Biophysical Journal
Article

Biophysical Society

Chromatin dynamics controls epigenetic domain formation

Marina Katava,[1] Guang Shi,[2] and D. Thirumalai[2,*]
[1]Laboratoire de Biochimie Théorique, CNRS, Université de Paris, Paris, France and [2]Department of Chemistry, The University of Texas, Austin, Texas

ABSTRACT In multicellular organisms, nucleosomes carry epigenetic information that defines distinct patterns of gene expression, which are inherited over multiple generations. The enhanced capacity for information storage arises by nucleosome modifications, which are triggered by enzymes. Modified nucleosomes can transfer the mark to others that are in proximity by a positive-feedback (modification begets modification) mechanism. We created a generic polymer model, referred to as 3DSpreader, in which each bead, representing a nucleosome, stochastically switches between unmodified (*U*) and modified (*M*) states depending on the states of the neighbors. Modification begins at a specific nucleation site (NS) that is permanently in the *M* state, and could spread to other loci that is dictated by chromatin dynamics. Transfer of marks among the non-nucleation loci occurs stochastically as chromatin evolves in time. If the spreading rate is slower than the chromatin relaxation rate, which is biologically pertinent, then finite-sized domains form, driven by contacts between nucleosomes through a three-dimensional looping mechanism. Surprisingly, simulations based on the 3DSpreader model result in finite bounded domains that arise *without* the need for any boundary elements. Maintenance of spatially and temporally stable domains requires the presence of the NS, whose removal eliminates finite-sized modified domains. The theoretical predictions are in excellent agreement with experimental data for H3K9me3 spreading in mouse embryonic stem cells.

SIGNIFICANCE Epigenetic spreading and maintenance play an important role in cellular differentiation and disease. A number of factors contribute to spreading mechanisms, which makes it difficult to enumerate all the possibilities using experiments alone. We created a generic computational model, accounting for chromatin dynamics and stochastic enzyme-catalyzed modifications of the epigenetic states. The enzyme reactions that modify and erase the epigenetic marks on the nucleosomes are modeled using reversible two-state kinetics. We find that discrete epigenetic domains are formed around the nucleation sites, which act as a source of epigenetic modifications. Surprisingly, finite heterochromatin domains form spontaneously by a looping mechanism without the need for boundary elements. We reproduce several experimental findings, including domain sizes in mouse embryonic stem cells.

INTRODUCTION

The inheritance of distinct phenotypes that are not encoded in the DNA sequence has been demonstrated in multicellular organisms. The resulting distinct morphological characteristics are maintained over multiple cellular divisions (1,2). Alterations in the epigenome, leading to distinct gene expression and subsequent phenotype variations, without any change to the underlying DNA sequence, are referred to as epigenetic modifications, which are carried over multiple cell divisions. As a consequence, some aspects of cellular memory are often associated with the term epigenetics (3–5). The study of the establishment and inheritance of genetic patterns is a burgeoning field, especially because epigenetic misregulation is implicated in aging and cancer. In eukaryotes, DNA condenses to form chromatin by wrapping around histone proteins. The physicochemical mechanisms governing genetic activity constitute a myriad of strategies that change the structure and organization of chromatin. These include chemical tagging of DNA (6,7) and histones (8), as well as regulation of transcription factors (9), RNA interference (10), chromatin remodeling proteins (11), and nuclear architecture (12). The epigenetic landscape emerging from these alterations enables the storage of more information than is possible using sequence alone and represents a powerful force in cellular differentiation and environmental adaptation (13).

Submitted October 26, 2021, and accepted for publication June 30, 2022.
*Correspondence: dave.thirumalai@gmail.com
Editor: Yi Qin Gao.
https://doi.org/10.1016/j.bpj.2022.07.001

The intricacies of initiation, spreading, and maintenance of histone modifications are unclear because of the involvement of several factors whose roles have not been quantitatively elucidated. Upon DNA replication, histones are redistributed in roughly equal proportions to the template and nascent DNA (14). Subsequently, nascent histones must re-establish the appropriate epigenetic attributes acquired before cell division ("memory"). Studies on chromatin inactivation have revealed important elements necessary for gene silencing, such as nucleation elements, which are specific DNA segments that bind protein complexes either directly or via RNA interference, and histone-modifying proteins, such as methyltransferases, which covalently modify histone tails (1). Although molecular identities of these mediators differ across eukaryotic species, certain unifying principles seem to underlie the observed homology between proteins that moderate them in yeasts, humans, mice, and flies (15,16). Experiments and theoretical studies suggest that a positive-feedback allosteric mechanism is important for the spreading of modifications, whereby the molecular complex binding to the nucleation site (NS) has an enhanced propensity to bind to nucleosomes that are already marked by an appropriate enzyme. The bivalent binding is thought to be the basis of the cooperative effect in spreading of the modifications (1).

Because of the involvement of several molecular components and inherent stochasticity in the modification process, a variety of mathematical models, which have provided insights into the formation of epigenetic domains and their self-perpetuation, have been proposed (2,17–26). Some models may be classified as one-dimensional (1D), in which spreading occurs along a lattice representing the chromatin with built-in implicit positive-feedback mechanism (21,22). Other models consider the possibility of spreading beyond near neighbors of a modified nucleosome, which implicitly accounts for long-range contacts (19,20) along the genomic length. More recently, models that consider the polymeric characteristics of chromatin explicitly (17,18,25,26) or implicitly (20) have been investigated.

We surmise that chromatin organization is strongly related to the spreading of histone modification, as shown by the strong correlation between the epigenetic states and the compartments observed in the Hi-C contact maps (27). To explore the interplay between chromatin structure and dynamics on the spreading of epigenetic marks, we introduce a polymer-based model that accounts for the kinetics of modifications, thereby coupling chromatin dynamics with stochastic chemical kinetics. The model captures the conformational dynamics of chromatin as well as the biochemical mechanism of spreading, implemented through distinct rules under which the enzyme reactions take place. The aim is to test the extent to which conformational dynamics influence the formation of epigenetic domains, and whether the spreading of epigenetic marks is achieved along the chromatin thread, or is controlled by non-adjacent nucleosomes (see Fig. 1). Our model, with a minimal number of parameters, shows that stable epigenetic domains emerge without explicit boundaries when the conformational rearrangement of chromatin dictates the spreading of epigenetic modifications.

We find that the NS acts as a positional signal that facilitates spreading to its surroundings. The bi-directional spreading from the NS was previously explored without considering chromatin conformation explicitly (20) or under conditions in which the formation of an epigenetic domain relies on attractive interactions between similarly modified loci (18,26). As a consequence, partial collapse of chromatin is required for epigenetic spreading. In our model, the cooperative effect of the writing and erasing processes, by which the probability of modification of a locus depends on epigenetic states of the neighbors, provides the sole positive reinforcement for spreading without the need for global conformational change of the chromatin. We show that finite modified domains cannot form without a looping mechanism, which brings nucleosomes that are well separated along the genome into proximity. The excellent agreement between the simulations and experiments on H3K9me3 spreading in mouse embryonic stem cells validates the proposed mechanism.

METHODS

Model

Based on the modification mechanisms sketched in Fig. 1, we used a polymer with $N = 300$ nucleosomes to probe epigenetic spreading. Each monomer of size σ represents 200 base pairs (bp), modeling one nucleosome and a linker DNA. Following our previous study (28), we choose the simplest bead on string model. Although features such as DNA linker length, nucleosome shapes, and orientations are important for the chromatin fold (29), we concentrate on the interplay between chromatin dynamics and enzyme kinetics on epigenetic spreading. The 300-monomer chain models 60 kb of DNA. The choice of chain length allows us to efficiently explore the spreading mechanism for a range of parameters. The effects of changing N are discussed in the supporting material. The modified and unmodified states of the nucleosome are denoted as M and U, respectively. The present model is a generalization of the chromatin copolymer model (CCM), which successfully predicted the organization and dynamics of interphase chromosomes (28).

The energy function used to simulate chromatin dynamics is described in the supporting material. The connectivity of the chain is modeled using the harmonic potential (see Eq. 1, Section 1 in the supporting material). A Kratky-Porod angular term (Eq. 2 in Section 1 in the supporting material) is used to define an angle potential between three successive nucleosomes. The persistence length l_p is defined through the parameter $l_k = 2l_p$. The nonbonded interactions are described using Lennard-Jones potential (Eq. 3 in Section 1 in the supporting material), where the parameter ε determines the strength of non-bonded interactions. The parameters in the model are given in Table S1 in the supporting material.

Physical properties of the chain

The value of ε determines whether the polymer adopts a random coil or is collapsed (see Section 6 in the supporting material for a detailed discussion). We first simulated chromatin in a good solvent using a fixed ε value that does not depend on the modification state of the nucleosome. This

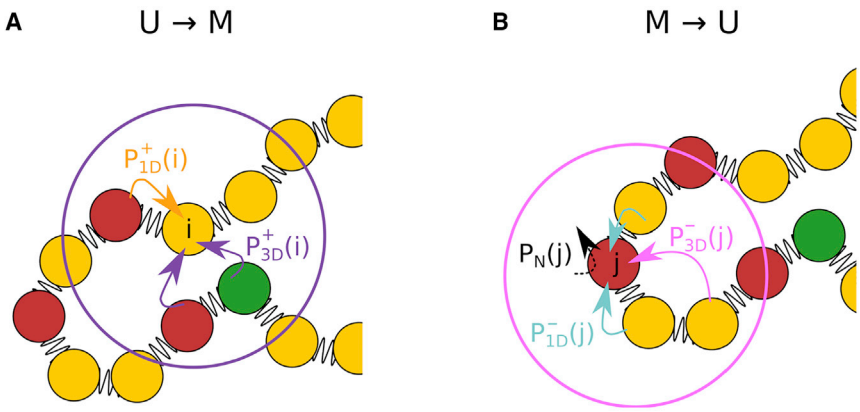


FIGURE 1 Transitions in the kinetic model. (*A*) The forward and backward reactions are shown on the left and right, respectively. The left depicts the enzyme-induced modification, $U \rightarrow M$, at the i^{th} nucleosome. The right sketches the removal of the mark, $M \rightarrow U$, from the j^{th} nucleosome. Modifications to unmarked (U) nucleosomes and removal of marks from the M-type nucleosomes are shown by arrows. Important mediators relevant to the spreading processes are shown in color: the nucleation site (NS) is in green, and U (M) nucleosomes are shown in yellow (red). Some or all elementary transitions are considered simultaneously, representing four separate spreading mechanisms (see Fig. S1). The forward reaction occurs if an unmarked nucleosome i has a neighbor nucleosome in the modified state with a probability $P^{+}_{1D}(i)$ (orange arrow) or with a probability $P^{+}_{3D}(i)$ if it is in the spatial vicinity of an NS/modified nucleosome (purple arrow). The backward reaction can always proceed with a probability $P_N(j)$, indicated by a black dashed arrow. The $M \rightarrow U$ reaction may also depend on the presence of other unmarked nucleosomes. In such cases, neighbor nucleosomes might induce the $M \rightarrow U$ transition in nucleosome j with a probability $P_{1D}{}^{-}(j)$ (cyan arrow), while U nucleosomes in the spatial vicinity of j may induce the reverse reaction with probability, $P_{3D}{}^{-}(j)$ (magenta arrow). (*B*) Asymmetric spreading from the NS (green) and non-NS sites (red). The forward rate for the non-NS sites is scaled by a factor of α compared with that for the NS site. Typically, we use $\alpha \ll 1$ in this study, which means that the ability of spreading of a non-NS site is much smaller than that of the NS site. To see this figure in color, go online.

scenario, with $\varepsilon_{MM} = \varepsilon_{UU} = \varepsilon$, explores the case where chromatin dynamics directly affects spreading but not vice versa. To explore the two-way coupling of chromatin dynamics and epigenetic spreading, we also simulated the spreading process by taking $\varepsilon_{MM} \neq \varepsilon_{UU}$.

Of particular relevance for epigenetic spreading is the persistence length of l_p, because it affects the kinetics of loop formation, which in turn controls three-dimensional (3D) spreading through the looping mechanism (25). The persistence length could vary depending on many parameters, including the organism type, chromatin length scale, and the epigenetic state. Flexible polymer models have been successful in reproducing experiments (30–32). We explored the impact of varying l_p on the spreading process (see the supporting material for details). Note that the chromatin dynamics and looping is not only determined by the l_p but also by ε. The results shown in this study is focused on the case that $l_p = \sigma$ because we aim to create a generic model, not adapted to a particular organism, but rather one that is used to explore different scenarios.

Epigenetic modifications

We denote the U state to represent the active chromatin, which transitions to the marked M state, inactive chromatin, in a single-step reaction $U \underset{k^-}{\overset{k^+}{\rightleftarrows}} M$, where k^+ and k^- are forward and backward enzyme rates, respectively. The forward reaction is a caricature of chemical modifications of histone 3 (H3) lysine 9 (K9) with methyl groups (me2/3), referred to as H3K9me2/3, a hallmark of inactive chromatin (15). The modification is catalyzed by Clr4 in *Schizosaccharomyces pombe* and Suv39h in humans (15,16). The reverse reaction takes into account the removal of markers corresponding to histone turnover. In addition, U can reverse the state of M nucleosomes in their vicinity, thus introducing the cooperative effect in the reverse reaction. The latter process may mimic the activity of enzymes that remove histone modifications (for example, Epe1 in *S. pombe* (33)). Both the two-state model (20,24) and three-state models (active, inactive, and unmarked) (17–19,26) have previously been used to study transitions between epigenetic states.

