UNIVERSITÀ DEGLI STUDI DI MILANO

FACOLTÀ DI SCIENZE E TECNOLOGIE

Corso di laurea magistrale in Fisica

Anno accademico 2016/2017

SPONTANEOUS DOMAIN FORMATION IN A SIMPLE MODEL OF DISORDERED HETEROPOLYMERS

Relatore interno: Prof. Guido TIANA
Relatore esterno: Prof. Marco Cosentino LAGOMARSINO
Correlatore: Prof. Marco GHERARDI

Autore: Matteo NEGRI
Matricola: 864807
PACS: 87.15.-v

Firma del relatore:

Contents

Chapter 1

Introduction

1.1 The biological problem

DNA is packed in the nucleus of eukaryotic cells in the form of a fiber of the alleged diameter of 30 nm, which is called *chromatin* (see figure 1.1); this fiber is the first level of spatial organization of the genome. The introduction of a new experimental technique in the 2000s (which is called 3C; for a review see [1]) made possible it to study the spatial organization of chromatin inside the nucleus.

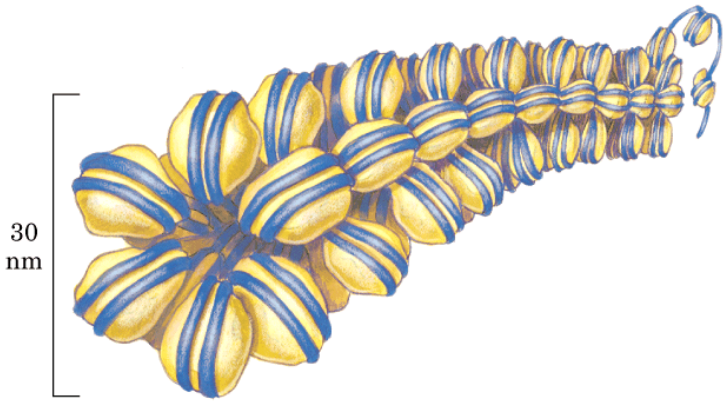


Figure 1.1: Chromatin is a fiber made of DNA (in blue) rolled around histones (in yellow) to form nucleosomes; the nucleosomes are rolled around themselves with an alleged diameter of 30 nm. *Image adapted from: Nelson, David L., Albert L. Lehninger, and Michael M. Cox. Lehninger principles of biochemistry. Macmillan, 2008.*

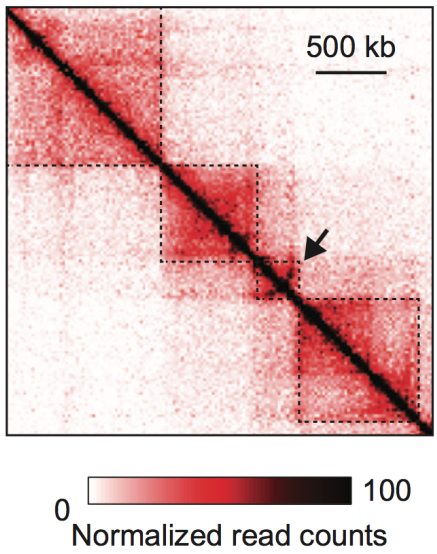
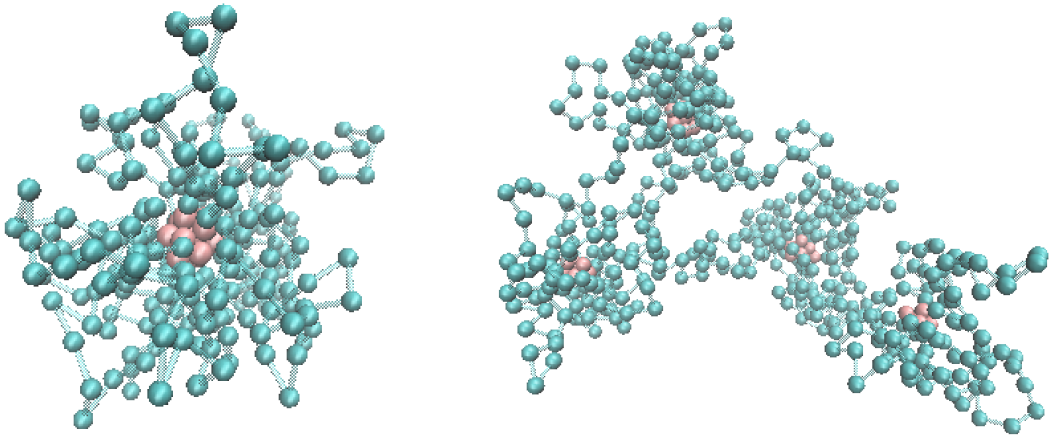


Figure 1.2: The contact matrix between chromatin sites shows spatial domains as blocks of typical size ∼1Mb. The matrix element (i, j) contains the probability (coded in red) to find the site i in the interaction range of the site j. The sites belonging to a certain domain rarely interact with sites in other domains. Using a canonical ensemble it is possible to infer the strength of the interaction from the contact frequency. *Image adapted from: Ringrose, Leonie. Epigenetics and Systems Biology. Academic Press, 2017.*

The datum provided by this kind of experiments is a contact matrix in which the element (i, j) contains the probability that the site of chromatin i is closer than a certain distance to the site j. With site of chromatin we mean a short region of DNA whose size is determined by experimental resolution; typically this size is thousands of DNA bases (∼ kb). Combining these informations with the genetic activity data allowed to understand that **the three-dimensional spatial conformation of the DNA is important for the regulation of gene transcription** [1].

Interphase is the longest period of the cell cycle, in which the cell performs its normal vital functions. During interphase chromatin has a non-random spatial organization made of spatial domains encapsulated in a hierarchical way. The most important spatial domain is the one with the typical size of 10^6 DNA bases (1 Mb), because it is the same size of the transcriptional mechanism that regulates genes with distant DNA sequences. These structures are called *Topological Associated Domains* (TADs): they are localized spatial domains in



(a) **(b)**

Figure 1.3: Collapse of a polymer of N monomers in localized spatial domains. The interacting monomers that form the core are pictured in red; the monomers that form the corona are pictured in light blue. a) Single rosette, $N = 257$. b) Multiple rosette with four domains, $N = 513$.

which the chromatin sites interact preferentially with sites of the same domain; the TADs appear as squared blocks in the contact matrix (see figure 1.2). Additionally, experimental contact matrices are characterized by a power law dependence of the interaction probability of a couple of sites P as a function of their genomic distance s, $P(s) \sim s^{-\alpha}$, where α has a range of values from 0.4 to 1.5 depending on the specific TAD.

Given the great interest in the mechanism of formation of the TADs, different classes of physical models have been developed (presented in [2, 3]) which describe the chromatin fiber as a polymer, choosing different interactions between the monomers depending on which property of the real system is described.

The actual molecular mechanism that mediates the interactions within chromatin is not currently understood and it is an open problem in biophysics; nevertheless, it is reasonable to think that the interaction is mediated by proteins that are bound to chromatin and that interact with each other. In the light of this, a trivial mechanism for the formation of TADs is that there are a number of different protein types equal to the number of TADs, and each type dimerize only with proteins of the same type. Since there are few thousands

TADs in mammalian genomes, this would require the presence of thousands of different types of DNA-binding proteins. It is more reasonable to think that only a small number of proteins is responsible for the interactions between the chromatin sites. In particular it has been proven that a protein called CTCF [4] is fundamental for the TADs formation [5]: this protein is thought to be responsible of a short-ranged attraction between distant sites.

Between the polymeric models proposed to describe the TADs formation with only few different type of interactions [6, 7, 8, 9], the simplest one is a polymeric chain in which only a monomer each n interacts with the others of the same type [10]. Monte Carlo simulations of this model show configurations in which the polymer is collapsed into a *multiple rosette* configuration: each rosette is made by a *core* of interacting monomers and by a *corona* of loops of non interacting monomers (see figure 1.3). The idea is that each rosette describes a TAD. Following [10] we refer to this model as the *heteropolymer with bridging interaction*, or *bridged heteropolymer*. However, from the published MC simulation it is not clear whether the rosette are a stable phase of the system, or just a metastable feature on the way to a standard, random globule.

This thesis studies the heteropolymer with bridging interaction with a Monte Carlo algorithm optimized for disordered systems, namely proteins. The model and its phenomenology will be described in section 1.3. The main goals of this work are two: the first is to show that the the multiple rosette configuration is a proper stable phase of the system; the second is to understand if the multiple rosette phase emerges even for a disordered distribution of the interactions.

Chapter 2 is dedicated to the first goal: the stability of the multiple rosette configuration. In section 2.1 we give a brief description of the algorithm; in section 2.2 we present the model; in sections 2.3 and 2.4 we reproduce some results from [10]; in section 2.5 we describe our results for the multiple rosette configuration.

In Chapter 3 we present the completely new study of the disordered heteropolymer with bridging interaction. In section 3.1 we give a qualitative theoretical argument to explain why we expect the disordered model to exhibit a multiple rosette state. Section 3.2 describes the results of the simulations of the disordered heteropolymer with bridging interaction; in this section we also address the self-averaging issue. Lastly, in section 3.3 we present a number of considerations on the contact matrices obtained from the simulations: we suggest that the disorder could be a fundamental ingredient for the description of the experimental data.

1.2 Useful concepts from polymer theory

Before presenting the bridged heteropolymer in the following section, it is useful to recall briefly some of the main concepts from classical polymer theory (for the complete detailed theory, see for example [11]). In particular here we present the phenomenology of the homopolymer, that will be useful to describe the possible phases of the bridged heteropolymer.

The homopolymer is a chain of monomers which have all the same interaction: hard core repulsion and short-ranged attraction. If the polymer is flexible, namely there are no many-body interactions that confer stiffness to it, the system undergoes a second-order phase transition between two different states:

- Above the critical temperature, the polymer is in the *coil* phase: its main characteristic is to occupy on average a spherical volume of radius $R_g \sim N^{3/5}$, where N is the number of monomers in the polymer. In this phase the sphere is almost and empty volume: computing the spatial density of monomers as $\rho = N/R_g$ we obtain $\rho \sim N^{-4/5}$, that tends to zero for $N \gg 1$.
- Under the critical temperature the polymer is in the *random globule* phase: it occupy a spherical volume of radius $R_g \sim N^{1/3}$, meaning that the spatial density of monomers inside the globule is constant: $\rho \sim N^0$.
- At a temperature called *theta point* the polymer has a configuration similar to a coil, but the radius of the spherical volume it occupies has a different law: $R_g \sim N^{1/2}$. This behaviour is the same of an ideal chain, namely a chain of monomers that have no interaction so that the chain chain can cross itself. It is instructive to note that the relation we just enunciated is analogous to a random walk.

In homopolymers the transition temperature is close to the theta point, their difference going to zero as $1/N^{1/2}$. The transition is usually referred to as the *theta point transition* (for a simple picture of this transition see figure 2.3 in chapter 2).

The radius of the spherical volume occupied on average by the polymer is called *radius of gyration* and is defined as the average square distance of the monomers to their centre of mass:

$$R_g := \sqrt{\frac{1}{N}\sum_i (r_i - \bar{r})^2}$$

where $\bar{r}$ is the coordinate of the centre of mass. This can also be expressed as the mean square distance between all the possible couples of monomers

$$R_g = \sqrt{\frac{1}{2N^2}\sum_{ij}(r_i - r_j)^2}$$

by expanding the term in parentheses and using the definition of $\bar{r}$.
To describe the theta transition is useful to define the *swelling parameter* α, which will be used also in chapter 2, as the ratio

$$\alpha = \frac{R_g}{R_0}$$

where R_0 is the radius of gyration of the polymer at the theta point; given the bond distance λ between two consecutive monomers, its expression reads $R_0 = \lambda(3/5N)^{1/2}$.

