THE UNIVERSITY *of* EDINBURGH

A multi–scale study of chromatin organisation and function: DNA topology, epigenetics and chromatin compaction

Marco Ancona

Doctor of Philosophy
The University of Edinburgh
18, July, 2022

Declaration of Authorship

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified. Parts of this work have been published in:

- Ancona, M., Bentivoglio, A., Brackley, C. A., Gonnella, G. and Marenduzzo, D. (2019). Transcriptional bursts in a nonequilibrium model for gene regulation by supercoiling. *Biophysical journal, 117(2)*, 369-376.
- Ancona, M., Michieletto, D. and Marenduzzo, D. (2020). Competition between local erasure and long-range spreading of a single biochemical mark leads to epigenetic bistability. *Physical Review E, 101*(4), 042408.
- Ancona, M. and Brackley, C. A. (2021). Simulating the chromatin mediated phase separation of model proteins with multiple domains. *arXiv preprint arXiv:2107.14518*.

Additional work not contributing to this thesis has been published in:

- Bentivoglio, A., Ancona, M., Brackley, C. A., Gonnella, G. and Marenduzzo, D. (2018). Non-equilibrium phase transition in a model for supercoiling-dependent DNA transcription. *Soft matter, 14*(18), 3632-3639.
- Ancona, M., Bentivoglio, A., Caraglio, M., Gonnella, G. and Pelizzola, A. (2021). Emergence of effective temperatures in an out-of-equilibrium model of biopolymer folding. *Physical Review E, 103*(6), 062415.

(Marco Ancona, 18, July, 2022)

Abstract

Marco ANCONA

A multi–scale study of chromatin organisation and function:
DNA topology, epigenetics and chromatin compaction

Understanding chromatin organisation at different length scales is still one of the most puzzling challenges in biophysics. Nowadays, it is clear that DNA or chromatin conformational changes can profoundly affect gene expression. Yet, the mechanisms underlying such conformational changes remain elusive. Several factors can intervene in gene regulation: supercoiling (SC), the extent of over– or under– twist of DNA double helix, can compact DNA in both bacteria and eukaryotes, yielding transcriptional over–expression or repression. Post-translational modifications of histone tails demarcate the "epigenetic" domains, which are therefore vital to establish the correct chromatin environment. Chromatin–binding proteins can form biological "condensates" via phase separation mechanisms. Recently, liquid–liquid phase separation (LLPS) has much been touted to motivate the formation of protein clusters in vivo, often referred to as 'nuclear bodies'. In addition, the so-called bridging-induced phase separation (BIPS), explains how protein aggregation can be mediated by chromatin only, even in the absence of protein-protein interaction. By using a multi-technique approach, in this thesis' work I investigate the structural and dynamical properties of DNA and chromatin at different length scales. Monte Carlo algorithms were implemented to simulate SC dynamics in a stochastic model for bacterial transcription. Similar techniques were used to show that an infection–like model can entail epigenetic bistability. Molecular dynamics simulations were employed to study the static and dynamical properties of model protein aggregates; the interplay between LLPS and BIPS was explored, showing properties which go far beyond the liquid state. Depending on the parameters, solid–like, glassy and fractal protein condensates can co–localise with chromatin.

Lay summary

Since its discovery in 1953, DNA organisation in cells has been studied in much detail. Different levels of DNA compaction are required in order to fit a 2 metre fibre into a nucleus of 2μm in size; moreover, a dense melt of DNA and proteins, the so–called *chromatin*, has to precisely fulfill all its 'duties' – first of all the production of vital proteins, a intricate multi–step machinery known as *gene expression*. To complicate the picture even more, nucleoplasm is very crowded: proteins and chromatin–associated complexes can diffuse, self–assemble and interact with DNA and chromatin, shaping nuclear organelles and chromatin domains. How different functional degrees of order and compaction can be reached in such a crowded environment? To answer this question, here I will dissect different aspects of DNA and chromatin organisation, and I will propose different physical models which inspect the relation between conformation changes of DNA and its functions. I investigate transcription regulation, epigenetics and static and dynamical properties of heterochromatin globules, by means of 1D stochastic models, Monte Carlo algorithms and 3D simulations in LAMMPS. My work reveals some interesting scenarios that might have vital implications for real cells.

Contents

List of Figures

List of Tables

List of Symbols

Chapter 1

bp	base pairs	dimensionless
N_{DNA}	DNA molecule length (in bp)	dimensionless
h_0	Relaxed DNA pitch	dimensionless
h	DNA pitch	dimensionless
Tw	Twist	dimensionless
Wr	Writhe	dimensionless
Lk	Linking number	dimensionless
σ_{Lk}	Specific linking number	dimensionless
Lk_{m}	Specific linking number of closed DNA rings	dimensionless
F	Supercoiling free energy	J
K	Elastic constant supercoiling free energy	J
$k_{ON,OFF}$	$ON \leftrightarrow OFF$ rates telegraph process	s^{-1}
k_i	transcription rate during ON state	s^{-1}
$f(t)$	pdf waiting times	s^{-1}
$\Phi(t)$	Sequence size function (ssf)	dimensionless
$\tau_{1,2}$	Inflection points of ssf	s^{-1}
τ_x	Burst duration	s^{-1}
β	Burst size	dimensionless
ξ	Burstiness parameter	dimensionless
σ	Local supercoiling density	dimensionless
$\mathcal{F}$	Supercoiling free energy density	$\mathrm{J\,m}^{-1}$
L	Lattice length	m
Δx	Lattice spacing	m
D	Supercoiling diffusivity	$\mathrm{m}^2\,\mathrm{s}^{-1}$
J_{tr}	Supercoiling flux density	$\mathrm{m\,s}^{-1}$
k_{topo}	Topoisomerases relaxation rate	s^{-1}
n	Number of genes	dimensionless
N	Number of RNAP	dimensionless
λ	Gene length	m
y_j	Promoter of the jth gene	m
J_i	Supercoiling flux generated by the ith RNAP	$\mathrm{m}^2\,\mathrm{s}^{-1}$

t_i	Time after initiation of the ith RNAP	s^{-1}
η_i	RNAP state (active or inactive)	dimensionless
J_0	Supercoiling flux generated at time $t = 0$	$\mathrm{m}^2\,\mathrm{s}^{-1}$
v	RNAP velocity	$\mathrm{m}\,\mathrm{s}^{-1}$
$k_{\mathrm{in,j}}$	Initiation rate of the jth gene	s^{-1}
α	Coupling parameter	dimensionless
k_0	Initiation parameter	s^{-1}
$\sigma_{p,j}$	Supercoiling at promoter of the jth gene	dimensionless
σ_p	Supercoiling at promoter	dimensionless
$\bar{J}$	Supercoiling average flux parameter	$\mathrm{m}^2\,\mathrm{s}^{-1}$
n_{topo}	Experimental number of topoisomerases per gene	dimensionless
$\sigma_{\mathrm{baseline}}$	Experimental baseline supercoiling in bacteria	dimensionless
$\bar{\sigma_p}$	Average negative supercoiling at promoter	dimensionless
σ_0	Supercoiling at time $t = 0$	dimensionless
$\bar{\sigma}(x)$	Time-averaged local supercoiling	dimensionless
$\bar{k_{\mathrm{in}}}$	Time-averaged initiation rate	dimensionless
$\bar{J_{\mathrm{tr}}}(x,t)$	Time-averaged supercoiling flux density	$\mathrm{m}\,\mathrm{s}^{-1}$
M	Rescaled flux	dimensionless
ω	inverse length scale of supercoiling screening	m^{-1}
$h(x)$	Auxiliary function	dimensionless
$\bar{x}$	Distance between (static) RNAP and promoter	m
X	Stochastic variable	$[X]$
μ_X	Expectation value of X	$[X]$
$\mathrm{Var}(X)$	Variance of X	$[X]^2$
Skew	Skewness	dimensionless
Kurt	Kurtosis	dimensionless

Chapter 2

N	Number of nucleosomes	dimensionless
s_i	State of ith nucleosome	dimensionless
n_s	Number of states for each nucleosome	dimensionless
n_j	Number of nucleosomes in the state j	dimensionless
n_M	Number of *modified* nucleosomes	dimensionless
n_U	Number of *unmodified* nucleosomes	dimensionless
$R_\pm$	$U \leftrightarrow M$ conversion rates (2-state)	s^{-1}
α	Rate of spontaneous conversion (2-state)	s^{-1}
η_M	Noise for modified nucleosomes	s^{-1}

F	recruitment-to-noise ratio	dimensionless
$m(t)$	Global methylated nucleosomes density	dimensionless
$a(t)$	Global acetylated nucleosomes density	dimensionless
$u(t)$	Global unmodified nucleosomes density	dimensionless
$R_{X\to Y}$	Conversion rates X, Y in U, M, A (3-state)	s^{-1}
$\alpha_{M,A}$	Methylation/acetylation rate by recruitment (3-state)	s^{-1}
$\beta_{M,A}$	Demethylation/deacetylation rate by recruitment (3-state)	s^{-1}
$\chi_{M,A}$	Spontaneous methylation/acetylation ssf	s^{-1}
$\gamma_{M,A}$	Spontaneous demethylation/deacetylation	s^{-1}
$a_{s,u}$	Saddle/unstable acetylation fixed point	dimensionless
$m_{s,u}$	Saddle/unstable methylation fixed point	dimensionless
a^*	Acetylation stable fixed point	dimensionless
m^*	Methylation stable fixed point	dimensionless
l	Distance	m
σ	Contact exponent in Lévy distribution	dimensionless
m_i	State (0 or 1) of the ith site/nucleosome on lattice	dimensionless
$q_{\lambda,\mu}$	Spontaneous methylation/demethylation rate	s^{-1}
λ	Methylation parameter by recruitment	dimensionless
μ	Demethylation parameter by recruitment	dimensionless
$H(\sigma+1)$	Normalisation factor	dimensionless
$q_{0\leftrightarrow 1}$	Methylation/demethylation rate	s^{-1}
z	Random number drawn by power law distribution	dimensionless
$m(x,t)$	Methylation field	dimensionless
$d(x,t)$	Demethylation field	dimensionless
D_A	Anomalous diffusivity	$m^{\sigma}\mathrm{s}^{-1}$
$\eta(x,t)$	Noise field	s^{-1}
$\Gamma^{'},^{''}$	Noise amplitudes	$\mathrm{m\,s}^{-1}$
$r_{\pm}$	Conversion rates	s^{-1}
$q^{'}_{0,1}$	Auxiliary rates	s^{-1}
$\bar{\sigma_p}$	Average negative supercoiling at promoter	dimensionless
κ	Effective methylation rate	s^{-1}
$\bar{\lambda}$	Rescaled methylation rate	s^{-1}
$\bar{\mu}$	Rescaled demethylation rate	s^{-1}
$\bar{m}$	Average methylation	dimensionless
d	Physical dimension of the system	dimensionless
α	Critical exponent associated with density decay	dimensionless

$S(t)$	Survival probability	s^{-1}
δ	Critical exponent associated with survival probability	dimensionless
Γ	Noise amplitude in the effective Langevin equation	$\mathrm{m\,s}^{-1}$
ζ	Noise field in the effective Langevin equation	s^{-1}
$f(\sigma)$	Monotonic auxiliary function	s^{-1}
$V(m)$	Effective mean methylation potential	dimensionless
A	Normalisation constant	dimensionless
Q	Noise amplitude in the mean field equation	s^{-1}
$J(m)$	Methylation flux in the equivalent Fokker–Planck (FP) equation	s^{-2}
$K(m)$	Auxiliary function in the mean field FP equation	s^{2}
$\tilde{\kappa}$	Rescaled effective methylation growth	dimensionless
m_{min}	Optimal methylation value in the monostable phase	dimensionless

Chapter 3

$\phi_{A,B}$	Fluid concentrations of type A and B	dimensionless
$\epsilon_{AA,BB,AB}$	Interaction strengths	J
$L(t)$	Length scale of fluctuations in LLPS	m
R_g	Radius of gyration j	m
L	Chromatin length	m
ν	Scaling exponent of $R_g(L)$	dimensionless
σ	Diameter chromatin bead	m
U_{FENE}	FENE potential	J
U_{WCA}	Weeks–Chandler–Andersen potential	J
R_0	Maximum extention FENE potential	m
k_B	Boltzmann constant	$\mathrm{J\,K}^{-1}$
T	Temperature	K
d_{ij}	Mean diameter between the ith and the jth beads	m
U_{BEND}	Kratky-Porod potential	J
K_{BEND}	Bending energy	J
θ_i	Angle of the ith polyemer bead triplet	rad
l_p	Persistence length	m
U_{LJ}	Truncated Lennard–Jones (LJ) potential	J

ϵ	Energy strength of attractive LJ	J
ϵ_{sp}	Energy strength of attractive LJ for specific binding sites	J
m_i	Mass of ith bead	kg
ξ_i	Friction ith bead	$\text{kg}\,\text{s}^{-1}$
η_i	Vector noise in Langevin equation	$\text{s}^{0.5}$
τ_{LJ}	Simulation time unit	s
N_{eq}^{ϵ}	Number of equilibrated runs	dimensionless
T_{eq}	Length equilibration run	s
τ_r	First relaxation time scale	s
T	Length of simulations (not Temperature, extrapolate from context)	s
t^*	Time at which a particle has committed to another position	s
$\delta r(t)^2$	Radial mean square displacement for single particle	m^2
Σ_i	Neighbourhood of particle i	dimensionless
N^*	Number of particles in cluster	dimensionless
$q_{lm}(i)$	Steinhardt bond order parameter	dimensionless
$n(\Sigma_i)$	Number of particles in the neighbourhood of i	dimensionless
$Y_l^m(\theta, \phi)$	Complex Laplace's spherical harmonics	dimensionless
$P_l^m(\theta)$	Legendre polynomial	dimensionless
$\bar{q}_l(i)$	Bond orientational order parameter associated with the 2nd coordination shell of i	dimensionless
$\psi_l(i)$	Bond orientational order parameter associated with i	dimensionless
ρ_i	Local density of the cell i	m^{-3}
V_i	Volume of the cell i	m^3
$\rho(x,t)$	Chromatin density field	m^3
$\phi(x,t)$	Protein density field	m^3
f	Landau free energy density	Jm^{-3}
$D_{1,2}$	'Elastic' constants in the Landau free energy	Jm^3
χ	Coupling constant in the Landau free energy	Jm^3
κ	Surface tension parameter in the Landau free energy	J m
g	Accumulation penalty parameter	Jm^6
ρ_0	Overall chromatin density	m^3
ϕ_0	Overall protein density	m^3
$M_{\rho,\phi}$	Chromatin and protein mobility in Model B equation	$\text{J}^2\,\text{s}^2\,\text{m}\,\text{kg}^{-1}$
$\mathcal{F}$	Landau free energy	J

t_u	Time unit	s
x_u	Space unit	m
t'	Rescaled time	dimensionless
x'	Rescaled time	dimensionless
Φ	Rescaled protein field	m^{-3}
$\mathcal{D}_0$	Rescaled diffusivity	dimensionless
$\mathcal{G}$	Rescaled accumulation penalty parameter	dimensionless
$\succ$	Rescaled coupling parameter	dimensionless
ρ'	Perturbed chromatin density field in the linear stability analyisis	m^{-3}
Φ'	Perturbed protein density field in the linear stability analyisis	m^{-3}
$\succ$	Rescaled coupling parameter	dimensionless
$\mathcal{D}$	Rescaled diffusivity and accumulation penalty	dimensionless
$\tilde{g}$	Fourier transform of function g	$[f]\,\mathrm{m}^3$
$\lambda_{+,-}(Q)$	Roots of characteristic polynomial	
Δ	Lag time	s
D_{eff}	Effective anomalous diffusivity	$\mathrm{m}^2\,\mathrm{s}$
α	Dynamical exponent MSD	dimensionless
μ_t	Mixing parameter	dimensionless
$n_{i,\mathrm{neigh}}$	Number of proteins in the Voronoi neighbourhood	dimensionless
$\tilde{g}$	Fourier transform of function g	$[f]\,\mathrm{m}^3$
$\lambda_{+,-}(Q)$	Roots of characteristic polynomial	
Δ	Lag time	s
D_{eff}	Effective anomalous diffusivity	$\mathrm{m}^2\,\mathrm{s}$
α	Dynamical exponent MSD	dimensionless
β	Stretched exponential exponent MSD	dimensionless
τ_{inf}	Second relaxation time scale MSD	s
δr_{cm}	Distance from the center of mass	m

Chapter 4

N_{max}	Maximum number of neighbours	dimensionless
M_p	Number pf patches	dimensionless
ϵ_P	Interaction energy for patchy model	J
$U\mathrm{CD}-\mathrm{C}$	HP1–chromatin attractive Morse potential	J
ϵ_{HC}	HP1–chromatin interaction energy	J
α	Shape parameter Morse potential	m^{-1}

r_{HC}	Cut–off distance HP1–chromatin potential	m^{-1}
$\mathcal{U}\mathrm{h} - \mathrm{NTE}$	HP1–HP1 attractive Morse potential	J
ϵ_{HH}	HP1–HP1 interaction energy	J
ϵ_P	Interaction energy for patchy model	J
r_{HH}	Cut–off distance HP1–HP1 potential	m^{-1}
l_x	Box size	m
ρ	Overall HP1 density	m^{-3}
N_{sb}	Number of sub–boxes	dimensionless
V_{sb}	Volume of sub–boxes	m^3
N_i	Number of HP1s in the $i$the sub–box	m^3
ρ_i	Proteins density in the ith sub–boxes	m^{-3}
ϕ_{sep}	Separation depth	dimensionless
ρ^*	Reference density for separation depth	m^{-3}
ρ_+	High–density reference value for separation depth	m^{-3}
ρ_-	Low–density reference value for separation depth	m^{-3}
f_{tot}	Fraction of HP1 bound to chromatin	dimensionless
f_{c}	Fraction of chromatin beads bound by HP1	dimensionless
τ	Simulation unit time	s
dr	Width of spherical shell for density calculation	J
ρ_{shell}	Shell density	m^{-3}
$N_{\mathrm{shell}}(r)$	Number of HP1s in the shell at distance r from the ceneter of the droplet	dimensionless
R_d	Droplet radius	m
ϵ	Energy strength of attractive LJ	J
ϵ_{sp}	Energy strength of attractive LJ for specific binding sites	J
D_f	Fractal dimension	dimensionless
M	Number of HP1 in sub–clusters for fractal dimension calculation	dimensionless

Preface

In 1953 J. D. Watson and F. H. C. Crick set a milestone in modern history by proposing an elegant explanation for the storage of genetic information, the double–helix model for DNA. The model was supported by fiber diffraction studies by M. H. F Wilkins, A. R. Stokes, and H. R Wilson, and R. Franklin and R. G. Gosling. This discovery marked a new era in biology, giving rise to modern molecular biology. Since then, numerous scientific techniques have been produced, such as rapid gene sequencing and monoclonal antibodies; also genetic fingerprinting and modern forensics, have been inspired from this first work.

Recent advances in technologies and computational power have allowed us to tackle several practical problems in biophysics, which would have been simply unfeasible back in the days. Optical spectroscopy and microscopy, EM (Electron Microscopy), X–ray and neutron diffraction methods, NMR (Nuclear Magnetic Resonance), ChIP (Chromatin Immunoprecipitation), chromosome conformation capture techniques using 3C, 4C–seq (Circularized Chromosome Conformation Capture), 5C (Chromosome Conformation Capture Carbon Copy) and Hi–C technologies, and other genome–wide methods of analysis, such as ChIA-PET (a combination of Hi–C with ChIP–seq) are the basis of modern structural molecular biology. In modern days, FISH (fluorescence *in situ* hybridization) imaging allows us to reveal chromosome structure and organization in living cells, at a large length scale. Single–particle tracking is used to track protein and DNA motion. Optical tweezers are an optimal tool to study mechanical properties of chromatin.

However, all existing experimental methods in molecular biology are substantially *incomplete*. Nowadays, each experimental procedure alone, or even a combination of them, can only define a provisional knowledge of living systems, depending on the particular focus of the experimental protocol. That is, an ideal molecular biophysical experiment would be capable to observe not only the positions of atoms or molecules *in vivo*, but also their related conformational changes in time and their role in vital biochemical reactions, which, in turn, are related to their function. It is quite easy to understand that this is simply unfeasible. Although we have reached unbelievable resolutions and computational power to process data, the amount of information we can

extract is still limited, and a full interpretation of phenomena is often out of our reach. The prescripted direction of this field is to lean towards a *complete* understanding of all biochemical processes that involve DNA, chromatin and associated factors; in a sense, this is the ultimate goal. It is under this precept that physics came upon to try to complement biological research. A multi–disciplinary effort might be needed to untangle this puzzling scenario. *We got to get in to get out (cit.)*.

The interplay between molecular biology and physics methods can look quite strange at first glance; the different 'angles' of the two studying approaches seem to be poles apart. On average, biologists tend to investigate with 'microscopes', trying to characterising the object of study in all its details, whilst physicists use 'telescopes', trying to infer the general picture or an (ideally) universal explanation of phenomena, irrespectively of the details. Although the latter slant looks more satisfying from a scientific standpoint, one could argue that living cells are really much more complicated than other physical systems. Since life has started on this Earth, in fact, evolution has struggled for billions of years to finely "optimise" the parameters and to build this complicated living machinery. *A living cell is definitely not a magnet*.

However, this (apparently odd) partnership has produced remarkable improvements in the field. Insights from polymer physics suggest abandoning the existing paradigm of regularly looped models of chromatin, and give space to a view where higher–order chromatin organization is considered. RW (Random walks), SAW (self–avoiding walks) and equilibrium globules represent the foundation for conceiving a good DNA or chromatin model. More detailed polymer structures try to hoist the prediction ability, attempting to map the high complexity of living systems into manageable models. Data–based modelling was a powerful tool to understand experiments, as is evident from the exponential increase of papers which present simulations as experiments validation. Different methods can be used refine the 'blurred' outcomes of experiments: from long–standing Monte Carlo methods and MD (Molecular Dynamics) simulations, to the modern inference analysis, such ML (Machine Learning) algorithms and deep neural networks; thus, physicist have been equipped of powerful tools that can be employed to link experimental data to models with a high degree of accuracy.

Within this vast framework, my PhD thesis represents an attempt to contribute to the general biophysical knowledge. In this thesis' work, I go through four problems that I have faced during my four–year PhD. Rather than a mere report, this thesis can be considered as a *memorandum* of the relevant background needed to grasp the main ideas behind the works described here. I tried to select and collect the most relevant features of the biological systems under study, which are useful to model DNA and chromatin effectively. Then, I

also report the main results, which are based on those background details. The overall purpose of this thesis is to show that different physical approaches and techniques can serve to understand or predict the behaviour of living system; DNA supercoiling, epigenetics, and chromatin–protein melts are studied by using typical statistical physics tools and models.

Each chapter is organised such that the literature review sections always precede the description of the models and the results. All chapters are endowed with an introduction and a final section of conclusions; often, old breakthrough papers have been discussed in the introduction, leaving the examination of more recent developments in the following sections. Although chapters are self–consistent, a common plot links all them; the main aim of this thesis is to show the ubiquitous connection between DNA (or chromatin) structural organization and dynamics with function, at a multi–scale level. That is, by using different methods, I show that a study of genomics can be performed at different length scales: from 10 (DNA) to 10^4 (chromatin) base pairs. This kind of bottom–up approach helps to build a step–by–step protocol through which chromatin can be modelled using different levels of coarse–graining.

This main topics of my research are organised in chapters, as it follows:

- In chapter 1, I explore the interplay between topology and transcription. In our 1D stochastic model, transcription and local supercoiling are coupled, generating a positive feedback loop that gives rise to a non–trivial statistics of transcriptional events. In particular, we focus on the characterisation of the emergent "bursty" dynamics, and we obtain a phase diagram that links the occurrence of intermittent transcription to the relevant parameters of the model.

- In chapter 2, I tackle the problem of how certain epigenetic marks are distributed along genomes as their patterns are important for ultimately determining cell fate. In our 1D stochastic model there is a competition between marked and unmarked states and where the kinetics is mediated by implicit enzymes that "write" and "erase" a gene-silencing epigenetic mark. The interplay of long-range infection/marking and local recovery/erasure processes can determine epigenetic bistability, under certain conditions.

- In chapter 3, I study the properties of protein clusters which form due to bridging–induced attraction (protein–chromatin interaction only). We consider two models for the chromatin fibre: an homopolymer and a polymer with strong binding sites. The two cases lead to different phases when the relevant interaction energies are relatively large. A detailed study of the single–particle dynamics and local ordering has been performed in both cases.

- In chapter 4, I describe two versions of a simple model protein inspired by heterochromatin protein 1 (HP1) and reminiscent of the well-studied "patchy particle" models. In our simulation both protein–protein and proteins–chromatin interactions are switched on. We map out the phase diagram of the system in terms of these two attractions, identifying different regimes with behaviour strikingly different from a standard phase-separating system.

Chapter 1

DNA and topology: supercoiling as regulating factor

The genetic information of all living organisms is encoded in double–stranded DNA; unfortunately, knowing the whole genetic sequence of a living specie is not sufficient to deduct the morphologic and functional characteristics of an organism – i.e. the *phenotype*. The word "phenotype" does not only refer to some "exterior" features of humans or mammals – blue eyes, ginger hair and the color of a dog's fur are typical phenotipical characters – but it has also a meaning for prokaryotes. Indeed, also *isogenic* (genetically identical) bacteria can differ for the abundance of mRNA and proteins produced by a single or multiple genes. In this way, identical bacteria can differentiate in multiple "molecular" phenotypes. Interestingly, such diversity can arise even in the same environment; the classic work by Novick e Weiner (1957) showed that a population of *E. Coli* can show cell–to–cell variability [1], as, under certain conditions, they can split into two sub–population, one rich and one poor in β–galactosidase.

This *all–or–none* behaviour suggests that gene expression entails a series of stochastic process. In the 1976, Spudich e Koshland [2] envisioned that Poissonian fluctuations in the "small" number of RNA messangers produced could cause such non–genetic variability. From that point on, several other experimental and numerical works [3–5] revealed several other sources of stochasticity (mainly due to stochastic nature of the biochemical reactions involved), all pointing to the fact that each cell shows a particular *individual* behaviour.

Transcription, the first step in the expression process, is thought likely to have a role in the expression variability and phenotypical diversity. This assertion is backed by experimental evidence: transcriptional *burts*, time intervals

of high transcriptional activity separated by longer dormant periods, were initially observed in eukaryotes, and more recently also in bacteria [6, 7].

What is, then, the biophysical mechanism underlying such "bursty" behaviour? Various attempts to model the ON $\leftrightarrow$ OFF switching of genes have been made, by means of two–state models [8, 9] and multi–state models [10, 11]. Nonetheless, only recently more sofisticated models were employed, in particular those that couple the transcriptional process with *supercoling* dynamics [12–16]. Weil and Vinograd (1963) pioneering work on Polyomavirus [17] showed the emergence of supercoils of the double helix. Afterwards, several other papers highlighted the importance of supercoiling in the transcriptional process; in particular, in the 1984, Liu et al. speculated that the RNA polymerase, the enzyme responsible of gene transcription, generates positive (negative) supercoiling ahead (behind) the transcribed gene. Such simple, qualitative picture was named *twin supercoiling domain model* [18], and currently represent the most accredited mechanism which links transcription and supercoiling generation.

The model in [12, 13] offers a natural framework to investigate whether a bursty dynamics can emerge and under which conditions. As we will see in section 1.4, this model consider a coupling between transcription *initiation* – the initial step of transcription, that begins at the promoter, a sequence upstream of the gene – and supercoiling diffusion. Transcription initiation is supercoiling–dependent, that is the larger negative supercoiling at the promoter, the higher the transcriptional rate; supercoiling dynamics is diffusive. The model includes also the action of topoisomerases, which are topological enzymes which can relax positive and negative supercoiling. Previously, this model was used to highlight the emergence of two regimes – a *relaxed* regime in which the transcriptional process is almost Poissonian and the *supercoiling–regulated* regime, where transcription is tightly regulated by supercoiling. Here, we mostly focus on the transcriptional time series, studying the statistical properties of transcriptional events and lag times between them. We show that, depending on the strength of topoisomerases and the flux of supercoiling injected by polymerases during transcription, the model allows us to distinguish a regime of bursty dynamics, and a regime in which a supercoiling wave propagates through the DNA ring. Quantitatively, the bursts parameters are consistent with experimental values [6, 7, 14].

In section 1.1 I describe the DNA structure. Then, in section 1.2, the basics of DNA topology are linked with DNA functions; I define linking nnumber, twist, writhe and supercoiling density, and I briefly discuss the action of the enzymes responsible for supercoiling relaxation, the topoisomerases. In section 1.3, I will focus on the transcriptional process; after speaking of the role of RNA polymerase, which is capable of producing RNA transcript by 'copying' the

genetic information stored in genes, I show some experimental evidence of bursty transcription *in vitro* and *in vivo*; in the same section the methods used in the project are described: therein, the sequence–size function and the notion of local supercoiling are defined. The aforementioned sections are part of the literature background. In section 1.4, I present my work, describing our one–dimensional stochastic model for supercoiling–regulated transcription. In this work, the mean field model in [12] has been extended to the case of periodic boundary condition and in the presence of topoisomerases. The key quantities and the parameter values used are discussed on the basis of experimental evidence. An extensive analysis of the parameter space produces a phase diagram in which the bursty and non–bursty phases are clearly distinguished. It is worth to mention that some preliminary results were also studied during my master thesis, but the major contribution occurred during the PhD.

1.1 DNA structure

DNA (Deoxiribonucleic acid) is a double-stranded biopolymer that holds all the genetic information of a living organism. Nucleotides are the building blocks of a single DNA strand: a sugar ring made of five carbons (numbered from $1'$ to $5'$), the deoxyribose, bind to a single phosphate group and a nitrogen base. Four nitrogen bases constitutes the genomic 'alphabet': Adenosine (A), Guanine (G), Cytosine (C) and Thymine (T). Nucleotides are linked via phosphodiester bonds, such that the $5'$ carbon atom of a unit is linked to the $3'$ carbon atom of the consecutive unit through the phosphate group, as to confer the $5' \rightarrow 3'$ directionality to the DNA filament, see Fig. 1.1(a). A complementary DNA strand ($3' \rightarrow 5'$) is coupled to the first one via hydrogen bonds between the bases, always pairing A and G (purines) with T and C (pyramidines) respectively. Complementarity of bases is crucial for the assembly of new RNA (ribonucleic acid) molecules during transcription, or repairing processes [19]. Units of coupled monomers are known as base pairs (bp).

In physiological conditions of pH, salt concentration and temperature (pH7, 0.2M NaCl, 37^o C), linear DNA in its most common form, B-DNA, appears as a double–helix molecule, with a right–handed chirality. The double helix has a diameter of 20 Å, the pitch is 34 Å. Bases are 3.4 Å thick (rise), and stacked as to form a 'spiral staircase'. Base–stacking is responsible for the thermal stability of the double helix [20]. The relative twist between consecutive base pairs is about 36^o in the azimuthal direction, oriented perpendicularly to the helix axis. In B-DNA, the relative orientation of the backbone–base bonds are responsible for the alternate motif of major and minor grooves along the DNA double helix [21, 22], see Fig. 1.1(b). The major groove is 12 Å wide and 8.5 Å deep, whereas

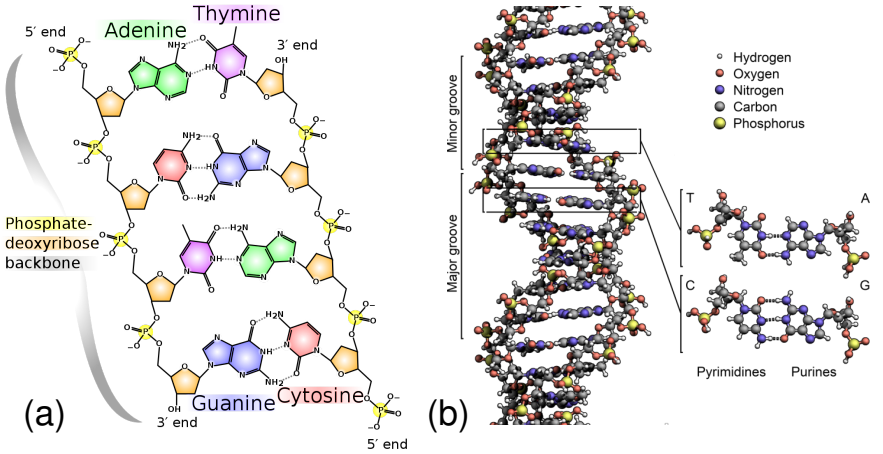

Figure 1.1. DNA structure. (a) Schematics of the primary structure of DNA. A and T are linked through two hydrogen bonds, whereas G and C through three hydrogen bonds. The two strands are antiparallel. **(b)** Secondary structure of DNA. Figures taken from [28].

the minor groove is narrower (6Å) and shallower (7 Å). The interplay between DNA groove structure and the specificity of the base sequences in the grooves is essential to create recognition patterns for certain DNA-binding protein, such as the TATA binding protein (TBP) and other transcription factors [23, 24]. A full helical turn consists of about 10.4–10.5 bp. Different DNA structures, A-, C-, D-, T- and Z-DNA amongst them, can arise in different solution conditions, or under mechanical stress, and can be either vital or detrimental for the cell sustainance. Z-DNA is arguably one of the most interesting: this left–handed double helix [25, 26] is often found in transcriptionally active DNA regions, at very high concentration of NaCl. Moreover, several chemicals can induce conversion of B-DNA to Z-DNA. Excess or loss of Z-DNA has also severe implications in disease. [27].

1.2 DNA topology: a link with function

Rather than a rigid, uniform structure, DNA exhibits great structural variability. Not only specific sequences can induce small intrinsic curvature to the helix axis, as ascertained by electrophoretic studies [29, 30], but DNA also manifests bending and torsional flexibility. The conformational variations associated with several DNA–binding proteins and complexes assess the strict relation between DNA structure (topology) and DNA function (genic expression). Thus, it is of enormous importance to study how DNA topology affects DNA regulation of genic expression.

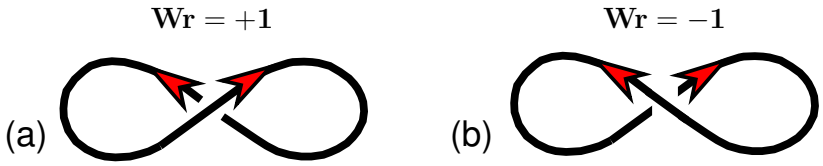

Figure 1.2. Positive and negative writhe. Diagram of **(a)** positive and **(b)** negative writhe. Figure taken from [31].

To this extent, DNA can be treated as a two-strands braid; let us consider the case where the two terminal ends are joined to form a closed ring. This simple system is instructive for two reasons: it allows us to define topological quantities which are properties of any pair of closed curves, and has a direct biological counterpart in bacterial chromosomes and plasmids, which are commonly closed–circular DNA molecules.

1.2.1 DNA supercoiling

As detailed above, DNA is a right–handed double–helix, in which the two filaments are wound around each other every 10.5 bp/turn. This number represents the pitch of the helix, h_0, under standard enviromental conditions. For a closed planar DNA molecule of length N_{DNA}, the ratio N_{DNA}/h represents the number of turns, being h the pitch of the relaxed DNA ring. In general $h \neq h_0$, as it depends on the external conditions; note that, even in standard conditions, $h \neq h_0$, since a residual torsion is needed to join the 3′ and 5′ terminal ends of each strand to close the ring. The *standard linking number* Lk is then defined as the nearest integer to the aforementioned fraction.

This number is a fundamental property of closed DNA molecules, as it is a topological invariant: both thermally– and mechanically–induced deformations of the helix cannot change this number. The number of turns can be modified by cutting either one (nicked DNA) or both (open DNA) DNA strands: several enzymes, such as topoisomerase and gyrases, can introduce or remove links by opening and swivelling the double helix, see section 1.3.2. However, deformation of a planar DNA ring can convert the extent of coiling around the helix axis, or *twist* (Tw), into bending excess of the axis in space, or *writhe* (Wr) (and *vice versa*), which eventually can generate extended intertwined DNA structures called *plectonemes*. In a plectoneme, DNA wraps multiple times around itself to form a braid. Twist and writhe are geometrical properties of closed DNA, and are related to the linking number Lk by the Călugăreanu-White-Fuller formula [32–34]:

$$\mathrm{Lk} = \mathrm{Tw} + \mathrm{Wr}. \tag{1.1}$$

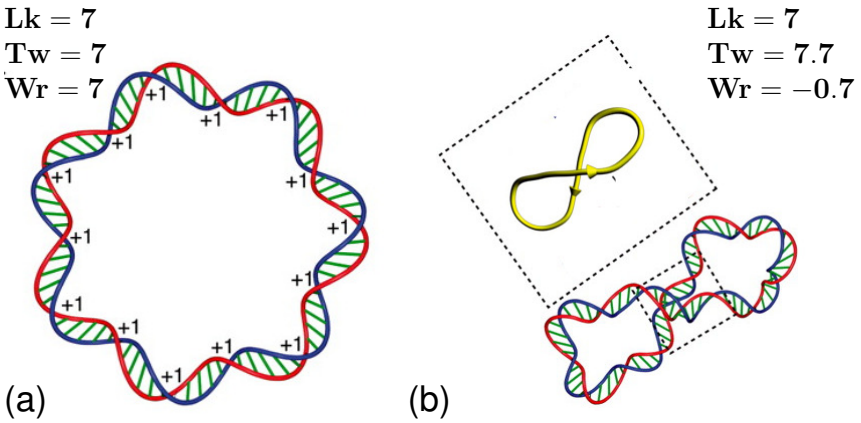

Figure 1.3. Twist–to–writhe conversion. (a) Circular DNA in its planar form. The linking number equals the twist. Lk can be calculated as half of the total number of intersection (to which $+1$ is assigned) between the two strands [35]. **(b)** Lk remains constant, while an excess of twist (Tw $= 7.7$) is balanced by negative writhe (Wr $= -0.7$). Figures adapted from [36].

Eq. (1.1) links a topological invariant of closed DNA to its geometrical properties. Note that, althogh Lk is an integer number, Tw and Wr can be rational and negative, see Fig. 1.3(a,b). A representative examples of twist–to–writhe conversion is shown in Fig. 1.3(c,d)

An excess/defect of linking number with respect to the reference linking number Lk_0 ($\mathrm{Lk}_0 = N/h_0$) quantifies the level of over/under–twist and the presence of positive/negative writhe. DNA molecules of the same length but with different Lk are referred to as *topoisomers*. The extent of torsional stress introduced can be measured as *specific linking difference*, also called *supercoiling density* (which we will refer to as supercoiling, for brevity):

$$\sigma_{\mathrm{Lk}} = \frac{\mathrm{Lk} - \mathrm{Lk}_0}{\mathrm{Lk}_0} = \frac{\Delta\mathrm{Tw} + \Delta\mathrm{Wr}}{\mathrm{Lk}_0}. \tag{1.2}$$

The specific linking difference σ_{Lk} is a global feature of the DNA molecule. Furthermore, it is independent of the length of the molecule, as it is normalized by Lk_0, thus it is useful to compare the level of supercoiling density between different topoisomers. Bacterial plasmids or the *E. Coli* chromosome have a baseline specific linking difference of -0.06 [35, 37].

Experiments with different relaxed topoisomers on agarose gels have revealed that the concentration $P(\mathrm{Lk}_x) \equiv [\mathrm{Lk}_x]$ of DNA rings having linking number $\mathrm{Lk}_x$is Gaussian [35]. Gel electrophoresis is commonly used to separate DNA molecules (often bacterial plasmids) on the basis of size and compactness; the DNA molecules can move inside the matrix gel under the innfluesnce of an electric field. Negatively supercoiled plasmids can migrate faster, due to their

writhed structure, whereas relaxed molecules, which are less compact have low mobility. Since $[\mathrm{Lk}_x]$ is Gaussian, we have:

$$\frac{1}{\Delta \mathrm{Lk}} \ln \frac{[\mathrm{Lk}_x]}{[\mathrm{Lk}_\mathrm{m}]} \propto -\Delta \mathrm{Lk}, \tag{1.3}$$

where $\Delta \mathrm{Lk} = \mathrm{Lk}_x - \mathrm{Lk}_\mathrm{m}$, being Lk_m the linking number of relaxed *closed* topoisomers. Therefore, the associated free energy $F[\sigma_{\mathrm{Lk}}]\, s$ reads:

$$F[\sigma_{\mathrm{Lk}}] = \frac{K}{{\mathrm{Lk}_\mathrm{m}}^2} \ln \frac{[\mathrm{Lk}_x]}{[\mathrm{Lk}_\mathrm{m}]} = K \left(\frac{\Delta \mathrm{Lk}}{\mathrm{Lk}_\mathrm{m}} \right)^2 \simeq K \sigma_{\mathrm{Lk}}^2. \tag{1.4}$$

Note that the addition of the constant term $\ln([\mathrm{Lk}_\mathrm{m}])$ does not change the free energy, which has been also scaled by $K/{\mathrm{Lk}_\mathrm{m}}^2$. The last approximation is valid for long DNA molecules, when $\mathrm{Lk}_m \simeq \mathrm{Lk}_0$. Eq. (1.4) confirms that the supercoiling of DNA is an elastic process of 'elastic' constant $2K$, at least for small ΔLk (since σ_{Lk} is dimensionless, K has the dimension of an energy).

Although Lk is conserved in a closed DNA molecule, some topological enzymes, or topoisomerases, are responsible of introducing or removing supercoiling *in vivo*. The topoisomerases of type I remove supercoiling by nicking DNA (i.e. cutting one strand), while , topoisomerases of type II cut both DNA strands. An entire family of topoisomerases of type I and II are present in both eukaryotes and prokaryotes; a list of topoisomerases is given in [35], while a more detailed review on their structure and function can be found in [38]. In bacteria, topoisomerase I removes negative supercoiling, as it can bind negatively supercoiled DNA regions, with a variation of the linking number of $\Delta \mathrm{Lk} = +1$. DNA gyrase, a particular topoisomerase of type II, introduces negative supercoiling, producing a linking difference of $\Delta \mathrm{Lk} = -2$. Both these topoisomerases do not require ATP.

1.3 Supercoiling and transcription: an inextricable relation

.

Several lines of evidence show that there is an intimate connection between supercoiling and transcription regulation. Transcription is the process through which DNA is read by the RNA polymerase (RNAP) to produce an RNA molecule, a single filament of linked nucleotides, which differs from single-stranded DNA by the presence of uracil (U) and ribose, instead of thymine and deoxyribose respectively. Messanger RNA (mRNA) is one of the most important functional transcripts, as it carries the instructions needed to protein synthesis. Here, I mainly focus on the properties of the transcriptional

process in bacteria, where there is only one type of RNAP which catalyse the production of mRNA. In most the prokaryotes about 40% is unconstrained plectonemic supercoiling [35], which is free to relax through the double helix and influence DNA structure. Differently, in eukaryotes both transcription and supercoiling are associated with several structural and functional proteins (such as transcription factors, promoter enhancers etc.); moreover, DNA is chromatinised (a more compact complex of DNA and proteins, such as histones): at this stage, chromatin supercoiling is not well understood.

1.3.1 The transcriptional process

Specific sequences along DNA, called *genes*, encode the genetic information to synthesize proteins. Genes represent the fundamental units of the inheritance process, as they are chiefly used as template to the assembly of RNA molecules. Bacterial genes are preceeded by the promoter, two upstream sequences located at -35 bp and -10 bp from the starting site of the gene (conventionally at the position $+1$). Promoters are directly recognized by bacterial RNAP.

The protein complex responsible for copying and catalysing transcripts is the RNAP; in bacteria, this enzyme has a core made of 4 subunits types: two α polypeptides serve to assemble the RNAP, the β and β' dimers function as binding *loci* for the promoter, and ω, which might act as to preserve the structural integrity of the complex [40]. The RNAP complex is completed by the sigma factor, that is able to recognise and bind specifically to the promoter [35].

The transcriptional process is schematised in Fig. 1.4. During the first stage of transcription, the initiation, RNAP proteins perform multiple attempts to bind the promoter; when the RNAP recognise and strongly binds to the promoter, the so–called *closed* DNA–RNAP complex forms. At this stage DNA is not denatured yet, rather DNA wraps around the polymerase by about 300^o [41], forming a left–handed superhelix. A negative writhe corresponding to a linking difference of $\Delta Lk = -1.25$ is associated with such conformation of DNA. Once the RNAP escapes the promoter (it requires < 10s [42]) the sigma factor is released, and a bubble of denatured DNA appears at the start of the gene: the *open* DNA-RNAP complex forms; a linking difference of $\Delta Lk = -1.7$ is associated with the formation of the open complex.

The second stage of transcription is elongation. The RNAP proceeds along the gene in the direction $5' \rightarrow 3'$, separating the double helix and assembling the new transcript; the translocation along the gene occurs an average speed of about 100 bp/s [19]. During elongation, RNAP rotation is hindered due to its large size and the drag generated by the crowded intracellular environment. Rather, DNA swivels when passing through the RNAP complex, being locally denatured. For every 10 bp which are transcribed, the linking number

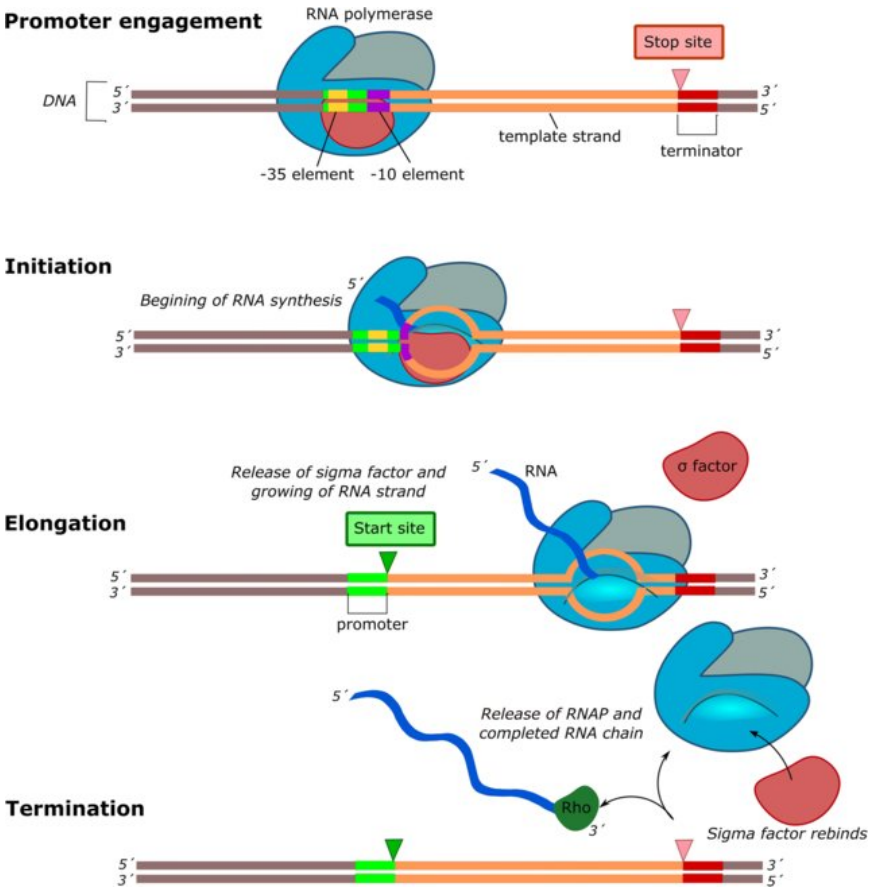

Figure 1.4. Transcriptional process. Schematics of the different stages of transcription. The promoter is engaged by the RNAP complex. The synthesis initiates by splitting the double helix: the nascent RNA molecule duplicate the $3' \rightarrow 5'$ template strand. The elongation starts when the sigma factor is released. During elongation, the RNAP moves along the gene, denaturing DNA at its passage (orange loop), allowing the RNA formation (blue stripe). The transcription ends when the RNAP encounters the termination sequence, and the RNA molecule dissociates from the polymerase. The *rho* factor (colored green) mediates the dissociation of the completed RNA. Figures taken from [39].

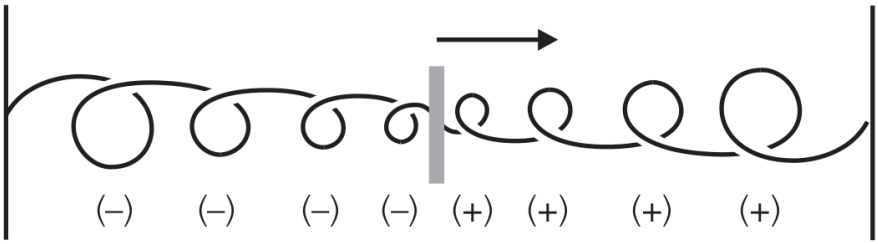

Figure 1.5. Twin supercoiled domain model: single polymerase. The DNA segment is tethered at both terminus, such that the molecule cannot rotate about its axis. Due to the moving polymerase, positive supercoiling accumulates ahead the polymerase, while negative supercoiling accumulates behind it. The arrow indicates the polymerase velocity. Figures taken from [35].

changes by $\Delta\mathrm{Lk} \approx +1$ ahead of the polymerase and $\Delta\mathrm{Lk} \approx -1$ behind it. This observations lead to the *twin supercoiled domain* model [18], which states the following: when an anchored DNA segment (the two terminal ends cannot rotate) is transcribed by a single polymerase, two supercoiling domains form ahead and behind the polymerase, as the RNAP act as a topological barrier (positive and negative supercoiling cannot recombine through the RNAP complex). If a single RNAP moves from left to right, the right domain is overwound, or positively supercoiled, the left domain is underwound, or negatively supercoiled, as shown in Fig. 1.5; conversely, in a circular DNA molecule, such as a plasmid, positive and negative supercoiling can annihilate. In the presence of multiple RNAP, several domains can form during simultaneous transcription; in particular, negative supercoiling accumulates in between two divergent genes (gene orientations points outwards), whereas positive supercoiling builds up in between two convergent genes (gene directions points inwards). For genes oriented in the same direction, or tandem genes, the extent of supercoiling generated in the region within the two moving polymerases vanishes, see Fig. 1.6.

The termination of transcription occurs when the RNAP encounters the terminator, which is a sequence formed by several consecutive A–T base pairs preceeded by two symmetrical sequences, which determine the formation of a stem loop in the nascent RNA chain. The dissociation of RNA from the DNA-RNAP complex is aided by the *rho* factor, which destabilises the RNA–DNA duplex [43].

It is important to note that the formation of the DNA-RNAP closed complex, as well as its conversion in the open complex and the following denaturation of the double helix, does not need ATP. Thus, although the negatively supercoiled state of a sole DNA molecule (or segment) does not correspond to lowest energy conformation (see also section 1.2.1), it represents the best condition

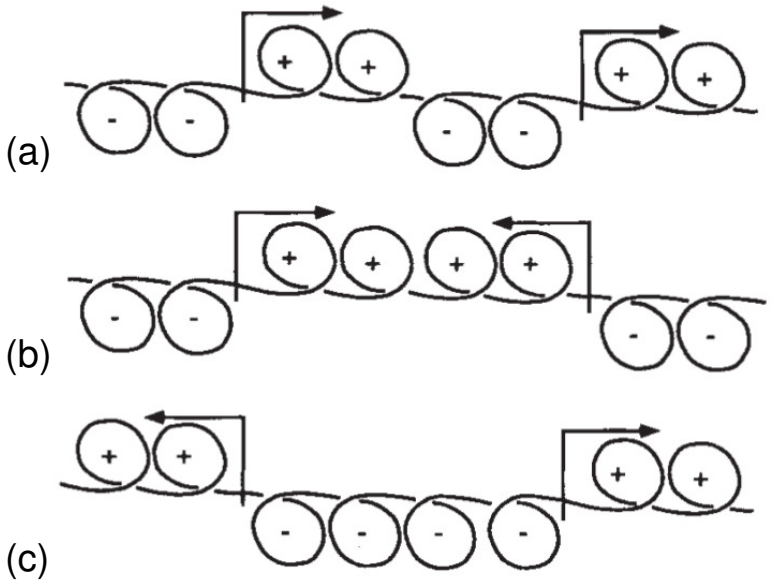

Figure 1.6. Supercoiling domains between a pair of transcribed genes. (a) Tandem genes. Positive supercoiling generated ahead of the left gene counterbalance the negative supercoiling generated behind the right gene. **(b)** Convergent genes. Positive supercoiling accumulates in between the two genes. **(c)** Divergent genes. Negative supercoiling accumulates in between the two genes. Figures taken from [35].

for an efficient transcriptional process. Indeed, bacterial chromosomes and plasmids are constantly held at a negative level of supercoiling. In Ref. [37], Brahms *et al*. show that by increasing the level of negative supercoiling in pBR22 plasmids (*E. Coli*) the transcriptional rate is augmented, reaching its peak at $\sigma_{\mathrm{Lk}} \simeq -0.06$. Nonetheless, higher levels of negative supercoiling are detrimental, favouring the formation of segment of Z-DNA with dramatic effects on the transcription. On the other hand, a supercoiling level equals to $+0.03$, inhibits transcription completely.

1.3.2 Topoisomerases can relax supercoiling generated by transcription

In living bacteria, the optimal level of supercoiling is maintained by topoisomerases. Although the precise microscopic nature of the relaxation of the built–up supercoiling due to topological enzymes is not clear at this time, the link between topoisomerases and transcription is certainly very strict. In bacteria, it is well known that negative supercoils generated behind the transcribing RNAP can be almost immediately removed by topoisomerases I, which binds

the polymerases through its zinc domains [44], to prevent the formation of detrimental Z-DNA segments that would hinder transcription. Indeed, each topoisomerases I can relax 1-–10 supercoils per second on average [45], faster than gyrase, which relaxes at most 0.5–1 supercoils per second [46]. A bacterial RNAP, which transcribes at a velocity of about 50–100bp/s, creates about 5–10 supercoils per second ahead and behind of it. Thus, in principal, a single topoisomerases I may be capable of removing all negative supercoils generated behind the RNAP, whereas gyrases might not be sufficiently fast to eliminate positive supercoils during transcription. However, recall that positive and negative supercoils generated in–between transcribed tandem genes tend to naturally annihilate, as stated by the twin supercoiled domain model. On the other hand, a rough estimation counts only ~ 0.02 active topoisomerases I per gene [14] (see also section 1.4.1). The distribution of topoisomerases along chromosomes is not well known, but it is reasonable to assume that topoismerases I act preferably in gene–rich regions of the DNA, as they relax negative supercoiles. Conversely, gyrases are more likely to bind in–between genes, since they can indirectly remove positive writhe (by introducing negative linking difference), which has been shown to form independently of genes [16, 47].

1.3.3 Transcription as a 'bursty' process

In *E. Coli*, there are 0.2–2 RNAP per gene [48]. The low number of polymerases can be a source of large fluctuations in the single–gene abundance of transcripts [49]; indeed, in the absence of external regulatory factors which induce gene activity or repression, the transcription activity is subject to the availability of free RNAPs. Moreover, the RNAP loading at a promoter is itself a stochastic process: the mechanism through which RNAPs find their specific binding sites (a combination of three–dimensional diffusion through the nucleoplasm or cytoplasm and one–dimensional diffusion along the genome [50, 51]) leads to a broad distribution of search times. After the RNAP engages the promoter, the actual initiation of transcription occurs when the σ factor is releases; the retention of the σ factor determines the so–called *abortive initiation*, that delays the transcription of a gene. Additionally, associated cofactors can catalyse or reprime gene transcription [52]. The random nature of these biochemical reaction contributes to the inherent stochasticity of the transcription initiation, which can take several hours in mammals [53] and tens of minutes in bacteria [6].

Intriguingly, transcription has been proved to be often 'bursty', that is, clusters of closely spaced transcriptional events are separated by longer dormant periods. Long–standing observations have assessed the intermittent nature of transcription in eukaryotes; cycles of high and low transcription activity may

be due to chromatin remodelling [54] and *cis*–regulatory DNA factors [55]. In both cases, refractory periods during which genes are silent last for hours. On the other hand, more recent experiments show that bacterial refractory times are significantly shorter (tens of minutes) and the distributions of inter–events lag–times are more compatible with a random ON $\leftrightarrow$ OFF dynamics of the promoter, also known as telegraph process [56]. In particular, Ref. [6] show striking evidence of transcriptional bursts in a low–transcribed bacterial gene. In this experiment, both the duration of bursts and silent periods are exponentially distributed, with mean values of about 6 min and 37 min, respectively. The measured average burst size (number of transcription in a single burst) was approximately of 2.2.

What is the mechanism that underpins such intermittent switchings between active and repressed state of genes in bacteria? Can supercoiling play an acting role in producing bursts of transcription? Golding *et al*. suggest multiple causes, but only Chong *et al*. [7] identify the interplay of supercoiling dynamics and topoisomerases activity as one of the most relevant causes for bursty transcription in bacteria. Specifically, the introduction of topoisomerases I *in vitro* reduces the transcriptional efficiency of 80%, while, after the induction of gyrases, the transcription rate is slowly restored to its baseline value. They argued that the different rates of supercoil relaxation by topoisomerase I and gyrase yield the intermittent behaviour of gene transcription *in vivo*.

