

# Unfolding of the chromatin fiber driven by overexpression of bridging factors

Isha Malhotra, Bernardo Oyarzún, and Bortolo Matteo Mognetti

*Université Libre de Bruxelles (ULB),*

*Interdisciplinary Center for Nonlinear Phenomena and Complex Systems,*

*Campus Plaine, CP 231, Blvd. du Triomphe, B-1050 Brussels, Belgium \**

## Abstract

Nuclear molecules control the functional properties of the chromatin fiber by shaping its morphological properties. The biophysical mechanisms controlling how bridging molecules compactify the chromatin are a matter of debate. On the one side, bridging molecules could cross-link faraway sites and fold the fiber through the formation of loops. Interacting bridging molecules could also mediate long-range attractions by first tagging different locations of the fiber and then undergoing microphase separation. Using a coarse-grained model and Monte Carlo simulations, we study the conditions leading to compact configurations both for interacting and non-interacting bridging molecules. In the second case, we report on an unfolding transition at high densities of the bridging molecules. We clarify how this transition, which disappears for interacting bridging molecules, is universal and controlled by entropic terms. In general, chains are more compact in the case of interacting bridging molecules since, in this case, interactions are not valence-limited. However, this result is conditional on the ability of our simulation methodology to relax the system towards its ground state. In particular, we clarify how, unless using reaction dynamics that change the length of a loop in a single step, the system is prone to remain trapped in metastable, compact configurations featuring long loops.

---

\* Isha.Malhotra@ulb.ac.be, Bortolo.Matteo.Mognetti@ulb.be

## INTRODUCTION

Nuclear DNA in eukaryotic cells is scaffolded by histones and other proteins forming a fiber known as chromatin. A myriad of molecules, mainly proteins and ribonucleic acids (RNAs), regulate the morphological properties of the chromatin by selectively tagging and bridging specific loci of the genome [1–4]. Different conformations of the chromatin fiber result in different functional states. For instance, a change in the concentration of a bridging molecule can open or fold specific segments of the fiber, allowing or not polymerase enzymes to diffuse towards a promoter, and therefore silencing or activating particular genes.

Coarse-grained models have been employed to study the three-dimensional structure of the chromatin. A typical class of models uses bead-and-spring chains (see Fig. 1) in which monomers (which may represent a segment of chromatin containing multiple histones) interact through selective interactions [5–8] which could change as a result of chemical reactions triggered by epigenetic regulators [9]. On one side, such models are used to sample the physical properties of the system. Moreover, bead-and-spring models are also used to reconstruct the most likely conformation of the fiber compatible with experimental results in chromosome-capture techniques [10] measuring the euclidean proximity between different chromatin loci (as expressed by connectivity maps) [11, 12].

Simulation approaches are challenged by the multiscale nature of the system [13–15] as well by the necessity of sampling the large variability of the possible fiber configurations. In this paper, we address the latter issue and deploy an ensemble of Monte Carlo (MC) moves that allow changing the connectivity state of a chain in a single step by binding/unbinding two monomers while simultaneously regrowing a fraction of the chain. Using our simulation scheme, we study the morphological properties (compaction, connectivity map, and loop size distribution) of chromatin fibers cross-linked by bridging factors (see Fig. 1a). We study how the concentration of the bridging molecule ( $\rho_B$ ) and the affinity between bridging molecules and tagged (or reactive) monomers (quantified by the association constant,  $K_a$ ) affect the phase behavior of the system. Intriguingly, as foreseen in Ref. [16], for high values of  $\rho_B$  (and a given value of  $K_a$ ), the fiber unfolds given that all reactive monomers carry a linker, and bridges cannot form (see Fig. 1b). Simulation results are backed by scaling analysis and theory. The overexpression of a bridging factor can affect the phenotype (e.g. [17]). However, we are not aware of molecular studies of the chromatin fiber at high values of

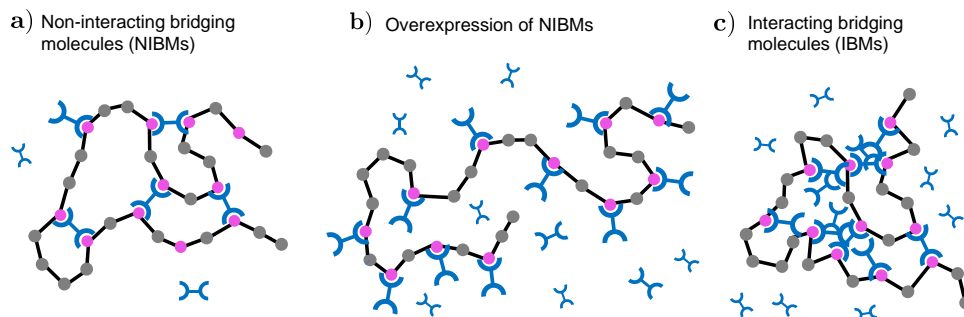


Figure 1. **a)** Non-interacting bridging molecules can bind tagged sites of the fiber (in purple) and drive compaction of the chromatin through the formation of loops. **b)** For high densities of the bridging molecules, each reactive monomer carries a B molecule, loops open, and the chromatin unfolds. **c)** Microphase separation of interacting B molecules can drive compaction of the fiber.

$\rho_B$ . The unfolding transition of Fig. 1b is somehow analogous to the resolubilization of virus crystals [18], proteins [19], or nucleic acids [20] at a concentration of multivalent ions (forming inter-molecular bridges) higher than some mM. Similarly, aggregates of ligand-presenting colloids cross-linked by short DNA oligomers dispersed in solution melt when increasing the concentration of the bridging molecule beyond a threshold value (ranging between  $10^{-8}$  M and  $10^{-4}$  M) which depends on the number of ligands per particle [21]. The proposed method, along with other simulation strategies implementing reaction moves in systems of multivalent chains [22–30], are ideal for studying the unfolding transition given that the high free energy barriers between competing states are sidestepped by dedicated MC moves. On the other hand, MC methods are not suitable to study the dynamics of the system unless the reaction timescales are much bigger than the backbone relaxation times [27].

Loop formation is not the only way of compactifying chromatin’s segments. Bridging molecules may interact through residual (e.g., multivalent) interactions [31, 32]. Importantly, while remaining soluble in the solution, the bridging molecules may condensate on the chromatin fiber and form finite-sized drops enveloping reactive monomers (see Fig. 1c) [16]. The condensation of bridging molecules results in effective interactions between reactive monomers, which are not valence limited, finally driving the folding of the fiber [33]. In the presence of interacting bridging molecules, we find that the re-entrant transition disappears. In particular, the fiber remains in a compact form as more bridges are found onto the chain.

This result only holds when comparing equilibrated states. When not changing the length of the loops with dedicated MC moves (which do not affect the equilibrium distribution of the system), chains folded by non-interacting bridges become much more compact. The results and methodology presented in this paper allow assessing the thermodynamics of competing mechanisms leading to domain formation in chromatin.