In the framework of a mean-field theory, the free energy F for the homopolymer can be written in terms of the swelling parameter α as it follows [12]:

$$\beta F(\alpha) = \frac{N^{1/2}B(\beta)}{\lambda^3\alpha^3} + \frac{C(\beta)}{\lambda^6\alpha^6} + \frac{1}{\alpha^2} + \frac{3\alpha^2}{2} \tag{1.1}$$

where $\beta = 1/k_BT$ and $B(\beta)$ and $C(\beta)$ are respectively the second and the third viral coefficient. While the first and the second term come from the virial expansion of the partition function, the third and the fourth ones are respectively the entropy of the globule state and of the coil state. The minimization respect to α of the equation 1.1 describes the phases of the transition: the coil state corresponds to $\alpha > 1$, the globule state corresponds to $\alpha < 1$ and the theta point corresponds to $\alpha = 1$.

Given a solution of many identical polymers above the theta point, we define the *overlapping threshold* as the global density of monomers c^* at which the polymeric coils start to be densely packed (see figure 1.4 b). This threshold divides the *dilute regime* $c < c^*$ from the *semi-dilute regime* $c > c^*$. We expect c^* to be comparable with the local density of monomers inside a single random coil. In the semi-dilute regime the solution is more concentrated and the coils overlap with each other (see figure 1.4 c).

This regime is particularly useful to define a fundamental concept in polymer physics: the concept of *blobs*. A blob is a spherical region of diameter ξ defined

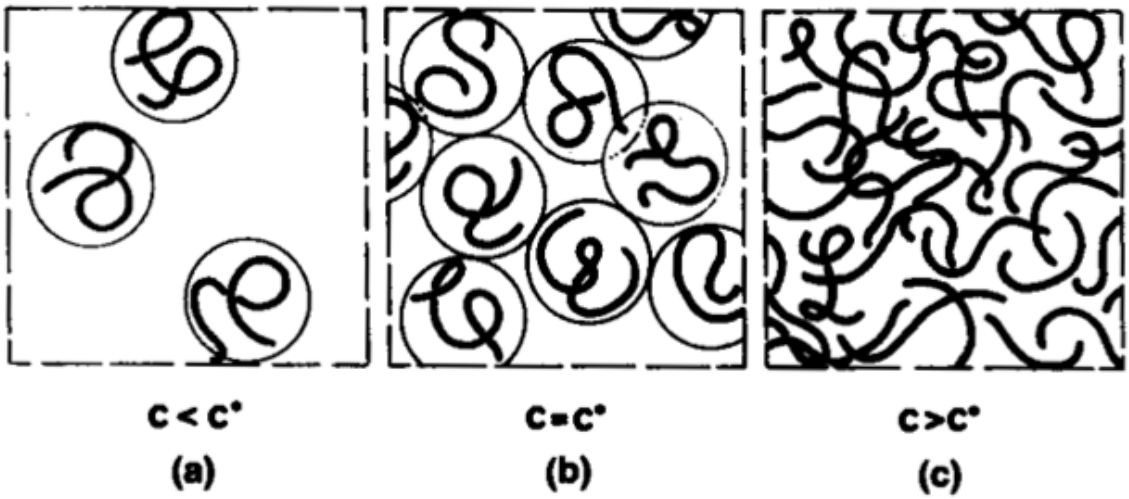


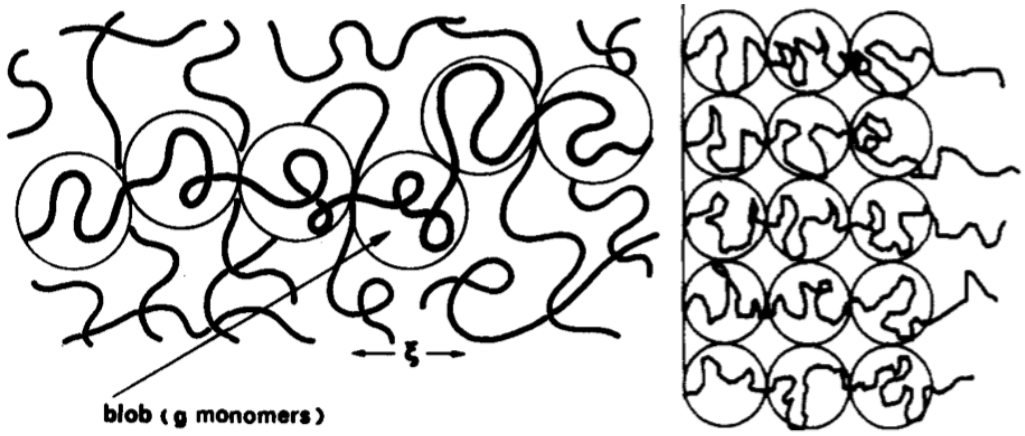
Figure 1.4: A solution of identical polymers above the theta point displays an overlapping threshold c^* at which the random coil are contiguous (b). If $c < c^*$ the solution is in the dilute regime and the global density is lower that the local density inside a coil (a). If $c < c^*$ the solution is in the semi-dilute regime: the solution is more concentrated and the coils overlap with each other (c).

by the correlation length between monomers: because of excluded volume effects, each blob contains predominantly monomers of a single chain. Within a blob, the polymer resembles a self-avoiding walk which gives a number g of monomers per blob of order $g = (D/\lambda)^{1/\nu}$, where ν is an exponent that depends on the system considered ($\nu = 3/5$ in the mean-field theory). Focusing on a single polymeric chain in the semi-dilute solution, we can picture the polymer as a chain of blob super-monomers of size ξ (see figure 1.5 a); since the blobs do not interact with each other, the chain of blobs can be regarded as an ideal chain on large scale . Additionally, this picture allows to compute the free energy F of a polymer in a semi-dilute regime simply as proportional to the number of blobs of the system N_{blob}, namely

$$\beta F = N_{\text{blob}},$$

with $\beta = 1/k_B T$ (see for example [11]).

The concept of blobs is particularly useful to describe semi-dilute solutions in specific situations such as densely packed polymers grafted to a surface. In fact in these situations the size of the blobs ξ is determined by the surface density σ of polymers that have one head bound to the surface itself (a plane



(a) **(b)**

Figure 1.5

surface covered with identical polymeric strands is outlined in figure 1.5 b): it holds that $\xi^2 = \sigma$.

In this thesis we will consider *star polymers*: they are made by a spherical *core* whose surface is covered with polymeric strands, which form a spherical *corona*. These strands are called *arms* of the star polymer. In particular in chapter 3 we will find an expression for the entropy of a star polymer with a certain random distribution of length of the arms. Following [10] we will use this expression to approximate the entropy of a rosette-like state of the bridged heteropolymer.

1.3 Bridging interaction

In this section we briefly sum up the results described in [10] for the heteropolimer with bridging interaction, namely a polymer in which only a monomer each n interacts with the others of the same type.

The collapse of the polymer to a single rosette state is regulated by the competition entropic cost of forming loops and the attractive interaction between bridgers. It is worth to note that this collapse displays an abrupt jump in the order parameter α, resembling a first order phase transition; this is an important difference with the theta-point collapse, which is a second order phase transition.

The density of bridgers plays an important role in the collapse of the polymer to a single rosette state. In fact, in the limit of only two bridging monomers placed at both ends of the polymer, we have a switch-like first-order phase transition; in the opposite limit where all monomers are bridgers we recall a homopolimeric second-order phase transition. A crossover is thought to happen between these two regimes.

The collapse of the polymer into a multiple rosette state is given by the competition between the entropic cost of forming a rosette with many loops and the energetic cost of forming multiple small cores - namely a surface tension contribution. The entropic term of a rosette with f loops can be estimated by the entropy of a star polymer with f arms: the main term is then proportional to $f^{3/2}$ [10, 13]. The surface tension of a globule made of f monomers is proportional to $f^{2/3}$. Combining these two terms gives a simple estimate of the free energy of a single rosette:

$$F_{\text{rosette}} \simeq k_B T f^{3/2} + \varepsilon f^{2/3} \tag{1.2}$$

where k_B is the Boltzmann constant, T is the temperature and ε is the interaction energy between two bridgers. A multiple rosette with q domains and p bridgers has rosettes with $f = p/q$ loops each; so the free energy $F = q\,F_{\text{rosette}}$ becomes

$$F(q) \simeq k_B T\, p^{3/2}\, q^{-1/2} + \varepsilon\, p^{2/3}\, q^{1/3} \tag{1.3}$$

The number q of rosettes that minimize the free energy is

$$q_{\text{eq}} \sim p \left(\frac{\varepsilon}{k_B T} \right)^{-6/5}$$

This qualitative argument suggest that can exist a temperature at which a multiple rosette state with $q_{\text{eq}} > 1$ is stable [10]. In section 3.1 we will extend this argument to a certain class of disordered models.

Chapter 2

Bridged heteropolymer

In this chapter we study the polymer with short-ranged bridging interaction originally proposed in [10], that is an ordered polymer in which the interacting monomer are equi-spaced.

In section 2.1 we describe the algorithm we use, pointing out how it is optimized for disordered systems (this will be useful especially in the chapter 3) and how we can argue that the system reaches the thermodynamic equilibrium.

In section 2.2 we describe the formalization of the model, while in sections 2.3 and 2.4 we test the algorithm on simple systems whose phenomenology is already known.

In section 2.5 we present the simulations of a multiple domain state. The main results of this section are that we computationally prove that the multiple domain is a stable phase and that large size systems can have a hierarchical organization.

2.1 Description of the algorithm

Our simulations use an off-lattice Monte Carlo algorithm with Metropolis accept-reject rule optimized for protein folding, called *Montegrappa* [14]. Using an optimized algorithm is necessary to ensure an effective sampling of the canonical ensemble.

Disordered systems (like proteins) show an extremely complex energetic landscape: the presence of several local minima in the free energy causes to a sys-

tem to be trapped in metastable configuration if it is simulated with a simple Monte Carlo algorithm. In this case the algorithm performs a wrong statistical sampling of the canonical ensemble.

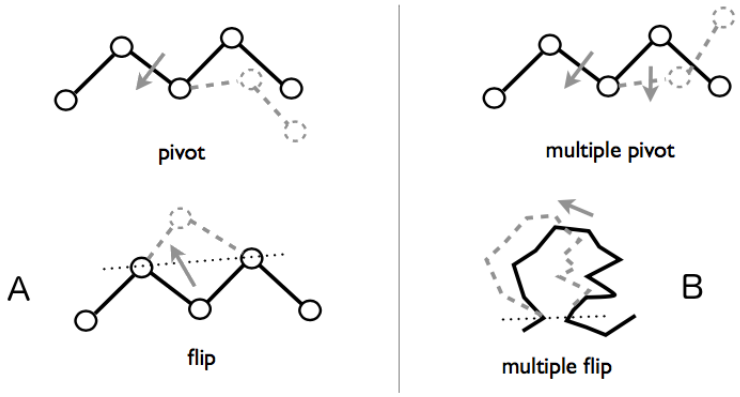


Figure 2.1: Available moves in Montegrappa. A) Local moves. A flip is the rotation of a bead chosen at random around the axis defined from the preceding and the following one. A pivot move changes a dihedral at random. B) Non-local moves. A multiple pivot move is an extension of pivot moves, which changes at random a set of consecutive dihedrals. A multiple flip consists in choosing two non-consecutive atoms of the backbone and moving the backbone atoms in between around the axis defined by the two.

The standard solution to avoid the described situation is to add to the usual set of local moves (see figure 2.1 A) a set of non-local moves, which attempt to modify the configuration of many particles at the same time (see figure 2.1 B). In addition to the standard solution, Montegrappa uses *parallel tempering* [15] (which is similar to the *simulated tempering* [16]): several replicas of the system are simulated at the same time, each one at a different temperature, and a new Monte Carlo move attempts to exchange the configuration of two replicas with similar temperature. This mechanism allows a configuration trapped in a local minimum to escape, thanks to the exchange with a replica at a higher temperature. This statistical ensemble is called *multicanonical ensemble*; with a technique called *multiple histogram* [17] it is possible to combine all the data in an efficient way to obtain thermodynamic averages in a usual canonical ensemble.

2.2 Description of the model

In this thesis we consider a coarse-grained model consisting in a polymer made of N consecutive monomers represented as hard spherical beads of radius R_{HC}. Each bead represents a region of the chromatin fiber of the size of the resolution of the 5C experiments, that is ~ 3 kb.

