Polymer models of the hierarchical folding of the Hox-B chromosomal locus

Carlo Annunziatella, Andrea M. Chiariello, Simona Bianco, and Mario Nicodemi*
Dipartimento di Fisica, Universitá di Napoli Federico II, INFN Napoli, CNR, SPIN, Complesso Universitario di Monte Sant'Angelo, 80126 Naples, Italy
(Received 6 June 2016; published 4 October 2016)

As revealed by novel technologies, chromosomes in the nucleus of mammalian cells have a complex spatial organization that serves vital functional purposes. Here we use models from polymer physics to identify the mechanisms that control their three-dimensional spatial organization. In particular, we investigate a model of the Hox-B locus, an important genomic region involved in embryo development, to expose the principles regulating chromatin folding and its complex behaviors in mouse embryonic stem cells. We reconstruct with high accuracy the pairwise contact matrix of the Hox-B locus as derived by Hi-C experiments and investigate its hierarchical folding dynamics. We trace back the observed behaviors to general scaling properties of polymer physics.

DOI: 10.1103/PhysRevE.94.042402

I. INTRODUCTION

In the nucleus of mammalian cells chromosomes have a complex, nonrandom spatial organization [1–5]. Their architecture is intimately linked to genome biological function and its disruption is often related to a variety of diseases, including congenital diseases and cancer [6–8]. New technologies, such as Hi-C methods [2,3,5,9], have opened the way to explore chromatin folding at the genomic scale. It is emerging that chromosomes tend to form megabase-long domains, called topologically associating domains (TADs), enriched for internal contacts. Studies using mouse and human cells have discovered that TADs are universal building blocks of chromosomes and are partially conserved between different cell types [10,11]. Chromosomes are also divided into 10-Mb-long regions belonging to the so-called A and B compartments, associated respectively with active and repressed gene states [9]. In fact, TADs and A and B compartment regions are thought to be only two levels of an entire hierarchy of higher-order domains, named metaTADs [12], spanning the genome from the sub-Mb to chromosomal scales (100 Mb) [12–14].

To explain the complex patterns revealed by experimental data, approaches from polymer physics have been introduced (see, e.g., the reviews in [15–21]). Models have been proposed, in particular, to focus on the effects of chromatin looping produced by its interactions with molecular factors, such as the dynamic loop model [22] and the more recent extrusion loop model, where CTCF binding site interactions mediated by cohesin are specifically discussed [23]. Here we consider a simple polymer physics model, the string and binder switch (SBS) model [24]. In this model nonrandom chromatin conformations are established through specific interaction of chromatin with diffusible DNA-binding molecules, driving folding by the formation of loops. This model can describe the mechanisms underlying the self-assembling of chromatin three-dimensional (3D) architectural domains, such as TADs, and the metaTADs hierarchy, as emerging by Hi-C and fluorescence *in situ* hybridization (FISH) data [12,25].

Here we focus on a specific genomic locus, a 2-Mb-long region around the Hox-B gene, a genomic region important during embryonic development. Based on classical scaling concepts of polymer physics, we show that the SBS model is capable at explaining with 95% accuracy the locus pairwise contact map in mouse embryonic stem cells (mESCs), as derived by Hi-C experiments. We also show that its folding dynamics has an intrinsic hierarchical structure, where domains of different scales are progressively aggregated.

II. MODEL AND SIMULATION DETAILS

A. Model

In the SBS model [24,25] a chromatin fiber is represented as a self-avoiding walk (SAW) polymer chain, composed of N beads (Fig. 1). Each bead can interact with diffusing particles (binders) with an interaction energy $E_{\rm int}$. The total number of binders is P. Chain and binders undergo Brownian motion and each particle of the system (bead or binder) obeys the Langevin equation [26]

$$m\frac{d^2}{dt^2}x(t) = -\zeta\frac{d}{dt}x(t) - \nabla V(x(t)) + \xi(t), \qquad (1)$$

where m is the mass of the particle, ζ is the friction coefficient of the particle in the solvent, ∇V is the force term generated by the potential $V(x)$, and $\xi(t)$ is the randomly fluctuating force having a correlation function

$$\langle \xi_i(t)\xi_j(t)\rangle = 2\zeta k_B T\delta_{ij}\delta(t-t'), \qquad (2)$$

where k_B is the Boltzmann constant and T is the temperature of the system. The friction coefficient of the particle ζ is related to the viscosity of the solvent η according to the Stokes equation $\zeta = 3\pi\eta\sigma$, where σ is the diameter of the particle, and to the diffusion coefficient through the relation $D = k_B T/\zeta$.

