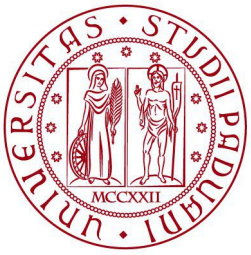


**Head Office:** Università degli Studi di Padova
**Department:** Physics and Astronomy "Galileo Galilei"

**Ph.D. Course in:** Physics
**Series:** XXXII

**Thesis title:**

# Magnetic Polymer Models for Epigenomic Spreading and Chromatin Organization

**Supervisor:** Prof. Enzo Orlandini

**Coordinator:** Prof. Franco Simonetto

**Ph.D. student:** Davide Colì

*"In the beginning the Universe was created. This has made a lot of people very angry and been widely regarded as a bad move"*

---

Douglas Adams, `The Restaurant at the End of the Universe`

**Abstract**

In this thesis we present a detailed investigation of the interplay between 3D organization of the chromatin and epigenomic spreading in Eukaryotic nuclei, via polymer physics models.

We begin with a review of the biology behind epigenetic processes, and of some basic physical models that have been proposed to describe them. We also examine the model presented in [1], and study in details its equilibrium and non-equilibrium dynamics via extensive molecular dynamic simulations. At equilibrium we confirm the existence of a first-order phase transition between a swollen and epigenetically disordered phase, and a compact ordered one. At non-equilibrium we prove the existence of two novel phases [2] where the polymer is organized in a compact-disordered or swollen-ordered fashion.

We extend the model by inserting "genomic bookmarking" [2], that is transcription factors permanently bound to the DNA and which enhance the positive feedback loop in the epigenetic machinerie. We also develop a more realistic, biologically inspired, model which successfully reproduces the distribution of epigenetic marks in a *Drosophila* chromosome.

We next discuss the model proposed in [1] and its relation with magnetic polymer models on a lattice [3]. We find a Landau-Ginzburg-like expression for the free-energy of a dense magnetic polymer in the mean-field approximation. The new free-energy, and *ad hoc* Montecarlo simulations, confirm the presence of a first-order phase transition between a swollen-disordered phase and a compact-ordered one at equilibrium. We also study the non-equilibrium kinetic of the mean-field model by employing a set of opportunely modified "Model A" equations.

At last, we develop a phenomenological mean-field model for a melt of chromatin fibers in a closed system akin to the Eukaryotic nuclei [4]. The phenomenological free-energy will depend on a conserved density field and a non-conserved magnetisation field. The equilibrium phases will be hence investigated analytically via a common-tangent construction construction, and include both uniform and demixed phases undergoing phase separation. The dynamics of the equilibrium phases is then studied via a set of "Model C" equations in order to estimate the critical coarsening exponents. Finally, we insert in the kinetic equations some non-equilibrium terms akin to a first-order reaction which arrest the phase separation.

# Acknowledgements

My first day of PhD feels like yesterday, but at the same time it feels ages have passed from when I was "just" a master student. These last three years have been a source great satisfactions, but at the same time they were background for some dark personal battles. I have to admit that without the help of other wonderful people, maybe this thesis would not even exist.

First and foremost, I have to thank my parents. Thank you for everything, really. Without your support I wouldn't be a Ph.D. student, I wouldn't be a physicist, I wouldn't be me.

Second, but maybe not less important in this context, I have to thank my supervisor, Enzo Orlandini. Thank you for putting up with me and my quirks during all these long years, and being a wise guide in the intricate road of academic research.

I want to thank Davide Michieletto and Davide Marenduzzo with whom I collaborated during these years. My visits to Edinburgh, and research in general, have been the most enjoyable even thanks to you. I also thank you for your support, even if silent, during the last period: I would probably have not been so patient with myself.

I want to thank Prof. Franco Simonetto, my PhD coordinator. Without your interest and understanding, who knows where would I be right now?

Thank you Alberto, Luca, and Sebastiano: good friends make days less gray. Thank you Andrea, Valentina, Federico, Matteo, and Gianluca: my PhD compatriots, united we are strong. Thank you Alessandra and Alba: let us keep up the good old philosophic calls.

Finally, a thesis is not only made of written words, but also of pictures. For this I re-thank by artistic trio: Alessandra, Sebastiano and Luca (a.k.a. The PowerPoint Ninja).

# Contents

# Introduction

*Would it save you a lot of time if I just gave up and went mad now?*

---

Douglas Adams, `The Hitchhiker's Guide to the Galaxy`

The term epigenetics usually refers to the study of stable and inheritable changes in the phenotype of an organism due to modifications in the gene expression, rather then in the genome itself [5]. Processes of this kind are actually quite common in the biology of higher organisms like eukaryotes. Let us think, for example, of a multicellular organism who presents a variety of tissues. Two cells from different tissues will have the same genome, but at the same time they are highly specialized and their function and form will be completely different from one another. This is not, by far, the only example of epigenetic, as modifications in phenotype are usually observed as a method to react to external stimuli, like what is observed in general in the polyphenism of insects [6], or in the temperature-dependence of the sex of some species of fishes and reptiles [7], or even with stressful situations in humans [8]. From a microscopic point of view, changes in gene-expression are thought to arise through a series of biochemical modifications, called *epigenetic marks*, of nucleosomes [9, 10] which compose a molecular pattern sometimes called *histone* or *epigenetic marks*. On the other hand, changes in gene-expression are also observed to be correlated with changes in the 3-dimensional structure of the whole chromatin complex [11]. However, while this correlations are known, the exact mechanism regulating the interaction between these three players (gene expression, histone modifications, and chromatin structure) is not fully understood, and has been one of the most studied problems in biology in the last years.

In this context, polymer-physics based models, with a statistical physic approach have been proven useful to uncover some of the fundamental mechanisms that shape chromatin organization, aiding a more insightful understanding on how genome architecture and biological processes are intertwined. Indeed, while biophysical models tend to oversimplify the dynamics by omitting various biological components, even basic homopolymer physics has proven quite effective at explaining some key features

of nuclear architecture in a variety of organisms [12], while heteropolymer models have been employed in order to discuss finer details of the spatial organization [13].

While the correlation between 3-dimensional structure of chromatin and histone code is clear, and well documented [11, 14, 15], most physical models for the study epigenetics focus either on finding the chromatin architecture given a fixed epigenetic landscape [13], or on studying effective models for the spreading of epigenetic marks [16, 17]. Recently, a new model, proposed by Michieletto et al. [1], has the objective to reconcile these two aspects of epigenomic dynamics. Here, the chromatin is modeled as a semi-flexible heteropolymer whose beads can assume different colors, representing each a particular epigenetic mark. Differently from the "static landscapes" model, the "color" of the beads can change, and spread, in analogy with the positive-feedback loop between two classes of enzymes "readers" and "writers" which interact and influence directly the dynamics of the epigenetic landscape in the chromatin.

In this work, we will thoroughly examine the nature of the coupling between these two aspects, by developing a biophysical model to study, and reproduce, the formation of stable, robust, epigenetic landscapes. The analysis will be based on some extensions of the model proposed by Michieletto et al., which will be thoroughly studied both at equilibrium [1] and at non-equilibrium level [2]. The analysis has been performed mainly by employing extensive molecular dynamic simulations using the software `LAMMPS` [18]. We will show the existence of a variety of stable steady phases for long-time dynamics, which all correspond to actual structures observed *in vivo* in chromatin. Notably at equilibrium we will prove the existence of a first-order phase transition between a swollen, epigenetically disordered configuration, and a compact, epigenetically ordered one, which is important because proves that the model allows for a bistable epigenetic switch that can retain memory of its state. The model will be then extended to include a completely new biophysical mechanism, called *genomic bookmarking* [2], in order to explain how to reconcile the *de novo* formation of inheritable, yet plastic, epigenetic domains, in the face of the rapid turnover of the underlying histone marks due to cell mitosis. Interestingly, we will challenge the problem of studying the epigenomic spreading on a *Drosophila* chromosome by employing this new extended model; the results based on simulations are in good agreement with the experimental observations *in vivo* of the chromosome.

Since the model can be written in terms of an effective Hamiltonian, it is possible to write down a free-energy which will be studied in the mean-field limit. By using a Ginzburg-Landau description [3] in analogy with previously studied models of magnetic polymer on a cubic lattice [19]. The mean-field theory will also be tested by ad-hoc Montecarlo simulations, and later extended to describe out-of-equilibrium properties by using "Model A" kinetic equations [20]. With this approach we can prove again the existence of a first-order phase transition at equilibrium, while at the same time reproducing almost faithfully the non-equilibrium diagram found in [2]. Finally, we examine the structure of the whole Eukaryotic nucleus, by considering a phenomenological Landau-Ginzburg-like free-energy for the case of many chromatin

fibers in a non-dilute close system [4]. By inserting non-equilibrium "switching" terms in the kinetic equations of the system ("Model C" dynamics [20]), we can see the presence of arrested (and tunable) phase separation of epigenomic domains, in qualitative agreement with what observed in experiments.

This thesis is organized as follows:

**Chapter 1** We give a brief review of epigenetics (see section 1.1), with a particular emphasis on chromatin structure, and the biochemistry involved in epigenetic marks. We then introduce some models that have been previously employed to discuss chromatin structure in presence of a fixed epigenetic landscape (see section 1.2). The opposite case is also considered, namely we present 0-dimensional ("mean-field") and 1-dimensional models for epigenomic spreading (see section 1.3) where the underlying chromatin structure is neglected. Finally, we present the model proposed by Michieletto et al. in [1], which explicitly couples 3-dimensional chromatin organization with epigenomic spreading (see section 1.4). Here, chromatin is modeled as a semi-flexible polymer chain whose beads can bear different epigenetic marks represented by colors. The interaction among the beads is modeled via a Potts-like interaction (see section 1.4.1). The dynamics is the result of two contributions: "Langevin dynamics" which deals with the spatial dynamics of the polymer chain, and a "recolouring dynamics" which deals with epigenomic spreading and evolves via a Metropolis-Hastings-like algorithm. The equilibrium dynamics presented in section 1.4.3, has already been discussed in [1], but we re-derived all the results by performing independent molecular dynamic simulations. Notably, our results confirm the presence of a first-order phase transition between a swollen-disordered phase transition and a compact-ordered one, by employing the method of Binder cumulants. Finally, we also derive the non-equilibrium long-time dynamics of the model presented in [2] (see section 1.4.4). In this case, we show the presence of two novel phases, not observed at equilibrium, where polymer organizes in a compact-disordered or swollen-ordered fashion, and we discuss the nature of the possible phase transitions between all these phases.

**Chapter 2** We present a possible extension for our core model by introducing a novel biophysical mechanism called "genomic bookmarking". Here we suppose the existence of some transcriptional factors, called "bookmarkers", which remain bound to chromatin during mitosis, and have a role in guiding epigenomic spreading. All the results presented in this chapter have also been discussed in [2]. Here, we study the formation of epigenetic domains by varying the spatial configurations of genomic bookmarkers on the chain (see section 2.1.1). We prove that if a local minimum bookmarker density is present, coherent epigenetic domains can form and that they are stable for long times (see section 2.1.2), as experimentally observed in actual *in vivo* in chromatin. In section 2.2 we

present further extensions of the model. First, we discuss the dynamics in presence of asymmetric interactions between the various epigenetic marks (see section 2.2.1). Second, we introduce a random process in the model to simulate cell replication, by excising part of the bookmarked genome. We find that, even with the excision process, our epigenetic domains remain stable during several cell cycles. Notice that this proves that genomic bookmarking processes are good candidate to explain the *de-novo* formation of epigenetic domains after mitosis. Finally, we present the results of molecular dynamic simulations in the context of a *Drosophila* chromosome. The results of these simulations are in good agreement with experimental findings.

**Chapter 3** By referring to the model first presented in section 1.4, we develop here a Landau-Ginzburg mean-field theory in analogy with magnetic polymers. The main results of this chapter have been discussed in [3]. We consider a discrete model by mapping the polymer chain into an interacting self-avoiding walk (ISAW, see section 3.1) on a (hyper-)cubic lattice; with this we developed a `C++` software where we were able to to compare our mean-field results with direct Montecarlo simulations (see section 3.2). In order to derive the mean-field equations, we take several preliminary steps. First, we study the dynamics of a Potts-like system with non-interacting inert states on a lattice (see section 3.3), and we discuss what kind of field-theory one has to choose in order to observe a phase transition. We also observe that, when a phase transition is present, its order depends not only on the number of interacting states as in the classic Potts model, but also on the number of non-interaction inert states. Second, we present a Landau-Ginzburg theory for a dense homopolymer on a lattice (see section 3.4), and show that our model correctly predicts a continuous phase transition between a swollen and a compact phase, as in the classic Flory $\theta$-collapse. Finally, the full mean-field theory for magnetic polymers with Potts-like interactions (see section 3.5) is discussed. We show that in agreement with what observed in section 1.4.3, the system presents a first-order phase transition between a swollen-disordered phase and a compact-ordered one; these results are qualitatively, and quantitatively, in agreement with Montecarlo simulations. As a last step, we have examined the "Model A" dynamics of our model, in order to study the relaxation to equilibrium dynamics of the fields (see section 3.6). Thanks to these kinetic equations, we develop a non-equilibrium mean-field kinetic model and study its steady states to derive a non-equilibrium phase diagram. Notably, for high density polymers, the phase diagram we find from these equations is qualitatively identical to what found in section 1.4.4.

**Chapter 4** While in the previous chapters we have discussed the behaviour of isolated chromatin fibers in dilute solutions, here we will develop a mean-field theory in order to examine our model when considering the whole nucleus of an Eukaryotic cell, i.e. a close system with finite volume where several fibers are present. Results of the mean-field theory can all be found in [4] and

they are compared with molecular dynamics simulations [2]. We introduce a new phenomenological free-energy density (see section 4.1) as a function of the conserved field density and of the non-conserved magnetisation. We can study the equilibrium phases of this phenomenological model (see section 4.2) by employing the so-called "common tangent construction". This analysis show the existence of three different phases: one with uniform density and incoherent epigenetic magnetisation; another with uniform density but coherent magnetisation, and finally one characterised by phase separation where the nucleus organizes in high-density clusters with coherent epigenetic mark. We study the dynamics of the phase separation process by employing a set of "Model C" equations (see section 4.3), and find the respective coarsening exponents. Finally, we extend the kinetic equations to include a first-order reaction (see section 4.4) between an "active" component of the chromatin which participates to the recolouring dynamics, and an "inactive" component which does not. With this extension the model displays an interesting arrested phase separation where the system organizes in high-density clusters of steady size, in qualitative agreement with what is observed experimentally in cell nuclei.

# Chapter 1

# The Physics of Epigenetics

*"What is the secret of life" I asked.*
*"I forget," said Sandra.*
*"Protein," the bartender declared. "They found out something about protein."*
*"Yeah," said Sandra, "that's it."*

---

Kurt Vonnegaut, `Cat's Cradle`

A common misconception among non-biologists is that the birth, the life, and death of a human, or any other living creature, is solely governed by its DNA. However, we don't need to look too far away to understand the limits of this kind of thinking. Take for example two cells from two different tissues in the human body: they will have completely different shapes, appearances, and functions, but still, the same exact DNA sequence can be found in their nuclei. The reason for this is simple: while the genome in itself is the same, different cells have different genes expressed. The changes in organisms caused by the modification of gene expression rather than modifications of the genetic code are studied by that branch of biology called *epigenetics*.

In this chapter, we will give a brief introduction to the biology of epigenetics, as well as a more concise and precise definition of the term. After this biological parenthesis, we will examine the possibility of developing simple physical models to reproduce and explain, the main features observed in this kind of phenomena.

## 1.1 The Biological Picture

In this section, we will give a small review of the biological mechanisms behind epigenetics. While this is enough to understand the biological background employed in the rest of this work, it is not by any means exhaustive. For a more in-depth analysis see for example [21].

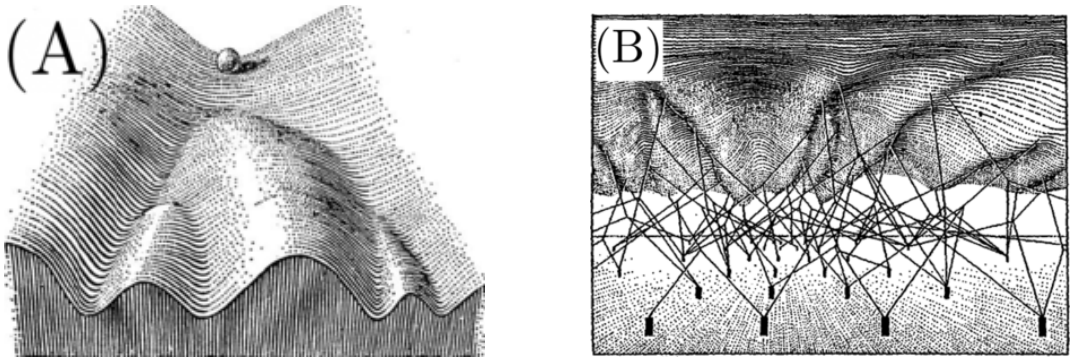



**Figure 1.1:** Pictorial depiction of the so-called "Waddington's epigenetic landscape".**(A)** The landscape represent by which the totipotent embryonic cell, represented by the marble at the top, differentiate while rolling down one of the possible epigenetic pathways; the end point at the end of the hill represent the possible cell fates, i.e. the different tissues present in an organism.**(B)** The landscape is determine by "hidden" interactions between the cell and the genome, here represented via some wires that shape the surface of the hill. Picture adapted from [23].

### 1.1.1 What is Epigenetics? Definition of the Term

While the study of epigenetics has exploded in the last decades, its basic concepts have been studied as long as a century. Indeed, the term "epigenetics" was first coined by Conrad Waddington [22] in 1942, and the origin of the word itself comes from Waddington's study of *epigenesis*, i.e. the process by which multicellular organisms develop from a single cell through a sequence cells differentiation. In the context of Waddington's studies, then, epigenetics is the study of how the genome guides the organism toward its development [23]. The ideas of Waddington are captured by the famous "Waddington's epigenetic landscape" (see fig. 1.1), where an initially totipotent embryonic cell, represented by a marble, rolls down a hill whose surface is shaped by the genome. While rolling, multiple "epigenetic pathways" are available, but eventually the cell will stop at one of the points at the base of the hill. These final stops of the marble represent the eventual cell fates, that is, tissue types.

The definition given by Waddington, however, is not the only one we can find in literature. A second definition was given by Nanney [24]. Indeed, while Waddington's epigenetic has a broad wide meaning, Nanney refers with the term to the existence of a non-genetic mechanism at the cellular level, which regulates the gene expression.

Given the existence in literature of various, non-equivalent definitions of epigenetics, no solid consensus of its precise meaning exists [25]. In recent years, a new precise definition has been proposed: "An epigenetic trait is a stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence" [5]. As this definition allows an operational treatment of the epigenetic subject, we choose to adopt it from now on.

### 1.1.2 Chromatin

Before we delve into the intricacies of the epigenetics processes, let us spend a few words on the substrate where they take place: the chromatin.

One of the main roles of DNA in the cell life, is the long-term preservation of the instructions needed to the cell, such as the instructions to build proteins and RNA. In the cell nucleus of eukaryotes, DNA is not a "naked molecule", but rather it is found in precise organized 3-d structure thanks to the help of some architectural proteins. The complex of DNA and these proteins it is called *chromatin* fiber, while the whole 3-dimensional structure is called *chromosome*.

The role of chromatin is an important one: left alone the multiple DNA molecules would tangle and could be easily damaged. Chromatin structure, instead, reinforces and packages DNA into denser configurations. Moreover, chromatin has an important role during mitosis (or meiosis) as it helps with the segregation of the chromosomes.

If we examine the microscopic structure of chromatin, we observe that it is formed by bead-like structures called *nucleosomes*. Each nucleosome complex is composed by a string of DNA wrapped around eight structural proteins, called *histones*, and it is linked to the next bead by a substring of DNA called *linker DNA* (see fig. 1.2). It is important to note that the nucleosomes are not electrically neutral, as the histones and DNA are both electrically charged and are the source of electrostatic fields in the nucleus [26] which contribute strongly to the folding of the chromatin [27].

It is important to point out that chromatin's role is not only structural, but it is an important character in gene expression. Indeed, if we examine the nucleus of some unicellular, eukaryote, organism (e.g. yeast), we will note that that it is almost homogeneously filled by the chromatin. On the other hand, in the nucleus of multicellular organisms, coexistence of regions with various density of chromatin is observed. This difference is due to the fact that while the genome in yeast is almost always all actively transcribing proteins, in a cell of a tissue of a multicellular organism some part of the genome have to be silenced as they must produce only a certain subset of all the possible proteins. Indeed, the density of chromatin is strongly correlated with the transcription activities of the genome in that area: denser and less accessible areas (called *euchromatin*) are silenced; less dense and more accessible areas (called *heterochromatin*) are, instead, actively transcribing.

Finally, in the last years, thanks to the development of new techniques for the study of chromosome conformations (e.g. Hi-c [29]), there has been a leap in the understanding of finer details of chromatin conformation. One of the most interesting finding is that chromatin organizes in smaller self-interacting regions (see fig. 1.3) called TADs (topologically-associated domains) [30, 31]. TADs can be thought as architectural chromatin units that define regulatory landscapes. As we already pointed out, chromatin organization is strictly correlated to gene expressions, and therefore TADs seem to play a central role in it. Indeed, it seems that TAD boundaries correspond to those of replication domains [32], and genes tend expressed (or repressed together) during cell differentiation when they are located in the same TAD.

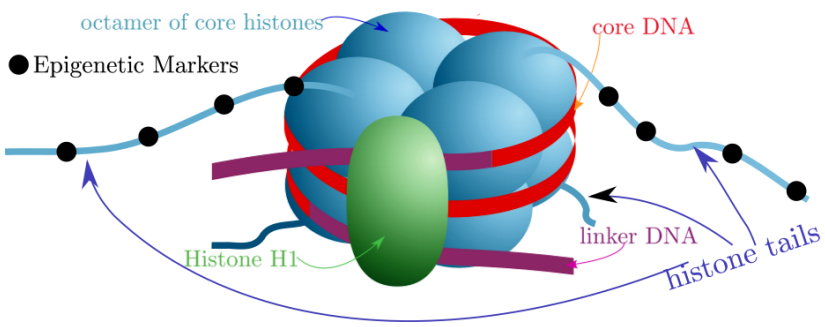



**Figure 1.2:** Scheme of nucleosome organizations. There are five types of histones: H1, H2A, H2B, H3, H4; the histones H2A,H2B, H3, H4 are called core histones. DNA wraps around eight core histones (two per type) and it is kept in this position by an H1 histone; the complex histone-DNA is often called NCP (nucleosome core particle). An NCP is linked to another one by a strand of DNA called linker DNA. The complex NCP-linker DNA is called nucleosome. Core histones, share the feature of having a long "tails" to which epigenetic markers can attach. Adapted from [28].

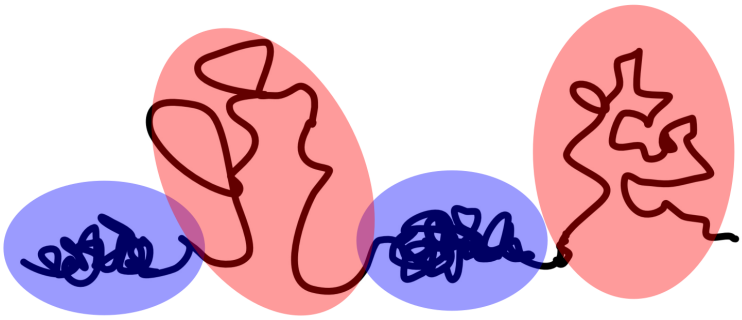


**Figure 1.3:** Pictorial Scheme of TADs organization in chromatin. You can observe the sequence of distinct sequence of domains of different shape and form. Namely they can be, for example, compact (highlighted in blue) and associated to silenced genome (heterochromatin), or swollen (highlighted in red) and associated to actively transcribing genome (euchromatin).

### 1.1.3 Epigenetics

As we briefly discussed in the introduction of this chapter, during the epigenesis a cell pass from being totipotent to having a specialized, and in some way restricted, function compatible with its fate. Therefore, the development and the determination of the cell type involves restrictions in its potency and result from a different gene expression. The main way to regulate gene expression, is via gene silencing, i.e. preventing certain genes from being over-expressed. General understanding is that gene silencing is achieved through chromatin condensation (*heterochromatinization*).

These kind of processes are well studied in the case of the eukaryote cell, but are not well known in the prokaryote case (see for example [33]); from now on we will discuss only about the epigenetic dynamics of the eukaryote cell.

One of the most important mechanisms, found both in prokaryotes and eukaryotes, that plays a fundamental role in heterochromatinization is the so-called DNA-methylation, that is the process that adds a methyl group to certain bases of the DNA. Methylated DNA can regulate gene expression by preventing the binding of transcriptional proteins to the gene, while at the same time it may be bound with a certain class of architectural protein called MBD. MBD proteins, itself, recruit chromatin-remodeling proteins that can modify histones, and therefore forming a bundle of compact, inactive, chromatin.

### Histone Modifications

A different class of epigenetic processes who do not affect directly the DNA molecule, is called "histone modifications". The core histones present a long 'tail' on one end of their amino acid structure, where several molecules, called epigenetic marking can attach as a form of post-translational modification (see fig. 1.2), including methylation, acetilation, phosphorylation, sumoylation, ubiquitylation, and ADP-ribosylation [34]. As several epigenetic markings can bind to an histone tail at the same time, a huge number of combinations is possible. However, not all of these combinations seem to be possible, and only a few of them seem to be relevant. For example, these relevant combinations can even be reduced to 5 classes in the case of Drosophila, the so called "five colors of chromatin" [35], while in humans it has been estimated that over 50 different chromatin states occur [36].

To deal with the great complexity of chromatin modifications, organisms have developed a wide variety of "helper" molecules (proteins, enzymes) to regulate them. Oversimplifying the problem, we can generally group these enzymes in three big groups "writers", "erasers", and "readers". Writers are enzymes that have evolved to modify specific amino-acids of the histones, or nucleotide bases of the DNA molecule. "Erasers" are the enzymes involved in removing such modifications. Finally, "readers" are proteins that bind specific chromatin modifications, providing the ability to the cell to sense the chromatin state of a given locus.

Multiple possible modifications can occur in nucleosomes; we will focus on two of the most common ones: histone methylation, and histone acetilation. Methylation and acetilation of chromatin components can serve a variety of functions, with both positive and negative effects on transcription, combining to ensure the desired gene expression pattern for any particular cell. For example, methylation of promoters with H3K9me3 on the histone H3K9 occurs at inactive genes [37], while H4K20me3 (on histone H4K20) has been linked to constitutive heterochromatin, e.g. actively transcribing genome [38]. Unlike other modifications, methylation does not change the overall charge, although it does alter the hydrophobic character and size of the modified residue. On the other hand, acetilation of histones results in the neutralization of its

positive charge, and leads to a decrease in association with the negatively charged DNA backbone. This reduces chromatin compaction, rendering the modified locus more accessible to transcription factors and thereby increasing gene expression [39].

### Dynamics of Post-Translational Modifications

An interesting biophysical mechanism proposed in the last years, deals with the *spreading* of epigenomic modifications of the nucleosomes [40]. Here, the epigenetic modifications are not a static feature of the nucleosome, but rather they are dynamical, and self-propagating. For example, in yeast, the spreading of silencing marks is observed via the formation of Sir-complexes [41]. The idea is that a histone bearing the mark H4K16ac (acetilation on histone H4) attract a protein Sir2p or Sir3p (readers); when bound the Sir proteins attract enzymes (writers) that deacetylate the neighbouring nucleosome, and thereby increasing the affinity for those nucleosome to recruit additional Sir complexes. Notably, more example of similar mechanisms have been observed on other readers relative to other nucleosome modifications: these include SU(VAR) [42], and MLL [43]. Interestingly, this positive feedback loop is self-reinforcing, and it is a candidate to explain how nucleosome modifications quickly reassert upon mitosis in the daughter cells. For this reason, the postulate of the existence of an "epigenomic spreading" based on a readers-writers positive-feedback loop, is usually employed as a fundamental starting point of most biophysical models.

Finally, we introduce here a last key-component of the gene-expression mechanism, and which seem to play a fundamental role in epigenetic spreading. Transcription factors (TFs) are proteins possessing domains that can bind directly to to the DNA molecule, more specifically to certain regions called "promoters" or "enhancers" of a specific gene. TFs can be both "activators" and boost the expression of a gene when bound to the promoter region, or can be "repressor" and silence it. Indeed, they possess a domain that interacts directly with some transcription factors, and consequently regulates the amount of messenger RNA (mRNA) produced by the gene. Now, notice that the dynamics of epigenomic spreading just discussed, cannot be completely random, but rather be guided from an external factor in order to achieve the desired gene expression. The role of guide, in this context, seem to be assume from transcription factors themselves. Indeed, recent studies show that the epigenomic configuration can be somehow predicted from knowing the TFs density profile [44] (see also chapter 2).

## 1.2 Modeling Chromatin as a Polymer

After the brief introduction to the biology of the epigenetic problem, we start by discussing the possibility of studying epigenetics in terms of physical processes. One approach offered by statistical physics is to consider an ensemble of configurations of the chromatin found in different experiments [45]. Statistical properties of such

an ensemble can tell us important information about the principles governing DNA packaging. The general aim will be to develop a simple physical model which reproduce an ensemble of configurations resembling the ones found in the experiments.

The possibility of studying the dynamics of the chromatin via simple polymer models has been widely investigated [46]. Indeed, this approach has the great perk of being simple enough to be studied via analytic, or numerical approaches, and, yet powerful enough to correctly estimate the chromatin structure at large scales, without regard of its fine structure [47].

As we discussed in the precious section, chromatin can be visualized as a succession of nucleosomes complex. So the first great temptation one can have is to disregard the peculiarities of each single nucleosome (DNA bases, epigenetic marks attached to the histone tails, etc.), and to consider the chromatin as a homopolymer.This is not generally true. Indeed, as we already noted, in the nucleus of multicellular organisms we observe the coexistence of regions with higher chromatin density and regions with lower chromatin density, which means that at least two "types" of chromatin fiber must be considered at the same time (euchromatin and heterochromatin). Nonetheless, in some unicellular organisms, e.g. yeast, chromatin is observed to (almost) homogeneously occupy all the space in the nucleus and its dynamics can be explained with a simple homopolymer model.

Regardless of the choice of model, some physical parameters must be extracted from direct experiments. At physiological salt concentrations, the chromatin structure has a diameter of $\sim 10\,\text{nm}$, with a linear density of around $0.5 - 0.6$ nucleosomes per nm of chromatin fiber, corresponding to approximately 110 bp (base pairs) per nm [48].

Unfortunately, the stiffness of the polymer or its persistence length $L_p$ is hardly accessible by experiments. Depending on the experimental procedure different results have been measured, ranging from $L_p = 30\,\text{nm}$ to $L_p = 200\,\text{nm}$ [49][50]. As modelization goes by, however, one must consider the steric interactions between nucleosomes and possible occupancy of linkers by other DNA-binding proteins. This fact leads the fiber to be less flexible, and suggests that $L_p$ consists of about five or six nucleosomes [51]. This also mean that each "bead" of our polymer will not be a single nucleosome, but it will be big enough to represent a few neighboring nucleosomes.

### 1.2.1 Chromatin as a Homopolymer

In the simple case of yeast, chromosomes are shorter than in multicellular organisms, and attached to the centrosome. Moreover, the whole genome is actively transcribing most of the time, with the exceptions of the regions that govern the cell sexual behavior, activated only during mitosis, and of the chromosome extremities, called telomeres. As the chromosomes are almost completely composed of euchromatin, they can be found in swollen coil configurations, and occupy most of the cell nucleus.

Fluorescent microscopy techniques have been used to study the dynamical behaviour of chromosomal loci, with a special interest to their mean square displacement.

Analysis have proved that the loci show anomalous diffusion [52] compatible with a Rouse Model [53], where the polymer dynamics is almost completely dominated by elastic interactions between first-neighbours beads [46].

Thanks to this study, we can conclude that the highly transcribed DNA in the nucleolus, yeast chromosomes behave as a polymer brush, and are essentially organized by simple physical principles [54].

### 1.2.2 Block-Copolymer model for Chromatin

Most of the model that consider chromatin as an homopolymer, cannot take in account the great variety of chromatin organization observed in more complex organisms. Namely, they fail to describe the coexistence of euchromatin and heterochromatin, each characterised by a different spatial density. In order to correctly depict the chromatin picture, one must delve in biology and introduce epigenomic in the model.

In this section, we introduce a new model proposed by [13] where the chromatin folding is driven by effective epigenomic dependent interactions between chromatin loci. Namely, rather than considering chromatin as a homopolymer, here chromatin is a heteropolymer where each polymer bead type correspond to a different epigenetic mark. Moreover, given that usually that epigenetic marks organize in domains of various sizes, we can simplify the model and restrict the analysis to block copolymers, that is heteropolymers composed by successive blocks of different monomers.

We model the chromatin fiber as an interacting self-avoiding bead and spring chain, where each bead represents a certain portion of the genome (usually composed of more nucleosomes). Beads are characterised by an epigenetic mark that here is represented by their color (see fig. 1.4), and do interact with each other. In addition to the standard excluded volume interaction modeled via Lennard-Jones potentials, we consider a short-range attractive interaction $E_{mn}$ between the monomer $m$ and $n$. Generally, the interaction depends on all the possible combinations of states the two monomers can have; here, however, we consider only two type of interactions:

1) non-specific interactions $U_{ns}$ which are effective interactions deriving from the compaction effects due to confinement into the nucleus;

2) specific interactions $U_s$ between monomers having the same epigenetic mark.

Hence $E_{mn} = U_{ns} + \delta_{mn} U_s$, where $\delta_{mn} = 1$ if the monomers $m$ and $n$ have the same epigenetic state, $\delta_{mn} = 0$ otherwise.

This modelization is motivated by many experimental evidence suggesting effective interactions between loci of identical chromatin state. Indeed, it has been shown that Polycomb group or HP1 proteins may create physical bridges between distant heterochromatin regions. Note, however, that for simplicity we are assuming that all these kind of processes give rise to the same interaction strength $U_s$.

The model has been studied both using numerical simulations and analytical tools thanks to Gaussian approximations in the Fokker-Planck equation [55], varying the

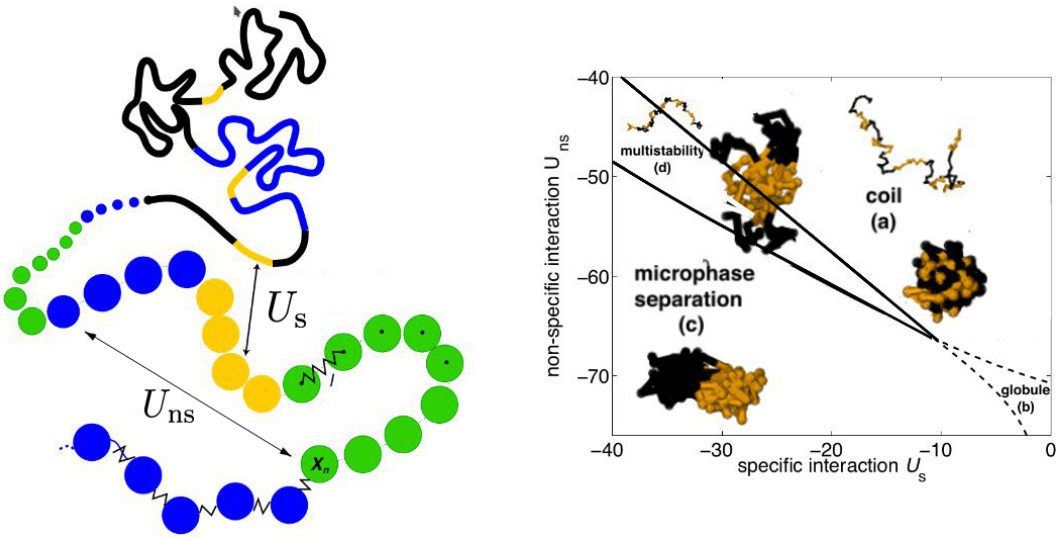



**Figure 1.4:** (a) Block copolymer model: the chromatin is modeled as a self- avoiding bead-spring chain where each monomer represents a portion of DNA and different colors represent different epigenetic states. The model integrates non-specific and specific short-range interactions to account respectively for the effective compaction of the chain and for epigenomically related affinities between monomers. (b) Phase diagram of the model, with snapshots of polymers in a configuration characteristic of the corresponding region (see text for details). Adapted from [13].

values of $U_s$ and $U_{ns}$ [13, 56]. For simplicity a block copolymer with an alternation of active (A) and inactive (B) epigenetic domains of the same size is considered. The resulting phase diagram (see fig. 1.4) prove the existence of four different equilibrium phases. For weak compaction $U_{ns}$ and specificity $U_s$, the system is in a coil phase with swollen chain conformation. If we increase $U_{ns}$, but keep the specifity low, the coil will collapse in a globule with no evident presence of epigenetic domains. On the other hand, if we keep both strong $U_{ns}$ and $U_s$ the system will present "micro phase separation" (MPS) and organize in a globule where all monomers of the same epigenetic state are densely packed into two distinct 3-dimensional domains. Finally between the coil region, and the MPS region, a multistability region lies at high $U_s$ and low $U_{ns}$; here multiple possible configurations are stable and are generally half way between the coil and the MPS configurations.

In the multistability region found in the phase diagram lies the main strength of the model. Here, indeed, one observes that epigenomic domains fold into topologically associated domains and interact transiently with each other. The specifity of interactions depends greatly on the properties of the domain itself; for example, small domains like most of the epigenomically active domains, show fast and short-lived interaction, while bigger domains may form long-lived metastable interactions. Direct comparison with experimental Hi-C maps of Drosophila suggest that biological sit-

uations are consistent with multistability. This implies that in the cell, chromatin organization is being dynamically and stochastically remodeled while conserving local key features of its topology.

## 1.3 Epigenomic-Spreading Models

In the previous section, we discussed some physical models which given an epigenetic landscape as an input, they return, as an output, the 3-dimensional configuration of the chromatin. These model clearly describe only half of the picture, as they treat epigenetic markings as stable. This assumption is generally false (see section 1.1.3) and it can be regarded as a valid approximation only when chromatin is, indeed, stationary.

As already discussed, gene expression, regulated by epigenetic modifications to the chromatin, undergoes dramatic changes during the cell lifetime, especially during its development. In fact, gene expression is stably propagated during the cell division (mitosis) and, sometimes, even during meiosis. We conclude that, in eukaryotes, cell replication involves both the synthesis of DNA, and chromatin assembly. At the same time when the DNA molecule is replicated, also the nucleosomes of the mother-cell chromatin will be distributed among the two daughter cells. Finally, as the daughter cells only will present half of the epigenetic markings of the mother, they will need to spread in order to reproduce faithfully the original nucleosome organization Unfortunately, the detailed mechanisms which regulate these kind of processes is not still clear [57].

In the next section we will discuss, some mathematical, and physical, models that try to understand and reproduce the dynamics of epigenetic spreading. Such models must present three important features in order to really illustrate the biological picture:

1) The model must present multistability. Indeed,while generally only an epigenomic configuration is observed during in a single cell life, looking at different cells of different tissues, they will obviously have different epigenomic organizations. In the simplest models, patterns of epigenetic marks allow to switch between several states that have a well-defined functional characterization.

2) The stationary states of the model must be *stable* against perturbations. Although some drastic events (like mitosis) may change gene expression, it generally remains the same for all the cell life and resists against thermal, biochemical and mechanical perturbations.

3) The model must reproduce in some way the heritability of the epigenome through different cell generations.

### 1.3.1 Zero-Dimensional Models

As a first step, we introduce the so-called zero-dimensional models where the nucleosome are imagined to be localized in a single point and therefore there is no concept of nucleosome distance or 3-d organization. In these models the identity of each nucleosome is lost, and the only variables that are relevant are the relative abundances $n_k = N_k/N$ of the epigenetic state $k$, where $N_k$ is the number of nucleosomes in the state$k$, and $N$ is the total number of nucleosomes. Note, however, that in treating these variables we must not forget that the total number of nucleosomes is conserved and therefore we have the constraint:

$$\sum_k n_k = 1 \ . \tag{1.1}$$

In this context, the spreading of epigenetic information is described by the time evolution of the densities $n_k$, while the mitosis can be modeled by an instantaneous process in which all the epigenetic species are diluted (e.g. halved) between the two daughter cells. Time evolution is described by a set of differential equations similar to the ones employed to describe chemical reactions; the time evolution of the abundance of the epigenomic specie $j$ is therefore:

$$\frac{\mathrm{d}}{\mathrm{d}t} n_j = \sum_{j \neq k} K^{(jk)} n_k - \sum_{j \neq k} K^{(kj)} n_j \ , \tag{1.2}$$

where $K^{(jk)}$ is is the rate of transition of nucleosomes from the state $k$ to the state $j$. The kinetic parameters $K$ are not generally constants, but rather depend on some other dynamical parameter, or on the abundances themselves. Note that $\sum_j \frac{\mathrm{d}}{\mathrm{d}t} n_j = 0$ and therefore equations (1.2) respect the constraint given by (1.1).

The properties of this class of models depend greatly on the number of epigenetic states, and on the parameterization of the chemical rates $K$.

#### Two-State Model.

We examine here the simplest possible epigenetic model that obeys equations in the form (1.2) [58]. In this model only two states are possible; the nucleosome can either be in a modified state $M$ or in an "anti-modified" one $A$ which means that the nucleosome has no epigenetic marking, or rather has a different one than $M$.

The epigenetic dynamics is vastly simplified, and it is inspired by the well known Voter model. Namely, we assume here that each nucleosome has a certain probability $1 - \alpha$ of spontaneously changing their epigenetic state (from $M$ to $U$ or vice versa), and that with a probability $\alpha$ two nucleosomes with the same mark will recruit a third and change its state to be the same as theirs. All this can be easily modeled employing ad hoc values for the kinetic constants:

$$\begin{aligned} K^{(MA)} &= \alpha n_M^2 + (1 - \alpha) \\ K^{(AM)} &= \alpha n_A^2 + (1 - \alpha) \end{aligned} \ . \tag{1.3}$$

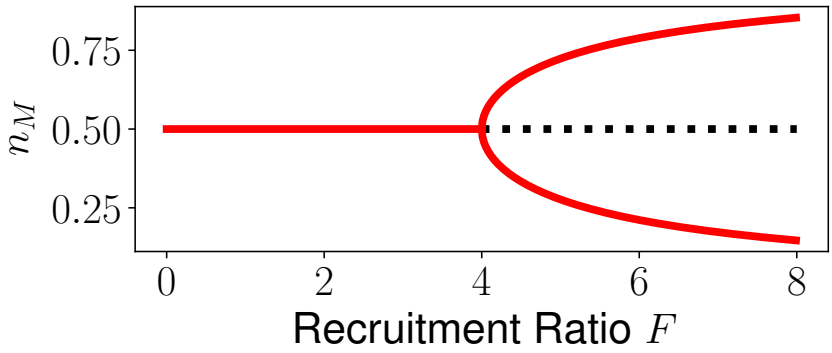



**Figure 1.5:** Steady states of differential equation (1.4); stable states are indicated via a red solid line, while unstable states are represented via a dotted black line. Note that at $F = 4$ the system has a bifurcation typical of bistable system, and indicating the presence of a second-order phase transition.

Now employing the constraint (1.1)($n_M + n_A = 1$) and substituting in (1.2), we obtain the differential equation:

$$\frac{\mathrm{d}}{\mathrm{d}t} n_M = \alpha \left(2n_M - 1\right) \left[n_M^2 - n_M + \frac{1}{F}\right] \quad , \tag{1.4}$$

where $F = \frac{\alpha}{1-\alpha}$ is the ratio of recruitment.

The analysis of the stationary point is easily done by setting equation (1.4) to zero and solving for $n_m$ (see also fig. 1.5). Results show that only one stationary point $n_M = 1/2$ exist if $F \leq 4$, while three are present if $F > 4$ which are $n_M = 1/2$ and $n_M = \frac{1}{2} \pm \frac{1}{2}\sqrt{\frac{F-4}{F}}$. Stability of the stationary point is deducted studying the sign of the derivative of the second term in (1.4). This study reveals that:

i. if $F \leq 4$, then the only stable stationary point is $n_M = \frac{1}{2}$, which indicates that the system do not show any form of coherence as half chromatin will be in a modified state $M$, and the other half in the anti-modified state $A$.

ii. if $F > 4$, instead, the incoherent state $n_M = \frac{1}{2}$ is unstable, while the system present bistability in the stationary points $n_M = \frac{1}{2} \pm \frac{1}{2}\sqrt{\frac{F-4}{F}}$. Note that in these states, we can observe an epigenetic order as one of the two states ($M$ or $A$) will prevail on the other.

We conclude that, while this model is extremely simple, it reproduces the multistability required by a proper modelization of the epigenetic processes. Could we have developed a simpler model? The answer is no, because if we choose a simpler form of the kinetic parameters (1.3) suppressing the cooperativity quadratic term, we will have a class of models that do not show our desired bistability.

**Three-State Models.**

We now engage a slightly more complex class of models, where histones can be found in three states, namely: Modified histones $M$, anti-modified histones $A$, unmarked histones $U$ [16].

In the previous two-state model, a single histone could "transform" directly its state into the other. However, if we have a modified histone, in order to become anti-modified, as a first step it has to lose all of its marks and become unmarked. With this assumption in mind, we will assume that the reaction $A \leftrightarrows M$ is impossible, and therefore $K^{(AM)} = K^{(MA)} = 0$. Moreover, we assume that the recruitment of modifying enzymes by surrounding modified or anti-modified nucleosomes occurs at a rate $\epsilon_X n_X$, and that random transitions between states occur at a rate $\gamma_X$.

Here, we briefly discuss the version of this model proposed by [59], where we assume that:

$$\begin{aligned} K^{(MU)} &= K^{(UA)} = \epsilon_M n_M + \gamma \\ K^{(AU)} &= K^{(UM)} = \epsilon_A n_A + \gamma \end{aligned} . \tag{1.5}$$

Recalling that $n_A + n_M + n_U = 1$ from (1.1), we will have the following differential equations:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} n_A &= (\gamma + \epsilon_A n_A)(1 - n_A - n_M) - (\gamma + \epsilon_M n_M) n_A \\ \frac{\mathrm{d}}{\mathrm{d}t} n_M &= (\gamma + \epsilon_M n_M)(1 - n_A - n_M) - (\gamma + \epsilon_A n_A) n_M \end{aligned} . \tag{1.6}$$

As a first step we examine the case of symmetric dynamics, where no epigenetic mark is favoured and $\epsilon_A = \epsilon_M \equiv \epsilon$. In this symmetric case the model present a clear analogy with a mean-field three-states Ising model, where each nucleosome mark correspond to a spin state $S$; namely the $M$ and $A$ marks correspond with the "active spins" $S = \pm 1$, while the unmarked histone can be identified with a null spin $S = 0$. In this context $\epsilon$ will correspond to the spin coupling constant, while $\gamma$ will be associated to the thermal noise of the system. We can also introduce in our system an epigenetic magnetisation defined as $m = n_M - n_A$.

If the analogy with the Ising model holds true, we expect to find a continuous phase transition in the epigenetic magnetisation $m$. This is indeed what we observe; namely, using the same method as in the previous section we observe the presence of three fixed points of the dynamics:

(i) $m = 0$ is always a fixed point, but it is stable only if $\epsilon \leq 3\gamma$.

(ii) $m = \pm(\gamma/\epsilon)\sqrt{(\epsilon/\gamma + 1)(\epsilon/\gamma - 3)}$ are stable fixed points of the dynamics if $\epsilon > 3\gamma$.

We find that this system undergoes a second order phase transition from an epigenetically disordered state with null magnetisation, to an ordered one characterised by bistability.

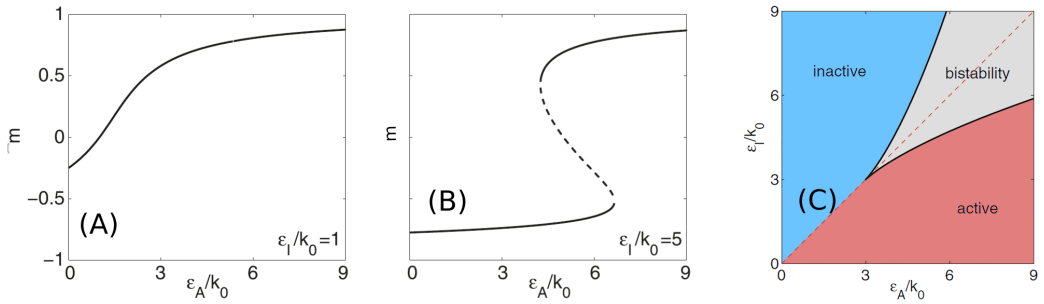



**Figure 1.6:** **(A-B)** Magnetisation as a function of $\epsilon_A$ keeping $\epsilon_M$ fixed. In (A) the system pass continuously from the active to the inactive region, while in (B), when $\epsilon_M$ is big enough, we observe the presence of a cusp catastrophe where the system bistable and alternates between the active, and inactive region. **(C)** Phase Diagram of the model as a function of $\epsilon_A$ and $\epsilon_M$. Adapted from [13].