In our model the bead i can interact with the bead j with an attractive short-ranged square well potential u_{ij} (see figure 2.2 A):

$$u_{ij} = \begin{cases} \infty & \text{if } d_{ij} < R_{\mathrm{HC}} \\ B_{ij} & \text{if } R_{\mathrm{HC}} < d_{ij} < R \\ 0 & \text{if } d_{ij} > R \end{cases}$$

where d_{ij} is the distance between the beads; R_{HC} is the hard-core radius preventing the beads to crossing each other; R is the range of the interaction and B_{ij} is the interaction energy, that depends on the types of the monomers i and j. Using square-well potentials makes the Monte Carlo calculations easier and faster than using smooth short-ranged potential. Note that in this model, if two beads are closer than the hard-core radius R_{HC} their interaction energy is infinite. This leads to such a large increase in energy that any associated conformation is effectively absent from the Boltzmann ensemble of the model polymer. This implements the basic fact that the chromatin fiber cannot cross itself.

Consecutive beads are set at a distance λ at the beginning of the simulation and the Monte Carlo moves do not change these distances. This is a useful trick to avoid the need of an additional potential to model the covalent bonds, which would slow down the simulation.

In order to represent bridging interactions along the polymer, we will use two types of beads: the polymer will be made of p interacting beads placed every k non interacting beads (see figure 2.2 B). We will refer to the interacting beads as *bridgers*. Therefore the interaction energy is

$$B_{ij} = \begin{cases} -\varepsilon & \text{if } i \text{ and } j \text{ are bridgers} \\ 0 & \text{otherwise} \end{cases}$$

where $\varepsilon > 0$ since the interaction between bridgers is always attractive.

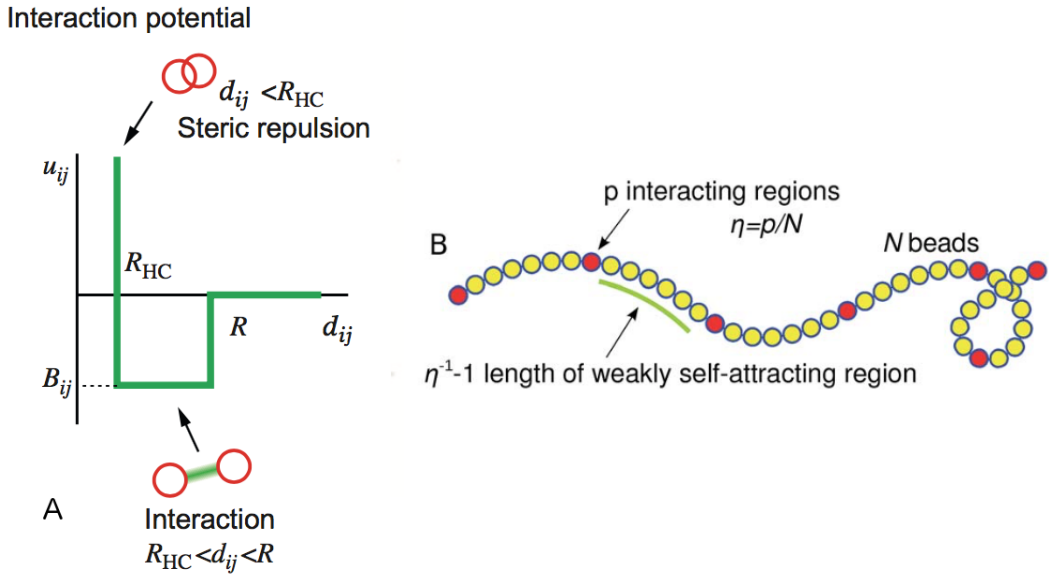


Figure 2.2: Illustration of the model. A) Square-well potential. The inter-particle potential comprises a hard-core repulsion between all discretization "beads" of the polymer (whose range is set by the parameter s), and a short-ranged attraction potential of range a and depth 3u for the homogeneous self-attraction (acting on all beads) and 3l for the sparse bridging interactions. B) Distribution of bridgers along the chain. A total of p bridging beads are placed across p equally spaced regions of length N/p.

The total energy of the system is the sum over all the two-body interactions

$$U = \sum_{i<j} u_{ij}.$$

The total energy defines the probabilities for the various configurations to occur in the equilibrium ensemble, since the Metropolis accept-reject rule reads

$$P(U_a \to U_b) = \min\left[1, e^{-\beta(U_a - U_b)}\right],$$

leading to the Boltzmann equilibrium distribution

$$P(U_a) = Z^{-1} e^{-\beta U_a}$$

where $\beta = 1/k_B T$ and Z is the partition function $Z = \sum_a P(U_a)$.

2.3 Testing the code: collapse of the homopolymer

As a simple test, in this section we study a polymer in which all monomers interact with each other with a short ranged square-well potential. This system is called *homopolymer* and we expect to find the conventional theta transition: at high temperature the polymer is in the random coil phase while al low temperature it is in the random globule phase (see figure 2.3).

To show this transition we study the *swelling parameter* α, that is the ratio between the radius of gyration of the polymer R_g at a given temperature and $R_0 = \lambda(3/5N)^{1/2}$, that is its value at the theta point:

$$\alpha = \frac{R_g}{R_0}.$$

The random globule phase is characterized by having $\alpha < 1$, while the random globule phase has $\alpha > 1$. The temperature at which $\alpha = 1$ defines the critical point in the Flory mean-field theory (cite) and it is called *theta point*. In this theory the theta point is defined by balancing the excluded volume with the attraction term in the second virial coefficient.

We simulated homopolymers of various length N, verifying the qualitative predictions that the theta point shifts towards higher temperature for increasing N; we also verified that the the radius of gyration in the coil phase is higher for higher N (see figure 2.4).

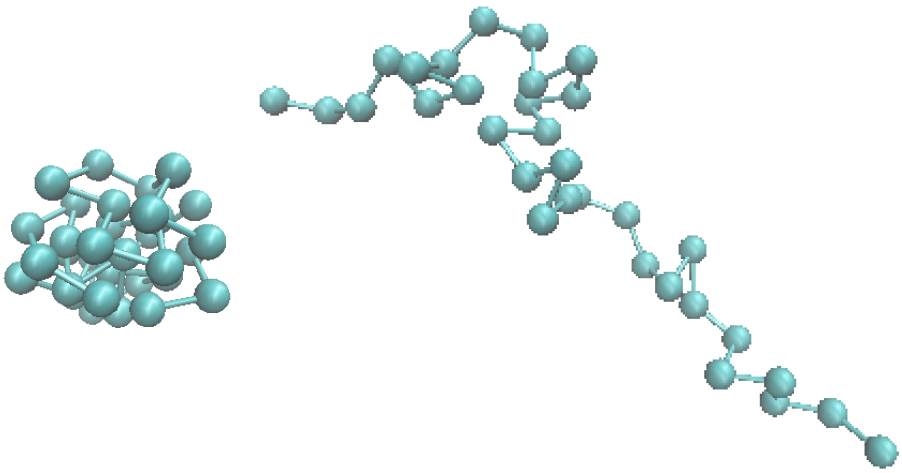


Figure 2.3: Snapshots of the two phases of the homopolymer. The random globule (on the left) has a gyration radius scaling as $R_g \sim N^{1/3}$ and constant density. The random coil (on the right) has a gyration radius scaling as $R_g \sim N^{3/5}$ and low density.

We set the initial state of all the replicas to a swollen configuration in which the beads are placed at the correct bond distance; then we perform enough Monte Carlo sweeps to reach the equilibrium state in all the replicas at different temperature. To argue that the results come from a proper equilibration of the system we use two rules. The first one is typical of the Monte Carlo simulations: we divide the trajectory in three equal sections, discard the first one and check the observables on the other two (see figure 2.5). If the thermodynamic averages of the two are compatible then we assume that the system has reached equilibrium.

Since we use parallel tempering, we need a way to check that none of replicas is stuck in a metastable configuration. An attempted exchange between two consecutive replicas is accepted every time the two replicas undergo energy fluctuations that overlap with each other; this means that if all replicas have overlapping energy fluctuations they can efficiently diffuse in the configuration space and overcome local metastable minima in the free energy. The way to ensure this fact is checking the trajectories of internal energy and adjusting the set of temperatures so that each trajectory overlaps with the adjacent ones (for example, in figure 2.5 the two lowest replicas does not overlap significantly, and so they are not reliable). This will be our second rule to argue the

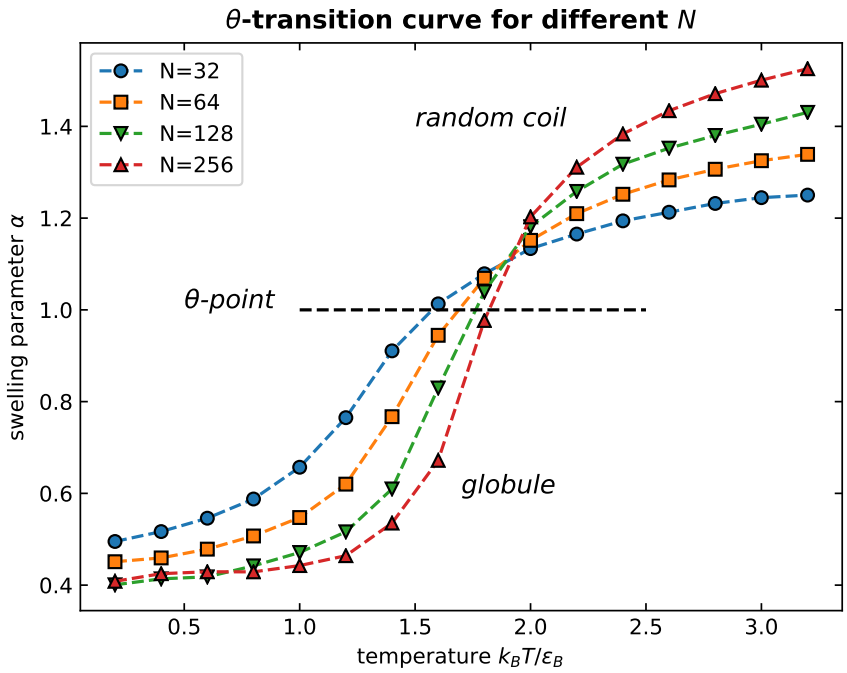


Figure 2.4: Collapse curve of the swelling parameter α as a function of temperature T, plotted for different lengths N of the homopolymer. The temperature axis is rescaled with the energy ε of the homogeneus attractive square-well potential. Simulations were performed with $N = 257$, $\varepsilon = 1$, $R = 1.44$, for 5×10^7 Monte Carlo sweeps.

proper equilibration of the system. Since a limited amount of computer cores are available, it is convenient to choose an uneven distribution of temperatures, closer to each other at low temperatures, where energy fluctuations are small, and more distant from each other at high temperatures, where thermal fluctuations favour exchanges.

2.4 Collapse into a single domain

In this section we study a polymer in which only p bridging beads interact, while the other are non-interacting beads. We set the parameters ε and R so that the polymer collapsed into a single domain. We simulated a polymer of

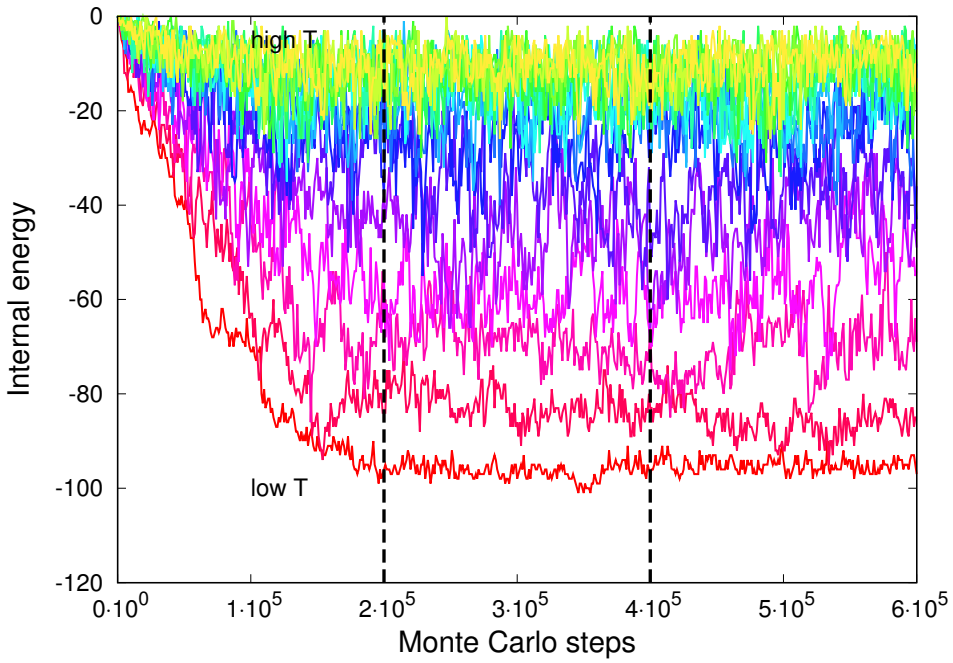


Figure 2.5: Internal energy of the replicas at different temperatures. The lower temperature replicas (in red) need more time to reach the equilibrium state than the higher temperature replicas (in yellow). The two lowest replicas does not overlap significantly, so they are not reliable. Simulations were performed with $N = 257$, $\varepsilon = 1$, $R = 1.44$, for 5×10^7 Monte Carlo sweeps.