The modifications achieved by specific DNA binding enzymes are modeled as stochastic events. We evaluate the transition probabilities between the two states at each time step (see Section 2 in the supporting material). The important parameters that determine the probability of spreading are $r = k^+/k^-$ and $\beta = k^+\tau_r$, where τ_r, the relaxation time of the chromatin is a characteristic time that is used to measure the spreading rate. In addition, we introduce the parameter α (Eq. 7 in supporting material), which rescales the forward rate. We use $\alpha = 1$ for spreading from the NS (Fig. 1). The α parameter is introduced to account for the possibility that spreading from the NS and from modified nucleosomes does not represent the same sequence of molecular events. A transition can occur only if at least one of the following five conditions is satisfied, shown schematically in Fig. 1:

1. 1D: $U \rightarrow M$ transition may occur only if at least one nucleosome at $i \pm 1$ is in the M state.
2. 3D: $U \rightarrow M$ transition may occur only if there is a minimum of one nucleosome, j, in state M that satisfies the criteria $|j - i| \geq 2$ and the distance r_{ij} to i^{th} nucleosome is less than the threshold, r_c.
3. Noise: $M \rightarrow U$ transition could occur at any nucleosome independent of the identity of all others.
4. 1D: $M \rightarrow U$ transition may occur only if least one nucleosome at $i \pm 1$ is in the U state.
5. 3D: $M \rightarrow U$ transition may occur only if there is a minimum of one nucleosome, j, in state U that satisfies $|j - i| \geq 2$ and $r_{ij} < r_c$.

Because we are interested in epigenetic spreading, we do not consider the spontaneous stochastic $U \rightarrow M$ transition (the analog of the third transition listed above). These transitions are allowed with probabilities given in Eq. 7 in the supporting material. Since multiple elementary steps could contribute to an epigenetic transition at each time step on a single nucleosome, the overall probability of changing the epigenetic state depends on the epigenetic states of its neighbors, except for the elementary transition 3, whose contribution is constant. We specifically explore two cases: 1) Scheme I: linear spreading model that includes elementary transitions 1, 3, and 4; and 2) Scheme II: linear and 3D spatial spreading model that includes all the listed elementary transitions 1–5. The details of the schemes are in Fig. S1 *a* and we mathematically formalize the overall forward ($P^+(t)$) and backward ($P^-(t)$) probability at each time step in Fig. S1 *b*.

At $t = 0$, all the nucleosomes are unmarked, except for the NS (shown in green in Fig. 1), chosen arbitrarily to be in the middle of the polymer. We also investigated the effects of altering the location of the NS (see Section 11 in the supporting material). Spreading starts at the NS, which enables bidirectional spreading along the nucleosomes, emanating outward from its location. A spreading trajectory is generated by stochastic changes in the nucleosome states, as outlined above in the five steps.

Since we consider only two epigenetic states, the nucleosome states can be characterized using Ising spin variables, $s(U) = -1$ and $s(M) = 1$ (34), except that the spin variables change stochastically with time depending on the instantaneous chromatin conformation. In such a system, there could be a

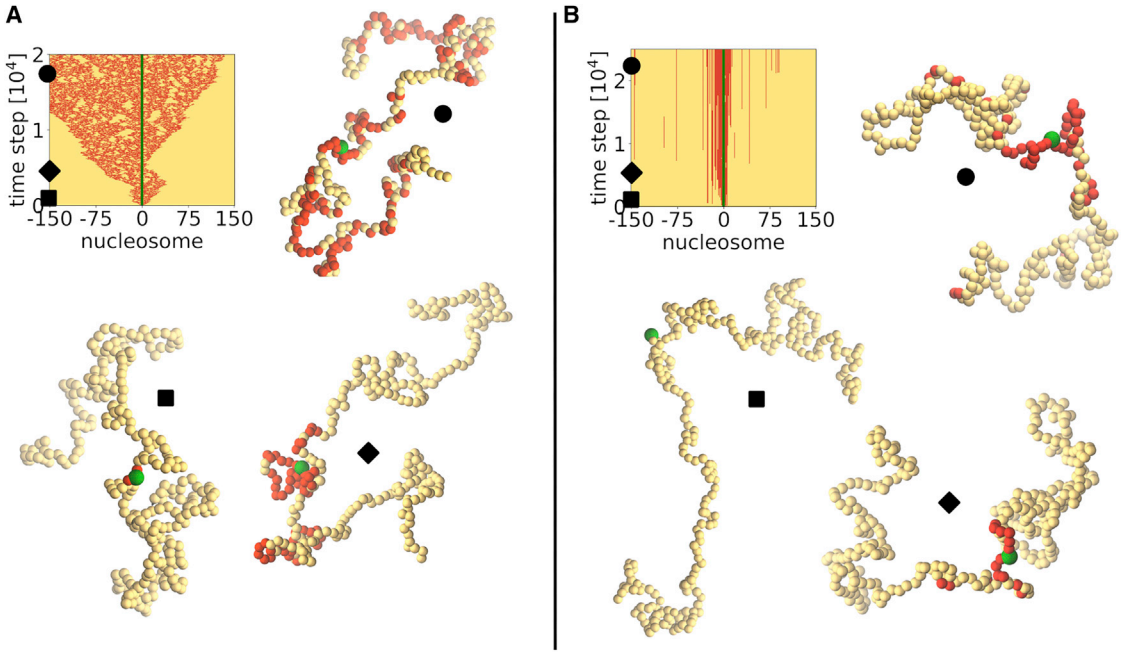


FIGURE 2 Visualization of single trajectories in the fast (*A*) and slow (*B*) spreading regimes. In the former case, the spreading occurs fast and independently of chromatin configuration, while in the latter case, the pattern of spreading may be determined by the dynamic rearrangement of the chromatin, as explained in the main text. The kymographs show the epigenetic identity of each nucleosome (marked/M/red, unmarked/U/yellow, NS/green) for a single trajectory, in both (*A*) and (*B*). Three distinct simulation time steps are chosen, marked with a symbol on the kymograph, and we extract polymer conformations with epigenetic identities of the nucleosomes to graphically represent them. For clarity, the size of the NS is enhanced. Accompanying Videos S1 and S2 are available online. To see this figure in color, go online.

cooperative transition in the spin states as chromatin evolves. The positive feedback, as the basis of cooperativity, arises because the probability of modifying a nucleosome increases if nucleosomes in the vicinity are already modified. This mimics the "reader-writer" model associated with the epigenetic spreading process, where modifying enzymes such as methyltransferases has a reader domain that binds to the previously modified nucleosome, and subsequently write the same modification to another nucleosome (35).

Epigenetic clock

The probabilities associated with epigenetic transitions in the model are coupled to the chromatin dynamics. The time that determines 3D spreading is associated with the lifetime of contact between two nucleosomes i and j with $|i-j| \geq 2$. There is a spectrum of lifetimes associated with loop formation depending on $|i-j|$. Our goal is to elucidate how chromatin dynamics are coupled to the enzyme-catalyzed reactions that modify the U state or erase the mark in the M state. Therefore, we chose the polymer relaxation time, τ_r, calculated from the time-dependent decay of the structure factor $F(q,t)$, evaluated at the wave vector $q = 2\pi/r_c$ (see Section 5 in the supporting material), to set the overall time scale (time unit of the epigenetic clock). We show in Fig. S5 that τ_r is a reasonable surrogate for the mean contact lifetimes during which modifications could occur by the looping mechanism. The dependence of the spreading on l_p and chain length N are described in the supporting material (Sections 9 and 10, respectively). Note that, in contrast to previous polymer models, there is an explicit connection between chromatin relaxation time (τ_r) and the dynamics of spreading. We set τ_r as the unit of time, and all other times are measured relative to τ_r.

Spreading rates

The limit of slow spreading (SS) is achieved by choosing $k^+\tau_r = \beta \ll 1$. In the SS limit, chromatin undergoes multiple cycles of relaxation, which implies that nucleosomes separated by large genomic distance have a substantial probability of being in proximity for spreading to occur by the looping mechanism. With our choice of $\beta = 0.01$, we find numerically that spreading by the 1D mechanism is not as important. We introduce the parameter γ in order to control the contribution of 1D spreading mechanism. In this case, the probability of spreading is given by $P^+(t = \gamma\tau_r)$. In the steady-state limit, the value of γ has no bearing on the results of the simulations because it is a 3D dominated mechanism. Thus, by changing β, one can examine a range of possibilities for epigenetic spreading. Based on the estimation provided in the discussion section below, we believe that the SS limit is biologically relevant, and we show these results in the main text.

For completeness, we also explore the opposite limit, fast spreading (FS). The results are shown in the supporting material. In the FS limit, $k^+\tau_r = \beta \gg 1$, which implies that modification occurs on time scales that are much shorter than τ_r, the chromatin relaxation time. In this limit, spreading occurs predominantly by a linear or 1D mechanism with the looping mechanism playing a less significant role. To achieve this limit in the simulations, γ (Fig. S1) is chosen so that $P^+(t = \gamma\tau_r)$ is large. In the FS limit, this is achieved by choosing $\beta\gamma > C(\approx 3 \sim 4)$.

Implementation

We start from an equilibrated conformation of the chromatin polymer with all the nucleosomes in the U state at $t = 0$. The simulations were also repeated with all nucleosomes in the M states in order to show that the steady-state behavior is independent of the initial conditions, which is assured because the chromatin is epigenetically ergodic (see Section 8 in supporting material for details). In addition, the simulations were performed for times that exceed τ_r by several orders of magnitude, so that the chromatin conformation is well sampled.

The conformation of the polymer is evolved by integrating the Langevin equations by choosing a suitable time step (described in Section 3 in the supporting material). At each time step, marking or unmarking of all the

nucleosomes is attempted with probabilities given in Fig. S1 *b*. A graphical representation of the modification algorithm is provided in Fig. 2 in the supporting material. A pictorial representation of the simulated trajectories for FS and SS limits are shown in Fig. S2, and in Videos S1 and S2.

Global epigenetic state

We use simple measures to characterize the global epigenetic state of chromatin as well as the modification state of the individual nucleosomes for both the FS and SS scenarios. By examining the dynamics in detail, we can quantitatively assess the contributions from 1D and 3D spreading separately for the two extreme cases. We determine the epigenetic state of chromatin using,

$$\langle S\rangle = \frac{1}{N_{\text{traj}}}\frac{1}{T}\sum_j\sum_t \frac{n_m^j(t) - n_u^j(t)}{N}, \tag{1}$$

where n_m^j (n_u^j) is the number of modified (unmodified) nucleosomes at time t in the j^{th} trajectory, and N is the total number of nucleosomes. The quantity $(n_m^j - n_u^j)/N$ is averaged over time T and N_{traj}. Unless otherwise stated, all the relevant quantities are calculated using $N_{\text{traj}} = 10$, where each trajectory starts with a different initial chromatin conformation.

Average epigenetic state of the nucleosomes

We also determined the average, $\langle s_i\rangle$, of each nucleosome, which is calculated using,

$$\langle s_i\rangle = \frac{1}{T}\frac{1}{N_{\text{traj}}}\sum_t\sum_j s_i^j(t), \tag{2}$$

where $s_i^j(t)$ is the epigenetic state of locus i in trajectory j at time t.

The fraction of time each nucleosome is modified is measured by f_{im},

$$f_{im} = \frac{1}{T}\frac{1}{N_{traj}}\sum_j\sum_t \delta^j_{s_i,+1}(t), \tag{3}$$

where $\delta^j_{s_i,+1}$ counts the number of occurrences of nucleosome i in state M in trajectory j, while T is the total number of simulation snapshots taken into account.

RESULTS

Chromatin topology drives domain formation

We explored the SS case by choosing $k^+ = 0.01/\tau_r$. In this limit, the chromatin polymer could form the allowed contacts through the looping process multiple times on the time scale $\approx 1/k^+$, which would allow for 3D as well as 1D spreading. Consequently, the inactivation profile would be determined by the contact probability between the nucleosomes separated by $|i-j|\geq 2$ (Fig. S6).

These expectations are tested by computing the epigenetic states of the nucleosomes as well as the associated inactivation profiles. We fixed k^+ and varied k^- to improve the odds of spreading, deciding finally on the choice $k^- = k^+/10{,}000$. For the chosen parameters, domain formation by the linear mechanism does not occur (Fig. 3, Scheme I). This is because the probability of linear spreading to neighboring nucleosomes, even away from the NS, is extremely low ($\approx 6\cdot 10^{-5}$) at each time step. Thus, we surmise that spreading must occur exclusively through the formation of 3D contacts. This is borne out in the panels for Scheme II in Fig. 3, which show clearly that the 3D spreading mechanism results in the formation of stable domains around the NS. It is worth noting that stable domain formation ($\approx$ 60 nucleosomes centered around the NS) does not require collapse (17) or partial collapse (18) of the chromatin polymer.

The spreading profile in the right panel of Fig. 3 for Scheme II shows that the peak in the f_{im} profile is localized around the NS. The boundary (or the interface) between the active and inactive domains is relatively soft (36), indicated by a continuous decrease of f_{im}, rather than by a step-like drop to 0, which would occur if the boundary were sharp. Thus, the percentage of inactive loci in the interface region between the active and inactive nucleosomes could indicate whether the boundary is efficient in preventing modified nucleosomes from spreading distally of its position.

An important finding in our work, with potential biological importance, is that, even without an explicit boundary element, the epigenetic domain size is finite, and is localized around the NS. The distribution of modifications around the NS is governed by the contact probability of the NS with surrounding residues, which depends on the chromatin conformational dynamics. Furthermore, the shape of the epigenetic domain shown in the right panel of Fig. 3 and the boundary formation is due to the asymmetry in the spreading rates from the NS (k^+) and the modified nucleosomes (αk^+ with α less than unity). This asymmetry is required for the formation of discrete domains without explicit boundary elements to halt the spreading process, and points to the importance of the NS that we explore further (see below).

