Painters in Chromatin : Theoretical models for 3D propagation of epigenetic marks

Amith Zafal Abdulla

Numéro National de Thèse : 2022ENSL0044

THESE

en vue de l'obtention du grade de Docteur, délivré par

l'ECOLE NORMALE SUPERIEURE DE LYON

Ecole Doctorale N° 52
Physique et Astrophysique de Lyon (PHAST)

Discipline : Physique

Soutenue publiquement le 02/12/2022, par :

Amith Zafal ABDULLA

Painters in chromatin: Theoretical models for 3D propagation of epigenetic marks

Peintres en chromatine : Modèles théoriques pour la propagation 3D des marques épigénétiques

Devant le jury composé de :

BARBI, Maria	Professeure Des Universités Sorbonne Université	Rapporteure
HOWARD, Martin	Professeur John Innes Center, UK	Rapporteur
CORPET, Armelle	Maitre De Conférences Institut NeuroMyogene	Examinatrice
THON, Genevieve	Professeur Univerisity of Copenhagen	Examinatrice
VICTOR, Jean-Marc	Directeur De Recherche Sorbonne Université	Examinateur
VAILLANT, Cedric	Directeur De Recherche ENS de Lyon	Directeur de thèse
JOST, Daniel	Directeur De Recherche ENS de Lyon	Co-directeur de thèse

I would like to dedicate this thesis to uppa, umma, ikka, itha and jibruu …

.
.
.
.
.
.
.

It was a dreary December morning in Lyon, and it started to drizzle as if nature couldn't express enough. I was walking up Avenue Debourg, and followed the the scent of roasted coffee beans to reach the 'La Cantina' café. "Bonjourno! my friend. Shall I make you the best coffee in the world?" shouted Francesco, the friendly barista with a wide smile. The smile felt contagious and they made me widen my dry lips; and return the gesture. As Francesco started making 'café au lait', I sat down looking at the orange and yellow hues, of a Claude Monet painting, attached to the wall behind him. Francesco started scooping crema into my coffee cup, detailing his patterns of love and laughter. The patterns gently drifted to swaying trees, to boats that have survived many storms, and to eager faces - all of which he craftfully painted on top of my coffee. They were all different patterns, but the same coffee, subtly changing its expression.

Different epigenetic patterns, but the very same DNA, artfully changing gene expression. With art, a nice coffee and a friend, I left with a smile to write this thesis; modelling propagation of epigenetic marks . . .

Acknowledgements

Tonight, the night of September the 20th, 2022 beholding a beautiful cityscape of Lyon (‘belle vue’ it is), I am writing this thesis thanks to the love, care and support of many. I fondly remember and embellish a few here ...

First and foremost, I am deeply grateful to my supervisors Daniel Jost and Cedric Vaillant for the great learning experience and for introducing me to the exciting world of 3D genome organization and epigenetics. In helping me grow, imparting the way of doing science and a numerous number of little things, down to working on weekends because of my woeful organization. I must appreciate their open-mindedness in gracefully receiving my stupid ideas, and providing much needed suggestions in working through the course of this thesis.

Maxime, for helping me settle in France, and on a lot things of other things: from taxes to liquid-liquid phase separation. It was a great trip to Annecy.. Hossein, for the numerous discussions on experimental techniques, HiC and life in France. Our weekend explorations of Lyon with Masha and Caroline were a lot of fun. I’d like to thank Jean-Michel for the discussions and suggestions. Dario, plenty of discussions, fond memories of late working hours, “Italian dinner”, football, badminton and for the “rescue plan” thanks a lot my friend. My sincere thanks to Nathan and Amir for the wonderful company. Genevieve, for the insightful discussions, St. Georges, drive to Millau, and many more, I am extremely thankful to you. And of course, for the mangoes!

I would like to extend my thanks to our collaborators in Montpellier, Giacomo Cavalli and Thierry Cheutin. Thank you Thierry especially for the unpublished data and figure.

My mom, who would patiently listen to all my epigenetic troubles and my father who’d still sing me the very same song that had put me to bed 20 years ago, just to put a smile on my face, for spending a whole month in Lyon with me while I was writing this thesis.

On a lot of my sleepless nights, my brother would stay awake on the phone keeping me company, my sister-in-law who would affectionately enquire “Enth patti Unni?”. Sammy, who in the middle of an insane amount of work would be ready for any co-op game. I am at a loss for words to express my gratitude.

Pour ma Tufafu, as always I’ll leave the space blank, for there is a lot. Thanks a lot my dear for the countless number of things. Thank you Claire for the beautiful Christmas memories.

Sreelekshmi, for being a constant pillar of support, for sitting alongside whenever I was down, I am deeply grateful to you. Isabelle, Julie, Emmanuelle, Corrine, Eric, Ali, Elham, Hungyo, Pascal, Camille, Elias, Lavi, Praneet, Sarath, Meenu and all the wonderful people at LBMC and physics, I sincerely appreciate the support.

I arrived in France three years ago and was hosted by Emmanuelle and Aldo at their rental, now its a place I’d go to when I miss home. I am extremely grateful for the warmth and care. Hamza, for the gaming nights and the stress busting support, Aakanksha, for the delicious food, thanks a lot. Francessco, Clara at the boulangerie, the people at Crous restaurant and the numerous smiles that kept me going, I am deeply grateful . . .

Abstract

A multitude of stable and heritable phenotypes arise from the same DNA sequence, owing to epigenetic regulatory mechanisms relying on the molecular cooperativity of "reader-writer" histone modifying enzymes. In this thesis, we focus on the fundamental mechanisms behind epigenome regulation and memory encoded by post-translational modifications of histone tails. Our aim is to develop general mechanistic frameworks of epigenomic regulation and memory integrating key biochemical and physical processes in order to investigate how such generic principles can be contextualized to specific biological systems.

After a review of the existing theoretical models (Chapter 1), we introduce a unified modeling framework, the "Painter model", describing the mechanistic interplay between sequence-specific recruitment of chromatin regulators, chromatin-state-specific reader-writer processes and long-range spreading mechanisms (Chapter 2). A systematic analysis of the model building blocks highlights the crucial impact of tridimensional chromatin organization and state-specific recruitment of enzymes on the stability of epigenomic domains and on gene expression. In particular, we show that enhanced 3D compaction of the genome and enzyme limitation facilitate the formation of ultra-stable, confined chromatin domains. The model also captures how chromatin state dynamics impact the intrinsic transcriptional properties of the region, slower kinetics leading to noisier expression. We apply our framework to analyze experimental data, from the propagation of γH2AX around DNA breaks in human cells to the maintenance of heterochromatin in fission yeast, illustrating how the painter model can be used to extract quantitative information on epigenomic molecular processes.

To go beyond the effective 3D description, we study explicit 3D polymer dynamics. Specifically, to investigate the caveats of simulating genes or regions of interest which are usually much smaller compared to the full length of a chromosome. Since the physics of long, topologically-constrained polymers may significantly deviate from those of shorter chains, we theoretically investigate the extent of the minimal genomic region that one should explicitly consider around a given locus in order to effectively capture the correct dynamical and structural properties of the domain of interest (Chapter 3). We show that this minimal

size depends on the overall epigenomic context and on the entanglement properties of the long polymer.

Finally, in Chapter 4, we present the ongoing work, integrating the painter model with explicit 3D polymer dynamics. We introduce a theoretical framework coupling 3D polymer dynamics and epigeneome regulation by diffusing HMEs, the “Living painter” model, that exhibits intriguing properties on the coupling between 3D genome folding and epigenetic spreading, reflecting the scope and extension of the thesis.

Table of contents

Chapter 1

General introduction on 4D Epigenomics: theoretical approaches

1.1 Biological context: Epigenetics and 3D genome organisation

The ability of organisms to precisely regulate gene expression is central to their development. Proper temporal and spatial expressions of genes in eukaryotes require activation of transcription during the appropriate developmental stages. In response to environmental and developmental cues, cells can adopt different gene expression patterns and differentiate into a variety of cell types. Once established, this pattern is frequently maintained over several cell divisions despite the fact that the initiating signal is no longer present or significantly weakened. This capacity of translating transient external stimuli into diverse and stable phenotypes without alteration of the genomic sequence is at the heart of the "epigenetic" regulation of gene expression [1]. Epigenetic processes are involved in the control of somatic inheritance and in the maintenance of cellular identity as well as in the transgenerational inheritance of traits by transmission via the germline [2].

A major class of mechanisms driving such epigenetic "memory" relies on chromatin-based processes regulating gene expression by the control of the local biochemical and structural properties of chromatin leading to different chromatin states more or less permissive to transcription [3]. Such control is mediated in part by biochemical modifications (Fig.1.1) of the DNA or of histone tails, the so-called epigenomic marks. The mechanisms of establishment and heritability of these chromatin states, either active or repressive, are governed by similar general rules involving the combined and self-reinforcing action

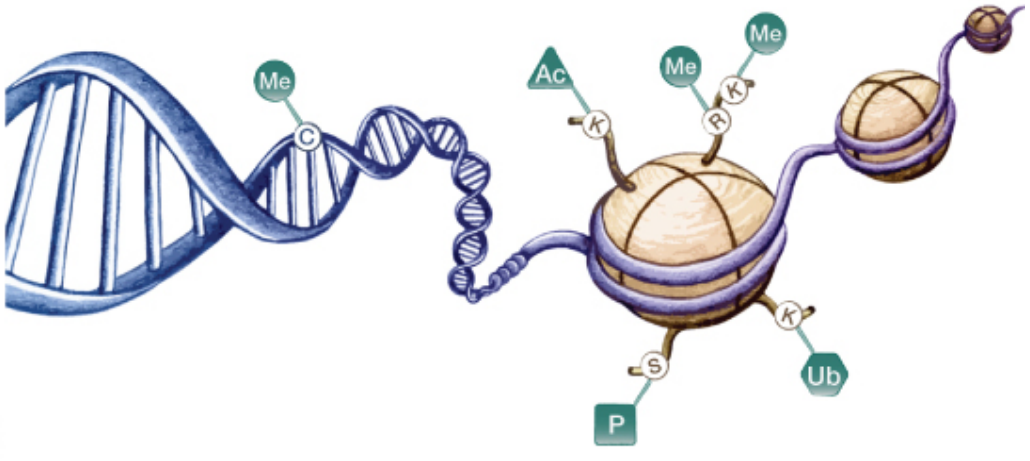

Fig. 1.1 The so-called epigenomic marks: biochemical modifications of the DNA (eg, methylation Me) or of histone tails (eg, methylation, acetylation Ac, phosphorylation P). *Illustration from france.promega.com.*

of specific chromatin-binding proteins that add or remove epigenomic marks [4–6] (Fig.1.2).

De novo assembly first proceeds by a nucleation ("forcing") stage via the targeting of specific enzymes at dedicated regulatory sequences by either DNA binding proteins or the RNAi-based pathways [4, 7–11]. Such nucleation elements are often composed by multiple binding sites for different DNA binding proteins that associate to their cognate sequence to stably recruit chromatin regulators. For example, in Drosophila, the Polycomb-based epigenetic repression of developmental genes rely on the targeting of the histone modifying enzymes (HMEs) PRC1 (monoubiquitination of H2AK118) and PRC2 (methylation of H3K27) complexes at specific "silencers" regions, the so-called Polycomb Response Elements, characterized by various combinations of binding sites for adaptor proteins (e.g. Gaga, Zeste, Pho) [12, 13]. Similarly, gene activation relies on the recruitment by Trithorax-group proteins (e.g. MLL2) of acetyltransferases (e.g. p300) and demethylases (e.g. UTX) at gene promoters and enhancers [14, 15].

Once initiated, the state is able to spread to the neighboring sequences and to form a stable chromatin domain [10] that can further propagate through replication and mitosis [16]. The ubiquitous ability of some chromatin regulators to be recruited by (e.g. Clr4) or to have a boosted activity in presence of (e.g. PRC2) the chromatin state they catalyze ("reading"capacity) and to spread this state to the neighboring sequences ("writing" capacity) introduce an effective positive feedback which is believed to be a key ingredient of epigenetic maintenance [10, 16, 17].

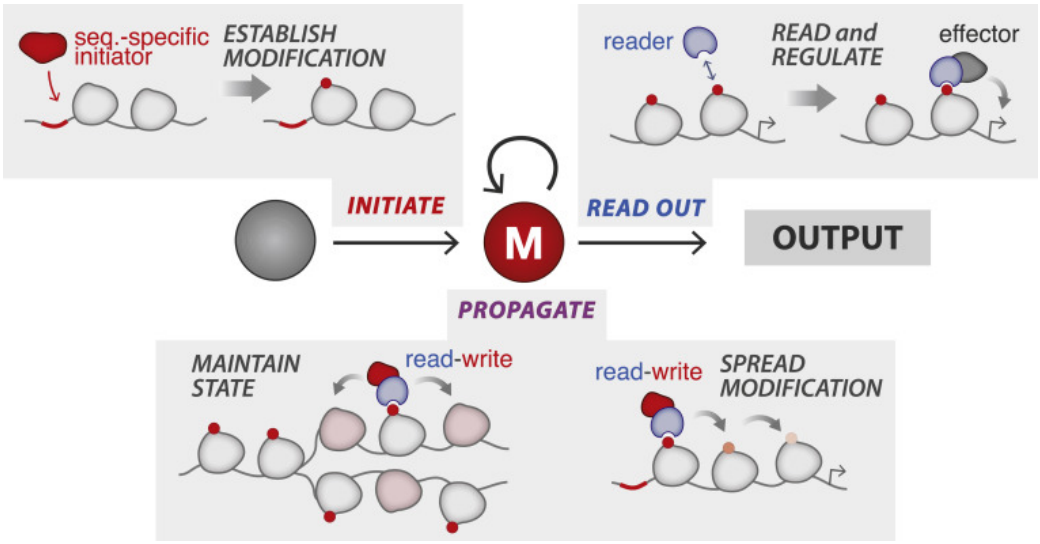

Fig. 1.2 General rules of epigenome assembly and regulation: initiation by sequence-specific factors and spreading and maintenance by reader/writer enzymes. *Adapted from [18].*

Such “reader-writer" principle of chromatin enzymes and its impact on epigenetic memory has been already investigated using simple mathematical models of chromatin state regulation with a focus on the dynamics of histone marks mediated by HMEs [11, 19–30] (see Sec. 1.2). In particular, in their seminal work, Dodd *et al.* [19] suggested that the maintenance of stable, extended active or repressed chromatin domains over generations, even in absence of nucleation signals, is made possible by the reader-writer property of HMEs coupled to their capacity to spread (or write) a mark at long-range along the genome. Indeed, such interplay leads to cooperativity and allows the effective formation of a large reservoir of modified nucleosomes to serve as templates to ensure full recovery after random perturbations such as transcription- [27, 29] or replication-mediated [23, 54] histone turnover.

Actually, the long-range spreading property reflects the polymeric nature of the genome that can bring in close spatial proximity two distant loci. Many experimental and theoretical studies have highlighted the correlation between spatial chromosome organisation and chromatin regulation [56, 55, 62–67] (see also Sec.1.3). Indeed, 3D chromatin domains, like A/B compartments and topologically-associated domains (TADs) observed in Hi-C experiments (Fig.1.3A) or chromocenters (Fig.1.3C), Polycomb foci and transcription factories observed by microscopy, are strongly associated with epigenomics [63, 62, 64], the formation of these 3D structures being often driven by architectural proteins (eg, HP1, PRC1) that binds specifically to given epigenomic marks [56, 142, 246, 57]. For example, in Drosophila, loci in the same TAD usually share the same epigenomic content and TADs

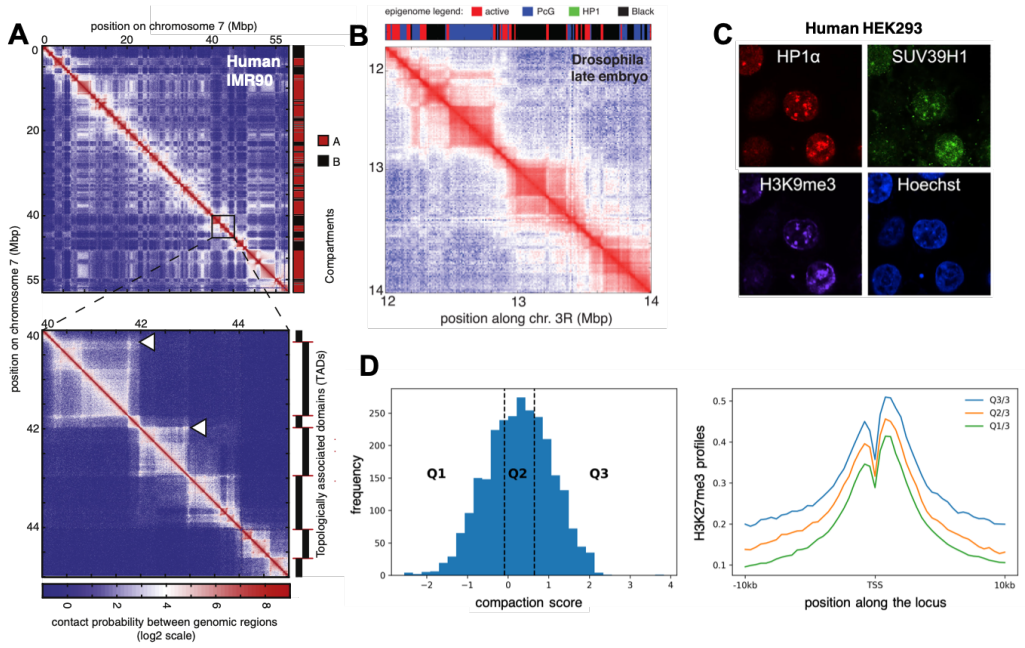

Fig. 1.3 Epigenome folding. (A) Experimental Hi-C map of chromosome 7 of the human cell line IMR90 [63]. On the right of the top panel: A/B compartments. On the right of the bottom panel: TAD segmentation. *Adapted from [72].* (B) Hi-C data of a 2Mbp-long region of chromosome arm 3R [55] (Drosophila, late embryo). On the top: local epigenomic state (see color legend) as obtained by Fillion et al [71] for the embryonic cell line Kc167. *Adapted from [204].* (C) Visualization in human cell line HEK293 of key components (histone mark: H3K9me3, HME: methyltransferase Suv39H1, architectural protein: HP1, DNA: Hoechst) of constitutive heterochromatin forming compact nuclear bodies (chromocenters). *Adapted from [57].* (D) Relation between compaction and H3K27me3 profiles in mouse embryonic stem cells. (Left) Genes targeted by Polycomb complexes are classified into 3 categories depending on the density of 3D contacts around the gene. (Right) Average H3K27me3 profiles around TSS for each categories. The more compact the gene is the more extended the profile is. *Adapted from [30].*

sharing the same epigenomic state tend to significantly contact at long-range [55, 64] (Fig.1.3B). Similarly, in mammals, loci having the same epigenetic marks tend to be more often colocalized [63, 62]. Interestingly, genomic regions localized inside compact 3D domains (i.e. domains with high numbers of self-contacts) exhibit epigenetic profiles that are more spread than loci inside weakly compacted compartments (Fig.1.3D).

All this suggests that there exists a coupling between the 3D chromosome organization and the epigenome regulation: genome folding may impact the long-range spreading of an epigenomic mark and a given epigenomic mark may impact genome folding via the recruitment of architectural proteins. This potential feedback loop between epigenomics and 3D genome strengthens the hypothesis that 3D genome folding is key to chromatin states

assembly and maintenance. In the following, we review the various theoretical approaches that have been developed to study such interplay between epigenomics and the 3D genome.

1.2 Theoretical Models of chromatin state spreading and maintenance

The first theoretical works studying the dynamics of chromatin states and formalizing their ability to carry epigenetic information (i.e. memory) were conducted by Sneppen's and Senguta's groups in 2007 [19, 47]. Since this pioneering work, many other contributions from these groups and other researchers have been proposed, either as complements, refinements/improvements, or as new theoretical frameworks and applications to other systems [11, 19, 20, 24, 21–23, 25–30]. This section will be devoted to a brief description of these various models.

As we have already pointed out, by definition, "epigenetic memory" requires multistability, i.e. the existence of stable alternative "phenotypes" that can be maintained across cell divisions in the absence of any external stimulus. For gene expression, this means different patterns of gene expression ("on"/"off") that can be maintained across many cell divisions. On the one hand, this can be achieved by autocatalytic transcriptional loops of transcription factors ("cytoplasmic" epigenetics), like in the cell type switching in *C. albicans* [45]. But, on the other hand, it has been clearly demonstrated that chromatin states, and in particular histone modifications, are associated with transcriptional activity, and that these modifications can be partially transmitted through replication, suggesting that they may also contribute to epigenetic regulation ("chromatin" epigenetics). Dodd and Sengupta's theoretical works aim at demonstrating that chromatin modifications, as a proxy for chromatin states, can indeed provide such stable memory. To this end, they proposed a minimal theoretical framework based on biological evidence about the molecular processes at work in the regulation of histone modifications. This is a very general framework that we will essentially adopt in this manuscript for the "Painter model" in Chapter 2.

1.2.1 Silencing dynamics in *S. pombe*: the H3K9me2/3 system

In their seminal paper entitled "Theoretical Analysis of Epigenetic Cell Memory by Nucleosome Modification", Dodd *et al.* focused on the dynamics of histone modifications associated with the "active" and "silenced"states in *S. pombe*. Silencing in *S. pombe* is controlled by the combined action of H3/H4 deacetylases (Clr3,Sir2), H3K9me2/3 methyl-transferase (Clr4), nucleosome remodelling factor (Mit1) and the HP1 (Swi6,Chp2) association with marked H3K9me2/3 nucleosomes, that promote the formation of a local

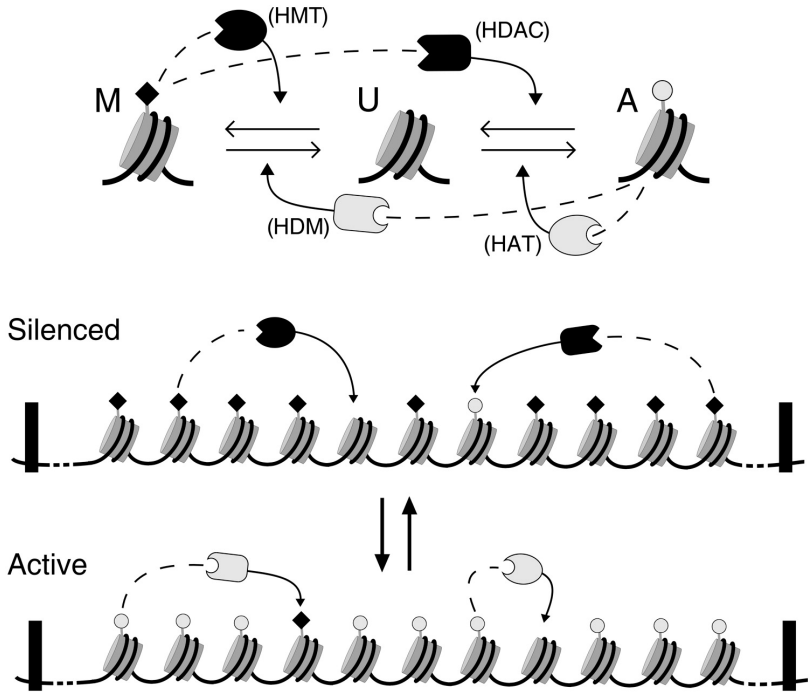

Fig. 1.4 Basic ingredients of the 3-state model: The three relevant nucleosome types — methylated (M, marked by a black diamond), unmodified (U), or acetylated (A, marked by a gray circle)—can be interconverted by recruitment of histone-modifying enzymes by nearby M or A nucleosomes (dotted lines) or by random "noisy" transitions. HMT indicates histone methyltransferases; HAT indicates histone acetyltransferases; HDM indicates histone demethylases; and HDAC indicates histone deacetylases. Note that within the DNA region delimited by the boundary elements (black rectangles) any nucleosome can stimulate the modification of any other (spreading *in trans*). *Figure and legend from [19] (Figure 1).*

repressive heterochromatin state. These silencing factors/effectors are primarily targeted to genome at specific sites via different pathways (RNAi, DNA binding proteins) from where spreading to neighboring sequences has been proposed to occur. Such heterochromatin is constitutively located at pericentromeric DNA repeats, sub-telomeric regions, and the mating type (mat) locus. Let us note that H3K9me2/3-HP1 is actually a well conserved chromatin-based silencing molecular system that is mostly targeted to repeated regions in most organisms, being often associated to/complemented by DNA methylation in higher organisms. The motivation and starting point of Dodd *et al.* analysis are based on the observation that, in a mutant strain of *S. Pombe* ($K\Delta :: ura4+$), mating type genes and a reporter gene ($ura4+$) inserted in the Mating type locus in the place of a nucleation sequence exhibit stochastic switching between very stable and heritable expression states : in the WT strain, this 20 kbp locus is constitutively silenced (repressed) by the H3K9me2/3/HP1 heterochromatin state; however when replacing the main heterochromatin nucleation sequence ("silencers") with the reporter genes, stochastic switching is observed. To account for this apparent epigenetic "bistability", for sake of simplicity, they introduce a minimal three-state model at nucleosome resolution, chromatin being modelled here as an array of fixed nucleosomes: an Acetylated state associated with the "active" state, a Methylated state (H3K9me2/3) associated with the "repressed" heterochromatin state, and a third "unmodified" state (Fig. 1.4). The acetylated state may represent a single type of histone acetylation (e.g. H3K9ac, H4K16ac) or a clustered state of various histone acetylations; similarly, the "methylated" state may be considered as a clustered state of both me2/me3 methylations on both histone tails. As the authors point out, the important point here is to consider two antagonistic states: H3K9 methylations can only be deposited by HMT on a non-acetylated nucleosome substrate and conversely methylation prevents any histone acetylation by HAT. Molecularly, conversion between the M state to the A state is (at least) a two step process that involve first deacetylation and then methylation; reversely, conversion of a M to a A state occurs via a demethylation step followed by an acetylation step. This a priori, justifies the introduction of an intermediate unmodified state. Again, the results obtained by Dodd *et al.* can be generalized to any three state model with two antagonistic states (see also [22] or [23])), whatever the precise and molecular definition of theses states.

Chromatin is modelled as an array of fixed nucleosomes, whose position along the array defines the position along the genomic domain (coarse-grained description of the genome at nucleosome resolution). For each nucleosome, the stochastic switching between the different states is controlled by conversion rates (Fig. 1.4): the action of the HME is implicitly taken into account by these conversion rates. To set these conversion rates, they distinguish between "random" and "recruited" conversion: (1) random conversion represents "noise" and accounts for the leaky activity of enzymes as well as histone turnover (for A, M to U conversions); (2) "recruited" conversion represents "positive feedback" and

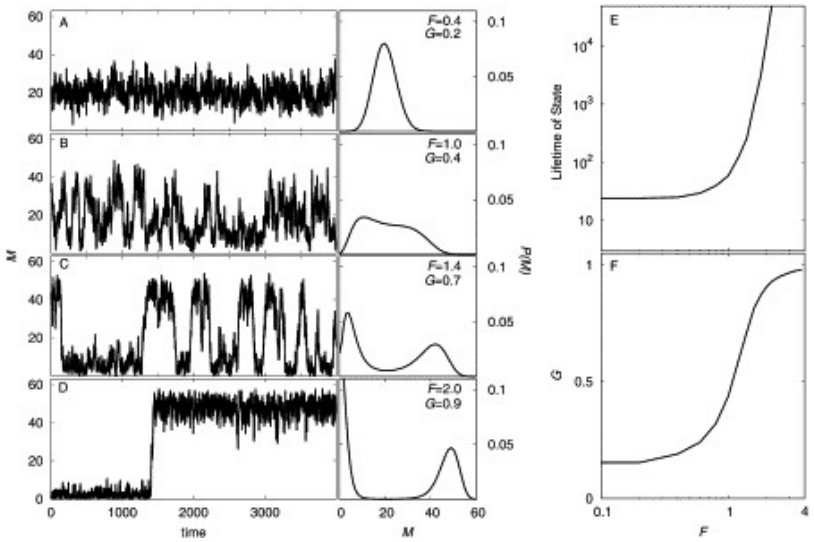

Fig. 1.5 (A–D) The left panels show samples of the time development of the number of M nucleosomes, M, over a range of feedback-to-noise ratios $F = 0.4, 1, 1.4$, or 2 ($F = r$ in the maintext). The total number of nucleosomes in the system is 60. Time is measured as average attempted conversions per nucleosome. The right panels show the corresponding probability distributions of M obtained from long simulations. (E) Relationship between F and the average length of time for which the system remains continuously in one or the other state. The high-M state is defined as $M > A$, and the high-A state as $A > M$. Transition to the high-M state is scored when $M > 1.5A$ and to the high-A state when $A > 1.5M$. (F) Relationship between F and the average "gap" between the numbers of M and A nucleosomes at any time point, $G = Average(|M-A|/|M+A|)$. *Figure and legend from [19] (Figure 2).*

accounts for the ability of enzymes to be recruited to a modified nucleosome (through specific association domains) and to promote the conversion of a second nucleosome to that modified state: for example, an M-state nucleosome can promote the conversion of another nucleosome's state from A to U or from U to M. It is assumed that these rates are the same along the network. The time evolution of the statistical properties of the system, i.e. the probability of having a given distribution of states, is formally described by a master equation [52] which is, in most cases, difficult to treat analytically but which can be solved "numerically" using kinetic Monte Carlo approaches. Importantly, to really investigate epigenetic memory (i.e. stability across cell division), they also account for the replication process where, in a semi-conservative manner, each A and M state is replaced by a U state with probability $1/2$ (this leads to a 2-fold dilution of total M and A nucleosomes). Using such a numerical scheme, from a set of stochastic trajectories (i.e. the dynamics of stochastic states along the nucleosome array), Dodd *et al.* derived statistical properties of the systems such as the statistical distribution of the fraction of nucleosomes in the M and A states ($M = N_M/N, A = N_A/N$; as well as the statistical distribution of $m = M - A$) as a function of time and as a function of the control parameters, namely the conversion rates. In the simplest "symmetric" case, where all "random" conversion rates are equal (k_o) and all "recruited" conversion rates are also equal (ε), there remains one control parameter, namely the ratio between the recruited and random rates $r = \varepsilon/k_o$. In the stationary regime, they show that the system undergoes a phase transition at a critical value r_c, from a non-coherent phase at a lower ratio $r < r_c$, where "noise" dominates and the macrostates A and M are unstable (Fig. 1.5 A) (the statistical distribution of $m = M - A = (N_M - N_A)/N$, $P(m)$ is centered on the zero value (Fig. 1.6(a)): on an average there is always the same number of M and A nucleosomes) to a coherent phase of bistability where the positive feedback dominates (Fig. 1.5 D), characterised by stable macrostates A and M for $r > r_c$ (the statistical distribution of $P(m)$ is now bimodal and peaks at the 2 coherent finite values $\pm m_*$ (Fig. 1.6(a)): with $m^* > 0$ (resp < 0) corresponding to a rather coherent M (resp A) macrostate). As pointed out by D. Jost in [22], this is a second-order phase transition similar to that characterising the 1D Ising chain with ferromagnetic coupling (i.e. coupling favouring the same spin orientation). Using a mean-field approach that essentially converts the stochastic master equations into a set of deterministic kinetic equations, D. Jost has indeed analytically derived the bifurcation diagram associated with this symmetric 3-state dynamics and in particular the value $r_c = 3$ of critical point. Using a Fokker-Planck approach, he obtained an analytical approximation of $P(m)$ that reproduces well the exact distributions (obtained by kinetic MC approaches), confirming in particular the results by Dodd and Sneppen. He also obtained an analytical expression for the stability of the macrostates (A or M) in terms of the mean first passage time $< \tau >$ showing that this stability, beyond the critical point r_c, increases exponentially with the strength of positive feedback r and the size

of the array (total number of nucleosomes) (Fig. 1.6(b,c)): the more nucleosomes are in the system, the more stable are the macrostates (thanks to cooperativity).

Importantly, Dodd *et al.* show that, in such 3-states model, strong bistability definitively requires an implicit or explicit cooperativity, where implicit cooperativity naturally emerges from the two-step recruited conversion (M to A requires two recruited conversions) (Figure 3 of [19]). They also address the crucial role of conversion at "long-range" (or spreading *in trans*) in the emergence of such bistability: indeed when only considering a Nearest-Neighbor (NN) model (spreading *in cis*) bistability is lost. Actually following the analogy with the 1D Ising chain in statistical physics, it has been demonstrated that for such 1D system, phase transition is only possible when considering long-range "interactions" between spins, the longer the range the sharper the transition. The longest range is achieved when obviously one considers that every nucleosome can be converted by recruitment of HME at any other nucleosome (infinite range), which is the situation primarily considered in [19, 22]. However, as we shall see below, considering a more realistic long-range spreading that accounts for chromatin looping (i.e., accounts for the polymeric nature of the chromatin) still maintains bistability and is a first step towards a better understanding of the coupling between 3D organisation, chromatin-state dynamics and memory.

1.2.2 Silencing dynamics in *S. Cerevisiae*: the SIR system

Along the same line, in their first article "Epigenetic chromatin silencing: bistability and front propagation" and subsequent related studies [46, 48, 21, 49], Sengupta's group introduced a theoretical model of silencing in yeast (*S. cerevisiae*). Such silencing is achieved by the combined action of deacetylation and binding of Sir3 proteins (an architectural protein that associates with nucleosome and that can oligomerize with adjacent sir3 proteins) and occurs constitutively at all sub-telomeres, as well as at the two mating loci HML and HMR [53]. Interestingly, deacetylation is performed by the Sir2 HDAC that coassociates with Sir3 on chromatin (on nucleosome, via Sir4), introducing a putative positive feedback in this silencing dynamics. Considering a minimal two-state model, the Acetylated (A, eg H4K16ac) versus the Sir-bound (S) state, they analysed the dynamical properties of a (fixed-positions) nucleosome array using a mean field treatment of the master equation. This leads to a deterministic kinetic (chemical) equation describing the time evolution of the average state "level" at each nucleosome position (ie, average acetylation level $\overline{A}$ and average Sir occupation $\overline{S}$), as a function of acetylation and Sir binding rates. By solving the stationary solutions and analyzing their stabilities, they derived bifurcation diagrams indicating transitions curves and points delineating mono-stable (silent, non-silent) and bi-stable regions in the 2-parameter space. As in the case of the three-state model investigated by D. Jost using similar mean field approach [22], this bifurcation diagrams present

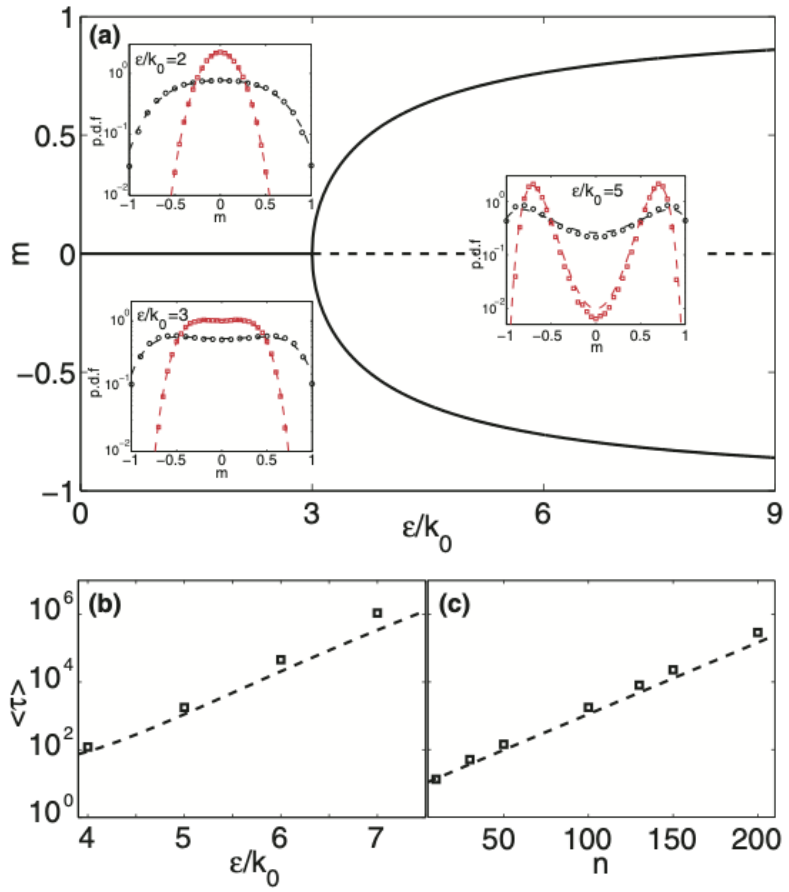

Fig. 1.6 3 states model (Fig. 1.4), symmetric regime. (a) Bifurcation diagram for $m = I - A$ (here I stands for Inactive and correspond to the M Dodds state) as a function of $r = \varepsilon/k_0$. Full (dashed) lines represent stable (unstable) fixed points of the dynamical system. Insets: Probability distribution functions (p.d.f.) of m for $n = 10$ (black circles) or $n = 100$ (red squares) computed from the Fokker-Planck approximation (dashed lines) or from stochastic simulations (squares). (b),(c) Mean first passage time $< \tau >$ (in k_o^{-1} units) to switch from m_A to m_I as a function of ε [(b) for $n = 100$] or of n [(c) for $\varepsilon/k_0 = 5$], computed from Eq. (9) [22] (dashed lines) or from simulations (squares). Standard errors on the estimation of $< \tau >$ are smaller than the symbol size. *Figure and legend from [22] (Figure 1).*

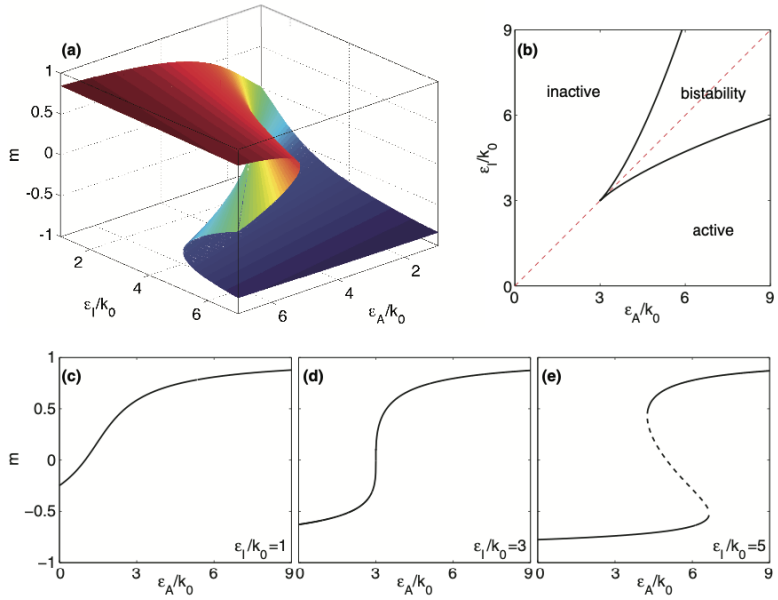

Fig. 1.7 3 states model (Fig. 1.4), asymmetric regime. (a) Cusp catastrophe surface representing the fixed points of the dynamical system as a function of ε_A/k_0 and ε_I/k_0. (b) Stability diagram and boundaries between the mono- and bistable regions. (c)–(e) Bifurcation diagrams for m as a function of ε_A/k_0 for fixed values of ε_I/k_0 [(c) = 1, (d) = 3, (e) = 5]. *Figure and legend from [22] (Figure 2).*

a "cusp catastrophe" surface (Fig. 1.7(a)) with bistability (Fig. 1.7(b,e)) and hysteresis (Fig. 1.7(e)) , which is indeed a common feature of dynamical systems with asymmetries in the control parameters values (here asymmetry between acetylation rate and Sir-binding rate values). As for Dodd *et al.* [19] and Jost [22], their analysis highlights the role of positive feedback and nonlinearity in generating bistability and, in a later study, inheritance [46]: here, cooperativity comes from Sir-bound state dependent deacetylation and non linearity is introduced *ad hoc* in the conversion rates (As shown in Dodd *et al.*, a two state model need explicit "non linearity"). In addition, in their first study and in subsequent article, they pinpointed and investigated the possible effect of SIR titration on the silencing dynamics and in particular on targeting and confining stable SIR silencing to finite size domains [21]. As we shall see in Chapter 2, in agreement with Sengupta's work, such titration in silencing factors is likely to be a major driving force in epigenetic inheritance of confined chromatin state domains.

Theoretical modelling of SIR mediated Silencing in yeast has been also carried out by Dodd's and Sneppen's group using similar approach as for *S. Pombe* (see above) [20, 31]. In [20], they carefully investigate the property of the (A,S) 2-states model (like in Sedighi's study [47]) by studying different conversion rules and including the effect of both nucleation sites (silencers) and barriers. They show, in particular, that bistability is still achieved if one

restrains long-range conversion only to Sir-mediated deacetylation (Acetylation remaining NN). They show also that the formation and maintenance of 1D confined domains of Sir-silencing (like the HMR and HML loci) can be obtained by the combination of (i) locally acting "silencers" (introduced via NN A to S conversion at specific locations), (ii) weak recruited conversion, (iii) presence of quite numerous ($1/10$ nucleosomes) barrier elements or "anti-silencer" (introduced via NN S to A conversion) in the flanking regions and (iv) global SIR titration (introduced via S state saturation) condition. In a more recent paper [31], they consider a more complex dynamics (as Sengupta's group in [49]) with additional states such as H3K79me (an antisilencing mark that prevents Sir3 nucleosome association and thus limits S state spreading in the model) and with SIR-SIR cooperativity between neighboring nucleosomes that strenghtens positive feedback in the Sir dependent deacetylation process.

1.2.3 PcG mediated silencing dynamics in plant, drosophila and mouse

In continuity to these pioneering works on silencing dynamics in *S. pombe* and *S. cerevisae*, the group of Martin Howard developed few years later a quantitative 3-state model of chromatin-based gene repression in the context of vernalization in flowering plants [11, 33, 27], in close collaboration with the Caroline Dean experimental group. Vernalization, by which plants perceive and retain a memory of winter that allows them to germinate or flower in spring, is a classic epigenetic process. In *Arabidopsis thaliana*, it involves Polycomb-based silencing of the floral repressor FLC during cold periods. FLC then generates stable silencing when temperatures rise: the repression occurs via (is accompanied by) the deposition of the H3K27me2/3 histone modification by the HMT complex PRC2 which is part of the Polycomb Group (PcG) proteins [17, 35]. "PcG" mediated gene repression is a very common and ancient mechanism of silencing in eukaryotes (unicellular and multicellular animals and plants), often targeted at developmentally regulated genes [39, 40]. It is involved in many processes such as chromosome X inactivation in mammals, Hox genes repression, stem cell plasticity. Interestingly, it has been shown in some species that PRC2 can also be targeted to and silence transposable elements (repeated sequences), which is generally constitutively performed by the H3K9me2/3 - mC heterochromatin system [38].

Using a similar three state model as in Dodd *et al.*, with M representing H3K27me3, and assuming a similar two-step conversion dynamics with recruited conversion (acting at "long-range": infinite range model) and noise, they build a quantitative model by reproducing H3K27me3 Chip data averaged along the FLC locus at different time along the cold-warm period and for different durations of the cold exposure period. They show that the observed H3K27me3 level can be accounted by considering, in addition to the M-state "recruited" conversion process, a site specific nucleation process of the M-state. During the cold, silencing is mostly nucleated and accumulated at this specific locus and then spreads toward

the FLC gene after return to the warm thanks to the M-state recruited conversion. Their results suggested that cells are in a bistable regime and that such localized nucleation drive stochastic switching toward a global M-state over the entire locus. Subsequent detailed experimental and theoretical works from the same group have further revealed a more complex (fine grained) picture of this epigenetic switch, with actually different PRC2 co-associated cofactors in the nucleation (cold) vs spreading process (warm) [36, 32].

Following this work on vernalization, Howard's group then developed a generic model of gene silencing by PcG that incorporates the antagonistic effect of transcription in the dynamics of silencing [27]. They considered a multi-state model (Fig. 1.8) with states representing different levels of methylation, from no methylation to trimethylated H3K27me3 states. The successive methylation steps result from random and long-range recruited conversions (only from trimethylated state, the main substrate of the "reader-writer" PRC2) whereas demethylation results from noise and local transcriptional activity (demethylation and histone exchange). On the one hand transcription is activated by trans-acting factor (Transcription factors) and on the other hand, transcription is repressed by the di- and trimethylated states. They show that such dynamics can lead to bistability, with two alternative states, the repressed state (high me2/3) and the active state (low me2/3). Interestingly, they further show that the slow dynamics of methylation enhance robustness to noise, in particular by filtering the fast perturbation of trans acting factors activity. Note that such interplay between transcription dynamics and chromatin state dynamics will be also discussed in the painter model described in Chapter 2. Recently, the Howard group, in collaboration with Margueron experimental team, applied such a framework to characterize epigenetic memory in various mouse cell lines [29]. Experimentally, after a long inhibition of PRC2 enzymatic activity leading to the loss of H3K27me3 marks and the activation of many PcG-target genes, they reintroduced PRC2 and monitored if an epigenetic memory of the inhibition was present after the introduction of PRC2. They observed that, in some cell lines, there exists a set of genes, initially repressed by the PcG machinery before the perturbation, that remain active after the reintroduction of PRC2. Application of the Howard group model to the data suggests that such a memory arises from the double-negative feedback between Polycomb-mediated silencing and active transcription that is posited in the model (Fig. 1.8) and that may lead to bistability. During the perturbation, some genes that were initially in the repressed state of bistability, are now highly expressed. When the perturbation is removed, they then fall/remain in the active state of bistability, driven by the inherent hysteresis of bistable system (Figs.1.7). Interestingly, embryonic cell lines (mESC) which are usually assumed to more 'dynamic' and 'plastic' does not exhibit such memory, consistent with predictions of another recent model of PcG regulation in mESC [30].

Based on the theoretical work of M. Howard's group on PcG-based silencing presented earlier, Leonie Ringrose has developed (in collaboration with M. Howard) a model of

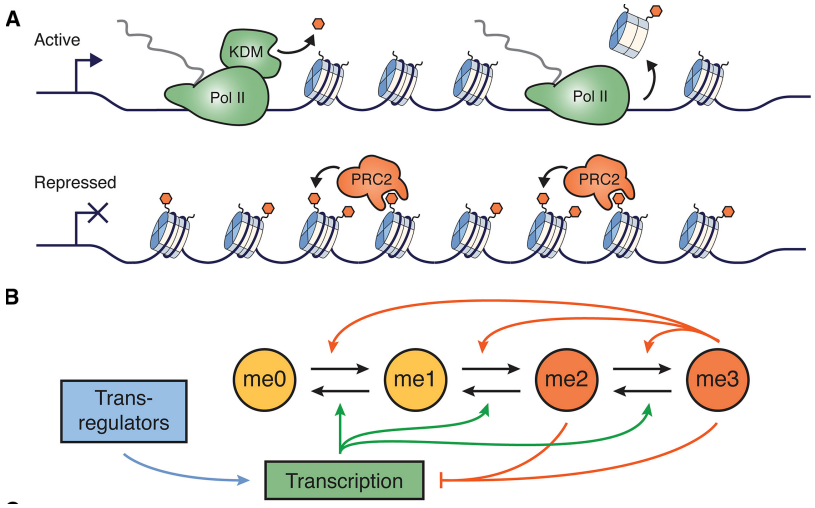

Fig. 1.8 (A) Schematic of alternative chromatin states. Active state characterized by presence of Pol II, which can carry H3K27-demethylases (KDM), and drive nucleosome exchange. Repressed state characterized by H3K27me3 (orange hexagons), which can positively feedback to recruit PRC2. (B) Diagrammatic representation of feedbacks in mathematical model. States me0 to me3 refer to methylation state of H3K27. Neutral marks me0/me1 indicated in yellow, repressive marks me2/me3 in orange. Black arrows represent state transitions; colored arrows represent feedback interactions. For clarity, histone exchange and H3K27me2-mediated recruitment of PRC2 are omitted. *Figure and legend from [27] (Figure 1).*

PcG-based silencing in the context of epigenetic gene silencing during fly (*Drosophila melanogaster*) development. Early in embryogenesis, genes are activated or repressed by targeting standard activators or repressors (DNA binding proteins that bind to the promoter and activate or repress gene transcription). At some point, these activators/repressors disappear but the expression state remains and is stably locked in this "initial" "on" or "off" state. This epigenetic memory is controlled at the chromatin level by the PcG/Trx system, where the PcG proteins direct silencing (see above) and the Trithorax (Trx) proteins, its counterpart, direct activation. These silencing/activation complexes are first nucleated in specific genomic domains called PRE/TRE along which they spread their chromatin state (repressive/active), which is then maintained through cell division. Reinig *et al.* proposed that state establishment at Polycomb/Tritorax response elements (PRE/TRE) (initiation stage) is instructed by the promoter state (i.e., bound or unbound by a repressor/activator) that drives the state to a given macrostate (initiation stage); after the disappearance of the repressors, the state is maintained through recruited conversions (positive feedback) (memory stage). To formalize this scenario, they use, for the PRE/TRE domain, a 3-state model (A,U,M) with noisy and recruited conversions, in an infinite range propagation mode. The dynamics of the state conversion is coupled to the state of the promoter which dynamically switches between a free and a bound state (2-state model) (Fig. 1.9). In their model, reversely, the global state of PRE/TRE is also conditioning the promoter states dynamics. This lead to a full coupling between promoter and PRE/TRE states: the more the active/repressed one, the more the active/repressed the other. Their model accurately recapitulates published studies of PRE/TRE-mediated epigenetic memory of both repression and activation; a key result is the crucial dependence of such bistability in cell cycle length, in agreement with the theoretical work of Zerihun et al. [23] (see below).

1.2.4 Chromatin-based epigenetic memory: crucial impact of cell cycle length

Since then, other approaches have been proposed to derive dynamical properties (macrostate stability and memory) of such 2 or 3 states generic models [176, 43, 42, 22, 23]. In Zerihun *et al.*, authors investigated the role of cell cycle duration in the epigenetic dynamics of a three state chromatin state dynamics. Considering a similar symmetric model as in [19] and [56], ie with long-range recruited conversion (actually infinite range) (Fig. 1.10(a)), they introduced a periodic replication process such as in Fig. 1.10(b) and derived the statistical properties of the nucleosomal array (ie statistical distribution of $m = I - A$) as function of (r, T), with $r = \varepsilon / k_o$ (recruited conversion rate over noise) and T the cell cycle duration (Fig. 1.11(a,b)). Note that in [22], all the derivation were done in a non replicating system, i.e., at the limit of very large T where the epigenetic system relaxed to its stationary state.

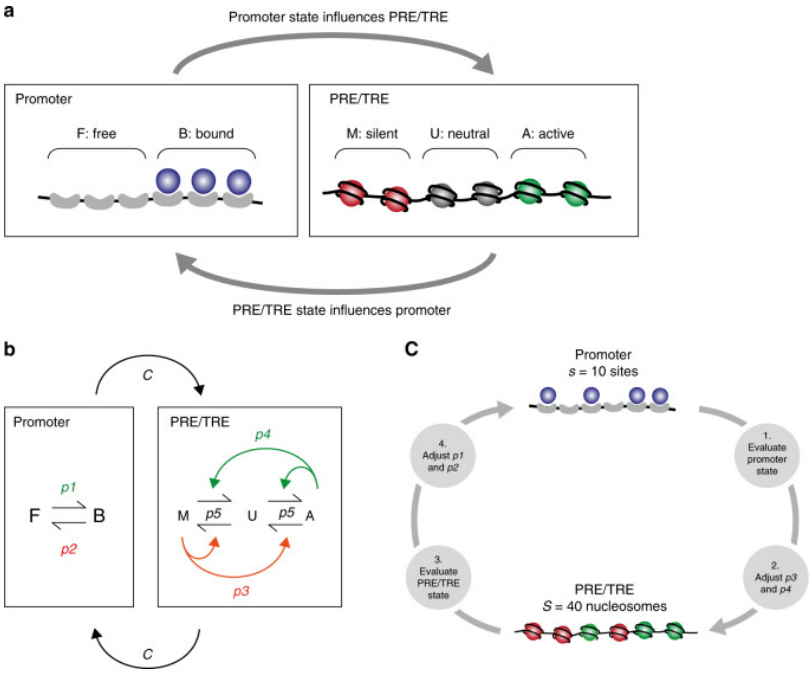

Fig. 1.9 A simple model for Polycomb/Trithorax regulation. (a) The promoter and PRE/TRE are shown schematically. Left: each promoter site can be either free (F) or bound (B). Right: each nucleosome in the PRE/TRE can be either silent (red, M), neutral (grey, U) or active (green, A). See main text for details. (b) The model is implemented stochastically, with probabilities for each of the transitions between F and B, and between A, U and M. For the promoter, p1: probability of transcription factor binding at a single promoter binding site. p2: probability of transcription factor unbinding at a single site. For the PRE/TRE, p3 and p4 (red and green arrows) denote feedback reactions in which nucleosomes in each of the M or A configurations convert other nucleosomes towards that configuration. The parameter p5 (black arrows) gives the probability of conversions between A, U and M that are independent of feedback. C: The promoter and PRE/TRE are coupled. (c) At each iteration of the simulation, the promoter state is evaluated and used to adjust the PRE/TRE parameters p3 and p4. Likewise the PRE/TRE state is evaluated and used to adjust the promoter parameters p1 and p2. These adjusted p1, p2, p3 and p4 values are used in the next iteration. *Figure and legend from [54] (Figure 1).*

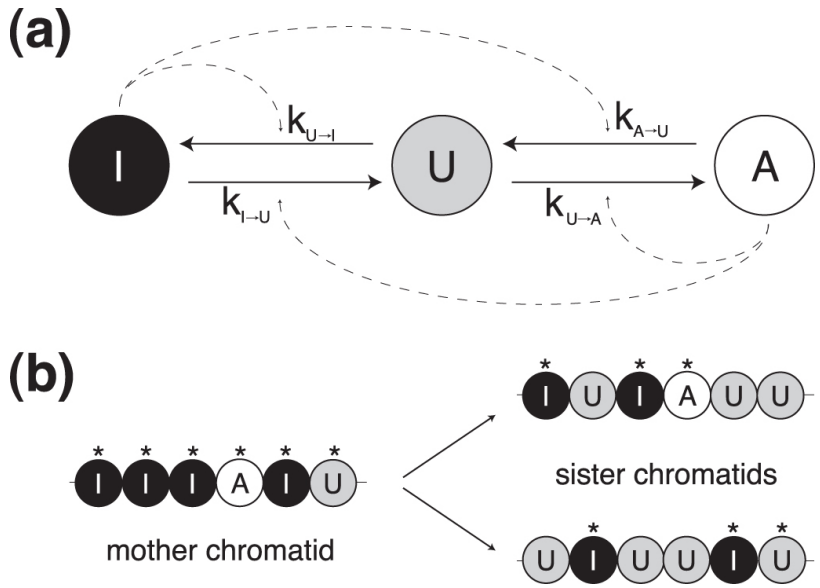

Fig. 1.10 3 state model and cell cycle length. (a) Three-states model for the local epigenetic state of one nucleosome. Transition rates between states depend on the total density of active and inactive marks (equations (1)–(4)). (b) Random distribution of the maternal marks (*) among the sister chromatids during replication. Empty spaces are filled with U-type nucleosomes. *Figure and legend from [23] (Figure 1).*

As shown by the phase diagram in Fig. 1.11(c), Zerihun *et al.* highlight the effect of having a finite cell cycle which limits the ability of the system to maintain a macro-state of modified nucleosomes: the critical recruited conversion rate ε_c value needed to have bistability is an increasing function of cell cycle duration. In other words, for a given recruited conversion rate, cell cycle duration can control the ability of the system to be in a bistable state, and thus its ability to form a stable and heritable domain.

1.2.5 A first step toward 3D: 1D model with effective 3D

In all of the above mentioned models, in their preliminary version, the dynamics of the local chromatin state does not depend on the 3D organization of chromatin. The models are essentially 1D models embedded in zero-dimensional space, where conversion of the cooperative state at a given nucleosome can occur only from the neighboring modified nucleosome (NN model) or from the set of modified nucleosomes (infinite range model). As shown by Dodd et al. [19], the NN speading model (or cis-spreading model) cannot lead to a stable coherent macrostate, whereas the infinite range model leads to bistability for conversion rates above a critical value [19, 47, 22]. In order to go beyond the rather unrealistic zero-dimensional model [19], Dodd *et al.* proposed to introduce in this model a long distance spreading term (trans spreading) that would account for chromatin looping.