To describe the system interaction potentials we employ a classical model of polymer physics [27]. Between any two particles (binders or beads), there is always a purely repulsive interaction described by a shifted, truncated Lennard-Jones (LJ) potential $V(r)$ that takes into account excluded-volume effects:

$$V(r) = \begin{cases} 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} + \frac{1}{4}\right] & \text{for } r < 2^{1/6}\sigma \\ 0 & \text{otherwise,} \end{cases} \qquad (3)$$

*Corresponding author: mario.nicodemi@na.infn.it

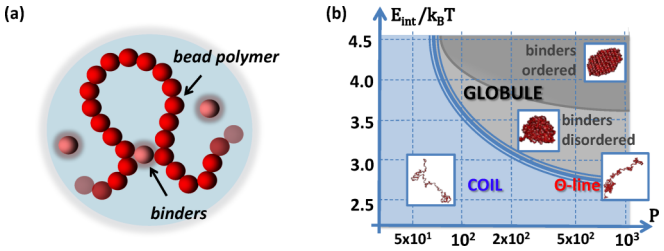

FIG. 1. (a) The SBS model is a SAW chain of beads interacting with molecular binders that can loop the polymer. In the model, P is the number of binders, which have a binding affinity $E_{\rm int}$ to the chain beads. (b) The system phase diagram in the $(P, E_{\rm int})$ plane has three main regions: The polymer can be open and randomly folded in its coil state or closed in a compact globular conformation above the Θ line; in the closed state, its binders can form a disordered lump or at higher $E_{\rm int}$ or P an ordered structure.

where $\epsilon = 1 k_B T$ is the energy unit, r is the distance between the particle centers, and σ is the particle diameter. Between two consecutive beads along the polymer chain there is an additional potential modeling a finitely extensible nonlinear elastic (FENE) potential spring [27]:

$$V_{\rm FENE}(r) = \begin{cases} -\frac{1}{2} K_{\rm FENE} \left(\frac{R_0}{\sigma}\right)^2 \ln\left[1 - \left(\frac{r}{R_0}\right)^2\right] & \text{for } r \leqslant R_0 \\ \infty & \text{otherwise.} \end{cases} \tag{4}$$

In our simulation we set R_0 equal to 1.6σ and $K_{\rm FENE}$, the strength of the FENE spring, equal to $30 k_B T$ [17]. This choice leads to an average bond length approximately equal to 0.97σ. Each polymer bead can interact with the binders through an attractive, truncated Lennard-Jones potential $V_{\rm int}(r)$ [27,28]:

$$V_{\rm int}(r) = \begin{cases} 4\epsilon_{b\text{-}b}\left[\left(\frac{\sigma_{b\text{-}b}}{r}\right)^{12} - \left(\frac{\sigma_{b\text{-}b}}{r}\right)^{6} \right. \\ \qquad \left. - \left(\frac{\sigma_{b\text{-}b}}{r_{\rm int}}\right)^{12} + \left(\frac{\sigma_{b\text{-}b}}{r_{\rm int}}\right)^{6}\right] & r < r_{\rm int} \\ 0 & \text{otherwise,} \end{cases} \tag{5}$$

where $\epsilon_{b\text{-}b}$ is the control parameter for the intensity of the polymer-binder interaction (it is a dimensionless parameter), $r_{\rm int}$ is the cutoff distance regulating the interaction range, and $\sigma_{b\text{-}b}$ is the sum of the radii of the interacting particles. Here we set $\sigma_{b\text{-}b} = 1\sigma$ and $r_{\rm int} = 1.3\sigma$. The minimum $E_{\rm int}$ of the interaction potential $V_{\rm int}$ is taken as the scale of the interaction. Here $E_{\rm int}$ is related to $\epsilon_{b\text{-}b}$ through the relationship

$$E_{\rm int} = \left| 4\epsilon_{b\text{-}b}\left[\left(\frac{\sigma_{b\text{-}b}}{r_{\rm int}}\right)^{6} - \left(\frac{\sigma_{b\text{-}b}}{r_{\rm int}}\right)^{12} - \frac{1}{4}\right]\right|. \tag{6}$$

B. Simulation details and physical units

In our molecular dynamics (MD) simulations, we first consider a homopolymer made of $N = 1000$ beads, where each bead can interact with all the binders in solution. We choose a 1000-monomer-long polymer as a large yet computationally feasible case study to investigate general aspect of chromatin folding at the scale of a few megabases.

We impose periodic conditions in the simulation box, which is chosen to have an edge size $D_{\rm box}$ approximately equal to the gyration radius of the corresponding SAW polymer ($D_{\rm box} \propto N^{0.588}$). This is a standard choice in MD simulation, in order to reduce boundary effects without increasing too much the required computational effort. Our MD simulations are performed by use of the LAMMPS code [29] and the equations of motion are integrated with the Verlet algorithm.

To map the MD dimensionless parameters into physical units, we proceed in the standard way [26]. To give a sense of scale of our simulations, in our first approach we aim to model chromatin folding at the TAD level [10,11]. We consider as a reference the mean genomic length L of a TAD, which is approximately 1 Mb. Then each bead contains $s_0 = L/N = 1$ kb. The physical diameter of the bead is estimated by imposing that the local genomic density is equal to the average nucleus density $\sigma \simeq (s_0/G)^{1/3} D_0$ [25], where D_0 is the nucleus diameter and G is the genome length. We consider a typical value of the nucleus diameter and genomic content in eukaryotes, $D_0 = 3.5$ μm and $G = 6.5$ Gb, to obtain a length unit $\sigma = 0.0187$ μm.