DNA replication and the NS

Next, we address the role of the NS in a pre-formed epigenetic domain and the impact of DNA replication. In our model, the NS is the only element whose epigenetic identity is unchanged, and functions as a reservoir for modifications. We performed simulations in which the conditions change over time. First, we performed simulations with NS present. At $t = 301{,}205\tau_r$, we removed the NS by changing it to a non-NS nucleosome in the M state, and continued running simulations in this condition for another $t = 301{,}205\tau_r$. With this alteration, the identity of the M state could change stochastically as the chromatin evolves. At $t = 602{,}410\tau_r$, we mimic DNA replication by randomly assigning the active state to half of the nucleosomes, repeating this assignment every $150.6\tau_r$. Note that the end-to-end distance of the polymer relaxes in about $602.4\tau_r$, indicating that the DNA replication time ($150.6\tau_r$) is shorter than the polymer relaxation time. We note parenthetically that the time for finite

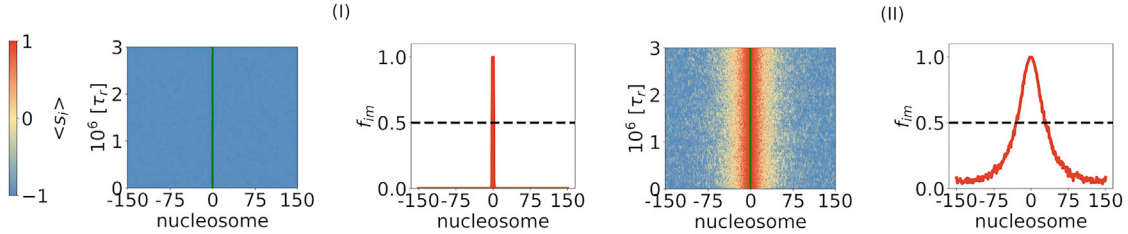


FIGURE 3 The formation of the epigenetic domain is determined by SS kinetics with $k^+ = 0.01/\tau_r$ for Scheme I, where spreading occurs only through 1D, and Scheme II, where 3D spreading dominates. (Left) Average (over an ensemble of trajectories) states of nucleosomes as a function of time. (Right) Mean values of the spin states of the nucleosomes. These are the profiles that show the extent of inactive domain formation. The value of $k^+/k^- = 10{,}000$ in all simulations, while $\alpha(\mathrm{I}) = 2.23 \cdot 10^{-4}$ and $\alpha(\mathrm{II}) = 2 \cdot 10^{-4}$. To see this figure in color, go online.

domain formation is roughly 50 times greater than the replication time (see Fig. 3). This implies that in our model it takes roughly 50 generations to establish finite domains. Experiments have shown that, in certain species, establishment of stable domains occurs in about 20 generations (37,38). We could have altered the DNA replication time in our simulations to reduce the number of generations needed for establishing stable domains. Nevertheless, our results are in qualitative agreement with experiments.

We follow the epigenetic identities of all the nucleosomes at all time steps and show them as a kymograph in the left panel of Fig. 4. We also calculate the inactivation profile of the chromatin for two distinct parts of the trajectory: 1) portion of the trajectory without NS (red curve in the right panel of Fig. 4), and 2) portion of the trajectory where DNA replications are repeated periodically (magenta curve in the right panel of Fig. 4). We define epigenetic domain to exist in the region above the level $f_{im}>0.5$. Fig. 4 shows the results for Scheme II in the SS regime. The figure shows that the finite-sized domain cannot be maintained once the NS is deleted (red curve). Stable, bounded domains that are maintained during DNA replication would only occur if the looping mechanism for modification is allowed and the NS does not change. Thus, the results in Fig. 4 show that the NS is required for domain maintenance, and even more so if DNA replication is permitted (magenta curve). Experimental evidence on *S. pombe* (39) does suggest that some nucleation events needed for heterochromatin formation are triggered at each cell cycle, indicating that maintenance requires the nucleation event.

Finite domain formation in an initially condensed or partially condensed chromatin

The results presented so far were obtained by setting $\varepsilon_{\mathrm{MM}} = \varepsilon_{\mathrm{UU}} = 0.1k_BT$. This choice places the chromatin in a good solvent, which could mimic the behavior in fission yeast (40). However, in some other eukaryotes, the heterochromatin is denser than euchromatin. In the context of our model, it would imply that the effective interaction $\varepsilon_{\mathrm{MM}} > \varepsilon_{\mathrm{UU}}$. To account for the higher chromatin density, we set the interaction strength $\varepsilon_{\mathrm{MM}} = 3\varepsilon_{\mathrm{UU}}$, which implies that, due to dynamical changes in the epigenetic landscape, the effective interactions between the nucleosomes are modified as in the previous polymer-based studies (17,18,26). We set $\varepsilon_{\mathrm{UM}}$ to the geometric mean, $\sqrt{\varepsilon_{\mathrm{MM}}\varepsilon_{\mathrm{UU}}}$. Since linear spreading does not depend on the solvent quality, we only explored Scheme II, where looping determines the outcome of the spreading process. We simulated the SS case ($k^+\tau_r = \beta = 0.01$).

To fully explore the range of solvent conditions, we considered two cases. In both, we chose $\varepsilon_{\mathrm{MM}} = 3\varepsilon_{\mathrm{UU}}$. Let us first consider $\varepsilon_{\mathrm{UU}} = 0.1k_BT$ and $\varepsilon_{\mathrm{MM}} = 0.3k_BT$. For a homopolymer, $\varepsilon_{\mathrm{MM}} = 0.3k_BT$ is close to but slightly

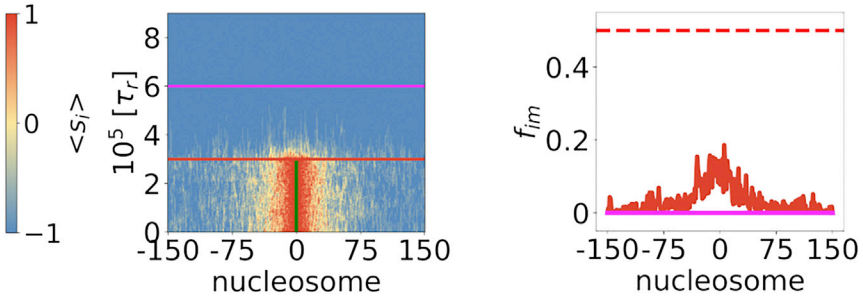


FIGURE 4 Role of nucleation in establishing and maintenance of epigenetic spreading for Scheme II in the SS regime. (Left) The state of each nucleosome in a trajectory, $\langle s_i \rangle$ (Eq. 2 in section "methods"). The trajectory is divided into three parts with distinct simulation conditions, whose transition is marked by two horizontal lines. Initially, the simulation starts with the NS present. Subsequently, the NS is removed (marked by a red horizontal line). After removing the NS, a new steady state is achieved. Thereafter, we introduce DNA replication (marked by magenta horizontal line) by stochastically replacing approximately 50% of M marks with U marks, mimicking equal redistribution of parent histones to both strands upon replication. Replications are repeated every $150.6\tau_r$. (Right) The presence of epigenetic domains provided the fraction of trajectory in the marked state for each locus is > 0.5. The red curve is computed from the trajectory between the NS removal and DNA replication. The magenta curve is computed from the trajectory after the DNA replication. The horizontal dotted red line represents the 50% mark. $k^+/k^- = 10{,}000$ and $k^+ = 0.01/\tau_r$. To see this figure in color, go online.

below the Θ point (see Fig. S4 in Section 6 in the supporting material), whereas for $\varepsilon_{UU} = 0.1k_BT$ chromatin behaves as disordered random coil. With this particular choice, we find (upper panels in Fig. 5) that the results are qualitatively similar to the case with $\varepsilon_{MM} = \varepsilon_{UU} = 0.1k_BT$ (Fig. 3). In particular, the domain size is bounded, as discovered in experiments on mouse stem cells and fibroblasts (36). This is outlined both by the pattern formed around the NS on the kymograph (upper left panel) as well as in the activity profile of the domain (upper right panel). Since the epigenetic spreading in this regime is mostly due to contacts between the NS and spatially nearby nucleosomes, spreading is localized.

The $\varepsilon_{MM} = 3\varepsilon_{UU}$ with $\varepsilon_{UU} = 0.1k_BT$ corresponds to a mixed case in which a fraction of modified nucleosomes experience moderately poor solvent conditions, whereas the U nucleosomes are in good solvent (see Fig. S4 in Section 6 in the supporting material). We then investigated the case when the modified nucleosomes are in bad solvent conditions and unmodified nucleosomes are close to but slightly below the Θ condition by choosing $\varepsilon_{MM} = 3\varepsilon_{UU} = 0.9k_BT$ while keeping the ratio $\varepsilon_{MM}/\varepsilon_{UU} = 3$ as before. The results in the lower panels in Fig. 5 are dramatically different from those in the upper panels. The spreading of the marks is centered around the NS as before. However, modification occurs without bound with the probability of the i^{th} nucleosome to be in the M state is ≈0.7 as the genomic length increases or decreases from the NS. In contrast to the results in Fig. 4, which shows that probabilities of being in the M or U state away from the NS are ≈0.5 (except for the ends), the results in the lower panels demonstrate a substantially higher probability of being in the M state. Taken together, the results in Fig. 5 suggest that $\varepsilon_{MM}/\varepsilon_{UU}$ is not as important as the solvent quality for predicting the outcome of the spreading dynamics. Although $\varepsilon_{MM}/\varepsilon_{UU} = 3$ in Fig. 5, the initial conformation of the chromatin is a disordered coil in the upper panel, whereas in the lower panels it is more compact. The final epigenetic state of chromatin can be anticipated by the quality of the solvent that the chromatin is in at the initial state, especially in the SS limit, where spreading occurs predominantly by the 3D mechanism.

Comparison with experiments on mouse embryonic stem cells

We explored the biological pertinence of our results by comparing them with the H3K9me3 enrichment profiles reported elsewhere (see Fig. 6 in (36)). In this study, the authors investigated the propagation of HP1α-induced H3K9me3 modification in the 10-kb *Oct4* locus in embryonic stem cells (36), and found that H3K9me3 propagated symmetrically to produce finite spatial domains. To rationalize the observations, they proposed a 1D model (24,36) in which spreading occurs with finite probability to the neighboring sites from an already modified nucleosome. The enrichment in H3K9me3 marks can be partitioned into two groups by using clustering (41); small domains that encompass 77.1% of the data (Fig. 6, left panel) and large domains that account for the rest (Fig. 6, right panel) (36). Since the enrichment profiles are equivalent to inactivation profiles reported throughout this paper, we make a comparison between the two in Fig. 6.

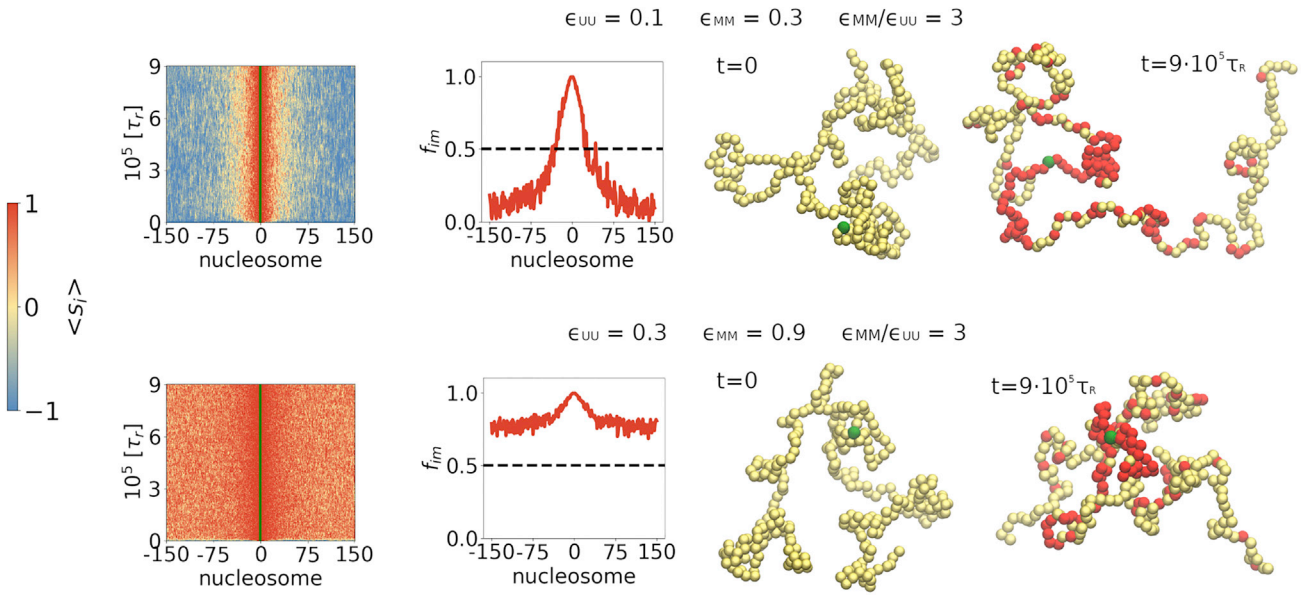


FIGURE 5 Effect of solvent quality on the spreading of epigenetic modifications using Scheme II, in which looping affects the spreading process. The ε_{UU} and ε_{MM} values of each are shown. For all simulations, $k^+/k^- = 10{,}000$ and $k^+ = 0.01/\tau_r$. The initial and the final conformations from a single trajectory are displayed on the right. To see this figure in color, go online.

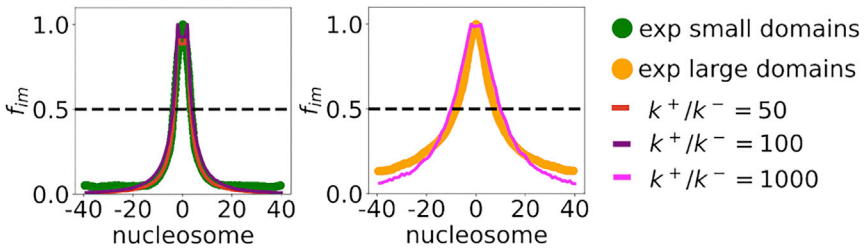


FIGURE 6 Comparison between simulations performed using Scheme II in the SS regime at different k^+/k^- values with experimental data for H3K9me3 domains in embryonic stem cells (41). The experimental data were divided into small and large domains by clustering. These domains differ in the size of the genomic length they cover, and we fit them using different k^+/k^- ratios that control the spreading process. We show that the small experimental domains correspond to a range of k^+/k^- parameters in our model (left), while a higher ($k^+/k^- = 1000$) value is needed to account for the larger domain size shown on the right. The parameters are chosen to be $k^+ = 0.01/\tau_r$ and $\alpha = 2k^-/k^+$. To see this figure in color, go online.