1.3.4 The sequence-size function and burst parameters

A classic model to describe transcriptional bursting is the *interrupted Poisson process* (IPP) [57], which describes transitions between an active (ON) and an inactive (OFF) state with Poissonian rates k_{ON} and k_{OFF}, as in a random telegraph process [56], together with transcription at a constant rate k_i whilst the system is in the ON state. The process can be characterised by the probability distribution function (*pdf*) $f(t)$ of waiting times – the time intervals between two consecutive transcriptional events – which is given by a double-exponential [57], where the two characteristic times are related to the interval between transcriptions in a single burst, and the interval between two consecutive bursts.

To determine whether a system is bursty, we need a tool that is capable to distinguish the dynamics regulated by two timescales, as in the telegraph process, from those regulated by a single timescale, as in a pure Poissonian process. Such tool is represented by the so-called *sequence-size function* (*ssf*) [57, 58], which is defined in terms of the distribution of waiting times $f(t)$:

$$\Phi(\tau) = \frac{1}{1 - \int_0^\tau f(t)dt}. \tag{1.5}$$

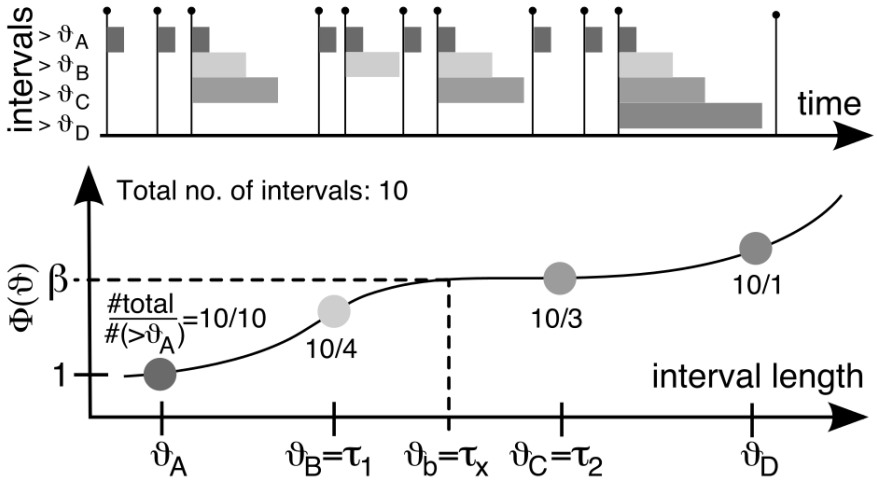

Figure 1.7. Schematics of the sequence-size function In the upper panel, we show a time series of events. Horizontal bars indicate the time interval larger than the thresholds, which are denoted by θ_i, with $i = A, b, B, C, D$. In the lower panel, we show the *ssf* in the presence of two timescales. The value of the ssf is evaluated as the total number of waiting times, divided by the number of those larger than the chosen threshold. The two timecales are $\tau_1 = \theta_B$ and $\tau_2 = \theta_C$, and corresponds with the two inflection points of the ssf. Therefore $\tau_x \equiv \theta_b = \tau_1/2 + \tau_2/2$. Correspondingly, $\beta = \Phi(\tau_x)$. Figure taken from [57].

This is the inverse of the probability of observing a waiting time larger than τ, or, equivalently, the average number of transcriptional events that are grouped into sequences interrupted by intervals longer than τ. If the dynamics is bursty, we expect two well-separated timescales for the decay of $f(t)$; correspondingly Φ will display a plateau and two inflection points. These points, τ_1 and $\tau_2 > \tau_1$, can be found as the zeros in the second derivative of Φ; these values also approximate the two timescales for $f(t)$ (this is not a strict equality, but rather an order-of-magnitude estimate). The value of Φ that equals the average number of transcriptions in a single burst is located between the two timescales, reasonably in the middle point $\tau_x \equiv (\tau_1 + \tau_2)/2$. Therefore, the *burst size*

$$\beta \equiv \Phi(\tau_x), \tag{1.6}$$

yields the average number of transcriptional events in a burst [57]. If the dynamics is not bursty, Φ will have no more than one inflection point. A schematics of *ssf* is presented in Fig. 1.7.

To quantify the "burstiness" of a transcriptional time series, we define the following parameter,

$$\xi = \frac{\Phi'(\tau_1) - \Phi'(\tau_2)}{\Phi'(\tau_1)}, \tag{1.7}$$

which measures the area under $\Phi''(\tau)$ between the two inflection points (when they exist), normalised by $\Phi'(\tau_1)$ so that the result remains between 0 and 1 (prime and double prime denote first and second derivatives respectively). ξ is zero when the dynamics are not bursty, and increases as the separation between the two characteristic timescales τ_1 and τ_2 becomes clearer: we refer to this parameter as the *burst significance*.

1.3.5 From the specific linking difference to the local supercoiling

As specified in section 1.2.1, the specific linking difference, or supercoling density, is a global property of a DNA molecule. However, the level of torsional stress as well as the conformational changes of the double helix can change along the DNA molecule. Key processes are the local denaturation of DNA in the proximity of the RNAP during transcription, and the nucleation and hopping of plectonemes [59, 60]. These phenomena can be triggered locally by biochemical reactions or mechanical stresses, and propagate dynamically through the whole DNA chain. In order to describe the supercoiling dynamics due to local modification of the double helix state, the notion of local supercoiling $\sigma(x,t)$ needs to be defined. To this end, a coarse-grained description is required: a 'block' of consecutive base pairs are clustered and labeled by x, which is a continuous positional variable along the DNA chain. At each time t of the dynamical evolution, the local twist and writhe difference ($\Delta\mathrm{Tw}(x,t)$ and $\Delta\mathrm{Wr}(x,t)$ [61] respectively) are summed up and normalised by the reference linking number Lk_0 of the coarse-grained block. A local version of Eq. (1.2) readly follows:

$$\sigma(x,t) = \frac{\Delta\mathrm{Tw}(x,t) + \Delta\mathrm{Wr}(x,t)}{\mathrm{Lk}_0}. \tag{1.8}$$

Now, the conservation law in a topologically constrained DNA molecule reads:

$$\sigma_{\mathrm{Lk}}(t) = \frac{1}{L}\int_L \sigma(x,t)\, dx \equiv \sigma_0, \tag{1.9}$$

where $\sigma_0 \equiv \sigma_{\mathrm{Lk}}(0)$ is the level of global supercoiling at which the DNA molecule of length L is prepered at the initial time t_0. Analogously, a local version of Eq. (1.4) can be associated with local supercoiling, where the free energy density is given by $\mathcal{F}(\sigma(x,t)) = K\sigma(x,t)^2$ and

$$F\left[\sigma(x,t)\right] = \int \mathcal{F}(\sigma(x,t))\, dx. \tag{1.10}$$

The notions of local supercoiling, twist and writhe allow us to conceive models that shed light on the physical principles which link local topology

to global DNA conformational states, and highlights how it can affect DNA biological function [12–16, 47, 62]. In Ref. [16], a mechanistical description of transcriptional elongation leads to bursting due to polymerase pausing along genes, and correlations between highly transcribed genes. In Ref. [63] an interesting stochastic field theory explains the twist–to–writhe conversion and plectoneme coarsening and hopping [60]. In Ref. [47], a detailed 3D coarse-grained model of circular DNA which embeds a physical transcribing polymerase predicts the local twist and writhe diffusion. In light of this, we will refer to local supercoiling as simply supercoiling throughout.

In the following, I will focus on a model, first introduced in Ref. [12], which is based on the interplay between stochastic transcriptional initiation and supercoiling diffusion. Such model provides a framework within we can understand upregulation of genes, bursty dynamics and supercoiling waves [14].

1.4 A stochastic model of supercoiling–dependent transcription

This model was first described in Ref. [12]. In that work, it was found a crossover from a relaxed regime where transcription is virtually Poissonian, to a supercoiling–regulated regime where transcription of neighbouring genes is highly correlated. In the present work we instead ask whether, and under what conditions, the coupling between supercoiling and transcription can lead to bursty dynamics.

1.4.1 The model

In this work, a circular DNA is modeled as a one-dimensional lattice of length $L = 15$ kbp with spacing $\Delta x = 15$ bp (which is approximately the footprint size of an RNAP). The one–dimensional scalar field $\sigma(x,t)$ denotes the local supercoiling density at a point x on DNA. Supercoiling dynamics are then described in continuous space by

$$\frac{\partial \sigma(x,t)}{\partial t} = \frac{\partial}{\partial x}\left[D\frac{\partial}{\partial x}\sigma(x,t) - J_{\mathrm{tr}}(x,t)\right] - k_{\mathrm{topo}}\sigma(x,t), \tag{1.11}$$

where the three terms on the right-hand side represent diffusion of supercoiling, supercoiling flux generated by transcription, and supercoiling dissipation due to topological enzymes, respectively. The boundary conditions are specified by

$$\sigma(0,t) = \sigma(L,t), \qquad J_{\mathrm{tr}}(0,t) = J_{\mathrm{tr}}(L,t), \tag{1.12}$$

at any time t. The first term in Eq. (1.12) arises from the global conservation of supercoiling; indeed, in a closed DNA loop and in the absence of topological enzymes, the total level of supercoiling σ_0 is conserved, as it is a topological invariant of the system. By imposing the continuity condition over the whole DNA molecule, supercoiling obeys "model B" (conserved) dynamics [64]. Further, as the free energy density of supercoiling $\mathcal{F}[\sigma]$ is quadratic (as explained in section 1.2.1), the chemical potential $(\partial\mathcal{F}/\partial\sigma)$ is linear in σ. Assuming constant mobility, this leads to the diffusion term in Eq. (1.11) [65].

The second term in Eq. (1.11) models the total flux of supercoiling introduced by the transcribing RNAPs. N polymerases can bind at the promoters of n genes – typically we consider a length of $\lambda = 66\Delta x \sim 1000$bp, which is a relevant size for bacterial genes. located at lattice positions y_j, $j = 1, \ldots, n$. Transcription initiates stochastically when an inactive RNAP binds at gene j with rate $k_{\text{in},j}$. The RNAP then elongates with velocity v (positive or negative depending on the direction of transcription) such that at time t_i after initiating it is located at position $x_i(t_i) = y_j + vt_i$ (where the index i labels the RNAP, thus t_i represents the time interval from initiation of the ith polymerase). As soon as the transcribing RNAP leaves the promoter, another free RNAP can bind at it and start a new transciption. Transcription terminates once the RNAP reaches the end of a gene. The total flux is then given by

$$J_{\text{tr}}(x,t) = \sum_{i=1}^{N} J_i(t_i)\delta(x - x_i(t_i))\eta_i(t), \tag{1.13}$$

where the sum is over all RNAPs, $J_i(t_i)$ is the flux generated by the ith RNAP, and the stochastic process $\eta_i(t)$ represents its state, taking a value of 0 if it is unbound, and 1 if it is actively transcribing. The flux of the ith RNAP is racked up in front of the travelling polymerase:

$$J_i(t_i) = J_0\left(1 + \frac{|v|t_i}{\Delta x}\right). \tag{1.14}$$

The sign of J_0 depends on the direction of gene transcription. Note that the integration of $\partial J_{\text{tr}}/\partial x$ over the lattice vanishes, due to the periodic boundary conditions, as required by the twin supercoiled domain model (positive supercoiling generated ahead of the RNAP equals negative supercoiling behind). I will explain the choice of Eq. 1.14 to model the supercoiling flux below.

The initiation rate is a function of the supercoiling at the promoter $\sigma_{y_j}(t)$, where y_j is the starting site of the jth gene.

$$k_{\text{in},j}(t) = k_0 \max\left\{1 - \alpha\sigma_{y_j}(t), 0\right\}, \tag{1.15}$$

where α is a coupling parameter (it represents the sensitivity of RNAP–DNA binding to the level of supercoiling). Thus, the supercoiling dynamics are coupled to the stochastic transcriptional kinetics in Eq. (1.11) through J_0 in Eq. (1.13), while the transcription kinetics are coupled back to supercoiling through the sensitivity α in Eq. (1.15). In the following, we will use the notation $\sigma_{p,j} \equiv \sigma_{y_j}(t)$ to indicate the local supercoiling at the promoter of the jth gene, or σ_p when we refer to the case of a single gene.

The third term on the right of Eq. (1.11) represents the loss of supercoiling due to the action of topoisomerases (such as, e.g., topoisomerase I and II, gyrase). This is introduced in the model in a minimal way, as a first–order reaction where both positive and negative supercoiling relax at the same rate k_{topo}, and with no spatial dependence. In general, this term does not conserve the total supercoiling.

Key model quantities and paramater values

An useful quantity is $\bar{J} = J_0 \left[1 + \lambda/(2\Delta x)\right]$, which represents the average value of the supercoiling flux generated during a transcription event. Then, the adimensional key quantities that control the model behaviour can be defined: $\bar{J}/D$ and k_{topo}/k_0. Lateron, we will explore the ability of the model to exhibit bursty behaviour at different points in the $(\bar{J}/D, k_{\text{topo}}/k_0)$ parameter space, for different gene arrangement cases.

The diffusion constant value has been chosen following Refs. [59, 60], which indicates a value of at most ~ 0.1 kbp^2/s for plectoneme diffusivity, when a DNA molecule is subjected to tensions of less than 1–2 pN. However, macromolecular crowding *in vivo* is likely to further slow down writhe/supercoiling diffusion – so for our simulations we consider values for D which are between ~ 4 and ~ 40 times smaller than the value quoted above. The smaller value was used for more systematic analyses as it enables more efficient simulations – in such way it is possible to use a larger value for the time step. Changing the diffusivity does not change the qualitative behaviour of the system, as long as $\bar{J}$ is changed accordingly.

The typical RNAP velocity in bacteria is ~ 100 bp/s [19], so that the time taken to transcribe a $\lambda = 1$ kbp gene is $\tau \sim \lambda/v \sim 10$ s. Then, through dimensional analysis, we expect an order–of–magnitude estimate for the $\bar{J}$ to be $\sim \lambda^2/\tau$ or $v\lambda \sim 0.1$ kbp^2/s – notably, this is the same order of magnitude as D. Thus, in our simulations we explore a range of values for $\bar{J}/D$ which is typically between 0.34 and 3.4.

Turning now to the dynamics of transcription initiation, measured RNAP initiation rate can vary widely, and typical values are in the range 1 s^{-1} to

1 hr^{-1} are observed (see [12, 66–68]). In *E. coli* there are an estimated 1000–10000 RNAP per cell [48], and about 5000 genes. To reflect this ratio, we take one RNAP per gene in our simulations, unless otherwise stated. The rate of topoisomerase action *in vivo* is equally difficult to estimate. However, we make a rough calculation to choose a reasonable range of value for k_{topo}. Ref. [69] counts ~ 500 topoisomerase I per cell in *E. coli*. Assuming that about half of these enzymes are bound, and that there are two genomes per cell on average, we arrive at $\sim$ 100 topoisomerase proteins being bound per genome, that is $n_{\mathrm{topo}} \sim 0.02$ topoisomerases per gene. Assuming additionally that each enzyme can on average relax 1–10 supercoil per second, and that the baseline bacterial supercoiling is such that $\sigma_{\mathrm{baseline}} \sim -0.05$, we get

$$k_{\mathrm{topo}} \approx \frac{n_{\mathrm{topo}}}{\sigma_{\mathrm{baseline}}\lambda}\frac{\Delta Lk}{\Delta t}, \tag{1.16}$$

that gives $k_{\mathrm{topo}} \sim 0.005$–$0.05\ \mathrm{s}^{-1}$, which is close to the physiological value of the baseline transcriptional rate of bacterial genes [6]. Due to this, and since the values of k_{topo} and k_0 are not known accurately, in our simulations we have systematically varied their ratio k_{topo}/k_0, between 0 and 1.4 (specifically we set $k_0 = 0.001\ \mathrm{s}^{-1}$, and varied k_{topo}).

To determine whether supercoiling can affect RNAP initiation at a promoter, we need to consider the time it takes for supercoiling generated by a previous transcription event to diffuse away, and the typical initiation rate ($k_{\mathrm{in},j}$ in our model). The mean field model described below in section 1.4.2 was used to estimate the extent of residual supercoiling at the promoter. For $k_{\mathrm{topo}} \simeq 0$:

$$|\overline{\sigma_p}| \simeq \frac{\bar{J}}{2D}k_0\tau. \tag{1.17}$$

If this quantity is larger than α^{-1} (see Eq. (1.29), setting $\omega \simeq 0$), then supercoiling can indeed increase the rate at which RNAP binds the promoter, leading to a positive feedback. Experiments in bacterial genes suggest that a supercoiling density of $\sigma_p \leq -0.01$ is sufficient to enhance RNAP binding, and so does the transcriptional process (see [12, 70]); thus, in our simulations we have set $\alpha = 100$. In our simulations, we start with a uniform initial condition $\sigma(x, t = 0) \equiv \sigma_0$, where $\sigma_0 = 0$. As detailed above in section, the supercoiling of bacterial chromosomes is often set at ~ -0.05; it is reasonable to assume that topo I and gyrases play an important role in maintaining such baseline negative supercoiling. On the contrary, by setting $\sigma_0 = 0$ in our model, the sole action of topoisomerases cannot shift the overall supercoiling density to negative values, as it relaxes positive and negative supercoiling in the same fashion. Therefore, the value of σ_0 should be seen as the baseline supercoiling ~ 0.05 in bacteria. Note also that the promoter position coincides with the

starting of gene y_i (a promoter at position $-2\Delta x$ with respect to the gene would be more realistic; however the results obtained for a moving polymerase are qualitatively the same).

Validation of the model

In order to understand the results in the next sections, a discussion on the assumptions of this model is in order.

First of all, observe that the model embeds a single diffusivity D of supercoiling, even if, in principal, one would expect different values of twist and writhe diffusivity; Arguably, twist diffuses very quickly [15], whereas writhe diffusion is expected to be much slower, as it involves more global DNA rearrangements. In ref. [47] simulations on supercoiled DNA rings show different order of magnitudes for twist and writhe diffusivity ($D_{\mathrm{Tw}} \approx 500 D_{\mathrm{Wr}}$). Moreover, twist diffusion is commonly much faster than the typical time scale of RNAP translocation along the gene ($\sim$ 10 s), thus is safe to assume that any twist generated by transcription can spread over the whole DNA domain istantaneously [15]. In a circular DNA, when only one RNAP is transcribing, positive and negative twist recombines, and therefore the net amount of twist vanishes. Moreover, it has been shown that DNA is unable to support much deviation of twist from its relaxed state: the theory in [59] indicates that DNA starts to writhe if the supercoiling density exceeds 0.01, which also corresponds to the supercoiling value that enhances transcription, as mentioned in the paragraph above. This suggests that the slower diffusion of writhe will dominate the supercoiling dynamics, thus, the underlying assumption here is that $D \approx D_{Wr}$.

Another aspect is that some of the mechanical effects of torsion of the double helix is neglected. In particular, it is well known that in some conditions the build up of net supercoiling can cause a drop in RNAP velocity along the gene and, eventually, its stalling [15, 71]. This has been shown in simulations where the DNA domain is bound by the transcribing RNAP and an anchoring barrier; in the absence of topoisomerases, twist spreads over the whole domain and exerts an effective drag on the moving RNAP. Importantly, significant deviations from the initial speed are expected for genes of length $\sim$ 3000 bp and larger [15] and highly transcribed (no supercoiling relaxation between transcriptional events). However, in this model, genes are 1000 bp long, no fixed barriers are present, and, more importantly, there is only 1 active RNAP per gene (on average); therefore, supercoiling can relax between transcription events, which is, as explained later, the condition needed to have transcriptional bursts *á la* Golding [6]. Therefore, to render the model manageable, the transcription velocity v has been kept constant.

A final complication is represented by the diffusion of supercoiling through the polymerase, which is present in our model. In real cells, such diffusion would require the unlikely rotation of the RNAP complex, as RNAP motion can be hindered due to the drag generated by the interaction with other factors which are present in the crowded surrounding [72]. Instead of introducing no–flux moving boundary conditions at the points $x_i(t_i)$, Eq. (1.14) resolves the issue, since it models the ramp–up of the flux as transcription progresses. Even if the diffusion process still competes with the accumulation of supercoiling, the residual diffusive leak can reasonably be taken into account, as to model the possible 'relaxation' of three–dimensional writhes in the proximity of the active polymerases. Additionally, since the dynamics is governed by writhe (twist is promptly converted into writhe), we assume that the amount of supercoiling injected ahead of and behind the polymerase at each time, J_0, is constant, as it is chiefly related to the undertwisting (denaturation) of ~ 10 bp of the double helix.

1.4.2 Mean field theory

The general properties of this model can be extracted by developing a mean field theory. In particular we solve an ordinary differential equation (ODE) with periodic boundary conditions and in the presence of topoisomerases. As said before, the following calculations generalise the mean field presented in Ref. [12].

The mean field approximation relies on the substitution $\sigma(x,t) \to \overline{\sigma(x)}$, where $\bar{\cdot}$ indicates the time–average. We assume that a *static* polymerase ($v = 0$) inject supercoiling flux in $x = 0$, for simplicity [we recall that in our simulation we use a *dynamical* polymerase, travelling along the gene at constant velocity, see Fig. 1.8(a)]. In steady state, for the case $N = n = 1$, Eq. (1.11) reads:

$$\frac{\partial^2 \overline{\sigma(x)}}{\partial x^2} - \frac{\bar{J}}{D} \frac{\overline{k_{\text{in}}}\tau}{\overline{k_{\text{in}}}\tau + 1} \frac{\partial \delta(x)}{\partial x} - \frac{k_{\text{topo}}}{D} \sigma(x) = 0 \,, \tag{1.18}$$

where we have made the mean field approximation

$$\frac{\overline{J_{tr}(x,t)}}{D} \to \frac{\bar{J}}{D} \frac{\overline{k_{\text{in}}}\tau}{\overline{k_{\text{in}}}\tau + 1} \delta(x) \equiv M\delta(x) \,, \tag{1.19}$$

with $\overline{k_{\text{in}}}\tau/(\overline{k_{\text{in}}}\tau + 1)$ the fraction of time that the system spends in the trascribing state, being $\overline{k_{\text{in}}}$ the mean transcriptional rate. Effectively, Eq. (1.18) describes the supercoiling profile generated by a static RNAP that produces continuously the riscaled flux in Eq. (1.19).

As the flux term acts only at $x = 0$, solving the model in the mean field approximation is equivalent to solving the following ODE with the boundary

conditions listed in Eq. (1.12),:

$$\begin{cases} \dfrac{\partial^2\sigma(x)}{\partial x^2} - \dfrac{k_{\text{topo}}}{D}\sigma(x) = 0 & \\ \left.\dfrac{\partial\sigma(x)}{\partial x}\right|_{x=0} = M\delta(x) & \text{Boundary condition} \\ \sigma\left(L/2\right) = \sigma\left(-L/2\right) & \text{Boundary condition.} \end{cases} \tag{1.20}$$

Note that both $\sigma(x)$ and $\sigma(-x)$ are solution of the ODE in Eq. (1.20). Therefore, the unique solution that satisfies Eq. (1.20) must be antisymmetric to fulfill the periodic boundary conditions, and should be continuous at any $x \neq 0$; then $\sigma(L/2) = \sigma(-L/2) = 0$.

The solution of Eq. (1.20) with the appropriate parity and boundary conditions is given by:

$$\sigma(x) = \frac{M}{2}\frac{\sinh\left[\sqrt{\dfrac{k_{\text{topo}}}{D}}\left(\dfrac{L}{2} - |x|\right)\right]}{\sinh\left[\sqrt{\dfrac{k_{\text{topo}}}{D}}\dfrac{L}{2}\right]}\text{sgn}(x), \tag{1.21}$$

where $\text{sgn}(x)$ is the *sign function*. From Eq. 1.22 it can be easily shown that in the limit $k_{\text{topo}} \to 0$ we obtain:

$$\sigma(x) = \frac{M}{2}\left(1 - \frac{2|x|}{L}\right)\text{sgn}(x). \tag{1.22}$$

The term proprtional to $1/L$ is the correction due to the periodic boundary conditions, that disappears for $L \to \infty$, recovering the solution in Ref. [12]. In the limit $L \to \infty$, with finite k_{topo}, we have

$$\sigma(x) = \frac{M}{2}\exp\left(-\omega|x|\right)\text{sgn}(x). \tag{1.23}$$

In Eq. (1.23), the quantity $\omega = \sqrt{\dfrac{D}{k_{\text{topo}}}}$ represents the inverse length scale over which supercoiling-mediated regulatory interactions are screened.

The validity of this mean field theory can be determined by comparing it to the time–average supercoiling profile in our single gene simulations. In Fig. 1.8(b), we show the average supercoiling profile at different $\bar{J}/D$ values, for $k_{\text{topo}}/k_0 = 0.2$.

We also aim to locate the crossover point between two different regimes of transcriptional activity. To this purpose, we find a self-consistent equation for

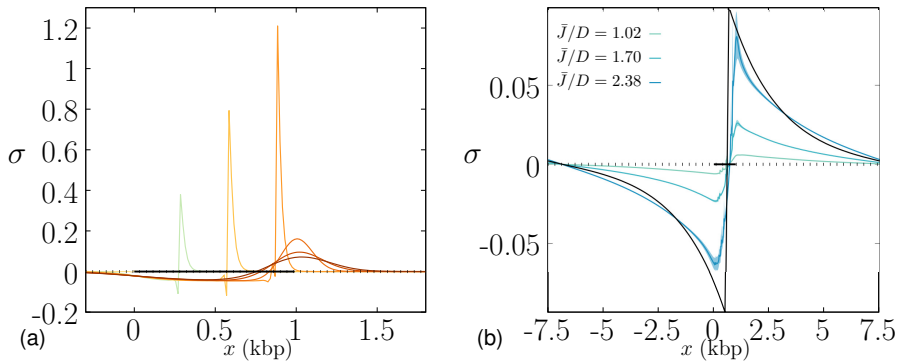

Figure 1.8. Supercoiling profiles. (a) Schematic of the supercoiling density generated by the progression of a transcribing RNAP along a gene, at different time intervals from initiation ($t_i = 3, 6, 9, 12, 15, 18$ s). The RNA polymerase creates positive supercoiling ahead of the gene, while it generates negative supercoiling behind. The supercoiling profile is obtained by solving Eq. (1.11) with $\bar{J}/D = 1.02$, $k_{\text{topo}}/k_0 = 0$ and other parameters as listed in section 1.4.1. The horizontal black segment indicates the location of the transcribed gene. **(b)** Supercoiling profiles at different $\bar{J}/D$, for $k_{\text{topo}}/k_0 = 0.2$, averaged over 251 frames every 1000 s. Black solid line represents the supercoiling profile predicted by the mean field in Eq. (1.21) for $M = 0.1$, with the substitution $|x| \to |x - 2\lambda/3|$.

the transcriptional rate $\overline{k_{\text{in}}}$, by inserting Eq. (1.21) calculated at the promoter y_i in Eq. (1.15). Using the same procedure shown in [12, 13] we find:

$$\overline{k_{\text{in}}} = \frac{h(x) + \sqrt{h^2(x) + 4k_0\tau}}{2\tau}, \tag{1.24}$$

where

$$h(x) = k_0\tau\left(1 + \frac{\alpha\bar{J}}{2D}\frac{\sinh\left[\omega(L/2 - |x|)\right]}{\sinh\left[\omega L/2\right]}\right) - 1. \tag{1.25}$$

The crossover can be found by imposing $h(x) \sim 0$, for which k_{in} starts to be significantly affected by the supercoiling (in particular, for $h > 0$, k_{in} increases much faster with $\bar{J}D$). Thus, if $\bar{x}$ is the distance of the static polymerase from the promoter, it follows:

$$\frac{\sinh\left[\omega L/2\right]}{\sinh\left[\omega(L/2 - \bar{x})\right]} = \frac{\bar{J}}{2D}\alpha k_0\tau. \tag{1.26}$$

Posing $A \equiv L/2$, $B \equiv A + |x|$, if $\omega A < \omega B \ll 1$ (small screening length), we can expand (1.26), and, under the condition $k_0\tau \ll 1$, we obtain

$$\frac{\bar{J}}{D}\alpha k_0\tau = 2 \cdot \frac{A + A^3\omega + o(\omega^2)}{B + B^3\omega + o(\omega^2)} \simeq \frac{2}{\left(1 + \frac{2\bar{x}}{L}\right)} + \frac{L\bar{x}}{2}\omega. \tag{1.27}$$

If $L \to \infty$, we should rewrite $h(x)$ using (1.23). We find

$$h(x) = k_0\tau \left[1 + \frac{\alpha \bar{J}}{2D}\exp(\omega \bar{x})\right] - 1. \tag{1.28}$$

In the limit $\omega \ll 1$, we have

$$\frac{\bar{J}}{D}\alpha k_0 \tau \simeq 2 + 2\omega\bar{x}\ . \tag{1.29}$$

Interestingly, simulations show that the point along the gene at which the time-averaged supercoiling profile crosses zero is $\sim 2\lambda/3$, independently of the parameters used (see Fig. 1.8(b)). The correct mean field profile of supercoiling for a moving polymerase is then computed by substituting $|x| \to |x - 2\lambda/3|$.

1.4.3 Numerical results

Simulations of the model described in Section 1.4.1 have revealed a key role of supercoiling in regulating gene expression. In Ref. [12], the level of transcription was monitored at different values of the relevant parameters, $\bar{J}/D$ and k_{topo}. In terms of the overall transcriptional rate, the model entails a crossover between two distinct regimes, as predicted by the mean field. For any kind of gene geometry (single gene, tandem genes, divergent/convergent genes), when the supercoiling flux created as a polymerase transcribes a gene is small ($\bar{J}\alpha k_0\tau/D \lesssim 2$), transcription is poorly correlated. When this flux is large ($\bar{J}\alpha k_0\tau/D \gtrsim 2$), the dynamics become highly correlated. For tandem genes, supercoiling drives transcriptional waves and regulates gene expression, promoting the transcription of genes which have a larger separation from their upstream neighbors. When considering genes with random orientations, transcription localizes at divergent gene pairs, which are highly upregulated. As expected, the action of topoisomerases, which locally relax supercoiling, downregulates transcription.

Another intriguing property of this model has been recently studied in Ref. [13]. In particular, if the baseline supercoiling is equal or larger than $\sigma_c = 1/\alpha$, the system undergoes a phase transition between an active transcribed phase and an absorbing phase where genes are "switched off" due to the background positive supercoiling.

In the following, I extend the analysis to focus on the stochastic nature of supercoiling dynamics, which in turn can display a 'bursty' behaviour, as shown in Fig. 1.9(a). We consider two different situations: (i) a single gene which can be transcribed by a single polymerase, in the presence of toposimerases, and (ii) 10 genes can be transcribed by 10 polymerases, with no topoisomerases.

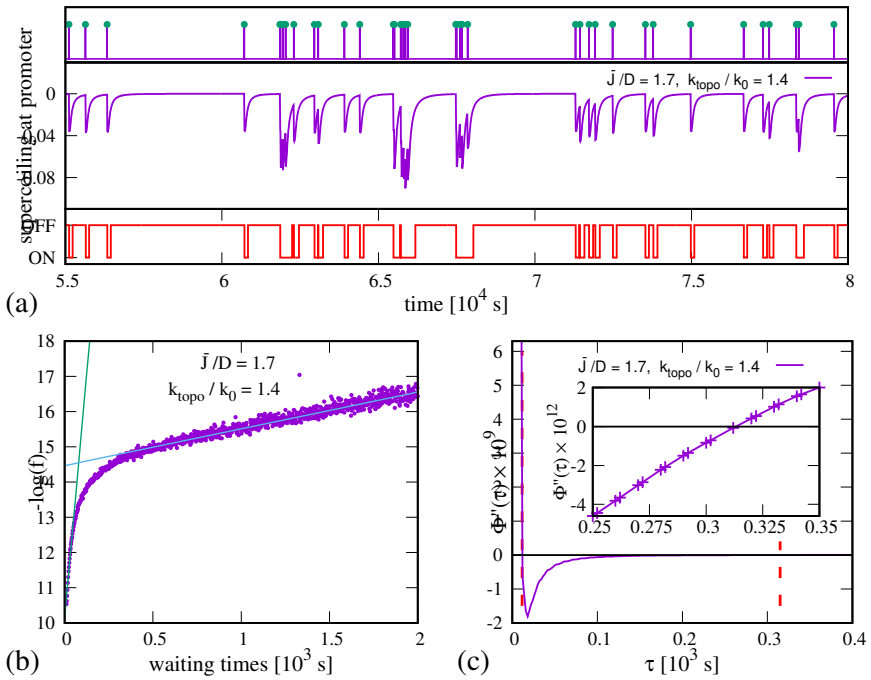

Figure 1.9. Transcriptional bursts and the sequence-size function. (a) Time series for a simulation with $\bar{J}/D = 1.7$, $k_{\text{topo}}/k_0 = 1.4$. Transcription events (top) are often grouped in bursts. Transcription initiation depends on the level of supercoiling at the gene promoter (middle). The state of the system (bottom) is defined as OFF if $\sigma_p \geq (1 - (k_0\tau_2)^{-1})/\alpha$, and ON otherwise. **(b)** Negative logarithm of the *pdf* of waiting times. The existence of two linear regions characterises the dynamics as bursty. **(c)** Second derivative of the *ssf*: the existence of roots at τ_1 and τ_2 demonstrates the presence of two timescales. Inset: zoom of the intersection of this second derivative with the x-axis close to τ_2, which is the intersection point. Red vertical dashed lines locate the position of τ_1 and τ_2.

We performed $4 \cdot 10^3$ indipendent simulations to sample the stationary distribution of waiting times, for each value of $\bar{J}/D$. Each simulation was equilibrated for $2 \cdot 10^6$ timesteps ($dt = 0.01$s), and run for another $2 \cdot 10^7$ timesteps to collect the data. For the case with 10 genes, the placement of gene along the lattice was chosen at random before running each q1 simulations.

A single gene

A clear signature of bursty behaviour can be found in the time series of supercoiling at the promoter, in a certain range of the relevant parameters. For instance, for $\bar{J}/D = 1.7$ and $k_{\text{topo}} = 1.4$, transcription events are grouped in bursts, and, correspondingly, the supercoiling at the promoter follows approximately a two–state dynamics, or telegraph process (bottom panel in Fig. 1.9(a)).

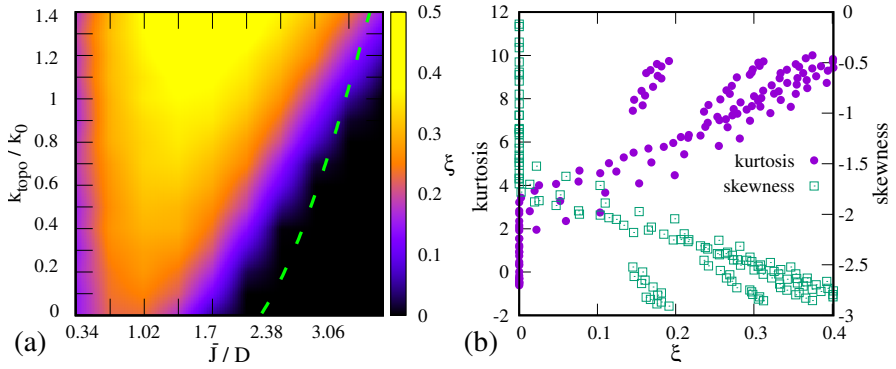

Figure 1.10. Burstiness for a single gene (a) Phase diagram showing burst significance as a function of model parameters. The burst significance ξ defined in (1.7) is plotted as a function of $\bar{J}/D$ and k_{topo}/k_0. A *non-bursty* regime is indicated by $\xi \sim 0$ (black) while $\xi > 0$ indicates a *bursty* regime (yellow-red). The dashed green line is the boundary predicted via mean field (see section 1.4.2). Note that for $\bar{J}/D = 0$, the dynamics is clearly non-bursty, as in the absence of supercoiling flux we have $\sigma = 0$ everywhere, and thus the transcriptional process is a Poisson process of rate k_0. As $\bar{J}/D$ increases, the dynamics become more bursty, with a maximum value at intermediate $\bar{J}/D$. For large values of $\bar{J}/D$ the dynamics is non-bursty again. **(b)** Non-Gaussian parameters of the distribution of the supercoiling at the promoter σ_p as a function of ξ, for different values of $\bar{J}/D$ and k_{topo}/k_0. The kurtosis (skewness) is correlated (anticorrelated) with ξ.

An analysis of $f(t)$ and $\Phi(t)$ shows that, for sufficiently large values of k_{topo}, $f(t)$ has two characteristic timescales (Fig. 1.9(b)), and $\Phi(\tau)$ has two well defined inflection points (Fig. 1.9(c)). Correspondingly, the distribution of supercoiling at promoter σ_p is bimodal [14]. For small k_{topo} and high values of the flux, there are no well defined timescales in $f(t)$ or inflection points in $\Phi(\tau)$; the distribution of supercoiling at promoter σ_p is unimodal and a direct inspection of the transcriptional dynamics show this is not bursty [14].

By computing ξ from simulations with different values of $\bar{J}/D$ and k_{topo}/k_0 we find two distinct regimes (Fig. 1.10(a)): the *non-bursty* regime, identified by $\xi = 0$, and the *bursty* regime, for $\xi > 0$ (a mean field theory gives a good prediction of the boundary between the two). In the region where the burst significance has the largest value ($k_{\text{topo}}/k_0 \sim 1.4$, $\bar{J}/D \sim 1.5$) the burst size is between 2 and 3, close to that measured in *E. coli in vivo*. Estimates of the other bursts parameters – burst duration and OFF–state duration – are in good agreement with experimental results [14]. We note though that the burst size depends on the model parameters: the system can produce bursts of significantly more than 2–3 events (at most ~ 10 on average in our simulations). However, this only occurs in the transition region between the non–bursty and bursty regimes (not shown). In the transition region, burst significance is smaller, which means that the separation between timescales is less marked.

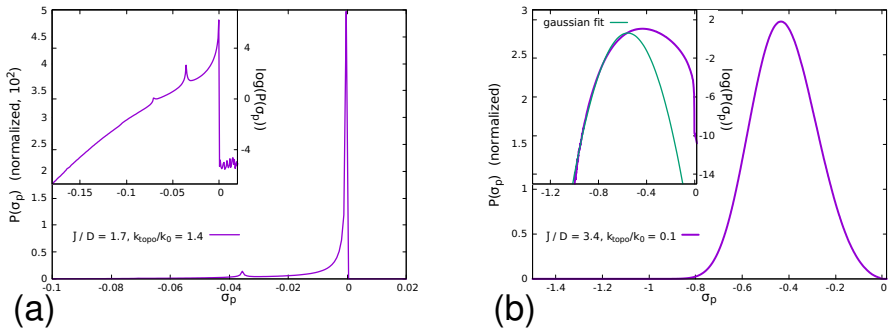

Figure 1.11. Probability distribution for supercoiling at promoter. (a) In the bursty phase, supercoiling at promoter is strongly peaked at $\sigma_p \sim 0$. Another clear peak appears for more negative value of supercoiling, due to occupation of the ON state. Inset: log-linear plot of the pdf in the main panel. (b) In the non-bursty phase the distribution is unimodal, with one Gaussian tail. The gene tends to more often be in a state with less negative supercoiling for longer time; this results in a non-Gaussian positive tail, with a nonzero kurtosis.

The results in Figure 1.10(a) also show that when the positive feedback between supercoil generation and transcription initiation is strong (for large $\bar{J}/D$ and $k_{\text{topo}} = 0$, identified as the *supercoiling-regulated regime* of Ref. [12]), the dynamics are never bursty; bursts are most significant when this feedback is much weaker (but non-zero). The reason for this seemingly surprising result is that if supercoiling upregulates transcription too much, the gene is essentially always on and the transcriptionally silent state is absent.

Our results show that topoisomerases action favours burstiness. In other words, although the dynamics can be bursty for $k_{\text{topo}} = 0$, burst significance is larger when $k_{\text{topo}} \neq 0$.

As bursting is generally due to switching back and forth between two transiently stable states, it is natural to ask whether there are any signatures of bistability in the stochastic transcriptional process we simulate. One such signature can be obtained from moments of the distribution of supercoiling at the promoter σ_p. For non-bursty behaviour, σ_p exhibits close–to–Gaussian fluctuations about an average value, whereas, for bursty transcription, this distribution is more markedly non-Gaussian and bistable (not shown). Quantitatively, burst significance correlates with the magnitude of non–Gaussianity parameters such as kurtosis and skewness (Fig. 1.10). Skewness and kurtosis are defined by means of the third and the fourth moments of a given distribution. With X a stochastic variable and μ_X and $\text{Var}(X)$ its mean and variance, skewness and kurtosis are defined as follows:

$$\text{Skew}(X) = E\left[\left(\frac{X - \mu_X}{\sqrt{(\text{Var}(X))}}\right)^3\right], \tag{1.30}$$

$$\mathrm{Kurt}(X) = E\left[\left(\frac{X-\mu_X}{\sqrt{(\mathrm{Var}(X))}}\right)^4\right] - 3. \tag{1.31}$$

A Gaussian distribution has $\mathrm{Skew}(X) = \mathrm{Kurt}(X) = 0$. For completeness, In Fig. 1.10(a,b) we present the typical probability distribution of supercoiling at the promoter, respectively in the bursty and non-bursty phases. Within our stochastic model, the supercoiling at promoter is directly linked to the probability of initiation, and therefore its distribution encodes all of the information about the process. As expected, for bursty dynamics we observe a bimodal distribution of σ_p, while for non-bursty dynamics we have a unimodal distribution, with fluctuations approximately Gaussian.

Multiple genes

We now switch to the study of arrays of 10 genes (10 polymerases). We consider only the case without topoisomerases (i.e. $k_{topo} = 0$).

In Figure 1.12 we consider the supercoiling–coupled transcriptional dynamics within an array of tandem – same direction of transcription – genes. We find that bursts are typically more significant than in the single gene case. For instance, for $\bar{J}/D = 1$ and $k_{\mathrm{topo}} = 0$, the single gene case was only weakly bursty, whereas for an array of 10 tandem genes the same parameters give rise to bursting which is around twice as significant. In the former case, indeed, we find $\xi \sim 0.23$, a mean burst size of ~ 1.53 and duration of bursts of ~ 1.5 min, whereas in the latter case $\xi \sim 0.4$–0.5 for the most bursty genes, burst size is ~ 2 and duration is ~ 3–4 min [14]. This is because transcription generates positive supercoils ahead of a gene, which act to down–regulate its downstream (right) neighbour [whilst upregulating the upstream (left) neighbour]. As a result, some genes may be transiently "switched off" – this can be appreciated, for instance, by inspecting the time series of supercoiling at the promoter, which at times can take sufficiently positive values such that $k_{\mathrm{in}}(t) = 0$. Just as in the case of a single gene with $k_{\mathrm{topo}} \neq 0$, the activity of each gene is effectively described by a two–state dynamics (ON↔OFF), and bursts can occur (Fig. 1.12(a)). As expected, the *pdf* of waiting times is well described by a double–exponential (see Fig. 1.12(b)), and $\Phi''(\tau)$ displays two zeros (see Inset Fig. 1.12(b)).

As for the single gene case, in the multi-gene system bursting does not occur for large values of $\bar{J}/D$. In this regime, the supercoiling-mediated intergenic interactions instead give rise to transcription waves which travel in the opposite direction to transcription (Fig. 1.12(c,d)). Transcription waves arise because transcription of a gene upregulates its upstream (left) neighbour: as a consequence, transcription of gene i is followed by that of gene $i-1$, then

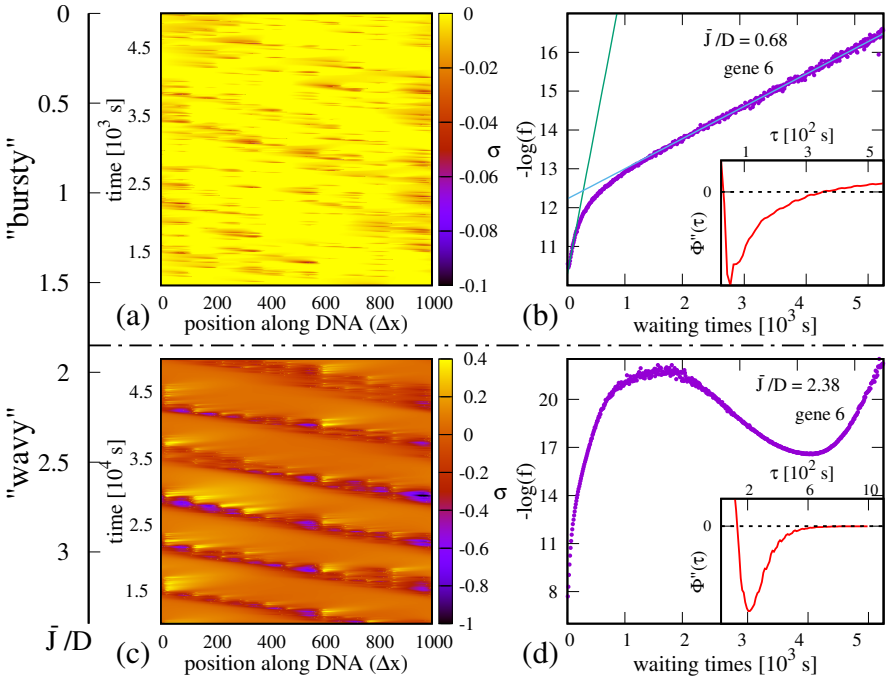

Figure 1.12. Bursty and wavy regimes for an array of 10 tandem genes. (a) Kymograph in the bursty regime ($\bar{J}/D = 0.68$). For clarity, we only show the negative supercoiling range. There are correlations between neighbouring genes, but no clear periodic pattern. Note that genes 1 and 10 turn red more often [compare to genomic map in Fig. 1.14(a)], since they are slightly upregulated. **(b)** Plot of the *pdf* of waiting times in the bursty regime, showing the emergence of two timescales. Inset: second derivative of the *ssf*, which displays two zeroes. **(c)** Kymograph in the supercoiling-regulated regime ($\bar{J}/D = 2.38$). For high values of the flux, the bursty dynamics are replaced by transcription waves. **(d)** *Pdf* of waiting times in the supercoiling-regulated regime. The new timescale associated with the wave modifies the shape of the distribution, giving rise to a "bump" at $\sim 4{\times}10^3$ s (local minimum in -log plot). Inset: second derivative of *ssf*. In the physically relevant range of waiting times ($\tau \lesssim 10^3$ s) the function asymptomatically approaches 0 without crossing the axis ($\tau_2 \to +\infty$).

$i-2$ and so on. We find that the wave velocity is $v \sim D/l$, independent of k_0 (Fig. 1.13(a), inset; $l = 100\Delta x$ is the mean separation between promoters in our simulations). Given our parameter choice, the wave speed is between 0.5–3.0 bp/s, and the time needed to trigger activity of the neighbouring upstream gene is between 6–12 minutes. The scaling of v can be understood by assuming that supercoiling propagates diffusively over the distance l between a gene and its upstream neighbour – simulations show that the prefactor in this relationship is slightly larger than 1. When in the "wavy" regime, the system can no longer be mapped onto a telegraph–like process, and bursts are no longer observed (accordingly $\xi = 0$, as $\Phi(t)$ does not have two inflection points, see

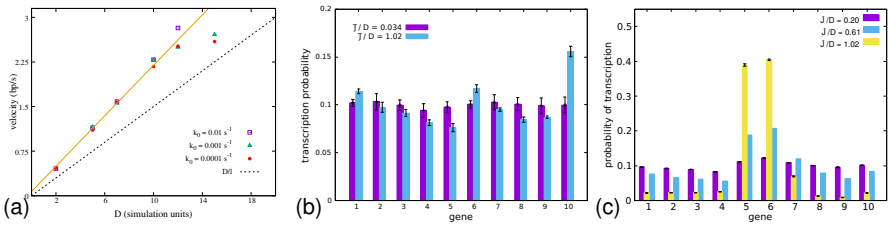

Figure 1.13. Wave velocity and transcription probabilities. (a) Wave velocity for different values of D ($\bar{J} = 25.5$) and k_0, so that we span a large range of $\bar{J}/D$ values, between 1.7–12.7. Values of D are given in simulation units (i.e., in units of $\Delta x^2/\Delta t$). Simulation data are well fitted by a straight line (orange), whereas the simple scaling theory discussed in the text underestimates the data slightly (dashed black line). **(b)** Tandem genes: histograms showing gene transcription probabilities in the bursty regime ($\bar{J}/D < 2$ (average over 10 runs). For sufficiently large values of $\bar{J}/D$, genes 1,6 and 10 are slightly upregulated. **(c)** Divergent genes: in the bursty regime the divergent pair of genes (5 and 6) are upregulated.

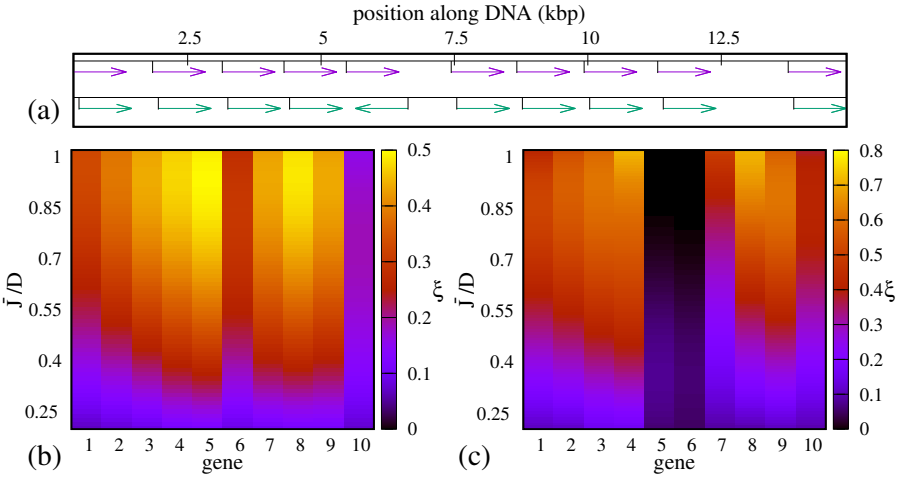

Figure 1.14. Burstiness for a multiple genes arrays. (a) Map of gene positions for tandem (top) and divergent geometry (bottom). **(b,c)** Plot of ξ for tandem geometry (b) and divergent geometry (c) [geometries used are in (a)]. The range of $\bar{J}/D$ (0.06–1.02) is chosen so that the system is in the bursty regime.

Fig. 1.12(d)).

Transcription waves only arise for arrays of tandem genes, and do not occur (or do so only transiently) for genes of differing orientation. In that case transcription-generated supercoiling upregulates pairs of divergent genes at the expense of other (convergent or tandem) genes which are down-regulated; in Fig. 1.13 we also show the transcription probability of genes for tandem genes (panel (b)) and in the presence of a pair of divergent genes (panel (c)). The transcription probability starts to differ as soon as the value of the flux is large enough to give rise to supercoiling mediated interaction (positive feedback

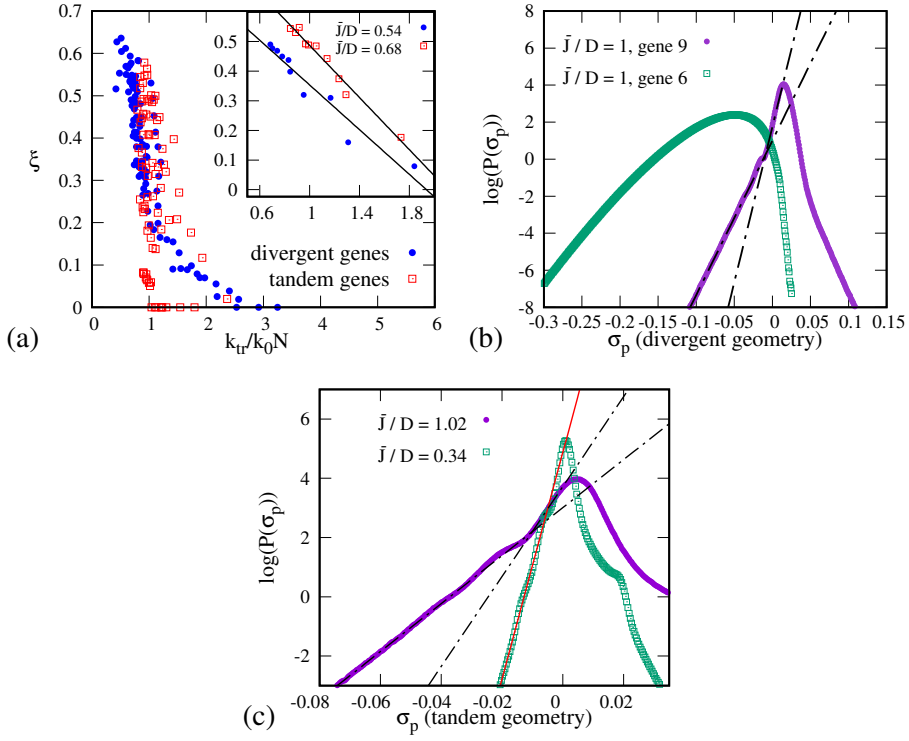

Figure 1.15. Bursty and wavy regimes for an array of 10 tandem genes. (a) The burstiness, ξ, is plotted for different value of the overall time-averaged transcription rate $k_{\rm tr}$ for tandem (red squares) and bidirectional (blue circles) genes (each point represents a gene; points are shown for different values of $\bar{J}/D$ between 0.06–1.02). Inset: for a given value of $\bar{J}/D$, ξ depends linearly on $k_{\rm tr}$ (each point represents a gene). **(b)** Log-linear plot of the distribution of σ_p. For bursty genes (gene 9, which is not part of the divergent pair) this distribution shows a singularity or bump. Such a singularity does not appear for highly transcribed genes (gene 6, in the divergent pair). **(c)** For a given gene (gene 4, tandem geometry) the slope of the negative tail of the distribution depends on $\bar{J}/D$. For small ξ (small $\bar{J}/D$) the bump disappears.

loop) between the two divergent genes.

In Fig. 1.14(a) we show the maps of gene positions for tandem array and in the presence of a divergent pair. In the former case, genes 1, 6 and 10 are upregulated by supercoiling as they have a larger space upstream of them, so are less affected by the repressive action of positive supercoils generated at their upstream neighbour. This occurs, albeit to a much lesser extent, even in the relaxed regime. As we will see, this relatively small upregulation is sufficient to yield a sizeable change in the burst significance. Since for relatively high value of the flux ($\bar{J}/D \sim 1$) the transcription across all genes is almost totally dominated by the pair of divergent genes, we find that the latter behave like a single upregulated gene. In terms of burstiness, this can render the situation qualitatively closer to that of a single gene. For either kind of gene

orientation, burstiness is gene–dependent, and the values of ξ for different genes are substantially different from each other (Fig. 1.14). We find that ξ is anticorrelated with the overall transcription rate, so that the genes which are expressed more (e.g. genes 1, 6 and 10 in the tandem setup or genes 5 and 6 in the divergent setup, compare with Fig. 1.13(b,c)) are less bursty (Fig. 1.15(a)). Burst size is also different in the tandem and bidirectional setups, being substantially smaller in the latter case ($\beta \lesssim 2$, not shown). Burst sizes measured experimentally in E. Coli [6] are closer to the value for single or tandem genes ($\beta \simeq 2.2$) – this is reasonable, as they refer to the transcription of operons which are normally made up from tandem genes controlled by a single promoter.

An analysis of the distribution of supercoiling values at the promoter shows that these differ qualitatively for the cases of multiple genes and a single gene. Unlike in the single gene case, the non–Gaussian parameters for the distribution of supercoiling at the promoter now only weakly correlate with the burst significance [14]. This is because the supercoiling-mediated interaction between genes give rise to non-Gaussian fluctuations even for non-bursty genes. Nevertheless, for both the tandem and divergent gene cases, bursting leaves a detectable signature in the tails of the distribution.

For the bursty transcription case, there is a singularity or a bump, while for non-bursty transcription the curve is smooth, as shown in Figure 1.15(b) for a divergent geometry and in Figure 1.15(c) for a tandem array. The singular point is located at $\sigma_p \approx (1 - (k_0\tau_x)^{-1})/\alpha$, where $\tau_x = (\tau_1 + \tau_2)/2$.

1.5 Conclusions and future perspectives

In summary, we have studied the occurrence of transcriptional bursts in a nonequilibrium model for supercoiling–regulated transcription, first introduced in [12]. For an isolated gene, we found that significant bursting occurs primarily in the presence of topological enzymes. This scenario is an important starting point for our model, as it can be relevant to the experimental investigation in Ref. [6] where the transcription of a gene on a bacterial plasmid was monitored. Indeed, we note that in the region of parameter space where bursts are most significant, the properties of the bursts generated in the model (size, duration, and interburst time) match with those found experimentally. However, it is also of interest to consider the case of multiple genes. This is because gene density is variable both across organisms and within genomes: for instance, in both yeast and bacteria gene density is high so that transcription is likely to affect neighbouring genes. This is also relevant for understanding synthetic DNA constructs containing multiple genes, which can

be used in biotechnology applications [73, 74]. Notably, topoisomerase action is not required for highly significant bursting in gene clusters because there, supercoiling can mediate transient inhibition of the neighbors of highly active genes. We considered both tandem and bidirectional gene geometries, and we found that the existence of bursting is intimately related to non–Gaussian fluctuations of the distributions of supercoiling at promoter. In conclusion, our model suggests that supercoiling can be a relevant factor that drives bursty transcription for non–highly transcribed genes, which could be crucial for explaining phenotypic diversity, especially in bacteria.