$N = 257$ beads.

In particular in this section we focus on varying the fraction $\eta = P/N$ of interacting beads, following the observations in [10]. For $\eta \sim 1$ we expect again a theta point collapse from a random coil to a random globule, that is a continuous transition. On the other hand, when $\eta \ll 1$, the system undergoes a switch-like transition between the rosette state and the coil state, displaying a more abrupt jump in the collapse curve (figure 2.6). The important observation is that there is a crossover between this likely first-order and the Flory-like homopolymer collapse behaviour when η sufficiently large; the smoothing of the collapse curves due to high η will be relevant in the following section to interpret the role of disorder.

To argue that this could be a proper first order phase transition we compare the results for very few bridgers with the theoretical two-state prediction de-

veloped in [10] (figure 2.7). The comparison is not satisfactory even in the case $p = 3$, while in [10] the agreement was good up to $p = 4$. The reason of this could be a different choice of the parameters, since we were not able to explore all the possibilities.

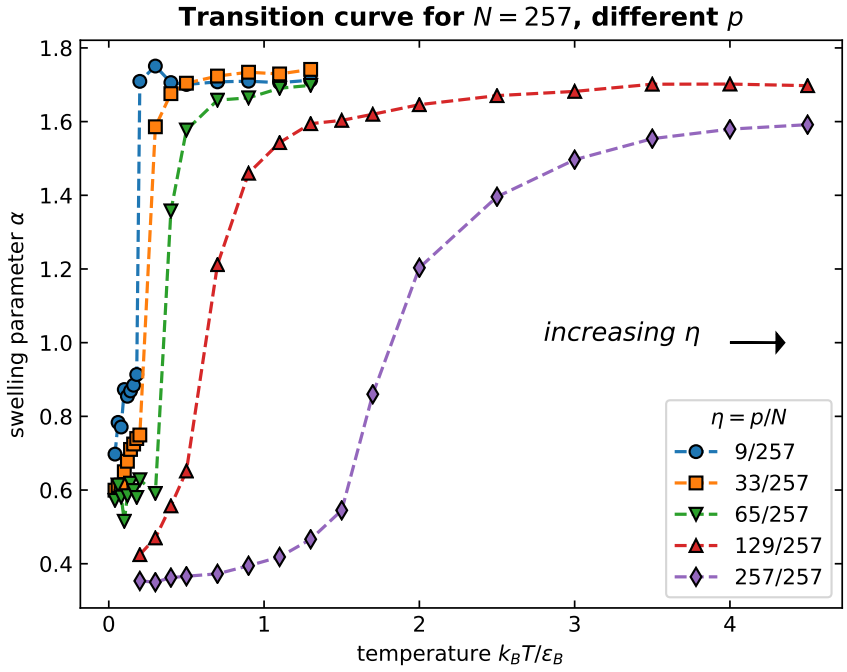


Figure 2.6: Collapse curve of the swelling parameter α as a function of temperature T, plotted for different concentration of bridging monomers $\eta = p/N$. The case $\eta = 9/257$ is a switch-like transition, instead $\eta = 257/257$ corresponds to a uniform collapse. Simulation were performed with $\varepsilon = 1$, $R = 1.44$, for 1×10^8 Monte Carlo sweeps.

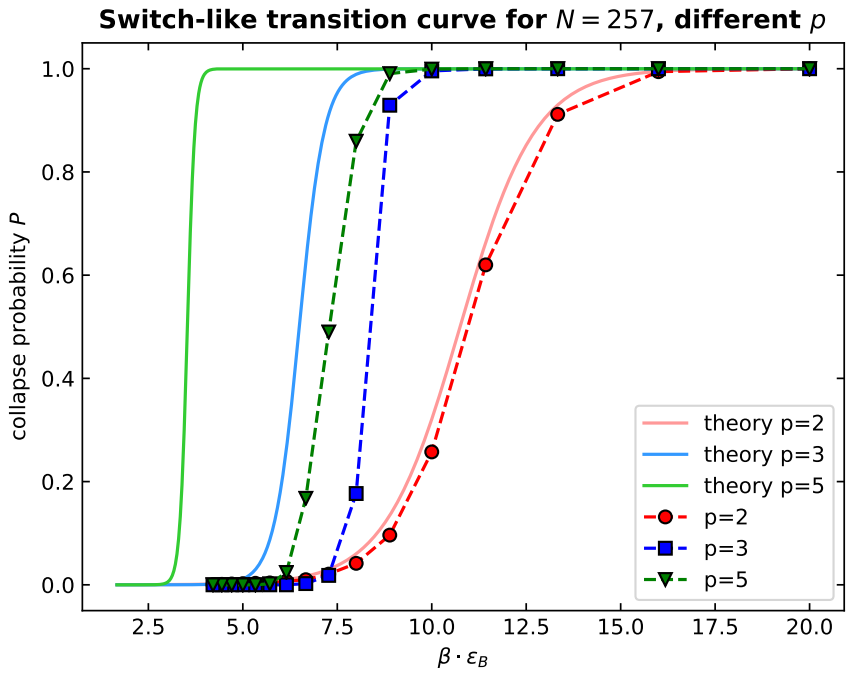


Figure 2.7: Collapse probability to a rosette-like state as a function of the temperature T, plotted for different number p of bridging monomers. We used small p to be within the two-state approximation; the collapsed state is charaterized by the maximum number of formed bonds between bridgers. Simulation were performed with up to $N = 256$ monomers, $\varepsilon = 1$, $R = 1.44$, for 3×10^7 Monte Carlo sweeps.

2.5 Collapse into multiple domains

The main goal of this section is to reproduce the collapse of the polymer with bridging interactions into a multiple rosette state. We show that this is a stable phase of the model using parallel tempering, as described in the subsection 2.3.

We simulated polymers from $N = 129$ up to $N = 1025$ beads, keeping constant the density of bridgers $\eta = p/N$. We observed that the polymer with $N = 129$ monomers collapses into a single domain, while the polymers ($N = 256$, $N = 513$ and $N = 1025$) collapse into a multiple domain state. The number of domains seem to depend linearly on p, as suggested in [10]. The collapse *for all*

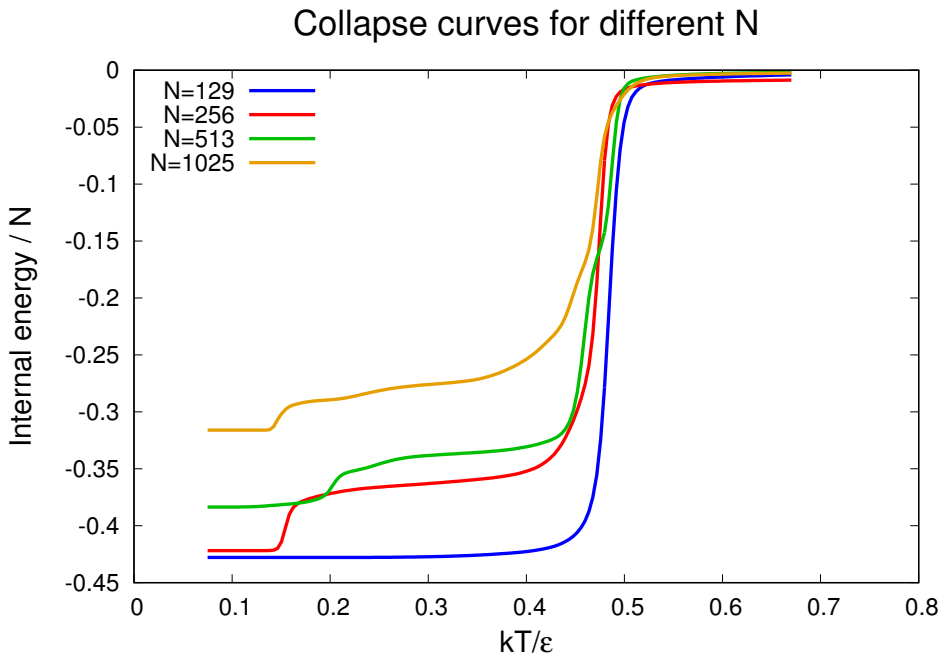


Figure 2.8: Collapse curves of internal energy for chains of different lenght N. The vertical axis has been rescaled dividing by N for each curve. In all these simulation the density of bridgers $\eta = p/N$ was kept constant to the value 1/16. The results for $N = 1025$ are preliminary since the thermalization is reliable for the collapse at $T \sim 0.47$ but not for the lower temperature states. Simulation were performed with up to $N = 1025$ monomers, $\varepsilon = 2.4$, $R = 0.77$, for 3×10^9 Monte Carlo sweeps.

values on N happens near a temperature of $T \simeq 0.47$.

Intriguingly we observe that the polymer with $N = 256$ monomers, after the first collapse at $T \simeq 0.47$, displays a second collapse at $T \simeq 0.15$ from the two domain phase to a one domain phase. While the first energy jump displays features similar to a first-order phase transition, as suggested in [10], the fusion of two domains resembles a nucleation-like phenomenon, so we guess that this could be similar to a second order phase transition.

A similar behaviour is observed for the polymer with $N = 513$ and $N = 1025$ monomers: they display a second collapse from the many domains state to a state with few less domains. We plotted the collapse curves of internal energy for all the different values of N, rescaling the internal energies with N so that

we could compare them (figure 2.8). We found that, except for $N = 1025$, each curve overlaps with the previous one in correspondence with its second collapse.

The low temperature phases are difficult to sample and for this reason the internal energy reconstruction is less reliable in the lower temperature, so a proper discussion on these observation requires much longer and computational-heavier simulation. Nevertheless we suggest that the real collapse curve of the polymer with $N = 513$ should display all the nucleation-like transitions observed in the underlying curves. We expected the polymer with $N = 1025$ to follow this behaviour; for this reason we argue that, given the large size of the system, this simulation is most likely far from equilibrium.

Summing up, we suggest that while increasing the size of the system new phases become available, and that the system can cross several hierarchical levels of organization before collapsing into a single domain. This could be relevant in the interpretation on the sub-TADs, described in the introduction.

Chapter 3

Disordered bridged heteropolymer

In this chapter we study the effect of disorder on the multiple domain state of the heteropolymer with bridging interaction.

In order to give a quantitative definition of the disorder in the model we consider a perturbation of the regular distribution of the bridgers described in [10]: each bridger is displaced from its position by a random number of beads extracted with a Gaussian distribution of variance σ; the variance is chosen to be the same for each monomer. We will consider σ as the parameter that describes the degree of disorder in the model. If $\sigma = 0$ we recover the case studied in Chapter 2.