To assess if our model, which involves an interplay of 1D and 3D spreading, could be used to reproduce the reported H3K9me3 enrichment profiles described above, we performed calculations using Scheme II. Using $k^+ = 0.01/\tau_r$, we find that, for a range of k^+/k^- values, we obtain good agreement with the enrichment profiles (Fig. 6) for both small and large domains. The excellent agreement between our predictions and experiments shown in Fig. 6 allows us to draw a few pertinent conclusions. 1) The competition between the forward and the backward rate remains crucial (36) when accounting for epigenetic domain width, even when 3D spreading is possible. For example, if k^+/k^- is large, domains without bound form. 2) The experimental data compare well with a model where the NS is situated in every epigenetic domain. In our model, the spreading rate from the NS is greatly enhanced compared with spreading from non-NS-modified nucleosomes. NSs are sequence-defined DNA fragments to which enzymes bind with high specificity and different conformational changes might arise by binding to different sequences, thus modulating the enzyme activity. 3) Although a linear spreading model can also reproduce the experiment curve (36), our results show that 3D contacts are not inconsistent with the experimental observation. Different mechanisms might be operative in heterochromatin formation in different species. The combination of 1D and 3D models, which integrates the rates of spreading and chromatin dynamics created here, could be a step toward fulfilling this goal. Indeed, there is experimental evidence pointing to the importance of both linear (nearest-neighbor) (42) spreading of H3K9me3 marks, as well as spatial spreading (43) of H3K27 marks, showing that chromatin silencing could proceed by both mechanisms, as accounted for in Scheme II of our model.

DISCUSSION

We developed a minimal polymer model to explore different scenarios for epigenetic domain formation with a focus on the coupling between chromatin dynamics and stochastic switching in individual nucleosome states. In our model, the interplay of the structural relaxation rate and the modification rates determines the efficacy of the spreading process. However, by setting our work in the context of the modeling literature on epigenetic spreading, we conclude that multiple scenarios for heterochromatin spreading are possible because the domain formation is an interplay of several time, length, and energetic scales. A similar picture emerges from experimental studies, encompassing different organisms and epigenetic modifications. It is only by exploring various scenarios and analyzing their validities against experimental data that we can determine the physics underlying epigenetic spreading, initiated by the complex enzymatic chemical reactions.

The main findings in our work are:

- In the SS limit, with the $U \rightarrow M$ modification rate that is much smaller than the relaxation rate of the polymer, finite domains are established on the time scales of $\approx 10,000\tau_r$. This time translates to about 50 generations, which is big, but is in qualitative accord with experimental results for silencing in certain species (37). The domain size is about 60 nucleosomes or roughly 12 kb. The width of the interface between the active and inactive domains is soft, involving a small number of nucleosomes (Fig. 3).
- The presence of the NS is essential for the formation of the modified domains. An already established domain cannot be maintained if NS is removed. It should be noted that there is evidence for the NS in experiments (36). We should add that one can envision other mechanisms where finite domains could be established without explicitly having NS in the model (25,26). However, it is gratifying and maybe surprising that we are able to predict finite domain spreading purely from SS of the marks in conjunction with looping as a way to transfer the epigenetic marks.
- An important finding in our work is that, in the limit of SS, finite modified domains form without boundary elements that are known to stop the spreading process. We find that finite domains form by mechanism II, which predominantly involves 3D chromatin looping. If the relaxation time associated with chromatin dynamics is shorter than the spreading time, then finite domains form. The formation of the finite domain does not depend on the spreading cutoff, because the rate is determined by the relaxation times associated with chromatin dynamics.

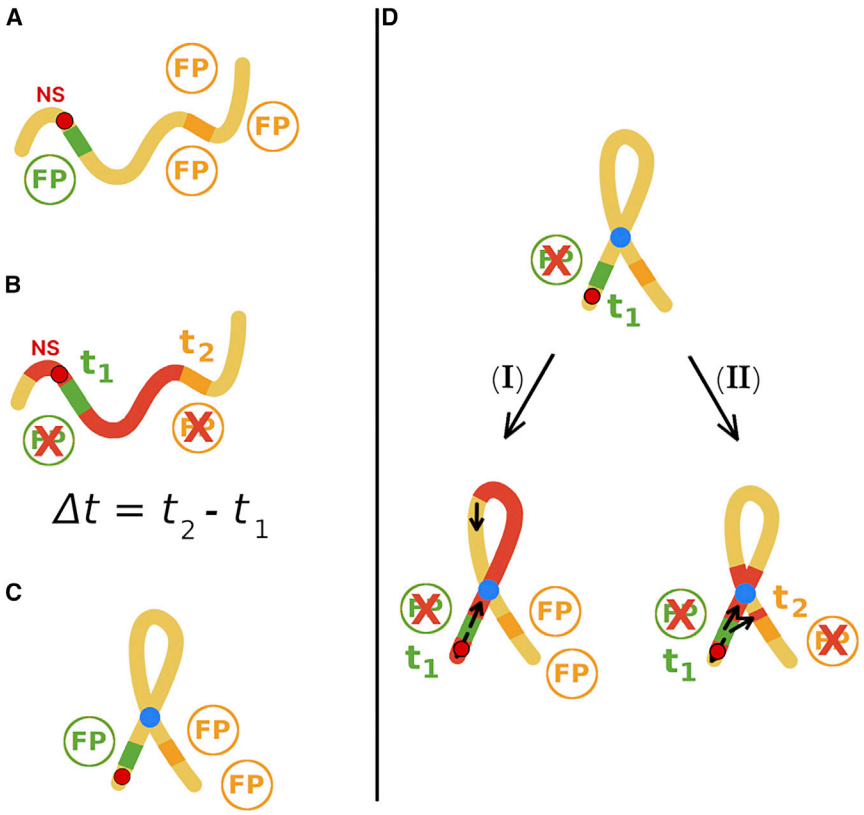


FIGURE 7 Scheme for the proposed experiment. (*A*) The gene for the orange fluorescent protein (OFP, in orange) is expressed if chromatin is unmarked (yellow). The repression of the green fluorescent protein gene (GFP) acts as a proxy for the binding of epigenetic spreading initiators to the NS, shown in a red circle with a black border. Because the nucleation event has not occurred, the GFPs and OFPs are produced. The presence of fluorescent proteins (FPs) could be detected by measuring fluorescence intensity at a given wavelength. (*B*) The binding of epigenetic spreading initiators to the NS represses the GFP expression at time t_1, which could be measured. Subsequently, epigenetic markers (red) spread until the OFP gene is inactivated, resulting in the disappearance of orange fluorescence at t_2, which is measurable. During spreading, chromatin can form loops but need not have a permanent structure. The time difference, $\Delta t = t_2 - t_1$, could be measured for many single cells to obtain the distribution $P(\Delta t)$. This represents the control experiment; a similar experimental setup has been employed in (44). (*C*) A permanent loop (blue) in chromatin structure may be constructed so that the spatial proximity of the NS and OFP gene is decreased for a given genomic NS-FP distance. (*D*) Using the FPs as reporters for spreading, only cells that lack green fluorescence are taken into account, as it marks the successful initiation of epigenetic spreading from the NS. If the heterochromatin spreading from the NS is strictly 1D, then $P(\Delta t)$ should not differ from the result in the control experiment. If 3D spreading is relevant, $P(\Delta t)$ would differ compared with the control experiment. The same conclusions should be achieved by placing the OFP gene at different genomic distances from the NS, for a given loop length. To see this figure in color, go online.

It is possible to form finite domains using polymer models (interactions between modified nucleosomes are attractive), which have many genomic bookmarks (GBMs) that, in effect, are the boundary elements, which stop epigenetic spreading (26). In this model, it was found that finite-sized domains, with spacing that is given by the genomic distance between the GBMs, form only if the linear density of the GBMs exceeds a critical value. The presence of a large number of GBMs is tantamount to using many boundary elements, which we find is unnecessary in the present model. It is only by quantitatively analyzing specific experiments that we can assess the validity of various models. We provide such experimental validation of our model in this paper by showing quantitative agreement with experiments.

- In an insightful article (25), it has been shown using a polymer model that, once formed, finite domains can be passed on to multiple generations upon DNA replication. In their model, the initial methylation patterns were taken from chromatin immunoprecipitation sequencing (ChiP-seq) data, and the maintenance of the silencing pattern is driven by heterochromatin protein 1 (HP1). Informed by experiments that HP1 preferentially binds to methylated tails in the histones, they showed the initial methylation pattern from the ChiP-seq data is not only maintained over nine generations but also spreads without having an NS. The HP1-mediated silencing is not unrelated to the GBM used elsewhere (26). In the present study, we were interested in exploring mechanisms for the establishment of finite domains. We do find, however, that the NS is necessary to maintain the finite domain upon DNA replication. We believe that if we imposed boundary elements, which would be the analog of preferentially binding of HP1 to methylated regions (25) or the use of GBM (26), there would not be a need to introduce the NSs. The present and previous studies are exploring different possibilities for spreading and maintenance of epigenetic marks, and only future experiments can differentiate between them.
- A major determinant of spreading is the solvent quality of the initial (fully unmarked) epigenetic state, rather than the relative strength of the $U-U$ and $M-M$ interactions. If the initial epigenetic state of the chromatin is a coil, contacts rarely form, resulting in localized spreading around the NS, which leads to a discrete bounded domain (upper panel in Fig. 5). If, initially, chromatin is poised to be just below the Θ point, we find that stable finite domains form (upper panel in Fig. 5) even if the interactions between the modified nucleosomes are attractive. On the other hand, if the initial chromatin state is such that it is

partially condensed, we find that epigenetic domains form without bound (lower panel in Fig. 5). Because the propagation of heterochromatin without bound is biologically untenable, our results suggest that either there ought to be multiple boundary elements that stop the spreading of the epigenetic marks or the environmental conditions for the unmodified nucleosomes should poise them close to the Θ conditions.

The data can be generated using the in-house computer code that has been deposited in Github. The URL is https://github.com/Katamar/epigenetic.

Experimental prospects

To test the prediction that stable finite domains can be driven by the 3D organization of chromatin without boundary elements, we propose an experiment based on fluorescent probes, similar to the one previously reported (44). The difference would be that the nucleation element and position-dependent fluorescent probe are placed within a permanent artificially engineered loop. The proposed experiment, shown schematically in Fig. 7, would help elucidate the extent to which chromatin looping plays a role in the formation of epigenetic domains. More generally, it could clarify whether epigenetic domain formation occurs by linear or spatial spreading or a combination of both these mechanisms, as is likely the case.

SUPPORTING MATERIAL

Supporting material can be found online at https://doi.org/10.1016/j.bpj.2022.07.001.

AUTHOR CONTRIBUTIONS

M.K. created the model. M.K. and G.S. implemented the model. M.K. performed the simulations. M.K.,G.S., and D.T. interpreted the findings. D.T. supervised the work. All authors discussed the results and contributed to the writing of the final manuscript.

ACKNOWLEDGMENTS

We thank Bassem Al-Sady and Ilya Finkelstein for advice and useful discussions. This work was initiated while M.K. was a postdoctoral fellow at the University of Texas. We acknowledge the National Science Foundation (CHE 19-00093) and the Collie-Welch Regents Chair (F-0019) for supporting this work.

DECLARATION OF INTERESTS

The authors declare that there is no conflict of interest.

REFERENCES

1. Moazed, D. 2011. Mechanisms for the inheritance of chromatin states. *Cell.* 146:510–518.
2. Cortini, R., M. Barbi, …, J. M. Victor. 2016. The physics of epigenetics. *Rev. Mod. Phys.* 88:25002.
3. Ptashne, M. April 2007. On the use of the word 'epigenetic'. *Curr. Biol.* 17:R233–R236.
4. Ptashne, M. 2013. Epigenetics: core misconcept. *Proc. Natl. Acad. Sci. USA.* 110:7101–7103.
5. Ringrose, L., and M. Howard. 2017. Dissecting chromatin-mediated gene regulation and epigenetic memory through mathematical modelling. *Curr. Opin. Syst. Biol.* 3:217–314.
6. Holliday, R., and J. E. Pugh. 1975. DNA modification mechanisms and gene activity during development. *Science.* 187:226–232.
7. Compere, S. J., and R. D. Palmiter. 1981. DNA methylation controls the inducibility of the mouse metallothionein-i gene in lymphoid cells. *Cell.* 25:233–240.
8. Bannister, A. J., and T. Kouzarides. 2011. Regulation of chromatin by histone modifications. *Cell Res.* 21:381–395.
9. Todeschini, A. L., A. Georges, and R. A. Veitia. 2014. Transcription factors: specific DNA binding and specific gene regulation. *Trends Genet.* 30:211–219.
10. Holoch, D., and D. Moazed. 2015. RNA-mediated epigenetic regulation of gene expression. *Nat. Rev. Genet.* 16:71–84.
11. Becker, P. B., and J. L. Workman. 2013. Nucleosome remodeling and epigenetics. *Cold Spring Harb. Perspect. Biol.* 5:a017905.
12. Dekker, J., and T. Misteli. 2015. Long-range chromatin interactions. *Cold Spring Harb. Perspect. Biol.* 7:a019356.
13. Heard, E., and R. A. Martienssen. 2014. Transgenerational epigenetic inheritance: myths and mechanisms. *Cell.* 157:95–109.
14. Alabert, C., and A. Groth. 2012. Chromatin replication and epigenome maintenance. *Nat. Rev. Mol. Cell Biol.* 13:153–167.
15. Beisel, C., and R. Paro. 2011. Silencing chromatin: comparing modes and mechanisms. *Nat. Rev. Genet.* 12:123–135.
16. Grewal, S. I., and D. Moazed. 2003. Heterochromatin and epigenetic control of gene expression. *Science.* 301:798–802.
17. Michieletto, D., E. Orlandini, and D. Marenduzzo. 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.
18. Jost, D., and C. Vaillant. 2018. Epigenomics in 3d: importance of long-range spreading and specific interactions in epigenomic maintenance. *Nucleic Acids Res.* 46:2252–2264.
19. Dodd, I. B., M. A. Micheelsen, …, G. Thon. 2007. Theoretical analysis of epigenetic cell memory by nucleosome modification. *Cell.* 129:813–822.
20. Erdel, F., and E. C. Greene. 2016. Generalized nucleation and looping model for epigenetic memory of histone modifications. *Proc. Natl. Acad. Sci. USA.* 113:E4180–E4189.
21. Berry, S., C. Dean, and M. Howard. 2017. Slow chromatin dynamics allow polycomb target genes to filter fluctuations in transcription factor Activity. *Cell Syst.* 4:445–457.e8.
22. Angel, A., J. Song, …, M. Howard. 2011. A Polycomb-based switch underlying quantitative epigenetic memory. *Nature.* 476:105–108.
23. Dodd, I. B., and K. Sneppen. 2011. Barriers and silencers: a theoretical toolkit for control and containment of nucleosome-based epigenetic states. *J. Mol. Biol.* 414:624–637.
24. Hodges, C., and G. R. Crabtree. 2012. Dynamics of inherently bounded histone modification domains. *Proc. Natl. Acad. Sci. USA.* 109:13296–13301.
25. Sandholtz, S. H., Q. MacPherson, and A. J. Spakowitz. 2020. Physical modeling of the heritability and maintenance of epigenetic modifications. *Proc. Natl. Acad. Sci. USA.* 117:20423–20429.