The MD time scale τ is fixed by considering the diffusion coefficient $D = \tau/\sigma^2$. Using a viscosity of 10 cP and a temperature $T = 300$ K, our time unit is $\tau = 0.0003$ s [28]. We employ an integration time step $\Delta t = 0.012$ [17] and we let the system evolve up to 5×10^8 steps, when stationarity is reached. Interestingly, we find that the typical time required to approach the equilibrium states from our simulations is of the same order of magnitude as those expected biologically [17,25,28,30] We perform also ensemble averages, up to 10^2 for each particular choice of the interaction energy $E_{\rm int}$ and binders number P considered in this study.

C. Initial configurations

In our simulations, the polymer chains are initially prepared in a random SAW configuration, while the binders are randomly located in the simulation box. To produce the starting random SAW polymer, we use the following standard approach [27]. We generate a random walk chain with a bond length equal to 0.97σ. Then, to remove any overlap between beads and binders, we make the system equilibrate, for several 10^6 time steps up to reach stationarity, with a soft potential rather than the above hard-core LJ repulsion:

$$V_{\rm soft}(r) = \begin{cases} A\left[1 + \cos\left(\frac{\pi r}{2^{1/6}\sigma}\right)\right] & \text{for } r < 2^{1/6}\sigma \\ 0 & \text{otherwise,} \end{cases} \tag{7}$$

where the factor A increases linearly in time. We check that the SAW state is approached by measuring the plateauing of the gyration radius R_g and its classical SAW scaling exponent.

III. RESULTS

A. Phase diagram

To characterize the thermodynamics phases of the model, we first focus on the simplest case where all polymer beads are identical (homopolymers) and can interact with only one type of binder. For the system described in the previous section, we construct the phase diagram as a function of its main parameters: the bead-binder interaction energy $E_{\rm int}$ and the binders number P (Fig. 1). The first phase transition occurring in the system is a classical coil-globule transition, where the open polymer folds into a compact configuration [25]. The level of compaction of the polymer