The model can be further improved by adding some important ingredients. In highly transcribed genes, the mechanical effects due to double helix torsion could be relevant, as highlighted in the theoretical works in [15, 16]. Therefore, in order to use this model for highly transcribed genes, operons and in the presence of topological barriers which prevent twist relaxation, a separate treatment of twist and writhe could be needed. In this respects the model described in [63] might represent a good starting point. A more complicated scenario can arise by modelling local action for topoisomerase I and gyrase, and different relaxation rates.

Chapter 2

Epigenetics: a route to overwrite cell fate

The word *epigenetics* refers to all those mechanisms which stands in between of genotype and phenotype [75]. In multicellular organisms, identical cells – which possess the same genome – can specialise [19, 76]; recently, it has been found that epigenetics plays a vital role in transitions between from unicellular to multicellular states of *Amoeba* [77]. The primary concept is always the same: cellular evolution, differentiation and specialisation are determined by 'enviromental' factors, which are non–genetic. In the early 40s, Waddington wrote in his famous paper: 'it is possible that an adaptive response can be fixed without waiting for the occurrence of a mutation' [76], which basically states the possibility that an interaction of the type individual–environment can develop diverse phenotypes *without* an actual modification of the genome. This is much in line with current knowledge: in modern terms, the *epigenome* represent "the biological interface between individual and environment"; in this respect, the epigenetic mechanisms can provide a key interpretation of phenotypical variability.

The contemporary view of epigenetics is often paired up with two concepts: "heritability" and "bistability" [78]. The key idea is that every cell should encompass mechanisms to guarantee the maintainance of some constitutive epigenetic traits, but also the ability of "adapting" to external stimuli through time. In the past, much effort has been spent to identify the factors that contribute to inheritance and phenotypical diversity. Four epigenetic mechanisms have been recoginsed [79]:

1. *DNA methylation*: the modification of dinucleotides CG 'islands' (regions where cytosine nucleotide is followed by a guanine along the same strand) which converts cytosine in 5–methilcytosine. After replication, the DNA methyltransferase recovers the methylation 'marks' in the duplicate, guaranteeing the inheritance in daughter cells. While widely present in plants

and mammals, such epigenetic modification is absent in yeast.

2. *Histone modifications and chromatin remodelling*: Different level of chromatin compaction and several chemical modifications of the histone tails contribute to a delicate equilibrium between enzymes. Indirectly, this epigenetic mechanism is responsible for regulating compactness of chromatin, which is closely related to chromatin function, since it determines the accessibility of transcriptional factors and RNAP. This plays a vital role in terms of chromatin functions: in fact, domains of compact chromatin are often transcriptionally silenced. Histone modification is the main object of study of this chapter.

3. *DNA–binding proteins*: function as transcriptional factors and modulate the gene expression.

4. *Non–coding RNA*: small RNA molecules, the so–called *small interfering RNA* (siRNA), or *microRNA*. Once produced, they can recruit the DNA-methyltransferase in the proximity of the gene, which catalyses methylation.

How do these mechanisms influence the phenotype? What is the underlying physics? Perhaps, the most suggestive interpretation (not definition) was given by the same Waddington in the 1957 [80]. In Fig. 2.1, the small marble represents an organism's phenotype, that, at the beginning, is in a pre–determined 'valley', that depicts the only available biophysical state (as it is the case of stem cells); then, it faces a series of viable developmental options, or epigenetic states, represented by multiple valleys. Although quite old, this picture is appealing to a physicist's eye. It mimics quite remarkably the omnipresent free energy landascape which leads to the mechanism of *spontaneous symmetry breaking* (SSB), through which several physical system can choose their fate between "equivalent" equilibrium states in their first steps of evolution [81]. From a physical standpoint, a system described by a Hamiltonian which is invariant (or symmetric) with respect a certain trasformation group, has multiple coexisting phase. SSB occurs when somehow the system is forced to select one, performing an "asymmetrical" choice. Such "symmetry breaking" can be induced by the initial condition, an external drive or a spontaneous fluctuation. Similarly to what happens in a ferromagnet, which can be described by Ising, or Potts models, a population of identical cells can exhibit different characters, as each cell can "adapt" to one of the multiple stable states [82]. In general, the most interesting aspect is to investigate how an epigenetic domain can form, spread, and stabilise, or under which conditions a chromatin segment can switch between different 'epigenetic' states. This leads to the study of laws that rule the spreading of histone modifications, or more generally epigenetic marks.

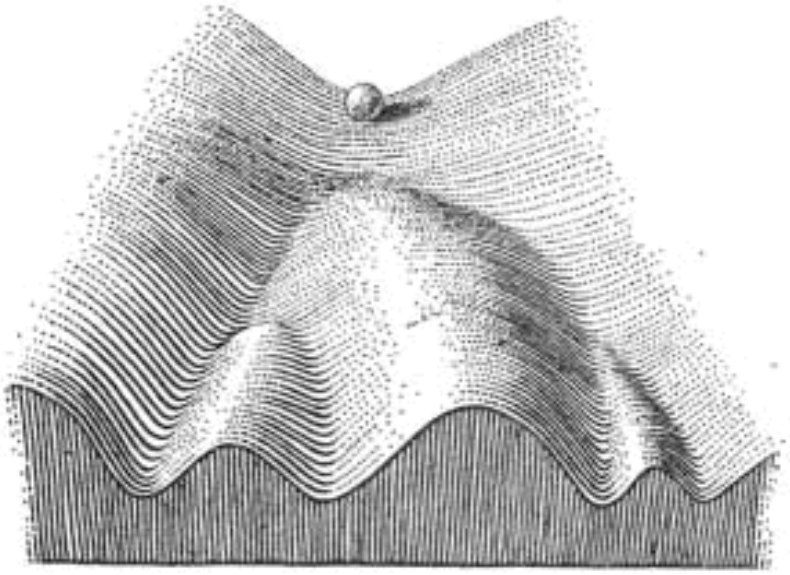

Figure 2.1. Waddington landscape. This scheme represents the so–called *chreodes*, that are the biological paths (or trajectories) of development. This landscape is reminescent of a time-dependent energy landscape, with the z axis the level of some kind of 'energy', the y axis time, and the x axis some representative variable (or variables) that describe the state of the system. At each 'developmental' time (starting from the position of the bead), the bead can choose its path, which is likely to pass through the 'energy' minima on the landscape. Figure taken from [83].

Several models have been conceived in the last decade, all trying to better understand the basis of epigenetic memory and bistability [82, 84–91]. These previous works have typically considered the interplay between two competing marks, that give rise to a competition which may in turn yield bistability between a globally active and inactive state [84, 91] via SSB. Such systems can then retain memory of their (active or inactive) state even in the presence of an external perturbation – such as the loss of half of the marks during DNA replication [82]. We highlight, though, that one-dimensional models exhibiting bistability in this way typically require the addition of one or more intermediate states [84, 86]. In this framework, our work tries to propose a reasonable mechanism for the estabilishment, spreading and switching of epigenetic patterns. Here, the novel elements are represented by: *(i)* the parametrisation of the extent of long–range contacts of the chromatin fiber, through the *contact exponent*; *(ii)* the definition of a specific mechanism for the *erasure* of a single histone modification.

This chapter is organised as follows. In section 2.1 I describe the bead–on–a–string structure of chromatin; in section 2.2, an overview of different post–translational histone modifications, describing how they can be deposited on the histone tails and removed. Then, I will focus on the role of epigenetic bistability and we describe the role of *Epe1* in yeast chromosomes, that lacks DNA methylation and often are marked by single epigenetic marks (section 2.3). In section 2.4 we briefly described the most relevant zero–dimensional models of epigenetics, which chiefly rely on the competition between two antagonist marks and on the role of cooperativity to explain bistability. In section 2.5 I remark the importance of three–dimensional models to describe the competition between marks, when coupled to polymer dynamics. So far, it is all part of the literature background. Finally, in section 2.6, I describe my work. In our infection–like, one–dimensional model, the cooperativity of methylation is introduced explicitely by modeling long–range contacts with *Lévy* distributions. Together with the 'erasing' mechanism which provides for the removal of epigenetic marks, long–range contacts can entail bistability in a large region of the parameter space.

2.1 Higher levels of DNA organization: the bead–on–a–string structure

In eukaryotes, DNA is enclosed in a small volume, the cell nucleus (5–8μm in diameter). Given the high stiffness of this long polymer (in humans, DNA is about 2 meters long), genomes would occupy a much larger volume than the nucleus in their random coil conformation; instead, DNA is organized into

multiple levels of "ordered packaging", which have to be retrieved for physiological purposes when required. To this aim, several architectural proteins can induce bending and looping of the double helix. In eukaryotes, about 75% of DNA is associated with histone proteins [92], to form the so–called chromatin. Five different classes of histones are ubiquitous in different species and cell types: histone 1 (H1), histone 2A (H2A), histone 2B (H2B), histone 3 (H3), and histone 4 (H4). More specifically, 8 subunits of four types of histones proteins, H2A, H2B H3, and H4 can interact with each others to form the heterodymers H2A–H2B and H3–H4 [93]. Two H3–H4 dymers, and H2A–H2B respectively, interact through a 4–helix bundle to create two tetramer, which in turn bind to form the histone octamer. The histone octamer is stabilized by the attractive interactions with the DNA molecule. The classical histone–DNA assembly consists of 1.7 left-handed turns of DNA wrapped around the octamer, to form the nucleosome core particle (NCP); In most species the NCP may also integrate one linker histone H1 at the DNA entry–exit site [85].

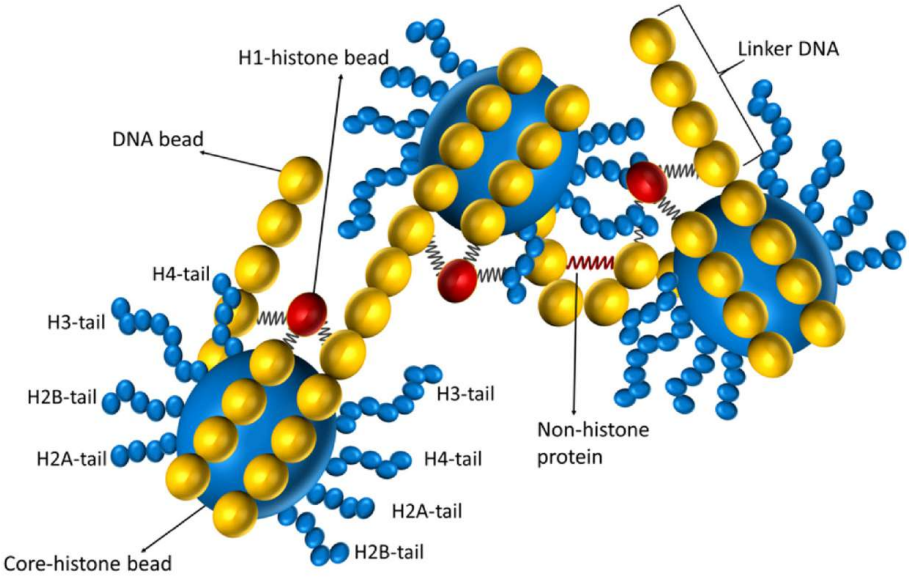

Figure 2.2. Schematics of the bead–on–a–string structure Depiction of the structure of linked nucleosomes.

Consecutive NCPs are connected by linker DNA, a portion of free DNA 20–60 bp long. Chromatin primary structure is thus constituted by a repeated sequence of one NCP and DNA linker ($\sim$ 200 bp), which forms the beads–on–a–string structure of eukaryotic chromatin, see schematics in Fig. 2.2. Importantly, NCPs are not electrically neutral, as histones carry a charge of $+144e$, while DNA contributes with $-294e$. Moreover, charge distribution is not spatially

uniform, as shown in Fig. 2.2. Electrical charges at the surface of DNA and histones are essential to provide higher level of chromatin compaction: the formation of nucleosomal arrays can be stabilized by electrostatic interaction between histones tails, which are the N–terminal sequences of H2A, H2B, H3, H4, which extend out of the globular histone core, see Fig. 2.2. In particular, the H3 and H4 tails, which are made of 35 and 20 residues respectively, are intrinsically disordered protein domains [85, 93], responsible of intranucleosome stability [94] and nucleosome stacking through internucleosomes tethering [95]. On the other hand, the H2A and H2B tails are required for proper nucleosom core reconstitution [96]. Although numerous works suggest the existence of different arrangements of nucleosomes in the 10–nm and 30–nm fibers [97, 98], the actual existence of an ordered organization of chromatin *in vivo* at these length scales is still a highly debated topic [92, 99]. An interesting physical perspective on the mesoscopic order of nucleosomes is offered by the theorethical framework in Ref. [100].

2.2 Post–translation modifications and epigenetics

Not only do histone tails play a role in structuring chromatin, they also serve as substrate of the post-translational modifications (PTMs). In particular, charged segments of the histone tails, the lysines K and arginines R, are the most targeted templates of such chemical modifications. The histone–tail PTMs, together with other DNA modifications, constitute a vast family of 'epigenetic marks'. A comprehensive list of the histone modifications and their functions is detailed in Ref. [101]. Grossly speaking, epigenetic marks can indirectly modulate gene expression through chromatin structural rearrangements. In fact, the most relevant function of histone modifications is the recruitment on non–histone proteins, which can *(i)* disrupt inter–nucleosomes contacts, hindering the formation of the 30–nm fiber, *(ii)* promote chromatin compaction, *(iii)* bind to gene promoters, activating a regulatory cascade of other modification events, which orchestrate several DNA–based processes (transcription, DNA repair, DNA replication, chromosome condensation).

The demarcation of epigenetic domains is therefore vital to estabilish the correct chromatin environment [101]. Mainly, there are two types of chromatin environments: *heterochromatin*, which is commonly compact and transcriptionally silent, and *euchromatin*, more swallen and transcritpionally active. Accordingly, PTMs distinct into those that correlates with heterochromatin formation and gene repression, and those which promote euchromatin formation and transcription. Acetylation (K–ac), methylation (K–me and R–me),

phosphorylation (S–ph and T–ph), and ubiquitination (K–ub) have been implicated in activation, while sumoylation (K–su), deimination, and proline isomerization have been implicated in repression [101]. However, it seems that the same modification can have both positive and negative effects on different DNA–related processes, depending on the particular context; for instance, the presence of different adjacent modifications or the binding of proteins can antagonise or promote the main role of a some modifications [102].

2.2.1 The reader–writer–eraser machinery

Thus, we face a complex scenario of several different modifications, possibly correlated, whose action can strongly be affected by the neighbouring modifications and enzymes. To simplify the general picture, we chiefly focus on general features of PTMs which are the most characterized to date [85]: *(i)* the trimethylation of the lysin 9 and 27 of the histone H3, called H3K9me3 and H3K27me3 *(ii)* lysin acetylation, as the H3K27ac mark. The former PTMs are associated with a set of 'writer' enzymes, such as lysine methyltransferases (KMTs) and 'eraser' proteins such as histone (lysine) deacetylases (HDACs or KDACs) or lysine demethylases (KDMs) [103]. Instead, writers devoted to acetylation are histone acetyltransferases (HAT) and lysine acetyltransferases (KAT) [85]. Most often, the writing 'skill' of a certain enzyme is accompanied by its 'reading' ability. A reader–writer complex is an enzyme that is accessorised of one or more reader domains, capable of recognizing PTMs, and a writer domain, which can install new PTMs. The regulation of the spreading of a modification across nucleosomes is primarly influenced by pre–existing PTMs, which function as substrate for readers; thus, the complex can activate the writer domains which can mark elsewhere along the chromatin chain, promoting the spreading of the epigenetic mark, possibly in a cooperative way. In Fig. 2.3 we present a schematic depiction on how PTMs can modulate the reader–writer activity [104]: *(i)* PTMs can result in an altered binding affinity of a writer or eraser: for instance, once a reader binds a pre–existing PTM, it favours (or disfavours) the installation of a new PTM in its proximity; *(ii)* PTMs can modify the structure of the nucleosome or the chromatin conformation, making the reader–writer more (or less) capable to epigenetically modify that particular nucleosome, or those in the vicinity; *(iii)* PTMs can elicit allosteric change in the enzyme, that is, they can directly modify the structure of the reader–writer complex, facilitating or hindering the installation of a new epigenetic mark on that nucleosome, or on the neighbour nucleosomes. All these mechanisms are non–exclusive. However, they contribute to build up an epigenetically stable environment, which serve, for instance, to the formation of the *constitutive*, or pericentric, heterochromatin, which has been found to be rich in mono–,

di–, and three–methylation, mostly on lysine and argynine residues of histone tails [105].

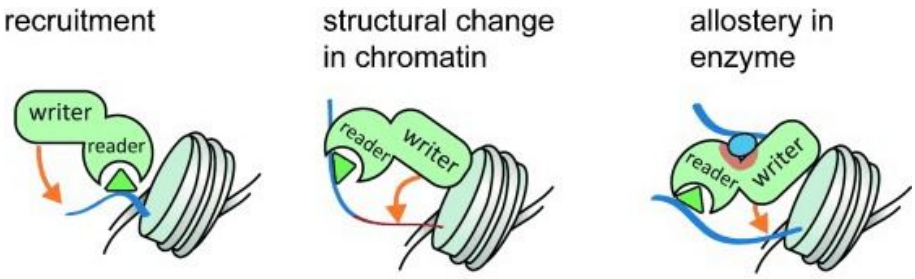

Figure 2.3. Schematics of writer–reader action. Left: reader can bind a pre–existing PMT, catalysing the installation of a new PTM. Center: pre–existing PTMs can trigger chromatin remodelling, by altering the structure of the histone. Right: pre–existing PTMs favour allosteric change in the writer–reader enzyme, which modifies the enzymatic activities. Figures taken from [104].

How does the estabilishment of epigenetic domains occur? In summary, the estabilishment *de novo* of an epigenetic domains starts with *nucleation*. This happens at specific chromatin sites, where specific enzymes favour the recruitment of reader–writer enzymes, which can start the subsequent spreading of the epigenetic mark. An instance is the *Clr3* complex in fission yeast, which targets specifically a sequence close to the *mat3* locus. Nucleation sites can be associated to certain molecular 'bookmarks' (such as the TFIID, a transcription factor which interacts with specific sequences called TATA box) [106]. Bookmarking is necessary to re–estabilish the epigenetic environment after the disruption caused by replication and cell mitosis [82]. The second step is represented by the *spreading*: the recruitment of writers along the chromatin chain is enhanced by the presence (or the absence) of pre–existing epigenetic mark, and possibly by the folding ability of chromatin. The 'nucleation and looping' mechanism assumed in Ref. [107] contemplates the propagation of the H3K9me3 mark via chromatin loops.

The epigenetic spreading cannot prolong indefinitely, otherwise a single epigenetic state could prevail over the whole genome, with deadly effects on the living organism. Hence, mechanisms that prevents the over–extension of epigenetic domains are required. Confinement of epigenetic domains can be estabilished through physical boundaries, or equilibrium between competing marks or enzymes. Such mechanisms of control are required for the entire cell life, or even multiple generations. Indeed, experiments suggests that, in yeast, the estabilishment of an epigenetic state can take generations [108], whereas in mice the spreading of heterochromatin occurs at a rate of $\sim$ 100 nucleosomes/day [109].

2.3 Evidence for epigenetic bistability

The maintainance of epigenetic domains, as for constitutive heterochromatin, is crucial, especially through several generations. For instance, a skin cell needs to retain the epigenetic memory of its state after division. Nevertheless, the dynamical character of PTMs, which can be deposited and removed by writers and erasers, also entails 'epigenetic multistability': a prototypical example is represented by the vernalization of Arabidopsis [110]. As shown in Fig. 2.4, the level of expression of the *flowering locus C* (FLC) during warm periods can change depending on the duration of the cold periods. Prolonged cold causes more repression of the FLC, which in turn promote flowering. During cold exposure, the H3K27me3 marks are deposited by the Polycomb repressive complex 2 (PRC2) at the nucleation region, aided by the PHD proteins. PHD protein can bind PRC2 and, during the warm period, activate the spreading of the mark over the neighbour regions, stabilising the FLC repression (see Fig. 2.4). The epigenetic character of this under–regulation is represented by the stability of the level of repression, which depends on the period of cold exposure. In such sense, this system is epigenetically multistable.

Other experiments on human and mice cells show that some domains possess both activating and repressive modifications [111]. The coexistence of different epigenetic mark leads to a differentiation in daughter cells. In other words, daughter cells might preserve one single modification type instead of both. This 'choice' can profoundly influence the fate of offspring cell at a functional level: for instance, H3K27me and H3K4me antagonise in mouse stem cells [112, 113] as they are implicated in silent chromatin and active chromatin, respectively. The enrichment of these opposing modifications within epigenetic bivalent domains correlated with low–level expression of developmental transcription factors [112].

In a recent work, K. Sneppen and L. Ringrose suggest that poised chromatin in Polycomb is bistable and not bivalent, that is, there is no coexistence between competing marks, rather domains switches between active and silent states [114]. The existence of bistable epigenetic domains is quite ubiquitous, not only in stem cells, but also in yeast: single–cell experiments suggest that stochastic transitions between two epigenetic states can occur either over the cell lifetime, or over a few generations [108, 115]. This is especially the case in "weakened" region of the genome, where, for instance, nucleation sites are missing or defected [116]. In section 2.6, I focus on the results of a simple one–dimensional model, which models the dynamics of epigenetic marks, especially in yeast, where a single PTM is expected to spread until reaching some boundaries, which can be physical proteins, such as CTCF, highly transcribed

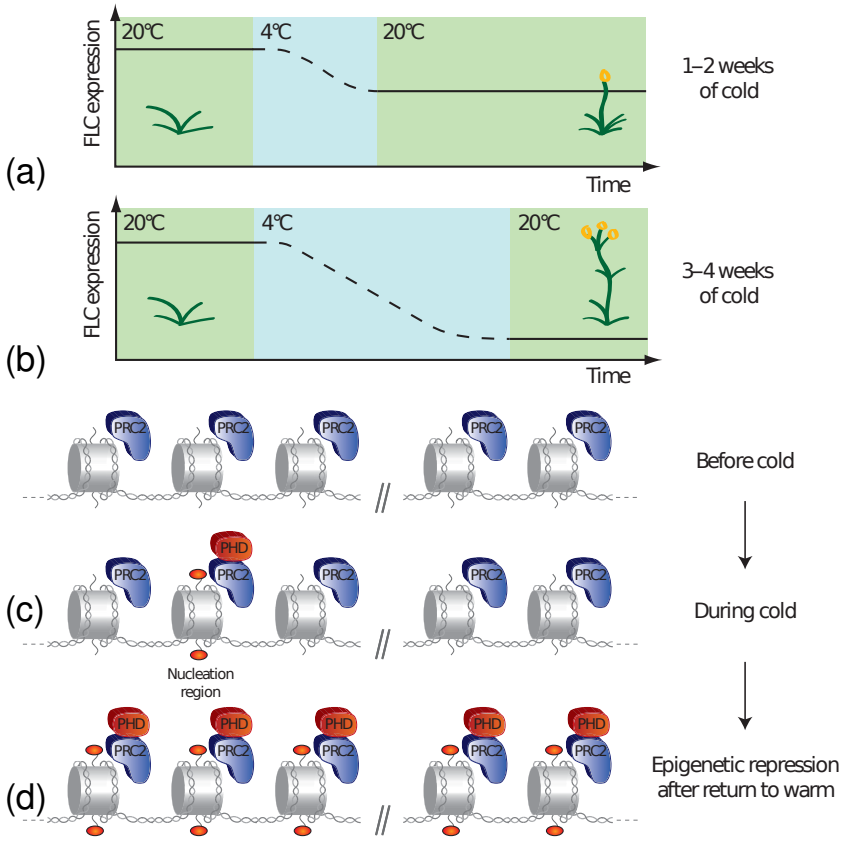

Figure 2.4. FLC vernalization and epigenetic state of chromatin. (a) After 1–2 weeks of cold, the level of FLC expression is diminished. During vernalization periods (3–4 weeks) FLC is repressed more, allowing the correct post–winter flowering. **(b)** Schematics of chromatin region before cold. Chromatin modifications are absent **(c)** During the cold period, PHDs binds the PRC2 proteins, starting the nucleation at the target site (the transcription start site of FLC). **(d)** At warm temperatures, the H3K27me3 and PHD-PRC2 complex spread all over the FLC gene body. Figures adapted from [110].

regions (euchromatin), or other complexes [117, 118]. In the next section, I explicitely discuss the Epe1, a particular domain of the JmjC protein which limits the spreading of heterochromatin by performing demethylase activity [119].

2.3.1 Epe1 can limit methylation spreading in Yeast

Differently from complex living organism, fission yeast lacks DNA methylation (not histone methylation, which is present), and a single type of methyltransferase is responsible for the self–propagation and maintainance of heterochromatic domains [119, 120]. In particular, Clr4 proteins can bind the H3K9me2/3 mark that they generate, through Swi6 protein which favour the recruitment of Clr4 and, thus, the spreading of PTMs [120]. Clr4 then act as 'writers' and 'readers' of this methylation mark.

Importantly, spreading of heterochromatin can be hindered by the Epe1 domain of the JmjC protein, which function as an 'eraser' of the methylation mark, in *Schizosaccharomyces pombe* [121]. Epe1 catalyses the histone demethylation reaction, even though more recent works point to the fact that it does not intervene directly with demethylase activity [122]. Rather, it colocalises with Swi6 moderating the repressive effect of silencing proteins, especially during the acetylation stage [122]. However, the real role of Epe1 is still uncertain, and a complete picture of the Epe1 significance in counteracting methylation spreading is still unknown. Nevertheless, it is commonly thought that Epe1–dependent removal of H3K9me ensures regulation and persistence of centromeric heterochromatin [119]. In yeast, competition between H3K9 methyltransferase and demethylase activities can be finely tuned to safely protect gene–rich regions from the propagating heterochromatic boundaries. In our model (section 2.6.1) we speculate that local demethylation can counteract the spreading of a single epigenetic mark, producing, in some cases, epigenetic bistability.

2.4 Zero–dimensional models: a framework for epigenetic bistability

Zero–dimensional models are characterized by the absence of any spatial information; in this models, the time evolution of the (global) epigenetic state of a chromatin domain is governed by rate–based equations. Commonly, the system consists of N nucleosomes that can be in n_s states, labeled by the variables $\{s_1, s_2, \ldots, s_N\}$, where s_i is the state of the ith nucleosome. In most models, the assumption is that the nucleosome state is specified by a single epigenetic mark on the histone tails, or multiple epigenetic marks contribute to

a single 'functional' state of the nucleosome, such as 'silent' or 'active'. In the simplest case $n_s = 2$. We define n_j as the number of nucleosomes in the state j, hence, it follows:

$$\sum_{j=1}^{n_s} n_j = N \,. \tag{2.1}$$

2.4.1 Two–state model

In Ref. [91] a sort of mean–field equation has been proposed to explain epigenetic bistability. The simplest case $n_s = 2$ was considered, so that Eq. (2.1) becomes $n_M + n_U = N$, where M and U indicate the *modified* and the *unmodified* states respectively. The evolutionary equation is given by:

$$\frac{\mathrm{d}n_M}{\mathrm{d}t} = R_+ \left[n_M\right] (1 - n_M) - R_- \left[n_M\right] n_M + \eta_M. \tag{2.2}$$

where η_M is a noise term which models the stochasticity of the reactions, and commonly depends on the domain size ($\sim N^{-1/2}$). The rates are given by:

$$\begin{aligned} R_+ \left[n_M\right] &= \alpha n_M^2 + (1 - \alpha), \\ R_- \left[n_M\right] &= \alpha (1 - n_M)^2 + (1 - \alpha). \end{aligned} \tag{2.3}$$

In this model, the first rate models the $U \to M$ conversion: the first term represents the recruitment, which occurs with probability α when two M sites 'infect' an U site, whereas the second term models the spontaneous conversion. The second rate represent the symmetric $M \to U$ conversion: it can be obtained from $n_M \to 1 - n_M$. The schematics of this model is depicted in Fig. 2.5(a). After some calculations, the Langevin equation becomes:

$$\frac{\mathrm{d}n_M}{\mathrm{d}t} = \alpha (2n_M - 1) \left[n_M (1 - n_M) - 1/F\right] + \eta_M. \tag{2.4}$$

where $F = \alpha / (1 - \alpha)$ is the recruitment–to–noise ratio. In the absence of the noise η_M, one can show that Eq. (2.4) entails bistability when $F > 4$ Ref. [91]. Calculations follow the same guidelines in section 2.6.4. Here, dynamical bistability is based on the assumption that a a cooperative mechanism for recruitment takes place, requiring two M (U respectively) sites to perform the conversion of a single U (M respectively). This assumption mimics the cooperativity of *Sir* complexes in spreading an epigenetic mark, as it occurs in budding yeast. Moreover, it is assumed that the writing and erasing mechanism (or, more generally, conversion to the U state) follow the same kinetic rules, see Eq. (2.3). It is worth to mention that a similar model has been studied in [116], where different sources of cooperativity were explored to support the emergence of epigenetic bistability.

Another important contribution towards the understanding of two–state epigenetic systems, such as in *S. Cerevisiae*, is represented by Ref. [123]. In this work, several different scenario were considered, from cooperativity to long–range reading/writing (which they refer to as local–local, global–local and global–global recruitment). It was shown that both elements can enhance epigenetic bistability.

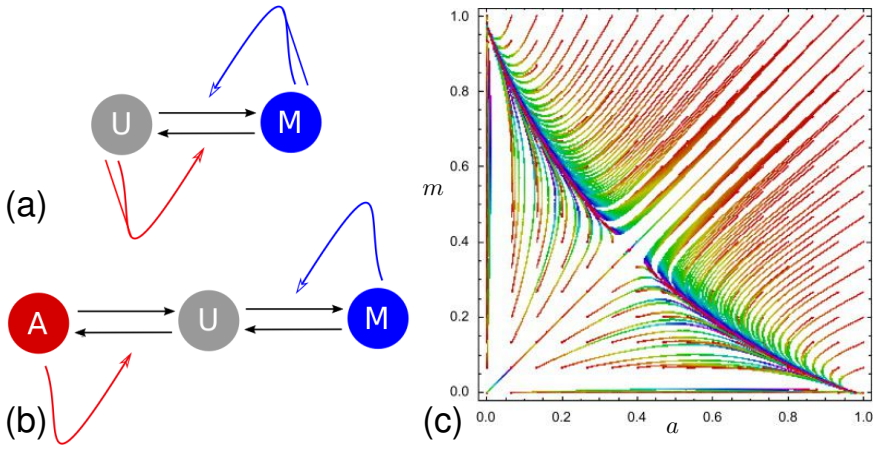

Figure 2.5. Models and phase diagram. (a) Schematics of the two–state model in [91]. The horizontal black arrows indicate the spontaneous interconversion, whereas the blue and red arrows represent the conversion by recruitment. Two M (U respectively) sites are required to mark (unmark respectively) a U site (M respectively). **(b)** Schematics of the three–state model in [84]. Here, interconversion between A and M sites are possible in a two–step process (via U). In its original version, no cooperativity is required for bistability. **(c)** Phase diagram of the three–state model in [124], as a function of m and a. Parameters used are: $\alpha_A = \alpha_M = 5$; $\beta_A = \beta_M = 3$; $\gamma_A = \gamma_M = 0.1$; $\chi_A = \chi_M = 0.01$. Figure taken from [124].

2.4.2 Three–state models

A prototypical three–state epigenetic model was proposed by Dodd *et al.* [84]. In this work, local and long–range symmetric conversions were considered, with the addition of a 'neutral' intermediate state. In particular, they consider a lattice of N nucleosomes, whose state is labelled by n_j, with $j = 1, \cdot, N$; each nucleosome can be in an 'antimodified' state (A), a 'modified' state (M) or an 'unmodified' state (U); broadly speaking, A and M are competing modifications (usually acetylation and methylation), whose functions can be chromatin activation and silencing, while U is an intermediate non–marked state. The model was conceived such that, in its non–cooperative version, the interconversion $A \longleftrightarrow M$ is forbidden. The process of attempted conversions is obtained by iterating the following steps:

(i) A random nucleosome, say the i–th, is selected for conversion. A recruited conversion is attempted with probability α (step *(iia)*); conversely, a spontaneous conversion is attempted with probability $1-\alpha$ (step *(iib)*).

(iia) Recruited conversion: A second random nucleosome, say the j–th, is selected (with uniform probability within the lattice); if n_j is either in the M or the A state, n_i is changed one step forward the state of n_j.

(iib) Noisy conversion: Nucleosome n_i is changed one step toward either of the other types with a probability of $1/3$.

The stochastic model is one–dimensional, however the correspondent zero–dimensional model was studied in Ref. [124]. The two relevant equations for the global methylation, unmarked and acetylation density, $m \equiv n_M/N$, $u \equiv n_U/N$ and $a \equiv n_A/N$ respectively, read:

$$\begin{aligned} \frac{\mathrm{d}m(t)}{\mathrm{d}t} &= R_{U\to M}u(t) - R_{M\to U}m(t), \\ \frac{\mathrm{d}a(t)}{\mathrm{d}t} &= R_{U\to A}u(t) - R_{A\to U}a(t), \end{aligned} \tag{2.5}$$

where the rates are

$$\begin{aligned} R_{U\to M} &= \alpha_M m(t) + \chi_M, \\ R_{M\to U} &= \beta_M a(t) + \gamma_M, \\ R_{U\to A} &= \alpha_A a(t) + \chi_A, \\ R_{A\to U} &= \beta_A m(t) + \gamma_A, \end{aligned} \tag{2.6}$$

The main result is that the interposition of an intermediate state between the A and the U state, provides bistability, even without cooperativity. An example of the flow lines in the (a, m) phase diagram when high A and M states are stable is shown in Fig. 2.5. If $\chi_M = \chi_A = \gamma_M = \gamma_A = 0$, two fixed points of the dynamical system are located at $(m^*, a^*) = (0, 1)$ and $(m^*, a^*) = (1, 0)$, whereas $(a_s, m_s) = 0, 0$ is an unstable fixed point, and the saddle point is located at

$$\begin{aligned} a_u &= \frac{\alpha_M \beta_A}{\alpha_A\beta_M + (\alpha_M + \beta_M)\beta_A} \\ m_u &= \frac{\alpha_A \beta_M}{\alpha_A\beta_M + (\alpha_M + \beta_M)\beta_A} \end{aligned} \tag{2.7}$$

They also verified that long–range spreading of the epigenetic mark enhances bistability [84].

2.5 Epigenetic competition is predicted by 3D models of "colored" chromatin

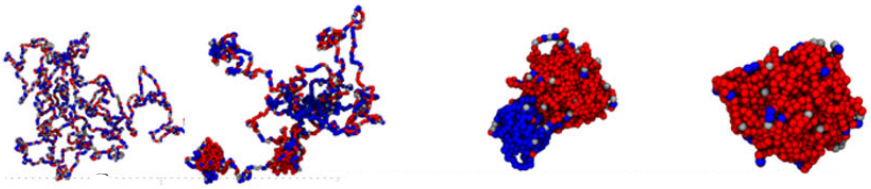

Figure 2.6. Competing epigenetic marks lead to estabilishment of an epigenetic territory. Snapshots taken at different timesteps from a random distribution of epigenetic marks. Epigenetic domains (blue or red) start to form, while the inert state (grey) tend to disappear; chromatin is more compact in the correspondence of marked region. Ultimately, one mark prevails, and the whole chromatin chain is collapsed. Figure adapted from [87].

Most recently, the three–dimensional dynamic organization of chromatin during epigenetic spreading and estabilishment has been explored by means of coarse-grained models [82, 87, 125]. Such models have the undeniable advantage of modelling the physical contacts between distant regions of chromatin, due to chromatin looping.

In these models, chromatin is modelled as a polymer, which can fold in the 3D space. As mentioned before, the long–range contacts can mediate the epigenetic spreading. Indeed, a Potts–like interaction was assumed between monomers 'in contact'. Similar to the model described in section 2.4.2, each monomer can be either in one of two competing epigenetic state (the authors identify the H3K9me3 and H3K27ac as competing marks), or in an inert, 'unmarked' state. In the presence of two competing symmetric epigenetic mark, the system undergoes a first–order–like transition between a coherent (either acetylated or methylated) compact phase and a disordered, more swallen state [87]. The selection of the dominating epigenetic mark occurs via breaking of the $\mathbb{Z}_2$ symmetry, similarly to what happens in the Ising model. The latter model offer a simple, possible explaination of the estabilishment of constitutive epigenetic domains. Nonetheless, both regions are equally compacted. It remains less understood which may be the actual mechanism that yields the difference between the more swollen chromatin, and the more crumpled heterochromatin, depending on the underlying epigenetic landscape.

To conclude this section, I should remark that all models described in sections 2.4 and 2.5 are based on the competition of two antagonistic marks, such as methylation annd acethylation, which are subjected to the same kinetic rules, which determine spreading or erasure of marks. The novelty in the model described below is that we consider the competition between a single

epigenetic mark and ann inert state. Rather, the estabilishment of bistability is demanded to the competition between the (long–range) writing and the (local) erasing mechanisms.

2.6 An infection–like 1D model predicts epigenetic bistability

Inspired by experimental evidence in yeast, we develop a one–dimensional model that can explain epigenetic bistability in short stretches of chromatin. In the two–state models in section 2.4.1 cooperative recruitment is required for bistability; instead, in the three–state models described in section 2.4.2, a robust bistability is obtained by means of an intermediate inert state between the two competing state (i.e. modified and antimodified), via spontaneous symmetry breaking. Here, instead, we consider a single biochemical mark and no cooperativity. Unlike previous models, the writing and the erasing mechanisms are different. In particular, we assume that the spreading of a mark occurs through long–range contacts mediated by the chromatin fibre. As detailed above, the long–range spreading of epigenetic marks can be essential for epigenetic inheritance and bistability; experiments have also shown that its contact probability is a power-law distribution [126, 127]:

$$P(l) \propto l^{-(\sigma+1)}, \qquad l \to +\infty, \tag{2.8}$$

where l is the distance along the chromatin fiber and $\sigma > 0$ is the contact exponent. Therefore, the writing probability can be described in terms of a *Lévy distribution* [84]. The *Lévy distribution* describes the so–called Lévy motions, or flights, defined by $0 < \sigma < 2$. Lévy flights are stochastic processes characterised by the occurence of long jumps, as the variance is divergent for $\sigma < 2$. In addition, for $\sigma \leq 1$ both the variance and the expectation value do not exist. Other details on the Lévy distributions are in Ref. [128]. Interestingly, the contact exponent defining this distribution can depend on cell type, stage of the cell cycle [129] and chromosome considered [130]. Moreover, by using the contact probability in Eq. (2.8), we are assuming that the timescale associated with chromatin relaxation is smaller than, or comparable to, the one related to the spreading of marks, as suggested in section 2.2.1. This is verified for small domains of about a hundred nucleosomes, if we compare the rate of heterochromatin spreading of about 100 nucleosome/day with the coherent motion of chromatin, which can be of about 10μm/day [131] (which is approximately the size of a nucleus). This means that, in some cases, it is realistic to consider the contact probability in Eq. (2.8) for random contacts of

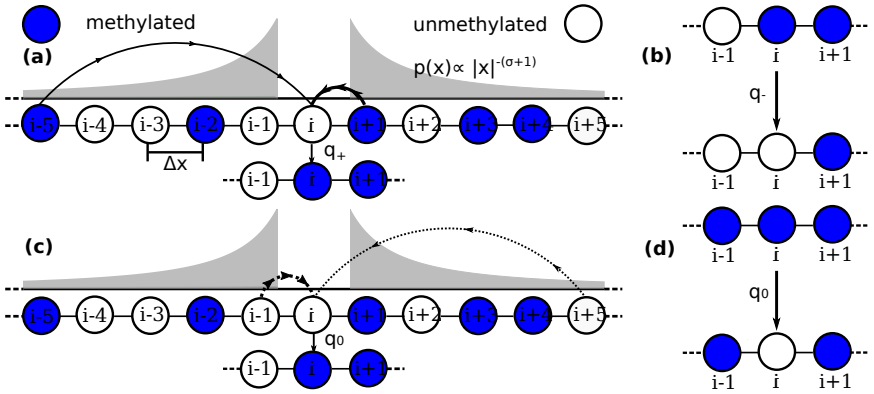

Figure 2.7. Microscopic rules. Representation of the transition rules for methylated/infected (or marked) sites (blue circles) and demethylated/recovered (or unnmarked) ones (white circles). **(a)** Site i, which is unmarked in this case, is in long–range contact with site j (here, $j = i - 5$). If j is marked, i can become methylated through a long–range "infection" at a rate $q_\lambda(1 + \lambda)$. **(b)** If j (here, $j = i + 5$) is unmarked, i can still become infected spontaneously, at rate q_λ. **(c)** To simulate the erasing process, a marked site i can be changed to the unmarked state at rate $q_\mu(1 + \mu)$ when at least one of the two neighbouring sites is unmarked. **(d)** A marked site i that is flanked by marked sites can spontaneously lose the mark at rate q_μ.

chromatin regions, rather than a contact map that mimics a fixed network of nucleosomes.

The erasure of methylation marks is modeled such that it is more likely to happen at the boundary of a methylated domain rather than in the bulk. Several scenarios can justify this assumption: *(i)* diffusion of the enzyme responsible for removing the epigenetic mark is hindered by the strongly crumpled conformation assumed by regions enriched in such mark [82]; *(ii)* the eraser enzyme possesses binding sites for both methylated and demethylated histones; *(iii)* erasing activity is more efficient at the boundary of methylated domains, where the concentration of mono– or di–methylated nucleosomes is higher (full methylation usually implicates three methylation marks deposited on the histone tail). Therefore, we describe mark erasure as a "local" process and we assume it is more probable in the presence of gradients in the density of methylation marks.

Due to the competition between writing and erasing mechanism, we find that a first–order–like transition separates a 'methylated' phase from a 'demethylated' one. The transition takes place due to the competition between long–range spreading and local erasing: for small systems, a wide region of the parameter space entails bistability, which is stronger in the proximity of the transition. This work shares some features with system in epidemics and in percolation–like models. In these contexts, there is often a competition between

'infection' and 'recovery' mechanisms; here, recovery is a jargon that refers to the transition to the 'native' unmarked state, as often used for two–state models [132]. For the sake of generality, we will also refer to the methylated and unmethylated states as infected and recovered states, respectively.

2.6.1 The model

The model is defined on a lattice of N sites, which can be either in the demethylated/recovered state or the methylated/infected state. Each site is associated with a dichotomous variable m_i, which can be respectively either 0 or 1. The transition rules are specified in the following way (see Fig. 2.7):

$$\begin{aligned} 0_i 1_j &\xrightarrow{q_\lambda(1+\lambda)} 1_i 1_j \\ 0_i 0_j &\xrightarrow{q_\lambda} 1_i 0_j \end{aligned} \tag{2.9}$$

$$\begin{aligned} 1_i 0_{i\pm1} &\xrightarrow{q_\mu(1+\mu)} 0_i 0_{i\pm1} \\ 1_{i-1} 1_i 1_{i+1} &\xrightarrow{q_\mu} 1_{i-1} 0_i 1_{i+1}, \end{aligned} \tag{2.10}$$

In words, if a site i is unmarked ($m_i = 0$), it can become marked at rate $q_\lambda(1+\lambda)$ when it enters in contact with a methylated site j ($m_j = 1$). Otherwise, if $m_j = 0$, it can convert spontaneously with rate q_λ. The site j is selected by drawing the distance $|i-j|$ from a normalized power–law distribution:

$$P(|i-j|) = \zeta(\sigma+1)|i-j|^{-(\sigma+1)} \qquad i \neq j. \tag{2.11}$$

where $H(\alpha) = (\sum_{n=1}^{\infty} n^{-\alpha})^{-1}$. Once the site j is selected, the conversion rate can therefore be written as

$$q_{0\to1} = q_\lambda(1+\lambda m_j). \tag{2.12}$$

Conversely, if site i is marked ($m_i = 1$), it can become unmarked at rate $q_\mu(1+\mu)$ if it is at a domain boundary, that is if $m_{i-1}m_{i+1} = 0$. Otherwise, if $m_{i-1}m_{i+1} = 1$ – i.e., if the site is flanked by two marked sites – it can be demethylated with a basal rate q_μ. In other words the sites in the domains bulk are more protected from erasure. The recovery rate can thus be written as

$$q_{1\to0} = q_\mu\left[1+\mu(1-m_{i-1}m_{i+1})\right]. \tag{2.13}$$

The Monte Carlo algorithm proceeds with random sequential updates, and each timestep (Monte Carlo sweep) consists of N conversion trials:

1) Extract a random integer number between 1 and N, which selects the i-th site. If $m_i = 1$, jump directly to step 4;

2) if $m_i = 0$, choose with probability $1/2$ one of the two directions to attempt a long–range infection event;

3) draw a random number z from a uniform distribution between 0 and 1, then consider the real number $r = z^{-1/\sigma}$, round it to the smaller positive integer $\lceil r \rceil$ so that $j = \mathrm{mod}\{i \pm \lceil r \rceil, N\} + \theta(-i \mp \lceil r \rceil)N$, which is needed to take into account the periodic boundary conditions; $\mathrm{mod}\{\cdot,\cdot\}$ indicates the modulo operation between two integers, θ is the Heaviside function, and the $\pm$ depends on the direction choosen in step 2;

4) perform a conversion with rate $(1 - m_i)q_{0\to1} + m_i q_{1\to0}$.

Note that the values of λ and μ are bound by the values $q_{\lambda,\mu}^{-1} - 1$, where $q_{\lambda,\mu} \leq 1$. To compare time in simulation units to physical time, one may calculate and compare any of the dimensionless quantities obtained by multiplying time with one of the rates. This means that a time step in our simulation equals the inverse long–range methylation rate (or the inverse demethylation rates at a domain boundary).

Phenomenological Langevin equation

From Eqs. (2.12) and (2.13) is it possible to write down a Langevin equation that qualitatively describes the dynamics of the methylation and demethylation fields, $m(x,t)$ and $d(x,t)$. The condition

$$\frac{1}{N}\left[\int m(x,t)\,dx + \int d(x,t)\,dx\right] = 1 \tag{2.14}$$

is always satisfied. Given the time increment dt, that corresponds to a Monte Carlo sweep (N attempted moves), the generic Langevin equation reads:

$$dm(x,t) = \left(D_A \nabla^\sigma m + r_+[m](1 - m(x,t)) - r_-[m]m(x,t)\right)dt + \eta(x,t)dt\ , \tag{2.15}$$

where D_A is the coefficient of anomalous diffusion, ∇^σ denotes a fractional derivative and the type of noise is not specified, as it depends on the structure of absorbing states. The generic structure of the noise term in a similar reaction–diffusion system is determined by its expectation value and correlation [133]:

$$\begin{aligned} \langle \eta(x,t) \rangle &= 0 \\ \langle \eta(x,t)\eta(x',t') \rangle &= \left[\Gamma' m(x,t) + \Gamma'' m(x,t)^2\right]\delta(x - x')\delta(t - t'), \end{aligned} \tag{2.16}$$

being Γ',$\Gamma'' > 0$. A simple fractional derivative integral representation is given in [132], where the Lévy distribution is the *kernel* of the following integral:

$$\nabla^{\sigma} m(\mathbf{x}, t) \equiv \int P(|\mathbf{x} - \mathbf{x}'|) m(\mathbf{x}', t)\, d\mathbf{x}'. \tag{2.17}$$

The production and annihilation rates (in the first–order approximation) are:

$$\begin{aligned} r_+[m(x,t)] &\simeq q_0 + q_1 m(x,t) \\ r_-[m(x,t)] &\simeq q_0' + q_1' m(x,t)\ . \end{aligned} \tag{2.18}$$

The term proportional to $m(x,t)$ in $r_+[m]$ can be understood as the short–range contribution of the Lévy distribution to the infection/methylation process (the long–range infection is taken into account by the fractional diffusion term in Eq. (2.17). Instead, as the erasure is more effective in regions with boundaries, in a coarse–grained approach the fraction of boundary sites is almost equal to the fraction of demethylated sites; hence the local erasure at the boudary is proportional to $1 - m(x,t)$, hence linear in $m(x,t)$. Combining Eq. (2.18) and Eq. (2.15) we obtain

$$\frac{dm(x,t)}{\partial t} = D_A \nabla^{\sigma} m + q_0 + \kappa m - g m^2 + \eta. \tag{2.19}$$

In general, this equation holds for low level of methylation ($m \sim 0$).

The first terms in Eqs. (2.12) and (2.13), q_λ and q_μ, are two basal rates which represent respectively spontaneous acquisition and deletion of methylation marks. Within the problem of understanding how epigenetic marks spread, it is natural to set these rates to non-zero values to account for biological noise and imperfect writing and erasure by the respective enzymes. Yet, it is instructive to consider first the situation in which either one or both the basal rates are zero: these cases lead to interesting physics, and help to better understand the behaviour of the general system. We should remark that none of the Langevin equations in this chapter has been solved numerically. However, they serve as a useful and simple tool to understand the physics of this model.

We shall remark that we will mainly consider the rescaled rates $q_\lambda \lambda = \bar{\lambda}$ and $q_\mu \mu = \bar{\mu}$ throughout the chapter. To describe the global methylation/ infection state, we use the order parameter $\bar{m} \equiv \langle m \rangle = 1/N \sum_{i=1}^{N} m_i$, which is the fraction of infected sites in the lattice.

2.6.2 Connection with percolation models

Different number of absorbing states can emerge if any of the basal rates is zero. If $q_\lambda \to 0$, there is one absorbing state ($\bar{m} = 0$); if $q_\lambda, q_\mu \to 0$, the absorbing states are two (either $\bar{m} = 0$, or $\bar{m} = 1$). As the absorbing states

are configurations from which it is not possible to escape, detailed balance is violated and the resulting system is out–of–equilibrium.

As we shall show, our model displays non-equilibrium phase transitions of different types, depending on the number of the absorbing states (either one or two), which reflects the symmetries involved in the corresponding effective action [134]. This model is respectively either in the *Long–Range Directed Percolation* (LRDP) non-universality class – and characterised by a second-order phase transition with a σ-dependent (*non-universal*) exponent – or, alternatively, can be mapped onto the *Long–Range Compact Directed Percolation* (LRCDP), – characterised by a first-order phase transition with a discontinuity in the average infected/methylated fraction of sites $\bar{m}$. It is well known that the standard Compact Directed Percolation (CDP) – i.e., the case with short-range interactions – falls in the same universality class of the *Standard Voter Model* (SVM) in $d = 1$. Moreover, critical exponents can be exactly computed, since the dynamics can be seen as an annihilation problem between two random walkers [135].

There are also studies of the annihilation of Lévy flights [136, 137] and of the Voter Model with long-range interactions/infection [138]: in these works, this model was investigated through the computation of the dynamical exponent α which characterise the decay of the density of infected sites, ρ, with time, i.e., $\rho \sim t^{-\alpha}$. This requires simulations which start with uniform density or fully disordered configurations. Here, instead, we focus on the survival probability exponent $\delta(\sigma)$ in simulations where the initial configuration has a single infected/methylated site (single seed simulations). Such condition is typically used to compute the dynamical exponent δ, since one expects a power law dependence of the survival probability with time – i.e., $S(t) \sim t^{-\delta}$ – for $t \gg 1$. Our aims are to: *(i)* check that in the limit of $q_\lambda \to 0$ (q_μ finite) the contact process belongs to the LRDP non-universality class; *(ii)* show that in the limit $q_\lambda, q_\mu \to 0$ our model falls in a different non–universality class, which is the LRCDP one; to this end we calculate the critical exponent δ.

Limit $q_\lambda \to 0$

In the limit where $q_\lambda \to 0$ and $\lambda \to +\infty$, such that the product $q_\lambda \lambda \equiv \bar{\lambda}$ is finite and $q_\mu > 0$, the system cannot recover from the totally demethylated state. The sole absorbing state is defined by $\bar{m} = 0$. The system can thus be recast as an infection model with long-range infection and a single absorbing state. The prototypical model for diffusive–reaction system with a single absorbing state and long–range infection is LRDP. This leads to $q_0 = 0$, and $\eta \equiv \sqrt{\Gamma m \zeta}$, where ζ is a typical uncorrelated white noise. The "square–root" multiplicative noise is justified by the Reggeon effective field theory in [134]. In particular,

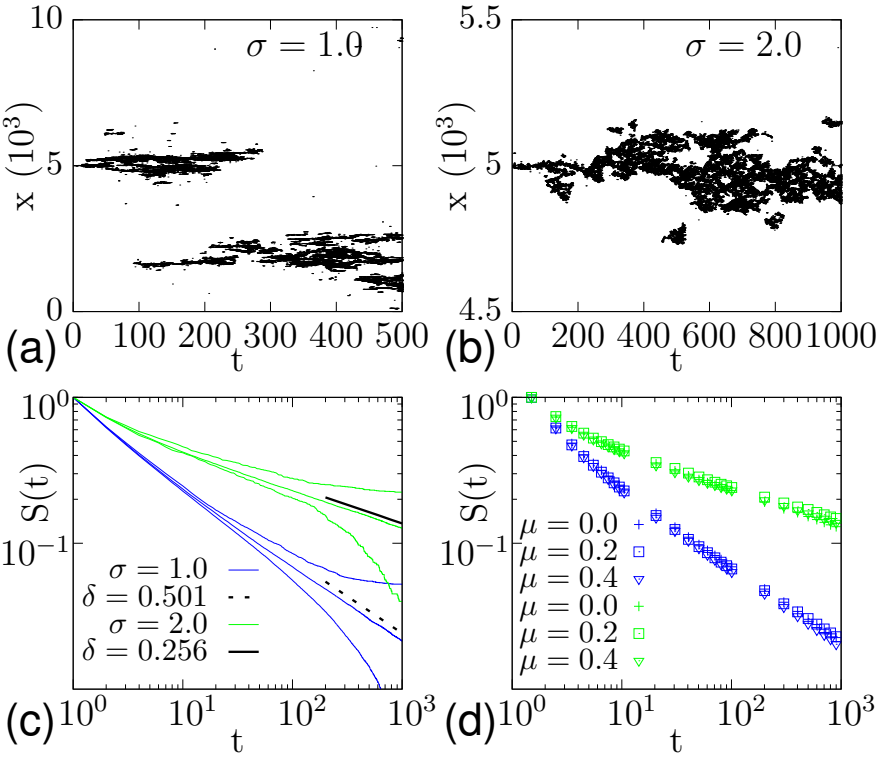

Figure 2.8. Kymographs and survival probability for LRDP. In the first two panels we show two typical kymographs at the transition point, and with $\bar{\mu} = 0.2$, for respectively **(a)** $\sigma = 1.0$ and **(b)** $\sigma = 2.0$. Note that there are no very compact domains. **(c)** Power-law trends of $S(t)$ for two values of the contact exponent ($\sigma = 1.0$, blue lines, and $\sigma = 2.0$, green lines), with $\bar{\mu} = 0$, and the corresponding value of the critical exponent δ (respectively dashed and solid black lines). For clarity, we also show the survival probability just above and below the transition, which display the typical decay and saturation. **(d)** By varying the value of $\bar{\mu}$ the survival probability does not change. The critical properties of the system, then, do not depend on the local erasing, but only on the global erasure rate.

the quadratic term in Eq. (2.16) would correspond to a quartic term in the effective action, $\propto \hat{m}^2 m$. Such term is irrelevant in the perturbation expansion, see Ref. [133]. Hence, in this limit Eq. (2.19) reads:

$$\partial_t m = \kappa m - g m^2 + D_A \nabla^\sigma m + \sqrt{\Gamma m}\zeta \,, \tag{2.20}$$

where, in general, $\kappa = \kappa(\bar{\lambda}, \bar{\mu}, q_\mu, \sigma)$ and Γ is the strength of the noise ζ. This equation predicts two different stable states in the mean field treatment: for $\kappa < 0$ the stable state is the absorbing one, defined by $\bar{m} = 0$, whereas for $\kappa > 0$, the stable state is $\bar{m} = \kappa/g$. Clearly the system is critical at $\kappa = 0$.

We simulate our system on a $N = 10^4$ lattice for 10^3 Monte Carlo sweeps, and we search for the critical point, i.e. we calculate the survival probability $S(t)$ which we expect to decay as $t^{-\delta}$ at some value of $(q_\mu, \bar{\mu}, \bar{\lambda})$. Below the transition, the system evolves toward the absorbing state, and $S(t)$ exhibits an exponential decay in the long–time limit. *Vice versa*, above the transition, the stable state is represented by a non–null mean methylation; even if some trajectory can suddenly fall into the absorbing state (as only a single seed is infected at the start), a finite fraction of replicas will steadely mantain the infection. Correspondingly, the survival probability will display a plateau at long times. Our results are in accordance with those found for LRDP in Ref. [132]. Two representative kymographs are shown in Fig. 2.8, for $\sigma = 1.0, 2.0$ at the critical point. Note that, when the infection process is long–range, multiple infected nuclei can be generated very far away from the original seed (see Fig. 2.8a). Conversely, when it is short–range, the infected domains are more gathered around the original seed position (see Fig. 2.8b). In any case, domains are not compact – i.e. there are holes within every single domain.