26. Michieletto, D., M. Chiang, …, D. Marenduzzo. 2018. Shaping epigenetic memory via genomic bookmarking. *Nucleic Acids Res.* 46:83–93.
27. Lieberman-Aiden, E., N. L. van Berkum, …, J. Dekker. October 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science.* 326:289–293.
28. Shi, G., L. Liu, …, D. Thirumalai. 2018. Interphase human chromosome exhibits out of equilibrium glassy dynamics. *Nat. Commun.* 9:3161.
29. Grigoryev, S. A. 2018. Chromatin higher-order folding: a perspective with linker DNA angles. *Biophys. J.* 114:2290–2297.
30. Socol, M., R. Wang, …, A. Bancaud. 2019. Rouse model with transient intramolecular contacts on a timescale of seconds recapitulates folding and fluctuation of yeast chromosomes. *Nucleic Acids Res.* 47:6195–6207.
31. Langowski, J. 2006. Polymer chain models of DNA and chromatin. *Eur. Phys. J. E Soft Matter.* 19:241–249.
32. Knoch, T. A. 2019. Simulation of different three-dimensional polymer models of interphase chromosomes compared to experiments–an evaluation and review framework of the 3d genome organization. *Semin. Cell Dev. Biol.* 90:19–42.
33. Ragunathan, K., G. Jih, and D. Moazed. 2015. Epigenetic inheritance uncoupled from sequence-specific recruitment. *Science.* 348:1258699.
34. Garel, T., H. Orland, and E. Orlandini. 1999. Phase diagram of magnetic polymers. *Eur. Phys. J. B.* 12:261–268.
35. Zhang, K., K. Mosch, …, S. I. Grewal. 2008. Roles of the clr4 methyltransferase complex in nucleation, spreading and maintenance of heterochromatin. *Nat. Struct. Mol. Biol.* 15:381–388.
36. Hathaway, N. A., O. Bell, …, G. R. Crabtree. 2012. Dynamics and memory of heterochromatin in living cells. *Cell.* 149:1447–1460.
37. Xu, E. Y., K. A. Zawadzki, and J. R. Broach. 2006. Single-cell observations reveal intermediate transcriptional silencing states. *Mol. Cell.* 23:219–229.
38. Obersriebnig, M. J., E. M. H. Pallesen, …, G. Thon. 2016. Nucleation and spreading of a heterochromatin domain in fission yeast. *Nat. Commun.* 7:11518.
39. Chen, E. S., K. Zhang, …, S. I. Grewal. 2008. Cell cycle control of centromeric repeat transcription and heterochromatin assembly. *Nature.* 451:734–737.
40. Matsuda, A., Y. Chikashige, …, Y. Hiraoka. 2015. Highly condensed chromatins are formed adjacent to subtelomeric and decondensed silent chromatin in fission yeast. *Nat. Commun.* 6:7753.
41. Bilodeau, S., M. H. Kagey, …, R. A. Young. 2009. SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state. *Genes Dev.* 23:2484–2489.
42. Müller, M. M., B. Fierz, …, T. W. Muir. 2016. A two-state activation mechanism controls the histone methyltransferase suv39h1. *Nat. Chem. Biol.* 12:188–193.
43. Leicher, R., E. J. Ge, …, S. Liu. 2020. Single-molecule and in silico dissection of the interaction between polycomb repressive complex 2 and chromatin. *Proc. Natl. Acad. Sci. USA.* 117:30465–30475.
44. Greenstein, R. A., S. K. Jones, …, B. Al-Sady. 2018. Noncoding RNA-nucleated heterochromatin spreading is intrinsically labile and requires accessory elements for epigenetic stability. *eLife.* 7:e32948.

Biophysical Journal, Volume *121*

Supplemental information

Chromatin dynamics controls epigenetic domain formation

Marina Katava, Guang Shi, and D. Thirumalai

Supporting Information: Chromatin dynamics controls epigenetic domain formation

Marina Katava, Guang Shi, D. Thirumalai

1 Energy Function

Energy function: The energy of the chromatin, U_T, is taken to be a sum of bond stretch (U_B), bond angle (U_{KP}), and interactions (U_{LJ}) between the nucleosomes that are separated by at least 3 bonds.

Bond stretch and bond angle potentials: The connectivity of the chromatin thread is taken into account using a harmonic potential,

$$U_{\mathrm{B}} = \frac{k_s}{2}(r - r_0)^2, \tag{1}$$

where r is the distance between the two consecutive nucleosomes, r_0 is the equilibrium bond length, and k_s is the spring constant. The bond angle is constrained using the Kratky-Porod potential in order to control the stiffness of the chain. We assume that,

$$U_{\mathrm{KP}} = \frac{k_B T l_k}{2\sigma}\Big[1 - \frac{\mathbf{t}_1 \cdot \mathbf{t}_2}{|\mathbf{t}_1||\mathbf{t}_2|}\Big], \tag{2}$$

where k_B is the Boltzmann constant, T is the temperature, $l_k/2$ is the intrinsic persistence length (l_p), and σ is the effective inter nucleosome distance. The length unit is σ. The variables $\mathbf{t_1}$ and $\mathbf{t_2}$ are bond vectors connecting nucleosomes (i, $i+1$) and ($i+1$, $i+2$), respectively.

Non-bonded potential: We used the Lennard-Jones (LJ) potential,

$$U_{\mathrm{LJ}} = 4\epsilon\Big[\Big(\frac{\sigma}{r}\Big)^{12} - \Big(\frac{\sigma}{r}\Big)^{6}\Big], \tag{3}$$

to model interactions between non-bonded loci. In the above equation, r is the distance between the nucleosomes, σ is roughly the size of the nucleosome, and ϵ, which sets the energy scale, is the strength of the interactions. The LJ interaction is truncated at 3σ ($\epsilon = 0$ for $r > 3\sigma$).

2 Modification Probabilities

The modification of the nucleosomes by enzymes is modeled using a two state kinetics,

$$\mathrm{U} \underset{k^-}{\overset{k^+}{\rightleftharpoons}} \mathrm{M}, \tag{4}$$

where k^+ and k^- are the forward and backward reaction rates, respectively. The solutions to the rate equations,

$$\begin{aligned} -\frac{d[U]}{dt} &= k^+[U] - k^-[M] \\ -\frac{d[M]}{dt} &= -k^+[U] + k^-[M]. \end{aligned} \tag{5}$$

are given by,

$$\begin{aligned} P^+(t) &= 1 - \frac{k^- + k^+ e^{-\lambda t}}{\lambda} \\ P^-(t) &= 1 - \frac{k^+ + k^- e^{-\lambda t}}{\lambda}, \end{aligned} \tag{6}$$

where $\lambda = k^+ + k^-$ and t is time. By defining $r = k^+/k^-$ and $\beta = k^+\tau_r$, where τ_r is the characteristic time scale, the modification probabilities become,

$$P^+(\tau_r) = 1 - \frac{1+\alpha r e^{-\beta(1+\alpha r)/\alpha r}}{1+\alpha r}; P^-(\tau_r) = 1 - \frac{\alpha r + e^{-\beta(1+\alpha r)/\alpha r}}{1+\alpha r}, \quad (7)$$

where $\alpha = 1$ for nucleation site. The relevant parameters that will control spreading are r, α, and β. Note that the probabilities given in Eq.7 reduce to $P^+ = k^+\tau_r$ and $P^- = k^-\tau_r$ when $\tau_r \to 0$. When τ_r is not small, Eq.7 give the correct values for transition probabilities.

Finally, the probabilities of the modification and un-modification of each nucleosome are computed based on Eq.7 with the following assumptions: (i) spreading from the bonded neighbor and non-bonded neighbors are independent and contribute equally to the transition (ii) each neighbor nucleosome contributes to the transition independently. Additional details on the computation of spreading probability are given in Figure S1 and Methods in the main text. We assume that epigenetic spreading occurs through a set of enzyme reactions, which presumably involve physical proximity between the enzymes and the nucleosomes. Due to the possible crowding effect would restrict the available space around the nucleosome including those that are nearby. Therefore, we assume that the 1D and 3D spreading cannot occur at the same time, and maybe treated as independent events. For this reason, we used the total transition probability that is given in Figure (S1b).

3 Langevin dynamics

Parameter	Description	Value
σ (Eq 3)	unit of length	1
ϵ (Eq 3)	strength of LJ interaction	$0.1k_BT$
d_0	equilibrium bond distance	1.0σ
k_s (Eq 1)	spring constant between connected loci	$\frac{3000\epsilon}{d_0^2}$
l_k (Eq 2)	double the chromatin persistence length	2σ
a^*	value of σ	$27nm$

Table S1: Parameters used in the Langevin dynamics simulations. * a is roughly the sum of the size of the nucleosome ($10nm$) and the contour length of the DNA ($\simeq 50bp \cdot 0.33nm/bp \approx 17nm$). Equation numbers in the first column refer to the main text.

We describe the time evolution of the chromatin by the Langevin equation,

$$m_i \frac{d^2\mathbf{r}_i}{dt} = \mathbf{F}_i - \xi \frac{d\mathbf{r}_i}{dt} + \mathbf{R}_i(t), \quad (8)$$

where $\mathbf{r}_i$ is the position of the i^{th} locus whose mass is m_i, $\mathbf{F}_i = -\frac{\partial U_T}{\partial \mathbf{r}_i}$ is the systematic force arising from U_T (see the main text), ξ is the friction coefficient, and $\mathbf{R}_i$ is the random force that satisfies

(a)

	1D forward	3D forward	1D reverse	3D reverse	Noise
	$P^{+}_{1D}(i)$	$P^{+}_{3D}(i)$	$P^{-}_{1D}(i)$	$P^{-}_{3D}(i)$	$P_N(i)$
I: 1D+F+R	Y	0	Y	0	Y
II: 3D+F+R	Y	Y	Y	Y	Y

(b)

i. $P^{+}(i) = P^{+}_{1D}(i) + P^{+}_{3D}(i) - P^{+}_{1D}(i)P^{+}_{3D}(i)$

ii. $P^{-}(i) = P_N(i) + \big(1 - P_N(i)\big)\big(P^{-}_{1D}(i) + P^{-}_{3D}(i) - P^{-}_{1D}(i)P^{-}_{3D}(i)\big)$

iii. $P^{\pm}_{1D} = 1 - \big[1 - P_{\pm}(t = \gamma\tau_r)\big]^{n^{1D}_{\pm}}$ $P^{\pm}_{3D} = 1 - \big[1 - P_{\pm}(t = \tau_r)\big]^{n^{3D}_{\pm}}$ $P_N(i) = P_{-}(t = \gamma\tau_r)$

$$n^{1D}_{\pm} = \sum_{j\in\{i-1,i+1\}} \delta_{s_j,\pm 1} \qquad n^{3D}_{\pm} = \sum_{|i-j|>2} \Theta(r_c - d_{ij})\delta_{s_j,\pm 1}$$

Figure S1: **(a)** Columns represent the elementary reactions and the rows show the schemes implemented in the simulations. The five elementary reactions are: 1D forward, 1D reverse, 3D forward, 3D reverse, and the noise. The noise process, with the associated probability $P_N(i)$ where i is the nucleosome label, affects only the modified nucleosomes. The two simulation schemes are: **I** (1D+F+R , (1D forward, 1D reverse, and the noise)), and **II** (3D+F+R, (1D forward, 1D reverse, 3D forward, 3D reverse, and the noise)). Y indicates that the probabilities for the five elementary reactions are computed using equation **iii** in **(b)**, and 0 means that the corresponding probability is zero. For each simulation scheme, the cumulative probability for the forward ($P^{+}(i)$) and reverse ($P^{-}(i)$) reactions is computed using Equations **(b)i** and **(b)ii**. For scheme **I** (1D+F+R), the probabilities $P^{+}_{3D}(i)$ and $P^{-}_{3D}(i)$ are zero, resulting in $P^{+}(i) = P^{+}_{1D}(i)$ and $P^{-}(i) = P_N(i) + (1 - P_N(i))P^{-}_{1D}(i)$, where $P^{\pm}_{1D}(i)$ and $P_N(i)$ are computed using Equations **(b)iii**. The probability for scheme **II**, which includes all the elementary reactions, is given by $P^{+}(i)$ (Equation **(b)i**) and $P^{-}(i)$ (Equation **(b)ii**). **(b)** Equations for all the relevant probabilities. (i) and (ii) are the cumulative probability for forward and reverse reaction. (iii) Displays the probabilities for the five elementary reactions. $t = \gamma\tau_r$ ($t = \tau_r$) is the time at which 1D and noise (3D) spreading is considered. The variables $n^{1D}_{\pm}$ ($n^{3D}_{\pm}$) are the number of bonded neighbors (non-bonded neighbors) whose state is U (-1) or M ($+1$) to i^{th} nucleosome. Non-bonded neighbors (js) satisfy the criterion $|i - j| \geq 2$ and $r_{ij} < r_c = 1.122\sigma$.

the fluctuation-dissipation theorem. An in-house code was developed to integrate the Langevin equation, with $T = 300K$, using the velocity-Verlet algorithm [1]. After equilibrating the polymer for times that exceed the relaxation time ($\tau_{R_{ee}}$) of the end-to-end vector of the polymer, we performed long simulations ($\gg 10\tau_{R_{ee}}$) so that reliable statistics for computing various quantities

of interest could be generated.