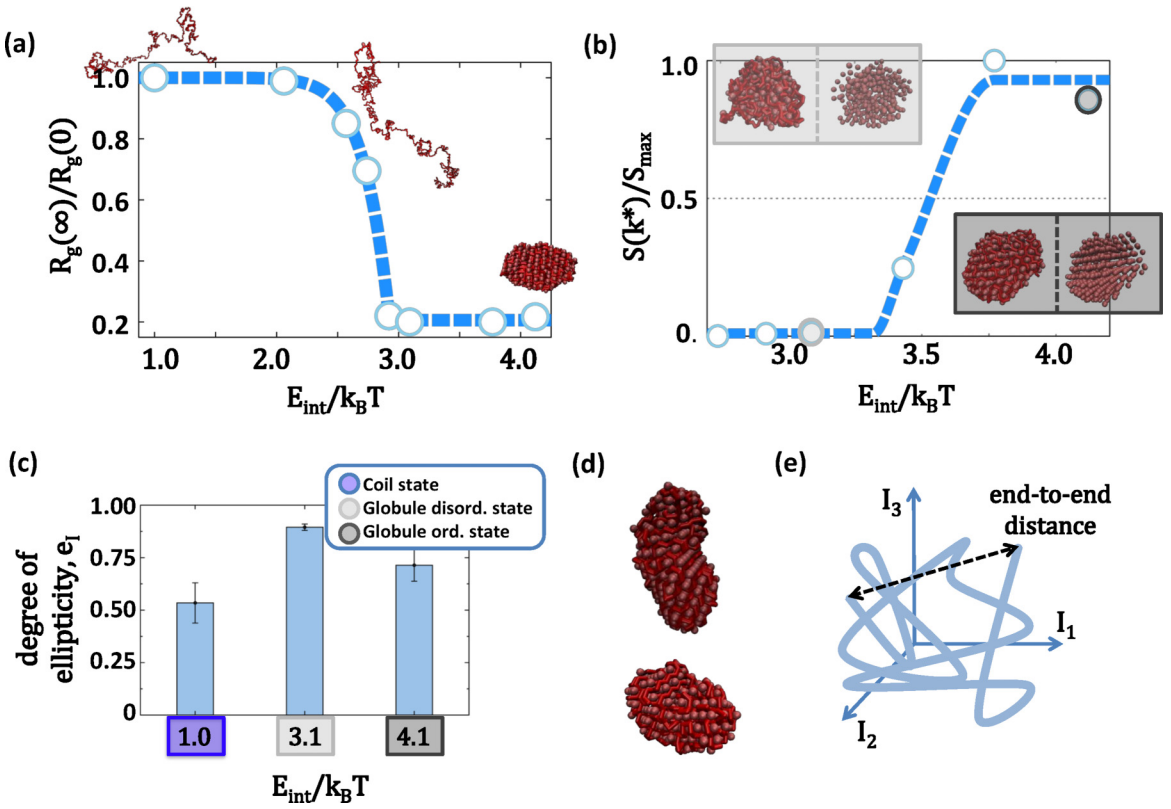

FIG. 2. (a) Equilibrium gyration radius of the SBS polymer R_g shown here as a function of the interaction energy $E_{\rm int}$ ($P = 524$). It has a drop at the coil-globule transition point. The snapshots of the system configurations at equilibrium are taken in the coil and globule states ($E_{\rm int} = 1.0k_BT$ and $E_{\rm int} = 4.12k_BT$) and around the Θ point ($E_{\rm int} = 2.57k_BT$). (b) The jump in the structure factor peak $S(k^*)$ signals that the binders undergo an ordering transition, albeit they do not interact with each other. The snapshots of the system at equilibrium show the entire system (polymer and binders) in the left panel, but only the binders in right panel, to highlight their ordering state. (c) The polymer ellipticity ratio $e_I = 2I_1/(I_2 + I_3)$ (where $I_1 < I_2 < I_3$ are the polymer principal momenta of inertia) should be equal to 1 in a spherical conformation. However, we find that it is smaller than 1 in all the different phases, here identified by the different values of $E_{\rm int}$. (d) Snapshots of different steady-state configurations for our homopolymer model in the ordered state clearly show that its conformation is generally nonspherical. (e) Schematic representation of the principal axes of inertia of the polymer and their reference system.

is captured by measuring, at equilibrium, its gyration radius $R_g = \sqrt{\sum_{i=1}^{N} m_i(\vec{r}_i - \vec{r}_{\rm c.m.})^2/M}$, where m_i and $\vec{r}_i$ are the mass and the position of the ith polymer bead, and M and $\vec{r}_{\rm c.m.}$ are, respectively, the total mass and the position of the center of mass of the polymer. The gyration radius R_g is high when the polymer is open in a SAW conformation, while it drops to lower values in the compact state (Fig. 2).

A second transition occurs when the polymer is in the globule state and corresponds to an ordering of the bound binders, albeit they have no direct interactions with each other [30]. At low interaction energy $E_{\rm int}$ or P, the binders attached to the already folded polymer form a disordered lump. However, they self-organize in ordered configurations if the interaction energy grows above a transition threshold (Fig. 2). To locate such a transition, we consider two structural quantities associated with the binders. The first one is the pair distribution function $g(r)$, defined by [26]

$$g(r) = \frac{V}{N_b^2}\left\langle \sum_i \sum_{i\neq j} \delta(r - r_{ij}) \right\rangle, \tag{8}$$

where N_b is the number of binders attached to the polymer, V is the volume, and δ is a Dirac delta function. The angular brackets indicate an ensemble average. The second quantity we consider is the structure factor $S(k)$, which is related to $g(r)$ [26] by the definition

$$S(k) = 1 + \frac{4\pi N_b}{V}\int_0^\infty r^2 \frac{\sin(kr)}{kr} g(r)dr. \tag{9}$$

The structure factor $S(k)$, when the bead system is in a disordered configuration, i.e., at low interaction energies with the polymer, is practically flat. Conversely, it has sharp peaks as a function of k when the polymer is in the globular state and $E_{\rm int}$ is high. The order parameter of the ordering transition is the ratio $S(k^*)/S_{\rm max}$, which has a jump at the transition point [26]. Here k^* is the wave vector corresponding to the second peak in $S(k)$ and $S_{\rm max}$ a normalization equal to the maximum of $S(k^*)$ (Fig. 2).

To investigate the shape of the equilibrated folded polymer in the different phases, we also calculate its inertia tensor T,

$$T_{jk} = \sum_i^N m_i\left(r_i^2\delta_{jk} - x_{ij}x_{ik}\right), \tag{10}$$

where j and k are the indices of the space axes, $j,k \in \{x,y,z\}$, i is a bead index, m_i is the mass of the ith bead, and x_{ij}