In Fig. 2.8c the power law behaviour of $S(t)$ at criticality is shown, and we find $\delta(\sigma = 1.0) \simeq 0.501$ and $\delta(\sigma = 2.0) \simeq 0.256$, which are compatible with the exponents found in [132]. (No survival probability for $\sigma \leq 0.5$ is shown, due to important finite-size effects). Finally, we show that the exponent is independent of the value of $\bar{\mu}$ (see Fig. 2.8d), which just modifies the value of q_μ at which the system is critical. In other words, models with only gradient-dependent (or "local") recovery are in the same universality class as models with global recovery terms at a fixed σ. This is consistent with the relevant terms in the Reggeon effective field theory [134] which are the same for every value of $\bar{\mu}$.

Limit $q_\lambda, q_\mu \to 0$

In the limit in which $q_\lambda, q_\mu \to 0$, and $\lambda, \mu \to +\infty$, in such a way that the quantities $q_\lambda \lambda, q_\mu \mu$ remain finite, the second reactions in Eqs. (2.9) and (2.10) are both suppressed, and the methylation/demethylation dynamics takes place only if the system is not fully demethylated/methylated. Indeed, in this case

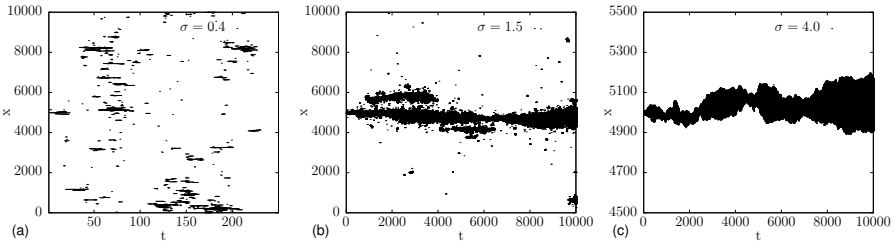

Figure 2.9. Kymographs at the transition point. Representative kymograph of the infection/methylation profile, when the system is at the critical point ($\kappa = 0$ in Eq. (2.21)), $S(t)$ is a power-law). **(a)** For $\sigma = 0.4$ no compact domains can live for long times, and the systems falls into the absorbing states after few hundreds timesteps. **(b)** For $\sigma = 1.5$ the main domain remains compact and lives for the entire duration of the simulation. Other domains can be created and erased. **(c)** For $\sigma = 4.0$ the only active compact domain is generated by one infected seed. it can live for a virtually infinite time.

there are *two* absorbing states defined by $\bar{m} = 0, 1$. The presence of two rather than one absorbing states modifies the universality class, so that the system is no longer equivalent to LRDP. In this case, the prototypical model is the LRCDP. In standard (short–range) compact directed percolation a particle–hole symmetry arises: $m \to 1 - m$. Such symmetry is retained also in the long–range version of the model. This symmetry is reflected in the noise term, that now is obtained by imposing $\Gamma' = \Gamma'' \equiv \Gamma$; therefore we have $\eta \equiv \sqrt{\Gamma m(1-m)}\zeta$. Again, if we define $q_\lambda \lambda = \bar{\lambda}$ and $q_\mu \mu = \bar{\mu}$, Eq. (2.19) becomes:

$$\partial_t m = \kappa m(1-m) + D_A \nabla^\sigma m + \sqrt{\Gamma m(1-m)}\zeta \,, \tag{2.21}$$

where κ is a generic function of the microscopical rates $\bar{\lambda}$ and $\bar{\mu}$. Unlike Eq. (2.20), the effective action built from Eq. (2.21) has another symmetry under the transformations $m \to 1 - m$ and $\hat{m} \to -\hat{m}$ at $\kappa = 0$, which is the reason why this variant of the model behaves differently from LRDP. Additionally, the dynamical equation in Eq. (2.21) is well defined for all m in $[0, 1]$. Eq. (2.21) predicts that for $\kappa < 0$ the completely demethylated state $\bar{m} = 0$ is stable, whereas for $\kappa > 0$ the stable state is $\bar{m} = 1$. At $\kappa = 0$ the system undergoes a discontinuous transition for every value of σ.

In Fig. 2.9 are shown three representative kymographs of the infection dynamics at the transition, for different values of the contact exponent σ. These kymographs are obtained by evolving our microscopic models, detailed in section 2.6.1, on a lattice of length $N = 10^4$, choosing $\bar{\mu}/\bar{\lambda}$ so that the system is at the transition between fully marked and unmarked phases. As already mentioned, the initial condition is a single infected seed located at $x = 5 \cdot 10^3$. For $\sigma = 0.4$ we see that no compact domains can form; instead, the

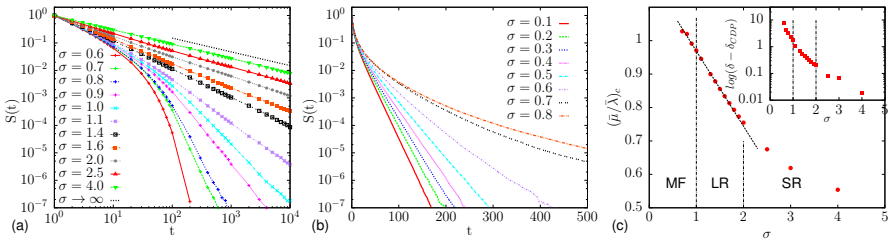

Figure 2.10. Survival probability and transition point. (a) In this panel we show the survival probability in a log–log plot, varying σ in the range $[0.6, 4.0]$. For $\sigma > 2$ the exponent tends to the value predicted by the CDP model ($\delta = 0.5$). As σ decreases the exponent δ increases, and the transient is longer. **(b)** For $\sigma < 0.6$ no power laws are observed. Instead, the decay is exponential (here we plot the survival probability in a log–linear scale). For $\sigma \gtrsim 0.6$ the exponential decay disappears, instead a power law behaviour takes place. **(c)** The location of the transition point increases as σ decreases. For $\sigma > 2$ (short–range regime, SR) the value of $(\bar{\mu}/\bar{\lambda})_c$ decays slowly to 0.5. For $1 < \sigma < 2$ (long–range regime, LR) $(\bar{\mu}/\bar{\lambda})_c$ the transition point scales linearly with the contact exponent. For $\sigma < 1$ (mean–field regime, MF) the location of the transition point is still compatible with the linear scaling, but cannot be computed for $\sigma < 0.6$. In the inset we show the divergence of the survival exponent δ as a function of σ.

infected/methylated domains are full of holes due to the long-range infection process, and at late times the system falls into the absorbing state $\bar{m} = 0$ (Fig. 2.9a). For $\sigma = 1.5$ a compact domains opens in the middle of the lattice: this can also create additional domains by infecting regions of the lattice far away from it, since the interactions are still long–range (see Fig. 2.9b). For $\sigma = 4.0$ the infection profile looks very similar to a typical CDP kymograph, with a single fluctuating compact domain which can virtually persist forever (Fig. 2.9c).

In Fig. 2.10a we show the survival probability as a function of time, for different values of $\sigma \geq 0.6$. For $\sigma > 2$ the slope approaches the value predicted for the SVM ($\delta = 0.5$). Indeed, one would expect that for $\sigma \geq 2$ the substitution $\nabla^\sigma \to \nabla^2$ holds and Eq. (2.21) becomes the well-known equation of CDP. This is, however, not the case: the system displays a smooth crossover between the long range and the short range behaviour, with the latter being re–established only at $\sigma \gtrsim 4.0$, as shown in Fig. 2.10c and its inset. A similar crossover has been found in other models with Lévy flights, such as the LRDP [132] and the LRVM [136]. Interestingly, we note that approximately the same value for the crossover which we obtain was found in Ref. [136] for the LRVM. Moreover, it is known that the SVM and CDP belong to the same universality class in $d = 1$, but not in other physical dimensions.

For smaller values of the contact exponent than those considered in Fig. 2.10a (i.e., for $\sigma < 0.6$) no power law can be detected in our simulations, see Fig. 2.10b. This fact may be due to finite-size effects, which can be strong for such long

interactions. Nevertheless, we observe that, after a very small transient, the survival probability decays exponentially, $S(t) \sim \exp(-k(\sigma)t)$. This exponential decay is absent for σ greater than $0.6 - 0.7$. Such crossover is highlighted in Fig. 2.10b, where we replot the survival probabilities for $\sigma = 0.7, 0.8$ in log-linear scale, which do not present any exponential decay. This change in behaviour is compatible with an analysis of the Reggeon field theory [134], which predicts a mean-field behaviour for $\sigma < 1$, as the real dimension d is above the upper critical dimension $d_c = \sigma$.

We close this section by discussing how we can map the microscopic rates $\bar{\lambda}$ and $\bar{\mu}$ to the effective parameter κ entering the phenomenological theory in Eq. (2.21). To do so, we note that, in the limit $\sigma \to +\infty$, the state of a lattice site, say the i-th, can only change according to its two neighbours; a single infection event occurs if a (recovered) boundary site is in contact with the methylated neighbour (probability $1/2$), otherwise it remains unmethylated (probability $1/2$). Conversely, an erasure event occurs at the methylated boundary site with probability 1. Hence, $\kappa = 0$ only if $\bar{\mu} = \bar{\lambda}/2$, and we conjecture the following form for $\kappa(\sigma)$:

$$\kappa(\sigma) = \left(1 - \frac{\bar{\mu}}{\bar{\lambda}} f(\sigma)\right) , \tag{2.22}$$

where $f(\sigma)$ is a monotonic function which satisfies $f(+\infty) = 1/2$. Such behavior is validated by simulations: in Fig. 2.10c we show the location of the transition point as a function of σ. For $\sigma < 2$ the transition point scales linearly with the contact exponent, whilst for $\sigma > 2$ it decays slowly towards $\bar{\mu}/\bar{\lambda} = 0.5$. This indicates that $\sigma \simeq 2$ is the crossing point between the short–range regime and the long–range behaviour. Interestingly, the value of δ seems to diverge for smaller and smaller values of σ (see inset in Fig. 2.10c); such divergence signals the crossover to the exponential decay regime shown in Fig. 2.10b.

2.6.3 Phase diagram of the system and connection with epigenetics

We now consider a more general case where all parameters are non-zero, so that there are no absorbing states in the system ($q_\mu, q_\lambda \neq 0$). This is likely to be more realistic as it includes the case where there is some generic biological noise. In line with previous models for bistability in epigenetic patterns [84, 91], we consider a small chromatin region, with $N = 100$ beads/nucleosomes. This is a realistic size to study, for instance, stochasticity in epigenetic domains in yeast [84, 139], or a single small chromatin domain of about 20 kbp in larger genomes (as in mammals).

To render the exploration of parameter space feasible, we set $q_\mu = q_\lambda = q_0$, and $\bar{\lambda} = \bar{\mu} = 1 - q_0$, and we study the behaviour of the model by varying the

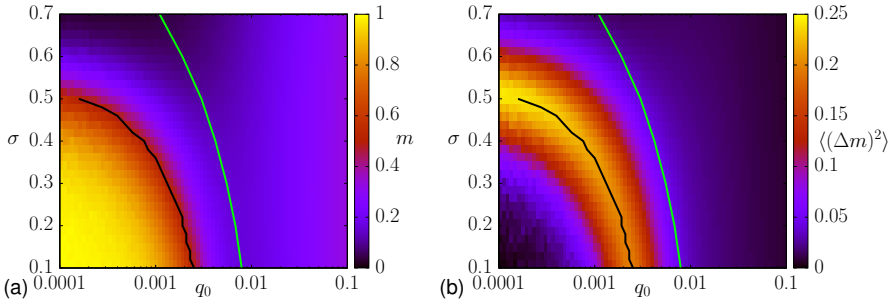

Figure 2.11. Phase diagram. (a) Phase diagram as a function of q_0 and σ for a system with $N = 100$. A transition between a marked and an unmarked regime can be seen; the black line correspond to systems where $\bar{m} = 0.5$. The green line instead separates a coherently unmarked regime from the disordered, or mixed, regime. **(b)** Variance of the fraction of marked sites in the system, $\langle(\Delta m)^2\rangle$. This quantity peaks close to the transition between the marked and unmarked regimes highlighting a region of enhanced bistability.

parameters q_0 and σ. Such assumption implicates that the baseline rate due to spontaneous conversions in real cell, as well as the writing and erasure rate due to recruitment, are symmetric. The symmetry of the competing epigenetic mechanism has been often considered in previous model, such as the ones described in section 2.4. Physically, q_0 can be viewed as a temperature-like parameter that regulates the baseline marking/unmarking rates, and therefore, the ratio noise–to–recruitment is $q_0/(1-q_0)$. Additionally, the biologically relevant values of σ are between 0 and 1 [140]. As this exponent should be associated with the looping probability of a polymer representing the chromatin fibre [141], the value of $\sigma = 0.5$ corresponds to looping of a random walk, $\sigma \simeq 1$ corresponds to looping of a self-avoiding walk [141], whereas $\sigma \simeq 0$ describes the decay of contact probability with genomic distance in a crumpled (or fractal) globule [142, 143].

In Fig. 2.11a we present the phase diagram obtained for different values of q_0 and for σ between 0.1 and 0.7. To find this phase diagram, we used a truncated Lévy distribution to simulate long-range infection – i.e. $P(|i-j|) = A|i-j|^{-(\sigma+1)}$, where $A = \sigma/(1-(N/2)^{-\sigma})$. This procedure was employed in order to limit boundary effects.

The phase diagram in Fig. 2.11a shows three distinct phases. First, for small value of q_0, although no real absorbing state is present, the system still tends to reach a typical state with large $\bar{m}$ (methylated regime), or with small $\bar{m}$ (demethylated regime). For sufficiently small values of σ, or equivalently sufficiently long interaction range, long–range methylation dominates over demethylation (bottom left region in the phase diagram). For larger σ, the fact that the erasure is more likely at domain boundaries tilts the balance in favour of demethylation, and $\bar{m} \simeq 0$ (top left region in the phase diagram). Finally, at

sufficiently large values of q_0, mean field theory applies. The latter predicts that methylation and demethylation should balance giving $\bar{m} = 1/2$ in steady state (as we consider $\bar{\lambda} = \bar{\mu}$). This is the mixed regime, where methylated and demethylated sites coexist in a disordered system, and it is found to the right of the green line in Fig. 2.11a. Note that, by using a truncated Lévy distribution, one can also extend the phase diagram in Fig. 2.11 to negative value of σ. Negative values of the contact exponent of finite domains of chromatin are often found in microscopy experiments and Hi-C data [144]. However, for $N = 100$, we do not expect any different qualitative behaviour from the diagram line $\sigma = 0.1$, since we are in the mean-field limit for which all the sites connect with each other. Moreover, such values for σ would be forbidden in the limit $N \to \infty$, since Eq. (2.8) would not be integrable. At larger N we expect a shift towards larger σ and q_0 the transition line between coherent and mixed regime. However, further analysis would be in order to infer the behaviour for $N \to \infty$.

2.6.4 Effective potentials and zero–dimensional equation

The line separating the coherent – either marked or unmarked – regime from the mixed one (which we call the coherence transition line) can be mapped out by analysing the probability distributions of $\bar{m}$ in steady state, for different parameter values, as shown in Fig. 2.12. From here on, we will use $m \equiv \bar{m}$, since we will always discuss the mean field without ambiguity.

In Fig. 2.12 we show the effective potentials $V \propto -\log(P(m))$– where $P(m)$ is the probability distribution function for m in steady state – in the different regions, for the representative values of $\sigma = 0.2$ and 0.6. In the methylated phase/regime (for $\sigma = 0.2$, $m \sim 1$) the global minimum is located at $m = 1$ (see Fig. 2.12a). For $\sigma = 0.6$, we are in the demethylated phase, and the global minimum is now located at $m = 0$. Increasing q_0 at small σ, the system first becomes demethylated (Fig. 2.12c, $\langle m \rangle \sim 0.3$), while for sufficiently large values of q_0 the effective potential has a global minimum for $0 < m < 1$ (Fig. 2.12d), and the system is in the mixed phase. The green line in Fig. 2.11a provides the boundary between the coherent regime (minimum at either $m = 0$ or $m = 1$) and this mixed phase/regime. It should be noted that, outside the mixed regime, we always find two local minima at $m = 0$ and $m = 1$, so that both these two states are always at least metastable in the coherent regime. Besides this widespread bimodality, there is a robust bistability region close to the transition between the methylated and the demethylated regime: this is apparent from Fig. 2.11b, which shows the variance $\langle (\Delta m)^2 \rangle$ as a function of q_0 and σ. Bistability arises due to the proximity to the methylated–demethylated transition. Because the $P(m)$ distributions are always bimodal, the transition

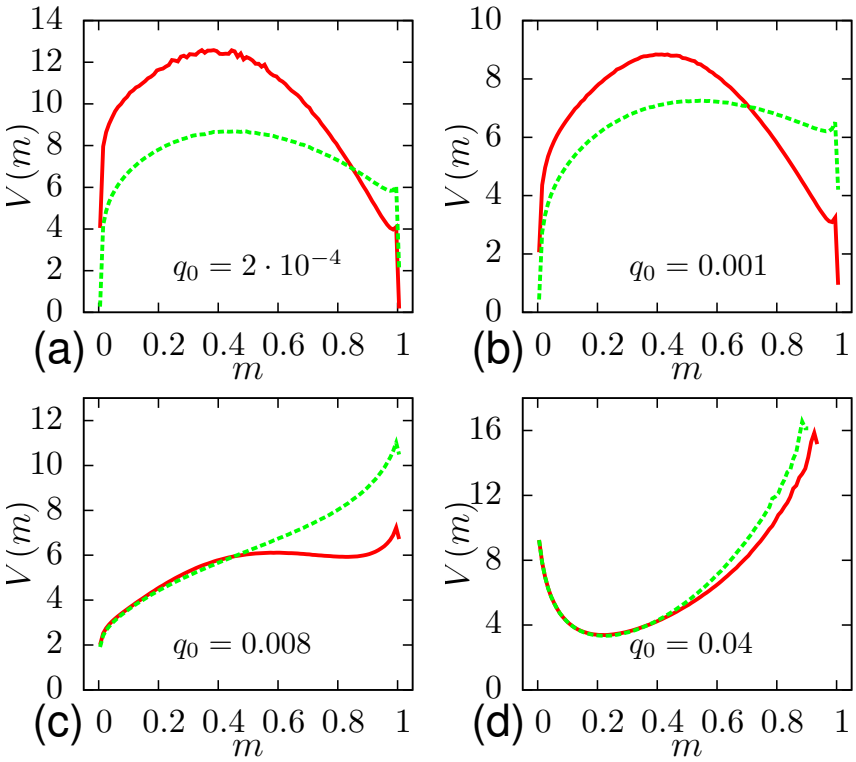

Figure 2.12. Methylation effective potentials. In this figure we show some representative plots of the effective potential $V(m)$, for $\sigma = 0.2$ (red solid line) and $\sigma = 0.6$ (green dashed line). For $\sigma = 0.2$, the increasing values of q_0 we consider cross both the marked/unmarked transition line and the coherence transition line. For $\sigma = 0.6$, the values of q_0 we consider cross only the coherence transition line. **(a)** For small q_0 the distribution is bistable, with the global minimum located at $m = 1$ and $m = 0$, respectively for $\sigma = 0.2$ and $\sigma = 0.6$. **(b)** The increase of q_0 produces a change in the weights associated with the two minima. **(c)** After the transition point the global minimum becomes $m = 0$ in both the cases even if for $\sigma = 0.6$ we are already beyond the coherence transition line. **(d)** For $q_0 = 0.04$ the system is in the mixed phase and the resulting unimodal distributions depend only very weakly on σ, since (2.8) would not be normalized.

is sharp and first-order like, hence coexistence (and bistability) naturally arise near the critical line. In the bistable region, therefore, our model predicts that epigenetic domains should be highly stochastic and may switch over time.

To better understand our simulation results, we now discuss a simple analytically tractable approximation, in line with the zero–dimensional model in section 2.4.1. Neglecting the spatial dependency of the order parameter (gradient term) in Eq. (2.21), we can write down a dynamical equation for the global methylation in our model, m, as follows,

$$\partial_t m = q_0(1-2m) + \kappa m(1-m) + \sqrt{Qm(1-m)}\zeta. \tag{2.23}$$

The first term in Eq. (2.23) allows *recovery* from the $m = 0$ and $m = 1$ states, which are then no longer absorbing state. Additionally, Q is the strength of the multiplicative noise, and ζ is a Gaussian random variable, with $\langle \zeta(t) \rangle = 0$ and $\langle \zeta(t)\zeta(t') \rangle = \delta(t-t')$. Note that we have neglected also the basal (additive) noise. Such additive noise would arise if one cosiders the finite number of writer–eraser complex, which renders the recovery from the $m = 0$ and $m = 1$ states stochastic. We want to solve the Fokker–Planck equation associated with Eq. (2.23), for the probability $P(m,t)$ that the system has a global methylation m at time t. This equation reads as follows

$$\begin{aligned} \partial_t P(m,t) &= \partial_m^2 \left[\frac{Qm(1-m)}{2} P(m,t) \right] \\ &- \partial_m \left[(q_0(1-2m)P(m,t) \right] \\ &- \partial_m \left[\kappa m(1-m)) P(m,t) \right] \\ &\equiv -\partial_m J(m). \end{aligned} \tag{2.24}$$

By imposing no-flux boundary condition, $J(m) = 0$, we obtain the following equation for the steady state distribution $P(m)$:

$$\partial_m K(m) = \frac{2\left[q_0(1-2m) + \kappa m(1-m)\right]}{Q\left[m(1-m)\right]} K(m)\ , \tag{2.25}$$

where we set $P(m) \equiv (Q/2)m(1-m)K(m)$. Then, the stationary probability distribution reads:

$$\begin{aligned} P(m) \sim \frac{1}{m(1-m)} \times \\ \exp\left(2 \int^m \frac{q_0(1-2m') + \kappa m'(1-m')}{Qm'(1-m')} dm' \right)\ . \end{aligned} \tag{2.26}$$

After some algebra, we find that the system is therefore described by the following effective potential $V(m)$:

$$V(m) \propto (1-q_0)\tilde{\kappa}m - (q_0 - Q)\log\left[m(1-m)\right] \ . \tag{2.27}$$

Since $\bar{\lambda} = \bar{\mu} = 1 - q_0$ in our microscopic model, we have assumed $\kappa = (1-q_0)\tilde{\kappa}$. Therefore, there are three parameters, which are fundamental in the mean-field treatment, namely q_0, κ and Q. For $q_0 < Q$, $P(m)$ is bimodal, since $V(m)$ is minimised for $m = 0, 1$ – where $V(m)$ diverges. Conversely, for $q_0 > Q$, $V(m)$ becomes unimodal and it is minimised for

$$m_{\min} = \frac{1}{2} - \frac{x}{\tilde{\kappa}} + \frac{|\tilde{\kappa}|}{\tilde{\kappa}}\sqrt{\frac{1}{4} - \left(\frac{x}{\tilde{\kappa}}\right)^2}, \tag{2.28}$$

where $x \equiv (q_0 - Q)/(1-q_0)$. Thus, the line $q_0 = Q$ determines a transition line separating the region where the probability is bimodal, and the system is coherent, from the unimodal region where there is a single stable state. Note that $\tilde{\kappa}$ determines the methylation level in the latter case: if $\tilde{\kappa} > 0$ then $m_{\min} > 1/2$, otherwise $m_{\min} < 1/2$. In all cases, $m_{\min}$ approaches $1/2$ as q_0 tends to 1. In Fig. 2.13 we present a sketch of the line separating the bimodal (epigenetically coherent) from the unimodal/mixed regime according to the analytical theory in Eq. (2.23). Note that such boundary is independent of the value of $\tilde{\kappa}$. Additionally, if we postulate that the sign of the linear term in Eq. (2.28) changes, for instance with σ, $\bar{\lambda}$, and $\bar{\mu}$, we can describe cases where the coherent regime may correspond to either a methylated or demethylated phase (top-left inset in Fig. 2.13).

2.7 Conclusion and future perspective

In summary, we have proposed here a simple one-dimensional stochastic model for the dynamics of mark spreading and erasure inspired by the problem of how the epigenetic marks spread along the genome. This is distinct from previous models (some of which we have reviewed in section 2.4.1), that typically consider the case where two or more different marks compete within the same region. Within our model, methylation is viewed as a long–range "infection" process mediated by the looping of the chromatin substrate. Importantly, while infection is long range, we assume that the erasure/recovery occurs locally and is enhanced at domain boundaries. In our microscopic model, we varied the exponent σ, regulating the global folding of the chromatin polymer (and hence the range of the infection process), as well as the baseline rate of spontaneous methylation/demethylation, q_0, which we have taken to be equal. The phase diagram in the (q_0, σ) plane shows two distinct transition line. First, there is a

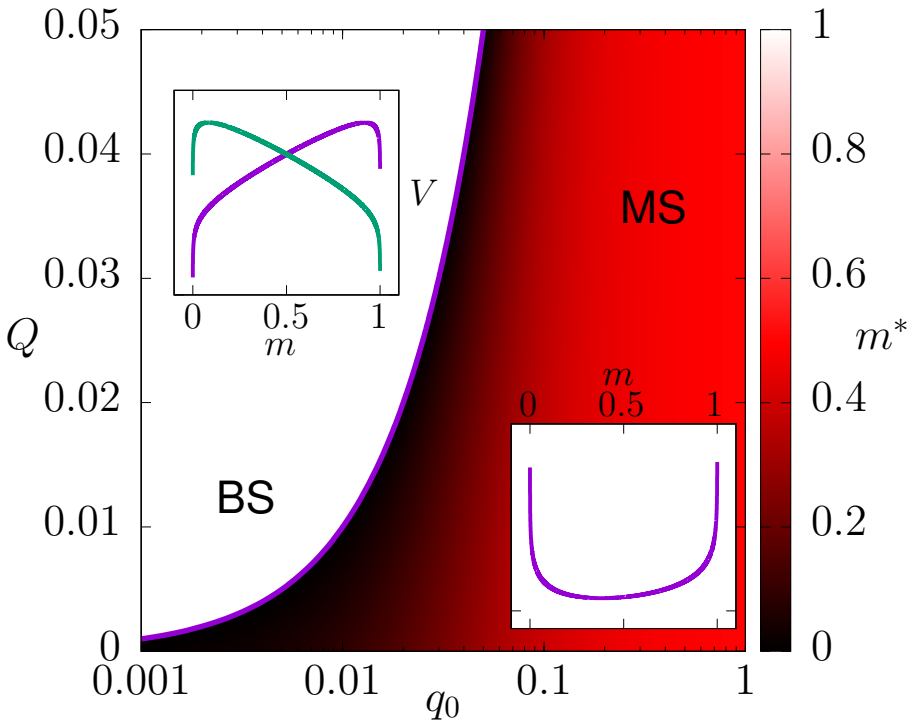

Figure 2.13. The bistable-unimodal transition, or the coherence transition line. This phase diagram displays the transition between a bistable (BS), or coherent, phase and a monostable (MS), or mixed phase, as they are predicted by the mean-field theory. In the first phase (white) the two minima of the effective potential $V(m)$ are $m = 0, 1$. In inset we show a typical bistable potential, for $q_0 = 0.001$, $Q = 0.01$, in both the cases $\tilde{\kappa} > 0$, green line, and $\tilde{\kappa} < 0$, purple line ($|\tilde{\kappa}| = 0.1$). In the latter phase only one minimum is present, and its location $m_{\min}$ varies from 0 close to the transition to 0.5, for $q_0 = 0$ (if $\tilde{\kappa}$ is negative). In inset we show a typical unimodal potential, for $q_0 = 0.1$, $Q = 0.01$ and $\tilde{\kappa} = -0.1$. Colors in the unimodal (mixed) regime are related to the value of $m_{\min}$ for a fixed value of $\tilde{\kappa} = -0.1$.

transition between a coherent regime (either methylated or demethylated), and a disordered, or mixed, regime (where marked and unmarked nucleosomes coexist with no domain formation). Second, in the coherent regime there is a transition between a methylated phase, when long–range infection is efficient (small σ), and a demethylated phase when infection is shorter–range (large σ). Increasing q_0 favours the demethylated phase in our parameter range: we interpret this as due to the boundary erasure term which tips the balance in the favour of demethylation, when σ is sufficiently large, and for realistic system size, say, of a hundred units/histones.

We have shown that the transition between coherent and mixed regimes can be understood on the basis of a simple stochastic differential equation which is analytically tractable, and which predicts there is a qualitative change in the

nature of the effective potential governing the steady state behaviour of the system. At low q_0, the potential has two local minima ($m = 0$, fully demethylated, or $m = 1$, fully methylated, whereas at larger q_0 noise dominates and there is a single minimum at intermediate m. The transition between methylated and demethylated phase, on the other hand, can be understood qualitatively as a transition between absorbing states in the limit where $q_0 \to 0$. In this limit, we have shown that the model is mappable onto a special case of the contact process, known as long–range compact directed percolation, which encompasses the voter model with a long–range interaction. If we modify rates such that only one absorbing states remain, the universality class changes and becomes, as expected, that of "standard" (i.e., non-compact) long–range directed percolation. We should remark that the mapping between this system and other paradigmatic models such as 'voter model', LRDP and LRCDP can inform on the fundamental nature of the epigenetic process: indeed, two mapped different process share the same scaling laws at criticality, irrespectively of the details of both systems.

The first-order nature of the transition (in the $q_0 \to 0$ limit) endows the system with bistability and hysteresis close to the critical line. This provides a pathway to the establishment of epigenetic memory — the phenomenon through which a chromatin region "remembers" its state even following a relatively strong perturbation. Interestingly, and in stark contrast with other models, bistability does not arise due to SSB and, in the unstable regime, also the unmarked phase can retain memory of its state.

It is thus tempting to speculate that cells may tune the σ and q_0 parameters, which are associated with the conformational changes of chromatin and the bare affinity of the writing/erasing enzymes, so as to control the variability in the epigenetic patterns and thus, in turn, the variability in gene expression within the same population. Recent experiments in yeast support the idea that this may be a biological bet-hedging strategy for survival against random attacks [145].

This model is a good starting point for more sophisticated models. Analogously to [82, 87], one could try to model this infection–like process onto a polymer in a three–dimensional setup. Here, the advantage would be to have long–range contacts directly into the model, without explicitly modeling the Lévy distribution as in the one–dimensional version.

Chapter 3

Solid and glassy behaviour of chromatin–binding proteins

We now take a step further into the study of the physics of genome. In eukaryotes, the chromatin fibre must be highly folded in order to fit inside the nucleus (10μm of diameter). However different levels of compaction and order can serve different purposes: as already mentioned in the previous chapter, tightly packaged chromatin regions are often transcriptionally repressed, whereas more swollen regions are usually transcriptionally active. Additionally, the presence of some "regularity" in the nucleosome organisation can have a crucial implications for epigenetic territories estabilishment and chromatin functions [146, 147]. In light of this, it is important to understand more about the level of chromatin compaction and the underlying biophysical mechanisms.

Whether regular chromatin structures can emerge *in vivo* at the 30–nm length scale was one major question which arised in the early 80, since J. T. Finch and A. Klug speculated that the vast majority of genome can be composed of 30–nm fibers, which is formed by coiling the 10–nm fibre into rather compacted ordered structures [148]. Later, Several works tried to solve this puzzle [146, 149–151], but modern experiments seem to exclude that an ordered compacted fibre can form *in vivo*. Nowadays, The common view has veered towards a different understanding of chromatin organization: rather than due to nucleosome stacking and intra–fibre contacts, 'heterochromatic' chromatin fibre can form and stabilise due to associated proteins, such as cohesin, heterochromatin protein 1, etc. which bind to chromatin and can determine its transcriptional state. Accordingly, focus has shifted towards the understanding of the features of chromatin–protein aggregates.

In the late 80s the idea of a "liquid–like" aggregation of the 10–nm fibre were already taking place [152]. More recent papers seems to support the formation of liquid chromatin drops [153–156], mainly highlighting that nucleosome are irregularly assembled at the 30–nm length scale. Concomitantly, a vast literature point to the fact that interacting chromatin–binding proteins

facilitate chromatin compaction via *liquid–liquid phase separation*. This latter has become the most touted argument to explain heterochromatin formation and co–localisation of chromatin and associated factors. However, relatively little is known about chromatin dynamical behaviour. From a physical standpoint, chromatin dynamics must be largely different from that of simple diffusing particles in liquid droplets: as a consequence of its polymeric nature, it should be always sub–diffusive. In principal, also chromatin–binding protein droplets could exhibit behaviours that are not strictly liquid, depending on their self–interacton or chromatin affinity [157].

Indeed, several cases of maturation of "liquid" protein droplets toward gel or solid state has also been recently reported [158, 159]. Amyloid–like fibers, and other proteins containing low complexity domains, can form gels, which are non-dynamic and rather stable [160]. Yeast protein Pab1 forms glassy droplets that play an adaptative role, promoting cell survival upon heat shock [161].

Similarly, one could ask if the same applies for chromatin–binding aggregates: are those clusters liquid or not? Is there a coexistence of a mobile and an immobile population, and if so, what determines such an internal distinction? How does this affect chromatin dynamics and structure? In this chapter (and lateron in chapter 4) we propose a simple mechanism that entails diverse dynamical regimes for chromatin–binding proteins and chromatin itself. The model is based on a physical phase separation induced by "molecular bridges", which are multivalent proteins capable of joining distal regions of the chromatin fibre. This "universal" mechanism can explain how chromatin compact together with some chromatin–associated proteins, which do not self–interact [162–164]. A similar phase separation was previously investigated by experimental models with gold nanoparticles that interact with double–stranded DNA [165, 166]. In this project, we extensively explore the internal dynamical properties of such proteins clusters, characterising proteins mobility and local ordering.

The chapter is organised as follows: first, in section 3.1, I will describe the "conandrum" of the 30–nm fibre; this section serves to show the modern view of chromatin organisazion: chromatin fibre is more likely to be in an "interdigidated state" than folding into regular structures. In section 3.2, I review the main aspects of liquid–liquid phase separation, starting from the classical problem of binary mixtures; then, I briefly list all living organelles which are thought to undergo LLPS, and, in particular, we briefly discuss on the role of cohesin, which can cluster without self–interacting. The aggregation of cohesin *in vitro* can be explained by another type of phase separation, which I will refer to as bridging–induced phase separation (BIPS): this will be the focus of section 3.3. First, we review this simple mechanism, which leads to several

phenomenological behaviours; then, we show that several states of matter can be associated with the structural and dynamical properties of living aggregates; finally, we show a piece of evidence related to BIPS of cohesin, *in vitro*. All the aforementioned sections are part of the literature background. In section 3.4.1, I describe the main feature of my work. A large section has been dedicated to methods and a mean field theory, which will serve to clarify the following sections; then, I report the main results obtained by simulating the dynamics of proteins (modelled as spheres) interacting with a chromatin fibre (modelled as a bead polymer) and undergo bridging–induced phase separation.

3.1 Chromatin organization: the conundrum of the 30–nm fibre

In section 2.1, I have described the bead–on–a–string model for chromatin, whose primary structure consists of a nucleosome and linker DNA, which repeats to form the extended 10–nm chromatin fibre. However, in mammalian cells, the 10–nm chromatin fibre is more likely to be compact, as several structural associated proteins favour looping, inter– and intra–fibre binding [154, 167]: linker histone [168], transcription factors, RNA polymerase II, CCCTC–binding factor (CTCF) [169], cohesin [170] and *heterochromatin protein 1* (HP1) [171] are known to co–localise with chromatin. During interphase (the phase between one mitotic phase and the next), chromatin assembles and organizes into chromosomes; different degrees of compaction are needed to either favour accessibility of some chromatin region (euchromatin) or compartimentalise silenced chromatin stretches into compact domains (heterochromatin). Thus, the following question arises: can the 10–nm fibre adopt an ordered secondary structure? At first glance, some ordering of the 10–nm fibre seems to be required in order to achieve chromatin organization [148]. Many structures were speculated, from the 'one–start' helix to the more complicated 'radial loop' structure, as shown in Fig. 3.1.

Nonetheless, the definitive conformation of chromatin at the 30–nm level remains elusive. First of all, the particular structure adopted is too sensitive to the length of the linker DNA [150, 173] and on the presence of the linker histone H1 [174]; *in vitro* experiment show that different salt concentrations produce different chromatin conformation [172] (see also Fig. 3.2). Simulations on model 10–nm fibre predict all possible regular structures, but only at null temperature [100]. More importantly, cryo–electron microscopy (cryo-EM), small–angle X–ray scattering (SAXS) and super–resolution microscopy seem to exclude the existence of regular, hierarchical structures of the 30-nm fiber, at physiological condition [172]. A combined–technique work by Nishino *et*

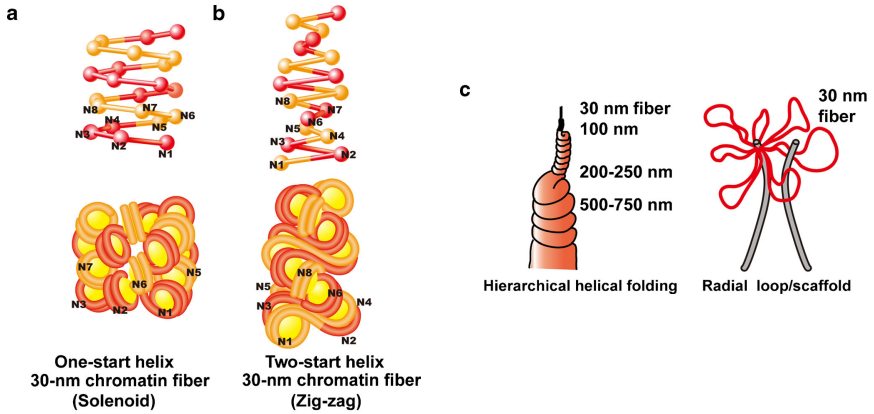

Figure 3.1. Classical models of ordered 30–nm chromatin fibre (a) One–start helix (solenoid) and **(b)** two–start helix (zig–zag) [atoms view (top) and cartoon (bottom)]. **(c)** Two classical higher order chromatin structure models: the hierarchical helical folding model (left) and the radial loop model (right). In the radial loop model, many loop structures of the 30-nm fiber (red) wrap around the scaffold structure (gray), which consists of condensin and topoisomerase IIα. Figures adapted from [154]

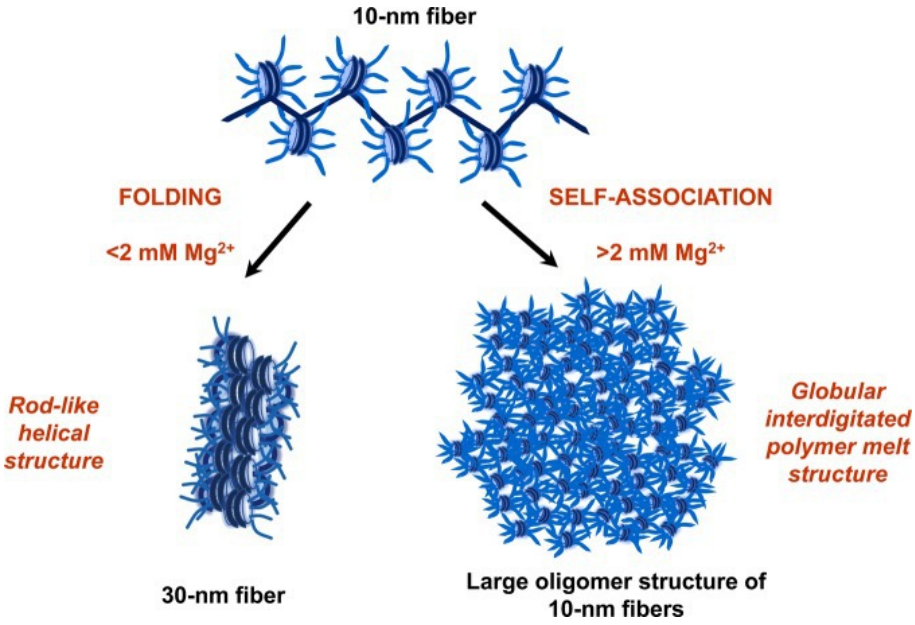

Figure 3.2. In vitro conformations of the chromatin fibre. 10–nm fibre [extended on top] can fold differently, depending on the salt concentration (Mg^{2+}). At low salt concentrations, chromatin folds into a 30–nm solenoid (bottom–left schematics). For higher salt concentration, a globular polymer melt structure forms (bottom–right schematics). Figure taken from [172].

al. [175] predicts controversial reults: they concluded that mitotic chromosomes are more likely to be arranged in a disordered, fractal way, even if the same work present evidence of prominent peaks at $\sim$ 11 and $\sim$ 6 nm in SAXS analysis for erythrocyte nuclei, as expected for a 30–mm two–start helix.

In summary, the absence of experimental findings of the 30–nm structural hierarchy led the common view to 'migrate' towards the idea that a disordered, interdigited 'polymer melt' characterise the nature of interphase and mitotic chromosomes *in vivo*, [schematics in Fig. 3.2]. In light of this, several biological papers suggest that chromatin, and the associated binding factors and proteins, can compartmentalise due to a liquid–liquid phase separation (LLPS). This latter concept is the object of the next Section.

3.2 Chromatin compaction and nuclear bodies: LLPS or something else?

Liquid–liquid phase separation (LLPS) is the physical mechanism that causes the demixing of coexisting miscible components. That is, in spite of the entropic loss entailed by the formation of two or more distinct phases, the energetic gain of phase separation is much more advantageous, in some region of the parameter space (component concentration, temperaure, pressure, pH, crowding agents, etc). Several biological and physical papers point to the fact the LLPS is responsible of the aggregation of proteins, and, possibly, to the compaction of region of chromatin. Clearly, such a mechanism would provide an general, biochemically independent explaination of how chromatin conformation and chromatin function. Here we present a simple case of LLPS, based on the most general theories of binary fluids and first–order phase transition [176, 177].

In a binary fluid, two fluids of type A and B, of concentrations ϕ_A and ϕ_B, with attractive self–interaction of strength ϵ_{AA} and ϵ_{BB} respectively, and an interaction between different species, of strength ϵ_{AB}. Thus, these interactions would separate the system into A–rich and B–rich regions, dependently on the temperature T. The critical temperature is a function of the microscopic interaction strength and concentrations. In general, at large temperatures the mixing entropy overcomes the tendency to separate, and the equilibrium phase is uniform and mixed. At lower T, demixing occurs.

In this latter scenario, when $\phi_A \ll \phi_B$ – which is always the case when considering a single macromolecular species (species A) in water (species B) – LLPS can occur through two mechanisms, depending on the initial conditions and on the quenching procedure. At intermediate temperatures, i.e. in the *coexistence* region, the two distinct phases estabilish via *nucleation and growth* of the species A, i.e. a random, thermal–driven fluctuation of the density of

component A in space leads to the formation of a droplet, or *nuclei*, through a condensation mechanism. If the droplet is large enough, the system escapes the metastable uniform phase, and the a A–rich domain grows. In a three–dimensional isotropic system, spherical domains are preferred, as to minimize the interfaces between the two phases.

At lower temperatures, in the region subtended to the spinodal curve, a fast phase separation occurs via the *spinodal decomposition*: even a small perturbation at very large frequencies, or small length scales, propagates through the system, creating several nuclei which grow into larger droplets. No nucleation barrier needs to be overcome in the phase transformation, since the uniform phase is unstable. When large droplets estabilish via spinodal decomposition, then a full phase separation can be reached due to *coalescence* of clusters. Another route to full phase separation is the so–called *Ostwald ripening*, where material is transported from small to large clusters, due to chemical potential gradients which estabilish between small and large droplets. As shown for instance in Ref. [176], all these mechanisms, during all stages, predict a scaling law of $L(t) \sim t^{1/3}$ for conserved order parameters (Model B), with L the length scale of fluctuations (the radius of the forming droplet).

An old–fashioned (but very interesting and detailed) review on the LLPS and, more in general, on the physics of phase transitions is in Ref. [177].

3.2.1 LLPS, chromatin compaction and connection with cohesin

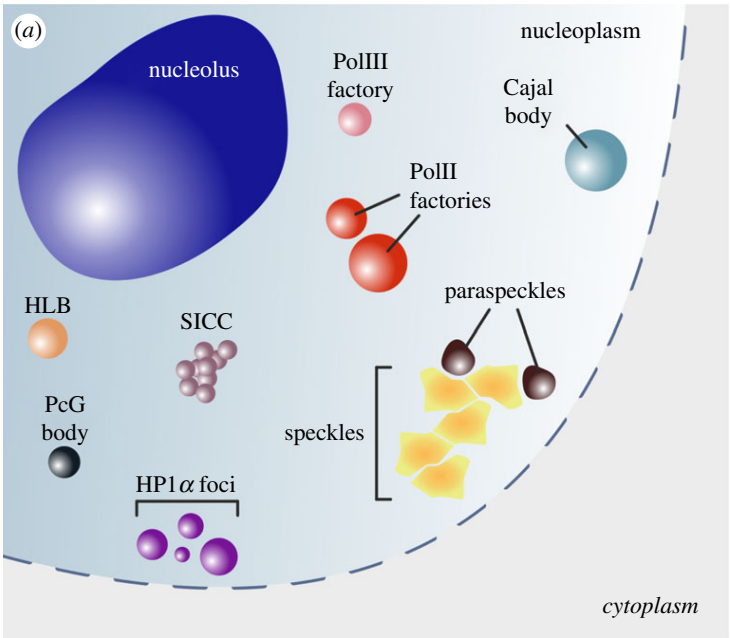

Figure 3.3. Membraneless organelles. Figure taken from [176].

LLPS has recently much been touted to explain the aggregation of cellular molecules into membraneless organelles [178–180]. In the nucleus, segregation of chromatin–related proteins and chromatin compaction are often tightly related. The development of modern techniques like genome–wide chromatin organization analysis (Hi-C) has improved our understanding of chromatin organization; chromosomes are generally divided into alternating Mbs–long compartments [126, 181], or 'territories' [182]: A–/B–compartments (active and repreessed chromatin regions, respectively) and TADs (topological associated domains). The former have been found to be quite dynamic, as, in mammals, 36% of the genomes changes compartment during development [182]. In Ref. [183], correlations between methylation patterns and A–/B–compartments have been revealed, suggesting that chromatin territories are not static.

However, LLPS does not require any interaction between DNA or chromatin, and chromatin–binding factors. Indeed, some of the molecules represented in Fig. 3.3 are thought to assembly without interacting with chromatin. The nucleolus is the paradigmatic example: an aggregate of several proteins and RNA molecules which is involved in the production of ribosomal units, which display a rapid recovery after fotobleaching, exchange of molecules with the environment and coalescence of smaller droplets into a larger one. Also paraspeckles and Cajal bodies can cluster without chromatin [184].

Although LLPS represent the prime candidate mechanism which explaines macromolecular crowding and assembly, it cannot be the relevant for agglomeration of non–interacting proteins. In this respect, *cohesin* is a vivid example. Cohesin has been implicated in a wide range of functions, such as pairing of homologous chromosomes during meiosis, formation and repair of DNA breaks and transcriptional control [185]. In particular, cohesin guarantees chromatin compaction due to *loop extrusion* [186]

Self–interaction of cohesins dimers have not been detected in yeast (except for Ref. [187], where the authors claim that in budding yeasts cohesin can oligomerize both on and off chromatin), and in human cells is not well known [188]. In spite of this, there is evidence that cohesin phase separate *in vitro*, or at least it co–localise with clustering proteins, such as transcription factors [189]. Since cohesin clustering function as support for the estabilshment of droplets of other proteins, another mechanism, different from LLPS, that explains the aggregation of such macromolecules in the proximity of DNA is in order. The *bridging–induced phase separation* (BIPS) seems to fill this gap: a new type of phase separation, which can entail the formation of spherical–like droplets without self–interaction among single constituents.

3.3 BIPS: a physical mechanism for chromatin–mediated phase separation

BIPS is a physical mechanism which leads to phase separation of multivalent chromatin–binding proteins or complexes that can only interact with chromatin. When binding to chromatin, proteins can bridge between different chromatin regions, as the chromatin fibre is capable of folding in the three–dimensional space. Thus, the local density of chromatin increases at the binding sites, creating attractive basins for other proteins which leads to further protein binding in those regions. Without any mechanism that hinder this positive feedback loop, the protein *foci* grow, and eventually coarsen until a single protein–rich droplet remains. Such a mechanism, often called polymer–polymer phase separation (PPPS), was firstly introduced in [130, 190, 191], then therotically studied in [162].

For non–specific binding, which means that proteins can bind everywhere along the fibre with the same affinity, a *macrophase* separation occurs; namely, independently of the number of proteins, different nuclei will grow and, eventually, coalesce into a single droplet. Further studies on the properties of BIPS have revealed that two different competing mechanisms can lead to *microphase* separation, namely the arrest of the droplet coalescence:

1. at equilibrium, the insertion of *specific*, equispaced, strong binding sites for proteins, joined by larger stretches, say of lenght l, of weakly binding chromatin. Proteins will accumulate close to the specific sites, promoting bridging between them, and the formation of loops of length $\approx l$. However, if the number of proteins is sufficiently small, clusters only grow up to a self–limiting size. This is due to the 'entropic' repulsion among loops, which prevent different clusters from coalescing [164, 192]. The entropic cost of forming a 'rosette' of n loops (which is superlinear in n) counteract the attraction term, which would cluster all specific binding sites [192] (through bridging).

2. proteins can switch between an active state and an inert state. Biochemical reactions can modify DNA-binding protein, affecting protein affinity for chromatin. Therefore, by modelling protein switching with rate α, we have effectively only a fraction of proteins bound to chromatin. Also, this mechanism competes with the growth and merging of different droplets, leading to microphase separation [163].

Both mechanisms provide simple explainations of some realistic phenomenologies seen *in vivo*; in fact, nuclear bodies only grow up to a finite size. Many

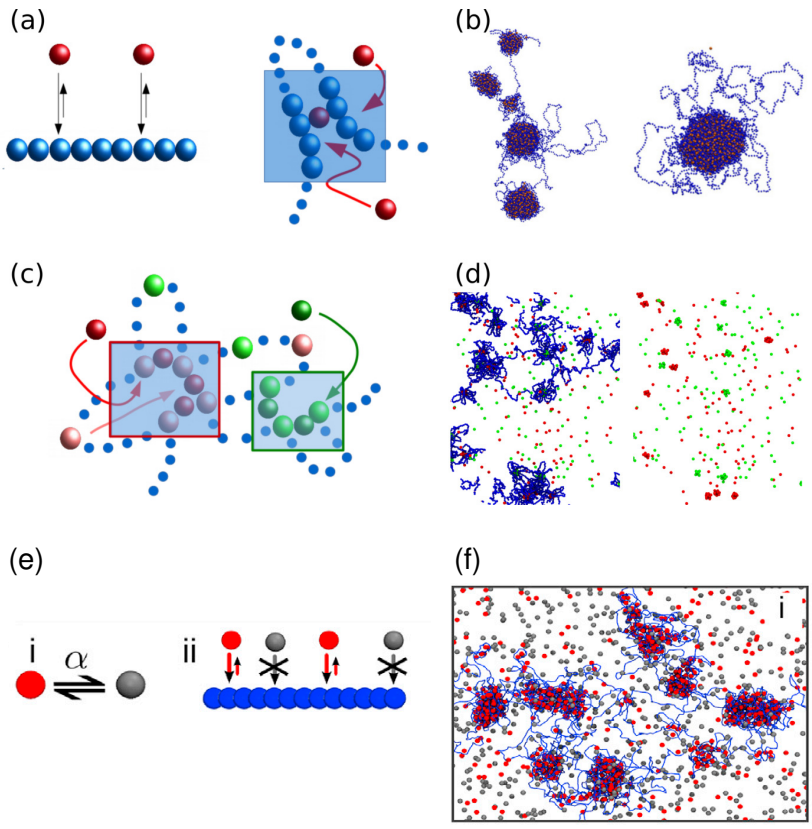

Figure 3.4. The bridging–induced attraction (micro)phase separation (a) A chromatin fibre is modelled as a bead–and–spring polymer, with spherical monomers (blue beads). Proteins (modelled as red spheres) bind to the chromatin fibre non–specifically. As proteins are multivalent, upon binding they create bridges which increase local chromatin density. Due to the density increase, more proteins can bind (right image). This effect has been called the "bridging-induced attraction". **(b)** As a result of the attraction, proteins undergo full phase separation. Left panel: at early times, large droplets form along chromatin and grow. Right panel: at equilibrium, all droplets merge into a single cluster. **(c)** Here, proteins interact both non–specifically (with low affinity to blue dots) and specifically (with strong affinity to either pink or light green). Red and dark green beads model two different transcription factors, which bind to pink and light green chromatin beads, respectively. **(d)** Now, bridging–induced attraction produces microphase separation, since steric interaction between different loops prevent coalescence of clusters. Right image shows only green and red clusters. **(e)** Proteins switch between active (red beads) and inactive (grey) states, with rate α (panel *(i)*). Red proteins interact with chromatin with some affinity, while grey proteins interact only sterically (panel *(ii)*). In this model, protein can also switch between active and inert states while bound. **(f)** Snapshot for $N = 4000$ switchable proteins, with switching rate $3 \cdot 10^{-4}$ inverse brownian times. Figures adapted from [163, 164]

transcription factors interact non–specifically, but weakly, with DNA and chromatin via electrostatic interactions, except for some specific DNA or chromatin sites. Those sites, which model specific cognate DNA sequences, mimic, for instance, the different regiones of the genome interested by specialised factories of RNA polII and PolIII [193, 194]. Additionally, several factors and proteins are subject the posttranslational modifications (e.g. through phosphorylation) which modify their conformation and disabling their ability to bind chromatin [195].

3.3.1 Nuclear organelles dynamics: liquid or immobile?

BIPS has been often associated to LLPS in terms of the liquid character of the chromatin–protein globules. The majority of FRAP experiments (fluorescence recovery after photobleaching) show that nuclear bodies are very dynamic. In FRAP experiments both internal mixing and the exchange of material with the soluble pool are observed, which measure how dynamic are proteins inside clusters [196]. Usually, one half of the droplet (or an internal spot) is photobleached, and then, the real–time mixing between the two halves is followed, by tracking the normalised intensity of the outcoming fluorescence.

Several lines of evidence seemed to suggest that many proteins rapidly and continuously enter and exit nuclear organelles: highly mobile, diffusive mechanism of movements were recognized in nuclear compartments [197, 198]. On the contrary, in Ref. [196] several relaxation timescales were identified for two internal components of Cajal bodies, being the longest timescale of $\sim$ 33min [199]. In HP1 droplets, a coexistence between mobile and immobile populations have been detected [200] and gel formation at long times [201]. In bacteria, H–NS are organised in ordered array [202], which suggest a solid–like structure of the H–NS clusters.

In general, contradictory results have been shown and different viewpoints proposed. Yet, whether such proteins are really liquid or not when bridging distant region of chromatin is still unknown. In the following paragraph, I present the striking experimental findings which relate cohesin droplets formation to the bridging–induced attraction. There, a liquid–like behaviour of cohesins, when they undergo full BIPS, has been shown through FRAP experiments.

3.3.2 BIPS might be in action during cohesin clustering

It is worth to mention that, only recently, the first experimental evidence of a possible BIPS has been shown in [203]. There, the authors show that cohesin clustering is ATP independent, involves many complexes ($\sim$ 720) and requires

DNA to form, as shown in Fig. 3.5). *In vitro*, the liquid behaviour of cohesin is supported by the following piecies of evidence:

- the shape of the droplet appear spherical–like;
- neighbouring clusters merge over time;
- after merging, clusters become spherical again.
- full and rapid recovery after FRAP, which demonstrate that the droplet can exchange cohesins with the environment.

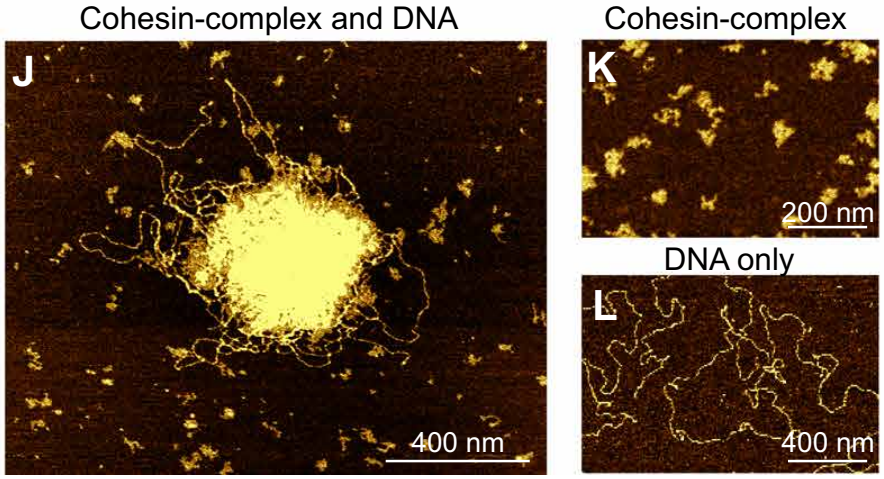

Figure 3.5. DNA compaction mediated by cohesin complex. Atomic force microscopy images of **(J)** DNA–cohesin agglomerate **(K)** cohesin complexes only and **(L)** DNA only. Note that DNA is required to have a single chromatin–cohesin cluster, compare with panel K. Figure taken from [203].