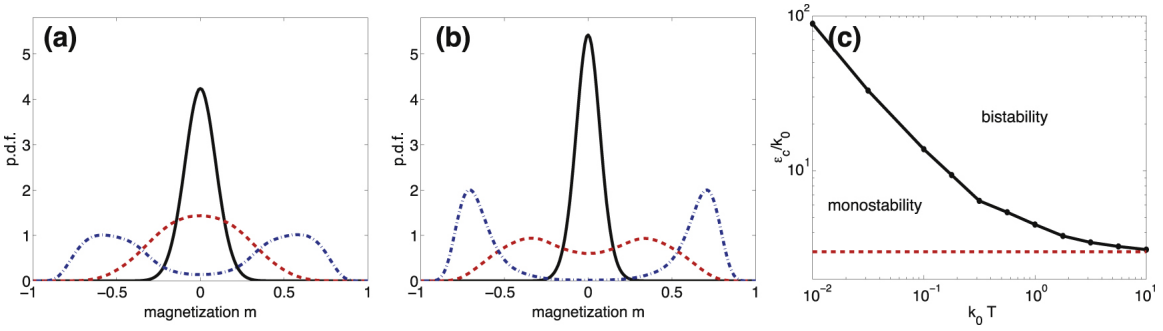

Fig. 1.11 Impact of cell cycle length on epigenetics. (a) and (b) Probability distribution function (pdf) of the magnetization $m = A - I$ for a population of asynchronous cells for different values of recruitment and cell cycle length T. (a) $k_0T = 1$ and $r = \varepsilon/k_0 = 1$ (full line), 4 (dashed line) and 6 (dotted and dashed line). (b) $\varepsilon/k_0 = 5$ and $k_0T = 0.1$ (full line), 1 (dashed line) and 20 (dotted and dashed line). (c) Phase diagram. Critical point ε_c as a function of T, separating the bistable region (right hand) to the monostable region (left hand). The dashed line corresponds to the long cell cycle limit $r_c = \varepsilon_c/k_0 \approx 3$. *Figure and legend from [23] (Figure 2).*

The objective was to highlight the role/necessity of long-range interactions in generating bistability and to emphasize that this long-range interaction emerges naturally from the generic property of the chromatin chain. To parameterize this long-range interaction, they naturally introduce a spreading term modulated by the contact probability that accounts for the colocalization of distant genomic sites, assuming a Gaussian chain at equilibrium. This is a mean-field approximation that is valid when the polymer dynamics is much faster than the state conversion dynamics (epigenetic dynamics), an approximation we will use in Chapter 2 and discuss in Chapter 4. In fact, although it was not the original intention of the authors, this study is a first theoretical demonstration that 3D chromatin organization can indeed affect the 1D dynamics of epigenetic marks through the propagation in trans that leads to a phase transition at a critical value from an "unmodified" macrostate (unmodified on average) to a bistable macrostate. This is in fact a very well known feature of 1D systems with cooperativity in statistical physics, as we have already pointed out before (e.g. the 1D Ising chain): in such systems, the phase transition requires a long-range interaction which is obviously ensured for the zero-dimensional system where the range is in fact “infinite”. If we now consider the long-range interaction induced by loops, bistability still occurs but it is less strong [19]: for a given nucleosome, the pool of nucleosomes that can potentially participate in the recruited conversion remains the whole system but now the conversion rate decreases with genomic distance, so that the efficiency of the recruited conversion decreases with genomic separation (there are indeed fewer nucleosomes that can potentially participate in the recruited conversion). In fact, theoretical developments in statistical physics have shown that as long as the decrease in the long-range interaction is less than $1/d^\gamma$, with $\gamma = 2$,

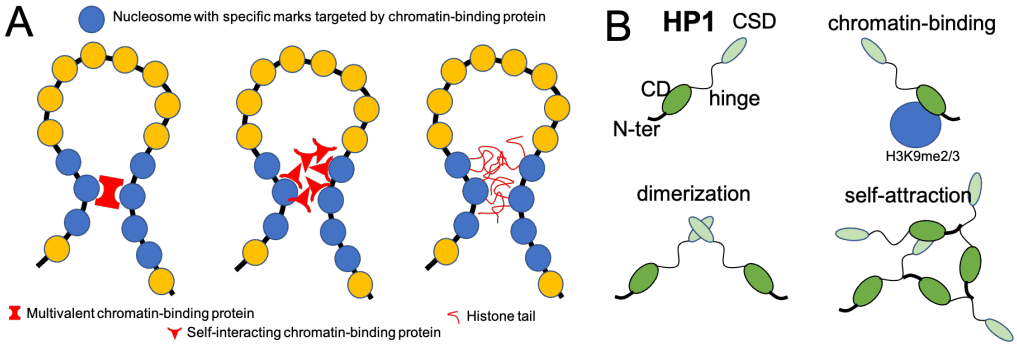

Fig. 1.12 (A) Three possible mechanisms for epigenomic-driven interactions via chromatin-binding proteins (left and center) or via histone tails (right). (B) Example of an architectural protein, HP1, and its various biochemical properties driving the formation of heterochromatin compartments.

the system still exhibits phase transition and bistability. Interestingly, the contact ability of chromatin decreases as a power law, in agreement with standard polymer physics, with exponents that are exactly in the range that ensures bistability: $\gamma < 2$. Note that for the Gaussian chain approximation used by Dodd et al. [19], $\gamma = 1.5 < 2$, value that leads to bistability.

A few other effective 3D descriptions are the 3D looping models of Erdel and Greene [25] and the stochastic model by Ancona *et al.* [28]. 3D looping models with sequence-dependent recruitment and long-range spreading mechanisms based on 3D contacts are by essence similar to our simple painter and reader-writer modes (in chapter 2). In [25], authors discussed the role of looping (3D contacts) and nearest neighbor spreading for epigenomic profiles, as well as the reader-writer feedback in memory and proposed that non-homogeneous state-specific recruitment may lead to stable confined domains.

1.3 Modeling of epigenome folding

As described in Sec.1.1 (Fig.1.3), many nuclear bodies, chromatin 3D compartments or contact domains are correlated with the 1D epigenomic information along the genomic sequence. Most of these observations indicate the possible colocalization (more contact than expected) of genomic regions sharing the same epigenomic content and/or the possible spatial segregation (less contact than expected) of regions sharing different epigenomic states.

Actually, there exists a growing body of evidence showing that some chromatin-binding architectural proteins or complexes, specific to some epigenomic marks, have the capacity (i) to bind to several chromatin regions; and/or (ii) to oligomerize or self-interact; thus, puta-

tively driving the formation of direct or effective physical interactions (or loops) (Fig.1.12A) between loci sharing the same chromatin state, possibly distant along the sequence. For example, the euchromatin-associated factor BRD4, that contains a bromodomain capable of binding acetylated histones and important in the regulation of super-enhancers and transcription [58], can *in vitro* and *in vivo* form liquid droplets [59]. Also, the constitutive heterochromatin-associated factor HP1 contains (Fig.1.12B) a chromoshadow domain (CSD) that allows the dimerization of HP1, a chromodomain (CD) that drives the binding to histones with H3K9me2/3 modifications and unstructured N-term and hinge regions that drive the *in vitro* formation of liquid condensates [142, 139, 57]. These three ingredients allowing the *in vivo* establishment of chromocenters [142, 57]. Interestingly, HP1 and heterochromatin are chemically excluded from BRD4 condensates [59]. There are also recent experiments suggesting that histone tails themselves may directly promote interactions between nucleosomes depending on the biochemical modifications [60, 61] (Fig.1.12A).

All this suggests the existence of physico-chemical mechanisms, driven or regulated by the epigenome, that significantly participate into the 3D structuring of the genome, by stabilizing or hindering short-range interactions between genomic loci.

To describe how such epigenomic-driven interactions may shape the 3D genome, several mechanistic biophysical models have been developed along the years. They usually model chromatin as a coarse-grain, bead-on-spring, polymer (Fig.1.13) where each monomer represents a given portion of chromatin (typically between 1 to 50kbp, depending on the model resolution). The dynamical and structural properties of the polymer are thus driven (i) either by explicitly simulating molecules that can bind specifically to some monomers, or by (ii) effective short-range direct interactions between monomers sharing the same epigenomic content.

The first class of models (i), sometimes termed string-and-binders-switch (SBS)-like models (Fig.1.13A), was mainly developed by Nicodemi and Marenduzzo groups [95–98]. In this case, the polymeric chain (one chromosome or piece of chromosome) is simulated with diffusing particles of several kinds that can bind to specific monomers whose corresponding genomic regions are enriched in some epigenomic marks or TF binding sites. In these frameworks, folding of the chain is driven by the capacity of the particles to bind several monomers at the same time, but the possible self-interactions between diffusing particles were not considered. However, very recent works start to integrate such self-attractions [245, 99].

The second class of models (ii), sometimes termed copolymer-like models (Fig.1.13B), was developed by my supervisors [56, 169, 90] followed by many other groups [100–104]. In this case, monomers directly interact via pair-wise potentials whose strengths of interaction

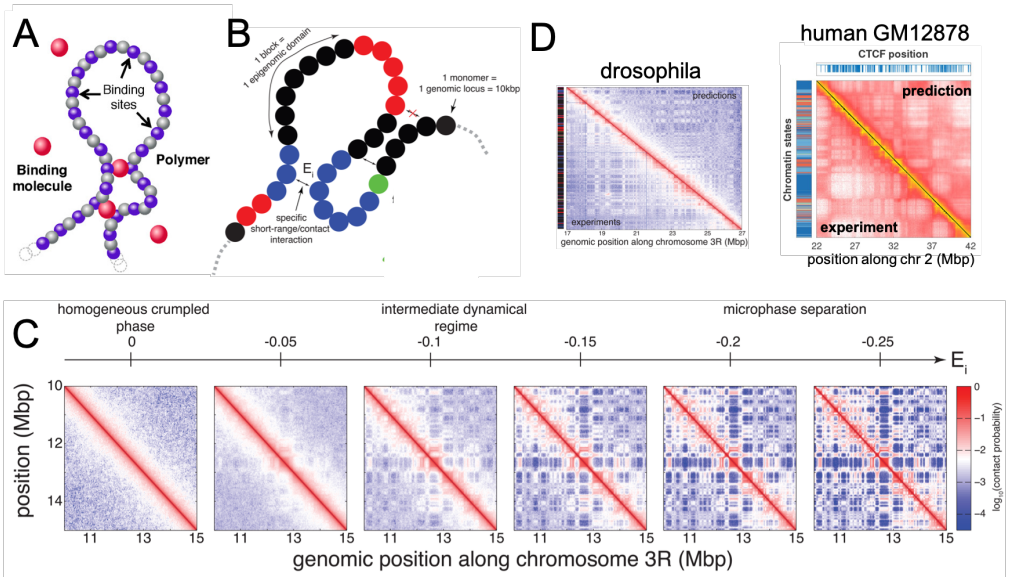

Fig. 1.13 (A) Scheme of the string and binders switch model. *Adapted from [95].* (B) Scheme of the block copolymer model. *Adapted from [204].* (C) Prediction of the block copolymer model as the strength of epigenomic-driven interactions is increased, from a homogeneous behavior (left) to full micro-phase separation (right). (D) Best fit obtained by copolymer model for Drosophila late embryo (left, *adapted from [204]*) or human cell line GM12878 (right, *adapted from [102]*).

depend on the local epigenomic content of each monomer. In its simple form, each monomer is defined by a single state, and only monomers of the same state can interact.

These two classes of models (and their variations) lead qualitatively to the same physical behavior. Since, loci sharing the same epigenomic state are usually arranged in long blocks (contiguous 10-100kbp-long regions characterized by a more or less homogeneous content of epigenomic marks) along the genome, formation of TADs may emerge from the full or partial collapse of each block [95, 203] and 3D compartments from a microphase separation of the different chromatin states [56, 90, 103] (Fig.1.13C), characteristics of block copolymers [105], where blocks of the same state colocalize forming more or less extended 3D domains depending on the strength of interaction [90] and of the linear arrangement of the 1D blocks [169]. This generic principle of microphase separation, when contextualized to specific systems, taking as inputs the corresponding 1D epigenomic landscape, generate predictions in very good agreement with Hi-C experiments (Fig.1.13D) and explain quantitatively the 3D compartmentalization observed in many species from Drosophila to mammals and plants. In many cases, it is found that the interaction strengths that best fit the data are weak (fraction of kT), consistent with a plastic, stochastic and dynamical chromatin organization [90, 106, 179].

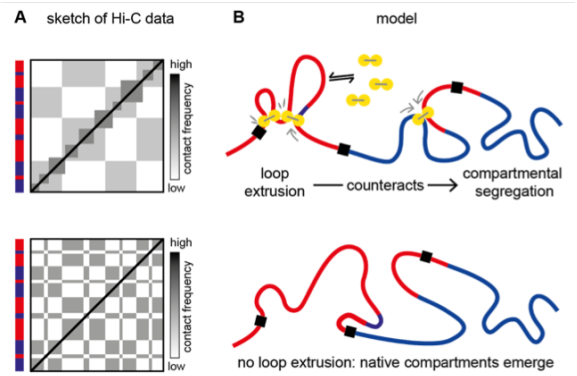

Fig. 1.14 Loop extrusion mechanism may compete with epigenomic-drive microphase separation. (A) Scheme of Hi-C signatures when both mechanisms are present (top,TADs and comparments) or with only microphase separation (bottom, only compartments). (B) Sketch of the mechanistic models when both mechanisms are considered (top) or when only epigenomic-driven interactions are accounted for (bottom). *Adapted from [107].*

As expected, epigenomic-driven interactions alone do not explain all the main features of chromosome organization. Recently, to improve our quantitative understanding of the nuclear chromosome organization, copolymer models were decorated with other key mechanisms [108] like SMC-mediated loop extrusion [107] (see below), interactions with the nuclear membrane [103, 109, 110] or with the nucleolus [111]. For example, Nuebler et al [107], studied the interplay between epigenomic-driven (micro)phase separation and loop extrusion. Indeed, in mammals, the formation of many TADs cannot be simply described by the folding of an epigenomic block but can be nicely interpreted as resulting from an active mechanism of loop extrusion [118, 241] (Fig.1.14) where cohesin rings are loaded to chromatin, extrude chromatin loops bidirectionally [112, 113] until unbinding or being stopped at oriented CTCF binding sites, enriched at TAD boundaries [63, 118]. Loop extrusion was shown to interfere with microphase separation by destabilizing it, in good agreement with experiments [114].

1.4 Modeling chromatin state dynamics in 3D

As we have seen epigenome driven interactions are central to the 3D organization of chromatin, 3D chromatin domains do strongly correlate with 1D epigenomic domains [63, 65]. Dynamical coupling of 1D epigenome and 3D organization had to be deciphered, for instance to quantitatively understand the correlation of repressive (heterochromatin) domains with chromatin compaction. These observations suggested a strong feedback between epigenome regulation and 3D chromatin folding which led to polymer models integrating

chromatin state transitions [66–70].

The first model of this class [67] invoked short-range interactions mediated by chromatin states (like states attracting and unlike states repelling) and an Ising-like chromatin state transition (mediated by the state of spatially close monomers). The model (Fig. 1.15i) proceeds by binding of multivalent proteins to a specific histone modification, and then it bridges the bound histone with histones in the same state, thereby 1D epigenome affecting 3D organization. The spreading process is directly linked to 3D configuration, as spreading (or recoloring) of states is modelled by Ising-like interaction with spatially proximal monomers. Polymer dynamics coupled to such local interactions enabled meta-stability of two chromatin states at critical point and phase transitions between the chromatin states away from the critical point (Fig. 1.15iv). Having different stochastic noise in epigenetic dynamics with respect to the polymer dynamics have shown to harbor a parameter range where several globular domains coexist qualitatively, reminiscent of TAD-like structures. In the following work [82], they have highlighted the significance of 'genomic bookmarking', bookmarking in the model refers to red(or blue) monomers which does not change state accounting for external signal or 'forcing' (like a transcription factor that binds to the DNA and recruits HMEs). They show that having a critical density of such bookmarks is essential for epigneomic domain stability (Fig. 1.15ii) and predict the full distribution of Polycomb marks in a Drosophila chromosome. Furthermore [238, 239], they have studied epigenetic microphase separation by using a field theoretic approach, wherein the system is described by two fields (i) density (local genomic density) and (ii) magnetization (local epigenomic state abundance). The approach that has been followed is to derive the free energy for a single chromosome with epigenetic and 3D dynamics from the partition function of a magnetic polymer (a self-avoiding polymer with epigenetic recoloring as described earlier) to study the equilibrium kinetics of the model. This leads to two possible equilibrium phases, a swollen disordered (with heterogeneous epigenetic marks) and a compact ordered phase (uniform epigenetic state) with a first order transition (a classification of phase transitions where the first derivative of free energy with respect to a thermodynamic variable is discontinuous, here the first derivative is magnetization which characterizes the overall epigenomic state of the system) between them (Fig. 1.15iii). Here, competing epigenetic domains are a natural consequence, but at equilibrium one dominant epigenetic domain takes over while a realistic scenario would be coexistence of multiple stable epigenomic domains.

My supervisors, Daniel Jost and Cedric Vaillant formulated the "Living chromatin" model in which epigenome driven interactions were modelled by a binding rate between two spatially close monomers of the same state implying that the compaction of the polymer at steady state would be determined by the ratio of binding and unbinding rates (k_b/k_u).

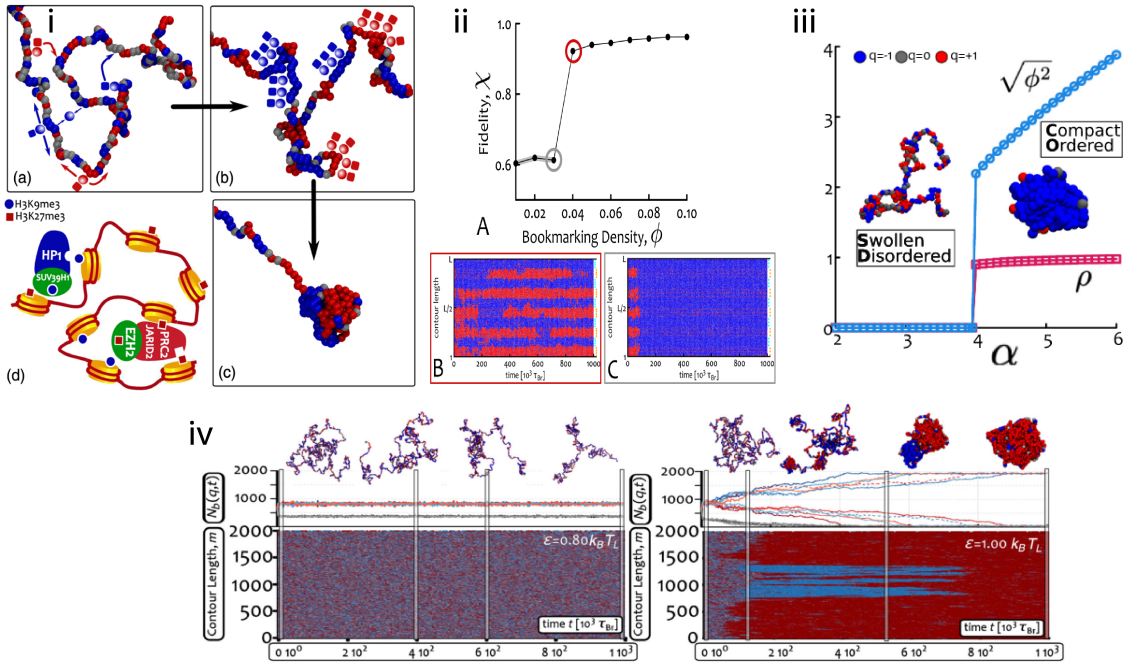

Fig. 1.15 (i) A 3D polymer model with "recoloring" for the propagation of epigenetic marks, (a)–(c) Multivalent binding proteins, or readers (shaded spheres), bind to specific histone modifications and bridge between similarly marked segments (distinguished here via their "color"). Histone-modifying enzymes, or writers (solid squares), are assumed here to be chaperoned by the bridge proteins. (d) Representation of Heterochromatin binding protein HP1 is known to recruit methyltransferase protein and the Polycomb Repressive Complex (PRC2) is known to comprise the histone H3 Lys 27 (H3K27) methyltransferase enzyme EZH while binding the same mark through the interaction with JARID2 *Figure and legend from [67], (Figure 1).* (ii) A critical density of bookmarks is required for stable domain formation, (A) Using the clustered pattern of bookmarks at different densities, we quantify the deviation from a 'perfect' block-like epigenetic pattern, (B and C) Kymographs representing the behaviour of the system at the points circled in red and grey in (A) *Figure and legend from [82] (Figure 4).* (iii) Equilibrium phase diagram for single chromosomes, The phase diagram is obtained by numerical minimization of free energy (red and blue symbols refer respectively to the order parameter and density profiles). The insets report snapshots from simulations *Figure and legend from [239] (Figure 3).* (iv) The two-state model above the critical point evolves into an epigenetically coherent state via a symmetry-breaking mechanism, Top row: Typical snapshots of 3D configurations adopted by the polymers as a function of time, below and above the critical point, Middle row: Time evolution of the total number of beads of a particular type, for four independent trajectories (the dashed one corresponds to the trajectory from which the snapshots are taken). Bottom row: Time evolution of the color of each polymer bead as a viewed as kymograph. *Figure and legend from [67] (Figure 2).*

This framework (Fig. 1.16) can be seen as the combination of block copolymer model of chromatin [56] and the seminal epigenome regulation model of Dodd *et al.* [19] discussed in sec. 1.2.1. Chromatin state transitions are steered by adjacent monomers in 3D, the transition rate of a monomer to a specific state (for example M-state) being determined by the number of spatially close monomers in M-state. Specifically, the model dissected the role of spreading via cis (along the chain) and trans (3D contacts) (Fig. 1.17B) which helps in understanding the feedback loop between 1D and 3D information. 1D epigenomic information along the genome driving 3D genomic contacts (thereby the 3D chromatin organization) while 3D organization driving epigenomic spreading via the trans spreading mechanism. And evidently, it was shown that epigenomic domains cannot be maintained without trans spreading stressing the significance of 3D organization on 1D epigenome domain stability. Such 1D-3D coupling was also shown to be essential for the stability of boundary between two antagonistic epigenomic domains. While the 'Living chromatin' framework captures the bi-stable chromatin state regime, like the epigenome recoloring model (discussed above, Fig. 1.15) the spreading mechanism is fundamentally different. Here, spreading in trans is not coupled to the copolymer dynamics, but only requires spatial proximity of the monomers while in the model developed by Michieletto *et al.* [67] the dynamics of epigenome and polymer are governed by the same Hamiltonian and spreading of chromatin state requires chemical bond between nearest monomers in 3D, as a prerequisite. This difference is reflected in the nature of the phase transition which is of second-order in the living chromatin model. The self-association of monomers in the same epigenomic state enables the formation of compartmentalized epigenomic domains. Living chromatin model shows that the avoidance of cis-spreading (biologically, it can be due to the presence of insulator proteins like CTCF) at the domain boundaries would significantly stabilize the boundary as well as the domains (Fig. 1.17C).

A model for spreading of chromatin states based on transient DNA looping was proposed by the group of Spakowitz [70] in which they describe looping('unlooping') dynamics, and employ it to describe loop mediated methylation. Such a mechanism as a prequisite requires chromosomal DNA to be able to adopt dynamically looped configuration and is likely to be applicable in the context of undifferentiated cells. Looped (or unlooped) state (Fig. 1.18A) is determined by a parameter called loop radius, a distance between two nucleosomes below which the state is looped (and above 'unlooped'). Looping rate is computed by determining the first passage time for looping (time taken to go from 'unlooped' to looped state) and assuming that looping kinetics are captured by Poisson process. Once, two nucleosomes are in the looped configuration, methylation can spread at a specified rate. Time evolution of the average methylation state is determined by a master equation, invoking the concentration of HP1 architectural proteins. This model heavily weighs on looping probability that determines effective spreading of methylation. Further,

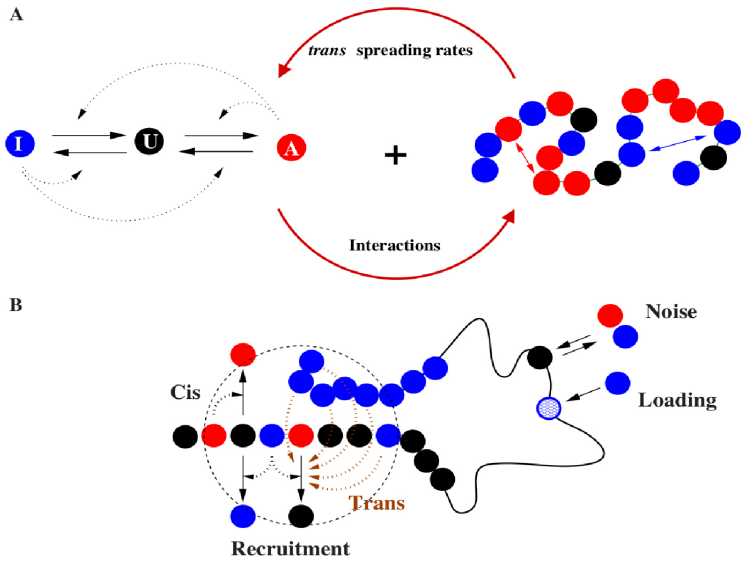

Fig. 1.16 The Living chromatin model. Each monomer can be in one of the three states: A, U and I; the inter-conversion dynamics between these states results from random or recruited (in cis or in trans) conversions. The chain is modeled by a semi-flexible self-avoiding bead-spring model with specific short-range attractions between monomers of the same epigenomic states (A or I). *Figure and legend from [66].*

they elaborate this framework by combining the effect of segregation (by starting from an experimentally obtained methylation state) and compaction [157] on chromosomal organization including the effect of architectural protein HP1. In this model HP1 proteins can bind and unbind to monomers in any state but preferentially binds to methylated monomers. The above mentioned segregation and compaction is achieved by attributing an experimentally derived energetic benefit for regions with higher HP1 concentration. Thus for a given HP1 concentration and a methylation profile, the model predicts which genomic regions compact to form heterochromatin. Based on the previously discussed loop-mediated spreading, methylation may spread from methylated/unmethylated monomers, the probability to be methylated for every monomer can be written as a function of local HP1 concentration, methylation and de-methylation rates. The feedback via HP1 proteins, wherein the proteins compact methylated region and promotes further spreading of methylation in the region, while in regions with less methylation would remain de-condensed directly links spreading of methylation and compaction of chromatin. As a consequence, it was shown that an optimal HP1 concentration is a prerequisite for robust maintenance of methylation profile (which is correlated to the initial profile) over several generations (Fig. 1.18B). This is an aspect that had not been probed by the previous models. In chapter 4, we discuss the effect of concentration of architectural proteins on phase separation properties.

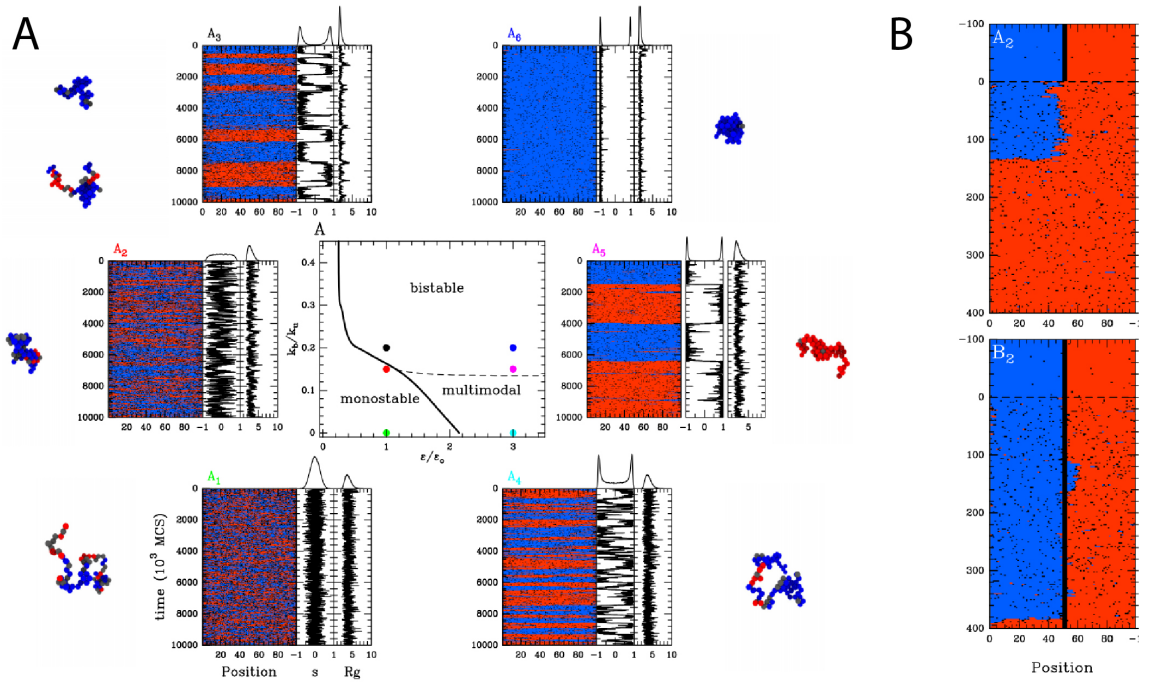

Fig. 1.17 (A) Illustrating the cis-trans spreading model and epigenomic driven self-attraction. The phase diagram as function of ratio of binding and unbinding rates (k_b/k_u), and spreading rate which demarcates regions of bi-stability. Examples of time evolution of epigenomic state and of radius of gyration are shown along with representative 3D configurations. (B) Maintenance of two neighbouring epigenomic domains. Typical time-evolution of the epigenome without 1D boundary (indicated as A_2 within the panel) and with a remaining 1D boundary (indicated as B_2 within the panel). *Figure and legend from [66].*

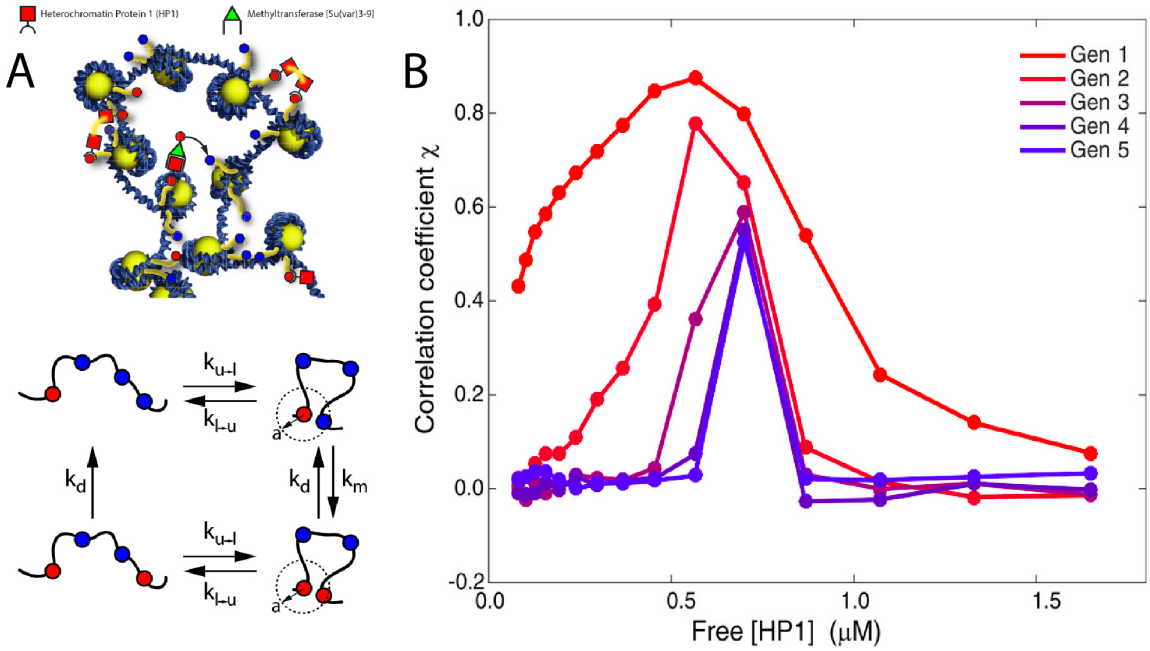

Fig. 1.18 (A) The top schematic depicts an array of nucleosomes and the interactions among methylation of H3K9, HP1α, and the methyltransferase SuVar3 – 9. The bottom schematic represents the kinetic processes that are involved in the conferral and removal of a methyl mark on a nucleosome via chromosomal looping and methyltransferase function. Each bead represents a nucleosome that is either methylated (red) or unmethylated (blue). *Figure and legend from [70] (Figure 1).* (B) Plot of the correlation coefficient (between the current generation profile and the initial profile) versus free HP1 concentration for five generations of reestablishment of the methylation sequence. *Figure and legend from [157] (Figure 4).*

Recently, the work by Thirumalai group [68] shows 3D looping in the slow epigenetic spreading regime produces finite domains without boundary elements, employing a 3D polymer on which epigenetic spreading is modelled. As in the Living chromatin framework, cis-trans spreading of chromatin state is delineated and spatial proximity of nucleosomes of same state determine the transition rate. When the process of spreading is fast compared to polymer dynamics, it occurs independently of the chromatin configuration (Fig. 1.19i). The model accounts for nucleation of HMEs by having a monomer whose epigenetic identity is unchanged. They investigated the interplay between spreading dynamics and chromatin motion, highlighting the role of nucleation in the maintenance of stable domains in a regime of fast epigenetic spreading. Also, another work incorporating generic concepts of 1D-3D coupling like self-association of monomers in the same chromatin state, 3D spreading of chromatin states and limitation of HMEs studied chromatin structure reorganization through cell cycle [69] (Fig. 1.19ii). Here, chromatin is modeled as a spherically confined copolymer with monomers of two types, one of which represent heterochromatin regions

and can self-attract. Here, spreading of epigenetic modification (marks) is modelled on a fixed 3D template and at every cell cycle, the 3D reorganizes based on 1D epigenome. Modifications spread to neighbors (in 3D space, within an interaction radius), at rate S, and are lost everywhere uniformly at rate L. This work does not focus on the dynamic coupling of 1D epigenome and 3D chromatin organization. This approach is based on considering cell cycle to have two alternating phases (i) "interphase", in which the polymer is fixed in a single configuration while spread (and loss) of marks happen and (ii) "metaphase", in which the polymer is compacted into a condensed state and then refolds into a new configuration, while marks are fixed (Fig. 1.19ii). This model also recapitulates the importance of enzyme limitation and 3D compaction discussed in chapter 2 and chapter 4.

In chapter 4, we present progressively how to advance from the living chromatin-like model to a much more finer description of the molecular processes by introducing the "Living painter" framework, coupling epigenome regulation by diffusing HMEs and 3D polymer dynamics.

1.5 Main objectives of the PhD thesis

As described in the previous sections, the reader-writer ability of HMEs coupled to long-range spreading and possibly genome folding are likely to be key players in the regulation of epigenetic marks. However, such interplay raises concerns about the maintenance of a stable 1D compartmentalization of the genome [25]. Indeed, as observed in many species, the states of chromatin along the genome are linearly organised into consecutive "domains" of finite sizes, leading globally to a 1D compartmentalisation of the genome into active and inactive domains with more or less well defined inter-domain boundaries [62, 71, 73]. Because of the requirement to regulate (activation or silencing) a given part of the genome without affecting its surrounding flanks, chromatin states have thus to be specifically targeted to and stably confined inside specific genomic domains. In standard mathematical models [19, 22, 54], long-term epigenetic memory goes with great difficulties to limit the expansion of the mark (see Sec. 1.2), questioning mechanistically how the 1D partitioning of the genome is established and above all maintained.

Several hypotheses have been proposed to address this question. Spreading may be slowed down by boundary elements [6, 20, 66, 27], usually termed "insulators", that restrict the local spreading of a mark along the genome. These insulators are often associated with specific DNA-binding proteins (e.g. CTCF, BEAF, CP190) or actively transcribed genes found at the boundaries between antagonistic chromatin states [76, 80, 74, 81, 75, 77–79].

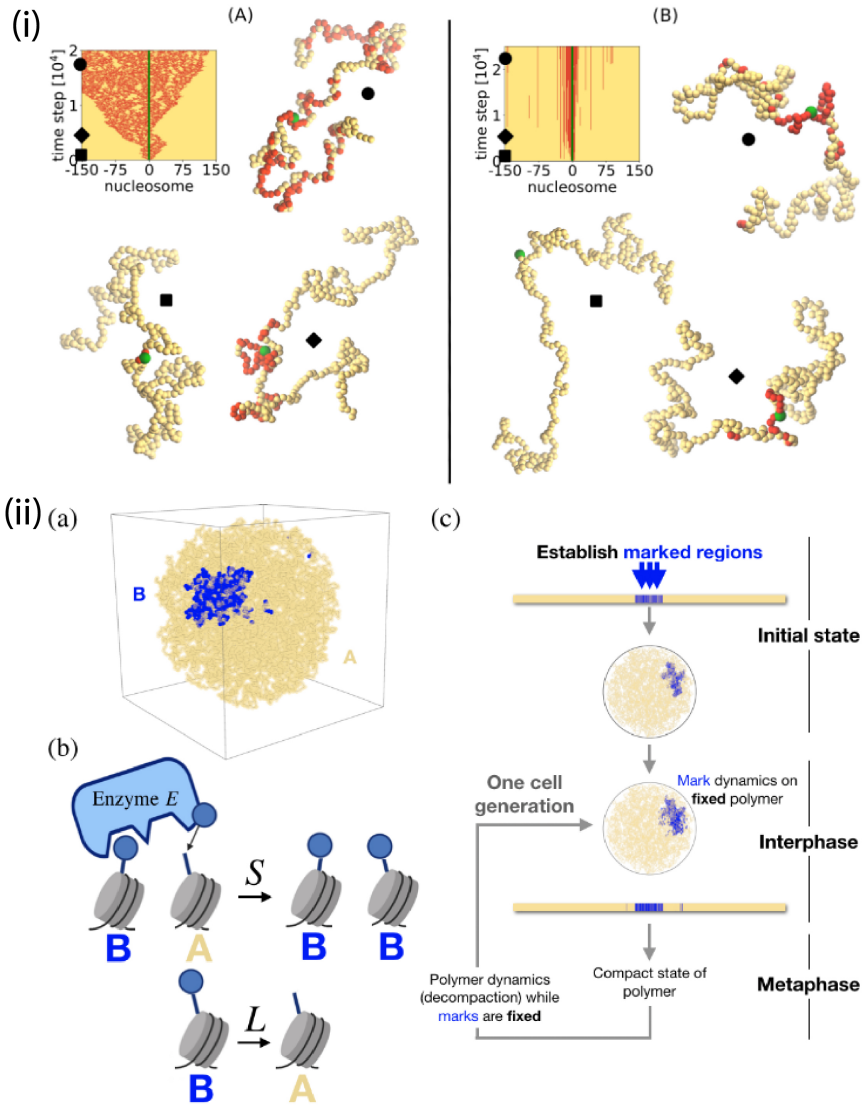

Fig. 1.19 (i) Visualization of single trajectories in the fast (A) and slow (B) spreading regimes. In the former case, the spreading occurs fast and independently of chromatin configuration, while in the latter case, the pattern of spreading may be determined by the dynamic rearrangement of the chromatin. The kymographs show the epigenetic identity of each nucleosome (marked/M/red, unmarked/U/yellow, nucleation site/green) for a single trajectory, in both (A) and (B). *Figure and legend from [68] (Figure 3).* (ii) Chromatin in the nucleus modeled as a spherically confined copolymer with monomers of two types, A (pale yellow) and B (blue), representing a varying pattern of histone marks. Monomers of type B, which represent regions bearing heterochromatic marks, self-attract. (b) Marks spread to neighbors (in 3D space, within an interaction radius), at rate S, and are lost everywhere uniformly at rate L. (c) The overall dynamics of our model consist of alternating phases of polymer dynamics and mark dynamics, coarsely stylizing the cell cycle. *Figure and legend from [69] (Figure 1).*

Spreading may also be limited by the formation of spatial compartments [67, 66, 82, 70, 68] as genome 3D compartmentalization is strongly correlated with 1D chromatin segmentation [55, 62, 65], thus providing 3D insulation and restricting long-range effects (see Sec.1.4).

Nevertheless, recent experimental studies suggest that, in vivo, the reader-writer mechanism, along with long-range spreading and insulation might not be strong enough to self-sustain by itself an epigenetic state and that a (compartmentalized) long-term memory may still be dependent on genomic bookmarking by keeping a weak "forcing" activity at regulatory sites [83–85, 10]. Such role of nucleation signals in the maintenance of a stable chromatin state and its interplay with the 3D genome organization has only been partially addressed theoretically [25, 28]. In this PhD thesis, I propose to develop general mechanistic frameworks of epigenomic regulation and memory integrating nucleation, reader-writer mechanisms, long-range spreading and genome folding, in order to systematically characterize such a crosstalk and to investigate how such generic principles can be contextualized to specific biological systems.

In Chapter 2, that gathers the main results of my PhD thesis, I introduce and fully characterize an end to end, simple modeling framework, the "Painter model", describing the mechanistic interplay between sequence-specific recruitment of chromatin regulators, chromatin-state-specific reader–writer processes and long-range spreading mechanisms, but neglecting the feedback of the epigenome on the 3D genome organization. In Chapter 3, towards an explicit integration of such a feedback and of the spatio-temporal dynamics of the polymer chromatin, I address a methodological question about how to precisely model a specific region of interest, capturing the correct biophysical properties, while minimizing computational efforts. In Chapter 4, I develop the "Living painter" model, progressively building from the painter model and explore the characteristics of the models that explicitly couple genome folding and the painter model, illustrating some of their key features. This chapter represents the scope and prospects of my thesis. And in the final part, I conclude the thesis and discuss possible perspectives.

Chapter 2

Painters in chromatin: epigenome regulation and memory

2.1 Introduction

As described in Chapter 1, epigenomic regulation is likely to depend on several key mechanisms including nucleation, reader-writer processes, long-range spreading and genome folding. In this chapter, we present a quantitative framework for epigenome regulation and systematically investigate the formation and memory of chromatin states. In particular, our modular approach allows to carefully examine the role of sequence-specific nucleation signals, and how they interplay with reader-writer mechanisms and 3D organisation. We also study the effect of chromatin state fluctuations on transcription.

After a deep analysis of the generic properties of each contributions, we show that such a framework allows to rationalize the maintenance of confined chromatin states and to address the role of epigenome regulation in gene expression. Finally, we contextualize our quantitative approach to several concrete experimental systems, from the formation of γH2AX domains around double-stranded breaks and heterochromatin domains around Transposable Elements (TE) insertion sites in mammals to the heterochromatin memory in fission yeast. The very good agreement between our simple model predictions and experimental data suggests that such multi-modal spreading model may provide a very promising framework to further investigate the link between genomic and epigenomic organisation during normal development, pathologies as well as during evolution.

This Chapter is largely inspired from the article that we published in *Nucleic Acids Research* (Abdulla, Vaillant, Jost. Painters in chromatin: a unified quantitative framework to systematically characterize epigenome regulation and memory. 2022. *Nucleic Acids Res*, **50**: 9083-9104). We also discuss newer analysis presenting intriguing perspectives of the framework.

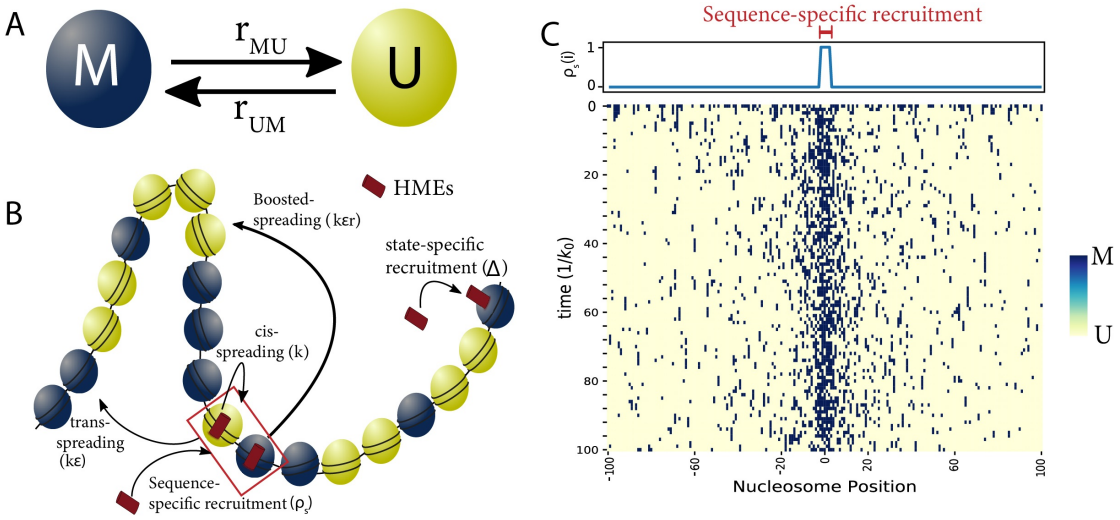

Fig. 2.1 "Painter" Model: (A) Each nucleosome can be in one of the two, unmodified "U" or modified "M" states. The switching between the two states is controlled by transition rates r_{UM}, r_{MU}. (B) Scheme of functionally distinct spreading mechanisms (coefficients of the terms are shown as in Eq. 2.2). The painter region where HMEs (brown) are recruited in a sequence-specific manner is framed in red. (C) Profile of $\rho_s(i)$, showing sequence-specific recruitment, $\rho_s(i) = 1$ depicting the nucleosomes framed in red in (B). Below an example of the stochastic time evolution of the epigenomic state simulated for a 201 nucleosomes region in the simple painter mode.

2.2 The "Painter model"

Chromatin is classically modeled by a unidimensional array of n nucleosomes. As a proxy for the local chromatin state, we assume that each nucleosome can fluctuate stochastically between a finite number of epigenetic states, each state corresponding to a specific combination of histone marks. To simplify, we consider a generic two-state model between an unmodified/neutral state (U) and a modified, either active or inactive, state (M) (Fig. 2.1A). The switching dynamics from M to U (respectively from U to M) is controlled by the transition rate r_{MU} (resp. r_{UM}).

r_{MU} integrates all the molecular processes that promote the removal of the histone modifications. This includes histone turnover [86], enzymatic removal of marks by "erasers" (such as histone deacetylases or demethylases) [87] and histone dilution between sister chromatids at replication forks [88]. In our theoretical study, in the simplest form, we lump all these processes into one constant $r_{MU} \equiv k_0$.

r_{UM} accounts for the spreading of the state M by dedicated "writer" enzymes: histone modifying enzymes (HMEs) are first recruited at some positions and then may "write" the M state to any neighboring U nucleosomes (Fig. 2.1B).

We assume HMEs can be mobilized to chromatin via two independent, additive, sequence- or state-specific, pathways. The quantity of enzymes bound at position i, $\rho_w(i)$, is then given by $\rho_w(i) = \rho_s(i) + \Delta\delta_{i,M}$ with $\rho_s(i)$ the sequence-specific contribution, $\delta_{i,M} = 1$ if i is in state M, $= 0$ otherwise, and $0 \leq \Delta \leq 1$ is the "reader" recruitment strength accounting for the capacity of some HMEs to co-associate with the same mark they catalyze [5]. For clarity, we only consider logical distributions for ρ_s with $\rho_s(i) = 1$ at given sequence-specific recruitment sites, called the "painter" regions.

Similarly, the activity $k_w(i)$ of a bound enzyme may be sequence- or state-specific: $k_w(i) = k_s(i)(1 + r\delta_{i,M})$ with $k_s(i)$ the normal enzymatic activity and $r > 1$ a boost factor accounting for the enhancement of enzymatic activity occurring for some HMEs in the presence of the same mark they catalyze [17]. For simplicity, we assume that $k_s(i) \equiv k$ is homogeneous.

We consider that a bound enzyme at position i may spread the state M not only at nucleosome i (with an on-site (*in cis*) activity $k_w(i)$) but also at longer-range (*in trans*) to any other nucleosome j at a rate $k_{trans}(i \to j) = \varepsilon k_w(i) P_c(i,j)$, proportional to the intrinsic enzyme activity $k_w(i)$, ε ($0 \leq \varepsilon \leq 1$) a multiplicative factor accounting for a putative modulation of the writing efficiency on off-site nucleosomes and $P_c(i,j)$ (≤ 1) the probability that i 'communicate' or 'contact' with j to allow spreading.

Altogether, the propensity for a nucleosome i to switch from U to M is thus given by the activities of locally bound HMEs and of HMEs bound to other nucleosomes that may contact with i:

$$\begin{aligned} r_{UM}(i) &= k_w(i)\rho_w(i) + \sum_{j \neq i} \varepsilon k_w(j) P_c(i,j) \rho_w(j) \qquad (2.1) \\ &= k \left[(\rho_s(i) + \Delta\delta_{i,M})(1 + r\delta_{i,M}) + \sum_{j \neq i} \varepsilon (1 + r\delta_{j,M})(\rho_s(j) + \Delta\delta_{j,M}) P_c(i,j) \right] \end{aligned}$$

2.2.1 Shape of spreading probability $P_c(i,j)$

The capacity for a HME recruited at position i to spread a mark at long-range is encoded into $P_c(i,j)$. P_c may translate various physical mechanisms that allow i to 'communicate' with j. The assumption is that this communication is faster than the polymer relaxation dynamics (see Chapter 4 for detailed comparison).

A natural mechanism is to consider that P_c captures the frequency that i and j are in spatial proximity such that a HME bound to i may catalyze a reaction in j (Fig. 2.17). In this case, assuming that the 3D chromatin organization equilibrates locally faster than the epigenomic landscape, $P_c(i,j)$ can be approximated by the contact probability between i and j [70] as observed in Hi-C experiments [92]. Indeed, for short genomic distances between i and j

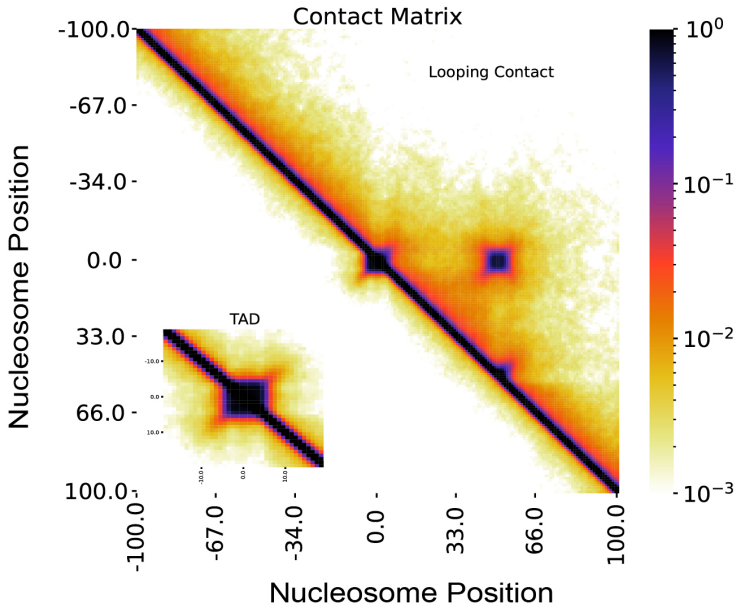

Fig. 2.2 Contact probability matrix for a compact domain (lower part) localized at positions [-4:4] (inset shows a zoom around the compact domain) and looping contact (upper part) between the painter region [-2:2] and nucleosomes at position [48:52].

($<$ 100kbp), theoretical and experimental studies suggest that, at this scale, 3D looping rates are $\sim$ sec^{-1} to min^{-1} [89, 94, 90, 93, 179] while the chromatin modification rates are often $\sim$ h^{-1} [91, 30]. In this chapter, for most of our theoretical results (except in Fig. 2.24C,D and some cases in Fig. 2.17, see below), we consider a homogeneous 3D contact probability $P_c(i,j) = 1/|i-j|^{\gamma}$, with γ reflecting the average compaction level of the chromatin fiber. Experimentally, γ has been shown to vary in the range $[0.5-1.5]$ depending on cell cycle stage, organism and/or cell fate [95, 115, 116]. In particular, we use the standard, intermediate value $\gamma = 1$. Note that the value of γ does not impact qualitatively the main conclusions of our work but may quantitatively have significant effects on the spreading process (Fig. 2.11) and thus must be carefully adjusted when considering specific experimental systems. We also consider contact matrices (for $P_c(i,j)$) representative of chromatin regions with (i) a central strongly-self-interacting domain of nine nucleosomes; or (ii) long-range loops between the painter region (positions $[-2:2]$) and five nucleosomes at positions $[48:52]$ (Fig. 2.2). In both cases, $P_c(i,j)$ was computed from polymer simulations using the lattice kinetic Monte-Carlo model developed in [90] (see Sec. 2.2.7). Also, we discuss three alternative mechanisms for P_c in the following sections:

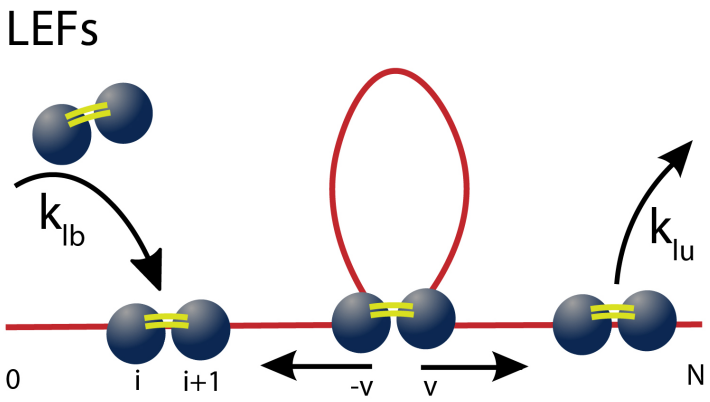

Fig. 2.3 Illustration of the loop extrusion model. Loop extruding factors (LEFs) can bind to the chromatin (in red) at position $i, i+1$ with a fixed rate k_{lb} and unbind at a rate k_{lu}. The two subunits of the LEFs move bidirectionally in opposite directions at velocity v leading effective extrusion of chromatin loops.

Nearest neighbor

HMEs bound at position i may impact only nearest-neighbor (NN) nucleosomes, i.e. $P_c(i,j) = 1$ if $|j-i| = 1$ and $P_c(i,j) = 0$ otherwise.

Diffusion-limited case

HMEs recruited at i, while unbinding and diffusing are more likely to rebind in the 3D vicinity, leading to a concentration gradient around i. To estimate $P_c(i,j)$ in the case where the effective spreading from position i to a distal nucleosome j occur by the unbinding and diffusing of the HME recruited in i, we follow a formalism consistent with Li et al. [143]. H. Berg [144] showed that the probability for a slowly-diffusing particle to hit an object of size a initially at a distance R from the particle is $P_d = a/R$. Within our context, R is thus the typical distance between i and j. $R(i,j)$ is taken to be the root mean squared distance between two loci of a simple self-avoiding polymer. $R(i,j) \sim s^{0.5}$ with $s = |j-i|$ the genomic distance between i and j. This leads to $P_d(i,j) \propto 1/|j-i|^{0.5}$, that is why we take the corresponding spreading probability to be $P_c(i,j) = 1/|j-i|^{0.5}$.

Loop extrusion case

HMEs bound at i may spread a mark to a distal nucleosome j only if they are placed in very closed proximity by a loop extruding factor (like cohesin or condensin) [117–120] translocating along the chromatin (Fig.1.14). We develop a simple mathematical model (Fig. 2.3) to estimate P_c in the case where spreading from a HME bound at i to a distal nucleosome j is only possible if they are placed in very close proximity by a loop extruding

factor (LEF like cohesin or condensin) translocating along the chromatin [117–120]. We assume loop extruding factors are composed by two motor subunits [145] that bind to chromatin on adjacent sites $(i, i+1)$ with a fixed rate k_{lb}. Once bound, they translocate bidirectionally away from one another at constant speed v. LEFs may also unbind from the chromatin at a rate k_{lu}. For simplicity, we assume a low density of LEFs on chromatin and thus neglect collisions between them. We also neglect the presence of barriers (like CTCF) that may stop or slow down LEF subunit translocation [146]. From these hypotheses, the dynamics of the probability to find a LEF between i and j ($P_l(i,j)$) simply follows

$$\frac{dP_l(i,i+1)}{dt} = k_{lb} - (2v+k_{lu})P_l(i,i+1) \tag{2.2}$$

$$\frac{dP_l(i,j)}{dt} = v(P_l(i+1,j)+P_l(i,j-1)) - (2v+k_{lu})P_l(i,j) \tag{2.3}$$

At steady state, this leads to $P_l(i,i+1) = k_{lb}/(2v+k_{lu})$, and in general, for $j \neq i+1$, $P_l(i,j) = v(P_l(i+1,j)+P_l(i,j-1))/(2v+k_{lu})$. Assuming an homogeneous system, ie that $P_l(i,j)$ only depend on the genomic distance s between i and j ($P_l(i,j) \equiv P_l(s)$ with $s = |j-i|$), we obtain the recurrence relation

$$P_l(s) = \left(\frac{2v}{2v+k_{lu}}\right) P_l(s-1) \tag{2.4}$$

$$\text{with} \quad P_l(1) = \frac{k_{lb}}{2v+k_{lu}} \tag{2.5}$$

Therefore $P_l(s) = \beta^s \alpha$ with $\beta = 2v/(2v+k_{lu})$ and $\alpha = k_{lb}/(2v)$, leading to $P_l(s) = \alpha e^{-s/s_0}$ with $s_0 \equiv -1/\log\beta > 0$ a typical genomic distance that characterize the processivity of the LEF. For example, in the limit of fast extruding LEFs ($2v \ll k_{lu}$), $s_0 \approx (2v)/k_{lu}$. That is why we take the corresponding spreading probability to be $P_c(i,j) = e^{-|j-i|/s_0}$.

2.2.2 Stochastic simulations

For a given set of parameters, the stochastic dynamics of the system is simulated using the standard Gillespie algorithm [121] implemented in Python (can be downloaded at https://github.com/physical-biology-of-chromatin/Painter-Model). In this chapter, we have simulated the dynamics of $n = 201$ nucleosomes corresponding to $\sim$ 40kbp-long genomic region. Starting from a random initial macro-state (i.e random choice between U or M states for each nucleosome), the system relaxes to a steady state. Each simulation corresponds to a "single cell" trajectory of the local epigenetic state (Fig. 2.1C). Unless specified (Fig. 2.20, 2.24), characterization of the system was done at steady-state. In each condition, at least 500 different trajectories were simulated. Times are given in $1/k_0$ unit that characterizes the typical turnover time of histone marks and is $\sim h$ [86, 30]

In our simulations, we start from a random initial state and at every discrete time step, 1 of the N possible state-conversion (M to U or U to M) occurs based on the defined probabilities of each state-conversion progressively driving the system to steady state.

2.2.3 Analytical solutions

In the case where state-specific terms are negligible ($r = \Delta = 0$, painter mode), the steady-state probability $P(M_i)$ of being in modified state M at any position i is simply given by :

$$P(M_i) = \frac{r_{UM}(i)}{r_{UM}(i) + r_{MU}(i)} \tag{2.6}$$

$$= \frac{(k/k_0)(\rho_s(i) + \varepsilon \sum_{j \neq i} \rho_s(j) * P_c(i,j))}{(k/k_0)(\rho_s(i) + \varepsilon \sum_{j \neq i} \rho_s(j) * P_c(i,j)) + 1}. \tag{2.7}$$

In the case where only state-specific recruitment is not present ($\Delta = 0, r > 0$)

$$P(M_i) = \frac{(k/k_0)(\rho_s(i) + \varepsilon \sum_{j \neq i} \rho_s(j) * P_c(i,j)(1 + r\delta_{j,M}))}{(k/k_0)(\rho_s(i) + \varepsilon \sum_{j \neq i} \rho_s(j) * P_c(i,j)(1 + r\delta_{j,M})) + 1} \tag{2.8}$$

where $\delta_{j,M} = 1$ if nucleosome j is in M-state, $= 0$ otherwise. Inside the painter region, $\overline{P_p}$ (the average value of $P(M_i)$ inside the painter region) can be computed by assuming that $\delta_{j,M} = \overline{P_p}$. Under this mean field approximation, $\overline{P_p}$ is given by analytically solving

$$\overline{P}_p = \frac{1 + \varepsilon \sum_{j \neq i} P_c(i,j)(1 + r\overline{P}_p)}{1 + \varepsilon \sum_{j \neq i} P_c(i,j)(1 + r\overline{P}_p) + (k_0/k)}. \tag{2.9}$$

giving,

$$\overline{P}_p = \frac{-(1 + c\varepsilon(1-r) + k_0/k)}{2\varepsilon rc} + \frac{\sqrt{(1 + c\varepsilon(1-r) + k_0/k)^2 + 4\varepsilon rc(1 + \varepsilon c)}}{2\varepsilon rc}. \tag{2.10}$$

where $c = \sum_{j \in painter} \rho_s(j) * P_c(i,j)$. For a painter region of size $N = 5$ nucleosomes, $c = 12.83$ if $P_c(i,j) = 1/|j-i|$. Outside the painter region, since $\rho_s(i) = 0$ (no sequence-specific recruitment), $P(M)$ follows Eq. 2.6 but by replacing ε by $\varepsilon(1 + r\overline{P}_p)$ (triangles in Fig. 2.14). This solution can be improved in the range of high fluctuations (Fig. 2.13B), if we first compute $P(M_i)\ \forall i \in painter$ from Eq. 2.8 by assuming that $\delta_{j,M} = P(M_i)$ and by solving the corresponding closed set of quadratic equations $\forall i \in painter$. Then, the average value of these solutions $\overline{P}^*_p = 1/L \sum_{i \in painter} P(M_i)$ is introduced and the probability of M-state

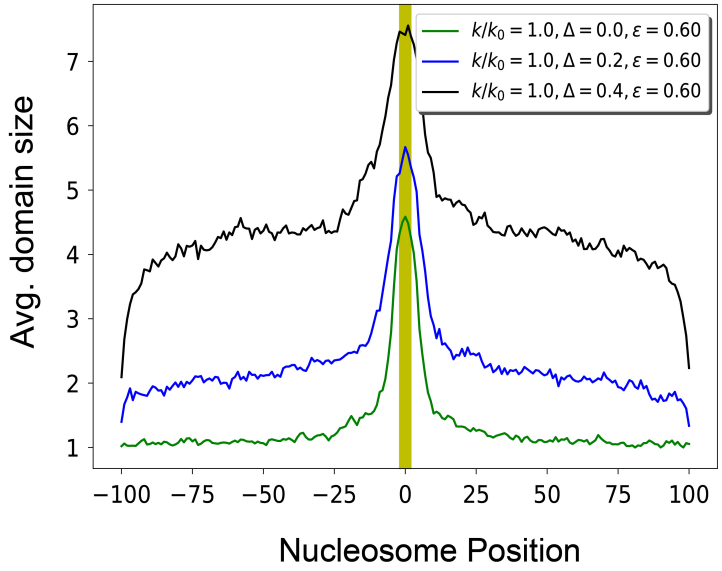

Fig. 2.4 Average size of 1D M-state domain as a function of genomic position. The green curve represents painter-mode ($r = \Delta = 0$, explained in Sec. 2.3.2). Blue and black curves illustrate the reader-writer mode ($\Delta > 0$, explained in Sec. 2.3.4) showing formation of 1D domains away from the painter region.

outside the painter region (circles in Fig. 2.14) is then given by replacing ε by $\varepsilon(1 + r\overline{P}_p^*)$ in Eq. 2.7.