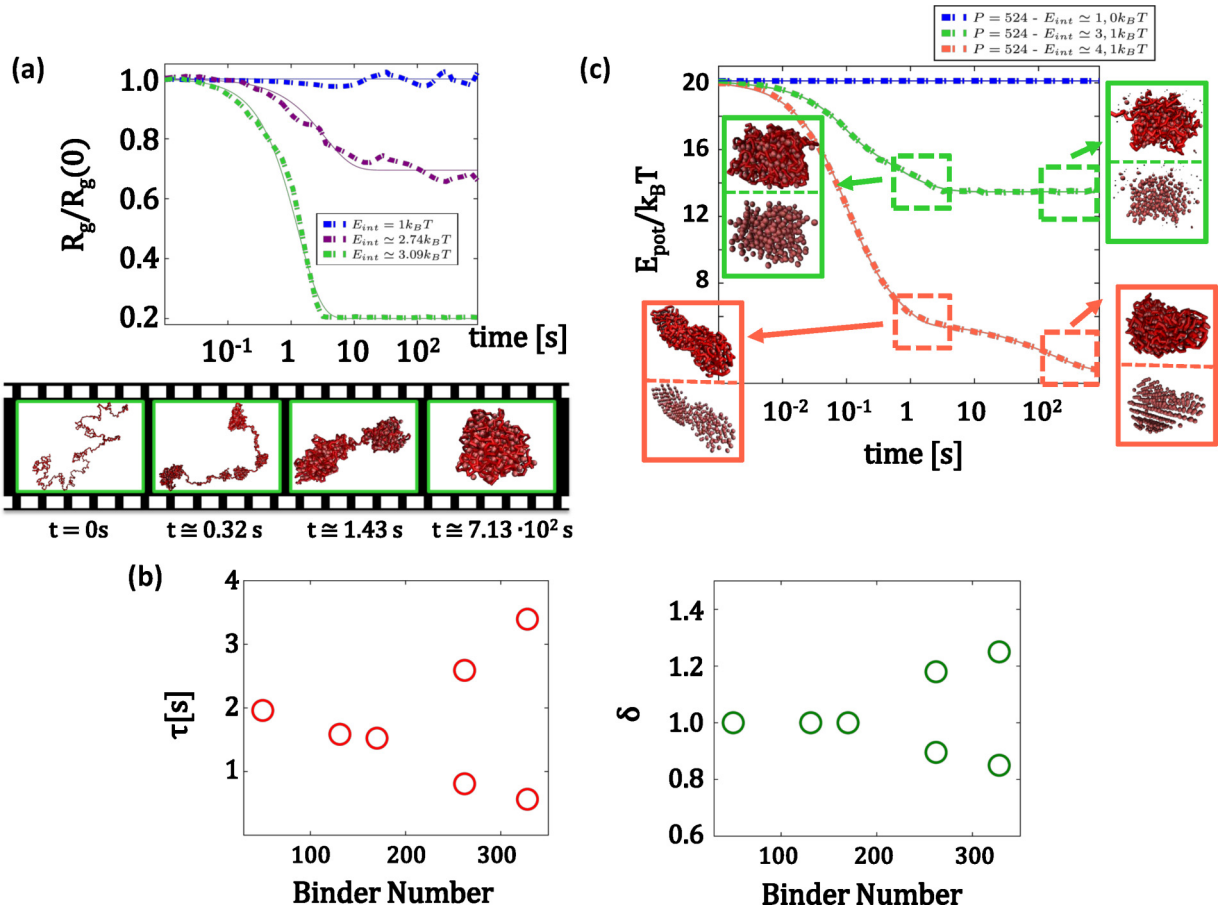

FIG. 3. (a) The polymer folding process from a SAW initial state is captured by the relaxation dynamics of its gyration radius $R_g(t)$. It is shown here for different values of the interaction energy [R_g is normalized with its SAW value $R_g(0)$]. The dotted lines are MD simulation data and the solid ones our stretched exponential fit. The bottom panel shows 3D snapshots of the polymer in time ($E_{\rm int} = 3.09k_BT$ and $P = 524$). (b) Time scale and exponent τ (left) and δ (right) of the stretched exponential fit of $R_g(t)$ shown as a function of the binder number P around the Θ point. (c) Total potential energy $E_{\rm pot}$ plotted as a function of time for different values of $E_{\rm int}$ ($P = 524$). For high values of $E_{\rm int}$ (e.g., $E_{\rm int} = 4.1k_BT$) there are different relaxation regimes: In the first stage of the dynamics the binders randomly aggregate on the polymer; in the second stage, they rearrange to form an ordered structure. The 3D snapshots (bottom only binders and top also polymer) at different time points help visualize the ordering transition at $E_{\rm int} = 4.1$.

is its jth coordinate. By diagonalizing T, we derive its three eigenvalues, which are the system principal momenta of inertia I_1, I_2, I_3. The ratio $e_I = 2I_1/(I_2 + I_3)$, where $I_3 \geqslant I_2 \geqslant I_1$, returns a measure of the degree of ellipticity of the polymer shape: In a perfectly spherical conformation $e_I = 1$, while the higher the level of ellipticity, the lower the e_I. We find that in the coil SAW state $e_I \simeq 0.5$, in the ordered globular state $e_I \simeq 0.7$, and in the disordered globular state $e_I \simeq 0.9$ (Fig. 2). Hence, even in the SAW state, the polymer is more elongated along one axis, which in this case we find to be statistically aligned with the end-to-end direction of the polymer. Our results on the asphericity of SAWs are in full agreement with previous findings from polymer physics [31]. Interestingly, experimental measures suggest that many chromosomal territories have regular ellipsoidlike shapes with an ellipticity falling within the range 0.7–0.9 [32].

B. Folding dynamics

The polymer folding process from a SAW configuration to a compact state is driven by the formation of loops produced by the binders. The details of the process depend on the specific choice of the system parameters, i.e., its interaction energy $E_{\rm int}$ and binders number P (Fig. 3). The gyration radius $R_g(t)$ has initially its SAW equilibrium value. Above the polymer Θ point, it gradually decreases in time t until a plateau is reached at the corresponding new equilibrium level (Fig. 2). At small values of P, the dynamics of $R_g(t)$ is well fitted by a single-exponential function $R_g(t) \sim \exp(-\gamma t)$. Around the Θ point the folding process becomes more complex and a combination of two stretched exponentials $r_1 \exp(-\gamma_1 t^{\delta_1}) + r_2 \exp(-\gamma_2 t^{\delta_2})$ is required to have an accurate fit of the data, like those shown in Fig. 3. The corresponding characteristic times $\tau_i = \int_0^\infty \exp(-\gamma_i t^{\delta_i}) dt$ and exponents δ_i are plotted in Fig. 3, which also includes 3D snapshots of the polymer during the folding dynamics. A similar dynamical behavior is found for the total potential energy $E_{\rm pot}$, i.e., the bead-binder LJ and bead-bead FENE interactions (Fig. 3). In particular, in the ordered globular phase, a third regime appears in the relaxation of $E_{\rm pot}$, corresponding to the ordering process of the binders' lump.