Even more interesting results can be found in the case where $\epsilon_A \neq \epsilon_M$ (see fig. 1.6). For every pair of parameters $(\epsilon_A/\gamma, \epsilon_M/\gamma)$, the dynamical system is either monostable or bistable (with an unstable fixed point). . Depending on the relative asymmetry between $\epsilon_A$ and $\epsilon_M$, the single (stable) fixed point of the monostable region corresponds to a modified or anti-modified epigenetic state. Bistability, with the coexistence of an active and an inactive coherent activity, is observed only for strong recruitment and small asymmetry. Interestingly, in the bistability region the system becomes ultrasensitive and weak asymmetries lead to important changes in the epigenetic state; this is the so-called cusp catastrophe [60] and it is typical of dynamical systems with asymmetry.

To understand the implications of this model, we note that differentiated cells exhibit a robust phenotype within the population and in time. The analysis of the dynamical models suggest that this is the case also in this model, as long the dynamical variables are well off in the monostability region and therefore stable against small perturbations in the recruitment rates $\epsilon_A, \epsilon_M$. At the same time, however, when modifications of the environment are important, cells and organisms can adapt by changing drastically their gene expression (e.g plant vernalization [61]), which is exactly the behaviour predicted by the system via the cusp catastrophe when strong asymmetries arise.

### 1.3.2 1-Dimensional Models of Epigenetic Spreading

The study of simple zero-dimensional (mean-field) models is often very useful as they can highlight the basic features of a complex dynamical system. However, if we want to understand the details of the spatial configurations showed by the dynamics, this kind of models are clearly not sufficient. In the case of epigenetics, active/silenced nucleosomes are observed to organize in "patches" with coherent epigenetic states, and

therefore a precise description of this spatial dynamics it is necessary. In this context, we introduce a model proposed in [17] which extend the ideas of the zero-dimensional models into a 1-dimensional model which allows large chromatin regions to exist stably and heritable in distinct expression states.

We recall that one of the most important feature we require from our epigenomic model is multistability; that is the ability for a dynamical system to have multiple stable stationary states under the same environmental conditions. One of the main lessons that we take from the study of zero-dimensional models is that in order to achieve the bistability typical of an epigenetic system, we need special, ad hoc, mechanisms. First requirement we make in order to have bistability is the existence of some positive feedback loop, that is a process that interacts in a cyclic way and that intensify the effects of a small disturbances which result in the increase in the magnitude of the perturbation. A positive feedback alone, however, is not sufficient to ensure bistability. A bistable system must also possess some type of non-linearity within the circuit, such that when a perturbation occur the intensity increases greatly and one epigenetic mark is vastly advantaged over the other ("ultrasensitivity") [62].

In practice, ultrasensitivity can be induced in the model, introducing a cooperation process between nucleosomes with the same mark. Cooperation can be either direct, e.g. two nucleosomes cooperating to recruit a third one [58], or indirect where each modified nucleosome catalyzes one of two separate modification reactions to fully convert a third nucleosome [16] (see section 1.3.1). More specifically, the cooperation between the nucleosome must act non-locally, inducing modifications of the marks of nucleosomes located far away on the DNA backbone. Thanks to this long-range interaction, any nucleosome is able to "sense" the majority nucleosome type within the patch, while on the other hand this would not be possible employing only short range interactions.

### The Model

We assume that nucleosomes can have only two epigenetic states: active (A) or silenced (S). The system is treated as a string of $N$ nucleosomes, and its evolution over time occurs through stochastic iteration of individual modification reactions. At each time-step, each reaction will be made with a chosen fixed probability. These probabilities allow adjustment of the strength of each reaction. In the model three kind of reactions are possible:

1) *Recruitment* reactions which ensure the existence of a positive feedback loop. The recruitment process depend strongly on the other nucleosomes present in the system. In the model cooperation is added assuming that two separate nucleosomes of the same state are required to recruit the enzymes needed to convert another nucleosome. Note, however, while some experiments suggest the existence of such a mechanism for the conversion A→S [63], no evidence has been found in the opposite case S→A.

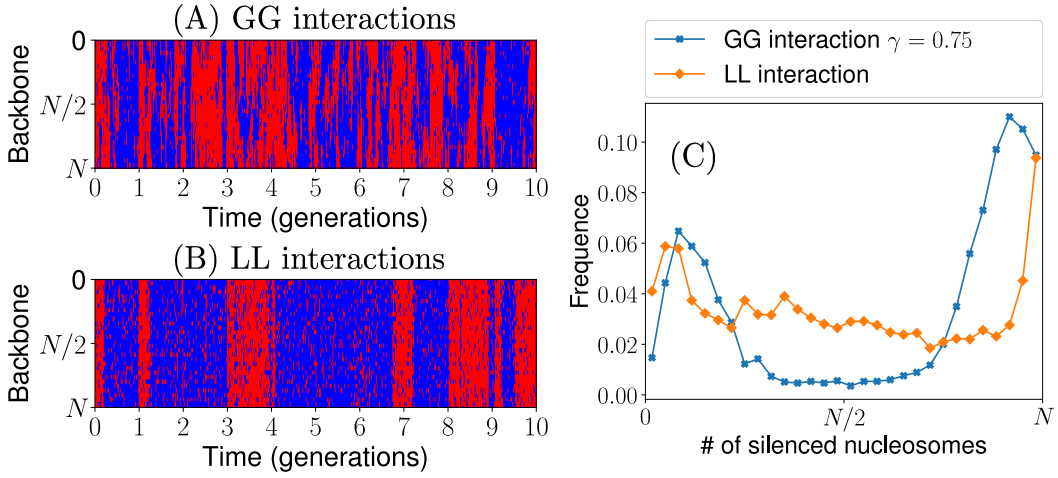



**Figure 1.7:** Comparison of 1-dimensional epigenetic spreading when only local interactions (LL) are present, or only global interactions are present (GG). We employ $N = 30$; $\lambda_1 = \lambda_2 = 0.37$; $\alpha_1 = \alpha_2 = 0.04$; $t_R = 10^3$. (a-b) Kymographs showing the space–time plots of an epigenomic region; blue represent silenced states, red are active states. (c) Histogram of the number silenced nucleosomes during the dynamics. Note that the GG dynamics show strong bistability between a state where all nucleosomes are active, and a state where all nucleosomes are silenced, while in the LL dynamics several intermediate cases are observed.

2) *Noise* which is reproduced in the model via random reaction that transform an active state in a silenced one, or vice versa

3) *DNA replication* is simulated by giving to each nucleosome a probability 0.5 of transforming into the active state A. This is to reproduce that each daughter strand will receive only half of the parent nucleosome marks.

At each time-step of the dynamics the recruitment process will be attempted with probability $\lambda_1$ for the reaction $S \to A$, and probability $\lambda_2$ for the reaction $A \to S$. In the same way, noise reactions happen with probabilities $\alpha_1$ for $S \to A$ and $\alpha_2$ for $S \to A$. DNA replication happen every $t_R$ timesteps.

The last detail that remains to be discussed is how recruitment is modeled. Among the various possible modelization of the process, we consider only local recruitment (LL) and global recruitment (GG), for more details see [17]. In local recruitment, two adjacent nucleosomes of the same mark will recruit a third adjacent nucleosome. In global recruitment, instead, two nucleosome $k_1$ and $k_2$ are selected such that their reciprocal distance $|k_1 - k_2|$ belong to a random exponential distribution with rate $\gamma$; if they have the same mark they will recruit a third nucleosome chosen again via the exponential distribution. In fig. 1.7 we compare the dynamics of two nucleosome systems, one with only local interactions, the other only with global interactions. As already discussed, no bistability can be observed if only local interactions are

employed, but if we introduce even a weak global interactions we observe that the system show avoidance of intermediate states, typical of bistable systems.

## 1.4 A Chromatin Folding Model for Epigenomic Spreading

In this section we will introduce a model first presented in [1], and that will be the foundation for every other model we will discuss from now on this thesis. The need for this model arises from what we discussed in the previous sections. Models presented in section 1.2, for example, focus in understanding the structure and organization of chromatin in the nucleus, but they do so by starting from a fixed, or an ensemble of epigenetic landscape, but they do not attempt to explain the underlying mechanisms regulating the dynamic of epigenomic itself. Models in section 1.3.1, instead, do the exact opposite by disregarding completely the concept of spatial organization and develop mean-field models to explain the spreading of epigenetic marks in the chromatin [16, 58]. Note, also, that the 1-dimensional model proposed by [17] tries to explain the existence of spatial structures in the epigenetic landscape. However, this model lacks a complete discussion of the 3-dimensional chromatin folding, while at the same time it introduces a long-range interaction between nucleosomes that cannot be sufficiently justified at this level.

All the models examined until now are therefore "lacking" as they do not challenge the well-established intimate relationship between chromatin architecture and gene expression (see section 1.1). For example, regions of the genome with repressive epigenetic modifications are usually correlated with more compact chromatin configurations (heterochromatin) [9, 17], or, on the other hand, regions actively transcribing usually present in more loose configurations (euchromatin) [64].

In light of this, we discuss here a polymer model with epigenetic switches, inspired by magnetic polymer models (see for example [19]), and that directly couples the 3D dynamics of chromatin folding to the 1D dynamics of epigenetics spreading.

In order to develop this model, we look at the details of epigenetic spreading in the cell. As already discussed, the main players in this process are the so-called "readers" and "writers" (see also section 1.1). "Writers" are the enzymes that are responsible for the establishment and propagation of a single specific epigenetic mark. "Readers", instead, are multivalent proteins that identify and bind to the epigenetic modifications made by the writers, bridging chromatin segments bearing the same histone marks. There is evidence that readers of a given mark recruit writers of that same mark [65, 66]. For example, regions actively transcribed by RNA-polymerase are found to be rich in active epigenetic marks; in this case the RNA-polymerase act as a reader protein that seem to recruit writers enzyme of the classes Set1 and Set2 [67]. Note that the writer-reader mechanism, as described, creates a positive-feedback loop (see fig. 1.8) which can lead, at least in principle, to a bistable dynamic system that can sustain epigenetic memory [17, 68] during successive cell generations (see

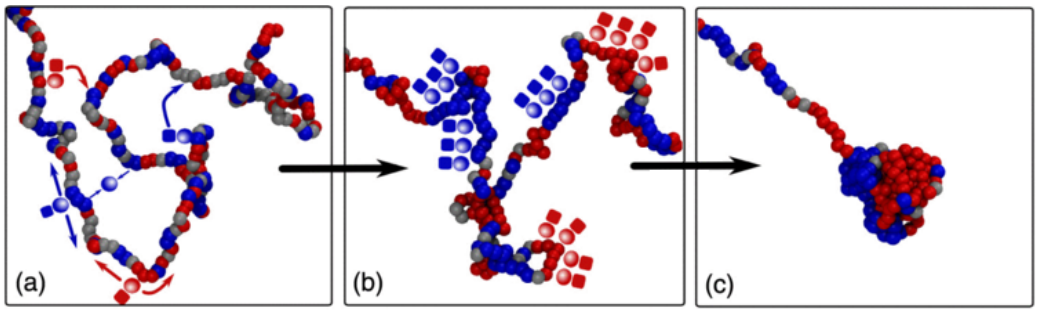



**Figure 1.8:** Representation of the epigenetic-spreading process induced by the readers-writers positive feedback loop. In this model, the chromatin polymer present only two possible histone modification identified by their color. **(a)**"Readers", here represented by shaded spheres, bind to specific histone modifications and bridge between similarly marked segments (distinguished here via their "color"). Histone-modifying enzymes, or writers (solid squares), are assumed here to be chaperoned by the bridge proteins. The positive-feedback mechanism and competition between different epigenetic marks results in a regulated spreading of the modifications **(b)**, which, in turn, drives the overall folding of the polymer **(c)**. Adapted from [1].

section 1.3.2).

While the model could be in principle studied with any number of epigenetic states, we initially focus to the simplest case of only two epigenetic states that symmetrically compete with each other (e.g., corresponding to "active" or "inactive" chromatin). The model predicts a phase transition between a swollen, epigenetically disordered phase, and a collapsed epigenetically coherent one with bistability. Interestingly, even in this simple situation, the model is different from the previous ones (see section 1.3) because the phase transition seem to be of the first order, rather than a continuous one. Indeed, the first-order nature of the phase transition arises thanks to the strong coupling between a 3-dimensional dynamics (chromatin folding) and a 1-dimensional dynamics (epigenetic spreading). In this fact lies one of the main strengths of the model: first order phase transitions are characterized by hysteresis cycles when quenching, and therefore the model permit a bistable epigenetic switch that retains memory.

### 1.4.1 The Model

In this section we will discuss the details of the model proposed in [1]. As already discussed, the model studies at the same the 3-dimensional chromatin architecture, and the dynamic of the epigenetic landscape. For this reason, we will generally consider two concurring dynamics:

(i) *Langevin dynamics*, which reproduces the spatial dynamics of the chromatin fiber;

(ii) *Recolouring dynamics*, which involve epigenetic marks and their spreading along the chromatin backbone.

Chromatin fiber is modelled by using a semi-flexible bead and spring chain of $N$ beads of size $\sigma$. As already discussed in section 1.2, beads do not represent a single nucleosome, but rather a few neighboring ones; namely, we consider $\sigma \approx 30\,\mathrm{nm}$, corresponding to approximately 15 nucleosomes [69]. Generalizing the work of [13], we consider the chromatin to be a copolymer, but not necessarily a block-copolymer. For simplicity, we assume that beads can be found in two possible epigenetic states: *active* (A) or *silenced* (S), or they can present no clear epigenetic state and are left *unmarked* (U). The beads interact with each other via a classical hard-core steric interact, or, if they have the same epigenetic mark, they can experience a short-range attractive interaction in order to mimic the "reader-writer" architecture we discussed in the precious section.

Finally, the epigenetic landscape itself can evolve and greatly influence the spatial dynamics of the polymer. Epigenetic states of the beads, usually represented by colours, evolve using a Potts-like dynamics that induces cooperative recruitment, and therefore, possibly, a bistable system.

### The Langevin Dynamics

Langevin dynamics deals with the time evolution of the spatial configurations of the chromatin fiber. The chain is composed of $N$ beads of size $\sigma$ and epigenetic marks $\{q_i\}_{i=1\dots N}$. Its time evolution is described by a set of $3N$ Langevin equations (hence the name "Langevin" dynamics) derived from an Hamiltonian $H_L$ at a temperature $T$, in the form:

$$m\ddot{\mathbf{r}}_i = -\gamma\dot{\mathbf{r}}_i - \partial_{\mathbf{r}_i} H_L\left(\{\,\mathbf{r}\,\},\{\,q\,\}\right) + \sqrt{2k_BT}\xi_i(t) \ , \tag{1.7}$$

where $m$ is the beads mass; $\gamma$ is a friction coefficient; $\xi_i$ is a classic Gaussian white noise such that $\langle\xi_i(t)\rangle = 0$, and $\langle\xi_i(t)\xi_j(t')\rangle = \delta_{ij}\delta(t-t')$ for all values of $i,j$ and for every time $t,t'$.

The Hamiltonian of the chromatin fiber $H_L$ is modeled as a classic Hamiltonian of a bead-spring chain, to which we add a flexibility term, and a bead-bead interaction term:

$$H_L = K + \mathrm{U_{KP}}\left(\{\,\mathbf{r}\,\}\right) + \mathrm{U_{HARM}}\left(\{\,\mathbf{r}\,\}\right) + \mathrm{U_{LJ}^{tot}}\left(\{\,\mathbf{r}\,\},\{\,q\,\}\right) \ , \tag{1.8}$$

where $K$ is the kinetic energy of the beads of our chain, $\mathrm{U_{KP}}$ is a Kratky-Porod potential modeling the bending rigidity of the chain, $\mathrm{U_{HARM}}$ is an harmonic potential ensuring the connectivity of the chain and, finally, $\mathrm{U_{LJ}^{tot}}$ is a repulsive/attractive interaction mediated by the epigenetic marks.

**Bead-Bead Interaction.** The interaction potential $\mathrm{U}_{\mathrm{LJ}}^{\mathrm{tot}}$ can be decomposed as a sum of two-body potentials:

$$\beta \mathrm{U}_{\mathrm{LJ}}^{\mathrm{tot}}\left(\{\,\mathbf{r}\,\},\{\,q\,\}\right) = \frac{1}{2}\beta \sum_{i=1}^{N}\sum_{j=1}^{N} U_{\mathrm{LJ}}(\|\mathbf{r}_i - \mathbf{r}_j\|\,; q_i, q_j) \ ,$$

where $\beta = \frac{1}{k_B T}$ and $U_{\mathrm{LJ}}$ is a two-particles potential. In this model we consider $U_{\mathrm{LJ}}$ to be a truncated and shifted Lennard-Jones potential (see fig. 1.9) which reads:

$$\beta U_{\mathrm{LJ}}(x; q_i, q_j) = \frac{4}{\mathcal{N}}\frac{\varepsilon(q_i,q_j)}{k_B T}\left[\left(\frac{\sigma}{x}\right)^{12} - \left(\frac{\sigma}{x}\right)^{6} - U_0(x_c)\right]\mathrm{H}\left[x - x_c(q_i,q_j)\right] \ , \quad (1.9)$$

where H is the Heaviside step function, $\varepsilon(q_1, q_2)$ is a free parameter that regulate the interaction strength, and $U_0$ is an auxiliary function which ensures that $U_{\mathrm{LJ}}(x_c; q_1, q_2) \equiv 0$. The cutoff value $x_c$ is what actually models the interaction between the various epigenetic marks; we set:

(a) $x_c(q_1, q_2) = 2^{1/6}\sigma$, if $q_1 \neq q_2$ or $q_1 = q_2 = U$, modeling only steric interaction between beads with different epigenetic marks, or with an unmarked bead;

(b) $x_c(q_1 = 1, q_2 = 1) = x_c(q_1 = 2, q_2 = 2) = 1.8\sigma$, modeling the effective attractive interaction between beads with the same epigenetic marks mediated by the "readers" enzymes.

The interaction strength $\varepsilon(q_1, q_2)$ is:

$$\frac{\varepsilon(q_1,q_2)}{k_B T} = \begin{cases} \alpha_L \equiv \dfrac{\epsilon_L}{k_B T} & \text{if } q_1 = q_2 = \{1,2\} \\ 1 & \text{otherwise} \end{cases} \ , \quad (1.10)$$

where $\alpha_L$ is a convenient, dimensionless, unit for the inverse of the temperature. Note the presence of the numerical constant $\mathcal{N}$ in (1.9). This constants is introduced for convenience as it ensures that the minimum of the attractive interaction is $-\epsilon_L$; it is easy to find that the numerical value of the constant is $\mathcal{N} = 4\left[\frac{1}{4} + 1.8^{-12} - 1.8^{-6}\right] \approx 0.8858$.

Finally, we want to point out a fundamental feature of this interaction. While the interaction itself is a short-range one, thanks to the mobility of the chain (see fig. 1.9) we can in principle have non-local, long ranged interactions as two beads far away on the backbone, can meet thanks to the polymer folding. As already pointed out many times, the presence of local and non-local interactions is a key-role to the birth of a bistable system, as we will discuss more in details in later sections.

**Stiffness Term.** In order to model the stiffness of a polymer, generally a three-body potential is introduced in the Hamiltonian. Several potentials are employed in literature, our choice lies in the so-called The Kratky-Porod potential [70] which

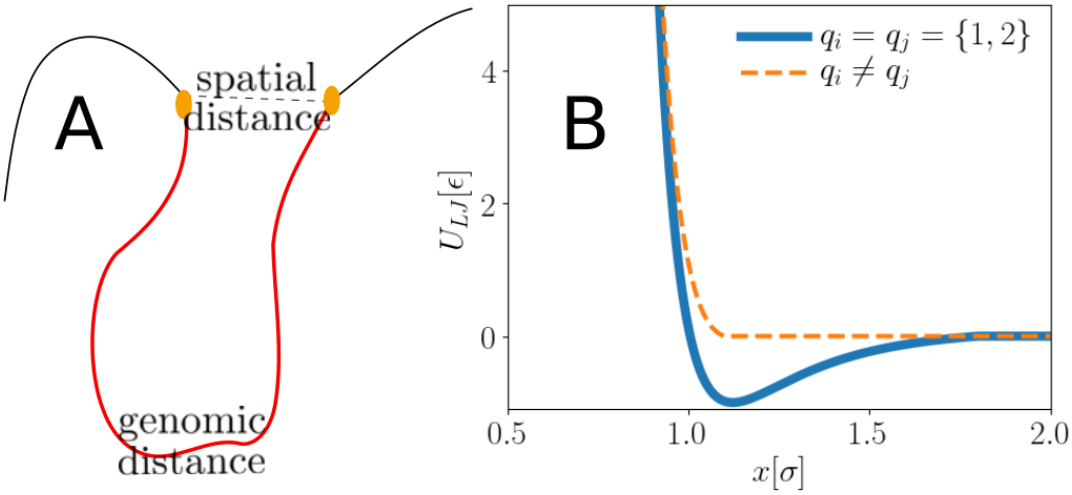



**Figure 1.9:** Bead-bead interaction is modeled via Lennard-Jones potential. (a) The truncated and shifted Lennard-Jones potential (1.9) with cutoff $x_c = 2^{1/6}\sigma$ model the steric interactions between beads with different epigenetic mark; the cutoff $x_c = 1.8\sigma$ instead present an attractive component which mimic the reader-writer dynamics for beads with the same mark. (b) While the interaction is short-ranged, the chromatin folding permits the interaction between beads far away on the chromatin backbone.

favour polymers to follow always the same direction in space. Kratky-Porod potential $\mathrm{U_{KP}}$ can be written in the form:

$$\beta \mathrm{U_{KP}} = \frac{\ell_k}{2\sigma} \sum_{i=1}^{N-2} (1 - \cos(\theta_i)) \quad , \tag{1.11}$$

where $\theta_i$ is the angle between the vector $\mathbf{r}_{i+1} - \mathbf{r}_i$ and the vector $\mathbf{r}_{i+2} - \mathbf{r}_{i+1}$, while parameter $\frac{1}{2}\ell_K$ is usually identified with the persistence length of the chain $\ell_P$. Experiments that try to measure the persistence length of the chromatin fiber find a wide range of possible values $\ell_P \approx 40 - 200\,\mathrm{nm}$ [47, 71], as it seems to depend greatly on the transcribing activity (heterochromatin or euchromatin), and even the concentration of the various architectural proteins. For simplicity, we choose to adopt the value $\ell_P = 3\sigma \approx 90\,\mathrm{nm}$.

**Connectivity Term.** The connectivity a of a polymer chain is usually ensured by a connectivity term in the Hamiltonian. Here, we choose to employ the classic Harmonic potential $\mathrm{U_{HARM}}$ which reads:

$$\beta \mathrm{U_{HARM}} = \frac{1}{2}\beta k_h \sum_{i=1}^{N-1} (\|\mathbf{r}_i - \mathbf{r}_{i+1}\| - r_0)^2 \quad , \tag{1.12}$$

where $r_0 = 2^{1/6}\sigma$ is the average distance between the beads, and $\beta k_h$ represent the connectivity strength and is set to $\beta k_h = 200\alpha_L$.

**Mapping to Physical Units.** Finally, we inquire the typical time-scales of the Langevin Dynamics. One can define the *Brownian time* $\tau_{\mathrm{Br}}$ as the average time required for a bead to diffuse its own size. Disregarding all the interactions, one can estimate the Brownian time as:

$$\tau_{\mathrm{Br}} \approx \frac{\sigma^2}{D} \ ,$$

where $D$ is the diffusion coefficient that can be found using the Einstein relations $D = k_B T/\gamma$, where $\gamma$ is the friction coefficient present in the Langevin equations (1.7). If we approximate the bead to a sphere, the friction coefficient $\gamma$ can be expressed with the Stokes' law $\gamma = 6\pi\eta\sigma$, where $\eta$ is the solution viscosity. Thanks to [69] we infer that viscosity is in the range $\eta \sim 100 - 200\,\mathrm{cP}$. Employing this numerical values, we can find the typical times at the standard temperature $T = 300\,\mathrm{K}$:

$$\tau_{\mathrm{Br}} = \frac{6\pi\eta\sigma^3}{k_B T_L} \approx 6 - 12\,\mathrm{ms} \ .$$

From now on, we will take the value $\tau_{\mathrm{Br}} = 10\,\mathrm{ms}$ as a reference.

### Recolouring Dynamics

After the discussion about the spatial dynamics of the chromatin fiber, let us discuss how the epigenetic landscape evolve. Recolouring dynamics deals with the creation and propagation of epigenetic marks. At a fixed spatial configuration $\{\mathbf{r}_i\}_{i=1\dots N}$, we define an epigenetic configurational energy $H_R$ as follows:

$$H_R = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} U_{\mathrm{epi}}\left(d_{ij}, q_i, q_j\right) \ , \tag{1.13}$$

where $U_{\mathrm{epi}}$ express the strength of the "epigenetic interaction" between two beads with reciprocal distance $d_{ij} = \|\mathbf{r}_i - \mathbf{r}_j\|$ and marks $q_i$, $q_j$. To understand what form the epigenetic interaction must assume, we recall that our main role is to reproduce the positive feedback loop observed in chromatin thanks to the reader-writer dynamics. The feedback loop is observed only nearby nucleosomes with the same epigenetic mark; therefore, we conclude that $U_{\mathrm{epi}} = 0$ if the $q_i \neq q_j$, or at least one of the two beads is still unmarked. On the other hand, as we want a *positive* feedback loop we must favour the interaction between beads with the same mark and that are not "too far away", hence $U_{\mathrm{epi}} < 0$ if $d_i j$ is smaller than a certain threshold $d_R$, and $q_i = q_j = \{A, S\}$. For simplicity, we exploit an epigenetic interaction similar to the attractive Lennard-Jones potential in (1.9):

$$\beta U_{\mathrm{epi}}(d_{ij}; q_i, q_j) = \frac{4}{\mathcal{N}}\alpha_R\left[\left(\frac{\sigma}{d_{ij}}\right)^{12} - \left(\frac{\sigma}{d_{ij}}\right)^{6} - U_0(x_c)\right] \tag{1.14}$$

with $d_{ij} < d_R \equiv 1.8\sigma$, and $q_i = q_j = \{A, S\}$. The variable $\alpha_R$ controls the interaction strength, and in general it is different ($\alpha_R \neq \alpha_L$) from the interaction strength introduced in (1.9).

It is important to understand why, in principle, $U_{\text{epi}}$ and $U_{LJ}$ are two different functions. They express the interaction strength relative to two different processes. $U_{LJ}$ control the dynamical constraints on the physical movement of the bead, due to the formation of reader-writer complexes where readers act like bridging proteins. The epigenetic interaction $U_{\text{epi}}$, instead, measure how likely it is for a certain bead to lose/transform its epigenetic mark due to the recruitment of other beads and/or the action of an *eraser* enzyme (see section 1.1).

Another peculiar feature is the similarity between the epigenetic energy $H_R$ (1.13) and the Hamiltonian of a Potts-like system with short range interactions between first neighbours. This analogy is quite strong and will be employed later (see chapter 3) to develop a mean-field version of this model. However, it is important to note that in a classic 1-dimensional spin system with short range interactions, due to the Mermin-Wagner theorem [72], no phase transition is possible. However, here, thanks to the chromatin folding effective long-range interactions between beads are indeed possible, and, in principle, phase transition could arise.

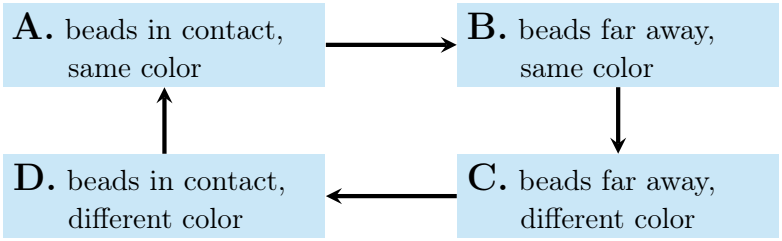
Finally, the recolouring dynamics is implemented using a Metropolis-Hastings algorithm with Hamiltonian $H_R$ at temperature $T$. Note that this implementation, in general, permits direct recruitment $A \to S$ where active beads can become silenced, or vice versa, without first becoming unmarked. While this is unusual, it can be explained via some sort of cooperation between erasers and writers enzymes, making the effective time where the nucleosome is left unmarked small and negligible. Furthermore, while the Langevin dynamics can be implemented at continuous time, recolouring dynamics is inherently discrete. This means, that we must introduce a typical recolouring time $\tau_R$ that regulate the dynamics speed. We choose $\tau_R = 10^3 \tau_B$, as the recolouring dynamics is much slower than the Langevin one, as it involves several chemical reactions needed to modify the epigenetic marks. In practice, Basically, the dynamics work as follow: (i) the chromatin fiber is created with some initial configuration; (ii) we implement the Langevin dynamics solving equations (1.7) for a time $\tau_R$; (iii) we implement the recolouring dynamics using the Metropolis Hastings algorithm; (iv) we repeat the process from step (ii).

**Detailed Balance.**

As just discussed, the epigenetic potential $U_{\text{epi}}$ and $U_{LJ}$, are in principle different as they model two different processes and therefore $\alpha_L \neq \alpha_R$. Now, note that the Langevin dynamics is implemented via Langevin equations derived from the Hamiltonian (1.8), and is therefore an equilibrium process. At the same time, we implement the recolouring dynamics via a Metropolis-Hastings algorithm that, again, reproduce equilibrium dynamics. Therefore, if $U_{\text{epi}} \equiv U_{LJ}$ or $\alpha_R = \alpha_L$, we have that the complete process (Langevin+recolouring) minimizes the same action potential, and, therefore, in this case we are clearly dealing with an equilibrium model. What happens, instead, in the general case? Can the equilibrium still be recovered?

One of the main features of an equilibrium system is that the principle of the

detailed balance holds, that is at equilibrium *the average rate of every process is equal to the average rate of its reverse process* [73]. It is difficult to verify directly the detailed balance, however we can employ an equivalent principle, called Kolmogorov's criterion [74], that claims that detailed balance is respected if and only if the probability for every possible cycle to happen is equal to the probability for the reversed cycle. Given this, we study the following cycle where two beads of the chromatin are manipulated in various ways:




If the detailed balance holds, then the Kolmogorov's criterion claims that the probability of the loop $p(A \to B \to C \to D \to A)$ is equal to the probability of the inverse loop $p(A \to D \to C \to B \to A)$.

To compute the probability of the forward loop $A \to B \to C \to D \to A$ we use a Metropolis-like algorithm:

1. The move $A \to B$ is energetically disadvantageous, the move is therefore accepted with probability $p(A \to B) = \exp[-\alpha_L]$.
2. In the move $B \to C$ there is neither loss nor gain of energy, therefore it is accepted with probability $p(B \to C) = 1$.
3. In the move $C \to D$ there is neither loss nor gain of energy, therefore it is accepted with probability $p(C \to D) = 1$.
4. The move $D \to A$ is energetically favorable, therefore it is accepted with probability $p(D \to A) = 1$.

We conclude that the probability of the forward loop is $p(A \to B \to C \to D \to A) = \exp[-\alpha_L]$.

The probability of the backward loop $A \to D \to C \to B \to A$ is computed in the same way:

1. The move $A \to D$ is energetically disadvantageous, the move is therefore accepted with probability $p(A \to D) = \exp[-\alpha_R]$.
2. In the move $D \to C$ there is neither loss nor gain of energy, therefore it is accepted with probability $p(D \to C) = 1$.
3. In the move $C \to B$ there is neither loss nor gain of energy, therefore it is accepted with probability $p(C \to B) = 1$.
4. The move $B \to A$ is energetically favorable, therefore it is accepted with probability $p(B \to A) = 1$.

We conclude that the probability of the backward loop is $p(A \to D \to C \to B \to A) = \exp[-\alpha_R]$.

The detailed balance is henceforth respected if and only if $\alpha_L = \alpha_R$. This means only if the two interactions ($U_{\text{epi}}$ and $U_{LJ}$ are identical, the dynamic will sample from an equilibrium distribution. In all the other cases, instead such consideration will not be possible as we will be in an out-of-equilibrium regime.

### 1.4.2 Simulation Details

The dynamics is studied by employing extensive numerical simulations. Langevin dynamics is evolved thanks to the integration of equations (1.7) using a velocity-Verlet scheme within the `LAMMPS` engine [18] in Brownian dynamics mode (NVT ensemble). The polymer has been enclosed in a large cubic box with side $L$ with periodic boundary conditions at border; the size of the box $L$ is chosen to be $L = 10^3\sigma$. Note that if we suppose that each monomer occupy a spherical volume $\frac{4}{3}\pi\left(\frac{\sigma}{2}\right)^3$, if we have $N = 10^4$ monomers, then the chain will occupy only the 0.5% of the total volume of the box. In our numerical simulation the value of $N$ is typically in the range $N \approx 100 - 1000$, therefore the system will evolve in a dilute regime where interactions between different chains can be neglected.

The initial configuration of the polymer is typically that of an ideal 3-dimensional random walk of $N$ steps in which each monomer assumes a random epigenetic mark $q = \{A, S, U\}$. Such a configuration is not viable, as it is likely that monomer overlaps. To solve this problem we do a preemptive "warm-up" run to equilibrate our chain. Initially we do evolve the system dynamics for $10^5\tau_{\text{Br}}$ timesteps in which the only interaction is an increasingly stronger steric soft-core repulsion between every pair of beads, while their colour is left unchanged. The soft-core potential $U_{\text{soft}}$ reads:

$$U_{\text{soft}}(d_{ij}) = A\left[1 + \cos\left(\frac{\pi d_{ij}}{d_C}\right)\right]\mathrm{H}\left[d_C - d_{ij}\right] \quad , \tag{1.15}$$

where $d_{ij}$ is the reciprocal distance between two beads; $A$ is the maximum interaction strength; $d_C$ is the cutoff that we set $d = 2^{1/6}\sigma$; H is the Heaviside step function.

Following the equilibration run, we have removed overlaps between monomers, and finally have a good ensemble of initial configurations for the simulation of the actual model. The main run typically consists of $10^6\tau_{\text{Br}}$ timesteps, in which $N$ recolouring moves are attempted every $\tau_R = 10^3\tau_{\text{Br}}$ timesteps. Each recolouring move is accepted or rejected employing a Metropolis-Hastings algorithm where the acceptance probability is given by:

$$p(q \to q') = \exp\left[-\frac{\Delta H_R}{k_B T}\right] \quad , \tag{1.16}$$

where $\Delta H_R$ is the variation of epigenetic energy (1.13) after the recolouring.

Finally, note that typically the chain size is $N = 100$ to speed up the simulation times, and for each set of interaction strength $(\alpha_L, \alpha_R)$ we perform 100 independent numerical simulations in order to have a good sampling.

### 1.4.3 Equilibrium Dynamics

As a first step in the study of the model, we focus on the equilibrium regime where the two interaction strengths $\alpha_L$ and $\alpha_R$ are identical. This is the simplest case we could study as it permits direct comparison with well-studied models, like spin models, or even coil-globule transition at $\theta$-point typical of homopolymers.

In order to study the dynamics, we introduce two state-variables to monitor both the chromatin folding, and the dynamics of the epigenetic landscape. The folding of the polymer is studied via the radius of gyration $R_g$, which is a measure of the polymer average extension [75]. The epigenetic state, is, instead, studied in analogy to a spin system where the activated states have spin $+1$, silenced states $-1$, and unmarked have spin 0. Thanks to this analogy, we define the "epigenetic magnetisation" $m$ as follows:

$$m = \frac{1}{N}\sum_{i=1}^{N}\left[\delta_{q_i,A} - \delta_{q_i,S}\right] \quad , \tag{1.17}$$

where $m = 1$ if the chromatin is completely active, $m = -1$ if it is completely silenced, $m = 0$ if no global epigenetic order is present.

First thing to notice is that, while the model permits each bead to assume three states (activated, silence, or unmarked), the unmarked state seem to be marginal in the dynamics. Indeed, in the ordered phases of the model, we find that the unmarked state rapidly disappears from the polymer at the advantage of the self-attractive ones, thus the model is effectively a two-state model. One way to "adjust" this, is to force the passage from the unmarked state during the recolouring, however [1] shows that while the dynamics displays some interesting features, at the equilibrium there are no substantial differences from our model.

We study the dynamics of a polymer composed of $N = 1000$ beads, by varying the interaction strength $\alpha$ in the model. We observe the existence of two fundamentally different equilibrium phases:

1. *Swollen-Disordered* (SD) phase (see fig. 1.10A) , observed at low values of $\alpha < \alpha_c \approx 0.95$ (weak interaction, high temperatures), where the chain is observed in an open state (high $R_g$), without any apparent epigenetic order ($|m| \approx 0$).
2. *Compact-Ordered* (CO) phase (see fig. 1.10B), observed at high values of $\alpha > \alpha_c$ (strong interaction, low temperatures). The chain collapses fairly quickly into a single globule (low $R_g$), and clusters of colors emerge. Clusters with different epigenetic state compete, and the system ultimately evolves into a single compact, epigenetically coherent, domain ($|m| \approx 1$).

Note that because the interactions between the two states are equal, the selection of which epigenetic mark dominates in the compact-ordered phase is via a $\mathbb{Z}_2$ symmetry breaking, which suggests the existence of a phase transition between the swollen-disordered and compact-ordered states.

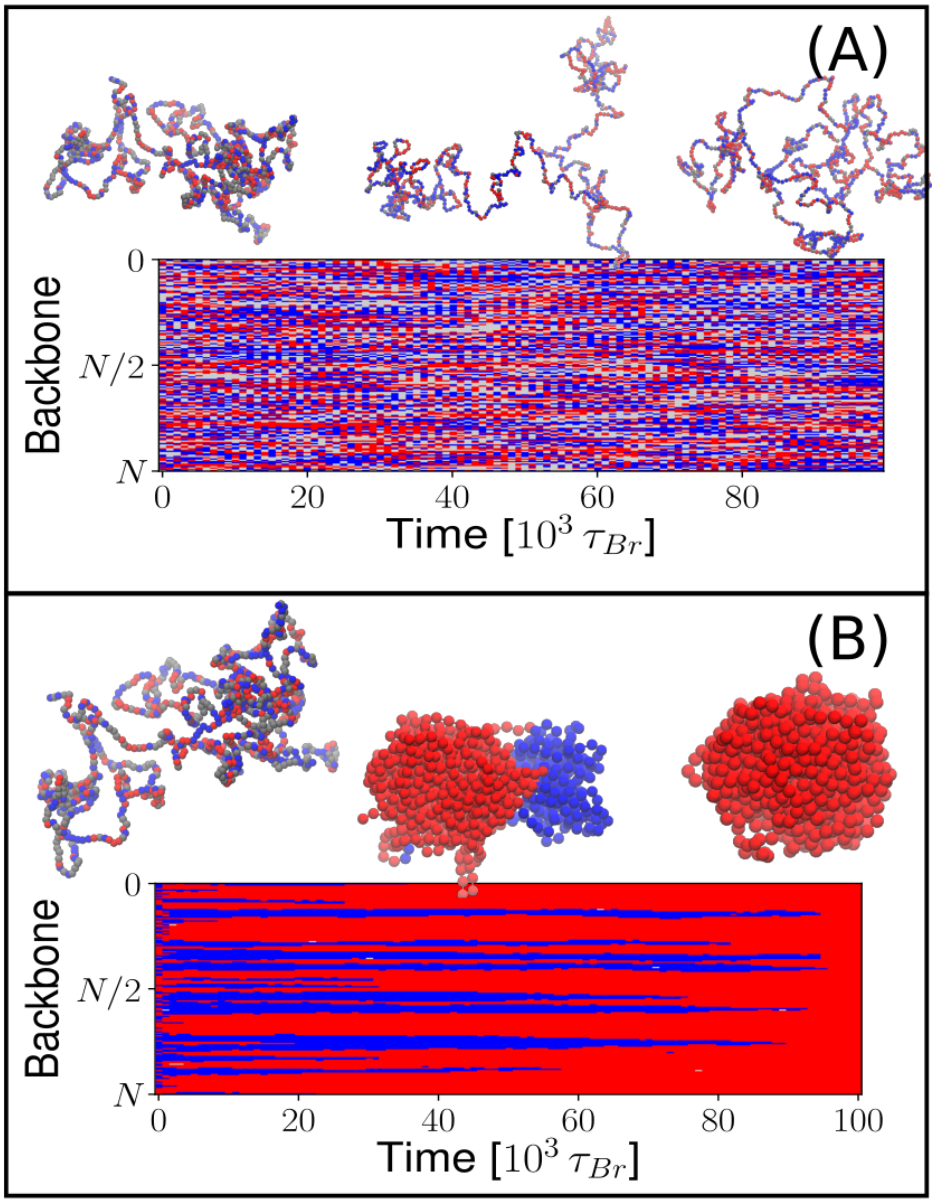



**Figure 1.10:** Time evolution of a polymer composed of $N = 1000$ beads at the equilibrium regime. On the bottom part of each figure, a kymograph illustrate the evolution of the epigenetic landscape, while on the top some snapshots show the spatial configurations in some moment of the dynamics. **(A)** Polymer in the swollen-disordered phase with $\alpha = 0.5$. No coherent epigenetic domains, and no ordered spatial structure, arise during the dynamics. **(B)** Polymer in the compact-ordered phase with $\alpha = 2.1$. The system start from a swollen-disordered configuration, but quickly organizes in two compact globules of different epigenetic mark. These two marks compete, but eventually the activated state (red marks) prevail.

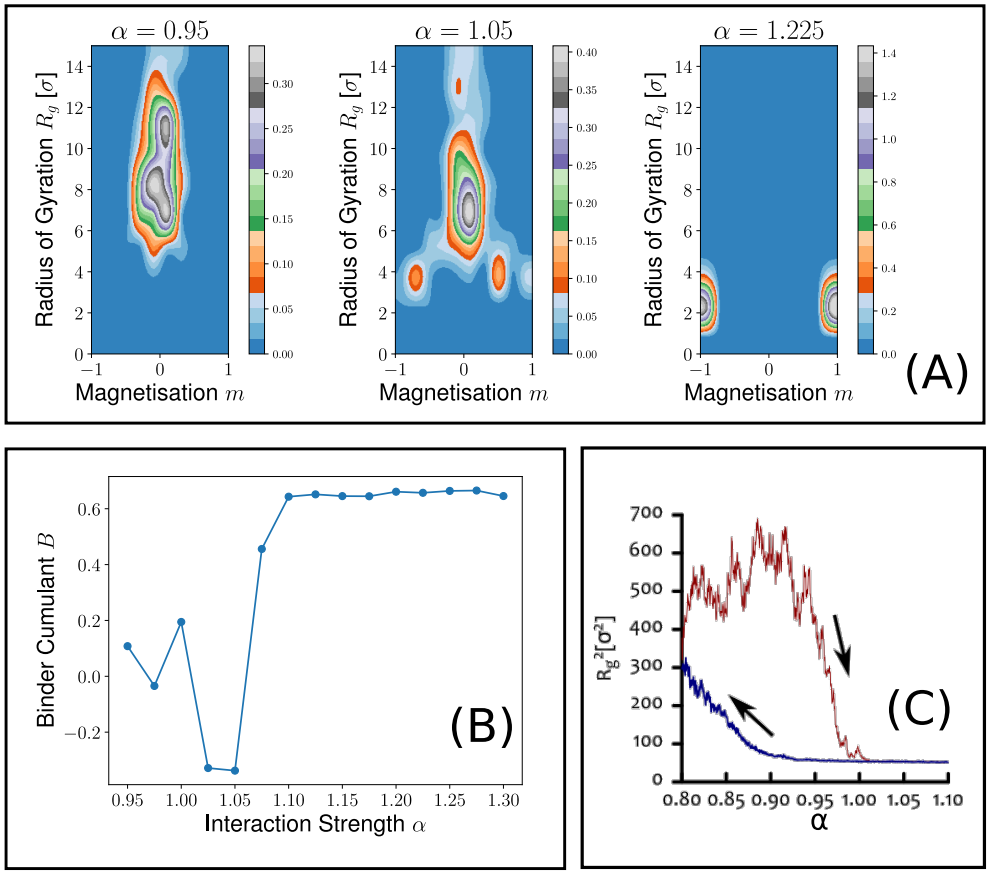



**Figure 1.11:** Numerical evidences of the presence of a first order phase transition between a SD phase, and a CO one. **(A)** The plots show the joint probability $P(m, R_g)$ for a chain of $N = 100$ beads, obtained from 200 independent numerical simulations. At $\alpha = 0.95$ the system show one maximum at high $R_g$ and $|m| \approx 0$ (SD); at $\alpha = 1.225$ two symmetric maximum are present at low $R_g$ and $m \approx \pm 1$ (CO); at $\alpha = 1.05$ three maxima are present, showing coexistence of the two phases. **(B)** Plot of the Binder cumulant as a function of $\alpha$ obtained from the same simulations as in (A). When $\alpha < 1.025$ the Binder cumulant is $B \approx 0$ typical of a disordered magnetic configuration; when $\alpha > 1.07$ we have $B \approx \frac{2}{3}$, compatible with the presence of a $\mathbb{Z}_2$ symmetry; when $\alpha = 1.025, 1.05$ we have $B < 0$ typical of first order transitions near the critical point. **(C)** Adapted from [1]. Plot of the radius of gyration of a chain of $N = 2000$ beads, as a function of the interaction strength $\alpha$ which we slowly increase from $\alpha = 0.8$ (below the transition) to $\alpha = 1.1$ (above the transition) in $10^6 \tau_{Br}$. From there, we decrease $\alpha$ gradually back to $\alpha = 0.8$ in the same amount of time (blue curve). The two curves, averaged over 5 different runs, show the presence of hysteresis typical of first-order transitions.

In order to study the existence, or the nature, of the phase transition we do some more thorough simulations with smaller chains ($N = 100$) at a wider range of interaction strength $\alpha \in [0.9, 1.3]$. For each value of $\alpha$, we do 200 independent simulation runs, in order to have a good sampling of the model. First, we note that the critical value $\alpha_c$, previously estimated at $\alpha_c \approx 0.95$, here is slightly higher and have $\alpha_c^\star \approx 1.02$, but this was expected due to finite-size effects [76]. Thanks to the sampling, we can construct the joint probability $P(m, R_g)$ (see fig. 1.11A), and study the distributions typical of the two phases. Namely, for the swollen-disordered phase, we expect a single maximum at large $R_g$ and small $m$, while for the compact-ordered we observe the presence of two ($\mathbb{Z}_2$ symmetry) maxima corresponding at small $R_g$ and $|m| \approx 1$. At the critical point ($\alpha \approx \alpha_c^\star$), three maxima are clearly visible, suggesting the presence of phase coexistence between SD and CO. The coexistence, and therefore the metastability, of two phases nearby the critical point, it is a hallmark of a first-order phase transition

Further proofs on the nature of the phase transition come from the study of the so called Binder cumulant [77] $B$ for the epigenetic magnetisation, which reads:

$$B(\alpha) = 1 - \frac{\langle m^4 \rangle}{3 \langle m^2 \rangle^2} \ , \tag{1.18}$$

where with $\langle - \rangle$ we mean the average taken over the simulation sample at a fixed interaction $\alpha$ value. In a magnetic spin system with $\mathbb{Z}_2$ symmetry, we observe that if there is a phase transition is of the first order at $\alpha_c$, then in the disordered phase we have $B(\alpha < \alpha_c) = 0$; in the ordered phase $B(\alpha > \alpha_c) = \frac{2}{3}$; finally in the neighborhood of $\alpha \approx \alpha_c$, the Binder cumulant will diverge and $B(\alpha \to \alpha_c) = -\infty$. In fig. 1.11B we plotted the values of $B(\alpha)$ in our simulations; we observe that nearby $\alpha_c^\star$, the value of $B$ is sharply negative, suggesting, again that our phase transition is of the first order.

Finally, [1] also verifies the presence of a first order phase transition, thanks to existence of an hysteresis curve (see fig. 1.11C). Here, we take an equilibrated chain below the transition point and slowly decrease the temperature (or increase interaction) until it reaches the compact-ordered phase, and then do the opposite process. By plotting the values of the radius of gyration as a function of the strength $\alpha$ we note that during the cooling, and the heating process, the model has two different trajectories, creating a hysteresis loop, typical, again, of a first order phase transition.

The existence of a first order phase transition in the model is striking as it greatly differs from the basic models we have used as building blocks. First, we have modeled the epigenetic landscape on the template of an a two-states Potts-like model which, as already discussed, do not show phase transition at 1D, and show a continuous transition at higher dimensionality. Second, we would be tempted to do comparison with the classic homopolymer in a solvent model; however, while this basic model show a transition between a swollen and a compact phase at the $\theta$-point, the transition is still continuous. The cooperation between folding, and magnetic-like interaction, give birth to a novel and interesting phenomenology [19, 78]. The existence of a first-order-like transition in this model provides, also, a marked difference with the previous ones,

which considered the epigenetic landscape as a 1-dimensional system (see section 1.3) where interactions acted mostly thanks to ad-hoc long-range interactions. These models usually shows a continuous phase transition; first-order phase transition can be obtained, but only in mean-field approximations (zero-dimensional models) where all nucleosomes interact with all nucleosomes.

Finally, we point out that, thanks to the presence of a first order phase transition, the model appears to be a good framework where to study real biological applications. Indeed, it naturally shows the properties we required from a good model describing the epigenetic landscape (see section 1.3). First, the model is strongly stable against perturbations, indeed the first-order transition implies the presence of hysteresis, that itself indicates the presence of "memory" in the system: when a nucleosome is switched on (or off) will rarely change its activation state. Second, as there is $\mathbb{Z}_2$ symmetry, the system is bistable in the ordered phase. All this strongly suggests that the features characterizing the epigenetic switches, may be in fact inherited from an effective first-order-like transition driven by the coupling between epigenetic dynamics and chromatin folding.