2.2.4 Observables

In this section, we introduce some pertinent physical measures/observables to analyze and characterize different modes of spreading.

1D domains

A fundamental property of the system we want to study is the spreading of the state M along the genomic region of interest. Therefore, as an observable, we will focus primarily on the steady-state probability $P(M_i)$ of being in the modified state M at any position i, which can be calculated analytically (see sec.2.2.3) or numerically. $P(M_i)$ generally can be compared to the Chip(-seq) experimental profile and corresponds to an average measure of M-state spreading, i.e. spreading at the "cell population" level. To go beyond this average behavior and gain more insight into the process of spreading at "single cell" level, we also extract the set of the 1D coherent M-state domains for each configuration, considering a coherent domain as a region/nucleosomal array formed by consecutive M-state nucleosomes (1D-connected M-state nucleosomes). For each nucleosome i, we then compute the size of the 1D domain it belongs to and further derive the mean value over all configurations

(Fig. 2.4) elucidating local 1D epigenomic domains around a painter region (region of sequence-specific recruitment of HMEs).

3D Contact domains

Importantly, in our model, we consider spreading in cis (NN model) but also in trans (in 3D) by introducing the spreading probability $P_c(i,j)$. We indeed aim at investigating the influence of 3D organization on the local state dynamics and its interplay with the different modes of spreading (painter-mode, boosted-painter, reader-writer). The local dynamics of the state at each nucleosome i and in particular the M-state dynamics now relies on the ability of every distant nucleosome j to spread the M-state mark (reader-writer mode) and not only of the NN $i-1$ and $i+1$. The more M-state nucleosomes, nucleosome i is "in contact" (at a spreading distance) with, the more stable its M-state is. Hence, for every 1D state configuration, and for every nucleosome i being in the M-state, another pertinent observable is the 3D connected M-state domain or 3D "contact" M-state domain it belongs to. Formally, such domain is formed by all M-state nucleosomes in contact with nucleosome i in 3D.

In our model, there is no explicit simulation of 3D organization such that we cannot, for every configuration, extract a contact domain associated to every nucleosome i by explicitly enumerating the M-state nucleosomes in the 3D neighborhood/in contact. Still, we can generate effective contact domains that are consistent with the mean-field P_c which indeed represents how frequent pairs of nucleosome are in "effective contact". For that, we randomly generated for every 1D state configuration an undirected graph (M_{net}) for the corresponding set of M-state nucleosomes (Fig. 2.5): for all pairs (i,j) of (M-state) nucleosomes in this set, an edge between them is inserted in M_{net} with a probability $P_c(i,j)$. Then, we infer the disconnected subgraphs of M_{net} that would represent the different clusters of connected M-state nucleosomes, i.e., the different effective contact domains associated with the 1D state configuration. Finally, for each node/nucleosome i, we compute the size of the "contact" domain it belongs to. This provides a measure of M-state contact domain at the "single cell level".

In the reader-writer mode, an M-state nucleosome may influence a U-state nucleosome via P_c. Thus, to behave coherently, a genomic region may not necessarily need to have all nucleosomes in the M-state, but rather that all M-states belong to the same graph and that all U-states can be reached via $P_c(i,j)$ and are connected to such network. To quantify this, for each configuration, we consider the largest M-state subgraph $M_{net}*$ in M_{net}. For each U-state nucleosome i, random links with any M-state nucleosome j of $M_{net}*$ are generated with probability $P_c(i,j)$. i is added to $M_{net}*$ if it has been linked to at least two M-state nodes of $M_{net}*$ (Fig. 2.5, Fig. 2.17C). The average size of such extended contact domain

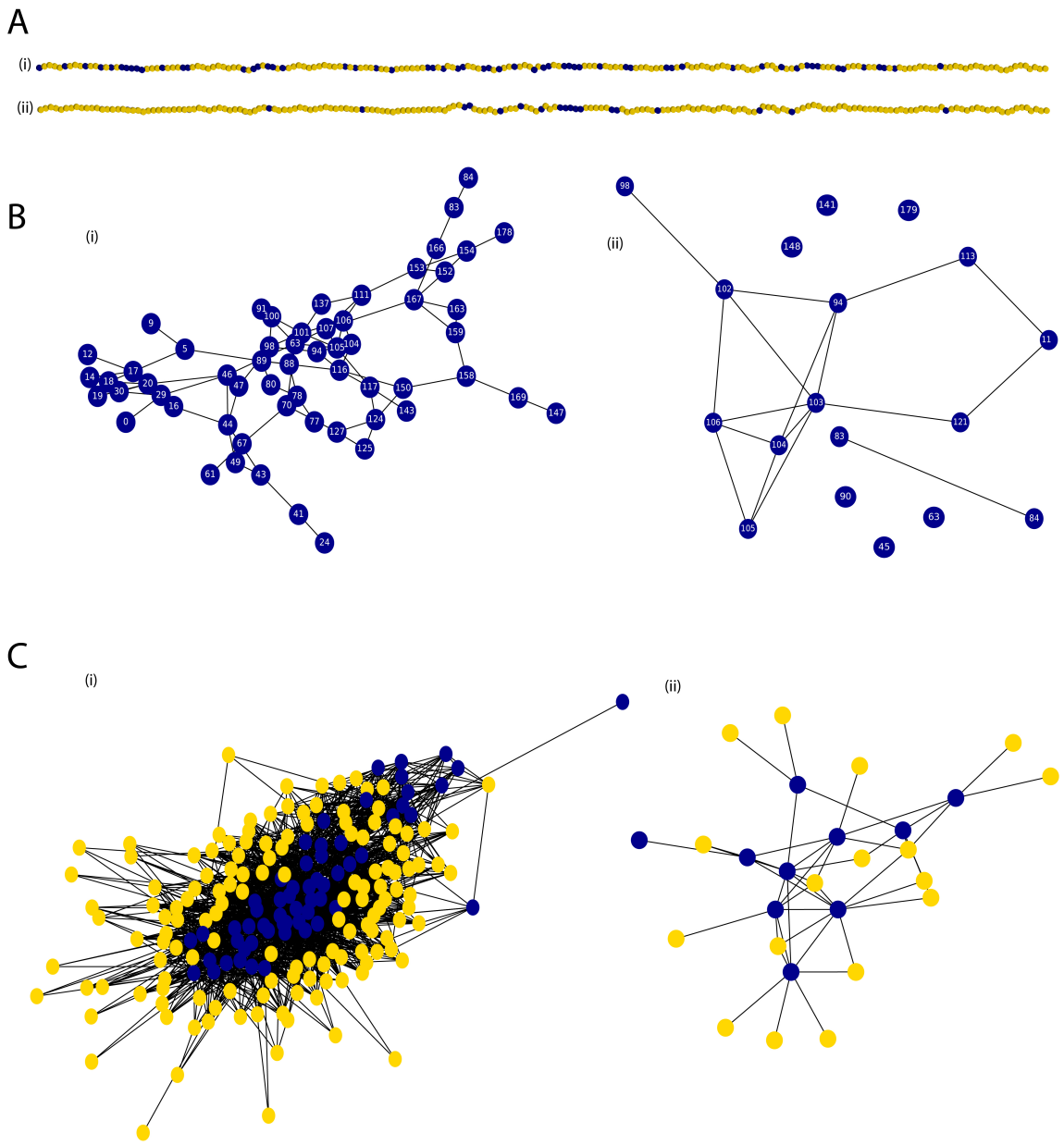

Fig. 2.5 (A) Two examples of the epigenomic state inside the region of interest (blue: M-state; yellow: U-state). Simulations were performed with $k/k_0 = 1$, $\varepsilon = 0.6$, $P_c(i,j) = 1/|i-j|$, and (i) $\Delta = 0.2$ and (ii) $\Delta = 0.0$. (B) For each example in (A), we plot one possible realization of the contact graph (M_{net}). In (i), all the M-state nodes are connected inside one subgraph of size 59. In (ii), there are multiple disconnected M-state subgraphs, the largest one being of size 10. (C) For the largest subgraph shown in (B), we plot one possible realization of an extended domain $M_{net}*$ including U-state nucleosomes. The blue nodes form the (modified) contact domain and the yellow nodes are unmodified nucleosomes in close proximity in 3D to the nucleosomes in largest M-state network $M_{net}(*)$. The extended domain size (total number of nodes) is (i) 198 and (ii) 27 in the two different cases (it is to be noted that the 3D extended domains illustrated in (i) and (ii) can be compared to 1D domains discussed in Fig. 2.4 blue and green curves respectively, elucidating how significant 3D genome organization is in the process of M-state spreading).

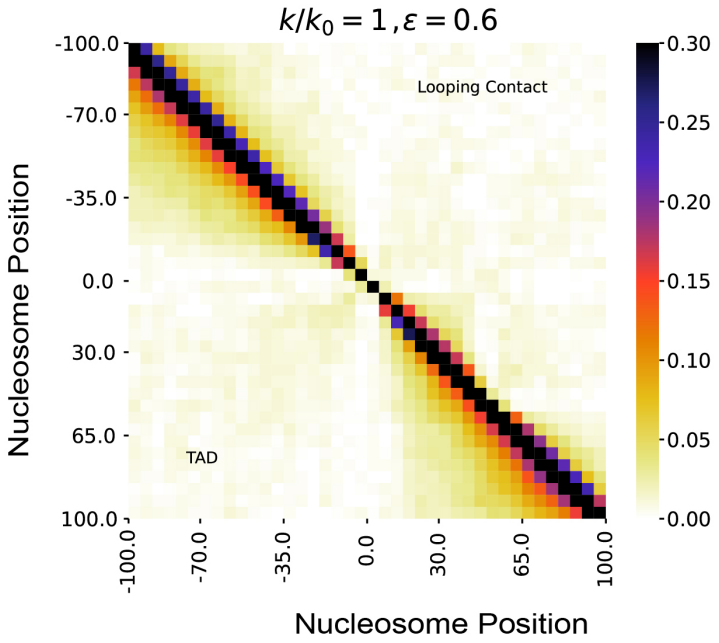

Fig. 2.6 Nucleosome state correlation in the compact domain (lower part) and looping (upper part) cases in the reader-writer mode (parameters:$k/k_0 = 1$, $\varepsilon = 0.6$) at the critical value $\Delta_c = 1.2$ for contact domain and $\Delta = 0.3$ for the loop case.

can thus be considered analogous to a percolation parameter [122], indicative of the extent of M-state spreading in the region.

The “reader-writer” mode coupled to spreading in trans, leads to a phase transition from a weakly modified state ($\overline{M} \sim 0$) to a fully modified state ($\overline{M} \sim 1$) when increasing recruitment strength Δ. To better capture how 3D drives this transition, it might be instructive to consider the evolution of the average size of extended contact domain (as defined before) as an observable (rather than $\overline{M}$). This characterization actually relates to the well known percolation theory[1]. In our system, network nodes are the the nucleosomes, links correspond to the effective 3D connections between pairs of nucleosomes (obtained from $P_c(i, j)$) and clusters are the contact domains.

Nucleosome state correlations

To compute the spatial (genomic) correlations between nucleosome states, we divide the genomic region into 40 bins of 5 nucleosomes. In one configuration, these bins are thus characterized by an M-state ranging from 0 to 5. Nucleosome state correlation matrix shown in Fig. 2.6, 2.10, 2.13, 2.15C represents the Pearson correlation between the M-

[1] In statistical physics and mathematics, percolation theory describes the behavior of a network when nodes or links are added. This is a geometric type of phase transition, since at a critical fraction of addition the network of small, disconnected clusters merge into significantly larger connected, so-called spanning clusters. *Cit. Percolation theory-Wikipedia*

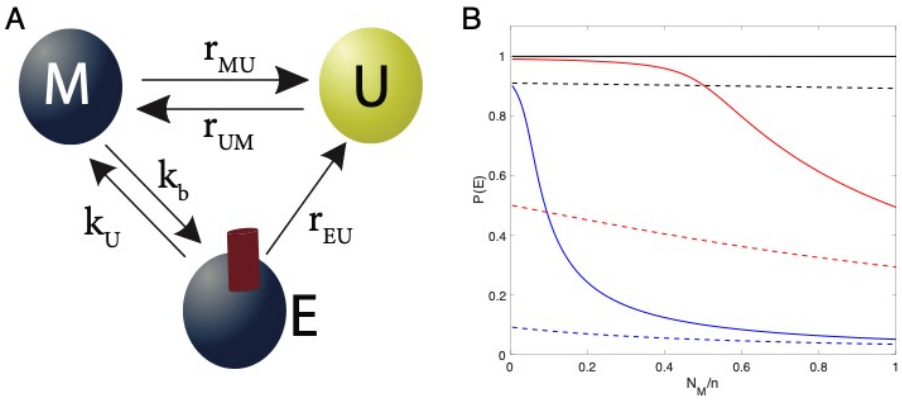

Fig. 2.7 Enzyme limitation. (A) Scheme of the enzyme limitation model. A modified nucleosome (M) can be in one of the two states: HME-free and HME-bound (E). The switching rate between these states is driven by the HME binding rate k_b and the corresponding unbinding rate k_u. We assume that the propensity to switch to the unmodified state U is similar in M or E states ($r_{EU} = r_{MU}$). U-state is assumed to be HME-free. (B) Probability $P(E)$ to be in the E-state for a modified nucleosome (Eq.2.11) as a function of the proportion of modified nucleosomes in the region N_M/n, for $N_e^{tot} - \sum_j \rho_s(j) = 1000$ (black), $= 100$ (red) and $= 10$ (blue) and $k_b/k_u = 1$ (full lines) and $= 0.01$ (dashed lines).

states computed for each pair of nucleosomes over all the configurations extracted from the simulations.

2.2.5 Enzyme limitation model

In the model described previously, we assumed that the concentration of HME is large enough to not have to consider the depletion of the pool of freely-diffusing HMEs (that impacts the sequence- and state-specific recruitment strength) by bound HMEs. Here, we consider a scenario where the number of enzymes is actually limited (Fig.2.7A).

In this model, the modified state can recruit HMEs explicitly and thus may switch between an enzyme-free state M and an enzyme-bound state E, the transition rates being dependent on the binding-unbinding kinetics of the enzyme k_b, k_u (Fig. 2.7A). Enzyme bound state has the ability to propagate modifications to neighboring nucleosomes in 3D. Assuming fast binding-unbinding enzyme kinetics, we can actually simplify the formalism into a two - state system. Indeed, we can write $k_b N_e(1 - P(E)) = k_u P(E)$ with $P(E)$ the probability for a modified state to be in the E state, $N_e = (N_e^{tot} - \sum_j \rho_s(j) - \sum_j \delta_{j,E})$ the current number of unbound enzymes where N_e^{tot} is the total number of HMEs, $\sum_j \rho_s(j)$ is the number of sequence-specifically bound enzymes and $\sum_j \delta_{j,E}$ the number of HMEs bound to M-states. By approximating $\sum_j \delta_{j,E} \approx \left[\sum_j \delta_{j,M/E}\right] P(E)$ the following expression

is obtained for $P(E)$:

$$P(E) = \left[2N_M/(N_e^{tot} - \sum_j \rho_s(j))\right]^{-1} \left[1 + \frac{N_M}{N_e^{tot} - \sum_j \rho_s(j)} + \frac{k_u}{k_b(N_e^{tot} - \sum_j \rho_s(j))} - \right.$$
$$\left. ((1 + \frac{N_M}{N_e^{tot} - \sum_j \rho_s(j)} + \frac{k_u}{k_b(N_e^{tot} - \sum_j \rho_s(j))})^2 - \frac{4N_M}{N_e^{tot} - \sum_j \rho_s(j)})^{1/2}\right] \tag{2.11}$$

with $N_M = \sum_j \delta_{j,M/E}$ the current total number of modified (M and E) states nucleosome in the region.

In the two-state formalism developed above, $P(E)$ is then used to re-normalize the state-dependent term in Eq. 2.2 to have accounted for enzyme limitation in the model. In this case, the modified state represents a lump state consisting of the M-state and the enzyme bound state E and such system is equivalent to our simple model (Eq. 2.2) but by replacing the state-dependent parameter Δ by $P(E)$ the probability that a HME is bound to a modified nucleosome. $P(E)$ explicitly depends on N_e^{tot}, the total number of HMEs, k_b and k_u the binding and unbinding rates (Fig.2.7B). In the limit of a non-limiting number of enzymes ($N_e^{tot} \gg 1$), $P(E)$ is almost constant with $P(E) \approx k_b N_e^{tot}/(k_u + k_b N_e^{tot}) \equiv \Delta$. For unstable enzymes ($k_b/k_u \ll 1$), $P(E)$ is less sensitive to fluctuations in N_M. In the rest, to simplify, we fix $k_b/k_u = 1$.

2.2.6 Transcription dynamics

Assimilating the modified state M to a silencing state, we assume that the instantaneous transcription rate α of a gene localized inside the region of interest is negatively impacted by the current proportion of M-state nucleosomes inside the promoter region. More precisely, we use the cooperative switch model of transcription (Fig. 2.8), proposed by Zerihun *et al.*[23].

$$\alpha(\overline{M}) = \alpha_0 \frac{\tanh((\overline{M} - \overline{M}^*)/d) - \tanh((-1 - \overline{M}^*)/d)}{\tanh((1 - \overline{M}^*)/d) - \tanh(-1 - \overline{M}^*)/d)} + \alpha_1 \tag{2.12}$$

where $\alpha_0 + \alpha_1$ is the maximal transcription rate, α_1 is the minimal rate (leaky transcription). $\overline{M} = (1/L)\sum \delta_{i,M}$, the current proportion of modified state in a L nucleosome-wide promoter region. $\overline{M}^*$ is the critical value of modification above which the gene is repressed and d defines the sharpness of the transition between activation and repression.

Gene expression dynamics is simulated along with the chromatin state dynamics ($\{r_{UM}(i), r_{MU}(i)\}$) within the same Gillespie simulations by considering, the transcription of mRNA molecules at rate α (Eq. 2.12), their degradation at rate β, their translation into proteins at rate Γ, the degradation of proteins at rate Ω and the maturation of proteins

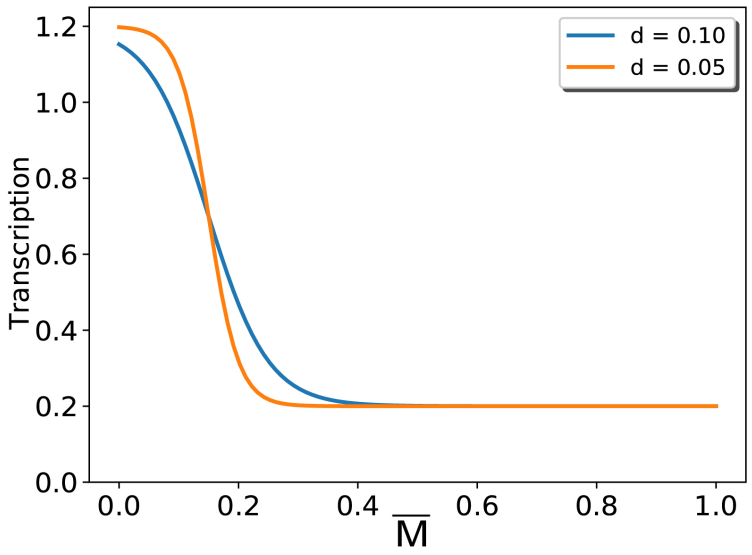

Fig. 2.8 Cooperative switch model: Transcription rate as a function of M state proportion ($\overline{M} = \frac{1}{L}\sum \delta_{i,M}$, Eq.2.12) with $\overline{M}^* = 0.15$ ($L = 5$), $\alpha_0 = 1$ min^{-1}, $\alpha_1 = 0.2$ min^{-1} for different steepness d.

at rate μ [124]. In Fig. 2.25, we fix $n = 5$, $L = 5$, $\alpha_0 = 1$ min^{-1}, $\alpha_1 = 0.2\alpha_0$, $d = 0.05$, $\overline{M}^* = 0.15$, $\beta = 0.1\alpha_0$, $\Gamma = \alpha_0$, $\Omega = 0.3\alpha_0$ and $\mu = 0.1\alpha_0$.

2.2.7 Polymer simulations

To estimate P_c in the cases of a central, strongly-self-interacting domain of nine nucleosomes and of a long-range loop between the painter region and a distal region, we perform simulations of a simple, isolated self-avoiding polymer (composed by 201 beads of diameter 10 nm) using the lattice kinetic Monte-Carlo model developed in [90] (also detailed in Chapter 3). To simulate the single 3D compact domain (Fig. 2.2 lower part), we use a self-attraction of $-1kT$ between the nine nucleosomes surrounding the painter region. The same framework is used for the looping case (Fig. 2.2 upper part) with attractive interaction ($-1kT$) between monomers of the painter region (positions: $[-2:2]$) and monomers at positions $[48:52]$. To obtain the $P_c(i,j)$ matrix in each case, we simulate 128 independent trajectories by first letting the system to equilibrate during $9*10^7$ Monte-Carlo time Step (MCS) before taking measurements every 10^5 MCS. From the ensemble of configurations, we thus estimate the contact probability $P_c(i,j)$ between any pairs of monomers as the probability that the relative distance between i and j is less than 40nm.

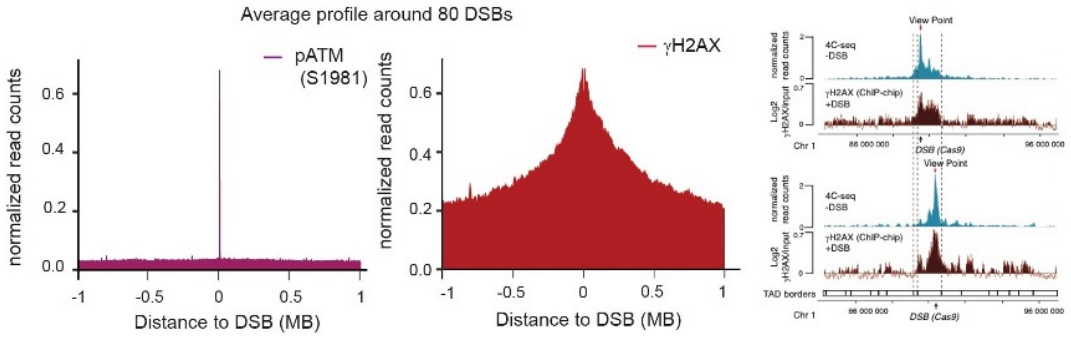

Fig. 2.9 (Left,Center) Average experimental Chip-seq profiles around DNA double strand breaks (DSB) of ATM kinase enzyme (left) and the average profile of H2A.X phosphorylation (center). The clear peaked ATM signal suggests that the enzyme is localized at the break, while phosphorylation extends over Mbps. (Right) Example of Chip-seq profiles around specific DSB in parallel with the 4C signal obtained from a point-of-view centered around the DSB. *Adapted from [120].*

2.2.8 Analysis of experimental data

In Fig. 2.26A, the average experimental profile $S(i)$ of H2A.X phosphorylation in human cells around double strand breaks is extracted from Arnould *et al.* [120] (Fig.2.9). We fit the experimental data by the chromatin state model in the painter mode ($r = \Delta = 0$) with one painter (ATM kinase) at position 0 (see Fig. 2.9,left) for a $\pm$1Mbp-region around the break by using Eq. 2.7 and assuming that $S(i) = AP(M_i) + B$ with A and B two constants. Parameter inference is done by minimizing the L2-distance between predictions and experiments.

Similarly, in Fig. 2.26B, the average density of H3K9me3 modifications flanking a transposable element in mESC cells is obtained from Robollo *et al.* [125]. Fit to experimental data using Eq. 2.7 is performed for a 10kbp-region with a centered painter region of size 7 nucleosomes corresponding to 1kbp transposable element. Parameter inference is done by minimizing the L2-distance between predictions and experiments.

In Fig. 2.26C the distribution of fluorescence of the reporter gene as a function of time and the corresponding fractions of “OFF” cells are extracted from Ragunathan *et al.* [10] for different strains. Modeling of the system is performed using the chromatin state+transcription model (see above) with $k = 0.05$, $\varepsilon = 0.6$, $L = 5$, $\alpha_0 = 1$, $\alpha_1 = 0.2\alpha_0$, $d = 0.05$, $\overline{M}^* = 0.1$, $\beta = 0.1\alpha_0$ (in min^{-1}) and the same parameters for protein dynamics as above. A scaling is applied to convert the number of proteins p into a fluorescence level F ($F = p + C$ with $C = 380$). The fraction of OFF cells is defined by the fraction of cells whose fluorescence is below 530 as defined experimentally [10]. The extracted experimental fluorescence at 6 hours is not aligned with the fluorescence peak observed at other time points. Fit to experimental data is performed by minimizing $\chi^2 = \sum_{t_i}[(OFF_{pred}(t_i) - OFF_{exp}(t_i))/OFF_{exp}(t_i)]^2$, where $OFF_{pred}(t_i)$ is the predicted fraction of OFF cells at time

t_i and $OFF_{exp}(t_i)$ the corresponding experimental observation, using a simple grid-search algorithm to scan parameters (Fig. 2.28).

2.3 Results

2.3.1 A generic model of epigenomic regulation

To investigate in detail the main mechanisms driving epigenomic regulation, we have developed a simple generic two-state model (Fig. 2.1A) where the local chromatin state can fluctuate between an unmodified, neutral state U and a modified state M that may correspond to an active (e.g. H3K4me3 or H3K27ac) or repressive (e.g. H3K27me3 or H3K9me2/3) state. We model chromatin as a unidimensional string of nucleosomes, each nucleosome being in state U or M. We consider a ∼40-kbp-long genomic domain (201 nucleosomes), a size that typically encompasses epigenetically-regulated regions like transposable elements [126], gene promoters and enhancers, the Mating type locus or subtelomeres in yeast [127].

The stochastic dynamics of the state is driven by histone-modifying enzymes (HMEs) that deposit ("writers", transition from U to M) or remove ("erasers", from M to U) specific histone modifications [5, 10, 16]. For example, PRC2 via its methyltransferase subunits EZH1 or EZH2 catalyzes the methylation of H3K27 [2, 128]. Our model integrates the key mechanisms acting on HME recruitment or activity, focusing on "writer" enzymes and lumping all the processes participating in histone mark removal into one effective turnover rate k_0 (see 2.2). Briefly (Fig. 2.1B), we have delineated the chromatin association of HMEs into sequence- and state-specific contributions: (a) writers, in association with DNA-binding proteins, may localize around specific genomic location with probability ρ_s, or (b) HME may be recruited to M-state nucleosomes with efficiency Δ. Recruited HMEs at position i can then write the epigenetic mark on-site with a rate k (action *in cis*) but may also spread it to a distal nucleosome j with a rate $(\varepsilon k)P_c(i,j)$ (action *in trans*) [15], where $P_c(i,j)$ accounts for the capacity of two nucleosomes to interact and may depend on the exact spreading mechanism (see 2.2). In the following, we consider that bound HMEs may catalyze reaction in their 3D neighborhood, as evidenced experimentally for several epigenetic systems including H3K27 methylation [15, 135], H3K9 methylation [129], γH2AX phosphorylation around DNA double strand breaks [130] and $P_c(i,j)$ is thus defined as the average 3D contact frequency between two genomic regions. HME enzymatic activity may be also boosted by allostery by a factor r if bound to a M-state nucleosome [17, 128, 131]. State-specific effects (recruitment and allosteric boost) are the so-called reader-writer mechanisms that are thought to be crucial for the establishment and maintenance of many epigenomic states [16].

In the following sections, we will systematically dissect the role of sequence-specific recruitment, reader-writer processes and spreading probability ($P_c(i,j)$) in epigenetic regulation using extensive simulations of our generic two-state model.

2.3.2 Spreading by sequence-dependent recruitment of enzymes: the “painter" mode

We first investigate the contribution of a simple mode of spreading where enzymes bind only to specific genomic locations and then may spread epigenetic marks in the 3D vicinity, in absence of reader-writer mechanisms. In this case, HMEs can be considered as “painters" that sit at specific places along the genome and “paint" the surrounding chromatin with a given modification. As sequence-dependent recruitment of enzymes is essential for *de novo* establishment of epigenetic domains [13, 132] and as many writers do not carry a “reader” subunit (that may drive state-dependent effects), such mode of regulation, hereafter called the “painter" mode, is likely to be a major way of epigenetic spreading.

To illustrate this model, we consider a simple case where painters are only recruited to a single 1kbp-long locus located in the middle of the 40kbp-long region (yellow area in Fig. 2.10A). In this situation, the probability $P(M)$ to have a modified nucleosome at a given position can be derived analytically (see Eq. 2.8) and only depends on the ratio (k/k_0) between the cis-spreading rate k and the turnover rate k_0 and on the ratio ε between the trans- and cis-activities (Fig. 2.10A,B). For all positions, we observe that $P(M)$ is a gradually increasing function of both parameters (Fig. 2.10B, Fig. 2.12). In the limit of very low trans-efficiency ($\varepsilon \approx 0$), modified states are essentially confined within the recruitment region due to the remaining on-site writing activity. For larger ε values, $P(M)$ is peaked at the painter region and decays at large genomic distances. The stronger ε and k/k_0, the wider the peak in $P(M)$ profiles and thus longer the range of spreading of the M-state (Fig. 2.10A).

Actually, the observed decay of $P(M)$ outside the recruitment zone is translating the decay of contact probability P_c with the painters bound to recruitment zone. More generally, in the limit of low contact probability or low spreading efficiency ($(k/k_0)P_c\varepsilon \ll 1$), $P(M)$ is directly proportional to the spreading probability P_c (see Eq. 2.8) which, in our case, is taken to be $\propto 1/s^{\gamma}$ with s, the genomic distance to the painter region and $\gamma = 1$ (Fig. 2.11). This illustrates the direct relationship between the local 3D organization and the profiles of epigenomic marks around recruitment sites [134].

In the painter mode, the local chromatin state only depends on the position of the bound painters and does not feedback on the recruitment of HMEs or on their activity. This lack of cooperativity leads to the absence of correlation between the nucleosome states at two different positions along the region (Fig. 2.10C). To quantify the efficiency of the spreading mechanism and its capacity at the single-cell level, to form more or less expanded, coherent

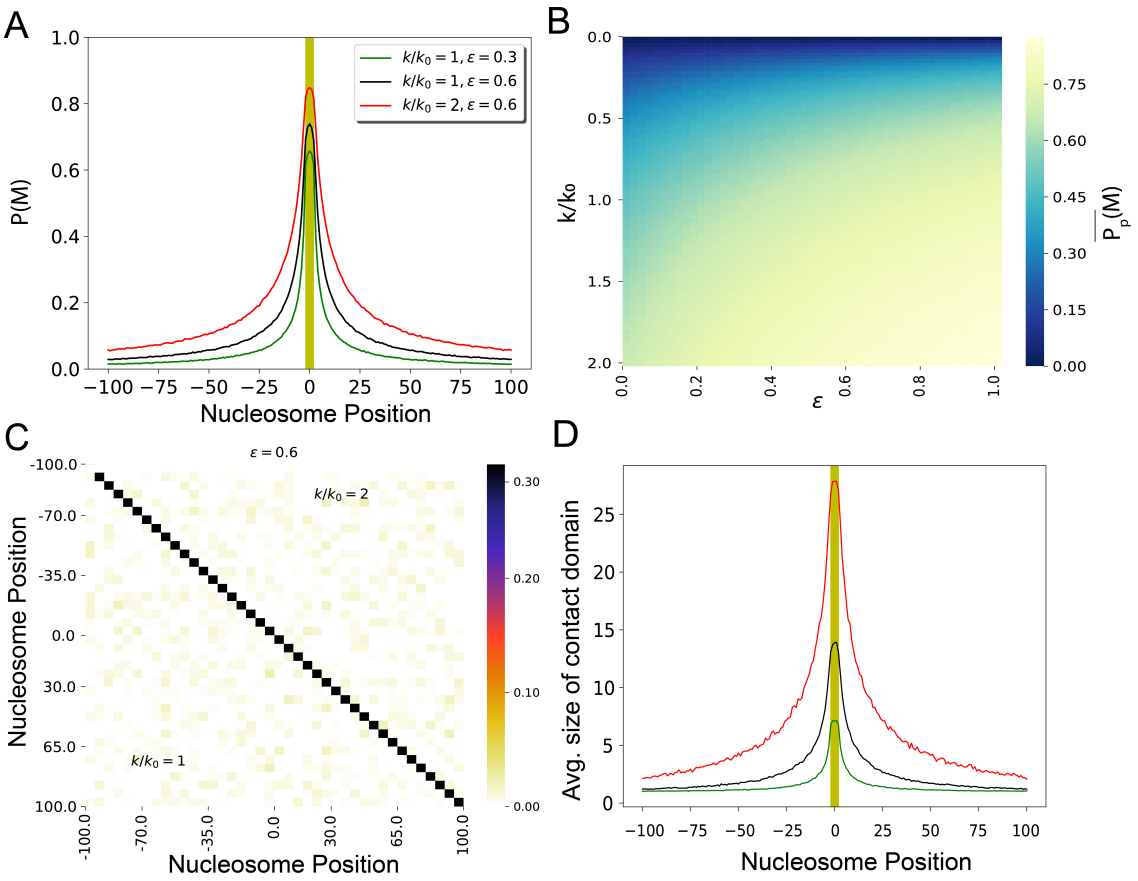

Fig. 2.10 Sequence-specific recruitment - Painter mode : (A) Virtual Chip-seq profiles of the M-state. Probability $P(M)$ to be modified as a function of the position around the 5 nucleosomes-long region where HMEs are recruited (yellow region) for different spreading efficiency ε and painter activity k/k_0. (B) Average probability $\overline{P}_p(M)$ of the M-state within the recruitment region as a function of ε and k/k_0. (C) Spatial correlation matrix between the nucleosome state at two positions (see 2.2.4) for $\varepsilon = 0.6$ and $k/k_0 = 1$ (lower part) or $= 2$ (upper part). (D) Average M-state contact domain size as a function of the nucleosome position for different ε and k/k_0 values.

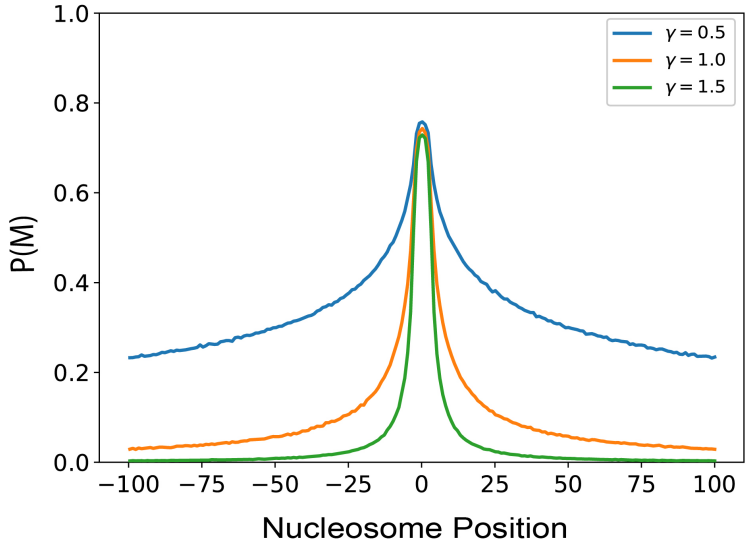

Fig. 2.11 Virtual Chip-seq profiles of the M-state as a function of the exponent γ in the painter mode. We show how chromatin compaction affects the process of spreading of M-state. A more compact chromatin region (lower γ-values) leads to more expanded profiles.

M-state domain, we estimate how nucleosomes in the M-state are effectively 'colocalized' in space. For each configuration and each M-state nucleosome, we compute the number of other M-state nucleosomes present in the same 3D contact domain (see Sec. 2.2.4). Fig. 2.10D shows that, on average, M-state nucleosomes are isolated from each other and only form part of small domains except around the painter region.

2.3.3 State-dependent enzymatic activity: the "boosted-painter" mode

In some systems such as Polycomb regulation (associated with the H3K27me3 modified state), it has been observed that the activity of the writers can be "boosted" by the presence of pre-existing modifications [17, 128, 131, 135, 136]. For example, the binding of the Polycomb writer PRC2 to H3K27me3 marked nucleosomes, triggers a boost in methyltransferase activity of its subunit EZH2 by allostery, via its other subunit EED. In our theoretical framework, this effect can be formalized by an increase in the writing rate of HMEs bound to a modified nucleosome via a multiplicative factor $r > 1$. For example, for mammalian PRC2, in vitro experiments suggested an allosteric boost of $r \lesssim 10$ fold in presence of H3K27me3 peptides [17, 136].

In Fig. 2.13, we characterize the impact of this boost on the simple painter mode described above (stably bound writers at a specific 1kbp-long region), still neglecting state-specific recruitment of HMEs ($\Delta = 0$). As shown in Fig. 2.13A, a weak ε value that essentially confines the M-state to the painter region in the simple painter mode ($r = 0$, green line), can be compensated by a strong boost term ($r = 20$, black line) with a significant

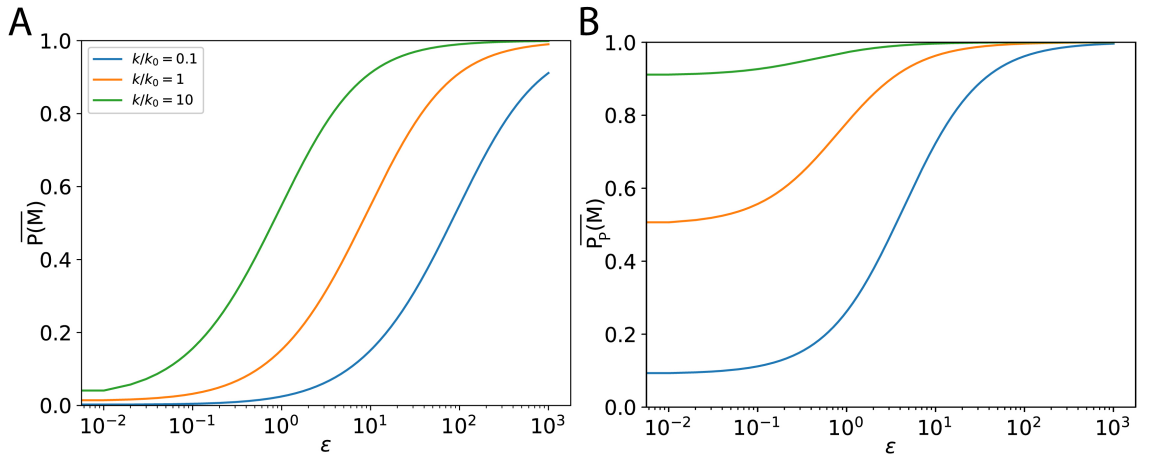

Fig. 2.12 Painter mode: the transition from low to high M-state as a function of the spreading efficiency (ε) for different k/k_0 and $P_c(s) = 1/s$. (A) Global average of $P(M)$. (B) Corresponding average value of $P(M)$ within the painter region.

overall enhancement of $P(M)$: the presence of modified nucleosomes in the painter region due to the "on site", cis-activity of painters (controlled by k/k_0) boosts globally the cis- and trans-spreading capacity of bound writers (via r) and favors further spreading inside and outside the painter region.

To get a better understanding of this "boosted-painter" mode, we first focus on the painter region by computing $\overline{P}_p(M)$, the average value of $P(M)$ inside this region (Fig. 2.13B). As expected, $\overline{P}_p(M)$ is an increasing function of k/k_0 and r. However, contrary to the simple painter mode (Fig. 2.12B), the transition from low to high M-state is a much sharper sigmoid function. Such switch-like behavior suggests a phase transition [28] inside the painter region that reflects the cooperative dynamics between writer-bound nucleosomes due to the state-dependent boost. The sharpness and position of the transition depends on k/k_0 and ε. High values lead (i) to smoother transition as the simple painter mode is significant enough to buffer the boost effect and (ii) to lower critical r-values as less boost is required to get high enzymatic activity.

In this spreading mode, the chromatin states of nucleosomes localized outside the painter region are now dynamically coupled to the ones inside, leading to positive spatial correlations that are maximal for r-values around the critical boost (Fig. 2.13C). Such coupling makes an exact analytical treatment of $P(M)$ intractable. However, using a mean-field approximation inside the painter region, we can derive an expression for $\overline{P}_p(M)$ (see section 2.2.3, Fig. 2.14). Outside the painter region, $P(M)$ is thus just as in the simple painter mode but for a boosted trans-activity $\varepsilon(1 + r\overline{P}_p(M))$ (blue and black circles in Fig. 2.13A). Hence, regarding the profile of epigenomic mark and of coherent 3D contact domain size (Fig. 2.13D), the boosted painter mode is indistinguishable from a simple painter mode with a greater "on site" and "off site" spreading rates (Blue dots, Fig. 2.13D). Only the presence

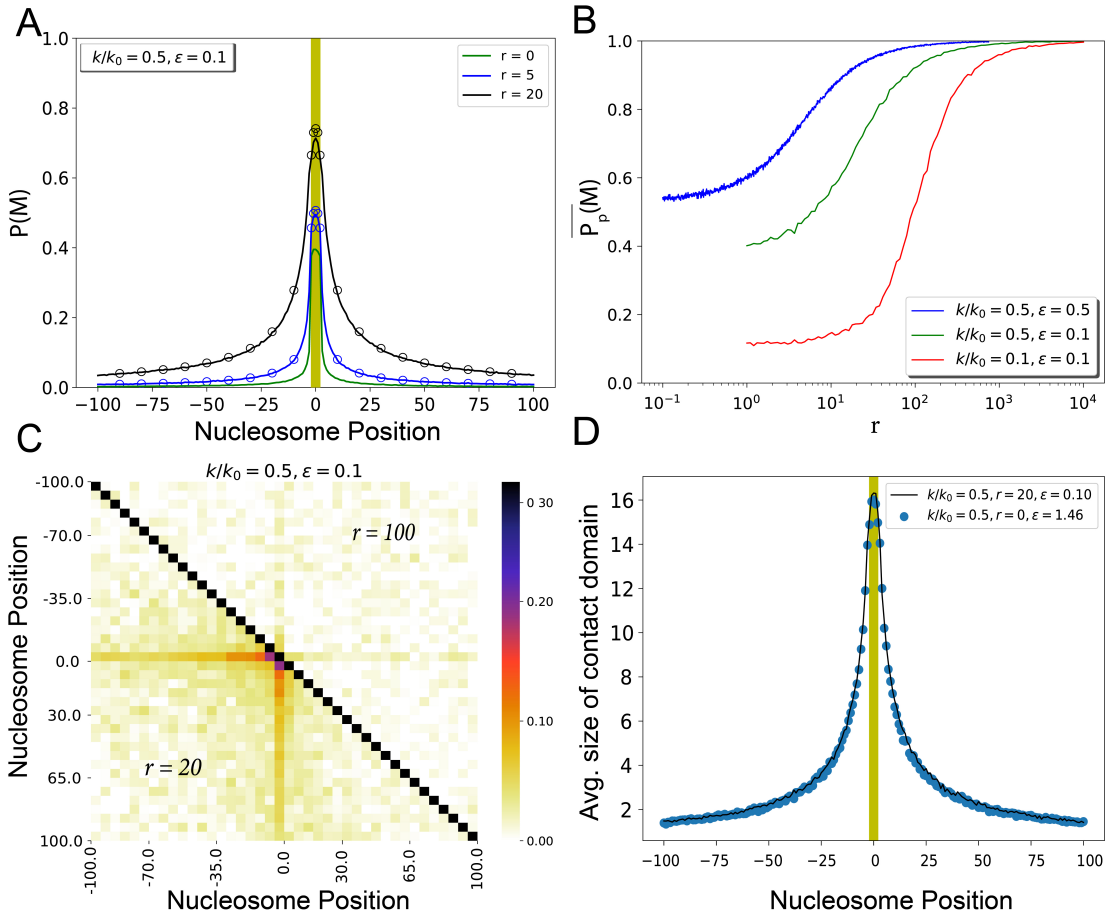

Fig. 2.13 State-specific enzyme activity - boosted painter mode : (A) Virtual Chip-seq profiles of the M-state. $P(M)$ as a function of the position around the 5 nucleosomes-long region where HMEs are recruited (yellow region) for different values of the boost factor r. The scatter points are the analytical solution of the model. (B) $\overline{P}_p(M)$ within the recruitment region as a function of r. (C) Spatial correlation matrix between the nucleosome state at two positions for $\varepsilon = 0.1$, $k/k_0 = 0.5$ and $r = 20$ (lower part) or $= 100$ (upper part). (D) Average M-state contact domain size as a function of the nucleosome position with boost $r = 20$ (full line). Dots correspond to a simple painter mode with an adjusted ε value to reproduce the $P(M)$ profile.

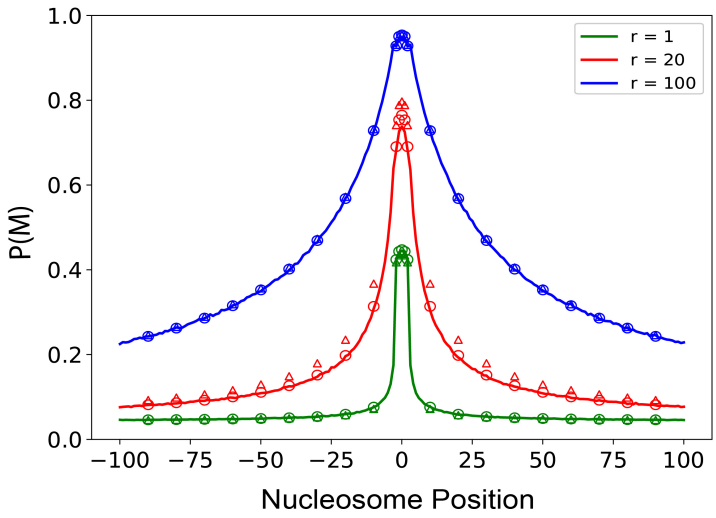

Fig. 2.14 Comparison between the analytical solutions and simulations in the boosted-painter mode. The different r values correspond to stable ($r = 100$) and highly fluctuating ($r = 20$) situations in Fig. 2.13B. The triangles are analytical solutions obtained by using $\overline{P}_p$ (from Eq. 2.10). The circles are obtained by using $\overline{P}^*_p$ to compute $P(M_i)$ outside the painter region which works better compared to the previous method.

of spatial correlations of nucleosome states with the recruitment region (Fig. 2.13C) might be a clear signature of this mode of propagation.

2.3.4 State-specific recruitment of enzymes: the “reader-writer" mode

Some writer enzymes have been shown to be recruited in a state-dependent manner to chromatin, having the ability to “read" (i.e. to be recruited at) a particular histone tail modification and “write" the same on another nucleosome in the 3D vicinity [5, 10, 2, 3]. For example, the methyltransferase Clr4 (associated with heterochromatin formation and H3K9me2/3 modifications in fission yeast) contains a chromodomain that may trigger its recruitment by H3K9me2/3 [10]. In our framework, we introduce this “reader-writer” mode by assuming that the probability of finding a HME bound at a M-state nucleosome is enhanced by a factor Δ (see Sec. 2.2). In the following part of this section, we focus on this effect coupled to the simple painter mode neglecting possible state-specific boost ($r = 0$).

In that case, thanks to the long-range action of M-state bound writers, the state dynamics of every nucleosome is coupled to the states of all the other nucleosomes which is well illustrated in Fig. 2.15C by the global increase in spatial correlations. Such reader-writer mode introduces a positive feedback in the global M-state dynamics which has been shown to promote the formation and inheritance of extended, stable M state domains [19, 56]. As shown in Fig. 2.15A, when combined with the simple painter mode, state-specific

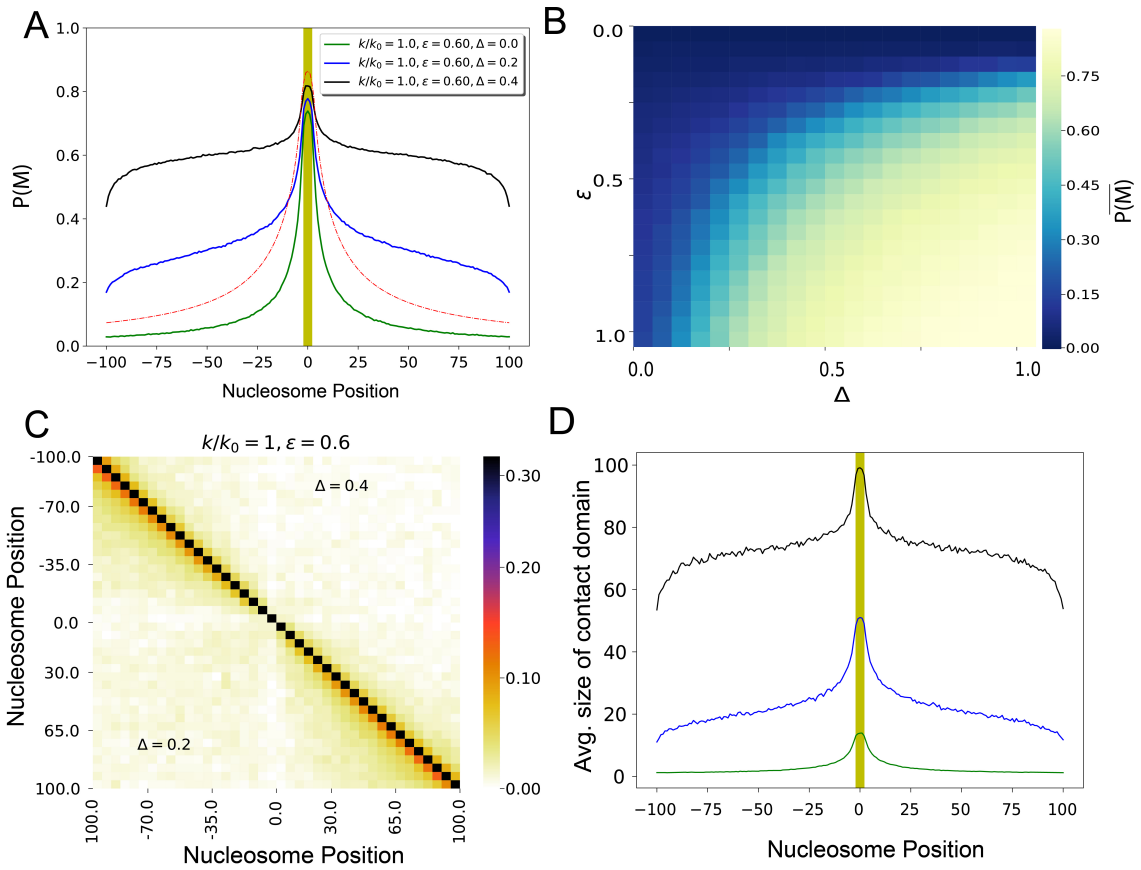

Fig. 2.15 State-specific recruitment of enzyme - reader-writer mode: (A) Virtual Chip-seq profiles of the M-state. $P(M)$ as a function of the position around the 5 nucleosomes-long region where HMEs are recruited (sequence specific) for different values of reader recruitment strength Δ (at the critical point $\Delta = 0.2$ and above $= 0.4$). (B) Average value $\overline{P(M)}$ of $P(M)$ inside the whole genomic region as a function of ε and Δ ($k/k_0 = 1$). (C) Spatial correlation matrix between the nucleosome state at two positions for $\varepsilon = 0.6$, $k/k_0 = 1$ and $\Delta = 0.2$ (lower part) or $= 0.4$ (upper part). (D) Average M-state contact domain size as a function of the nucleosome position for different Δ values. Colors as in (A).

recruitment strongly modifies the spreading pattern by notably increasing M-state occurrence away from painter region. In particular, we observe heavy tails for $P(M)$ (black and blue lines in Fig. 2.15A) that qualitatively differs from the typical $1/s^{\gamma}$ contact decay observed for the sole painter mode (red dotted curve in Fig. 2.15A). Such signature (heavy tails, deviation from simple painter model) in experimental profiles of epigenomic marks may thus be suggestive of a dominant reader-writer mechanism. In this regime, we also observe that the reader-writer process facilitates the formation of expanded 3D M-state contact domains away from the sequence-dependent recruitment region (fig. 2.15D) unlike the simple and boosted painter modes.

To quantify the resulting global increase in $P(M)$, we systematically compute $\overline{P(M)}$, the average value of $P(M)$ inside the whole genomic region (Fig. 2.15B). For a given trans-spreading activity (ε), we observe a sharp transition when state-specific recruitment efficiency (Δ) augments: from a pure simple painter profile localized around the painter region ($\overline{P(M)} \ll 0.5$) to a globally modified state ($\overline{P(M)} \gg 0.5$). As in the boosted-painter mode, this also suggests a phase transition but here, it reflects the cooperative dynamics between *all* nucleosomes of the region (see also next section). Around the critical Δ-value (e.g. $\Delta_c \approx 0.2$ for $k/k_0 = 1$ and $\varepsilon = 0.6$), fluctuations in $P(M)$ are maximal and lead to high inter-nucleosome correlations (Fig. 2.15C), for instance when $\Delta >> \Delta_c$ or $\Delta << \Delta_c$ inter-nucleosome correlation drops, see Fig. 2.16.

2.3.5 The spreading probability drives the percolation of the chromatin landscape

A key element of the different writing modes investigated in the previous sections is the capacity of recruited HMEs to spread an epigenetic signal [5, 10, 2, 3]. This property depends on the spreading probability $P_c(i,j)$ that captures the ability of two nucleosomes at positions i,j to interact. In the previous sections, we assumed that $P_c(i,j) = 1/|j-i|^{\gamma}$ with $\gamma = 1$, a generic scaling law accounting for a mechanism of spreading via 3D contacts as already observed for some HMEs like Polycomb group proteins (H3K27 methylation) [15], Clr4 methyltransferase (H3K9) [10], ATM kinase (γH2AX) [120] and characteristic of the polymeric nature of chromatin [147]. In this section, we explore the role of the shape of $P_c(i,j)$ in epigenomic regulation. This shape may depend on the specific spreading mechanism and on the experimental system or genomic region under study.

Here, we consider five alternative forms for $P_c(i,j)$. (1) Two forms that still correspond to a 3D contact spreading process but with contextualized, heterogeneous $P_c(i,j)$ (Fig. 2.2): one for a small 3D compact domain localized around the recruitment zone and another for a region with a strong loop between the painter area and a locus 10kbp away, mimicking, for example, a repressed MAT locus in yeasts and a promoter-enhancer loop in mam-

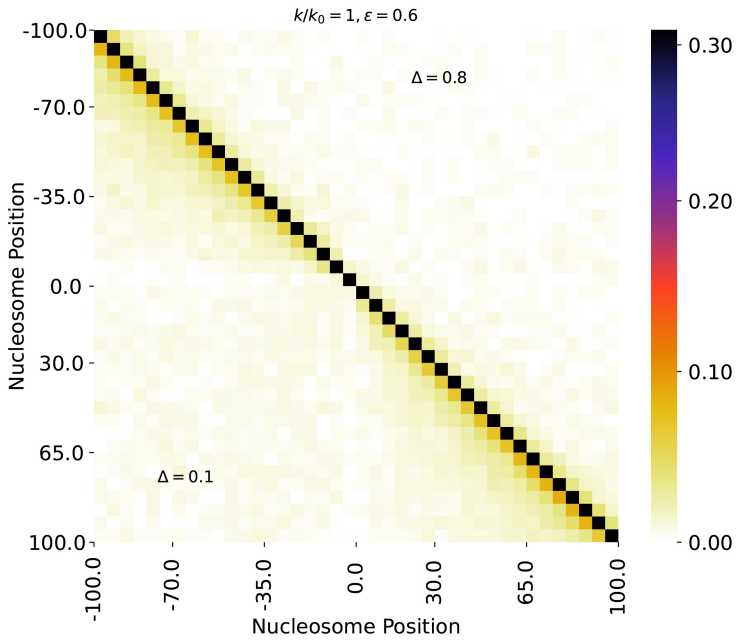

Fig. 2.16 Correlation matrices between the nucleosome state at two positions along the genomic region in the reader-writer mode, for $\varepsilon = 0.6$, $k/k_0 = 1$ and $\Delta = 0.1$ (lower part) and $\Delta = 0.8$ (upper part), $P_c = 1/s$. Reduced correlations are observed when Δ is far from the critical value ($\Delta_c \approx 0.2$ for such ε, k/k_0 values)

mals, respectively. (2) A shape accounting for a spreading to only nearest-neighbor (NN) nucleosomes that might be relevant in scenarios where an epigenomic state propagates unidimensionally along the chromatin [143, 25, 170]. (3) A form compatible with an effective contact spreading mechanism where HMEs recruited at a given position may diffuse in 3D to nearby positions and thus impact the state of distal nucleosomes ($P_c(i,j) = 1/|j-i|^{0.5}$). (4) A scenario where two nucleosomes may influence each other only if they have been placed in very close proximity ($P_c(i,j) = e^{-|j-i|/s_0}$) by loop extruding factors (eg, cohesin or condensin) that are molecular motors translocating along the chromatin and implicated in TAD formation in mammals [117–120] .

In the simple painter mode (Fig. 2.17, $\Delta = 0$), spreading from the central recruitment zone is limited by the shape of $P_c(i,j)$: long-range effective interactions like in the diffusion scenario lead to more extended $P(M)$ profiles compared to localized spreading as the NN case. In the 3D loop case, the M-state is able to stably propagate distally thanks to enriched 3D contacts with the painter area of sequence-specific recruitment.