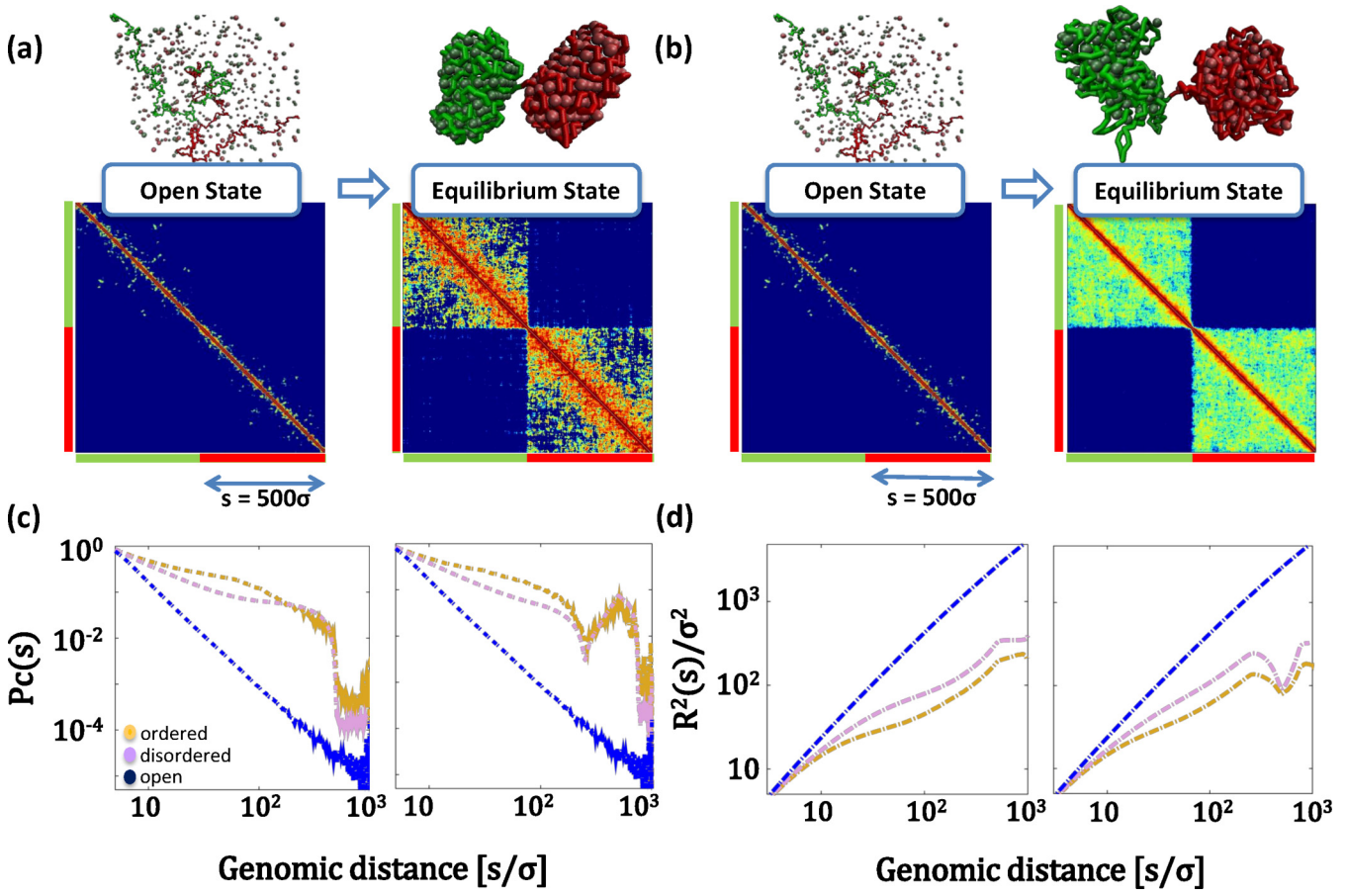

FIG. 4. The pattern of the equilibrium pairwise contact frequency matrix of our block-copolymer model with two types of sites (red and green) depends on the thermodynamic state of the system. While in the initial coil state the contact matrix is uniform, in the globular phase two distinct domains are formed, having different strength in (a) the ordered binder state ($E_{\rm int} = 4.1k_BT$) and (b) the disordered binder state ($E_{\rm int} = 3.1k_BT$). That is also shown in the 3D snapshots above. (c) System pairwise average contact probability $P_c(s)$ vs the contour separation s of polymer bead pairs. Two cases are considered: On the left is a model with only one red and one green domain [as in (a)] and on the right is a model with two red and two green alternating domains. The different colors correspond to the three different thermodynamics states. In the globular states (ordered and disordered), $P_c(s)$ has apparent crossovers around the domain boundaries. They are visible at a genomic distance equal to $s = N/2$ in the two-block model and also at $s = N/4, 3N/4$ in the four-block model. (d) Pair mean square distance $R^2(s)$ vs s in the two cases describe before, with its crossovers around domain boundaries.