The chapter is divided into two parts. In the first one (section 3.1) we argue with a theoretical approach that the multi domain phase can exist also in the disordered model. Following [10] we approximate the rosette-like state to a simple star polymer, so that we can extend the results [18, 13] to estimate the entropy of a disordered star polymer. These results are based on the blob description of the star polymer, that consist in a series of concentric shells made by f blob each, where f is the number of arms (see figure 3.1). We were not able to use the Gaussian distribution described in the previous paragraph, so we made an ansatz and developed the calculations for a certain class of power law distributions. Since the power laws represent a higher degree of disorder than the Gaussian distributions, we can still qualitatively argue that if the disordered multiple domain state is possible in the case of power law disorder, it is also possible for the Gaussian disorder.

In the second part of this chapter (sections 3.2 to 3.4) we describe the simulation we did, showing that not only the multiple domain state appears in the disordered model, but also that the Gaussian disorder has the surprising effect of stabilizing the multiple domain phase.

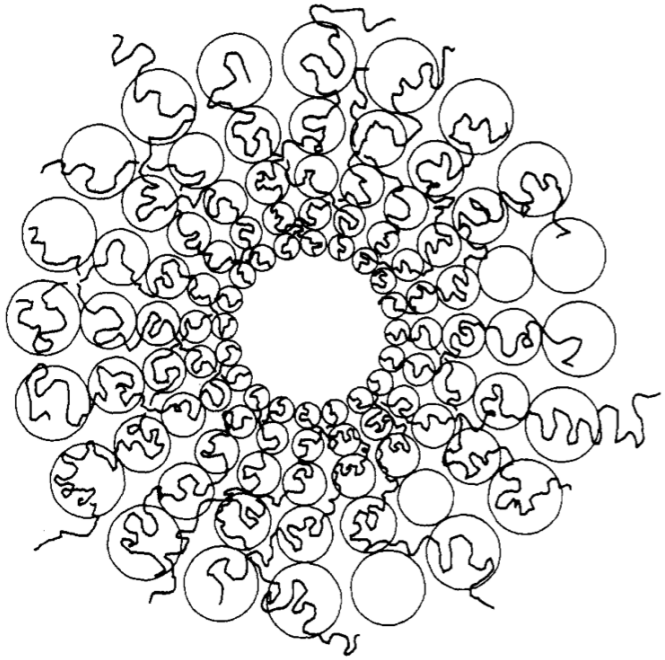


Figure 3.1: A star polymer is made of f arms attached to a rigid core (not drawn in the picture.) The blob description of the star polymer consist in a series of concentric shells made by f blob each.

3.1 Entropy of a disordered star polymer

As said before, we approximate a single rosette-like state with f loops to a star polymers made of f polymeric strands called *arms*, which are bound to a rigid spherical core that does not contribute to the entropy.

We follow [10, 13] to give an estimate of the entropy of a star-shaped polymer with arms of different length. The calculations are based on the blob model (see for example [11]) applied to star-shaped polymers [18].

The goal of this section is to calculate the total number of blobs N_{blob} in the star polymer, since the free energy F is proportional to that number:

$$\beta F = N_{\text{blob}} \tag{3.1}$$

where $\beta = 1/k_b T$. Given that we consider only excluded volume interactions between arms, equation 3.1 describes a purely entropic contribution to the free energy. This term will replace the first term in equation 1.2: if the entropy of a disordered star polymer is substantially different from the one of the regular star, then we expect that the multiple rosette state may not be stable any more. Otherwise, if the entropy disordered star polymer is only a perturbation of the regular case, we can expect the multiple rosette state to still be an equilibrium state of the model.

3.1.1 Hypotheses

In the case of a regular star we picture the semidilute region as made of concentric spherical shells of thickness $\xi(r)$, each one crossed by f arms (see figure 3.2). This means that each shell is made of f blobs of dimension $\xi(r)$ [18].

From these assumptions we can write an expression for ξ: the volume of one shell, divided by f, must be the volume of a blob; so we can write

$$\frac{r^2 \xi}{f} = \xi^3$$

and then the expression for the blob size ξ:

$$\xi(r) = f^{-1/2} r \tag{3.2}$$

To account for the different length of the arms, we consider the number of arms f as a decreasing function of the radius: $f(r)$. We give an ansatz for this function:

$$f(r) = f_0 \left(\frac{r}{b} \right)^{-\gamma} \tag{3.3}$$

where b is the radius of the core of the star and $\gamma \geq 0$.
Combining 3.2 and 3.3 we have the equation

$$\xi(r) = f_0^{-1/2} b \left(\frac{r}{b} \right)^{1+\frac{\gamma}{2}} \tag{3.4}$$

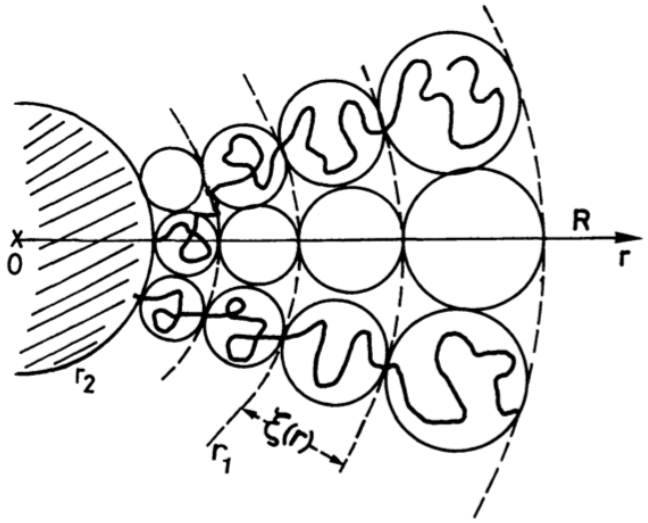


Figure 3.2: Representation of the blob model. The core of the star has radius r_2, while the radius of the corona is R; the dashed lines represent the radii of the shells. In order to account for the radial distribution of the monomers, the dimension of each blob ξ increases from the centre to the outside.

3.1.2 Positions of the shells

We are interested in the radius of each shell, because we can relate it to the monomer concentration in the way explained below. Since the tickness of a shell is $\xi(r)$, we can relate the radii of two consecutive shells with the following rule:

$$r_{m+1} - r_m = \xi(r_m)$$

that can be approximated to a simple differential equation

$$\frac{dr(m)}{dm} = \xi(r(m))$$

whose solution, using the expression 3.4, is

$$r(m) = b\left[1 - \frac{\alpha}{2} f_0^{-1/2} m\right]^{-\frac{2}{\gamma}} \tag{3.5}$$

for $\gamma \neq 0$.
If $\gamma = 0$, that is the case of the regular star, the solution is exponential:

$$r(m) = b\, e^{f_0^{-1/2} m} \tag{3.6}$$

3.1.3 Monomer concentration

Since a blob behaves by definition as a swollen polymer, his dimension is related to the number of monomers g it contains by $\xi = ag^{\nu}$, where a is the distance between two consecutive monomers. We set $\nu = 3/5$. Therefore we invert this relation and we use 3.4 to obtain $g(r)$:

$$g(r) = \left(\frac{\xi(r)}{a}\right)^{5/3} = \left(f_0^{-1/2}\frac{b}{a}\right)^{5/3}\left[\frac{r}{b}\right]^{\frac{5}{3}(1+\frac{\gamma}{2})} \tag{3.7}$$

Given the number of monomers in one blob $g(r)$, we can calculate the radial monomer concentration $c(r)$ from the total number of monomers in one shell fg divided by the volume of the shell:

$$c(r) = \frac{f(r)g(r)}{4\pi r^2 \xi(r)}$$

Using 3.3, 3.4 and 3.7 it becomes

$$c(r) = \frac{1}{4\pi r^2}\frac{1}{b}f_0^{2/3}\left(\frac{b}{a}\right)^{5/3}\left[\frac{r}{b}\right]^{\frac{2}{3}(1-\gamma)} \tag{3.8}$$

3.1.4 Radius of the star polymer

Since we know the total number of monomers N_{tot} in the star polymer, we can integrate the concentration to obtain the expression of the radius R of the star polymer:

$$\begin{aligned} N_{tot} &= \int_b^R 4\pi r^2 c(r) dr \\ &= f_0^{2/3}\left(\frac{b}{a}\right)^{5/3}\int_b^R \left[\frac{r}{b}\right]^{\frac{2}{3}(1-\gamma)}\frac{dr}{b} = \\ &= f_0^{2/3}\left(\frac{b}{a}\right)^{5/3}\frac{3}{5-2\gamma}\left[\left(\frac{R}{b}\right)^{\frac{5-2\gamma}{3}} - 1\right] \end{aligned} \tag{3.9}$$

Now we observe that we must exclude $\gamma > 5/2$, because it corresponds to an inverse relation $R(N_{tot})$ that diverges for a certain finite value of N. So we set $\gamma < 5/2$ and then we invert the relation 3.9 to obtain $R(N_{tot})$:

$$R(N_{tot}) = b\left(1 + \frac{5-2\gamma}{3}\left(\frac{a}{b}\right)^{5/3} f_0^{-2/3} N_{tot}\right)^{\frac{3}{5-2\gamma}}$$

We neglect the term $+1$ in the parentheses:

$$R(N_{tot}) \simeq b\left(\frac{5-2\gamma}{3}\left(\frac{a}{b}\right)^{5/3} f_0^{-2/3} N_{tot}\right)^{\frac{3}{5-2\gamma}} \tag{3.10}$$

3.1.5 Number of shells

We evaluate $r(S)$ using 3.5, where S is the total number of shells in the star polymer; then we compare it to the expression 3.10 for R :

$$r(S) \equiv R$$

Now we can finally obtain an expression for S

$$b\left[1-\frac{\gamma}{2} f_0^{-1/2} S\right]^{-\frac{2}{\gamma}} = \left(\frac{5-2\gamma}{3}\left(\frac{a}{b}\right)^{5/3} f_0^{-2/3} N_{tot}\right)^{\frac{3}{5-2\gamma}}$$

This leads to

$$S = \frac{2}{\gamma} f_0^{1/2}\left[1-\left(\frac{5-2\gamma}{3}\left(\frac{a}{b}\right)^{5/3} f_0^{-2/3} N_{tot}\right)^{-\frac{\gamma}{2}\frac{3}{5-2\gamma}}\right] \tag{3.11}$$

We observe that this expression reduces to the one for the regular star [13] in the limit $\gamma \to 0$

$$S \sim f_0^{1/2} \log\left[\left(\frac{5}{3}\left(\frac{a}{b}\right)^{5/3} f_0^{-2/3} N_{tot}\right)^{-\frac{3}{5}}\right] \tag{3.12}$$

where we used the limit formula

$$\lim_{\gamma \to 0} \frac{x^\gamma - 1}{\gamma} = \log x$$

3.1.6 Number of blobs

In order to finally obtain the expression for the number of blobs in the star, we need to calculate the series

$$N_{blob} = \sum_{m=1}^{S} f(r_m). \tag{3.13}$$

For $f(r_m)$ we combine the expression of the scaling of f

$$f(r) = f_0 \left(\frac{r}{b}\right)^{-\gamma}$$

and the expression of the radius of the m-th shell

$$r(m) = b\left[1 - \frac{\gamma}{2} f_0^{-1/2} m\right]^{-\frac{2}{\gamma}}.$$

We obtain

$$f(r(m)) = f_0 \left[1 - \frac{\gamma}{2} f_0^{-1/2} m\right]^2.$$

Plugging this last equation into equation 3.13 we find:

$$N_{\text{blob}} = \sum_{m=1}^{S} f_0 \left[1 - \frac{\gamma}{2} f_0^{-1/2} m\right]^2$$

$$= \sum_{m=1}^{S} f_0 \left[1 - \gamma f_0^{-1/2} m + \frac{\gamma^2}{4} f_0^{-1} m^2\right].$$

In the end we approximate the series with an integral, so that we obtain:

$$N_{\text{blob}} \simeq f_0 \left[S - \gamma f_0^{-1/2} \frac{S^2}{2} + \frac{\gamma^2}{4} f_0^{-1} \frac{S^3}{3}\right] \tag{3.14}$$

Again we observe that in the limit $\gamma \to 0$ this expression reduces to the one for the regular star

$$N_{\text{blob}} \simeq (f_0)^{3/2},$$

since we drop the logarithmic term in equation 3.12. This expression is the one used to estimate the entropy of a rosette-like state in [10].