## METHODS

### The coarse-grained model

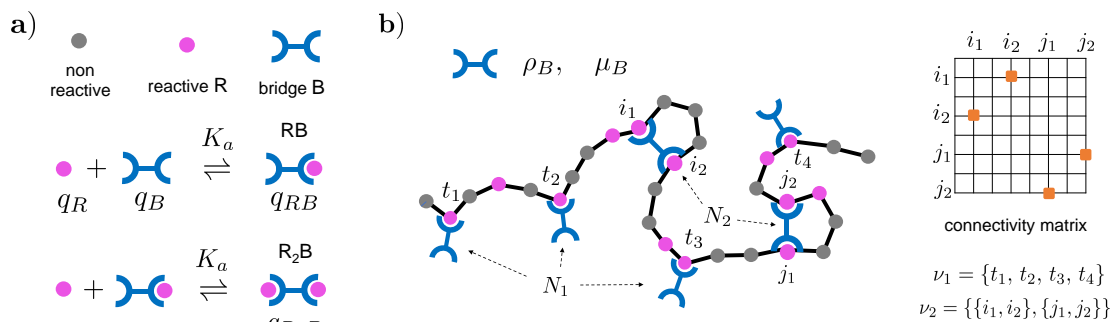


Figure 2. **a)** Formation of dimers (RB) and trimers (R<sub>2</sub>B) starting from reacting units (B molecules and R monomers) in diluted conditions. **b)** B molecules bind chains carrying R monomers and lead to the formation of loops.  $\mu_B$  and  $\rho_B$  are, respectively, the chemical potential and the density of the B molecules in solution.  $\nu_1$  is the list of R monomers forming an RB complex while  $\nu_2$  the list of pairs of monomers cross-linked by B molecules.  $\nu_2$  can be represented using connectivity maps that, when averaged, provide the likelihood of finding two monomers cross-linked by a B molecule.

Fig. 2 presents the model. We consider fully flexible freely jointed chains made by  $N$  monomers with bond length equal to  $\sigma$ .  $N_R$  monomers out of  $N$  can reversibly bind bridging (B) molecules dispersed in solution. We evenly distribute the reactive (R) monomers along the chain and define the degree of functionalization as  $f = N_R/N$ .  $N_1$  and  $N_2$  are, respectively, the number of R monomers carrying a bridge, not forming a loop, and the number of B molecules cross-linking two reactive monomers. In this study, we model bridging molecules implicitly without tracking their exact position in the system. Instead, after binding a bridge, an  $R$  monomer changes the way of interacting with the rest of the chain.

Binding/unbinding events between B and R are directly controlled by the chemical potential of the B molecules  $\mu_B$  and the association constant  $K_a$  (see below). For interacting bridges (see Fig. 1c), the attachment of B molecules to the fiber becomes cooperative.

The microstates of the system are then fully specified by the list of reactive monomers carrying B molecules ( $\nu_1$ ), the list of pairs of monomers cross-linked ( $\nu_2$ ), and the Euclidean positions of the  $N$  monomers  $\{r\} = \{\mathbf{r}_1, \mathbf{r}_2 \cdots \mathbf{r}_N\}$  (see Fig. 2). The partition function providing the statistical weight of each microstate reads as follows (if  $\beta = 1/k_B T$  where  $k_B$  is the Boltzmann's constant)

$$Z = \sum_{\nu_1, \nu_2} (q_R)^{N_R - N_1 - 2N_2} (q_{RB})^{N_1} \left( \frac{q_{R_2B}}{\Omega_0} \right)^{N_2} e^{\beta \mu_B (N_1 + N_2)} \mathcal{Z}_{\nu_1, \nu_2}, \quad (1)$$

$$\mathcal{Z}_{\nu_1, \nu_2} = \int d\{r\} e^{-\beta U_{\nu_1 \nu_2}(\{r\})} \prod_{\{i, j\} \in \nu_2} \delta(|\mathbf{r}_i - \mathbf{r}_j| - \sigma). \quad (2)$$

In Eq. 1,  $q_X$  ( $X = R, RB$ , and  $R_2B$ , see Fig. 2a) are the internal partition functions of the reactive monomers (respectively, when free, carrying a linker, and cross-linked). At the same time,  $\mathcal{Z}_{\nu_1, \nu_2}$  is the contribution accounting for the possible configurations of the chain backbone at a given  $\nu_1, \nu_2$ . A pair of monomers ( $i$  and  $j$ ) cross-linked by a bridging molecule ( $\{i, j\} \in \nu_2$ ) are constrained to stay at a fixed distance equal to  $\sigma$ . In Eq. 2, the integral over the chain's backbone is constrained by the fixed bond condition, while  $U_{\nu_1 \nu_2}$  specifies the interactions between non-linked monomers as detailed below. Using the language of Ref. [30], Eqs. 1, 2 define a probability distribution on a stratification: a hierarchy of nested manifolds in  $2(N - 1) - N_2 + 3$  dimensions (where 3 refers to the center of mass degrees of freedom).

As highlighted in a thread of investigations developing reaction ensemble Monte Carlo methods [34, 35], the internal partition functions are linked to the association equilibrium constant ( $K_a$ ) measured in diluted mixtures of B molecules and R monomers, the latter not constrained to stay on a chain (Fig. 2a). Specifically, the partition function of a single molecule of type  $X$  ( $X = R, B, RB$ , and  $R_2B$ ) is  $V \cdot q_X$ , where  $V$  is the volume of the system corresponding to the configurational space accessible to the center of mass of  $X$  treated as a classical variable. In diluted conditions, we can calculate the chemical potential ( $\mu_X$ ) from the partition function of  $N_X$  molecules as follows

$$Z_X(N_X) = \frac{(V q_X)^{N_X}}{\sigma_X N_X!} \Rightarrow \mu_X = k_B T \log \frac{\rho_X \sigma_X}{q_X} \quad (3)$$

where  $\rho_X$  is the density of the molecule  $X$  and  $\sigma_X$  the symmetry order of the coarse-grained representation of  $X$  ( $\sigma_R = \sigma_{RB} = 1$ ,  $\sigma_{R_2B} = 2$ , and  $\sigma_B = 1$  as we treat B molecules implicitly). We now consider the two reactions leading to the formation of RB and  $R_2B$  molecules:  $R+B \rightleftharpoons RB$  and  $RB+R \rightleftharpoons R_2B$ . In equilibrium conditions, the sum of the chemical potentials of the educts should be equal to the same quantity calculated for the products:  $\mu_R + \mu_B = \mu_{RB}$  and  $\mu_R + \mu_{RB} = \mu_{R_2B}$ . When using Eq. 3, these equations allow expressing the internal partition functions in term of the association constant  $K_a$  as follows

$$\frac{q_{RB}}{q_R q_B} = \frac{\rho_{RB}}{\rho_R \rho_B} \equiv K_a \quad \frac{q_{R_2B}}{q_{RB} q_R} = 2 \frac{\rho_{R_2B}}{\rho_{RB} \rho_R} \equiv 2K_a \quad (4)$$

where we assume that the two terminals of the B molecules react with the same strength when binding the first and the second R monomer. When considering reactive monomers tethered to a chain, we can use Eqs. 4 in Eq. 1 to calculate the statistical weight ( $\pi_{N'_1, N'_2}$ ) of binding a B molecule to the chain ( $N_1 \rightarrow N_1 + 1$ ) and of forming a loop ( $N_1 \rightarrow N_1 - 1$  and  $N_2 \rightarrow N_2 + 1$ ) relative to the one of a reference state ( $\pi_{N_1, N_2}$ )

$$\frac{\pi_{N'_1=N_1+1, N'_2=N_2}}{\pi_{N_1, N_2}} = \rho_B K_a \frac{\mathcal{Z}'}{\mathcal{Z}} \quad \frac{\pi_{N'_1=N_1-1, N'_2=N_2+1}}{\pi_{N_1, N_2}} = 2 \frac{K_a}{\Omega_0} \frac{\mathcal{Z}'}{\mathcal{Z}}. \quad (5)$$