Furthermore, they draw a phase diagram of cluster formation at various salt condition and coehesin concentrations. They find that 3–kbp is the critical DNA length to trigger the phase separation, beyond which the radius of gyration of scales as $R_g \sim L^{\nu}$, being $\nu \approx 0.45$. Simulation show that such exponent is compatible with proteins with effective valence close to 2–3, while multivalent proteins would give $\nu \approx 0.33$, as predicted by the Flory theory for collapsed globules.

Similarly, several other proteins could virtually behave as cohesin, especially *in vivo*. In Ref. [204] three other prototypical chromatin compartments (the nucleolus, clusters of active RNA polymerase II and pericentric heterochromatin domains) have been discussed, in relation to the enrichment of proteins and RNA in those regions, which can mediate chromatin compaction. It is therefore speculated that BIPS can occur in all these cases, assuming the role

of universal physical paradigma for phase separation of chromatin–binding factors *in vivo*.

In the next section, I will show that equilibrium BIPS entails multiple different dynamical regimes, for non–specific chromatin and in the presence of strong binding sites. The results of this work can thus help us to decipher this puzzling scenario.

3.4 Dynamical heterogeneity is associated with solid–like and glassy structures in BIPS clusters

In this chapter, we will study the internal dynamics of quasi–spherical droplets, which form due to the bridging–induced attraction (BIA) between multivalent proteins and chromatin, in the absence of chromatin–chromatin or protein-protein attractive interactions. We consider two different scenarios: in the first case, proteins interact non–specifically with all chromatin beads, mimicking the behaviour of proteins which bind to some heterochromatic regions of the genome; in the second case, a chromatin bead with a high affinity is placed every 30kbp along the chromatin chain, modelling the specific binding sites for transcription factors and cohesin loading.

3.4.1 Simulation scheme

In our coarse-grained approach, we model the chromatin fibre as bead-and-spring polymer, which consists of L consecutive spherical beads of diameter $\sigma = 30$nm, held together via finitely extensible non-linear elastic (FENE) bonds. The FENE potential associated to two adjacent polymer beads located at the position $\mathbf{r_i}$ and $\mathbf{r_{i+1}}$ reads:

$$\begin{aligned} U_{\mathrm{FENE}}(r_{i,i+1}) =& U_{\mathrm{WCA}}(r_{i,i+1}) + \\ &- \frac{K_{\mathrm{FENE}} R_0^2}{2} \log\left[1 - \left(\frac{r_{i,i+1}}{R_0}\right)^2\right]. \end{aligned} \tag{3.1}$$

where $r_{i,i+1} \equiv |\mathbf{r_i} - \mathbf{r_{i+1}}|$, k_B is the Boltzmann constant and T the temperature. We set the FENE bond energy $K_{\mathrm{FENE}} = 30 k_{\mathrm{B}} T/\sigma^2$, and the maximum extension $R_0 = 1.6\sigma$; Hereon, we use energy units of $k_B T$ throughout. The first term in (3.1) produces a pure repulsive force which prevents beads from overlapping. This is provided by the Weeks-Chandler-Andersen (WCA) potential:

$$\frac{U_{\mathrm{WCA}}(r_{ij})}{k_B T} = \begin{cases} 4\left[\left(\frac{d_{ij}}{r_{ij}}\right)^{12} - \left(\frac{d_{ij}}{r_{ij}}\right)^{6}\right] + 1 & \frac{r_{ij}}{d_{ij}} < 2^{1/6}, \\ 0 & \text{otherwise}, \end{cases} \tag{3.2}$$

where $d_{ij} = \sigma$ is the mean of the diameters of the beads i and j (here beads are all of the same size σ). Non-adjacent beads also interact via the WCA.

The bending rigidity of chromatin is modelled via the Kratky–Porod potential:

$$U_{\mathrm{BEND}}(\theta_i) = K_{\mathrm{BEND}}\left[1 - \cos(\theta_i)\right], \tag{3.3}$$

where θ_i is the angle formed between any triplet of beads $i-1$, i and $i+1$, $i = 1 \dots L$, and K_{BEND} is the bending energy. *In vitro* experiments show that the 30–nm chromatin fibre has a persistence length of between 40–200nm [205]. A reasonable estimation for our model is thus $l_p \simeq 120$nm. In light of this, we choose $l_p = 4\sigma$, from which $K_{\mathrm{BEND}} = l_p k_{\mathrm{B}} T/\sigma$ follows.

In our system, N active proteins can bind the chromatin fibre. Proteins are modelled as spherical beads of diameter σ. For simplicity, we assume that the proteins are the same size as the chromatin beads (a realistic assumption, as each is likely to be a protein complex, as is the case for human cohesin [206]). Each protein can interact with a chromatin bead via a truncated Lennard-Jones potential

$$U_{\mathrm{LJ}}(r_{ij}) = 4\epsilon\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6} - \left(\frac{\sigma}{r_c}\right)^{12} + \left(\frac{\sigma}{r_c}\right)^{6}\right], \tag{3.4}$$

where ϵ is the interaction strength, r_{ij} is the distance between the i-th protein and the j-th polymer bead, while r_c is a cutoff distance which set the interaction range. In our simulations we set $r_c = 1.9\sigma$, while we study our system at different ϵ. Observe that, when $r_c = 2^{1/6}\sigma$ we obtain the WCA potential, which corresponds to Eq. (3.2) for $\epsilon = 1k_{\mathrm{B}}T$. In the specific case we always set $\epsilon = 3k_{\mathrm{B}}T$ (before the transition point in the non–specific case), while the specific sites interact with proteins with a larger energy $\epsilon_{sp} > \epsilon$.

Langevin dynamics

We use the LAMMPS software [207] to perform Langevin dynamics. The dynamics of each polymer and protein beads of mass m_i ($i = 1 \dots N + L$) at position $\mathbf{r_i}$ is determined by the equation

$$m_i \frac{d^2\mathbf{r}_i}{dt^2} = -\nabla_{\mathbf{r}_i} U_i - \xi_i \frac{d\mathbf{r}_i}{dt} + \sqrt{2k_B T \xi_i}\boldsymbol{\eta}_i(t), \tag{3.5}$$

where U_i is the sum of all the potentials at the position $\mathbf{r}_i$. The friction ξ_i sets the diffusion constant $D_i = k_{\mathrm{B}}T/\xi_i$ for each bead; we set $m_i = 1$ and $\xi_i = 1$. The vector $\boldsymbol{\eta_i}$ is represents white noise with components which satisfy

$$\langle \eta_{i\alpha}(t) \rangle = 0 \ \text{ and } \ \langle \eta_{i\alpha}(t)\eta_{j\beta}(t') \rangle = \delta_{ij}\delta_{\alpha\beta}\delta(t - t'). \tag{3.6}$$

where $\eta_{i\alpha}$ is component α of the noise vector and δ_{ij} and $\delta(t)$ are the Kronecker and Dirac delta functions respectively. (3.5) is solved using a velocity-Verlet algorithm with timestep $0.01\tau_{\mathrm{LJ}}$, where τ_{LJ} is the simulation time unit defined by $\tau_{\mathrm{LJ}} = \sqrt{m\sigma^2/k_{\mathrm{B}}T}$. As we set $k_{\mathrm{B}}T = 1$, we have $\tau_{\mathrm{LJ}} = 1$.

Since most of our dynamical measurements should be calculated in stationary state, we run preliminary simulations to ensure that our system reached equilibrium.

For each replica, the initial conditions were generated as follows: the polymer chain is built as a 3–dimensional random walk in a cubic box of $120\sigma \times 120\sigma \times 120\sigma$; then, a long simulation was run to ascertain that the polymer was optimally swollen in dilute conditions; to check that the polymer is in an equilibrium configuration before switching on the interaction with proteins, we check that the polymer radius of gyration R_g is constant over time, on average. Thus, proteins were randomly positioned inside the simulation box, and a short run with a 'soft' repulsive potential between proteins and polymer beads was run to remove unphysical overlaps, followed by another short run where polymer beads and proteins interact via the potential in (3.4) with $\epsilon = 0$ (no attraction).

Simulation procedure and parameters

Starting from these initial conditions, a first quench to $\epsilon = 3k_{\mathrm{B}}T$ was performed, and the simulations were run for $10^5\tau_{LJ}$. In such a way, we prevent the formation of non–spherical clusters in the non–specific case and other structures, such as regularly structured platelets in the presence of specific binding sites. This may be stable or long–lived metastable states which we do not take under consideration in our study. Then, we quenched the system to $\epsilon > 3k_{\mathrm{B}}T$, and a long equilibration run is performed for T_{eq}. With such a scheme $N^{\epsilon}_{\mathrm{eq}}$ independent initial conditions for the real simulations were produced, and $N^{\epsilon}_{\mathrm{eq}}$ simulations were run. Measurements were taken during the latter, which were run for at least $\sim 8\tau_r$, where τ_r is the relaxation timescale shown below [section 3.4.5, Fig. 3.12]; details on the equilibration runs and real simualtions are in Tables 3.1 and 3.2.

In simulations without polymer (see below), we confine particles in a cubic box of $50\sigma \times 50\sigma \times 50\sigma$; then, two consecutive short runs with a soft repulsive potential and with the interaction in (3.4) between particles with $\epsilon = 0$ were performed. Then, those simulations were carried on to be equilibrated for T_{eq}, already quenched at the selected interaction energy value; this is because no other structures but the liquid and solid, spherical–like droplet have been detected. A final equilibrated run was performed for at least $T \sim 5.5\tau_r$,

Table 3.1. Details of non–specific polymer simulations.

ϵ ($k_B T$)	$\mathbf{T_{eq}}$ (τ_{LJ})	$\mathbf{T}$ (τ_{LJ})	$\mathbf{T}$ (τ_r)	$\mathbf{N^{\epsilon}_{eq}}$
3.5	$2.1 \cdot 10^6$	$1.01 \cdot 10^5$	81.0	50
3.6	$2.1 \cdot 10^6$	$1.01 \cdot 10^5$	65.6	50
3.7	$2.1 \cdot 10^6$	$1.01 \cdot 10^5$	53.4	50
3.8	$4.1 \cdot 10^6$	$1.001 \cdot 10^6$	133.7	6
4.0	$2.1 \cdot 10^6$	$1.001 \cdot 10^6$	61.8	36
4.5	$2.1 \cdot 10^6$	$1.001 \cdot 10^6$	7.9	50

Table 3.2. Details specific binding sites simulations.

ϵ ($K_B T$)	$\mathbf{T_{eq}}$ (τ_{LJ})	$\mathbf{T}$ (τ_{LJ})	$\mathbf{T}$ (τ_r)	$\mathbf{N^{\epsilon}_{eq}}$
6	$3 \cdot 10^5$	$1.01 \cdot 10^5$	257.2	49
8	$3 \cdot 10^5$	$1.01 \cdot 10^5$	111.0	49
9	$3 \cdot 10^5$	$1.01 \cdot 10^5$	52.9	49
9.5	$3 \cdot 10^5$	$1.01 \cdot 10^6$	36.1	49
10	$3 \cdot 10^5$	$1.001 \cdot 10^6$	27.4	49
11	$3 \cdot 10^5$	$1.001 \cdot 10^6$	15.4	49
12	$3 \cdot 10^5$	$1.001 \cdot 10^6$	8.7	50

Table 3.3. Details monospecies LJ particles simulations.

ϵ ($K_B T$)	$\mathbf{T_{eq}}$ (τ_{LJ})	$\mathbf{T}$ (τ_{LJ})	$\mathbf{T}$ (τ_r)	$\mathbf{N^{\epsilon}_{eq}}$
2.8	$5 \cdot 10^4$	$1.1 \cdot 10^4$	622.4	50
2.9	$5 \cdot 10^4$	$1.1 \cdot 10^4$	452.5	50
3.0	$5 \cdot 10^4$	$1.1 \cdot 10^4$	57.7	50
3.2	$5 \cdot 10^4$	$1.001 \cdot 10^6$	16.8	37
3.4	$5 \cdot 10^4$	$1.001 \cdot 10^6$	5.6	50

from which some quantities were calculated. Details on the parameters of simulations are in Tables 3.3.

Starting from the same initial conditions, another set of independent simulations was run to build the distributions shown in Figs. 3.13(c) and 3.17(b). For each value of the interaction strength ϵ, $N^{\epsilon}_{\mathrm{eq}}$ simulations were run for $T = 1000t^*$, where t^* was chosen such as the average mean–square displacement $\delta r(t^*)^2 \simeq 0.2$.

3.4.2 Measured quantities

Before going to the results, let us describe the methods used to analyse the equilibrum configurations obtained by our simulations. In the following we will show that the internal dynamics of proteins is correlated to some *static*

local properties of clusters. In particular, we will probe the *local bond ordering*, by means of the parameter ψ_6, which is a rotational invariant quantity, previously used in similar context [208]. To define this parameter, the *direct voronoi tesselletion* has been employed, a procedure to identify the nearest neighbours in the first coordination shell.

Direct Voronoi tessellation

To quantify the local bond ordering we perform the direct Voronoi tessellation of the 3D space. The direct Voronoi tessellation is needed to identify only those particle which constitute the *first coordination shell* of a given particle within the cluster, or *Voronoi neighbors*, which is, in turn, needed to calculate the degree of local ordering or detect local structures [209, 210].

Technically, the Voronoi tessellation is a partition of $\mathbb{R}^d$, where d is the spacial dimension, into regions of non–intersecting domains with distinct boundaries, as shown in Fig. 3.6, relative to a determined prescribed metric, or distance. If $d = 3$, such regions are the Voronoi polyhedra. Each particle is therefore a Voronoi generator. These methods have been widely applied, and consequentley refined, in different contexts. More details on the Voronoi tessellation are in [211, 212]. Herein, we briefly summarise the nearest neighbours search algorithm via direct Voronoi tessellation, following [210].

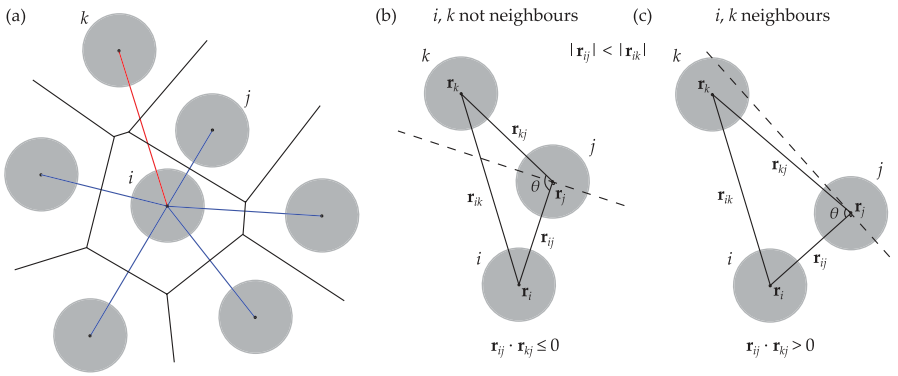

Figure 3.6. Direct Voronoi tessellation. (a) Example of the Voronoi partition of the two–dimensional space. Blue segments represent the minimum Euclidean distance between i and the direct Voronoi neighbours; conversely, k is not a neighbour, as the minimum Euclidean distance that links the i–th and the k–th Voronoi cells crosses also the j–th Voronoi region. **(b)** Proteins i and k are not neighbours, since $\theta > 90^o$. **(c)** In this scenario, Proteins i and k are neighbours, as $\theta < 90^o$. Figure taken from [210].

Suppose that protein i, which is located at position $\mathbf{r}_i$, is inside the cluster. First, we select all proteins which belong to the the neighborhood Σ_i of the ith protein. A protein i' belongs to Σ_i if $r_{ii'} < r_{\text{thre}}$ ($r_{\text{thre}} = 2.5\sigma$). This first selection is needed to reduce the computational cost of the search. A protein k is a

Voronoi neighbour of i if and only if for all proteins j closer to i than k, the angle subtended by the vectors r_{ij} and r_{kj} is less than $\pi/2$ radians. Equivalentely, invoking the law of cosines

$$r_{ik}^2 = r_{ij}^2 + r_{jk}^2 - 2r_{ij}r_{jk}\cos(\theta) \ , \tag{3.7}$$

the condition for k to be a Voronoi neighbour of i reads:

$$r_{ik}^2 < r_{ij}^2 + r_{jk}^2. \tag{3.8}$$

Thus, the algorithm proceed as follows:

1. loop over the N^* protein (in our simulation $N^* \leq N$ is the total nnumber of proteins inside the cluster, by selecting the i–th protein (or Voronoi generator);
2. find all proteins within a sphere of radius r_{thre} centered on r_i. Those proteins belong to Σ_i;
3. sort Σ_i in terms of the distance from i, starting from the closest;
4. loop over all proteins j in Σ_i;
5. loop over all proteins $k > j$, and check the condition 3.8. Eliminate k from Σ_i if Eq. (3.8) is not satisfied.

As shown in Fig. 3.6(a), this search algorithm rejects also those Voronoi generators whose corresponding Voronoi cell is in contact with the ith Voronoi cell, but the minimum distance segment between i and k crosses another voronoi cells. In Fig. 3.6(b–c) we show a schematics of the two cases, namely where Eq. (3.8) *is not* and *is* satisfied, respectively.

Local orientation order parameter

Once we have recognised the subsets of neighbours for each particle i, namely Σ_i, we can calculate the local bond–ordering parameter. Several bond–ordering parameters have been proposed [213] and widely used to detect local ordered structures in liquids [208, 214], glass–formers [209, 215] and crystals [214–217], all based on the following superposition of spherical harmonics originally introduced by Steinhardt *et al.* [208]:

$$q_{lm}(i) = \frac{1}{n(\Sigma_i)} \sum_{j \in \Sigma_i} Y_l^m(\theta_{ij}, \phi_{ij}) \ , \tag{3.9}$$

where θ_{ij} and ϕ_{ij} are the azimuthal and polar angles, which determine the orientation of the ij vector (bond) with respect to a reference frame cenetered in

the ith protein; Y_l^m are the (complex) Laplace's spherical harmonics of degree l and order m, which are defined as follows:

$$Y_l^m(\theta_{ij},\phi_{ij}) = \sqrt{\frac{(2l+1)}{4\pi}\frac{(l-m)!}{(l+m)!}}P_l^m(\cos(\theta_{ij}))e^{im\phi_{ij}} \ . \tag{3.10}$$

$n(\Sigma_i)$ is the number of direct Voronoi neighbours of protein i and P_l^m the Legendre polynomials. Observe that θ_{ij} and ϕ_{ij} depend on the arbitrarly chosen orientation of the reference frame. Therefore, in order to have an order parameter indepedent of the alignment of the system, we need to build a rotational invariant quantity, starting from q_{lm}.

To effectively discriminate different local crystal structures from the local order which can appear in liquid–like domains we should somehow consider not only the first coordination shell (nearest neighbours), but also the second coordination shell. More specifically, one would measure the 'correlation' of the local order in the first and the second coordination cells. In light of this, we consider the scalar product $\mathbf{q}_l(i)\cdot\mathbf{q}_l(j)$, where $\mathbf{q}_l \equiv q_{lm}$, and j is a nearest neighbour of i. The aforementioned scalar product is specified as follows:

$$\mathbf{q}_l(i)\cdot\mathbf{q}_l(j) = \sum_{m=-l}^{l} q_{lm}(i)\, q_{lm}^*(j) \ . \tag{3.11}$$

where $q_{lm}^*(j)$ indicates the complex conjugate of $q_{lm}(j)$. Using the addition rule of spherical harmonics, and substituting Eq. (3.9) in Eq. (3.11), one obtains:

$$\begin{aligned}\mathbf{q}_l(i)\cdot\mathbf{q}_l(j) &= \frac{1}{n(\Sigma_i)}\frac{1}{n(\Sigma_j)}\sum_{m=-l}^{l}\sum_{\substack{j'\in\Sigma_i\\ k\in\Sigma_j}} Y_l^m(\theta_{ij'},\phi_{ij'})\, Y_l^{m\,*}(\theta_{jk},\phi_{jk}) \\ &= \frac{2l+1}{4\pi}\sum_{j'\in\Sigma_i}\sum_{k\in\Sigma_j} Y_l^0(\delta_{ij'k}) \ .\end{aligned} \tag{3.12}$$

Here, j and j' are neighbours of i, whereas k is neighbour of j (k can also be equal to i); $\delta_{ij'k}$ is the angle between $\mathbf{r}_{ij'}$ and $\mathbf{r}_{jk}$, as shown in Fig. 3.7 (δ_{ijk} in the figure), such that

$$\cos(\delta_{ij'k'}) = \cos(\theta_{ij'})\cos(\theta_{jk'}) + \sin(\theta_{ij'})\sin(\theta_{jk'})\cos(\phi_{ij}-\phi_{jk'}) \ . \tag{3.13}$$

Then, we normalise q_{lm} of spherical harmonics [the normalisation renders the measure independent of the particular definition of $Y_l^m(\theta,\phi)$]:

$$\hat{q}_{lm}(i) = \frac{q_{lm}(i)}{\mathbf{q}_l(i)\cdot\mathbf{q}_l(i)} \ . \tag{3.14}$$

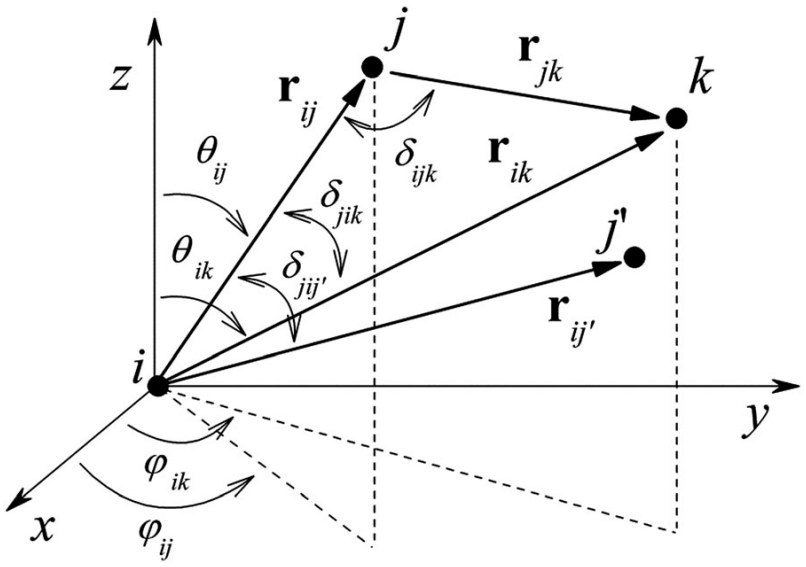

Figure 3.7. Orientation of bond vectors connecting Voronoi neighbours Figure taken from [213].

The dot product between the normalised vectors reads:

$$\begin{aligned}\hat{\mathbf{q}}_l(i)\cdot\hat{\mathbf{q}}_l(j) &= \frac{\mathbf{q}_l(i)\cdot\mathbf{q}_l(j)}{[\mathbf{q}_l(i)\cdot\mathbf{q}_l(i)]\,[\mathbf{q}_l(j)\cdot\mathbf{q}_l(j)]} \\ &= \frac{\sum_{j'\in\Sigma_i}\sum_{k\in\Sigma_j} Y_l^0(\delta_{ij'k})}{\left[\sum_{j,j'\in\Sigma_i} Y_l^0(\delta_{ijj'})\right]\left[\sum_{k,k'\in\Sigma_j} Y_l^0(\delta_{kjk'})\right]}\,,\end{aligned} \tag{3.15}$$

where, j and j' are neighbours of i, k and k' are neighbours of j (included i). Note that $Y_l^0(\omega) \propto P_l(\cos^2(\omega))$, where P_l is a Legendre polynomial, which only depends on the squared cosine of the angle ω. Therefore, the scalar product in Eq. (3.15) is rotationally invariant. Eq. (3.15) describes the correlations between the degree of orientational orders of the bonds which link proteins i and j to their neighbours, being j a direct neighbours of i. To define a quantity associated with protein i only, we average Eq. (3.15) over $j \in \Sigma_i$:

$$\bar{q}_l(i) = \frac{1}{n(\Sigma_i)}\sum_{j\in\Sigma_i}\hat{\mathbf{q}}_l(i)\cdot\hat{\mathbf{q}}_l(j)\,. \tag{3.16}$$

Now, this parameter quantifies correlations between bond–ordering of the first and second coordination shell, both *centered* in i. A more precise parameter which describes the local degree of orientational order *around* the i-th protein is:

$$\psi_l(i) = \frac{1}{n(\Sigma_i)+1}\left[\bar{q}_l(i) + \sum_{j\in\Sigma_i}\bar{q}_l(j)\right]\,, \tag{3.17}$$

which is obtained by averaging again on all proteins of the first coordination shell of i, including protein i.

Different choice of l can be made. In Ref. [213] the probability distributions of ψ_4, ψ_6 and ψ_8 are plotted for LJ liquid, BCC, FCC and HCP. The authors noted that ψ_6 can better discriminate disordered from ordered phase. In disordered phases, the distribution of ψ_6 is peaked at about 0.2, whereas larger values of ψ_6, commonly larger that 0.5 are associated with local order. In light of this, we will hereon use ψ_6 to quantify the extent of local order inside the droplet.

3.4.3 Mean field theory for BIPS

In Ref. [163] a mean field theory for BIPS with switching proteins has been developed. A simple limit of that theory is when all particle are *active*, that is the equilibrium case. Here we review a simple calculation, that, in the spirit of the Model B field theory [176], predicts the macrophase separation of active chromatin bridges.

These mean field theories always rely on the coarse–grained approximation, by which a discrete set of degrees of freedom (particles) can be described by a local field. In terms of the local density:

$$\rho_i = \frac{1}{V_i} \sum_{\substack{p=1 \\ p\in V_i}}^{N} \delta(\mathbf{x} - \mathbf{x_p}) \to \rho(\mathbf{x}) \,, \tag{3.18}$$

where V_i is the ith sub–volume in a suitable partition of the three–dimensional volume V, ρ_i is the associated density and $\delta(x)$ is the Dirac delta. If the volume V and the number of constituents, N and L, are sufficiently large (virtually infinite, with finite overall densities N/V and L/V), the coarse–grained approximation leads to a correct description of the macroscopic properties of the system.

In light of this, we describe the distribution of chromatin via the probability density field $\rho(\mathbf{x}, t)$ and the density of proteins by $\phi(\mathbf{x}, t)$. The Landau free energy $\mathcal{F}$ is specified by a suitable free energy density f:

$$f = \frac{D_1}{2}\rho^2 + \frac{D_2}{2}\phi^2 - \chi\rho\phi + \frac{\kappa}{2}\nabla\rho\cdot\nabla\rho + \frac{g}{4}\rho^4 \,. \tag{3.19}$$

The first two terms describe an 'elastic' energy of chromatin and proteins respectively, the third term represent the (negative) energy gain due to chromatin–protein interaction, the fourth term is the surface tension and the last term penalizes strong accumulation of chromatin (i.e. self–repulsion due to short–range steric interaction). Note that the last two terms also provide surface tension and prevent unphysical strong accumulation of proteins, as the two

fields are coupled; thus, terms proportional to $(\nabla\phi)^2$ and ϕ^4 are not required in the Landau expansion of the free energy density; indeed, note that the choice of imposing surface tension and accumulation avoidance for the protein field would lead to the same free energy after exchanging the two fields, $\rho \leftrightarrow \phi$. As neither the chromatin nor protein overall density can change ($\rho_0 = L/V$ and $\phi_0 = N/V$ are constant), the following constraint must hold:

$$\int_V (\rho(\mathbf{x},t) - \rho_0) dV = \int_V (\phi(\mathbf{x},t) - \phi_0) dV = 0\,, \tag{3.20}$$

being ρ_0 and ϕ_0 the value of the chromatin and protein overall density respectively. Then, the Cahn–Hillard equation (Model B) applies:

$$\begin{aligned} \frac{\partial \rho}{\partial t} &= M_\rho \nabla^2 \frac{\delta \mathcal{F}}{\delta \rho} \\ \frac{\partial \phi}{\partial t} &= M_\phi \nabla^2 \frac{\delta \mathcal{F}}{\delta \phi}\,, \end{aligned} \tag{3.21}$$

where $\delta/\delta\rho$ and $\delta/\delta\phi$ denotes the functional derivatives with respect to $\rho(\mathbf{x},t)$ and $\phi(\mathbf{x},t)$ respectively, and M_ρ and M_ϕ are the mobility coefficients. Inserting Eq. (3.19) into Eq. (3.21), we obtain:

$$\begin{aligned} \frac{\partial \rho}{\partial t} &= M_\rho \nabla^2 \left[D_1 \rho - \kappa \nabla^2 \rho - \chi \phi + g \rho^3 \right] \\ \frac{\partial \phi}{\partial t} &= M_\phi \nabla^2 \left[D_2 \phi - \chi \rho \right] \end{aligned} \tag{3.22}$$

To simplify the equation, we search for suitable combinations of the coefficients to rescale time and space. That is, we define the following time and space units:

$$t_u = \frac{\kappa}{M_\phi D_1 D_2} \qquad\qquad x_u = \sqrt{\frac{\kappa}{D_1}}\,. \tag{3.23}$$

By defining $t' = t/t_u$ and $x' = x/x_u$, rescaling the field ϕ such that $\Phi = M_\rho \chi \phi / M_\phi D_2$ and omitting primes, one obtains:

$$\begin{aligned} \frac{\partial \rho}{\partial t} &= \mathcal{D}_0 \left(\nabla^2 \rho - \nabla^4 \rho \right) - \nabla^2 \Phi + \mathcal{G} \nabla^2 \rho^3 \\ \frac{\partial \Phi}{\partial t} &= \nabla^2 \Phi - \mathcal{X} \nabla^2 \rho\,, \end{aligned} \tag{3.24}$$

where $\mathcal{D}_0 \equiv M_\rho D_1 / M_\phi D_2$, $\mathcal{G} \equiv M_\rho g / M_\phi D_2$ and $\mathcal{X} \equiv M_\rho \chi^2 / M_\phi D_2^2$ are dimensionless quantities. Note that the effective diffusivity $\mathcal{D}_0$ controls both diffusion and surface tension, different from the case studied in Ref. [163], where protein switching disfavours sharp interfaces.

3.4.4 Linear stability analysis

To study the system at the onset of instability, we linearise Eq. (3.24) by using $\rho' = \rho - \rho_0$ and $\Phi' = \Phi - \Phi_0$, where ρ' and Φ' are the deviation from the uniform solutions ρ_0 and Φ_0. This analysis allows us to study the response of the uniform phase to small perturbation in the density fields (ρ, Φ). After linearisation

$$\begin{aligned} \frac{\partial \rho'}{\partial t} &= \mathcal{D}\left(\nabla^2\rho' - \nabla^4\rho'\right) - \nabla^2\Phi' \\ \frac{\partial \Phi'}{\partial t} &= \nabla^2\Phi' - \mathcal{X}\nabla^2\rho' \,, \end{aligned} \tag{3.25}$$

where $\mathcal{D} \equiv \mathcal{D}_0 + 3\mathcal{G}\rho_0$. Introducing the Fourier transform:

$$\tilde{g}(\mathbf{q}) = \int \mathbf{e}^{-i\mathbf{q}\cdot\mathbf{x}} g(\mathbf{x}) d^3x \,. \tag{3.26}$$

Eqs. (3.25) become:

$$\begin{aligned} \frac{\partial \tilde{\rho}'}{\partial t} &= -\mathcal{D}\left(Q + Q^2\right)\tilde{\rho}' + Q\tilde{\Phi}' \\ \frac{\partial \tilde{\Phi}'}{\partial t} &= \mathcal{X}Q\tilde{\rho}' - Q\tilde{\Phi}' \,, \end{aligned} \tag{3.27}$$

with $Q \equiv |q|^2$. The general solution of Eqs. (3.27) (which is in the form $\dot{X} = \mathbb{A}X$) is a combination of two exponentials whose frequency is given by the eigenvalues of the matrix $\mathbb{A}$. In turn, the solution of Eqs. (3.25) is a superposition of several modes of different frequencies Q. Therefore, we search for the roots of the characteristic polynomial $\det(\mathbb{A} - \lambda I)$. The dispersion relation reads:

$$\lambda_\pm(Q) = -\frac{Q}{2}\left[\mathcal{D} + \mathcal{D}Q + 1 \pm \sqrt{(\mathcal{D} + \mathcal{D}Q + 1)^2 - 4(\mathcal{D} + \mathcal{D}Q - \mathcal{X})}\right] . \tag{3.28}$$

Note that $\lambda_\pm(0) = 0$, $\lambda_\pm(-\infty) \to -\infty$ and $\lambda_+ < 0$ for every Q. The instable solution exists if and only if $\lambda_-(Q) > 0$ for some Q values (as the perturbation grows as $\mathrm{e}^{\lambda(\mathbf{q})t}$ at those frequencies). Thus, positive frequencies exist if $\mathcal{X} > \mathcal{D}$. In physical quantities $\chi > \sqrt{D_1 D_2 - 3D_2 g\rho_0^2}$. Observe that the same qualitative result would apply if the chromatin–protein interaction term in Eq. (3.19) were replaced by a self–interaction term of the type $-\chi'\rho^2$ (effectively, a 'negative' elastic contribution to the free energy), which leads to the standard equation of Model B phase separation for binary fluids [176]. Therefore Eqs. 3.22 and 3.27 describe, for a suitable set of parameters, a macrophase separation similar to what occurs in a LLPS. Below, we see that the growth law of clusters is compatible with the exponent of Model B ($L \sim t^{1/3}$).

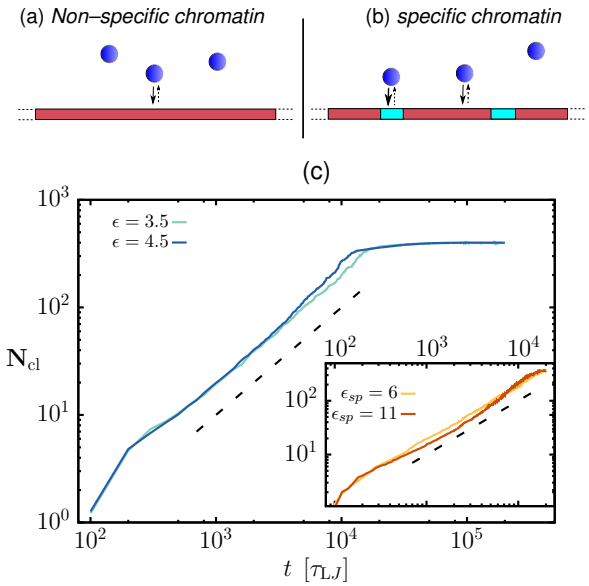

Figure 3.8. Chromatin model and domain growth law. Schematics of the two chromatin models. (a) Chromatin can interact non–specifically with proteins (b) Specific binding sites (cyan) are regularly interposed between stretches of 10 chromatin beads (red). Proteins interact strongly with those ($\epsilon_{sp} \geq 6k_BT$), weakly with the rest $\epsilon = 3k_BT$. In both models protein unbinding is due to thermal fluctuations. (c) Number of particles in the largest cluster, N_{cl}, is plot as a function of simulation time t, for $\epsilon = 3.5$ (liquid regime) and $\epsilon = 3.5$ (solid–like regime). After an initial transient, the cluster growth law is linear in time ($N_{cl} \sim t$), as expected for an Ostwald ripening process. Inset: the same behavior appears for specific binding sites, for both small and large specific interaction energies.

3.4.5 Non–specific interactions lead to liquid–to–solid–like phase transition

In Fig. 3.8(a–b) we show a schematics of the two chromatin models. In the first model [panel (a)], chromatin is modelled as a homopolymer and proteins can interact with chromatin with an affinity of ϵ. In the second model [panel (b)], strong binding sites are located every 30 kbp; proteins interact with those sites with a larger affinity ϵ_{sp}, whereas the interaction stregth with the remaining chromatin is set to $\epsilon = 3k_BT$. In Fig. 3.8(c) we show the number of proteins which belong to the largest cluster, as a function of time, for both models, in a system with $N = 400$ (proteins) and $L = 1000$ chromatin beads. Clusters begin to form due to spinodal decomposition (at sufficiently large energies), and, later on, they merge. in this phase, the number of protein in clusters N_b grows as $L^3 \sim t^1$, which gives the same exponent expected for Model B, $L \sim t^{1/3}$. In the final stage, one single droplet form, containing the whole compacted chromatin fibre and a large fraction of proteins.

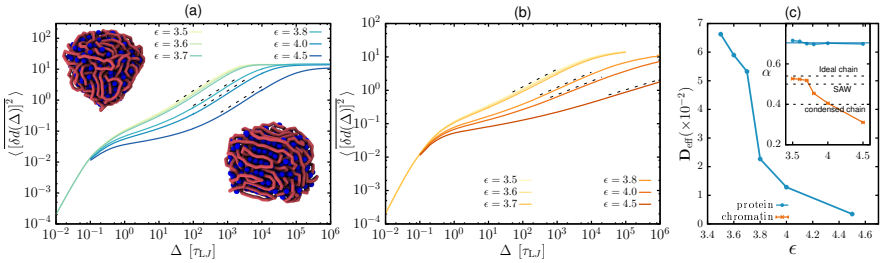

Figure 3.9. Mean square displacements and effective diffusion. (a) The total pMSD for proteins is plotted as a function of the lag time Δ. Four behaviors can be observed: *(i)* at small times, proteins can diffuse within the cages; *(ii)* then an inflection point (or plateau at large energies) indicates that proteins are constrained in their cages; *(iii)* at intermediate times, proteins escape their caged and sub–diffuse within the droplet; *(iv)* curves saturates as cluster proteins remains bound to chromatin. Snapshots show two equilibrium configurations extracted from simulations at $\epsilon = 3.5 k_B T$ and $\epsilon = 4 k_B T$. (b) The total pMSD for chromatin beads is plotted as a function of Δ. The same regimes as in panel (a) are displayed. At large energy values, our simulations are too short to evidence the saturation of chromatin pMSD, as its dynamics are far slower than protein ones. (c) Effective diffusivity is extracted by the fitting line shown in panel (a–b) (dashed lines). The jump of D_{eff} at $\epsilon = 3.8$ signals a liquid–to–solid transition. Accordingly, the value of the $\alpha \simeq 0.52$ exponent for chromatin change from to smaller values; conversely, for proteins $\alpha \approx 0.7$.

Although the qualitative behavior of BIPS is alike in both cases, both the internal local structures and dynamics differ in the two scenario at equilibrium, as we will show in the following sections.

Non–specific interactions lead to liquid–to–solid–like phase transition

We first study a system of proteins which interact non–specifically with a modelled chromatin fiber. In steady state, a single high–density spherical–like droplet appears, which can exchange proteins with the low–density phase. To probe the internal dynamics of proteins in the high–density phase, we consider the total pairwise mean square displacement (pMSD)

$$\langle\, \overline{[\delta d(\Delta)]^2}\, \rangle = \frac{1}{T-\Delta} \sum_{i<j} \int_0^{T-\Delta} \left[d_{ij}(t+\Delta) - d_{ij}(t) \right]^2 dt \qquad (3.29)$$

where d_{ij}(t) indicates the distance between the ith and jth proteins which belong to the cluster at time t, the bar $\bar{\cdot}$ indicates a time average over the simulation time T, while $\langle \cdot \rangle$ represents the average over all the $N_t^* \leq N$ proteins *within* the cluster at time t. We should here remark that in order to identify the proteins which belong to clusters, we employed the DBSCAN algorithm [218]. DBSCAN is an unsupervised density-based clustering algorithm which allows identification different clusters under the assumption that all clusters are of

equal density. The algorithm needs two parameters: a threshold distance which defines the neighborhood of a protein, l, and the minimum number of points which form a cluster, minPTS. Three categories of points are defined: the *core points*, which are points which have at least minPTS neighbours, meaning they are within a distance l; the *reachable points*, which are within distance l from a core point; the *noise points* which are not reachable from any other point. Starting from a core point p, all points that are reachable from p form a cluster. In this work we specified $l = 2.5\sigma$ and minPTS $= 4$. Our results are chiefly independent on the particular choice of the DBSCAN parameters, as we have only one cluster in all simulations.

In Fig. 3.9(a) we show the pMSD in (3.29) as a function of the lag time Δ for different values of the interaction strength ϵ, averaged over $N_{\rm eq}^{\epsilon}$ independent replicas ($N_{\rm eq}^{\epsilon}$ is the number of equilibrated runs, see section 3.4.1). After an initial diffusive regime for very short times ($\Delta \lesssim 10^{-1}\tau_{LJ}$), a plateau appears at intermediate times; the appearance of a plateau is related to the presence of effective "cages", which transiently constrain the motion of proteins, that effectively result less mobile. Caging is typical in systems at high density which behave as viscous liquids or glasses [219–221]. In our model, the plateau is particularly marked for $\epsilon \geq 3.8$, when the droplet rearranges in partially ordered structures, as shown in the snapshot in Fig. 3.9(a). At larger times, a 'mobile' regime is restored. The pMSD curves saturate when all proteins have moved of a distance of the radius of the droplet on average.

The abrupt deviation of the pMSD curves at $\epsilon = 3.8$ is an hallmark of a liquid–to–solid transition. The same qualitative behaviour emerges from chromatin dynamics. In Fig. 3.9(b) we show the pMSD calculated as in (3.29) for chromatin beads, as a function of Δ, with $N_t^* = L$, for any t. For $\epsilon < 3.8$ we find $\delta r^2 \sim \Delta^{\alpha}$, with $\alpha \simeq 0.52$, which is compatible with the exponent predicted from the Rouse model for ideal chains $\alpha = 0.5$, and not distant from the exponent predicted for a self–avoiding walk (SAW, $\alpha = 0.54$). Notably, for $\epsilon \geq 3.8$ also the chromatin dynamics varies with the interaction strength, with the extracted exponents significantly smaller than 0.5, see inset in fig. 3.9(c). It is worth to mention that the exponent extracted (at least for $\epsilon \leq 4$) are chiefly comparable with some recent experimental and numerical results [222, 223].

To visualize the abrupt nature of the crossover, we evaluate an effective diffusivity, which is defined as

$$D_{\rm eff} = \lim_{\Delta\to\infty} \frac{\langle \overline{\delta d(\Delta)^2} \rangle}{\Delta^{\alpha}} \,. \tag{3.30}$$

In practice, we fit the numerical curves with the function $D_{\rm eff}\Delta^{\alpha}$ in the region immediately before the saturation, as indicated in Figs. 3.9(a–b), assuming that this exponent reflects the trend of the pMSD in an infinitely large droplet

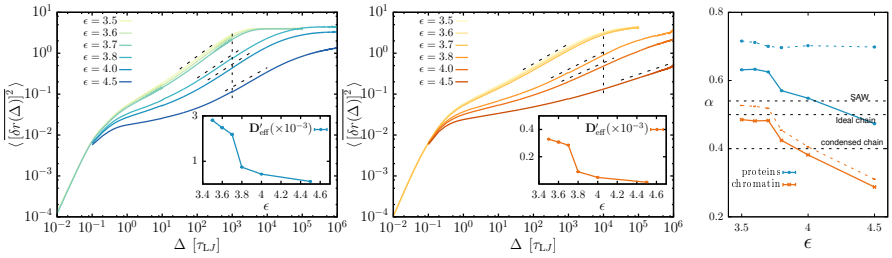

Figure 3.10. Radial MSD and alternative effective diffusivity for non–specific interactions. (a) The radial mean–square displacement curves for proteins is similar to those in Fig. 3.9. An inflection point at intermediate Δ (especially for large energies) indicates a regime in which proteins are 'caged'. These curves are well separated at $\epsilon = 3.8$; in the inset, we show the measure of diffusivity in (3.31), that is discontinuous at $\epsilon \simeq 3.8 K_B T$. (b) The radial MSD curves for polymer beads behaves qualitatively as those obtained for proteins. Polymer dynamics is slower, as predicted by $D'_{\rm eff}$, which is about one order of magnitude smaller. (c) Here we show the exponent α obtained by the linear fits in Fig. 3.10(a–b) (solid lines). Note that the slopes of rMSD curves for protein depends on the interaction energy in the solid phase (i.e. $\epsilon \geq 3.8$). Instead, the exponent extracted for chromatin rMSD is slightly below with respect to the α obtained by fitting pMSD curves [compare with dashed curves, which reproduce the plots in Fig. 3.9].

(no saturation occurring). We also remark that the exponent is approximately constant for all values of interaction strengths, which ensures that $D_{\rm eff}$ is correctly defined for every ϵ. In Fig. 3.9(c) we show the effective diffusion $D_{\rm eff}$ as a function of the interaction energy ϵ. Both curves display a sharp crossover at $\epsilon = 3.8$, which suggest that our system undergoes a discontinuous liquid–to–solid–like transition in the thermodynamic limit ($N, L \to \infty$). In the inset in Fig. 3.9(c), the dynamical exponent obtained by a linear fit of the pMSD curves in Fig. 3.9(a) are plotted as a function of ϵ. Protein dynamics are always *sub–diffusive*, with $\alpha \simeq 0.75$ for any ϵ.

Another measure of the internal droplet dynamics is the radial mean square displacement (rMSD), which measures the mean squared displacement $\langle \overline{\delta r_i^2} \rangle$, where r_i is the distance from the center of mass of the droplet:

$$\langle \overline{[\delta r(\Delta)]^2} \rangle = \frac{1}{T-\Delta} \sum_{i=1}^{N_i^*} \int_0^{T-\Delta} [r_i(t+\Delta) - r_i(t)]^2 \, dt \,, \tag{3.31}$$

the rMSD are shown in Fig. 3.10(a–b), for both proteins and chromatin. Note that the discontinuous nature of the transition is retained, as the curves are largely separated at $\epsilon = 3.8$. On the other hand, the value of the exponent α slightly differs from the ones found though the pMSD, see Fig. 3.10(c). In particular, for proteins α varies with energy for proteins. This can be explained by the following argument: in the solid phase, two protein subpopulation can coexist (as we will see much in detail lateron), thus n proteins will belong to the

slow component, $N - n$ to the fast ones. Therefore, about $n^2/2$ pairs contribute poorly to the total pMSD, while about $N^2/2 - Nn/2$ pairs move as 'freely' as in the liquid phase. If n is relatively small (as it is the case for $\epsilon \leq 4.5$, see Fig. 3.13 below), then the contribution of the $n^2/2$ immobile pairs is negligible, at least in our range of parameters. Observe that we have not computed the standard mean square displacement (MSD) of proteins: despite it is simple to subtract the MSD of the center of mass, it is much more complicated to rule out the effect of the droplet rotation. Nonetheless, an inspection of the standard MSD shows a discontinuous transition at $\epsilon \simeq 3.8$.

An estimation of the effective diffusivity from the radial displacement cannot be extracted, as the dimension of this quantity changes with the exponent α (as it does with ϵ). Naively, one can define an effective diffusion D'_{eff} by considering the largest Δ^* available from the curves, and defining $D'_{\text{eff}} \equiv \langle \overline{\delta r(\Delta^*)} \rangle / \Delta^*$. This measure gives results which are compatible with those in Fig. 3.9, e.g. it retains the discontinuity at $\epsilon = 3.8$. We should also remark that, although the pMSD is more effective than rMSD in extracting the effective diffusivity, in the following we will also use the rMSD as a dynamical property associated to each single protein.

Simulated FRAP

In order to characterize the dynamical properties of the two phases, we simulate a FRAP (fluorescence recovery after photobleaching) experiment, by 'staining' all proteins within one half of the droplet at time $t = 0$ and observing the internal mixing. From movies, we observe that the internal mixing progressively slows down as the energy increases. To describe more quantitatively the internal recovery, we calculate the following 'mixing' parameter

$$\mu_t = 1 - \frac{2}{N} \sum_i \sum_{j \neq i} \frac{(1 - \delta_{x_i,x_j}) \Theta(d - d_{ij}(t))}{n_{i,\text{neigh}}(t)} , \tag{3.32}$$

where x_i is 1 if the protein is stained, 0 otherwise, δ_{x_i,x_j} is the Kronecker delta, and $n_{i,\text{neigh}}(t) = \sum_j \Theta(d - d_{ij}(t))$ is the number of proteins within a distance d from the ith protein, at time t; $\Theta(x)$ is 1 if $x > 0$, 0 otherwise. Ideally, this parameter ranges from 1 (totally demixed) to 0 (totally mixed) in an infinitely large binary mixture. The finiteness of the system and the presence of a small fraction of particles which can be exchanged between the droplet and the low–density phase can slightly modify this range. However, as we will see below, is it still possible to extract at least one relaxation timescale.

If the internal mixing occurs due to standard diffusion through the interface which divides the two halves, the recovery is expected to be exponential, i.e $\mu_t \sim \exp(-t/\tau_r)$, where τ_r is the relaxation timescale: such a behaviour

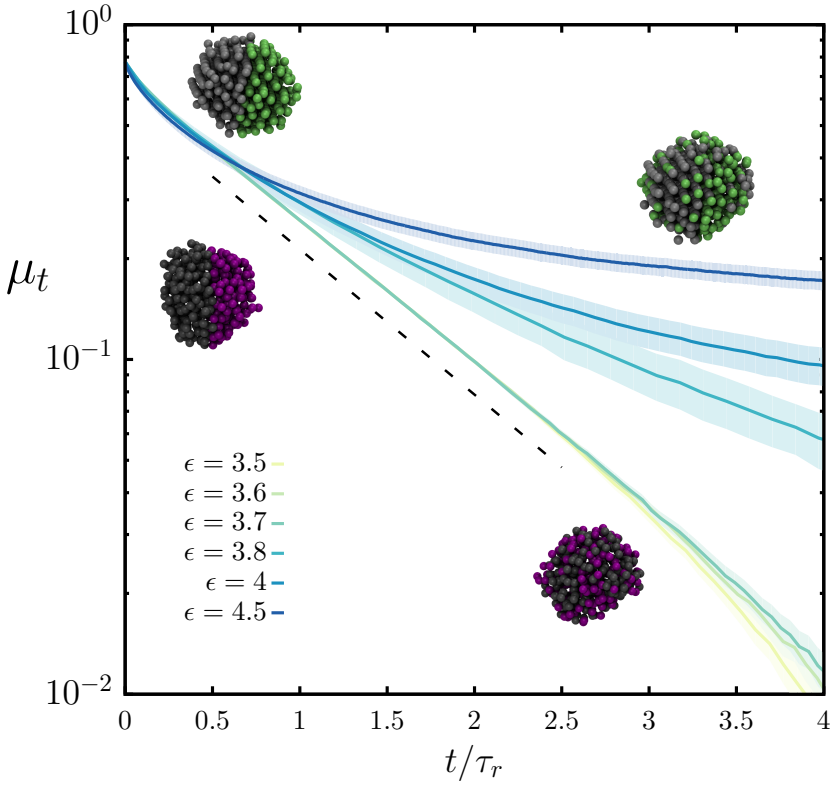

Figure 3.11. FRAP mixing parameter. Here, we show the behaviour of μ_t as a function of time, rescaled by τ_r. The value of τ_r is obtained with the procedure described in For $\epsilon < 3.8$ (liquid dynamics), the recovery occurs in a diffusive fashion, and the relaxation curve decays exponentially (dashed line) and sit on top of each other. Correspondingly, we show two representative snapshots of the FRAP experiment *in silico*, at $t = 0$ and $t = 5 \cdot 10^5 \tau_{\mathrm{LJ}}$. Note that two halves recombine, resulting in a well mixed droplet. For $\epsilon \geq 3.8$ (solid dynamics), relaxation curves do not overlap, and decay slower than an exponential. Similar snapshots show a partial recombination of the two halves, even if local ordered structures are preserved.

is expected for liquid dynamics. To the contrary, if the dynamics are sub–diffusive, the relaxation might not be exponential in time; this is the case in glassy and dynamically heterogeneous systems, where the structural relaxation is known to follow a stretched exponential law, namely $f(t) \sim \exp\left[-(t/\tau_r)^\beta\right]$, with $\beta < 1$. For $\epsilon < 3.8$ the parameter μ_t decays exponentially, for a few relaxation times (τ_r). Then, the mixing parameter can become larger than 0, due to spurious effects introduced by protein exchange with the soluble pool; in the log–linear plot this cause a drop of the curves for large times ($t \gg \tau_r$). Note that all curves collapse after rescaling time. Conversely, for $\epsilon \geq 3.8$ the decay is clearly non–exponential, see Fig. 3.11.

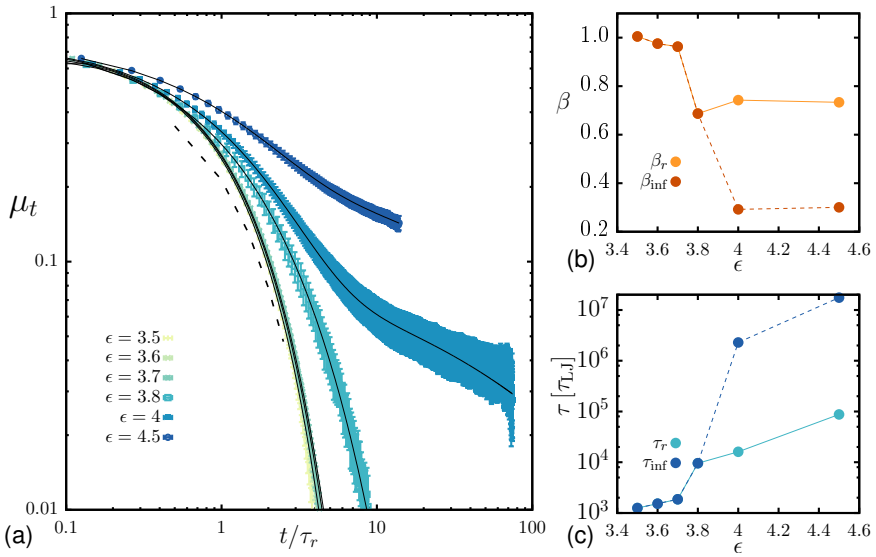

Figure 3.12. Log–log plot of mixing parameter and relaxation times for non–specific interactions. (a) Here, we present the mixing parameter vs. time, in a double logarithmic scale. We note that, for $\epsilon > 3.8$, a heavy tail appears, showing that a long–time relaxation occurs. (b–c) The exponent β and relaxation timescales as a function of energy. Note that the liquid–to–solid transition reflects into the discontinuities of β_r, $\beta_{\rm inf}$, τ_r and $\tau_{\rm inf}$.

We fit all curves with either a single ($B = 0$) or a double stretched exponential (excluding the drop at large times) by using the fitting expression:

$$f(t) = A\mathrm{e}^{-\left(\frac{t}{\tau_r}\right)^{\beta}_r} + B\mathrm{e}^{-\left(\frac{t}{\tau_{\rm inf}}\right)^{\beta_{\rm inf}}} \tag{3.33}$$

In Fig. 3.12(a) we show the curves in Fig. 3.11 on a log–log plot, with the error bars calculated as standard error of the mean value. For $\epsilon < 3.8$, the fit with a single stretched exponential [Eq. (3.33) with $B = 0$] gives a value of β close to 1 (as reported in Fig. 3.12(b)), as expected for an exponential decay. For $\epsilon = 3.8$, the mixing parameter is well fitted by a single stretched exponential with $\beta < 1$, whereas for $\epsilon > 3.8$, a second stretched exponential is required to fit the curves at large times. Therefore, for $t \gg \tau_r$ a larger relaxation timescale, $\tau_{\rm inf}$ is detected; the first relaxation timescale is likely to be related to the partial recombination due to the presence of liquid/disordered regions; in fact, although regular structures appear in large regions of clusters when in the solid–like phase, a partial recombination still occurs, as demonstrated by the snapshots in Fig. 3.11. Instead, the second larger relaxation timescale could be associated with the disruption of the solid structure. Note also that this typical time is equal or larger than the simulation time, see Fig. 3.12(c), and that the associated curves in Fig. 3.33(a) do not decay to zero. Thus, the

extracted values of $\beta_{\rm inf}$ and $\tau_{\rm inf}$ are not reliable; nonetheless, the discontinuous character of the transition is ascertained by β_r and τ_r.

In Fig. 3.12(b–c) we show the fitted values of β_r and $\beta_{\rm inf}$, τ_r and $\tau_{\rm inf}$, respectively. Observe that the discontinuous nature of the liquid–to–solid–like transition is retained by β and the relaxation times of clusters. All these quantities display a jump at $\epsilon = 3.8$.