In the reader-writer mode (Fig. 2.17, $\Delta > 0$), as observed in the previous section, state-specific recruitment allows a facilitated spreading and modifies the shape of $P(M)$. For a given reader-writer strength Δ, scenarios with longer-range $P_c(i,j)$ are more impacted. However, for all scenarios, it exists a critical Δ value (Δ_c) above which the M-state spreads

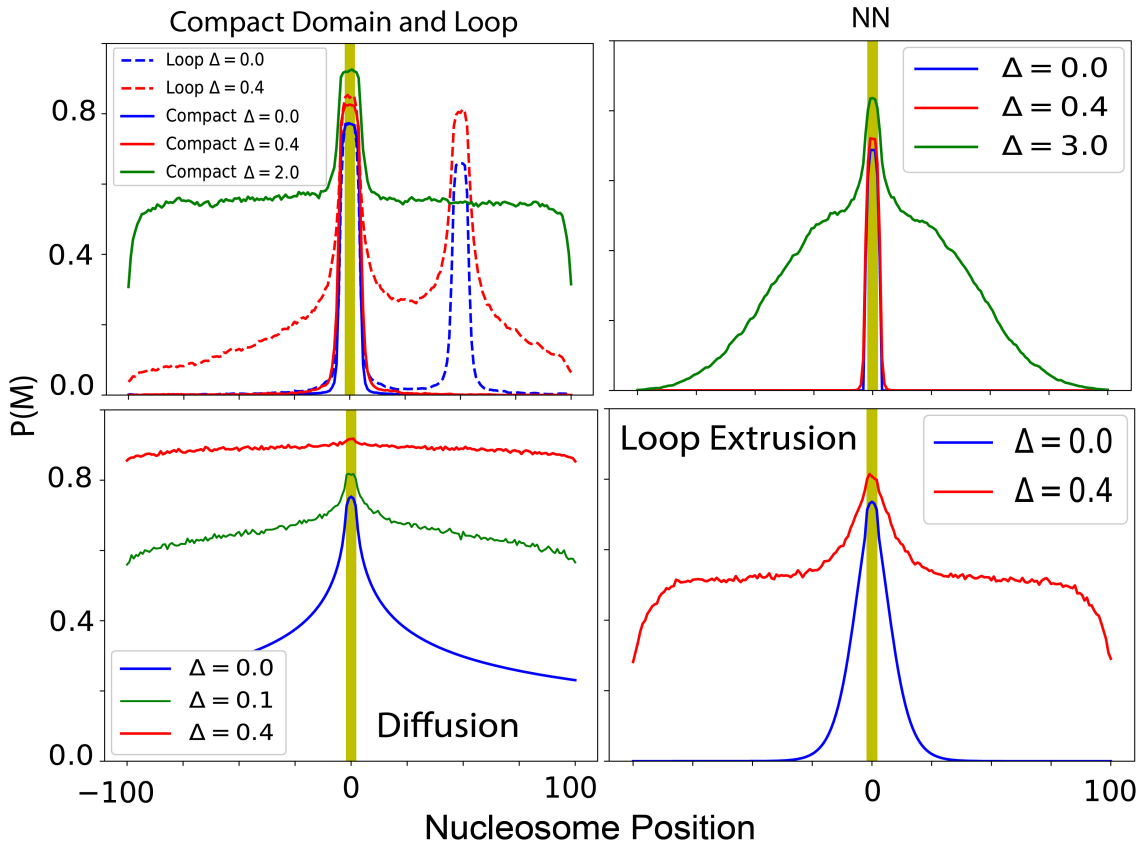

Fig. 2.17 Spreading Probability and $P(M)$. Virtual Chip-seq profiles of the M-state. $P(M)$ as a function of the position around the 5 nucleosomes-long region where HMEs are recruited (yellow) for different $P_c(i,j)$ and for $k/k_0 = 1$, $\varepsilon = 0.6$ with $\Delta = 0$ (painter mode, Blue) and $\Delta > 0$ (reader-writer mode, Red and Green). (1)$P_c(i,j)$ corresponding to a compact domain and loop (dotted lines). (2) Nearest neighbor spreading. (3) Diffusion limited spreading. (4) $P_c(i,j)$ driven by loop extrusion mechanism.

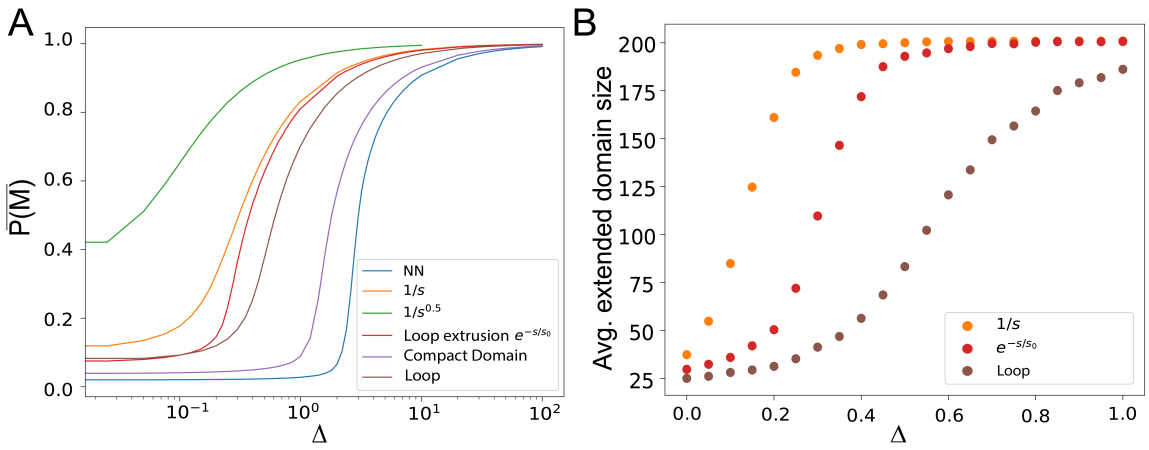

Fig. 2.18 Spreading probability and percolation. (A) Average value ($\overline{P(M)}$) of $P(M)$ inside the whole genomic region in the reader-writer mode as a function of Δ ($k/k_0 = 1$, $\varepsilon = 0.6$) for different $P_c(i,j)$. (B) Average size of the most extended generalized domain as a function of Δ for three situations described in (A).

over the whole region (Fig. 2.18A), more local $P_c(i,j)$ shapes (eg, NN or compact domain cases) exhibiting larger Δ_c values. Such transition driven by the reader-writer capacity of HMEs resembles actually to a generic percolation transition [28] where the system starts to be fully connected (or percolated) [122]. In our context, it corresponds to a situation where all the M-state nucleosomes form a unique expanded contact domain from which all the remaining U-state nucleosomes can be reached via $P_c(i,j)$ to allow spreading (see section 2.2.4, Fig. 2.5). To quantify this, for each configuration, we look how the most expanded M-state contact domain is connected to the unmodified regions. Fig. 2.18B shows the evolution of the average size of such extended generalized contact domains (containing U- and M-state nucleosomes) for three different scenarios. The global epigenomic state becomes percolated when this size approaches 200, meaning that the extended domain covers the whole genomic region (Fig. 2.19). Interestingly, a full overall M-state ($\overline{P(M)} \sim 1$) is not required to reach percolation, in particular for long-range spreading mechanism. For example, in the $P_c(i,j) = 1/|j-i|$ case, percolation transition occurs at $\Delta_c \approx 0.2$ where $\overline{P(M)} \approx 0.3$ (Fig. 2.19).

2.3.6 The reader-writer mode may lead to epigenetic memory

Having characterized how the various writing modes establish typical epigenomic profiles from the painter region, we now investigate how these different modes impact the "epigenetic memory" of these states. By definition, epigenetic memory stands for the abil-

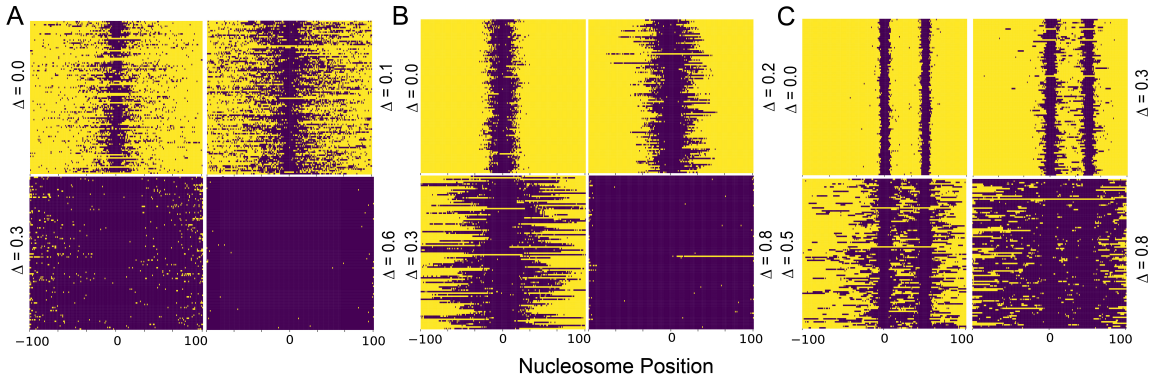

Fig. 2.19 Illustration of percolation of M-state. (A) For four different Δ values when $P_c(s) = 1/s$ (brown curves in Fig. 2.18), most extended generalized domains observed in different configurations (100 examples) are stacked together to show cell to cell variability (all nucleosomes belonging to a domain are colored in blue). (B) Same as (A) but for loop extrusion case, $P_c(s) = e^{-s/s_0}$. (C) Same as (A) but for loop case (Fig. 2.2).

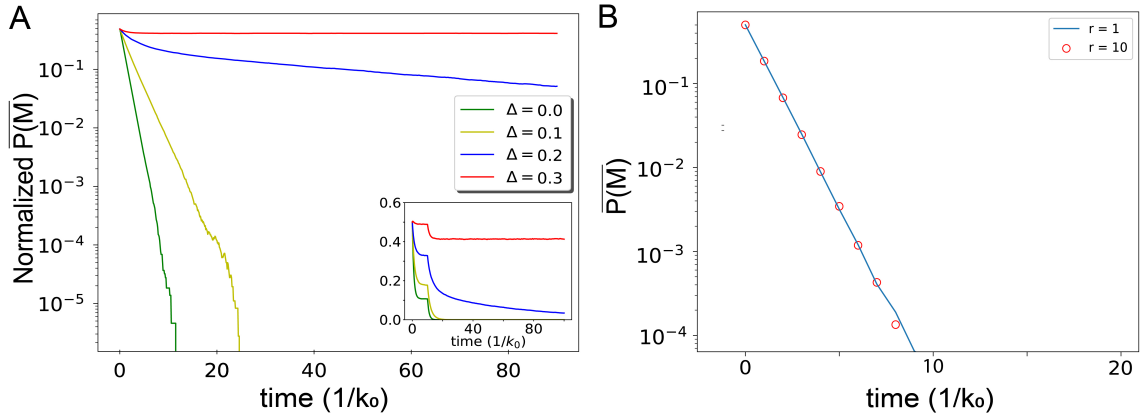

Fig. 2.20 Epigenetic memory: (A) Normalized time-evolution of $\overline{P(M)}$ after the unbinding of sequence recruited painters at $t = 0$ for different values of Δ and with $k/k_0 = 1, \varepsilon = 0.6$. Inset shows the non-normalized evolution. (B) Memory in the booster-painter mode. To investigate epigenetic memory, we study the evolution of $\overline{P(M)}$ after the sequence dependent HMEs unbind at $t = 0$. Even though boosted-painter mode has state-dependence within the painter region, once the sequence-dependent HMEs unbind there is no more cooperativity hence, the decay of $\overline{P(M)}$ is the same for different boost factor r values unlike in the reader-writer mode where there is a strong dependency in the parameter Δ (see (A)).

ity of maintaining a given transcriptional or chromatin state in the absence of the initial, sequence-dependent stimulus (e.g. transcription factors or HMEs).

For such purpose, focusing on a 3D spreading mechanism ($P_c(i,j) = 1/|j-i|$), we follow the dynamics of the mean genomic M-state $\overline{M} \equiv 1/n \sum \delta(i,M)$, once the sequence-specific recruited painters have been released (for examples of individual trajectories, see Fig. 2.21A). We then consider the time evolution of the ensemble-average $\overline{P(M)}$ of $\overline{M}$ ($\overline{P(M)} = \langle \overline{M} \rangle$) for different parameters (Fig. 2.20A). To be more general, we also track the time evolution $\overline{P(M)}_t$ of $\langle \overline{M} \rangle$ in absence of recruitment but for arbitrary imposed initial state with fixed $\overline{P(M)}_{t=0}$ (Fig. 2.21B).

In the simple painter mode ($\Delta = 0$, green line in Fig. 2.20A and left panel in Fig. 2.21B), the M-state relaxes very quickly to the U state whatever the initial state. Indeed, in absence of recruited writers in the painter area, only transitions from M to U are possible and the global state decays exponentially with a characteristic time $1/k_0$. Similar behaviors are observed in the booster-painter mode (Fig. 2.20B).

In the reader-writer mode, for low Δ values, the decay remains fast and exponential ($\Delta = 0.1$ in Fig. 2.20A, Fig. 2.21A) with a relaxation time that increases with Δ (orange dots in Fig. 2.21C,D). Above a critical value, state-dependent recruitment is strong enough to stably maintain a M-state after HMEs unbind from the painter region ($\Delta = 0.3$ in Fig. 2.20A, Fig. 2.21A) and to keep the memory of an initial (even small) M-state enrichment (Fig. 2.21B). The initial profile along the genome is lost and gives rise to an uniform spreading of the M-state (Fig. 2.22). In this case, even though the M-state is maintained, it is not correlated with the state at $t = 0$. In general, time correlation of M-state is quickly lost which means the reader-writer mechanism would be able to maintain the level of repression but not exactly the shape of the profile.

Interestingly, for Δ-values just below this critical point, the relaxation dynamics is better described by a two time-scale exponential (Fig. 2.21D) kinetics ($\Delta = 0.2$ in Fig. 2.20A, Fig. 2.21): a fast initial decay following the trend of the low-Δ regime (blue dots in Fig. 2.21C) and a slow decay at larger time (inset in Fig. 2.21C). In this regime, random $M \to U$ conversions dominate but the state-dependent recruitment allows self-maintenance of a coherent M-state for long time periods before reaching a global, absorbing U-state ($\overline{M} = 0$) from which the system cannot escape as also observed by [28]. Such maintenance time-period is very stochastic ($\Delta = 0.2$ in Fig. 2.20C) while, in the low and high Δ-regimes, convergence to steady-state behaves quite uniformly.

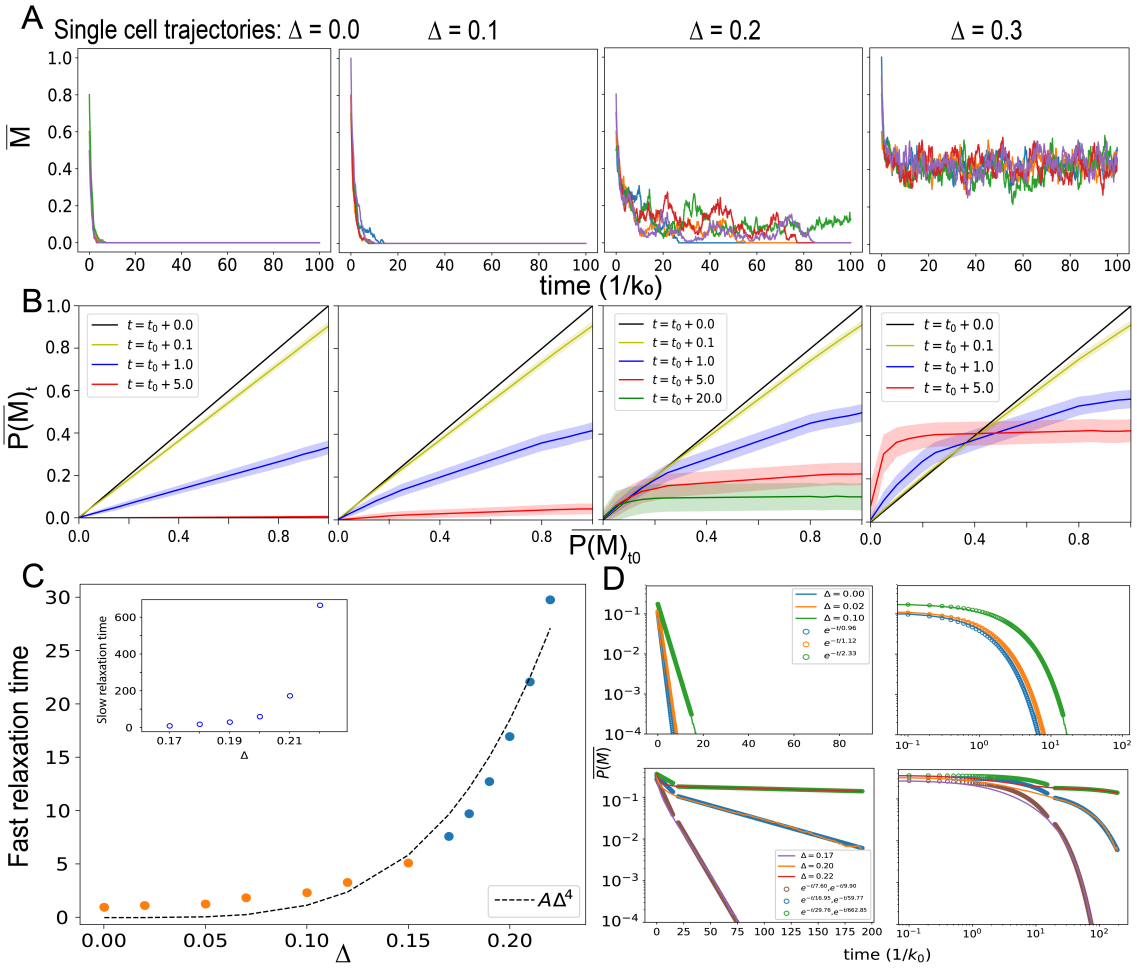

Fig. 2.21 (A) Examples of "single cell" time-evolution of the M state decay ($\sum \delta_{i,M}/201$ vs t) for different values of Δ. (B) $\overline{P(M)}_t$, the probability to be modified at a time t after the unbinding of painters, as a function of the imposed initial state with fixed $\overline{P(M)}_{t=0}$, for different values of Δ. (C) Fast (main) and slow (inset) relaxation times as a function of Δ obtained by fitting the curve in (A) by one or two time-scale exponential decay (see (D)). The fast time scales as $\approx \Delta^4$. (D) Decay of the average M-state value in the region shown for different Δ values in log-lin scale (i,iii) or log-log scale (ii,iv). Lines represent simulation results, dots correspond to the fit. (i,ii) For $\Delta << \Delta_c = 0.2$, the decay is purely exponential. (ii) (i) in log-log scale. (iii,iv) For $\Delta \leqslant \Delta_c$ the decay can be fitted by two time scale exponentials: a fast exponential decay at small time for $t \ll t*$ ($\overline{P(M)} = D_0 e^{-t/\tau_0}$) and a much slower exponential at large time for $t \gg t*$ ($\overline{P(M)} = D_1 e^{-t/\tau_1}$ with $\tau_1 > \tau_0$). Exact fit values are shown in the plot.

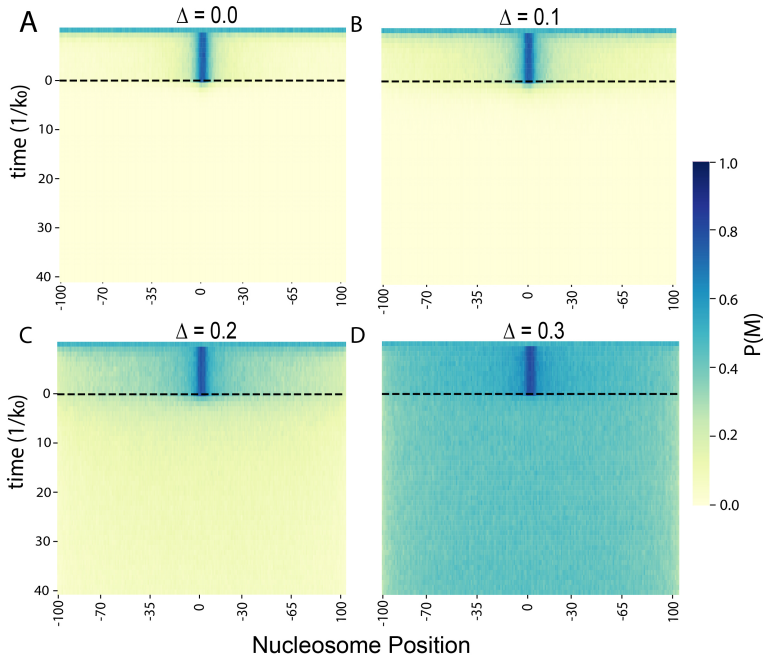

Fig. 2.22 Time evolution of $P(M)$ after the unbinding of the sequence-dependent painters at $t = 0$ (dotted black lines) for different values of Δ with $k/k_0 = 1$ and $\varepsilon = 0.6$ as shown in Fig. 2.20A inset. Here $\forall\Delta$, the final state is a uniform state with no sequence-dependency.

2.3.7 Maintenance of a confined chromatin state requires chromatin compaction and enzyme titration

In the previous section, we showed that epigenetic memory was possible for strong-enough state-specific recruitment. However, in this regime of parameter, the spreading of the M-state along the genome is not constrained [68], leading to unconfined memory [25] (Fig. 2.24A). Of course, experimentally, epigenomic domains are confined to specific regions in the genome and do not spread ubiquitously [10, 16]. In this section, we investigate what could be the minimal changes to our model to support confined memory and the maintenance of a localized M-state even in absence of sequence-specific recruitment.

Our simple epigenomic model assume implicitly that the number of HMEs is not limiting in the system. Recent quantitative proteomics experiments however suggest that such assumption might not be satisfied for some epigenomic marks like the Polycomb/H3K27me3 system during fly embryogenesis [137]. Therefore, a mechanism that could possibly restrict the spreading of M-state might be enzyme limitation. In the model, under the approximation of fast binding-unbinding enzyme kinetics at M-state nucleosomes, it is possible to account for enzyme titration via an effective Δ parameter that depends explicitly on the total number of HMEs N_e^{tot} and on the current fraction of bound HMEs, high fractions being associated with low effective state-dependent recruitment. Limiting the number of HMEs thus reduces the spreading efficiency (Fig. 2.23A). For a constant $k_b/k_u = 1$, $P(M)$

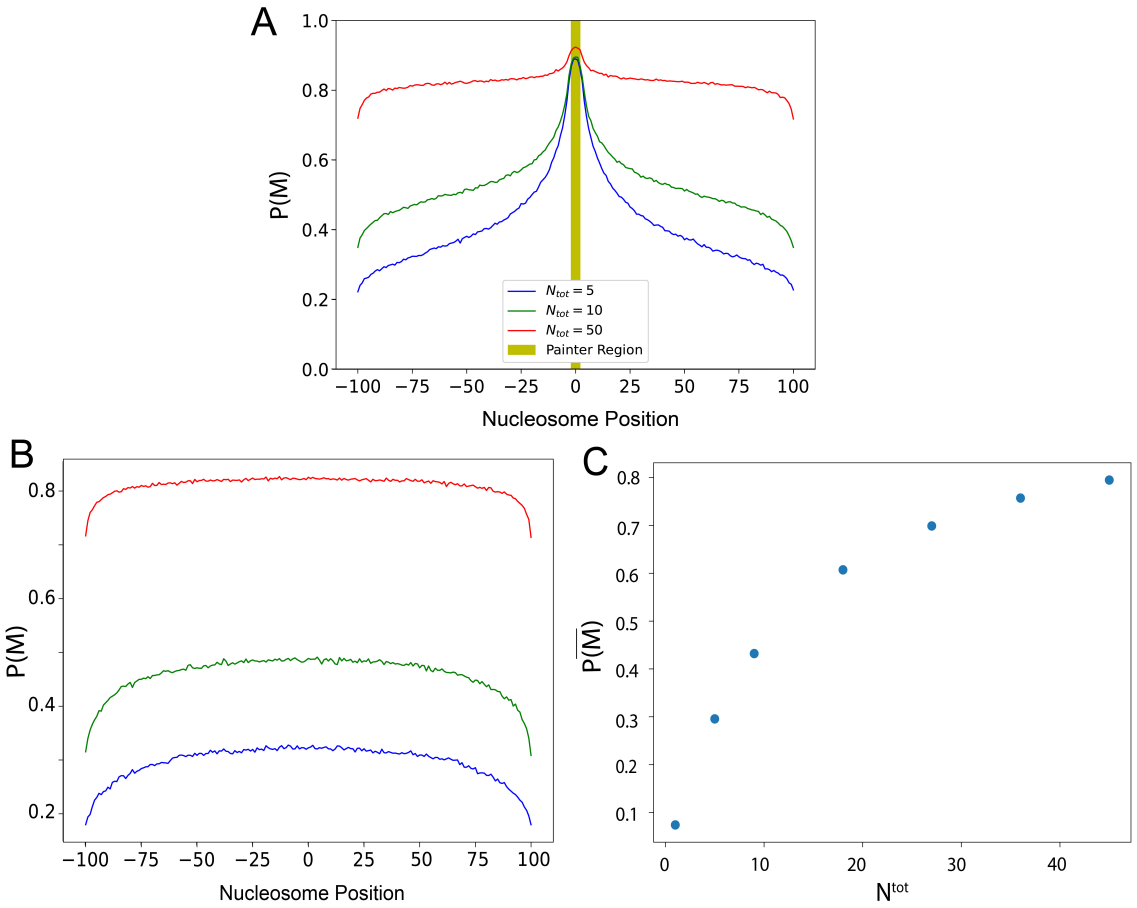

Fig. 2.23 Enzyme limitation. (A) Virtual Chip-seq profiles of the M-state (HME bound+HME-free modified state) with sequence-dependent recruitment at the painter region for different total number of HMEs N_e^{tot} with $k/k_0 = 2$, $\varepsilon = 0.9$ and $k_u/k_b = 1$. $P_c(s) = 1/s$. (B) The same profiles as in (A) but in absence of sequence-dependent recruitment, showing the unconfined uniform spreading of the M-state for all N^{tot} values. $P_c(s) = 1/s$. (C) Variation of $\overline{P(M)}$ as a function of N_e^{tot} in absence of sequence-dependent recruitment.

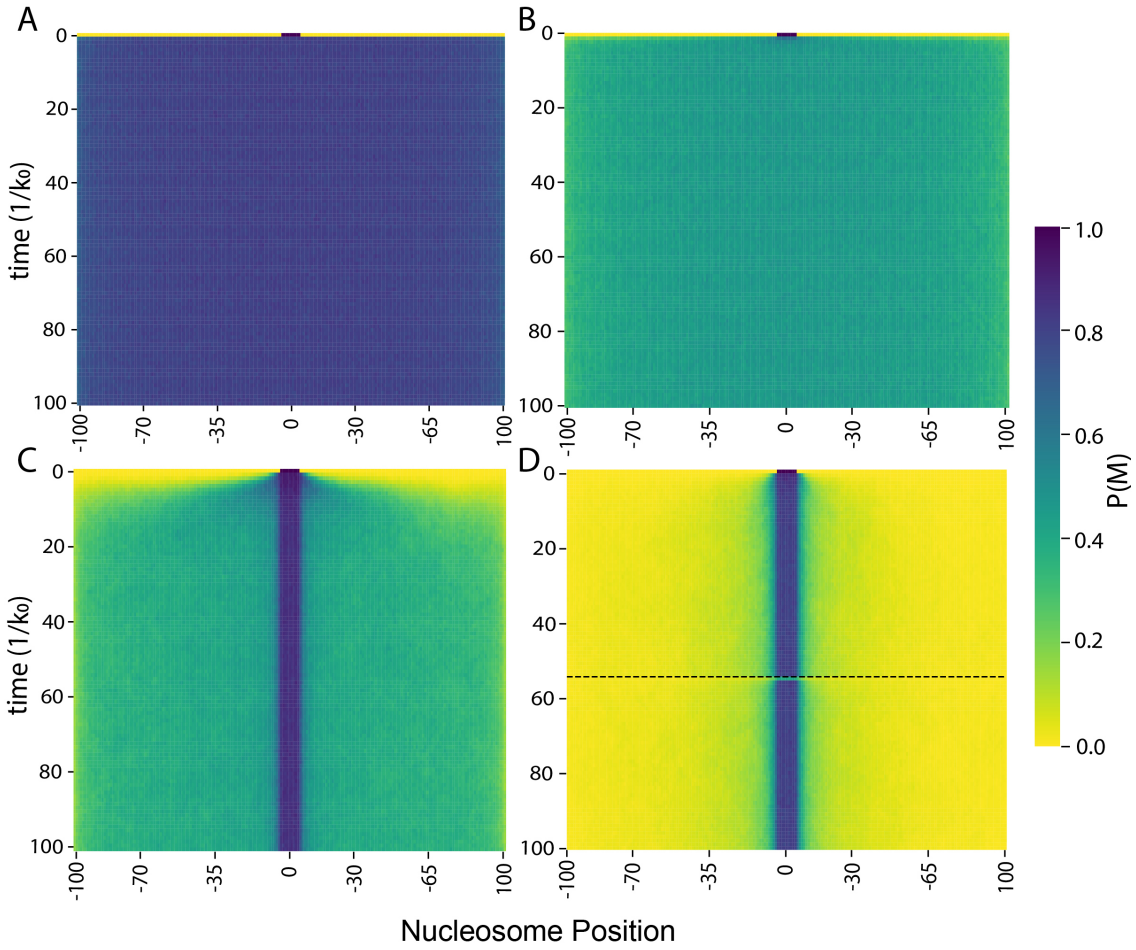

Fig. 2.24 Confined epigenetic memory. (A) Evolution of $P(M)$ in the enzyme non-limiting regime $N^{tot} = 45$. The initial configuration is a domain of modified states around nucleosome position 0 and there is no sequence-specific recruitment for $t > 0$. $k/k_0 = 2, \varepsilon = 0.9$. (B) As in (A) but with enzyme limitation, $N_e^{tot} = 9$, $k_b/k_u = 1$ (C) As in (A) but with 3D compaction of the domain (Fig. 2.2 shows the corresponding contact probability matrix), in large $N_e^{tot} = 45$ limit. (D) As in (A) but with enzyme limitation ($N_e^{tot} = 9$, $k_b/k_u = 1$) and 3D compaction. Modified state domain remains confined even under replicative dilution ($T_{cyc} = 55/k_0$, dashed black line).

in a region is dependent on the total number of HMEs. In the presence of sequence-dependent recruitment, enzyme limitation can confine the domains around the painter region (Fig. 2.23A) while in the absence of sequence dependent recruitment, we have a uniform steady state profile of M-state whose absolute value is dependent on N_e^{tot} (Fig. 2.23B,C). However in ranges of parameters that support epigenetic memory, the long-range-spreading activity of bound HMEs still lead to unconfined epigenomic domains (Fig. 2.24B).

As long-range spreading via the trans-activity of HMEs seems to promote unconfined memory, another possible mechanism to restrict spreading might thus be to modulate the 3D communication between loci [66]. Some epigenomic marks are associated with architectural proteins that may indeed impact chromatin organization. For example, PRC1 and HP1 that can bind H3K27me3- and H3K9me2/3-marked chromatin respectively, are known to promote chromatin compaction in vitro [138, 139] and in vivo [140–142]. To explore how 3D compaction of domains would influence epigenetic memory, we introduce a nucleosome-nucleosome contact probability $P_c(i,j)$ that is consistent with the formation of a compact domain around the painter region (see 2.2) with P_c inside the domain being stronger than outside. In the limit of high N_e^{tot} (Fig. 2.24C), such 3D organization clearly facilitates the maintenance of a stable M-state inside the compacted region even in absence of sequence dependent stimulus, but exhibits "flooding" of the M-state outside this region with time due to the residual spreading between nearest-neighbor sites. However, when we couple 3D compaction and enzyme limitation, the initial domain remains very stable with limited flooding even under strong perturbation like replication (Fig. 2.24D), suggesting that both ingredients can lead to confined epigenetic memory.

2.3.8 Chromatin state dynamics regulates transcriptional noise

Histone modifications are likely to play a role in the regulation of genome accessibility thereby impacting transcriptional activity of genes [3]. To understand how gene expression may be affected by the dynamics of chromatin states, we consider a simple toy model where the transcriptional state of a gene depends on the current mean M-state proportion $\overline{M} = \frac{1}{L}\sum \delta_{i,M}$ over a 1kbp-wide ($L = 5$) region that would represent the gene promoter. We consider in this section that the state M is a repressive state (e.g. constitutive heterochromatin via H3K9me2/3) and that the transcription rate is inhibited by the presence of modified nucleosomes at the promoter ($\overline{M} \geq \overline{M}^*$) in a switch-like manner [23](see 2.2.6).

To only focus on the epigenetics-transcription relation, we consider a simple painter mode (as in Fig. 2.10) with $k/k_0 = 2$ and investigate the transcriptional properties of a gene as a function of the proximity of its promoter to the painter region (Fig. 2.25). We observe that the steady-state average number of mRNA per cell $\langle mRNA \rangle$ for this gene is larger for more distant promoters (Fig. 2.25C), qualitatively mirroring the decrease in $P(M)$

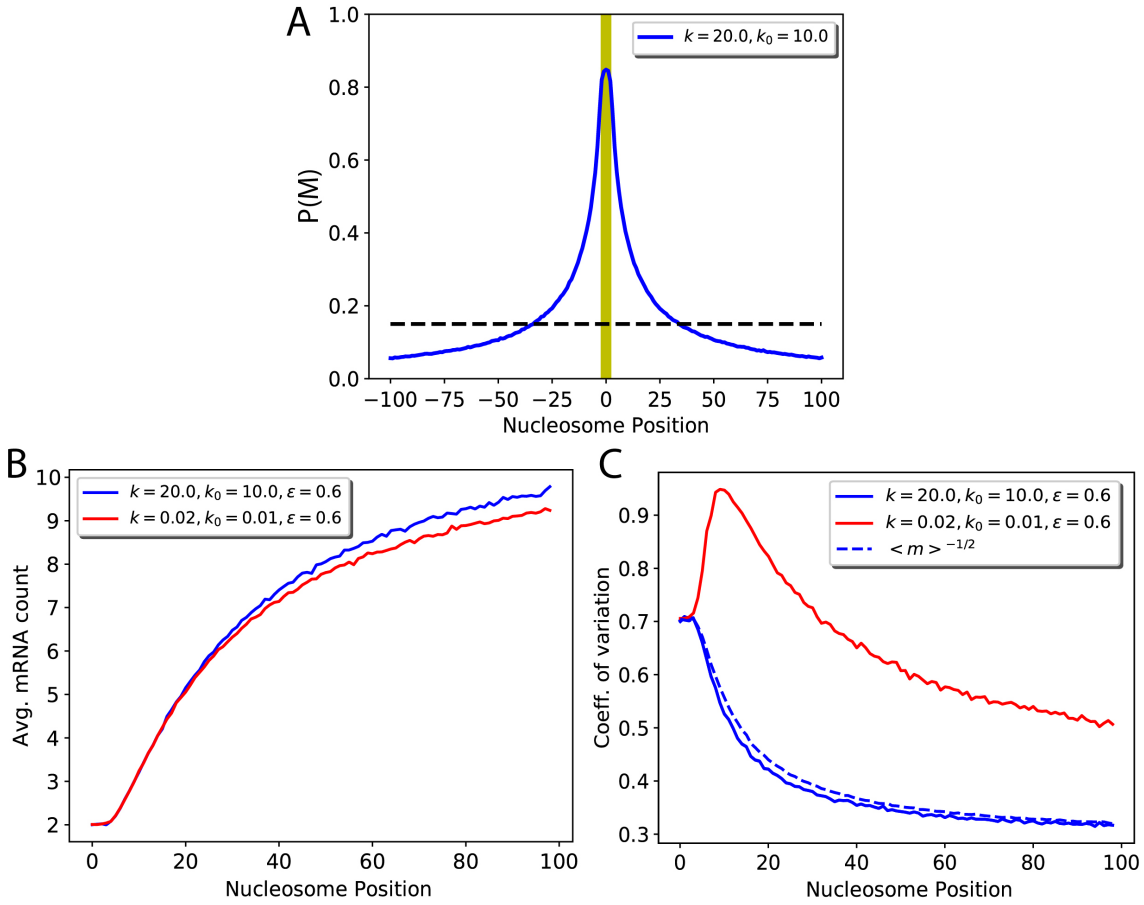

Fig. 2.25 Chromatin state dynamics and transcription: (A) $P(M)$ profile for $k/k_0 = 2, \varepsilon = 0.6$. Black dotted line indicates the value of $P(M)$ at the transcriptional switch ($P(M) = M^*$, see Fig. 2.8). (B) Average mRNA count for the two different epigenomic dynamics as a function of the distance of the gene from the painter region. (C) The coefficient of variation of the mRNA count for the same parameters as in (C). The dashed line shows $1/\sqrt{\langle mRNA \rangle}$ with $\langle mRNA \rangle$ taken from panel (B).

(Fig. 2.25A). Interestingly, for the same *relative* k/k_0 value (and thus the same average M-state profile, see Eq. 2.7, Fig. 2.25A), the average transcription level depends on the *absolute* values of k and k_0 that control the kinetics of M state dynamics, comparatively to the transcription dynamics that is itself driven by the transcription and mRNA degradation rates (see 2.2.6). We detect that slow epigenetic dynamics (red line in Fig. 2.25B) tend to favor a slightly more repressed state compared to fast dynamics (blue line in Fig. 2.25B).

Such kinetic effect is more striking when considering the intrinsic stochastic fluctuations of gene expression by estimating the coefficient of variation (CV) (Fig. 2.25C), defined as the ratio between the standard deviation of the corresponding steady-state distribution of mRNA number per cell and its average value, a high CV meaning a noisy, highly fluctuating gene. For fast dynamics (blue line in Fig. 2.25D), relative fluctuations decrease with the distance from the painter region. CV follows $\sim 1/\sqrt{\langle mRNA \rangle}$ (dashed blue line in Fig. 2.25C), characteristic of the intrinsic transcriptional noise found in elementary gene expression models [148]: lowly expressed gene being relatively more noisy. For slow dynamics (red line in Fig. 2.25C), we instead observe a sharp increase up to a distance s^* where fluctuations become maximal, followed by a gradual decrease at larger distances; fluctuations remaining always larger than in the fast dynamics case. Indeed, in our transcription rate model (see 2.2.6), close to the transition point $\overline{M}^*$, small variations of $\overline{M}$ result in large deviations for the transcription rate. If the epigenetic dynamics is faster than mRNA production and degradation rates, the transcription level cannot adjust to the rapid fluctuations of $\overline{M}(t)$ that are filtered by the slow mRNA dynamics [149]. However, if the chromatin dynamics is much slower, mRNA level adapts to the current, slowly fluctuating M-state. Hence expression will stochastically switch between a repressed ($\overline{M}(t) > \overline{M}^*$) and expressed state ($\overline{M}(t) < \overline{M}^*$) leading to large fluctuations of expression. The peak observed in CV in this case thus translates the interplay between such ultra-sensitivity to epigenetic fluctuations with the more standard dependence of the noise to $\langle mRNA \rangle$ (fast chromatin dynamics case). Biologically, this strong propagation of epigenetic fluctations to the gene expression level may be related to the well known variegation phenomenon where genes inserted near heterochromatin domains may have variable transcriptional pattern, depending on the distance of their insertion sites to the heterochromatin domains [150].

2.3.9 Applications to diverse biological contexts

We have studied systematically the behavior of our generic model of epigenomic regulation and its consequences on epigenetic memory and transcription. In the following, we will describe three applications of this framework to specific biological situations.

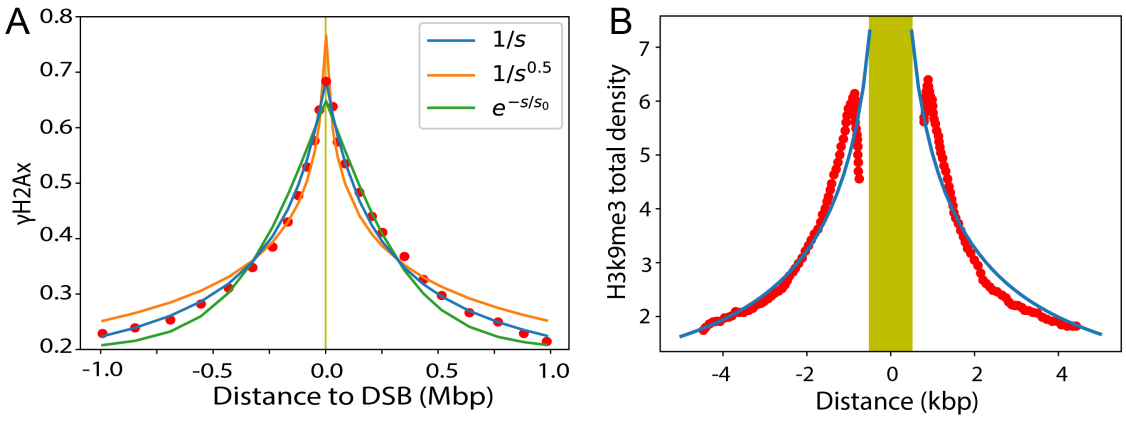

Fig. 2.26 (A) Experimental phosphorylation profile (dots) around a DNA double strand break [120] and the corresponding model prediction (i) $S(i) = 0.58P(M_i) + 0.11$, $(k/k_0)\varepsilon = 1213$ $P_c(i,j) = 1/|i-j|$ (blue) (ii) $S(i) = 0.85P(M_i)$, $(k/k_0)\varepsilon = 29.6$, $P_c(i,j) = 1/|i-j|^{0.5}$ (orange) (iii) $S(i) = 0.9P(M_i) + 0.2$, $(k/k_0)\varepsilon = 1$, $P_c(i,j) = e^{-|i-j|/s_0}$, $s_0 = 1050$ (green) (B) Experimental H3K9me3 density (dots) flanking a transposable element in mouse ES cells [125] and the corresponding model prediction $S(i) = 10.75P(M_i) + 0.1$ (full line).

Phosphorylation of *H2AX* around double-stranded breaks

During our analysis of the simple painter mode (Fig. 2.10), we showed that there is a direct relationship between the distribution of marks around the painter binding sites and the local 3D chromatin organization. This 3D↔1D relationship has been evidenced experimentally in the context of DNA double-stranded breaks (DSB) in humans [120]. After a break had occured, variant histones H2AX present in the chromatin are phosphorylated (the so-called γH2AX mark) by the ATM kinase recruited at the DSB site. The correlative analysis of the 4C and γH2AX Chip-seq signals around DSBs clearly reveal high similarities (Fig. 2.9, right), suggesting that the spreading by painter mode is at work in that system. In Fig. 2.26A, we extracted the average γH2AX profile observed around DSBs (red dots) from [120] that extends over Mbps (Fig. 2.9, center). Using the analytical formula for a simple painter model (Eq. 2.7) with a painter region localized at the DSB and $P_c(s) \sim 1/s$ the typical average contact probability found in human cells [92], we were able to fit very well the experimental data (blue line in Fig. 2.26A, L2-distance $= 1.4 * 10^{-4}$)(see 2.2.8). This leads to one identifiable parameter $(k/k_0)\varepsilon = 1213$. In the original article, Arnould *et al.* suggested that the loop extrusion mechanism might be directly implicated in the spreading mechanism [120]. To test this alternative hypothesis, we perform a similar inference using the loop extrusion-like spreading probability ($P_c(s) \sim \exp[-s/s_0]$) which leads to a less precise fit even if the model has an additional parameter (processivity $s_0 \sim 150 - 200$kbp) (green line in Fig. 2.26A, L2-distance$= 5.1 * 10^{-4}$). Note that the diffusion-like spreading

mechanism ($P_c(s) \sim 1/s^{0.5}$) is also inconsistent with the experimental data (orange line in Fig. 2.26A, L2-distance$= 9.0 * 10^{-4}$). These results differ from a similar analysis performed by Li *et al.* [143] neglecting turnover rate of the mark, suggesting that Tel1 (yeast homolog of ATM) spreads phosphorylation around DSB via a loop extrusion-like mechanism. But within our framework, the best prediction (minimum L2-distance) is given by the spreading-by-3D-contact mechanism ($P_c(i,j) = 1/|j-i|$)).

Heterochromatin formation around retrotransposons

An entirely different system where the simple painter mode may also be operative is the spreading of H3k9me2/3 modifications over few kbps around retrotransposons in mouse ES cells by Setdb1 [125, 151], a lysine methyltransferase recruited by the KRAB–Zinc Finger Protein KAP1 [152]. As in the previous example, we can fit the experimental Chip-seq profiles (red dots in Fig. 2.26B) by a simple model with a painter zone corresponding to the typical size range of a transposable element $\sim 1kbp$ [126] (blue line in Fig. 2.26B). In this case, we found $(k/k_0)\varepsilon = 0.78$. Note that, these predictions were done without invoking the reader-writer mechanism which might be relevant in other heterochromatin contexts (see below).

Compared to γH2AX, the spreading of H3K9me2/3 is much less efficient leading to smaller domains (kbp- vs Mbp-wide). This may translate fundamental differences in the spreading and turnover rates as histone methylation maintenance in mammalian cell lines could be slow ($\sim$ 10-20 hours) [91, 30] while establishment of γH2AX foci is much faster ($\sim$ 1-30 min) [153, 154].

Memory of heterochromatin in fission yeast

Finally, we apply our chromatin state dynamical model to quantitatively characterize the stability of epigenetic memory for fission yeast heterochromatin. In *Schizosaccharomyces pombe*, establishment of H3K9me2/3 domains is in part regulated by the balance between the methyltransferase activity of Clr4, its state-dependent recruitment by H3K9me histones and the demethylase-like action of Epe1 [10, 155]. In [10], a heterochromatin domain is established via the sequence-specific recruitment of Clr4 at an ectopic locus containing a fluorescent reporter gene, leading to the repression of the gene. Epigenetic memory of this domain and involved mechanisms are then characterized by releasing sequence tethering and by tracking the progressive re-activation of the gene as a function of time (full lines in Fig. 2.29A). In particular, they considered four different strains: (i) $TetR-Clr4-I, epe1+$ with a mutated Clr4 ($Clr4-I$) with no reader-writer property and a normal demethylase-like activity ($epe1+$), the model analogue being $\Delta = 0, k_0 = x$; (ii) $TetR-Clr4-I, epe1\Delta$ same as (i) but without demethylase-like activity ($epe1\Delta$) (model equivalent - $\Delta = 0, k_0 < x$);

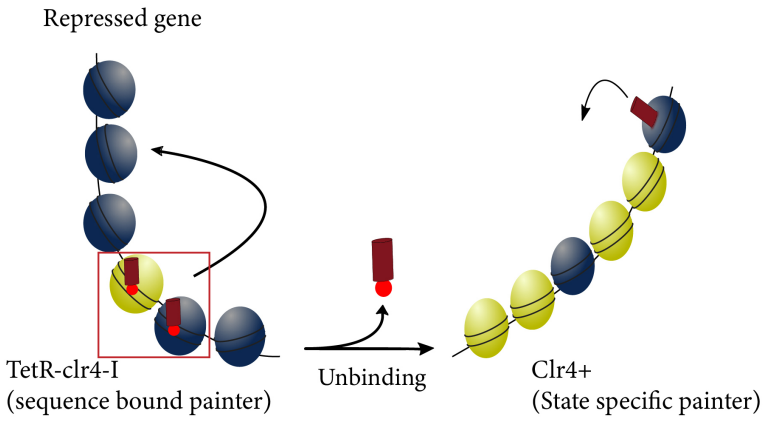

Fig. 2.27 Scheme describing the experiment on engineered yeast to study heterochromatin memory (Fig. 2.29A,B). Initially, the HMEs recruited at the target region establish repression (repressed gene state; (blue) modified nucleosomes) and then subsequently, once the HME unbinds, the gene expression is monitored over a period of 100 hours in the four different mutants (see Fig. 2.29A,B). After the sequence-dependent painter unbinds, all mutants gradually loose repression but at different rates (Fig. 2.29B) as the level of methylated nucleosomes decreases. In some mutants where $Clr4+$ is present (red and magenta, Fig. 2.29B), after removing the sequence-specific recruitment, the reader-writer mechanism fights against this loss of repression by maintaining for some time a methylated state modifications.

(iii) $TetR-Clr4-I, Clr4+, epe1+$ with wt-like Clr4 and Epe1 activities ($\Delta > 0, k_0 = y$); and (iv) $TetR-Clr4-I, Clr4+, epe1\Delta$ as in (iii) but in $epe1\Delta$ background ($\Delta > 0, k_0 < y$). They observed that memory of a repressed state was enhanced by the reader-writer module of Clr4 and the absence of Epe1 (dots in Fig. 2.29B).

To rationalize these experiments with our quantitative framework, we simulate the experimental memory assay like in Fig. 2.20 but for various values of Δ and k_0 and by tracking the transcriptional activity of the region using the model described in Fig. 2.25. More precisely, we follow the experimental protocol and monitor the time-evolution of the distribution of fluorescence of the reporter gene within the cell population (Fig. 2.29A), from which we can define the proportion of cells that are still repressed ("OFF") by heterochromatin (Fig. 2.29B). The predicted time-evolution of this proportion for each scanned (Δ, k_0) value is then quantitatively compared to the experimental data obtained for each of the four strains (Fig. 2.28). For strain (i), gene activation is fast (Fig. 2.27, black dots in Fig. 2.29B) and data are compatible with a simple painter mode with high turnover rate ($k_0 > 0.007\ min^{-1}$, $\Delta = 0$) (Fig. 2.28A, and black line in Fig. 2.29B), consistent with the presence of Epe1 and the absence of reader-writer recruitment which, we demonstrated in Fig. 2.20, is essential for epigenetic memory. For strain (ii), the gene de-repression is slower than in (i) (blue dots in Fig. 2.29B) but can still be well captured by a simple painter mode ($\Delta = 0$) (blue line in Fig. 2.29B) associated with decrease of k_0 by at least

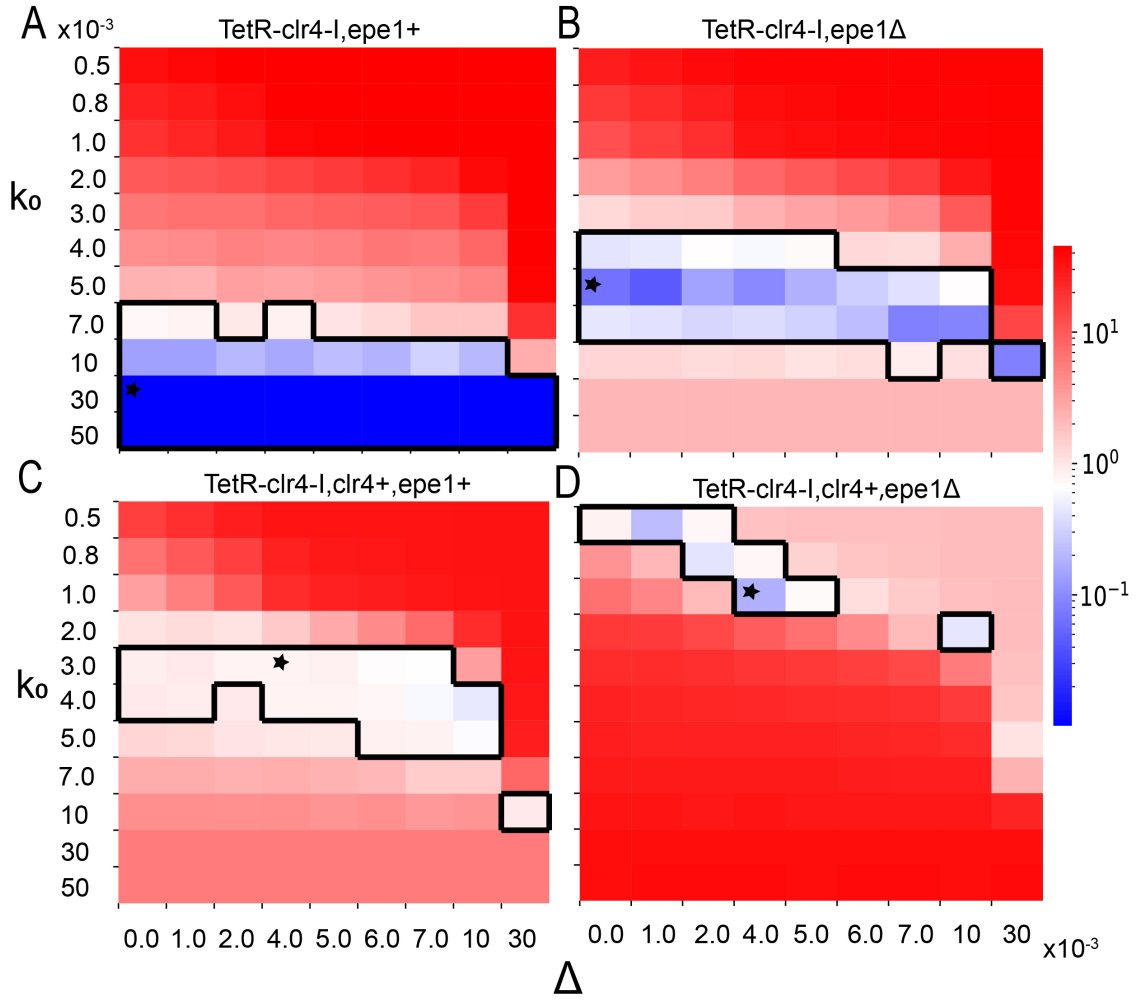

Fig. 2.28 Phase diagram showing the χ^2 value that quantifies the predictive power of the model predictions relatively to the experimental data for each of the four mutants. The region within the highlighted contour represents the region of the parameter space with an acceptable fit ($\chi^2 < 1$). Star marks the best fitting parameters with respect to the experimental constraints. These parameters are used in Fig. 2.29.

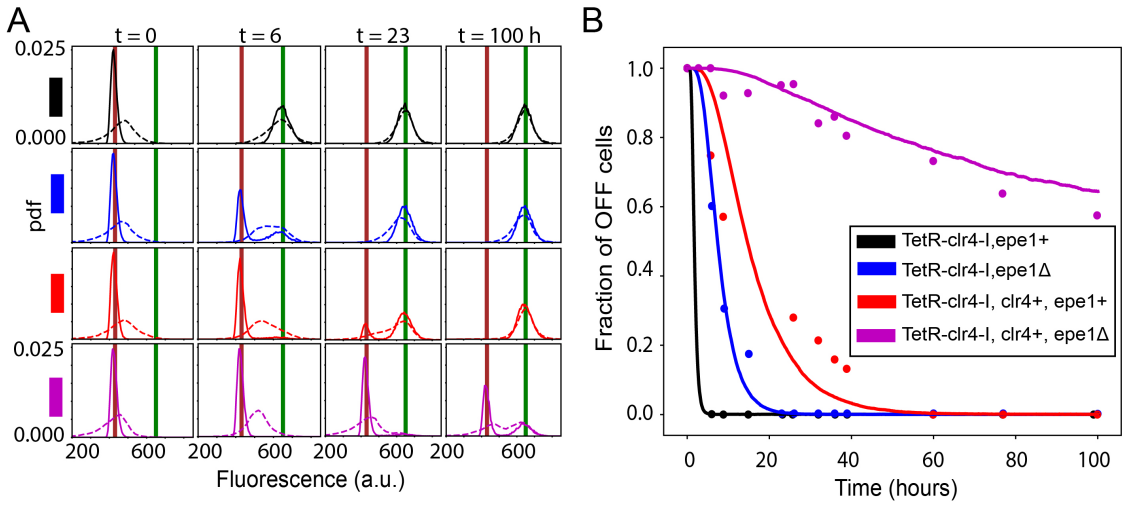

Fig. 2.29 (A) Heterochromatin maintenance in fission yeast. (A) Distributions of fluorescence expressed from a reporter gene at specific time points after painter release at $t = 0$ for different strains (experiments: dashed lines [10], model predictions: full lines). Brown and green vertical lines mark the mean fluorescences in the OFF and ON states respectively. (B) Experimentally-measured (dots) and predicted (full lines) time-evolution of the proportion of OFF cells in the population for the various strains.

6 fold (Fig. 2.28B), consistent with absence of Epe1 in this strain and thus with less turnover. In both backgrounds (iii)($epe1+$) and (iv)($epe1\Delta$), the reader-writer capacity of Clr4 strongly enhances memory and slows down gene activation (red and purple dots in Fig. 2.29B). This behavior can only be reproduced (red and purple lines in Fig. 2.29B) by a reader-writer mode with $\Delta = 0.004$ (Fig. 2.28C,D). Interestingly, for strain (iv) where memory is maximal, the model predicts that the biological system is in a parameter range just below the critical regime of the reader-writer mode (like the blue line in Fig. 2.20A). In this regime, the epigenetic state is long-lastly maintained until rapid removal of the marks ($\Delta = 0.2$ in Fig. 2.20C), leading to a stochastic switch in transcriptional activity and bimodal distributions for protein levels (Fig. 2.29A, purple at t = 100h), as observed experimentally [10].

Polycomb regulation in mESCs

In parallel to the development of the Painter model, a similar but much more detailed formalism to specifically address the regulation of Polycomb-target genes in mouse embryonic stem cells was developed by Kapil Newar [30]. The formalism is briefly described below.

The Polycomb system via the methylation of the lysine 27 of histone H3 (H3K27) plays central roles in the silencing of many lineage-specific genes during development. Recent experimental evidence suggested that the recruitment of histone modifying enzymes like the

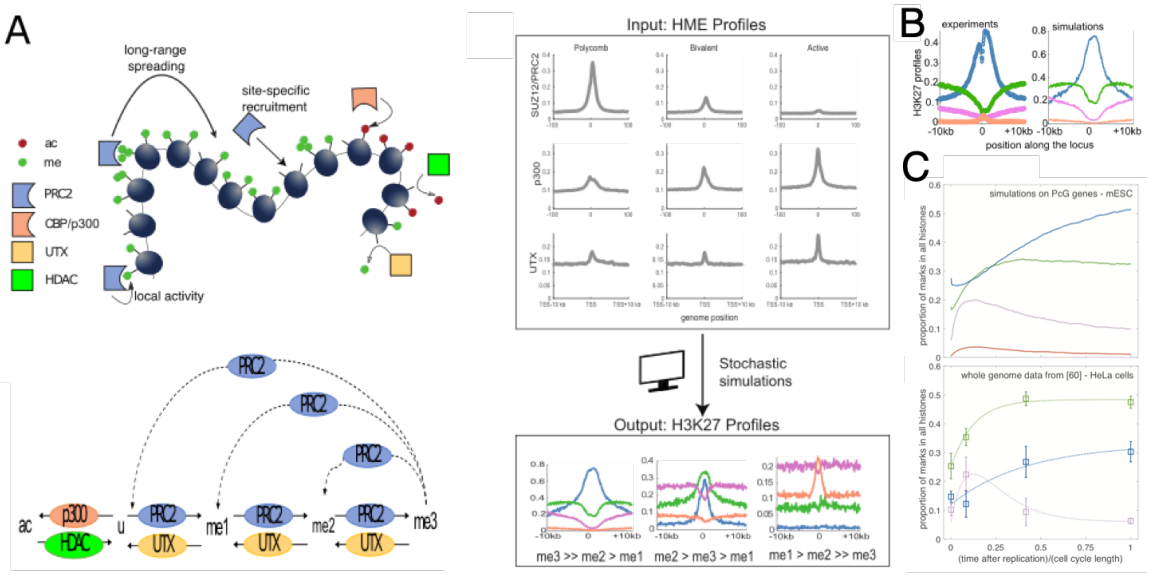

Fig. 2.30 Model of Polycomb regulation in mESC. (A) Scheme of the mathematical model describing the dynamics of the biochemical modifications of H3K27 thanks to the action of several histone modifying enzymes. Recruited PRC2s has a boosted acitivity if bound to trimethylated nucleosomes. (B) Experimental versus predicted Chip-Seq profiles of H3K27 modifications around Polycomb-target genes. (C) Experimental versus predictions cell-cycle dynamics of the modifications. *Adapted from [30].*

Polycomb repressive complex 2 (PRC2) at specific sites and their spreading capacities from these sites are key to the establishment and maintenance of a proper epigenomic landscape around Polycomb-target genes. Here, to test whether such mechanisms, as a minimal set of qualitative rules, are quantitatively compatible with data, we developed a mathematical model that can predict the locus-specific distributions of H3K27 modifications (acetylation, mono-, di- and trimethylation) knowing the distribution of HMEs like the methyltransferase PRC2 or the demethylase UTX (Fig. 2.30A). Actually, the model is an extended version of the "boosted-painter" mode where the trimethylation state locally boosts the activity of sequence-recruited PRC2.