The previous expressions clarify how  $\rho_B$  and  $K_a$  are the only parameters required to parametrize the model. Instead, the configurational contributions  $\mathcal{Z}'/\mathcal{Z}$  are sampled using Monte Carlo simulations of the coarse-grained model, as detailed in the next section.  $\Omega_0$  (see Eqs. 1, 5) is the configurational volume available to the orientational degree of freedom of an  $R_2B$  dumbbell ( $\Omega_0 = 4\pi\sigma^2$ ). In Eq. 1, we divide  $q_{R_2B}$  by  $\Omega_0$ , given that both  $\mathcal{Z}$  (Eq. 2) and  $q_{R_2B}$  (Eq. 3) account for all possible orientations of an  $R_2B$  molecule. Notice that  $\Omega_0$  does not divide  $q_{RB}$  in Eq. 1 as  $\mathcal{Z}$  does not depend on the orientation of the B molecules attached to the chain.

We now discuss the interactions between non-linked monomers  $U_{\nu_1 \nu_2}(\{r\})$  (Eq. 2). We decompose  $U_{\nu_1 \nu_2}(\{r\})$  into the sum of pair potentials,  $U_{\nu_1 \nu_2} = \sum_{i,j} u_{ij}(r_{ij})$ ,  $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ , where  $i$  and  $j$  are not linked. We consider precursors in good solvent regimes. In particular, reactive and non-reactive monomers interact through purely repulsive potential  $u_{i,j} = u^R$ . However, when considering interacting bridges (see Fig. 1c), interactions between reactive monomers bound to a B molecule (either belonging to an RB or  $R_2B$  complex) feature an attractive well modeled using a cut and shifted Lennard-Jones potential,  $u_{i,j} = u^A$ .

Specifically, we have

$$u^X(r) = \begin{cases} 4 \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^{12} \right] - 4 \left[ \left( \frac{\sigma}{\lambda_X} \right)^{12} - \left( \frac{\sigma}{\lambda_X} \right)^{12} \right] & r < \lambda_X \\ 0 & r > \lambda_X \end{cases} \quad (6)$$

with  $X = A, R$  and  $\lambda_R = 2.5\sigma$ ,  $\lambda_A = 2^{\frac{1}{6}}\sigma$ .

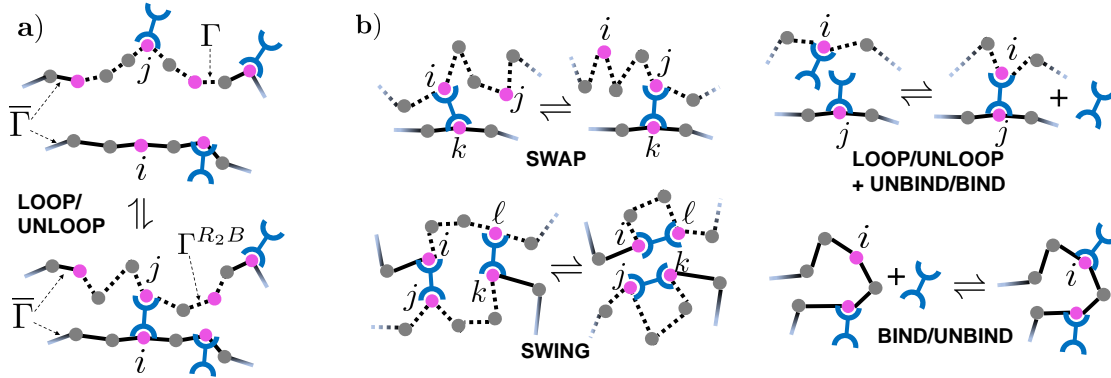


Figure 3. **a)** Loop/unloop MC moves react complementary monomers ( $i$  and  $j$ ) while updating a segment of the chain encompassing  $j$  (highlighted using dotted line). **b)** MC moves involving multiple reaction events or changing the number of B molecules attached to the chain.

### The simulation strategy

First, we consider the MC moves that change the connectivity state of the chain  $\nu_2$  by reacting complementary monomers. When attempting to form a new loop (for instance, by reacting monomers  $i$  with  $j$  in Fig. 3a), we start from deleting from the system a fraction of the chain ( $\Gamma$ ) in the proximity of  $j$  (dashed segment in the top of Fig. 3a). We then generate a new segment configuration ( $\Gamma^{R_2B}$ ) with  $i$  linked with  $j$  (Fig. 3a bottom). This process is done as proposed in [29] using methods growing chains with fixed endpoints [36–39]. In the reverse move, we delete a segment of the chain ( $\Gamma^{R_2B}$ ) containing a randomly selected monomer forming a loop ( $j$ ) and regrow a segment ( $\Gamma$ ) without any loop. Acceptance rates are calculated as done in Configurational Bias Monte Carlo [40–42]. Importantly, we calculate the relative statistical weights of the two configurations of Fig. 3a using the second of Eqs. 5. The acceptance rates become then a function of the association constant  $K_a$ . We refer to SM Sec. 2.3 for the details of the algorithm.

We also used MC moves that implement multiple binding/unbinding events simultaneously. In the Swap MC move (Fig. 3b), monomer  $k$  detaches from  $i$  and sequentially binds a second complementary monomer  $j$  [29]. In the Swing MC move (Fig. 3b), two pairs of reacted monomers exchange their partners [29]. In this work, we also use an MC move in which two reactive monomers carrying a B molecule react while simultaneously detaching one of the two bridges (loop/unloop+unbind/bind in Fig. 3b). B molecules are reversibly attached to the chain using the binding/unbinding moves (Fig. 3b). Contrary to the swing and the swap move, the binding/unbinding moves do not update the configuration of the backbone  $\{r\}$ . The details of the MC moves of Fig. 3b are reported in SM Sec. 2.2 (bind/unbind) and SM Secs. 2.4, 2.5, and 2.6 (respectively, for the loop/unloop+binding/unbinding, the swap, and the swing).