It is interesting to note that, at least for small values of γ, the entropy of a rosette-like state as a function of the number of arms f is not affected by the disorder in the length of the arms. This fact suggest that is possible for the multiple domain state to survive as a stable phase even in the model with disordered distribution of bridgers.

3.1.7 Discussion on the ansatz

In order to justify the ansatz 3.3 on the scaling of the number of blob per shell, in this section we show that $f(r)$ can be linked to the distribution $p(N)$ of the number of monomers in a single arm. In particular we will find that, given a certain value of the exponent γ, there is a power law distribution $p(N)$ that produces $f(r) = f_0(r/b)^{-\gamma}$.

Let's interpret $f(r)$ as the fraction of arms that reach *at least* a distance r from the core; namely $f(r)$ is the probability that one arm arrives at least at a distance r. This probability is nothing but the cumulative distribution of the probability $p(r)$ that one arm reaches the distance r:

$$f(r) = 1 - \int_0^r p(r')dr'.$$

We can obtain the probability $p(r)$ deriving one time the cumulative distribution:

$$p(r) = -\frac{\partial}{\partial r} f(r). \tag{3.15}$$

Since we are interested in the probability in function of the number of monomers N we need to change the variable with the relation

$$p(N)dN = p(r)dr,$$

that gives

$$p(N) = p(r) \left(\frac{dN}{dr} \right)^{-1} \Bigg|_{r(N)}. \tag{3.16}$$

To find the relation $N(r)$ we integrate the number $g(r)$ of monomers in one blob:

$$N(r) = \int_b^r g(r')dr',$$

where $g(r) = (f(r)^{-1/2} r)^{1/\nu}$. Furthermore we observe that

$$\frac{dN}{dr} = g(r).$$

Thanks to these relations, the expression 3.16 becomes

$$p(N) = f'(r)(f(r)^{-1/2} r)^{1/\nu} \tag{3.17}$$

Plugging into the equation 3.17 the ansatz 3.3, we find the expression of the number of monomers contained in a sphere of radius r

$$N(r) \sim r^{\frac{1}{\nu}(1+\frac{\gamma}{2})+1}$$

and the expression of the probability that one arm has N monomers

$$p(N) \sim r(N)^{-1-\gamma-\frac{1}{\nu}(1+\frac{\gamma}{2})}. \tag{3.18}$$

Using the equation 3.17 into the expression 3.18 we finally obtain the power law distribution $p(N)$ that corresponds to the ansatz $f(r) = f_0(r/b)^{-\gamma}$:

$$p(N) \sim N^{-1-\frac{\gamma}{\frac{1}{\nu}\left(1+\frac{\gamma}{2}\right)+1}}. \tag{3.19}$$

3.2 Collapse into multiple domains

This section describes the results of the simulations of the polymeric model with disordered distributions of the bridging interactions along the chain. As stated above, we consider perturbations of the regular distribution studied in the previous chapter. In particular we consider a displacement of each interacting bead of a random number of positions, extracted with a Gaussian distribution of variance σ.

This particular choice of the distribution is made to use σ as a parameter of the phase diagram: we want to identify the range of temperatures in which the multiple domain phase is stable as a function of the "amount of disorder", defined as σ.

The simulations are performed at fixed density of bridgers $\eta = p/N$. The unperturbed distance between bridgers η^{-1} is the parameter that determines the actual distribution as a function of σ: for $\sigma < \eta^{-1}$ the displacements are mostly localized around the regular positions, so we expect a small perturbation of the behaviour described for the ordered system. On the contrary, when $\sigma > \eta^{-1}$ the distribution of the bridgers tends to the uniform distribution, that is the highest degree of disorder possible in this model.

The result of our simulations is that the multiple domain phase appears to be stable up to $\sigma = 2\eta^{-1}$ (see figure 3.3). This means that the polymer collapses into a multi-rosette state both for small perturbations and almost uniform distribution. This result agrees with the qualitative prediction proposed in section 3.1: since the disorder does not change the leading term in the entropy of the rosettes, we expect the collapse to be at least qualitatively the same. Indeed we find that the collapsed phase for all values of σ exhibits two domains, just as the model without disorder.

Interestingly, the disorder has the unexpected effect of stabilizing the multiple domain phase: the range of temperatures in which the multiple domain phase is stable broadens proportionally to σ. In order to have a clear information on the critical temperatures, we plotted in figure 3.4 the specific heats relative to the transitions shown in figure 3.3. As we can see in figure 3.4 there are two peaks in the specific heat, one for the collapse from coil to two-domain state (high temperature) and one for the transition from two-domain state to one-domain state (low temperature). We observed that these peaks shift respectively towards higher temperatures and towards lower temperatures.

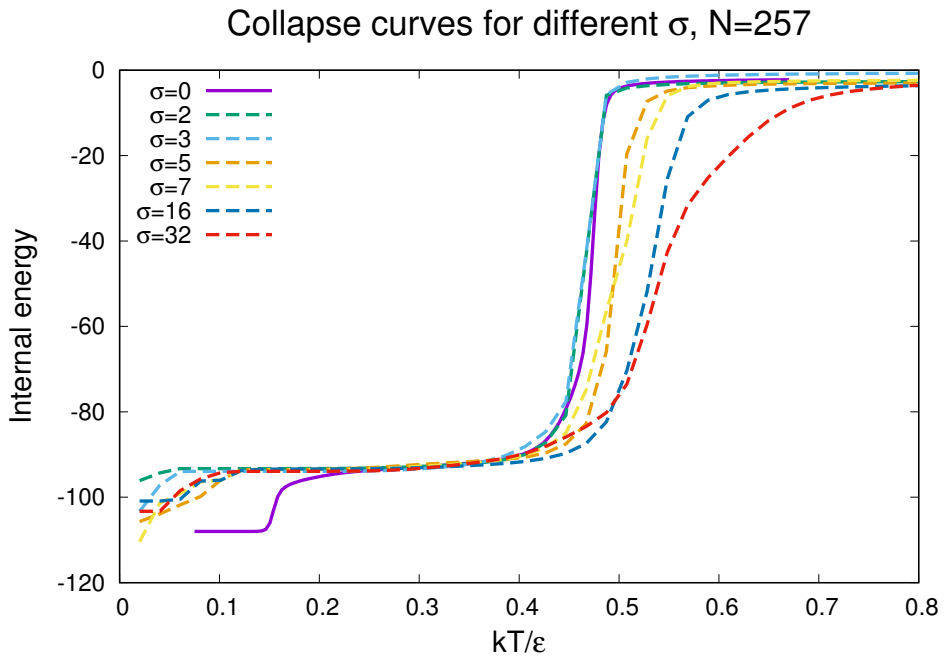


Figure 3.3: Collapse curves of internal energy of a polymer of length $N = 257$ monomers. Each dashed curve relates to a different value of the variance σ; i used value of σ up twice the value of the distance between monomers in the deterministic distribution (solid line). Since the internal energy has proven to be a self-averaging quantity, only one simulation for each value of σ was performed (see section 3.2.1). These simulations were performed with $N = 257$, $\eta = 17/257$, $R = 0.77$, $1.8 \cdot 10^9$ Monte Carlo sweeps.

Additionally we point out that the disorder smoothens the collapse curve of the higher transition (see figure 3.3); this behaviour is reflected in the specific heat, since the height of the peak decreases with σ, while it becomes wider. This could suggest that the transition coil to multiple domain state is no longer switch-like, but rather continuous.

This behaviour resembles the effect of increasing the density $\eta = p/N$ in the ordered model described in section 2.4: the model with high density of interacting beads undergoes a Flory-like transition. While this is trivial for $\eta \simeq 1$ (in this case the model is a homopolymer), we observe that even for $\eta \simeq 0.5$ or less the loops between monomers are most likely too short to produce an entropic cost sufficient to make the transition switch-like. We suggest that the effect of the disorder can be imputed to the formation of high density zones

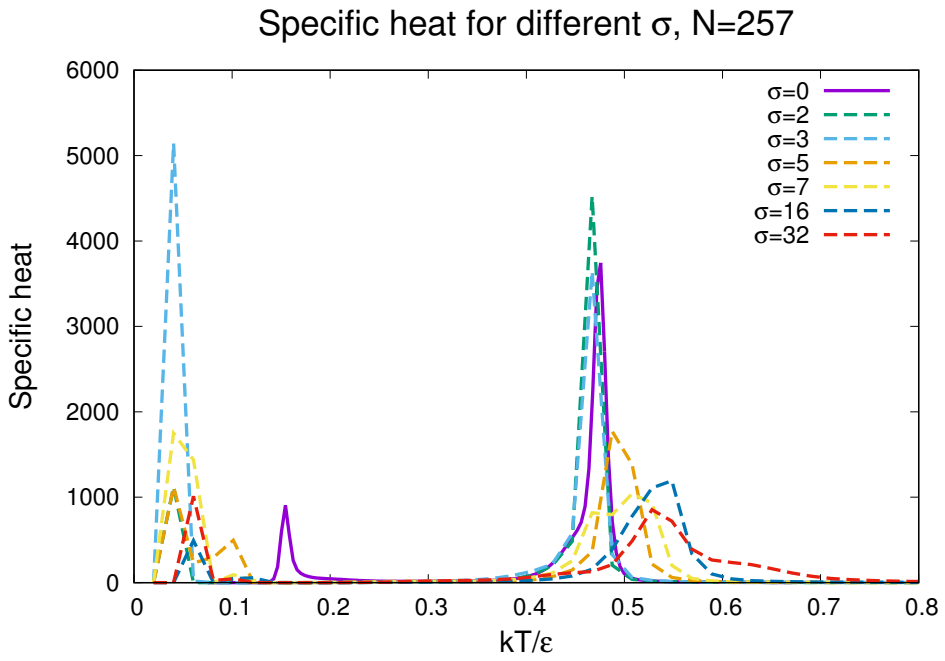


Figure 3.4: The collapse of the disordered bridged heteropolymer displays two peaks in correspondence of the transition from coil to the two-domain state and of the transition from the two-domain state to the single domain one. For increasing σ the first peak shifts towards higher temperatures, while the second one shifts towards lower temperatures. These simulations were performed with $N = 257$, $\eta = 17/257$, $R = 0.77$, $1.8 \cdot 10^9$ Monte Carlo sweeps.

that could drastically reduce the entropic cost of forming loops, causing the collapse to be a second order transition.

3.2.1 Self averaging behaviour

Since the distribution of interactions in the model is stochastic, each realization of the system can in principle give rise to different equilibrium states. This means that the value of a certain observable at equilibrium depends in general on the index of the replica r. If the value the observable does not depend on r, the observable is considered *self-averaging* and it is representative of all the different replicas.

In a Monte Carlo simulation the physical observables are estimated by the

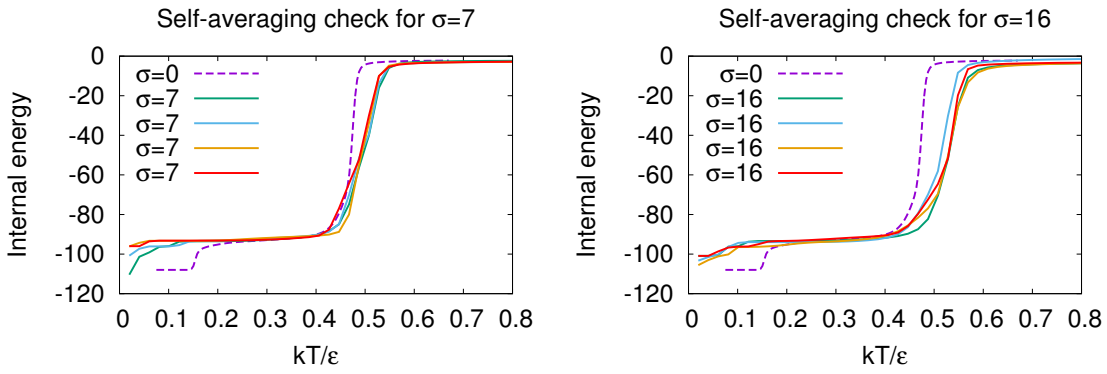


Figure 3.5: Four simulations were performed for each of two values of the variance $\sigma = 7, 16$, and different realization of the disorder were used in each run (each realization is coded with a different color). The curves of the different realizations are in good agreement with each other for both values of σ, so we conclude that the internal energy is self-averaging. These simulations were performed with $N = 257$, $\eta = 17/257$, $R = 0.77$, $1.8 \cdot 10^9$ Monte Carlo sweeps.

statistical sampling; for the sake of concreteness here we consider the simplest estimator, that is the arithmetic average of a certain observable O:

$$[O] := \frac{1}{m} \sum_m O\left(C(m)\right) \tag{3.20}$$

where $O(C(m))$ is the observable calculated on the configuration C at the Monte Carlo time-step m.