Within the biological context of mouse embryonic stem cells where lots of quantitative data are available, our model showed very good agreement with experimental profiles of H3K27 acetylation and methylation around Polycomb-target genes in wild-type and mutants (Fig. 2.30B). In particular, we have demonstrated the key role of the reader-writer allosteric module of PRC2 and of the competition between the binding of activating and repressing enzymes in shaping the H3K27 landscape around transcriptional start sites. The predicted dynamics of establishment and maintenance of the repressive trimethylated H3K27 state suggest a slow accumulation, in perfect agreement with experiments (Fig. 2.30C).

2.4 Discussion and conclusion

In this work, we have developed a unified mathematical two-state model of epigenomic regulation, integrating key mechanisms like reader-writer processes. Our simple framework states that the establishment and maintenance of an epigenetic state results from the recruitment and spreading activity of histone modifying enzymes (HMEs). Recruitment of HMEs can be mediated by specific genomic sequences or by the local epigenomic state. Spreading encompasses an on-site action of the HMEs and a long-range, trans activity modulated by the local chromatin state.

In particular, we systematically studied three generic modes of regulation that recapitulate most of the experimentally-known epigenetic systems: (i) a simple painter mode (Fig. 2.10) in which HMEs are targeted to specific genomic regions and spread via their trans-activity around these binding sites; this mode may be representative of the regulation of small, local epigenomic domains like acetylation marks (e.g. H3K27ac or H3K9ac) around promoters and enhancers; (ii) a boosted-painter mode (Fig. 2.13) in which recruited HMEs have an enhanced activity if bound to specifically-modified regions; this mode may account for the allosteric boost observed for PRC2 in presence of H3K27me3 [17]; (iii) a reader-writer mode (Fig. 2.15) in which HMEs may also be recruited by specific epige-

netic signals; this mode may capture the regulation of extended chromatin domains like heterochromatic regions (e.g. H3K9me2/3).

In the simple and boosted-painter modes, we found that there is a direct, simple relationship between the binding profiles of HMEs, the chromatin organization and the profile of epigenomic state (Eq. 2.7). Verification of this relation using experimental measurements of these three information (e.g. bulk Chip-seq and Hi-C data) in wild-type-like conditions may be a strong evidence for one of these two spreading modes [30]. A perfect illustration of this is our application to γH2AX around DSBs in which the experimental profile is well fitted by the model (Fig. 2.26A) suggesting a simple or boosted-painter mode with a spreading-by-3D-contact scenario rather than a more complex cohesin-mediated mechanism [120].

Distinguishing the boosted- from the simple-painter mode would then require additional information like, for example to estimate the spatial correlations between chromatin state that have very different signatures between the two modes (Fig. 2.10C *vs* 2.13C). However, this information is currently very difficult to access experimentally as it would require a single-cell assay to estimate covariations of chromatin states; but recent progresses in single-cell Chip-seq experiments [160, 161] may open new venues.

In the booster-painter and reader-writer modes that both involve a “reader" capacity of HME, we observed phase-transitions and critical behaviors via respectively the strengths of boosting activity (r) and of state-dependent recruitment (Δ) (Figs. 2.13B, 2.15B). These behaviors are driven by the effective positive feedback loops and cooperative effects [19] that emerge from the enhanced spreading efficiency of some HMEs (more activity or more recruitment) if the histone modification they catalyze is already present locally. These effects are more important in the reader-writer mode and lead to qualitatively different epigenomic profiles, with extended domains around the binding peaks of HME (Fig. 2.15A), and distal spatial correlations (Fig. 2.15C). In particular, we found that these profiles are very sensitive to the underlying spreading capacity of HMEs [156], longer-range mechanisms facilitating the percolation of the system into large contact domains at low state-dependent recruitment (Fig. 2.17C) [28].

Eventually, these domains may be maintained in the absence of genomic bookmarking (Fig. 2.20A) for strong enough state-dependent recruitment. However, our analysis of heterochromatin memory in fission yeast (Fig. 2.26C,D) suggests that such self-sustainable memory may not be occurring for H3K9me2/3 chromatin domains in wild-type conditions and that (weak) nucleation, sequence-dependent signaling might still be needed for maintaining a stable epigenetic landscape [66, 68]; even if a reduction in the effective histone turnover rate (e.g. by a lower demethylase activity as in the $epe1\Delta$ mutants in [10] or by longer cell-cycles [23, 54]) might trigger the system near to the critical point and may lead to long-term memory [28], but potentially also to high sensitivity to external cues [56]. As

shown by previous modeling works [25, 28], we also confirmed that purely local spreading mechanisms (as the NN scenario) would require too strong feedback and state-dependent recruitment to stabilize extended, coherent epigenomic domains during a long time period, even if being able to maintain them over few cell cycles [158].

One theoretical issue however with standard reader-writer models with long-range spreading is the difficulty to stabilize finite-size epigenomic domain without genomic bookmarking and to avoid unconfined spreading at long-time scale, as already pointed out by Erdel and Greene [25] and Dodd and Sneppen [20]. While strong insulators or barriers preventing local spreading [20, 66] and/or the formation of compact 3D domains [66] may slowdown but do not prevent uncontrollable spreading, we have shown that the combination of 3D compartmentalization with enzyme titration may lead to ultra-stable confined memory of small domains even under strong perturbations like replication (Fig. 2.24D), as also previously observed by Sandholtz *et al.* using an explicit polymer model of epigenetic regulation with controlled HMEs concentrations [157]. This may be particularly relevant for the Polycomb regulatory system as it involves HMEs (PRC2) in limited numbers, as measured recently during fly embryogenesis [137] and evidenced also in mammals [162], and is associated with PRC1, a protein complex promoting the compaction of H3K27me3-tagged regions [138, 55]. Similarly, maintenance of confined silenced domains in yeast may rely on the combination of their spatial compartmentalisation via Sir3-mediated self-attraction [163] and Sir4-driven tethering to the nuclear envelop [164]) and of the titration of the Sir2 deacetylase [165, 166]. Note that, in our model, 3D compartmentalization was accounted via an attractive interaction between M states, mimicking PRC1, HP1 or Sir3 modus operandi. However, additional architectural mechanisms such as the spatial co-association of boundary elements are likely to be at work: in mammals, the stability of Polycomb domains may be reinforced by contacts between CTCF insulator sites [167] possibly via the cohesin-SA2-mediated loop-extrusion process [168]; in yeasts, colocalization of TFIIIC-binding boundary elements participate to chromatin domain integrity [75, 80]. Additionally, the presence of antagonistic chromatin states domains may act as competitive barriers to epigenomic spreading [6] but also may favor 3D compartmentalisation by strengthening phase separation [169]. Actually, all kinds of processes that reinforce *intra-* vs *inter-* contact domains is expected to promote confined memory by limiting long-range "contamination": in a condition of limited number of HMEs, increasing compartmentalization will further increase their local concentration (by sequestration), leading to a stronger stability and weaker pervasive long-range spreading (Fig. 2.24D), in agreement with recent experimental observations in fission yeast suggesting that a "critical density" of H3K9me3 is required for the stable inheritance of confined heterochromatin [159].

To go beyond the mere description of epigenomic regulation, we also proposed a simple mathematical model describing the impact of chromatin state dynamics on gene expression

[23]. Using this generic and modular framework, we quantified the role of epigenetic fluctuations in transcriptional noise. It allowed us to investigate positional variegation as well as epigenomic memory in terms of gene expression in addition to histone marks. In particular, this makes possible to analyze and interpret the transcriptional outputs of epigenetic-related experiments. For example, using this framework, we quantitatively described heterochromatin assays by fitting gene reporter expression distribution versus time and strains. Introducing a feedback of transcription on the epigenomic dynamics via the recruitment of HMEs [27, 29] or changes in turnover rates [170] may allow a finer description of the interplay between epigenomics and transcription.

In our work, we focused on a single painter site, but to go further the painter model may provide a very interesting framework to specifically address the pivotal role of the recruitment(painter) sites in globally shaping the epigenome. In particular, to investigate how their sizes, strengths or genomic distributions influence epigenomic regulation [156]. Interestingly, within some epigenomic domains, some sequences can recruit HMEs while being unable to nucleate chromatin states on their own [7]. These elements may actually act as secondary recruitment sites after the initial nucleation events at the strong, autonomous sites [171] and, in the presence of numerous weak sites, a stable domain can be formed even in the absence of strong sites [172]. Based on our results, an appealing hypothesis would be that, the numerous dispersed sites along the genomic domain might compensate their "individual" weakness by maintaining "collectively and cooperatively" a high local concentration of HMEs via long-range 3D "communication".

For example, to briefly extend the analysis, we consider a region of 200 nucleosomes containing two sequence specific recruitment regions separated by 50 nucleosomes (Fig. 2.31Top). To explore the contributions of the two recruitment regions, we consider three combinations (inset Fig. 2.31A) (i) $\rho_{s1}(i) = 1, \rho_{s2}(i) = 0$ (blue, Fig. 2.31A) (ii) $\rho_{s1}(i) = 0, \rho_{s2}(i) = 1$ (orange, Fig. 2.31A) (iii) $\rho_{s1}(i) = 1, \rho_{s2}(i) = 1$ (red, Fig.2.31A). In the simple painter mode (Fig.2.31A), the distant painter regions behave almost independently as $P(M)$ outside the painter regions (red curve) is equivalent to the probability of being modified by the each painter separately (green curve), and $P(M)$ at the left of the left painter is only weakly affected by the presence of the right painter. This is valid in the reader-writer case as well (Fig. 2.31B). However, when one of the painter is weak (Fig. 2.31bottom, $(\rho_s)^{right} = 0.2$), the weaker painter is slightly reinforced/buffered by the strong painter, both in simple painter and the reader-writer mode.

A promising application of Eq. 2.7 in the simple/boosted-painter modes is that, it can be employed to address the inverse problem of inferring the HME binding sites knowing the 3D

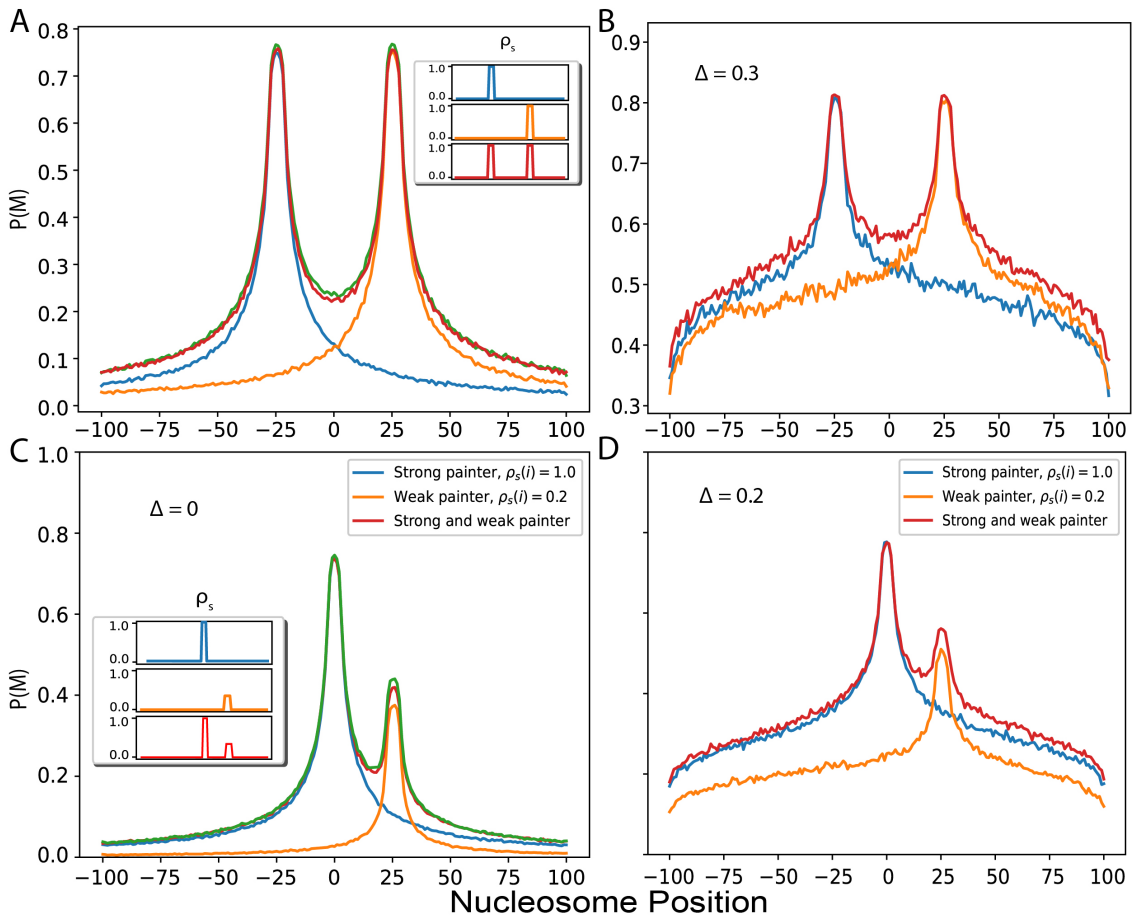

Fig. 2.31 Virtual Chip-seq profiles of the M-state for a multiple painter region. (Top) Probability $P(M)$ to be modified as a function of the position in presence of only the left painters (blue, P_{s1} at position [-28,-23]), only the right painters (orange, P_{s2} at position [23,28]) and both painters (red, $P_{s1,s2}$). Legend shows the corresponding different $\rho_s(i)$ profiles. (A) Simple painter mode $k/k_0 = 1, \varepsilon = 0.6$. (B) Reader-writer mode with $\Delta = 0.3$. (Bottom) Multiple painters with different HME recruitment strength, with painters on the left at position [-2,2] having a strength of recruitment $(\rho_s)^{left} = 1$ (strong) and on the right at position [23,28] with a strength $(\rho_s)^{right} = 0.2$ (weak). In the presence of the strong painter, the weak painter is slightly reinforced/buffered. (C) Simple painter mode $k/k_0 = 1, \varepsilon = 0.6$. (D) Reader-writer mode with $\Delta = 0.2$. Green line indicates the $P(M)$ when the two painters act independently $1 - (1 - P(M_{1/left})) \times (1 - P(M_{2/right}))$

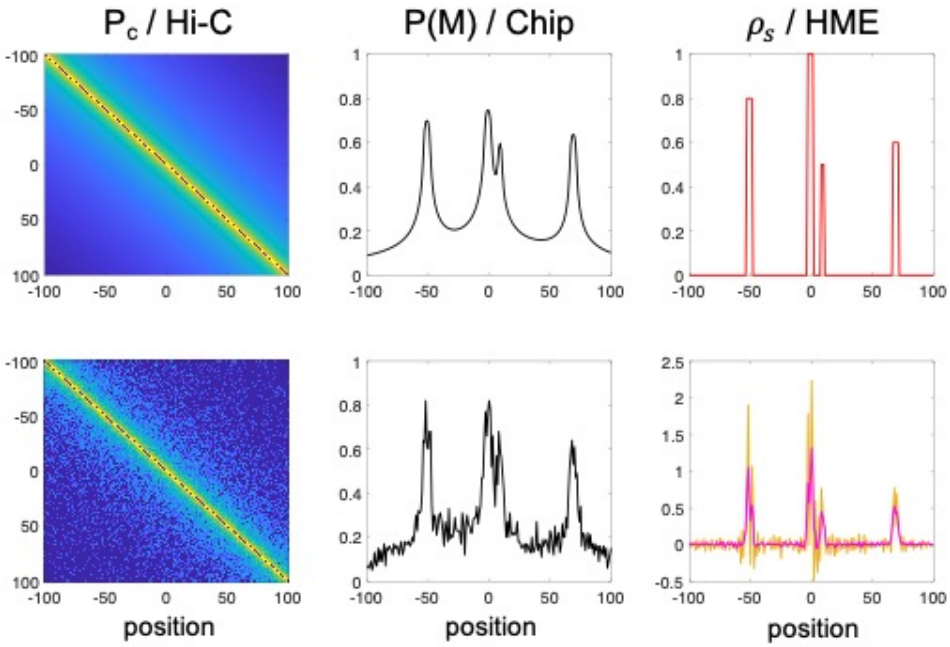

Fig. 2.32 Inference of HME binding sites. (Top) Ideal synthetic data (contact probability P_c on the left and position and strength of HME ρ_s on the right) used to compute a virtual Chip-seq profile of the M-state (center) with $k/k_0 = 1, \varepsilon = 0.6$. (Down) Noisy contact and chipseq data generated from (Poissonian noise added to the data) the ideal profiles (top) used to infer the position of HMEs (left) using the ridge technique for $\Gamma = 0.1$ (orange) or $\Gamma = 1$ (purple), assuming that we know the correct k/k_0 and ε values. Mis-estimation of these parameters may lead to defect on the shape or height of the inferred peaks but not on their positions.

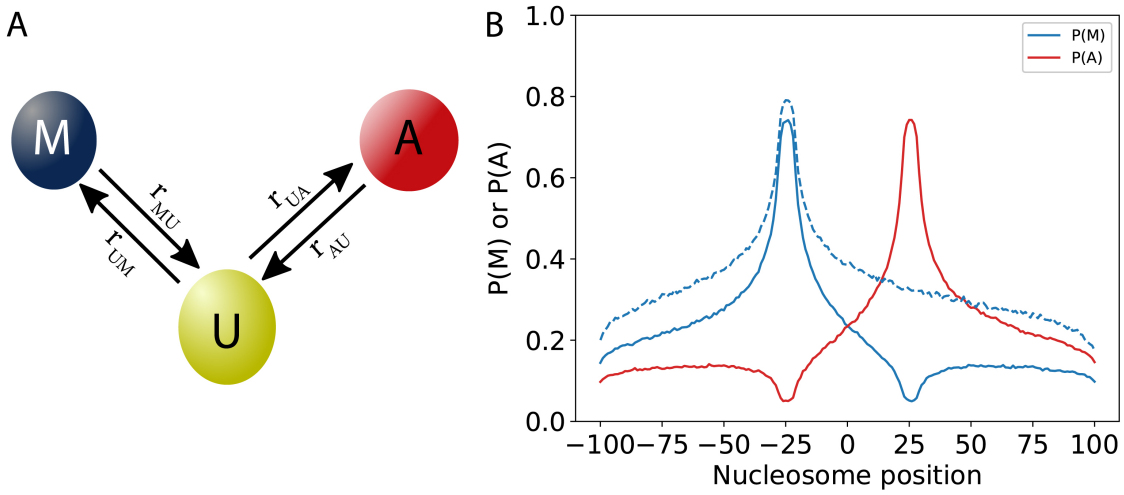

Fig. 2.33 Antagonistic marks. (A) Extended “Painter" model. (B) Virtual Chip-seq profiles of the M-state and A-state. Probability $P(M)$ to be in M-state and $P(A)$ to be in A-state as a function of the position. M-state painter localized at 5 nucleosome region around position -25 and similar A-state painter localized around position 25. Dotted line indicates the $P(M)$ profile spreading in the absence of antagonistic A-state and exhibits more extended spreading of M-state.

chromosome organization (e.g. via Hi-C) and the Chip-seq profile of epigenetic states. For example, in the simple painter mode, Eq. 2.7 can be written in matrix form as $A\rho_s = Y$, where $A_{ij} = \varepsilon * P_c(i,j)$ for $i \neq j$, $A_{ii} = 1$ and $Y_i = P(M)_i/((k/k_0)(1-P(M)_i))$. A priori, this linear equation can be solved easily to extract $\rho_s^* = A^{-1}Y$, the HME profile. However, inverting the possibly noisy matrices (experimental biases, lack of sampling/sequencing, etc.) can dramatically amplify such noise. To limit the noise propagation, standard regularization techniques may be used [133]. For instance, the ridge technique leads to $\rho_s^* = (A^TA + \Lambda I)^{-1} * (A^TY)$. For every positive value of regularization parameter Λ, we have a unique ρ_s solution. In Fig. 2.32, we illustrate the concept on synthetic data where we calculate the amount of HMEs along the sequence (bottom, right, ρ_s^*) for a given noisy probability to be modified ($P(M)$) and 3D organization ($P_c(i,j)$).

Similarly, related approaches still based on Eq. 2.7 can also be used to the estimate chromatin folding properties from HME and histone mark Chip-seq profiles. Actually, this latter strategy was recently used by Redolfi *et al.* [134], in the so-called Dam-C technique, to extract 4C-like contact probability information based on the spreading of Dam-mediated DNA methylation from a painter region.

One strong hypothesis that we made in order to focus on the painter and reader-writer processes and their consequences on epigenetic memory, was to neglect the interplay between several chromatin modifications [173, 174]. Indeed, the regulation of one silencing

epigenomic mark may often result from the competition with antagonistic activation marks. For example, in Fig. 2.33, we show the possibility of extending the painter model to a three state system, where there are two competing modifications M and A (Fig. 2.33A). The rates r_{UA} and r_{AU} have exactly the same form as r_{UM} and r_{MU} respectively (see Eq. 2.2). In Fig. 2.33B, we simulate a system with a painter zone for M and another for A, separated by 50 nucleosomes. The competition with the other mark now leads to asymmetric profiles around the painter region with limited propagation towards the antagonistic region. Indeed, multiple chromatin modifications also create natural barriers beyond which one modification dominates the other, preventing unconfined spreading.

The tug-of-war between active and inactive marks is driven by reader-writer processes complemented with reader-eraser mechanisms that actively remove opposite marks [175, 173]. This may lead to bistability between active and inactive chromatin states and participate in reinforcing epigenetic memory [19, 176, 27]. As for writers, recruitment of eraser enzymes may be sequence- or state-specific. For example, the UTX/JMJD3 demethylase is recruited by Trithorax-group proteins at active developmental genes to remove the repressive Polycomb H3K27me3 mark [14, 177]. In addition, the regulation of histone turnover may also participate in the erasing mechanisms. Indeed, genomic regions tagged with inactive modifications like H3K9me2/3 are know to display low histone turnover rate, while transcribed regions tends to exhibit higher rates [86, 170, 178]. All this may be responsible of complementary feedback loops in the epigenomic regulatory dynamics that finely tune the different levels of epigenomic modifications. For example, transcription may mediate such a loop not only by depending on the current epigenomic state, as we tentatively modeled in Fig. 2.25, but also, in return, by impacting epigenomic stability via an increase in histone turnover rate [27, 29].

In addition to the recruitment of opposite HMEs, spreading barriers may also arise from other mechanisms. For example, in Fig. 2.34, we simulate a situation where a region have a strong turnover rate (eg, due to highly active transcription), thus creating a simple 1D barrier. In the simple painter mode, induction of such a barrier only affects the M-state within the barrier since the process of M-state spreading is purely happening via 3D contacts $P_c(i, j)$ from the painter region. While in the case of reader-writer mode, the presence of 1D barrier reduces modifications in and around the barrier, since the barrier prevents recruitment of HMEs in the region.

Another simplification made in our work was to assume that the 3D chromatin organization (that mediates the long-range spreading) was fast compared to epigenomic-associated rates and was independent of the current chromatin state (see chapter 4). While the former hypothesis may be satisfied for short genomic distances, for large, Mbp-wide, chromatin domains like pericentromeric, H3K9me2/3-tagged regions in higher eucaryotes, the dynamics

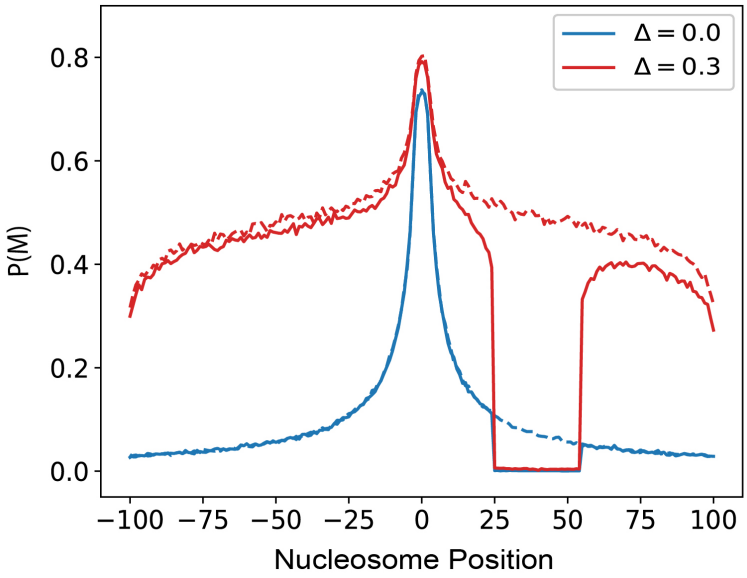

Fig. 2.34 1D barrier: Virtual Chip-seq profiles of the M-state. Dotted lines show $P(M)$ with a 1D barrier, a region of very high turnover rate ($100\,k_0$) between position 25 and 55, and the solid lines show $P(M)$ without the barrier.

of contact may be slow [90, 179] leading to less efficient spreading [70, 66]. As discussed above, heterochromatin domains (e.g. H3K9me3, H3K27me3) are often associated with architectural proteins (e.g. HP1, PRC1) that may impact on their compaction. Therefore, it introduces again a dynamical feedback loop between long-range epigenomic spreading and 3D chromatin organization: nucleation of a chromatin state drives its compaction that in turn facilitate spreading and maintenance [64, 139]. Recently, we and other groups investigated more carefully this coupling by developing a model that explicitly account for both 3D and 1D dynamics of chromatin [66, 67, 157, 68], highlighting the key role played by genome folding on epigenomic regulation. Combining the formalism developed here with such more detailed framework would allow to better characterize the structure-function relationship of chromatin including the formation and maintenance of confined chromatin domain.

To conclude, the formalism developed here is generic and modular, as it provides a simple description of epigenomic regulation in terms of HMEs recruitment and enzymatic activity. It can be easily upgraded to include more chromatin modifications (Fig. 2.33) and cross-talks and feedback loops and can thus be contextualized to a variety of specific chromatin regulatory systems as in the PcG work I participated to [30] (Sec.2.3.9).

A very promising application of our painter model might be to explore how epigenomic (and corresponding transcriptional) dynamics might influence genome evolutionary processes. Of particular interest is the challenging question of the role of epigenome in the

control of the integration and selection of transposable elements (TEs). TEs are mobile genomic elements that can be inserted at new genomic positions or can modify their location either via a “copy and paste" or in a “cut and paste" mechanism [126]. They comprise as much as $40-60\%$ of mammalian genomes. For active retrotransposons, a specific class of TEs, host cells have developed defense strategies in order to limit their invasion and expansion and to maintain genome integrity. Most of these elements are targeted by epigenomic silencing mechanisms (e.g. H3K9me2/3) [180–182] that limit their expression and thus, in *fine*, restrict their capacity of “parasitic" transposition [183]. This silencing is usually achieved by site specific recruitment of HMEs such as Setdb1 and Su(var)3-9s that nucleate and further spread the H3K9me2/3 marks [180, 184, 185] within TE elements and beyond into the flanking region, as shown in Fig. 2.26B with the simple painter mode in mESC. Very similar spreading patterns have also been observed around several other TE families in two Drosophila species, *D. melanogaster* and *D. simulans* [186]. As expected and confirmed by our theoretical approach that couples epigenomic and transcriptional dynamics (Fig. 2.25C), such long-range spreading of repressive chromatin state from TE elements has been shown to mediate long-range silencing of flanking genes [187]. This is actually reminiscent of the well known phenomenon of position effect variegation (PEV) observed in various organisms, where the incidental repositioning of a normally expressed/repressed genes, by translocation or transposition, next to a hetero/euchromatin fuzzy domain boundary induces its stochastic repression/expression depending on the genomic distance between the insertion and the boundary [150, 188, 189]. On one hand, there is a generic selection against TE transposition into euchromatin due to their deleterious effects for host cells that might be not only related to potential genetic alterations but also to epigenomic alteration via PEV. As shown in [186], the species-specific spreading ability of TEs is likely to be a driving force of their counter-selection. On the other hand, TEs are main contributors of genome evolution, by promoting genomic diversity and allowing regulatory innovations. Co-opted TEs play a crucial regulatory role in various nuclear processes, in particular in gene regulation during development [190, 191] as well as in 3D genome organisation [192]. TEs are constitutively silenced in most cell types and this silencing is required not only to limit their transposition but also to maintain proper cell-type 3D organisation and gene expression pattern. Hence, deregulation of TE has been associated to pathologies and is a clear hallmark of cancer [193, 194]. Overall this clearly indicates the need to develop a quantitative model based on the painter framework that would describe the epigenomic control of TEs and how they affect in turn the epigenome dynamics of flanking genomic regions. This will pave the way for a better understanding of the role of epigenome dynamics in TE-based genome plasticity during evolution as well as in pathologies.

Our model highlights the crucial impact of tridimensional chromatin organization on the stability of epigenomic domains and on gene expression. To explore the interplay between 3D chromatin organization and 1D epigenomic information [66], it is thus pertinent to explicitly model the chromatin polymer dynamics.

Chapter 3

Towards correctly simulating local chromatin organization

3.1 Introduction

To explore different facets of 3D chromatin organization, here we utilize the well-established idea of presenting genome folding as a polymer physics problem [195]. Polymer physics models have provided a concrete framework to understand chromatin dynamics [196] and to explore the physical mechanisms driving 3D genome organization [197, 95]. As suggested in the conclusion of Chapter 2, a methodological continuity of the painter model work is now to extend/refine our effective 3D model by explicitly taking into account the dynamics of 3D chromatin organization. In particular, this would allow to build quantitative 1D/3D coupling models to describe real systems, such as the dynamics of Polycomb repression at Hox genes [198]. The development of such a model is an interesting short-term perspective of the current work and would be explored in Chapter 4. However, investigating systematically such a model would require efficient and proper simulation of the behavior of a region of interest for long time period and for various parameter sets. In this chapter, we address such methodological issue by studying how to optimally model the spatio-temporal dynamics of chromatin domains.

Indeed, regions of biological interest for our chromatin state modelling, such as Hox gene domains in Drosophila or Mating type loci in yeast, typically span only a small fraction of the full chromosome length (dozens to hundreds of kbp). With recent experimental advancements, such as Micro-C, now providing high resolution 3D data sets [199], it may be now possible at the experimental level, to extract finely the coupling between 1D and 3D for these regions, at (few) nucleosome resolution (200bp-1kbp) with good statistics. Hence, it is important to investigate the conformational properties of such small specific regions by building “fine-scale” 3D models of chromatin.

When considering a region of interest, it is tempting to limit the simulations to a minimal chromatin domain encapsulating the corresponding genomic segment. However, restricting the system to the region of interest relies on the assumption that the spatial conformation of this region is not affected by the rest of the chromosome(s) [90], i.e., the chromatin inside the regions does not interact significantly with other chromatin fragments outside the region. This means that such a strategy requires, as a prerequisite, to validate this 3D "independence/isolation" hypothesis.

Knowing that polymer properties are sensitive to finite-size effects [200–202], a more accurate choice would be to simulate the region with the entire chromosomes, which would ensure to account for all the external interactions. But, over biologically-relevant timescales, it is likely to constitute a computationally-intensive task, especially when working at few nucleosome resolution. Here, we wish to address the question of how to minimize this computational effort while precisely reproducing local chromatin mechanics and conformational statistics.

From a physics perspective, getting such an accurate description of the system is important when we estimate mechanical properties or physical parameters of specific chromatin region from polymer simulations [203]. It is therefore of importance to address the question of how these estimated quantities may depend on the total polymer length. Thus, we seek to theoretically investigate what would be the extent of the minimal genomic region that one should explicitly consider around a given locus in order to effectively capture the correct dynamical and structural properties of the domain of interest.

Typically, there may exist two major contributions of the total polymer length simulated around a region of interest: (i) a pure polymeric contribution: the subchain (corresponding to the targeted region) is embedded inside a (often long) polymeric chain (the chromosome) that may impact the dynamical and structural properties of the subchain ; (ii) an environmental contribution: the subchain may have specific interactions with other regions along the long polymeric chain (or other chromosomes) that may also impact its spatio-temporal dynamics. The former contribution has been partially explored by my supervisors [204], showing that there might be strong differences between an isolated domain embedded inside a neutral long chain and the domain within its native, interaction context. However, the latter contribution remains under-investigated.

In this chapter, we first explore the -polymeric- contribution by systematically varying the length of the polymer L being simulated and quantifying the variations in structural and dynamical properties of the genomic region of interest. We show that the variations may be directly linked to excluded volume and entanglement properties of the long chromosome. We can quantitatively predict the length L of the polymer to be simulated while studying a particular gene/locus, to rightly capture the structure and dynamics of this region. Finally,

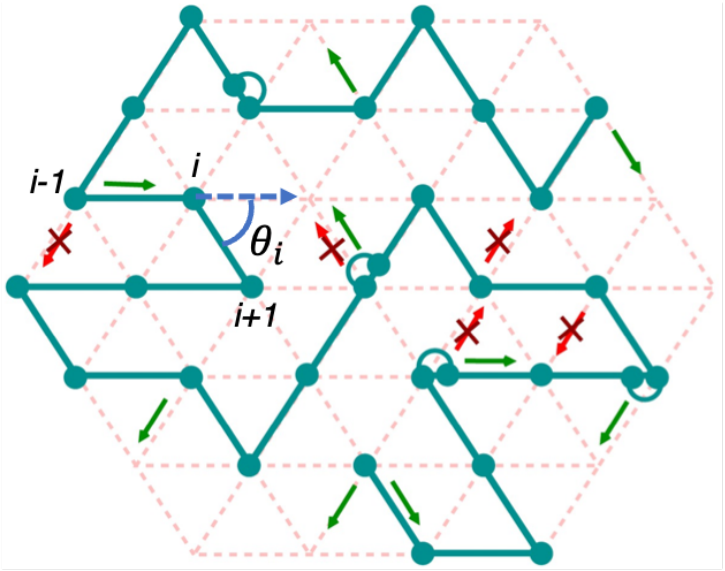

Fig. 3.1 Scheme of a lattice polymer configuration (2D projection of a 3D fcc lattice): Solid green line represents the polymer chain and dotted line, the lattice. Semicircular arcs indicate doubly occupied lattice sites. Each lattice site is allowed to contain a maximum of two beads if and only if they are consecutive to each other along the chain. Some of the allowed and forbidden moves are shown in green and red respectively. *Adapted from [90].*

we combine our results on the polymeric contribution with the environmental one to derive some thumb rules on how to choose the correct length to simulate.

3.2 Models and methods

3.2.1 Polymer lattice model

We model chromatin as a semi-flexible, self-avoiding polymer on a face centered cubic lattice (fcc) using the lattice kinetic Monte-Carlo model developed in [90, 169]. The polymer chain is composed of N beads (or monomers), each of size b and representing n bp. Thus, the contour length of the simulated chromatin segment would be given by $L_c = Nb$ encompassing $L = Nn$ bp. The polymer evolved on an fcc lattice of dimension $S \times S \times S$ containing $4S^3$ lattice sites (Fig. 3.1) that can be possibly occupied by monomers. Each lattice site can be occupied by a maximum of two monomers if and only if they are consecutive along the polymeric chain. Double occupancy of consecutive monomers accounts for the effect of contour length fluctuations [212] and allows the chain to still efficiently move in dense system. Two non-consecutive monomers cannot be at the same lattice site owing to excluded volume interactions.

Periodic boundary conditions are used to approximate polymer confinement and the effect of other chains (concentration); occupancy being characterized by the chromatin volumic fraction, given by $\phi = N/4S^3$. It is to be noted that the periodic boundary condition does not confine the polymer to the finite volume of the simulation box instead the polymer is free to extend over large distances.

The bending rigidity of the chain is accounted using standard Hamiltonian [217]:

$$E_{bend} = \sum_i k_{int}(1 - \cos\theta_i) \tag{3.1}$$

where θ_i is the angle between two successive bonds around monomer i (Fig. 3.1) and k_{int} is a measure of bending stiffness and is related to the chain Kuhn length l_k by [90]

$$(13\frac{l_k-1}{l_k+1}+12)x^4+(52\frac{l_k-1}{l_k+1}+24)x^3+26\frac{l_k-1}{l_k+1}x^2+(52\frac{l_k-1}{l_k+1}-24)x+13\frac{l_k-1}{l_k+1}-12=0 \tag{3.2}$$

$$k_{int} = -2kTln(x).$$

In the last part of the chapter, to account for epigenome-driven interactions, we introduce self-attractions between monomers that share the same epigenomic state [56, 90]. This is accounted via the Hamiltonian:

$$E_{epi} = \sum_{<i,j>} J_p(e_i)\delta_{e_i,e_j} \tag{3.3}$$

where the sum runs over all the pairs (i, j) of monomers occupying nearest-neighbor lattice sites, e_i is the epigenomic state of monomer i, $\delta_{e_i,e_j} = 1$ if $e_i = e_j$ ($= 0$ otherwise), $J_p(e)$ (in units of kT) is the strength of self-interaction between monomers of state e. For instance, the blue monomers in Fig. 3.5A can be thought of as in same epigenetic state e_M, interacting with a strength $J_p(e_M) \equiv J$, while the yellow monomers are in epigenetic state e_U, which doesn't interact ($J_p(e_U) = 0$) with each other (analogous to the unmodified state U described in chapter 2) .

In the following sections (except in sec. 3.3.5 and sec. 3.3.6), we will work at a fine-scale model with $n = 1$kbp, $b = 20$nm and $k_{int} = 3.217$kT which correspond to a chromatin fiber of diameter b, linear compaction ≈ 50bp/nm and with a Kuhn length of $l_k = 100$nm, consistent with recent estimation in yeast [218, 219]. We maintain a chromatin volume fraction of $\phi \equiv b^3\rho_{bp}/(\sqrt{2}n) \approx 0.05$, giving a bp-density of about $\rho_{bp} = 0.009bp/nm^3$ [220].

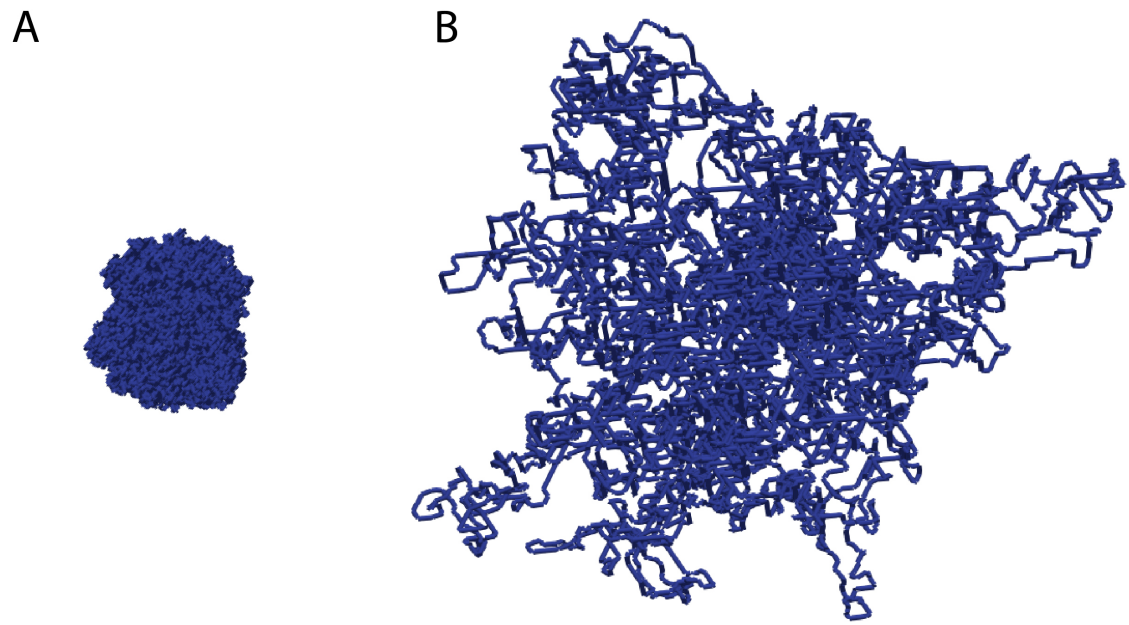

Fig. 3.2 Snapshots extracted from a simulation of a 20,000 bead-long polymer. (A): initial knot free configuration. (B) Relaxed state after 10^8 MCS.

3.2.2 Simulations

The fcc lattice has a valency of 12, i.e., each lattice site has 12 nearest-neighbors (NN). The dynamics of the chain is simulated using local moves on the lattice allowing the monomers to randomly hop between nearest-neighbor sites [213]. More precisely, one trial move consists in randomly picking a monomer and in attempting to move it to one of its randomly chosen nearest-neighbor, under the constraints that the chain connectivity is maintained (two consecutive monomers along the chain occupy the same or NN sites) and that the double-occupancy/self-avoidance rule (see above) is conserved (except some cases in 3.3.4). The trial move from an old configuration o to a new one n is then accepted with a probability $accept(o \to n)$ according to the Metropolis criterion on the total Hamiltonian $E_{tot} \equiv E_{bend} + E_{epi}$:

$$accept(o \to n) = min\,(1, \exp[-(E_{tot}(n) - E_{tot}(o))]) \tag{3.4}$$

We define one Monte-Carlo step (MCS) as N trial monomer moves. Note that this local kinetic Monte-Carlo scheme prevents chain crossing and thus conserves the current topological constraints.

In each situation and trajectory, we initialize the system (Fig. 3.2A) by a knot free configuration generated by the “hedgehog algorithm" [216]. Starting from a straight chain

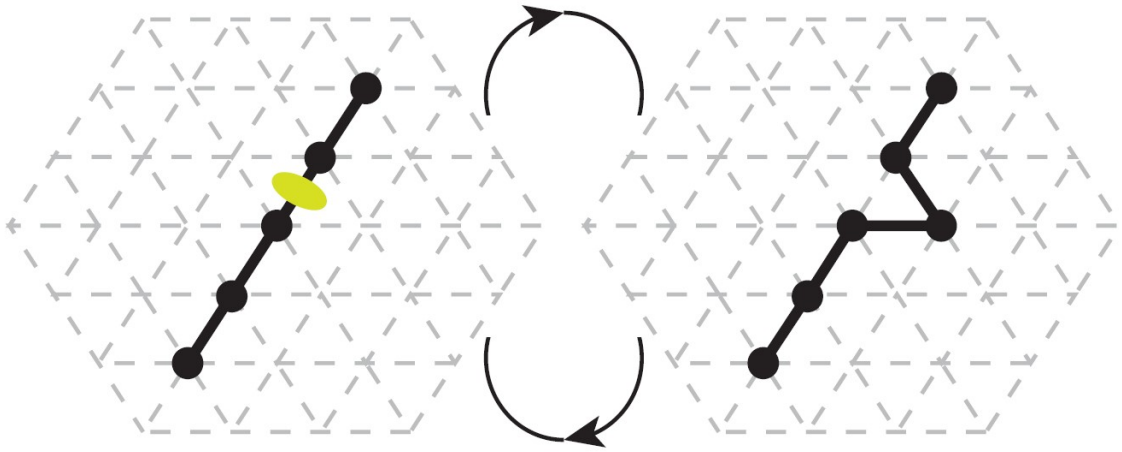

Fig. 3.3 The hedgehog algorithm to construct the initial configuration. Starting from a rod-like chain (left) containing few monomers, the structure is iteratively built by randomly choosing a link (green oval) and by inserting a new monomer at a site close to the two already placed monomers of the link (right) such that the new configuration still verifies lattice rules (excluded volume and connectivity). This step is repeated until the entire chain is grown (Fig. 3.2A). *Adapted from [221].*

with a few monomers (Fig. 3.3, left), a link is randomly chosen and a monomer is inserted at site close to the two monomers such that the new configuration satisfies excluded volume and chain connectivity criterion (Fig. 3.3, right), the process is repeated until the entire chain is grown [221]. The dynamics of the chain is then simulated during 10^8 MCS, recording data every 10^5 MCS. For each parameter sets, we simulated at least 128 trajectories.

Note that the choice of knot-free initial configurations is motivated by the observations that (1) chromosomes during interphase are likely to be weakly knotted and entangled [214, 215]; (2) due to the slow large-scale reorganization kinetics of such long, topologically-constrained polymers (impossibility for two strands to cross each other), this initial state ensures that the chromosomes generally remain in a weakly-entangled/crumpled regime over the course of the simulations, as suggested by Rosa and Everaers [196].

Numerical simulations were performed using the custom made code developed in the Jost group which can be downloaded at https://github.com/physical-biology-of-chromatin/LatticePoly.

3.2.3 Measured observables

For each data point, we compute several structural quantities like (1) the pairwise squared distance between monomers i and j, $R^2(i,j)$ and (2) the radius of gyration of a subchain

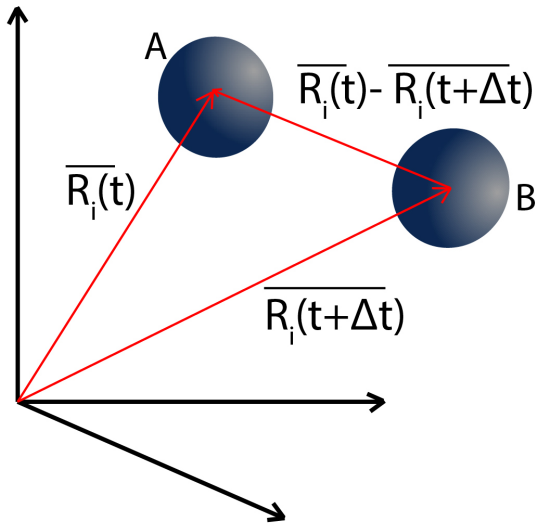

Fig. 3.4 An illustration of the displacement of a monomer and the quantity MSD. A monomer i is at position A at time t, $\vec{R}_i(t)$ and at time $t+\Delta t$, it is at position B, $\vec{R}_i(t+\Delta t)$. The displacement of the monomer i for the duration Δt is $\vec{R}_i(t)-\vec{R}_i(t+\Delta t)$, this displacement squared and averaged over an ensemble of Δt at different time ts would give the mean squared displacement (MSD) of the monomer i for Δt time duration.

$[i_0 : j_0]$:

$$Rg([i_0 : j_0]) = \left[\left(\frac{1}{2(j_0 - i_0 + 1)^2}\right) \sum_{i,j\in[i_0:j_0]} < R^2(i,j) >\right]^{1/2} \tag{3.5}$$

which quantifies the overall size of the region of interest (Fig.3.5A).

To quantify the mobility of monomers, we compute the mean squared displacement of each monomer i (Fig. 3.4)

$$MSD_i(\Delta t) =< (R_i(t) - R_i(t+\Delta t))^2 > \tag{3.6}$$

where the average is performed over time t and trajectories. $R_i(t)$ is the position of i in 3D at time t, and Δt is a time-lag. The MSD typically corresponds to the squared distance travelled by a monomer during the time lag Δt. Similarly, we estimate the MSD for the center of mass of a subchain $[i_0 : j_0]$ by replacing R_i by $R_{cm} = \left[1/(j_0 - i_0 + 1)\sum_{i\in[i_0:j_0]} R_i\right]$, the position of the center of mass in the equation above. And also, the relative MSD between two monomers i and j by using $R_i - R_j$ in the formula. This quantity relates to the relative motion of one monomer versus another.

3.2.4 Contextualizing the simulations to specific experimental systems

At the end of the Chapter, we employ the block copolymer framework to contextualize our conclusions to investigate specific regions in drosophila and yeast. The resolution of the model is turned to 800bp (i.e., 15nm) in order to stick to the experimental resolution of the data, the bending modulus is thus set to 4.05kT to yield a Kuhn length of 100nm [218, 219]. We use a bp-density of $\rho_{bp} = 0.009 bp/nm^3$ for Drosophila and $\rho_{bp} = 0.005 bp/nm^3$ for yeast [90].

We obtained the processed Hi-C maps from [222] and [223] at 800-bp resolution, for drosophila and yeast, respectively. Experimental Hi-C data were then normalized as in [179] to transform contact frequencies into contact probabilities in order to be comparable with model predictions. Then, we select one region of interest for each species: for drosophila, we consider a $\sim$ 307 kb region containing a topologically associating domain (TAD) [224, 225] (from 17955 to 18262 kbp) in chromosome 3R (lower part, Fig. 3.21), and for yeast a $\sim$ 50 kb-long region (from 334274 to 378470 bp) in chromosome 12 containing the gene REA1 (lower part, Fig. 3.18). From the experimental Hi-C maps, we identify the relevant compartments in drosophila and domains in yeast using the "hic_data.find_compartments" tool from the TADbit suite [227], using Principal Component Analysis (PCA) of the observed over expected matrices. In the context of block copolymer model, in the yeast case, we assign each monomer within the gene/TAD to the “blue” state — which may self-interact via a nearest-neighbor attractive potential variable of depth J — and attribute all the remaining monomers to the non-attractive “yellow” state ($J_p = 0$). In the Drosophila case, each monomer along chromosome 3R corresponding to one compartment (Positive value of the PCA), is assigned a blue, self-interacting state and the other monomers (negative values) to yellow, neutral states. For each J, we simulate 48 replicates.

To compare model predictions to experiments, from the simulations, we compute for each pairs of monomers the average contact probabilities. A parameter involved in this analysis is the radius of capture R_c, ie the maximum 3D physical distance between two monomers to be considered as being viewed as in contact in Hi-C experiments [221] (any two monomers i and j are marked to be in contact if $R(i, j) < R_c$). This is an unknown parameter like the strength of interaction, indeed experimentally it is likely that such quantity may tightly depend on the Hi-C protocol [179]. The structural quantities like the mean squared end-to-end distance ($< R_e^2 >$) and the Hi-C maps of the regions were computed by averaging over the replicates and over the 10^7 MCS following a relaxation step of 10^7 MCS.

To fit the experimental Hi-C data, we thus systematically vary R_c from 15 to 150 nm and the interaction strength J from 0 to $-0.4\,kT$. The best-fit parameter combination (J and

R_c) is obtained by minimizing a χ^2 score, defined as

$$\chi^2 = \left(\frac{1}{2N_t}\right) \sum_{(i,j)\in t} [(HiC_{pred}(i,j) - HiC_{exp}(i,j)]^2 \tag{3.7}$$

where $HiC_{pred}(i,j)$ is the simulated contact frequency between monomers i and j and $HiC_{exp}(i,j)$ the corresponding experimental data, t is the ensemble of pairs of monomers on which the fit is based, N_t represents the size of t. In our case, t is limited to all pairs of monomers with a genomic distance larger than 4kbp, as our coarse-grain model is not expected to correctly capture local, fine-scale details below 4kbp. The corresponding parameter values are verified to also maximize the Pearson correlation between the experimental and simulated Hi-C matrix (Fig. 3.20, 3.17), thus evidencing the accuracy of our optimization protocol.

3.3 Results

3.3.1 The system under study

To reiterate, in this chapter we ask a rather simple question: what would be the minimal length of DNA to simulate around a specific region of interest to fully capture the correct structural and dynamical properties of this region?

To address it, we consider a domain of interest d of size L_d embedded inside a larger polymer of total size L (Fig. 3.5A). So our question may be reformulated as how L impacts the behavior of the domain. To specifically understand the effects of changing length, for a given domain size L_d, we simulate different systems with different L, keeping the volume fraction ϕ constant to conserve chromatin density. For simplicity, we place the domain of interest in the middle of the larger polymer but this position does not impact the results, unless it is placed at the very end of the polymer (Fig. 3.5B).

In the rest, we vary L from L_d to 20Mbp, the typical size of a chromosome arm in Drosophila [226]. We consider the simulations performed for $L_{ref} = 20$Mbp as the reference one and basically ask whether the structural and dynamical reference properties of d are conserved when L is decreased.

3.3.2 Total polymer length determines domain compaction for a confined self-avoiding walk

As a first step, we study how L affects the structural characteristics of a region of interest d of size $L_d = 160$kbp, much smaller than the length of our reference case ($L_{ref} = 20$Mbp) in

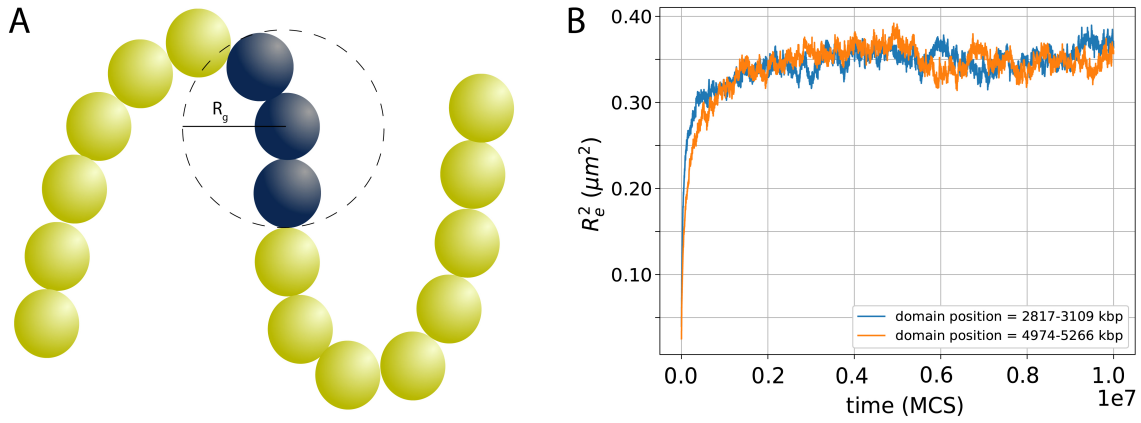

Fig. 3.5 (A) An illustration of the polymer being simulated, individual monomers being shown in yellow and the region of interest d is highlighted with monomers in blue (an illustration of radius of gyration R_g of the region of interest is shown). (B) Evolution of R_e^2 of the region of interest (domain) of length $L_d = 292$kbp in a polymer of length $L = 5926$kbp. The relative position of the domain with respect to the polymer is varied, blue curve, when the domain is in the middle of the polymer and orange, when the domain is shifted towards the right but it is to be noted that the structural variable evolves is the same manner.

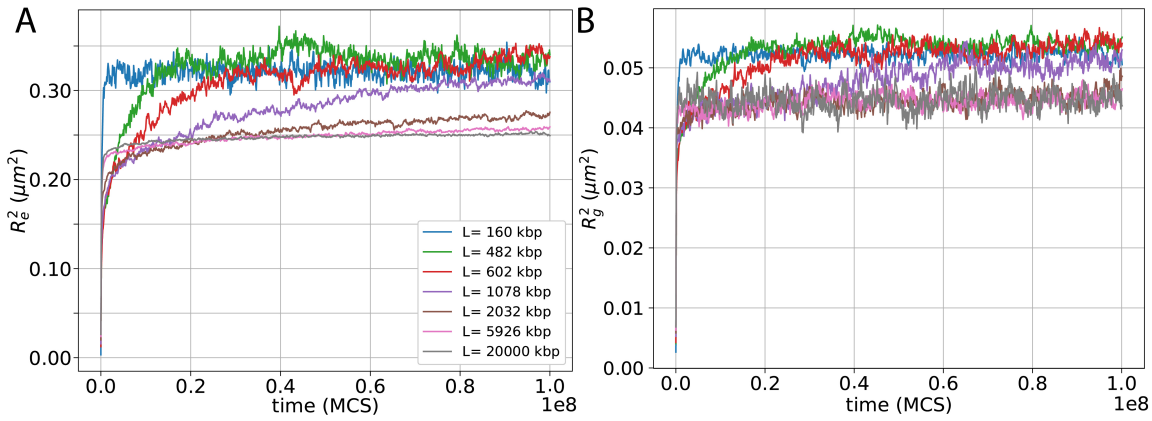

Fig. 3.6 (A) Time evolution of the squared end-to-end distance R_e^2 of a 160kbp domain for different total polymer length L. (B) Same but for the squared radius of gyration R_g^2.

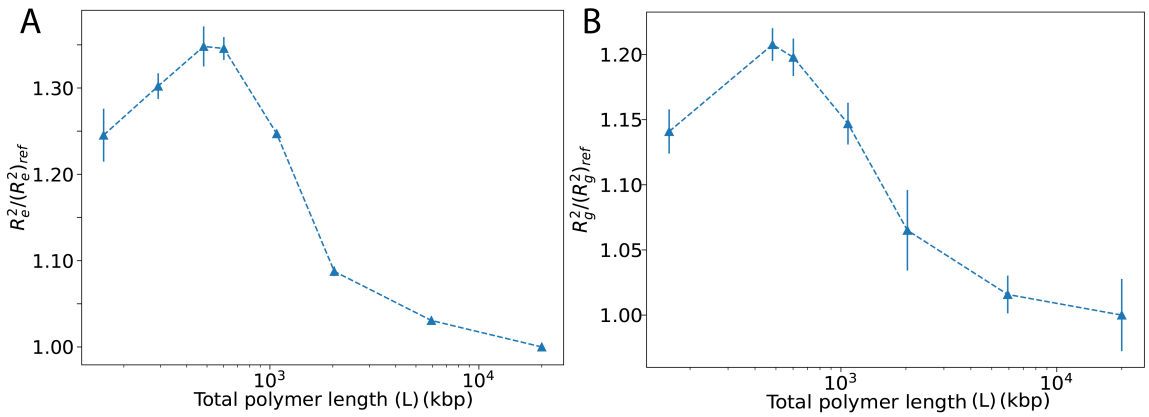

Fig. 3.7 (A) Deviation of R_e^2 of 160kbp domain from the reference value (at $L_{ref} = 20$ Mbp) as a function of the total polymer length L. The error bars indicate the standard deviation of R_e^2. (B) Same as in (A) for R_g^2.

a simple case of a homopolymer purely driven by self-avoidance, starting from a compact, post-mitotic-like, knot-free configuration.