To further relax the system, we also employ standard MC moves that update the backbone of the polymer without affecting  $\nu_1$  and  $\nu_2$ . Specifically, we use pivot and double pivot moves that rotate fractions of the chains. Segments of the chain are also regrowth using the CBMC method (as used in the loop/unloop move) without changing the connectivity state of the system. A dedicated MC move displaces, and reorients reacted complexes ( $R_2B$ ) while regrowing a fraction of the surrounding network. In each MC cycle, we randomly perform one of the previous moves.

## Comparison with existing methodologies

Simulation schemes to study the statistical properties of chains carrying reactive units are finding applications in nanotechnology, for instance, to self-assemble polymeric nanoparticles [43]. Molecular Dynamics simulations have been used to study the morphological properties of the nanoparticles with an emphasis on finding protocols leading to maximal chain compaction [43, 44]. In most of these studies, cross-linking between complexes is irreversible. Refs. [26, 45] introduced a three-body potential that allows implementing swap-like moves in Molecular Dynamics simulations. Recently, Ref. [30] introduced a general MC scheme which allows adding/removing holonomic constraints reversibly. Inter-molecular, reversible linkers are also used to enforce topological entanglement when studying polymer melts using soft potentials [46]. In this respect, we note that the current version of our simulation method, the chains are crossable. On the one hand, the crossability of the chromatin fiber



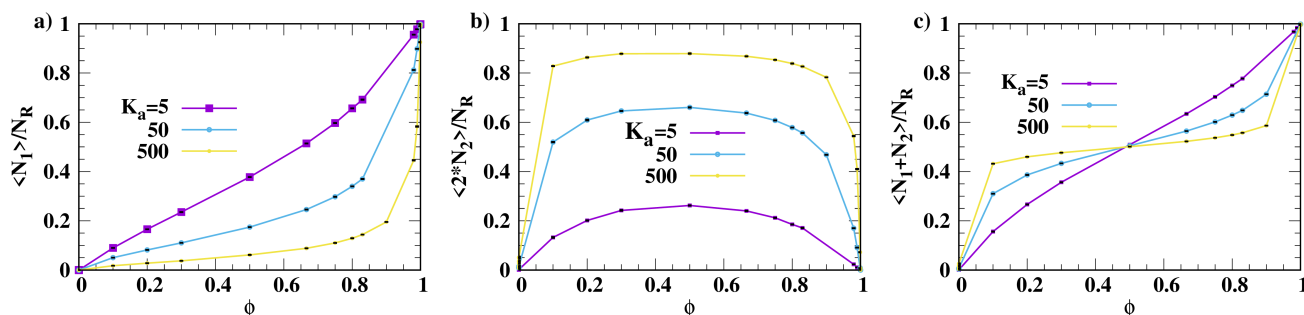


Figure 4. **a-c)** Fraction of reactive monomers carrying a linker, forming a loop, and number of B molecules attached to the chain *per* R monomer as a function of  $\phi$ .  $N = 1000$  and  $f = 0.5$ . In all cases, the errorbar is smaller than the size of the symbols.

is guaranteed by the action of dedicated enzymes. Secondly, topological constraints could be included in our methodology using linking numbers.

Chains carrying monomers featuring selective interactions are currently used to study condensation of multivalent proteins into membraneless bodies [22–24]. These contributions study the molecular determinants underlying the aggregation of multivalent proteins comprising short folded domains linked by intrinsically disordered regions. In particular, the LASSI package [24] employs lattice models in which folded domains are mapped into beads forming reversible linkages while disordered regions into strings of beads interacting through non-specific interactions. The lattice model is parametrized by atomistic simulations and is sequence-dependent [47]. Relevant to the findings of this work, non-specific interactions between linkers could lead to microphase separation in multicomponent systems [23] and affect the physical mechanism underlying chain aggregation [22]. The MC moves presented in the previous section could be readily adapted to the LASSI setting. Finally, simulations based on field-theoretic methods [48] have also been used to study supramolecular polymer physics [25].

## RESULTS AND DISCUSSION

### Non-Interacting Bridging Factors (NIBs): Monofunctional Chains

In this section, we quantify how the bridges' density ( $\rho_B$ ) and the association strength ( $K_a$ ) regulate the number of reactive monomers carrying linkers ( $N_1$ ) as well as the num-

ber of loops ( $N_2$ ). We then correlate  $N_1$  and  $N_2$  with the morphological properties of the chains, namely the gyration radius, the distribution of the loop length, and the averaged connectivity map.

When forbidding the formation of loops, the bindings of bridges to different reactive monomers become independent events, and the probability that a reactive monomer carries a B molecule becomes equal to (see the discussion after Eq. 8)

$$\phi = \frac{\langle N_1 \rangle_{\text{noloops}}}{N_R} = \frac{K_a \rho_B}{1 + K_a \rho_B}, \quad (7)$$

where  $\langle \dots \rangle_{\text{noloops}}$  denotes a constrained average with  $N_2 = 0$  as obtained using Eq. 1. Motivated by this observation, in the following, we use the variable  $\phi$  to discuss our results. Fig. 4a and SM Fig. S1a study the fraction of reactive monomers carrying a linker without forming a loop ( $\langle N_1 \rangle / N_R$ ) as a function of  $\phi$  when changing  $K_a$  (Fig. 4a) and  $f$  (SM Fig. S1a). As expected,  $\langle N_1 \rangle$  increases with  $\phi$  from  $N_1 = 0$  (for  $\phi = 0$ ) to  $N_1 = N_R$  (for  $\phi = 1$ ). When increasing the association constant,  $\langle N_1 \rangle$  starts to deviate from the linear behavior predicted by Eq. 7 due to loop formation. In particular, at a given  $\phi$ , loops become more favorable at high values of  $K_a$  given that the formation of loops at a given  $N_1 + N_2$  is only controlled by the association constant and free-energy terms (discussed below) which do not depend on  $\rho_B$  (see Eq. 1). In Fig. 4b and SM Fig. S1b, we study the number of loops as a function of  $\phi$ . As anticipated above,  $\langle N_2 \rangle$  increases with  $K_a$  (Fig. 6b) and the degree of functionalization  $f$  (if  $f \leq 0.5$ , see SM Fig. S1b). Importantly, the number of loops is non-monotonic in  $\phi$  and goes to zero when  $\phi$  tends towards  $\phi = 1$ . This observation underlies the unfolding of the chain when overexpressing bridging molecules. Entropic terms regulate the opening of a loop in favors of two R monomers carrying two bridging molecules. The process is purely entropic as the number of reacted monomers (and therefore the number of  $K_a$  terms entering the Boltzmann distribution, see Eq. 1) in the two competing microstates does not change. For high values of  $\rho_B$ , the system attempts to maximize the number of bridging molecules present on the chain, therefore opening loops.