C. Block-copolymer models

To investigate the formation of more complex architectures, including different topological domains, we consider a block-copolymer model with two different bead types, each interacting with its specific cognate binders, visually represented in red and green (Fig. 4). In our MD simulations, each subpolymer is 500 beads long, so the whole polymer is still made of 1000 beads in total. We set the number of binders $P = 460$ high enough to drive the polymer in compact folded states and considered three values of interaction energy ($E_{\rm int} = 0$, $3.12k_BT$, and $4.12k_BT$) corresponding to the three phases described above.

To compute the average contact probability $P_c(s)$, as a function of the contour distance s (i.e., the genomic separation) between two loci, and the entire pairwise contact matrix, we proceed in the following way. For each independent replica of the system we consider an equilibrated configuration at a given time step. Then we set a contact threshold distance $\lambda\sigma$, where σ is the length unit, and consider the distance r_{ij} between each pair i and j ($i \neq j$, where i, j are bead indices along the chain). If $r_{ij} < \lambda\sigma$, then we count a contact between the monomers i and j. In this way, we obtain a contact matrix for a particular time step. Here we show the results for $\lambda = 3.5$, but we check that small changes in λ do not alter our results. To derive the polymer pairwise contact matrix, to be compared with Hi-C data, we consider the average of these matrices over many independent conformations.

The equilibrium pairwise contact frequency matrix of our block copolymer in the coil state has a uniform pattern. Conversely, in the globular state two distinct domains spontaneously form, one composed of the red beads and one composed of the green beads (Fig. 4). It is also interesting to consider the average pairwise contact probability $P_c(s)$ as a function of s, which has apparent crossovers around the domain boundaries, at a genomic distance $s = N/2$ (Fig. 4). The mean square distance $R^2(s)$ has a similar mark (Fig. 4).

Finally, we consider, in the same conditions, a block copolymer having four distinct blocks, each block formed by 250 beads (two red and two green, 1000 beads in total), so as to simulate the formation of four different TADs. Now,