### 1.4.4 Out-Of-Equilibrium Dynamics

In this section we will discuss the dynamics of the model, when the epigenetic interaction $\alpha_R$, and the Lennard-Jones interaction $\alpha_L$ are different, and the model is therefore in an out-of-equilibrium regime. The dynamics is again studied using extensive numerical simulations [2] of the system and exploring the phenomenology when $\alpha_L \neq \alpha_R$, with a certain regard to the cases where $\alpha_L \gg \alpha_L$, or $\alpha_L \ll \alpha_R$. While in the equilibrium regime, only a swollen-disordered phase, and compact-ordered phase are possible, here at least two other phases (swollen-ordered and compact-disordered) are possible. These two new possibilities not only have a certain biological relevance, but also prove that the plasticity of the model to reproduce various epigenetic landscapes.

As a first step, we study the dynamics of a chain composed of $N = 100$ beads by varying the values of $\alpha_L$ and $\alpha_R$ in the range $\alpha_L, \alpha_R \in [0, 4]$. For each pair of values $(\alpha_L, \alpha_R)$ we usually compute only 16 independent runs, as we are only interested in the final steady state of the dynamics, and generally do not compute statistic on it. This is not the case, however, for points nearby the equilibrium line $\alpha_L = \alpha_R$, where we do several simulations in order to verify the presence (or not) of coexistence between the phases, as observed in the previous section.

Results of the analysis (see fig. 1.12) show that, depending on the values of $(\alpha_L, \alpha_R)$, the chain can display four different phases at the steady state:

1. The *Swollen-Disordered* (S.D.) phase, already observed at the equilibrium regime, arise when the epigenetic and attractive interactions are both weak. Here the system presents in an open configuration characterised by high radius of gyration, and no apparent epigenetic order.

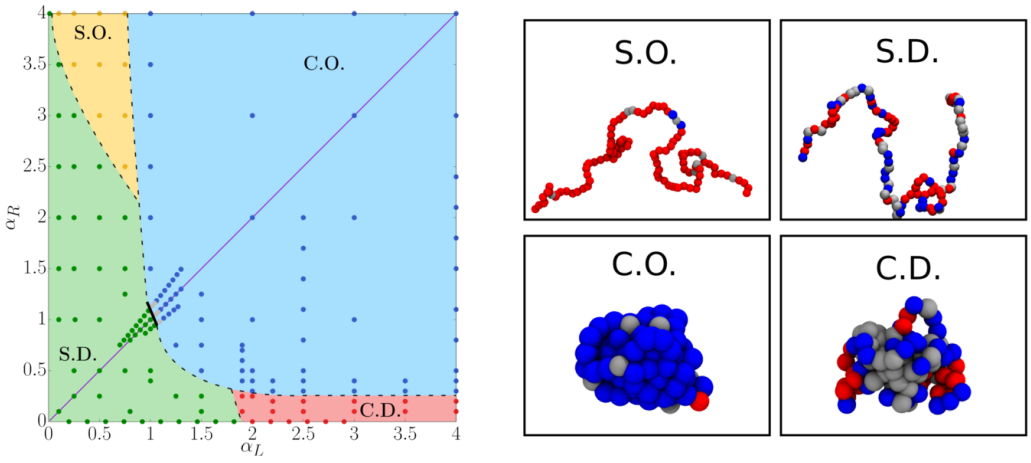



**Figure 1.12:** The phase diagram of the model in the out-of-equilibrium regime, in the space $(\alpha_L, \alpha_R)$ for a chain with $N = 100$ beads. We can observe four different phases: (1) swollen-disordered (S.D.); (2) compact-disordered (C.D.); (3) swollen-ordered (S.O.); (4) compact-ordered (C.O.). Each dot correspond correspond to a set of parameters examined via an actual numerical simulation; grey dots represent set of parameters where coexistence between C.O. and S.D. is observed. Dashed lines are qualitative borders delimitating the phases, while the solid one stand for a possible first-order-like phase transition. On the right side, snapshots illustrative of each of the four phases are shown.

2. The *Compact-Ordered* (C.O.) phase, already observed at the equilibrium regime, arise when the epigenetic and attractive interactions are both strong. Here the system presents in a strongly compact configuration characterised by low-radius of gyration, and highly epigenetic ordered, usually display only one big domain characterised by a single epigenetic mark. Notably, this kind of configuration is typically observed in the inactive copy of the X chromosome in female mammals, which is entirely transcriptionally silent, and inherited in the cell division [9].

3. The *Compact-Disordered* (C.D.) phase arise when there is weak epigenetic interaction (low $\alpha_R$), and strong attractive interaction (high $\alpha_L$). The system is found in a compact configuration characterised by small radius of gyration, but no epigenetic order. Note, however, that the typical radius of gyration found in this phase is bigger than what observed in the C.O. phase. What we observe is reminiscent of the so-called black chromatin [79] found in Drosophila, where no coherent epigenetic mark is observed, but still tends to localise in 3D, possibly due to the linker histone H1.

4. The *Swollen-Ordered* (S.O.) phase arise when there is a strong epigenetic interaction (high $\alpha_R$), and weak attractive interaction (low $\alpha_L$). The system is found in an open configuration characterised by an high gyration radius, while the epigenetic landscape show *local* epigenetic order. Here, the landscape is organized

in various domains of coherent epigenetic mark, which can be found both in open configurations or in small globuli. If the epigenetic interaction is strong enough, all the domains will fuse into a single domain of coherent epigenetic mark, found in a swollen configuration. Notably, this kind of configuration resemble what is observed when the chromatin is actively transcribing (euchromatin) [9].

The existence, and characterisation, of phase transitions between the domains depicted in the phase diagram, is still object of debate. First, note that, nearby the equilibrium line, we still observe phase coexistence suggesting the permanence of a first-order-like phase transition between S.D. and C.O phase. However, when we stray away from the equilibrium the coexistence seems to cease to exist, indicating a possible continuous phase transition. As we will discuss in a moment, in the other cases some heuristic reasoning suggest the presence of either a continuous phase transition, or no phase transition at all. Let us now discuss the main properties, of the novel phases found in the out-of-equilibrium regime.

### Swollen-Ordered Phase

Here, we discuss the main features of the phenomenology of the Swollen-Ordered phase. This phase is observed at small attractive interactions strength ($\alpha_L \lesssim 0.62$) and high epigenetic interaction ($\alpha_R \gtrsim 2.4$). The chain present in an open swollen configuration, but still display local epigenetic order.

Let us consider a chain in the S.D. phase at low $\alpha_L$ value which slowly quenches into the S.O. region increasing only the value of $\alpha_R$. If the value of $\alpha_L$ is small enough, one could suppose that the each bead can interact only with its first neighbours on the chromatin backbone, if this is true the chain dynamics should be similar to what happens in the 1D Ising model, and therefore there should be no phase transition between S.D. and S.O. On the other hand, however, thanks to chromatin folding sporadic interactions with distant beads are possible, and indeed, if $\alpha_L \gtrsim 0.3$ we observe the formation of localized, but still swollen, spatial structures (see fig. 1.13A). Given this, we can not determine if a phase transition occurs, or not, and further studies are necessary.

Second, we consider the model in the case of no-attractive interaction ($\alpha_L \to 0$), and strong epigenetic interaction. Note that if $\alpha_L \to 0$, not only we don't have attractive interactions among beads of the same mark, but also in the Hamiltonian (1.8) the terms regarding connectivity strength would vanish. The result would be a weakly connected polymer, similar to a really diluted gas where beads would never interact; therefore regardless of the value of $\alpha_R$ the system will always be in the S.D. phase if $\alpha_L \to 0$. Let us start from a chain in these conditions, and slowly increase the value of $\alpha_L$. As $\alpha_L$ increase, the average distance between the beads will decrease, and they can start interacting again. Because we kept the value of $\alpha_R$ high, they will quickly form locally ordered epigenetic domains, while still retaining a completely swollen configuration (see fig. 1.13B). If we still increase the attraction

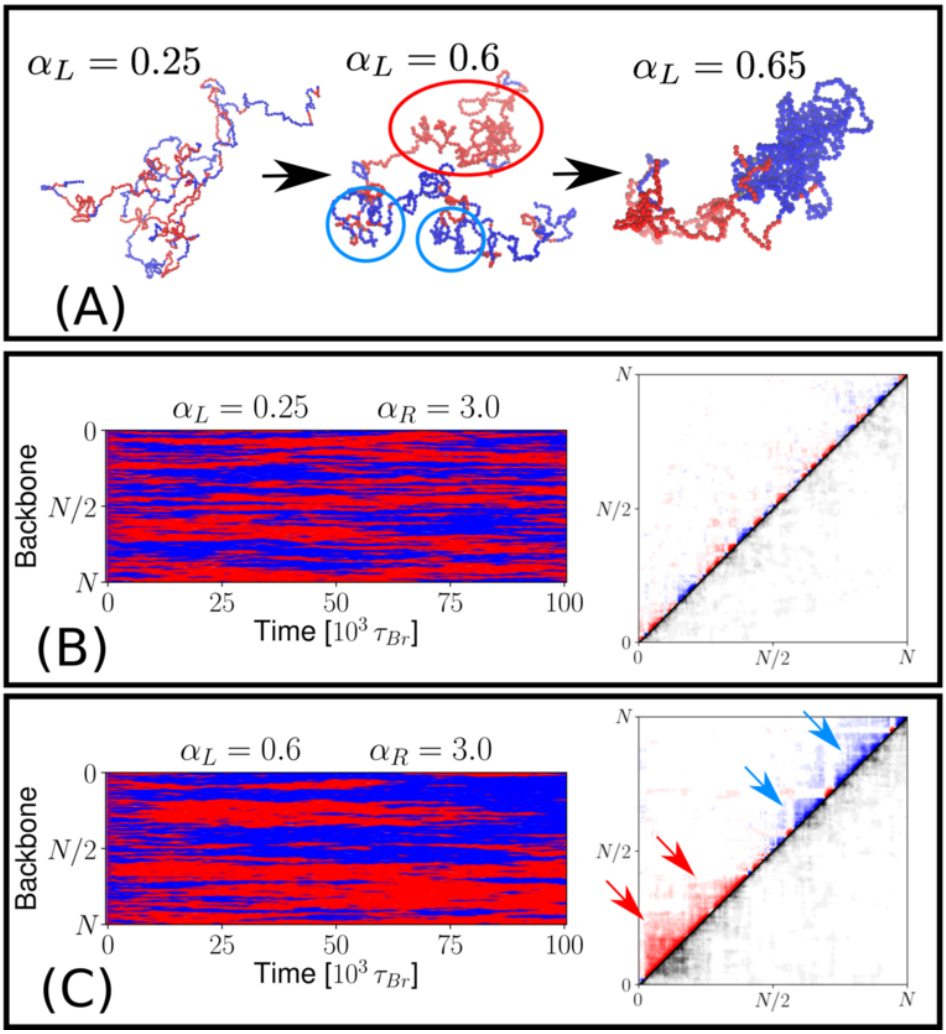



**Figure 1.13:** Model Dynamics in the S.O. phase for a polymer with $N = 1000$ **(A)** Snapshot of a polymer at the steady state with different values of $\alpha_L$, and $\alpha_R = 3.0$. At $\alpha_L = 0.25$ the polymer is completely swollen, while still showing local epigenetic order; $\alpha_L = 0.6$ the system show TAD-like formations (highlighted by some circles); at $\alpha_L = 0.65$ we are in the C.O. region and the TAD-like structures now are found in compact regions. **(B)** Kymograph (on the left) for a polymer at $\alpha_L = 0.25$ and $\alpha_R = 3.0$ show the existence of metastable, locally ordered regions. On the right a contact map averaged over the last $10^5 \tau_{Br}$ steps of the dynamics: the upper half is color-coded to separately show the probability of red-red, blue-blue contacts. **(C)** Same as (B), but this time at $\alpha_L = 0.6$; this time the contact map clearly show the existence of TAD-like structures highlighted by some arrows.

strength, we will note that the system starts to form spatial structure similar to the TADs (topologically associating domains) observed in chromatin [80], where beads of the same mark seem to aggregate, without necessarily forming compact domains (see fig. 1.13C). The local density of TADs increase as $\alpha_L$ increases, until they reach a point where the swollen configuration becomes unfavorable, and they will collapse in globule in the C.O. phase. Note, however, that while the analogy between the S.O. structures, and the TADs is striking, there is a fundamental difference between the two: TADs are stable, while our architectures are metastable and will change in time. This issue will be treated more in detail, and somehow solved, in the next chapter.

### Compact-Disordered Phase

Finally, we discuss briefly here then phenomenology of the compact-disordered phase, observed at high attractive interactions ($\alpha_L \gtrsim 1.9$), and weak epigenetic interaction ($\alpha_R \lesssim 0.3$). The system here present in a semi-compact configuration, as its radius of gyration is still bigger than the typical one found in the C.O. phase, and disordered epigenetic landscape.

As a first step, we study the system in the limit of $\alpha_R \to 0$, where there is no epigenetic interaction, and the landscape is subject to completely random recolouring every $\tau_R$ steps. Studying the graphic of the radius of gyration as a function of $\alpha_L$ (see fig. 1.14) we observe that the system pass continuously from a swollen configuration with high $R_g$ at $\alpha \lesssim 2$, to a compact configuration at $\alpha \gtrsim 2$ with low $R_g$. Therefore, it is tempting to claim that there is a second-order-like phase transition between S.D. and C.D. when $\alpha_R \to 0$. Indeed, the dynamics of this system is reminiscent of what happens in the case of randomly colored heteropolymers in a solvent [81, 82]. Here, the polymer undergoes a second-order phase transition between a coil-regime and a globule-regime, similar to the $\theta$-transition observed in homopolymer. Note however that while in the classic case the heteropolymer structure is kept fixed, in our model is inherently out-of-equilibrium as the epigenetic landscape is continuously recoloured in a random fashion. We thus conclude that the phase transition appear to be continuous, but due to the out-of-equilibrium nature of the model, further verification are needed.

Finally, we examine here the possible phase transition between the compact-disordered phase, and the compact-ordered one. In the limit of high interaction strength $\alpha_L$, a chain organizes quickly in a solid, compact configuration. In this limit the typical time to reach a stable, spatial, configuration are much faster than the typical recolouring time $\tau_R$, and the reciprocal distances between the beads are somehow frozen. Hence, each bead will, in average, interact always with the same beads, and the system will behave like a 3D spin system with two-state Potts-like dynamics, that is characterised by a second order phase transition [83] (see also chapter 3). Thanks to this heuristic, we conclude that even in the case of C.D. to C.O., the model will undergo a second order phase transition.

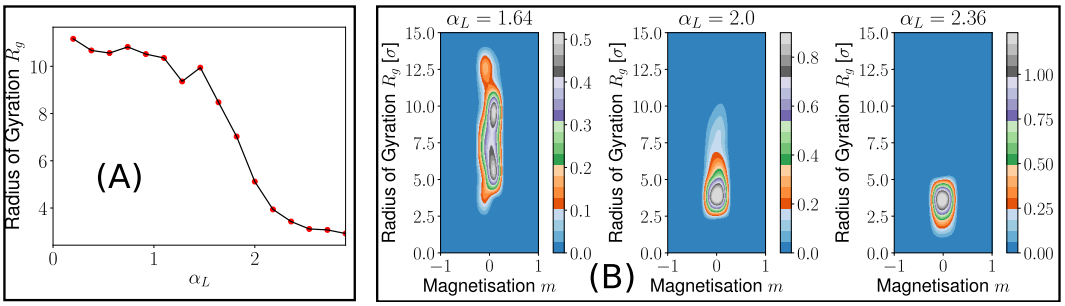



**Figure 1.14:** Possible continuous phase transition of a chain of $N = 100$ beads in the limit of random recolouring ($\alpha_R \to 0$). **(A)** Radius of gyration as a function of the attractive interaction strength $\alpha_L$; note that the system pass continuously from a swollen configuration with $R_g \approx 10$, to a compact configuration with $R_g \approx 2.5$. **(B)** Joint probabilities $P(m, R_g)$ at various values of $\alpha_L$ show no coexistence suggesting, again, a second-order phase transition.

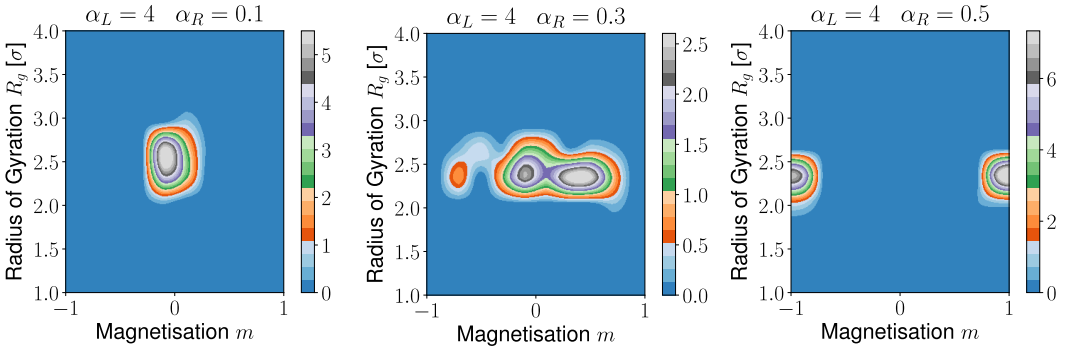



**Figure 1.15:** Possible continuous phase transition from the C.D. to C.O. when keeping $\alpha_L$ to a very high value ($\alpha_L = 4$) and varying $\alpha_R$. Initially we have that the system is found in a compact-disordered configuration (low $R_g \approx 2$, and $m \approx 0$; near the transition point the system show bifurcation at $\alpha_R = 0.3$, and finally is in a compact-ordered ($|m| \approx 1$ configuration $\alpha_R = 0.5$.

# Chapter 2

# Epigenetic Bookmarking

*"Is that vodka?" Margarita asked weakly.*
*The cat jumped up in his seat with indignation.*
*"I beg pardon, my queen," he rasped, "Would I ever allow myself to offer vodka to a lady? This is pure alcohol!"*

---

Mikhail Bulgakov, `The Master and Margarita`

In chapter 1 we introduced and discussed a model [1] to describe the dynamics of chromatin folding, and at the same time, to reproduce the main features of epigenetic marks spreading on the chromatin backbone. While the model, indeed, reproduce the main properties we expect from a model of epigenetic dynamics(e.g. multistability, stability against perturbations, and memory), it fails to reproduce the richness of epigenetic landscapes observed in real cells. Indeed, the dynamics always tend to favour the formation of big coherent epigenetic domains, while real chromatin fibers in living cell show complex architectures characterised by a multitude of different epigenetic marks. In this chapter, we propose an extension to the model of section 1.4 to deal with this problem thanks to the introduction of a new class of biophysical mechanisms that can establish and maintain robust yet plastic epigenetic domains via the so-called genomic-bookmarking mechanism [2]. We also extend the model to include a greater number of "colors" and prove that genomic-bookmarking alone explain the distribution of marks found by experiments on a *Drosophila* chromosome.

First, let us recall the main features of of the model discussed in chapter 1 that we later extend to include genome bookmarking. Here, the chromatin fiber is described by a semi-flexible polymer chain whose beads represent group of neighbouring nucleosomes. Each bead can have a different *colour* representing one of the possible combinations of epigenetic modifications applied to the nucleosomes. To simplify the system, we chose to adopt a three-state model where beads can either be "blue" e.g. methylated (silenced genome), "red" e.g. acethylated (actively transcribing genome), or "gray"

e.g. no modification is present on the nucleosomes. Beads interact with each-other via a short-range interaction which is attractive for beads with the same epigenetic mark, and steric otherwise. The time evolution of the spatial positions of the polymer beads is defined by a set of Langevin equations (1.7), while the "recolouring dynamics", entrusted with the role to regulate epigenetic spreading, is defined by the spin-like Hamiltonian (1.13) and evolved thanks to a Metropolis-Hastings algorithm.

Results of numerical simulations show that at the equilibrium regime, the model undergoes a first order phase transition from a swollen-disordered to a compact-ordered phase (see section 1.4.3), where the system is found in densely compact globule characterised from a single epigenetic domain. While this situation resemble what observed in inactivated X chromosome of mammalian females, it is of difficult generalisation to other biological observations. At the out-of-equilibrium regime (see section 1.4.4) the system show more interesting configurations, namely in the so-called swollen-ordered phase where TAD-like structures appear. Unfortunately, this structures are metastable and tend to be short-lived, while the opposite situation is found in cells.

In order to extend the model, we will employ three assumptions on the biophysical mechanisms involved in gene expression; all three assumptions are still object of debate in biology, but have consistent literature backing them up. First, as in the basic model, we assume the existence of a positive-feedback because of the "reader-writer" interaction. This is based on the observation that certain readers, binding to the chromatin, recruit writers of a specific mark, and, again, nucleosome of the new "written" mark will recruit reader proteins, ans so on. For instance, it is observed that HP1 (a reader binding to heterochromatin) recruits SUV39h1, a writer for the epigenetic mark H3K9me3 (methylation of histone H3, associated to gene silencing) [84]. Another example is that the Polycomb-Repressive-Complex PRC2 (a reader) contains the enhancer-of-zeste EZH2 (a writer) that spreads the epigenetic mark H3K27me3 also associated with gene silencing [85, 86].

Our second assumptions, regards the existence of genomic-bookmarking (GMB) mechanisms. The hypothesis, and the studies, on GBM come from the crucial property of epigenetic memory during mitosis. Indeed, it is not clear how, after mitosis, the new daughter cells faithfully reproduce the transcription program of their mother. One of the hypothesis, is that a mechanism concerning the inheritability of DNA-methylation, and the propagation of histone modifications can re-establish gene transcription after mitosis [87]. However, observations suggest that these two mechanism alone are not sufficient to maintain transcription profile in the cell life. For this reason, it has been suggested that some transcription factors, called GMB factors, can bind to mitotic DNA in their cognant site and remain dynamically associated with chromatin through mitosis. Example of GMB factors are Sox2 in mouse embryonic stem cells [88], or Polycomb-Group-Proteins (PGP) bound to Polycomb-Response-Elements (PRE) in *Drosophila* [89, 90].

Finally, the last assumption is that the recruitment of read-write machineries is

coupled to specific genomic-bookmarking binding. This means, that GMB actively participate to the positive-feedback loop relative to a certain, and precise, epigenetic mark, and reinforcing it. These three assumptions allow our model to reconcile short-term turnover of post-translational modifications with long-term epigenetic memory and plasticity.

## 2.1 A Polymer Model for Genomic Bookmarking

In this section we will try to understand the main features of the model described in section 1.4, when genomic bookmarking is considered too. As in the previous chapter, we model the chromatin fiber via a semi-flexible heteropolymer chain composed of $N$ beads. As already pointed out, we consider a three states model with interacting states A, B and unmarked, non-interacting, state U. In addition to this, we consider the possibility that a bookmarking transcription factor, which can either be an activator or a repressor [9], is bound to one bead with either an A or B mark; in this case we will indicate these particular beads with the states $A^\star$ or $B^\star$ respectively. The study of this model will be, again, carried out by employing numerical integration by velocity-Verlet scheme of equations (1.7), and a Metropolis-Hasting algorithm on Hamiltonian (1.13); the simulations have been performed thanks to the `LAMMPS` software [18].

We now discuss more in details how genomic-bookmarking has been introduced in the model. A complete model of how bookmarks guide the re-formation of the previous epigenetic pattern, and therefore of a coherent gene expression, after mitosis remains unknown [91]. Here, as already pointed out, we assume that the configurations and the cognate states of genomic-bookmarking factors depend on the particular tissue the cell depends on. Furthermore, we assume they remain dynamically associated to chromatin during mitosis [88, 89], and that when re-entering in the inter-phase of the cell life, they can recruit appropriate read/write machineries and help by rebooting the previous genomic expression.

In practice, in our polymer model, we account for bookmarking by assuming that beads with mark $A^\star$ and $B^\star$ can not change their mark (colour), and therefore they do not participate to the recolouring dynamics. This simple trick ensures that in the neighborhood of the bookmarked bead its colour is favoured by the recolouring dynamics, as highlighted by the Hamiltonian (1.13). Notice also that, in the basic model (see section 1.4), the choice of which mark dominates the local spreading is decided via $\mathbb{Z}_2$ symmetry breaking. Here, instead, by choosing a particular mark for our bookmarked beads, we locally break the symmetry and favor that mark over the others.

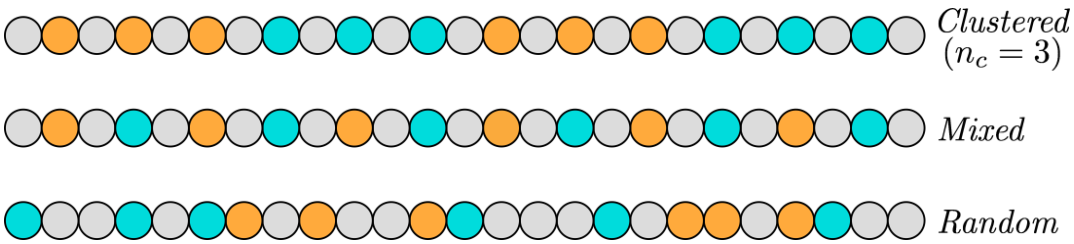



**Figure 2.1:** Pictorial illustration of the three bookmark configurations we employ in our numerical simulations: *clustered*, *mixed*, and random configurations; bookmarks are indicated in orange for the A$^\star$ mark, in cyan for the B$^\star$ mark.

### 2.1.1 Phenomenology

A complete study of the dynamics of the polymer model with genomic-bookmarking is much more complicated than in the basic case. Here other than the two interaction strengths $\alpha_L$ (attractive interaction), and $\alpha_R$ ("epigenetic" interaction), we have two more degrees of freedom: the number $M$ of genomic-bookmarks on a chain of size $N$, and the positional configuration of such bookmarks. For this reason, we will limit our study here to give a qualitative understanding of what happens when certain bookmark configurations are employed.

The considerations about the equilibrium regime made in section 1.4.1 are still valid here; namely, if we have the same attractive interaction $\alpha_L$ and epigenetic interaction $\alpha_R$ ($\alpha_L = \alpha_R$), we are at the equilibrium, otherwise ($\alpha_L \neq \alpha_R$) we are in an out-of-equilibrium regime. Indeed, using a "normal", non-bookmarked, bead, one can repeat exactly the same procedure to prove that detailed balance holds true if and only if $\alpha_L = \alpha_R$. On the other hand, a bookmarked bead participates only to the Langevin dynamics, and therefore, obeying equations (1.7) we are sure that its dynamics will be of equilibrium, as long the others will be themselves at equilibrium.

In order to qualitatively understand how the dynamics is influenced by the introduction of bookmarks, we set some numerical simulations of chains composed of $N = 1000$ beads and with bookmark density $\phi \equiv M/N = 0.1$. Qualitatively, the dynamics of this model at the extremes of the phase diagram that is really small (or big) values of the interactions $\alpha_L$, $\alpha_R$ won't differ from the basic model, and we will still observe phases similar to the one found in fig. 1.12. For this reason, rather than studying minutely the dynamics for every possible values of $(\alpha_L, \alpha_R)$, we will focus on some limited regions of the phase diagram. Namely, we are interested in the qualitative differences with the basic model at the equilibrium dynamics, and nearby the critical point ($\alpha_L = \alpha_R \approx 0.95$), in order to examine the dynamics of the new epigenetic domains. At the same time, we are notably interested in verifying that employing GMB mechanisms, the TADs observed in the swollen-ordered phase in the basic model will become stable, and retain long-term epigenetic memory.

Finally, we have to decide which bookmark positional organization will be employed. As there are infinite possibilities, we restrict the modelization to three cases (see also

fig. 2.1):

(i) *Clustered* configuration: bookmarked beads are equally spaced along the fibre; their mark alternates after every $n_c > 1$ consecutive bookmarks between $A^\star$ and $B^\star$. Note that $n_c$ defines the cluster size.

(ii) *Mixed* configuration: similarly to the clustered configuration bookmarks are equally spaced along the fibre, but now their mark is alternated every other bookmark ($n_c = 1$).

(iii) *Random* configuration: the position and mark of each bookmarked bead is taken from a random uniform distribution, while still keeping the constraint $\phi = 0.1$.

These configuration are studied by simulating the system at various interaction strengths $\alpha_R$ and $\alpha_L$, both at the equilibrium and the out-of-equilibrium regime. Results are shown in fig. 2.2, fig. 2.3, 2.4, and fig. 2.5. We want to examine the stability, and structure, of epigenomic domain that forms during the dynamics. Stability of the epigenomic domains is investigated by looking at the kymographs of the chains during their evolution. Chromatin architecture can be studied both looking at the snapshots of the dynamics, and constructing the so-called "contact maps" of the dynamics. A contact map, usually, shows the probability $P(i,j)$ for the $i$-th monomer, and the $j$-th monomer to be at a distance $r_{ij} = \|\mathbf{r}_i - \mathbf{r}_j\|$ shorter than a certain cutoff $\sigma_{\text{cut}}$. In our figures we plot two versions of the same contact map: on the lower side, in shades of grey, we have the contact probability $P(i,j)$ with cutoff $\sigma_{\text{cut}} = 8\sigma$; on the upper side, instead, we plot the value of $\tilde{P}(i,j)$ defined as :

$$\tilde{P}(i,j) = \begin{cases} P(i,j) & \text{if} \quad q_i = q_j = \text{A} \\ -P(i,j) & \text{if} \quad q_i = q_j = \text{B} \\ 0 & \text{otherwise} \end{cases} . \tag{2.1}$$

This version of the contact map helps us to keep track of the various, topological, genetic domains. Positive values of $\tilde{P}$ are represented in red, while negative in blue, in accordance with our color convention. Contact maps are numerically estimated by considering the last $10^5 \tau_{\text{Br}}$ of the dynamics.

### Clustered Configuration

The first bookmark configuration we consider is the clustered one. Namely, we study for completeness two cases: the case of small clusters $n_c = 10$, and big clusters, $n_c = 50$. Results can be found at fig. 2.2 and fig. 2.3. The system has been studied, in both cases, at equilibrium when $\alpha_L = \alpha_R = 0.98$, corresponding in the unbookmarked case to the compact-ordered (CO) phase; when $\alpha_L = \alpha_R = 0.92$, corresponding to the swollen-disordered (SD) phase; and at out-of-equilibrium when $\alpha_L = 0.6$ and $\alpha_R = 3$, corresponding to the swollen-ordered phase.

At $\alpha_L = \alpha_R = 0.98$, in both the case of $n_c = 10$ and $n_c = 50$, the system organizes itself in two compact globuli where red and blue domains separately coalesce in 3D

(macro-phase-separation), to give a checker-board appearance of the contact map, reminiscent of the pattern formed by A/B compartments in Hi-C maps [29]. The two globuli appear to be stable over long periods of time, in open contrast to what observed in the unbookmarked model (section 1.4.3). Finally, while this formation is reminiscent of TADs, we want to point out that the density contact-map show the existence of contacts, even if limited, between the two compartments, making the power of the analogy limited.

At $\alpha_L = \alpha_R = 0.92$, we had that the unbookmarked system would lie in the swollen-disordered phase. Here instead, in the case of $n_c = 50$, we find that the presence of bookmarks alter completely the dynamics and lead, eventually, to the formation of of two globuli and is, instead, akin to the CO phase, as observed in the previous case. When $n_c = 10$, instead, the system is indeed in an open configuration, but creates small globuli with coherent epigenetic mark. Unfortunately these architectures are rather noisy and unstable, and therefore not akin to TADs.

Finally, in the out-of-equilibrium regime, when $\alpha_L = 0.6$, $\alpha_R = 3$, we have, this time, the formation of stable, coherent, independent, architectures akin to TADs. In the $n_c = 10$ these architectures present in great number, and both in globular and open configurations. In the $n_c = 50$, instead, two globuli again form, but this time they are less "compact" (as shown by the compact map), and are separated by along strand of chromatin.

### Mixed Configuration

The second bookmark configuration we consider is the "mixed" one, always keeping a bookmark ratio $\phi = 0.1$; this configuration correspond to a clustered one with $n_c = 1$. We test this configuration at the same conditions and same values of $\alpha_R, \alpha_L$ of the clustered configuration. Results are shown in fig. 2.4.

Contrary to what observed in the clustered case, the mixing configuration seem to be disruptive to the formation of coherent epigenetic domains. At the equilibrium we observe a behaviour similar to what observed in the unbookmarked model: at $\alpha_L = \alpha_R = 0.98$ the system organizes in a single compact epigenetic domain characterised by a single color chosen via $\mathbb{Z}_2$ symmetry breaking; at $\alpha_L = \alpha_R = 0.92$ the system seems to be in a swollen-disordered phase. Out-of-equilibrium dynamics ($\alpha_L = 0.6$, $\alpha_R = 3$) show the formation of short-lived epigenetic domains, and the contact map do not seem to indicate the formation of underlining architectures. Compare this result to what observed in the unbookmarked case in the swollen-ordered phase: the presence of a mixed bookmark configuration upsets the formation of TAD-like structures.

The study of this configuration strongly suggests that the formation of long-lived, stable, epigenetic domains is strictly correlated with the configuration of the bookmarked beads. Namely, as the differences observed between the clustered and the mixed configurations point out, a single isolated bookmark is not enough to sustain an epigenetic domain (see next section for further discussions), while a "cluster" of successive, cooperating bookmarked beads of the same mark, can lead to the

construction of architectures akin to TADs.

### Random Configuration

The last typology of configuration we consider is the "random" one, with $\phi = 0.1$, where the position and the "color" of the bookmarked is chosen via a uniform random distribution. The configurations are studied at the same conditions and the same values of $\alpha_R$, $\alpha_L$ as in the previous cases. Results relative to a specific random configuration are show in fig. 2.5; in other random configurations studied the results were qualitatively similar.

At $\alpha_L = \alpha_R = 0.98$, similarly to what observed in the mixed configuration, the system coalesces in a single compact globuli characterised by a single epigenetic domain. Notice however, that the choice of the determinant epigenetic mark is not chosen here by a symmetry breaking (a random distribution is inherently non-symmetric), but rather from possible local (or global) in balances caused by the bookmark distribution. At $\alpha_L = \alpha_R = 0.92$ short-lived epigenetic domains arise, and the system present in a swollen-configuration. Finally, in the out-of-equilibrium regime ($\alpha_L = 0.6$, $\alpha_R = 3$), long-lived epigenetic domains akin to TADs form.

Note that this last, out-of-equilibrium, result is in open contrast with what found in equilibrium simulations with fixed epigenomic landscape [13, 92]. Notice also that these results are not in contrast with the considerations made in the mixed case. Indeed, random configurations permit the existence of consecutive bookmarked beads with the same mark, in fact they are rather common. More interestingly, the study of this ensemble of configurations tells us that there is no need for the location of bookmarks to follow strict rules for they to form stable architectures, but rather they have large degrees of freedom. This is important, as the effective bookmark configurations one could observe *in vivo* in a cell surely follow some kind of logic, but at a first glance they would probably look as a mid-case between the clustered configuration and the random one. However, thanks to this study, we are able to guarantee that even in a "more biological" configuration, the model is able to sustain the creation of stable, epigenetically coherent, domains.

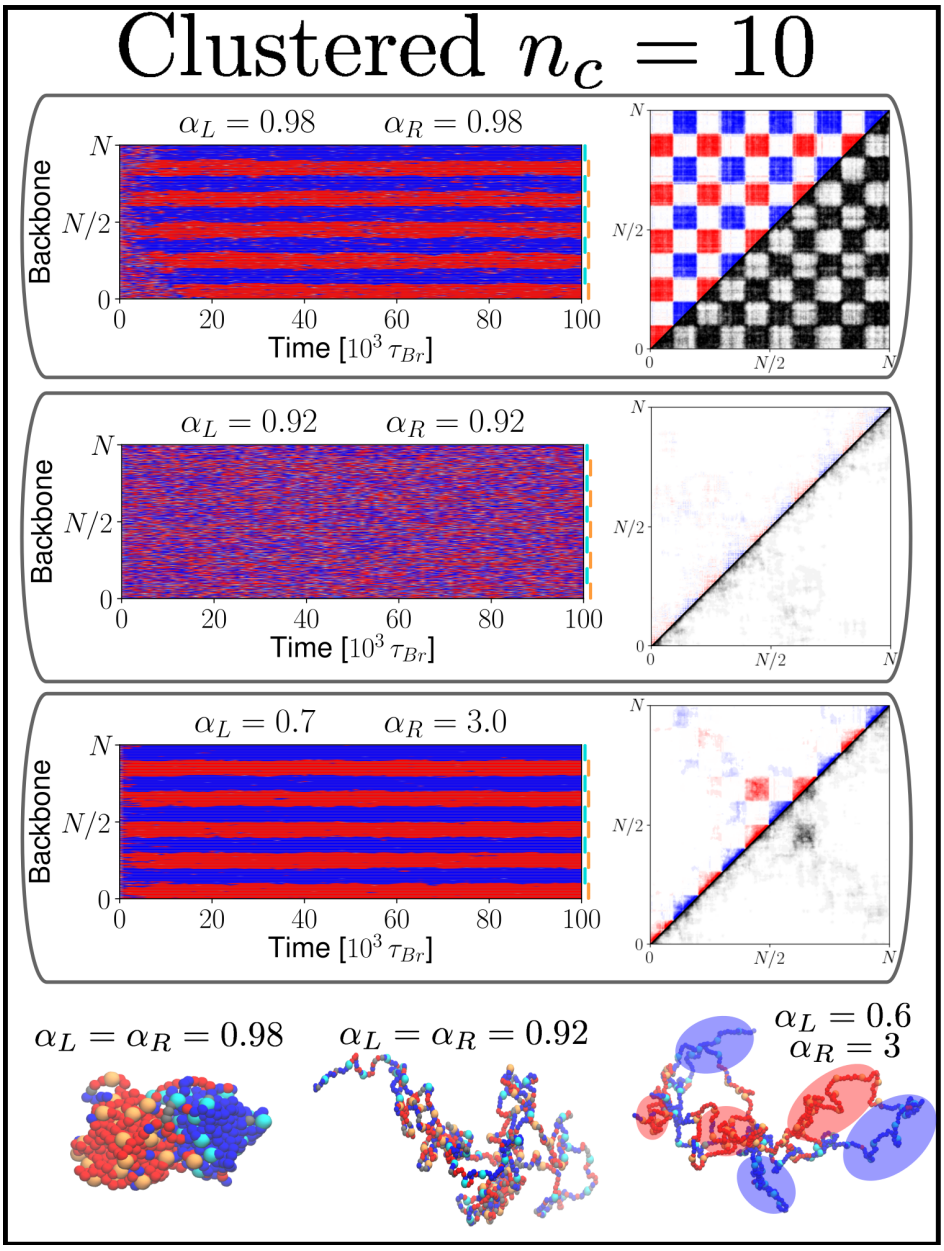



**Figure 2.2:** Results of numerical analyses for the bookmarked model in the clustering ($n_c = 10$) configuration. Results are shown both at equilibrium ($\alpha_L = \alpha_R$) and out-of equilibrium regime. Beside of kymographs colored dots are positioned: orange dots show the position of "red bookmarks", while cyan dots show the position of the "blue" ones. In the snapshots bookmark are highlighted using the same color pattern. In particular note that: if $\alpha_L = \alpha_R = 0.98$ we observe the formation of two stable globules akin to the CO phase of the previous chapter; if $\alpha_L = \alpha_R = 0.92$ as observed in the kymograph, all tentative of forming an ordered epigenetic phase are metastable and the configuration is akin to the SD phase; if $\alpha_L = 0.6$, and $\alpha_R = 3$, as observed in the kymograph and contact map, the system form stable epigenetically ordered formation akin to TADs well visible in the snapshot.

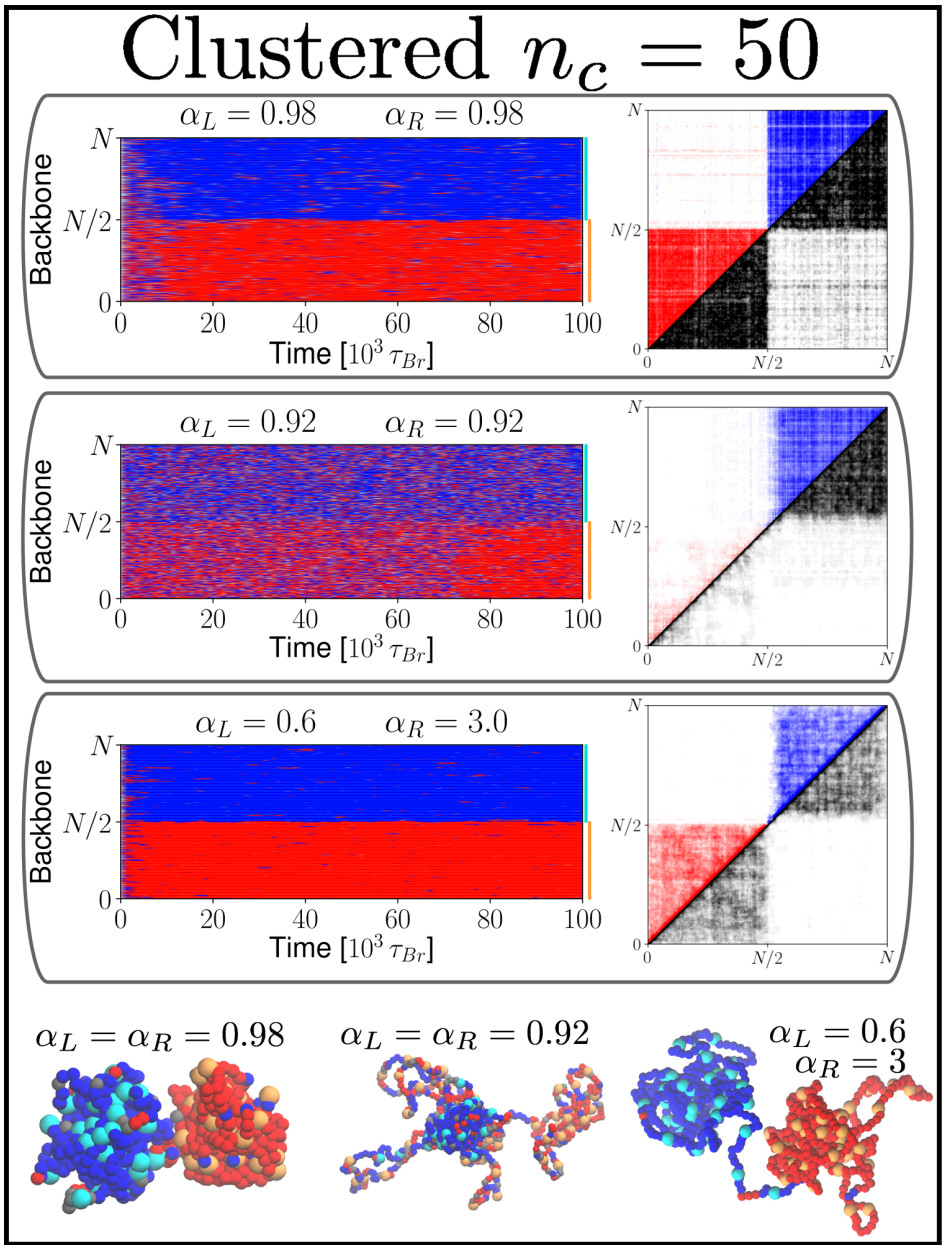



**Figure 2.3:** Results of numerical analyses for the bookmarked model in the clustering ($n_c = 50$) configuration. Results are shown both at equilibrium ($\alpha_L = \alpha_R$) and out-of equilibrium regime. Color conventions are the same as in fig. 2.2. In particular note that: if $\alpha_L = \alpha_R = 0.98$ we observe the formation of two stable globules akin to the CO phase of the previous chapter; if $\alpha_L = \alpha_R = 0.92$, both kymograph and contact map, suggest that if simulations time were longer we would have observed the same situation as in $\alpha_L = \alpha_R = 0.98$. Finally, in the out-of-equilibrium case ($\alpha_L = 0.6,\ \alpha_R = 3$) we again observe the formation of two compact globuli, but this time the density contact map (in grey) suggest that there are no contact between the two. This situation is akin to what observed in two repressed, consecutive, TADs.

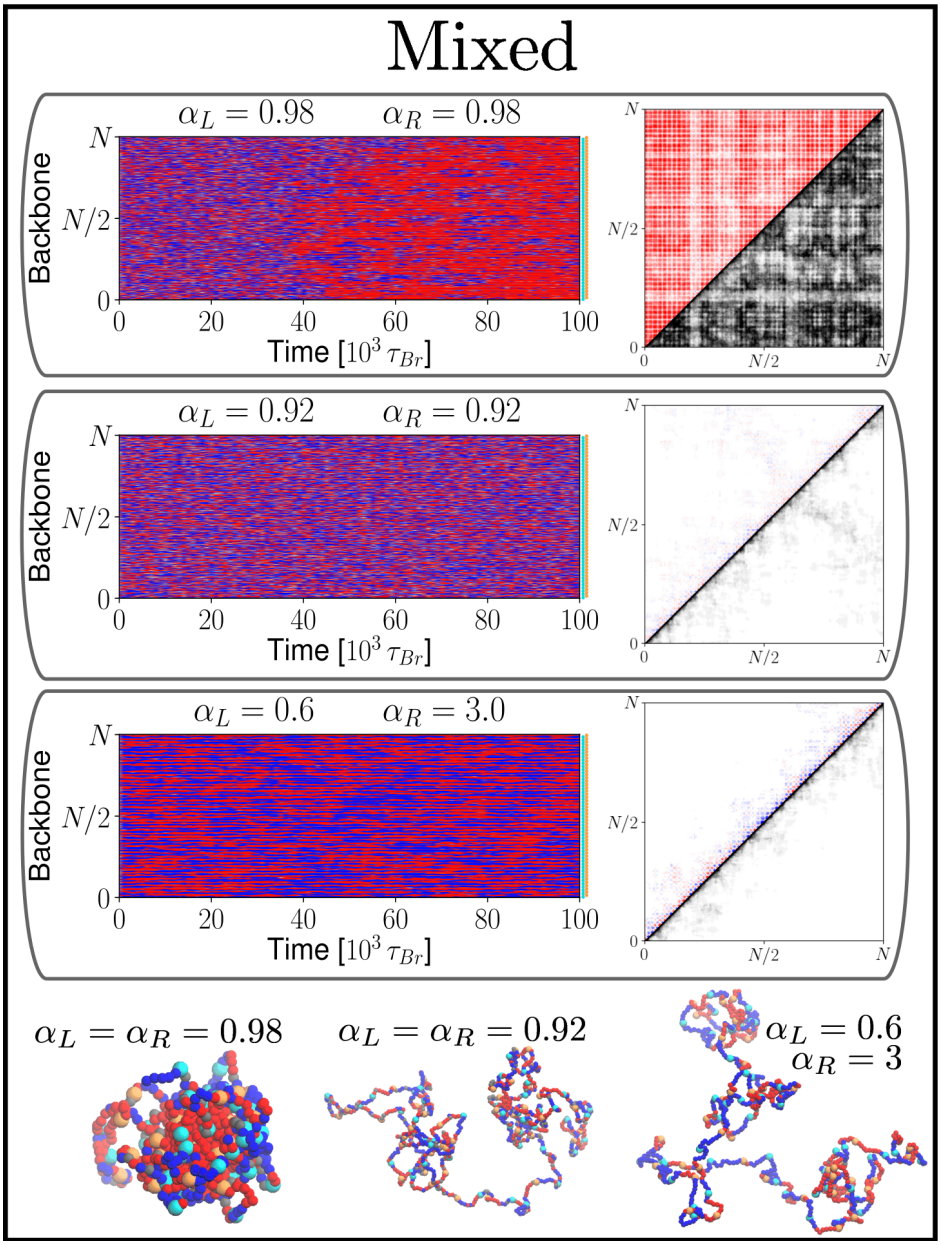



**Figure 2.4:** Results of numerical analyses for the bookmarked model in the mixed configuration. Results are shown both at equilibrium ($\alpha_L = \alpha_R$) and out-of equilibrium regime. Color conventions are the same as in fig. 2.2. In particular note that: if $\alpha_L = \alpha_R = 0.98$ we the system organizes in a single globule (CO phase) whose colour is decided by $\mathbb{Z}_2$ symmetry breaking; if $\alpha_L = \alpha_R = 0.92$, no spatial or epigenetic order is observed (SD phase). Finally, in the out-of-equilibrium case ($\alpha_L = 0.6$, $\alpha_R = 3$) small ordered epigenetic domains appear, however they are short-lived and, therefore, not comparable to TADs.