Intriguingly, for all values of  $K_a$  and  $f$  reported in Fig. 4b and SM Fig. S1b, the plots of  $\langle N_2 \rangle$  as a function of  $\phi$  are symmetric with respect to the axis  $\phi = 1/2$ . To understand this result and make new predictions, in Fig. 5, we present a pathway to estimate the free energy of two systems at different  $\rho_B$  (corresponding to a given  $\phi$  and  $\phi' = 1 - \phi$  with  $\phi' > \phi$ ) and same  $K_a$ . Taking as a reference state a polymer without any B molecule, we decompose the

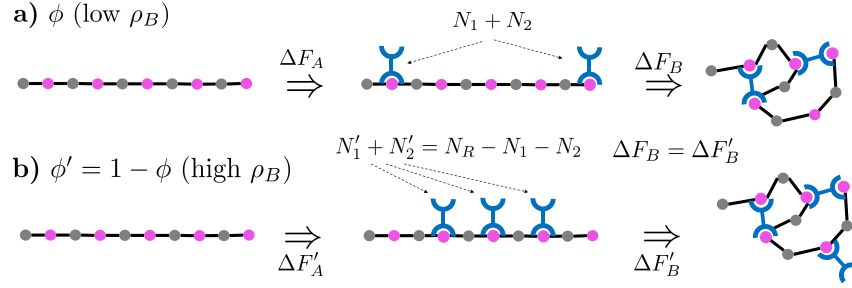


Figure 5. A two-step pathway to calculate the free energy of the system with a given  $N_1$  and  $N_2$ .  $\Delta F_A$  is the free energy of attaching  $N_1 + N_2$  B molecules to the fiber, while  $\Delta F_B$  accounts for the loops' contribution to the free energy.

free energy of the system into the contribution of putting  $N_1 + N_2$  B molecules on the chain ( $\Delta F_A$ ) and the one arising from making  $N_2$  loops ( $\Delta F_B$ ) without attaching any extra linker to the polymer.  $\Delta F_A$  can be calculated as

$$\beta \Delta F_A = -\log \left[ \binom{N_R}{N_1 + N_2} (K_a \rho_B)^{N_1 + N_2} \right]. \quad (8)$$

In Eq. 8, the binomial term counts the ways of distributing  $N_1 + N_2$  bridges within  $N_R$  R monomers, and  $K_a \rho_B$  is the statistical weight of each microstate as obtained using Eq. 1 and Eqs. 4. (Notice that Eq. 7 follows from Eq. 8.) Using Stirling's approximation, we obtain

$$\beta \Delta F_A = (N_1 + N_2) \log \frac{N_1 + N_2}{N_R} + (N_R - N_1 - N_2) \log \frac{N_R - N_1 - N_2}{N_R} - (N_1 + N_2) \log(K_a \rho_B). \quad (9)$$

The calculation of  $\Delta F_B$  is not feasible. Mean-field estimates of  $\Delta F_B$  are also tricky because, in contrast to systems with ligand-presenting colloids, reactions between complementary monomers can hardly be treated as independent events. Even though analytic expressions of this term are not available, we can prove that  $\Delta F_B$  is the same for the two systems considered in Fig. 5. This claim follows from the fact that, in the two systems, the numbers of R, empty monomers and R monomers carrying a bridge are exchanged. It follows that the combinatorial and configurational terms entering the calculation of  $\Delta F_B$  are the same in the two cases, as B molecules do not affect the local morphology of the chain. Therefore, in general, we can write the following equality

$$\Delta F_B(K_a, N_2, N_1 + N_2) = \Delta F_B(K_a, N_2, N_R - N_1 - N_2). \quad (10)$$

The previous equation explains the reason why  $\langle N_2 \rangle$  is symmetric with respect to the axis  $\phi = 1/2$ : At a given value of  $N_1 + N_2$ , the two terms of Eq. 10 are minimized by the same value of  $N_2$ .

We can push our analysis a step further and predict that the total number of linkers on the chain ( $\langle N_1 + N_2 \rangle$ ) is an antisymmetric function of  $\phi$ . The most likely number of  $N_1$  and  $N_2$  follows from saddle point equations

$$\frac{\partial(\Delta F_A + \Delta F_B)}{\partial N_1} = 0 \quad \frac{\partial(\Delta F_A + \Delta F_B)}{\partial N_2} = 0. \quad (11)$$

When developing the first of these equations using Eq. 8, we obtain

$$\frac{\partial(\Delta F_A + \Delta F_B)}{\partial N_1} = 0 \Leftrightarrow \log \frac{N_1 + N_2}{N_R - N_1 - N_2} + \frac{\partial \Delta F_B}{\partial N_1} = \log(K_a \rho_B) = \log \frac{\phi}{1 - \phi}, \quad (12)$$

where we use Eq. 7 to express  $K_a \rho_B$  in terms of  $\phi$ . Under the transformations  $N_1 + N_2 \rightarrow N_R - N_1 - N_2$  and  $\phi \rightarrow 1 - \phi$ , all the terms of the second equation change sign. It follows that  $\langle N_1 + N_2 \rangle$  is an antisymmetric function of  $\phi$  along the axes  $\phi = 1/2$  and  $N_1 + N_2 = N_R/2$ . We verify this prediction in Figs. 4c, SM Fig. S1c using simulation results. In particular, the total number of B molecules on the chain increases with  $\phi$ . Similar to Fig. 4c and SM Fig. S1c, when increasing the values of  $K_a$  and  $f$ , the plots deviate from a linear behavior (corresponding to  $N_2 = 0$ ). For  $\phi < 1/2$ ,  $\langle N_1 + N_2 \rangle > \phi$  as two R monomers can cooperatively stabilize a B molecule through the formation of a loop. Instead, for  $\phi > 1/2$ ,  $\langle N_1 + N_2 \rangle < \phi$  as multiple B molecules are shared by pairs of R monomers. When  $\phi \rightarrow 0$  and  $\phi \rightarrow 1$ , the behavior of  $\langle N_1 + N_2 \rangle$  is dominated, respectively, by  $\langle N_2 \rangle$  and  $N_R - \langle N_2 \rangle$ .

Importantly the previous arguments rely on the fact that we did not consider non-specific interactions between linker molecules and the chain's backbone. For instance, in the presence of steric interactions, Eq. 10 (and similarly Eq. 8) would not be valid as more bridges on the chain would increase the configurational cost of folding chains. Although we expect the general trend to remain unaffected by non-specific interactions, the simulation methodology can easily be adapted to account for these terms.

