed: 10th May 2022.

[10] Sarah A. Sabatinos and Susan L. Forsburg. *Molecular Genetics of Schizosaccharomyces pombe*, pages 759–795. Elsevier Inc., Department of Molecular and Computational Biology, University of Southern California, Los Angeles, California, USA, 2010.

[11] Cathérine Dupont, D Randall Armant, and Carol A Brenner. Epigenetics: definition, mechanisms and clinical perspective. In *Seminars in reproductive medicine*, volume 27, pages 351–357. © Thieme Medical Publishers, 2009.

[12] Anna A Kalashnikova, Ryan A Rogge, and Jeffrey C Hansen. Linker histone h1 and protein–protein interactions. *Biochimica Et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1859(3):455–461, 2016.

[13] Andrés Penagos-Puig and Mayra Furlan-Magaril. Heterochromatin as an important driver of genome organization. *Frontiers in Cell and Developmental Biology*, page 982, 2020.

[14] Ke Zhang, Kerstin Mosch, Wolfgang Fischle, and Shiv IS Grewal. Roles of the clr4 methyltransferase complex in nucleation, spreading and maintenance of heterochromatin. *Nature structural & molecular biology*, 15(4):381–388, 2008.

[15] Robert J Klose, Eric M Kallin, and Yi Zhang. Jmjc-domain-containing proteins and histone demethylation. *Nature reviews genetics*, 7(9):715–727, 2006.

[16] David J Owen, Prisca Ornaghi, Ji-Chun Yang, Nicholas Lowe, Philip R Evans, Paola Ballario, David Neuhaus, Patrizia Filetici, and Andrew A Travers. The structural basis for the recognition of acetylated histone h4 by the bromodomain of histone acetyltransferase gcn5p. *The EMBO journal*, 19(22):6141–6149, 2000.

[17] Takatomi Yamada, Wolfgang Fischle, Tomoyasu Sugiyama, C David Allis, and Shiv IS Grewal. The nucleation and maintenance of heterochromatin by a histone deacetylase in fission yeast. *Molecular cell*, 20(2):173–185, 2005.

[18] Alexandra B Lantermann, Tobias Straub, Annelie Strålfors, Guo-Cheng Yuan, Karl Ekwall, and Philipp Korber. Schizosaccharomyces pombe genome-wide nucleosome mapping reveals positioning mechanisms distinct from those of saccharomyces cerevisiae. *Nature structural & molecular biology*, 17(2):251–257, 2010.

[19] Michaela J Obersriebnig, Emil MH Pallesen, Kim Sneppen, Ala Trusina, and Genevieve Thon. Nucleation and spreading of a heterochromatic domain in fission yeast. *Nature communications*, 7(1):1–11, 2016.

[20] Jan Fabio Nickels, Ashleigh Katrine Edwards, Sebastian Jespersen Charlton, Amanda Møller Mortensen, Sif Christine Lykke Hougaard, Ala Trusina, Kim Sneppen, and Geneviève Thon. Establishment of heterochromatin in domain-size-dependent bursts. *Proceedings of the National Academy of Sciences*, 118(15), 2021.

[21] Genevieve Thon, Klavs R Hansen, Susagna Padrissa Altes, Deepak Sidhu, Gurjeet Singh, Janne Verhein-Hansen, Michael J Bonaduce, and Amar JS Klar. The clr7 and clr8 directionality factors and the pcu4 cullin mediate heterochromatin formation in the fission yeast schizosaccharomyces pombe. *Genetics*, 171(4):1583–1595, 2005.

[22] Ira M Hall, Gurumurthy D Shankaranarayana, Ken-ichi Noma, Nabieh Ayoub, Amikam Cohen, and Shiv IS Grewal. Establishment and maintenance of a heterochromatin domain. *Science*, 297(5590):2232–2237, 2002.

[23] Genevieve Thon and Tove Friis. Epigenetic inheritance of transcriptional silencing and switching competence in fission yeast. *Genetics*, 145(3):685–696, 1997.

[24] Davide Michieletto, Enzo Orlandini, and Davide Marenduzzo. Polymer model with epigenetic recoloring reveals a pathway for the de novo establishment and 3d organization of chromatin domains. *Physical Review X*, 6(4):041047, 2016.

[25] Ian B Dodd, Mille A Micheelsen, Kim Sneppen, and Genevieve Thon. Theoretical analysis of epigenetic cell memory by nucleosome modification. *Cell*, 129(4):813–822, 2007.

[26] Daniel Jost and Cédric Vaillant. Epigenomics in 3d: importance of long-range spreading and specific interactions in epigenomic maintenance. *Nucleic acids research*, 46(5):2252–2264, 2018.

[27] N Haddad, D Jost, and C Vaillant. Perspectives: using polymer modeling to understand the formation and function of nuclear compartments. *Chromosome Research*, 25(1):35–50, 2017.

[28] PyTorch. A gentle introduction to `torch.autograd`. https://pytorch.org/tutorials/beginner/blitz/autograd_tutorial.html, 2022. Accessed: May 1, 2022.

[29] Davide Colì, Enzo Orlandini, Davide Michieletto, and Davide Marenduzzo. Magnetic polymer models for epigenetics-driven chromosome folding. *Physical Review E*, 100(5):052410, 2019.

[30] Chris A Brackley, Stephen Taylor, Argyris Papantonis, Peter R Cook, and Davide Marenduzzo. Nonspecific bridging-induced attraction drives clustering of dna-binding proteins and genome organization. *Proceedings of the National Academy of Sciences*, 110(38):E3605–E3611, 2013.

[31] Davide Michieletto, Michael Chiang, Davide Coli, Argyris Papantonis, Enzo Orlandini, Peter R Cook, and Davide Marenduzzo. Shaping epigenetic memory via genomic bookmarking. *Nucleic acids research*, 46(1):83–93, 2018.

[32] Kyosuke Adachi and Kyogo Kawaguchi. Chromatin state switching in a polymer model with mark-conformation coupling. *Physical Review E*, 100(6):060401, 2019.

[33] Namiko Mitarai. Diffusive and stochastic processes: Lecture notes. Course material for the course 'Diffusive and Stochastic Processes' at University of Copenhagen, apr 2021.

[34] R. J. Barlow. *A Guide to the Use of Statistical Methods in the Physical Sciences*. John Wiley & Sons, Department of Physics and Astronomy, University of Manchester, 1989.