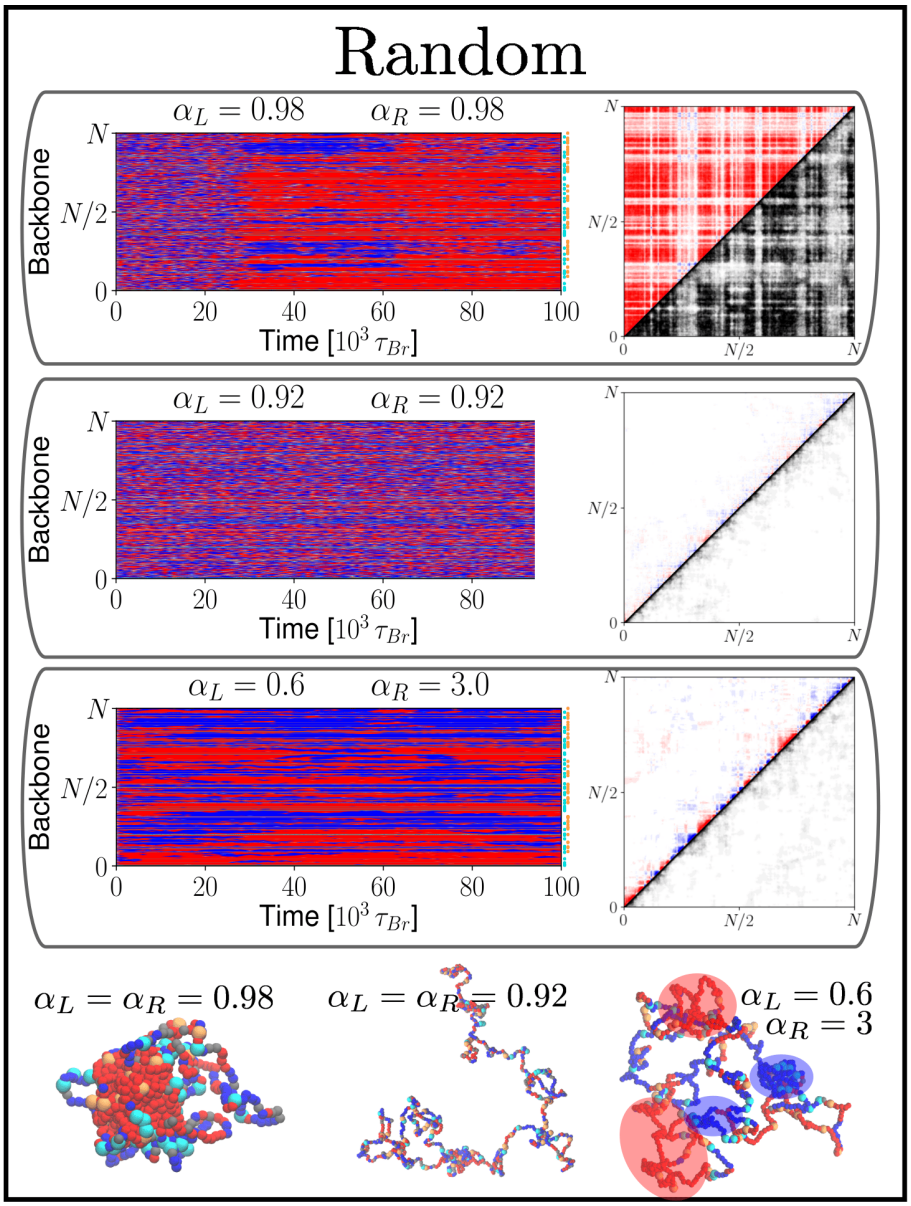



**Figure 2.5:** Results of numerical analyses for the bookmarked model in a random configuration. Results are shown both at equilibrium ($\alpha_L = \alpha_R$) and out-of equilibrium regime. Color conventions are the same as in fig. 2.2. In particular note that: if $\alpha_L = \alpha_R = 0.98$ we the system organizes in a single globule (CO phase) whose colour is decided by $\mathbb{Z}_2$ symmetry breaking; if $\alpha_L = \alpha_R = 0.92$, no spatial or epigenetic order is observed (SD phase). Finally, in the out-of-equilibrium case ($\alpha_L = 0.6,\ \alpha_R = 3$) ordered and stable epigenetic, akin to TADs, appear.

### 2.1.2 Critical Bookmarking Density

In the previous section we noted that the mixed bookmarking configuration show that the equilibrium behaviour similar to what observed in unbookmarked chains, failing to create a number of stable and coherent epigenetic domains of different marks. This, together with the observations regarding other configurations, leads to the conclusion that a minimum local bookmarking density $\phi$ is required in order to coadiuvate the formation of TAD-like structures.

To address this question, we employ some numerical simulations over a chain of size $L = 1000$ where we systematically vary the bookmarking density. As bookmarking configuration, we choose the clustered one from the previous section, as we verified it is the most efficient when creating (or maintaining) epigenetic domains. Here, $\phi$ varies from $phi = 0$ (unbookmarked) to $\phi = 0.1$; in order to simplify the analysis we focus on the case where each epigenetic domain has a size of 100 beads, i.e. we choose $n_c = 100\phi$. Notice that a domain of size 100 beads correspond to around $3 \times 10^5$ DNA bases (as seen in section 1.4), which is in the range of typical domains found with Hi-C techniques [29, 93]. To facilitate, again, the analysis we set our simulation at the equilibrium when $\alpha_L = \alpha_R = 0.98$ which corresponds to a compact-ordered phase; indeed, as we observed in the previous section, the eventual epigenetic domains are much more distinguishable in this phase of the dynamics. Finally, as initial conditions for our simulation we set a typical equilibrium configuration of the $n_c = 10$ clustered configuration at the same strength values (see fig. 2.2) where the system is already organized in two separate globuli, in order to monitor better the behaviour of the epigenetic domains when some bookmarks are being eliminated. Notice, however, that we have equilibrium conditions, and therefore the final steady state of the dynamics do not depend on the actual initial configurations.

In order to monitor the dynamics, we define here a new order parameter: the "fidelity" $\chi$ which measures how well the epigenetic domains are maintained. First, let us define $P(i)$ as the probability for the $i$-th monomer of a chain to be "red"; we compare this probability with the ideal result $\Pi(i)$ that is the result we would have if the domains were perfectly maintainers and can be defined as:

$$\Pi(i) = \frac{1}{2}\left[\operatorname{sgn}\left(\sin\left(\frac{\pi i}{n_d}\right)\right) + 1\right] = \begin{cases} 1 & \text{if } i \text{ in red domain} \\ 0 & \text{otherwise} \end{cases} , \tag{2.2}$$

where $n_d$ is the number of beads in a domain ($n_d = 100$ in our case). Defined the two probabilities $P(i)$ and $\Pi(i)$, we define their "average distance" $\Delta^2$ as:

$$\Delta^2 = \frac{1}{L}\sum_{i=1}^{L} (P(i) - \Pi(i))^2 \quad . \tag{2.3}$$

The fidelity parameter $\chi$ is then defined as $\chi = 1 - \Delta^2$; notice that $\chi \approx 1$ only if $P(i) \approx \Pi(i)$ i.e. there is ideal block formation, while $\chi \approx \frac{1}{2}$ if the domain coalesces into a single domain as in the unbookmarked case.

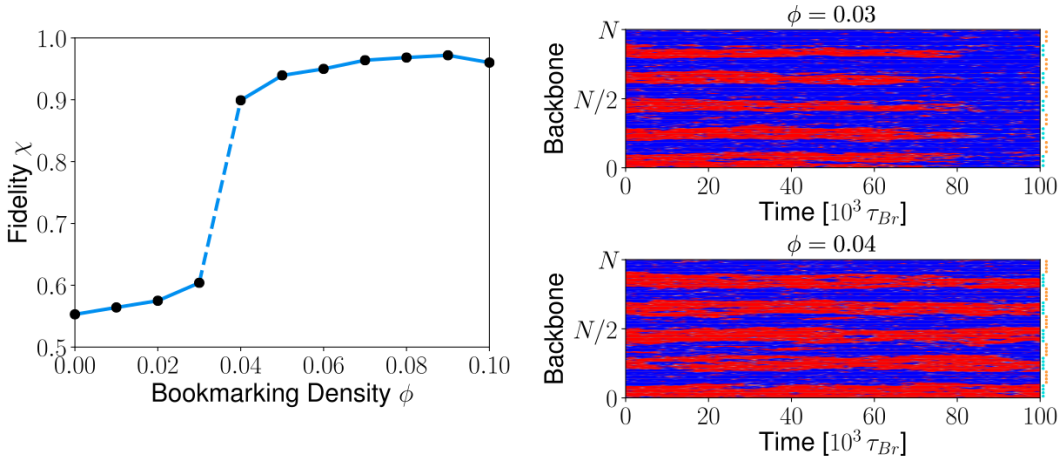



**Figure 2.6:** **(A)** Fidelity $\chi = 1 - \Delta^2$ as a function of the bookmarking density $\phi$. Note the abrupt jump from a single-globule "phase" ($\chi \approx 0.5$) to a two-globuli phase ($\chi \approx 1$ at $\phi_c = 0.04$. To estimate the parameter, we employ numerical simulation of chains with $L = 100$ at $\alpha_L = \alpha_R = 0.98$; as initial configuration of the dynamics we employed the last configuration found in fig. 2.2 when $\alpha_L = \alpha_R = 0.98$. We employed a clustered bookmark configuration with $n_c = 100\phi$. Parameter $\Delta$ is estimated via equation (2.3) and $P(i)$ is estimated by considering the last $10^5$ steps of the dynamics. **(B-C)** Kymographs of the dynamics when $\phi = 0.03$ and $\phi = 0.04$. In first case we observe that the initial epigenetic domains slowly shrink in size until they disappear and the dynamic is overcome by a single mark; in the second case, we observe that the system, although noisy, can mmaintain the epigenetic domains for a long time.

Results are shown in fig. 2.6. We observe that there the fidelity $\chi$ show a behaviour similar to a phase transition, with critical density $\phi_c \approx 0.04$. For low density $\phi < \phi_c$, a single mark takes over the dynamics and a single compact-globuli is formed. For higher densities ($\phi \geq \phi_c$), instead, we have a high fidelity $\chi \approx 1$, and the chains remain organized in two globuli of different epigenomic marks.

The estimated critical bookmark density value $\phi_c \approx 0.04$ is quite realistic, since it corresponds to about 1 or 10 nucleosomes in about 400, as not all nucleosomes coarse-grained in a "bookmark bead" need to be bookmarked. Therefore, we argue that, crucially, not all the genome must have this critical density of bookmarks, but only regions required to robustly develop a specific domain of coherent histone marks in a given cell-line. Indeed, notice that a similar behaviour was observed in the random configuration (see fig. 2.5) when we considered an out of equilibrium dynamics: not all the genome organized in TAD-like structures, but only small regions with higher density of coherent bookmark beads.

## 2.2 Towards a more Realistic Model

In this section we expand our model to describe more realistic, and biologically relevant, scenarios. Namely, we will try and reproduce the main features of the epigenetic landscape found *in vivo* in *Drosophila* cells. To do so, we have to carry out several modification to the model, comprehensive of relative interaction strengths among the various states, and the number of states itself, while at the same time introducing new out-of-equilibrium features. Results will show that our model predicts a "ChIP-seq", which maps global binding sites for the transcription factors on the DNA, compatible with the one found *in vivo*.

### 2.2.1 Asymmetric Interactions among Epigenomic Marks

In chapter 1 and the previous sections, we have studied the dynamics of epigenomic spreading when two competing marks A and B are present, without giving them a particular biological identity. Also, we always assumed that the two marks behave exactly in the same way: this feature was actually important for our preliminary studies as it permits us to study both the dynamics of the spreading in presence of competition, and the presence of phase transitions due to $\mathbb{Z}_2$ symmetry breaking in the "epigenomic magnetisation".

Let us now consider an "asymmetric ,model" where the blue-blue interaction, and red-red interaction, are different. In fact, generally, the reader-writer machinerie that starts the positive feedback loop is unique for each epigenetic mark, and therefore, generally, different proteins have different binding energies. Now, in order to construct a more realistic model, we choose to study the dynamics of two common marks, from now on: red marks will encode Polycomb repression; blue marks encode constitutive heterochromatin.

We suppose that blue-blue interaction is larger than the red-red one; in this case at the equilibrium the thermodynamic symmetry of the system is broken and the blue mark eventually takes over all non- bookmarked regions. However, if there are bookmarks for the red mark, they will locally favour the red state, and stable red domains may form (see fig. 2.7A). We test this hypothesis via some numerical simulations of a chain of $L = 1000$ beads were we set $\alpha_L^{\text{blue}} = 1.05\alpha_L^{\text{red}}$ and $\alpha_R^{\text{blue}} = 1.05\alpha_R^{\text{red}}$; we insert $n_B = 10$ bookmarks in a central region of the chain of size 100 beads such that the local bookmark density is $\phi = 0.1$. We study the system when $\alpha_L^{\text{red}} = 0.6$, $\alpha_R^{\text{red}} = 3$, because, as we have noticed in the previous section, when this parameters are employed, there are more possibilities of observing TAD-like behaviours. Starting from a swollen and, epigenetically disordered fibre, where red, blue and grey beads are equal in number, we observe that blue marks quickly invade all the non-bookmarked regions and convert red beads into blue ones, mimicking what observed during the heterochromatin spreading [84], but the central segment, aided by the presence of red bookmarks, remains stable and red.

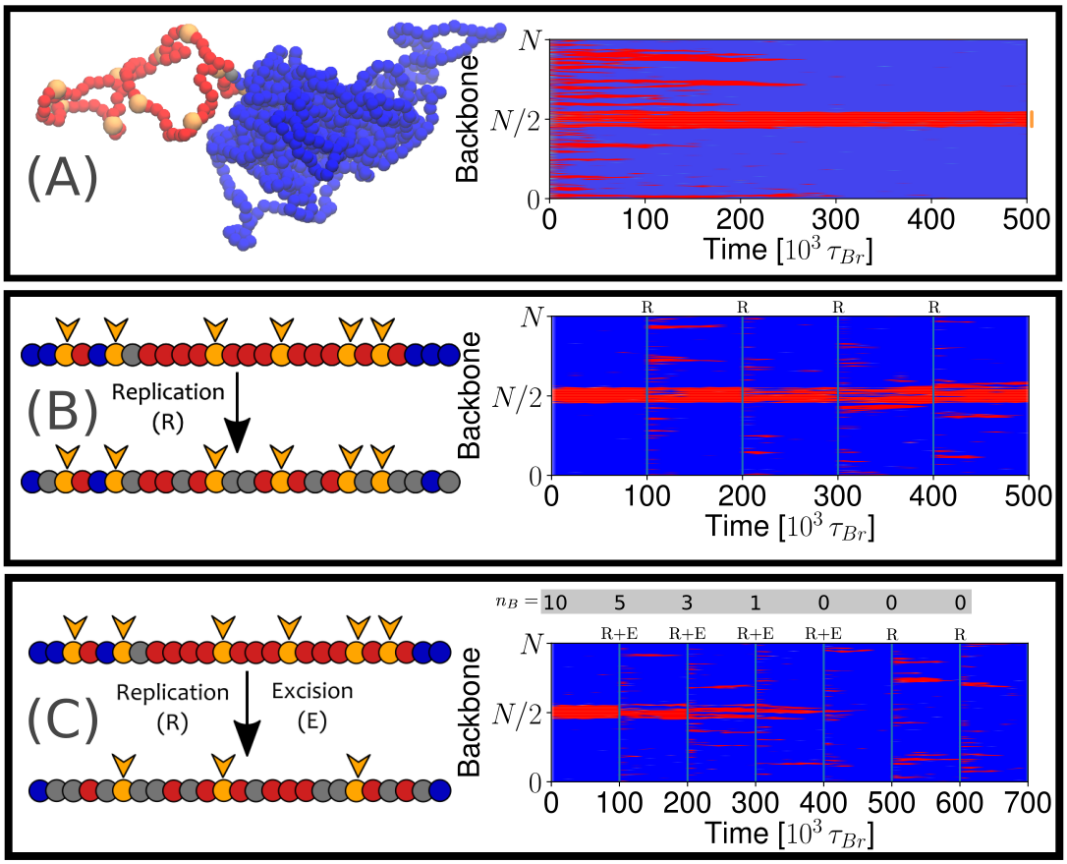



**Figure 2.7:** **(A)** Snapshot of a typical configuration, and kymograph of the dynamics of a chain with $L = 1000$ beads with asymmetric dynamics where $\alpha_L^{\text{blue}}/\alpha_L^{\text{red}} = \alpha_R^{\text{blue}}/\alpha_R^{\text{red}} = 1.05$, and $\alpha_L^{\text{red}} = 0.6$, $\alpha_R^{\text{red}} = 3.0$. $n_b = 10$ bookmarks (orange beads) are positioned in the middle of the chain such that we have a local density $\phi = 0.1$. Kymograph shows that the favored blue mark spreads into all non-bookmarked areas, while at the same time the central red-bookmarked area remains stable. **(B)** On the left: exemplification of a conservative cell replication process: the epigenetic landscape is modified by a random process where each non-bookmarked bead has $1/2$ probability of becoming grey. On the right a typical kymograph of this dynamics: replication process happen every $10^5\tau_{Br}$; dynamical parameters are the same in A; we employed the last configuration of A as the initial configuration of this run. Notice that while the replication process introduces a strong noise to the process, the central, bookmarked, red-domain remain stable. **(C)** Same as B, but here we also introduce excision of bookmarks; at each replication step, we remove about half of the bookmarks (number of bookmarks $n_b$ is above the kymograph in the grey panel). Notice that the red domain remains stable for a long time, even when the bookmark density drops below the critical value $\phi_c \approx 0.04$ (see section 2.1.2).

### 2.2.2 Cell Replication Process

The second thing we have to do to proceed in the direction of a more realistic, biological, model is to test whether or not, the epigenetic patterns established via genomic-bookmarking are stable against strong perturbations. This is important, because during mitosis, not all the epigenetic marks of the mother cells are kept in the daughter cells, resulting, at the beginning of the daughter cell life, in a vastly different epigenetic landscape.

We introduce in the model a new random process every $10^5\tau_{Br}$ time-steps where each bead has 1/2 probability of being recoloured into a grey bead. This is to mimic what happens during a semi-conservative replication, while however keeping the spatial architecture unchanged [94]. We study the dynamics of this new process in the same asymmetric conditions as in section 2.2.1, but we start our numerical simulations from a stable configuration as in fig. 2.7A. Results (see fig. 2.7B) show that our model is robust against extensive perturbations, and that already-established patterns formed via genomic-bookmarking remains stable against various rounds of cell replication. Importantly, the combination of epigenetic memory we observed in section 1.4, and genomic-bookmarking allows the cell to propagate after mitosis both "global" and "local" epigenetic patterns.

Another observation coming from the biology is that, even with a low density of transcription factors, epigenetic memory do not vanishes immediately, but rather deteriorates gradually. Indeed, let us consider some experiments [95], where the role of some transcription factors akin to bookmarks, called Polycomb-response-elements (PREs), in epigenetic memory was investigated. In this experiment, the Polycomb-mediated gene repression was perturbed by artificially inserting or deleting a certain quantity of PREs. PREs are correlated with the presence of regions of the genome marked with H3K27me3; excision of these factors result in a slow, gradual, decreasing of the region dominated by H3K27me3. We observed a similar phenomenon in fig. 2.6: even when the bookmark density was below the critical value $\phi_c \approx 0.04$ (e.g. $\phi = 0.03$), the red domains persisted for a long time until, eventually, they fade away.

Let us now test this claim in the asymmetric case, and in the presence of a replication process. Namely, in fig. 2.7C we show the results of a simulation where to each replication we accompanied by the random excision of about half of the bookmarks. At each cell cycle, we dilute the number of bookmarks, until, at some point, they can no longer sustain the local red states, and the region is consequently "invaded" by the favored blue mark. However, just as in the symmetric equilibrium case of fig. 2.6, the local epigenetic memory of the red domain seems to persist even when the bookmark ratio is lower than the critical value, in accordance with the experimental results.

Finally, we want to remark that the basic results of these two sections show that our model, while simplistic, is a useful stepping stone for a more complex, realistic, biological model, as we will see in the next section, where we will deal with the modeling of the chromosome of a *Drosophila*'.

### 2.2.3 Modelling of a *Drosophila* Chromosome

After we studied how our model deals with asymmetric interactions, and verified that gives realistic results when replication/excision processes are introduced in the system, we are finally ready to challenge a more difficult, realistic, scenario. Namely, we will perform some numerical simulations to produce the spreading of certain epigenetic marks on the right arm of chromosome 3 of a *Drosophila* S2 cell. We are particularly interested in studying the competition of the two repressed states found in these chromosomes: one mediated by Polycomb proteins, the other representing the classic heterochromatin.

#### Possible Epigenetic States

As a first step, we must set the parameters of our simulations, most important of all is the number of "states" a bead can adopt; we base our assumptions on the numerical analysis of [96]. Here, thanks to chromatin immunoprecipitation (ChIP) techniques [97], which is a procedure used to determine whether a given protein binds to or is localized to a specific DNA sequence *in vivo*, they study the localization of 18 different histone modifications. Thanks to a machine-learning approach, they reduce this 18-states model, to have only 9 "essential" states (generally a combination of the original 18 modifications) and proceed to study their features always while keeping the other states hidden (Hidden Markov Model). Not all the states are present in all the chromosomes, namely state 5 and state 8 are associated only to certain ones and notably to the X chromosome. We will therefore consider:

**State 1.** It is associated to active promoters and enhancers; we will represent this state via red beads.

**States 2-3-4.** All of these states are found in corrispondence of actively transcribing regions; as we are not interested in the details of their dynamics we will consider these states as a single "transcribed" state represented in green.

**State 6.** It is associated with Polycomb repressed regions rich in the H3K27me3 mark; they are represented in purple.

**State 7.** Heterochromatin (HET) rich in the H3K9me3 mark; represented in blue.

**State 9.** Nucleosomes in this state do not appear in ChIP sequencing and are generally silent during transcription; regions rich with this kind of genes are called *gene deserts* and constitute around 25% of the whole genome. We represent this state in grey.

To these 5 states found in [96], we add an unmarked state (represented in white), and we complete the model by adding the bonding sites (represented in orange) of the PSC/PRE transcription factors which act as bookmarks for the Polycomb-mediated marks.

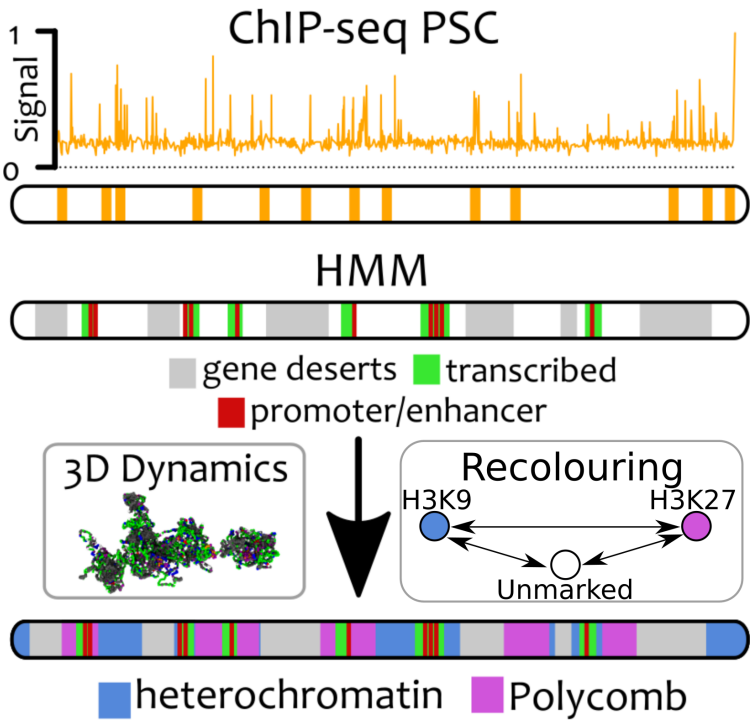



**Figure 2.8:** Pictorial illustration of the spatial positioning of the non-recolouring states. The location of PSC/PRE (bookmarks, in orange) are mapped onto beads using ChIP-Seq data from [89]. The other states position (gene deserts in grey, transcribed genes in green, promoters in red) are taken from [96], The remaining beads ($\sim 20\%$) are initially unmarked (white). When we start the recolouring machine, coupled with the 3D dynamics, the unmarked beads will eventually recolour and become either heterochromatin (blue) or Polycomb (purple).

### Aim, and Initialization, of the Model

After we have decided how many, and what, states the beads of our chain can have, we must clarify our intent in order to set our numerical simulations. We want to verify that the tools we have developed by studying genomic-bookmarking can be applied in realistic situations. Notably, these tools are better applied to repressive states, as repressed regions tend to appear in compact configurations *in vivo*, compatibly with our results. For this reason, we choose to focus our study on the dynamics of the two repressed states: Polycomb-mediated (state 6), and HET (heterochromatin, state 7). Namely, as initial configuration of our chain we will fix the position of all the other states (states 1-2-3-4-9), and of the bookmarked beads mediated by PSC. The remaining chromatin fiber (about 20% of the total) will be left unmarked in white (see fig. 2.8). During our simulations the unmarked states will be recoloured in blue (heterochromatin) or purple (Polycomb) using the same recolouring dynamics we have always employed in the previous sections, while all the other states will remain

fixed. At the end of our simulations, we will compare the distribution (represented by a ChIP-sequence *in silico*) of Polycomb-mediated mark H3K27me3, with actual experimental results.

Finally we have to model the interactions among the various states (see table 2.1). Namely, unmarked states experience steric interactions with the others; Polycomb and heterochromatin have asymmetric interaction and HET is slightly favored, but, at the same time, we increase the affinity with between Polycomb and their bookmarks PSC. We further introduce a strong interaction between promoters, and an interaction between promoter beads and transcribed genes to favour formation of loops observed in transcribed area of the genome [98]. Finally, an attractive interaction between gene desert beads is considered, in order to mimic their compaction by H1 linker histone [99].

| | Polycomb | HET | PSC | gene desert | promoter | transcribed |
|---|---|---|---|---|---|---|
| Polycomb | 0.6 | 0 | 0.9 | 0 | 0 | 0 |
| HET | 0 | 0.72 | 0 | 0 | 0 | 0 |
| PSC | 0.9 | 0 | 0.9 | 0 | 0 | 0 |
| gene desert | 0 | 0 | 0 | 0.6 | 0 | 0 |
| promoter | 0 | 0 | 0 | 0 | 3.0 | 0.6 |
| transcribed | 0 | 0 | 0 | 0 | 0.6 | 0.6 |

**Table 2.1:** Values of the interaction strengths $\alpha$ (see (1.7) and (1.13)) between the various states a bead can assume.

### Results

We consider a chain of $L = 9037$ beads, colour with initial Configuration for the states 1-2-3-4-9 taken from [96], while we map the position of PSC via ChIP-sequence in [89]. The remaining chromatin is left unmarked.

We evolve the system to the steady state, and we estimate the ChIP-sequence for the H3K27me3 mark as the probability of finding a Polycomb mark at a certain genomic position. Bookmarked beads are accounted in this computation only if they are in the neighborhood of a non-bookmarked bead with such a mark. Thanks to an ensemble of independent simulation, we are able to provide an "*in silico*" ChIP-sequence track for Polycomb marks which can be compared with *in vivo* ChIP-sequence data [96] (see fig. 2.9B). We compute the Pearson correlation coefficient $\rho$ to compare the results of our model with the experimental results. Notably, we obtain $\rho = 0.46$, which is to be compared against a $\rho^\star = 0.006$ we would obtain with a random data-set, and conclude that the two ChIP-sequences are in excellent agreement.

Notice from fig. 2.9C that not all bookmarked segments (in orange) are populated by Polycomb marks (in purple). What we observe, instead, is that the spreading of the Polycomb marks require an appropriate 3D structure. A single bookmark, as observed in section 2.2.2 and section 2.1.2, is not sufficient to create a coherent epigenetic domain, but rather it needs a "favorable" epigenetic landscape and contacts

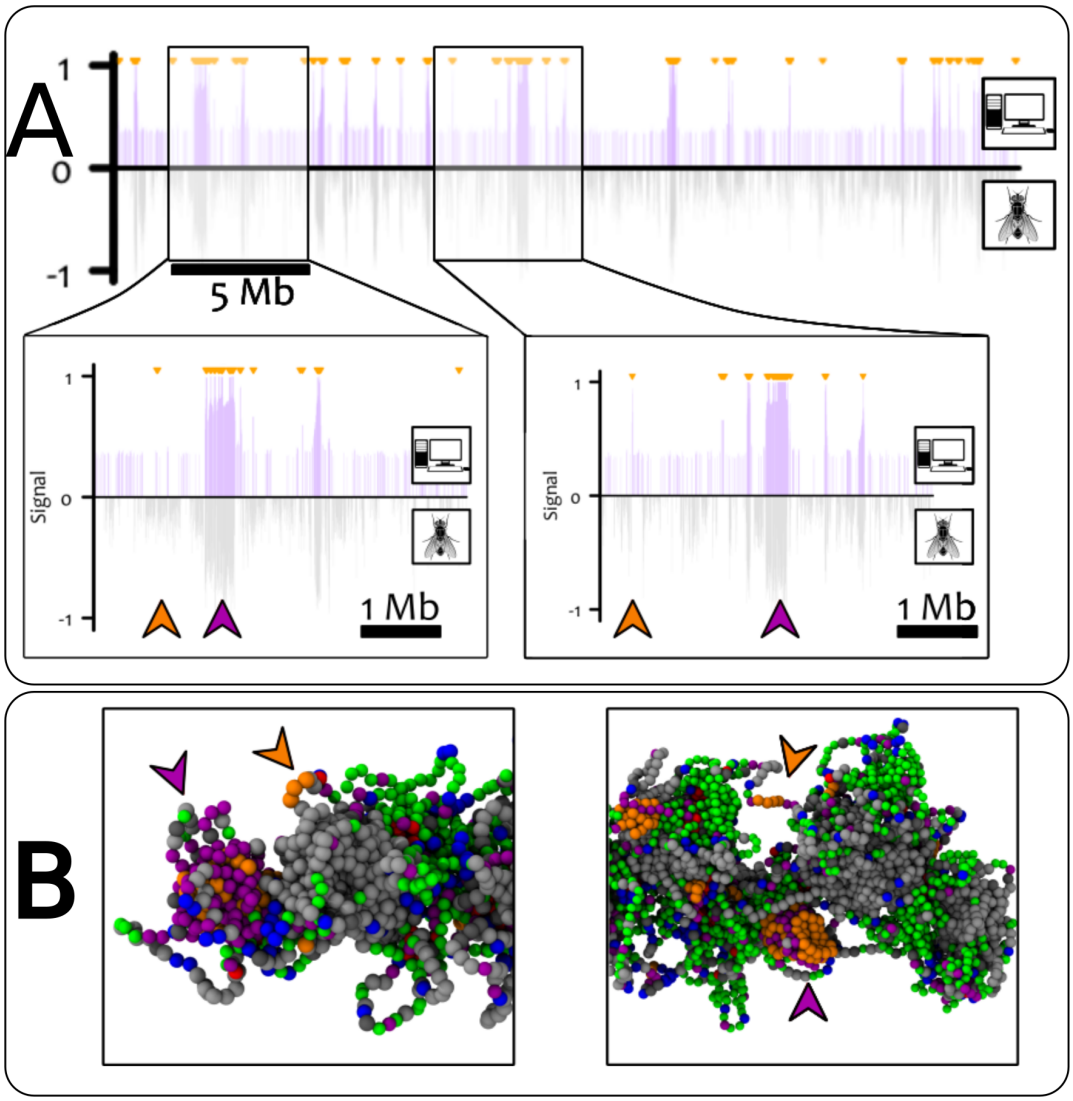



**Figure 2.9:** (**A**) Comparison between *in silico* (purple) and *in vivo* (grey) ChIP-sequences of the Polycomb-mediated mark H3K27me3. On the top, small orange arrows indicate the PRE/PSC (bookmarks) positions. Results show great agreement between the two sequences. (**B**) Snapshots of the dynamics. Notice that presence of various bookmarks favor the formation of H3K27me3 domains (indicated by the big purple arrow), while a single bookmark, if not coadiuvated by the surrounding epigenetic landscape (see big orange arrow) is not enough to guarantee the formation of coherent domains.

with other bookmarks. Again, this is consistent with 3D chromatin conformation being crucial for the spreading and establishment of epigenetic patterns.

## 2.3 Discussion

In this chapter we have discussed a possible extension to the simple model presented in section 1.4 to reproduce the local epigenetic order observed in real *in vivo* chromosomes. The first model predicted that a single epigenetic mark eventually spreads on all the chromatin, but in real chromosomes distinct stable epigenetic domains, characterised by different marks, coexist at the same time. We show that this feature can be reproduced by introducing genomic bookmarking in our model.

Bookmarkers are transcription factors which recruit the reader/writers machinerie of a specific epigenetic marks, and that we suppose permanently bound to DNA. In the model, bookmarkers-bound chromatin is modeled by a bead bearing a specific mark that does not partecipate to the recolouring dynamics, hence enhancing the probability of spreading its mark on the other non-bookmarked beads.

Our results strongly suggest that a critical density of bookmarks of a certain mark, can form stable epigenetic domains of that same mark by spreading on its neghbouring beads. Its dynamics is stopped either by thermal noise, or by the competition with other bookmarks of different mark. Notably, our model does not require other external factors to reproduce boundaries between the domains, as the spreading results to be self-regulated by the bookmarks configuration.

Epigenetic spreading is also regulated when we have a single type of bookmark, but the other epigenetic marks are thermodynamically favored. This is in accordance with experimental observations where heterochromatin spreads on large domains, until it reaches actively transcribing areas. We have also shown that the bookmarked domains are robust against large perturbations of the system, e.g. what observed during mitosis. The domains result to be stable for several cell cycles even if we suppose that during mitosis a number of bookmarks is excised, and remain stable whenever the bookmark density remains above a certain critical value. Notably, this means that the model permits the formation of *de novo* epigenetic domains, in response to external stimuli, by removing bookmarks, or inserting newly activated ones.

Finally, we have tested our model in a realistic problem, by recreating the distribution of the H3K27me3 mark on a *Drosophila* chromosome, by starting from the position of PSC proteins which act as Polycomb bookmarks. The results of the simulations are in great agreement with experimental data. Our simulations show also that not all bookmarks end up in H3K27me3 domains, but rather it depends on the 3D organization of the chromatin, in agreement with recent experiments.

While the model satisfyingly reproduces realistic epigenetic landscapes, it still has room for improvement. For example, we have shown that the model can simulate the dynamics of repressive states, but it is not probably suited for the study of actively transcribing chromatin. In [2], D. Michieletto shows that active marks may be better modelled as resulting from a co-transcriptional positive feedback loop, where promoters are interpreted as bookmarks for active marks. With this new extension, it is possible in theory to simulate even more complex realistic scenarios. For example, it

would be interesting to employ the model to estimate the transcription rate of specific genes by measuring in the simulations for what fraction of the dynamics the beads, corresponding to those genes, bear the active mark and are spatially located nearby a promoter. Results of this kind of simulations would allow a direct comparison with RNA-seq experiments, and, possibly, to a fine tuning of all the parameters employed in our model.

Our model predicts the epigenetic landscape on the chromatin given the distribution of the bookmarks, but does not say anything about the dynamics of the bookmarks themselves. Indeed, it is not clear how starting from a stem cell, several cell lineages are established: how do bookmarking transcription factors know the whereabouts of their cognate site? It is possible that, due to morphological and/or environmental cues, the production of specific boookmarking factors is triggered and interact with a pre-existing bookmark pattern. However, all this, probably, masks the presence of new biophysical mechanisms which necessitate more discussion both on the biological and physical front.

# Chapter 3

# Mean-Field Theory of Magnetic Polymers

*You don't have to test everything to destruction just to see if you made it right.*

---

Neil Gaiman, Terry Pratchett, `Good Omens: The Nice and Accurate Prophecies of Agnes Nutter, Witch`

In chapter 1 (section 1.4) we have introduce a novel model to study the coupling of chromatin folding with epigenomic spreading in the nucleus of eukaryotes. There, we model the chromatin fiber as a semi-flexible, spring-and-beads, polymer. To each bead of the chain we assign a "colour" which represents the epigenetic states of the nucleosome. The spatial dynamic of the chain obeys a set of Langevin equations (1.7), while the "recoloring" dynamics is described by a Potts-like Hamiltonian eq. (1.13). Thus far (chapter 1 and chapter 2), we analyzed this model by employing a set of numerical simulations of the system. While this method is powerful and it allows a in-depth qualitative, and quantitative, study of the system, it gives no certain answer regarding certain fundamental questions in statistical mechanics. For example, we have observed that at the equilibrium (section 1.4.3) there is a transition between a swollen-disordered phase to a compact-ordered one; while we have strong evidences that there is indeed a phase transtion, and it is of the first order, numerical simulations alone cannot be conclusive.

In this chapter, we focus on field-theoretic approaches to study analytical and numerically solvable models of magnetic polymers following the work of [3]. Namely, we will map the chromatin model developed in chapter 1 into a lattice model, where we can develop a mean-field Landau-Ginzburg-like theory at equilibrium. Analytical results of our polymer models, will be compared with equivalent Montecarlo simulations

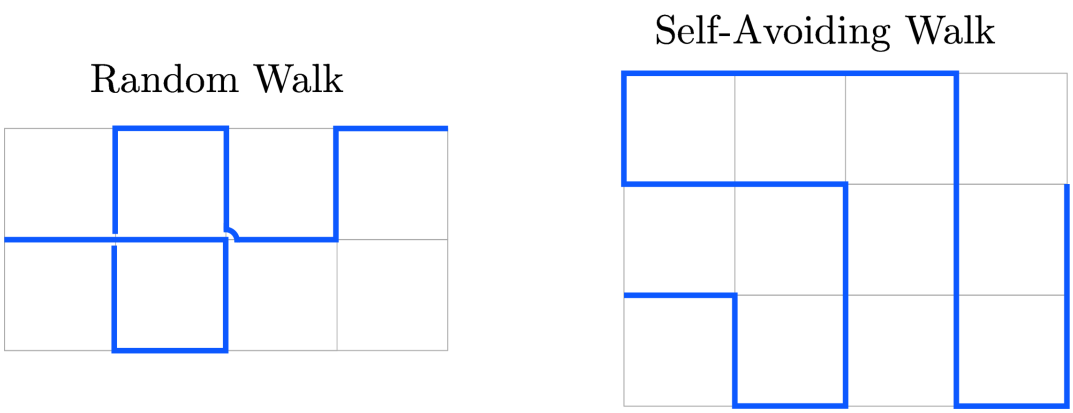



**Figure 3.1:** Pictorial compariosn of a classic random walm (on the left) and a self-avoiding random walk (on the right) on a square 2D lattice. Differently from a non-constrained random walk, in the self-avoiding walk (SAW) the path cannot pass two times from the same node.

in order to infer the effectiveness, and the limits, of a mean-field approach to this kind of problems. Furthermore, we derive the kinetics equations for the equilibrium [100], and generalize them to include out-of-equilibrium dynamics.

In order to derive the mean-field equations for our core model (see section 1.4), we will proceed as follows: first, we will study the dynamics of a classic magnetic system on a lattice with interaction (3.3); econd, we will study a Ginzubrg-Landau theory for the equilibrium of an interacting homopolymer; finally, we will "merge" these two approaches and develop a complete mean-field theory for the model. Notably, within the first two approaches we will show the existence of a second-order phase transition akin to what is observed respectively in the Ising model and the coil-globule collapse of the Flory-theory. In the last model, instead, a first-order phase transition is observed confirming the numerical results of section 1.4.3.

## 3.1 Basics of the Model

A theoretical approach to the study of the dynamics of polymer is quite difficult since several dynamicle variables (position and velocity of each beads) and model-dependent parameter (beads-beads interaction; connectivity term, flexibility terms) must be considered at the same time. For this reason, if one is uniquely interested in the equilibrium steady states of the system, it is useful to consider a more simplified, models. Lattice polymer models, for example, have been studied since the very beginnings of polymer physics [101], and have been shown to be fruitful to describe conformational phase transitions in these systems.

As a first attempt to the construction of a lattice polymer model, one can map a polymer made by $N$ monomers into a $N-1$ steps random walk on a lattice. The random walk will define a path onto the lattice and describe the spatial configuration

of the polymer, whose vertices are placed on the sites of the lattice crossed by the path. The main drawback of this model is that a random walk cross itself several times, violating the physical constraint of excluded volume. A useful, and more realistic, extension is the so-called self-avoiding walk (SAW, see fig. 3.1), that is a path on a lattice that never cross itself. Mapping polymers into self-avoiding walks has been proven useful many times during history, as the critical behaviour of SAWs belong to the same universality class as the one of "standard" polymers [102]. For this reasosn, results for SAWs at the termodynamic limit, most of the time agrees qualitatively and quantitatively with what observed with real polymers [103, 104].

We want to develop a Landau-Ginzburg mean field theory for the equilibrium dynamics of the polymer model described in section 1.4. Let us now highlights the main points of the model. First we must deal with the spatial dynamics: the dynamics of a semi-flexible chain which, as already pointed out, can be treated by mapping the polymer spatial configuration into a SAW on a cubic lattice. Second, we must consider the beads-beads interactions present in the model. Namely, each beads can be in three different states represented by the colors red, blue and grey, and interacting one to another via short ranged (nearest neighbours) interactions. Red and blue beads are "active" states, that is they interact with each other with an interaction akin to the classic Potts model, e.g. beads of the same color attract each other, but they do not interact with other colors. Grey beads are instead "inert" states and do not interact with any other beads. The Hamiltonian of the model on a $N$-steps SAW $\gamma$ can be written as:

$$H_\gamma = \frac{1}{2}\sum_{i,j=1}^{N} \Delta^\gamma_{i,j}\mathcal{I}(i,j) \ , \tag{3.1}$$

where $\mathbf{r}_i$ is the position of the $i$-th monomer on the lattice, $\Delta^\gamma$ is the contact matrix of the SAW $\gamma$, such that:

$$\Delta^\gamma_{a,b} = \begin{cases} 1 & \text{if monomers } a \text{ and } b \text{ are nearest neighbours} \\ 0 & \text{otherwise} \ , \end{cases} \tag{3.2}$$

and $\mathcal{I}$ is the interaction between the two monomers:

$$\mathcal{I}(i,j) = \begin{cases} -\epsilon & \text{if } q_i = q_j = \text{red or blue} \\ 0 & \text{if } q_i \neq q_j \text{ or } q_i = q_j = \text{grey} \ . \end{cases} \tag{3.3}$$

Since the solution of the full problem is long and requires many intermediate steps, for clarity we choose to separate it into smaller "sub-problems". First, we will study the system with interaction (3.3), but by considering a SAW who completely fills the lattice (also called Hamiltonian path), where we can replace the contact matrix $\Delta^\gamma$ in (3.1) with the contact matrix of the lattice $\Delta$. Second, we will consider an Hamiltonian of the kind of (3.1), but with an easier interaction where $\mathcal{I}(i,j) = -\epsilon$ always, that is the polymer is effectively an homopolymer. Finally, once these two problems are solved it will be easier to merge them together, and find the desired mean-field theory associated to the full problem.

## 3.2 Montecarlo Simulations of Self-Avoiding Walks

Before considering the field-theoretical approach, let us describe the Montecarlo method we have employed to performthe simulations of the model.

Given the nature of the system, our method will be composed of two distinct parts, one that deals with the evolution of the magnetic states, and the other which deals with the conformational changes of the polymers. While the first part is straight-forward and can be implemented as done in standard spin-lattice models via a Metropolis-Hastings method, the second is more tricky and must be approached with caution.

The catalogue of methods one can employ to study the configurations of self-avoiding walks is quite vast [105]. However, since we are dealing with interacting SAWs with the additional degree of freedom given by the magnetic interaction, it is convenient to adopt a method in the class of *dynamic methods*. Namely, let $H_\gamma[\mathbf{S}]$ be the energy of SAW with spatial configuration $\gamma$, and spin configuration $\mathbf{S}$, then we define the partition function at temperature $T$ as:

$$\mathcal{Z} = \sum_{\{\gamma\}} \sum_{\{S\}} \exp\left[-\beta H_\gamma[S]\right] \quad , \tag{3.4}$$

where $\beta = 1/(k_B T)$, and the sums are over every possible SAW and spin configuration. We start our simulation from a random SAW and, by using an appropriate stochastic, Markovian, process, we evolve this chain until it has reached the equilibrium distribution described by the partition function (3.4). This stochastic process can be achieved by using a Metropolis-Hastings algorithm using Hamiltonian $H_\gamma$, and by attempting "moves" which evolve the spatial configuration of the SAW.

In our simulations we choose to employ a combination of *local moves*, and *pivot moves*, in order to guarantee the ergodicity of the process [105]. They consist in:

**Local Moves** A local move (see fig. 3.2A) is one that alters the configuration of a few consecutive beads at the same time. In our algorithm, whenever we propose a local move, we choose a random consecutive segment of 4 beads, and replace its configuration with another chosen randomly which preserves the end points of the segment. The move is then accepted only if the new configuration does not overlap with the pre-existing remaining chain.

**Pivot Moves** A pivot move (see fig. 3.2B) is a "non-local move", as it changes the configuration of a great number of beads at the same time. In the case of pivot moves, we choose randomly a site along the chain as a pivot point, and apply randomly one of the possible symmetries of the lattice (e.g. rotations and reflections), on all the sites following our pivot point. As before, the new configuration is accepted only if there are no overlapping with the remaining chain.

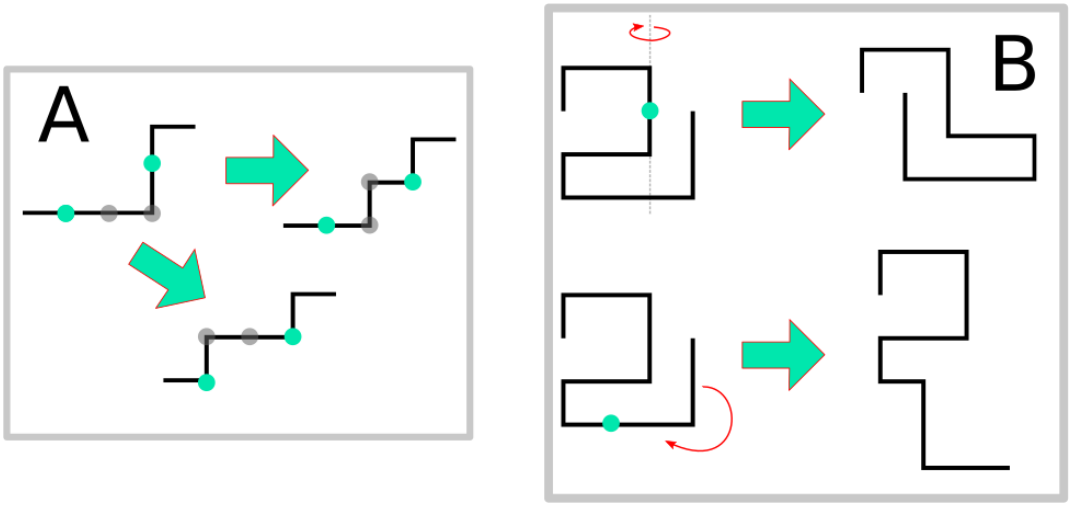



**Figure 3.2:** Pictorial illustration of local and pivot moves on a 2D square lattice. **(A)** When performing a *local move* we choose randomly 4 consecutive sites of the SAW. We keep the first and the last sites fixed, while we move the other two to obtain another SAW configuration. **(B)** We perform a *pivot* move by choosing randomly a site of the SAW (the pivot), and performing on all the following sites one geometric transformation. On the top figure we perform a reflection with respect to the $y$ axis; on the bottom one we perform a 90 degree clockwise rotation.

We are finally ready to fully characterize our algorithm. We start our simulations from a random $N$-step SAW, with random magnetic configuration. On this chain we attempt 1 pivot move, and $N-1$ local moves. We evaluate the difference in energy $\Delta H$ betweent the $H$ before the $N$ moves, and the one after; we accept (all) the configurational moves with probability $p = \min(1, \exp[-\beta \Delta H])$. After this, we perform $N$ spin-flipping moves and accept/reject them following the usual Metropolis-Hastings algorithm for spin models. The union of the $N$ configurational moves and the $N$ spin moves composes a single Montecarlo step. In order to achieve the equilibrium distribution, we perform $10^2 N$ Montecarlo steps.

To speed-up the convergence of the algorithm, we employ the *Multiple Markov Chain* strategy [106, 107]. Namely. we run the Markov chains to simultaneously find the equilibrium distributions for a set of different temperatures $\mathbf{T} = \{T_1, T_2, \ldots, T_k\}$. For each of these simulations, we perform in parallel a Montecarlo step, as just described. Then an adjacent pair, $T_i$, $T_{i+1}$ among the $T$ values is chosen, and a "swap move" is attempted, where the chain configuration at temperature $T_i$ with energy $H_i$ is swapped with the one at temperature $T_{i+1}$ with energy $H_{i+1}$. This swapping move is accepted with probability $p$:

$$p = \min\left(1, e^{(\beta_{i+1}-\beta_i)(H_{i+1}-H_i)}\right) \quad , \tag{3.5}$$

where $\beta_j = 1/(k_B T_j)$. The whole swapping process is also a Markov chain. Therefore,

the complete process (Montecarlo steps+swap strategy) is itself ergodic and has a unique limit in the so-called multi-canonical ensemble [108] which is actually the producr distribution of the separate canonical ensembles at temperature $T_1, T_2, \ldots, T_k$. The main advantage of this kind of strategy is that it leads to an algorithm where metastable states can be overcome [109], as the chains in these states can be swapped away at other higher temperatures. Thanks to this the times needed to reach equilibrium for our Montecarlo simulations becomes noticeably shorter, and, at the same time, we are simulating the dynamics at a wide array of temperatures.

## 3.3 Lattice-Wide Mean-Field Models

In this section we will develop a mean-field theory for a magnetic spin-model with nearest neighbours interaction described by (3.3). While this model is similar to the classic Potts model, it is interesting to examine in detail, as the presence of an inert state forces us to use a quite different approach to the one commonly employed.

For the remainder of this section, we will always use the convention of employing a $d$-dimensional hypercubic lattice with $M$ sites and coordination number $z = 2d$.