First, we follow the dynamics of decompaction of the initial configuration and of convergence towards a (pseudo)stationary state for structural properties of domain d. In particular, we measure the ensemble-averaged end-to-end squared distance (R_e^2) of d (i.e. $R^2(i,j)$ between the two boundaries of d) as a function of time (Fig. 3.6A). Such observable corresponds to the slowest Rouse mode of a polymer chain [209] and thus well captures the full convergence towards equilibrium. We also compute the time evolution of the squared radius of gyration R_g^2 (Fig. 3.6B).

For small L ($<$ 500kbp), we observe a rapid expansion of the domain and a fast convergence (less than 10^7 MCS) towards an equilibrium state. For large L ($> 2Mbp$), there is still a first rapid decompaction step but followed by a slower convergence ($\sim 5.10^7$ MCS) towards a pseudo-stationary state with R_e^2 being almost constant but still very slowly and slightly expanding at long time-scale ($> 10^8$ MCS). Between these two regimes ($L = 1.078$Mbp in Fig. 3.6) the system converges to equilibrium but very slowly.

Interestingly, for $L \lesssim 1$ Mbp all the curves converge towards approximately the same value ($R_e^2 \sim 0.32\ \mu m^2$, $R_g^2 \sim 0.05\ \mu m^2$). While, for $L \gtrsim 2$Mbp, they converge towards a more compact pseudo-steady-state characterized by smaller internal distances ($R_e^2 \sim 0.25\ \mu m^2$, $R_g^2 \sim 0.04\ \mu m^2$). This transition to a more compact state by increasing L is quantified in Fig. 3.7, where the deviation from the reference value is indicated. We indeed observe a significant deviation compared to the reference case at small $L \lesssim 1$Mbp of about $20-30\%$ which drops to 5% for $L \gtrsim 2$Mbp.

To further illustrate the variation in structural features when L varies, we plot the 3D distance squared $R^2(s)$ between two loci inside the domain as a function of the genomic

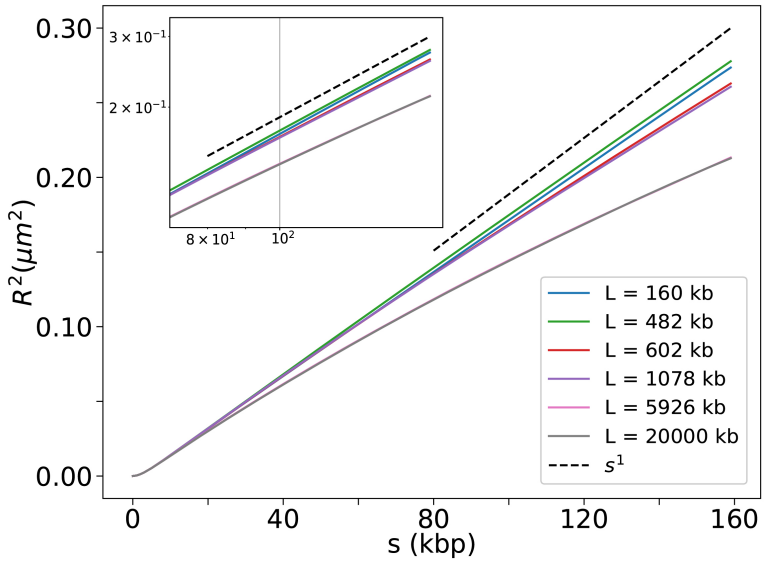

Fig. 3.8 3D distance squared between monomers of the domain R^2 as a function of genomic distance s. Inset shows the zoom of the large genomic distances ($s \gtrsim 50$ kbp) in log-log scale. Note that the curves for $L = 5926$kbp and $L = 20000$kbp overlap.

distance s between them ($s = |j - i|$) (Fig. 3.8) at equilibrium or pseudo-steady state. We observe that the transition between the low and high L-regimes happens at all scales even if it is significant only for large genomic distance ($s \gtrsim 50$ kbp).

Next, to generalize our observations, we then ask to what extent such L-transition depends on the size of the domain of interest. In Fig. 3.9, we plot the L-transition of R_e^2 for three different domain sizes L_d. Strikingly, we find that for all L_d values, there is a transition and that it occurs almost at the same L ($\sim 1-2$Mbp). The amplitude of the transition however depends on L_d, larger domains exhibit larger amplitudes, consistent with the previous observation that distances between larger genomic distances are more impacted by L.

Before any theoretical interpretation (see below, Sec.3.3.4), we can already conclude that, for a given L_d, there are indeed L values much smaller than the size of the entire chromosome arm (the reference state) that may fully capture the correct structural properties of the small region, at least in a simple homopolymer context. Hence, taking a value just above the transition provides already an empirical way of choosing a "minimal" L, here 2 Mbp.

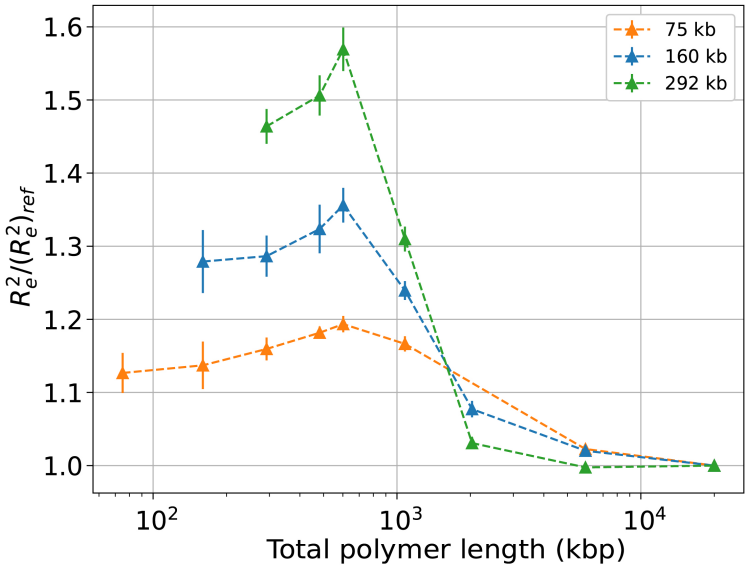

Fig. 3.9 Deviation of R_e^2 from the reference value (at $L_{ref} = 20Mbp$) as a function of total polymer length L as in Fig. 3.7A for different domain sizes.

3.3.3 Total polymer length affects domain mobility for a confined self-avoiding walk

In this section, we study how L may impact the mobility of monomers inside the domain d and more generally the dynamical properties of the region for $L_d = 160$kbp. For this, we monitor the average mean squared displacement (MSD) of individual monomers of the domain ($g_1(\Delta t)$, Fig. 3.10A), the MSD of the center of mass of the domain ($g_3(\Delta t)$, Fig. 3.10B) and the MSD of the end-to-end vector of the domain (Fig. 3.11). Note that whatever be the L value, these three quantities reach a steady-state very rapidly.

When the size of the full chain equals the size of the domain ($L = L_d$, blue curves in Fig. 3.10), we observe the behavior of a simple Rouse polymer [209]. The center of mass has a simple Brownian motion with $g_3(\Delta t) \approx D\Delta t$. Individual monomers exhibits a crossover between a sub-diffusive regime ($g_1(\Delta t) \propto \Delta t^{1/2}$) at small timescales and, at large timescales, a diffusive one, following the center of mass of the whole polymer.

When L increases, for g_3 (Fig. 3.10B), we observe that the motion of the center of mass depends on L, larger L leading to smaller MSD for the same time-lag. Indeed, the global motion of the domain is impacted by the rest of the chain, longer chains having slower diffusion coefficients [209]. Interestingly, beyond the structural L-transition described above ($L \gtrsim 2$Mbp), g_3 does not significantly depend on L, at least at time-scales where the motion of the center of mass of the full chain remains negligible ($\Delta t \lesssim 10^9$ MCS). In this case, g_3 of

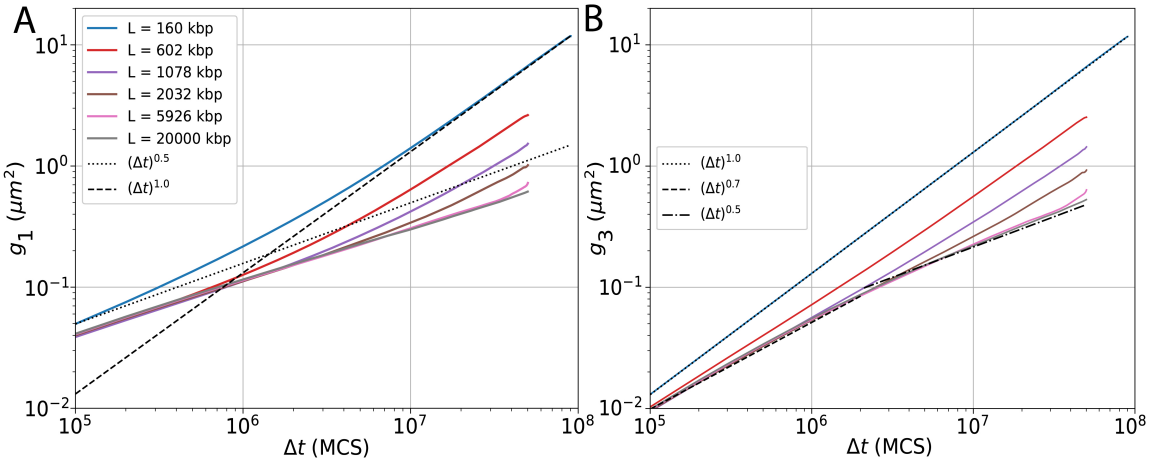

Fig. 3.10 Mean squared displacement (MSD) as a function of time-lag Δt for: (A) Individual monomer MSD inside the domain $g_1(\Delta t)$ and (B) the center of mass MSD of the domain $g_3(\Delta t)$.

the domain is sub-diffusive ($g_3 \propto \Delta t^{0.7}$ for $\Delta t \lesssim 2.10^6$ MCS and $g_3 \propto \Delta t^{0.5}$ for $\Delta t \gtrsim 2.10^6$ MCS) as expected for a sub-chain embedded into a larger one [179]. For g_1 (Fig. 3.10A), we observe that at small timescales, the monomer mobility is Rousean ($g_1(\Delta t) \propto \Delta t^{0.5}$) and does not depend on L. Only at larger timescales, L starts to impact g_1 as crossover towards the motion of the center of mass of the domain (g_3) is occurring. Again, for $L \gtrsim$ 2Mbp, g_1 becomes relatively independent of L.

MSD of the end to end vector of the domain quantifies the relative mobility of the monomers within the domain and may, biologically-speaking, characterize the motion between regulatory elements like promoters and enhancers [233]. For all L values, such MSD is composed by a sub-diffusive regime (MSD$\propto \Delta t^{0.5}$) at small time lags Δt that crosses over a plateau (MSD=2 $< R_e^2 >$) for larger time lags when configurations at time t and $t + \Delta t$ are fully decorrelated.

The monomers near the ends of a polymer have more degree of freedom (or are subject to fewer constraints). Therefore, when $L = L_d$, as opposed to when $L > L_d$, maximum mobility is reached (blue curve, Fig. 3.11) as expected. For $L > L_d$, the MSD is slowed down and it takes much longer time to reach the plateau. Interestingly, we observe a similar behavior as we observed for the equilibration time for R_e^2 (Fig.3.6A): for $L \lesssim 1-2$Mbp, the decorrelation time (time to reach the plateau) is an increasing function of L; for $L \gtrsim 1-2$Mbp, MSD reaches a slightly rising false-flat, similar for every L. Intriguingly, for L values beyond the structural L-transition where we see higher domain compaction ($L \gtrsim$ 2Mbp), the relative mobility of the ends of the domain is weakly but significantly increased (pink and brown curves in Fig. 3.11), indicating that the variation in structural and dynamical properties of

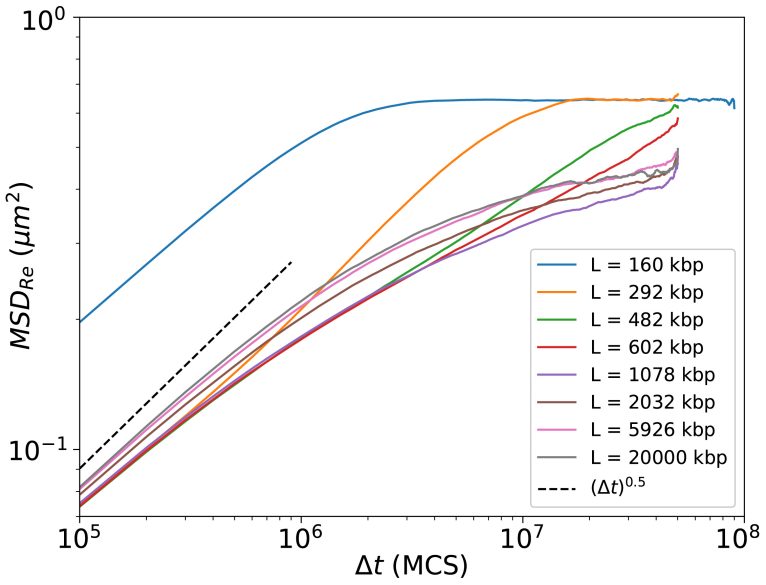

Fig. 3.11 Time evolution of MSD of end to end vector of the domain for different polymer length L.

the domain may be coupled: slightly more compact organization leading to slightly faster relative dynamics.

To conclude, the reference dynamical properties for a given L_d (i.e., those obtained for $L = L_{ref}$) may be fully captured by simulations of smaller polymer chains, providing that L is beyond the L-transition ($L \gtrsim 2$ Mbp in our model) like for the structural properties.

3.3.4 Total polymer length modulates the topological regime for a confined self-avoiding walk

In this section, we seek to theoretically interpret the influence of L on the compaction and mobility of small embedded domains in the simple case of a confined self-avoiding homopolymer.

In simple polymer models like Gaussian chains (Rouse model), Worm-like chains, phantom chains or random walks where excluded volume is neglected, the (steady-state) statistical properties of a subchain should not depend on the rest of the chain (i.e. of L). In Fig.3.12, we indeed verify that, if we remove the excluded volume constraint in our simulations, $< R_e^2 >$ is independent of L. Moreover, we also observe that the internal dynamics is also conserved (Fig. 3.13). This suggests that the L-transition is related to excluded volume and topological constraints.

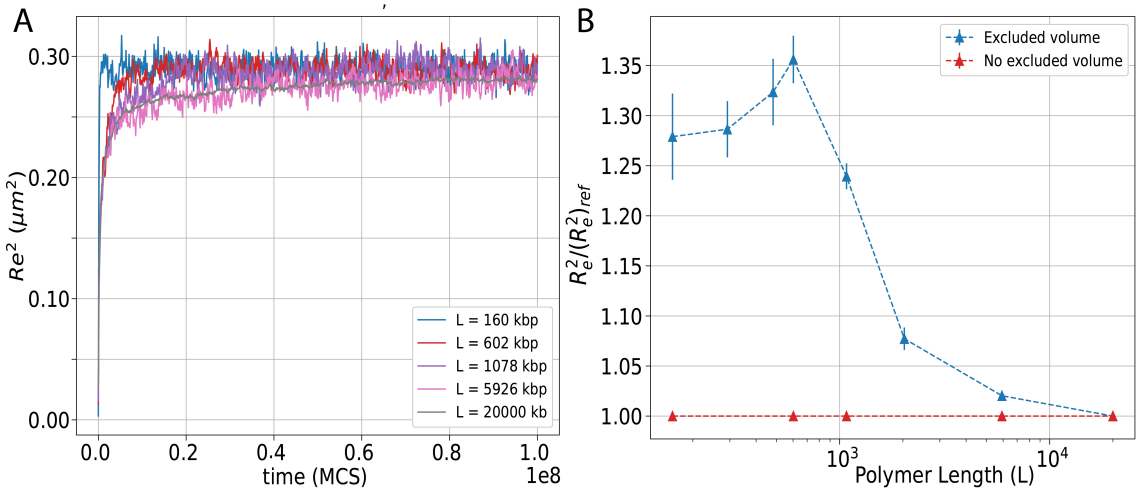

Fig. 3.12 (A) Time evolution of a R_e^2 of 160kbp domain for different total polymer length L, in absence of excluded volume interactions. (B) Deviation of R_e^2 of 160kbp domain from the reference as a function of L in presence (blue) or absence (red) of steric hindrance.

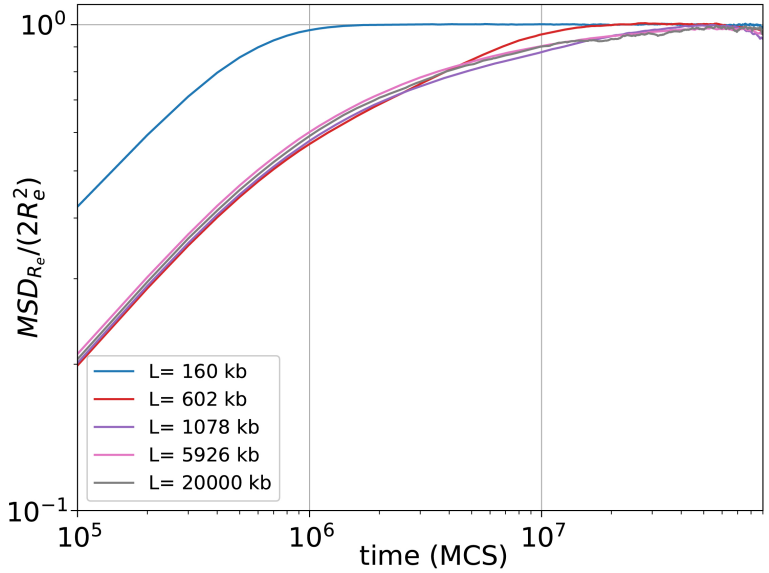

Fig. 3.13 Time evolution of MSD of end to end vector of the domain (normalized by the plateau value) for different polymer length L, without excluded volume interactions.

Indeed, long confined polymers like chromosomes [205] whose sub-segments generally cannot cross one another (like a macroscopic string) are subject to topological constraints which may considerably affect the polymer conformation and dynamics [206]. The importance of such constraints depends on the ratio between the contour length of the polymer $L_c = Nb$ and the so called entanglement length, L_e [228]. L_e could be related with the tube diameter in the reptation model [209] or to the crossover time between a Rouse-like motion and a reptation-like motion [228]. It depends on the Kuhn length, l_k characterizing the rigidity of the chain and on the volumic density ρ_{bp} (in bp/nm^3) and could be phenomenologically estimated via the relation [210]:

$$L_e \approx l_k \left(\frac{19}{\rho_k l_k^3} \right)^2 \tag{3.8}$$

with $\rho_k = (\rho_{bp}/n)(b/l_k)$, the volumic density in Kuhn segments. The same description can be thought of by defining an overlap parameter, $\Omega(N_{bp}) = \frac{\rho_{bp}}{N_{bp}} < R^2(N_{bp}) >^{3/2}$ [211], where $R^2(N_{bp})$ is the squared 3D distance corresponding to a genomic distance of N_{bp}. The length (in bp) at which Ω reaches the characteristic threshold of 19 determines the entanglement length (in bp). With the model parameters investigated above ($l_k = 100$nm and $\rho_{bp} = 0.009 bp/nm^3$), we have $L_e = 10.9\mu m \equiv 556$ kbp. If $L \gg L_e$, topological constraints are strong and the large-scale dynamics of the polymer might be very slow, meaning that the characteristic timescales of equilibration (or of decorrelation) of the lower Rouse modes of the chain (ie corresponding to large-scale variables, like the end-to-end vector) could be very large, and that the polymer keeps memory of the initial topological state of the chain (eg, presence or not of knots, large-scale organization) for a very long time [211, 90], way larger that the total simulation time (that corresponds approximately to ~ 1 hour of real time in our case).

In our simulations, as we increase L, the system transitions from a regime of weak topological constraints ($L/L_e \lesssim 1$) to a regime of strong constraints ($L/L_e \gg 1$). We thus wonder if the variation in domain structure and dynamics observed above (Sec. 3.3.2 and 3.3.3) is actually driven by such transition in topological regimes.

To do so, we perform simulations and analyze the properties of the end-to-end vector of the domain d ($L_d = 160$kbp) for polymer models with different entanglement lengths L_e. In addition to the default one ($L_e = 556$kbp) studied above, we explore 3 other situations with lower L_e by varying the dimensions of the simulation box, ie modifying the monomer volumic density and thus L_e (Eq.3.8). For each situation, we also reach a (pseudo)-steady-state (Fig. 3.14A) that allows us to precisely define R_e^2. In Fig. 3.14B, we observe that the L-transition is also present for the other L_e values. Importantly, the amplitude of the

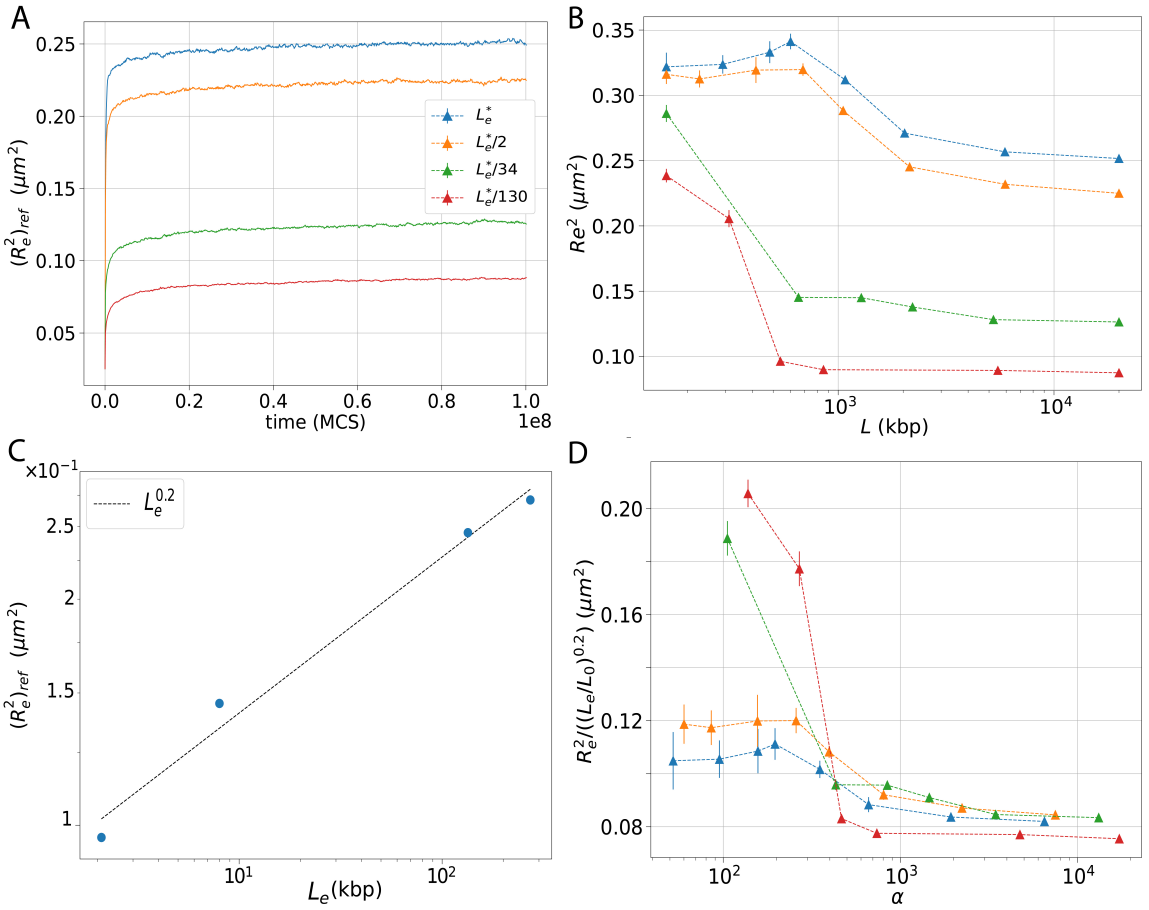

Fig. 3.14 (A) The reference end-to-end distance squared as a function of time for different entanglement length L_e. (B) R_e^2 of 160 kbp domain as a function of total polymer length L for different L_e values. (C) The pseudo-steady-state value $(R_e^2)_{ref}$ as a function of L_e. (D) R_e^2 normalized by $(L_e/L_0)^{0.2}$ as a function of α, where $\alpha = L/(L_0(L_e/L_0)^{0.2})$, for polymers of different entanglement length L_e. $\alpha \gtrsim 1000$ delimits the region where structural properties are stable irrespective of L.

transition is larger for lower L_e and the critical value of L where the system shifts from less to more compact is smaller for lower L_e values. We also observe that, not only the end-to-end distance, but all internal distances are affected by L_e (Fig. 3.15). The differences seen at large genomic distances between $L = L_d$ and $L = L_{ref}$ are stronger for smaller L_e but also the two curves starts to deviate from each other at shorter genomic distances, meaning that topological effects become more prominent at smaller genomic scale (red curve, $s \lesssim$ 8kbp) for smaller L_e. All this confirms that the L-transition is driven by a change of topological regime for the whole chain, and that large domains (or genomic distance) compared to the entanglement length ($L_d/L_e, s/L_e \gtrsim 1$) would be impacted more.

To quantify this, in Fig. 3.14C, the pseudo-steady-state value $(R_e^2)_{ref}$ of R_e^2 in the reference case ($L =$ 20Mbp) is plotted as a function of entanglement length L_e, and empirically we obtain a scaling law for the end-to-end distance squared after the transition $(R_e^2)_{ref} \propto L_e^{0.2}$. As expected, strong topological constraints lead to more compact domains. We define $\alpha = L/(L_0(L_e/L_0)^{0.2})$, a dimensionless quantity to estimate the extent of topological constraints in the system with $L_0 =$ 1kbp, an arbitrary default value. Phenomenologically, we find that $\alpha \gtrsim 1000$ corresponds to the values of L for which the structure of the domain is not affected anymore by increasing L (Fig. 3.14D). In other words, more practically, a polymer of length $L > L_\alpha \equiv 1000 * (L_0(L_e/L_0)^{0.2})$, would preserve the structural properties of the domain, like being simulated within the whole (reference) chromosome.

Altogether, our observations suggest that the L-transition is related to excluded volume and topological constraints. In the regime where topological constraints are relevant ($\alpha \gtrsim$ 1000 or $L \gtrsim L_\alpha$), the subchain maintains its topological state during the whole simulation, i.e., a knot-free topology in our case. The system is thus at a quasi-equilibrium but for a fixed topology. In the knot-free case, this limits the structural fluctuations of the subchain and leads to more compact structures, as typically observed for crumpled or 'fractal' polymers [147, 206, 211] that are often associated with compact, space-filling curves [92].

3.3.5 Total polymer length impacts the coil-to-globule transition

The modeling of genome organization and in particular of the 3D compartmentalization of the epigenome (chromatin states, e.g. euchromatin vs heterochromatin) has led to the introduction of copolymer models of chromatin. Beyond the excluded volume constraint, these models [56, 66, 90, 64, 203] include chromatin-state dependent self-attraction (J) between loci sharing the same epigenomic content (see Models and Methods). J effectively accounts for different physico-chemical mechanisms such as charge modulated nucleosome-nucleosome interaction [231], self-associating proteins bridging [138, 142, 139] or solvent quality [232]. As a general (universal) behaviour, a homopolymer with self-attraction J between all its monomers experiments a coil-globule transition when increasing J, from an

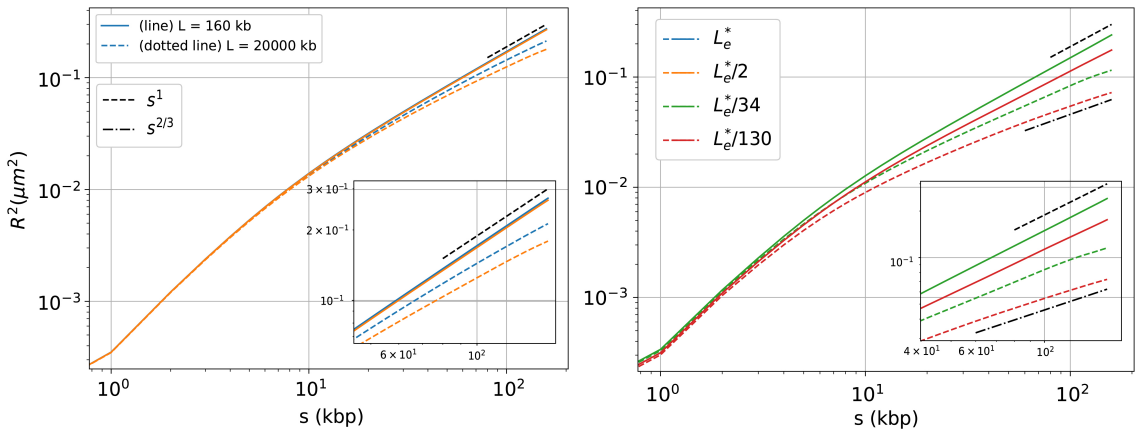

Fig. 3.15 Average squared 3D distance $R^2(s)$ between monomers of the domain ($L_d =$ 160kbp) as a function of genomic distance s. Solid lines correspond to $L = L_d$ and dotted lines, $L = L_{ref} =$ 20000kbp. Inset shows a zoom at large genomic distances ($s \gtrsim 50$ kbp) in log-log scale. The four colors indicate four different entanglement lengths with $L_e^* =$ 556kbp.

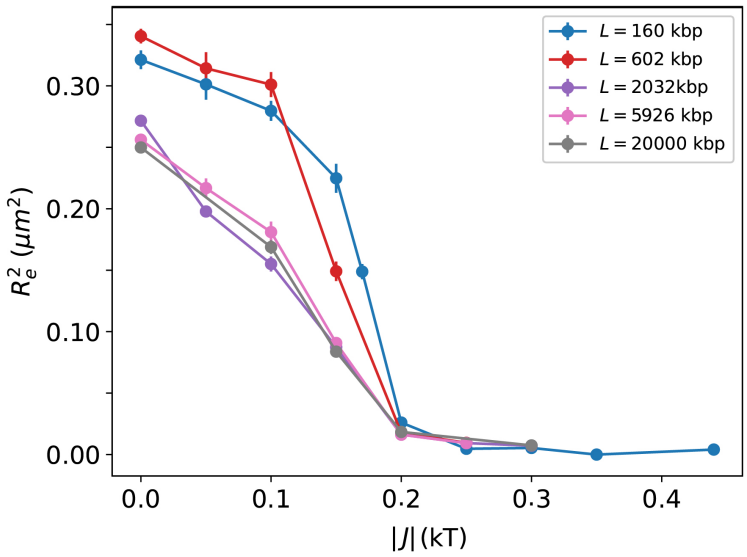

Fig. 3.16 R_e^2 for a 160 kb region as a function of interaction strength $|J|$ for different total polymer lengths L.

expanded coil-like conformation for $|J| < |J_c|$ to a compact globule-like conformation for $|J| > |J_c|$ [229, 230], where J_c is the transition point (called the "theta" point). In this case, the collapse transition depends on the (finite) size of the chain [201, 202] with, for example, sharper transitions (susceptibility) for larger sizes.

Here, in continuity with previous sections, we propose to study how the embedding of a self-interaction domain inside a larger neutral chain of size L influences its conformational properties.

Generic considerations in long, reference polymers

Here, we thus employ the copolymer model to study how the total polymer length would affect the coil-globule transition of a particular region of interest, taking as reference the situation for $L_{ref} = 20Mbp$. To simplify, we consider a region of size $L_d = 160$ kb of a particular epigenomic state embedded into a neutral chain of size L. For all L, as we vary the strength of self-attraction J, the subchain gradually collapses (Fig. 3.16) with R_e^2 of the region following a standard coil-globule transition, starting from an initial expanded state (higher R_e^2) to a more compact (lower R_e^2). Beyond the theta-collapse, in the globular state, R_e^2 is independent of L. However, for $|J| < 0.2kT$, during the collapse, R_e^2 becomes more sensitive to L. Consistent with our observations in the pure self-avoiding case that larger L lead to more compact subdomains (Fig. 3.7A), larger L would be more collapsed for a given J value. Interestingly, the coil-globule transition curves are almost similar for all $L \gtrsim L_\alpha \approx 2$Mbp that corresponds to the critical L value of the L-transition observed in the pure self-avoiding system.

To conclude, we find that when the sub-chain is collapsed, the effect of L is not significant, while in a coil or partially collapsed state, simulations for small L values deviate from the reference one unless α is larger than 1000 to ensure that the correct topological regime is established.

Specific considerations in short, reference chromosome

In this subsection, we contextualize our analysis to a specific contact domain in yeast (region $334274 - 378470$ bp of chromosome 12, $L_d = 45.6$kbp) embedded into a "small", neutral reference chromosome of size $L_{ref} = 1.1$Mbp, to illustrate the role of L_{ref}.

We simulate two scenarios: (i) one with $L = L_d$ and (ii) one with $L = L_{ref}$ the size of chromosome 12. For each scenario, we systematically vary J and predict in each case the Hi-C map (for various radius of contact R_c). For both scenarios, the best fitting radius of capture is $R_c = 60nm$. Interestingly, the best fitting strength of interaction is the same when $L = L_d$ ($J^{(i)} = -0.25$ kT) and $L = 1.1$ Mb ($J^{(ii)} = -0.25$ kT) (Fig. 3.17) which suggests that the total polymer length doesn't drastically affect the 3D organization of the domain

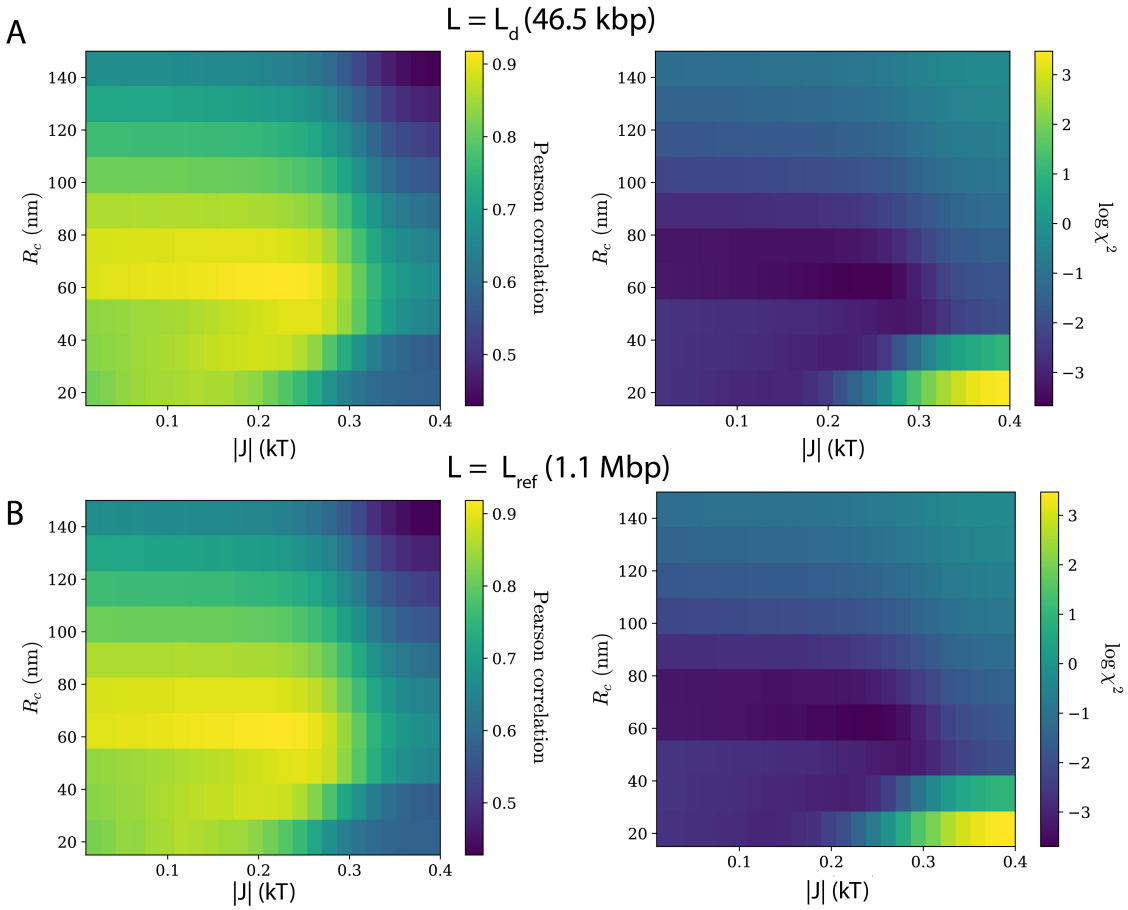

Fig. 3.17 Phase diagram (R_c and J) quantifying the predictive power of the simulation relative to experimental Hi-C data using two different metrics: (i) Pearson correlation (left) and (ii) χ^2 as a function of two parameters (right). (A) For $L = L_d = 46.5$kbp, the best fitting parameters (minimum χ^2) being $R_c = 60$nm, $J = -0.25$ kT ;(B) For $L = 1.1$Mbp, the best fitting parameters (minimum χ^2) being $R_c = 60$nm, $J = -0.25$ kT.

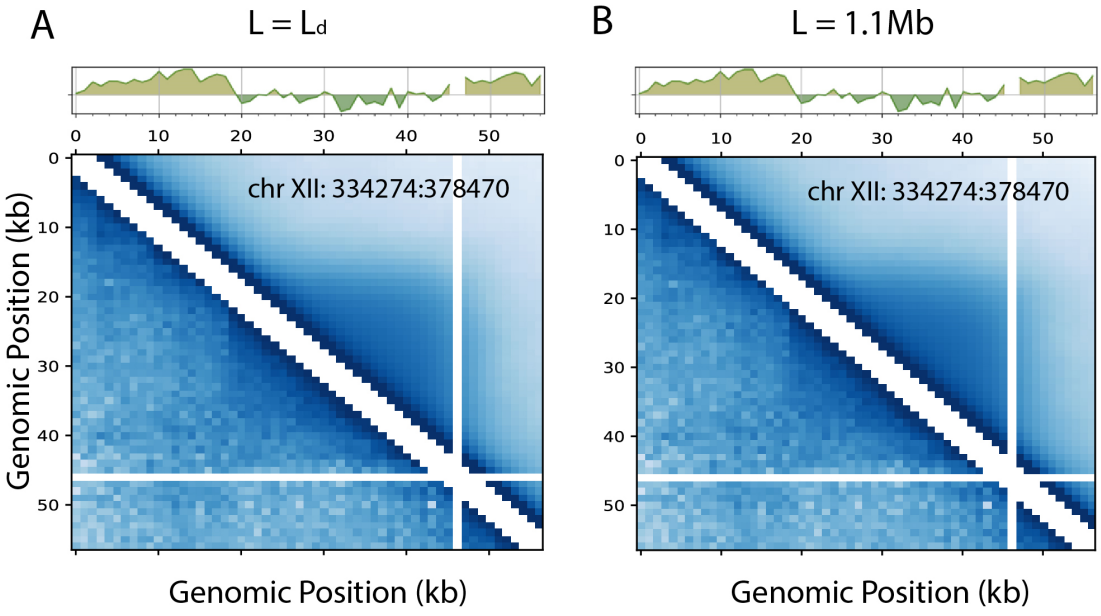

Fig. 3.18 Simulated (upper part) and experimental (lower part) Hi-C of a yeast domain for simulation with (A) $L = L_d = 46.5$kbp and $J = -0.25kT$ (corresponding to Fig.3.17A); with (B) $L = 1.1$ Mbp and $J = -0.25kT$ (Fig. 3.17B). The monomers within the gene (interacting M-state) and outside (non-interacting U-state) are shown at the top (PCA analysis). Pearson correlation value:$= 0.92$ for $L = 1.1$Mbp and $L = L_d = 46.5$kbp

in yeast. In Fig. 3.18, the predicted Hi-C maps of the region in both scenarios are shown, exhibiting both high correlations (0.92) with the experimental map.

Relation to L_e

The total length of the polymer L plays a crucial role in the structure and dynamics of the domain, in particular in the case of systems where strong topological constraints $L_{ref}/L_e >> 1$ are present in the reference case, like the genomes of drosophila [90] and higher eukaryotes. Within our phenomenological analysis, it means that topological constraints may be needed to be accounted for when $\alpha(L_{ref}) > 1000$, in particular when the region is not fully collapsed or globular (see Fig.3.16) as often observed in biological context [90].

In the generic consideration above with a long reference chromosome, $\alpha(L_{ref}) \gg 1000$, and we expect that the minimal length of polymer to simulate to preserve the structural reference properties (as in the full chromosome), is $L \gtrsim L_\alpha \approx 2-4$ Mbp.

In the specific yeast case with small reference chromosome ($L_e = 2.2$ Mbp, $L_\alpha \approx 5$Mbp), $\alpha(L_{ref}) = 236 \ll 1000$, so we expect topological constraints not to be too relevant and thus that L has only a limited impact.

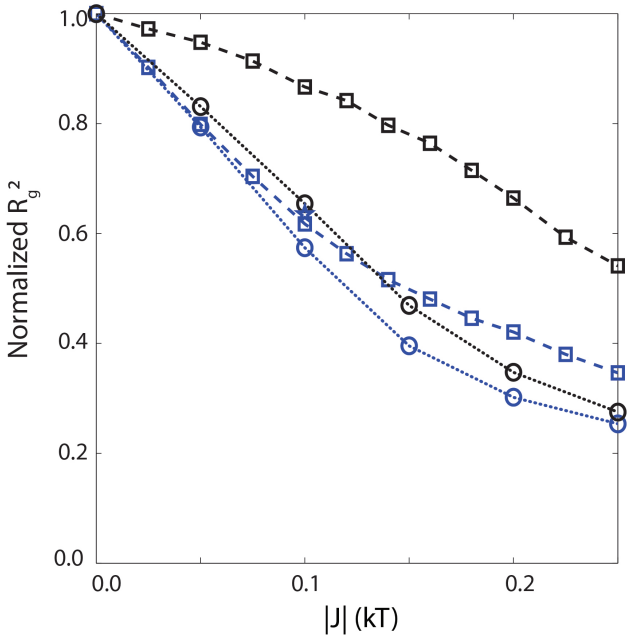

Fig. 3.19 Evolution of normalized $(R_g)^2$ as a function of the strength J of self-attraction for (i) a 330kbp-long domain (blue squares) isolated from other domain of the same epigenomic state along the sequence; (ii) a 290kbp-long domain (black squares) surrounded along the genome by many other domains of the same state. Data were normalized by the corresponding values in the neutral case ($J = 0$). These simulations correspond to a block copolymer model of chromosome arm 3R in Drosophila accounting for self-interactions between loci sharing the same epigenomic state [90]. Circles represent situations when the interactions were only between the monomers of the domain of interest (no interaction with the other loci of the same epigenomic state dispersed along the genome). *Adapted from [204].*

3.3.6 Total polymer length and epigenomic context drive genome folding

In the analysis presented above, we have not accounted for the environmental contribution, like, for example, for putative epigenomic interactions between the domain of interest and the rest of the genome which might potentially affect the local structure of the region. We have considered regions which are completely isolated to focus on how topological constraints impact the 3D organization of the region. But it is evident that the overall epigenomic context and associated interactions would be pertinent to fully describe the structure and dynamics of the domain. For example, in the book chapter published in 2019 [204], my supervisors partially explored this question by investigating how the radius of gyration of a $\sim$ 300 kbp-long domain may depend on the strength J of self-attraction if the domain was isolated inside a long, neutral polymer (a situation very close to our analysis in Sec.3.3.5 in

the reference case) or if the domain was simulated as being part of a larger block copolymer thus potentially interacting with other domains of the same epigenomic state along the chromosome (Fig. 3.19). They observed that if the domain was already isolated (i.e., far from other regions of the same state along the genome), only considering the self-interacting domain inside a neutral chain, does not impact the structure of the domain strongly (blue lines in Fig. 3.19). However, for a domain that is close to other regions of the same state (black lines in Fig. 3.19), accounting for the context have a strong impact, simulations of a fully isolated domain leading to more compaction than contextualized simulations.

To go beyond this, and try to couple my analysis of the total polymer length L with the environmental epigenomic context, we simulate a domain of interest corresponding to the region $17955-18262$ kbp in chromosome 3R of Drosophila ($L_d = 307$ kbp), which is denoted as a 'black' epigenomic domain (silent chromatin) by the classification done in [71]. In its native context, this domain is embedded into a polymer of size $L_{ref} = 32.1$Mbp and the domain can interact with all the other black domains along the chromosome.
We simulate three scenarios: (i) $L = L_d$; (ii) $L = 6.2$ Mbp, a length ($\approx 2 * L_\alpha$) that should capture the polymeric, reference case based on our theoretical analysis; and (iii) $L = L_{ref}$ Mbp. For each case, we vary J and compute Hi-C maps for different radius of capture R_c (Sec.3.2.4), and estimate the parameters that best reproduces the experimental HiC of the region in the each scenario (Fig. 3.20). In contrast to our previous analysis, we now are considering a domain that is in a dense interacting environment, potentially having lot of interactions with domains of the same epigenomic state along the chain.

For the three cases, the best fitting radius of capture is $R_c = 60nm$. Interestingly, we find two different optimal J values ($J = -0.17$kT for scenario *i* and $= -0.135$kT for scenario *ii* and *iii*), confirming that simulating $L = L_d$ and $L = 32.1$Mbp is quantitatively different also for a highly interacting domain embedded inside a long chromosome with interacting domains. It has to be noted that in all the cases, the best fitting parameters predict the Hi-C map with good accuracy (Fig. 3.21).

We note that changing the total polymer length significantly alters the 3D organization of the region of interest (about 25 percent change in the interaction strength J).Interestingly, we find that the length predicted by our analysis on topological constraints is sufficient (Fig. 3.20B, Fig. 3.21) to accurately account for the right epigenomic and topological regime.

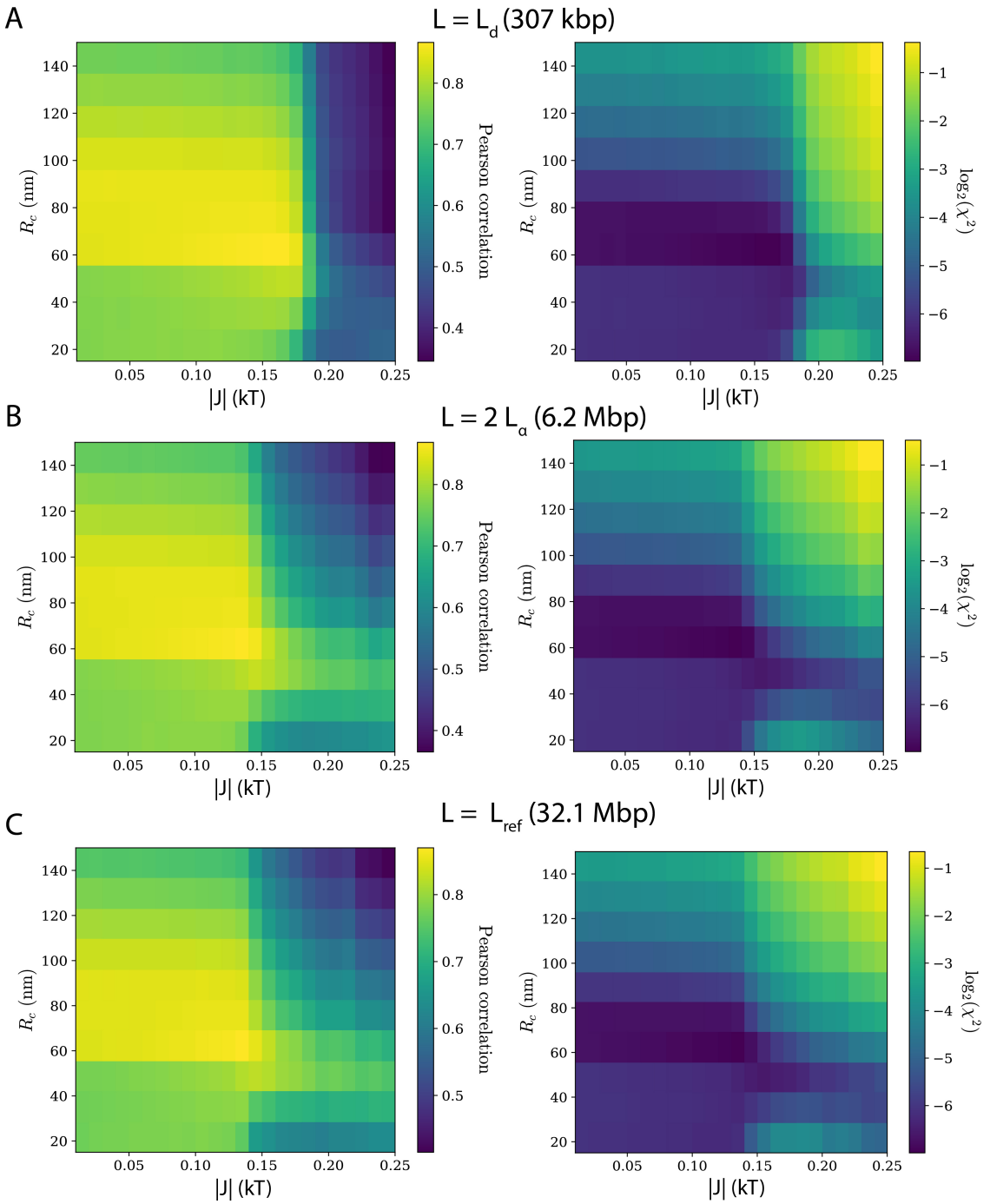

Fig. 3.20 Phase diagram (R_c and J) quantifying the predictive power of the simulation relative to experimental Hi-C data using two different metrics: (i) Pearson correlation (left) and (ii) χ^2 as a function of two parameters (right). (A) For $L = L_d = 307$kbp, the best fitting parameters (minimum of χ^2) being $R_c = 60$nm, $J = -0.17$ kT ; (B) For $L \approx 2 * L_\alpha = 6.2$ Mbp the best fitting parameters (minimum of χ^2) being $R_c = 60$nm, $J = -0.135$ kT ; (C) for $L = L_{ref} = 32.1$ Mbp, the best fitting parameters (minimum χ^2) being $R_c = 60$nm, $J = -0.135$ kT. These values are in agreement with maximum Pearson correlation.

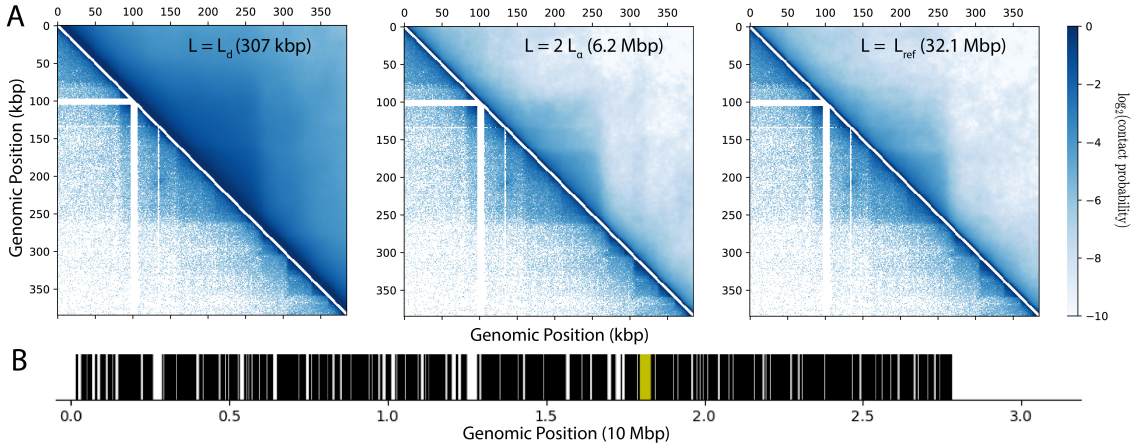

Fig. 3.21 Simulated (upper part) and experimental (lower part) Hi-Cs of a drosophila domain for simulations with (Left) $L = L_d = 307$kbp and $J = -0.17$kT (optimal case in Fig.3.20A). (Middle) $L = 6.2$ Mbp and $J = -0.135$kT (Fig. 3.20B). (Right) $L = 32.1$Mbp and $J = -0.135kT$ (Fig. 3.20C). The monomers within the TAD and outside are shown at the top (PCA analysis). It is to be noted that the 3D organization of the domain is very similar in the case of $L = 6.2$Mbp and $L = 32.1$Mbp. Pearson correlation values obtained for these parameters are 0.86, 0.87 and 0.87 respectively. The monomers of the interacting epigenomic state (M) and the non-interacting U-state are shown at the bottom (PCA analysis). (B) Shows the distribution of black domains along chromosome 3R, the domain we have considered is highlighted in yellow. *Data from [71].*

3.4 Discussion and conclusion

In this chapter, we have illustrated the role of total polymer length on the structural and dynamical properties of a specific region of interest. Our analysis using a simple self-avoiding homopolymer gives quantitative predictions on the minimal length of DNA to be simulated around a specific region of interest to fully capture the correct polymeric properties of the region.

We have systematically studied how increasing the total polymer length L would affect the structural properties ($R_e^2, R_g^2, R^2(s)$) of a region within. In particular, we have quantified how expanded or open the region (Fig. 3.7, 3.8) would be in comparison with a reference case corresponding to a long chromosome ($L = 20$ Mb). It has to be noted that for long polymers ($L = 20$ Mb), the time evolution of R_e^2 of the domain indicates the presence of a very slowly expanding (pseudo)steady-state, while when the total polymer length is shorter, the structure equilibrates much faster (Fig. 3.6). In addition to the structure, the length of DNA around the region also affects the dynamics of the region, shown by the difference in mobility (Fig. 3.7) when the total polymer length is varied. We have observed that without

excluded volume interactions, the structure and dynamics of the domain is not affected by the total polymer length (Fig. 3.12, 3.13), suggesting that steric interactions are responsible for the different behaviors observed when varying L. Furthermore, we quantified the extent of the corresponding topological constraints by a single parameter (α) that depends on the entanglement length. This parameter helps in estimating the minimal length L_α of DNA to be simulated to capture the correct - pure- polymeric properties. These results would significantly help in extracting accurate physical measures of different domain using simple polymer models [203].

We used a simple copolymer framework to understand how the total polymer length may affect the folding properties of the domain of interest. Beyond the Θ collapse, we showed that the globule-state is unaffected by the total polymer length (Fig. 3.16). However, before the collapse (a biologically-relevant regime [90, 95, 56]), we found that L may impact the folding and should be chosen carefully like in the pure self-avoiding case. Finally, we have contextualized the model to experimental data, in drosophila [222] (with strong topological constraints) and yeast [223] (weakly constraint). In drosophila, L may alters the 3D organization of the region (Fig. 3.20) while in yeast, it doesn't play a crucial role (Fig. 3.17). This conclusion is further compounded by the fact that the interacting domains (which mostly corresponds to genes [199]) in yeast are very small ($<$ 10kbp) and thus topogical effects have less impact on the organization (Fig. 3.9).

Overall, our results suggest that the pure polymeric - self-avoidance related - contribution, and thus the proper spatio-temporal dynamics of a domain may be well accounted by: (i) $L \gtrsim L_\alpha$ for a long reference chromosomes ($L_{ref} \gg L_\alpha$) and a coil or partially-collapsed state for the domain of interest ; (ii) $L \approx L_d$ for a globule-state for the domain or a very short domain ($L_d \ll L_e$) or a short reference chromosome ($L_{ref} \ll L_\alpha$).

As stated in the Introduction of this Chapter, there may be two main contributions of the total polymer length: the polymeric effect that we mostly focus on in this chapter and the epigenomic context. In the last part of our results, we explored how accounting for the epigenomic environment may impact the 3D organization of the region of interest. Previous works and our preliminary analysis suggest that (1) accounting for the context (compared to a situation where a domain is embedded inside a neutral polymer) is crucial as it may impact the predicted structural (and dynamical, not shown) properties (Fig.3.19); (2) if L is large enough to capture the polymeric behavior, it should be also large enough to contain enough 'context' to capture the correct behavior. A more systematic analysis of the context by studying more L values or different epigenomic environments would help rationalizing the importance of the context contribution versus the polymeric one. In addition to the epigenomic context, similarly, it is also likely that association with nuclear lamina [235] or other nuclear bodies might also potentially bring in additional geometric constraints that

may contribute to the behavior of a domain of interest and that may depend on the size L of the simulated polymer.

One key result of this work is the requirement of maintaining the system in the correct topological regime to ensure that polymeric properties are conserved. However, this result is based on the physical hypothesis of the model that the polymer strands cannot cross each other (or at least only rarely), thus maintaining the system in the same - knot-free - topological state during the simulations of long polymers [196]. Experiments have estimated that indeed chromosomes are mostly knot-free or only weakly knotted [214, 215] during G1. However, it is not clear if, *in vivo*, such a state actually emerges from the decompaction of a knot-free, mitotic-like initial configuration that would remain knot-free thanks to rare strand crossing events. There is actually evidence that the activity of the topoisomerase II, a protein that may catalyze strand crossing, may be important in the higher-order organization of the genome [236, 237]. The knot-free state will then rather emerges from the combination of topo II activity and of active loop extruding factors [240] like cohesins or condensins which are thought to play crucial role in TAD formation [118]. It thus would be very interesting to introduce both ingredients (some chain crossing in a loop extrusion polymer model [119]) to investigate how our conclusions are modified.

The lattice polymer model enables us to model explicit 3D chromatin dynamics. Now, we have (1) practical rules to model biological systems at an optimal size to minimize the computational burden while capturing the correct polymeric behavior and (2) a quantitative framework (lattice polymer) to explore, how to integrate the Painter model (Chapter 2) with an accurate 3D polymer dynamics in order to study the coupling between 3D chromatin organization and epigenome regulation (Chapter 4). For instance, to model the Daschung (dac) locus ($L_d \sim 80$ kb in chromosome 2L [198]) which is a small isolated Polycomb domain, our analysis would suggest that polymer simulations with a total polymer length $L > 2-3$Mb (orange curve, Fig. 3.9) would correctly capture the structural and dynamical properties of the region.

Chapter 4

Towards a quantitative model coupling epigenome regulation and 3D polymer dynamics

4.1 Introduction

As described in Chapter 1, there is an explicit coupling between epigenome regulation and 3D genome organization: epigenomic information may spread in 3D via the action of HMEs (like Clr4 or PRC2) and epigenomic landscape may impact genome folding via the recruitment of architectural proteins (like HP1 or PRC1). These two mechanisms may lead to a positive feedback loop where the epigenomic-driven compaction of the genome organization facilitates the spreading in 3D which leads to higher compaction and so on. Therefore, one interesting perspective of my work would be to upgrade the painter model (detailed in Chapter 2) to account for the spatio-temporal dynamics of the chromosome and how it impacts the spreading and maintenance of an epigenomic signal, more realistically.

This feedback was already investigated by my supervisors via the "Living chromatin" model [66] (described in Chapter 1) that couples an explicit description of the 3D polymer dynamics with a 3-state model with antagonistic marks for the epigenome (as in Fig.2.33). However, such model was very simplified and generic. For example, it does not integrate the complete description of epigenomic regulation that we have developed in Chapter 2 in terms of recruitment (sequence- or state-dependent), spreading and allostery. It also does not explicitly account for the role of HMEs and architectural proteins, while the concentration, distribution, kinetics and binding properties of HMEs and architectural proteins may play an important role in regulating the epigenome and organizing the 3D genome and while quantitative estimations of these properties start to be available [137, 248, 249, 246, 139,

245]. It is therefore of high interest to continue the characterization of this feedback using more advanced models based on the results of my PhD thesis.