Dynamical heterogeneity and bond–ordering parameter

To further investigate the single–protein dynamics, we consider some dynamical quantities associated with each protein (or chromatin) bead, and their statistics. First, we consider $\delta r_t(t^*)^2$, where t^* is fixed and defined as the lag time at which proteins commit to a new position, such that $\delta r^2 \equiv \langle \delta r_i(t - t^*) \rangle \simeq 0.2$ on average, for every i (lateron, we will use δr^2 and $\langle \delta r_i(t - t^*) \rangle$ interchangeably, for the sake of simplicity). In Fig. 3.13(a) we show two trajectories of single proteins. The two trajectories are representative of two distinct sub–populations: the first trajectory (blue) is always close to $\delta r^2 \simeq 0$, and it is referred to a *slow* protein, whereas the other one (red) presents large fluctuations, and it is representative of the *fast* sub–population. Note that the rMSD at fixed lag times is able to distinguish between the two populations, while the standard MSD would not, as it probes droplet translation and rotation, that 'hides' the internal protein motion [see three–dimensional trajectories showed in Fig. 3.13(a)]. The separation between the two sub–populations emerges if we consider the following time–averaged quantity, for each particle (we drop the index i for simplicity):

$$\overline{\delta r_t(t^*)^2} = \frac{1}{t - t^*} \int_0^{t-t^*} \left[r(t' + t^*) - r(t') \right]^2 dt' \tag{3.34}$$

This quantity allows us to track the *time–averaged* mobility of single proteins as time t increases. In Fig. 3.13(b) we show the behavior of $\overline{\delta r_t(t^*)^2}$ for a sample of proteins which remain inside the droplet for the whole duration of a single simulation run ($T = 1000t^*$), for two values of the interaction energy, $\epsilon = 3.5$,4. At time $t = t^*$, the distribution of the squared displacements δr^2 (δd^2) are strongly non–Gaussian for any value of ϵ, as proteins are confined inside the cluster (not shown). Conversely, the time–averaged rMSD trajectories differs for the two cases: in the liquid phase ($\epsilon = 3.5$) all trajectories tend to the mean value $\delta r^2 \simeq 0.2\sigma$, whereas in the solid phase ($\epsilon = 4$) the trajectories can be separated into 'mobile' and 'immobile'; the proteins that belongs to the first fraction are more than 10 times faster than the those which are immobile (or slow). The threshold rMSD is $\overline{\delta r_t(t^*)^2} = 0.05\sigma$, since this correspond to a local minimum in the associated bistable distributions, as described below.

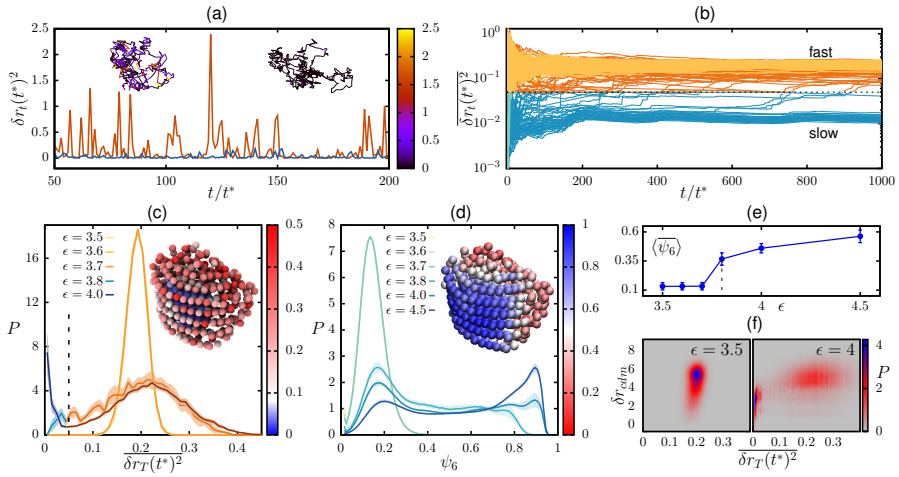

Figure 3.13. Dynamical heterogeneity and bond–ordering parameter. (a) Two $\delta r_t(t^*)^2$ time series are shown for two different proteins inside a cluster for a simulation at $\epsilon = 4$. The blue trajectory is representative of the slow population, being always very close to $\delta r^2 = 0$. On the contrary, the red trajectory show wild fluctuations which signify that the fast protein easily escape cages and 'travels' within the droplet. The two snapshots represents the actual trajectories of the two proteins in 3D. Colors signal the relative radial displacement at each time increment. Although one (left snapshot) is much more mobile than the other (right snapshots), the two proteins cover approximately the same volume space, due to droplet translation and rotation. (b) Time–averaged rMSD is probed for 94 proteins within a droplet, which forms in two different simulations at $\epsilon = 3.5$,4. While in the liquid phase all trajectories converge to the expected values $\delta r^2 \simeq 0.2$, in the solid phase, proteins display dynamical heterogeneity. (c) Probability distribution of $\overline{\delta r_T(t^*)^2}$, obtained averaging over particles and $N_{\rm eq}^{\epsilon}$ replica. For $\epsilon < 3.8$ distributions are unimodal, whereas for $\epsilon \geq 3.8$ they are long–tailed or bimodal. The fraction of slow proteins is represented by the blue portion of the distribution; the vertical dashed line at δr^2 indicates the minimum of the distribution at $\epsilon = 4$. The snapshots depicts the average mobility of proteins, as colors represent the value of $\overline{\delta r_T(t^*)^2}$. (d) Probability distribution of ψ_6 correspondent to those in panel (c). Here, distributions show structural heterogeneity in the solid–like phase. The snapshot displays that more structured portion of the droplet have a larger ψ_6 (blue) than less ordered ones (red). (e) Order parameter for local ordering $\langle\overline{\psi_6}\rangle$. The plot shows a small jump in the value of the order parameter (~ 0.25) at the transition, and a smooth increase as the energy goes up, signalling a weak first–order transition. (f) Joint probability distribution $P(\delta r_{\rm cm}, \overline{\delta r_T(t^*)^2})$. In the liquid phase, the distribution is unimodal. When in the solid phase, the distribution is bimodal, and peaked at low values of mobility and distances from the center of mass of the droplet ($\delta r_{\rm cm} \sim 2\sigma$).

Dynamical heterogeneity is particular of this system, since there is a co-existence of regions well structured in hexagonal–like arrays [resembling the hexagonal order in standard mono–species Lenard-Jones colloids] and defective disordered regions. This partial ordering is maintained by the chromatin fiber, which locally organizes itself into ordered arrays [see snapshots in

Fig. 3.9(a)]. Differently from a single–constituent liquid just beyond the freezing point, in which 'cages' continuously form and break, here transient cages are conserved for much longer, as chromatin motion is slower than protein rearrangements; nevertheless, proteins can slowly be exchanged between the two sub–population, as shown in Fig. 3.13(b). This behaviour explains the partial recombination which occurs at relatively high interaction energy.

This picture is also supported by the probability distributions of the time–averaged rMSD in Eq. (3.34), calculated at $t = T$. The distribution $\overline{\delta r_t(\Delta)^2}$, should be Gaussian if $\Delta \ll \tau_r \ll t$ and the dynamics are homogeneous. In Fig. 3.13(c) we show the distribution $P[\overline{\delta r_T(t^*)^2}]$ for proteins squared displacements. Distributions are clearly Gaussian in the liquid phase, while a largely non–Gaussian tail ($\epsilon = 3.8$) or a second peak ($\epsilon = 4$) appear at small values of the squared displacement. Conversely, the distribution of displacements of chromatin beads calculated at t^* are always unimodal (not shown). Moreover, for $\epsilon = 3.8, 4$, distributions are peaked at $\overline{\delta r_T(t^*)^2} \sim 0$, suggesting that the structured chromatin regions remain immobile for the whole simulation time T, which is the same timescale at which proteins exhibit heterogeneous dynamics. Observe that the bistable distributions in Fig. 3.13 display a local minimum at $\sim 0.05\sigma$, which is the separation between the slow and the fast populations.

In this system, heterogeneous dynamics is related to the appearance of ordered regions in the droplet. To quantify the level of ordering, we compute the local bond–ordering parameter, ψ_6 [see section 3.17]. In Fig. 3.13(d) we show the distribution of ψ_6, for different values of the interaction energy. We note that in the liquid phase ($\epsilon < 3.8$) the distribution are unimodal, and centered at $\psi_6 \sim 0.12$, as expected for a liquid (or a disordered system); moreover, distributions for different ϵ perfectly overlap. Beyond the transition point a long tail ($\epsilon = 3.8$) or a second peak at $\psi_6 \sim 0.9$ ($\epsilon > 3.8$) appears. Therefore, dynamical heterogeneity correlates with *structural* heterogeneity. The local bond–ordering parameter can also inform on the nature of the transition: $\langle\overline{\psi_6}\rangle$ is constant for $\epsilon < 3.8$, then jumps to larger values at the transition, as shown in Fig. 3.13(e). Differently from mono–species LJ particles, where distribution in the solid phase are unimodal and the transition is sharper (not shown), here $\langle\overline{\psi_6}\rangle$ increases smoothly with energy in the solid phase, which is compatible with a *weak* first–order phase transition.

The position of the 'immobile' fraction inside the droplet differs slightly from the population of ordered proteins. As shown in the snapshots in Figs. 3.13(c-d), slow proteins are confined in the bulk of the cluster, while the more ordered proteins are located *on* the surface. This is because the proteins on the surface can be only partially caged, and they can move more freely along the radial direction. We consider the probability associated with the time–averaged rMSD and the distance from the center of mass δr_{cm}, comparing

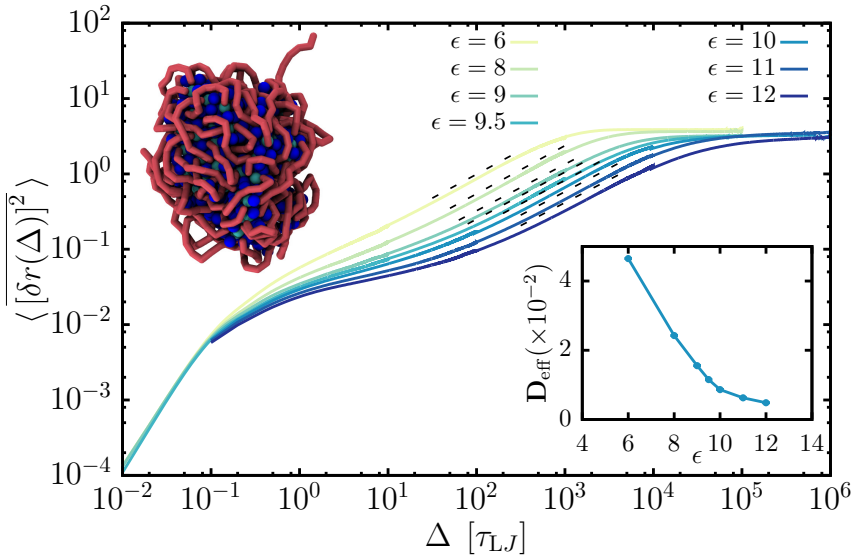

Figure 3.14. radial MSD (specific). Here, we show The total pMSD for proteins as a function of the Δ. Similarly to Fig. 3.9 curves display an inflection point or a plateau at the timescale of caging. Then, a sub–diffusion regime leads to the saturation of the curv, due to the effective confinement of particles in clusters. The snapshot show a typical equilibrium configuration at $\epsilon = 11 k_B T$. In the inset, the effective diffusivity D_{eff} is plotted as a function of ϵ_{sp}. No jump is detected, rather it smoothly decreases to smaller values as the energy increases.

the multivariate distributions in the liquid and solid phases in Fig. 3.13(f). As expected, in the solid, phase the slow proteins ($P[\overline{\delta r_T(t^*)^2}] \sim 0$) are (on average) in the bulk ($\sigma < \delta r_{cm} < 4\sigma$), while more mobile proteins are closer to the surface ($\delta r_{cm} > 4\sigma$).

3.4.6 Specific binding sites lead to glassy dynamics

Now, we switch to the second chromatin model. We study the dynamics of proteins when interacting to a polymer with strong binding sites with affinity $\epsilon_{\text{sp}} \geq 6 k_B T$. The remaining chromatin is only weakly 'sticky' ($\epsilon = 3 k_B T$). For a smaller number of proteins and longer weakly–interacting polymer stretches, a *microphase* separation can occur, due to the entropic repulsion among loops.

Here, we still consider $N = 400$ and $L = 1000$, and the weakly–interacting chromatin stretches are about 30 kbp long, namely 10 beads in our coarse–grained model. Also in this case the system undergoes a full phase separation, at every interaction value; to avoid the formation of non–spherical structures at high energies, we perform an initial simulation run at $\epsilon_{\text{sp}} = \epsilon = 3 k_B T$, then we quench the system at larger ϵ_{sp}. As a consequence, a spherical–like droplet forms, being all the specific binding sites 'absorbed' in the bulk of the

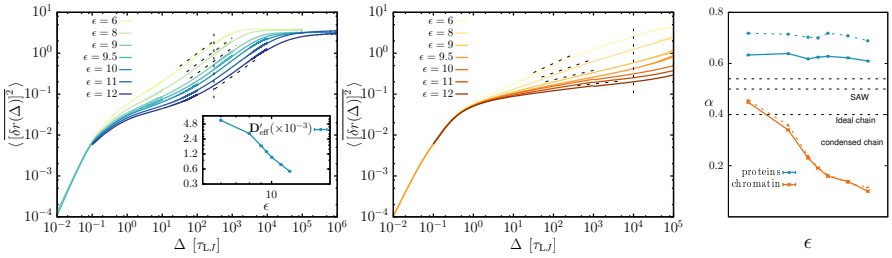

Figure 3.15. Radial MSD and alternative effective diffusivity in the presence of strong binding sites. (a) The rMSD curves resemble those in 3.14 in the main text. Protein caging becomes more relevant at large energy values; although the dynamics is slower as ϵ increases, there is no jump in the effective diffusivity D'_{eff}. (b) Here we show the rMSD curves for polymer beads. The effective diffusivity of chromatin beads is continuous as in 3.14 (c) Differently from the non–specific case, the linear fits of rMSD curves in 3.14 give a value of α which is approximately constant for any energy value ($\alpha \sim 0.62$, solid curve), but it is underestimated compared with those obtained by fitting the pMSD plots (dashed curves).

protein droplet, while several weakly–binding chromatin loops extend out of the cluster, as shown in the snapshot in Fig. 3.14.

We calculate the pMSD for different values of the specific interaction energy, as displayed in Fig. 3.14. Similarly to the non–specific case, after a pseudo–diffusive regime, a plateau can appear (especially at large energy), which is compatible with a caging dynamics, then the curve grows as proteins 'break' the cages, until it saturates, since we limit the pMSD to those proteins which remain inside the droplet during the whole simulation run. Interestingly, in this case there is no clear separation between the curves, as supported by the effective diffusivity, which continuously decreases with the interaction energy. Thus, there is no signal of a discontinuous transition. Accordingly, also the other measure of diffusivity vary continuously with energy. in Fig. 3.15, we present the plots of rMSD for both proteins and polymer (pMSD curves are not shown for polymer, as it behaves exactly like rMSD). Notably, both protein pMSD and the rMSD curves give a sub–diffusion exponent which is approximately constant with ϵ_{sp} ($\alpha = 0.7$ and $\alpha = 0.62$, respectively), see Fig. 3.15(c).

Mixing parameter and simulated FRAP

In Fig. 3.16 we show the behaviour of the mixing parameter μ_t as a function of time, in a log–log scale. Similar to the non–specific case, the presence of a majority of 'weak' binding sites ($\epsilon = 3K_{\text{B}}T$) makes the surface of the droplet more liquid,causing a faster drop in the liquid regime. In the glassy regime, τ_r grows linearly without discontinuity. The relaxation curves for $\epsilon_{\text{sp}} = 11k_{\text{B}}T$ and $\epsilon_{\text{sp}} = 12k_{\text{B}}T$ can be fitted with a double stretched exponential: in this case

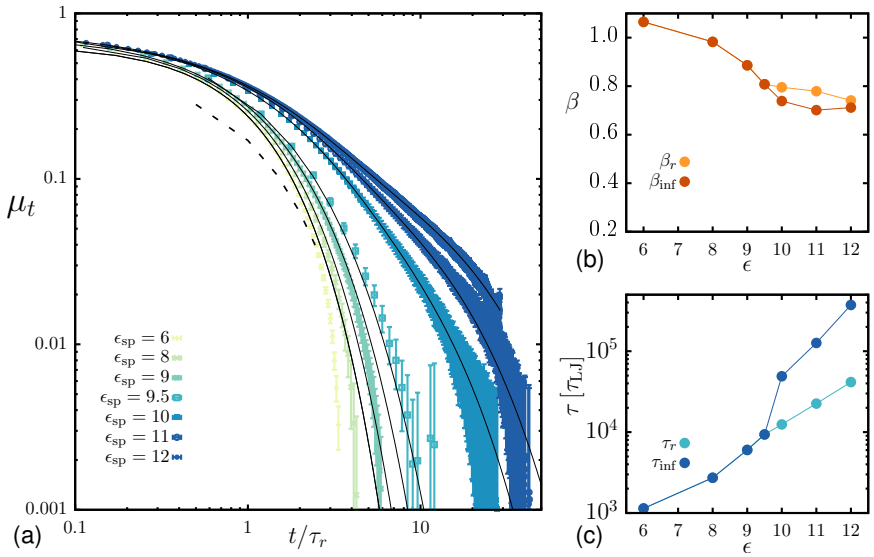

Figure 3.16. Log–log plot of mixing parameter and relaxation times in the presence of strong specific binding sites. (a) The mixing parameter presents a single realxation timescale up to $\epsilon_{sp} > 9$; for larger specific energies, a second relaxation timescale can be extracted. (b–c) The exponent β and relaxation timescales as a function of energy. Differently from Fig. 3.12, both β_r and β_{inf} change value smoothly, from $\beta \sim 1$ to $\beta \sim 0.7$. Accordingly, τ_r does not display any discontinuity, whereas τ_{inf} shows a jump at $\epsilon_{sp} = 10$.

$\beta_r \simeq \beta_{inf}$, and $\tau_{inf} > \tau_r$, but less than one order of magnitude larger. This behaviour might be caused by some residual local ordered structures which can persiste for the whole simulation time, as shown below in Fig. 3.17.

Dynamical heterogeneity and glassy structure

To further investigate the protein dynamics, we monitor the time–averaged rMSD trajectories during rescaled time, see Eq. (3.34); again, t^* is the time at which proteins have committed to a new position such as $\delta r^2 \simeq 0.2$ on average. In Fig. 3.17(a) we show several trajectories for single simulations performed at $\epsilon_{sp} = 6k_BT$ and $\epsilon_{sp} = 11k_BT$. For $\epsilon_{sp} = 6k_BT$, a single population of proteins is detected, with average mobility close to the expected average value of 0.2, at large t/t^*. At $\epsilon_{sp} = 11k_BT$, the system displays dynamical heterogeneity: a separation between fast and slow proteins can be identified, even though more particles can be exchanged between the two sub–populations then in the non–specific case, due to the ability of particles to escape more easily from the cages, since the majority of chromatin beads interact only weakly with proteins.

Strikingly, even though no sharp transition is predicted by the effective diffusivity measurements, there is a crossover between a homogeneous and a heterogeneous dynamical phases. This result is supported by the distributions

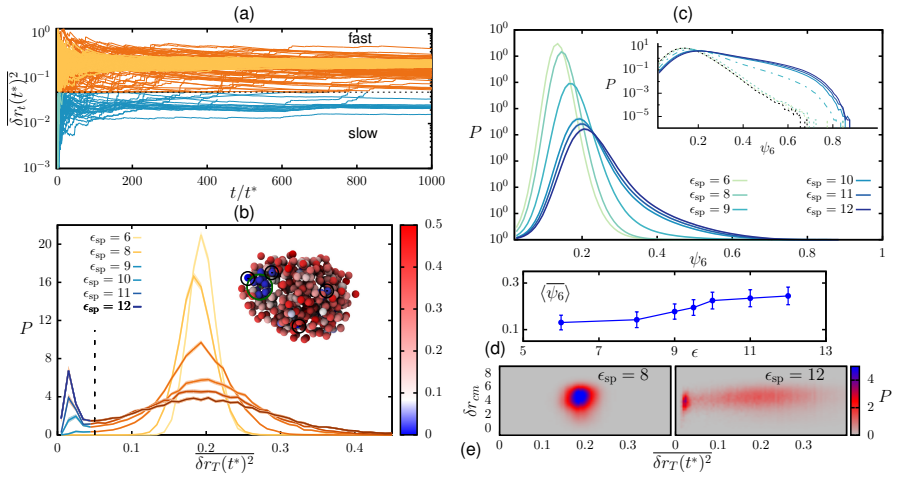

Figure 3.17. Dynamical heterogeneity and structural disorder entail glassy dynamics. (a) Alike Fig. 3.9(a) two time–averaged rMSD trajectories are shown for a sample of 94 proteins within the same droplet, for simulations at $\epsilon_{sp} = 6k_BT$ and $\epsilon_{sp} = 12k_BT$. In the liquid phase all trajectories converge to the expected values $\delta r^2 \simeq 0.2$, whereas in the glassy phase, proteins separates into a fast (orange trajectories) and slow (blue trajectories) fractions. Again, the separating threshold is set to $0.05\sigma^2$. (b) Probability distribution of $\overline{\delta r_T(t^*)^2}$. For $\epsilon < 9k_BT$ distributions are unimodal, and no particles fall into the slow sub–population. For $\epsilon \geq 9k_BT$ the region $[0, 0.05\sigma^2]$ starts to be populated, until a high peak arises at about $0.02\sigma^2$ at large specific energies. In the snapshot, we can see a typical equilibrium configuration (polymer is not shown), where red proteins are the more dynamical, whereas blue proteins are 'immobile'. We pinpoint the position of some sparse immobile proteins with black circles, while the green circle locates a locally ordered hexagonal–like sub–cluster of slow particles.

of $\overline{\delta r_T(t^*)^2}$ in Fig. 3.17(c). For $\epsilon_{sp} = 6k_BT$ and $\epsilon_{sp} = 8k_BT$ distributions are unimodal, with a non–zero probability only in the region $\overline{\delta r_T(t^*)^2} > 0.05\sigma^2$. Conversely, at $\epsilon \geq 9k_BT$, the region $\overline{\delta r_T(t^*)^2} > 0.05$ associated with 'immobile' proteins becomes progressively to be populated, and at $\sim 0.02\sigma^2$ peak arises for $\epsilon_{sp} \geq 10k_BT$. We note in the snapshots in Fig. 3.17(c) that only few of the proteins which belongs to the slow fraction are associated with a partial local ordering (observe the green circle, that highlights a two–layer hexagonal–like cell of slow particles), and the whole cluster appears globally disordered.

To quantify the level of local ordering, we calculate the distributions of ψ_6, that we show in Fig. 3.17(c). We find that the probability distributions are always unimodal and peaked at $\psi_6 \sim 0.2$, as usually found in disordered structures in other contexts, and differently from the non–specific case, where there is a coexistence between large disordered and ordered regions. The lack of local ordering is typical of glasses, which can display an heterogeneous dynamics with no nucleation of solid–like regions. This is supported by the nucleation plots (not shown): for the liquid–to–solid transition, we observe

the nucleation of a solid by monitoring the timeseries of the energy, during equilibration (note that this occurs more frequently at larger energies); in the specific case, no nucleation have been observed for any value of the specific interaction energy, rather the energy of the systems fluctuate about a constant value through time. However, most of the time, regions of local ordering transiently form and disrupt, while only sporadically they last for the whole simulation. By comparing the distribution of ψ_6 observed in the non–specific case in the liquid phase and those in Fig. 3.17, we note that the system crosses over between two regimes [see the inset of Fig. 3.17(c)]: the distributions obtained for $\epsilon_{\rm sp} = 6k_{\rm B}T$ and $\epsilon_{\rm sp} = 8k_{\rm B}T$ overlap with the distribution at $\epsilon = 3.5$ (liquid phase). Conversely, for $\epsilon_{\rm sp} > 9$ the long tail is much more populated, and the crossover is located at approximately $\epsilon_{\rm sp} = 9$, as vividly shown in the inset of Fig. 3.17(c). Also, the order parameter $\langle\overline{\psi_6}\rangle$ increases smoothly as ϵ_{sp} increases, see Fig. 3.17(d). Dynamical heterogeneity with a low degree of orientational order is often related to a glassy dynamics. Indeed, there is no clear formation of a long–range order in clusters at large specific interaction energies, and a faster exchange between the two dynamical sub-populations occurs. The liquid–to–glass transition of clusters is accompanied by a particular distribution of proteins and specific binding sites inside the droplet: the joint probability $P[\overline{\delta r_T(t^*)^2}, \delta r_{cm}]$ displays a slightly different distributions of slow and fast proteins compared to those in Fig. 3.13(f). In fact, in the glassy phase both fast and slow proteins prefer to stay in an intermediate shell located between 4σ and 6σ, as shown in Fig. 3.17(e).

3.5 Conclusions and future perspectives

In summary, we have studied the dynamical and structural properties of clusters which form due to bridging–induced attraction, first introduced in [162]. In our simulation, the equilibrium configuration is always represented by a single, large droplet made of proteins and chromatin.

For a chromatin modelled as an homopolymer, we find a weak first–order transition between a liquid and a solid–like phases. The effective diffusivity obtained from mean–square displacement measures display a jump at the transition point, as well as other relevant quantities the relaxation timescale associated with internal recombination and the order parameter $\langle\overline{\psi_6}\rangle$, associated with the extent of local ordering of the droplet. The distributions of the time–averaged mobility defined in Eq. (3.34) show that there is coexistence between a fast and a slow fraction of proteins in the solid–like phase. Additionally, the dynamical heterogeneity is accompanied by structural heterogeneity; the latter can be explained by recognising that the solid structure is held together by the

chromatin fiber, and the protein caging is provided by the ordered structure of chromatin. Therefore, an ordered chromatin region corresponds to an ordered structure of protein in clusters, and *vice versa*.

Interestingly, in the presence of strong binding sites, we find a different dynamical phase when the interaction strength between proteins and specific sites becomes large. Unlike the non–specific case, the effective diffusion and the relaxation timescale related to internal mixing vary in a continuous fashion as the specific interaction energy increases. Dynamics are still heterogeneous, but this is not associated with structural heterogeneous local order. This is typical of glasses, where dynamics are heterogeneous but structures are inherently disordered. A comparison between joined distributions for the non–specific and the specific interaction cases, show that in the former case, the slow fraction prefers to locate in the bulk of the droplet, being separated from the remaining proteins; in the latter both slow and fast populations are both positioned in an intermediate shell inside the droplet.

We should also remark that a first investigation of this system does not show any *aging* effect in both models, which is a typical feature of glasses. Though, further efforts in this direction are in order to definitely exclude such effects in droplets of phase separating proteins.

This project highlights the possible emergence of states of matter different from the liquid one, which has been widely assumed for aggregated of chromatin–binding factors. This might have crucial implications for protein function, for instance a slower dynamics could hinder further recruitment of other chromatin factors. Different lines of research can depart from this work. A simple investigation of the dynamics in the presence of different types of chromatin–binding proteins would be interesting; Another line can be a study of the dynamics in a non–equilibrium scenario where proteins can switch between an active and an inert state, mimicking some PTM and phosphorilation of certain protein domains. From a physical perspective, one could try to vary the fraction of specific sites – e.g. starting from an homopolymer to a polymer with 1/10 ratio of 'strong' *versus* 'weak' binding sites, going through a block copolymer: in such way, a full phase diagram which would depend on the energy of specific sites and the strong *vs.* weak sites ratio could be drawn, revealing interesting crossover between different phases.

Chapter 4

The chromatin mediated phase separation of proteins with multiple domains

Bridging–induced phase separation (BIPS) has been recently recognised in living cells by experiments on cohesin [203]. However, liquid–liquid phase separation (LLPS) is another mechanism that certainly can play a role in protein aggregation. A common notion is that flexible, low complexity and *intrinsically disordered protein* (IDP) [or proteins with *intrinsically disordered regions* (IDRS)] facilitate LLPS [224]. Many IDPs, and several proteins which possess IDRs, have indeed been found to readily phase separate. One example is *heterochromatin protein I* HP1, a heterochromatin–associated protein highly conserved in eukaryotes, constituted by two globular and three IDRs, which is known to dimerise and co–localise with heterochromatin foci [171].

The HP1 was originally identified in *Drosophila melanogaster* embryo nuclei [225]. Several past works have studied the heterochromatic distribution of this protein, showing that *Drosophila* HP1 is rich at the cetromeres and telomers of unfixed mitotic chromosomes [226]. More recent papers have reported that some HP1 paralogs can undergo phase separation *in vitro* [200, 227], and forms liquid droplets upon H3K9me3 binding [200, 228, 229]. Despite this, the exact function of HP1 in heterochromatin formation and gene silencing remains elusive; in light of the experimental evidence, possibilities are that it directly drives chromatin compaction, that it sterically occludes binding of activating proteins, or that it recruits further gene silencing machinery.

The idea behind this work is to study the interplay between LLPS and BIPS, considering how they could drive protein–chromatin foci localisation and compaction *in vivo*. In our simulations, we modeled HP1 dimers as a three-dimensional rigid body made of multiple beads, each representing a different HP1 (dimer) domain. Such an approach has been inspired by works on patchy

particles [230–234]. Patchy particles consist of hard cores accessorised with a discrete number of attractive sites, or patches, on their surfaces. Comparing our simulations with typical results obtained from systems of patchy colloids is useful to understand the equilibrium (or quasi–equilibrium) properties and the dynamics of our model HP1s, when interacting with chromatin and other HP1s. Our primary aim was to investigate the possible behaviour of HP1, or general multivalent proteins, and the biological implications of its interaction with chromatin. Nonetheless, the study revealed also interesting results from a strict physical perspective. In standard LLPS, a growth law between the typical length scale associated with the instability (namely, the droplet radius) and the number of proteins is $R \sim N^{1/3}$, being N the number of constituent composing the high–density phase. Unexpectedly, in a particular region of the parameter space of our system, we find that R^3 grows sub–linearly with N, and the density of the droplet slightly grows as N increases. Such a result poses new interest into the physical properties of those systems which undergo BIPS.

This chapter is organised as follows. In section 4.1 I briefly review the HP1 structure and its main functions. In section 4.2, I link HP1 structure to its ability to undergo phase separation; therein, I remark that the specific interactions of HP1 domains with chromatin and other HP1 domains, can give rise to different binding 'modes' and chromatin compaction. In section 4.3 I describe some relevant properties of patchy particles and gelation, which are linked to a recent work which studies phase separating patchy particles via BIPS [235]. This is all part of the literature background. In section 4.4 I report methods and results obtained by simulating (via LAMMPS software) our self–interacting model HP1s when interacting with chromatin, which is the essential part of my work.

4.1 HP1 structure and functions

HP1 is ubiquitous and highly conserved in different organisms, ranging from *S. pombe* (Swi6) to mammals [171], in which three paralogs are present: HP1α, HP1β and HP1γ. In general, the HP1 complex consists of two globular, ordered domains, the *chromoshadow domain* (CSD) and the *chromodomain* (CD), and three flexible IDRs, the *hinge* region (HR), the N–terminal extension (NTE) and the C–terminal extension (CTE). The three paralogs are highly homologous (for instance the CD and CSD of HP1α and HP1γ show 71% and 87% sequence identity, respectively [236]), but can perform different functions. A schematics of HP1 in its linear and three–dimesional structure is shown in Fig. 4.1.

The CSD is responsible of homodimersation of HP1 proteins. In fact, HP1 are often found as dimers *in vivo* [236, 237], through CSD–CSD cross–binding (as shown in Fig. 4.1(b), right panel). The CSD also functions as a hub for

binding different ligands which are, in turn, involved in the transcriptional process. The CTE plays a role in determining the affinity of the dimeric interaction CSD–CSD, as well as the binding strength with some ligands. A more detailed discussion on the role of CSD and its ligands is in [236].

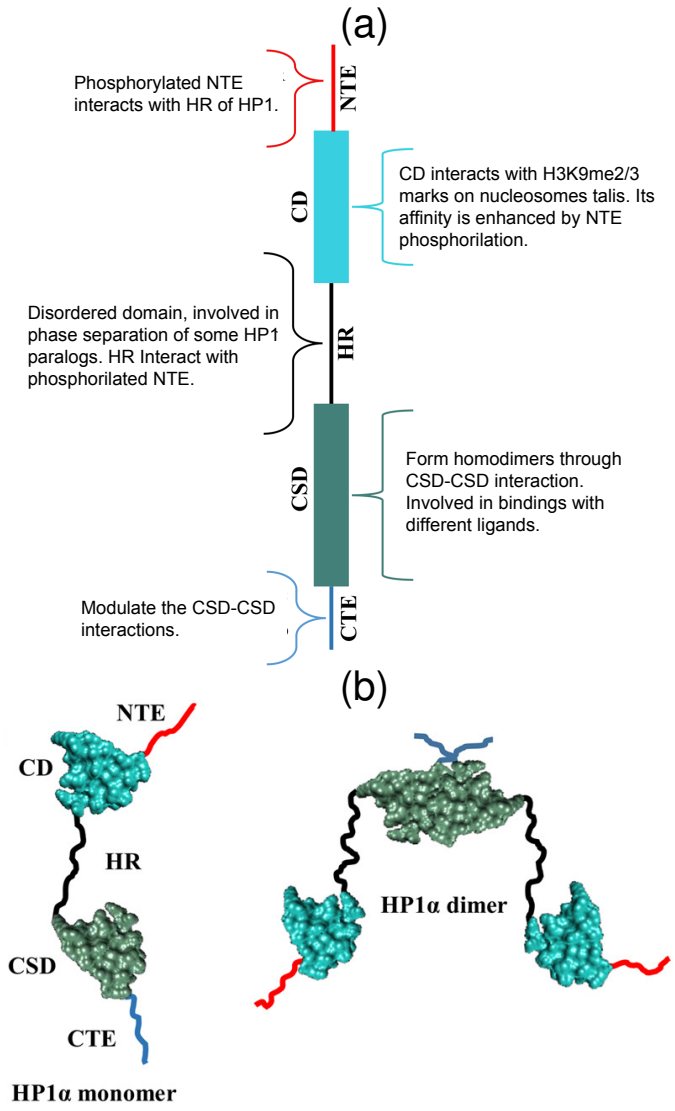

Figure 4.1. HP1 domains and dimer structure. **(a)** Schematics and details of HP1 domains. **(b)** Left: three–dimensional structure in cartoon view of a single HP1. Right: the two CSD are bound to form the dimer; here we show the three–dimensional structure in cartoon view of a HP1α dimer. Figure adapted from [238].

The CD contains a hydrophobic cage which recognises and binds methylated histone proteins with high specificity but low affinity. In particular, many HP1 proteins have been found to bind H3K9me2/3, but with different strength [239–241]. The CD affinity can also be enhanced by some chemical

modification of other HP1 domains, as it is the case of mouse HP1α: there, phosphorilation of serine residues in the NTE increases the CD affinity to H3K9me3 by 5–fold. This suggest that the presence of negative charge in the NTE determines the level of affinity for methylated chromatin: the more negatively charged the NTE, the stronger HP1α binds to methylated domains, which have been found in rich proportion in heterochromatic regions. Therefore, it may be speculated that tuning the CD interaction strength with chromatin could have large effects in heterochromatin constitution and spreading [236]. *In vivo* studies in *Drosophila* point to this direction [242, 243]. On the other hand, CD mutations seems to abolish the activity of HP1 [171].

The hinge is disordered and less conserved among HP1 proteins. Its variability in length and peptide sequence renders its role not well understood yet [236]. Its most relevant role seems to be the ability of some paralogs to phase separate, in the presence of some chemical modifications (phosphorilation) of the NTE, which allow NTE–hinge interactions [227]. In fig. 4.1(a) the role of each HP1 domain is summarised.

From our point of view, the two relevant features are HP1's ability to innteract with chromatin, and to oligomerise and, possibly, create droplet. Thus, it seems that LLPS and BIPS can both be relevant in this case.

4.1.1 HP1 is a multivalent protein that contains IDRs

Intra–cellular phase separation is often driven by intrinsically disordered proteins/regions (IDPs/IDRs) [224]. Classically, folded proteins are defined by the presence of a native state, which is the preferred low–energy state of a protein, which guarantees the proper functions of the protein itself. Conversely, IDPs/IDRs lack such native structures, appearing more mobile and assuming a large amplitude of conformational fluctuations [244]. At one hand, IDRs prevent proteins from folding into a compact tertiary native structure [245]; on the other hand, sequences of IDPs and IDRs are enriched in exposed charges that can promote weak interactions with other substrates, such as homologous proteins (homotypic interactions) or different proteins or complexes (heterotypic interactions).

Together with disorder, which favour the non–specific, collective aggregation of IDPs, also multivalency play an important role in triggering LLPS. Three specific elements are essential in order to determine phase separation: *(i)* the number of sites or motifs that mediate attractive interactions, *(ii)* the strength of each individual binding site, and *(iii)* the relative positions of those sites. When multiple weakly interacting sites are located on a disordered, flexible protein domain, coacervation (separation of a solution into two liquid phases) and LLPS can occur.

HP1 contains three IDR, the hinge, the NTE and the CTE. In particular, the hinge and the phosphorilated NTE are responsible of oligomerisation *in vitro* [227, 246]. In Ref. [246] the authors present experimental evidence that disordered HP1 domains are essential for HP1–DNA phase separation. However, not much is known of the nature of the hinge–NTE interaction. Different types of interaction can lead to different types of macroscopic behaviour: a weak, longer range interaction, for instance, might be associated with a liquid behaviour of HP1 droplet. This is compatible with the presence of multiple weak binding spots on the disordered domain [246]. On the contrary, stronger, short range interactions can arrest the phase separation dynamics, leading to gel–like structures. This could represent the case when a flexible domain undergoes "folding–upon–binding", or simply whenever there are very specific short range interacting sites in the flexible region [247].

4.2 HP1 undergo phase separation and can drive DNA compaction

Experimental lines of evidence suggests diverse roles for HP1. Traditionally, HP1 was identified as the central component of heterochromatin, which can bind the methylation mark on histone H3 lysine 9. Therefore, HP1 is mainly related to gene repression (as heterochromatin, either constitutive or facultative, is commonly associated with transcriptional silencing). However, HP1 plays other roles in addition to gene repression: it promotes chromosome segregation [248], regulates H3 deposition and cohesin action [249] and provides mechanical stability during interphase [250]. However, a deeper understanding of HP1 functions *in vivo* requires a profound knowledge of the mechanisms which promote its aggregation and of the dynamical properties.

4.2.1 HP1: self–assembly and phase separation

Recently, HP1 has shown to self–assemble in different living species. Studies on protein Swi6 in *S. Pombe* (the yeast homologue of HP1) have shown oligomers of these proteins beyond dimers [251]. In human cells, HP1α phase separation is mainly caused by the phosphorilation of NTE [227], as phosphorilation allows NTE–hinge interactions: such modification is essential to switch the auto–inhibited (closed) form to the active (open) state of HP1α, which, in turn, promotes self–interaction at a higher level through NTE–hinge binding.

Differently from cohesin (see section 3.5), this HP1 paralog has been found to clusterise *in vitro*. In Fig. 4.2 we report the experimental evidence of phase separation of HP1a in *Drosophila*, as shown in [200]. In Fig. 4.2(a) the purified

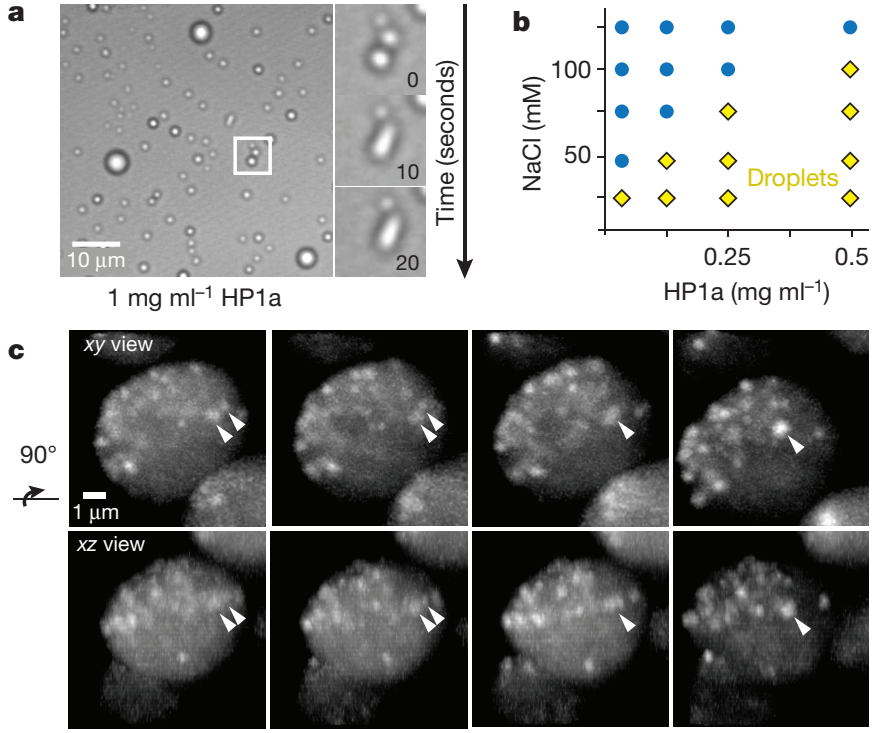

Figure 4.2. Phase separation of *Drosophila* HP1a. (a) Purified Drosophila HP1a forms liquid phase droplets in vitro that undergo fusion. **(b)** Phase diagram of HP1a droplet formation. At low level of salt, or large concentration of HP1a, droplets can form (yellow rhombuses). At high level of salt, or small concentration of HP1a no droplets are detected (blue circles). **(c)** Different planar views of *drosophila embryos* nuclei, where the HP1a are stained by the green fluorescent protein (GFP). GFP-–HP1a forms liquid droplets that fuse and round up. Figure adapted from [200].

HP1a from Drosophila is shown to phase separate *in vitro*, generating highly spherical droplets which merge over time; the LLPS of HP1a occurs at sufficiently high concentration of HP1 and low level of salt [see Fig. 4.2(b)]. *In vivo*, similar fluorescent spots reveal the coexistence of several HP1a droplets, which can merge over time [Fig. 4.2(c)]. As detailed in Chapter 3, these features are typical of LLPS, which would mean that some of the HP1 functions are prominently related to its 'liquid' behaviour once it phase separates.

Although LLPS is considered the main mechanisms for protein aggregation, it does not signify that protein clusters remain in a liquid state for ever. Indeed, immobile condensates or cluster regions have been identified under physiological conditions in nucleoli, stress granules and also heterochromatin (as also reviewed in Chapter 3, see section 3.2). In Ref. [252] the authors demonstrate that liquid droplets of FUS protein, a prion–like protein containing IDRs (associated with the neurodegenerative disease ALS) convert with time from a liquid

to an aggregated, solid state; purified nuclear proteins can phase separate into immiscible droplets, that contain non–coalescing phases which are remarkably similar to nucleoli *in vivo* [253]; pab1 in yeast, a marker of stress granules, phase separates and forms hydrogels *in vitro* upon exposure to physiological stress conditions [254]; droplets of pub1 in budding yeast, which condensate via LLPS, can convert into gel–like structures under pH changes, and solids upon temperature changes [255].

Similarly, the formation of aggregates different from liquids involves also HP1. HP1α, that forms droplet via LLPS, can age and form gels *in vitro*, possibly due to regions with varying mobilities within HP1α molecules. In Ref. [157], specific serine residues of HP1α have been recognised to contribute to gelation: FRAP experiments detect a progressive slowing down of the mixing dynamics, much in line with the results in chapter 3 and in Ref. [256]. Interestingly, the same paper shows that the addition of chromatin (an array containing 12 nucleosomes with a lysine methylation) significantly slows down the rigidification process as evidenced by the retention of mobility in FRAP experiments.

4.2.2 HP1: chromatin binding and compaction

Since the CD can specifically bind H3K9me3 (for both mammalian HP1 and yeast Swi6), it is speculated that HP1 can compact heterochromatin. Not only can the CD bind methylated nucleosomes, but also CSD–CSD dimerisation can stabilise the affinity to chromatin, binding the H3 histone [257]. Although the cooperation of different domains is likely to enhance HP1–chromatin binding, the interaction between CD and methylation marks is thought to be the prime factor that yields the formation of the HP1–chromatin complex [258]. Such speculation is validated by Hiragami–Hamada *et al.*, which found that phosphorylation dissociated HP1α from the nucleosome core, leaving only the CD–H3 interaction as the only relevant interaction [241] for heterochromatin compaction and maintainance [259]. Cryo-electron microscopy [258] and detailed molecular simulations [260] have indicated that HP1 can readily sit between adjacent nucleosomes; this binding mode suggests that (at least under dilute conditions) BIPS can be hindered, as HP1 might 'coat' chromatin, rather than bridging distant nucleosomes and favour chromatin looping, as discussed below. However, the ability of real HP1 dimers to form bridges between distant chromatin regions remains unclear. In the next sections we will describe the macroscopic behaviour of droplets of model proteins, which depends on their bridging ability.

Experiments have thus assessed a relation between HP1 phase separation and chromatin compaction, with a variety of dynamical behaviour, including dynamical heterogeneity, i.e coexistence of a mobile and an immobile fraction

of HP1 in the vicinity of heterchromatin [200]. These findings, along with the results in chapter 3, inspired our work, which tries to answer to some unexplored questions, by examining the interplay between LLPS and BIPS for composite protein geometries. In particular, what are the structural characteristic of phase–separated aggregates when both LLPS and BIPS are in effect? Can different population of HP1 be distinguished depending on their interaction with chromatin? And which kind of dynamics emerges by changing the relevant interactions (HP1–HP1 and HP1–chromatin binding)? In the next section, I will briefly review some interesting properties of 'patchy' particles, which have been much exploited to reveal diverse states of matter and form of aggregation that also differs from standard liquids or crystals (i.e gels or glasses). In Section 4.4, I will report the main result of this project.

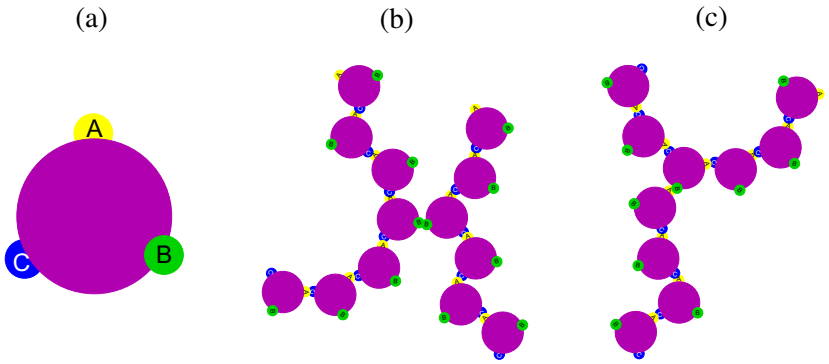

Figure 4.3. Patchy particles. **(a)** Sketch of a patchy particle with three patches, which may have different attraction ranges and strengths ($M = 3$). **(b)** X–junction: a bond connects two particles of two linear chains, as to form an X–shaped structure. **(c)** Y–junction: an end particle of one linear chain is bound to an inner particle of another linear chain to form an Y–shaped structure. Figure taken from [261].

4.3 'Patchy' molecular bridges: gelation and chromatin coating

In the past, patchy particles have much been studied as the simplest example of consituents which do not necessarly nucleate a denser phase; they consists of a simple colloid, decorated by a small number of identical 'sticky' spots, as shown in Fig. 4.3, in fact, after quencing (a sudden change in the external controllable parameters, i.e. temperature, pressure etc.), various mixtures of self–assembled clusters can arise, from chains to rings, or closed loops, and branched clusters. Therefore, patchy particles represent the prototypical examples of aggregates that differs from those typical of vapour–liquid condensation or LLPS as a consequence of directional interactions. We remark that there are other routes

to gelation for simple colloids: non–equilibrium arrested phase separation and equilibrium gels of colloids resulting from competition between short–range attraction and long–range repulsion are vivid examples [232].

4.3.1 Equilibrium gels of patchy particles

Aggregates of patchy particles can produce *equilibrium gels*. A *real* gel is characterised by an inherent disorder, similar to those found in liquids, but displays a large typical length scale signal which reflects the fractal properties of the structure; this means that in a wide regime of lengths, a fractal dimension can be extracted by measuring some structural quantities, such as the radius of gyration as a function of the cluster size, or the pairwise correlation $g(r)$. Gels can be more elastic (as gelatin) or more rigid (as silica gels). Additionally, patchy particles gels can form at very small volume fractions.

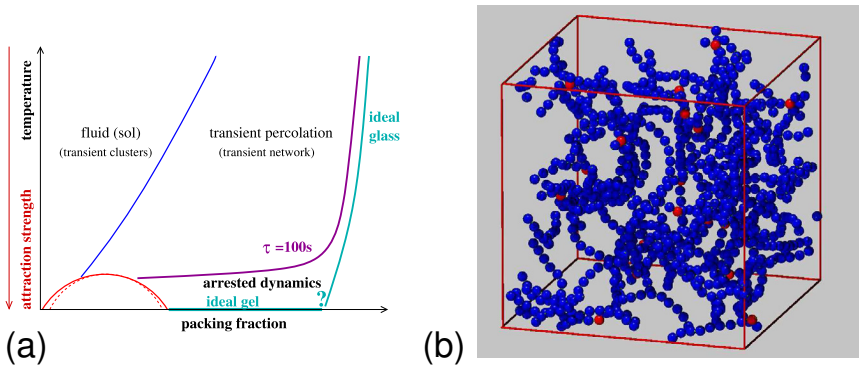

Figure 4.4. Phase diagram and gelation. (a) At $T = 0$, an ideal equilibrium gel phase emerges, and crosses over the glass phase. At low temperature (or high interaction energy) the dynamics is arrested, therefore quasi–ideal gels cann form, with large relaxation times. The question mark refers to the unknown details of the crossover from gel–to–glass dynamics. **(b)** A snapshot from simulations of a gel made of a mixture of two– and three–coordinated particles ($M = 2.025$, $\phi = 0.033$, being ϕ the volume fraction) at very low temperature. Red particles (three neighbours) give rigidity to the gel, while blue particles (two neighbours) provide persistence to the length of network chains. Figure adapted from [232].

Theoretically, an ideal equlibrium gel can only obtained at $T = 0$, as the bond lifetime is infinity. Practically, by limiting the number of neighbours per particle and by including directional interactions between colloids, one can produce a quasi–ideal gel, e.g. a saturated network structure which is characterised by relatively low packing fractions (empty spaces, no bulk liquid), and very long lifetimes. The works in Refs. [262–264] (which have been reviewed in [232]) study the behaviour of colloids interacting with a pre–defined maximum number $N_{\max}$ of neighbours. In such a system, the smaller the coordination number $N_{\max}$, the smaller the volume fraction at which the system

gels, as shown in the schematics in Fig. 4.4(a). Note that the actual nature of the transition/crossover between gels and attractive glasses is not known yet. An ideal gel clearly differs from a glass far from the transition point (or crossover). First, the low coordination number prevents cages from forming at low T. Second, gels have a very large structural length scale, much larger than that typical of glasses. Moreover, non–ergodicity properties depends on the length–scale of observation; indeed, in gels, the dynamics at small length scales appears ergodic, in contrast to what is found in glasses, where all length scales are non–ergodic.

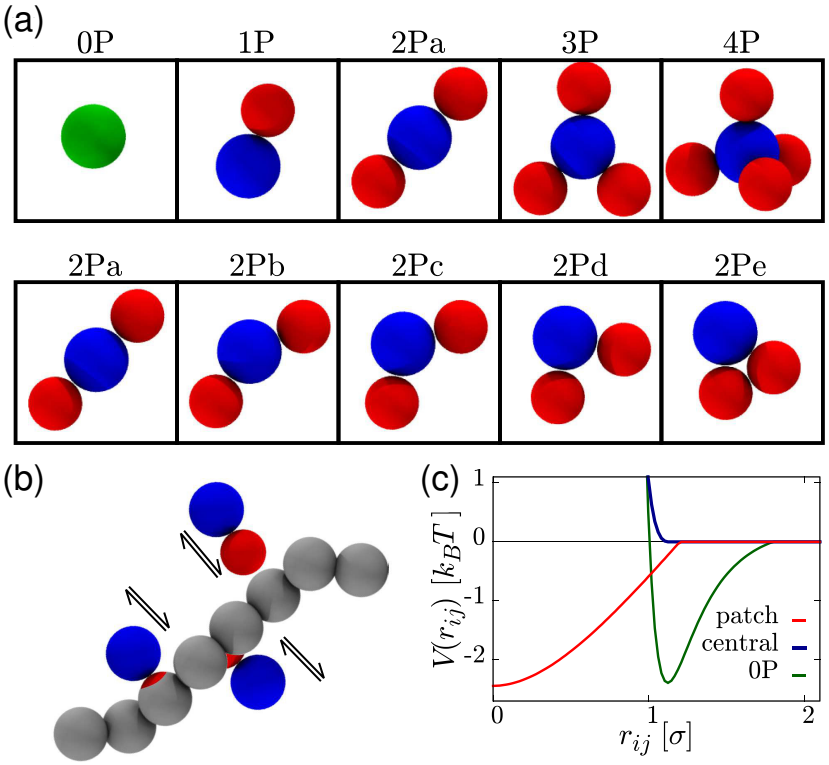

Figure 4.5. Schematics of 'patchy' proteins. (a) Schematics of model protein. Top row: XP indicates the patchy protein consisting of X patchies linked to a core colloid. Bottom row: different angles between the two patches of a 2P protein. **(b)** Schematic showing how patches (red) interact with polymer beads via a purely attractive interaction. The central core (blue) interacts repulsively with proteins. **(c)** Interacting potentials of patch (red curve), core particle (blue curve) and colloid with isotropic attractive interaction. Figure adapted from [235].

The $N_{\max}$ model can be refined by introducing explicit, geometrically organised patches, which can interact via a short–range potential (usually a square–well). For patchy particles, the key parameter is the number of patches M_p. Upon decreasing M_p, the phase separation occurs at very low densities,

and ceases to exist if $M_p \leq 2$ [261]. Binary mixtures of patchy particles are useful to study the phase behaviour of systems with and effective fractional M_p. The appearance of the so–called empty liquids is also interesting, namely states with an arbitrarily small occupied volume at temperatures below the liquid–gas critical temperature [231]. A typical gel structure which percolates the whole system is represented in Fig 4.4(b).

In real cells, proteins can behave as patchy particles to many extents [235, 265]: HP1 domains, for instance, can interact with nucleosomes through the CD, while hinges bind the NTE of other HP1. Therefore, in terms of HP1–HP1 interactions, this protein can behave as a patchy particle with $M_p = 4$, provided a strong short–range interaction that prevents multiple NTE for interacting with a single hinge (and *vice versa*). A very useful collection of the methods and simulations on patchy particles is given in [266].

4.3.2 BIA in protein–inspired patchy particles

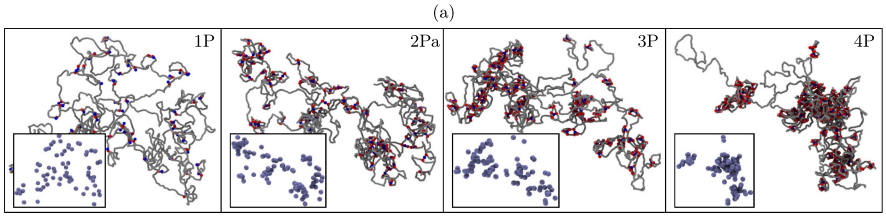

Figure 4.6. The BIA for model proteins with different valence. Results from simulations of $N = 100$ proteins interacting with an $L = 2000$ bead polymer were performed in a square box of size 90σ. Figure adapted from [235].

In Ref. [235] a study of chromatin–binding 'patchy' protein phase separation is presented; this work focuses on the multivalence protein ability of bridging distal regions of chromatin. In particular, diverse minimal geometries have been considered, as shown in Fig. 4.5(a), obtained by changing the number of binding sites (e.g. red beads) and the angle between them.

Thus, the phase behaviour of such patchy proteins was investigated in the absence of protein self–interaction, in order to probe the bridging ability of these model proteins. First, the author explored the ability to cluster and compact the polymer by varying the number of (equidistantly spaced) patches, i.e 1P, 2Pa, 3P, 4P. The snapshots in Fig. 4.6 show qualitatively the conformation of a 2000–bead long polymer which interacts with 100 proteins of different type. Note that for the case of 1P proteins, no clusters are detected, and the polymer remains swallen. In the other cases, proteins organise in clusters, which tend to grow faster as the interaction energy increases. Namely, the cooperative effect of BIA is stronger if the interaction energy and the valence are large.

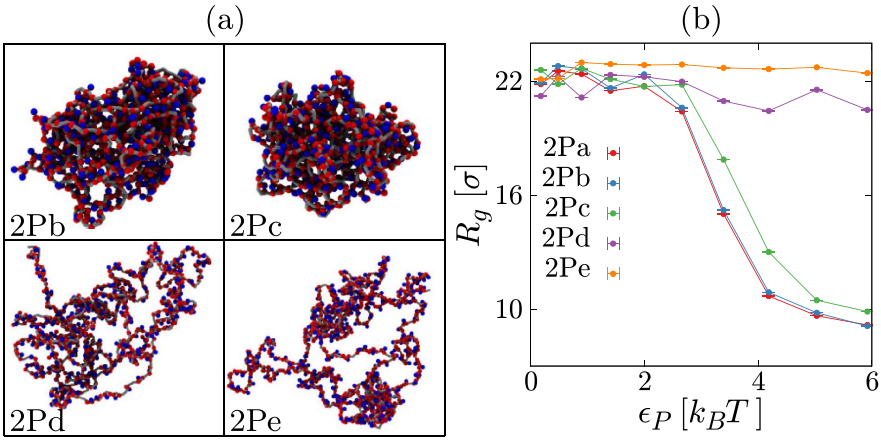

Figure 4.7. BIA for valence–2 model proteins with different shape. Polymer compaction by valence-2 model proteins ($N = 500$ proteins interacting with an $L = 1000$ bead polymer in a cube of size 70σ). (a) Snapshots showing typical equilibrium configurations in simulations (only proteins which are bound to the polymer are shown). (b) The radius of gyration is plotted as a function of the interaction energy for each model protein. Figure adapted from [235].