### 3.3.1 The Potts Model

Before we delve in the details of our model, let us review what the approach to the standard $\nu$-states Potts model [83]. We recall that the Hamiltonian, in this case, can be written as:

$$H_{\text{Potts}} = -\frac{\epsilon}{2} \sum_{i,j=1}^{M} \Delta_{ij} \delta_{q_i,q_j} \quad , \tag{3.6}$$

where $\Delta$ is the contact matrix of the lattice, $\delta$ is the Kronecker delta, and $q_i$ is the "color" $q_i = \{1, \ldots, \nu\}$ of the $i$-th site. The problem, here, is that it is quite difficult to treat a system whose elements can have a multitude of values with a mean-field approach. The main strategy that has been employed is to restrict the study to only a certain direction in the state-space. Namely, we choose a "privileged" color, e.g. $q = 1$, and study its abundance, while constraining the other $\nu - 1$ colors to have always the same abundance. More in details, we introduce a two-state model where we map the colors $q$ into the spins $S$ which take the values:

$$S(q) = \begin{cases} 1 & \text{if } q = 1 \\ -\frac{1}{\nu-1} & \text{otherwise} \quad . \end{cases} \tag{3.7}$$

The delta interaction in (3.6) must be replaced with a novel interaction $\mathcal{I}_{\text{Potts}}(S_i, S_j)$. Indeed, while the cases involving the privileged spin $S = 1$ remain the same, the interaction $\mathcal{I}_{\text{Potts}}(-\frac{1}{\nu-1}, -\frac{1}{\nu-1})$ must actually take into account that the spin $S = -\frac{1}{\nu-1}$ is a mixture of colors and therefore must be averaged over all the possible

underlying interactions, that is:

$$\mathcal{I}_{\text{Potts}}(-\tfrac{1}{\nu-1},-\tfrac{1}{\nu-1}) = \langle \mathcal{I}(q,q)\rangle_{q\neq 1} = -\frac{\epsilon}{\nu-1} \ , \tag{3.8}$$

which is easily computed by using that in the state $S=-\frac{1}{\nu-1}$ all the states $q\neq 1$ are equally probable. The interaction can be explicitly written in terms of the spin values:

$$\mathcal{I}_{\text{Potts}}(S_i,S_j) = -\epsilon\frac{(\nu-1)S_iS_j+1}{\nu} \ , \tag{3.9}$$

and used to compute a partition function at temperature $T$:

$$\mathcal{Z}_{\text{Potts}} = \sum_{\{S\}} \exp\left[\alpha\sum_{i,j}\Delta_{i,j}\frac{(\nu-1)S_iS_j+1}{\nu}\right] \ , \tag{3.10}$$

where we are summing over all possible spin configurations $\{S\}$ and $\alpha=\frac{\epsilon}{k_BT}$. Notice that in the sum over all the spin configurations the spins $S=-\frac{1}{\nu-1}$ must be counted $\nu-1$ times in order to count for the various color states they are representing.

Finally, we want to discuss the equilibrium state of the Potts model along the restriction where all the states $q\neq 1$ are equally abundant. It is easy to see from Hamiltonian (3.6) that at the equilibrium two possible ordered states are possible:

1. the system is "completely coherent", as all the sites present the state $q=1$;
2. the system is in a "mixed" configuration, divided in $\nu-1$ coherent domains of $q\neq 1$, all equal in size.

Notice that the difference in energy between the two states is an interface term, and therefore vanishes as we go into the thermodynamic limit $M\to\infty$, and therefore the two equilibrium states are equally probable. This suggests the presence of a symmetry breaking when varying the temperature $T$ as the system will have to "choose" one of the two equivalent equilibrium states. This kind of behaviour is typical of systems which presents a phase transition at some critical temperature $T_c$. This is confirmed by a study of the partition function (3.10), which can be done using a mean-field approach via an Hubbard-Stratonovich transform; this procedure gives the classic results expected from a Potts model [83].

### 3.3.2 Inserting inert states

We now consider an $\nu$ state Potts model and add $\mu$ inert states, such that in total we have $\nu+\mu$ states. As a first step, we will try to redo the reasoning employed in the standard Potts model in this more complex case. We will proceed as before: let us consider a two state spin-model where $S=1$ correspond to states with $q=1$ and $S=-\frac{1}{\nu+\mu-1}$ correspond to all the other states. The main difference with the

previous model stands in the expression for the interaction between spins $\mathcal{I}^{\star}_{\text{Potts}}$ which this time will read:

$$\mathcal{I}^{\star}_{\text{Potts}}(-\tfrac{1}{\nu+\mu-1},-\tfrac{1}{\nu+\mu-1}) = \langle \mathcal{I}^{\star}(q,q)\rangle_{q\neq 1} = -\epsilon\frac{\nu-1}{(\mu+\nu-1)^2} \ , \tag{3.11}$$

as only $\nu-1$ states are interacting this time. As before, we can write this interaction explicitly in the spins $S$ as follows:

$$\mathcal{I}^{\star}_{\text{Potts}}(S_i,S_j) = -\epsilon\frac{\varsigma}{(\nu+\mu)^2}\left(S_i+\frac{\mu}{\varsigma}\right)\left(S_j+\frac{\mu}{\varsigma}\right) - \epsilon\frac{\nu-1}{\varsigma} \ , \tag{3.12}$$

where $\varsigma$ is a parameter which reads:

$$\varsigma = (\nu+\mu)^2 - \nu - 2\mu \ .$$

Using this new interaction one can build a mean-field theory, just like in the standard Potts case. However, in this case, we have a problem, and the theory is not useful to study the critical properties of the model. Let us look at the equilibrium states of this model along the restriction where all the states $q \neq 1$ are equally abundant. The states will be the same as in the previous case, however, the "mixed case" will present inert states and therefore is characterised by a smaller energy of the "completely" coherent case. This means, that when we vary the temperature $T$, this time, we do not have symmetry breaking and, therefore, no phase transition can be observed.

### Mean-Field Theory

Let us, nevertheless, try to develop a mean-field theory from interaction (3.12), and discuss the equilibrium results. As a first step, to ease the notation, we introduce the parameters $a, b, c, Q$:

$$a = \frac{\varsigma}{(\nu+\mu)^2} \ , \qquad b = \frac{\mu}{\varsigma} \ , \qquad c = \frac{\nu-1}{\varsigma} \ , \qquad Q = \nu+\mu-1 \ .$$

With this choice of parameters, we can write the partition function for the model at temperature $T$:

$$\mathcal{Z} = \sum_{\{S\}} \exp\left[\frac{\beta\epsilon}{2} a \sum_{i,j}(S_i+b)\Delta_{ij}(S_j+b)\right] \ , \tag{3.13}$$

where $\sum_{\{S\}}$ sums over all possible spin configurations and counts each spin state with value $S = -\frac{1}{Q}$ multiple ($Q$) times. Notice, also, that we have neglected the constant term $\sum_{i,j} c\Delta_{ij}$.

A mean-field theory can be easily obtained from a partition function in the form (3.13) by taking an Hubbard-Stratonovich transform [110] in the field $\phi$:

$$\mathcal{Z} \sim \int_{\mathbb{R}^M} \exp\left[-\sum_{i,j}\frac{1}{2\beta\epsilon}\frac{1}{a}\phi_i\Delta^{-1}_{ij}\phi_j\right]\sum_{\{S\}}\exp\left[\sum_i \phi_i(S_i+b)\right]\mathrm{d}\phi \ , \tag{3.14}$$

where we have disregarded the constant term $\sqrt{(\det\Delta)/(2\pi)^M}$, and we define $\mathrm{d}\phi = \mathrm{d}\phi_1\,\mathrm{d}\phi_2\cdots\mathrm{d}\phi_M$. As a first step, we perform the summation over the spin configurations, an obtain:

$$exp\left[\sum_i \phi_i(S_i+b)\right] = \prod_i \left(e^{\phi_i(1+b)} + Qe^{\phi_i\left(-\frac{1}{Q}+b\right)}\right) \ ,$$

therefore mapping (3.14) into:

$$\mathcal{Z} \sim \int_{\mathbb{R}^M} \exp\left[-\sum_{i,j}\frac{1}{2\beta\epsilon}\frac{1}{a}\phi_i\Delta_{ij}^{-1}\phi_j + \sum_i \log\left(e^{\phi_i(1+b)} + Qe^{\phi_i\left(-\frac{1}{Q}+b\right)}\right)\right]\mathrm{d}\phi \ . \tag{3.15}$$

On this new expression of $\mathcal{Z}$ we perform both a saddle-point approximation, and employ a mean-field hypothesis. Namely, we assume that the integral in (3.15) can be approximated by the maximum value of its integrand (saddle-point approximation), and that in this maximum we have $\phi_i \equiv \phi$ for all $i$ (mean-field hypothesis). Using this procedure, (3.15) simplifies and can be approximated to:

$$\mathcal{Z} \sim \exp\left[-\frac{M}{a}\frac{\phi^2}{2\alpha z} + M\log\left(e^{\phi(1+b)} + Qe^{\phi\left(-\frac{1}{Q}+b\right)}\right)\right] \ , \tag{3.16}$$

where $\alpha = \beta\epsilon$, and we have employed that $\sum_{i,j}\Delta_{ij}^{-1} = \frac{M}{z}$ [110]. We are finally ready to compute the mean-field expression for the free-energy per site $f = -\frac{k_BT}{M}\log\mathcal{Z}$:

$$\frac{1}{k_BT}f = \frac{1}{a}\frac{\phi^2}{2\alpha z} - \log\left(e^{\phi(1+b)} + Qe^{\phi\left(-\frac{1}{Q}+b\right)}\right) \ . \tag{3.17}$$

**Physical Relevance of the Field $\phi$**

Before looking at the properties of free-energy (3.17), let us discuss the physical meaning of the field $\phi$. To do so, we add in the partition function (3.13) a new linear term due to a fictitious external field $h$:

$$\mathcal{Z}_{\text{ext}} = \sum_{\{S\}}\exp\left[\frac{\beta\epsilon}{2}a\sum_{i,j}(S_i+b)\Delta_{ij}(S_j+b) + \beta h\sum_i S_i\right] \ . \tag{3.18}$$

Notice, that we can find the order parameter magnetisation $m = \frac{1}{M}\langle\sum_i S_i\rangle$ in the ensemble defined by $\mathcal{Z}$, by taking a first derivative of $\mathcal{Z}_{\text{ext}}$ in the external field $h$:

$$m = \frac{1}{\beta}\left(\frac{\partial}{\partial h}\log\mathcal{Z}_{\text{ext}}\right|_{h=0} \ .$$

By employing the same methodology as before, we compute (3.18) in the mean-field approximation and obtain:

$$\mathcal{Z}_{\text{ext}} \sim \exp\left[-\frac{M}{a}\frac{\phi^2}{2\alpha z} + M\log\left(e^{\phi(1+b)+\beta h} + Qe^{\phi\left(-\frac{1}{Q}+b\right)-\beta\frac{h}{Q}}\right)\right] \ . \tag{3.19}$$

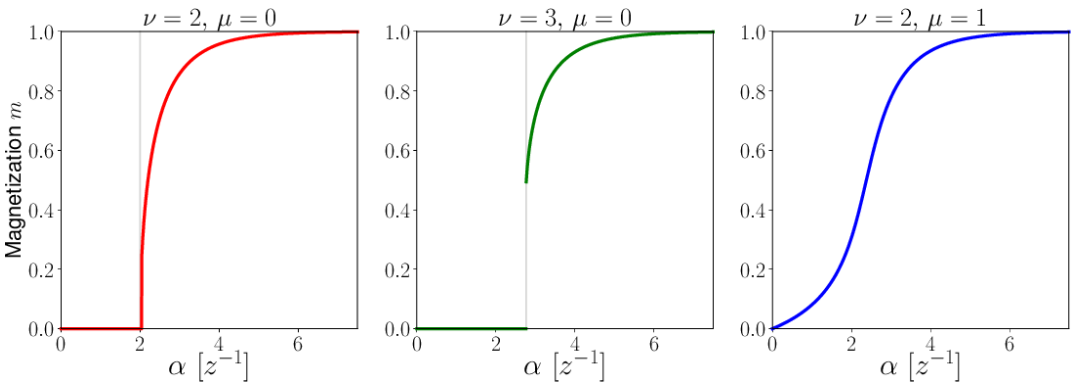



**Figure 3.3:** Plots of magnetization $m$ at the equilibrium as a function of $\alpha$. Plots obtained by numerically minimizing free-energy (3.17) and by inserting the $\phi$ value found in (3.20). Results show that phase transition occurs only when $m = 0$, while in the other cases we observe only that the magnetisation grows, as $\alpha$ grows. Notice that in accordance with the classic Potts model, we find a continuous phase transition for $\nu = 2, \mu = 0$, and a first-order for $\nu = 3, \mu = 0$ [83].

Therefore, we can find an expression for the magnetisation in the mean-field limit:

$$m = \frac{e^{\phi(1+b)} - e^{\phi\left(-\frac{1}{Q}+b\right)}}{e^{\phi(1+b)} + Qe^{\phi\left(-\frac{1}{Q}+b\right)}} \ , \tag{3.20}$$

which indicates that the field $\phi$ describes the average magnetisation of the system, and, more precisely, the abundance of the privileged spin $S = 1$ .

#### Equilibrium properties of the model

To study the properties of free energy (3.17), we expand it around $\phi = 0$. One can notice the presence of a linear term in $\phi$

$$\frac{1}{k_B T} f \sim -\log(Q+1) - b\phi + o(\phi^2)$$

which mimics the effect of an external magnetic field of intensity $b$. This term precludes the model from displaying a phase transition whenever we have $b > 0$. Notice, also, that $b = 0$ if and only if there are no inert states ($\mu = 0$), and the standard Potts model is recovered, confirming our previous discussion (see fig. 3.3).

### 3.3.3 Two-Fields approach

Here, we discuss an alternative approach to the one presented in the previous section, that preserve the critical properties of the system even when inert states are present. Let us consider a model with $\nu$ active states, and $\mu$ inert ones. As in the previous

approach, we restrict the phase space to consider only states where one color $q = 1$ is privileged, while the other $\nu - 1$ *active* color have always the same abundance. Notice the difference with the previous case: here we explicitly separate active states from inert ones in two states, while before we integrated all the other $\nu + \mu - 1$ states into a single state. With this assumption, we can map a model with $\nu + \mu$ colors, into a three-states spin model where: the spin $S = 1$ represents the privileged color $q = 1$; the spin $S = -1/(\nu - 1)$ represent all the other $\nu - 1$ active spins; the spin $S = 0$ represents all the $\mu$ inert states.

### Mean-Field Model

As a first step, we are particularly interested in the case of $\nu = 2$, $\mu = 1$ case, as it corresponds to our original interaction described in Hamiltonian (3.1). Namely, the three colors are described by the three spin values $S = \{-1, 0, +1\}$. With this choice, it is easy to see that interaction (3.3) can be written as:

$$\mathcal{I}(S_i, S_j) = -\frac{\epsilon}{2}\left[S_i^2 S_j^2 + S_i S_j\right] \quad . \tag{3.21}$$

Using this interaction we can explicitly write down a partition function for the model:

$$\mathcal{Z} = \sum_{\{S\}} \exp\left[\frac{\beta\epsilon}{2}\sum_{i,j} S_i^2 \Delta_{ij} S_j^2 + \frac{\beta\epsilon}{2}\sum_{i,j} S_i \Delta_{ij} S_j\right] \quad , \tag{3.22}$$

where we are summing over every possible spin configuration $\{S\}$. A field study of this system is possible by taking two successive Hubbard-Stratonovich transforms on the terms $S_i\Delta_{ij}S_j$ and $S_i^2\Delta_{ij}S_j^2$ by introducing the respectively the fields $\phi$ and $\psi$:

$$\mathcal{Z} \sim \int_{\mathbb{R}^{2M}} \exp\left[-\sum_{i,j}\frac{1}{2\beta\epsilon}\left(\psi_i \Delta_{i,j}^{-1}\psi_j + \phi_i\Delta_{i,j}^{-1}\phi_j\right)\right] \sum_{\{S\}} \exp\left[\sum_i \left(S_i\phi_i + S_i^2\psi_i\right)\right] \mathrm{d}\phi\,\mathrm{d}\psi \tag{3.23}$$

where we have disregarded the constant multiplicative term $(\det\Delta)/(2\pi)^M$, and we define $\mathrm{d}\phi = \mathrm{d}\phi_1\,\mathrm{d}\phi_2\cdots\mathrm{d}\phi_M$, and $\mathrm{d}\psi = \mathrm{d}\psi_1\,\mathrm{d}\psi_2\cdots\mathrm{d}\psi_M$. As a first step, we notice that we can perform easily the summation over every spin configuration:

$$\sum_{\{S\}} \exp\left[\sum_i \left(S_i\phi_i + S_i^2\psi_i\right)\right] = \prod_i \left(e^{\phi_i+\psi_i} + e^{-\phi_i+\psi_i} + 1\right) \quad , \tag{3.24}$$

which maps equation (3.23) into:

$$\mathcal{Z} \sim \int_{\mathbb{R}^{2M}} \exp\left[-\sum_{i,j}\frac{1}{2\beta\epsilon}\left(\psi_i \Delta_{i,j}^{-1}\psi_j + \phi_i\Delta_{i,j}^{-1}\phi_j\right) + \sum_i \log\left(2e^{\psi_i}\cosh\phi_i + 1\right)\right] \mathrm{d}\phi\,\mathrm{d}\psi \tag{3.25}$$

We now, finally, introduce the mean-field hypotheses via a saddle-point approximation. Namely, we suppose that the value of $\mathcal{Z}$ can be well approximated by the maximum of the distribution defined by the exponential in (3.25) (saddle-point approximation), and that in this maximum we have that $\psi_i \equiv \psi$ and $\phi_i \equiv \phi$ for all $i$ (mean-field hypothesis). Within this framework, the partition function reads:

$$\mathcal{Z} \sim \exp\left[-M\frac{\phi^2}{2\alpha z} - M\frac{\psi^2}{2\alpha z} + M\log\left(2e^{\psi}\cosh\phi + 1\right)\right] \quad , \tag{3.26}$$

where $\alpha = \beta\epsilon$, and we have used the relation $\sum_{i,j}\Delta^{-1}_{i,j} = M/z$. We finally estimate the free-energy per site $f = -\frac{k_B T}{M}\log\mathcal{Z}$:

$$\beta f = \frac{\phi^2 + \psi^2}{2\alpha z} - \log(2e^{\psi}\cosh\phi + 1) \quad . \tag{3.27}$$

**Physical Relevance of the Fields $\phi$, $\psi$.**

Before looking in details at the equilibrium properties of the model, let us discuss the physical relevance of the two fields $\phi$ and $\psi$. First, we include in the partition function (3.22) two new terms to account for fictitious external fields $h$ and $w$, one acting on the active spins, and the other on the inert spins:

$$\mathcal{Z}_{\text{ext}} = \sum_{\{S\}}\exp\left[\frac{\beta\epsilon}{2}\sum_{i,j}S_i^2\Delta_{ij}S_j^2 + \frac{\beta\epsilon}{2}\sum_{i,j}S_i\Delta_{ij}S_j - \sum_i\beta h S_i - \sum_i\beta w(1-S_i^2)\right] \quad . \tag{3.28}$$

Notice that we can obtain an estimate for the average value of the "magnetisation" $m = \frac{1}{M}\langle\sum_i S_i\rangle$, and of the inert-states density $\varrho = \frac{1}{M}\langle\sum_i(1-S_i^2)\rangle$, in the ensemble defined by (3.22), simply using the relations:

$$m = -\frac{1}{M\beta}\left(\frac{\partial\mathcal{Z}_{\text{ext}}}{\partial h}\bigg|_{h,w=0}\right) \qquad \varrho = -\frac{1}{M\beta}\left(\frac{\partial\mathcal{Z}_{\text{ext}}}{\partial w}\bigg|_{h,w=0}\right) \quad . \tag{3.29}$$

We can follow the same procedure used on $\mathcal{Z}$, to estimate a mean-field approximation for the partition function $\mathcal{Z}_{\text{ext}}$ (3.28); we obtain:

$$\mathcal{Z}_{\text{ext}} \sim \exp\left[-M\frac{\phi^2}{2\alpha z} - M\frac{\psi^2}{2\alpha z} + M\log\left(2e^{\psi}\cosh(\phi-\beta h) + e^{-\beta w}\right)\right] \quad . \tag{3.30}$$

By employing the relations (3.29), one obtains the following estimates for magnetisation $m$ and inert density $\varrho$:

$$\begin{aligned}\varrho &= \left\langle\frac{1}{2e^{\psi}\cosh(\phi)+1}\right\rangle \\ m &= \left\langle\frac{2e^{\psi}\sinh(\phi)}{2e^{\psi}\cosh(\phi)+1}\right\rangle \quad .\end{aligned} \tag{3.31}$$

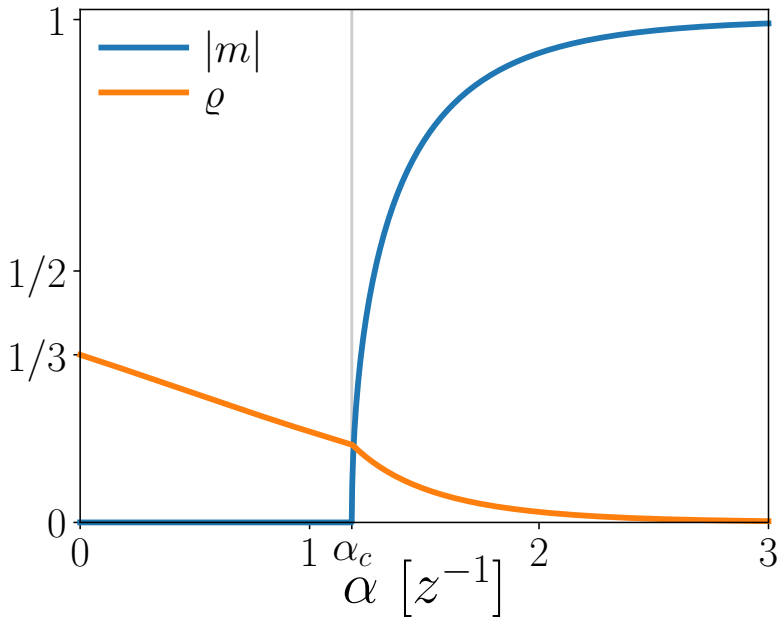



**Figure 3.4:** Equilibrium phases of the model described by partition function (3.22). For each value of $\alpha$ the order parameters $m$ and $\varrho$ have been found by first numerically minimizing free-energy (3.27), and then substituting the so-found field values in (3.31). We observe a second-order phase transition between a magnetically disordered state characterised by $m = 0$, to an ordered state with $|m| > 0$ at the critical inverse temperature $\alpha_c z = \frac{1}{2}e^{-1} + 1$.

We conclude that the field $\phi$ is intimately related to the magnetisation, as the sign of $\phi$ and $m$ will be the same; the field $\psi$, instead, helps $\phi$ in regulating the total abundance of the inert states. Notice, indeed, that when $\phi = \psi = 0$ we have $\varrho = \frac{1}{3}$ and $m = 0$ indicating a disordered system where all the three states (2 active and one inert) are equally probable. On the other hand when $\psi$ and $\phi$ are large, $\varrho$ will quickly decrease, while at the same time the magnetisation will increase, suggesting the presence of an ordered state where the density of inert states is negligible.

### Equilibrium properties of the model

We are finally ready to study the equilibrium phases described by the free-energy (3.27) by minimizing $f$ in the two fields $\phi$ and $\psi$. By studying the equations $\frac{\partial f}{\partial \phi} = 0$ and $\frac{\partial f}{\partial \psi} = 0$, we find that the equilibrium value of $\psi$ depends only on $\phi$ by the following relation:

$$\psi^{\mathrm{eq}}\left[\phi^{\mathrm{eq}}\right] = \frac{\phi^{\mathrm{eq}}}{\tanh\left(\phi^{\mathrm{eq}}\right)} \quad .$$

By substituting in (3.27), and expanding in series around $\phi^{\text{eq}} \sim 0$, one obtains:

$$f\left[\phi^{\text{eq}}\right] \approx -\log(3) + B[\alpha]\phi^2 + C[\alpha]\phi^4 + O(\phi^6) \ , \tag{3.32}$$

where $B[\alpha]$ and $C[\alpha]$ are:

$$B[\alpha] = \frac{5}{6\alpha z} - \frac{5e}{3(1+2e)} \qquad C[\alpha] = \frac{1}{30\alpha z} + \frac{76e-87}{180(1+2e)^2} \ .$$

By minimizing this expression, we obtain an approximation for the equilibrium value of $\phi^{\text{eq}}$ as a function of $\alpha$:

$$\phi^{\text{eq}} \approx \begin{cases} 0 & \text{if } \alpha z \leq \frac{1+2e}{2e} \\ \pm\sqrt{-\frac{B[\alpha]}{2C[\alpha]}} & \text{if } \alpha z > \frac{1+2e}{2e} \end{cases} \ . \tag{3.33}$$

This show the presence of second-order phase transition between a disordered state at high temperature ($\alpha \leq \alpha_c$) and an ordered one at smaller temperatures ($\alpha > \alpha_c$), with critical inverse temperature $\alpha_c = \frac{1+2e}{2e} z^{-1}$. Numerical minimizations of free energy (3.27) confirm the presence of a continuous phase transition (see fig. 3.4), akin to what observed in the classic 2-states Potts model. Interestingly, while the magnetisation remains always $m = 0$ in the incoherent state $\alpha < \alpha_c$, the inert-states density $\varrho$ starts at value $1/3$ when $\alpha = 0$, and is always decreasing when $\alpha$ increases. This suggests that the three states (two actives and one inert) are equally probable only when $\alpha = 0$, but whenever $\alpha > 0$ the active states will always be favored, even when no magnetic coherence arises.

### 3.3.4 Generalization of the Two-Fields approach

In this section, we will briefly discuss a generalization for interaction (3.21) to include the case where $\nu \geq 2$ active states are present, and a generic number $\mu \geq 0$ of inert states is introduced. We use an approach similar to what discussed in section 3.3.1 where we studied the classic Potts model. Namely, first we choose a privileged active color (e.g. $q = 1$) to which we associate the spin state $S = 1$; the other $\nu - 1$ active states will be all mapped into a single spin value $S = -\frac{1}{\nu-1}$ and assumed equiprobable; finally the $\mu$ inert states are all mapped in the spin state $S = 0$. By following the approach described by (3.8), we can write a new form for the interaction $\mathcal{I}$ for the case of a Potts model with $\mu$ inert states as:

$$\mathcal{I}(S_i, S_j) = -\epsilon a(S_i^2 - bS_i)(S_j^2 - bS_i) - \epsilon c S_i S_j \ , \tag{3.34}$$

where $a, b, c$ are numerical parameters which depend exclusively on the number of active states $\nu$:

$$a = \frac{(\nu-1)^2}{\nu} \ , \qquad b = \frac{\nu-2}{\nu-1} \ , \qquad c = \frac{\nu-1}{\nu} \ . \tag{3.35}$$

### Mean-Field Model

We proceed as before; as a first step we write down the expression for the partition function at temperature $\beta = 1/(k_B T)$:

$$\mathcal{Z} = \sum_{\{S\}} \exp\left[\sum_{i,j} \frac{\beta\epsilon a}{2}(S_i^2 - bS_i)\Delta_{i,j}(S_j^2 - bS_j) + \sum_{i,j} \frac{\beta\epsilon c}{2} S_i \Delta_{i,j} S_j\right] \quad , \tag{3.36}$$

where we are summing over any spin configuration, but counting $\nu - 1$ times the spin-state $S = -\frac{1}{\nu-1}$, and $m$ times the spin-state $S = 0$. We now perform the usual Hubbard-Stratonovich transforms in the fields $\phi$ and $\psi$:

$$\begin{aligned}\mathcal{Z} \sim \int_{\mathbb{R}^{2M}} \exp\left[-\sum_{i,j} \frac{1}{2\beta\epsilon}\left(\frac{1}{a}\psi_i \Delta_{i,j}^{-1}\psi_j + \frac{1}{c}\phi_i \Delta_{i,j}^{-1}\phi_j\right)\right] \times \\ \times \sum_{\{S\}} \exp\left[\sum_i \left(S_i\phi_i + (S_i^2 - bS_i)\psi_i\right)\right] \mathrm{d}\phi\, \mathrm{d}\psi \quad .\end{aligned} \tag{3.37}$$

The summation over all the spin configuration can now be easily computed, and reads:

$$\sum_{\{S\}} \exp\left[\sum_i \left(S_i\phi_i + (S_i^2 - bS_i)\psi_i\right)\right] = \prod_i \left(e^{\phi_i + \frac{1}{\nu-1}\psi_i} + (n-1)e^{\frac{-\phi_i+\psi_i}{\nu-1}} + m\right) \quad , \tag{3.38}$$

where we have used that $b = \frac{\nu-2}{\nu-1}$.

The procedure for finding a mean-field free-energy is identical to the one discussed in the previous section. In this case we obtain that the free-energy per site $f = -\frac{1}{M} k_B T \log \mathcal{Z}$ takes the form:

$$\frac{1}{k_B T} f = \frac{\nu}{(\nu-1)^2}\frac{\psi^2}{2\alpha z} + \frac{\nu}{\nu-1}\frac{\phi^2}{2\alpha z} - \log\left(e^{\phi + \frac{1}{\nu-1}\psi} + (\nu-1)e^{\frac{-\phi+\psi}{\nu-1}} + \mu\right) \quad . \tag{3.39}$$

### Physical Relevance of the Fields $\phi$, $\psi$.

By following the same procedure of the previous section, we can map the two fields $\phi$ and $\psi$ into: the magnetisation $m$, and the inert-states density $\varrho$. We find that:

$$\begin{aligned} m &= \frac{2e^{\frac{1}{\nu-1}\psi}\sinh\phi}{e^{\phi+\frac{1}{\nu-1}\psi} + (\nu-1)e^{\frac{-\phi+\psi}{\nu-1}} + \mu} \quad , \\ \varrho &= \frac{\mu}{e^{\phi+\frac{1}{\nu-1}\psi} + (\nu-1)e^{\frac{-\phi+\psi}{\nu-1}} + \mu} \quad . \end{aligned} \tag{3.40}$$

Notice that when we are in the classic Potts model with no inert states ($\mu = 0$), equations (3.40) shows correctly that $\varrho = 0$ for every value of the fields.

**Equilibrium Properties of the System.**

The equilibrium properties of the model are obtained by minimizing $f$ (3.39) in the two fields $\phi$ and $\psi$. As in the previous section, we find that at the equilibrium the field $\psi$ depends only on $\phi$ by the following relation:

$$\psi^{\text{eq}}\left[\phi^{\text{eq}}\right] = (\nu - 1)\frac{\phi^{\text{eq}}}{\tanh\left(\phi^{\text{eq}}\right)} \ .$$

Now we insert this estimation of $\psi$ in (3.39), and expanding it in series:

$$f \sim A[\alpha;\mu,\nu]\phi^2 + B[\alpha;\mu,\nu]\phi^3 + C[\alpha;\mu,\nu]\phi^4 + O(\phi^5) \ , \tag{3.41}$$

where:

$$\begin{aligned}
A[\alpha;\mu,\nu] &= -\frac{\nu}{3\alpha z} + \frac{\nu}{\nu-1}\frac{1}{2\alpha z} - \frac{e\nu(2\nu+1)}{6(\nu-1)(\mu+e\nu)} \\
B[\alpha;m,n] &= -\frac{e\nu(\nu-2)}{6(\nu-1)^2(\mu+e\nu)} \\
C[\alpha;m,n] &= \frac{\nu}{30\alpha z} + \frac{e\nu\left[-3\mu(4\nu^3+13\nu^2-43\nu+31)+e\nu(8\nu^3-39\nu^2+114\nu-98)\right]}{30(\nu-1)^3(\mu+e\nu)^2}
\end{aligned} \ . \tag{3.42}$$

An analytic study of this kind of expansion is in general quite difficult, but some considerations can still be made. First, notice that the cubic term in $\phi$ ($B[\alpha;\nu,\mu]$) is always negative and non-null whenever we have $\nu > 2$, which is typical of systems showing first-order phase transition. Indeed, just like in the classic Potts model, numerical minimizations of free-energy (3.39) show that whenever we have more than two active states ($\nu \geq 3$), we have always a first-order phase transition (see fig. 3.5A) from a magnetically incoherent state ($|m| = 0$) to a magnetically coherent one with ($|m| > 0$) for every value of $\mu$.

The last case we have to examine is the case $\nu = 2$, that is when we have only two active states. It is easy to prove from (3.39) that $f(-\phi,\psi) = f(\phi,\psi)$ for every value of $\psi$. This indicates that expansion in series of $f$ will have only even powers of $\phi$. For this very reason, a general study of the phase transition in $\phi$ for an arbitrary value of $\mu$ is difficult. Nevertheless, from the expansion (3.41), it turns out that $C[\alpha;\nu,\mu] \geq 0$ always if $\mu \leq 2$, while it can become negative otherwise, indicating, again, a possible first-order phase transition for $\nu = 2, \mu > 2$. Notably, numerical minimizations suggest that this argument is only partially right: we observe that if $\nu = 2, \mu \leq 3$ the phase transition is of the second order, while if $\nu = 2, \mu \geq 4$ the phase transition seems always to be of the first order (see fig. 3.5B). By comparison with the classic Potts model: we have shown that the order of the phase transition depends not only on the number of active states, but also on the possible presence of noisy inert states.

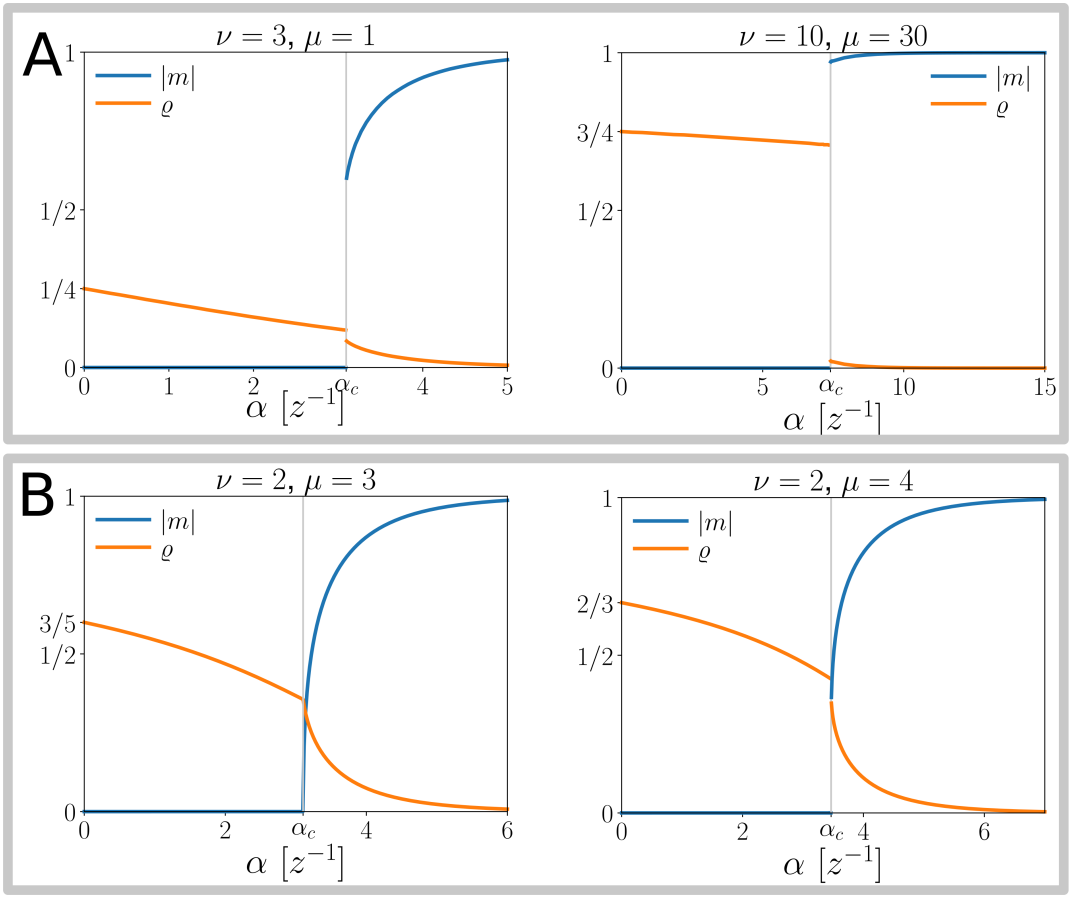



**Figure 3.5:** Order parameters $\mu$ and $\varrho$ (see (3.40)) as a function of $\alpha$ at equilibrium. **(A)** If $\nu \geq 3$ we have a first-order phase transition between an incoherent $\mu = 0$ phase and an ordered one $|\mu| > 0$, for whatever arbitrary value of $\mu$. **(B)** If $\nu = 2$ we can have both a continuous phase transition if $\mu \leq 3$, or a first order one if $\mu \geq 4$. Notice also that in both situations we have that the inert density $\varrho$ decreases monotonously with $\alpha$ increasing, with its maximum value at $\alpha = 0$ coinciding with $\varrho(0) = \frac{\mu}{\nu+\mu}$.

## 3.4 Landau-Ginzburg Model for an Homopolymer

In this section we present a Landau-Ginzburg theory to study the equilibrium properties of interacting homopolymers, modeled as an ISAW on a lattice. This kind of models have been widely studied as they are the simpler way to construct a mean field theory to describe the coil-globule collapse observed in polymers in solutions [101, 111]. However, a complete formulation of the problem via a Landau-Ginzburg-like free-energy is missing. Indeed, one can satisfyingly describe the problem only in the limit of the so-called "Hamiltonian paths", that is self-avoiding walks that completely fill the lattice. Nearby this limit, we can write down a free-energy which qualitatively (but not quantitatively) describe correctly the coil-globule transition.

In order to find an expression for the mean-field free energy we divide the analysis in two parts. First we study the system in absence of any interaction, i.e. we study the so-called "enumeration" of SAW in a lattice, via an heuristic argument that can be rigorously proved in the case of Hamiltonian paths [112]. Second, in the limit of Hamiltonian paths, we introduce an interaction between the monomers and study the coil-globule transition.

### 3.4.1 Enumeration of Self-Avoiding Walks

Let us consider an $N$ step self-avoiding walk on a lattice with $M$ sites. We are interested in estimating the number of such walks $\mathcal{Z}_{\text{SAW}}$ in the thermodynamic limit:

$$N \to +\infty \text{ and } M \to +\infty \qquad \text{with } 0 \leq \rho \equiv \frac{N}{M} \leq 1 \ , \tag{3.43}$$

where we will call $\rho \equiv N/M$ the packing *density* of the SAW. An exact answer to this problem is actually still an open problem both in mathematics and physics, and only estimates in certain limits are possible.

Here we present an easy heuristic method which, despite this, gives a surprisingly good first approximation for $\mathcal{Z}_{\text{SAW}}$ in terms of the density $\rho$ on a lattice with $M$ sites and coordination number $z$. We write the partition function $\mathcal{Z}_{\text{SAW}}$ as follows [113]:

$$\mathcal{Z}_{\text{SAW}} \sim \frac{M!}{(M-N)!} \left(\frac{z}{M}\right)^{N-1} \ . \tag{3.44}$$

Let us examine the various terms in this expression. First, the term $M!/(M-N)!$ is the number of ways in which we can sequentially (the order counts) pick $N$ sites on a lattice of $M$ sites. However, without any restriction, the $N$ sites picked are generally far away from each other and do not describe a walk; for this reason we correct this number by inserting the factor $\left(\frac{z}{M}\right)^{N-1}$ because only the first bead has actually $M$ possible choices, while the remaining ones must be chosen among the $z$ first neighbours.

We are interested in an expression for a free-energy in the thermodynamic limit (3.43). We define the free-energy per monomer at the temperature $T$:

$$f_{\text{SAW}} = -\frac{k_B T}{N} \log \mathcal{Z}_{\text{SAW}} \ . \tag{3.45}$$

Using that $M = N/\rho$ and the Stirling approximation for the factorials, it is easy to prove that in the thermodynamical limit:

$$\frac{1}{k_B T} f_{\text{SAW}} = -\log\left(\frac{z}{e}\right) + \frac{1-\rho}{\rho} \log(1-\rho) \ . \tag{3.46}$$

While the procedure is heuristic, the result is the same as in more rigorous mean-field studies [112], in the limit of Hamiltonian paths, e.g. $\rho \to 1$. Namely, these works suggest that the result is correct within an approximation of order $\frac{1}{z}$.

### 3.4.2 Mean-Field Theory of an Interacting SAW

Thanks to these results, we can develop a mean-field theory for an interacting self-avoiding walk, in order to model the behaviour of homopolymers in a solution. As a first step, we write an expression for the energy of an homopolymer modeled by an $N$-steps ISAW $\gamma$ on a lattice composed of $M$ sites with coordination number $z$:

$$H_{\text{ISAW}} = -\frac{\epsilon}{2}\sum_{i=1}^{N} n_c(i;\gamma) \ , \tag{3.47}$$

where $n_c(i;\gamma)$ is the number of nearest-neighbours of the $i$-th monomer of $\gamma$ occupied by other monomers. Given, this Hamiltonian, we can write the partition function at temperature $T$:

$$\mathcal{Z}_{\text{ISAW}} = \sum_{\{\gamma\}} \exp\left[\frac{\beta\epsilon}{2}\sum_{i=1}^{N} n_c(i;\gamma)\right] \ , \tag{3.48}$$

where we are summing on every possible SAW $\gamma$, and $\beta = 1/(k_B T)$.

An exact expression for (3.48) is not known. However in the limit of large packing density $\rho \equiv N/M$ ($\rho \to 1$), it is possible to study the free-energy density $f_{\text{ISAW}} = -k_B T \mathcal{Z}_{\text{ISAW}}$ by doing a series expansion in $z^{-1}$ [114]. Here, we present a possible derivation of the first terms of this serie via some mean-field arguments.

Let us consider the function $n_c(i;\gamma)$ which gives us the number of "contacts" of the $i$-th monomer with the other monomers on the SAW. Our first mean-field approximation is to disregard the dipendence from the particular monomer, and we assume that $n_c(i;\gamma) = \tilde{n}_\gamma \ \forall i \in \gamma$. Second, we assume $\rho \sim 1$, that is the polymer occupies most of the lattice; win this way we can assume that all the polymers $\gamma$ have about the same average number of contacts $n_\gamma$. With this approximation, we are saying that the value of $n_\gamma$ uniquely depends on the geometric properties of the lattice ($M$, $z$) and the size of the polymer $N$; more specifically we assume $n_\gamma \equiv n_c(\rho; z)$. With this assumption (3.48) can be approximated as:

$$\mathcal{Z}_{\text{ISAW}} = \exp\left[\frac{\beta\epsilon}{2} n_c(\rho; z)\right] \mathcal{Z}_{\text{SAW}} \ , \tag{3.49}$$

where we have used that $\mathcal{Z}_{\text{SAW}} = \sum_{\{\gamma\}} 1$. Finally, we must estimate the average number of contacts $n_c(\rho; z)$ as a function of the density. Again, a general form is not easy to find, but trivially $n_c(1; z) = z$, therefore, if we have $\rho \sim 1$ we can approximate this function to:

$$n_c(\rho; z) = \rho z \ . \tag{3.50}$$

Notice that while this approximation works in the Hamiltonian path limit, in the opposite case $\rho \to 0$, we would have that $n_c(\rho; z) \to 0$ which is not true, as we expect that $n_c(\rho, z) \to 2$ to ensure the connectivity the chain. Nevertheless, when we have dense walks this crude approximation is actually confirmed by more rigorous approaches [114].

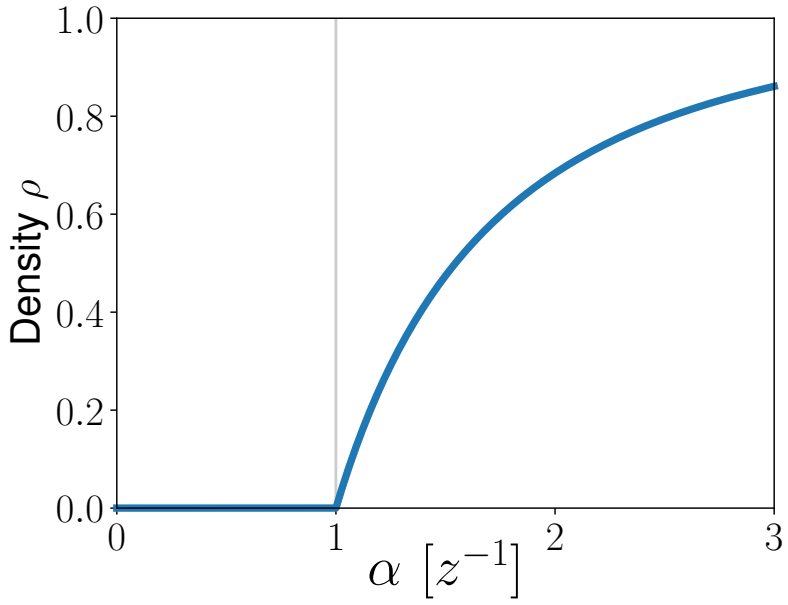



**Figure 3.6:** Plots of the equilibrium value of the density $\rho$ as a function of the temperature parameter $\alpha$. The plot has been obtained via numerical minimization of free-energy (3.51). The plot shows the existence of continuous-phase transition between a swollen phase ($\rho = 0$) and a globular phase where $\rho > 0$ at $\alpha = 1/z$, in qualitative accordance with the classic Flory theory.

Using (3.50), (3.49) and (3.46), we are able to write an expression for the free-energy $f_{\text{ISAW}} = -\frac{k_B T}{N} \log \mathcal{Z}_{\text{ISAW}}$:

$$\frac{1}{k_B T} f_{\text{ISAW}} = \frac{1-\rho}{\rho} \log(1-\rho) - \frac{1}{2} \alpha z \rho \ , \tag{3.51}$$

where $\alpha = \beta\epsilon$ and we have disregarded the constant term in (3.46).

We are now ready to study the equilibrium dynamics of an interacting self-avoiding walk described by (3.51). To have a qualitative understanding of the model, we expand $f_{\text{ISAW}}$ in series of $\rho$:

$$\frac{1}{k_B T} f_{\text{ISAW}} \sim -1 + \frac{1}{2}(1 - \alpha z)\rho + \frac{1}{6}\rho^2 + O\left(\rho^3\right) \ . \tag{3.52}$$

At the equilibrium, the system will assume the value of $\rho$, which minimize the free energy $f_{\text{ISAW}}$. One can then estimate the equilibrium density $\rho_{\text{eq}}(\alpha)$ as:

$$\rho_{\text{eq}}(\alpha) \approx \begin{cases} 0 & \text{if } \alpha \leq z^{-1} \\ \frac{1}{24}(\alpha z - 1) & \text{if } \alpha > z^{-1} \end{cases} \ . \tag{3.53}$$

This result suggests the presence of a second-order phase transition at $\alpha = 1$ from a swollen state characterised bey $\rho = 0$ to a more dense phase with $\rho > 0$. Numerical

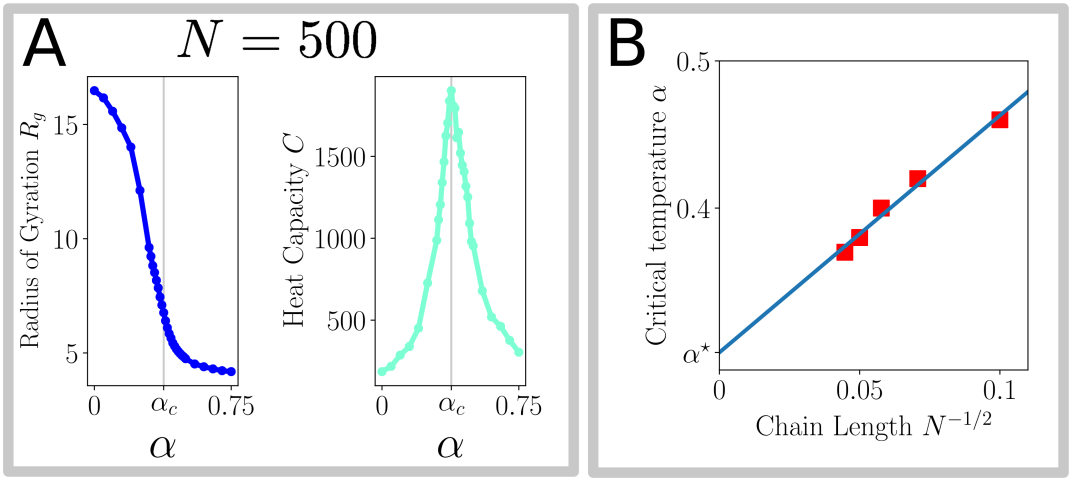



**Figure 3.7:** Results of Montecarlo simulation of chains with Hamiltonian (3.47). **(A)** Plots of radius of gyration $R_g$ and reduced heat capacity $C$ as a function of the temperature parameter $\alpha$, obtained via simulations of chains with $N = 500$. Results show a second-order phase transition between a swollen phase (high $R_g$) and a compact one (low $R_g$) at the inverse temperature $\alpha_c \approx 0.37$ located thanks to the peak in the plot of $C$. **(B)** We estimated the critical inverse temperature $\alpha_c$ for chains of length $N = 100, 200, 300, 400, 500$ and plotted as a function of $N^{-1/\nu}$ with $\nu = 2$ [115]. Interpolation of the results gives us an estimate of the critical temperature at the thermodynamic limit $\alpha^\star \approx 0.298$.

minimizations of (3.51) confirm the presence of second-order phase transition at $\alpha_c = 1/z$ (see fig. 3.6). We have shown that our mean-field estimation for the free energy of an homopolymer is consistent with what classically found in the study of polymer solutions, as it shows the existence of a continuous phase transition in the coil-globule collapse.