4 Epigenetic modification algorithm flowchart

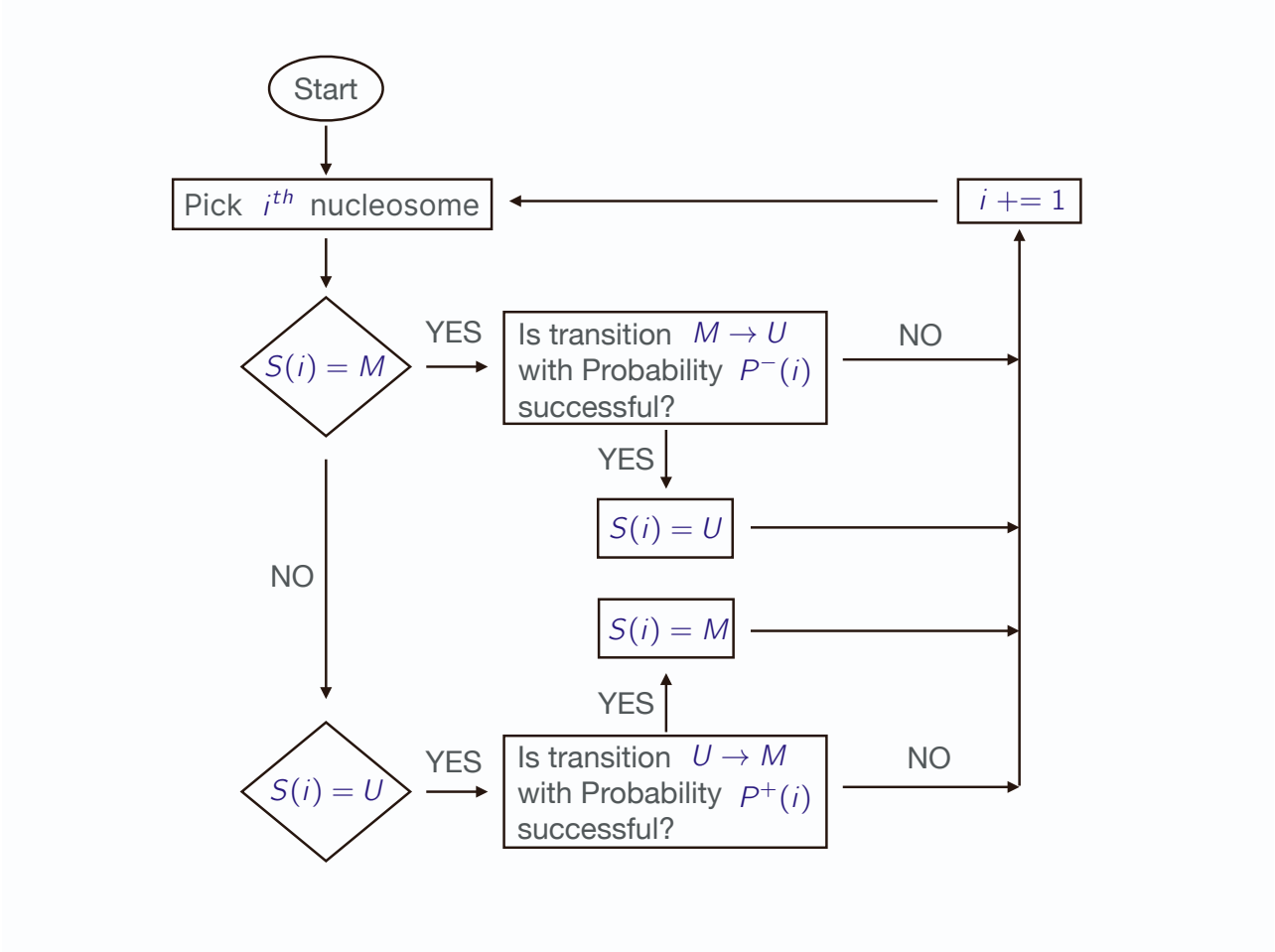


Figure S2: Flowchart of the algorithm for epigenetic modifications at each time step. The probability values are given in Figure S1.

5 Structural relaxation time

We calculated the structural relaxation time, τ_r, from the decay of the structure factor,

$$F(q,t) = \frac{1}{N}\left\langle \sum_{j=1}^{N} \exp^{-iq\cdot(r_j(t)-r_j(0))} \right\rangle, \tag{9}$$

with $q = \frac{2\pi}{r_c}$. In the above equation, $r_j(t) - r_j(0)$ is the displacement of nucleosome r_j. From the decay of $F(q,t)$ (Figure S3) we estimated the characteristic time, τ_r. The time τ_r, which is an estimate for the relaxation of the chromatin polymer, is assumed to set the over all time scale. All other rates that are relevant to epigenetic spreading are set relative to τ_r. It is natural to use τ_r, especially for 3D spreading to monitor the modification process. Note that τ_r is a function of N, l_p, as well as solvent quality.

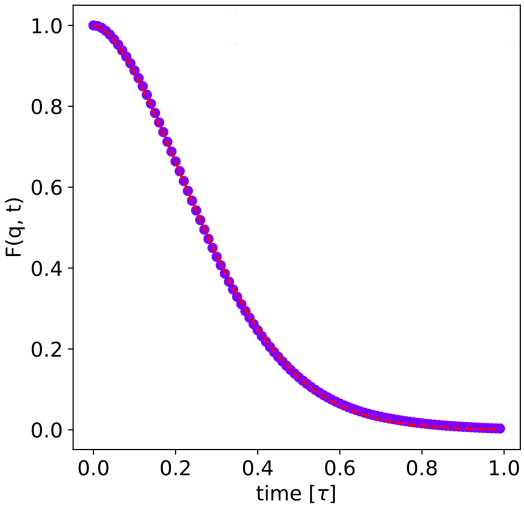
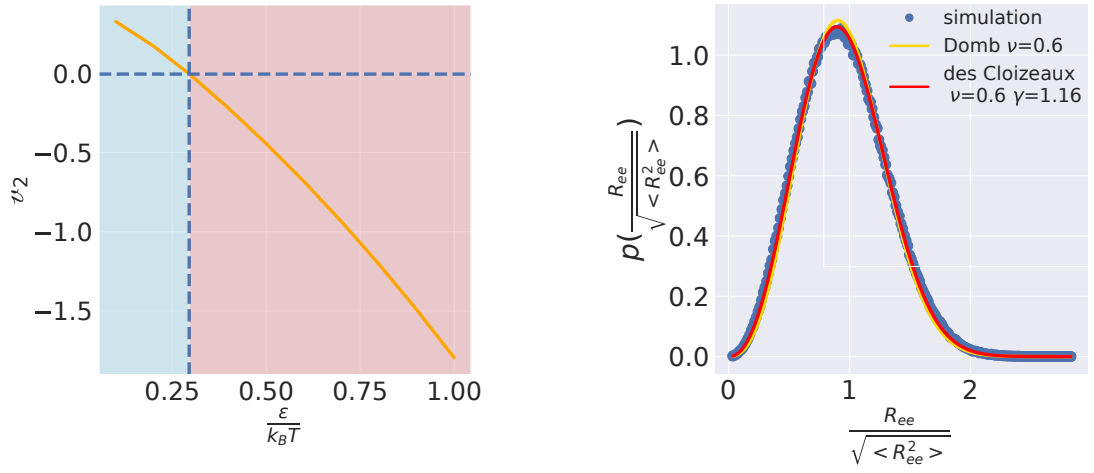


Figure S3: Intermediate scattering function, evaluated at the wave vector $q = \frac{2\pi}{r_c}$ ($r_c = 1.122\sigma$), as a function of time. $F(q,t)$ was averaged over 15 independent trajectories by removing the overall rotational and translational degrees of freedom. The value of τ_r is extracted using the fit, $F(q,t) = \exp^{-(t/\tau_r)^\beta}$ with β=1.7, yielding $\tau_r \approx 0.3\tau$ where τ is the natural time governing Eq. 8. For our purposes the precise value of the τ_r is irrelevant. What matters are the spreading probabilities given in Figure S1.

6 Solvent quality



(a) (b)

Figure S4: (a) Second virial coefficient as a function of the Lennard-Jones parameter ϵ, characterizing the interaction between the nucleosomes. The blue (red) region corresponds to "good solvent" ("poor solvent"). (b) Comparison of the end-to-end distance distributions computed from simulations ($\epsilon = 0.1 k_B T$) with the rigorous theoretical prediction for a polymer in a good solvent [2].

The non-bonded interactions are determined by the Lennard-Jones interaction strength, ϵ (Eq. 3 in the SI), between the nucleosomes that are separated by at least two bonds from each other.

The ϵ parameter is an effective interaction strength between the loci. The dimension of the chain depends on the second virial coefficient,

$$v_2 = 2\pi \int_0^\infty r^2 dr[1 - \exp^{-\beta U_{LJ}(r)}], \tag{10}$$

where $U_{LJ}(r)$ is given in Eq (3) in the main text, $\beta = 1/k_BT$. If $v_2 > 0 (< 0)$, then the chromatin could be extended (random coil). In Figure S4, we show v_2 as a function of $\beta\epsilon$. The θ-point at which $v_2 = 0$ corresponds to $\epsilon = 0.3k_BT$ (Figure S4). We varied ϵ in our simulations to assess the impact of solvent quality on the nature of epigenetic spreading. For $\epsilon = 0.1k_BT$, which is in the good solvent region, chromatin behaves as a Flory random coil. We expect that the distribution $P(x)$, with $x = R_{ee}/ < R_{ee}^2 >^{1/2}$, should follow the universal behavior with $P(x) \sim x^g \exp^{-x^\delta}$, where $\delta \approx 1/(1-\nu)$, where ν is the Flory exponent (≈ 0.6 in 3D). The excellent agreement between theory and simulation confirms that the chromatin polymer is indeed a random coil. We also show results for $\epsilon \geq 0.3k_BT$, see Figure **??** in main text.

7 Contact times and τ_r

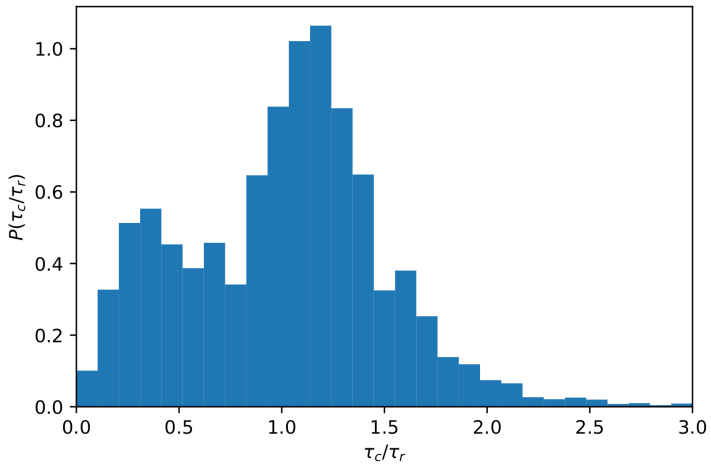


Figure S5: Distribution of contact duration times τ_c. A contact forms if two nucleosomes are within $r_c = 1.122\sigma$ The mean value is $\langle \tau_c \rangle = 0.84\tau_r$.

There are three time scales that characterize our epigenetic polymer model (i) The rate of the forward reaction rate k^+(Eq 4 in the main text), (ii) The second is k^-, the backward reaction rate. Both k^+ and k^- are defined in the main text. (iii) The third is, τ_r, the chromatin relaxation time. Below we describe how these timescales are chosen, and assigned physical meaning.
As shown in Figure **??**, epigenetic spreading could occur either linearly (1D) (i and $i \pm 1$) or by non-bonded nucleosomes that come into proximity through loop formation (3D). The looping time could be substantial, making the modification probability through the 3D mechanism less efficient than by 1D. To account for the separation in time scales, we allow for 1D modifications to occur on time scale $t = \gamma\tau_r$. In other words, the probabilities of modification through 1D mechanism are computed using equations in Figure S1 (b). Spreading in 3D occurs if two loci,

separated by at least 2 bonds, come into contact. There is a spectrum of looping times that depend on the separation $|i-j|$ between the nucleosomes. A relevant time for modification is the contact life time (τ_c) during which 3D spreading could occur. The distribution of $P(\tau_c/\tau_r)$ contact life times, expressed in units of τ_r, (Figure S5) shows that the average $<\tau_c>\approx 0.84\tau_r$. To simplify the computations, we calculated the probabilities of 3D spreading using equations shown in Figure S1 (b). The use of $\langle\tau_c\rangle$ as a proxy for looping, independent of the genomic separation between the loci, simplifies the computations without qualitatively altering the results. We assume that spreading is slow, which implies that the polymer relaxes much faster than the time scale in which the enzyme reaction ($\frac{1}{k^+}$) occurs. This is implemented by choosing $k^+ = \frac{0.01}{\tau_r}$.