Second, the author compared the different phase behaviour of 2P proteins, by varying the angle between the two patches. Mainly, based on the particular model used, the cooperativity of BIA depends on their ability to bridge distant regions of chromatin. Reasonably, proteins with two patches on opposite sides (2Pa) display the larger tendency to cluster, while proteins with two very close patches (2Pe) tend to 'coat' the polymer: the BIA is not in effect. The author identifies three different binding modes for valence–2 proteins: the 'dangling', 'coating' and 'bridging' modes. A dangling protein is bound to polymer via a single patch, a coating protein is bound with both patches to adjacent polymer beads, a bridging protein is bound with both patches to distal polymer beads, favouring plymer compaction, as displayed in Fig. 4.7(a). Conversely, for 2Pd and 2Pe proteins the polymer never compacts, see Fig. 4.7(b).

The work in Ref. [235] is quite instructive to understand the structural properties of aggregates of real proteins, depending on the valence and the internal protein geometry.

4.4 BIPS and LLPS interplay of HP1–inspired model proteins

Here, we consider simple coarse-grained model proteins which resemble HP1 in solution with a model chromatin fibre. Our aim is to explore the parameter

space of the model, to understand under what conditions aggregates containing both proteins and chromatin form, and to measure their structural and dynamical properties. Importantly, our model is simple enough to allow simulations of large systems at many different points in the parameter space, but retain details of the domain structure of the protein (explicitly incorporating protein-protein and protein-DNA interaction domains). Specifically, we consider two separate models which are related to two microscopic possibilities: (i) the low-complexity domains give rise to weak and effectively longer ranged multivalent protein-protein attractions; (ii) the interactions between flexible domains are short ranged and have a limited valence such that exactly two domains can interact at a time. The former could mimic, for example, a scenario where these flexible domains adopt an extended coil configuration, meaning multiple coils can overlap with multiple weakly interacting contact points; we model such "multivalent" interactions using a longer range interaction potential between the spheres representing the hinge and NTE, such that several NTEs can simultaneously interact with a hinge and *vice versa* (determined by the geometry and steric hindrance). The latter case may arise, instead, when a disordered protein domain forms a globular secondary structure when interacting with the correct binding partner [267]. In this case, we consider limited valence interactions, using a shorter ranged potential such that at most one hinge and one NTE can interact at a time. Since an HP1 dimer has two hinges and two NTEs, in the "limited valence" model, a given dimer can bind to at most four others at once. We should note that, though our model is inspired by HP1, due to its simplicity, we expect our results to be applicable more widely.

4.4.1 Simulation scheme

Each HP1 dimer is represented by a rigid body consisting of seven spheres arranged as shown in Fig. 4.8: the blue sphere represents the CSD from each HP1 making up the dimer (i.e., two CSDs), then the black, green and orange spheres represent the hinge, CD and NTE domains respectively, two of each per dimer. Our coarse–grained approach does not attempt to model the full details and exact dimensions of the dimer; nevertheless, we aim to capture the main features of the physics at the mesoscale. For simplicity, all HP1 component spheres have a diameter 0.5σ, which gives a rough size of $1-1.5\sigma \approx 10-15$nm for a dimer (compared with $13-22$ nm for a real HP1 dimer, depending on post-translational modifications [227]). The CSD and CD, which are globular domains, are estimated to have a diameter of about 3 nm in reality [268]; the sizes of the flexible hinge and NTE domains are more difficult to estimate, but one might expect them to be larger than the folded domains. Choosing to

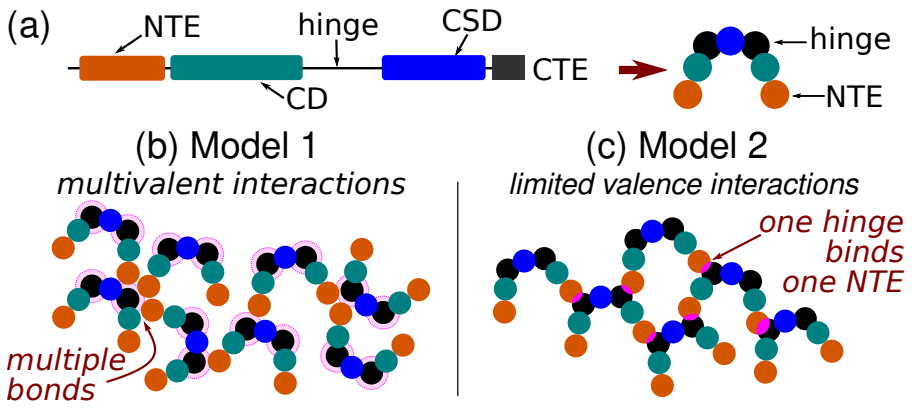

Figure 4.8. A simple coarse-grained protein model inspired by HP1. (a) Left: schematic showing the domain structure of the HP1 protein as detailed in the text. Right: schematic representation of the model HP1 dimer. (b) and (c) Two alternative models for interactions between HP1 dimers.

represent the domains as spheres of the same size is therefore an approximation, but we do not expect this will qualitatively change the observed behaviour.

Interactions between HP1 component beads and chromatin beads are modelled as follows. CDs interact attractively with chromatin beads through the potential

$$U_{\mathrm{CD-C}}(r) = \begin{cases} \epsilon_{\mathrm{HC}} \left[(\mathrm{e}^{-2\alpha r} - 2\mathrm{e}^{-\alpha r}) \right. \\ \qquad \left. -(\mathrm{e}^{-2\alpha r_{\mathrm{HC}}} - 2\mathrm{e}^{-\alpha r_{\mathrm{HC}}}) \right] & r \leq r_{\mathrm{HC}}, \\ 0 & \text{otherwise}, \end{cases} \tag{4.1}$$

where r is the separation between the centers of the CD and the chromatin bead, ϵ_{HC} is the energy which determines the strength of the interaction, α is a "shape" parameter, and r_{HC} is the cutoff distance which sets the range of the interaction. We set $\alpha = 5$ and $r_{\mathrm{HC}} = 0.9\sigma$. The CSD, hinge, and NTE interact sterically with the chromatin beads, through the WCA potential given in Eq. (3.2); we set $d_{ij} = 0.75$ for the CSD, and $d_{ij} = 0.5$ for the hinge and NTE. Firstly, as our model CSD effectively represent *two* dimerized CSD, therefore it is reasonable to model a larger steric CSD-chromatin interaction; secondly, as the hinge and the NTE are flexible/disordered, they can easily adapt their shape to allow CD and chromatin to interact, which results in an effective smaller radius of our modelled beads. In practice, this allows partial overlap of these beads with the chromatin beads, which is essential to permit the CD and chromatin beads to interact.

Interactions between HP1s are modelled similarly. The hinge and NTE domains in different HP1 dimers interact via the potential

$$U_{\mathrm{h-NTE}}(r) = \begin{cases} \epsilon_{\mathrm{HH}} \left[(\mathrm{e}^{-2\alpha r} - 2\mathrm{e}^{-\alpha r}) \right. \\ \qquad \left. -(\mathrm{e}^{-2\alpha r_{\mathrm{HH}}} - 2\mathrm{e}^{-\alpha r_{\mathrm{HH}}}) \right] & r \leq r_{\mathrm{HH}} \\ 0 & \text{otherwise.} \end{cases} \tag{4.2}$$

The two different models of the HP1–HP1 interaction are specified by different sets of parameters: for the *multivalent* model we set $\alpha = 0.5$ and $r_{\mathrm{HH}} = 1.3$; for the *limited valence* model we set $\alpha = 5$ and $r_{\mathrm{HH}} = 0.6$. All other HP1 component beads interact sterically via the WCA, except for the CSD and the NTE which can overlap (there is no interaction) in order to permit the correct binding of NTE domains with hinges. The functional forms of $U_{\mathrm{CD-C}}(r)$ and $U_{\mathrm{h-NTE}}(r)$ are the same as the commonly used Morse potential.

We note that the minima of the functions $U_{\mathrm{CD-C}}$ and $U_{\mathrm{h-NTE}}$ is at $r = 0$, i.e., the beads can overlap. In practice, it is not always possible to achieve the separation $r = 0$ due to steric interactions between the other component beads. For the HP1-chromatin interaction, allowing the CD and chromatin beads to overlap, along with the short range of the interaction, ensures that a given CD can interact with at most one chromatin bead at a time (e.g., a single CD cannot form a bridge between two chromatin beads). Since the CD spheres are smaller than the polymer beads, it is possible for more than one CD to interact with a given polymer bead at the same time. These are reasonable choices, since a chromatin bead represents several nucleosomes, and the CD is thought to interact with nucleosome surface charges. We also note that the functional forms of $U_{\mathrm{CD-C}}(r)$ and $U_{\mathrm{h-NTE}}(r)$ are the same as the commonly used Morse potential. In the limited valence case, the short range of $U_{\mathrm{h-NTE}}(r)$ with no repulsive core means that one NTE can interact with exactly one CSE at a time and vice versa. For the multivalent case the longer range and small value of α (which leads to a broader shape) is such that multiple NTEs can simultaneously interact with one CSE and vice versa.

The values of α have been chosen arbitrarily to model differently the two HP1 models, as they cannot be extracted from experimental evidence. The parameter α effectively models the steepness (i.e. the strength of the attraction forces) of the attractive basin in the potential in (4.2); The larger α, the stronger are the attractive forces, producing a more 'selective' binding. Similarly, we choose the values of r_{HH} and r_{HC}.

The dynamics of the polymer beads and HP1s (rigid body translation and rotation) are governed by Langevin dynamics in Eq. (3.5); we perform extensive simulations using the LAMMPS molecular dynamics software [207].

Below we present simulations of a system containing $N = 1000$ model HP1s

and an $L = 1000$ bead polymer. For simplicity, we consider a homogeneous polymer where all beads can bind HP1 (i.e., it represents a section of H3K9me3 modified chromatin). For the multivalent HP1 model we ran long simulations to obtain equilibrium configurations. For the limited valence HP1s the system displayed long-lived non-equilibrium metastable (gel–like) states. We confine all components of the system in a cubic box of size $l_x = 35\sigma$ (approximately equal to the radius of gyration of the polymer as predicted by the worm-like chain model). While the confinement reduces the entropy of the system by forbidding some extended polymer configurations, it also prevents the polymer from interacting with its periodicity including a "wall potential", i.e. an effective repulsive interaction located at $x, y, z = \pm 17.5\sigma$ (with our simulation box centered on $(x, y, z) = (0, 0, 0)$). These parameters represent a dilute regime. While the confinement reduces the entropy of the system by forbidding some extended polymer configurations, it also prevents the polymer from interacting with its periodic image (test simulations showed that, in the presence of periodic boundaries, HP1 could bridge distant chromatin regions across the boundaries, and the polymer could become trapped in extended unphysical configurations).

To map the density in our simulation with real densities, we suggest a possible mapping of the polymer beads to 1kbp of chromatin with bead diameter $10 - 20$nm. This leads to a HP1 concentration of $\sim 4 - 39$mMolar. We would point out, however, that within most of the regimes studied, the behaviour is independent of density. Although this density is larger than typical densities measured in cells, here it is the ratio of the number of proteins to the length of polymer (or number of available binding sites on the polymer) which is important, whilst our results are chiefly independent of the density chosen.

4.4.2 Model 1: Multivalent protein-protein interactions

With this version of the model, the proteins behave like a standard phase separating system [64, 176], at least when the HP1-chromatin interaction energy ϵ_{HC} is small. When ϵ_{HC} is larger there is more interesting behaviour. We summarise the emerging regimes in the simulation snapshots in Fig. 4.9 and in the phase diagrams in Fig. 4.10.

When $\epsilon_{\mathrm{HC}} < 12 k_B T$, a phase transition between a uniform mixed phase and a separated phase takes place as ϵ_{HH} increases. Above a critical value a roughly spherical cluster, or "droplet", of HP1 forms. Dynamically, after quenching from a small to large value of ϵ_{HH}, many clusters form and these grow/coalesce until a single droplet remains at equilibrium. We call this the *dense droplet* regime. By measuring the density ρ of HP1s inside and outside of

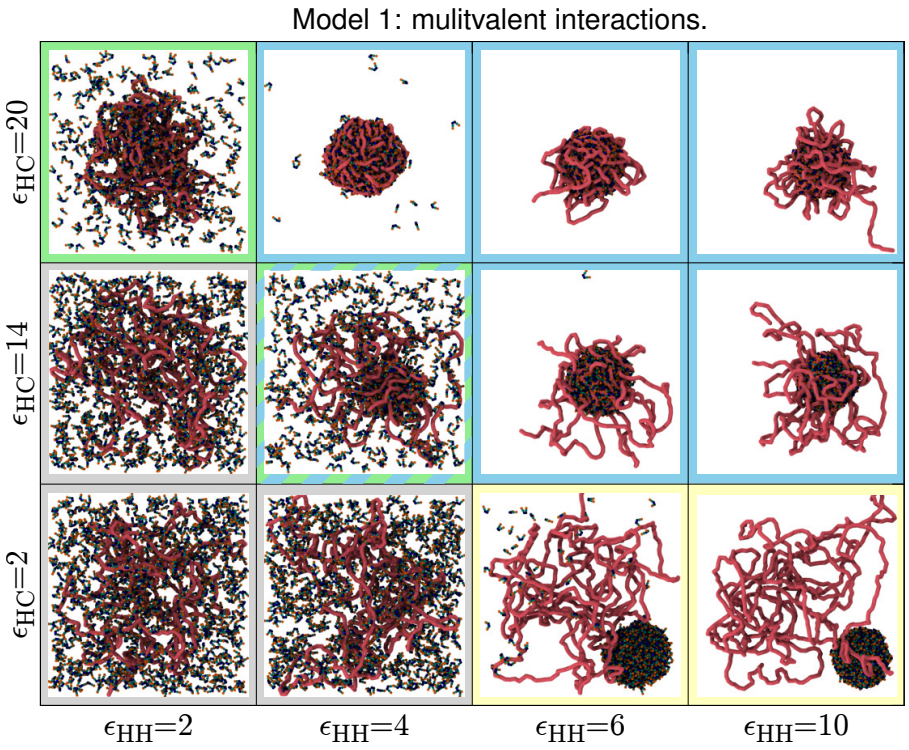

Figure 4.9. Protein-chromatin and multivalent protein-protein interactions lead to phase separation. Snapshots of equilibrium configurations for $N = 1000$ multivalent HP1 dimers interacting with an $L = 1000$ bead polymer representing a 1 Mbp chromosome fragment for different values of the HP1–HP1 and HP1–chromatin interaction energy.

the droplet (the procedure explained below in section 4.4.3), we can also map out a sketched phase diagram on the ρ–ϵ_{HH} plane [Fig. 4.10(b) top].

For small HP1-HP1 interaction energies, $\epsilon_{HH} < 5k_BT$, there is no droplet. The HP1-chromatin attraction leads to HP1s becoming bound to the polymer, and there is a smooth increase of the fraction bound as ϵ_{HC} increases. For large ϵ_{HC} there are sufficient HP1s bound such that the region occupied by the chromatin has a higher than average protein density, while the surroundings have a lower than average protein density [green region in Figs. 4.10(a)]. In this sense there is a phase separation, however this regime is profoundly different from the dense droplet phase. Here, a significant fraction of the proteins remain unbound, while the remaining HP1s tend to "coat" the polymer. Hence, we refer to it as the *coating* regime.

When both ϵ_{HH} and ϵ_{HC} are large [blue region in Figs. 4.10(a)] a protein droplet forms, but now the polymer is also absorbed into it. Or in other words, the droplet compacts the polymer. We call this the *absorbing droplet* regime. Interestingly, the polymer is absorbed to a different degree depending on the

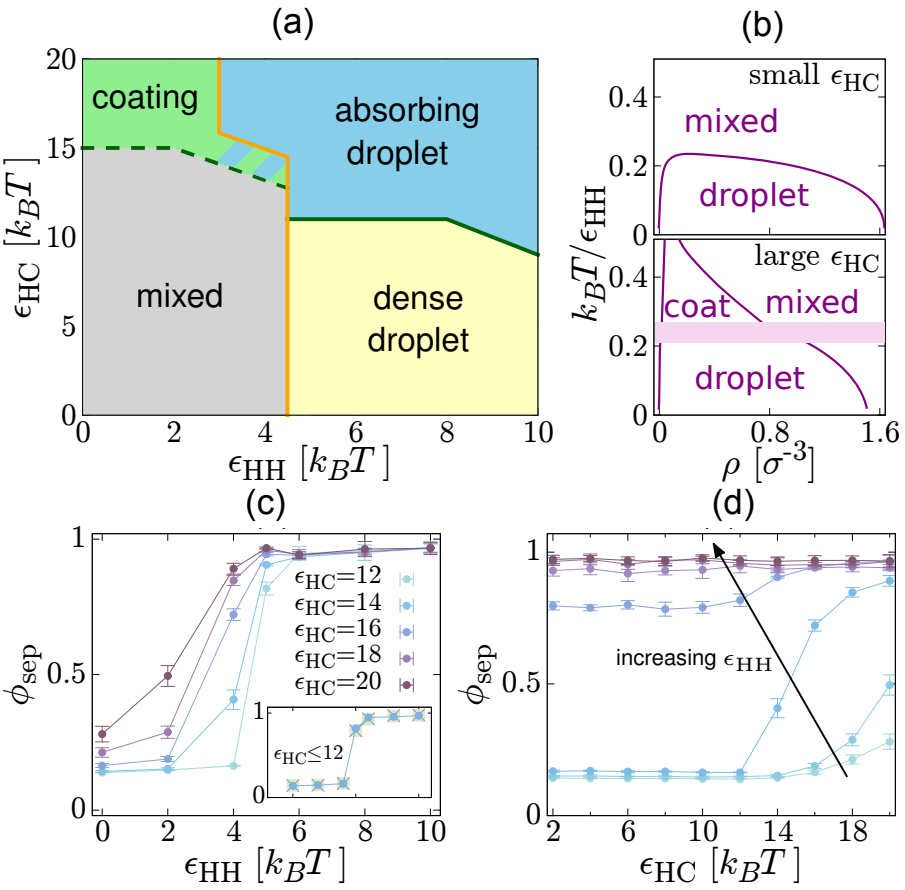

Figure 4.10. Phase diagrams and separation depth. (a) Phase diagram showing the different behaviours of the multivalent HP1s with different parameter values. Border colours in Fig. 4.9 indicate the relevant region in (a). (b) Phase diagram on the ρ-ϵ_{HH} plane for small ϵ_{HC} (top) and large ϵ_{HC} (bottom). ρ is the overall number density of HP1s. In the bottom plot the shaded bar covers a region where there is a chromatin associated droplet, but the protein density inside and outside the droplet depends on the overall protein density (see text). These are sketch plots based on measurements of the HP1 density inside and outside of droplet. (c) Phase separation depth ϕ_{sep} is plotted as a function of ϵ_{HH} for different values of ϵ_{HC} as indicated (units are k_BT). Each point is obtained from an average over 4 simulations of duration $5\times10^3\tau$ (τ is the simulation time unit). Error bars show standard error in the mean; lines are a guide to the eye. The inset shows a similar plot for ϵ_{HC} =2,6,10 and $12k_BT$ where points overlap. (d) ϕ_{sep} is plotted as a function of ϵ_{HC} for different values of ϵ_{HH} between 0 and $10k_BT$ (darker colours for larger values, as indicated by the arrow and increasing in steps of $2k_BT$).

precise values of the interaction energies [compare snapshots at $\epsilon_{HC} = 20k_BT$ and different ϵ_{HH} in Fig. 4.9, where different amounts of chromatin extend out from the droplet]. As before, measurements of HP1 density inside and outside of the droplet allow the construction of the ρ–ϵ_{HH} phase diagram for large ϵ_{HC}, on which we can also identify the coating regime [Fig. 4.10(b) bottom]. Briefly, at large HP1 concentration and small ϵ_{HH}, the non-bound HP1 density

reaches the same value of the bound HP1 density, thus the system is in a mixed phase; when HP1 concentration decreases, HP1s prefer to co-localise with chromatin, giving rise to a region with larger density along chromatin: this is the coating regime; increasing ϵ_{HH}, HP1 can phase separate on their own, forming a protein-chromatin droplet. Moreover, there is a further new region on this phase diagram (the shaded stripe) where a droplet forms only due to HP1–chromatin interactions (i.e., ϵ_{HH} is not large enough for a protein droplet to form on its own). We discuss this in more detail below.

Separation depth

To characterise these regimes more quantitatively [and to determine the positions of the lines in Fig. 4.9], we measure the local protein density by splitting the simulation box into N_{sb} sub–boxes of volume V_{sb}. If there are N_i HP1s in the ith sub–box, the local density is $\rho_i = N_i/V_{\mathrm{sb}}$. To quantify the level of phase separation we then consider a 'separation depth' parameter [269] defined as

$$\phi_{\mathrm{sep}} = \frac{1}{N_{\mathrm{sb}}} \sum_{i=1}^{N_{\mathrm{sb}}} \frac{\rho_i - \rho}{\rho^* - \rho}, \tag{4.3}$$

where $\rho = N/l_x^3$ is the overall number density of HP1s, and ρ^* is a reference density which takes the value ρ_+ when $\rho_i > \rho_+/2$ and $\rho_- = 0$ otherwise. This measures the mean local deviation from the uniform overall number density ρ, normalized by the expected deviation for a strongly phase separated system. We use $\rho_+ = 0.5$ and $N_{\mathrm{sb}} = 125$, chosen so as to be optimal for distinguishing the different regimes, and leading to $\phi_{sep} \to 1$ on droplet formation. Figure 4.10(c) shows how ϕ_{sep} varies with ϵ_{HH}, for different values of ϵ_{HC}. For $\epsilon_{\mathrm{HC}} \leq 12 k_B T$ the points sit on top of each other [Fig. 4.10(c) inset], and we observe a sharp crossover (at $\epsilon_{\mathrm{HH}} \approx 4.5 k_B T$) from $\phi_{\mathrm{sep}} \sim 0.15$ in the mixed phase to $\phi_{\mathrm{sep}} \sim 1$ in the dense droplet phase. As noted above, for these values of the energy the model behaves qualitatively the same as, e.g., interacting Brownian colloids [232], and we expect a first-order phase transition in the thermodynamic limit (Model B). We use a value of $\phi_{\mathrm{sep}} = 0.5$ to set the position of the orange line in Fig. 4.10(a). As ϵ_{HC} increases, this line shifts to the left—we discuss this interesting regime further below. Figure 4.10(d) shows that for small ϵ_{HH} the separation depth is independent of ϵ_{HC} throughout the uniform phase ($\phi_{\mathrm{sep}} \sim 0.15$), before increasing at larger ϵ_{HC} in the coating or absorbing droplet regimes. We identify the value of ϵ_{HC} at which ϕ_{sep} starts to increase ($\phi_{\mathrm{sep}} \gtrsim 0.15$) to draw the green dashed line. For $\epsilon_{\mathrm{HH}} > 6 k_B T$ the separation depth $\phi_{\mathrm{sep}} \sim 1$, independently of ϵ_{HC}; i.e., this parameter cannot differentiate between droplets and absorbing droplets.

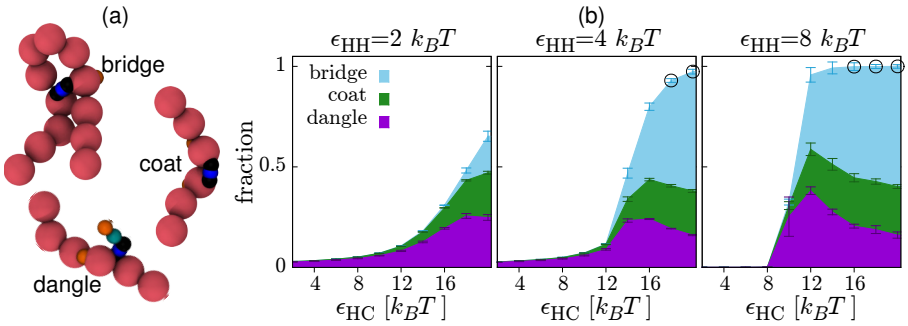

Figure 4.11. Protein-chromatin binding modes. (a) Simulation snapshots of a single model HP1 and a short section of chromatin. The three different binding modes are depicted. (b) Plots showing the fraction of the $N = 1000$ proteins bound to the chromatin in each mode for different interaction energies. The height of each coloured region indicates the proportion of proteins, with the regions stacked on top of each other. In this way the height of the total coloured region indicates the fraction of proteins bound in any mode f_{tot}. Values are obtained from averaging over 4 simulations of duration $5 \times 10^3 \tau$, and error bars show the standard error in the mean. Black circles around points indicate where bridging is the dominant binding mode (more than half of the bound proteins).

Binding modes

We now consider the nature of the interactions between the HP1 dimers and the chromatin. Similarly to 'patchy' model proteins in section 4.3.2, each model HP1 can bind the polymer in three different modes [Fig. 4.11(a)]. In the "dangling" mode, HP1s bind to chromatin through one CD; when both CDs bind to chromatin at adjacent $|i - j| < 2$ polymer beads, the HP1 binds in the "coating" mode; f inally, if the CDs interact with polymer beads which are separated along the chain ($|i - j| \geq 2$), then the protein is "bridging". As detailed in section 4.3.2, the shape of the protein determines its likelihood to bind in each mode: bridging incurs an entropic penalty (due to polymer looping), so unless the shape of the protein specifically disfavours coating, the coating mode is favourable. This is the case here: in the absence of protein–protein interactions we mainly observe coating.

In Fig. 4.11(b) we plot the fraction of bridging, coating and dangling proteins as a function of ϵ_{HC}. If we consider the total fraction of proteins bound to the polymer f_{tot}, at $\epsilon_{HH} = 2k_BT$, f_{tot} increases smoothly with ϵ_{HC}. Coating and dangling are the dominant binding modes; the BIA is therefore not in effect, and we do not observe BIPS or chromatin compaction. At large ϵ_{HH}, where there is a droplet, f_{tot} increases very sharply as ϵ_{HC} is increased and the polymer becomes absorbed into the droplet [the curve becomes steeper from left to right in the panels of Fig. 4.11(b)]. This could indicate the presence of a first–order phase transition in the thermodynamic limit. Within the absorbing droplet regime we also observe that the fraction of bridging proteins increases with

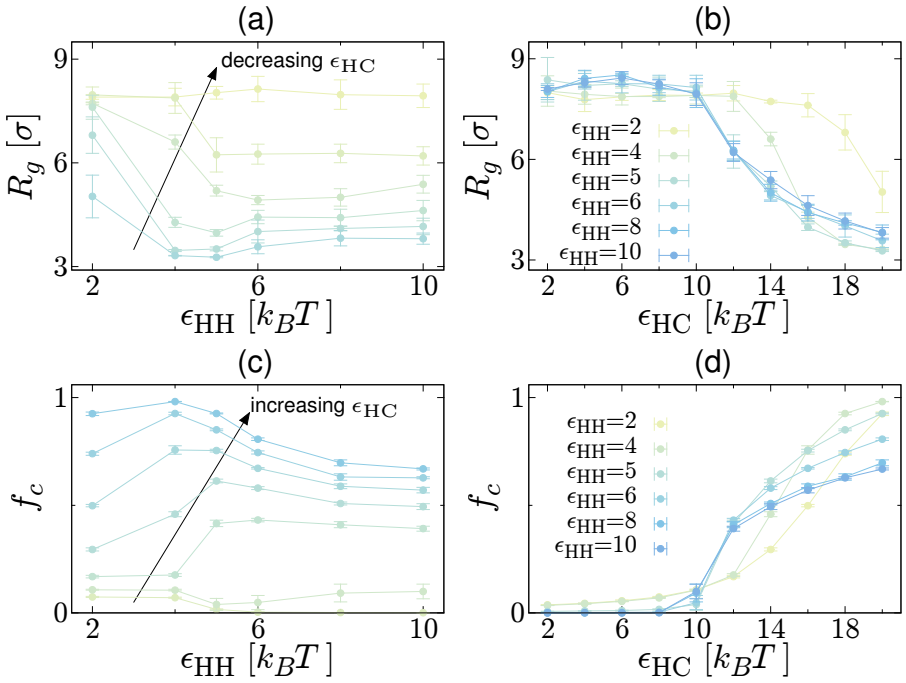

Figure 4.12. HP1-chromatin interactions, chromatin compaction, and droplet dynamics. (a-b) Plots showing how the radius of gyration of the polymer representing the chromatin segment depends on the interaction energies. In (a) from top to bottom curves are for ϵ_{HC} values between 10 and $20k_BT$ increasing in steps of $2k_BT$. Points are obtained from an average of 4 independent simulations; error bars show the standard error in the mean, and connecting lines are a guide to the eye. (c-d) Plots showing how the fraction of chromatin beads which are bound by proteins f_c, depends on the interaction energies. In (c) from bottom to top curves are for ϵ_{HC} values between 8 and $20k_BT$, increasing in steps of $2k_BT$.

ϵ_{HC}, and it becomes the dominant mode of binding when both interactions are strong. The main driver of this is that as ϵ_{HC} increases, more of the polymer becomes absorbed inside the droplet, and so the likelihood of two distant regions being close enough together for bridges to form increases.

Radius of gyration and number of chromatin beads interacting with HP1

One proposed function of HP1 *in vivo* is to compact heterochromatin. The ability of our model proteins to compact chromatin can be probed by measuring its radius of gyration, defined as

$$R_g^2 = \frac{1}{L}\sum_{i=1}^{L}(\mathbf{r}_i - \bar{\mathbf{r}})^2, \tag{4.4}$$

where $\mathbf{r}_i$ is the position of the ith chromatin bead, and $\bar{\mathbf{r}} = (1/L)\sum_i \mathbf{r}_i$.
Figs. 4.12(a–b) show how R_g depends on the interaction energies. Interestingly,

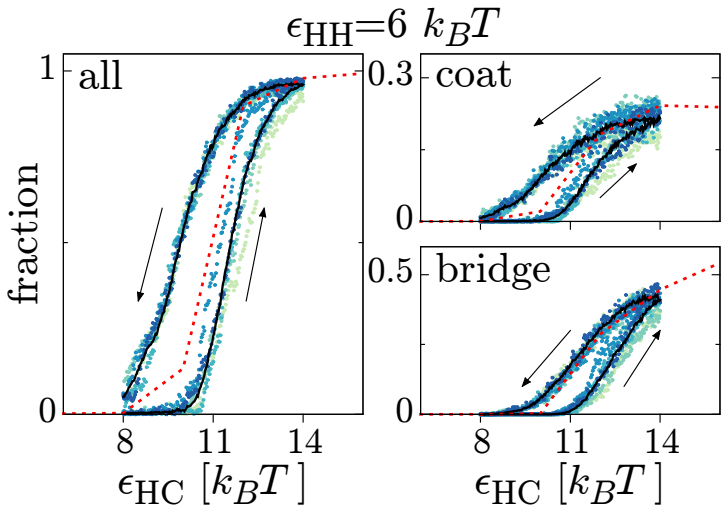

Figure 4.13. Hysteresis loops. (a) Snapshots from simulations with $\epsilon_{HH} = 4k_BT$ and $\epsilon_{HC} = 20k_BT$ but with different numbers of proteins N as indicated. (b) Plot showing how the protein densities within the high and low density phases (inside and outside the droplet) vary with the number of proteins. (c) Plot showing how the radius of gyration of the polymer R_g and radius of the droplet R_d vary with N, shown in log-log scale. The black line has a slope 1/3, which is how the droplet radius would scale in a standard phase separating system.

R_g can vary non–monotonically as ϵ_{HH} increases; similar behaviour is observed in the fraction of polymer beads bound by proteins, f_c [Fig. 4.12(c,d)]. The reason for this non–monotonicity is strikingly apparent in the top row of snapshots in Fig. 4.9: in the leftmost snapshot the polymer is swollen, in the second from the left it is fully absorbed into a protein droplet (small R_g and large f_c), but in the two right–hand snapshots the polymer is only partially absorbed into the droplet (R_g increases again, while f_c decreases). That the amount of absorbed polymer varies so widely within the absorbing droplet regime is likely due to competition between different contributions to the free energy. While HP1–chromatin binding represents a reduction in free energy, this is offset by the reduction in entropy due to the compaction/confinement of the polymer within the droplet. Increasing ϵ_{HC} increases the amount of chromatin absorbed as the entropic loss is overcome. On the other hand, the presence of the polymer within a droplet will reduce the number of HP1–HP1 interactions due to steric effects; so increasing ϵ_{HH} *decreases* the amount of chromatin absorbed (effectively the polymer is 'squeezed out' of the droplet).

Finally in this section, we consider intermediate values of the HP1–HP1 interaction strength, $\epsilon_{HH} \approx 4k_BT$, where we observe the most interesting behaviour. Here, in the absence of chromatin interactions there is no droplet formation and ϕ_{sep} is small. However, we note that as ϵ_{HC} increases, the orange

line in Fig. 4.10(a) moves to the left, so droplets *can* form at $\epsilon_{\mathrm{HH}} \approx 4k_BT$ *if* the protein-chromatin interaction energy is large enough. In other words, HP1-chromatin attraction promotes protein aggregation. This can be understood as follows: when ϵ_{HC} is large enough, a significant number of HP1s become localised to the polymer and these tend to bind in the coating mode. Then, intermediate HP1–HP1 interactions are sufficient to allow extended chromatin–HP1–HP1–chromatin bridges to form. The BIA is therefore in effect, leading to chromatin compaction and protein clustering; we note that this is the *only* region of the phase diagram where the BIA is in effect and a true BIPS is observed. When both ϵ_{HH} and ϵ_{HC} have intermediate values we observe an absorbing (BIPS) protein droplet *and* coating of the chromatin which emerges from the droplet [cross–hatch shaded region in Fig. 4.10(a)].

Hysteresis loops

In Fig. 4.11(b), we showed that there is an abrupt change in quantities such as the total fraction of proteins bound to the polymer f_{tot}, the fraction of polymer beads bound by proteins f_c, and the polymer radius of gyration R_g, as ϵ_{HC} is increased and the system moves from the droplet to the absorbing droplet regime. This hints that there may be a first order phase transition in the thermodynamic limit. To elucidate this further, we performed simulations where we slowly vary the parameter values in time and looked for evidence of hysteresis.

In Fig. 4.13, we show a hysteresis loop for the fraction of proteins bound to the polymer (in total and in different modes) as ϵ_{HC} is slowly increased from $8k_BT$ to $14k_BT$ before being decreased again. we ran 12 independent repeat simulations for $4\times10^4\tau$, each starting from a different equilibrium configuration for $\epsilon_{\mathrm{HH}} = 6k_BT$ and $\epsilon_{\mathrm{HC}} = 8k_BT$ (droplet regime). For the first $2 \times 10^4\tau$ of each simulation ϵ_{HC} is increased by an increment of $3\times10^{-2}k_BT$ every $10^2\tau$, until it reaches $\epsilon_{\mathrm{HC}} = 14k_BT$ (the absorbing droplet regime). Then, over the second $2\times10^4\tau$ of the simulation ϵ_{HC} is reduced in the same fashion (until $\epsilon_{\mathrm{HC}} = 8k_BT$). We keep the protein–protein interaction energy constant throughout at $\epsilon_{\mathrm{HH}} = 6k_BT$. the system displayed hysteresis as the polymer became absorbed and then re–emerged from the droplet. For this intermediate value the droplet is highly dynamic; one would expect a slower response for larger ϵ_{HH}. Similar results holds for R_g and the fraction of dangling proteins (not shown). Therefore, there is also hysteresis in terms of the polymer configuration, as it retains memory of its previous state for a significantly long time after the system crosses the transition.

These observations suggest that in the limit of a large droplet the system would show a first–order transition as ϵ_{HC} increases, to a phase where the

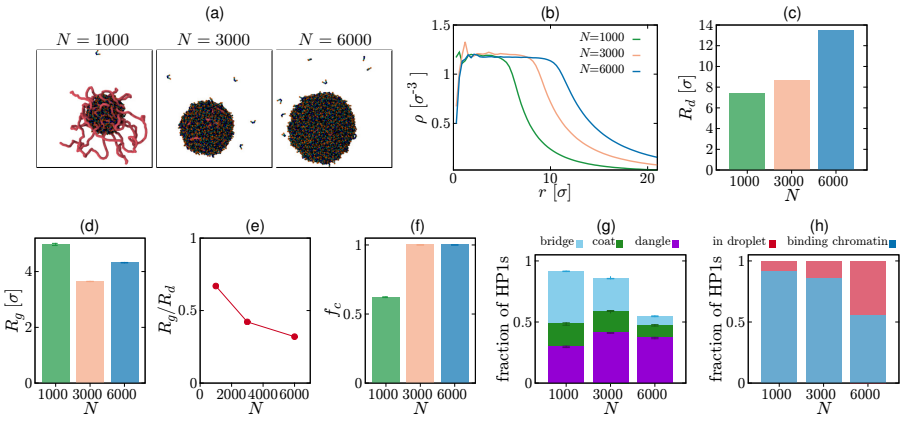

Figure 4.14. Varying the number of proteins for parameters where a droplet would also form without chromatin. Plots showing the effect of varying the number of proteins for parameter values $\epsilon_{\mathrm{HH}} = 6k_BT$, $\epsilon_{\mathrm{HC}} = 14k_BT$ within the absorbing droplet phase. (a) Snapshots from simulations with $N = 1000$, 3000 and 6000 proteins. (b) Density of proteins within a probing sphere of radius r centred on the the centre of mass of the droplet. (c) Bar plot showing the radius of the protein droplet in simulations with different values of N. (d) Bar plot showing the radius of gyration of the polymer. (e) Plot showing how the ratio R_g/R_d varies with N. (f) Bar plot showing the fraction of polymer beads bound by proteins. (g) Stacked bar plot showing the fraction of the total number of proteins which are bound to the polymer in each of the three modes. Bars are stacked on top of each other so, for example, the distance between the bottom and top of the green region gives the fraction of proteins bound in the coating mode. The total height shows the total fraction of proteins bound to the polymer. (h) Stacked bar plot showing the fraction of the total number of proteins which are in the droplet but not binding to chromatin beads (red) and in the droplet *and* binding to chromatin beads (blue).

polymer is fully absorbed; in our small system we instead observe an extended co–existence regime where the polymer is only partially absorbed (see also below).

4.4.3 Varying protein density

We now consider the effect of the overall protein density for the multivalent HP1s. As expected, for large ϵ_{HH}, we observed the same behaviour as a standard (Model–B) phase separation. Fig. 4.14 shows results for the case where $\epsilon_{\mathrm{HH}} = 6k_BT$ and $\epsilon_{\mathrm{HC}} = 14k_BT$; for these parameters the protein droplet would form even in the absence of chromatin. As can be observed from the snapshots in Fig. 4.14(a), increasing the number of proteins leads to a larger droplet which absorbs a larger fraction of the polymer. We confirm quantitatively that the density of HP1s within the droplet is independent of the total number of HP1s (i.e. the overall density). The procedure to calculate the density in Fig. 4.14(a)

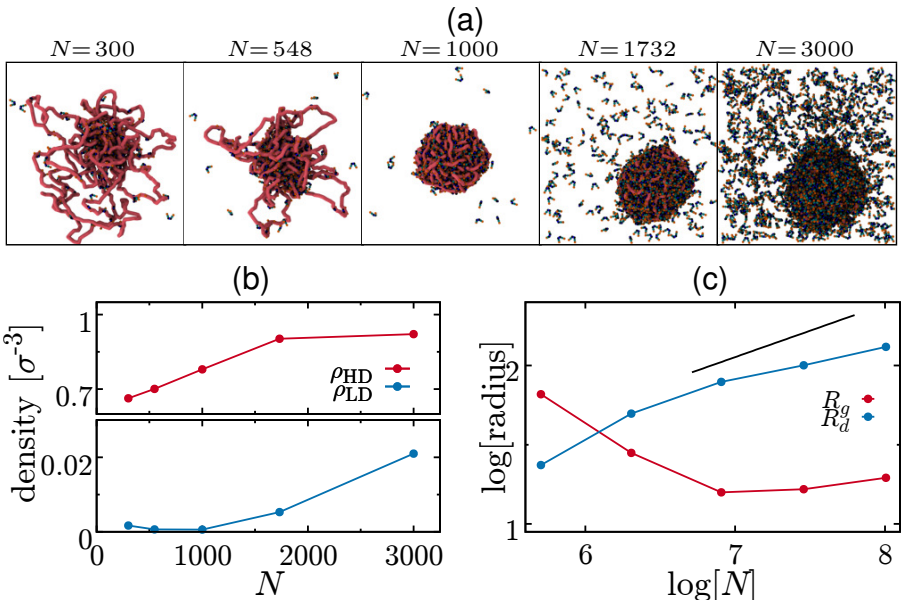

Figure 4.15. Varying the number of proteins for parameters where a droplet do not form without chromatin. (a) Snapshots from simulations with $\epsilon_{\mathrm{HH}} = 4k_BT$ and $\epsilon_{\mathrm{HC}} = 20k_BT$ but with different numbers of proteins N as indicated. (b) Plot showing how the protein densities within the high and low density phases (inside and outside the droplet) vary with the number of proteins. (c) Plot showing how the radius of gyration of the polymer R_g and radius of the droplet R_d vary with N, shown in log-log scale. The black line has a slope 1/3, which is how the droplet radius would scale in a standard phase separating system.

has been also used to produce the plots of separation depths in Fig. 4.10(a), and it reads as follows.

In order to calculate the protein densities, e.g., within or outside of a protein droplet, we consider a 'probe sphere' of radius r centred on the centre of mass of the largest droplet (protein cluster). We then progressively increase r, and calculate the density of proteins within the probe sphere, and within a spherical shell of width dr. We then average over time and repeat simulations (finding a new droplet centre of mass each time). The local protein density, is calculated as

$$\rho_{\mathrm{shell}}(r) = \frac{N_{\mathrm{shell}}(r)}{4\pi r^2 dr}, \tag{4.5}$$

where $N_{\mathrm{shell}}(r)$ is the number of proteins within the spherical shell of width $dr = 0.3\sigma$ and radius r.

We plot the density within the probe sphere ρ_{ps} as a function of its radius r in Fig. 4.14(b), while the droplet radius for the three different values of N is shown in Fig. 4.14(c) [taken to be the position of the half maximum point in the $\rho_{\mathrm{ps}}(r)$ curve]. From this, we see that when HP1–HP1 attractions drives droplet formation the behaviour is consistent with standard (Model–B) phase

separation (ρ_{HD} is independent of N and droplet radius R_d increases as $N^{1/3}$).

Fig. 4.14(d–h) reveal a complicated relationship between the droplet size and chromatin absorption/compaction. This stems from the balance between the energetic gain which arises from HP1s binding chromatin, the entropic loss due to HP1 bound chromatin being confined to the volume of the droplet, and any energetic loss due to HP1–HP1 bonds being broken to accommodate HP1–chromatin 'bonds'. For the $N = 6000$ case, the polymer is completely absorbed within the droplet, and we note that the volume which the polymer coil occupies is significantly smaller than the volume of the droplet. That is to say, the polymer is compacted to a greater extent that it would be due to simply being confined within the droplet. This is clear if one considers the ratio R_g/R_d, which steadily decreases with N [Fig. 4.14(e)]. To understand this, we varied the interaction strengths ϵ_{HH} and ϵ_{HC} by a small amount (such that we stay in the same regime) and observed the effect on the polymer radius of gyration (data not shown). We found that increasing ϵ_{HH} led to greater compaction of the polymer. This is consistent with expectations if we consider the protein droplet to be an effective solvent within which the polymer is dissolved (e.g., as considered in Flory–Huggins theory). On the other hand, increasing ϵ_{HC} also led to greater compaction of the polymer; in the Flory-Huggins theory, increasing polymer-solvent attraction leads to *swelling* of a polymer. Clearly the ability of our model HP1s to form bridges means that treating the droplet as a solvent gives an incomplete picture.

A strikingly different behaviour is observed for intermediate ϵ_{HH} (the region where the BIA is in effect, i.e, where a droplet only forms due to the presence of the polymer). Figure 4.12(e) shows snapshots for $\epsilon_{\mathrm{HH}} = 4k_BT$ and $\epsilon_{\mathrm{HC}} = 20k_BT$ with different numbers of proteins. It is immediately clear that the density of proteins within the two phases varies with N [see also Fig. 4.15(b)]. This can be rationalised as follows. For small N a protein droplet forms on the polymer via the BIA. This droplet is rather 'loose', and as N increases, more space within the droplet becomes filled with proteins and the density (ρ_{HD}) increases. At the same time more polymer becomes absorbed and the droplet grows [R_g decreases, and the droplet diameter R_d increases, Fig. 4.15(c)]. When $N \approx 1000$ all of the polymer is absorbed, and R_g reaches a minimum; as N and R_d increase further the polymer can swell slightly. At some point the droplet density reaches a maximum, and adding further proteins instead leads to an increase in the density of proteins outside the droplet. The droplet still grows with N, but more slowly than in a standard phase separation (where $R \sim N^{1/3}$).

In summary, for the narrow range of parameters where phase separation only occurs in the presence of the polymer, we find the surprising result that the density of the phases (ρ_{HD} and ρ_{LD}) depends on the overall protein density

[shaded band in Fig. 4.10(c)]. This has important implications for protein-chromatin interaction *in vivo*: for instance, this numeric results are in line with recent experimental evidence in mice, where HP1 foci density increases as more HP1 are expressed, rather than their size [270].

4.4.4 Model 2: Limited valence protein-protein interactions

In this model the HP1 dimer–dimer interactions have a limited valence, i.e., exactly one hinge domain can interact with exactly one NTE domain at a time, similarly to what happens with classic patchy particles. While in the previous section we specifically considered equilibrium configuration, here we study the metastable states obtained when the system is quenched by instantaneously switching on both protein–protein and protein–chromatin interactions. Specifically, we start from an equilibrium configuration for $\epsilon_{\mathrm{HH}}, \epsilon_{\mathrm{HC}} = 0$, switch on interactions and run for $10^4\tau$ (where τ is the simulation time unit); after this time the measured quantities (f_c, ϕ_{sep}, etc.) have stopped systematically varying. Steady state values of these quantities are then obtained by averaging over a further $10^4\tau$ simulation.

Typical snapshots are shown in Fig. 4.16(a). Similar behaviour is observed as for the multivalent interaction model. At low ϵ_{HH} we have the same mixed and coating regimes. For low ϵ_{HC}, as ϵ_{HH} increases, we go from the mixed phase to an aggregate phase. Unlike the multivalent model, here the aggregates are not spherical; instead multiple irregularly shaped clusters form. We also see small closed clusters of HP1s where all (or most) hinge and NTE domains are bonded When both ϵ_{HC} and ϵ_{HH} are large, many of the aggregates become associated with the polymer, which becomes compacted. Some smaller clusters remain detached from the polymer.

As before, we measure the separation depth ϕ_{sep} as a function of the two interaction energies. Figure 4.16(b) shows that the behaviour is again similar to the multivalent model in that ϕ_{sep} increases with ϵ_{HH}. However, the largest ϕ_{sep} values are smaller than in the multivalent case, consistent with several protein aggregates of different size forming, rather than a single phase separated droplet. There is also a regime where proteins aggregate only when the interaction with the chromatin is strong enough, though it is less clear than for the multivalent model. Specifically, for $\epsilon_{\mathrm{HH}} = 8k_BT$ there is a cluster only when ϵ_{HC} is large, but ϕ_{sep} only reaches intermediate values [Figs. 4.16(a) and (c)]. For $\epsilon_{\mathrm{HH}} = 12k_BT$, ϕ_{sep} has an intermediate value just less than 0.6 for a broad range of ϵ_{HC} values [Figs. 4.16(c)], behaviour which is not observed in the multivalent model. This arises because while clusters do form, there are many of them; they are also highly dynamic, continually forming, dissolving, merging and breaking apart [232].

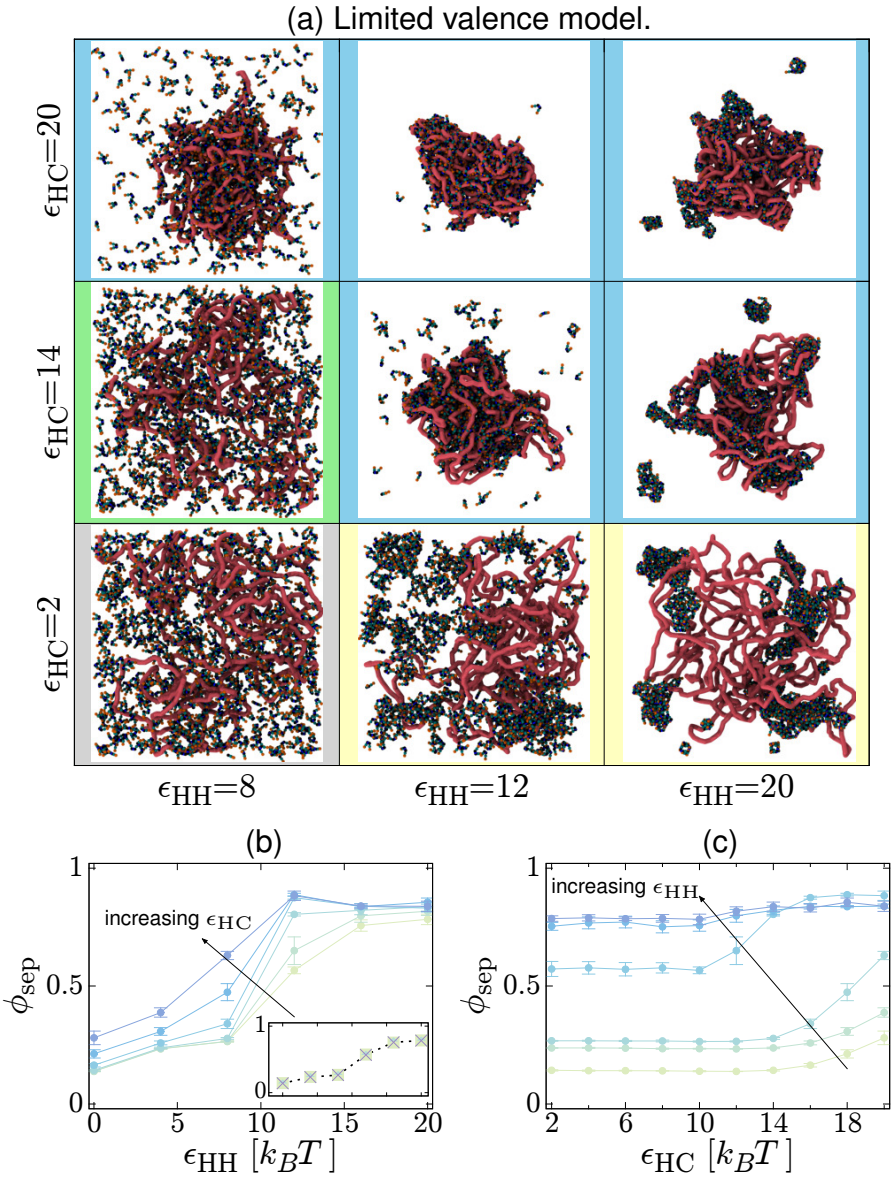

Figure 4.16. The limited valence HP1 displays similar behaviour regimes. (a) Snapshots are shown for simulations of the the limited valence HP1 model with different HP1-HP1 and HP1-chromatin interaction energies. Border colours indicate similarity to the different regimes observed for the multivalent model in Fig. 4.9. (b) Plot showing how the separation depth parameter varies with ϵ_{HH} for different values of ϵ_{HC} for the limited valence model. Data for ϵ_{HC} between 10 and 20k_BT increasing in steps of 2k_BT are shown in the main plot. The inset shows that points for $\epsilon_{HC} = 6, 8$, and 10k_BT sit on top of each other. (c) Similar plot showing ϕ_{sep} as a function of ϵ_{HC}. Curves are for different values of ϵ_{HH} between 0 and 20k_BT increasing from bottom to top in steps of 4k_BT.

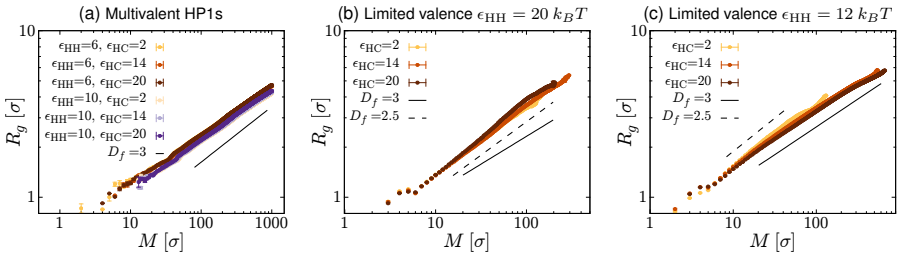

Figure 4.17. Alternative quench schemes lead to different long-lived metastable configurations for limited valence model proteins. Snapshots are shown from simulations of the limited valence model using parameters $\epsilon_{\mathrm{HH}} = 16k_BT$ and $_{\mathrm{HC}} = 16k_BT$. (a) Configuration obtained from the end of a $2 \cdot 10^4\tau_{\mathrm{LJ}}$ long simulation where for the first $10^4\tau_{\mathrm{LJ}}$ the HP1-HP1 attraction was switched off; the HP1-chromatin attraction was kept switched on for the full duration. (b) Configuration obtained from the end of a $2 \cdot 10^4\tau_{\mathrm{LJ}}$ simulation where for the first $10^4\tau_{\mathrm{LJ}}$ the HP1-chromatin attraction was switched off and instead the HP1-HP1 attraction was on for the full duration. (c) Configuration from a simulation of duration $4 \cdot 10^3\tau_{\mathrm{LJ}}$ with a higher density of HP1s in the absence of polymer ($N = 1000$ proteins in a smaller system of side $l_x = 21$) using periodic boundary conditions (instead of 'walls' as in all other simulations).

Clusters have fractal dimension

Measurements of clusters and sub–clusters show that these have a fractal dimension less than 3, as would be expected in a gel. In the snapshots of the limited valance model shown in Fig. 4.18 we observe irregularly shaped protein clusters with a structure strikingly different to the spherical droplets formed by the multivalent model. To quantify this difference, here we estimate the fractal dimension D_f of the clusters. In simulations, the fractal dimension of cluster, e.g. from a diffusion limited cluster aggregation (DLCA) process [271], is typically obtained from a scatter plot of the cluster mass (or number or particles) *versus* radius. For spherical clusters one would expect a scaling $R \sim N^{1/3}$, while fractal clusters give $R \sim N^{1/D_f}$ where $D_f < 3$ for a 3D system.

In our simulations, we typically observe a single or small number of clusters, meaning it is difficult to obtain enough measurements to determine D_f. Another common method is to extract the fractal dimension from the structure factor $S(q)$ [272], but this is again difficult to obtain from our simulations of a small number of clusters in a confined geometry. To estimate D_f, we instead consider sub–clusters, measuring their mass and radius of gyration. We use the following scheme: we consider the ith HP1 together with all of its bound neighbours (defined as any HP1 whose centre of mass is within 1.1σ of HP1 i; different threshold values do not significantly alter the result). We denote this set of proteins a level 1 sub–cluster, and record the number of proteins M and radius of gyration R_g associated with this set. This is repeated for all HP1s

in the system. We then consider level 2 sub–clusters, consisting of HP1 i, its bound neighbours, and all of the bound neighbours of neighbours; again, we record M and R_g for $i = 1 \dots N$. Level 3 sub–clusters include neighbours of neighbours of neighbours, etc. We continue increasing the level until there are no further unique sub-clusters, taking care not to double count. Fig. 4.18 shows plots of R_g against M for all possible sub–clusters; each point represents the mean R_g of all sub-clusters (of any level) with a given number of HP1s.

For the multivalent model, for all parameters where there are protein droplets, the sub–cluster plots are roughly linear on a log-log scale, and have similar slope. A linear fit to a function $R_g = aM^{1/D_f}$ gives $D_f \approx 3.2$, close to the expected $D_f = 3$ for spherical droplets.

For the limited valence model, we find that sub–cluster plots are not always linear over the whole curve, and the exponent depends on the parameters. For large $\epsilon_{\mathrm{HH}} = 20 k_B T$ the plots are roughly linear with fractal dimension $D_f \approx 2.5$, which is insensitive to the value of ϵ_{HC}. This is close to the value $D_f = 2$ observed in simulations of patchy particles [273]. For smaller $\epsilon_{\mathrm{HH}} = 12 k_B T$ there is not a single power law relationship between R_g and M, but for large clusters $D_f \approx 3$. The reason for this difference is likely due to the difference in the protein dynamics. For $\epsilon_{\mathrm{HH}} = 12 k_B T$, protein in clusters can dynamically rearrange to satisfy the maximum number of bonds, tending to adopt more space–filling shapes; at larger ϵ_{HH}, HP1–HP1 bonds persist for long times, leading to dynamically arrested fractal clusters.