Finally, we want to verify how good are our predictions, by comparing the mean-field result with the ones we obtain by employing Montecarlo simulations. We perform the simulations by following the procedure described in section 3.2, and considering chains of increasing size in the range $N = 100 - 500$, on a cubic lattice ($z = 6$). Due to some technicalities, it is difficult to measure directly the parameter $\rho$; therefore, as a measure of the state of the system, we employ the radius of gyration $R_g$ of the chain. Finally, we define the "reduced" heat capacity $C$ as :

$$C = \langle (H_{\text{ISAW}} - \langle H_{\text{ISAW}} \rangle)^2 \rangle \quad , \tag{3.54}$$

that is the variance of the chain energies at equilibrium. Heat capacity $C$ diverges in the thermodynamic limit [110] as $\alpha$ approaches its critical value $\alpha_c$. Using this property, we can estimate the value of $\alpha_c$ for each chain length $N$ (see fig. 3.7A) , and interpolate a critical value $\alpha^\star$ by using finite-size scaling analysis: $\alpha_c = \alpha^\star + kN^{-1/\nu}$,

where $\nu$ is the cross-over exponent and in mean-field has been estimated to be $\nu = 2$ [107, 115] (see fig. 3.7B).

Using this procedure we obtain $\alpha_c^\star \approx 0.30 \pm 0.02$, that is compatible with the value found in [107]. Notice, however, that the Montecarlo result does not agree with the mean-field result that in a cubic lattice is $\alpha_c = 1/6$. This is probably due to a bad estimate of the average number of contacts $n_c(\rho; z)$ made in (3.50).

## 3.5 A Mean-Field Model for Magnetic Polymers

In this section we develop a mean-field theory for a polymer, modeled as SAW on a lattice, with magnetic bead-bead interactions. As in section 3.3, we will develop the theory for a generic Potts-like model with $\nu$ active interacting states, and $\mu$ magnetically inert states, by focusing on the $\nu = 2, \mu = 1$ model corresponding to the one studied in the previous chapters (see section 1.4 and chapter 2) via molecular dynamics simulations.

As discussed in the previous sections, when dealing with a magnetic system with several states $q = \{1, 2, \dots, \nu + \mu\}$ it is convenient to choose a privileged state (e.g. $q = 1$), while forcing the other states to have be equally probable. There are two main ways to do this:

1. (One-field approach) We map the privileged state $q = 1$ into the spin $S = 1$; we suppose *all* the other states (both active and inert) be equally probable and are mapped into the spin state $S = -\frac{1}{\nu+\mu-1}$ (see section 3.3.2).
2. (Two-fields approach) We map the privileged state $q = 1$ into the spin $S = 1$; we suppose that the active states are equally probable and mapped into the spin state $S = -\frac{1}{\nu-1}$; the $\mu$ inert states are all mapped into the null spin $S = 0$ (see section 3.3.3 and section 3.3.4).

We have shown that the first approach is not effective when describing the criticality of a model with inert states ($\mu \geq 1$). However, when dealing with self-avoiding walks, we will show that qualitatively all the main properties of the model can be explained by employing only the average magnetisation of the chain.

For these reasons, we will still first study the $\nu = 2, \mu = 1$ model using the first (one-field) approach, and we will show that the results are compatible with what obtained with the two-field one. Finally, we will generalize the two-field approach to study the interaction in the general case of $\nu \geq 2$, $\mu \geq 0$ states.

### 3.5.1 One-Field Approach

We present here, a first method to study the dynamics of magnetic polymer at equilibrium when we have $\nu = 2$ interacting states, and $\mu = 1$ inert states [3]. The energy of a chain modeled by an $N$-step SAW on a lattice of $M$ sites is the same as

in (3.1), and for a particular configuration $\gamma \in$ SAW it reads:

$$H_\gamma = \frac{1}{2}\sum_{i,j} \Delta^\gamma_{i,j} \mathcal{I}(S_i, S_j) \ , \tag{3.55}$$

where we recall that $\Delta^\gamma_{i,j}$ is the contact matrix of the SAW $\gamma$ (3.2), and $\mathcal{I}$ is the magnetic interactions between spins. The particular form of $\mathcal{I}$ depends on the approach chosen; as a first step we use the "one-field approach" where we map a privileged state in the spin $S = 1$, while the other two are mapped in the spin $S = -\frac{1}{2}$. Following (3.12), in the case of $\nu = 2$, $\mu = 1$, the magnetic interaction $\mathcal{I}$ reads:

$$\mathcal{I}(S_i, S_j) = -\frac{5}{9}\epsilon \left(S_i + \frac{1}{5}\right)\left(S_j + \frac{1}{5}\right) - \frac{1}{5}\epsilon \ . \tag{3.56}$$

Hence a general form for the partition function of the model at temperature $T$ is:

$$\mathcal{Z} = \sum_{\{\gamma\}}\sum_{\{S\}} \exp\left[\frac{\beta\epsilon}{2}\sum_{i,j}\left(\frac{5}{9}(S_i + \tfrac{1}{5})\Delta^\gamma_{i,j}(S_j + \tfrac{1}{5}) + \frac{1}{5}\Delta^\gamma_{i,j}\right)\right] \ , \tag{3.57}$$

where the sums are over every possible SAW $\gamma$, and every possible spin configuration $\{S\}$. Interestingly, while the constant term in (3.12) is usually discarded in lattice-wide models, here it is relevant as it is multiplied by the contact matrix $\Delta^\gamma$.

As a first step, we treat the magnetic component of partition function (3.57), by performing an Hubbard-Stratonovich transformation in the field $\phi$, and summing over every possible spin configuration $\{S\}$. Similarly to (3.15), we obtain:

$$\mathcal{Z} \sim \sum_{\{\gamma\}} \int_{\mathbb{R}^N} \exp\left[\sum_{i,j}\left(-\frac{9}{10}\frac{1}{\alpha}\phi_i \left(\Delta^\gamma_{ij}\right)^{-1}\phi_j + \frac{1}{5}\alpha\Delta^\gamma_{ij}\right) + \sum_i \log\left(e^{\frac{6}{5}\phi_i} + 2e^{-\frac{3}{10}\phi_i}\right)\right] \ , \tag{3.58}$$

where $\alpha = \beta\epsilon$. Let us disregard for now the sum over self-avoiding walks configurations, and consider only the magnetic terms. As done in the previous sections, we employ both saddle-point, and mean-field, approximations to obtain an estimate for the magnetic component of $\mathcal{Z}$:

$$\mathcal{Z}^\gamma_{\text{magnetic}} \sim \exp\left[-\frac{9}{10}\phi^2 \sum_{i,j}\left(\Delta^\gamma_{ij}\right)^{-1} + M\log\left(e^{\frac{6}{5}\phi} + 2e^{-\frac{3}{10}\phi}\right)\right] \ , \tag{3.59}$$

which is in similar in form to what studied before, but presents the peculiar term $\sum_{i,j}\left(\Delta^\gamma_{ij}\right)^{-1}$. This term, generally, depends on the particular choice of SAW $\gamma$ and it is difficult to estimate. However, we know that if we have an Hamiltonian path (i.e. a SAW where $N = M$), then the result of this sum is the same as the one we find for lattice contact matrix $\sum_{i,j}\Delta^{-1}_{ij} = M/z$. This results gives us an insight on the physical interpretation of our summation term: we can estimate it as $M$ over the

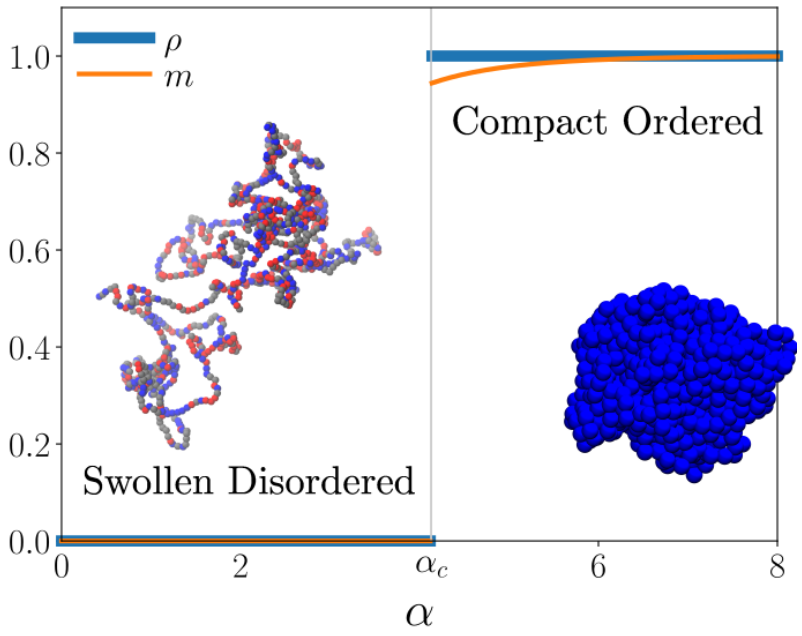



**Figure 3.8:** Density $\rho$ and magnetisation $m$ as a function of $\alpha$ at equilibrium described by free energy (3.60); magnetisation has been computed via formula (3.20). Notice that at $\alpha_c \approx 3.95$ we observe that both order parameters display a finite jump indicating a first order phase transitions between a swollen-disordered phase and a compact-ordered one. The insets report snapshots from molecular dynamics simulations as described in section 1.4.3 and are representative of the swollen-disordered and compact-ordered phase.

average number of contacts that each monomer has. As discussed in section 3.4.2, if we consider SAW with high density $\rho \sim 1$ (we recall that $\rho = N/M$), then we can approximate the average number of contacts as $\rho z$, and therefore:

$$\sum_{i,j} \left(\Delta_{ij}^{\gamma}\right)^{-1} \approx \frac{M}{\rho z} \ .$$

Finally, only the term $\sum_{\{\gamma\}} \exp\left[\frac{1}{5}\alpha\Delta_{ij}^{\gamma}\right]$ in (3.57) remains to be discussed. Notice, however, that this term describes the partition function of an interacting SAW whose monomers interact via an attractive interaction of intensity $-\frac{1}{5}\epsilon$. This is exactly the situation we have studied in section 3.4.2; using (3.51), and in the limit of $\rho \sim 1$, we have:

$$\frac{1}{k_B T} \log\left(\sum_{\{\gamma\}} \exp\left[\frac{1}{5}\alpha\Delta_{ij}^{\gamma}\right]\right) \sim -N\frac{1-\rho}{\rho}\log(1-\rho) + \frac{N}{10}\alpha\rho z \ .$$

We are now ready to write an estimate of the mean-field free-energy per monomer

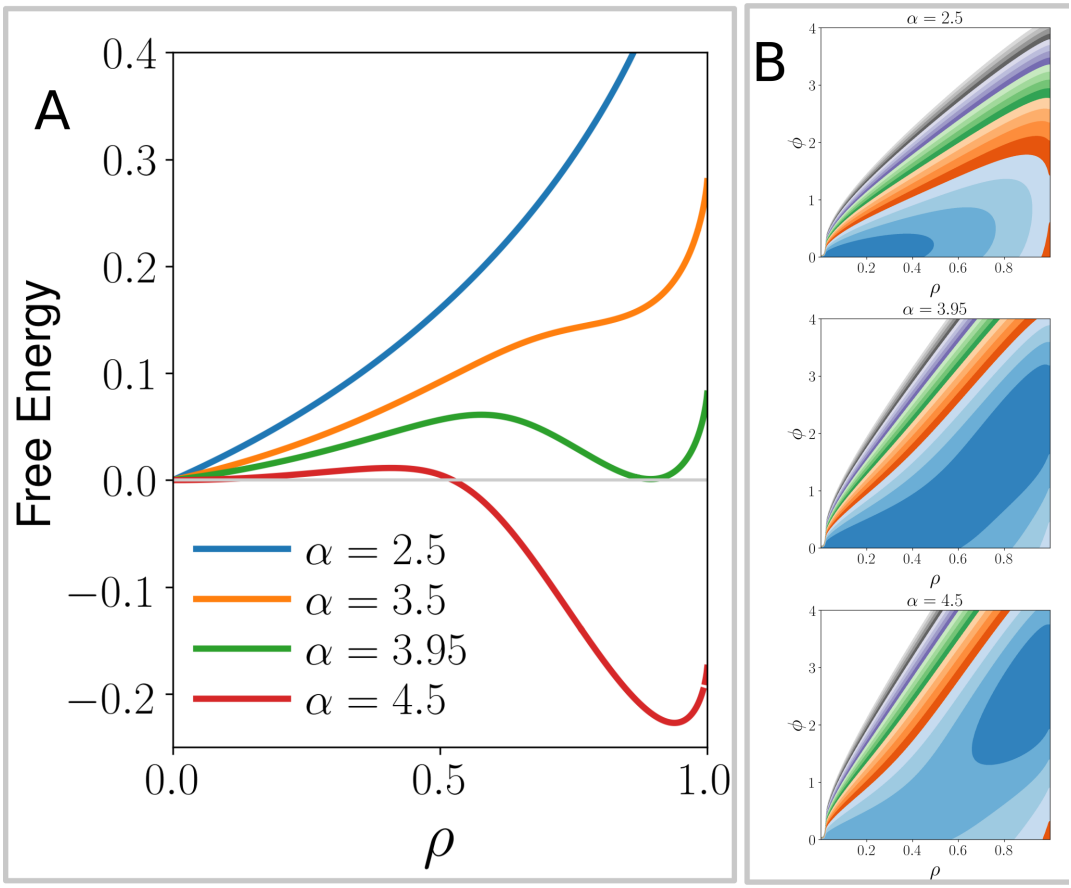



**Figure 3.9:** **(A)** Reduced free energy $f^\star(\rho;\alpha)$ as a function of $\rho$ for various values of $\alpha$. $f^\star$ has been found by numerically solving equation (3.61) to find the function $\phi^\star(\rho;\alpha)$ and using it in the free energy (3.60). Notably we observe that increasing the value of $\alpha$ a minima far away from $\rho = 0$ is created, inducing a first-order phase transition at $\alpha \approx 3.95$. **(B)** Contour plots of free-energy (3.60) as a function of the two fields $\phi$ and $\rho$ shows that for $\alpha < \alpha_c \approx 3.95$ the system is found at low $\phi$ and $\rho$ values (swollen chain, incoherent magnetisation), while if $\alpha > \alpha_c$ the system present high density and big values of $\phi$ suggesting a compact chain with coherent magnetisation.

$f = -\frac{k_B T}{N} \log \mathcal{Z}$ as:

$$\frac{1}{k_B T} f = \frac{1-\rho}{\rho} \log(1-\rho) - \frac{1}{10}\alpha\rho z + \frac{9}{10}\frac{\phi^2}{\alpha z} - \log\left(e^{\frac{6}{5}\phi} + 2e^{-\frac{3}{10}\phi}\right) \ . \tag{3.60}$$

The equilibrium phases of this free-energy can be estimated by minimizing it in terms of the two fields $\rho$ and $\phi$. Namely, by imposing $\frac{\partial f}{\partial \phi} = 0$, one obtains:

$$\frac{9\phi}{5\alpha\rho z} - \frac{\frac{6}{5}e^{\frac{6}{5}\phi} - \frac{3}{5}e^{-\frac{3}{10}\phi}}{e^{\frac{6}{5}\phi} + 2e^{-\frac{3}{10}\phi}} = 0 \ . \tag{3.61}$$

It would be tempting here to use this expression for $\rho$ as a function of $\phi$ at equilibrium and substitute in (3.60), to do a series analysis as we did in the previous sections. However, due to the boundary constraints on $\rho$ ($0 \leq \rho \leq 1$), this kind analysis can be tricky, and for this reason we decided to do the opposite: numerically compute from (3.5.1) the value of $\phi$ as a function of $\rho$, and substitute it in the free energy to obtain the reduced free energy $f^\star(\rho;\alpha)$ at the equilibrium. By plotting this reduced free energy (see fig. 3.9), we observe that it has always a minimum in $\rho = 0$. On the other hand, for high values of $\alpha$ ($\alpha \gtrsim 3.4$) a new global minimum at higher density is formed, suggesting a first-order phase transition between a swollen phase (low $\rho$) to a compact phase (high $\rho$) at $\alpha = \alpha_c \approx 3.95$. Furthermore, if we numerically minimize the free energy (see fig. 3.8), and employ the definition of magnetisation found in (3.20), we observe that there is a first-order phase transition between a swollen-disordered phase, and a compact ordered one, confirming our results obtained by molecular dynamic simulations in section 1.4.3.

### 3.5.2 Two-Fields Approach

We now briefly present a mean-field theory for a magnetic polymer with a generic number $\nu \geq 2$ of interacting states, and generic number $\mu \geq 0$ of inert states. To do so, we will follow what we called the "two-fields approach" (see section 3.3.4). By considering an interaction of the form (3.34); the partition function will read:

$$\mathcal{Z} = \sum_{\{\gamma\}} \sum_{\{S\}} \exp\left[ \frac{\beta\epsilon}{2} \sum_{i,j} \left( a(S_i^2 - bS_i)\Delta_{ij}^{\gamma}(S_j^2 + bS_j) + cS_i\Delta_{i,j}^{\gamma}S_j \right) \right] \quad , \tag{3.62}$$

where $a, b, c$ are parameters which depend only on the number of active states $\nu$ as in (3.35). As in the previous cases, we perform two successive Hubbard-Stratonovich transformations to introduce the fields $\phi_i$ and $\psi_i$:

$$\begin{aligned} \mathcal{Z} \sim \sum_{\gamma} \int_{\mathbb{R}^{2N}} \exp\Bigg[ \frac{1}{2\alpha} \sum_{i,j} \left( \frac{1}{a}\psi_i \left(\Delta_{ij}^{\gamma}\right)^{-1} \psi_j + \frac{1}{c}\phi_i \left(\Delta_{ij}^{\gamma}\right)^{-1} \phi_j \right) + \\ - \sum_i \log\left( e^{\frac{\psi_i}{\nu-1}} (e^{\phi_i} + (\nu-1)e^{-\frac{\phi_i}{\nu-1}}) + \mu \right) \Bigg] \mathrm{d}\phi\, \mathrm{d}\psi \quad . \end{aligned} \tag{3.63}$$

By proceeding as in the previous section, we obtain a mean-field expression of the free-energy $f(\phi, \psi, \rho; \alpha)$ as:

$$\begin{aligned} \frac{1}{k_B T} f = \frac{1-\rho}{\rho} \log(1-\rho) + \frac{\nu}{(\nu-1)^2} \frac{\psi^2}{2\alpha\rho z} + \frac{\nu}{\nu-1} \frac{\phi^2}{2\alpha\rho z} + \\ - \log\left( e^{\frac{\psi}{\nu-1}} \left( e^{\phi} + (\nu-1)e^{-\frac{\phi}{\nu-1}} \right) + \mu \right) \quad . \end{aligned} \tag{3.64}$$

As a first step, let us study the model when $\nu = 2$ and $\mu = 1$ in order to further verify the correctness discussed in the "one-field" approach of the previous section. In

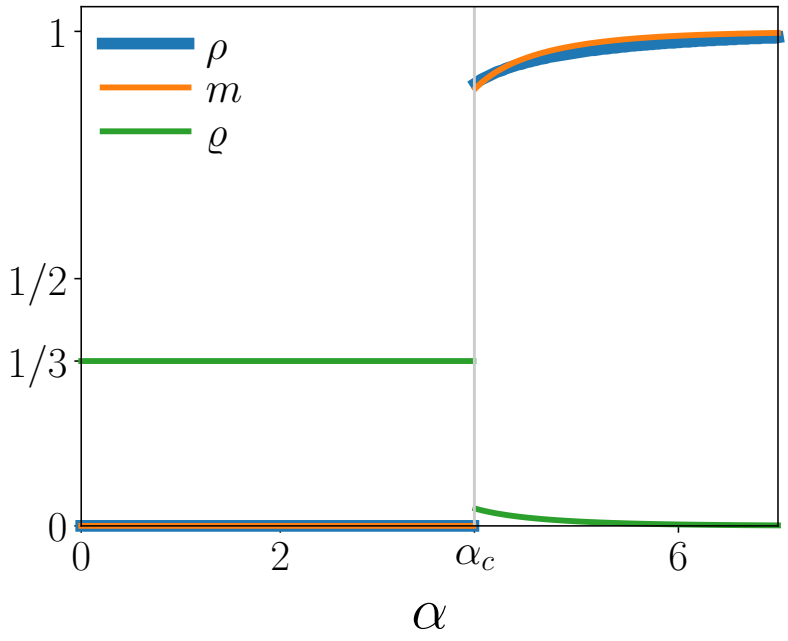



**Figure 3.10:** Results of the minimization of the free energy (3.65) as a function of $\alpha$; the order parameters magnetisation $m$, and inert states density are found by employing equations (3.40). We observe a first-order phase transition between a swollen-disordered state ($|m| = \rho = 0$), to a compact ordered phase ($\rho > 0, |m| > 0$), with critical inverse temperature $\alpha_c \approx 3.95 z^{-1}$.

this particular case, free-energy (3.64) reads:

$$\frac{1}{k_B T} f = \frac{1-\rho}{\rho} \log(1-\rho) + \frac{\psi^2}{\alpha \rho z} + \frac{\phi^2}{\alpha \rho z} - \log\left(2e^{\psi}\cosh(\phi) + 1\right) \quad . \tag{3.65}$$

By numerically minimizing this free energy we obtain the plots shown in fig. 3.10. We observe, again, the presence of a first-order phase transition at $\alpha_c \approx 3.95 z^{-1}$, in good agreement with what observed in the "one-field approach", suggesting its validity.

### Comparison with Montecarlo Simulations

As a final step, we compare our results with what observed Montecarlo simulations of the model. A complete Montecarlo experiment of the system must evolve not only the spatial configuration of an ISAW, but also its spin configuration. In order to consider these two processes, for a chain of length $N$, we define a single "Montecarlo step" as composed by the two following sub-steps [107]:

1) a configurational Montecarlo step, where we modify the spatial configuration of our SAW, by following the procedure described in section 3.2;

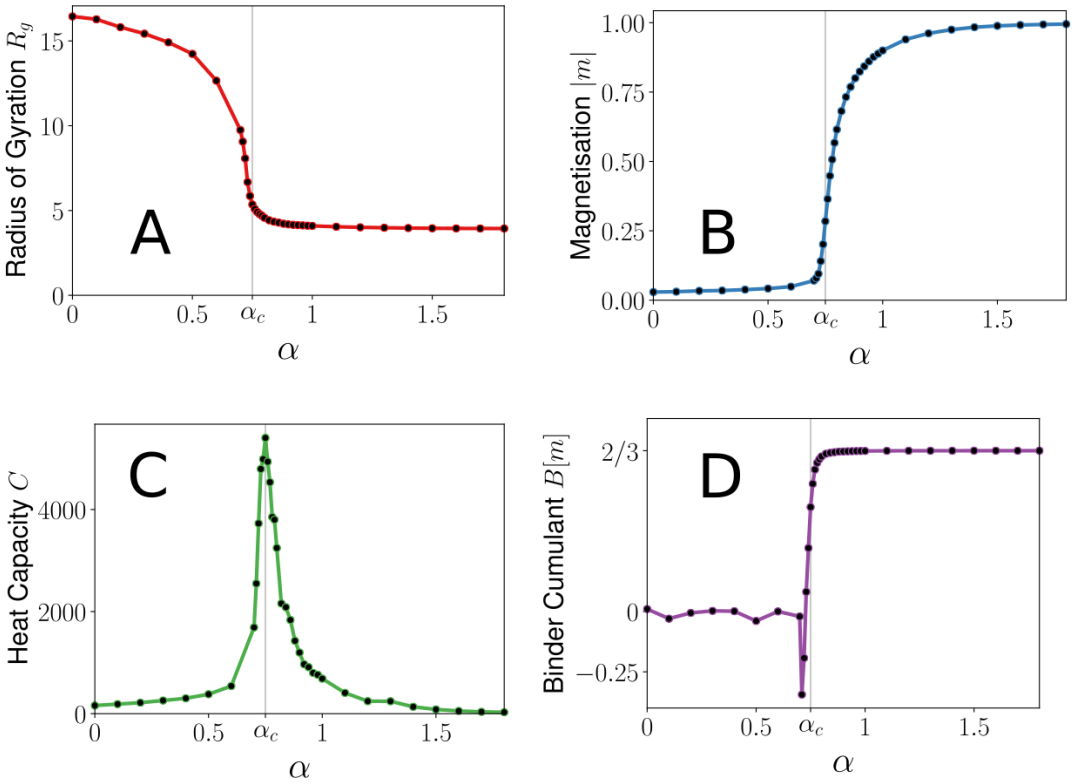



**Figure 3.11:** Results of Montecarlo simulations for a chain of length $N = 500$, whose beads can have three states $q = \{1, 2, 3\}$ with interaction (3.3). **(A-B)** The plots of radius of gyration $R_g$ and magnetisation $m$ as a function of $\alpha = \beta\epsilon$ show the presence of a phase transition between a swollen-disordered phase (high $R_g$, $|m| \approx 0$), to a compact-ordered phase (small $R_g$, $|m| > 0$. **(C)** The critical temperature parameter $\alpha_c \approx 0.75$ between the two phases can be estimated via the peak of the reduced heat capacity, defined as the variance of the energy at equilibrium. **(D)** The Binder cumulant of the magnetisation (3.66) sharply becomes negative nearby the critical point $\alpha \approx \alpha_c$, before becoming positive again and reaching the value $\frac{2}{3}$. This behaviour indicates that the phase transition is of the first order.

2) $N$ independent color flips, where we randomly change the color of a single site, and accept the change using the usual Metropolis-Hastings criterion [20].

Each of these Montecarlo steps is then repeated for a number of times (about $10^3 \times N$), until the system reach a steady state.

We simulate the $\nu = 2, \mu = 1$ model for chains of length $N$ in the range $N = 50 - 500$. The Montecarlo approach, together with the molecular dynamics simulations presented in section 1.4.3 confirm the presence of a first order phase transition. Namely,

to examine the order of the phase transition, we estimate the Binder cumulant $B(m)$:

$$B(m) = 1 - \frac{\langle m^4 \rangle}{3 \langle m^2 \rangle^2} \ , \tag{3.66}$$

whose properties have already been discussed in chapter 1; we observe that as $\alpha$ approaches the critical value $\alpha_c$, the Binder cumulant becomes sharply negative, as expected in first-order phase transitions. Finally, as done in section 3.4.2, we evaluate an effective critical value $\alpha_c(N)$ for each finite size chain size $N$ by estimating the position of the peak of the reduced heat capacity $C(\alpha)$. From these estimates we can extrapolate a value for the critical inverse temperature at the thermodynamic limits $\alpha^\star$, by interpolating the law $\alpha_c(N) = \alpha^\star + bN^{-1/2}$. We find the value $\alpha^\star \approx 0.63 \pm 0.01$, to be compared with the mean-field value $\alpha_c^{\mathrm{mf}} = 3.95 z^{-1}$, which in the cubic lattice yield $\alpha_c^{\mathrm{mf}} = 0.658$, quite close to the Montecarlo estimate.

### Generic Number of States, and Limits of the Mean-Field Model

After having examined in details the case $\nu = 2$, $\mu = 1$, let us now discuss what happens if we consider an arbitrary number of active states $\nu \geq 2$, and inert states $\mu \geq 0$. Given the difficulties we find in minimizing analytically the free-energy (3.64), we rely mostly on numerical minimization.

In the case of a two-states classic Potts model ($\nu = 2, \mu = 0$), we observe a peculiar behavior in the mean-field dynamics. Namely, we observe that by varying the value of $\alpha$, the system undergoes two different phase transitions. One at $\alpha z \approx 2$ that is a continuous phase transition from a swollen-disordered phase ($|m| = 0$, $\rho = 0$) to a compact-disordered ($|m| = 0$, $\rho > 0$) configuration with a continuous-phase transition akin to the classic $\theta$-point transition of the Flory model. A second, first order, phase transition is observed at $\alpha z \approx 3.1$ between the compact-disordered phase and a compact-ordered one ($|m| > 0$, $\rho > 0$). In all the other cases ($\nu \geq 2$, $\mu > 0$) we always observe only a first-order phase transition between the swollen-disordered and compact-ordered phases, much similar qualitatively to the one displayed in fig. 3.10.

In order to verify the correctness of the mean-field previsions at $\nu = 2$, $\mu = 0$, we performed both some Montecarlo studies with chain of length $N = 500$, and molecular dynamics simulations akin to the ones discussed in section 1.4. In both these analysis, we can only find, as always, a first-order phase transition between the swollen-disordered and compact-ordered phase, with no trace of compact-disordered configurations. We conclude that, in this case, the mean-field model approximation fails to describe properly the equilibrium properties of the system. Indeed, all the hypothesis we made to compute free-energy (3.65) were based on the assumption that the SAW density $\rho$ is close to $\rho \sim 1$. This means that our free-energy is not suitable to study small density phases as exemplified by the $\nu = 2, \mu = 0$ case. Notably, one of the possible corrections one could do is on the estimate of the number of contacts (3.50), which indeed is well justified if $\rho \sim 1$, but loses all meaning when $\rho \sim 0$ where

we have $n_c(0) = 0$, against the very definition of SAW which ensures that the number of contacts must always be $n_c \geq 2$.

## 3.6 Out-Of-Equilibrium Dynamical Equations

In the previous section, we have developed a mean-field theory to determine the equilibrium phase at various temperatures, displayed by a magnetic polymers. The equilibrium phases are characterized by the introduction of some mean-field variables: the chain density $\rho$, and the fields $\phi$ and $\psi$ (only in the "two-fields approach"). In this section, we are interested in characterizing the dynamics of these variables, rather than their equilibrium value, and, possibly, try to extend the mean-field theory out of equilibrium, in order to discuss the results of molecular dynamics presented in section 1.4.4.

Before we present our dynamical model, we must decide which set of mean-field variables we are going to consider. We can either use the free-energy (3.60) found with the "one-field approach" with mean-field variables $(\phi, \rho)$, or the free-energy (3.65) found with the "two-fields approach" and variables $(\phi, \psi, \rho)$. We have shown in the previous section that the two free-energies are equivalent when studying the dynamics of magnetic SAWs, therefore for simplicity we consider the one obtained by "one-field approach" as there are only two fields to treat.

The dynamics will be described by a set of "model A" equations [20], which are found in analogy with Hamiltonian mechanics. Namely, we assume to have a spatially extended domain $V$ characterized by free energy density as in (3.60), and that the dynamics of its fields $\phi(\mathbf{x})$ and $\rho(\mathbf{x})$ is determined by a functional Hamiltonian $\mathcal{H}$ in the form:

$$\mathcal{H}[\rho, \phi] = \int_V \left[ f(\rho, \phi; \alpha) + \kappa_\phi (\nabla\phi)^2 + \kappa_\rho (\nabla\rho)^2 \right] \mathrm{d}\mathbf{x} \ , \tag{3.67}$$

where $\kappa_\phi, \kappa_\rho$ are surface-tension like coefficients which quantify the rates of diffusion of our fields $\phi$ and $\rho$. The dynamics of the fields, will be, therefore determined by the Hamilton-like equations:

$$\begin{cases} \dfrac{1}{\Gamma_\phi}\dfrac{\partial\phi}{\partial t} = -\dfrac{\delta\mathcal{H}}{\delta\phi} \\ \dfrac{1}{\Gamma_\rho}\dfrac{\partial\rho}{\partial t} = -\dfrac{\delta\mathcal{H}}{\delta\rho} \end{cases} , \tag{3.68}$$

where $\Gamma_\phi, \Gamma_\rho$ are mobility-like coefficients with a role much similar to the one mass have in the dynamics of particles in classic Hamiltonian systems. Using the expression for free energy found in (3.60), equations (3.68) reads:

$$\begin{cases} \dfrac{\partial\phi}{\partial t} = \Gamma_\phi \left( -\dfrac{9}{5}\dfrac{\phi}{\alpha\rho z} - \dfrac{3}{e^{\frac{3}{2}\phi} + 2} + \dfrac{6}{5} + \kappa_\phi \nabla^2\phi \right) \\ \dfrac{\partial\rho}{\partial t} = \Gamma_\rho \left( \dfrac{9}{10}\dfrac{\phi^2}{\alpha\rho^2 z} + \dfrac{\alpha z}{10} + \dfrac{\log(1-\rho)}{\rho^2} + \dfrac{1}{\rho} + \kappa_\rho \nabla^2\rho \right) \end{cases} . \tag{3.69}$$

Before studying these equations, let us examine how we introduced the out-of-equilibrium dynamics in section 1.4.4. We assumed that during the Langevin dynamics associated to the field $\rho$, and the recoloring dynamics associated with the field $\phi$, evolve generally thanks to interactions with different intensities. We try to mimic this procedure, by introducing two independent quantities $\alpha_\phi = \beta\epsilon_\phi$ and $\alpha_R = \beta\epsilon_R$ which will substitute $\alpha$ in equations (3.69):

$$\begin{cases} \dfrac{\partial\phi}{\partial t} = \Gamma_\phi \left( -\dfrac{9}{5}\dfrac{\phi}{\alpha_\phi \rho z} - \dfrac{3}{e^{\frac{3}{2}\phi} + 2} + \dfrac{6}{5} + \kappa_\phi \nabla^2 \phi \right) \\ \dfrac{\partial\rho}{\partial t} = \Gamma_\rho \left( \dfrac{9}{10}\dfrac{\phi^2}{\alpha_\rho \rho^2 z} + \dfrac{\alpha_\rho z}{10} + \dfrac{\log(1-\rho)}{\rho^2} + \dfrac{1}{\rho} + \kappa_\rho \nabla^2 \rho \right) \end{cases} . \tag{3.70}$$

We have decoupled the effective inverse temperature $\alpha$ into two parameters $\alpha_\phi, \alpha_\rho$, which separately affect the dynamics of the polymer ($\alpha_\rho$) or of the epigenetic field ($\alpha_\phi$). Notice that, if we try to apply definition (3.68) to this last set of equations, we obtain that a corresponding Hamiltonian $\mathcal{H}$ exists if and only if $\alpha_\phi = \alpha_\rho$; therefore, we conclude that in the case of $\alpha_\phi \neq \alpha_\rho$ the dynamics equations (3.70) do not describe the dynamics of an equilibrium system.

We study equations (3.70) by numerically integrating them over a $50 \times 50$ grid, by employing a mid-step method with time-step $\Delta t = 0.01$. The choice of the mobility and surface tension coefficients $\Gamma_{\phi/\rho}, \kappa_{\phi/\rho}$ affects only details of the kinetics of the system, but not the qualitative properties of the steady-state for long times. We set their values to $\Gamma_\phi = 10^{-1}$, $\Gamma_\rho = 10^{-3}$, and $\kappa_\phi = \kappa_\rho = 2$. We tested the system by employing several initial conditions of the fields $\phi$ and $\rho$, and found out that if we set $\rho \sim 1$ than for certain values of the parameters $(\alpha_\phi, \alpha_\rho)$, the system flows to local minima. To prevent this, we employ random initial conditions broadly distributed where the average value of the density is $\bar{\rho} = 0.6$.

By using this procedure, we look at the long-time steady states for the system in terms of the fields $\rho$ and $\phi$, and varying the values of the interaction parameters $(\alpha_\rho, \alpha_\phi)$. The results allow to draw a non-equilibrium phase diagram of our system shown in fig. 3.12C.

Coherently with the results of the previous section, we observe that along the equilibrium line $\alpha_\phi = \alpha_\rho$, the system goes from a swollen disordered phase ($\rho = 0$, $\phi = 0$), to a compact-ordered one ($\rho > 0$, $\phi > 0$). Outside the equilibrium line, not only we still observe the two phases, but also a third phase for large $\alpha_\rho$ and small $\alpha_\phi$ appears. This phase is characterized by a compact configuration ($\rho > 0$) and null magnetization ($\phi = 0$), and it is therefore reminiscent of the compact-disordered phase observed in the out-of-equilibrium phase diagram found with molecular dynamics (see section 1.4.4).

Notably, in chapter 1 we also observed a fourth, swollen-ordered phase ($\rho = 0$, $\phi > 0$) in correspondence of the region with small $\alpha_\rho$ and large $\alpha_\phi$. In this context, instead, we find that this region of parameter values is still dominated by the compact-ordered phase. This should not surprise us since in our calculation we always employed

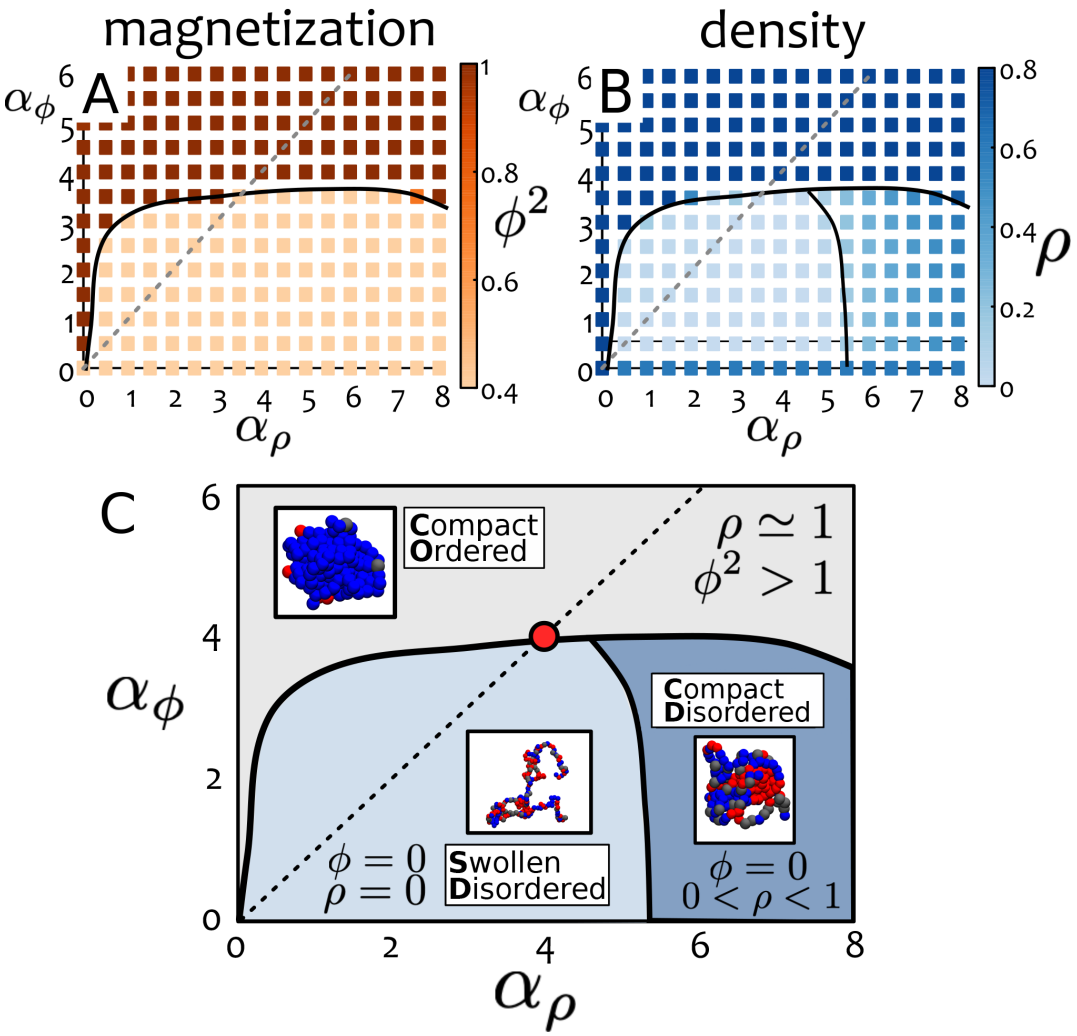



**Figure 3.12:** **(A-B)** Steady states of the field $\phi$ and density $\rho$ as a function of $\alpha_\phi, \alpha_\rho$. Results are obtained by integrating equations (3.70) on a $50 \times 50$ grid with $\Gamma_\phi = 10^{-1}$, $\Gamma_\rho = 10^{-3}$, and $\kappa_\phi = \kappa_\rho = 2$. **(C)** By comparing the two plots for $\phi$ and $\rho$ we are able to draw a qualitative phase diagrams for our dynamics, which shows the existence of three distinct phases: compact-ordered, compact-disordered and swollen-disordered. The red circle denotes the first-order transition observed at in equilibrium. Insets show snapshots of representative configurations from molecular dynamics of the magnetic polymer as in (1.4.4).

the hypothesis of $\rho \sim 1$, and therefore should be not very accurate at low-densities.

### Swollen-Disordered to Compact-Disordered Phase Transition

Let us briefly examine the phase transition between the swollen-disordered and compact-disordered phases, which is observed by keeping $\alpha_\phi$ small, and varying the parameter $\alpha_\rho$. In order to simplify the argument, we study equations (3.70) in the limit of $\alpha_\phi \ll 1$. In this limit, the dynamics of the field $\phi$ is dominated by the linear term in(3.70):

$$\frac{\partial \phi}{\partial t} \sim -\frac{9\Gamma_\phi}{5\alpha_\phi z \rho}\phi \ ,$$

which, if $\rho > 0$, it implies that $\phi(t) \sim \exp[-\frac{9\Gamma_\phi}{5\alpha_\phi z\rho}t]$. This means that $\phi$ reaches exponentially fast the $\phi = 0$ whenever $\rho > 0$, and $\alpha_\phi \ll 1$. On the other hand, if $\rho = 0$, the only possibility for equations (3.70) not to diverge is that $\phi = 0$ too. We conclude that if $\alpha_\phi \ll 1$, then we can safely assume that $\phi \equiv 0$ during the whole dynamics. In this case, we have to deal only with the dynamics of the field $\rho$ which will read:

$$\frac{\partial \rho}{\partial t} = \Gamma_\rho \left( \frac{\alpha_\rho z}{10} + \frac{\log(1-\rho)}{\rho^2} + \frac{1}{\rho} + \kappa_\rho \nabla^2 \rho \right) \ , \tag{3.71}$$

Notice that this expression is the "model A" equation for a field theory described by free-energy (3.51) with $\alpha = \frac{1}{5}\alpha_\rho$, which describes a self-interacting homopolymer. We conclude therefore that in the limit $\alpha_\phi \ll 1$ the system shows a coil-globule transition akin to what observed for homopolymers, suggesting that the non-equilibrium phase transition we observe is continuous.

### Compact-Disordered to Compact-Ordered Phase Transition

Finally, we examine the compact disordered to compact-ordered phase transition when $\alpha_\rho \gg 1$. In this limit, the most relevant term for the dynamics of the field $\rho$ is the term:

$$\frac{\partial \rho}{\partial t} \sim \Gamma_\rho \frac{\alpha_\rho z}{10} \ ,$$

which indicates that in the first instants of the dynamics the density grows almost linearly, until it reaches a value $\rho \approx 1$. In this limit, the dynamics of the magnetization field $\phi$, will read:

$$\frac{\partial \phi}{\partial t} = \Gamma_\phi \left( -\frac{9}{5}\frac{\phi}{\alpha_\phi z} - \frac{3}{e^{\frac{3}{2}\phi} + 2} + \frac{6}{5} + \kappa_\phi \nabla^2 \phi \right) \ . \tag{3.72}$$

Notice that this equation is what we would obtain if we studied the "model A" equations of the $\nu = 2, \mu = 1$ model as presented in section 3.3.2. This indicates that, in the one-field approach considered, no phase transition will be detected by only using the dynamics described in (3.72). However, if we would have used a two-field approach, we could have found a second-order phase transitions between the compact-disordered phase, and the compact-ordered one.

## 3.7 Discussion

In this chapter we have proposed a field theoretical approach to study the dynamics of magnetic inspired models of epigenetic systems. Our approach has shown successfully that the model studied in section 1.4 can be mapped into an interacting SAW with magnetic-like interactions between its vertices. Namely, we have established via analytical means that there exist a first-order phase transition between the swollen epigenetically disordered phase, and the compact epigenetically coherent one, confirming the results of molecular dynamics simulations presented in the previous chapters.

While our main objective was to study the dynamics of magnetic polymers, we also developed a complete theory for spin-lattice models with Potts-like interactions in the presence of inert states. Notably, we extended the classic Potts theory, by noticing that the number of magnetically inert states can dramatically influence the dynamics, and changes the order of transition.

The methodology we have described in this chapter, being simple enough it can be applied to a broader set of models. For example, we have only studied the case where interacting states interact all identically via a Potts-like interaction. It would be interesting to explore the possibility of modeling asymmetric models which perhaps show non-null interaction between different states. Another fascinating possibility would be to expand the mean-field theory in order to embrace an approximation of higher order, perhaps via an RPA approach [116]. In this context, it would be interesting to study the formation of localized domains by inserting the genomic bookmarking described in chapter 2.

In summary the results presented in this chapter show that a field theory approach to a complex model, like the magnetic polymer ones, can identify the general mechanisms involved in the interaction between epigenomic spreading, and chromatin 3D organization. Moreover, thanks to our strategy, we can push our system out-of-equilibrium to study via analytic means the properties of the model. Such results, would be almost impossible to derive via a far-from-equilibrium system without an underlying effective free-energy.

# Chapter 4

# Epigenomic Phase Separation in the Cell Nucleus

*Never trust the sound of rain upon a river rushing through your ears*

---

Porcupine Tree, `Arriving Somewhere but not here`

In the previous chapters, we have investigated the behaviour of the model proposed in section 1.4 in several ways, ranging from numerical simulations (chapter 1 and chapter 2) to mean-field techniques (chapter 3). Namely, we have studied the dynamics of epigenetic patterns, using a Potts-like model coupled with the chromatin dynamics in order to fully understand the intimate relationship between gene expression, and chromatin architecture. We have found that the model, even in the most basic form, suggest the existence of a first-order-like phase transition between an epigenetically disordered phase, and an ordered one, granting therefore the existence of an "epigenetic memory".

In this chapter, we aim to expand the model to study chromatin organization and epigenetic spreading at the scale of cell nucleus, always inspired by analogies with mahnetic polymer models. Namely, we extend the mean-field theory presented in chapter 3 to include in the discussion the presence of several chromatin fibers in non-dilute environment, by studying a new Landau-Ginzburg field theory [20] where the dynamics of epigenetic marks is linked to that of genome folding within the nucleus. Our theory considers, as in the previous chapters, a 1D chain of Potts-like spins which is allowed to fluctuate in 3D, but the study will be handled via classic field-theory techniques, along with some numerical simulations.

Similarly to what observed in chapter 1, at the equilibrium regime we already capture some features seen *in vivo* in cells, such as segregation of different epigenetic marks, but we fail to explain the experimentally observed coexistence of diverse

epigenetic and genomic domains in eukaryotic nuclei [21]. For this reason, we modify the model to consider non-equilibrium processes, due to generic energy-consuming biochemical and biophysical processes in the nucleus. We discover that a simple first-order reaction leads to arrested, and tunable, phase separation of epigenomic domains, in qualitative agreement with experiments [117].

## 4.1 Construction of a Coarse-Grained Model

In chapter 3 we introduced techniques to study the equilibrium dynamics of a 3D self-avoiding chain with epigenomic spreading via mean-field techniques. Here, we try to expand the study to the case of a multitude $N$ of such chains contained in a limited volume $V$. Namely, we will "construct", via phenomenological considerations, a new free energy density $f$ in the magnetization and density fields. The analysis of this phenomenological free-energy will be mostly conducted via field theory techniques. However, it is useful to employ,at the same time, numerical simulations of the molecular dynamics (here called "Brownian Dynamics") described by the model, in order to verify that the phenomenological model is indeed qualitatively correct.

### Brownian Dynamics

Here, we briefly describe the numerical simulations employed when studying the dynamics of a melt of polymer. As already discussed in section 1.4.2, chromosomes are modelled using semi-flexible bead-spring chains. Each bead is marked with an epigenetic state $q = \{\mathrm{A}, \mathrm{B}, \mathrm{U}\}$, and the dynamics of the chains are described by a set of Langevin equations (1.7) derived by the Hamiltonian (1.8) with attractive interaction $\epsilon_L$. After evolving the dynamics of a $N$-beads long chain for a certain time $\tau_R$, we evolve the color of the beads using $N$ Metropolis moves at temperature $T$ and Hamiltonian (1.13) and considering epigenomic interaction strength $\epsilon_R$ (in general $\epsilon_L \neq \epsilon_R$, see section 1.4.4). This process is repeated several times, until the system achieves a steady state. The molecular dynamics simulations are run at constant-temperature (NVT ensemble) employing the `LAMMPS` engine [18] with a Velocity-Verlet scheme.

To model the whole nucleus dynamics, we perform simulations of a melt of annealed polymers at different monomer densities $\rho = N/V$ and different values of interaction strengths $\alpha_R = \alpha_L$ (equilibrium regime). Usually, we consider $N = 50$ polymers with 256 beads each, and the range of parameters employed are $\rho = 0.1 - 0.8\,\sigma^{-3}$ (with $\sigma$ defining the beads size), and $\alpha_L = 0.75 - 1.1$.

### Phenomenological Field Model

Here we will develop a field-theoretical approach to describe a solution of polymers with epigenomic interactions. In order to derive a consistent field theory, one possible way is to: (i) derive an Edwards-like Hamiltonian [118] from the original one (1.8);

(ii) account for the presence of a melt of polymers [119]; (iii) introduce a Potts-like interaction and recolouring dynamics. While such a procedure is in theory possible, it is of difficult application. Thus, we decided to follow a simpler way: rather than deriving, an exact form for the free-energy density $f$, we introduce some ansatzs and employ a phenomenological form for $f$.