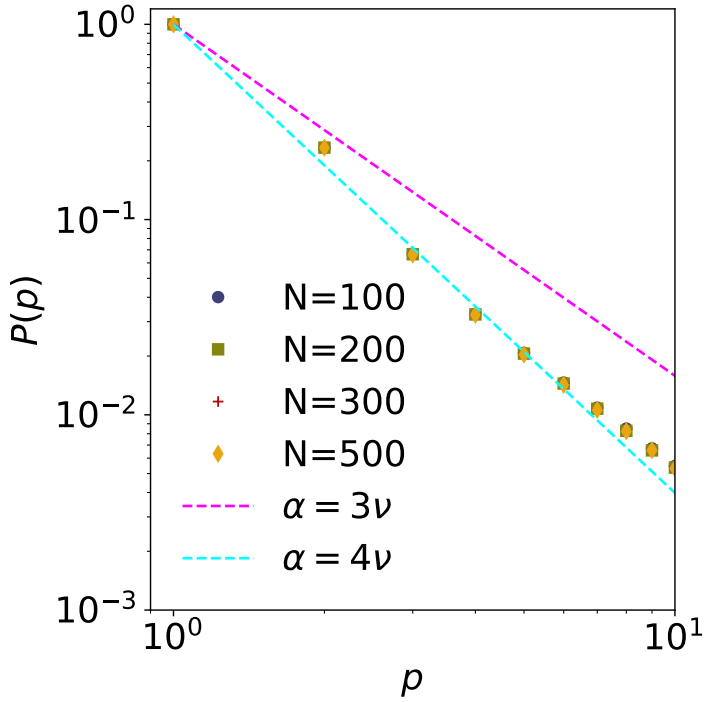


Figure S6: Dependence of the contact probability $P(p)$ as a function of the number of bonds that separate two nucleosomes, p. The dashed lines represent theoretical scaling $P(p) \sim p^{-\alpha}$, where α is a multiplicative value of the Flory exponent $\nu = 3/5$.

8 Epigenetic ergodicity

In order to asses if the time scales governing spreading result in a non-equilibrium state, we introduce the epigenetic ergodicity measure by generalizing a measure introduced in the context of glasses [3]. Multiple pairs of independent simulations were used to calculate the epigenetic ergodicity, defined as,

$$d(t) = \frac{1}{N}\sum_{j=1}^{N}[s_{a,j}(t) - s_{b,j}(t)]^2, \tag{11}$$

where $s_{a,j}(t)$ is the time-average of the epigenetic state of nucleosome j in trajectory a over time t, $s_{b,j}(t)$ is the corresponding quantity in trajectory b. If the system is ergodic, each trajectory

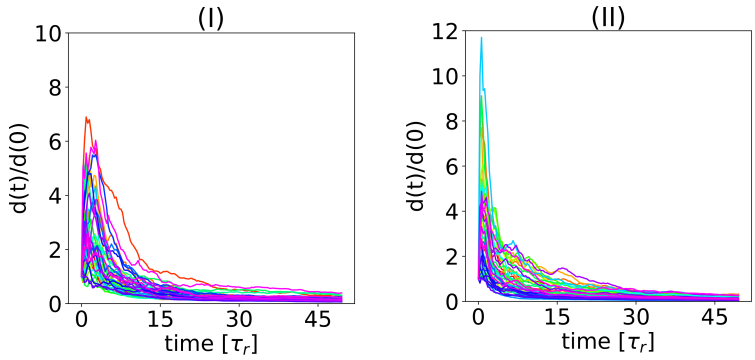


Figure S7: Epigenetic spreading is ergodic. We compared 10 independent trajectories for both biological mechanisms. The epigenetic ergodic measure vanishes at long times. The parameter values in simulations are $k^+/k^- = 50$, $k^+ = 100/\tau_r$, with total simulation time $15,000\tau_r$.

would explore the entire epigenetic phase space, and the time-average for both the trajectories should converge at long times. Thus, in an ergodic, or quasi ergodic system, the quantity $d(t)$ should vanish at long t. Figure S7 shows that $d(t)$ vanishes at long times for all the spreading processes on time scales that are far less than the time needed for steady-state spreading to be established. Thus, for the range of time scales considered here, and for $N = 300$, epigenetic ergodicity is established. It is conceivable that epigenetic ergodicity could be broken, resulting in glass-like epigenetic states for different N, and solvent quality governing the spreading dynamics. It is unclear if this could confer any biological advantage for epigenetic memory.

9 Influence of chromatin persistence length on epigenetic spreading

We anticipate two extreme scenarios for the influence of the persistence length, l_p, on the spreading mechanisms. If $L/l_p >> 1$ (L is the contour length of the polymer) then 3D transient loop-driven spreading, as found here and elsewhere [4, 5], is likely. In the opposite stiff chain limit $L/l_p << 1$, we expect that spreading would be predominantly determined by the 1D mechanism. To verify these expectations, we performed simulations by varying the bending stiffness (l_k) in the bond angle potential. For different values of l_k, we extracted the persistence length using,

$$\langle cos\theta(p)\rangle = \exp^{-\frac{p}{l_p}}, \tag{12}$$

where θ is the angle between two bond vectors separated by a distance p along the contour of the polymer. The exponential fit to simulations, with l_k ranging from $(2-200)\sigma$, is shown in Figure S8. A crossover between the flexible chain ($L \gg l_p$) to a rigid chain is expected when $L \ll l_p$. Figure S8 shows that the best fit is obtained when l_p spans several bond distances, while the rigid and flexible limit exhibit stronger deviations from the nominal estimate, $l_p = l_k/2$. The exponential decay $\langle cos\theta(p)\rangle$ (Eq. (12)), is valid strictly for a polymer without excluded volume interactions, and the deviations of the exponential scaling have previously been reported for large

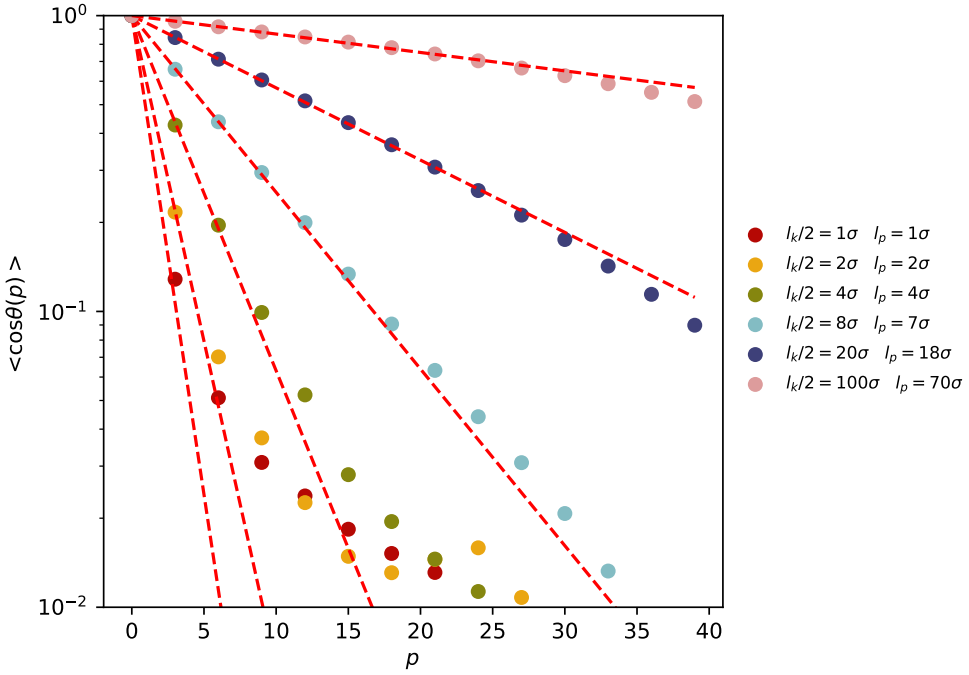


Figure S8: Bond vector correlation function $< cos\theta(p) >$ as a function of the bond distance p along the polymer for different intrinsic stiffness values. Red dashed lines are fits to the curves. The values of the persistence length l_p (listed on the right) are extracted using Eq 12. The fits are increasingly more accurate as l_p increases.

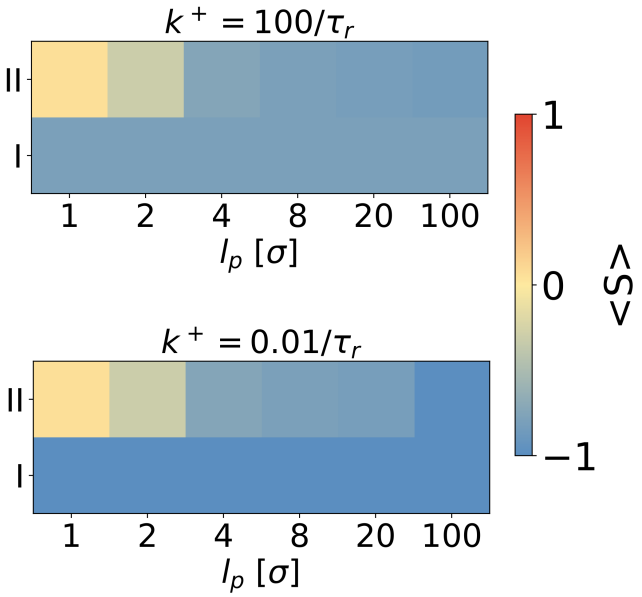


Figure S9: Heat maps showing the average value of the epigenetic order parameter $< S >= \frac{n_m - n_u}{N}$ as a function of the persistence length. It is clear that the 3D result converges to the 1D result as the persistence length increases.

p when excluded volume conditions are taken into account [6, 7].

We characterize the global epigenetic state using $< S >$ for two values of k^+ as a function of l_p. As anticipated, the 3D model converges to the 1D limit (Figure S9) once the chain stiffness

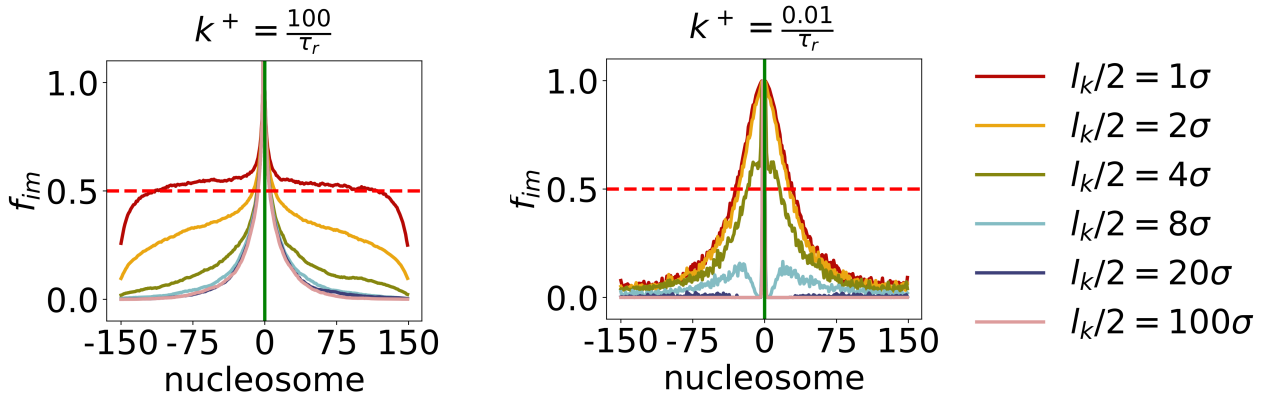


Figure S10: The effect of persistence length, $l_p = l_k/2$, on spreading for Scheme II depicted as the fraction f_{im} of modified state of nucleosome i. Scheme II is tested under fast, $k^+ = \frac{100}{\tau_r}$, and slow, $k^+ = \frac{0.01}{\tau_r}$, spreading conditions. We used $k^+/k^- = 50$ for fast spreading, and $k^+/k^- = 10,000$ for slow spreading, which are the same values used in the SI and the main text.

increases substantially ($\frac{l_p}{\sigma} \geq 8$). For instance, when spreading is fast ($k^+ = 100/\tau_r$) or slow ($k^+ = 0.01/\tau_r$), the value of $< S >$ for I and II mechanisms are roughly the same at $\frac{l_p}{\sigma} = 8$, as shown in Figure S9. Thus, regardless of the enzyme rates for modifying a nucleosome, the 1D and 3D mechanisms converge when l_p is sufficiently large. The reason is that in the stiff chain limit, the energy penalty to bend the polymer is high, thus effectively preventing the formation of looping contacts, which is needed for 3D spreading.

Upon closer observation of 3D spreading, the simulated domains in Figure S10 reveal that enhanced flexibility improves the propensity for stable domain formation. In particular, we observe a stabilizing effect of enhanced flexibility on the epigenetic pattern around the nucleation site. These results show that there is an interplay between 3D and 1D spreading, which is determined by L/l_p.

10 Finite size effects

Polymer length N affects relaxation timescales, which in turn changes polymer looping kinetics [8], with possible effects on the 3D spreading mechanism. We assessed the effects of changing N from (100-1000) for the fast-spreading mechanism. In the slow-spreading regime, the 1D spreading in Scheme **I** (left panel in Fig S11) shows very little local spreading due to low spreading probability at each time step. However, Fig S11 Scheme **II** reveals that the inactivation domain profiles behave identically at equal distances from the nucleation site, independent of the length of the polymer. This is because the nucleation site is a major contributor to domain formation, and 3D contact formation of the nucleation site with other residues determines the domain shape. The fraction of modification decreases with genomic distance from the nucleation site, underlined by a lower probability of contact formation at larger distances (Figure S6). Thus, similar domain shapes for chains of different lengths reflect a given probability contact scaling with genomic distance, which is constant in these simulations.

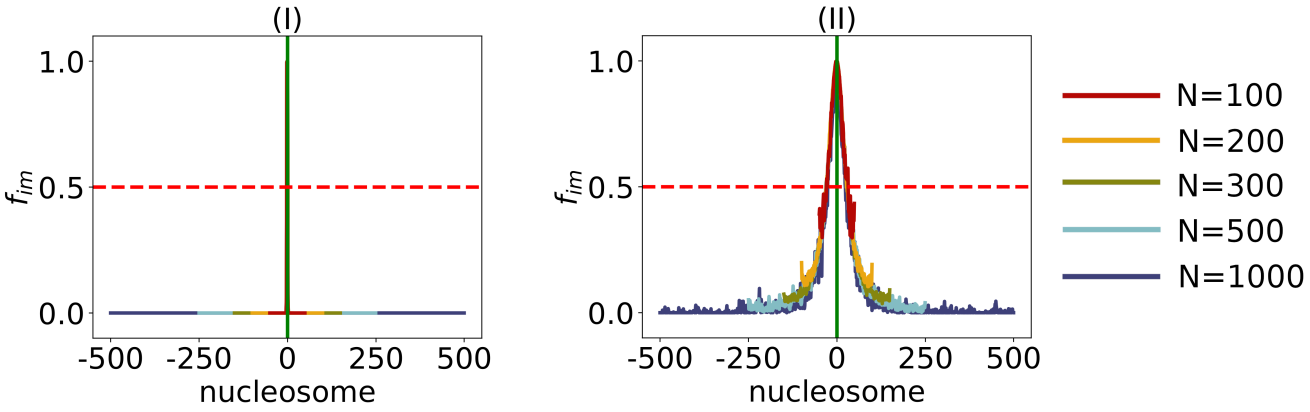


Figure S11: Tests for biological models as a function of locus length, N, with parameters $k^+ = \frac{0.01}{\tau_r}$, $k^+/k^- = 10,000$. The left (right) panel shows modification obtained using Scheme **I** (Scheme **II**). The modification probabilities for the Schemes are given in Figure S1.