In general $[O]$ depends on r:

$$[O]_r := \frac{1}{m} \sum_m O\left(C_r(m)\right)$$

where $C_r(m)$ are the configurations of the replica r as a function of the Monte Carlo time-step m.

In order to obtain actual informations on the mean behaviour of the replicas we consider the quantity

$$\langle O \rangle := \frac{1}{r} \sum_r [O]_r$$

that is the arithmetic average on different replicas of the estimated observables. Now we can ask how much the value of the observable $[O]_r$ on a certain replica is different from the mean behaviour $\langle O \rangle$. If the observable measured on a single replica is comparable with the mean value of all replicas, namely if $[O]_r \simeq \langle O \rangle$, we consider the observable O to be self-averaging. If the observable O is self-averaging, its value $[O]_r$ does not actually depend on the replica index r. Importantly, in this case we can one realization of the stochastic distribution of interactions as representative of the actual distribution.

As noted in section 2.1, in this thesis we use a more complicated technique to compute thermodynamic averages than the simple arithmetic average 3.20; nevertheless we must discuss the self-averaging properties of the results.

We simulated four different realizations of the distribution of interactions described in section 3.2, for two different distributions (that is: two values of σ). The results are shown in figure 3.5: we show that the internal energy is self-averaging.

3.3 Contact matrices

In this section we present the average contact matrices relative to the simulations described in section 2.5 (regular distribution of bridgers) and section 3.2 (disordered distribution of bridgers) We present simulations performed with the same number of monomers $N = 256$, at fixed number of interacting monomers $p = 16$.

The contact matrices we present are made by averaging on the Monte Carlo time the instantaneous contact map of different configurations. This means that the element (i, j) of a contact matrix contains the probability that the distance d_{ij} between the monomer i and the monomer j is less than a threshold d^*; we set the threshold to be the same value as the range of interaction R: $d^* = R$.

A proper discussion on the contact matrices of this systems would require an order parameter for the transition from coil to the multiple domain phase; the order parameter should be an observable able to count the number of domains in the contact matrices and possibly to identify their borders. Such observable is absent in the current state of the literature; for this reason we will limit the discussion in this section only to qualitative observations.

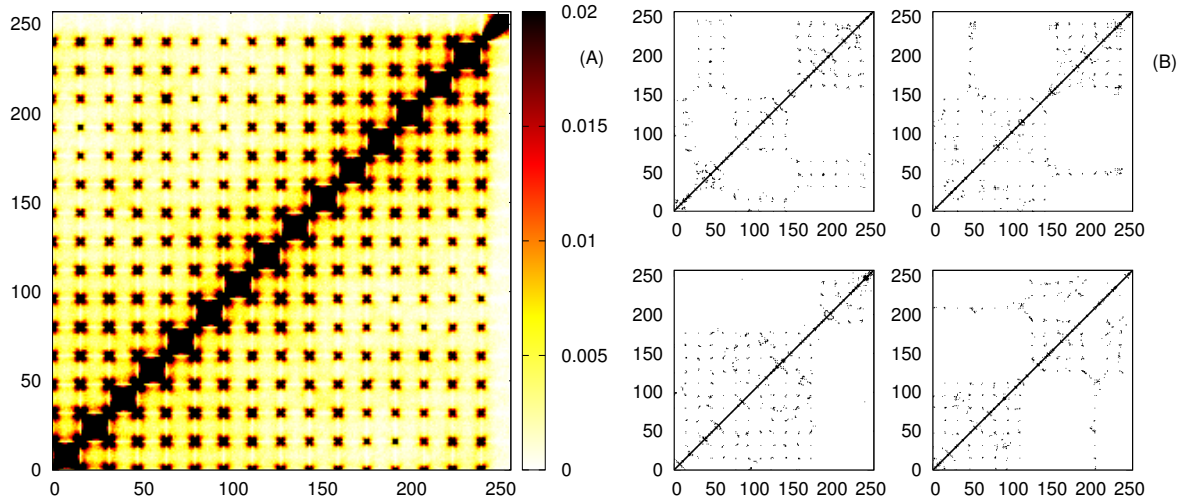


Figure 3.6: $\sigma = 0$. Contact matrix for a polimer made of $N = 257$ monomers and $p = 16$ bridgers distant 16 monomers from each other. The system is at the temperature $kT/\varepsilon = 0.3$, at which two domains are visible. These simulations were performed with $N = 257$, $\eta = 17/257$, $R = 0.77$, $1.8 \cdot 10^9$ Monte Carlo sweeps.

At first we consider the average contact matrix for the regular distribution (figure 3.6 A). Though a slight structure can be intuited, there are no clear borders nor blocks. In order to interpret the behaviour of the system we need to look at the instantaneous contact map of different configurations (figure 3.6 B): we can argue that there are actually configurations with two blocks, but the borders of these blocks are very changing. Different blocks in the contact matrices correspond to different allocations of the bridgers in the two rosettes of the polymer (see for example figure 3.7). The configurations appear to change sufficiently frequently to not produce a stable average configuration in which two block are actually visible.

Then we consider a single realization of the disordered distribution of bridgers with $\sigma = 3$. It is clear that the average contact matrix shows a two-blocks structure with well-defined borders (figure 3.8 A); it is also visible a sub-block in the central region (from monomer 50 to 150), interpretable as a region of few bridgers that are frequently exchanged between the two rosettes.

We suggest that the difference between these two behaviours could be interpreted as a localization phenomenon induced by disorder. The system with a

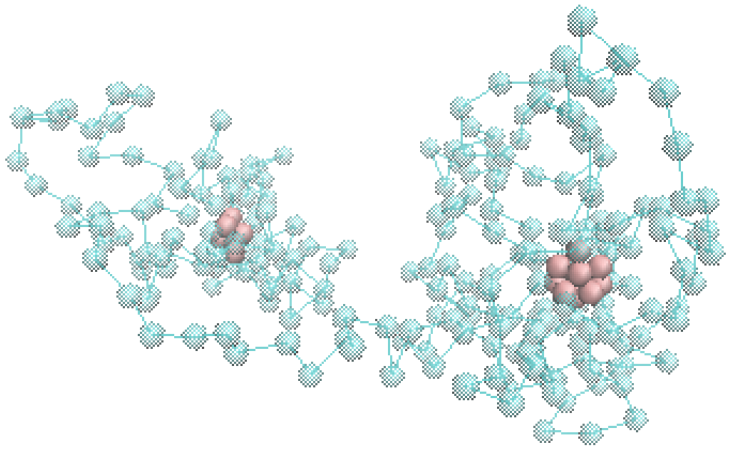


Figure 3.7: Typical configuration of the polymer relative to the previous contact matrix. The domain on the left is made of 6 bridgers, the one on the right is made of 10.

regular distribution of bridgers is invariant under translation along the chain; this means that, at thermodynamic equilibrium, the different spatial domains can exchange bridgers and - in this way - translate along the chain. For a two-domains phase this behaviour should be visible as a frequent reallocation of the bridger in the rosette, similar to the behaviour observed in figure 3.6 B. The addition of disorder can break translational symmetry, reducing the possibility of the domains to exchange monomers and move their borders.

The idea is that the disorder can introduce enough inhomogeneities in the system to force the domains to remain segregated: we suggest that the border between two consecutive domains could form in correspondence of a large gap in the distribution of the bridgers. While this is not clear in figure 3.8, it is more evident in figure 3.9.

In figure 3.9 we present the average contact matrix for a single realization of the disordered distribution of bridgers, with $\sigma = 16$. We can observe that the two-blocks structure still has well-defined borders and in particular these border seem to correspond to the large gaps at the monomer indexes ~ 25 and ~ 110 (this is also visible in the typical snapshots of the Monte Carlo trajectories, in figure 3.9 B).

However, the effect of localization induced by the disorder appears to be non self-averaging, because different realizations of the same distribution $\sigma = 16$ does not show a clear compartmentalization into two fixed spatial domains; for example see figure 3.10, in which the reallocation of bridgers typical of the regular system is still visible.

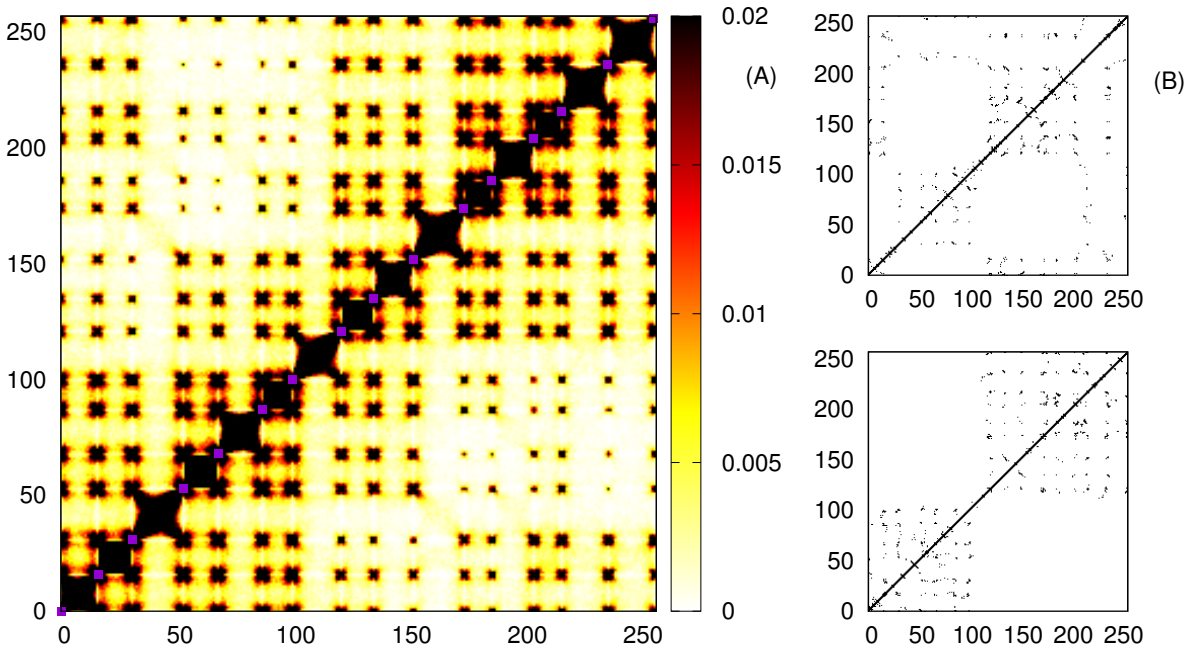


Figure 3.8: $\sigma = 3$. Contact matrix for a polimer made of $N = 257$ monomers and $p = 17$ bridgers distant 16 monomers from each other. The system is at the temperature $kT/\varepsilon = 0.3$, at which two domains are visible. These simulations were performed with $N = 257$, $\eta = 17/257$, $R = 0.77$, $1.8 \cdot 10^9$ Monte Carlo sweeps.