Alternative quenches: closed loops and percolating gels

As noted above, the limited valence model behaves similarly to patchy particles in that the system can adopt long–lived metastable states with multiple fractal clusters (including "closed loops" where all bonds are satisfied). The observed structures therefore depend on the initial condition or the quenching procedure used. To highlight this, in Fig. 4.18(a–b) we show configurations obtained with two different quenches. In Fig. 4.18(a), after starting from an equilibrium configuration for $\epsilon_{\mathrm{HH}}, \epsilon_{\mathrm{HC}} = 0$, first the HP1–chromatin attraction was switched on, then later the HP1–HP1 attraction was switched on. This generated structures where most of the proteins were associated with the polymer, and were spread roughly uniformly along it; a few small (closed loop) clusters were not associated with the polymer. In Fig. 4.18(b), first the HP1–HP1 attraction was switched on, and then later HP1–chromatin attractions were switched on. In this case the proteins tend to sit in larger clumps associates with smaller sections of the polymer; there were much larger polymer regions without proteins bound. This latter morphology arises because the large HP1 clusters form first, only later becoming associated with the polymer. In all other

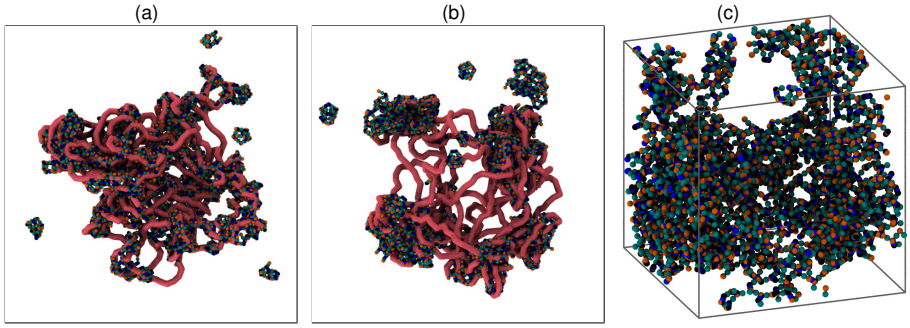

Figure 4.18. Fractal dimension of HP1 sub–clusters. Logarithmic scale plots showing radius of gyration R_g *vs.* the number of proteins M in sub–clusters for multivalent and limited valence simulations in droplet/cluster regimes. Each point shows the mean R_g of all unique sub–clusters with a given M; averages are also over time for a single simulation. Error bars show the standard error in the mean. We expect a power-law relationship with $R_g \sim M^{1/D_f}$, where D_f is the fractal dimension. (a) Points show data obtained from simulations of the multivalent model in the droplet or absorbing droplet regimes, with ϵ_{HH} and ϵ_{HC} as indicated (units are k_BT). Lines show the slope for the indicated values of D_f (b) Points show data obtained from simulations of the limited valence model with large $\epsilon_{\mathrm{HH}} = 20k_BT$ and ϵ_{HC} as indicated (clustering regimes). (c) Points show data from simulations of the limited valence model with smaller $\epsilon_{\mathrm{HH}} = 12k_BT$.

limited valence simulations in this work we switched on both interactions at the same time. To demonstrate that the limited valence HP1 can form a gel, we also performed a simulation with periodic boundaries and a smaller box size (higher HP1 density); a snapshot is shown in Fig. 4.18(c).

4.4.5 Discussion and future perspectives

In this chapter we have studied the behaviour of simple model proteins interacting with a bead–and–spring polymer model for chromatin. Differently from the setup in chapter 3, we consider a more complicated system, where rigid bodies composed of spheres represent different protein domains which interact attractively with each other or with chromatin. The domain structure was based on that of HP1, but our goal was to obtain insight on the interplay between LLPS and BIPS in general.

In this model we were capable of drawing a phase diagram, by varying protein–protein and protein–chromatin interactions; two different type of HP1–HP1 interaction were investigated: a 'multivalent' interaction (multiple HP1 can bind together) and a 'limited valence' interaction (only one HP1 domain can stick to another).

In the multivalent model different regimes were identified: we found that in the absence of protein–chromatin interactions, increasing the protein–protein

interaction strength ϵ_{HH} led to liquid droplet formation (Model B). Increasing protein-chromatin attractive interactions lead to a sharp crossover to a regime where the chromatin is absorbed into the droplet (with indications that there is a first–order phase transition in the thermodynamic limit). Importantly, the level of chromatin absorption depended on both interaction energies, and the number of proteins/size of the droplet. For most of the parameters studied, a significant fraction of the chromatin "looped out" from the droplet [the looping statistics of a similar situation have been studied in Ref. [274]]. This suggests that precise parameter tuning would be required for protein–protein attraction (LLPS) alone to mediate chromatin associated protein droplet formation and chromatin compaction/isolation *in vivo*.

An interesting regime in our multivalent protein simulations is for intermediate values of ϵ_{HH}, where a droplet only forms if ϵ_{HC} is large enough. In other words, phase separation is promoted by interaction with chromatin; this can be viewed as chromatin–HP1–HP1–chromatin bridges enabling the BIA to drive protein clustering. In this regime we also see a dependence on the overall protein density ρ which is fundamentally different to standard LLPS. The density of proteins within and outside the droplet depends on ρ, and the droplet volume grows sub–linearly as ρ increases. This behaviour originates from the formation of a "loose" protein cluster on the chromatin for small ρ, which can "fill up" as proteins are added to the system; at larger ρ, sites on the chromatin become saturated, so as more proteins are added these instead remain unbound (increasing the density in the protein poor region). This is reminiscent of recent work showing that varying the overall concentration of the nucleophosmin protein (a key component of nucleoli, which form via LLPS) leads to variation in its density both inside and outside the nucleolus; in that system there are multiple phase-separating components which leads to a complicated high–dimensional phase diagram [275]. This result is also compatible with a recent study, that showed that over–expression of HP1 in mouse *does not* lead to an increase in the size of foci, but instead the protein density within the foci increases [270].

The limited valence model showed similar regimes to the multivalent case, but instead of a spherical droplet the proteins formed fractal clusters [similar to the structures formed by patchy particles [230, 231]]. The limited valence HP1s could also form a gel in simulations with a higher density and periodic boundaries, as shown in experiments [157].

To summarise, in this chapter I present some useful result on the equilibrium properties of self–interacting HP1 which can bind chromatin; in particular we found that *(i)* the effective valence of these proteins and *(ii)* and the nature of their interactions are essential to promote the emergence of different high–density phases, also very different from the standard liquid droplet.

Future works might be directed to the study of "good" bridgers (e.g. simple spheres). A phase diagram of BIPS in the presence of simple spheres interacting with chromatin has not been presented yet. In recent works, it has also been shown that cluster formation via the BIA can be dramatically altered by non–equilibrium chemical reactions which stochastically switch the proteins back and forward between a binding and non–binding state [modelling post–translational modifications [141]]. It would be interesting to study the equilibrium properties of clusters in such contexts, using the same approach presented here (i.e. by adding also an attractive interaction between switching proteins). It could be relevant to study the case where multiple strong binding sites, which are spatially separated, or dynamical (mimicking the histone modifications changes as genes are activated), act as nucleation points for chromatin–binding proteins.

Overall conclusions

One of the main goal of molecular biology is to explain the complex relationship between chromatin structure and function. Many techniques can be employed to reveal chromatin organisation at large scales, while specific mechanochemical experiments can reveal DNA structure at a small length scale; other methods can be used to track transcription patterns. Experiments and numerics suggest that complex networks of transcription units and genes may be highly sensible to small changes in local regions [276, 277] in some conditions. The complexity of regulatory networks is still under debate, and many questions on the interplay between DNA and chromatin conformation and gene expression are still left unanswered.

In this thesis' work, I show some examples on how physical modelling can help to unveil possible connections between DNA and chromatin conformation and functioning. Not only the thesis follows a "bottom–up" approach, as chapters are sorted by increasing length scales of interest, but also two are the main methods of investigation exploited here: in the first two chapters, simple one–dimensional models of DNA and chromatin fibre have been proposed, and Monte Carlo algorithms have been employed to replicate the stochasticity of real biochemical reactions; in the other chapters, extensive MD simulations were used to simulate brownian dynamics of chromatin fibre and associated proteins, and a realistic 3D view of protein aggregates and chromatin conformation has been shown.

In general, we observe that positive feedback loops can facilitate the emergence of non–trivial macroscopic behaviours. In chapter 1, I show that the coupling between supercoiling dynamics with transcription initiation introduces non–linearity in a simple model of bacterial gene arrays. This results in the emergence of a bursty regime: when the coupling is relatively weak, transcription is highly intermittent and distributions of supercoiling at promoter are strongly non–gaussian; on the contrary, a collective supercoiling dynamics are at play when polymerases strongly affect the supercoiling at gene promoters: supercoiling waves arise for single gene and tandem geometry, whereas in the presence of a pair of divergent genes, the transcription across all genes is almost totally dominated by them. In chapter 2, I show that long–range contacts can favour the estabilishment and maintainance of epigenetic domains.

After nucleation, methylation marks can spread across a genome region when chromatin is crumpled, whilst for more swollen chromatin fibre, short–range contacts dominates and local erasure prevails, leaving the chromatin domain either in a non–methylated or a mixed epigenetic state.

In chapter 3 we investigate in detail the dynamics of proteins which undergo BIPS. Bridging–induced attraction provides a simple mechanism to explain the growth of protein clusters and heterochromatin compaction; I show the emergence of solid–like and glassy behaviours, for a uniformly interacting chromatin fibre and in the presence of strong binding sites interposed between weakly–interacting stretches of chromatin, respectively. In chapter4 4, I revise the biochemistry of the HP1, an ubiquitous heterochromatin-associated factor, and I study the interplay between LLPS and BIPS of these proteins. Different macroscopic behaviours have been obtained by changing the relevant energies at play: coating, dense and absorbing droplet phases enrich the already copious behaviour of bridging–induced phase separating systems; remerkably, a regime where the systems behaves different from simple LLPS arises: the density of droplet increases as the volume fraction becomes larger. Gel–like structure can form due to the limited valence of model HP1s.

I would like to ackowledge my supervisors, Davide and Chris, for helping me to write this thesis and their contribution as co–authors of my works; all other co–authors of my publications, Beppe G., Alessandro P., Davide M., Alessandro B., Michele C.; also, I would like to thank all my friends and my family.

Bibliography

[1] Aaron Novick and Milton Weiner. "Enzyme induction as an all-or-none phenomenon". *Proceedings of the National Academy of Sciences of the United States of America* 43 (1957), pp. 553–566.

[2] John L Spudich and Daniel E Koshland. "Non-genetic individuality: chance in the single cell". *Nature* 262 (1976), pp. 467–471.

[3] Harley H McAdams and Adam Arkin. "Stochastic mechanisms in gene expression". *Proceedings of the National Academy of Sciences* 94 (1997), pp. 814–819.

[4] Mukund Thattai and Alexander Van Oudenaarden. "Intrinsic noise in gene regulatory networks". *Proceedings of the National Academy of Sciences* 98 (2001), pp. 8614–8619.

[5] Attila Becskei and Luis Serrano. "Engineering stability in gene networks by autoregulation". *Nature* 405 (2000), pp. 590–593.

[6] Ido Golding et al. "Real-time kinetics of gene activity in individual bacteria". *Cell* 123 (2005), pp. 1025–1036.

[7] Shasha Chong et al. "Mechanism of transcriptional bursting in bacteria". *Cell* 158 (2014), pp. 314–326.

[8] Johan Paulsson. "Models of stochastic gene expression". *Physics of life reviews* 2 (2005), pp. 157–175.

[9] Brian Munsky, Gregor Neuert, and Alexander Van Oudenaarden. "Using gene expression noise to understand gene regulation". *Science* 336 (2012), pp. 183–187.

[10] Andrew Mugler, Aleksandra M Walczak, and Chris H Wiggins. "Spectral solutions to stochastic models of gene expression with bursts and regulation". *Physical Review E* 80 (2009).

[11] Tianshou Zhou and Jiajun Zhang. "Analytical results for a multistate gene model". *SIAM Journal on Applied Mathematics* 72 (2012), pp. 789–818.

[12] CA Brackley et al. "Stochastic model of supercoiling-dependent transcription". *Physical review letters* 117 (2016).

[13] A Bentivoglio et al. "Non-equilibrium phase transition in a model for supercoiling-dependent DNA transcription". *Soft matter* 14 (2018), pp. 3632–3639.

[14] Marco Ancona et al. "Transcriptional Bursts in a Nonequilibrium model for gene regulation by supercoiling". *Biophysical journal* 117 (2019), pp. 369–376.

[15] Stuart A Sevier and Herbert Levine. "Mechanical properties of transcription". *Physical review letters* 118 (2017).

[16] Stuart A Sevier and Herbert Levine. "Properties of gene expression and chromatin structure with mechanically regulated elongation". *Nucleic acids research* 46 (2018), pp. 5924–5934.

[17] Roger Weil and Jerome Vinograd. "The cyclic helix and cyclic coil forms of polyoma viral DNA". *Proceedings of the National Academy of Sciences of the United States of America* 50 (1963), pp. 730–738.

[18] Leroy F Liu and James C Wang. "Supercoiling of the DNA template during transcription". *Proceedings of the National Academy of Sciences* 84 (1987), pp. 7024–7027.

[19] Bruce Alberts et al. *Molecular biology of the cell*. 6th ed. New York and Abingdon, UK: W.H. Freeman Co, 2007.

[20] Peter Yakovchuk, Ekaterina Protozanova, and Maxim D. Frank-Kamenetskii. "Base-stacking and base-pairing contributions into thermal stability of the DNA double helix". *Nucleic Acids Research* 34 (2006), pp. 564–574.

[21] Jeremy M Berg, John L Tymoczko, and Lubert Stryer. *Biochemistry*. Ed. by Macmillan. 1st ed. New York, 2002.

[22] Oliver Henrich et al. "Coarse-grained simulation of DNA using LAMMPS". *The European Physical Journal E* 41 (2018), pp. 1–16.

[23] Peter L Privalov et al. "What drives proteins into the major or minor grooves of DNA?" *Journal of molecular biology* 365 (2007), pp. 1–9.

[24] Sarah White et al. "Recognition of the four Watson–Crick base pairs in the DNA minor groove by synthetic ligands". *Nature* 391 (1998), pp. 468–471.

[25] Richard R Sinden. *DNA structure and function*. Ed. by Macmillan. 1st ed. New York, 1994.

[26] Richard E Dickerson. "DNA structure from A to Z". *Methods in enzymology* 211 (1992), pp. 67–111.

[27] Subramaniyam Ravichandran, Vinod Kumar Subramani, and Kyeong Kyu Kim. "Z-DNA in the genome: from structure to disease". *Biophysical reviews* 11 (2019), pp. 383–387.

[28] *Zephyris*. URL: https://en.wikipedia.org/wiki/File:DNA_Structure\%2BKey\%2BLabelled.pn_NoBB.png.

[29] Joan C Marini et al. "Bent helical structure in kinetoplast DNA". *Proceedings of the National Academy of Sciences* (1982), pp. 7664–7668.

[30] Hen-Ming Wu and Donald M Crothers. "The locus of sequence-directed and protein-induced DNA bending". *Nature* 308 (1984), pp. 509–513.

[31] Yair Augusto Gutiérrez Fosado. *Double-stranded coarse grained model for DNA : applications to supercoiling and denaturation*.

[32] James H White. "Self-linking and the Gauss integral in higher dimensions". *American journal of mathematics* 91 (1969), pp. 693–728.

[33] F Brock Fuller. "The writhing number of a space curve". *Proceedings of the National Academy of Sciences* 68 (1971), pp. 815–819.

[34] F Brock Fuller. "Decomposition of the linking number of a closed ribbon: a problem from molecular biology". *Proceedings of the National Academy of Sciences* 75 (1978), pp. 3557–3561.

[35] Andrew D Bates, Anthony Maxwell, Tony Maxwell, et al. *DNA topology*. 1st ed. 2005.

[36] Guillaume Witz and Andrzej Stasiak. "DNA supercoiling and its role in DNA decatenation and unknotting". *Nucleic Acids Research* 38 (Dec. 2009), pp. 2119–2133.

[37] JG Brahms et al. "Activation and inhibition of transcription by supercoiling". *Journal of molecular biology* 181 (1985), pp. 455–465.

[38] James J Champoux. "DNA topoisomerases: structure, function, and mechanism". *Annual review of biochemistry* 70 (2001), pp. 369–413.

[39] Ana G Abril et al. "Prokaryotic sigma factors and their transcriptional counterparts in Archaea and Eukarya". *Applied microbiology and biotechnology* 104 (2020), pp. 4289–4302.

[40] Renjith Mathew and Dipankar Chatterji. "The evolving story of the omega subunit of bacterial RNA polymerase". *Trends in microbiology* 14 (2006), pp. 450–455.

[41] Claudio Rivetti, Martin Guthold, and Carlos Bustamante. "Wrapping of DNA around the E. coli RNA polymerase open promoter complex". *The EMBO journal* 18 (1999), pp. 4464–4475.

[42] Kate L Henderson et al. "Mechanism of transcription initiation and promoter escape by E. coli RNA polymerase". *Proceedings of the National Academy of Sciences* 114 (2017), E3032–E3040.

[43] Catherine A Brennan, Alicia J Dombroski, and Terry Platt. "Transcription termination factor rho is an RNA-DNA helicase". *Cell* 48 (1987), pp. 945–952.

[44] Bokun Cheng et al. "Direct interaction between Escherichia coli RNA polymerase and the zinc ribbon domains of DNA topoisomerase I". *Journal of Biological Chemistry* 278 (2003), pp. 30705–30710.

[45] Ksenia Terekhova et al. "Bacterial topoisomerase I and topoisomerase III relax supercoiled DNA via distinct pathways". *Nucleic acids research* 40 (2012), pp. 10432–10440.

[46] Purnananda Guptasarma. "Cooperative relaxation of supercoils and periodic transcriptional initiation within polymerase batteries". *Bioessays* 18 (1996), pp. 325–332.

[47] Yair AG Fosado et al. "Nonequilibrium dynamics and action at a distance in transcriptionally driven DNA supercoiling". *Proceedings of the National Academy of Sciences* 118 (2021).

[48] HDPP Bremer, Patrick P Dennis, et al. "Modulation of chemical composition and other parameters of the cell by growth rate". *Escherichia coli and Salmonella: cellular and molecular biology* 2 (1996), pp. 1553–69.

[49] Mads Kaern et al. "Stochasticity in gene expression: from theories to phenotypes". *Nature Reviews Genetics* 6 (2005), pp. 451–464.

[50] Otto G Berg, Robert B Winter, and Peter H Von Hippel. "Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory". *Biochemistry* 20 (1981), pp. 6929–6948.

[51] Stephen E Halford and John F Marko. "How do site-specific DNA-binding proteins find their targets?" *Nucleic acids research* 32 (2004), pp. 3040–3052.

[52] Arjun Raj and Alexander Van Oudenaarden. "Nature, nurture, or chance: stochastic gene expression and its consequences". *Cell* 135 (2008), pp. 216–226.

[53] Björn Schwanhäusser et al. "Global quantification of mammalian gene expression control". *Nature* 473 (2011), pp. 337–342.

[54] Claire V Harper et al. "Dynamic analysis of stochastic transcription cycles". *PLoS Biol* 9 (2011).

[55] David M Suter et al. "Mammalian genes are transcribed with widely different bursting kinetics". *Science* 332 (2011), pp. 472–474.

[56] Crispin W Gardiner et al. *Handbook of stochastic methods*. Vol. 3. springer Berlin, 1985.

[57] Maciej Dobrzyński and Frank J Bruggeman. "Elongation dynamics shape bursty transcription and translation". *Proceedings of the National Academy of Sciences* 106 (2009), pp. 2583–2588.

[58] Niraj Kumar, Abhyudai Singh, and Rahul V Kulkarni. "Transcriptional bursting in gene expression: analytical results for general stochastic models". *PLoS Comput Biol* 11 (2015).

[59] John F Marko. "Torque and dynamics of linking number relaxation in stretched supercoiled DNA". *Physical Review E* 76 (2007).

[60] Marijn TJ van Loenhout, MV De Grunt, and Cees Dekker. "Dynamics of DNA supercoils". *Science* 338 (2012), pp. 94–97.

[61] Randall D Kamien. "Local writhing dynamics". *The European Physical Journal B-Condensed Matter and Complex Systems* 1 (1998), pp. 1–4.

[62] Yair Augusto Gutierrez Fosado, Davide Michieletto, and Davide Marenduzzo. "Dynamical scaling and phase coexistence in topologically constrained dna melting". *Physical review letters* 119 (2017).

[63] Giada Forte et al. "Plectoneme dynamics and statistics in braided polymers". *Physical Review E* 99 (2019).

[64] Paul M Chaikin, Tom C Lubensky, and Thomas A Witten. *Principles of condensed matter physics*. Vol. 10. Cambridge university press Cambridge, 1995.

[65] Marco ANCONA and Giuseppe GONNELLA. "Proprietà statistiche e burst di trascrizione in un modello stocastico per la trascrizione genica dipendente dal supercoiling" ().

[66] S-T Liang et al. "Activities of constitutive promoters in Escherichia coli". *Journal of molecular biology* 292 (1999), pp. 19–37.

[67] Vicent Pelechano, Sebastián Chávez, and José E Pérez-Ortín. "A complete set of nascent transcription rates for yeast genes". *PloS one* 5 (2010).

[68] Dean A Jackson, ANA Pombo, and Francisco Iborra. "The balance sheet for transcription: an analysis of nuclear RNA metabolism in mammalian cells". *The FASEB Journal* 14 (2000), pp. 242–254.

[69] Yasushi Ishihama et al. "Protein abundance profiling of the Escherichia coli cytosol". *BMC genomics* 9 (2008), pp. 1–17.

[70] Harold Weintraub, Pei Feng Cheng, and Kathleen Conrad. "Expression of transfected DNA depends on DNA topology". *Cell* 46 (1986), pp. 115–122.

[71] Adam M Corrigan et al. "A continuum model of transcriptional bursting". *Elife* 5 (2016).

[72] Peter R Cook. "The organization of replication and transcription". *Science* 284 (1999), pp. 1790–1795.

[73] Catherine Naughton et al. "Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures". *Nature structural & molecular biology* 20 (2013), pp. 387–395.

[74] Fedor Kouzine et al. "Transcription-dependent dynamic supercoiling is a short-range genomic force". *Nature structural & molecular biology* 20 (2013), pp. 396–403.

[75] Aaron D Goldberg, C David Allis, and Emily Bernstein. "Epigenetics: a landscape takes shape". *Cell* 128 (2007), pp. 635–638.

[76] Conrad H Waddington. "Canalization of development and the inheritance of acquired characters". *Nature* 150 (1942), pp. 563–565.

[77] Simon Yuan Wang et al. "Role of epigenetics in unicellular to multicellular transition in Dictyostelium". *Genome biology* 22 (2021), pp. 1–30.

[78] Adrian Bird. "Perceptions of epigenetics". *Nature* 447 (2007), p. 396.

[79] Francesco Bottaccioli and Anna Giulia Bottaccioli. "La rivoluzione epigenetica". *Sistema salute. Rivista di educazione sanitaria e di promozione della* (2012), pp. 319–335.

[80] Conrad Hal Waddington. *The strategy of the genes*. Routledge, 2014.

[81] Luca Peliti. *Statistical mechanics in a nutshell*. Princeton University Press, 2011.

[82] Davide Michieletto et al. "Shaping epigenetic memory via genomic bookmarking". *Nucleic acids research* 46 (2018), pp. 83–93.

[83] *creode*. URL: https://it.wikipedia.org/wiki/Creode.

[84] Ian B Dodd et al. "Theoretical analysis of epigenetic cell memory by nucleosome modification". *Cell* 129 (2007), pp. 813–822.

[85] Ruggero Cortini et al. "The physics of epigenetics". *Reviews of Modern Physics* 88 (2016).

[86] Scott Berry, Caroline Dean, and Martin Howard. "Slow chromatin dynamics allow polycomb target genes to filter fluctuations in transcription factor activity". *Cell systems* 4 (2017), pp. 445–457.

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

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

[89] Daniel Jost. "Bifurcation in epigenetics: implications in development, proliferation, and diseases". *Physical Review E* 89 (2014).

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

[91] Mille A Micheelsen et al. "Theory for the stability and regulation of epigenetic landscapes". *Physical biology* 7 (2010).

[92] Behrouz Eslami-Mossallam, Helmut Schiessel, and John van Noort. "Nucleosome dynamics: Sequence matters". *Advances in colloid and interface science* 232 (2016), pp. 101–113.

[93] Karolin Luger et al. "Crystal structure of the nucleosome core particle at 2.8 Å resolution". *Nature* 389 (1997), pp. 251–260.

[94] James Allan et al. "Participation of core histone "tail" in the stabilization of the chromatin solenoid." *The Journal of cell biology* 93 (1982), pp. 285–297.

[95] Pu-Yeh Kan, Tamara L Caterino, and Jeffrey J Hayes. "The H4 tail domain participates in intra-and internucleosome interactions with protein and DNA during folding and oligomerization of nucleosome arrays". *Molecular and cellular biology* 29 (2009), pp. 538–546.

[96] Aurélie Bertin et al. "Structure and phase diagram of nucleosome core particles aggregated by multivalent cations". *Biophysical journal* 93 (2007), pp. 3652–3663.

[97] Gary Felsenfeld and James D McGhee. "Structure of the 30 nm chromatin fiber". *Cell* 44 (1986), pp. 375–377.

[98] Philip JJ Robinson et al. "EM measurements define the dimensions of the "30-nm" chromatin fiber: evidence for a compact, interdigitated structure". *Proceedings of the National Academy of Sciences* 103 (2006), pp. 6506–6511.

[99] Mikhail Eltsov et al. "Analysis of cryo-electron microscopy images does not support the existence of 30-nm chromatin fibers in mitotic chromosomes in situ". *Proceedings of the National Academy of Sciences* 105 (2008), pp. 19732–19737.

[100] Renata Garces, R Podgornik, and Vladimir Lorman. "Antipolar and anticlinic mesophase order in chromatin induced by nucleosome polarity and chirality correlations". *Physical review letters* 114 (2015).

[101] Tony Kouzarides. "Chromatin modifications and their function". *Cell* 128 (2007), pp. 693–705.

[102] Christopher R Vakoc et al. "Histone H3 lysine 9 methylation and HP1γ are associated with transcription elongation through mammalian chromatin". *Molecular cell* 19 (2005), pp. 381–391.

[103] Christian Beisel and Renato Paro. "Silencing chromatin: comparing modes and mechanisms". *Nature Reviews Genetics* 12 (2011), pp. 123–135.

[104] Beat Fierz. "Synthetic chromatin approaches to probe the writing and erasing of histone modifications". *ChemMedChem* 9 (2014), pp. 495–504.

[105] Nehmé Saksouk, Elisabeth Simboeck, and Jérôme Déjardin. "Constitutive heterochromatin formation and transcription in mammals". *Epigenetics & chromatin* 8 (2015), pp. 1–17.

[106] Kevin D Sarge and Ok-Kyong Park-Sarge. "Gene bookmarking: keeping the pages open". *Trends in biochemical sciences* (2005), pp. 605–610.

[107] Katharina Müller-Ott et al. "Specificity, propagation, and memory of pericentric heterochromatin". *Molecular systems biology* 10 (2014), p. 746.

[108] Eugenia Y Xu, Karl A Zawadzki, and James R Broach. "Single-cell observations reveal intermediate transcriptional silencing states". *Molecular cell* 23 (2006), pp. 219–229.

[109] Nathaniel A Hathaway et al. "Dynamics and memory of heterochromatin in living cells". *Cell* 149 (2012), pp. 1447–1460.

[110] D Richards, S Berry, and M Howard. "Illustrations of mathematical modeling in biology: Epigenetics, meiosis, and an outlook". *Cold Spring Harbor symposia on quantitative biology*. Vol. 77. Cold Spring Harbor Laboratory Press. 2012, pp. 175–181.

[111] Bradley E Bernstein et al. "Genomic maps and comparative analysis of histone modifications in human and mouse". *Cell* 120 (2005), pp. 169–181.

[112] Bradley E Bernstein et al. "A bivalent chromatin structure marks key developmental genes in embryonic stem cells". *Cell* 125 (2006), pp. 315–326.

[113] Véronique Azuara et al. "Chromatin signatures of pluripotent cell lines". *Nature cell biology* 8 (2006), pp. 532–538.

[114] Kim Sneppen and Leonie Ringrose. "Theoretical analysis of Polycomb-Trithorax systems predicts that poised chromatin is bistable and not bivalent". *Nature communications* 10 (2019), pp. 1–18.

[115] Callum J O'Kane and Edel M Hyland. "Yeast epigenetics: the inheritance of histone modification states". *Bioscience reports* 39 (2019).

[116] Mohammad Sedighi and Anirvan M Sengupta. "Epigenetic chromatin silencing: bistability and front propagation". *Physical biology* 4 (2007), p. 246.

[117] Shiv IS Grewal and Songtao Jia. "Heterochromatin revisited". *Nature Reviews Genetics* 8 (2007), pp. 35–46.

[118] Shinji Honda et al. "The DMM complex prevents spreading of DNA methylation from transposons to nearby genes in Neurospora crassa". *Genes & development* 24 (2010), pp. 443–454.

[119] Pauline NCB Audergon et al. "Restricted epigenetic inheritance of H3K9 methylation". *Science* 348 (2015), pp. 132–135.

[120] Ke Zhang et al. "Roles of the Clr4 methyltransferase complex in nucleation, spreading and maintenance of heterochromatin". *Nature structural & molecular biology* 15 (2008), p. 381.

[121] Nabieh Ayoub et al. "A novel jmjC domain protein modulates heterochromatization in fission yeast". *Molecular and cellular biology* 23 (2003), pp. 4356–4370.

[122] Sara Isaac et al. "Interaction of Epe1 with the heterochromatin assembly pathway in Schizosaccharomyces pombe". *Genetics* 175 (2007), pp. 1549–1560.

[123] Ian B Dodd and Kim Sneppen. "Barriers and silencers: a theoretical toolkit for control and containment of nucleosome-based epigenetic states". *Journal of molecular biology* 414 (2011), pp. 624–637.

[124] Diana David-Rus et al. "Inheritance of epigenetic chromatin silencing". *Journal of theoretical biology* 258 (2009), pp. 112–120.

[125] Davide Michieletto et al. "Nonequilibrium theory of epigenomic microphase separation in the cell nucleus". *Physical review letters* 123 (2019).

[126] Erez Lieberman-Aiden et al. "Comprehensive mapping of long-range interactions reveals folding principles of the human genome". *science* 326 (2009), pp. 289–293.

[127] A Colliva et al. "Ising-model description of long-range correlations in DNA sequences". *Physical Review E* 91 (2015).

[128] Vasily Zaburdaev, Sergey Denisov, and Joseph Klafter. "Lévy walks". *Reviews of Modern Physics* 87 (2015).

[129] Johan H Gibcus et al. "A pathway for mitotic chromosome formation". *Science* 359 (2018).

[130] Mariano Barbieri et al. "Complexity of chromatin folding is captured by the strings and binders switch model". *Proceedings of the National Academy of Sciences* 109 (2012), pp. 16173–16178.

[131] Alexandra Zidovska, David A Weitz, and Timothy J Mitchison. "Micron-scale coherence in interphase chromatin dynamics". *Proceedings of the National Academy of Sciences* 110 (2013), pp. 15555–15560.

[132] Haye Hinrichsen and Martin Howard. "A model for anomalous directed percolation". *The European Physical Journal B-Condensed Matter and Complex Systems* 7 (1999), pp. 635–643.

[133] Martin J Howard and Uwe C Täuber. "Real 'versus imaginary' noise in diffusion-limited reactions". *Journal of Physics A: Mathematical and General* 30 (1997).

[134] Marco Ancona, Davide Michieletto, and Davide Marenduzzo. "Competition between local erasure and long-range spreading of a single biochemical mark leads to epigenetic bistability". *Physical Review E* 101 (2020).

[135] L Peliti. "Renormalisation of fluctuation effects in the A+ A to A reaction". *Journal of Physics A: Mathematical and General* 19 (1986), p. L365.

[136] Daniel Vernon and Martin Howard. "Branching and annihilating Lévy flights". *Physical Review E* 63 (2001).

[137] Daniel C Vernon. "Long range hops and the pair annihilation reaction A+ A 0: Renormalization group and simulation". *Physical Review E* 68 (2003).

[138] DE Rodriguez, MA Bab, and EV Albano. "Effective multidimensional crossover behavior in a one-dimensional voter model with long-range probabilistic interactions". *Physical Review E* 83 (2011).

[139] Fabian Erdel and Eric C Greene. "Generalized nucleation and looping model for epigenetic memory of histone modifications". *Proceedings of the National Academy of Sciences* 113 (2016), E4180–E4189.

[140] Soya Shinkai et al. "PHi-C: deciphering Hi-C data into polymer dynamics". *NAR genomics and bioinformatics* 2 (2020).

[141] Chris A Brackley et al. "Nonequilibrium chromosome looping via molecular slip links". *Physical review letters* 119 (2017).

[142] A Grosberg et al. "Crumpled globule model of the three-dimensional structure of DNA". *EPL (Europhysics Letters)* 23 (1993), p. 373.

[143] Leonid A Mirny. "The fractal globule as a model of chromatin architecture in the cell". *Chromosome research* 19 (2011), pp. 37–51.

[144] Mario Nicodemi and Ana Pombo. "Models of chromosome structure". *Current opinion in cell biology* 28 (2014), pp. 90–95.

[145] Sito Torres-Garcia et al. "Epigenetic gene silencing by heterochromatin primes fungal resistance". *Nature* 585 (2020), pp. 453–458.

[146] Andrew Bassett et al. "The folding and unfolding of eukaryotic chromatin". *Current opinion in genetics & development* 19 (2009), pp. 159–165.

[147] Shantanu Kadam, Tripti Bameta, and Ranjith Padinhateeri. "Nucleosome sliding can influence the spreading of histone modifications". *arXiv preprint arXiv:2110.05905* (2021).

[148] John T Finch and Aaron Klug. "Solenoidal model for superstructure in chromatin". *Proceedings of the National Academy of Sciences* 73 (1976), pp. 1897–1901.

[149] CL Woodcock, L-LY Frado, and JB Rattner. "The higher-order structure of chromatin: evidence for a helical ribbon arrangement." *The Journal of cell biology* 99 (1984), pp. 42–52.

[150] Brian D Athey et al. "The diameters of frozen-hydrated chromatin fibers increase with DNA linker length: evidence in support of variable diameter models for chromatin." *The Journal of cell biology* 111 (1990), pp. 795–806.

[151] Ken van Holde and Jordanka Zlatanova. "Chromatin fiber structure: Where is the problem now?" *Seminars in cell & developmental biology*. Vol. 18. Elsevier. 2007, pp. 651–658.

[152] Jacques Dubochet et al. "Cryo-electron microscopy of vitrified specimens". *Quarterly reviews of biophysics* 21 (1988), pp. 129–228.

[153] Kazuhiro Maeshima, Saera Hihara, and Mikhail Eltsov. "Chromatin structure: does the 30-nm fibre exist in vivo?" *Current opinion in cell biology* 22 (2010), pp. 291–297.

[154] Kazuhiro Maeshima et al. "Chromatin as dynamic 10-nm fibers". *Chromosoma* 123 (2014), pp. 225–237.

[155] Yasumasa Joti et al. "Chromosomes without a 30-nm chromatin fiber". *Nucleus* 3 (2012), pp. 404–410.

[156] Stephen E Farr et al. "Nucleosome plasticity is a critical element of chromatin liquid–liquid phase separation and multivalent nucleosome interactions". *Nature communications* 12 (2021), pp. 1–17.

[157] Bryce E Ackermann and Galia T Debelouchina. "Heterochromatin protein HP1α gelation dynamics revealed by solid-state NMR spectroscopy". *Angewandte Chemie International Edition* 58 (2019), pp. 6300–6305.

[158] W Michael Babinchak and Witold K Surewicz. "Liquid–liquid phase separation and its mechanistic role in pathological protein aggregation". *Journal of molecular biology* 432 (2020), pp. 1910–1925.

[159] Priyanka Dogra et al. "Intermolecular Charge-Transfer Modulates Liquid–Liquid Phase Separation and Liquid-to-Solid Maturation of an Intrinsically Disordered pH-Responsive Domain". *Journal of the American Chemical Society* 141 (2019), pp. 20380–20389.

[160] R Flamia et al. "Transformation of amyloid-like fibers, formed from an elastin-based biopolymer, into a hydrogel: an X-ray photoelectron spectroscopy and atomic force microscopy study". *Biomacromolecules* 8 (2007), pp. 128–138.

[161] Sreenivas Chavali, Alexander Gunnarsson, and M Madan Babu. "Intrinsically disordered proteins adaptively reorganize cellular matter during stress". *Trends in biochemical sciences* 42 (2017), pp. 410–412.

[162] Chris A Brackley et al. "Nonspecific bridging-induced attraction drives clustering of DNA-binding proteins and genome organization". *Proceedings of the National Academy of Sciences* 110 (2013), E3605–E3611.

[163] Chris A Brackley et al. "Ephemeral protein binding to DNA shapes stable nuclear bodies and chromatin domains". *Biophysical journal* 112 (2017), pp. 1085–1093.

[164] Chris A Brackley and Davide Marenduzzo. "Bridging-induced microphase separation: photobleaching experiments, chromatin domains and the need for active reactions". *Briefings in functional genomics* 19 (2020), pp. 111–118.

[165] Anatoly A Zinchenko, Kenichi Yoshikawa, and Damien Baigl. "Compaction of single-chain DNA by histone-inspired nanoparticles". *Physical review letters* 95 (2005).

[166] Anatoly A Zinchenko et al. "Single-chain compaction of long duplex DNA by cationic nanoparticles: modes of interaction and comparison with chromatin". *The Journal of Physical Chemistry B* 111 (2007), pp. 3019–3031.

[167] Sharon Pepenella, Kevin J Murphy, and Jeffrey J Hayes. "Intra-and inter-nucleosome interactions of the core histone tail domains in higher-order chromatin structure". *Chromosoma* 123 (2014), pp. 3–13.

[168] Christopher L Woodcock, Arthur I Skoultchi, and Yuhong Fan. "Role of linker histone in chromatin structure and function: H1 stoichiometry and nucleosome repeat length". *Chromosome Research* 14 (2006), pp. 17–25.

[169] Zhonghui Tang et al. "CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription". *Cell* 163 (2015), pp. 1611–1627.

[170] Kerstin S Wendt et al. "Cohesin mediates transcriptional insulation by CCCTC-binding factor". *Nature* 451 (2008), pp. 796–801.

[171] Joel C Eissenberg and Sarah CR Elgin. "The HP1 protein family: getting a grip on chromatin". *Current opinion in genetics & development* 10 (2000), pp. 204–210.

[172] Jeffrey C Hansen et al. "The 10-nm chromatin fiber and its relationship to interphase chromosome organization". *Biochemical Society Transactions* 46 (2018), pp. 67–76.

[173] Andrew Routh, Sara Sandin, and Daniela Rhodes. "Nucleosome repeat length and linker histone stoichiometry determine chromatin fiber structure". *Proceedings of the National Academy of Sciences* 105 (2008), pp. 8872–8877.

[174] JEAN O THOMAS. "The higher order structure of chromatin and histone H1". *Journal of Cell Science* 1984 (1984), pp. 1–20.

[175] Yoshinori Nishino et al. "Human mitotic chromosomes consist predominantly of irregularly folded nucleosome fibres without a 30-nm chromatin structure". *The EMBO journal* 31 (2012), pp. 1644–1653.

[176] Michael E Cates and Elsen Tjhung. "Theories of binary fluid mixtures: from phase-separation kinetics to active emulsions". *Journal of Fluid Mechanics* 836 (2018).

[177] Kurt Binder. "Theory of first-order phase transitions". *Reports on progress in physics* 50 (1987).

[178] Steven Boeynaems et al. "Protein phase separation: a new phase in cell biology". *Trends in cell biology* 28 (2018), pp. 420–435.

[179] April L Darling, Boris Y Zaslavsky, and Vladimir N Uversky. "Intrinsic disorder-based emergence in cellular biology: Physiological and pathological liquid-liquid phase transitions in cells". *Polymers* 11 (2019).

[180] Diana M Mitrea and Richard W Kriwacki. "Phase separation in biology; functional organization of a higher order". *Cell Communication and Signaling* 14 (2016), pp. 1–20.

[181] Job Dekker, Marc A Marti-Renom, and Leonid A Mirny. "Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data". *Nature Reviews Genetics* 14 (2013), pp. 390–403.

[182] Jesse R Dixon et al. "Topological domains in mammalian genomes identified by analysis of chromatin interactions". *Nature* 485 (2012), pp. 376–380.

[183] Jean-Philippe Fortin and Kasper D Hansen. "Reconstructing A/B compartments as revealed by Hi-C using long-range correlations in epigenetic data". *Genome biology* 16 (2015), pp. 1–23.

[184] Spiros Palikyras and Argyris Papantonis. "Modes of phase separation affecting chromatin regulation". *Open biology* 9 (2019).

[185] Kim Nasmyth. "Cohesin: a catenase with separate entry and exit gates?" *Nature cell biology* 13 (2011), pp. 1170–1177.

[186] Iain F Davidson et al. "DNA loop extrusion by human cohesin". *Science* 366 (2019), pp. 1338–1345.

[187] Siheng Xiang and Douglas Koshland. "Cohesin in space and time: architecture and oligomerization in vivo." *bioRxiv* (2020).

[188] Di Shi et al. "The acetyltransferase Eco1 elicits cohesin dimerization during S phase". *Journal of Biological Chemistry* 295 (2020), pp. 7554–7565.

[189] Jian Yan et al. "Transcription factor binding in human cells occurs in dense clusters formed around cohesin anchor sites". *Cell* 154 (2013), pp. 801–813.

[190] Davide Marenduzzo, Cristian Micheletti, and Peter R Cook. "Entropy-driven genome organization". *Biophysical journal* 90 (2006), pp. 3712–3721.

[191] Renko de Vries. "Influence of mobile DNA-protein-DNA bridges on DNA configurations: coarse-grained Monte-Carlo simulations". *The Journal of chemical physics* 135 (2011).

[192] D Marenduzzo and E Orlandini. "Topological and entropic repulsion in biopolymers". *Journal of statistical mechanics: theory and experiment* 2009 (2009).

[193] Meng Xu and Peter R Cook. "Similar active genes cluster in specialized transcription factories". *Journal of Cell Biology* 181 (2008), pp. 615–623.

[194] Argyris Papantonis and Peter R Cook. "Transcription factories: genome organization and gene regulation". *Chemical reviews* 113 (2013), pp. 8683–8705.

[195] Tina L Tootle and Ilaria Rebay. "Post-translational modifications influence transcription factor activity: a view from the ETS superfamily". *Bioessays* 27 (2005), pp. 285–298.

[196] Korie E Handwerger, Christine Murphy, and Joseph G Gall. "Steady-state dynamics of Cajal body components in the Xenopus germinal vesicle". *The Journal of cell biology* 160 (2003), pp. 495–504.

[197] Tom Misteli, Javier F Caceres, and David L Spector. "The dynamics of a pre-mRNA splicing factor in living cells". *Nature* 387 (1997), pp. 523–527.

[198] Robert D Phair and Tom Misteli. "High mobility of proteins in the mammalian cell nucleus". *Nature* 404 (2000), pp. 604–609.

[199] Sabine Snaar et al. "Mutational analysis of fibrillarin and its mobility in living human cells". *Journal of Cell Biology* 151 (2000), pp. 653–662.

[200] Amy R Strom et al. "Phase separation drives heterochromatin domain formation". *Nature* 547 (2017), pp. 241–245.

[201] Yi-Hsuan Lin, Julie D Forman-Kay, and Hue Sun Chan. "Theories for sequence-dependent phase behaviors of biomolecular condensates". *Biochemistry* 57 (2018), pp. 2499–2508.

[202] Remus T Dame, Maarten C Noom, and Gijs JL Wuite. "Bacterial chromatin organization by H-NS protein unravelled using dual DNA manipulation". *Nature* 444 (2006), pp. 387–390.

[203] Je-Kyung Ryu et al. "Bridging-induced phase separation induced by cohesin SMC protein complexes". *Science Advances* 7 (2021).

[204] Karsten Rippe. "Liquid–Liquid Phase Separation in Chromatin". *Cold Spring Harbor Perspectives in Biology* (2021).

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

[206] Michael T Hons et al. "Topology and structure of an engineered human cohesin complex bound to Pds5B". *Nature Communications* 7 (2016), pp. 1–11.

[207] Steve Plimpton. "Fast Parallel Algorithms for Short-Range Molecular Dynamics". *J. Comp. Phys.* 117 (1995), pp. 1–19.

[208] Paul J Steinhardt, David R Nelson, and Marco Ronchetti. "Bond-orientational order in liquids and glasses". *Physical Review B* 28 (1983).

[209] Thomas Speck, Alex Malins, and C Patrick Royall. "First-order phase transition in a model glass former: Coupling of local structure and dynamics". *Physical review letters* 109 (2012).

[210] Alex Malins et al. "Identification of structure in condensed matter with the topological cluster classification". *The Journal of chemical physics* 139 (2013).

[211] Der-Tsai Lee. "On k-nearest neighbor Voronoi diagrams in the plane". *IEEE transactions on computers* 100 (1982), pp. 478–487.

[212] Qiang Du, Vance Faber, and Max Gunzburger. "Centroidal Voronoi tessellations: Applications and algorithms". *SIAM review* 41 (1999), pp. 637–676.

[213] Hossein Eslami, Parvin Sedaghat, and Florian Müller-Plathe. "Local bond order parameters for accurate determination of crystal structures in two and three dimensions". *Physical Chemistry Chemical Physics* 20 (2018), pp. 27059–27068.

[214] Pieter-Reináten Wolde et al. "Simulation of homogeneous crystal nucleation close to coexistence". *Faraday discussions* 104 (1996), pp. 93–110.

[215] Hajime Tanaka. "Roles of bond orientational ordering in glass transition and crystallization". *Journal of Physics: Condensed Matter* 23 (2011).

[216] Wolfgang Lechner and Christoph Dellago. "Accurate determination of crystal structures based on averaged local bond order parameters". *The Journal of chemical physics* 129 (2008).

[217] Hossein Eslami, Neda Khanjari, and Florian Müller-Plathe. "A local order parameter-based method for simulation of free energy barriers in crystal nucleation". *Journal of chemical theory and computation* 13 (2017), pp. 1307–1316.

[218] Pankaj Mehta et al. "A high-bias, low-variance introduction to machine learning for physicists". *Physics reports* 810 (2019), pp. 1–124.

[219] David S Simmons et al. "Generalized localization model of relaxation in glass-forming liquids". *Soft Matter* 8 (2012), pp. 11455–11461.

[220] Beatriz A Pazmino Betancourt et al. "Quantitative relations between cooperative motion, emergent elasticity, and free volume in model glass-forming polymer materials". *Proceedings of the National Academy of Sciences* 112 (2015), pp. 2966–2971.

[221] Francesco Puosi, Antonio Tripodo, and Dino Leporini. "Fast vibrational modes and slow heterogeneous dynamics in polymers and viscous liquids". *International journal of molecular sciences* 20 (2019).

[222] Guang Shi et al. "Interphase human chromosome exhibits out of equilibrium glassy dynamics". *Nature communications* 9 (2018), pp. 1–13.

[223] Soya Shinkai et al. "Dynamic nucleosome movement provides structural information of topological chromatin domains in living human cells". *PLoS computational biology* 12 (2016).

[224] Clifford P Brangwynne, Peter Tompa, and Rohit V Pappu. "Polymer physics of intracellular phase transitions". *Nature Physics* 11 (2015), pp. 899–904.

[225] Tharappel C James and SC Elgin. "Identification of a nonhistone chromosomal protein associated with heterochromatin in Drosophila melanogaster and its gene". *Molecular and cellular biology* 6 (1986), pp. 3862–3872.

[226] Rebecca Kellum, Jordan W Raff, and Bruce M Alberts. "Heterochromatin protein 1 distribution during development and during the cell cycle in Drosophila embryos". *Journal of cell science* 108 (1995), pp. 1407–1418.

[227] Adam G Larson et al. "Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin". *Nature* 547 (2017), pp. 236–240.

[228] Robin C. Allshire and Hiten D. Madhani. "Ten principles of heterochromatin formation and function". *Nature Reviews Molecular Cell Biology* 19 (2018), pp. 229–244.

[229] Serena Sanulli, John D. Gross, and Geeta J. Narlikar. "Biophysical Properties of HP1-Mediated Heterochromatin". *Cold Spring Harbor Symposia on Quantitative Biology* 84 (2019), pp. 217–225.

[230] P.I.C. Teixeira and J.M. Tavares. "Phase behaviour of pure and mixed patchy colloids – Theory and simulation". *Current Opinion in Colloid & Interface Science* 30 (2017), pp. 16–24.

[231] Emanuela Bianchi et al. "Theoretical and numerical study of the phase diagram of patchy colloids: Ordered and disordered patch arrangements". *The Journal of chemical physics* 128 (2008).

[232] Emanuela Zaccarelli. "Colloidal gels: equilibrium and non-equilibrium routes". *Journal of Physics: Condensed Matter* 19 (2007).

[233] Diana Fusco and Patrick Charbonneau. "Crystallization of asymmetric patchy models for globular proteins in solution". *Phys. Rev. E* 88 (2013).

[234] Zhenli Zhang and Sharon C. Glotzer. "Self-Assembly of Patchy Particles". *Nano Letters* 4 (2004), pp. 1407–1413.

[235] CA Brackley. "Polymer compaction and bridging-induced clustering of protein-inspired patchy particles". *Journal of Physics: Condensed Matter* 32 (2020).

[236] Daniele Canzio, Adam Larson, and Geeta J Narlikar. "Mechanisms of functional promiscuity by HP1 proteins". *Trends in cell biology* 24 (2014), pp. 377–386.

[237] Serena Sanulli, John D Gross, and Geeta J Narlikar. "Biophysical properties of HP1-mediated heterochromatin". *Cold Spring Harbor Symposia on Quantitative Biology*. Vol. 84. Cold Spring Harbor Laboratory Press. 2019, pp. 217–225.

[238] Amarjeet Kumar and Hidetoshi Kono. "Heterochromatin protein 1 (HP1): interactions with itself and chromatin components". *Biophysical reviews* 12 (2020), pp. 387–400.

[239] Peter R Nielsen et al. "Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9". *Nature* 416 (2002), pp. 103–107.

[240] Steven A Jacobs and Sepideh Khorasanizadeh. "Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail". *Science* 295 (2002), pp. 2080–2083.

[241] Kyoko Hiragami-Hamada et al. "N-terminal phosphorylation of HP1α promotes its chromatin binding". *Molecular and cellular biology* 31 (2011), pp. 1186–1200.

[242] Tao Zhao and Joel C Eissenberg. "Phosphorylation of Heterochromatin Protein 1 by Casein Kinase II Is Required for Efficient Heterochromatin Binding inDrosophila". *Journal of Biological Chemistry* 274 (1999), pp. 15095–15100.

[243] Tao Zhao, Tomasz Heyduk, and Joel C Eissenberg. "Phosphorylation site mutations in heterochromatin protein 1 (HP1) reduce or eliminate silencing activity". *Journal of Biological Chemistry* 276 (2001), pp. 9512–9518.

[244] Erik W Martin and Alex S Holehouse. "Intrinsically disordered protein regions and phase separation: sequence determinants of assembly or lack thereof". *Emerging Topics in Life Sciences* (2020).

[245] Robin Van Der Lee et al. "Classification of intrinsically disordered regions and proteins". *Chemical reviews* 114 (2014), pp. 6589–6631.

[246] Madeline M Keenen et al. "HP1 proteins compact DNA into mechanically and positionally stable phase separated domains". *Elife* 10 (2021).

[247] Daniela Bonetti et al. "How robust is the mechanism of folding-upon-binding for an intrinsically disordered protein?" *Biophysical journal* 114 (2018), pp. 1889–1894.

[248] Abdelhamid M Azzaz et al. "Human heterochromatin protein 1α promotes nucleosome associations that drive chromatin condensation". *Journal of Biological Chemistry* 289 (2014), pp. 6850–6861.

[249] Nobuhiro Nonaka et al. "Recruitment of cohesin to heterochromatic regions by Swi6/HP1 in fission yeast". *Nature cell biology* 4 (2002), pp. 89–93.

[250] Gohei Nishibuchi et al. "N-terminal phosphorylation of HP1α increases its nucleosome-binding specificity". *Nucleic acids research* 42 (2014), pp. 12498–12511.

[251] Joel C Eissenberg and Sarah CR Elgin. "HP1a: a structural chromosomal protein regulating transcription". *Trends in Genetics* 30 (2014), pp. 103–110.

[252] Avinash Patel et al. "A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation". *Cell* 162 (2015), pp. 1066–1077.

[253] Marina Feric et al. "Coexisting liquid phases underlie nucleolar subcompartments". *Cell* 165 (2016), pp. 1686–1697.

[254] Joshua A Riback et al. "Stress-triggered phase separation is an adaptive, evolutionarily tuned response". *Cell* 168 (2017), pp. 1028–1040.

[255] Sonja Kroschwald et al. "Different material states of Pub1 condensates define distinct modes of stress adaptation and recovery". *Cell reports* 23 (2018), pp. 3327–3339.

[256] Marco Ancona and Chris A Brackley. "Simulating the chromatin mediated phase separation of model proteins with multiple domains". *arXiv preprint arXiv:2107.14518* (2021).

[257] Yanli Liu et al. "Peptide recognition by heterochromatin protein 1 (HP1) chromoshadow domains revisited: Plasticity in the pseudosymmetric histone binding site of human HP1". *Journal of Biological Chemistry* 292 (2017), pp. 5655–5664.

[258] Shinichi Machida et al. "Structural Basis of Heterochromatin Formation by Human HP1". *Molecular Cell* 69 (2018), pp. 385–397.

[259] Kyoko Hiragami-Hamada et al. "Dynamic and flexible H3K9me3 bridging via HP1β dimerization establishes a plastic state of condensed chromatin". *Nature communications* 7 (2016), pp. 1–16.

[260] Shuhei Watanabe et al. "Interactions of HP1 Bound to H3K9me3 Dinucleosome by Molecular Simulations and Biochemical Assays". *Biophysical Journal* 114 (2018), pp. 2336–2351.

[261] PIC Teixeira and JM Tavares. "Phase behaviour of pure and mixed patchy colloids—Theory and simulation". *Current Opinion in Colloid & Interface Science* 30 (2017), pp. 16–24.

[262] Robin J Speedy and Pablo G Debenedetti. "The distribution of tetravalent network glasses". *Molecular Physics* 88 (1996), pp. 1293–1316.

[263] E Zaccarelli et al. "Model for reversible colloidal gelation". *Physical review letters* 94 (2005).

[264] Emanuela Zaccarelli et al. "Gel to glass transition in simulation of a valence-limited colloidal system". *The Journal of chemical physics* 124 (2006).

[265] Diana Fusco et al. "Characterizing protein crystal contacts and their role in crystallization: rubredoxin as a case study". *Soft matter* 10 (2014), pp. 290–302.

[266] Nicoletta Gnan, Francesco Sciortino, and Emanuela Zaccarelli. "Patchy Particle Models to Understand Protein Phase Behavior". *Protein Self-Assembly*. 2019, pp. 187–208.

[267] Kenji Sugase, H. Jane Dyson, and Peter E. Wright. "Mechanism of coupled folding and binding of an intrinsically disordered protein". *Nature* 447 (2007), pp. 1021–1025.

[268] Gwen Lomberk, Lori Wallrath, and Raul Urrutia. "The heterochromatin protein 1 family". *Genome biology* 7 (2006), pp. 1–8.

[269] Natalia Vladimirova, Andrea Malagoli, and Roberto Mauri. "Diffusion-driven phase separation of deeply quenched mixtures". *Physical Review E* 58 (1998).

[270] Fabian Erdel et al. "Mouse Heterochromatin Adopts Digital Compaction States without Showing Hallmarks of HP1-Driven Liquid-Liquid Phase Separation." *Molecular cell* 78 (2020), pp. 236–249.

[271] Swetlana Jungblut, Jan-Ole Joswig, and Alexander Eychmüller. "Diffusion- and reaction-limited cluster aggregation revisited". *Physical Chemistry Chemical Physics* 21 (2019), pp. 5723–5729.

[272] Hua Wu, Marco Lattuada, and Massimo Morbidelli. "Dependence of fractal dimension of DLCA clusters on size of primary particles". *Advances in Colloid and Interface Science* 195-196 (2013), pp. 41–49.

[273] Debra J. Audus, Francis W. Starr, and Jack F. Douglas. "Valence, loop formation and universality in self-assembling patchy particles". *Soft Matter* 14 (2018), pp. 1622–1630.

[274] Chase P. Broedersz et al. "Condensation and localization of the partitioning protein ParB on the bacterial chromosome". *Proceedings of the National Academy of Sciences* 111 (2014), pp. 8809–8814.

[275] Joshua A. Riback et al. "Composition-dependent thermodynamics of intracellular phase separation". *Nature* 581 (2020), pp. 209–214.

[276] Iain Mathieson. "The omnigenic model and polygenic prediction of complex traits". *The American Journal of Human Genetics* 108 (2021), pp. 1558–1563.

[277] CA Brackley et al. "Complex small-world regulatory networks emerge from the 3D organisation of the human genome". *Nature communications* 12 (2021), pp. 1–14.