In Fig. 6a, we study the averaged squared radius of gyration  $\langle R_g^2 \rangle$  as a function of  $\phi$  for  $f = 0.5$ . We observe that  $\langle R_g^2 \rangle$  is non-monotonic in  $\phi$  with the chains that reswell when the system tends towards  $\phi = 1$ . This behavior mirrors the trends observed for the number of bridges (see Fig. 4b) and clarifies how intra-molecular linkages drive the compaction of the fiber. In particular,  $\langle R_g^2 \rangle$  decreases when increasing the value of  $K_a$  as more loops become

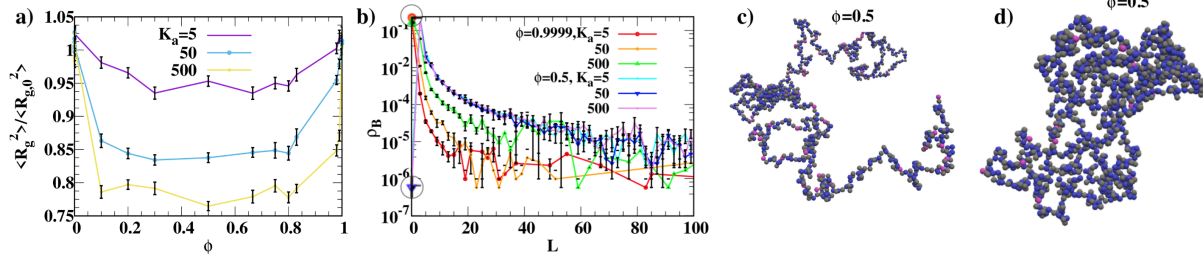


Figure 6. **a)** Averaged radius of gyration  $\langle R_g^2 \rangle$  of chains folded by NIBMs as a function of  $\phi$ .  $\langle R_{g,0}^2 \rangle$  is the averaged gyration radius in a system without B molecules. **b)** Loop length distribution for different values of  $K_a$  and  $\phi$ . In **a)** and **b)**, we use  $N = 1000$ ,  $f = 0.5$ , and calculate errorbars using 50 independent simulations consisting of  $5 \cdot 10^5$  MC cycles. **c)** and **d)** are snapshots with  $K_a = 500$ , when, respectively, using and not using swap and swing MC moves (Fig. 3b). Grey, pink, and blue beads represent, respectively, non reactive monomers, not-looped R monomers, and looped monomers.

stable. In Fig. 6b, we report the probability of forming a loop made by  $L$  segments for different values of  $K_a$  and  $\phi$ . The  $L = 0$  value refers to the probability for an R monomer to be unlooped. Consistently with Fig. 6a, increasing  $K_a$  decreases the probability of finding monomers unpaired. Moreover, the system attempts to minimize the length of loops as a result of the higher configurational cost of forming longer loops. As a result, the loop length distribution sharply decreases with  $L$ . Longer loops are expected when reducing the non-specific repulsion between monomers ( $u^R$  in Eq. 6). However, the system can feature persistent longer loops unless employing MC moves that change the length of a loop in a single step [29]. This result is shown in Fig. 6c and 6d, where we report two fiber structures when using (Fig. 6c) or not (Fig. 6d) the swap and the swing MC move (Fig. 3b). In the second case, we obtain a much more compact structure with longer loops that persist during the simulation. Longer loops will also occur when using semiflexible backbones. However, we stress how the unfolding of the fiber at high  $\rho_B$  is not model dependent and will be found in all systems featuring non-interacting bridging molecules. Consistently with the fact that thermodynamic states only feature short loops, SM Figs. S2, S3 show how the results of Figs. 4, 6a are not affected by the chain's length.

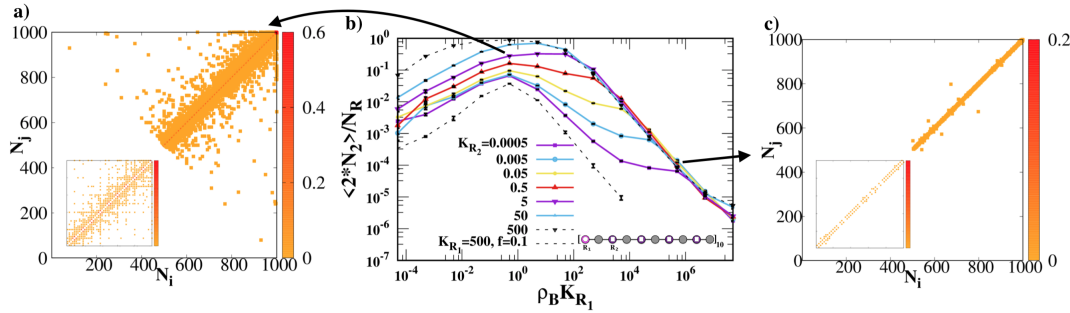


Figure 7. **a)** and **c)** Averaged connectivity maps of bifunctional chains with  $f = 0.5$  and  $K_{R_2} = 5$ . In **a)**,  $\rho_B K_{R_1} = 0.5$ , while in **c)**  $\rho_B K_{R_1} = 5 \cdot 10^4$ . The inset shows the connectivity map of the segment made of the first 100 beads. **b)** Fraction of reactive beads forming a loop as a function of  $\rho_B$  at different  $K_{R_2}$  for  $K_{R_1} = 500$ .

### Non-Interacting Bridging Factors (NIBs): Bifunctional Chains

In this section, we consider chains carrying two types of R monomers ( $R_1$  and  $R_2$ ) reacting with a single type of B molecules with different affinities (defined as  $K_{R_1}$  and  $K_{R_2}$ ). We aim to study if some degree of dispersity in the specific interactions between R monomers and B molecules could broaden the range of  $\rho_B$  corresponding to compact chains. We consider chains made of repeated blocks in which half of the monomers are reactive. Each block comprises one monomer of type  $R_1$  and four of type  $R_2$ , interposed with inert monomers (see inset of Fig. 7b). In Fig. 7b, we report the fraction of reactive monomers forming a bridge as a function of  $\rho_B$  while keeping the association constant of the strongest monomer ( $R_1$ ) equal to  $K_{R_1} = 500$ . We change the association constant between B and  $R_2$  ( $K_{R_2}$ ) from  $K_{R_2} = 0$  to  $K_{R_2} = 500$ . The two limiting cases (dashed lines in Fig. 7b) correspond to monofunctional systems with, respectively,  $f = 1/10$  ( $R_2$  monomers stop being reactive) and  $f = 1/2$  ( $K_{R_1} = K_{R_2}$ ).

We recover the same phenomenology found in the monofunctional case. In particular, intramolecular bridges disappear, both for large and small values of  $\rho_B$ . An inspection of the connectivity maps (see Fig. 7a) shows how, for intermediate values of  $\rho_B$ , the chains predominantly form loops between  $R_1$  and  $R_2$  monomers. Reacting  $R_1$  with  $R_2$  allows minimizing the configurational costs of forming long loops while prioritizing the binding of B molecules to  $R_2$  monomers. When increasing  $\rho_B$ , the system attempts to maximize the number of