In this prospective chapter, I will introduce how to develop models of increasing complexity by integrating step-by-step the ingredients described above, towards a quantitative "Living painter" model that explicitly integrate the 3D chromosome and histone modifying enzyme (HME) dynamics, which can be used to systematically dissect generic or specific biological problems in the future. In particular, we will focus on precisely describing the step-wise increments of our mathematical framework and on illustrating some of their features, notably how the feedback loop may impact the epigenomic landscape. This chapter also recapitulates one of the key findings of the painter model and the living painter model rightly marks the significance of limited enzyme regime in the formation of confined chromatin state domains. Note that this chapter is exploratory and does not aim to provide a systematic investigation of the ingredients of the model, but introduces several concepts that will build the ground for short-term perspectives of my work (see Sec. 5).

4.2 Modelling epigenome regulation in 3D

In this section, we describe different models of increasing complexity, translating the simple painter model (detailed in Chapter 2) into the "Living painter" model in order to explicitly investigate the coupling between epigenome regulation and 3D chromosome organization.

4.2.1 Coupling the painter model with 3D polymer dynamics

First, we propose a simple Living Chromatin-like model (Fig. 4.1) that combines the Painter model for epigenome regulation (see Chapter 2) and the lattice polymer model for 3D polymer dynamics (see Chapter 3) to explicitly study the coupling between the two processes, as proposed in the "Living Chromatin" framework [66]. In the original painter model, 3D chromosome organization was effectively accounted for via the spreading probability term $P_c(i, j)$ between two genomic positions i and j. Here, we introduce the polymer dynamics explicitly.

We consider a polymer composed by N monomers each consisting of n bp and or size b (same notation as in Chapter 3). Following the framework developed in chapter 2, an epigenomic state is associated to each monomer. This state can dynamically fluctuate between two flavors: modified M and unmodified U. The spatio-temporal dynamics of the polymer follows the same principles as described in Chapter 3 (Sec. 3.2): self-avoiding and semi-flexible copolymer chain evolving on a FCC lattice, only monomers currently in the M-state are assumed to be able to self-attract with a strength of interaction J (Eq. 3.3).

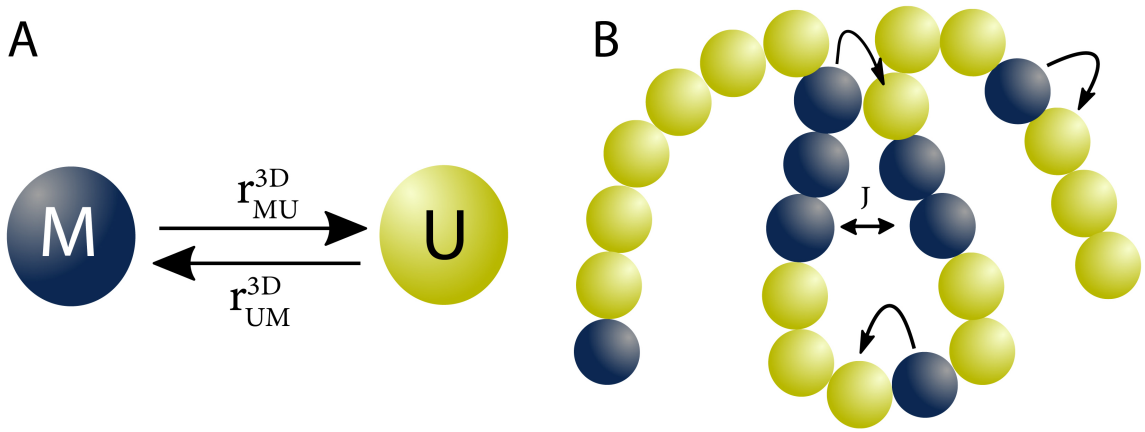

Fig. 4.1 An illustration of Living chromatin-like model, coupling the painter model (Chapter 2, Fig 2.1) and 3D polymer dynamics. (A) Nucleosomes/monomers can fluctuate between a modified M and unmodified U states driven by the transition rates r^{3D}_{UM}, r^{3D}_{MU}. (B) Polymer structure and dynamics are driven by short-range self attractive potentials of strength J between M monomers. U monomers are assumed to be neutral. Arrows indicate implicit spreading of M-state.

The epigenomic state dynamics of monomer i is driven by two transition rates $r^{3D}_{UM}(i)$ and $r^{3D}_{MU}(i)$ (similar to Eq. 2.2):

$$\begin{aligned} r^{3D}_{UM}(i) &= k\left[(\rho_s(i)+\Delta\delta_{i,M})(1+r\delta_{i,M})+\sum_{j\in\nu(i)}\varepsilon(1+r\delta_{j,M})(\rho_s(j)+\Delta\delta_{j,M})\right] \quad (4.1)\\ r^{3D}_{MU}(i) &= k_0 \end{aligned}$$

Here, the notations are similar to the painter model in Chapter 2 (Sec. 2.2). Briefly, k is the intrinsic activity of HMEs; r, the state-dependent boosted activity; Δ, the state-dependent recruitment of HMEs; ε, a factor that accounts for putative modulation of HMEs activity on trans-spreading; $\rho_s(i)$, the sequence-dependent recruitment of HMEs; k_0, the lumped turnover rate. Note that, for the spreading term in r^{3D}_{UM} (right part), the sum runs over $\nu(i)$, the current ensemble of monomers j that occupy nearest neighbors positions on the lattice to that of monomer i, accounting explicitly for the current 3D neighborhood.

Overall, this model is very similar to the original Living Chromatin model but contextualized to the painter model of Chapter 2. It does not explicitly account for HMEs for the epigenomic spreading or architectural proteins for the spatial organization, i.e., implicitly assuming that they are not limiting factors and that their intrinsic dynamics (diffusion, binding/unbinding rates) is fast and can be neglected.

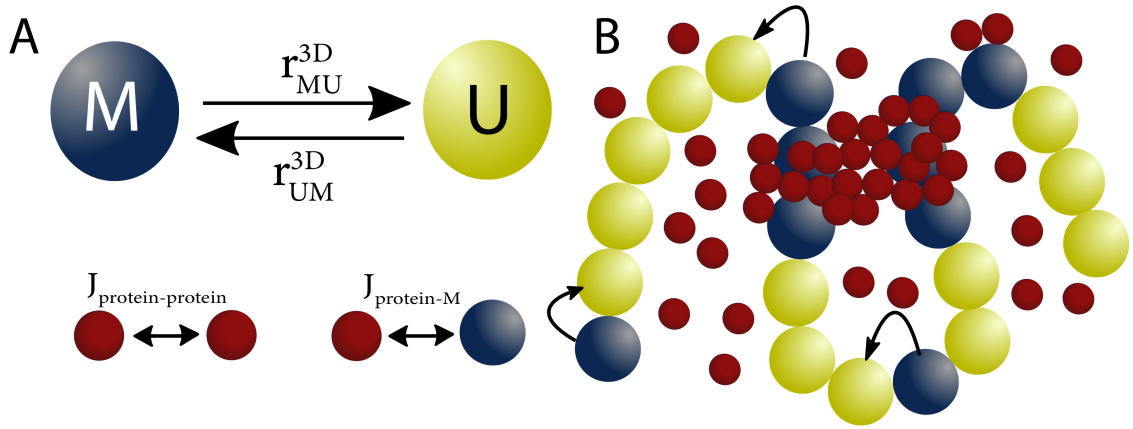

Fig. 4.2 Self-interacting proteins (brown, PRC1 in Fig. 4.8) driving chromatin folding. Here, the painter model coupled with the 3D polymer dynamics (as in the previous section) establishes the epigenetic landscape of the region (eq.4.1). However, the interaction between M-state is driven by self-attracting proteins ($J_{protein-protein}$) which can interact with the M-state ($J_{protein-M}$) as well.

4.2.2 Lattice-gas model for architectural proteins

Here, we modify the framework described in the previous section by explicitly accounting for the architectural proteins that mediates the epigenomic-driven interactions between M-state monomers.

While the transition rates to switch between M and U states are assumed to be the same as in Eq. 4.1, we introduce diffusible molecules, that correspond to the architectural proteins associated with the M state (like HP1 for H3K9me2/3 marks or PRC1 for H3K27me3 [141]). These proteins can self-interact and can also specifically interact with M state monomers. Following the framework developed by Maxime Tortora in the Jost group [245], we model them as a lattice gas evolving on the same FCC lattice as the polymer (Fig. 4.2). More precisely, we consider $N_{protein}$ molecules that diffuse on the lattice. One lattice site can be simultaneously occupied by both the chromatin polymer chain and at most one molecule. The dynamics of the proteins is driven by two contributions: (1) short-range, contact interactions between molecules; (2) short-range contact interactions with M-state monomers. The first contribution is quantified by the Hamiltonian:

$$E_{protein-protein} = -\frac{J_{protein-protein}}{2} \sum_{i \in \mathscr{P}} \sum_{l \in \mathscr{V}(i)} \sigma_l \tag{4.2}$$

where $J_{protein-protein}$ is the strength of self-attraction between proteins, the double sum runs over the ensemble $\mathscr{P}$ of proteins and the ensemble $\mathscr{V}(i)$ of 12 lattice sites nearest-neighbor

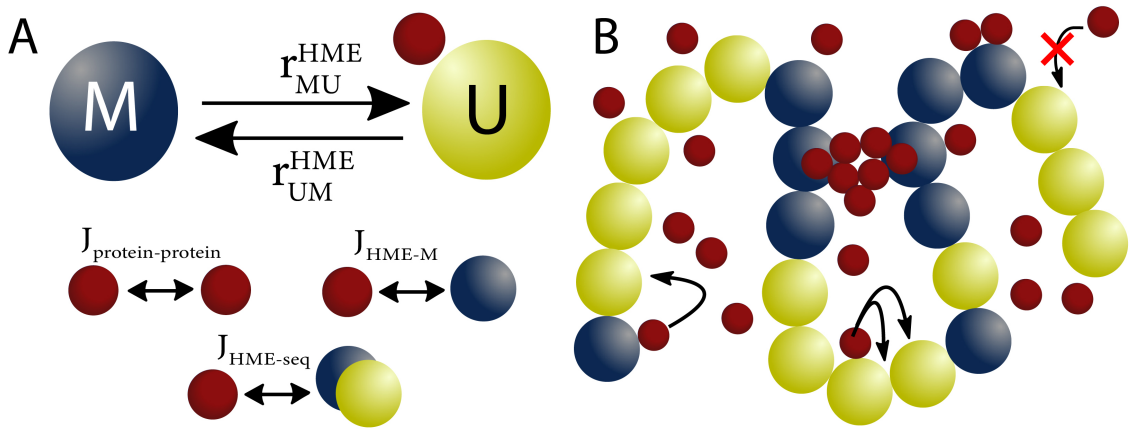

Fig. 4.3 A schematic representing the "Living Painter" model. Diffusing HMEs (painters) driving chromatin state spreading. In this formalism, painters spread modified state when they are spatially close to unmodified state. The painters/HMEs are capable of self-interacting ($J_{protein-protein}$) accounting for the architectural proteins, interacting with M-state (J_{HME-M}, accounting for state dependent recruitment of HMEs) and also the sequence ($J_{HME-seq}$) for sequence recruitment of HMEs.

of that occupied by protein i, and $\sigma_l = 1$ if a protein is on lattice site l ($= 0$, otherwise). The second contribution follows the Hamiltonian

$$E_{protein-M} = -J_{protein-M} \sum_{i \in \mathscr{P}} \sum_{l \in \mathscr{V}(i)} \pi_l^M \tag{4.3}$$

where $J_{protein-M}$ is the strength of interaction between proteins and monomers, and the occupancy number π_l^M quantifies the number of M monomers present at a given lattice site l.

Overall, this model is based on the same epigenome dynamics as the previous, living-chromatin-like one, but relies on a finer description of epigenomic-driven interactions by explicitly accounting for the presence of architectural proteins associated with the M-state.

4.2.3 "Living Painter"; spreading of epigenetic marks by diffusing HMEs

In both the previous models, the spreading action of HMEs was implicit: it is actually M state monomers that propagate their state to nearest-neighbor lattice sites, and HMEs were not directly modeled. Here, we develop a model that now explicitly account for the kinetics and action of HMEs within the nucleus (Fig. 4.3), thus providing better insights into the process of epigenomic spreading and maintenance.

To simplify, we consider that HMEs and architectural proteins are part of the same complex (like the Polycomb complex that may group PRC1 and PRC2) and we model the kinetics of such complex as a lattice gas of diffusing particles (as in the previous model). In addition to (i) $E_{protein-protein}$ (Eq.4.2) that accounts for complex-complex interaction mediated by architectural proteins; and (ii) $E_{protein-M}$ (Eq.4.3) that accounts for the state-dependent recruitment (reader-writer mode) of the complex at M-state monomers; the lattice gas dynamics is also driven by (iii) an Hamiltonian accounting for the sequence-dependent recruitment of HMEs:

$$E_{HME-seq} = -J_{HME-seq} \sum_{i \in \mathscr{P}} \sum_{l \in \mathscr{V}(i)} \pi_l^{\rho_s} \tag{4.4}$$

where $\pi_l^{\rho_s}$ quantifies the number of sequence specific loci at a given lattice site l. This would correspond to the painter region (ρ_s) in the effective 3D model described in Chapter 2.

Regarding the epigenomic dynamics, since we explicitly describe HMEs, the transition from U to M of a monomer i would now be constrained by the presence of the diffusing particles around it. More precisely:

$$\begin{aligned} r_{UM}^{HME}(i) &= k \sum_{l \in \mathscr{V}(i)} \sigma_l (1 + r\delta_{l,M}^{NN}) \\ r_{MU}^{HME}(i) &= k_0 \end{aligned} \tag{4.5}$$

where $\delta_{l,M}^{NN} = 1$, if lattice site l is nearest neighbor of the lattice site occupied by a M-state monomer. Note that the differential binding of the complex to the sequence and chromatin state is modulated by the strength of interaction $J_{HME-seq}$ and $J_{protein-M}$, which would impact the residence time of HMEs on the chromatin. But, the spreading dynamics of the enzymes is assumed to be similar irrespective of whether they are freely-diffusing, sequence-bound or state-bound ($\varepsilon = 1$).

Overall, this model couples an explicit description of architectural proteins (driving genome folding) with a finer depiction of HME (driving the spreading). It can be viewed as a generalization (including the feedback with the 3D organization) of the enzyme-limitation model developed in Chapter 2 (Sec.2.2.5, 2.3.7).

4.3 Preliminary analysis of the models

In this section, we illustrate the behavior of above mentioned models using simple examples.

4.3.1 Simulations

Simulations of the three models are performed using a Kinetic Monte-Carlo scheme. We consider a polymer of size $N = 2032$ with $n = 1$kbp and $b = 20$nm ($l_k = 100nm$), corresponding to a $L \approx 2$ Mbp region ($L \approx L_\alpha$ to conserve polymeric properties while minimizing the computation burden). The size of the lattice $S = 21$ is imposed such that the volumic fraction is fixed to $\Phi = 0.055$ and periodic boundary conditions are used. Note that, since the spatial resolution of our chromatin fiber model is $n = 1$ kbp (~ 5 nucleosomes), M and U states actually represent a coarse-grained, binary description of the local epigenomic state.

As in Chapter 3 (Sec.3.2.2), the polymer dynamics follows the same Monte-Carlo scheme with local trial moves driven by: $E_{tot} = E_{bend} + E_{epi}$ for the first model, $E_{tot} = E_{bend} + E_{protein-M}$ for the second and $E_{tot} = E_{bend} + E_{protein-M} + E_{HME-seq}$ for the third one.

For the second and third models, the $N_{protein}$ proteins or complexes are initially uniformly distributed on the lattice in random arrangement and evolve through a standard lattice-gas scheme. In each trial move, a protein is selected at random and displaced to a random neighbouring site with an acceptance criterion (Metropolis, Eq.3.4) associated with the Hamiltonian $E_{prot} = E_{protein-protein} + E_{protein-M}$ for the second model and $E_{prot} = E_{protein-protein} + E_{protein-M} + E_{HME-seq}$ for the third one.

The dynamics of the epigenomic state is driven by the two rates r_{UM} and r_{MU}. Within the kinetic Monte-Carlo framework, an "epigenomic" trial move consists in randomly choosing one monomer i, and then in trying to change its epigenomic state with a probability $p = r_{UM}(i)\,dt$ if its current state is U or $p = r_{MU}(i)\,dt$ if it is M. dt ($\ll 1/k, 1/k_0$) represents the infinitesimal duration of one Monte-Carlo time step (MCS).

In the first model, one MCS is then composed by N polymer trial moves and N epigenomic trial moves. In the second and third models, one MCS is in addition completed by N_{trial} protein trial moves. N_{trial} can be adjusted to account precisely for the diffusion constant of the diffusing proteins, in comparison to the diffusion of the polymer beads. The time mapping between simulation time (MCS) and real time can be done by imposing that the MSD of polymer loci $g_1(\delta t) \approx 0.15\mu m^2$ for a real time of ~ 100 seconds as typically observed experimentally [89, 221, 250]. In our case (Fig.3.10A), it leads to 1 MCS= $\delta t \approx 80\mu sec$. To simplify, we fix $N_{trial} = N_{protein}$. This leads to a diffusion constant of proteins $\sim b^2/\delta t \approx 3\mu m^2/sec$, a reasonable range of diffusion constants for freely diffusing particles in the nucleus [251].

For a given set of parameters, we first let the system relax for 10^5 MCS in the absence of any interaction and spreading mechanism, and then monitor the evolution of the polymer conformation and spreading of chromatin state during 10^6 MCS. In each condition, at

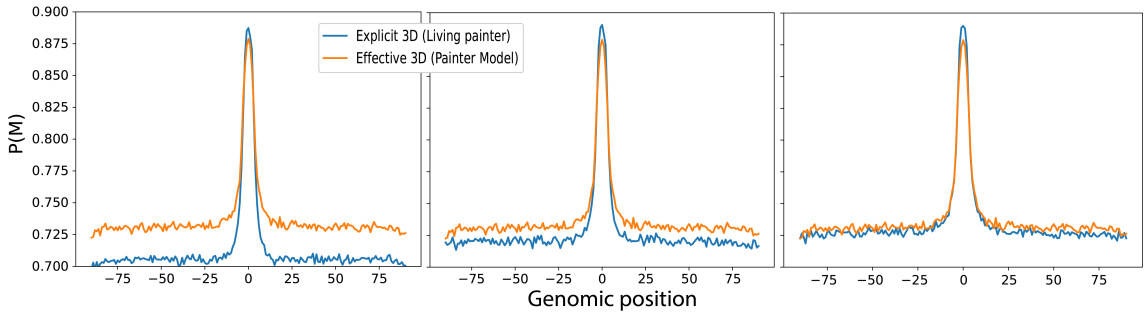

Fig. 4.4 Virtual Chip-seq profiles of the M-state. Probability to be modified $P(M)$ for a toy model (one painter area of 5 monomers in the middle of the chain, $k_{int} = 0$) computed using the simple Painter model (orange, Chapter 2) and the living-chromatin like model (blue) for different values of k with $k/k_0 = 1$ ($J = 0$, $r = 0$, $\Delta = 1$ and $\varepsilon = 1$): $k\delta t = 0.1$ (left), $= 0.01$ (center) and $= 0.001$ (right). To make predictions in the simple painter model, we use the contact matrix $P(i, j)$ computed from the simulations of a homopolymer.

least 60 different trajectories were simulated. Numerical simulations were performed using the custom made code developed in the Jost group which can be downloaded at https://github.com/physical-biology-of-chromatin/LatticePoly/tree/Painter.

4.3.2 Living-chromatin-like model

Polymer dynamics vs epigenomic dynamics

As described in Sec. 4.2, the first model (living-chromatin-like model with implicit HMEs and architectural proteins) is a generalization of the painter model described in Chapter 2 where we account explicitly for the polymer dynamics. Actually, the painter model correspond to a limiting case where the epigenomic state does not impact the 3D chromosome organization ($J = 0$), where the polymer dynamics is fast compared to the epigenomic kinetics ($k\delta t \ll 1$), such that the spreading communication between two monomers can be approximated by $P(i, j)$ the probability of contact between them. In Fig. 4.4 (right), we indeed verify that, in this limit, both models predict the same profiles.

Interestingly, when epigenetic rates are faster (left and middle in Fig.4.4), spreading is less efficient and $P(M)$ is lower. Indeed, in these cases, communication between monomers is less efficient leading to an effective reduction of ε, the factor that accounts for putative modulation of off-site spreading. This indicates that the ratio of timescales between epigenetic fluctuations and polymer dynamics play a crucial role in determining the steady-state profile of modifications, i.e, the spread of histone modifications, and has to be adjusted carefully according to the system under consideration. In our case, as $k \sim k_0 \sim min^{-1} - h^{-1}$,

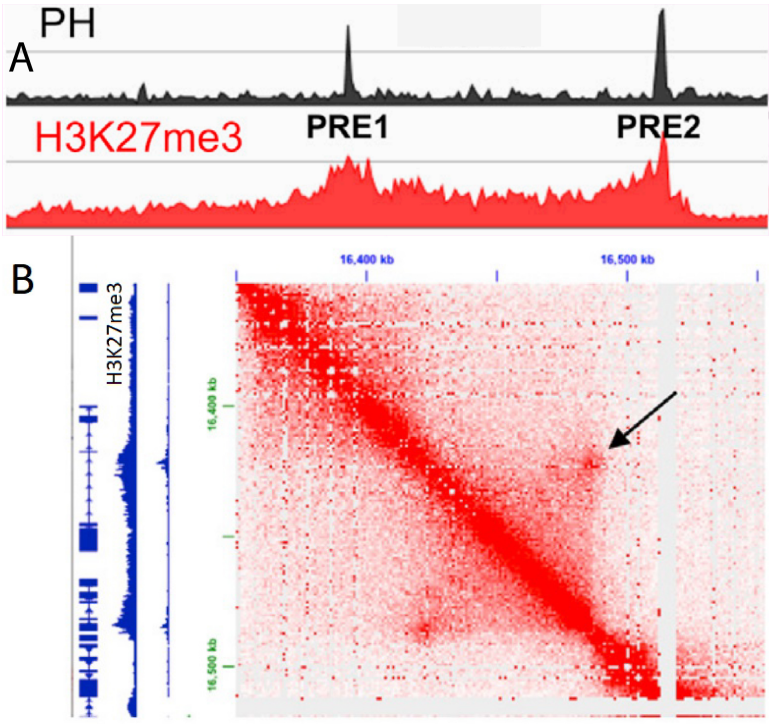

Fig. 4.5 (A) Experimental Chip-seq profiles of PH (PRC1 component) and H3K27me3 around the Dac-locus. Recruitment sites of PRC complexes are called PREs. (B) The Hi-C map of the same region showing the contact between the two PREs. On the left, the H3K27me3 profile (scaled to the Hi-C map) is also shown. *Adapted from [198].*

we would remain in the fast polymer dynamics regime in particular for monomers close to the painter zone where the dynamics of contact with it remains fast compared to $1/k$.

Role of the feedback loop

In the following sections, to contextualize our framework, we consider that the M-state is H3K27me3, that the architectural proteins correspond to PRC1 and that the HMEs are PRC2, both complexes being central in the regulation of Polycomb target genes [15]. We consider the spreading of H3K27me3 around two painter zones separated by 80 kbp, hence mimicking the situation around the Dacshung locus in Drosophila, a small Polycomb-repressed region important during fly embryogenesis [252, 253]. The two painter areas corresponds to PRE elements that are specific motifs targeted by PRC1/PRC2 complexes [198] which creates a 100 kbp domain marked by H3K27me3 (Fig. 4.5).

Using the Living-Chromatin-like model, we first simulate the systems for different value of the reader-writer state-recruitment term Δ in absence of the feedback of the epigenomic on genome folding ($J = 0$). In this case, very similar to the reader-writer mode of the

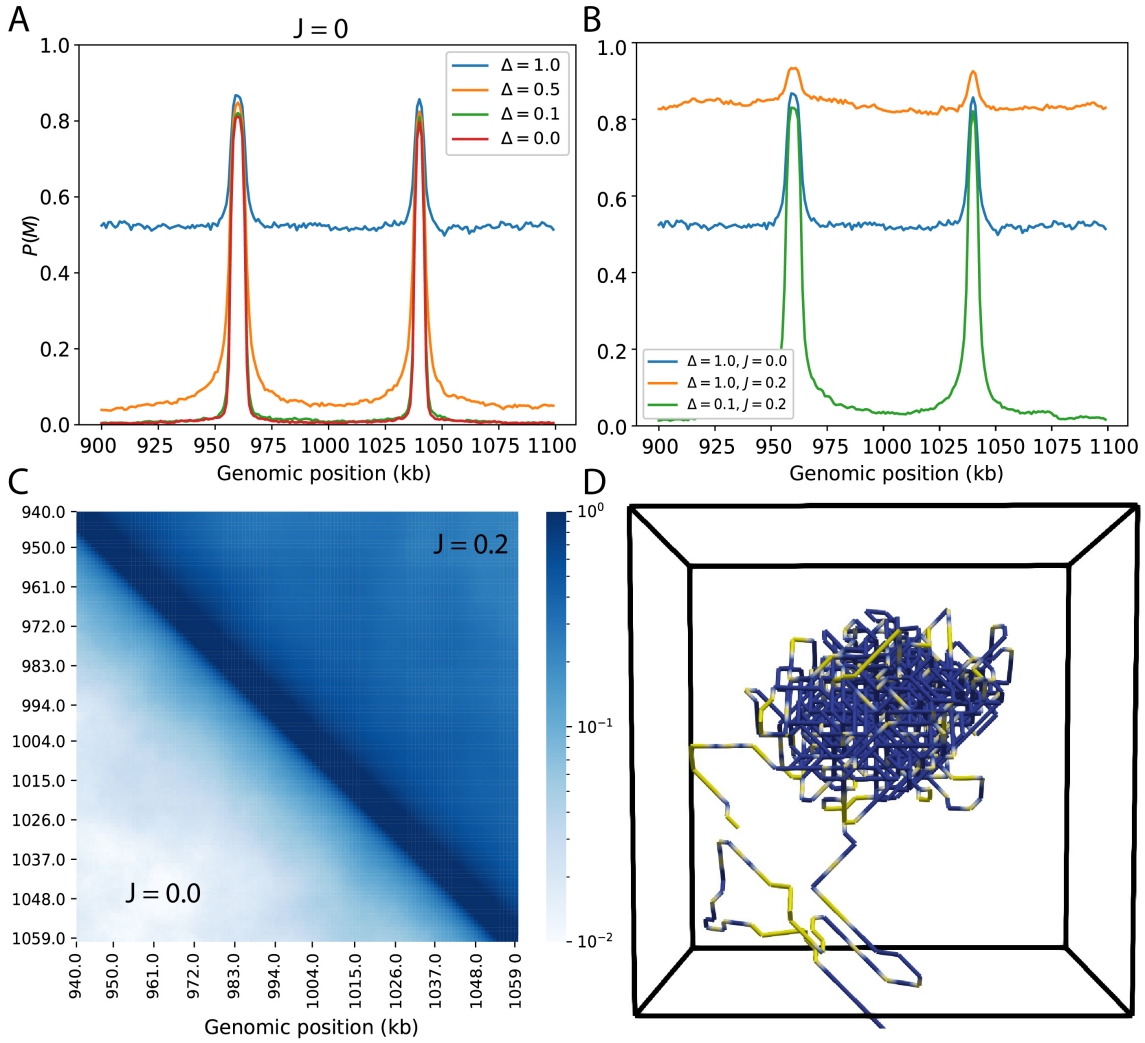

Fig. 4.6 Living-Chromatin-like model. (A) Virtual Chip-seq profiles of the M-state around the dac-like region ($950-1050$ kbp) when there is no feedback on 3D $J=0$. The painter model parameters are $k=k_0=0.001$, $\varepsilon=1$ and Δ values are indicated in the legend. (B) Same as (A) but when the feedback is introduced ($J=0.2$). (C) The contact probability matrix (virtual Hi-C) around the dac-like region when there is no feedback (lower, $\Delta=1$) and the normalized observed/expected contact probability for the case when $J=0.2$, $\Delta=1$. The long-range contacts are amplified by factor of 35 (between the painter regions). (D) Snapshot of a chromatin configuration for $\Delta=1$ and $J=0.2$. Modified chromatin states are shown in blue and unmodified in yellow.

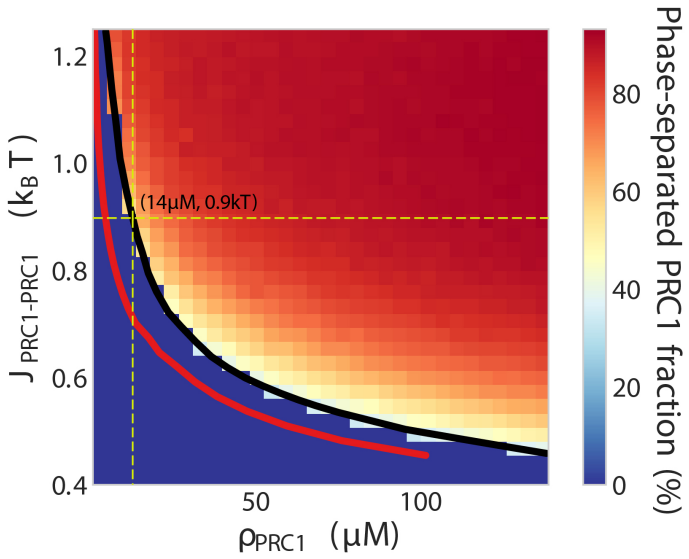

Fig. 4.7 Phase diagram of the lattice gas model in absence of chromatin as a function of the PRC1 density and self-attraction strength. Spontaneous phase separation occurs above the spinodal line (black line). The binodal line (red line) shows the concentration in the dilute phase. The yellow line denotes the concentration ($\rho^{crit}_{PRC1} \approx 14\mu M$ and the corresponding $J_{PRC1-PRC1} = 0.9$ kT) above which phase separation is observed *in vitro*.

painter model (Chapter 2), increasing Δ leads to significant increase in the probability of M-state, even further away from the painter region. Above a critical Δ, the chromatin state modifications spread in an unconfined manner ($\Delta = 1.0$, blue curve in Fig. 4.6A). Next, we introduce the 3D feedback on chromatin state spreading by having self-interacting modified state ($J > 0$). To illustrate this, we choose $J = 0.2kT$, a value that may push an homogeneously modified system in or close to a globule state (Fig. 3.16). For the two investigated cases ($\Delta = 0.1$ and $\Delta = 1$, Fig. 4.6B), we observe an overall increase in $P(M)$ coupled to a higher compaction of the chain (Fig. 4.6C,D). For example, for $\Delta = 1$, we observe $\overline{P(M)} = 0.84$ (for $J = 0$, $\overline{P(M)} = 0.57$) which drives the domain to a highly compact state, long-range contacts being almost 35 times enhanced. All this suggests, consistently with the previous analyses of living-chromatin-like models [66–68], that the 3D feedback loop greatly enhances the capacity of spreading by creating a compact environment where 3D contacts are more numerous.

It is to be noted that the compact 3D organization does not constrain the spread of modified state and leads, as observed in the reader-writer mode of the painter model, to unconfined memory.

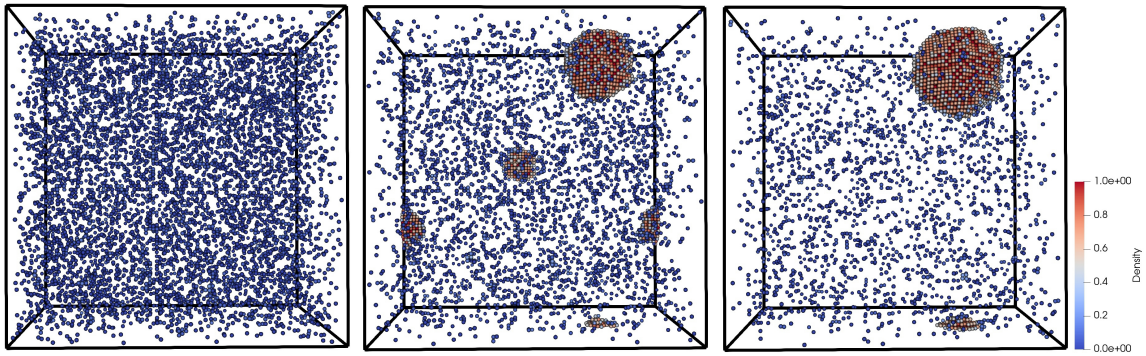

Fig. 4.8 Snapshots from simulations showing the typical time-evolution of the system, starting from a well mixed single phase (left) towards the coexistence of two phases (right): one dense liquid phase (reddish) and dilute liquid phase (blue). The self-attraction strength of PRC1 proteins is set to 0.9kT and the PRC1 concentration to $\rho_{PRC1} \approx 14\mu$M.

4.3.3 Accounting explicitly for architectural proteins

In this section, we investigate the behavior of the second model introduced in Sec.4.2 where the architectural proteins are now explicitly accounted for (Sec.4.2.2). We are using the same dac-like system with two painter areas introduced before.

Liquid-liquid phase separation of self-interacting proteins

First, we investigate the behavior of the architectural proteins (PRC1) alone in absence of the polymer chromatin by varying the concentration of PCR1 (ρ_{PRC1}) and the strength of self-attraction between PRC1 ($J_{protein-protein}$), in order to characterize the phase separation process of PRC1. The concentration of proteins ρ_{PRC1} is given in molar concentration by $\rho_{PRC1} = (N_{PRC1}/N)(1/N_A v_{site})$, where N_A is Avogadro number and $v_{site} = b^3/\sqrt{2}$, the effective volume of each lattice site. We use the lattice-gas model to calibrate the phase separation regime for PRC1 proteins, similarly to the work of Tortora *et al.* [245] for HP1 proteins. Initially, the system starts from a well mixed configuration (Fig. 4.8A), and, for a given ρ_{PRC1}, the formation of droplets is observed above a critical self-attraction strength (the spinodal concentration, black line Fig. 4.7). The simulated behavior is fully consistent with standard single-component liquid-liquid phase separation. For example, Fig. 4.8 shows the time progression of the process, starting from a well mixed initial states that spontaneously phase separates, leading, at steady state, to a stable dense liquid phase (reddish) and a dilute liquid/gas phase (blue). *In vitro* studies on PRC1 phase separation [246] suggest a critical concentration $\rho^{crit}_{PRC1} \approx 14\mu M$ above which phase separation occurs at physiological salt concentration. Based on the phase diagram (Fig. 4.7), we thus infer that the minimum

self-attraction strength required to observe droplet formation at this concentration is ≈ 0.9kT. In the next, we fixed $J_{protein-protein}$ to this value.

Coupling with polymer and epigenomic dynamics

Now, we introduce the 3D chromatin and epigenetic dynamics. To perform a preliminary investigation, we fix the heterotypic interaction strength between M-state monomer and PRC1 ($J_{protein-M}$) to half the self-attraction strength ($J_{protein-M} \approx J_{protein-protein}/2$) and the concentration of PRC1 to $\rho_{PRC1} \approx 14\mu$M. For a weak reader-writer recruitment ($\Delta = 0.1$), we find that these interaction parameters, give rise to a very similar 1D M-state profile as seen in living chromatin-like model (Fig. 4.9A). However, interestingly, the 3D organization of the region is quite different (Fig. 4.9C) driven by the self-attracting proteins binding to the two PRE regions (Fig. 4.9B) and the formation of a phase-separated droplet (Fig. 4.9D) of PRC1 that specifically bridges the two PRE regions.
When $\Delta = 1$ (strong reader-writer), both models lead to the same $P(M)$ profiles (Fig. 4.10A) and very similar 3D organization (Fig. 4.10C) as the number of architectural protein simulated is not limiting and still allows to create a droplet large-enough to encompass the entire dac-region and almost all the surrounding polymer (Fig. 4.10D). Note that, like in the living-like chromatin state, we do not observe confined memory in this example.

4.3.4 "Living Painter" model: spreading by diffusing HMEs

In this section, we explore the spreading of M-state predicted by the "Living Painter" description (Sec. 4.2.3) where the diffusion and binding of HME/architectural proteins are explicitly modeled.

In the absence of phase separation

First, we investigate the situation where diffusing molecules do not self-interact ($J_{protein-protein}$=0). Note that, even in this case, such molecules may impact the 3D polymer organization as they can bind to multiple monomers via the heterotypic interactions $J_{HME-seq}$ and $J_{protein-M}$.

Given the complexity of the parameter space (ρ_{HME}, $J_{HME-seq}$, $J_{protein-M}$, etc.), we make some assumptions to help with the preliminary exploration of the model. We fix $|J_{HME-seq}| = 2kT > |J_{HME-M}|$ to account for the strong binding of PRC2 at PREs (Fig. 4.5) and $\rho_{HME} = 7\mu$M. In the absence of reader-writer term ($J_{HME-M} = 0$), M-states and HMEs localize around the two PREs (Fig. 4.11A,B) and $P(M)$ is slightly enhanced around the painter regions when $r = 10$ (boosted painter). In the presence of the reader-writer term ($J_{HME-M} = 1kT$), we capture the standard behavior with extended heavy tails and unconfined spreading. In the case of boosted and simple painter-mode, the small but non-

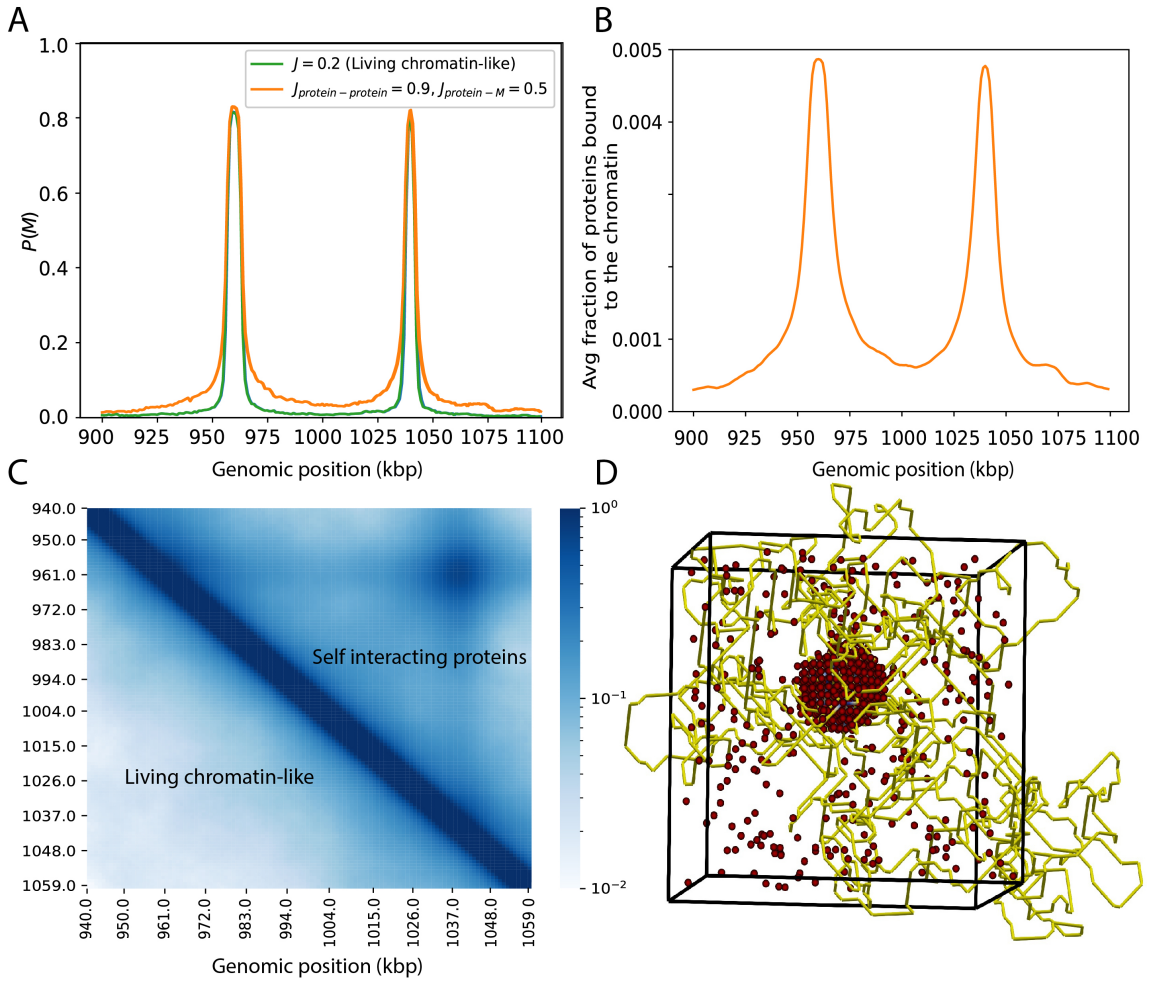

Fig. 4.9 Living-Chromatin model with explicit architectural proteins: low Δ value. (A) Virtual Chip-seq profiles of the M-state in the dac-like region ($950-1050$ kbp) in the living chromatin-like without (green, copolymer model $J = 0.2$) and with self-interacting protein (orange, $J_{protein-protein} = 0.9$, $J_{protein-M} = 0.5$, $\rho_{PRC1} \approx 14\mu$M). The painter model parameters are $k/k_0 = 1$ with $k\delta t = 0.001$, $\varepsilon = 1$ and $\Delta = 0.1$.(B) Virtual Chip-seq of PRC1 bound to the chromatin as a function of genomic position (same parameters as in panel A). (C) The contact probability matrix (virtual Hi-C) for the two cases shown in (A). (D) Snapshot of system configuration corresponding to the self-interacting protein model with same parameters as in (A). Modified chromatin states are shown in blue (not visible since the region is within the droplet) and unmodified in yellow.

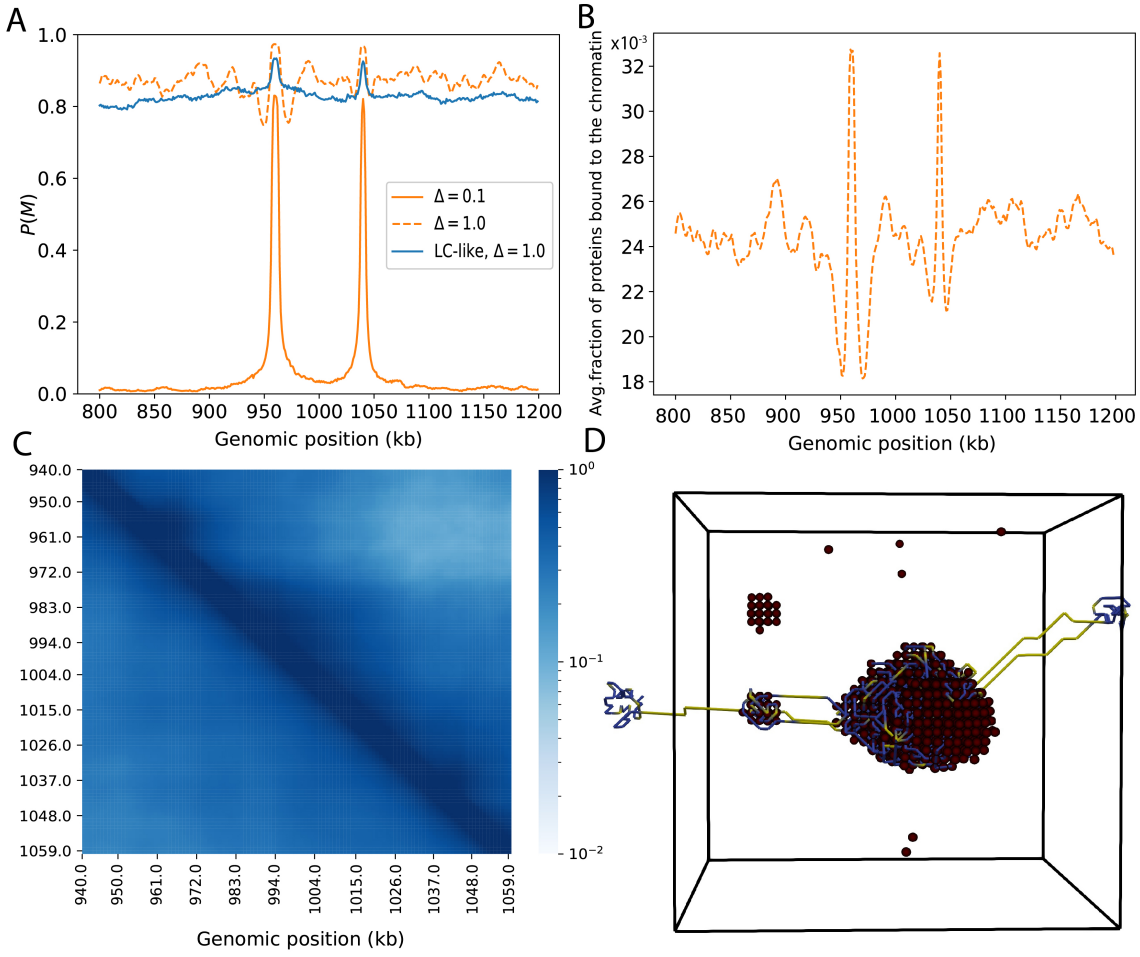

Fig. 4.10 Living-Chromatin model with explicit architectural proteins: high Δ value. (A) Virtual Chip-seq profiles of the M-state in the dac-like region (950 − 1050 kbp) in the living chromatin-like with self-interacting protein (orange, $J_{protein-protein} = 0.9$, $J_{protein-M} = 0.5$, $\rho_{PRC1} \approx 14\mu$M) for $k/k_0 = 1$ with $k\delta t = 0.001$, $\varepsilon = 1$, $\Delta = 0.1$ (solid) and $\Delta = 1$ (Dashed). For $\Delta = 1$, comparison with LC-like copolymer model (blue, $J = 0.2$). (B) Virtual Chip-seq of PRC1 bound to the chromatin as a function of genomic position. (C) Virtual Hi-C (top) With explicit architectural proteins, (bottom) living chromatin-like model (D) Snapshot of system configuration corresponding to $\Delta = 1$ with explicit architectural proteins as in (A).

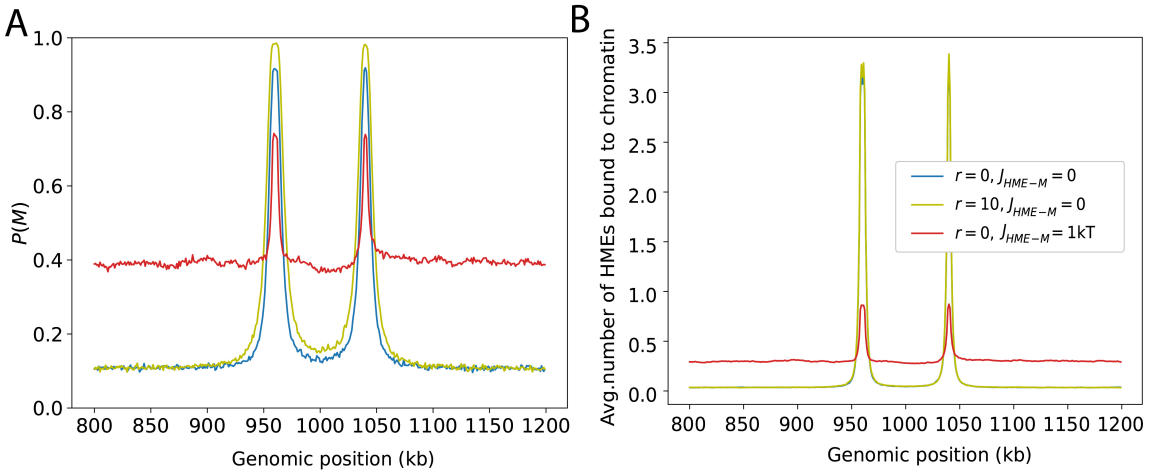

Fig. 4.11 Living Painter model: different modes of spreading in absence of self-attraction. (A) Virtual Chip-seq profiles of the M-state around the dac-like region ($950-1050$ kbp) for different spreading modes (i) painter-mode (blue) (ii) boosted-painter (yellow) and (iii) reader-writer mode (red). (B) Average number of chromatin bound HMEs in each case. Parameters used are $k/k_0 = 1$ with $k\delta t = 0.001$, $J_{HME-seq} = 2$kT, $\rho_{HME} = 7\,\mu$M, $J_{protein-protein} = 0$.

zero modifications ($P(M) \approx 0.1$) further away from the painter region is probably due to the random encounter of diffusing HMEs with the distant sites, like the background level which may be seen in experimental Chip-seq profiles (Fig. 4.5). Indeed, the profile of bound HMEs, shown in Fig. 4.11B, exhibits a non-zero value away from the painter areas even in the absence of state-specific recruitment.

Enzyme limitation

As we have shown in Chapter 2 (Sec. 2.3.7), the key ingredients that might lead to confined chromatin state memory are enzyme limitation and 3D compaction of the domains. Here, still assuming $J_{protein-protein} = 0$, we systematically reduce the concentration of HMEs, and observe a coherent transition from unconfined spreading of M-state to a regime where M-states are confined around the Dac-locus (Fig. 4.12A) but with a shape of $P(M)$ profile very different from the simple painter model. The Chip-seq of HMEs also follow the trend, as the number of enzymes become limited, HMEs preferentially bind at the PRE region compared to the rest of the locus, and the number of HME bound at the PRE is even higher than when HMEs are present in excess (Fig. 4.12B).

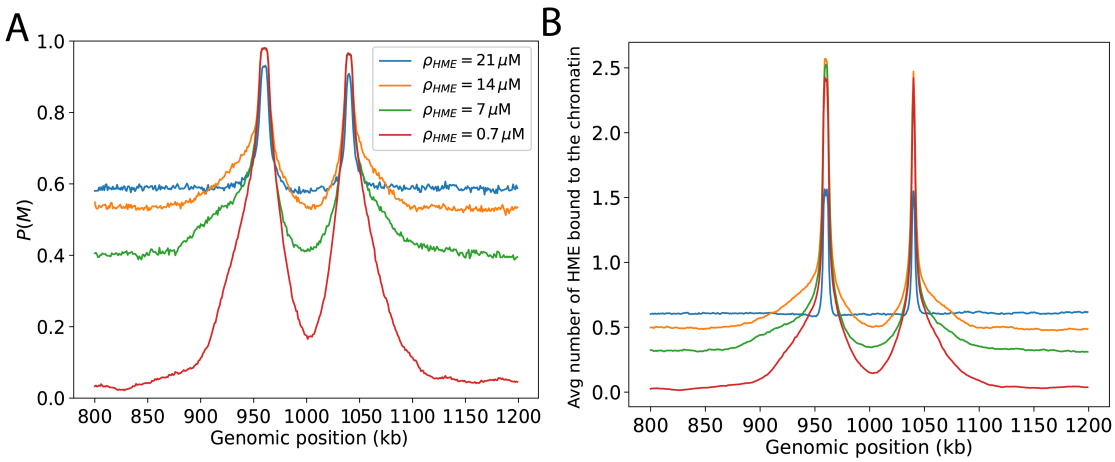

Fig. 4.12 Living Painter model: effects of enzyme limitation in absence of self-attraction. (A) Virtual Chip-seq profiles of the M-state in the dac-like region ($950-1050$ kbp) for different HME concentrations ρ_{HME}. (B) Corresponding amount of HMEs bound to chromatin for each case. Parameters used are $k/k_0 = 1$ with $k\delta t = 0.001$, $r = 10$, $J_{HME-seq} = 2\,\mathrm{kT}$, $J_{HME-M} = 1\,\mathrm{kT}$, $J_{protein-protein} = 0$.

Nucleation v/s phase separation

An interesting aspect of the "Living painter" model would be to address epigenetic memory in all its mechanistic detail. In this section, we investigate a parameter regime where the model would favour the formation of large phase separated droplet (excess of HME/architectural proteins $\rho_{HME} = 21\mu$M) but with the possibility of having a (meta)stable small multi-droplet system where diffusing HMEs/proteins are stably bound to chromatin and form multiple foci.

The parameter driving this property is the ratio of heterotypic interaction (diffusing particle and chromatin) to homotypic interaction (self-attraction of diffusing particles). In a regime where the heterotypic interaction is much less than the homotypic interaction, we are likely to see a single phase separated droplet (Fig. 4.13). In the other extreme, having strong heterotypic interactions would favour nucleation of these particles on the chromatin (as opposed to self-association) leading to the formation of several smaller droplet-like structures along the chain at different positions where some M-states have been nucleated. Average Chip-seq profiles of modifications or proteins might not be informative about these processes (Fig. 4.13A,B) since the nucleation of droplets being stochastic (either from the painter areas or randomly by diffusing HMEs), their positions vary from trajectory to trajectory. This holds true for Hi-C of the region as well (Fig. 4.13C). Rather, single-cell experimental information would be quite different between the two cases.

Advances in super-resolution microscopy has increased the resolution of nuclear imaging and thus can be used to investigate in detail the mechanistic scenario of Polycomb foci

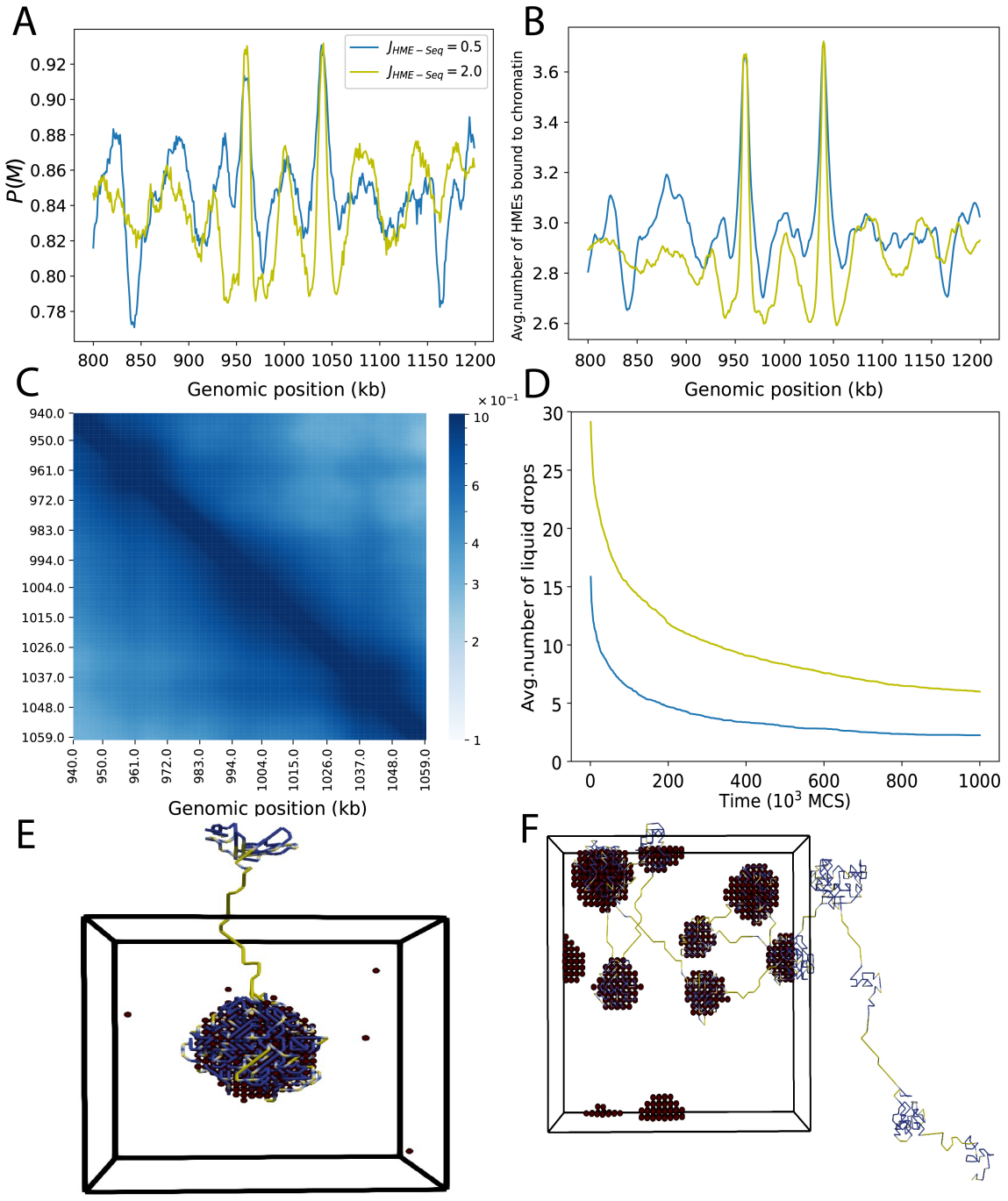

Fig. 4.13 Living painter model: Nucleation vs Self-association in the phase-separated regime. (A) Virtual Chip-seq profiles of the M-state in the dac-like region (950 – 1050 kbp) for two different HME-chromatin interaction strength at high HME concentration $\rho_{HME} = 21\mu$M. (Blue line) Homotypic interaction > heterotypic interaction: $J_{HME-seq} = J_{HME-M} = 0.5$ kT. (Yellow line) Homotypic interaction < heterotypic interaction: $J_{HME-seq} = 2$kT, $J_{HME-M} = 1$ kT. In both the cases homotypic interaction is set to 0.9 kT. (B) Average number of HMEs bound to chromatin for both the case in (A). (C) Virtual Hi-C map: (top) $J_{HME-seq} = J_{HME-M} = 0.5$ kT, (bottom) $J_{HME-seq} = 2$kT, $J_{HME-M} = 1$ kT. (D) Average number of liquid drops formed as a function of time. (E) A representative snapshot of configuration for $J_{HME-seq} = J_{HME-M} = 0.5$ kT. (F) A representative snapshot of configuration for $J_{HME-seq} = 2$kT, $J_{HME-M} = 1$ kT.

formation. For example, we monitor the number of droplets formed as a function of time, a quantity that may be available experimentally: a scenario leading to a large phase-separated droplet would exhibit an asymptote converging rapidly to 1 (Blue in Fig. 4.13D) contrary to a multi-droplet scenario leading to a high number of meta stable droplets (Yellow in Fig. 4.13D). Note that, as shown in [245], the dynamics of droplets fusion become very slow in a heterotypic system (proteins in interaction with a polymer) as it is coupled to the polymer dynamics (which is slow) and therefore may explain why the multi-droplet system is (meta)stable over a very long time period.

4.4 Discussion and conclusion

In this chapter, we have discussed the development of the "Living painter" model and presented the ingredients of the model, showing the progressive transition from the original painter model to a more complex framework accounting explicitly for HMEs and architectural proteins. The chapter doesn't present a complete analysis of the models discussed, rather it reflects the scope and extension of this thesis. Nevertheless, our preliminary analysis of the model leads to several interesting observations.