10.1 The epigenetic switch

The figure Figure S12 shows that the interplay between the three parameters, k^+, k^-, and α with $\gamma = 0.03$, determines the global epigenetic state. For spreading events driven by the nucleation site, $\alpha = 1$, while for those triggered by modified nucleosomes (Ms), $\alpha < 1$. We consider the fast spreading case here in order to test parameter sensitivity, as it converges faster, and a wider range of parameters could be tested in a reasonable time. In this case, the spreading rate (k^+) is much faster than the polymer relaxation rate (τ_r^{-1}), $k^+ = 100/\tau_r$. The k^- value is a fraction of the forward rate, which means the reversal of M occurs at a slower rate compared to modification of the U state. We follow the modification status of all the nucleosomes over the course of the simulation time in order to assess if the chromatin is predominantly in the global active (unmarked with $\langle S \rangle \approx -1$) or inactive (marked with $\langle S \rangle \approx 1$) state.

Figure S12 (A) shows the global epigenetic state $\langle S \rangle$ as an average over 10 trajectories with different initial states, as a function of the rescaling parameter α, which modifies the forward rate distally of nucleation site (see main text). A range of α and k^+/k^- are probed, and the global epigenetic state $\langle S \rangle$ is computed for each parameter set. The data shown in subfigure (A) is then rearranged and shown in subfigure (B) as follows: (i) For a given k^+/k^-, α values are rescaled to $\alpha k^+/k^-$. This yields a different set of $\alpha k^+/k^-$ values for each k^+/k^-. (ii) With the new scale, $\alpha k^+/k^-$, only $< S >$ data that falls within the range of $\alpha k^+/k^- \in [0.5, 3.0]$ is taken into account. For each k^+/k^- ratio separately, the data is divided into bins with width 0.5. (iii) $< S >$ values are associated to $\alpha k^+/k^-$ bins. If multiple $< S >$ values are in the same, their average is computed. (iv) Each $\alpha k^+/k^-$ bin is associated with a single $< S >$ value, as shown Figure S12 (B).

Plots of the global epigenetic state and the nucleosome-dependent state in Figure S12 (B) allow us to draw a few pertinent conclusions. (i) Modifications are highly improbable if k^+/k^- is less than a certain value (≈ 10) regardless of the value of α. The chromatin remains in the global U state with $\langle S \rangle \approx -1$, regardless of whether the modifications occurs by Scheme **I** or Scheme **II** (see the left panels in Figure S12). In general, αk^+ has to exceed k^- for spreading

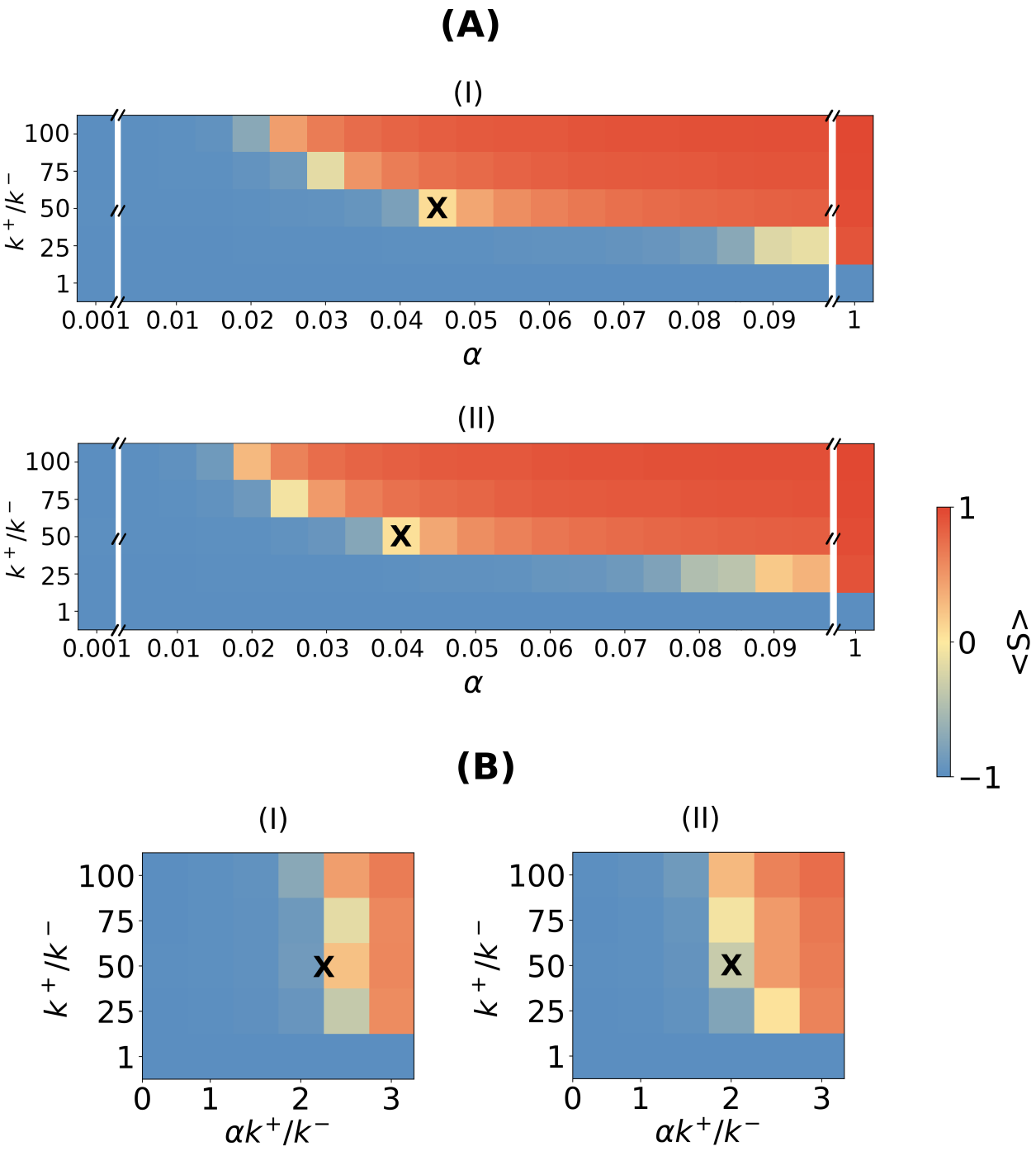


Figure S12: Epigenetic switching depends on modulation of the forward vs back rate distally from the nucleation site. Heatmaps show the overall epigenetic state, characterized by $\langle S \rangle$. Scheme **I** and Scheme **II** parameters used in main text are marked by **X**. (A) Epigenetic switching shown as a function of α, quantified using the mean value, $< S >$. (B) The global epigenetic state is a delicate balance between forward and backward rates, and the domain acts as an epigenetic switch (see the main text). The parameters are chosen to ensure that the global average $\langle S \rangle \approx 0$. In all the panels, $k^+ = 100/\tau_r$ (fast spreading), the total length of the simulations is $15,000\tau_r$, and $k^- = k^+/50$.

to occur. At higher values of k^+/k^- there is a global $U \rightarrow M$ transition as α is increased. (ii) The transition from a predominantly active to predominantly inactive state occurs over a narrow range of k^+/k^-, and α. For instance, for both mechanisms, **I** and **II**, the switch from $\langle S \rangle < 0$ to $\langle S \rangle > 0$ occurs over a narrow range of α, as is evident from Figure S12 (B). The critical point in the switch from active to inactive state is determined by k^+/k^-. The transition hinges on the competition between the forward (favors spreading) and the reverse reaction (inhibits spreading) distally from the NS, and is controlled by $\alpha k^+ \approx k^-$. Under this condition, the modification results in the concurrent formation of similar-sized patches containing U and M nucleosomes, resulting in $\langle S \rangle \approx 0$. This is most evident in spreading by mechanism **II** (Figure S12 (B)), showing a switch from active state (yellow color, $\langle S \rangle < 0$) to the inactive state (red color, $\langle S \rangle > 0$) through the mixed state (white squares, $\langle S \rangle \approx 0$). The delicate state of the domain, depending on a narrow $\alpha(k^+/k^-)$ range produces an epigenetic switch, reminiscent of gene control expression in cells, whereby cellular fate depends on the relative concentrations of the epigenetic regulators. Shifting the balance in one direction or the other could result in dramatically different outcomes in terms of gene expression patterns [9]. (iii) If $\alpha k^+/k^- < 1$ then the reverse reaction occurs predominantly by random histone turnover ('noise' in the list of transitions) that does not depend on the epigenetic identity of neighbors. On the other hand, if $\alpha k^+/k^- > 1$ the reverse reaction is facilitated due to positive feedback, which pushes the domain state towards the global U state. To mitigate this effect, stronger enhancement of the forward reaction is needed to achieve silencing levels comparable to the reverse noise dominated mechanism.

11 NS location

We investigated the effect of changing the location of the NS on the spreading process. Given that mechanistically spreading occurs bi-directionally from the NS along the chromatin chain, we expected similar results if the NS location is changed. The simulations confirm that it is the NS that drives spreading both in Schemes **I** and **II**. The results are similar to the ones in the main text except that the spreading profiles are centered around the NS (Figure S13).

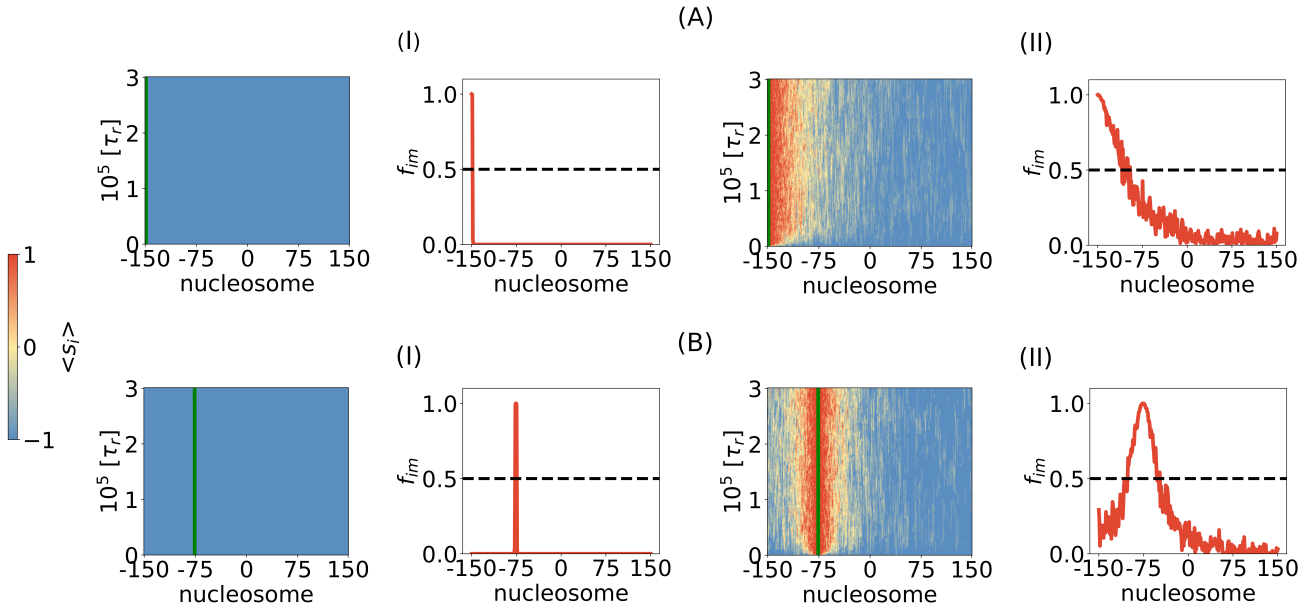


Figure S13: In (A) , the NS is in position $i = -150$, while in (B) it is in position $i = -75$. The relevant parameter values are $k^+ = \frac{0.01}{\tau_r}$, $k^+/k^- = 10,000$.

References

[1] William C Swope, Hans C Andersen, Peter H Berens, and Kent R Wilson. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *The Journal of Chemical Physics*, 76(1):637–649, 1982.

[2] Michael E Fisher. Shape of a self-avoiding walk or polymer chain. *The Journal of Chemical Physics*, 44(2):616–622, 1966.

[3] D Thirumalai, Raymond D Mountain, and TR Kirkpatrick. Ergodic behavior in supercooled liquids and in glasses. *Physical Review A*, 39(7):3563, 1989.

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

[5] Fabian Erdel and Eric C Greene. Generalized nucleation and looping model for epigenetic memory of histone modifications. *Proceedings of the National Academy of Sciences*, 113(29):E4180–E4189, 2016.

[6] J P Wittmer, H Meyer, J Baschnagel, A Johner, S Obukhov, L Mattioni, M Müller, and A N Semenov. Long range bond-bond correlations in dense polymer solutions. *Phys. Rev. Lett.*, 93(14):147801, October 2004.

[7] Hsiao-Ping Hsu, Wolfgang Paul, and Kurt Binder. Polymer chain stiffness versus excluded volume: A monte carlo study of the crossover towards the wormlike chain model. *Euro Phys. Lett.*, page 28003, 2010.

[8] Ngo Minh Toan, Greg Morrison, Changbong Hyeon, and D Thirumalai. Kinetics of loop formation in polymer chains. *J. Phys. Chem. B*, 112(19):6094–6106, May 2008.

[9] Rudolf Jaenisch and Adrian Bird. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature genetics*, 33(3):245–254, 2003.