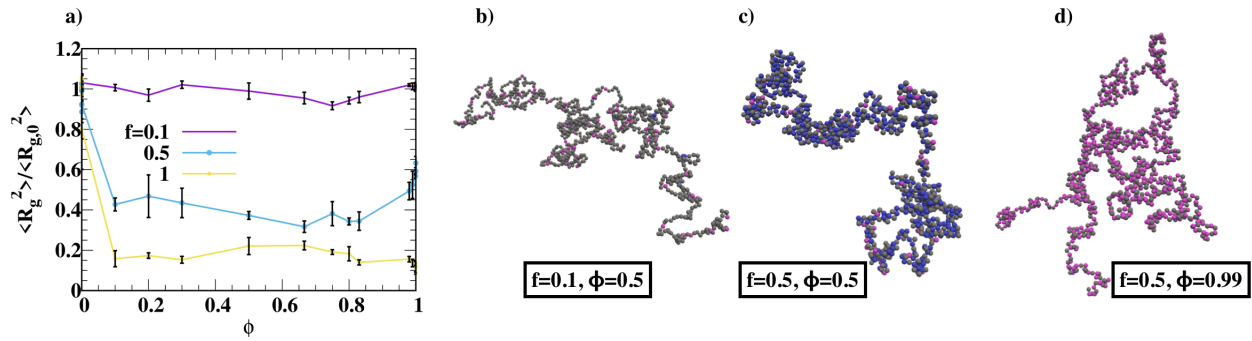


Figure 8. **a)** Averaged radius of gyration  $\langle R_g^2 \rangle$  of chains folded by IBMs as a function of  $\phi$  for  $K_a = 500$ .  $\langle R_{g,0}^2 \rangle$  is the averaged gyration radius in a system without bridging molecules. We have calculated error bars using 50 independent simulations consisting of  $5 \cdot 10^5$  MC cycles. **b)-d)** Typical snapshots for  $K_a = 500$  and different  $f, \phi$ . Grey, pink and blue beads represent, respectively, non reactive monomers, not-looped R monomers, and looped monomers.

B molecules on the chains. In these conditions, the connectivity maps (see Fig. 7c) do not feature the texture typical of copolymer architectures (like in Fig. 7a). In particular, the few short loops involve with equal probability  $R_1$  and  $R_2$ . Interesting, all curves (with  $K_{R_2} > 0$ ) collapse onto the monofunctional case (with  $f = 0.5$ ) in the large  $\rho_B$  limit. This general finding clarifies how purely entropic terms control the unfolding transition. However, the unfolding transition is not entropic in the case of bridging molecules formed by dimers (like in the case of associating YY1 [49] or CTFC binding factors [50]). Indeed, in the latter case, two passivated R monomers would carry two pairs of dimers as compared to a single dimer entering a loop. Therefore, we predict how large dimerization constants would accelerate the unfolding transition.

### Interacting Bridging Molecules (IBMs)

In this section, we consider the case in which B molecules interact as due, for instance, to multivalent interactions [16]. Here, NIBMs can condensate around multiple R sites that can therefore cluster. Following the strategy of Ref. [9], we drive the clustering of R monomers carrying B molecules using the LJ potentials introduced in Eq. 6.

Fig. 8a reports the averaged gyration radius of the IBM model as a function of  $\phi$  for different



degrees of functionalization  $f$ . As compared to Fig. 6,  $\langle R_g^2 \rangle$  rapidly decreases with  $\phi$  and then exhibits a plateau for  $\phi \gtrsim 0.1$ . The sharp decrease of  $\langle R_g^2 \rangle$  in  $\phi$  is due to a cooperative effect in which B molecules are colocalized by R monomers and stabilized by their mutual interaction. We verify this claim in SM Fig. S4, where we compare the fraction of R monomers carrying a B molecule in the NIBM and the IBM model. More B molecules are found on the chain in the case of the IBM since, in this model, R monomers can simultaneously interact with multiple partners. This observation also explains the fact that chains are more compact in the NIBM model (compare Fig. 8a with Fig. 6a). Similar findings have been reported in synthetic systems, where scholars struggled to self-assemble compact polymeric nanoparticles using intra-molecular bridges [43].

Fig. 8a does not exhibit the unfolding transition, and the chains stay compact even for  $\phi = 1$ . This result follows from the fact that monomers carrying B molecules are not passivated but can still interact as a result of the interactions between B molecules. This result is confirmed by snapshots showing how for  $f = 0.5$  and different values of  $\phi$ , the morphology of the fiber is comparable (see Fig. 8c and 8d and SM Fig. S5 for the same configurations in which we only report beads bound to a B molecule). On the other hand, smaller values of  $f$  increase the size of the chain as R monomers interact with fewer partners (see Fig. 8a and 8b).

## CONCLUSION

We presented a chromatin model consisting of a chain carrying reactive monomers folded by bridging (B) molecules dispersed in solution (at a fixed density,  $\rho_B$ ) reversibly attaching the chromatin fiber. We studied the effect of  $\rho_B$  and the association constant between B molecules and the reactive monomers on the morphological properties of the chain. For intermediate values of  $\rho_B$ , the chromatin folds due to the formation of loops comprising B molecules cross-linking two reactive monomers. Instead, overexpression of bridging factors can lead to the unfolding of the fiber. We highlighted the generality of this unfolding transition using thermodynamic considerations. Importantly, the unfolding transition is peculiar to non-interacting, bridging molecules. Instead, for models featuring B molecules that tempt to phase separate when colocalized on the fiber, chains are much more compact (even for high values of  $\rho_B$ ) as a result of the fact that interactions between reactive monomers are not valence-limited.



Our results are supported by Monte Carlo methodologies, which, as compared to Molecular Dynamics simulations, allow reconfiguring the topology of the networks through dedicated moves, therefore enabling efficient sampling between configurations featuring different connectivity states. Equilibrium configurations feature almost exclusively short loops. Short loops are thermodynamically stable because they minimize excluded volume interactions and configurational costs. However, we noticed that it is somehow difficult to relax the system towards the ground state. In particular, longer loops (resulting in much more compact chains) will persist unless using dedicated MC moves implementing multiple reactions in a single step.

The results presented in this paper allow inferring the microscopic mechanisms driving compaction from experimental results.

## AUTHOR CONTRIBUTIONS

BMM, IM and BO designed the research. BO and IM developed the program. IM carried out all simulations, analyzed the data. IM and BMM wrote the article.

## ACKNOWLEDGMENTS

We thank K. Rippe and F. Erdel for pointing our attention towards Ref. [17] and for suggesting the study of Fig. 7. Financial support was provided by an A.R.C grant of the *Fédération Wallonie-Bruxelles*. Computational resources have been provided by the Consortium des Equipements de Calcul Intensif (CECI), funded by F.R.S.-FNRS under Grant No. 2.5020.11. The authors declare no conflict of interest.

- 
- [1] S. J. McBryant, V. H. Adams, and J. C. Hansen, *Chromosome Research* **14**, 39 (2006).
  - [2] A. J. Ruthenburg, H. Li, D. J. Patel, and C. D. Allis, *Nature reviews Molecular cell biology* **8**, 983 (2007).
  - [3] S. A. Lambert, A. Jolma, L. F. Campitelli, P. K. Das, Y. Yin, M. Albu, X. Chen, J. Taipale, T. R. Hughes, and M. T. Weirauch, *Cell* **172**, 650 (2018).