We describe the system by two fields $n(\mathbf{x}, t)$ and $m(\mathbf{x}, t)$; they are respectivelt the density and "magnetization" fields. The former records the local genomic density at a given spatio temporal location, the latter the local abundance of a given epigenetic state. These fields should be interpreted as local averages over a volume element centred in $\mathbf{x}$, large enough to smooth out microscopic fluctuations of the chromatin structure, but much smaller than the typical nucleus size. The equilibrium properties are described by the following free-energy density $f$:

$$\beta f\left[m(\mathbf{x}), n(\mathbf{x})\right] = am(\mathbf{x})^2 + bm(\mathbf{x})^4 + cn(\mathbf{x})^2 + dn(\mathbf{x})^3 - \chi m(\mathbf{x})^2 n(\mathbf{x}) \ , \qquad (4.1)$$

where $\beta = \frac{1}{k_B T}$ is the inverse of the temperature. The phenomenological parameters of the uncoupled system are constant and set to be $a > 0$, $b > 0$, $c > 0$, $d > 0$. The parameter $\chi > 0$, governing the coupling between the epigenetic profile and the chromatin organization, is temperature dependent; more specifically $\chi(T)$ decreases as the temperature increase and $\chi(T) \to 0$ as $T \to +\infty$. Finally, we consider the system contained in a fixed volume $V$; therefore the local density $n$ obeys the following constraint:

$$n_0 \equiv \frac{1}{V} \int_V n(\mathbf{x}, t)\, \mathrm{d}\mathbf{x} = \text{constant} \qquad \forall t \ , \qquad (4.2)$$

where $n_0$ is the average (and constant) density of the system.

In writing equation (4.1) we have employed only two ansatzs so far:

(1) the free-energy $f$ must respect the $\mathbb{Z}_2$ symmetry observed in section 1.4 [1] of the magnetization field (symmetry between the epigenetic marks A/B);

(2) as the magnetization vanishes ($m \to 0$) the Potts-like interaction are non-existent; therefore the model must describe the behaviour of the density field via a standard virial expansion for non-ideal gases [20]. Namely, when $m = 0$ the ground state of the system must be a nucleus homogeneously filled by DNA.

Finally, we choose a minimal coupling term, $\chi m^2 n$ to model the feedback between chromatin folding ($n > 0$) and epigenetic ordering ($m^2 > 0$). In this context, $\chi$ can be thought of as parametrising the self-attraction of equal epigenetic marks, which is mediated by the usual reader-writer machinerie (see section 1.1).

Free-energy (4.1) is in apparent disagreement with the single-chain mean-field energy we have found in the previous chapter (see section 3.5), as in the latter the interaction term between the two fields is inversely proportionaly to the density fields. However, notice that the single-chain model is valid only in the limit of strongly diluted system where $n_0 \ll 1$. Indeed, the coupling term $\chi m^2 n$ becomes ininfluent and therefore other, different, terms should be taken into consideration in this limit.

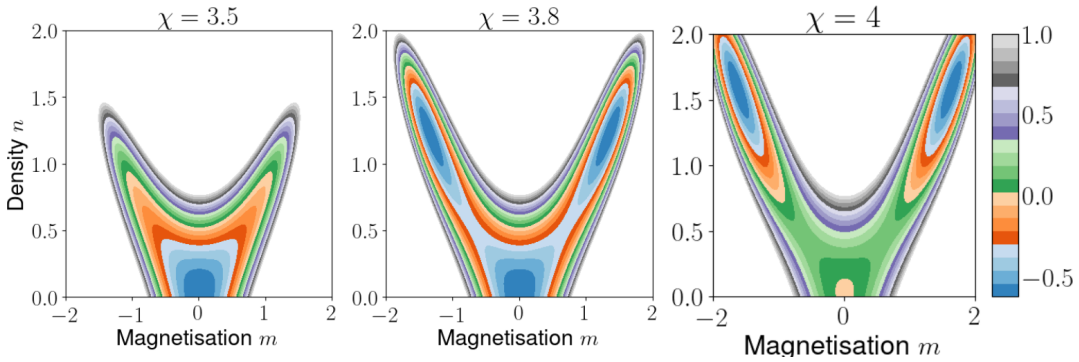



**Figure 4.1:** Countour plots of the free energy $f$ (4.1), with $a = b = c = d = 1$, when no constraint (4.2) is considered. At $\chi = 3.5$ the only minimum is at $m = n = 0$, but increasing the interaction stremgth to $\chi = 3.8$ two other metastable minima (with $m^2 > 0$ and $n > 0$ appear and we have phase coexistence typical of a first-order phase transition. Finally at higher temperatures ($\chi = 4$) the null phase $m = n = 0$ is now become metastable in favour of the ordered phases.

On the other hand, we will prove in section 4.2.3 that even without doing so, the model described by (4.1) correctly describes qualitatively the behaviour of the single-chain system when taking the limit $n_0 \to 0$.

## 4.2 Equilibrium of the Coarse-Grained Model

In this section we will study the behaviour of a model described by the free-energy density (4.1) (with $a = b = c = d = 1$) at the equilibrium. As a first step, let us ignore the constraint given by a finite volume (4.2); doing this we are considering the system in a dilute regime similar to what studied in chapter 3. In this case one can disregard the locality of the free-energy density $f$ and therefore the equilibrium phases are easily found by solving the system of equations $\frac{\partial f}{\partial m} = 0$ and $\frac{\partial f}{\partial n} = 0$, and studying the stability of the steady states found while varying the value of the interaction strength $\chi$.

While the procedure itself is quite simple, explicit results are difficult to present here as they involve solutions of fourth-order polynomial equations. Nevertheless, exact calculations (made thanks to the `Mathematica` software) show the existence of a first order phase transition at $\chi_c \approx 3.8$ (see fig. 4.1) from a "swollen-disordered" phase characterised by $m = n = 0$, and a "compact-ordered" phase with $m^2 > 0$ and $n > 0$, compatibly with what found in section 1.4 and chapter 3. Note that this result does not (qualitatively) depend on the values of the phenomenological constants $a$, $b$, $c$, $d$.

Let us now consider the finite volume constraint (4.2). In this case, we can not disregard the fact that $f$ is a local density, and thus we must define the total

free-energy functional $\mathcal{F}$ as a function of the magnetization $m$ and density $n$ fields:

$$\mathcal{F}[m,n] = \int_V f[m(\mathbf{x}), n(\mathbf{x})]\, \mathrm{d}\mathbf{x} \ . \tag{4.3}$$

Generally, the equilibrium properties of the system are found by minimizing this functional. Here, however, we have a constraint on the density field $n$, and thus we have to resort to employ Lagrange multipliers techniques. Namely, we must minimize the functional:

$$\mathcal{G}[m,n] = \mathcal{F}[m,n] - \mu \int_V [n(\mathbf{x},t) - n_0]\, \mathrm{d}\mathbf{x} \ ,$$

where $\mu$ is a Lagrange multiplier. Minimization of this functional is done by solving the equations $\frac{\delta \mathcal{G}}{\delta m} = 0$ and $\frac{\delta \mathcal{G}}{\delta n} = 0$, which is reduced to the set of equations:

$$\begin{cases} \dfrac{\delta f}{\delta m}[m(\mathbf{x}), n(\mathbf{x})] = 0 \\ \dfrac{\delta f}{\delta n}[m(\mathbf{x}), n(\mathbf{x})] = \mu \\ \dfrac{1}{V}\displaystyle\int_V n(\mathbf{x})\, \mathrm{d}\mathbf{x} = n_0 \end{cases} , \tag{4.4}$$

where with $\frac{\delta f}{\delta -}$ we denote the functional derivative of the free-energy density $f$ with respect to either the field $m$ or $n$. Note that from these equation we can infer the physical meaning of the Lagrange multiplier $\mu$ as the chemical potential.

The first equation of (4.4) gives us the equilibrium values $m^\star$ of the non-conserved field $m$ as a function of the conserved field $n$ and the interaction parameter $\chi$. This equation is easily solved and reads:

$$m^\star[n(\mathbf{x},t);\chi] = \begin{cases} 0 & \text{if } n(\mathbf{x},t) \le \dfrac{a}{\chi} \\ \pm \dfrac{\sqrt{\chi\, n(\mathbf{x},t) - a}}{\sqrt{2b}} & \text{if } n(\mathbf{x},t) > \dfrac{a}{\chi} \end{cases} . \tag{4.5}$$

Thanks to these solutions, we can restrict the problem to consider only the conserved density field. For this purpose, we define the effective free-energy density $f^\star \equiv f^\star[n(\mathbf{x},t)] \equiv f[m^\star(\mathbf{x},t), n(\mathbf{x},t)]$; this new quantity reads:

$$f^\star[n(\mathbf{x});\chi] = \begin{cases} cn^2 + dn^3 & \text{if } n(\mathbf{x}) \le \dfrac{a}{\chi} \\ -\dfrac{a^2}{4b} + \dfrac{a\chi}{2b} n + \left(c - \dfrac{\chi^2}{4b}\right) n^2 + dn^3 & \text{otherwise} \end{cases} . \tag{4.6}$$

This procedure simplifies Eqs. (4.4) which become:

$$\begin{cases} \dfrac{\delta f^\star}{\delta n}[n(\mathbf{x})] = \mu \\ \dfrac{1}{V}\displaystyle\int_V n(\mathbf{x})\, \mathrm{d}\mathbf{x} = n_0 \end{cases} . \tag{4.7}$$

Notice that this equation is always satisfied by the trivial uniform solution $n(\mathbf{x}) \equiv n_0 \ \forall \mathbf{x}$, and chemical potential $\mu = f^\star[n_0; \chi]$.

### 4.2.1 Common-Tangent Construction

To look for non trivial solutions we first notice that the expression for $f^\star$ in (4.6) does not contain derivatives of $n$, hence the first of Eqs. (4.7) consist of a polynomial equation to be solved at fixed $\mathbf{x}$. In other words, for each value of $\mu$, the field function $n(\mathbf{x}, t)$ can assume only a finite number of values $\{n_1, n_2, \ldots, n_k\}$. In turn, these entail the coexistence of different phases (as they are characterised by different values of $n$).

Coexistence is achieved when two phases in the system, $n(\mathbf{x}) = n_-$ and $n(\mathbf{x}) = n_+$ are characterised by the same pressure $P = f^\star - n\frac{\delta f^\star}{\delta n}$ and chemical potentials $\mu$. These conditions can be represented graphically by the common tangent construction [20] (see fig. 4.2, which states that the tangents of the function $f^\star(n)$ at points $(n_-, f(n_-; T))$ and $(n_+, f(n_+; T))$ must have the same slope $\mu$ and the same intercept $P$. The decomposition of the average concentration $n_0$ into a mixture of $n_-$ and $n_+$ with $n_- \leq n_0 \leq n_+$ occurs through the so-called "binodal" decomposition [20]. Namely, for each value of $\chi$ we must solve the system of equations:

$$\begin{cases} \left.\dfrac{\partial f}{\partial n}\right|_{n_-} = \left.\dfrac{\partial f}{\partial n}\right|_{n_+} \\ f(n_-) - n_- \left.\dfrac{\partial f}{\partial n}\right|_{n_-} = f(n_+) - n_+ \left.\dfrac{\partial f}{\partial n}\right|_{n_+} \end{cases} . \tag{4.8}$$

As we noted, in general there could exist three (or more) phases $n_1, n_2, n_3$ that satisfy at the same time the common tangent construction (4.8), i.e. such that $\mu(n_1) = \mu(n_2) = \mu(n_3)$ and $P(n_1) = P(n_2) = P(n_3)$. However, in the specific case of a free-energy in the form (4.6) this is impossible. Indeed, given a generic third order polynomial function $g(x) = a + bx + cx^2 + dx^3$ there are no two points $x_1$ and $x_2$ such that the common tangent construction is satisfied. In (4.6) we are "lucky" because the function $f^\star$ is defined piece-wise, and therefore we have a hope to find two densities $n_- < a/\chi$ and $n_+ > a/\chi$ that respect the construction.

Finally, if the domain of the free energy present boundaries, e.g. is a given interval, one can have some particular situations at its border. In our case, the domain is, in fact, limited by the physical assumption that the density field is always positive $n(\mathbf{x}) \geq 0$. Indeed, the slope of the tangent $\mu \equiv \delta f / \delta m$ is not well defined when the density $n(\mathbf{x}, t) = 0$; in this case we assume that the tangent slope can take any arbitrary, and when we apply the common-tangent construction one must only assume that the tangent from $n_- \equiv 0$ and $n_+$ has only the same intercept (pressure), with no constraint on $\mu$.

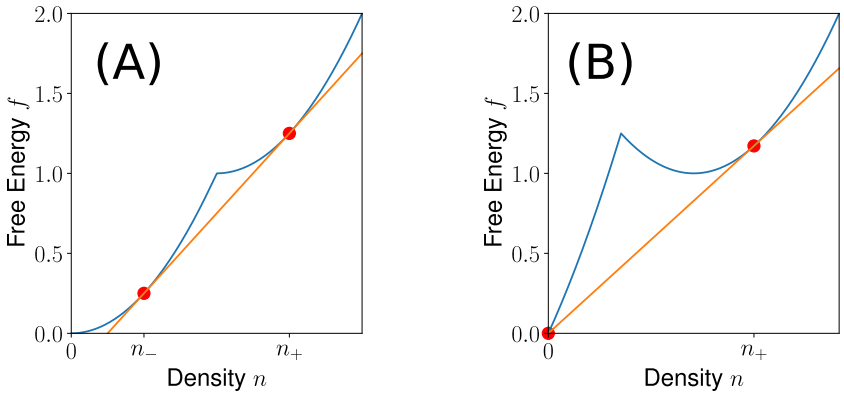



**Figure 4.2:** Pictorial illustration of the common tangent technique. **(A)** Given a free energy $f$ (in blue) we find two points $n_-$, $n_+$ such that the tangent (in orange) to the graph of $f$ is the same. **(B)** If we define the system for $n >= 0$, in the point $n = 0$ the tangent is not well defined; therefore we just need to find another point $n_+$ such that the tangent to $f$ passes from the origin.

### 4.2.2 Phase Diagram Construction

After we developed the method to investigate the behaviour at equilibrium of the system, let us apply it to our system. First, as we already noticed, we know that the homogenous solution $n(\mathbf{x}) \equiv n_0$ is always a solution of equations (4.7). Thus, if for some value of $(\chi, n_0)$ we find a *demixed* solution thanks to the common tangent construction, we need some method to discern which of the two situations is the stable one found at the equilibrium.

The main method we employ is to find the so called *spinodal region* in which the homogeneous solution $n(\mathbf{x}) = n_0$ is unstable; to verify the stability of this solution we just have to check the second derivative of $f$ and impose $\frac{\delta^2 f^\star}{\delta n^2}\Big|_{n(\mathbf{x})=n_0} < 0$. This inequality is quickly solved and its solution reads:

$$\frac{a}{\chi} < n_0 < \frac{\chi^2 - 4bc}{12bd} \quad . \tag{4.9}$$

Inside this region of values the homogeneous solution is linearly unstable and the system spontaneously demixes into low density $(n_-)$ and high density $(n_+)$ phases. Beware, however, that if the homogeneous solution is not unstable, it does not mean it is stable, as it could be metastable. To check for metastability we simply compare the free-energies given by equation (4.3). Namely, if we have a demixed configuration $(n_-, n_+)$ to be compared with the homogeneous solution we must check the inequality:

$$\alpha f\left(m^*(n_+), n_+; \chi\right) + (1-\alpha) f\left(m^*(n_-), n_-; \chi\right) \lesseqgtr f\left(m^*(n_0), n_0; \chi\right) \quad ,$$

where $\alpha$ is the fraction of the system in the $n_+$ phase such that:

$$\alpha n_+ + (1-\alpha) n_- = n_0 \quad , \tag{4.10}$$

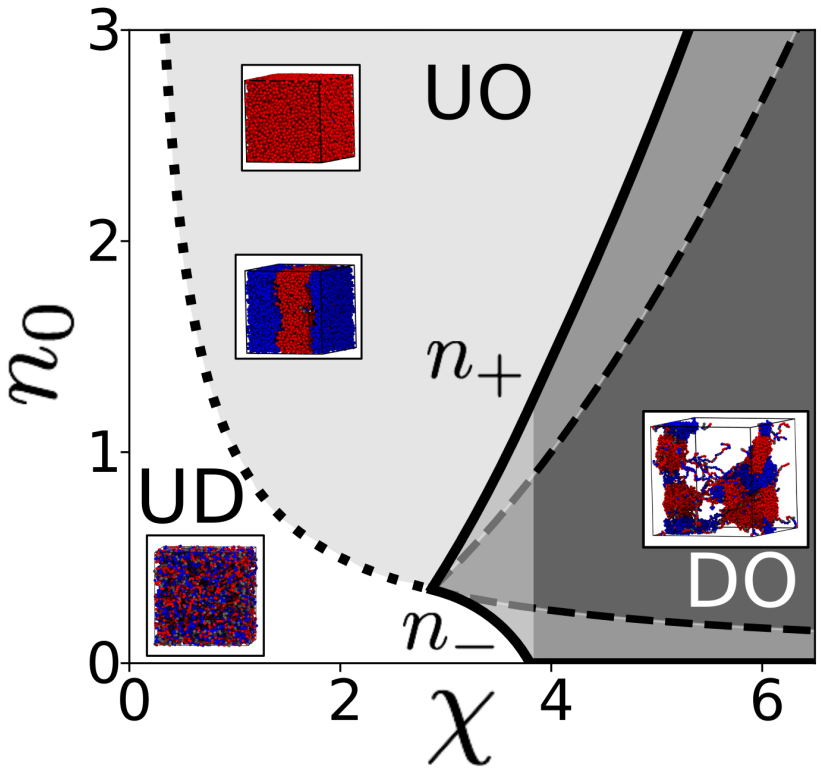



**Figure 4.3:** Equilibrium phase diagram associated to the free energy Eq. (4.1) and obtained using the common tangent construction. The three phases are: (UD) uniform density ($n = n_0$) and epigenetically disordered ($m^2 = 0$); (UO) uniform density ($n = n_0$) and epigenetically ordered ($m^2 > 0$); (DO) demixed and epigenetically ordered ($n = n_+$, $m^2 > 0$ and $n = n_-$,$m^2 = 0$). A fourth partially-demixed ordered (PDO) phase is characterised by weaker variations in density ($n_- > 0$) and denoted by a white shading within the DO phase. The dotted line marks the critical value of the coupling $\chi_c(n_0)$, the solid lines identify the boundaries of the coexistence region (binodals) and the dashed lines mark the spinodal region where the uniform solution is linearly unstable [20, 120]. Insets report representative snapshots from Brownian Dynamics simulations of dense solution of annealed copolymers.

which reads: $\alpha = \frac{n_0 - n_-}{n_+ - n_-}$.

By applying this procedure to our free energy for every value of $(\chi, n_0)$ we can finally obtain an equilibrium phase diagram of our model. Notice that the qualitative features of our phase diagram do not depend on the values of the phenomenological constants $a, b, c, d$ in (4.1); for this reason, from now on we will focus on the case $a = b = c = d = 1$. The phase diagram as a function of the coupling interaction $\chi$ and average density $n_0$ is reported in fig. 4.3; the binodal lines $n_-(\chi)$ and $n_+(\chi)$ are drawn with thick black lines and denote the solutions to the common-tangent construction for a given $\chi$. We find that the equilibrium system is characterised by

the presence of three possible phases:

1. *Uniform Disordered (UD) Phase*: the genome organizes spatially as a uniform ($n(\mathbf{x}) \equiv n_0$) and epigenetically incoherent ($m(\mathbf{x}) = 0$) colouring phase.
2. *Uniform Ordered (UO) Phase*: the genome is spatially uniform ($n(\mathbf{x}) \equiv n_0$), but the epigenetic marks are all coherently coloured ($m(\mathbf{x})^2 \neq 0$).
3. *Demixed Ordered (DO) Phase*: The genome is characterised by local aggregation into high density clusters ($n(\mathbf{x}) > n_0$), each characterised by coherent colouring ($m(\mathbf{x})^2 > 0$). In low density zones ($n(\mathbf{x}) < n_0$), epigenetic marks are incoherent ($m(\mathbf{x})^2 = 0$).

Notice that that the binodal line $n_-(\chi)$ in fig. 4.3 has an angular point at $\chi \approx 3.8$; this suggest the Demixed-Ordered phase is actually divided in two "sub-phases":

3.1 PDO (Partial Demixed Ordered) phase: genome organizes in domains of typical densities $n_+$ and $n_-$, with $n_+ > \frac{a}{\chi} > n_- > 0$; colouring is coherent only in the domains where $n(\mathbf{x}) = n_+$.

3.2 actual DO phase: genome organizes in droplets of typical density $n_+ > 0$, dispersed in a phase with $n_- = 0$; in this case there is colouring order ($m(\mathbf{x})^2 \neq 0$) in the domains where $n(\mathbf{x}) > 0$.

### 4.2.3 Nature of the Phase Transitions

In this section, we will discuss about the existence, and order, of the phase transition between our four phases. For our convenience, we will divide the discussion in three parts: (1) transition between uniform phases (UD $\leftrightarrow$ UO); (2) transitions between uniform and demixed phases (UD $\leftrightarrow$ PDO, UO $\leftrightarrow$ PDO, UO $\leftrightarrow$ DO); (3) transition between demixed phases (PDO $\leftrightarrow$ DO). Namely, we will prove that all of these transitions, with a little doubt on PDO $\leftrightarrow$ DO, are continuous phase transitions (see fig. 4.4), in opposition with the first-order phase transition we found in the dilute regime in chapter 1 or chapter 3.

#### Transition between Uniform Phases

In a equilibrium uniform phase with average genomic density $n_0$, the magnetization will be itself homogeneous and assume the value given by (4.5). Let us study the quench from the Uniform-Disorderd phase to the Uniform-Ordered one. We can do this quenching by changing both the value of the coupling $\chi$ and density $n_0$; in whichever way we choose to do this we will have a transition when $n_0\chi = a$. It is easy to verify from (4.1) that the order parameter $m$ is continuous along the critical line $n_0 = a/\chi$, but its gradient $\left(\frac{\partial m}{\partial n_0}, \frac{\partial m}{\partial \chi}\right)$ is not. This proves that the UD $\leftrightarrow$ UO transition is continuous (see fig. 4.4A).

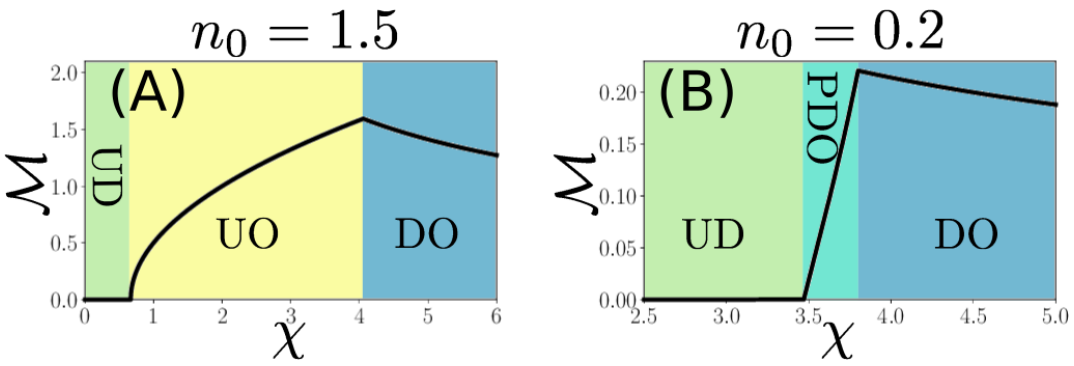



**Figure 4.4:** Order parameter $\mathcal{M}$ (4.11) as a function of the coupling $\chi$ at fixed average density $n_0$. Thanks to the phase diagram fig. 4.3 we highlight regions of the plot with different colors depending on the phase the system is in. Plots indicate that, when passing from a phase to another, the average magnetization is continuous, but, at the critical point, its derivative is not. This suggest a second-order phase transition between the various phases.

### Transition between Demixed and Uniform Phases

In order to investigate the possible phase transition between the other phase, we introduce a new order parameter $\mathcal{M}$ defined as the average magnetization of the system. Namely, we define:

$$\mathcal{M}(\chi, n_0) = \frac{1}{V} \int_M |m[\mathbf{x}]| \, \mathrm{d}\mathbf{x} \ . \tag{4.11}$$

Notice that the parameter $\mathcal{M}$ takes simple form in both the uniform, and demixed phases:

$$\mathcal{M}(\chi, n_0) = \begin{cases} m^\star[\chi, n_0] & \text{in the uniform phases} \\ \alpha \, |m^\star[\chi, n_+]| & \text{in the demixed phases} \end{cases} , \tag{4.12}$$

where $m^\star$ is the equilibrium magnetization (4.5), and $\alpha$ is the ratio of the system with active chromatin (4.10). Note that when we pass from the Uniform-Disordered phase to a Demixed one, we cross the binodal (see fig. 4.3) in the $n_-(\chi)$ branch, when this happens the system undergoes a phase transition at the critical point $P_c$. In $P_c$ we have that $n_- \to n_0$, and thus $\alpha \to 0$. This implies that the average magnetization $\mathcal{M} \to 0$ as we approach the critical point; on the other hand in the UD phase we always have $\mathcal{M} = 0$, and therefore the order parameter $\mathcal{M}$ is continuous during the transition. The other possible case is when we pass from the Uniform-Ordered phase to a demixed one. In this case, we have that nearby the critical point $n_+ \to n_0$ and therefore $\alpha \to 1$; by using a similar procedure as before, it is easy to prove that even in this case $\mathcal{M}$ is continuous during the transition.

Let us now show that the first derivative of the average magnetization $\mathcal{M}$ is not continuous at the critical points, and, therefore, that there is a continuous phase

transition between the uniform phases and the demixed ones. Let us derive (4.12) over $\chi$:

$$\frac{\partial \mathcal{M}}{\partial \chi} = \begin{cases} \left.\dfrac{\partial m^\star}{\partial \chi}\right|_{n_0} & \text{in the uniform phases} \\ \dfrac{\partial \alpha}{\partial \chi} m^\star[\chi, n_+] + \alpha \left.\dfrac{\partial m^\star}{\partial \chi}\right|_{n_+} & \text{in the demixed phases} \end{cases} , \tag{4.13}$$

where for convenience we have assumed $m^\star > 0$. Notice that the "uniform" term and demixed term, in the usual $\alpha \to 1$ and $n_+ \to n_0$, or $\alpha \to 0$ and $m^\star \to 0$ limits, are the same if and only if $\frac{\partial \alpha}{\partial \chi} m^\star[\chi, n_+] = 0$. However $|m^\star[\chi, n_+]| > 0$ always as $n_+ > \frac{a}{\chi}$ (see section 4.2.1); on the other hand if we take the derivative over $\chi$ of equation (4.10), we obtain:

$$\frac{\partial \alpha}{\partial \chi} = \frac{1}{n_+ - n_-} \left[ \alpha \frac{\partial n_+}{\partial \chi} + (1-\alpha) \frac{\partial n_-}{\partial \chi} \right] , \tag{4.14}$$

and the term in the parenthesis is never zero as $\frac{\partial n_+}{\partial \chi} > 0$ always as seen in fig. 4.3. This shows that the derivatives of $\mathcal{M}$ are never the same when passing from whatever uniform phase to whatever demixed one, and proves the existence of a second-order phase transition (see fig. 4.4).

### Transition between Demixed phases

Finally, we examine the transition between the Partial-Demixed-Ordered phase (PDO) and the actual Demixed-Ordered phase (DO). As in the previous case, we employ the order parameter $\mathcal{M}$: from (4.13) we have that both in PDO and DO, the derivative of $\mathcal{M}$ is

$$\frac{\partial \mathcal{M}}{\partial \chi} = \frac{\partial \alpha}{\partial \chi} m^\star[\chi, n_+] + \alpha \left.\frac{\partial m^\star}{\partial \chi}\right|_{n_+} .$$

Notice that nearby the critical line, the second term of the derivative is the same in the two phases, and therefore the eventual difference must be in the first one. From (4.14) we see that the derivative of alpha depend itself on the term $\frac{\partial n_-}{\partial \chi}$. From fig. 4.3 and from extensive studies employing the `Mathematica` software, it seems that the derivative $\frac{\partial n_-}{\partial \chi}$ is discontinuous at $\chi \approx 3.8$. This implies, again, that even $\frac{\partial \mathcal{M}}{\partial \chi}$ is discontinuous and that the phase transition is of the second order (see fig. 4.4B).

### Comparison with the Single-Chain Model

In this small section, we will discuss how our findings in the coarse-grained model are compatible with the phenomelogy of the single-chain model discussed in section 1.4 and section 3.5.

First, from the phase diagram shown in fig. 4.3, we see that if we keep $n_0 = 0$ while varying the value of $\chi$, the system will pass directly from a uniform-disordered phase to a demixed-order one, without passing from the intermediate partially demixed

one. Notice that this is exactly what is observed in the single-chain model where the system passes from a swollen (here uniform) disordered phase to a compact (here demixed) ordered one.

The only apparent discrepance from the two models is regarding the order of the transition. In the study of the single-chain model, we have shown the existence of a first-order phase transition between the swollen-disordered and the compact-ordered phases at equilibrium. In the coarse grained model, instead, we have proven that the uniform to demixed transition is continuous when $n_0 > 0$. Let us now show here that in the limit of $n_0 \to 0$ the transition ceases to be continuous, and become of the first order. The average magnetisation $\mathcal{M}$ (4.11) naturally tends to 0 as $n_0 \to 0$, and it is therefore not a good order parameter in this regime. To solve this problem, we notice that we can estimate the total magnetisation $\mathbf{M}$ of the system by averaging $\mathbf{M} = \mathcal{M}/n_0$. If we do so, employing equation (4.12), we have that:

$$\mathbf{M}(\chi) = \begin{cases} 0 & \text{in the uniform phases} \\ \dfrac{\alpha}{n_0} \left| m^\star[\chi, n_+] \right| & \text{in the demixed phases} \end{cases} , \tag{4.15}$$

where we have used that $m^\star[\chi, n_0] = 0$ if $n_0 \ll 1$. Now, since we are in the limit of $n_0 \ll 1$, we skip completely the PDO phase, and always have that in the demixed phase $n_- \equiv 0$, and therefore $\alpha = \frac{n_0}{n_+(\chi)}$. We conclude that in the limit of $n_0 \to 0$ we can write the total magnetisation as:

$$\mathbf{M}(\chi) = \begin{cases} 0 & \text{in the uniform phases} \\ \dfrac{1}{n_+(\chi)} \sqrt{\dfrac{\chi n_+(\chi) - a}{2b}} > 0 & \text{in the demixed phases} \end{cases} . \tag{4.16}$$

This shows that $\mathbf{M}$ display a jump discontinuity when we are in a strongly dilute system ($n_0 \to 0$), compatibly with what observed in the single-chain model.

## 4.3 "Model C" and Dynamical Scaling

In recent years techniques employing chromosome contrast captures techniques, like Hi-c, permits us to study the details of chromatin (and epigenetic landscape) dynamics in the mammalian nuclei. For this reason, here we try to study the main properties of our model, when we incorporate the dynamics. Namely, the dynamics will be described by a set of "model C" equations [20], which will be integrated and studied their time dependence after quenching from Uniform-Disordered Phase to all the other possible phases.

### Model C Equations

In order to develop a dynamical, equilibrium, theory, we proceed in analogy with Hamiltonian mechanics. Namely, we define a "finite-temperature" Hamiltonian $\mathcal{H}[m, n]$

[20] which essentially is a free energy (4.1) to which we add two diffusive terms:

$$\mathcal{H}[m,n] = \int_V \left[ f(m,n) + \kappa_m \|\nabla m\|^2 + \kappa_n \|\nabla n\|^2 \right] \quad , \tag{4.17}$$

where the coefficients $\kappa_m$, $\kappa_n$ regulates the diffusion strength. The Hamilton equations relative to an Hamiltonian $\mathcal{H}$ with a non-conserved field (e.g. magnetization $m$) and a conserved-field (e.g. density $n$) read:

$$\begin{cases} \dfrac{1}{\Gamma_m}\dfrac{\partial m}{\partial t} = -\dfrac{\delta \mathcal{H}}{\delta m} \\ \dfrac{1}{\Gamma_n}\dfrac{\partial n}{\partial t} = \nabla^2 \dfrac{\delta \mathcal{H}}{\delta n} \end{cases} \quad , \tag{4.18}$$

where $\Gamma_m$, and $\Gamma_n$ are two variables that regulate the speed of the dynamics and have a role akin to the one of the mass in the classic Hamiltonian theory. For an Hamiltonian (4.17), these equations will read:

$$\begin{cases} \dfrac{\partial m}{\partial t} = -\Gamma_m \dfrac{\delta f}{\delta m} + D_m \nabla^2 m \\ \dfrac{\partial n}{\partial t} = \Gamma_n \nabla^2 \dfrac{\delta f}{\delta n} - D_n \nabla^4 n \end{cases} \quad , \tag{4.19}$$

where for our convenience we have defined $D_m \equiv \Gamma_m \kappa_m$ and $D_n \equiv \Gamma_n \kappa_n$, and have employed the identity $\frac{\delta}{\delta m}\|\nabla m\|^2 = -\nabla^2 m$. These are the so-called "Model C"-equations which deal with the dynamics of two coupled fields when one is conserved (density $n$) and the other is not (magnetization $m$). Model C equations are to be compared with the so-called "Model A" equation which deals only with a non-conserved field (first equation of (4.19)), and with "Model B" equation which deals only with a conserved field (second equation of (4.19)). By using the explicit form of free energy density in (4.1), equations (4.19) become:

$$\begin{cases} \dfrac{\partial m}{\partial t} = \Gamma_m \left( 2\chi m n - 2am - 4bm^3 \right) + D_m \nabla^2 m \\ \dfrac{\partial n}{\partial t} = \Gamma_n \nabla^2 \left( 2c\rho + 3d\rho^2 - \chi m^2 \right) - D_n \nabla^4 n \end{cases} \quad . \tag{4.20}$$

As one can imagine, an explicit solution for these equation cannot be found analytically. For this reason, we integrated numerically these equations, using a `C++` software, by discretizing the problem on a 2D $M \times M$ grid, with periodic boundary conditions at borders, and employing a mid-point method for the time evolution. The initial conditions of the fields $m$ and $n$ are generally compatible with the Uniform-Disordered phase. Notice, however, that a perfectly uniform solution do not permits diffusion in equations (4.20) and we would not observe the dynamics; therefore, at the initial conditions, we add a small Gaussian noise with variance $10^{-2} n_0$.

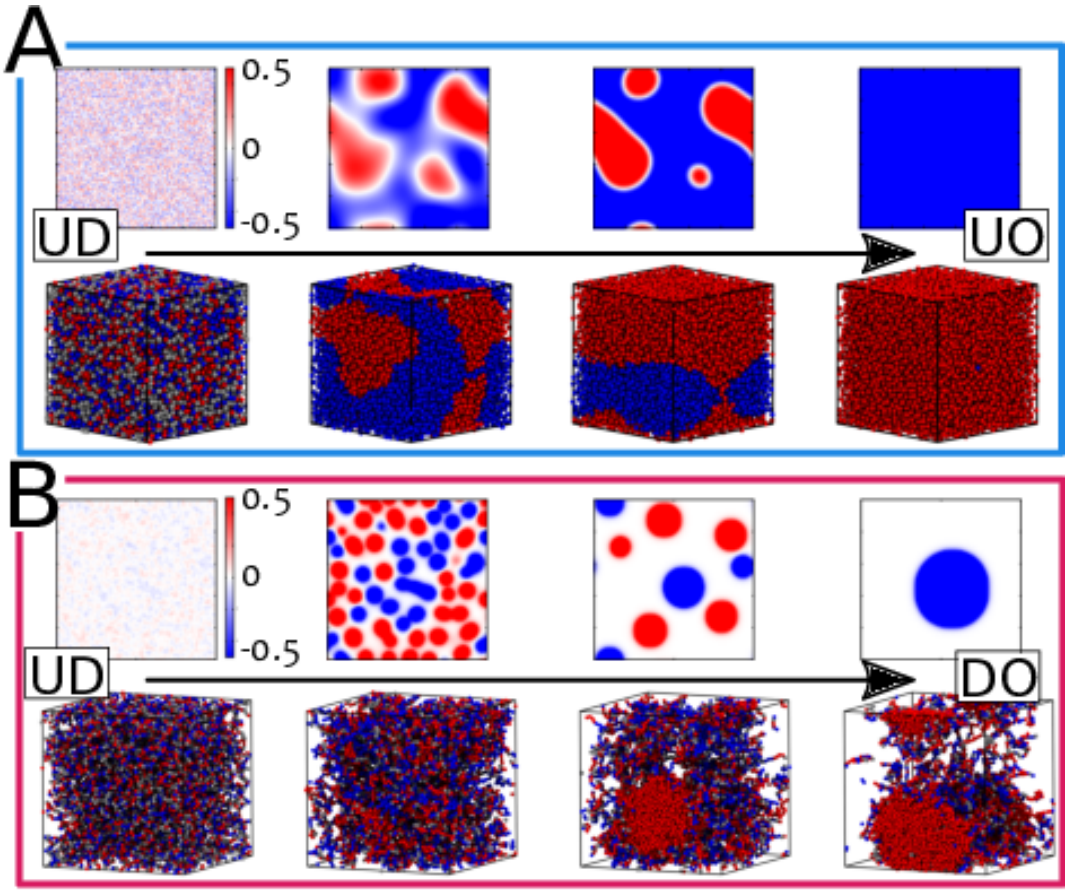



**Figure 4.5:** **A** Snapshots following the quench $Q_1$ from Uniform Disordered (UD) to Uniform Ordered (UO) ($\chi = 0 \to 1$; $n_0 = 2$). **B** Snapshots following the quench $Q_4$ from Uniform Disordered (UD) to Demixed Ordered (DO) ($\chi = 0 \to 6$) at $n_0 = 0.2$). Top rows show the evolution through numerical integration of equations (4.20) on a $512 \times 512$ grid. Bottom rows show the evolution through Brownian-Dynamics simulations (UD$\to$ UO, $\rho = 0.8\sigma^{-3}$, $\alpha_L = \alpha_R = 1$; UD$\to$DO, $\rho = 0.1\sigma^{-3}$, $\alpha_L = \alpha_R = 0.9$)

The numerical solutions of equations (4.20) helps us monitoring the time evolution of the density and magnetization fields during several possible quenching trajectories in the phase space. We quench from a typical configuration observed in the UD phase at $\chi = 0$ (homogeneous density $n(\mathbf{x}) = n_0$ and null magnetization $m(\mathbf{x}) = 0$) and go to one of the ordered phase. Namely, we focused our study on four representative quenches:

1. $Q_1$: UD $\to$ UO ($\chi = 0 \to 1$; $n_0 = 2$);
2. $Q_2$: UD $\to$PDO ($\chi = 0 \to 3.5$; $n_0 = 0.5$);
3. $Q_3$: UD $\to$ DO with large average density ($\chi = 0 \to 6$; $n_0 = 2$);
4. $Q_4$: UD $\to$ DO with small average density ($\chi = 0 \to 6$; $n_0 = 0.2$).

Qualitative study of the solutions of (4.20) in these conditions show (see fig. 4.5) that the system reorganize quickly both in the fields $n$ and $m$. Namely, in the quench $Q_1$ where we remain in a uniform phase, we observe that $n(\mathbf{x})$ remains uniform for the whole dynamics, while $m(\mathbf{x})$ coalesces in several clusters of different (opposite) magnetization. Eventually one of the two epigenetic marks will win and the system will present as a single colored domain, whose color is chosen by a $\mathbb{Z}_2$ symmetry breaking. In the other cases, we quench into a demixed phase. Here, the system undergoes a phase separation between the two densities $(n_-, n_+)$; namely, we observe the formation of high density clusters (density $n_+$) forming in a system with "background" density $n_-$. Each high-density cluster presents coherent magnetization, but different clusters may have different magnetization sign. However, the dynamics show that, eventually, all the clusters will coalesce into a single big circular high-density domain characterized by a single magnetization sign chosen, again, via $\mathbb{Z}_2$ symmetry breaking.

### Typical Domain Length Scale

A standard way to investigate the coarsening observed in the quench dynamics, is to study the time dependence of typical size $L$ of cluster-like structure. The typical domain length $L$, relative to a field (e.g. $m$) has been computed by introducing the so-called spherically-averaged structure factor $S$ [121, 122]. Let us go step by step. The procedure is identical for the two fields, but we will focus on the magnetization field $m$ as it shows coarsening both in the uniform and in the demixed phases. First, let us define $\widetilde{m}(\mathbf{k}, t)$ as the Fourier transform of the magnetization field $m$; we define the spherically averaged structure factor as:

$$\begin{aligned} S(k,t) &= \langle \widetilde{m}(\mathbf{k},t)\, \widetilde{m}(-\mathbf{k},t) \rangle_k \\ &= \frac{1}{2\pi} \int_0^{2\pi} \widetilde{m}(k,\phi,t)\, \widetilde{m}(k,\pi+\phi,t)\, \mathrm{d}\phi \end{aligned} \quad , \tag{4.21}$$

where $k$, $\phi$ are the polar coordinates of the vector $\mathbf{k}$, and with $\langle - \rangle_k$ denoting an average over a shell in $\mathbf{k}$ space at fixed norm $k$. Once computed the structure factor $S$, the average domain length of the domain is defined as the inverse of the first moment of the probability distribution proportional to $S$. Namely, the typical size $L$ reads:

$$L(t) = 2\pi \frac{\int_0^{+\infty} S(k,t)\, \mathrm{d}k}{\int_0^{+\infty} k S(k,t)\, \mathrm{d}k} \quad . \tag{4.22}$$

The length-scale $L(t)$ has been computed for each time for each quench; the Fourier transforms $\widetilde{m}$ and $\widetilde{n}$ have been computed using an FFT (Fast Fourier Transform) algorithm.

We are interested in how the typical domain size $L$ evolves during time in our quenches (see fig. 4.6). It is known that the dynamics of $L$ tends asymptotically to have the form $L(t) \sim t^\gamma$ [20]; we try here to estimate the value of $\gamma$ for each quench.

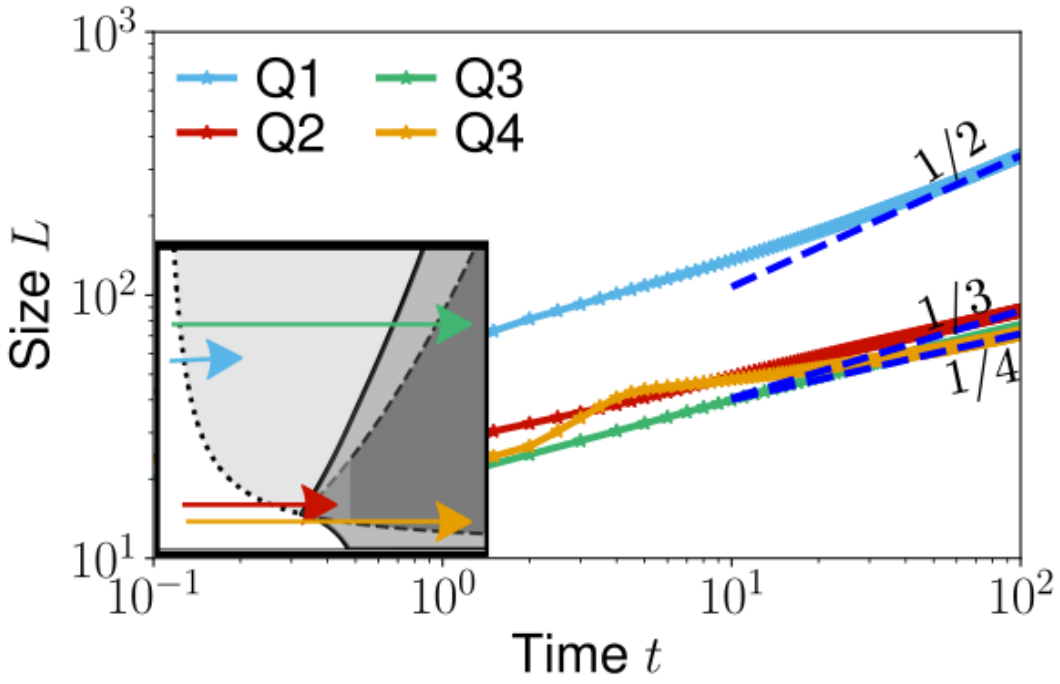



**Figure 4.6:** Evolution of typical epigenetic domain size $L$ following a quench from the uniform disordered phase. The growth displays a power law that is compatible with Model A dynamics ($\alpha = 1/2$) when the density field remains uniform (quench $Q_1$), but it is significantly slowed down when both fields are re-organized ($\alpha = 1/4$). In the inset we schematically show the paths of the quenches in the phase diagram. Here we evolved Eqs. (4.20) with fixed $a = b = c = d = 1$ and $\Gamma_m = \Gamma_n = \kappa_m = \kappa_n = 1$ in a grid of size $512 \times 512$. The trajectories $L(t)$ displayed in the plot are obtained by averaging the trajectories of 16 independent runs for each quench.

The quench into the uniform-ordered state $Q_1$ follows a different scaling with respect to all the others. In $Q_1$ the density field $n$ remains uniform, while the magnetization forms clusters with typical domain size $L$ scaling as $L \sim t^\gamma$ with $\gamma \approx 0.46$, compatible with the coarsening speed showed by model A dynamics $\gamma_A = \frac{1}{2}$. In all the other quenches the reorganization of both epigenetic and density fields leads to a slower coarsing rates that yields a smaller growth exponent $\gamma \approx 0.23 - 0.3$. Our results show a slower growth than the one expected from a "model B" dynamics ($\gamma_B = \frac{1}{3}$). However the slowing down of the dynamics in Model C is a well known phenomenon [123], and it is probably due to the interplay between the two fields.

### Biological Relevance of the Equilibrium Model

Let us now look at the possible implications of our findings in biological contexts. Our theory suggests that certain perturbations can trigger nuclear, macroscopic phase

separation driven by the epigenetics landscape. This is achieved either by moving the system into the spinodal region, or by overcoming the metastability of the UO phase outside this region. While compactification of large genomic regions is indeed observed for example in the inactivation of the X chromosome in female mammalian [124], the model (as seen in the quenches $Q_2$, $Q_3$ or $Q_4$) predicts an uncontrolled spreading of a single epigenetic mark to the whole nucleus. This situation, is actually never observed in real biological situations, meaning that the model is not complete, and perhaps, that the reader-writer machinerie cannot be described at the equilibrium, and that it has inherently out-of-equilibrium features.

## 4.4 Epigenetic Switching

As noticed in the previous section, an equilibrium dynamics lead to the indefinite coarsening of coherent domains both in magnetization and in density. This is not what is observed in realistic biological situations where coarsening is observed, but is not indefinite and chromatin agglomerates are characterized by a certain typical size: this class of phenomenon where coarsening is at a certain point blocked is called "arrested phase-separation". Given these reasons, we need to modify the model in order to take into account this observation. A typical strategy in this kind of situation is to stray away from equilibrium dynamics, and introduce out-of-equilibrium mechanisms. Example of these kind of mechanisms have been widely studied in order to examine the organization of chromatin [125], centrosomes [126] and chromosomes in general [127]. Some of these model, indeed, predict the presence of arrested phase separation, for example by affecting the local chromatin state and its accessibility to proteins [128], by preventing the deposition of histones with a given epigenetic mark [129].

Inspired by these works, we introduced a similar mechanism in our model that we will call "epigenetic switching". Namely, we propose that chromatin can dynamically convert itself from an "active state" which actively participate to the epigenomic spreading, and an "inactive state" which is magnetically neutral and do not participate to the recoloring dynamics. On the biological point of view, we may see the epigenetic switching as a way to simplify the role of ATP in re-modeling the chromatin; in this optic the inactive chromatin represent that parts of the genome that can not be accessed by proteins, or are poor in nucleosomes. Regarding equation (4.20), we are assuming that the local chromatin density field $n(\mathbf{x})$ is actually divided in two components $n_a(\mathbf{x})$, that is the local density of the active component, and $n_i(\mathbf{x})$, the density of the inactive one. We model the "switching" between an active and an inactive state via a simple first-order chemical reaction, which has been known to induce arrested phase separation (see for example [130]). Due to these considerations

equations (4.20) are mapped into their out-of-equilibrium counterpart:

$$\begin{cases} \dfrac{\partial m}{\partial t} = \Gamma_m \left(2\chi m n_a - 2am - 4bm^3\right) + \kappa_m \nabla^2 m \\ \dfrac{\partial n_a}{\partial t} = \Gamma_n \nabla^2 \left(2cn_a + 3dn_a^2 - \chi m^2\right) - \kappa_n \nabla^4 n_a + \sigma_a n_i - \sigma_i n_a \\ \dfrac{\partial n_i}{\partial t} = \Gamma_n \nabla^2 \left(2cn_i + 3dn_i^2\right) - \kappa_n \nabla^4 n_i - \sigma_a n_i + \sigma_i n_a \end{cases} , \tag{4.23}$$

where the amplitudes of $\sigma_{a/i}$ describe the rates at which the density fields $(n_a, n_i)$ are activated/inactivated, i.e. the rates at which chromatin re-modeling factors act on the genome. Notice, that, apart from the chemical reaction term, the equation for the active field $n_a$ are similar in form to the one for $n$ in (4.20); on the other hand the dynamic of the inactive density $n_i$ is merely diffusive to reflect its passive and inert behavior.