In the first part, we recapitulate the properties of living chromatin-like models by integrating the painter model discussed in chapter 2 with the 3D polymer dynamics detailed in chapter 3. It is to be noted that in chapter 2, the description is at a finer scale (nucleosomal), but here the resolution is at 1kbp hence the chromatin states are also an effective coarse grained representation. We show how having compact 3D domains enhance the spreading of chromatin state and how 3D compaction alone might not lead to confined chromatin state domains. In the regime of excess enzymes, it leads to unconfined spreading of M-state (Fig. 4.6). We have discussed the phase separation properties of diffusing proteins and contextualized it to PRC1 (Fig. 4.7) using *in vitro* experimental data [246]. Accounting explicitly for architectural proteins further improves the virtual Hi-C, better capturing the 3D organization of the dac-locus (the experimentally observed contact between the two PREs is better captured, Fig. 4.9C).

As a next step, we integrate within the lattice-gas model, the explicit modelling of HMEs to develop the living painter model. Our preliminary analysis of the model, characterizes the process of spreading of chromatin state (Fig. 4.11), showing how the process of spreading can be dissected into sequence-dependent and chromatin state-dependent processes like in the painter model. We explicitly show how limited enzyme regime would lead to confined chromatin state spreading in the case of a small compact region (red line in Fig. 4.12). The

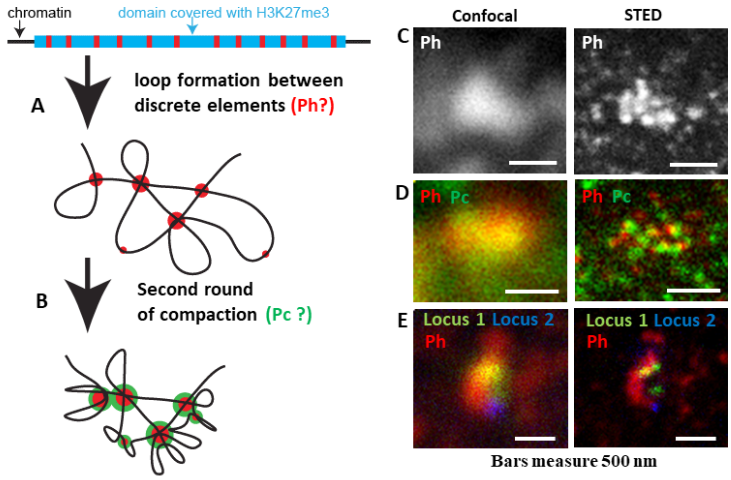

Fig. 4.14 Preliminary results from the Giacomo Cavalli lab (courtesy of Thierry Cheutin, unpublished results). (A) Hox genes nano-compartments assembly driven by Ph/Pc Scaffolding Proteins. 3D Assembly of Polycomb-repressed genes might rely on the clustering of PREs (red) by Ph binding/bridging. (B) Followed by the clustering of H3K27me3 chromatin (cyan) by Pc binding/bridging. (C) Images of Ph. (D) Images of Pc. (E) Images of PREs sites. Comparing the nuclear distribution of these elements at conventional (confocal) and high resolution (STED) microscopy.

precise conditions (i.e., what exactly is the contribution of 3D) and the critical point at which spreading would start to 'flood' (unconfined spreading) are yet to be quantitatively identified.

Finally, we discuss the impact of phase separation of diffusing HMEs and the two contrasting regimes of the model: (i) where the system favors the formation of a single phase separated droplet of HMEs around a particular chromatin region (heterotypic interactions are much less than homotypic interactions) and (ii) where the system stably maintains a multi-droplet system having favored nucleation of HMEs on chromatin (homotypic interactions are much less than heterotypic interactions). For example, the recent experimental results of our collaborator Thierry Cheutin in Cavalli's lab, using super-resolution microscopy (Fig. 4.14), suggest that Ph/Pc, two components of PRC1 in drosophila, establish loops between the multiple nucleation sites (PREs) of one Hox complex by creating nuclear nanoclusters. This motivates further characterization of the nucleation mediated multi-droplets system. However in that specific biological situation, the mechanism driving such multi-droplets phase might be different from the one we investigated in Fig. 4.13. Experimental investigation on both, PRE recruitment strength and Ph/Pc hetero- and homoptypic interactions, would probably help us to provide a better understanding of the precise mechanism at work in these hox genes domains.

Moreover, it would be quite interesting to study percolation transitions (the measure we have used in chapter 2 (see Sec. 2.2.4) in the living painter framework, since actual configurations can be utilized to understand the transition. Specifically, 3D contact domains might be an observable informative about the two distinct regimes of nucleation driven droplet-like structures and phase separation driven single droplet formation (Fig. 4.13). In this chapter, we have only done a preliminary analysis which opens a lot of new avenues to explore. The model currently does not distinguish between architectural proteins and histone modifying enzymes, decoupling HMEs and structural proteins (bridgers) is one such aspect to be explored.

Chapter 5

Conclusion and Perspectives

In this thesis, we have developed a generic (applicable to diverse biological contexts) mathematical framework for epigenome regulation and maintenance, and the progression of the thesis is working towards a precise description of the molecular mechanisms involved in chromatin state spreading and maintenance. In Chapter 1, we revisit the notion of epigenetics, particularly chromatin-based epigenetic regulation of gene expression and the significance of 3D genome organization. We have provided a comprehensive overview of the biological context and theoretical approaches to model chromatin state spreading and maintenance broadly classifying them as 1D models, effective 3D and models with 1D-3D coupling ("Living Chromatin" models). This classification also represents the chronological development of the field.

The main part of the thesis (Chapter 2) focuses on introducing and characterizing an effective 3D model of epigenome regulation, the "Painter" model. We started by translating biological evidences and observations into "building blocks" of a mathematical description; sequence-specific chromatin state spreading, allosteric boost in spreading and the reader-writer mechanism, coupled with long-range spreading in 3D forms the basis of painter model. A systematic dissection of these processes provide novel insights, even suggesting some prospective experimental direction, like measurements of nucleosome state correlation (Sec. 2.2.4) might prove useful in understanding the mechanism driving the spread of chromatin state. We have addressed the impact of local 3D genome organization on chromatin state spreading by incorporating different molecular mechanisms that mediate inter-nucleosome contacts in 3D like simple nearest-neighbour contact, diffusion and loop extrusion, showing unequivocally that percolation of chromatin state is crucially dependent on the 3D feedback. Studying different molecular mechanisms (varying 3D contacts) has led to identifying enzyme limitation and 3D compaction as key ingredients for a confined, stable chromatin state in the absence of sequence-dependent recruitment. Our work also

investigates the impact of chromatin state dynamics on transcription, slow chromatin state dynamics leading to noisier gene expression. This analysis allows us to accurately interpret the experiments performed by Ragunathan *et al.* [10] and to make quantitative predictions of the kinetics of heterochromatin decay in fission yeast.

We have developed an end-to-end (from chromatin state to gene expression) modular framework, which can be easily extended to include finer details as shown in Sec. 2.4. The modular nature of the framework (being able to integrate different nucleosome-nucleosome contacts and spreading mechanisms) can be exploited to make locus dependent spreading of chromatin states. A simple mathematical framework, the painter model accurately predicts chromatin state spreading in a diverse biological contexts from yeast to human cells (Sec. 2.3.9).

The pertinence of 3D genome organization in gene regulation has been established by previous works (discussed in Chapter 1) but little is known about the 1D-3D coupling. To explore the dynamic coupling between epigenome regulation and 3D genome organization, we went beyond the effective 3D description and studied explicit 3D polymer dynamics. In Chapter 3, we refined the 3D simulation method by attempting to accurately describe the structure and dynamics of biologically relevant regions of interest (genes, TADs) which are often very small compared to the size of the chromosome. Evidently shown by analysis of experimental (and theoretical) data, not maintaining the system in correct topological regime would lead to significant variations in physical measurements of the region of interest. A quantitative investigation of what length of the chromosome should be simulated in-order to correctly account for the epigenomic context (epigenomic interaction of the domain with the environment) and topological regime of the region (like when the domain is in the full chromosome) led to a length prediction dependent on the extent of topological constraints, generally much less than the full chromosome length. We hope our analysis would facilitate the design of computationally efficient and physically accurate polymer simulations in the future.

To further develop the painter model, incorporating finer description to investigate experimental systems like Polycomb repression in Drosophila, we integrated the painter model with explicit 3D polymer dynamics. In Chapter 4, we have shown how the painter model can be scaled up to have 3D polymer dynamics and also to have diffusing painters (histone modifying enzymes) regulating epigenome dynamics, the so-called "Living Painter" model. We have discussed the progressive development of the Living Painter model in the chapter. Our preliminary analysis recapitulates the findings of painter model, that 3D compaction alone might not confine chromatin state spreading. The Living Painter framework provides a

more accurate description to investigate the aspect of enzyme limitation and its contribution to confined chromatin states. We show that a limited number of diffusing painters do facilitate confined chromatin state domains.

I would prefer to present Chapter 4 as a reflection of my current research interests. A glimpse into the possibilities that can be explored by the Living painter model is provided by the analysis contrasting the two regimes, one favoring nucleation of HMEs creating local multi-droplet system while the other favoring phase separation of such proteins. A complete phase diagram and analysis of percolation transitions is a necessary step in the characterization of the model. This promises a framework to understand fine-scale epigenomic memory, clustering of Polycomb repressive regions and a quantitative mechanistic understanding of gene regulation coupling epigenomics and 3D chromatin organization. For example, in a near future, I plan to apply such a framework in order to characterize the spatio-temporal dynamics of Polycomb regulation during fly embryogenesis in collaboration with Thierry Cheutin from Cavalli's lab.

References

[1] Cavalli, G. and Misteli, T. (2013) Functional implications of genome topology. *Nat Struct Mol Biol* **20**, 290–299.

[2] Gaydos, L. J., Wang, W. and Strome, S. (2014) Gene repression H3K27me and PRC2 transmit a memory of repression across generations and during development. *Science* **345**, 1515–1518.

[3] Allis, C., Jenuwein, T. and Reinberg, D. (2007) *Epigenetics.* Cold Spring Harbor Laboratory Press, NY.

[4] Beisel, C. and Paro, R. (2011) Silencing chromatin: comparing modes and mechanisms. *Nat Rev Genet* **12**, 123–135.

[5] Zhang, T., Cooper, S. and Brockdorff, N. (2015) The interplay of histone modifications - writers that read. *EMBO Rep* **16**, 1467–1481.

[6] Fourel, G., Magdinier, F. and Gilson, E. (2004) Insulator dynamics and the setting of chromatin domains. *Bioessays* **26**, 523–532.

[7] Fourel, G., Lebrun, E. and Gilson, E. (2002) Protosilencers as building blocks for heterochromatin. *Bioessays* **24**, 828–835.

[8] Reyes-Turcu, F. E. and Grewal, S. I. (2012) Different means, same end-heterochromatin formation by RNAi and RNAi-independent RNA processing factors in fission yeast. *Curr Opin Genet Dev* **22**, 156–163.

[9] Hathaway, N. A., Bell, O., Hodges, C., Miller, E. L., Neel, D. S. and Crabtree, G. R. (2012) Dynamics and memory of heterochromatin in living cells. *Cell* **149**, 1447–1460.

[10] Ragunathan, K., Jih, G. and Moazed, D. (2015) Epigenetics. epigenetic inheritance uncoupled from sequence-specific recruitment. *Science* **348**, 1258699.

[11] Angel, A., Song, J., Dean, C. and Howard, M. (2011) A polycomb-based switch underlying quantitative epigenetic memory. *Nature* **476**, 105–108.

[12] Kassis, J. A. and Brown, J. L. (2013) Polycomb group response elements in Drosophila and vertebrates. *Adv Genet* **81**, 83-118.

[13] Dorafshan, E., Kahn, T. G. and Schwartz, Y. B. (2017) Hierarchical recruitment of Polycomb complexes revisited. *Nucleus* **8**, 496-505.

[14] Wang, S. P., Tang, Z., Chen, C. W., Shimada, M., Koche, R. P., Wang, L. H. *et al.* (2017) A UTX-MLL4-p300 transcriptional regulatory network coordinately shapes active enhancer landscapes for eliciting transcription. *Mol Cell* **67**, 308–21.e6.

[15] Schuettengruber, B., Bourbon, H., Di Croce, L. and Cavalli, G. (2017) Genome regulation by Polycomb and Trithorax: 70 years and counting. *Cell* **171**, 34-57.

[16] Probst, A. V., Dunleavy, E. and Almouzni, G. (2009) Epigenetic inheritance during the cell cycle. *Nat Rev Mol Cell Biol***10**, 192-206.

[17] Margueron, R. *et al.* (2009) Role of the polycomb protein Eed in the propagation of repressive histone marks. *Nature* **461**, 762-767.

[18] Park, M. *et al.* (2019) Engineering Epigenetic Regulation Using Synthetic Read-Write Modules. *Cell* **176**, 227-238.

[19] Dodd, I. B., Micheelsen, M. A., Sneppen, K. and Thon, G. (2007) Theoretical analysis of epigenetic cell memory by nucleosome modification. *Cell* **129**, 813–822.

[20] Dodd, I. B. and Sneppen, K. (2011) Barriers and silencers: A theoretical toolkit for control and containment of nucleosome-based epigenetic states. *J Mol Biol* **414**, 624-637.

[21] Dayarian, A. and Sengupta, A. (2013) Titration and hysteresis in epigenetic chromatin silencing. *Phys Biol* **10**, 036005.

[22] Jost, D. (2014) Bifurcation in epigenetics: implications in development, proliferation, and diseases. *Phys Rev E Stat Nonlin Soft Matter Phys* **89**, 010701.

[23] Zerihun, M. B., Vaillant, C. and Jost, D. (2015) Effect of replication on epigenetic memory and consequences on gene transcription. *Phys Biol* **12**, 026007.

[24] Erdel, F., Muller-Ott, K. and Rippe, K. (2013) Establishing epigenetic domains via chromatin-bound histone modifiers. *Ann N Y Acad Sci* **1305**, 29–43.

[25] Erdel, F. and Greene, E.C. (2016) Generalized nucleation and looping model for epigenetic memory of histone modifications. *Proc Natl Acad Sci U S A***113**, E4180–E4189.

[26] Cortini, R., Barbi, M., Caré, B., Lavelle, C., Lesne, A., Mozziconacci, J. and Victor, J.-M. (2016) The physics of epigenetics. *Rev Mod Phys* **88**, 025002.

[27] Berry, S., Dean, C. and Howard, M. (2017) Slow chromatin dynamics allow polycomb target genes to filter fluctuations in transcription factor activity. *Cell Systems* **4**, 1–13.

[28] 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. *Phys Rev E* **101**, 042408.

[29] Holoch, D., Wassef, M., Lövkvist, C. *et al.* (2021) A cis-acting mechanism mediates transcriptional memory at Polycomb target genes in mammals. *Nat Genet* **53**, 1686–1697.

[30] Newar, K., Abdulla, A. Z, Salari, H., Fanchon, E. and Jost, D. (2022) Dynamical modeling of the H3K27 epigenetic landscape in mouse embryonic stem cells. *PLoS Comput Biol* **18**(9), e1010450.

[31] Sneppen, K. and Dodd, I. B. (2015) Cooperative stabilization of the SIR complex provides robust epigenetic memory in a model of SIR silencing in Saccharomyces cerevisiae. *Epigenetic* **10**, 293-302.

[32] Costa, S. and Dean C. (2019) Storing memories: the distinct phases of Polycomb-mediated silencing of Arabidopsis FLC. *Biochem Soc Trans.* **47**, 1187–1196.

[33] Song, J., Angel, A., Howard, M. and Dean, C. (2012) Vernalization – a cold-induced epigenetic switch *Journal of Cell Science* **125**, 3723–3731.

[34] Lövkvist, C., Mikulski, P., Reeck, S., Hartley, M., Dean, C. and Howard, M. (2021) Hybrid protein assembly-histone modification mechanism for PRC2-based epigenetic switching and memory *eLife* **10**, e66454.

[35] Margueron, R. and Reinberg, D. (2011) The Polycomb complex PRC2 and its mark in life. *Nature* **469**, 343–349.

[36] Yang, H., Berry, S., Olsson, T. S. G, Hartley, M., Howard, M. and Dean, C. (2017) Distinct phases of Polycomb silencing to hold epigenetic memory of cold in Arabidopsis. *Science* **357**, 1142–1145.

[37] Reinig, J., Ruge, F., Howard, M. and Ringrose L. (2020) A theoretical model of Polycomb/Trithorax action unites stable epigenetic memory and dynamic regulation. *Nat Commun.* **11**, 4782.

[38] Déléris, A., Berger, F. and Duharcourt, S. (2021) Role of Polycomb in the control of transposable elements. *Trends in Genetics* **37**, 882–889.

[39] Kondo, T., Ito, S. and Koseki, H. (2016) Polycomb in Transcriptional Phase Transition of Developmental Genes. *Trends in Biochemical Sciences* **41**, 9–19.

[40] Steffen, P. A. and Ringrose, L. (2014) What are memories made of ? How Polycomb and Trithorax proteins mediate epigenetic memory. *Nat Rev Mol Cell Biol.* **15**, 340–356.

[41] Micheelsen M. A., Mitarai, N., Sneppen, K. and Dodd, I. B. (2010) Theory for the stability and regulation of epigenetic landscapes. *Phys Biol.* **7** 026010.

[42] Sood, A. and Zhang, B. (2021) Quantifying the Stability of Coupled Genetic and Epigenetic Switches With Variational Methods. *Front Genet.* **11**, 636724.

[43] Sood, A. and Zhang, B. (2020) Quantifying epigenetic stability with minimum action paths. *Phys Rev E.* **101**, 062409.

[44] Sneppen, K. and Mitarai, N. (2012) Multistability with a metastable mixed state *Phys. Rev. Lett.* **109**, 100602.

[45] Sriram, K., Soliman, S. and Fages, F. (2009) Dynamics of the interlocked positive feedback loops explaining the robust epigenetic switching in Candida albicans. *Journal of Theoretical Biology* **258**, 71–88

[46] David-Rus, D., Mukhopadhyay, S., Lebowitz, J.L. and Sengupta, A. M. (2009) Inheritance of epigenetic chromatin silencing. *J. Theor. Biol.* **258**, 112–120.

[47] Sedighi, M. and Sengupta, A. M. (2007) Epigenetic chromatin silencing: bistability and front propagation. *Phys. Biol.* **4**, 246–255.

[48] Mukhopadhyay, S., Nagaraj, V. H. and Sengupta, A. M. (2010). Locus dependence in epigenetic chromatin silencing. *BioSystems* **102**, 49–54.

[49] Nagaraj, V. H., Mukhopadhyay, S., Dayarian, A. and Sengupta, A. M. (2014) Breaking an Epigenetic Chromatin Switch: Curious Features of Hysteresis in Saccharomyces cerevisiae Telomeric Silencing. *PLoS ONE* **9**: e113516.

[50] Mukhopadhyay, S. and Sengupta, A. M. (2013) The role of multiple marks in epigenetic silencing and the emergence of a stable bivalent chromatin state. *PLoS Comput Biol.* **9**, e1003121.

[51] Haerter, J. O., Lövkvist, C., Dodd, I. B and Sneppen K. (2014) Collaboration between CpG sites is needed for stable somatic inheritance of DNA methylation states. *Nucleic Acids Res.* **42**, 2235–44.

[52] van Kampen, N. G. (1981) Stochastic Processes in Physics and Chemistry NHPL Eds.

[53] Gartenberg, M.R. and Jeffrey S. Smith, J.S. (2016) The Nuts and Bolts of Transcriptionally Silent Chromatin in Saccharomyces cerevisiae *Genetics* **203**, 1563–1599.

[54] Reinig, J., Ruge, F., Howard, M. *et al.* (2020) A theoretical model of Polycomb/Trithorax action unites stable epigenetic memory and dynamic regulation. *Nat Commun* **11**, 4782.

[55] Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A. and Cavalli, G. (2012) Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* **148**, 458–472.

[56] Jost, D., Carrivain, P., Cavalli, G. and Vaillant, C. (2014) Modeling epigenome folding: formation and dynamics of topologically associated chromatin domains. *Nucleic Acids Res* **42**, 9553–9561.

[57] Wang, L.*et al.* (2019) Histone Modifications Regulate Chromatin Compartmentalization by Contributing to a Phase Separation Mechanism. *Mol Cell.* **74**(4), 646-659.

[58] Narita, T. *et al.* (2021) Enhancers are activated by p300/CBP activity-dependent PIC assembly, RNAPII recruitment, and pause release. *Mol Cell* **81**(10), 2166-2182.

[59] Sabari, B. R. (2018) Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361**(6400).

[60] Gibson, B. A. *et al.* (2019) Organization of Chromatin by Intrinsic and Regulated Phase Separation. *Cell* **179**(2), 470-484.

[61] Zhang, M. *et al.* (2022) Molecular organization of the early stages of nucleosome phase separation visualized by cryo-electron tomography. *Mol Cell* **82**(16), 3000-3014.

[62] Ho, J. W. K., Jung, Y. L., Liu, T., Alver, B. H., Lee, S., Ikegami, K., Sohn, K.-A., Minoda, A., Tolstorukov, M. Y., Appert, A., Parker, S. C. J., Gu, T., Kundaje, A., Riddle, N. C., Bishop, E., Egelhofer, T. A., Hu, S. S., Alekseyenko, A. A., Rechtsteiner, A., Asker, D., Belsky, J. A., Bowman, S. K., Chen, Q. B., Chen, R. A.-J., Day, D. S., Dong, Y., Dose, A. C., Duan, X., Epstein, C. B., Ercan, S., Feingold, E. A., Ferrari, F., Garrigues, J. M., Gehlenborg, N., Good, P. J., Haseley, P., He, D., Herrmann, M., Hoffman, M. M., Jeffers, T. E., Kharchenko, P. V., Kolasinska-Zwierz, P., Kotwaliwale,

C. V., Kumar, N., Langley, S. A., Larschan, E. N., Latorre, I., Libbrecht, M. W., Lin, X., Park, R., Pazin, M. J., Pham, H. N., Plachetka, A., Qin, B., Schwartz, Y. B., Shoresh, N., Stempor, P., Vielle, A., Wang, C., Whittle, C. M., Xue, H., Kingston, R. E., Kim, J. H., Bernstein, B. E., Dernburg, A. F., Pirrotta, V., Kuroda, M. I., Noble, W. S., Tullius, T. D., Kellis, M., MacAlpine, D. M., Strome, S., Elgin, S. C. R., Liu, X. S., Lieb, J. D., Ahringer, J., Karpen, G. H. and Park, P. J. (2014) Comparative analysis of metazoan chromatin organization. *Nature* **512**, 449–52.

[63] Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., Sanborn, A. L., Machol, I., Omer, A. D., Lander, E. S. and Aiden, E. L. (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–80.

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

[65] Haddad, N., Vaillant, C. and Jost, D. (2017) Ic-finder: inferring robustly the hierarchical organization of chromatin folding. *Nucleic Acids Res* **45**, e81.

[66] Jost, D. and Vaillant, C. (2018) Epigenomics in 3D: importance of long-range spreading and specific interactions in epigenomic maintenance. *Nucleic Acids res* **46**, 2252–2264.

[67] Michieletto, D., Orlandini, E. and Marenduzzo, D. (2016) Polymer model with epigenetic recoloring reveals a pathway for the de novo establishment and 3D organization of chromatin domains. *Phys Rev X* **6**, 041047.

[68] Katava, M., Shi, G. and Thirumalai, D. (2022) Chromatin dynamics controls epigenetic domain formation. *Biophys J* **121**(15), 2895-2905.

[69] Owen, J. A., Osmanovic, D. and Mirny, L. A. (2022) Design principles of 3D epigenetic memory systems. *bioarXiv*, 2022.09.24.509332.

[70] Sandholtz, S. H., et al. (2019) Physical modeling of the spreading of epigenetic modifications through transient DNA looping. *J Phys A: Math Theor* **52**, 434001.

[71] Filion, G.J., van Bemmel, J.G., Braunschweig, U., Talhout, W., Kind, J., Ward, L.D., Brugman, W., de Castro, I.J., Kerkhoven, R.M., Bussemaker,H.J. *et al.* (2010) Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. *Cell* **143**, 212–224.

[72] Jost, D., Vaillant, C. and Meister, P. (2017) Coupling 1D modifications and 3D nuclear organization: data, models and function. *Curr Op Cell Bio* **44**, 20–27.

[73] Roudier, F., Ahmed, I., Bérard, C., Sarazin, A., Mary-Huard, T., Cortijo, S., Bouyer, D., Caillieux, E., Duvernois-Berther, E., Al-Shikhley, L. *et al.* (2011) Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. *EMBO J.* **30**, 1928–1938.

[74] Kueng, S., Oppikofer, M. and Gasser, S. M. (2013) SIR Proteins and the Assembly of Silent Chromatin in Budding Yeast *Annu Rev Genet* **47**, 293-324.

[75] Thurtle, D. M. and Rine, J. (2014) The molecular topography of silenced chromatin in *Saccharomyces cerevisiae Genes Dev* **28**, 245-258.

[76] De, S. *et al.* (2020) Defining the Boundaries of Polycomb Domains in Drosophila *Genetics* **216**, 689-700.

[77] Kirkland, J. G., Raab, J. R. and Kamakaka, R.T. (2013) TFIIIC bound DNA elements in nuclear organization and insulation *Biochim Biophys Acta* **1829**, 418-424.

[78] Bonev, B., Mendelson, C. N., Szabo, Q., Fritsch, L., Papadopoulos, G. L., Lubling, Y., Xu, X., Lv, X., Hugnot, J. P., Tanay, A. and Cavalli, G. (2017) Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* **171**, 557-572.e24.

[79] Hou, C., Li, L., Qin, Z. S. and Corces, V. G. (2012) Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. *Mol Cell* **48**, 471-84.

[80] Charlton, S. J. *et al.* (2020) Integrity of a heterochromatic domain ensured by its boundary elements *Proc Natl Acad Sci U S A* **117**, 21504-511.

[81] Donze, D., Kamakaka, R. T. (2001) RNA polymerase III and RNA polymerase II promoter complexes are heterochromatin barriers in Saccharomyces cerevisiae. *EMBO J.* **20**, 520–31.

[82] Michieletto, D., Chiang, M., Colì, D., Papantonis, A., Orlandini, E., Cook, P. R. and Marenduzzo, D. (2017) Shaping epigenetic memory via genomic bookmarking *Nucleic Acids res* **46**, 83–93.

[83] Cheng, T. H. and Gartenberg, M. R. (2000) Yeast heterochromatin is a dynamic structure that requires silencers continuously. *Genes Dev* **14**, 452-463.

[84] Laprell, F., Finkl, K. and Muller, J. (2017) Propagation of Polycomb-repressed chromatin requires sequence-specific recruitment to DNA. *Science* **356**, 85–88.

[85] Coleman, R. and Struhl, G. (2017) Causal role for inheritance of H3K27me3 in maintaining the OFF state of a *Drosophila* HOX gene. *Science* **356**, eaai8236.

[86] Deal, R. B., Henikoff, J. G. and Henikoff, S. (2010) Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. *Science* **328**, 1161-1164.

[87] Jenuwein, T. and Allis, C.D. (2001) Translating the histone code. *Science* **293**, 1074-1080.

[88] Margueron, R. and Reinberg, D. (2010) Chromatin structure and the inheritance of epigenetic information. *Nat Rev Genet* **11**, 285-296.

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

[90] Ghosh, S. K. and Jost, D. (2018) How epigenome drives chromatin folding and dynamics, insights from efficient coarse-grained models of chromosomes. *PLoS Comput Biol* **14**, e1006159.

[91] Alabert, C., Barth, T. K., Reveròn-Gòmez, N., Sidoli, S., Schmidt, A., Jensen, O. N., Imhof, A. and Grot, A. (2015) Two distinct modes for propagation of histone PTMs across the cell cycle. *Genes Dev* **29**, 585–590.

[92] Lieberman-Aiden, E., Berkum, N. L. V., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R, Sabo, P. J, Dorschner, M. O., Sandstrom, R., Bernstein, B., Bender, M. A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L. A., Lander, E. S., Dekker, J. (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-93.

[93] Kumari, K., Prakash, R. and Padinhateeri, R. (2022) Heterogeneous interactions and polymer entropy decide organization and dynamics of chromatin domains. *bioRxiv* 2021.02.17.431616.

[94] Gabrielle, M. et al. (2022) Dynamics of CTCF- and cohesin-mediated chromatin looping revealed by live-cell imaging. *Science* **376**, 496-501.

[95] Barbieri, M., Chotalia, M., Fraser, J., Lavitas, L. M, Dostie, J., Pombo, A. *et al.* (2012) Complexity of chromatin folding is captured by the strings and binders switch model. *Proc Natl Acad Sci U S A.* **109**(40), 16173–16178.

[96] Esposito, A. *et al.* (2022) Polymer physics reveals a combinatorial code linking 3D chromatin architecture to 1D chromatin states. *Cell Rep.* (38)13, 110601.

[97] Brackley, C. A *et al.* (2013) Nonspecific bridging-induced attraction drives clustering of DNA-binding proteins and genome organization. *Proc Natl Acad Sci U S A.* **110**(38), E3605-E3611

[98] Buckle, A. *et al.* (2018) Polymer Simulations of Heteromorphic Chromatin Predict the 3D Folding of Complex Genomic Loci. *Mol Cell* **72**(4), 786-797.

[99] Ancona, M. and Brackley, C. A. (2022) Simulating the chromatin-mediated phase separation of model proteins with multiple domains. *Biophys J* **121**(13), 2600-2612.

[100] Pierro, M. D., Cheng, R. R., Lieberman-Aiden, E., Wolynes, P. G. and Onuchic, J. N. (2017) De novo prediction of human chromosome structures: Epigenetic marking patterns encode genome architecture. *Proc Natl Acad Sci U S A.* **114**(46) 12126-12131.

[101] Shi, G., Liu, L., Hyeon, C. *et al.* (2018) Interphase human chromosome exhibits out of equilibrium glassy dynamics. *Nat. Commun.* **9**, 3161.

[102] Qi, Y. and Zhang, B. (2019) Predicting three-dimensional genome organization with chromatin states. *PLoS Comput Biol* **15**(6),e1007024.

[103] Falk, M., Feodorova, Y., Naumova, N. *et al.* (2019) Heterochromatin drives compartmentalization of inverted and conventional nuclei. *Nature* **570**, 395–399.

[104] Rao, S. S. P. *et al.* (2017) Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305–320.

[105] Bates, F. S. and Fredrickson, G. H. (1990) Block copolymer thermodynamics: Theory and experiment. *Annual review of physical chemistry* **41**, 525–557.

[106] Pierro, M. D. *et al.* (2018) Anomalous diffusion, spatial coherence, and viscoelasticity from the energy landscape of human chromosomes. *Proc Natl Acad Sci U S A.* **115**(30), 7753-7758.

[107] Neubler, J. *et al.* (2018) Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proc Natl Acad Sci U S A.* **115**(29), E6697-E6706.

[108] Fujishiro, S. and Sasai, M. (2022) Generation of dynamic three-dimensional genome structure through phase separation of chromatin. *Proc Natl Acad Sci U S A.* **119**(22), e2109838119.

[109] Sati, S. *et al.* (2020) 4D Genome Rewiring during Oncogene-Induced and Replicative Senescence. *Mol Cell* **78**(3), 522-538.

[110] Chiang, M. *et al.* (2019) Polymer Modeling Predicts Chromosome Reorganization in Senescence. *Cell Rep* **28**(12), 3212-3223.

[111] Di-Stefano, M., Nützmann, H., Marti-Renom, M. A. and Jost, D. (2021) Polymer modelling unveils the roles of heterochromatin and nucleolar organizing regions in shaping 3D genome organization in Arabidopsis thaliana. *Nucl Acids Res* **49**(4), 1840–1858.

[112] Davidson, I. F. *et al.* (2019) DNA Loop Extrusion by Human Cohesin. *Science*, **366**, 1338–45.

[113] Banigan, E. J. *et al.* (2020) Chromosome organization by one-sided and two-sided loop extrusion. *eLife* **9**, e53558.

[114] Schwarzer, W., Abdennur, N., Goloborodko, A. *et al.* (2017) Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51–56.

[115] Cattoni, D.I., Cardozo Gizzi, A.M. and Georgieva, M. *et al.* (2017) Single-cell absolute contact probability detection reveals chromosomes are organized by multiple low-frequency yet specific interactions. *Nat Commun* **8**, 1753.

[116] Gibcus, J. H. *et al.* (2018) A pathway for mitotic chromosome formation. *Science* **359**, eaao6135.

[117] Alipour, E. and Marko, J. F. (2012) Self-organization of domain structures by DNA-loop-extruding enzymes. *Nucleic Acids Res.* **40**, 11202–11212.

[118] Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N. and Mirny, L. A. (2016) Formation of chromosomal domains by loop extrusion. *Cell Rep.*, **15**, 2038–2049.

[119] Ghosh, S. K. and Jost, D. (2020) Genome organization via loop extrusion, insights from polymer physics models. *Brief Funct Genomics.* **23**, 119-127.

[120] Arnould, C. *et al.* (2021) Loop extrusion as a mechanism for formation of DNA damage repair foci. *Nature* **590**, 660-665.

[121] Gillespie, D.T. (1977) Exact stochastic simulation of coupled chemical reactions. *J Phys Chem* **81**, 25.

[122] Hinrichsen, H. and Howard, M. (1999) A model for anomalous directed percolation. *Euro Phys B* **7**, 635–643.

[123] Breskin, L., Soriano, J., Moses, E. and Tlusty, T. (2006) Percolation in living neural networks. *Phys. Rev. Lett* **97**, 188102.

[124] Koern, M. *et al.* (2005) Stochasticity in gene expression: from theories to phenotypes. *Nat Rev Genet* **6**, 451-464.

[125] Robello, R. *et al.* (2011) Retrotransposon-induced heterochromatin spreading in the mouse revealed by insertional polymorphisms. *PLoS Genet* **7**, 9.

[126] Wells, J. N. and Feschotte, C. (2020) A field guide to eukaryotic transposable elements. *Annu Rev Genet* **54**, 539-561.

[127] Grunstein, M. and Gasser, S.M. (2013) Epigenetics in *Saccharomyces cerevisiae*. *Cold Spring Harb Perspect Biol* **5**, a017491.

[128] Sanulli, S. *et al.* (2015) Jarid2 methylation via the PRC2 complex regulates H3K27me3 deposition during cell differentiation. *Mol. Cell*, **57**, 769-783.

[129] Zhang, K., Mosch, K., Fischle, W. *et al.* (2008) Roles of the Clr4 methyltransferase complex in nucleation, spreading and maintenance of heterochromatin. *Nat Struct Mol Biol* **15**, 381–388.

[130] Caron, P. *et al.* (2015) Non-redundant functions of ATM and DNA-PKcs in response to DNA double-strand breaks. *Cell Rep* **13**, 1598–1609.

[131] Simon, J. A. and Kingston, R. E. (2013) Occupying chromatin: Polycomb mechanisms for getting to genomic targets, stopping transcriptional traffic, and staying put. *Mol Cell*, **49**, 808-824.

[132] Brothers, M. and Rine, J. (2022) Distinguishing between recruitment and spread of silent chromatin structures in *Saccharomyces cerevisiae*. *eLife* **11**, e75653.

[133] Hoerl, A. E. and Kennard, R. W. (1970) Ridge Regression: Applications to Nonorthogonal Problems. *Technometrics* **12**(1), 69–82.

[134] Redolfi, J., Zhan, Y., Valdes-Quezada, C., Kryzhanovska, M., Guerreiro, I., Iesmantavicius, V., Pollex, T., Grand, R. S., Mulugeta, E., Kind, J., Tiana, G., Smallwood, S. A., de Laat, W. and Giorgetti, L. (2019) DamC reveals principles of chromatin folding in vivo without crosslinking and ligation. *Nat Struct Mol Biol* **26**, 471-480.

[135] Oksuz, O. *et al.* (2018) Capturing the onset of PRC2-mediated repressive domain formation. *Mol Cell* **70**, 1149–1162.

[136] Lee, C. H. *et al.* (2018) Distinct stimulatory mechanisms regulate the catalytic activity of polycomb repressive complex 2. *Mol Cell* **70**, 435–48.e5.

[137] Bonnet, J. *et al.* (2019) Quantification of proteins and histone marks in Drosophila embryos reveals reveals stoichiometric relationships impacting chromatin regulation. *Dev Cell* **51**, 632–644.

[138] Isono, K., Endo, T. A., Ku, M., Yamada, D., Suzuki, R., Sharif, J., Ishikura, T., Toyoda, T., Bernstein, B. E. and Koseki, H. (2013) SAM domain polymerization links subnuclear clustering of PRC1 to gene silencing. *Dev Cell* **26**, 565–577.

[139] Larson, A. G., Elnatan, D., Keenen, M. M., Trnka, M. J., Johnston, J. B., Burlingame, A. L., Agard, D. A., Redding, S. and Narlikar, G. J. (2017) Liquid droplet formation by $HP1\alpha$ suggests a role for phase separation in heterochromatin. *Nature* **547**, 236–240.

[140] Boettiger, A., Bintu, B., Moffitt, J. *et al.* (2016) Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* **529**, 418–422.

[141] Szabo, Q., Jost, D., Chang, J. M., Cattoni, D. I., Papadopoulos, G. L., Bonev, B., Sexton, T., Gurgo, J., Jacquier, C., Nollmann, M., Bantignies, F. and Cavalli, G. (2018) TADs are 3D structural units of higher-order chromosome organization in Drosophila. *Sci Adv* **4**, eaar8082.

[142] Strom, A. R., Emelyanov, A. V., Mir, M., Fyodorov, D. V., Darzacq, X. and Karpen, G. H. (2017) Phase separation drives heterochromatin domain formation. *Nature* **547**,241–245.

[143] Li, K., Bronk, G., Kondev, J. and Haber, J. E (2020) Yeast ATM and ATR kinases use different mechanisms to spread histone H2A phosphorylation around DNA double-strand break. *Proc Natl Acad Sci U S A* **117**, 21354-21363.

[144] Berg, H. (1993) Random Walks in Biology. *Princeton University Press*

[145] Jeppsson, K., Kanno, T., Shirahige, K. *et. al.* (2014) The maintenance of chromosome structure: positioning and functioning of SMC complexes. *Nat Rev Mol Cell Biol* **15**, 601–614.

[146] Wutz, G. *et. al.* (2017) Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *EMBO J.* **36**(24), 3573-3599.

[147] Mirny, L. A. (2011) The fractal globule as a model of chromatin architecture in the cell. *Chromosome Research* **19**(1), 37-51.

[148] Walczak, A. M., Mugler, A. and Wiggins, C. H. (2012) Analytic methods for modeling stochastic regulatory networks. In: Liu X., Betterton M. (eds) Computational modeling of signaling networks. Methods in molecular biology (Methods and Protocols), vol 880. Humana Press, Totowa, NJ.

[149] Simpson, M. L., Cox, C. D. and Sayler, G. S. (2004) Frequency domain chemical Langevin analysis of stochasticity in gene transcriptional regulation. *Journal of Theoretical Biology* **229**, 383-394.

[150] Elgin, S. C. R. and Reuter, G. (2013) Position-effect variegaion, heterochromatin fomration and gene silencing in Drosophila. *Cold Spring Harb Perspect Biol* **5**, a017780.

[151] Montavon, T., Shukeir, N., Erikson, G. *et al.* (2021) Complete loss of H3K9 methylation dissolves mouse heterochromatin organization. *Nat Commun* **12**, 4359.

[152] Wolf, D. and Goff, S. P. (2009) Embryonic stem cells use ZFP809 to silence retroviral DNAs. *Nature* **458**, 1201–1204.

[153] Liu, Y., et al. (2020) Very fast CRISPR on demand. *Science* **368**(6496):1265–1269.

[154] Markova, E., Schultz, N. and Belyaev, I. Y. (2007) Kinetics and dose-response of residual 53BP1/gamma-H2AX foci: co-localization, relationship with DSB repair and clonogenic survival. *Int J Radiat Biol.* **83**(5), 319-29.

[155] Audergon, P. N. C. B., Catania, S., Kagansky, A., Tong, P., Shukla, M., Pidoux, A. L. and Allshire, R. C. (2015) Restricted epigenetic inheritance of H3K9 methylation. *Science* **348**, 132-135.

[156] Wakim, J. G., Sandholtz, S. H and Spakowitz, A. J (2021) Impact of chromosomal organization on epigenetic drift and domain stability revealed by physics-based simulations. *Biophys J* **120**, 4932–4943.

[157] Sandholtz, S. H, MacPherson, Q. and Spakowitz, A. J (2020) Physical modeling of the heritability and maintenance of epigenetic modifications. *Proc Natl Acad Sci U S A* **117**, 20423-20429.

[158] Ramakrishnan, N., Pillai, S. R. B and Padinhateeri, R. (2022) High fidelity epigenetic inheritance: Information theoretic model predicts threshold filling of histone modifications post replication. *PLoS Comput Biol* **18**(2), e1009861.

[159] Cutter DiPiazza, A. R., Taneja, N., Dhakshnamoorthy, J., Wheeler, D., Holla, S. and Grewal, S. (2021) Spreading and epigenetic inheritance of heterochromatin require a

critical density of histone H3 lysine 9 tri-methylation. *Proc Natl Acad Sci U S A* **118**, e2100699118.

[160] Grosselin, K., Durand, A., Marsolier, J. *et al.* (2019) High-throughput single-cell ChIP-seq identifies heterogeneity of chromatin states in breast cancer. *Nat Genet* **51**, 1060–1066.

[161] Prompsy, P., Kirchmeier, P., Marsolier, J. *et al.* (2020) Interactive analysis of single-cell epigenomic landscapes with ChromSCape. *Nat Commun* **11**, 5702.

[162] Weber, C. M., Hafner, A., Kirkland, J. G. *et al.* (2021) mSWI/SNF promotes Polycomb repression both directly and through genome-wide redistribution. *Nat Struct Mol Biol* **28**, 501–511.

[163] Gartenberg, M. R. and Smith, J. S. (2016) The nuts and bolts of transcriptionally silent chromatin in saccharomyces cerevisiae. *Genetics* **203**, 1563–1599.

[164] Meister, P. and Taddei, A. (2013) Building silent compartments at the nuclear periphery: a recurrent theme. *Curr Opin Genet Dev* **23**, 96–103.

[165] Smith, J. S., Brachmann, C. B., Pillus, L. and Boeke, J. D. (1998). Distribution of a limited Sir2 protein pool regulates the strength of yeast rDNA silencing and is modulated by Sir4p. *Genetics* **149**, 1205-1219.

[166] Michel, A. H., Kornmann, B., Dubrana, K. and Shore, D. (2005). Spontaneous rDNA copy number variation modulates Sir2 levels and epigenetic gene silencing. *Genes Dev* **19**, 1199-1210.

[167] Dowen, J. M., Fan, Z. P., Hnisz, D., Ren, G., Abraham, B. J., Zhang, L. N., Weintraub, A. S., Tong Ihn Lee, J. S. and Zhao, K. (2014) Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* **159**, 374–387.

[168] Cuadrado, A., *et al.* (2019) *Cell Rep* **12**, 3500-3510.

[169] Olarte, J., Haddad, N., Vaillant, C., and Jost, D. (2016) The folding landscape of the epigenome. *Phys Biol* **13**, 026001.

[170] Chory, E. J., Calarco, J. P., Hathaway, N. A., Bell, O., Neel, D. S. and Crabtree, G. R. (2019) Nucleosome Turnover Regulates Histone Methylation Patterns over the Genome. *Mol Cell* **73**, 61–72.

[171] Radman-Livaja, M. *et al.* (2011) Dynamics of Sir3 spreading in budding yeast. *EMBO J* **30**, 1012–1026.

[172] De, S., Mitra, A., Cheng, Y., Pfeifer, K. and Kassis, J. A. (2016). Formation of a polycomb-domain in the absence of strong polycomb response elements. *PLoS Genet* **12**, e1006200.

[173] Sneppen, K. and Ringrose, L. (2019) Theoretical analysis of Polycomb-Trithorax systems predicts that poised chromatin is bistable and not bivalent. *Nat Commun* **10**, 2133.

[174] Alabert, C., Loos, C., Voelker-Albert, M., Graziano, S., Forné, I., Reveron-Gomez, N., Schuh, L., Hasenauer, J., Marr, C., Imhof, A. and Groth, A. (2020) Domain Model Explains Propagation Dynamics and Stability of Histone H3K27 and H3K36 Methylation Landscapes. *Cell Rep* **30**, 1223-1234.e8.

[175] Xu, Y., Zhang, S., Lin, S., Guo, Y., Deng, W., Zhang, Y. and Xue, Y. (2017) WERAM: a database of writers, erasers and readers of histone acetylation and methylation in eukaryotes. *Nucleic Acids Res* **45**, D264–D270.

[176] Micheelsen, M. A. *et al.* (2010) Theory for the stability and regulation of epigenetic landscapes. *Phys Biol* **7**, 026010.

[177] Agger, K., Cloos, P. A. C, Christensen, J., Pasini, D., Rose, S., Rappsilber, J. *et al.* (2007) UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. *Nature* **449**, 731–4.

[178] Torné, J., Ray-Gallet, D., Boyarchuk, E., Garnier, M., Le Baccon, P., Coulon, A. *et al.* (2020) Two HIRA-dependent pathways mediate H3.3 de novo deposition and recycling during transcription. *Nat Struct Mol Biol* **27**, 1057–68.

[179] Salari, H., Di Stefano, M. and Jost, D. (2022) Spatial organization of chromosomes leads to heterogeneous chromatin motion and drives the liquid- or gel-like dynamical behavior of chromatin. *Genome Res* **32**(1), 28-43.

[180] Groh, S. and Schotta, G. (2017) Silencing of endogenous retroviruses by heterochromatin. *Cell Mol Life Sci* **74**, 2055–2065.

[181] Varshney, D. *et al.* (2015) Selective repression of SINE transcription by RNA polymerase III. *Mobile Genetic Elements* **5:6**, 86-91.

[182] Choi, J. *et al.* (2020) DNA methylation and histone H1 jointly repress transposable elements and aberrant intragenic transcripts *Molecular Cell* **77**, 1-14.

[183] Slotkin, R. K. and Martienssen,R. (2007) Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genetics* **8**, 272–285.

[184] Coluccio, A., Ecco, G., Duc, J. *et al.* (2018) Individual retrotransposon integrants are differentially controlled by KZFP/KAP1-dependent histone methylation, DNA methylation and TET-mediated hydroxymethylation in naïve embryonic stem cells. *Epigenetics Chromatin* **11**, 7.

[185] Rowe, H. M. *et al.* (2010) KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* **463**, 237-240.

[186] Lee, Y. C. G and Karpen, G. H. (2017) Pervasive epigenetic effects of Drosophila euchromatic transposable elements impact their evolution. *eLife* **6**, e25762.

[187] Groner, A. C. *et al.* (2010) KRAB–Zinc Finger Proteins and KAP1 can mediate long-range transcriptional repression through heterochromatin spreading. *PLoS Genet* **6**, e1000869.

[188] Abramov, Y. A. *et al.* (2016) The differences between cis- and *trans*-gene inactivation caused by heterochromatin in Drosophila. *Genetics* **202**, 93–106.

[189] Kitada, T., Kuryan, B. G., Tran, N. N., Song, C., Xue, Y., Carey, M. and Grunstein, M. (2012) Mechanism for epigenetic variegation of gene expression at yeast telomeric heterochromatin. *Genes Dev* **26**, 2443-55.

[190] Chuong, E. B., Elde, N. E. and Feschotte, C. (2017) Regulatory activities of transposable elements: from conflicts to benefits. *Nat Rev Genet* **18**, 71-86

[191] Erwin, J. A., Marchetto, M. C. and Gage, F. H. (2014) Mobile DNA elements in the generation of diversity and complexity in the brain. *Nat Reviews Neuro* **15**, 497-506.

[192] Lu, J. Y. *et al.* (2021) Homotypic clustering of L1 and B1/Alu repeats compartmentalizes the 3D genome. *Cell research* **31**, 613-630.

[193] Tubio, J. M. C. *et al.* (2014) Mobile DNA in cancer. Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. *Science* **345**, 1251343.

[194] Burns, K. H. (2017) Transposable elements in cancer. *Nat Reviews Cancer* **17**, 415-424.

[195] van den Engh, G., Sachs, R. and Trask, B. J. (1992) Estimating genomic distance from DNA sequence location in cell nuclei by a random walk model. *Science* **257**(5075), 1410.

[196] Rosa, A. and Everaers, R. (2008) Structure and dynamics of interphase chromosomes. *PLoS Comput. Biol.* **4**, e1000153.

[197] Hyeon, C. and Thirumalai, D. (2011) Capturing the essence of folding and functions of biomolecules using coarse-grained models. *Nature Communications* **2**(487).

[198] Ogiyama *et al.* (2018) Polycomb-Dependent Chromatin Looping Contributes to Gene Silencing during Drosophila Development. *Mol Cell* **71**, 73–88.

[199] Hsieh, T. S. *et al.* (2015) Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. *Cell* **162**(1), 108-119.

[200] Conte, M. *et al.* (2022) Dynamic and equilibrium properties of finite-size polymer models of chromosome folding. *Phys Rev E* **104**, 054402.

[201] Caré, B. R. *et al* (2014) Finite-size conformational transitions: A unifying concept underlying chromosome dynamics. *Commun. Theor. Phys.* **62**, 607.

[202] Caré, B. R., Emeriau, P. E., Cortini, R. and Victor, J. M. (2015) Chromatin epigenomic domain folding: size matters[J]. *AIMS Biophysics* **2**(4), 517-530.

[203] Lesage, A., Dahirel, V., Victor, JM. *et al.* Polymer coil–globule phase transition is a universal folding principle of Drosophila epigenetic domains. *Epigenetics Chromatin* **12**, 28.

[204] Vaillant, C. and Jost, D. (2019) Modeling the Functional Coupling between 3D Chromatin Organization and Epigenome. *CRC Press*.

[205] Alberts, B. *et al.* (2007) Molecular Biology of the Cell. *Garland Science, New York* fifth ed.

[206] Halverson, J. D., Smrek, J., Kremer, K. and Grosberg, A. Y. (2014) From a melt of rings to chromosome territories: the role of topological constraints in genome folding. *Rep. Prog. Phys.* **77**, 022601.

[207] Annunziato, A. (2008) DNA Packaging: Nucleosomes and Chromatin. *Nature Education* **1**(1), 26.

[208] Emanuel, M., Radja, N. H., Henriksson, A. and Schiessel, H. (2009) The physics behind the larger scale organization of DNA in eukaryotes. *Phys. Biol.* **6**, 025008.

[209] Doi, M. and Edwards, S. F. (1986) The Theory of Polymer Dynamics. *Oxford University Press, New York*.

[210] Uchida, N., Grest, G.S. and Everaers, R. (2008) Viscoelasticity and primitive path analysis of entangled polymer liquids: From F-actin to polyethylene. *J. Chem. Phys.* **128**, 044902.

[211] Jost, D., Rosa, A., Vaillant, C. and Everaers, R. (2018) A Polymer Physics View on Universal and Sequence-Specific Aspects of Chromosome Folding. *Nuclear Architecture and Dynamics, AP press.* Chapter 6, 149-169.

[212] Newman, M. E. J., Strogatz, S. H. and Watts D. J. (2001) Random graphs with arbitrary degree distributions and their applications. *Phys Rev E.* **64**, 026118.

[213] Hugouvieux, V., Axelos, M. and Kolb, M. (2009) Amphiphilic multiblock copolymers: From intramolecular pearl necklace to layered structures. *Macromolecules* **42**, 392–400.

[214] Zhang, B. and Wolynes, P. G. (2015) Topology, structures, and energy landscapes of human chromosomes. *Proc Natl Acad Sci U S A.* **112**,6062–6067.

[215] Tavares-Cadete, F., Norouzi, D., Dekker, B. *et al.* (2020) Multi-contact 3C reveals that the human genome during interphase is largely not entangled. *Nat Struct Mol Biol* **27**, 1105–1114.

[216] Earnshaw, W.C. and Laemmli, U. K. Architecture of metaphase chromosomes and chromosome scaffolds. *The Journal of Cell Biology* **96**(1), 84–93.

[217] Auhl, R., Everaers R., Grest G., S., Kremer, K. and Plimpton, S. J. (2003) Equilibration of long chain polymer melts in computer simulations. *The Journal of Chemical Physics* **119**(24), 12718–12728.

[218] Herbert, S., Brion, A., Arbona, J. *et al.* (2017) Chromatin stiffening underlies enhanced locus mobility after DNA damage in budding yeast. *EMBO J.* **36**(17), 2595-2608.

[219] Socol, M. *et al.* (2019) Rouse model with transient intramolecular contacts on a timescale of seconds recapitulates folding and fluctuation of yeast chromosomes. *Nucl Acids Res* **47**(12), 6195–6207.

[220] Bantignies, F., Roure, V., Comet, I., Leblanc, B., Schuettengruber, B., Bonnet, J. *et al.* (2011) Polycomb-dependent regulatory contacts between distant Hox loci in Drosophila. *Cell* **144**(2), 214–226.

[221] Jost, D. (2022) Polymer Modeling of 3D Epigenome Folding: Application to Drosophila. *Methods Mol Biol.* 2301, 293-305.

[222] Ing-Simmons, E., Vaid, R., Bing, X. Y. *et al.* (2021) Independence of chromatin conformation and gene regulation during Drosophila dorsoventral patterning. *Nat Genet* **53**, 487–499.

[223] Costantino, L., Hsieh, T. S., Lamothe, R., Darzacq, X. and Koshland, D. (2020) Cohesin residency determines chromatin loop patterns. *eLife* **9**, e59889.

[224] Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li., Y., Shen, Y. *et al.* (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**(7398), 376–380.

[225] Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**(7398), 381–385.

[226] Adams, M.D. *et al.* (2000) The Genome Sequence of Drosophila melanogaster. *Science* **287**, 2185-2195.

[227] Serra, F., Baù, D., Goodstadt, M., Castillo, D. Filion, G. and Marti-Renom, M.A. (2017) Automatic analysis and 3D-modelling of Hi-C data using TADbit reveals structural features of the fly chromatin colors. *PLOS Comp Bio* **13**(7), e1005665.

[228] Pütz, M. *et al* (2000) What is the entanglement length in a polymer melt?. *Europhys Lett.* **49**, 735.

[229] Grosberg, A. Y. and Kokhlov, A. R. (1994) Statistical Physics of Macromolecules. *Am. Inst. of Phys.*, Melville, NY, .

[230] Wu, C. and Wang, X. (1998) Globule-to-Coil Transition of a Single Homopolymer Chain in Solution. *Phys. Rev. Lett.* **80**, 4092.

[231] Sun, J., Zhang, Q. and Schlick, T. (2005) Electrostatic mechanism of nucleosomal array folding revealed by computer simulation. *Proc Natl Acad Sci U S A* **102**(23), 8180-8185.

[232] Davey, C. A. *et al.* (2002) Solvent Mediated Interactions in the Structure of the Nucleosome Core Particle at 1.9 Å Resolution. **319**(5), 1097-1113.

[233] Chen, H., Levo, M., Barinov, L. *et al.* (2018) Dynamic interplay between enhancer–promoter topology and gene activity. *Nat Genet* **50**, 1296–1303.

[234] Meaburn, K. and Misteli, T. (2007) Chromosome territories. *Nature* **445**, 379–381.

[235] Ulianov, S. V., Doronin, S. A., Khrameeva, E. E. *et al.* Nuclear lamina integrity is required for proper spatial organization of chromatin in Drosophila. *Nat Commun* **10**, 1176.

[236] Neguembor, M. V. *et al.* (2021) Transcription-mediated supercoiling regulates genome folding and loop formation. *Mol Cell* **81**(15), 3065-3081.

[237] Uusküla-Reimand, L., Hou, H., Samavarchi-Tehrani, P. *et al.* (2016) Topoisomerase II beta interacts with cohesin and CTCF at topological domain borders. *Genome Biol* **17**, 182.

[238] Michieletto, D. *et al.* (2019) Nonequilibrium Theory of Epigenomic Microphase Separation in the Cell Nucleus. *Phys Rev Lett.* **123**, 228101.

[239] Colì, D., Orlandini, E., Michieletto, D. and Marenduzzo, D. (2019). Magnetic polymer models for epigenetics-driven chromosome folding. *Phys Rev E*, **100**(5), 052410.

[240] Orlandini, E., Marenduzzo, D. and Michieletto, D. (2019) Synergy of topoisomerase and structural-maintenance-of-chromosomes proteins creates a universal pathway to simplify genome topology. *Proc Natl Acad Sci U S A* **116**(17), 8149-8154.

[241] Sanborn, A. L. *et al.* (2015) Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc Natl Acad Sci U S A* **112**(47), E6456-E6465.

[242] Bonev, B. and Cavalli, G. (2016) Organization and function of the 3D genome. *Nat Rev Genet* **17**, 661–678.

[243] Therizols, P. *et al.* Chromatin decondensation is sufficient to alter nuclear organization in embryonic stem cells. *Science* **346**, 1238–1242.

[244] Gonzalez-Sandoval, A. *et al.* Perinuclear anchoring of H3K9-Methylated chromatin stabilizes induced cell fate in C. elegans embryos. *Cell* **163**, 1333–1347. (2015)

[245] Tortora, M. M. C, Brennan, L., Karpen, G. and Jost. D. (2022) Liquid-liquid phase separation recapitulates the thermodynamics and kinetics of heterochromatin formation. *bioRxiv* 2022.07.11.499635

[246] Plys, A. J. *et al.* (2019) Phase separation of Polycomb-repressive complex 1 is governed by a charged disordered region of CBX2. *Genes Dev.* **33**, 799–813.

[247] Seif, E. *et al.* (2020) Phase separation by the polyhomeotic sterile alpha motif compartmentalizes Polycomb Group proteins and enhances their activity. *Nat. Commun.* **11**, 5609.

[248] Colombo, M., Pessey, O. and Marcia, M. (2019) Topology and enzymatic properties of a canonical Polycomb repressive complex 1 isoform. *FEBS Lett.* **593**(14), 1837–1848.

[249] Choi, J., Bachmann, A., Tauscher, K. *et al.* (2017) DNA binding by PHF1 prolongs PRC2 residence time on chromatin and thereby promotes H3K27 methylation. *Nat Struct Mol Biol* **24**, 1039–1047.

[250] Tortora, M. M. C, Salari, H. and Jost, D. (2020) Chromosome dynamics during interphase: a biophysical perspective. *Curr Opin Genet Dev* **61**, 37-43.

[251] Milo, R., Jorgensen, P., Moran, U. and Weber, G. (2010) Springer M. BioNumbers–the database of key numbers in molecular and cell biology. *Nucleic Acids Res.* **38**(Database issue), D750-3.

[252] Atallah, J., Vurens, G., Mavong, S., Mutti, A., Hoang, D., and Kopp, A. (2014) Sex-specific repression of dachshund is required for Drosophila sex comb development. *Dev. Biol.* **386**, 440–447.

[253] Gutiérrez, L., Oktaba, K., Scheuermann, J.C., Gambetta, M.C., Ly-Hartig, N. and Muller, J. (2012). The role of the histone H2A ubiquitinase Sce in polycomb repression. *Development* **139**, 117–127.

[254] Wang, X. *et al.* (2017) Molecular analysis of PRC2 recruitment to DNA in chromatin and its inhibition by RNA. *Nat Str Mol Bio* **24**(12), 1028–1038.