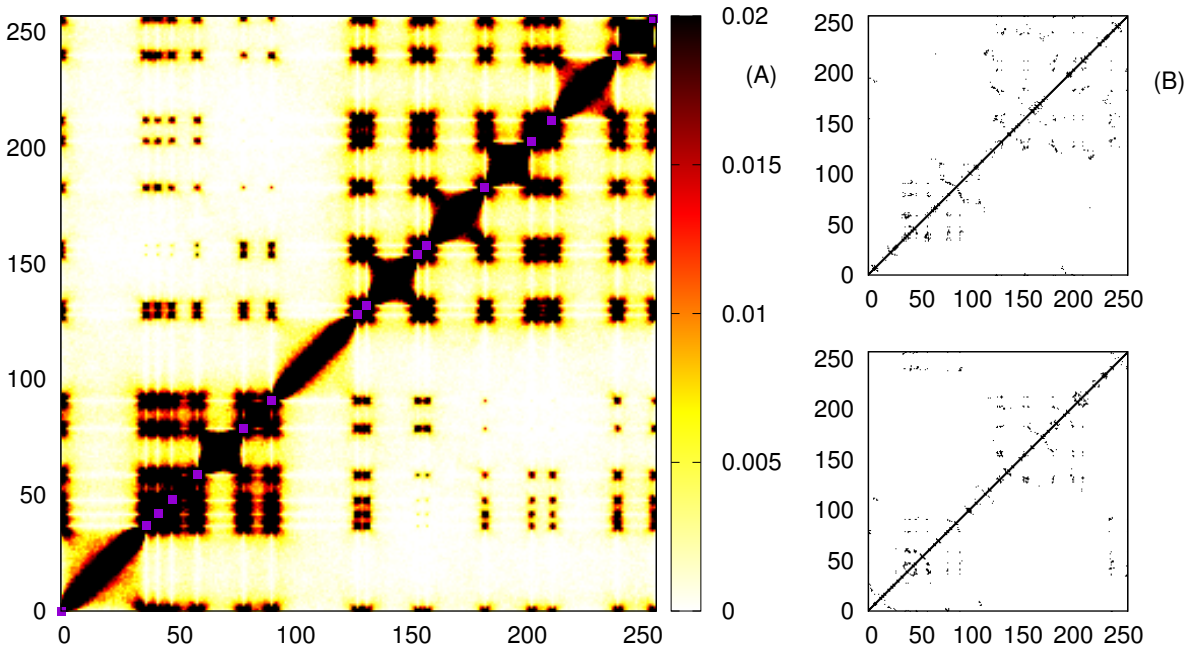


Figure 3.9: $\sigma = 16$, $r = 1$. Contact matrix for a polimer made of $N = 257$ monomers and $p = 17$ bridgers distant 16 monomers from each other. The system is at the temperature $kT/\varepsilon = 0.3$, at which two domains are visible. These simulations were performed with $N = 257$, $\eta = 17/257$, $R = 0.77$, $1.8 \cdot 10^9$ Monte Carlo sweeps.

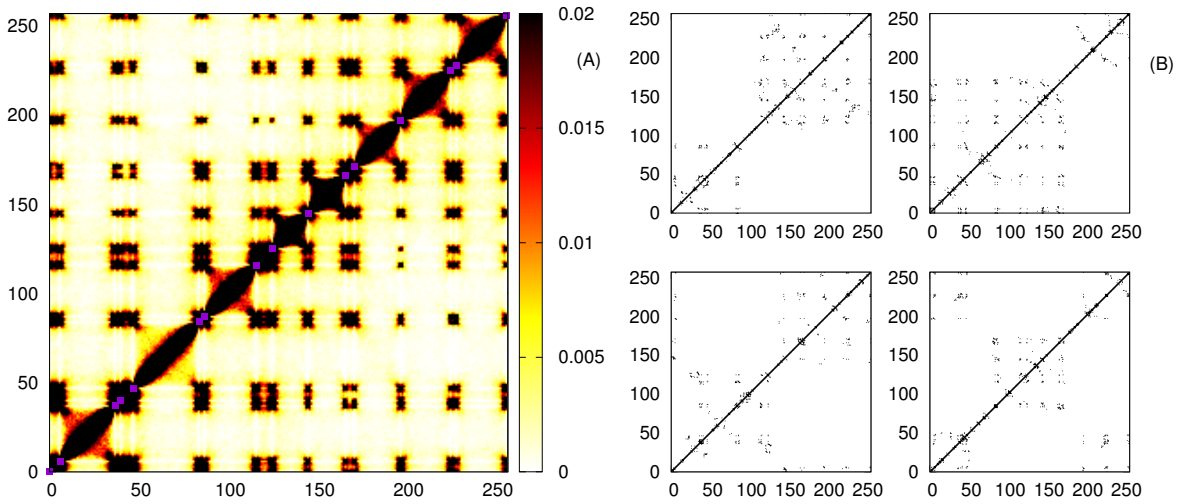


Figure 3.10: $\sigma = 16$, $r = 0$. Contact matrix for a polimer made of $N = 257$ monomers and $p = 17$ bridgers distant 16 monomers from each other. The system is at the temperature $kT/\varepsilon = 0.3$, at which two domains are visible. These simulations were performed with $N = 257$, $\eta = 17/257$, $R = 0.77$, $1.8 \cdot 10^9$ Monte Carlo sweeps.

Conclusions

We have used a homopolymer with bridging interaction model to describe the formation of spatial domains in the chromatin.

We first considered the equi-spaced distribution of bridgers proposed in [10]: we have proved with parallel tempered Monte Carlo simulations the stability of the multiple rosette state, excluding the possibility of it being a metastable feature on the way to the standard random globule. This feature of the model is a subtancial difference with the class of models that use non-equilibrium polymeric model to describe the spatial organization of chromatin, such as the *fractal globule* model [6]. Proving the equilibration of the model used in this thesis is important, because some evidences show that the chromatin fiber diffuses rapidly enough on the TADs scale to actually reach equilibrium states [3].

Then we have extended the homopolymer with bridging interaction to a more realistic case than the regular distribution of interactions, namely we have introduced disorder in the model. We have considered a random perturbation of the position each bridger by a random number of beads extracted with a Gaussian distribution. Importantly, we computationally proved that the multiple rosette state is stable both for small amount of disorder and for a uniform random distribution (in contrast with [19]).

Surprisingly, we have observed that the disorder plays a significant role in the stabilization of the multiple domain phase: the more the disorder in the system, the wider the range of temperatures in which the phase is stable.

Lastly, we point out that the disorder appears to cause a localization-like behaviour. The model without disorder is invariant under translations along the polymer; for this reason the spatial domains are able to travel along the chain. The addition of the disorder seems to break translational invariance, forcing the domains to be fixed at a certain position along the polymer.

Future developments

The polymeric model described in this thesis provides a simple mechanism to describe the formation of spatial domain: a short-ranged attraction that binds together distant monomers is sufficient to produce a stable multiple rosette configuration. Nevertheless this is a minimalist model and there are several features of the experimental data which still remain unexplained, such as:

1. The rosettes displayed by the bridged heteropolymer all have the same size, which is clearly not the case in the chromatin data [1]. A question to be answered is which distribution of the interacting monomers on the polymer can display spatial domains of different size.

2. The experimental contact matrices are characterized by a power law dependence of the contact probability P as a function of the genomic distance s: $P(s) \sim s^{-\alpha}$ (see section 1.1). The model described in this thesis still has to be tested to reproduce the possible values of α.

Additionally, the phenomenon of localization that we have described in section 3.3 could be a fundamental ingredient to describe the formation of TADs. Experimental data show contact matrices with well-defined blocks [1], which do not seem compatible with the idea of rosettes that translate along the chromatin fiber. The symmetry breaking effect induced by disorder could be necessary to describe TADs fixed in certain positions on the genome. Additionally, the idea of localization caused by disorder is important in many areas of physics, among which the notorious Anderson localization ([20], reviewed for example in [21]). For these reasons it will be interesting to study the localization of spatial domains without the border effects caused by the finite size of the simulation; a way to do this could be simulating ring polymers. It must be noted that Monte Carlo simulations do not provide a physical dynamic but only a statistical sampling of the canonical ensemble. Nevertheless it is still possible to obtain cinematic informations (such as the characteristic time of the translation) from a Monte Carlo trajectory by using only local moves, which can be interpreted as a diffusive dynamic [22].

Bibliography

[1] A. Denker and W. De Laat, “The second decade of 3c technologies: detailed insights into nuclear organization,” *Genes & development*, vol. 30, no. 12, pp. 1357–1382, 2016.

[2] M. V. Imakaev, G. Fudenberg, and L. A. Mirny, “Modeling chromosomes: Beyond pretty pictures,” *FEBS letters*, vol. 589, no. 20PartA, pp. 3031–3036, 2015.

[3] L. Ringrose, *Epigenetics and Systems Biology*. Academic Press, 2017.

[4] M. Merkenschlager and E. P. Nora, “Ctcf and cohesin in genome folding and transcriptional gene regulation,” *Annual review of genomics and human genetics*, vol. 17, pp. 17–43, 2016.

[5] J. Zuin, J. R. Dixon, M. I. van der Reijden, Z. Ye, P. Kolovos, R. W. Brouwer, M. P. van de Corput, H. J. van de Werken, T. A. Knoch, W. F. van IJcken, *et al.*, “Cohesin and ctcf differentially affect chromatin architecture and gene expression in human cells,” *Proceedings of the National Academy of Sciences*, vol. 111, no. 3, pp. 996–1001, 2014.

[6] L. A. Mirny, “The fractal globule as a model of chromatin architecture in the cell,” *Chromosome research*, vol. 19, no. 1, pp. 37–51, 2011.

[7] 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, no. 15, pp. 9553–9561, 2014.

[8] C. A. Brackley, J. Johnson, S. Kelly, P. R. Cook, and D. Marenduzzo, “Simulated binding of transcription factors to active and inactive regions folds human chromosomes into loops, rosettes and topological domains,” *Nucleic acids research*, vol. 44, no. 8, pp. 3503–3512, 2016.

[9] L. I. Nazarov, M. V. Tamm, V. A. Avetisov, and S. K. Nechaev, “A statistical model of intra-chromosome contact maps,” *Soft Matter*, vol. 11, no. 5, pp. 1019–1025, 2015.

[10] V. F. Scolari and M. C. Lagomarsino, “Combined collapse by bridging and self-adhesion in a prototypical polymer model inspired by the bacterial nucleoid,” *Soft matter*, vol. 11, no. 9, pp. 1677–1687, 2015.

[11] P.-G. De Gennes, *Scaling concepts in polymer physics*. Cornell university press, 1979.

[12] I. Lifshitz, A. Y. Grosberg, and A. Khokhlov, “Some problems of the statistical physics of polymer chains with volume interaction,” *Reviews of Modern Physics*, vol. 50, no. 3, p. 683, 1978.

[13] T. Witten and P. Pincus, “Colloid stabilization by long grafted polymers,” *Macromolecules*, vol. 19, no. 10, pp. 2509–2513, 1986.

[14] G. Tiana, F. Villa, Y. Zhan, R. Capelli, C. Paissoni, P. Sormanni, E. Heard, L. Giorgetti, and R. Meloni, “Montegrappa: an iterative monte carlo program to optimize biomolecular potentials in simplified models,” *Computer Physics Communications*, vol. 186, pp. 93–104, 2015.

[15] R. H. Swendsen and J.-S. Wang, “Replica monte carlo simulation of spin-glasses,” *Physical Review Letters*, vol. 57, no. 21, p. 2607, 1986.

[16] E. Marinari and G. Parisi, “Simulated tempering: a new monte carlo scheme,” *EPL (Europhysics Letters)*, vol. 19, no. 6, p. 451, 1992.

[17] A. M. Ferrenberg and R. H. Swendsen, “Optimized monte carlo data analysis,” *Physical Review Letters*, vol. 63, no. 12, p. 1195, 1989.

[18] M. Daoud and J. Cotton, “Star shaped polymers: a model for the conformation and its concentration dependence,” *Journal de Physique*, vol. 43, no. 3, pp. 531–538, 1982.

[19] I. Junier, O. Martin, and F. Képès, “Spatial and topological organization of dna chains induced by gene co-localization,” *PLoS computational biology*, vol. 6, no. 2, p. e1000678, 2010.

[20] P. W. Anderson, “Absence of diffusion in certain random lattices,” *Physical review*, vol. 109, no. 5, p. 1492, 1958.

[21] B. Kramer and A. MacKinnon, “Localization: theory and experiment,” *Reports on Progress in Physics*, vol. 56, no. 12, p. 1469, 1993.

[22] H. Risken, “Fokker-planck equation,” in *The Fokker-Planck Equation*, pp. 63–95, Springer, 1996.