- [4] K. Müller-Ott, F. Erdel, A. Matveeva, J.-P. Mallm, A. Rademacher, M. Hahn, C. Bauer, Q. Zhang, S. Kaltofen, G. Schotta, *et al.*, Molecular systems biology **10** (2014).
- [5] M. Barbieri, M. Chotalia, J. Fraser, L.-M. Lavitas, J. Dostie, A. Pombo, and M. Nicodemi, Proceedings of the National Academy of Sciences **109**, 16173 (2012).
- [6] D. Jost, P. Carrivain, G. Cavalli, and C. Vaillant, Nucleic acids research **42**, 9553 (2014).
- [7] M. Di Pierro, B. Zhang, E. L. Aiden, P. G. Wolynes, and J. N. Onuchic, Proceedings of the National Academy of Sciences **113**, 12168 (2016).
- [8] G. Tiana and L. Giorgetti, *Modeling the 3D conformation of genomes* (CRC Press, 2019).
- [9] D. Michieletto, M. Chiang, D. Coli, A. Papantonis, E. Orlandini, P. R. Cook, and D. Marenduzzo, Nucleic acids research **46**, 83 (2018).
- [10] S. S. Rao, M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov, J. T. Robinson, A. L. Sanborn, I. Machol, A. D. Omer, E. S. Lander, *et al.*, Cell **159**, 1665 (2014).
- [11] S. Wettermann, M. Brems, J. Siebert, G. Vu, T. Stevens, and P. Virnau, Computational Materials Science **173**, 109178 (2020).
- [12] F. Serra, M. Di Stefano, Y. G. Spill, Y. Cuartero, M. Goodstadt, D. Baù, and M. A. Marti-Renom, FEBS letters **589**, 2987 (2015).
- [13] Q. MacPherson, B. Beltran, and A. J. Spakowitz, Proceedings of the National Academy of Sciences **115**, 12739 (2018).
- [14] O. Perišić, R. Collepardo-Guevara, and T. Schlick, Journal of molecular biology **403**, 777 (2010).
- [15] R. Collepardo-Guevara, G. Portella, M. Vendruscolo, D. Frenkel, T. Schlick, and M. Orozco, Journal of the American Chemical Society **137**, 10205 (2015).
- [16] F. Erdel and K. Rippe, Biophysical journal **114**, 2262 (2018).
- [17] C. Koesters, B. Unger, I. Bilic, U. Schmidt, S. Bluml, B. Lichtenberger, M. Schreiber, J. Stockl, and W. Ellmeier, Immunology letters **109**, 165 (2007).
- [18] R. Asor, O. Ben-nun Shaul, A. Oppenheim, and U. Raviv, ACS nano **11**, 9814 (2017).
- [19] F. Zhang, M. Skoda, R. Jacobs, S. Zorn, R. A. Martin, C. Martin, G. Clark, S. Weggler, A. Hildebrandt, O. Kohlbacher, *et al.*, Physical review letters **101**, 148101 (2008).
- [20] T. T. Nguyen, I. Rouzina, and B. I. Shklovskii, The Journal of chemical physics **112**, 2562 (2000).

- [21] J. Lowensohn, B. Oyarzún, G. N. Paliza, B. M. Mognetti, and W. B. Rogers, *Physical Review X* **9**, 041054 (2019).
- [22] T. S. Harmon, A. S. Holehouse, M. K. Rosen, and R. V. Pappu, *Elife* **6**, e30294 (2017).
- [23] T. S. Harmon, A. S. Holehouse, and R. V. Pappu, *New Journal of Physics* **20**, 045002 (2018).
- [24] J.-M. Choi, F. Dar, and R. V. Pappu, *PLoS computational biology* **15** (2019).
- [25] G. H. Fredrickson and K. T. Delaney, *The Journal of chemical physics* **148**, 204904 (2018).
- [26] F. Sciortino, *The European Physical Journal E* **40**, 3 (2017).
- [27] S. H. Sandholtz, B. G. Beltran, and A. J. Spakowitz, *Journal of Physics A: Mathematical and Theoretical* **52**, 434001 (2019).
- [28] B. Oyarzún and B. M. Mognetti, *The Journal of chemical physics* **148**, 114110 (2018).
- [29] B. Oyarzún and B. M. Mognetti, *Molecular Physics* **116**, 2927 (2018).
- [30] M. Holmes-Cerfon, arXiv preprint arXiv:1912.10176 (2019).
- [31] A. G. Larson, D. Elnatan, M. M. Keenen, M. J. Trnka, J. B. Johnston, A. L. Burlingame, D. A. Agard, S. Redding, and G. J. Narlikar, *Nature* **547**, 236 (2017).
- [32] A. R. Strom, A. V. Emelyanov, M. Mir, D. V. Fyodorov, X. Darzacq, and G. H. Karpen, *Nature* **547**, 241 (2017).
- [33] C. A. Brackley, S. Taylor, A. Papantonis, P. R. Cook, and D. Marenduzzo, *Proceedings of the National Academy of Sciences* **110**, E3605 (2013).
- [34] C. H. Turner, J. K. Brennan, M. Lísal, W. R. Smith, J. K. Johnson, and K. E. Gubbins, *Molecular Simulation* **34**, 119 (2008).
- [35] T. W. Rosch and E. J. Maginn, *Journal of Chemical Theory and Computation* **7**, 269 (2011).
- [36] M. Dijkstra, D. Frenkel, and J. P. Hansen, *J. Chem. Phys.* **101**, 3179 (1994).
- [37] C. D. Wick and J. I. Siepmann, *Macromolecules* **33**, 7207 (2000).
- [38] B. Chen and J. I. Siepmann, *The Journal of Physical Chemistry B* **104**, 8725 (2000).
- [39] F. A. Escobedo and J. J. de Pablo, *The Journal of Chemical Physics* **102**, 2636 (1995).
- [40] J. I. Siepmann and D. Frenkel, *Molecular Physics* **75**, 59 (1992).
- [41] G. C. A. M. Mooij, D. Frenkel, and B. Smit, *Journal of Physics: Condensed Matter* **4**, L255 (1992).
- [42] D. Frenkel and B. Smit, *Understanding molecular simulation: from algorithms to applications*, Vol. 1 (Academic press, 2001).

- [43] A. J. Moreno and F. Lo Verso, Single-Chain Polymer Nanoparticles: Synthesis, Characterization, Simulations, and Applications (2017).
- [44] A. J. Moreno, F. Lo Verso, A. Sanchez-Sanchez, A. Arbe, J. Colmenero, and J. A. Pomposo, *Macromolecules* **46**, 9748 (2013).
- [45] S. Ciarella and W. G. Ellenbroek, arXiv preprint arXiv:1912.08569 (2019).
- [46] A. Ramírez-Hernández, B. L. Peters, L. Schneider, M. Andreev, J. D. Schieber, M. Müller, and J. J. de Pablo, *The Journal of chemical physics* **146**, 014903 (2017).
- [47] K. M. Ruff, T. S. Harmon, and R. V. Pappu, *The Journal of chemical physics* **143**, 12B607\_1 (2015).
- [48] G. Fredrickson *et al.*, *The equilibrium theory of inhomogeneous polymers*, Vol. 134 (Oxford University Press on Demand, 2006).
- [49] A. S. Weintraub, C. H. Li, A. V. Zamudio, A. A. Sigova, N. M. Hannett, D. S. Day, B. J. Abraham, M. A. Cohen, B. Nabet, D. L. Buckley, *et al.*, *Cell* **171**, 1573 (2017).
- [50] C.-T. Ong and V. G. Corces, *Nature Reviews Genetics* **15**, 234 (2014).