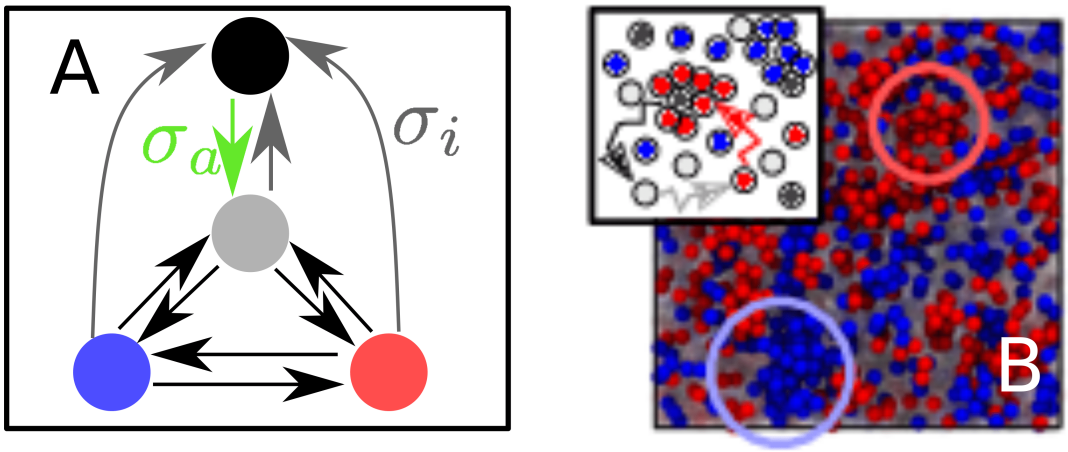



**Figure 4.7:** **(A)** Pictorial illustration of the switching scheme. Any part of the genome can become at any time inactive (in black), but when reactivating it is forced to be in the unmarked (gray). state. **(B)** Snapshot from a Brownian-Dynamics simulation performed using the methods described in section 4.1, but inserting also the possibility for a bead to become inactive (represented in black) and do not participate to the recoloring. Beads can become inactive or become active with the same switching rate $\kappa_S$ $[\sigma_a = \sigma_i]$. We set monomer density $\rho = 0.8\sigma^{-3}$, $\alpha = 0.9$, and $\kappa_S = 10^{-4}\tau_B^{-1}$. The arrows denote a possible loop arising in steady state of the simulations.

Notice also that, given the presence of a first-order reaction between the two, the density fields $n_a$ and $n_i$ are not singularly conserved. However, in order to be coherent with the conditions of the equilibrium model, a total conservation law is needed, and in this case it takes the form

$$\frac{1}{V}\int_V \left[n_i(\mathbf{x};t) + n_a(\mathbf{x};t)\right] \mathrm{d}\mathbf{x} = n_0 \equiv \text{constant} \qquad \forall t \ . \tag{4.24}$$

Indeed, these constraints are surely respected in equations (4.23), as we find that:

$$\frac{\partial}{\partial t}(n_a + n_i) = \nabla \cdot \mathbf{K}\,, \tag{4.25}$$

where:

$$\mathbf{K} = \Gamma_n \nabla \left[2c(n_a + n_i) + 3d(n_a^2 + n_i^2) - \chi m^2\right] - \kappa_n \nabla \left[\nabla^2 (n_a + n_i)\right]\,. \tag{4.26}$$

As long as we assume "normal" boundary conditions at the border of the volume $V$ (e.g. reflecting or periodic), we will always have that $\int_V \nabla \cdot \mathbf{K} = 0$; thus from equation (4.25) we obtain that $\frac{\partial}{\partial t}(n_a + n_i) = 0$ and that the total density is conserved.

We want to point out that, given the scheme described by equations (4.23), we have that the system is inherently out of equilibrium when $\sigma_i \neq 0,\ \sigma_a \neq 0$. There are several ways to infer this. First, note that when an active bead is inactivated it loses its magnetization. On the other hand, when a bead is activated, it will always have $m = 0$ because its dynamics was decoupled from $m$ (see fig. 4.7A). This breaks the time-reversal symmetry and, therefore, the detailed balance is not respected, putting the system out-of-equilibrium. Another way to see this is to notice that the reaction amplitudes $\sigma_a$, $\sigma_i$ are constants and do not depend on any of the density fields.

### Arrested Phase Separation and Typical Cluster Size

We are finally ready to study, and integrate, equations (4.23). Results show that, as imagined, the addition of a chemical reaction between the active and inactive state indeed induces an arrested phase separation (see fig. 4.8). We note there is comparatively much less density variations with respect to equilibrium, as the inactive component $n_i$ tend to occupy uniformly all the space surrounding the magnetically active clusters. Furthermore, For large $\sigma_a/\sigma_i$, chromatin droplets interactions start being relevant: we observe that clusters with the same magnetization form "super-beads" which come together forming branched aggregates (see fig. 4.8C), as this configuration minimizes interfaces in the fields $m$ and $n_i$.

We also examine the dependence of the average cluster size $L$ as a function of the reaction rates. To simplify the analysis, we study the case where $\sigma \equiv \sigma_i = \sigma_a$, and study how $L$ varies as a function of $\sigma$. By studying the numerical solutions of equations (4.23), we infer that $L \sim \sigma^{-1/4}$, in agreement with the observations of [131].

### Out-of-Equilibrium Phase Diagram

Here, we study the steady states of the dynamics of equations (4.23) by fixing the the values of $n_0$ and $\chi$, but varying the values of the reaction rates $\sigma_{a/i}$. The equations are studied via numeric integration on a $100 \times 100$ grid, and employing the parameters $\Gamma_{m/n} = \kappa_{m/n} = 1$.

Results are presented in fig. 4.9 and show that varying the value of the reaction rates can push a system which at equilibrium ($\sigma_a = \sigma_i = 0$) would show a certain

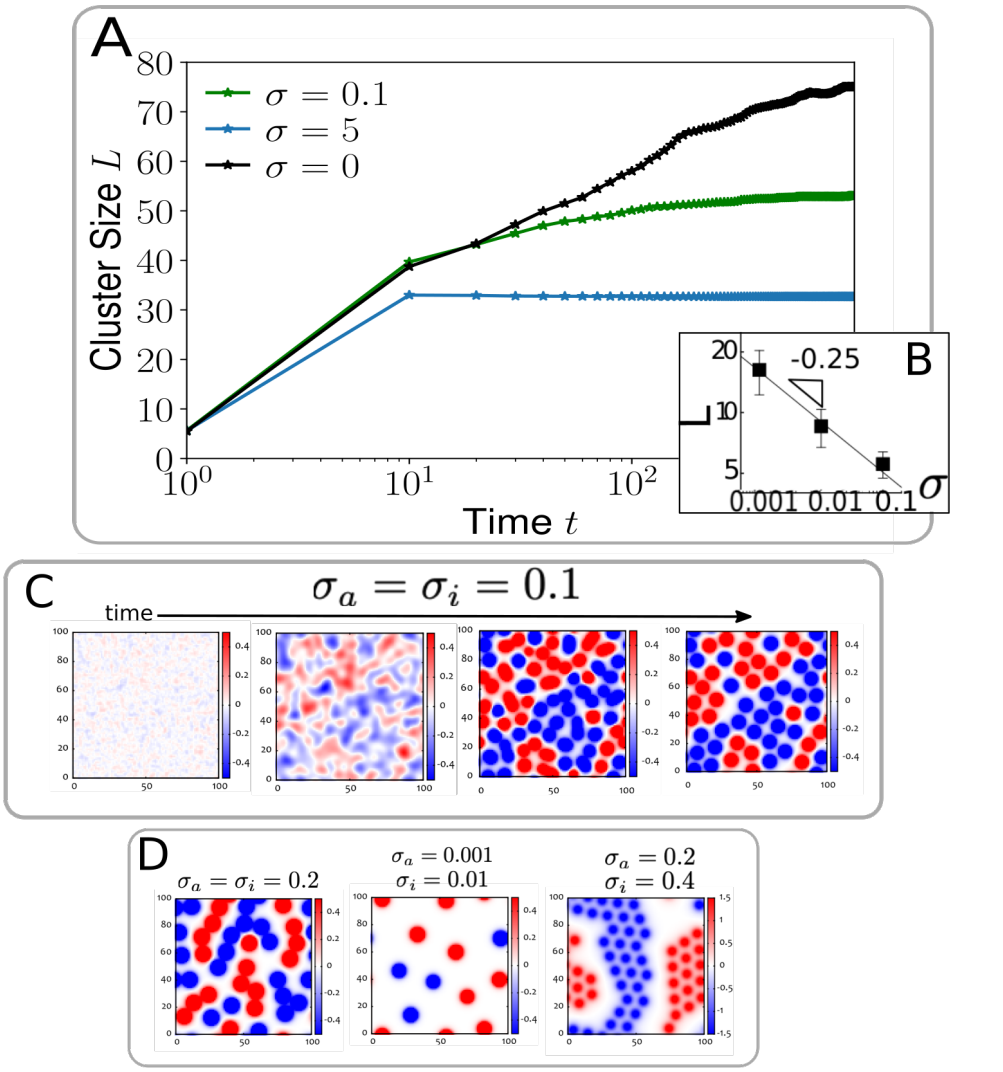



**Figure 4.8:** **(A)** Size of a typical cluster $L$ as a function of time for various values of $\sigma \equiv \sigma_i = \sigma_a$; notice that if $\sigma = 0$ we are at the equilibrium regime and coarsening continues indefinitely, however if $\sigma > 0$ the size $L$ reaches a stable, finite, value which depend on the $\sigma$ value. $L$ is obtained by integrating equations (4.23) on a $256 \times 256$ grid and by applying the method described by equations (4.22), and averaging over 16 independent runs. We set $a = b = c = d = 1$, and $\chi = 6$, $n_0 = 0.5$. **(B)** Average cluster size $L$ at the steady state as a function of $\sigma$. The plot (in log-log) shows that $L$ decreases as $\sigma$ increases following the power-law $L \sim \sigma^{-1/4}$. Results have been found by averaging over 16 independent runs and with the values varying $\sigma$ in the range $\sigma = 0.1 - 15$. **(C)** Time-dependent snapshots obtained evolving (4.23) at $\sigma_a = \sigma_i = 0.1$ with $\chi = 6$ and $n_0 = 1$. **(D)** Steady state configurations obtained with same parameters except for the switching rates, which are given in the figure.

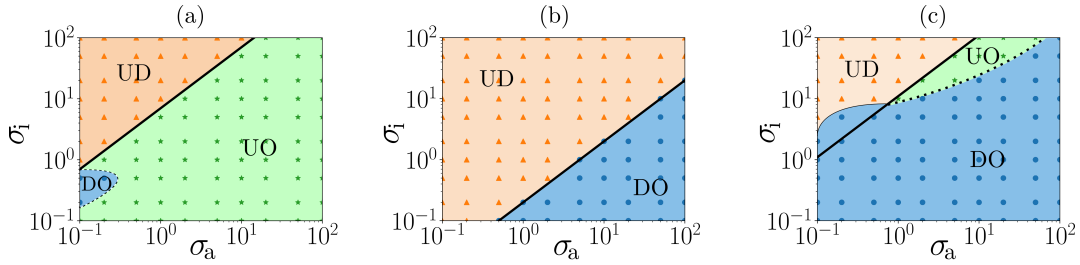



**Figure 4.9:** Non-equilibrium phase diagrams obtained by integrating numerically equations (4.23) on a 2-dimensional grid of side $L = 100 \times 100$, with $a = b = c = d = 1$, $\Gamma_m = \Gamma_n = 1$, $\kappa_m = \kappa_n = 1$. All the figures are obtained at fixed values of $\chi$ and $n_0$ .The thick solid line is found via equation (4.28) as an heuristic way to separate the ordered phases (UO, DO) from the disordered one (UD). Notably, this approximation works completely fine in the first two plots, but files in the last one. The dashed line, and the thin solid line, are instead qualitative lines drawn on behalf of the numerical integrations which are meant to separate one phase from the other. In the figures, different colors highlight different phases (orange for UD, blue for DO, green for UO).

phase, to other regions of the phase diagram. Notably, we note that states that systems an ordered phase at equilibrium, can be driven to the uniform disordered phases, or another ordered one, simply by varying the value of $\sigma_{a/i}$. An heuristic reason for this can be found by looking at the steady state of equations (4.23). In this limit, one can found an easy expression for the average active density:

$$\langle n_a \rangle = \frac{1}{V} \int_V n_a \mathrm{d}\mathbf{x} \approx \frac{\sigma_a n_0}{\sigma_i + \sigma_a} \,, \tag{4.27}$$

where we have used that $\int_V \nabla^2 n_a = 0$, and that $\int_V (n_a + n_i) = n_0 V$. In section 4.2, we had shown that at the equilibrium an ordered state (both demixed and uniform) can arise only if we have an active local density $n_a > a/\chi$. Using this last relation, and equation (4.27), one can conclude that the ordered phases (uniform or demixed) are strongly favored if

$$\sigma_i < \left(1 - \frac{n_0 \chi}{a}\right) \sigma_a \,, \tag{4.28}$$

in very good quantitative agreement with the observations from the numerical evolution of the system (see fig. 4.9).

## 4.5 Discussion

In this chapter, we have proposed a Landau-Ginzburg mean-field theory to describe the coupling between spatial dynamics and epigenetic landscape within the cell nucleus. Namely, we discuss, in light of phenomenological observations, free-energy density in the conserved density field $\rho$, and non-conserved field magnetization.

We discuss the equilibrium dynamics of the model by using the "common-tangent construction", and find its equilibrium phases as a function of the average density $n_0$ and the temperature-dependent coupling parameter $\chi$. We find that the system can either organizes in three different phases: (i) Uniform-Disordered (UD) phase characterized by uniform density and no epigenetic order; (ii) Uniform-Ordered (UO) phase characterized by uniform density an local epigenetic order; (iii) Demixed-Ordered (DO) phase where the system show phase separation and organizes in epigenetically ordered, high-density clusters similar to chromatin agglomerates observed in experiments.

We study the kinetics of our equilibrium model, by using a set of "Model C" equations which describe the approach to equilibrium of the density and magnetization fields. We find that quenching from the UD phase to the DO one, the high-density clusters formed by the system grow indefinitely, until a single epigenetic mark spreads and overtakes the others. The coarsening dynamics of these domains is in good agreement with classic results of Model C kinetics. Unfortunately, these results are not realistic as in cell nuclei several epigenetic domains, of different marks, can coexist at the same time.

We extend the model by considering a first-order reaction which converts a "magnetizable" chromatin state that can be epigenetically marked into an "inert" state that cannot be modified, and vice versa. The resulting kinetic equations describe a non-equilibrium system which shows arrested phase separation, where the unlimited spreading of epigenetic marks is blocked, and stable epigenetic domains coexist. We study the typical cluster size as a function of the reaction rates, and find that that our results are in good agreement with what observed in similar models.

It would be interesting to refine the model by specializing the non-equilibrium terms introduced, or implementing new higher order reactions, to describe certain biological process like RNA production, phosphorylation, or ATP-controlled switching. Furthermore, the predictions of a more refined model could be tested by *in vivo* experiments by controlling the rates of post-translational modification of certain protein complexes that are known to remodel chromatin.

# Conclusions

*"Begin at the beginning," the King said, very gravely,*
*"and go on till you come to the end: then stop."*

---

Lewis Carrol, `Alice in Wonderland`

The arguments presented in this thesis represents a great part of my work during my Ph.D. studies and concerns the use of statistical mechanics techniques for the and modeling of interacting heteropolymers and magnetic polymers. We briefly present here the main results of this thesis.

In chapter 1 we have first reviewed the basic biology principles behind epigenetics. Then we have discussed basic physical models which deals either with the 3D organization of chromatin given a fixed epigenetic landscape, or eith the spreading of epigenetic marks. Finally, we have presented the starting minimal mode which couples the 3D chromatin dynamics with epigenomic spreading. Here, chromatin is modeled as a semi-flexible polymer whose beads bear a "color" representing a specific epigenetic mark and which interact among each other via a Potts-like interaction. At equilibrium we have shown the existence of a first-order phase transition between a swollen-disordered phase and a compact-ordered one, compatibly with the robustness and memory showed by epigenetic domains. When we drive the system out-of-equilibrium two novel phases characterised by swollen-ordered and compact-ordered configurations arise; all the possible phase transitions of the non-equilibrium phase diagram drawn are discussed.

In chapter 2 we have extended the model of the previous chapter to include "genomic bookmarking". We have shown that, depending on the bookmarkers configurations, several coherent, robust, epigenetic domains of different marks can form and coexist at the same time. We have also inserted in the model a novel random process that mimics the mitosis process and verify that the epigenetic domains are robust against large perturbations, and even resist to the excission of bookmarkers for several cell cycles. Finally, by using some experimental data, we have challenged the problem of simulating a chromosome of a *Drosophila* stem cell, finding results in great agreement

with experimental observations.

The model as constructed works well with repressive epigenetic marks, but it would be interesting to extend it to consider also actively transcribing genome, perhaps by employing a co-transcriptional model. In this context, one could use the model to estimate the rate of expression of certain genes with what found in RNA-seq experiments.

In chapter 3 we have developed a mean-field theory of the original model, in analogy of magnetic polymer models on a lattice. First, we have studied the equilibrium phases of a spin-lattice model with Potts-like interaction where however "inert", non-interacting states are pesent. Notably, by generalizing the classic Potts result, we have shown that the equilibrium properties of the model depends not only on the number of active states, but also on the number of inert ones. Second, we have discussed mean-field theory for an homopolymer and found a Flory-like continuous phase transition. By developing a general theory of dense magnetic polymers we show that there is a discontinuous phase transition between a swollen-disordered phase and a compact-ordered one, in agreement with what observed in simulations presented in chapter 1. Finally, we have developed a "Model A" kinetic theory of the model, and we have discussed its non-equilibrium extension. By studying the steady states of the non-equilibrium model we have proven the existence of a compact-disordered phase akin again to what observed numerically in chapter 1.

The mean field-theory has been developed for high-density polymers, and therefore an extension for low-density phases is needed. Furthermore, a possible interesting extension of the model consist in considering higher order approximations to describe phenomena localized in space. In this context, it would be interesting to model the genomic bookmarking as an external non homogeneous magnetic field localized on certain sites of the chain.

In chapter 4 we have developed qa field-theoretical model, describing the dynamics of several chromatin fibers in a dense closed system (the eukaryotic nucleus) with epigenetic interactions. We have first found a phenomenological expression for a free-energy density in the density conserved field, and the magnetization non-conserved field. We have studied the equilibrium of this model by employing the common-tangent construction and have shown that in certain conditions the system is in a demixed phases, where it organizes in high-density clusters characterised by coherent epigenetic marks. We have also discussed the kinetics of this model by employing a set of "Model C" equations and shown that the high-density clusters grow indefinitely in contrast to all experimental observations. Finally, we have modified the model by inserting non-equilibrium terms that induce a first-order reaction between an "active" state of the chromatin which actively partecipate to the epigenetic dynamics, and an "inert" state of the chromatin which does not. Notably, these terms induce an arrested phase separation where the clusters have a stable size at the steady state and coexist with

other clusters of different epigenetic marks.

This model has great potentials as it can be easily refined to include more specific biological non-equilibrium terms, due for example to RNA production, and its predictions could be in principle verified by experiments.

The main results presented in this work have been published in the following peer-reviewed papers:

- D. Michieletto, M. Chiang, D. Colì, A. Papantonis, E. Orlandini, P. R. Cook, and D. Marenduzzo, "Shaping epigenetic memory via genomic bookmarking," *Nucleic Acids Research*, vol. 46, pp. 83–93, nov 2017
- D. Colì, E. Orlandini, D. Michieletto, and D. Marenduzzo, "Magnetic polymer models for epigenetics-driven chromosome folding," *Phys. Rev. E*, vol. 100, p. 052410, Nov 2019
- D. Michieletto, D. Colì, D. Marenduzzo, and E. Orlandini, "Nonequilibrium theory of epigenomic microphase separation in the cell nucleus," *Phys. Rev. Lett.*, vol. 123, p. 228101, Nov 2019

# Bibliography


[1] D. Michieletto, E. Orlandini, and D. Marenduzzo, "Polymer model with epigenetic recoloring reveals a pathway for the de novo establishment and 3d organization of chromatin domains," *Physical Review X*, vol. 6, dec 2016.

[2] D. Michieletto, M. Chiang, D. Colì, A. Papantonis, E. Orlandini, P. R. Cook, and D. Marenduzzo, "Shaping epigenetic memory via genomic bookmarking," *Nucleic Acids Research*, vol. 46, pp. 83–93, nov 2017.

[3] D. Colì, E. Orlandini, D. Michieletto, and D. Marenduzzo, "Magnetic polymer models for epigenetics-driven chromosome folding," *Phys. Rev. E*, vol. 100, p. 052410, Nov 2019.

[4] D. Michieletto, D. Colì, D. Marenduzzo, and E. Orlandini, "Nonequilibrium theory of epigenomic microphase separation in the cell nucleus," *Phys. Rev. Lett.*, vol. 123, p. 228101, Nov 2019.

[5] S. L. Berger, T. Kouzarides, R. Shiekhattar, and A. Shilatifard, "An operational definition of epigenetics," *Genes & Development*, vol. 23, pp. 781–783, Apr. 2009.

[6] U. R. Ernst, M. B. V. Hiel, G. Depuydt, B. Boerjan, A. D. Loof, and L. Schoofs, "Epigenetics and locust life phase transitions," *Journal of Experimental Biology*, vol. 218, pp. 88–99, jan 2015.

[7] D. Crews, "Sex determination: where environment and genetics meet," *Evolution and Development*, vol. 5, pp. 50–55, jan 2003.

[8] C. Park, J. D. Rosenblat, E. Brietzke, Z. Pan, Y. Lee, B. Cao, H. Zuckerman, A. Kalantarova, and R. S. McIntyre, "Stress, epigenetics and depression: A systematic review," *Neuroscience & Biobehavioral Reviews*, vol. 102, pp. 139–152, jul 2019.

[9] B. Alberts, A. D. Johnson, J. Lewis, D. Morgan, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell (Sixth Edition)*. W. W. Norton & Company, 2014.

[10] R. Stadhouders, G. J. Filion, and T. Graf, "Transcription factors and 3d genome conformation in cell-fate decisions," *Nature*, vol. 569, pp. 345–354, may 2019.

[11] S. I. S. Grewal, "Heterochromatin and epigenetic control of gene expression," *Science*, vol. 301, pp. 798–802, aug 2003.

[12] A. Rosa and C. Zimmer, "Computational models of large-scale genome architecture," in *International Review of Cell and Molecular Biology*, pp. 275–349, Elsevier, 2014.

[13] D. Jost, P. Carrivain, G. Cavalli, and C. Vaillant, "Modeling epigenome folding: formation and dynamics of topologically associated chromatin domains," *Nucleic Acids Research*, vol. 42, pp. 9553–9561, aug 2014.

[14] J. R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, and B. Ren, "Topological domains in mammalian genomes identified by analysis of chromatin interactions," *Nature*, vol. 485, pp. 376–380, apr 2012.

[15] C. A. Brackley, S. Taylor, A. Papantonis, P. R. Cook, and D. Marenduzzo, "Non-specific bridging-induced attraction drives clustering of DNA-binding proteins and genome organization," *Proceedings of the National Academy of Sciences*, vol. 110, pp. E3605–E3611, sep 2013.

[16] I. B. Dodd, M. A. Micheelsen, K. Sneppen, and G. Thon, "Theoretical analysis of epigenetic cell memory by nucleosome modification," *Cell*, vol. 129, pp. 813–822, may 2007.

[17] I. B. Dodd and K. Sneppen, "Barriers and silencers: A theoretical toolkit for control and containment of nucleosome-based epigenetic states," *Journal of Molecular Biology*, vol. 414, pp. 624–637, dec 2011.

[18] S. Plimpton, "Fast parallel algorithms for short-range molecular dynamics," *Journal of Computational Physics*, vol. 117, pp. 1–19, Mar. 1995.

[19] T. Garel, H. Orland, and E. Orlandini, "Phase diagram of magnetic polymers," *The European Physical Journal B*, vol. 12, pp. 261–268, Nov. 1999.

[20] P. M. Chaikin and T. C. Lubensky, *Principles of Condensed Matter Physics*. Cambridge University Press, 2000.

[21] R. Cortini, M. Barbi, B. R. Caré, C. Lavelle, A. Lesne, J. Mozziconacci, and J.-M. Victor, "The physics of epigenetics," *Reviews of Modern Physics*, vol. 88, Apr. 2016.

[22] C. H. Waddington, "Canalization of Development and the Inheritance of Acquired Characters," *Nature*, vol. 150, pp. 563–565, Nov. 1942.

[23] C. H. Waddington, *The strategy of the genes. A discussion of some aspects of theoretical biology.* London: George Allen & Unwin, Ltd., 1957.

[24] D. L. Nanney, "Epigentic Control Systems," *Proceedings of the National Academy of Sciences*, vol. 44, pp. 712–717, July 1958.

[25] D. Haig, "The (dual) origin of epigenetics," *Cold Spring Harbor Symposia on Quantitative Biology*, vol. 69, pp. 67–70, Jan. 2004.

[26] M. Gebala, S. L. Johnson, G. J. Narlikar, and D. Herschlag, "Ion counting demonstrates a high electrostatic field generated by the nucleosome," *eLife*, vol. 8, June 2019.

[27] Z. Yang and J. J. Hayes, "The divalent cations Ca2+ and Mg2+. play specific roles in stabilizing histone–DNA interactions within nucleosomes that are partially redundant with the core histone tail domains," *Biochemistry*, vol. 50, pp. 9973–9981, Nov. 2011.

[28] L. Stryer, *Biochemistry (4th edition).* W.H. Freeman & Company, 1995.

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

[30] E. P. Nora, B. R. Lajoie, E. G. Schulz, L. Giorgetti, I. Okamoto, N. Servant, T. Piolot, N. L. van Berkum, J. Meisig, J. Sedat, J. Gribnau, E. Barillot, N. Blüthgen, J. Dekker, and E. Heard, "Spatial partitioning of the regulatory landscape of the x-inactivation centre," *Nature*, vol. 485, pp. 381–385, Apr. 2012.

[31] W. de Laat and D. Duboule, "Topology of mammalian developmental enhancers and their regulatory landscapes," *Nature*, vol. 502, pp. 499–506, Oct. 2013.

[32] B. D. Pope, T. Ryba, V. Dileep, F. Yue, W. Wu, O. Denas, D. L. Vera, Y. Wang, R. S. Hansen, T. K. Canfield, R. E. Thurman, Y. Cheng, G. Gülsoy, J. H. Dennis, M. P. Snyder, J. A. Stamatoyannopoulos, J. Taylor, R. C. Hardison, T. Kahveci, B. Ren, and D. M. Gilbert, "Topologically associating domains are stable units of replication-timing regulation," *Nature*, vol. 515, pp. 402–405, Nov. 2014.

[33] M. J. Blow, T. A. Clark, C. G. Daum, A. M. Deutschbauer, A. Fomenkov, R. Fries, J. Froula, D. D. Kang, R. R. Malmstrom, R. D. Morgan, J. Posfai, K. Singh, A. Visel, K. Wetmore, Z. Zhao, E. M. Rubin, J. Korlach, L. A. Pennacchio, and R. J. Roberts, "The epigenomic landscape of prokaryotes," *PLOS Genetics*, vol. 12, pp. 1–28, 02 2016.

[34] M. A. Dawson and T. Kouzarides, "Cancer epigenetics: From mechanism to therapy," *Cell*, vol. 150, pp. 12–27, jul 2012.

[35] G. J. Filion, J. G. van Bemmel, U. Braunschweig, W. Talhout, J. Kind, L. D. Ward, W. Brugman, I. J. de Castro, R. M. Kerkhoven, H. J. Bussemaker, and B. van Steensel, "Systematic protein location mapping reveals five principal chromatin types in drosophila cells," *Cell*, vol. 143, pp. 212–224, oct 2010.

[36] J. Ernst and M. Kellis, "Discovery and characterization of chromatin states for systematic annotation of the human genome," *Nature Biotechnology*, vol. 28, pp. 817–825, jul 2010.

[37] H. Kimura, "Histone modifications for human epigenome analysis," *Journal of Human Genetics*, vol. 58, pp. 439–445, jun 2013.

[38] L. Balakrishnan and B. Milavetz, "Decoding the histone h4 lysine 20 methylation mark," *Critical Reviews in Biochemistry and Molecular Biology*, vol. 45, pp. 440–452, aug 2010.

[39] M. D. Shahbazian and M. Grunstein, "Functions of site-specific histone acetylation and deacetylation," *Annual Review of Biochemistry*, vol. 76, pp. 75–100, jun 2007.

[40] M. Ptashne, "On the use of the word 'epigenetic'," *Current Biology*, vol. 17, pp. R233–R236, Apr. 2007.

[41] B. Yang and A. L. Kirchmaier, "Bypassing the catalytic activity of SIR2 for SIR protein spreading in saccharomyces cerevisiae," *Molecular Biology of the Cell*, vol. 17, pp. 5287–5297, Dec. 2006.

[42] J. R. Danzer, "Mechanisms of HP1-mediated gene silencing in drosophila," *Development*, vol. 131, pp. 3571–3580, June 2004.

[43] J. Wysocka, T. Swigut, T. A. Milne, Y. Dou, X. Zhang, A. L. Burlingame, R. G. Roeder, A. H. Brivanlou, and C. D. Allis, "WDR5 associates with histone h3 methylated at k4 and is essential for h3 k4 methylation and vertebrate development," *Cell*, vol. 121, pp. 859–872, June 2005.

[44] D. Benveniste, H.-J. Sonntag, G. Sanguinetti, and D. Sproul, "Transcription factor binding predicts histone modifications in human cell lines," *Proceedings of the National Academy of Sciences*, vol. 111, pp. 13367–13372, Sept. 2014.

[45] M. Vendruscolo, "Determination of conformationally heterogeneous states of proteins," *Current Opinion in Structural Biology*, vol. 17, pp. 15–20, Feb. 2007.

[46] A. R. Khokhlov, A. Y. Grosberg, and V. S. Pande, *Statistical Physics of Macromolecules (Polymers and Complex Materials)*. American Institute of Physics, 2002.

[47] M. Tark-Dame, R. van Driel, and D. W. Heermann, "Chromatin folding - from biology to polymer models and back," *Journal of Cell Science*, vol. 124, pp. 839–845, mar 2011.

[48] J. Langowski, "Polymer chain models of DNA and chromatin," *The European Physical Journal E*, vol. 19, pp. 241–249, mar 2006.

[49] Y. Cui and C. Bustamante, "Pulling a single chromatin fiber reveals the forces that maintain its higher-order structure," *Proceedings of the National Academy of Sciences*, vol. 97, pp. 127–132, jan 2000.

[50] K. Bystricky, P. Heun, L. Gehlen, J. Langowski, and S. M. Gasser, "Long-range compaction and flexibility of interphase chromatin in budding yeast analyzed by high-resolution imaging techniques," *Proceedings of the National Academy of Sciences*, vol. 101, pp. 16495–16500, nov 2004.

[51] A. Routh, S. Sandin, and D. Rhodes, "Nucleosome repeat length and linker histone stoichiometry determine chromatin fiber structure," *Proceedings of the National Academy of Sciences*, vol. 105, pp. 8872–8877, jun 2008.

[52] H. Hajjoul, J. Mathon, H. Ranchon, I. Goiffon, J. Mozziconacci, B. Albert, P. Carrivain, J.-M. Victor, O. Gadal, K. Bystricky, and A. Bancaud, "High-throughput chromatin motion tracking in living yeast reveals the flexibility of the fiber throughout the genome," *Genome Research*, vol. 23, pp. 1829–1838, sep 2013.

[53] P. E. Rouse, "A theory of the linear viscoelastic properties of dilute solutions of coiling polymers," *The Journal of Chemical Physics*, vol. 21, pp. 1272–1280, jul 1953.

[54] S. Huet, C. Lavelle, H. Ranchon, P. Carrivain, J.-M. Victor, and A. Bancaud, "Relevance and limitations of crowding, fractal, and polymer models to describe nuclear architecture," in *International Review of Cell and Molecular Biology*, pp. 443–479, Elsevier, 2014.

[55] H. Risken, *The Fokker-Planck Equation*. Springer Berlin Heidelberg, 1989.

[56] E. G. Timoshenko, Y. A. Kuznetsov, and K. A. Dawson, "Conformational transitions of heteropolymers in dilute solutions," *Physical Review E*, vol. 57, pp. 6801–6814, jun 1998.

[57] D. M. MacAlpine and G. Almouzni, "Chromatin and DNA replication," *Cold Spring Harbor Perspectives in Biology*, vol. 5, pp. a010207–a010207, jun 2013.

[58] M. A. Micheelsen, N. Mitarai, K. Sneppen, and I. B. Dodd, "Theory for the stability and regulation of epigenetic landscapes," *Physical Biology*, vol. 7, p. 026010, jun 2010.

[59] D. Jost, "Bifurcation in epigenetics: Implications in development, proliferation, and diseases," *Physical Review E*, vol. 89, jan 2014.

[60] S. H. Strogatz, *Nonlinear Dynamics And Chaos: With Applications To Physics, Biology, Chemistry And Engineering (Studies in Nonlinearity)*. CRC Press, 1994.

[61] D. Horvath, "Common mechanisms regulate flowering and dormancy," *Plant Science*, vol. 177, pp. 523–531, dec 2009.

[62] J. E. Ferrell, "Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability," *Current Opinion in Cell Biology*, vol. 14, pp. 140–148, apr 2002.

[63] M. Onishi, G.-G. Liou, J. R. Buchberger, T. Walz, and D. Moazed, "Role of the conserved sir3-BAH domain in nucleosome binding and silent chromatin assembly," *Molecular Cell*, vol. 28, pp. 1015–1028, dec 2007.

[64] J. C. Eissenberg and L. L. Wallrath, "Heterochromatin, position effects, and the genetic dissection of chromatin," in *Progress in Nucleic Acid Research and Molecular Biology Volume 74*, pp. 275–299, Elsevier, 2003.

[65] A. M. Näär, B. D. Lemon, and R. Tjian, "Transcriptional coactivator complexes," *Annual Review of Biochemistry*, vol. 70, pp. 475–501, jun 2001.

[66] M. J. Carrozza, B. Li, L. Florens, T. Suganuma, S. K. Swanson, K. K. Lee, W.-J. Shia, S. Anderson, J. Yates, M. P. Washburn, and J. L. Workman, "Histone h3 methylation by set2 directs deacetylation of coding regions by rpd3s to suppress spurious intragenic transcription," *Cell*, vol. 123, pp. 581–592, nov 2005.

[67] G. E. Zentner and S. Henikoff, "Regulation of nucleosome dynamics by histone modifications," *Nature Structural & Molecular Biology*, vol. 20, pp. 259–266, mar 2013.

[68] K. Sneppen, M. A. Micheelsen, and I. B. Dodd, "Ultrasensitive gene regulation by positive feedback loops in nucleosome modification," *Molecular Systems Biology*, vol. 4, p. 182, jan 2008.

[69] A. Rosa and R. Everaers, "Structure and dynamics of interphase chromosomes," *PLoS Computational Biology*, vol. 4, p. e1000153, aug 2008.

[70] K. Kremer and G. S. Grest, "Dynamics of entangled linear polymer melts: A molecular-dynamics simulation," *The Journal of Chemical Physics*, vol. 92, no. 8, p. 5057, 1990.

[71] D. Murugesapillai, M. J. McCauley, R. Huo, M. H. N. Holte, A. Stepanyants, L. J. Maher, N. E. Israeloff, and M. C. Williams, "DNA bridging and looping

by HMO1 provides a mechanism for stabilizing nucleosome-free chromatin," *Nucleic Acids Research*, vol. 42, pp. 8996–9004, jul 2014.

[72] N. D. Mermin and H. Wagner, "Absence of ferromagnetism or antiferromagnetism in one- or two-dimensional isotropic heisenberg models," *Physical Review Letters*, vol. 17, pp. 1133–1136, nov 1966.

[73] G. N. Lewis, "A new principle of equilibrium," *Proceedings of the National Academy of Sciences*, vol. 11, pp. 179–183, Mar. 1925.

[74] C. Gardiner, *Stochastic Methods: A Handbook for the Natural and Social Sciences (Springer Series in Synergetics)*. Springer, 2009.

[75] P. J. Flory, "Thermodynamics of high polymer solutions," *The Journal of Chemical Physics*, vol. 10, pp. 51–61, jan 1942.

[76] M. E. Fisher and M. N. Barber, "Scaling theory for finite-size effects in the critical region," *Physical Review Letters*, vol. 28, pp. 1516–1519, June 1972.

[77] K. Binder, "Critical properties from monte carlo coarse graining and renormalization," *Physical Review Letters*, vol. 47, pp. 693–696, aug 1981.

[78] T. Garel and H. Orland, "Mean-field model for protein folding," *Europhysics Letters (EPL)*, vol. 6, pp. 307–310, June 1988.

[79] P. V. Kharchenko, A. A. Alekseyenko, Y. B. Schwartz, A. Minoda, N. C. Riddle, J. Ernst, P. J. Sabo, E. Larschan, A. A. Gorchakov, T. Gu, D. Linder-Basso, A. Plachetka, G. Shanower, M. Y. Tolstorukov, L. J. Luquette, R. Xi, Y. L. Jung, R. W. Park, E. P. Bishop, T. K. Canfield, R. Sandstrom, R. E. Thurman, D. M. MacAlpine, J. A. Stamatoyannopoulos, M. Kellis, S. C. R. Elgin, M. I. Kuroda, V. Pirrotta, G. H. Karpen, and P. J. Park, "Comprehensive analysis of the chromatin landscape in drosophila melanogaster," *Nature*, vol. 471, pp. 480–485, Dec. 2010.

[80] A. Pombo and N. Dillon, "Three-dimensional genome architecture: players and mechanisms," *Nature Reviews Molecular Cell Biology*, vol. 16, pp. 245–257, Mar. 2015.

[81] A. Y. Grosberg and E. Shakhnovich, "Theory of phase transitions of the coil-globule type in a heteropolymer chain with disordered sequence of links," *Sov.Phys. JETP*, vol. 64, pp. 1284–1290, Dec. 1986.

[82] T. Garel, H. Orland, and E. Pitard, "Protein Folding and Heteropolymers," in *Series on Directions in Condensed Matter Physics*, pp. 387–443, World Scientific, Dec. 1997.

[83] F. Y. Wu, "The potts model," *Reviews of Modern Physics*, vol. 54, pp. 235–268, Jan. 1982.

[84] N. A. Hathaway, O. Bell, C. Hodges, E. L. Miller, D. S. Neel, and G. R. Crabtree, "Dynamics and memory of heterochromatin in living cells," *Cell*, vol. 149, pp. 1447–1460, jun 2012.

[85] A. Angel, J. Song, C. Dean, and M. Howard, "A polycomb-based switch underlying quantitative epigenetic memory," *Nature*, vol. 476, pp. 105–108, jul 2011.

[86] A. Collinson, A. J. Collier, N. P. Morgan, A. R. Sienerth, T. Chandra, S. Andrews, and P. J. Rugg-Gunn, "Deletion of the polycomb-group protein EZH2 leads to compromised self-renewal and differentiation defects in human embryonic stem cells," *Cell Reports*, vol. 17, pp. 2700–2714, Dec. 2016.

[87] N. Lodhi, Y. Ji, and A. Tulin, "Mitotic bookmarking: Maintaining post-mitotic reprogramming of transcription reactivation," *Current Molecular Biology Reports*, vol. 2, pp. 10–15, Mar 2016.

[88] S. S. Teves, L. An, A. S. Hansen, L. Xie, X. Darzacq, and R. Tjian, "A dynamic mode of mitotic bookmarking by transcription factors," *eLife*, vol. 5, p. e22280, nov 2016.

[89] N. E. Follmer, A. H. Wani, and N. J. Francis, "A polycomb group protein is retained at specific sites on chromatin in mitosis," *PLoS Genetics*, vol. 8, p. e1003135, Dec. 2012.

[90] F. Ciabrelli, F. Comoglio, S. Fellous, B. Bonev, M. Ninova, Q. Szabo, A. Xuéreb, C. Klopp, A. Aravin, R. Paro, F. Bantignies, and G. Cavalli, "Stable polycomb-dependent transgenerational inheritance of chromatin states in drosophila," *Nature Genetics*, vol. 49, pp. 876–886, apr 2017.

[91] E. Heard and R. A. Martienssen, "Transgenerational epigenetic inheritance: Myths and mechanisms," *Cell*, vol. 157, pp. 95–109, mar 2014.

[92] M. Barbieri, M. Chotalia, J. Fraser, L.-M. Lavitas, J. Dostie, A. Pombo, and M. Nicodemi, "Complexity of chromatin folding is captured by the strings and binders switch model," *Proceedings of the National Academy of Sciences*, vol. 109, pp. 16173–16178, sep 2012.

[93] S. S. Rao, M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov, J. T. Robinson, A. L. Sanborn, I. Machol, A. D. Omer, E. S. Lander, and E. L. Aiden, "A 3d map of the human genome at kilobase resolution reveals principles of chromatin looping," *Cell*, vol. 159, pp. 1665–1680, Dec. 2014.

[94] S. Berry, C. Dean, and M. Howard, "Slow chromatin dynamics allow polycomb target genes to filter fluctuations in transcription factor activity," *Cell Systems*, vol. 4, pp. 445–457.e8, apr 2017.

[95] F. Laprell, K. Finkl, and J. Müller, "Propagation of polycomb-repressed chromatin requires sequence-specific recruitment to DNA," *Science*, vol. 356, pp. 85–88, mar 2017.

[96] P. V. Kharchenko, A. A. Alekseyenko, Y. B. Schwartz, A. Minoda, N. C. Riddle, J. Ernst, P. J. Sabo, E. Larschan, A. A. Gorchakov, T. Gu, D. Linder-Basso, A. Plachetka, G. Shanower, M. Y. Tolstorukov, L. J. Luquette, R. Xi, Y. L. Jung, R. W. Park, E. P. Bishop, T. K. Canfield, R. Sandstrom, R. E. Thurman, D. M. MacAlpine, J. A. Stamatoyannopoulos, M. Kellis, S. C. R. Elgin, M. I. Kuroda, V. Pirrotta, G. H. Karpen, and P. J. Park, "Comprehensive analysis of the chromatin landscape in drosophila melanogaster," *Nature*, vol. 471, pp. 480–485, dec 2010.

[97] R. Blecher-Gonen, Z. Barnett-Itzhaki, D. Jaitin, D. Amann-Zalcenstein, D. Lara-Astiaso, and I. Amit, "High-throughput chromatin immunoprecipitation for genome-wide mapping of in vivo protein-DNA interactions and epigenomic states," *Nature Protocols*, vol. 8, pp. 539–554, feb 2013.

[98] S. Kadauke and G. A. Blobel, "Chromatin loops in gene regulation," *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, vol. 1789, pp. 17–25, jan 2009.

[99] T. Sexton, E. Yaffe, E. Kenigsberg, F. Bantignies, B. Leblanc, M. Hoichman, H. Parrinello, A. Tanay, and G. Cavalli, "Three-dimensional folding and functional organization principles of the drosophila genome," *Cell*, vol. 148, pp. 458–472, feb 2012.

[100] P. C. Hohenberg and B. I. Halperin, "Theory of dynamic critical phenomena," *Reviews of Modern Physics*, vol. 49, pp. 435–479, jul 1977.

[101] P. J. Flory, "The configuration of real polymer chains," *The Journal of Chemical Physics*, vol. 17, pp. 303–310, mar 1949.

[102] P.-G. Gennes, *Scaling Concepts in Polymer Physics*. NCROL, dec 1979.

[103] Y. Miyaki, Y. Einaga, and H. Fujita, "Excluded-volume effects in dilute polymer solutions. 7. very high molecular weight polystyrene in benzene and cyclohexane," *Macromolecules*, vol. 11, pp. 1180–1186, Nov. 1978.

[104] A. Yamamoto, M. Fujii, G. Tanaka, and H. Yamakawa, "More on the analysis of dilute solution data: Polystyrenes prepared anionically in tetrahydrofuran," *Polymer Journal*, vol. 2, pp. 799–811, Nov. 1971.

[105] A. D. Sokal, "Monte carlo methods for the self-avoiding walk," *Nuclear Physics B - Proceedings Supplements*, vol. 47, pp. 172–179, Mar. 1996.

[106] C. J. Geyer and E. A. Thompson, "Annealing markov chain monte carlo with applications to ancestral inference," *Journal of the American Statistical Association*, vol. 90, no. 431, pp. 909–920, 1995.

[107] M. C. Tesi, E. J. J. van Rensburg, E. Orlandini, and S. G. Whittington, "Interacting self-avoiding walks and polygons in three dimensions," *Journal of Physics A: Mathematical and General*, vol. 29, pp. 2451–2463, may 1996.

[108] F. Wang and D. P. Landau, "Efficient, multiple-range random walk algorithm to calculate the density of states," *Physical Review Letters*, vol. 86, pp. 2050–2053, Mar. 2001.

[109] M. E. J. Newman, *Monte Carlo Methods in Statistical Physics.* Clarendon Press, apr 1999.

[110] K. Huang, *Statistical Mechanics, 2nd Ed.* Wiley India Pvt. Limited, 2008.

[111] M. L. Huggins, "Solutions of long chain compounds," *The Journal of Chemical Physics*, vol. 9, pp. 440–440, may 1941.

[112] H. Orland, C. Itzykson, and C. de Dominicis, "An evaluation of the number of hamiltonian paths," *Journal de Physique Lettres*, vol. 46, no. 8, pp. 353–357, 1985.

[113] A. M. Nemirovsky and M. D. Coutinho-Filho, "Packing flexible polymer chains on a lattice," *Physical Review A*, vol. 39, pp. 3120–3129, mar 1989.

[114] A. M. Nemirovsky, J. Dudowicz, and K. F. Freed, "Thermodynamics of a dense self-avoiding walk with contact interactions," *Journal of Statistical Physics*, vol. 67, pp. 395–412, Apr 1992.

[115] B. Duplantier, "Tricritical polymer chains in or below three dimensions," *Europhysics Letters (EPL)*, vol. 1, pp. 491–498, may 1986.

[116] A. Altland, *Condensed Matter Field Theory.* Cambridge University Press, apr 2010.

[117] T. Cremer, M. Cremer, B. Hübner, H. Strickfaden, D. Smeets, J. Popken, M. Sterr, Y. Markaki, K. Rippe, and C. Cremer, "The 4d nucleome: Evidence for a dynamic nuclear landscape based on co-aligned active and inactive nuclear compartments," *FEBS Letters*, vol. 589, pp. 2931–2943, may 2015.

[118] S. F. Edwards, "The statistical mechanics of polymers with excluded volume," *Proceedings of the Physical Society*, vol. 85, pp. 613–624, Apr. 1965.

[119] F. Schmid, "Self-consistent-field theories for complex fluids," *Journal of Physics: Condensed Matter*, vol. 10, pp. 8105–8138, Sept. 1998.

[120] Y. Fosado, D. Michieletto, and D. Marenduzzo, "Dynamical scaling and phase coexistence in topologically constrained DNA melting," *Physical Review Letters*, vol. 119, sep 2017.

[121] V. M. Kendon, M. E. Cates, I. Pagonabarraga, J.-C. Desplat, and P. Bladon, "Inertial effects in three-dimensional spinodal decomposition of a symmetric binary fluid mixture: a lattice boltzmann study," *Journal of Fluid Mechanics*, vol. 440, pp. 147–203, aug 2001.

[122] R. Wittkowski, A. Tiribocchi, J. Stenhammar, R. J. Allen, D. Marenduzzo, and M. E. Cates, "Scalar $\varphi 4$ field theory for active-particle phase separation," *Nature Communications*, vol. 5, July 2014.

[123] J. Kockelkoren and H. Chaté, "Late stages of coarsening in model c," *Physica D: Nonlinear Phenomena*, vol. 168-169, pp. 80–92, aug 2002.

[124] S. F. Pinter, R. I. Sadreyev, E. Yildirim, Y. Jeon, T. K. Ohsumi, M. Borowsky, and J. T. Lee, "Spreading of x chromosome inactivation via a hierarchy of defined polycomb stations," *Genome Research*, vol. 22, pp. 1864–1876, Sept. 2012.

[125] L. Hilbert, Y. Sato, H. Kimura, F. Jülicher, A. Honigmann, V. Zaburdaev, and N. Vastenhouw, "Transcription establishes microenvironments that organize euchromatin," *bioRxiv*, 2017.

[126] D. Zwicker, M. Decker, S. Jaensch, A. A. Hyman, and F. Jülicher, "Centrosomes are autocatalytic droplets of pericentriolar material organized by centrioles," *Proceedings of the National Academy of Sciences*, vol. 111, no. 26, pp. E2636–E2645, 2014.

[127] C. Brackley, J. Johnson, D. Michieletto, A. Morozov, M. Nicodemi, P. Cook, and D. Marenduzzo, "Nonequilibrium chromosome looping via molecular slip links," *Physical Review Letters*, vol. 119, sep 2017.

[128] C. Brangwynne, P. Tompa, and R. Pappu, "Polymer physics of intracellular phase transitions," *Nature Physics*, vol. 11, pp. 899–904, 11 2015.

[129] S. Henikoff, "Nucleosome destabilization in the epigenetic regulation of gene expression," *Nature Reviews Genetics*, vol. 9, pp. 15–26, jan 2008.

[130] S. C. Glotzer, E. A. D. Marzio, and M. Muthukumar, "Reaction-controlled morphology of phase-separating mixtures," *Physical Review Letters*, vol. 74, pp. 2034–2037, mar 1995.

[131] C. A. Brackley, B. Liebchen, D. Michieletto, F. Mouvet, P. R. Cook, and D. Marenduzzo, "Ephemeral protein binding to DNA shapes stable nuclear bodies and chromatin domains," *Biophysical Journal*, vol. 112, pp. 1085–1093, mar 2017.