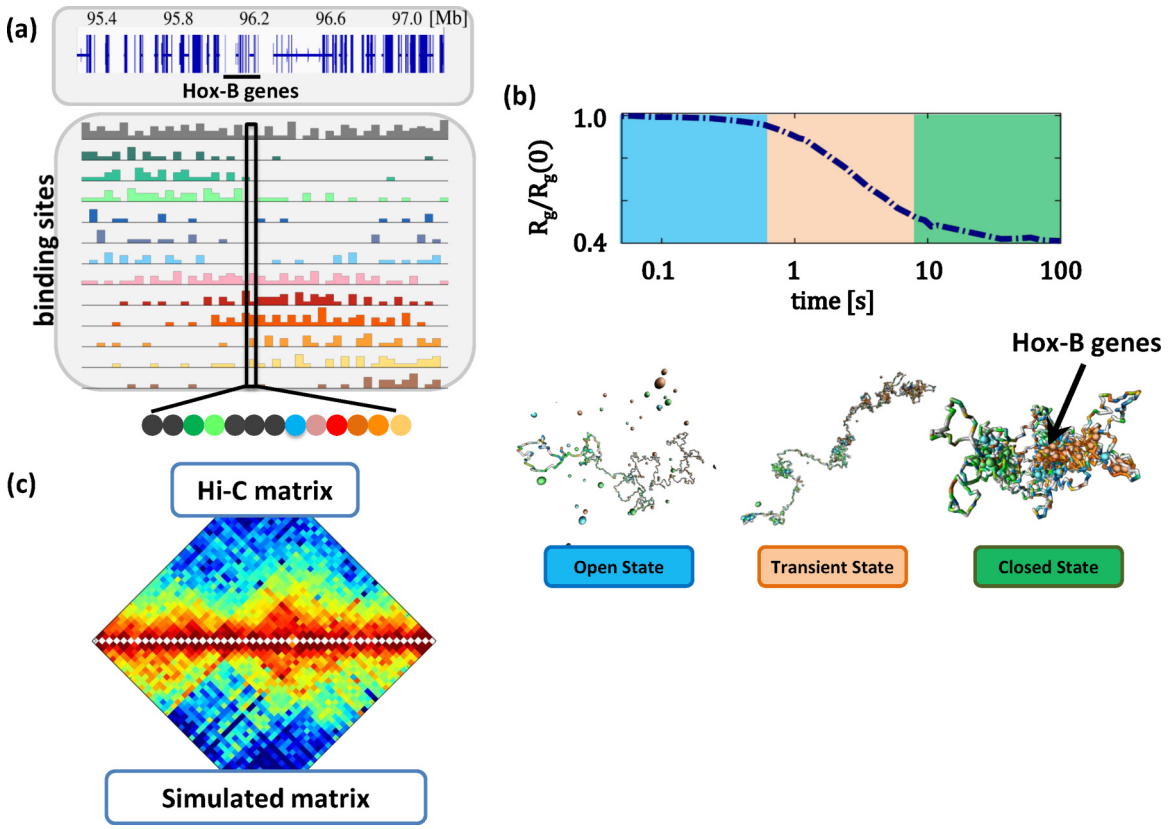

FIG. 5. (a) Scheme of the Hox-B genomic locus considered in our model. The histograms show the location and abundance of the different types (colors) of binding sites of the SBS polymer model that best explain the Hi-C contact map of the locus in mESCs. (b) Folding dynamics of the locus, as represented by $R - g$, from an open SAW initial configuration (left) to the final compact state (right). Here $P = 39$ and $E_{\text{int}} = 4.12k_BT$. The bottom panel shows 3D snapshots of the locus in time. The positions of the Hox-B genes are highlighted in the equilibrium configuration. (c) The Hi-C experimental contact matrix of the locus in mESCs (top) and the one inferred by the model described in (a) (bottom). They have a 95% Pearson correlation, highlighting that our model describes a relevant part of the folding mechanisms.

four different domains are formed along the polymer (Fig. 4), as visible in the contact probability and the mean square distance. Additionally, binders of the same type can bring in close physical proximity the different blocks of the same type, giving rise to a more complex hierarchical organization: There are therefore lower- and higher-order structures deriving from intradomain and interdomain interactions, respectively. Interestingly, a hierarchical pattern is experimentally found in Hi-C maps [12,30].

D. Model of the Hox-B locus in mESCs

Next we aim at understanding whether our model can explain the folding of specific, real genomic loci, rather than the average features of chromosomal conformations. We consider, in particular, the case of the Hox-B locus in mESCs, where Hi-C experimental data are available. Our considered locus is 1.92 Mb long (chr11:95280000–97200000), binned at 40-kb resolution, and centered around the Hox-B gene cluster (Fig. 5). We generalized our block-copolymer model in order to allow for different types of binding sites (colors); each type of bead along the chain can interact only with its corresponding type of binders. The bead types and their positions along the model polymer are obtained by a simulated annealing Monte Carlo optimization procedure that minimizes the differences between the experimental Hi-C contact matrix and the theoretical contact matrix predicted by polymer physics. The method we employ uses a standard simulated annealing procedure with a cost function that also includes a Bayesian term (a chemical potential) to avoid overfitting [30]. We use published Hi-C data from Dixon *et al.* [10]. Twelve different bead types (each visually represented by a different color) are identified by the optimization procedure to describe the locus (Fig. 5).

The polymer we consider is a chain made of $N = 576$ beads. So the elementary bead contains about 3.3 kb. Each bead interacts with its specific binder. In our simulations we sample values of the total binder number P ranging from zero to 195, to explore the system conformations in the previously described thermodynamics phases. The interaction between the binders and beads is modeled by an attractive LJ potential as described before, using the following parameters: $\epsilon_{b\text{-}b} = 12k_BT$, $\sigma = 1$, and $r_{\text{int}} = 1.5$. We consider an ensemble of 10^2 independent equilibrated polymers, each starting from a SAW configuration as described above. To reach equilibrium, we run the simulations up to 2.5×10^8 time steps. The folding dynamics, as seen before, has a hierarchical nature as visualized in Fig. 5, where also 3D snapshots of the locus are given at different time points during folding.

In the case of our Hox-B polymer model, the procedure to compute the pairwise contact matrix is applied to a mixture of the contact matrices of the different states in order to maximize the Pearson correlation coefficient between the model predicted and Hi-C pairwise contact frequency matrices. To check the robustness of our approach we also consider a higher threshold value $\lambda = 8$ and a variant of the procedure where only contacts between monomers of the same color are retained in the calculation of the contact matrix, finding a best combination made of 72% open SAW state and 28% compact state. In this way we get a Pearson correlation coefficient of $r = 0.95$ (Fig. 5) between the model predicted and Hi-C data. Overall, our results support a picture whereby the SBS polymer model can describe the folding of the Hox-B locus and can provide systematic access to all information on the system conformations, well beyond the pairwise contact matrix available from the experiments.

IV. CONCLUSION

We have discussed the SBS model, a simple polymer model of chromatin where 3D conformation is established through attachment of diffusible factor (binders) to the polymer. The model explains how different genomic conformation can arise spontaneously with switchlike nature and the mechanisms underlying chromatin self-organization: The modification of chromatin architecture can be regulated by simple parameters, such as protein number (i.e., their concentration) and binding affinity. The polymer model can fold in different stable architectural classes corresponding to its thermodynamics emergent phases: the coil state and the globular compact polymer state, where the binders in turn can undergo an ordering transition. While a given locus can be folded differently across a population of cells, its stable folding conformations fall in classes corresponding to the system thermodynamics phases, as dictated by polymer physics. Hence, basic scaling concepts could be helpful to understand chromosome conformations.

The SBS model is a schematic representation of chromatin and the many complications that arise in real situations. Polymer confinement, crowding, entanglement, and many additional effects, such as an interplay between equilibrium and off-equilibrium phenomena, are likely to have important implications as found in the investigation of other complex fluids (see, e.g., [33–39] and references therein). The SBS model has been previously used to try to model symmetry-breaking events in the architecture of the Xist locus upon X-chromosome inactivation [40,41] and to model chromosome recognition at mitosis and meiosis [42]. Here, in particular, we showed within the SBS model framework that the formation of chromatin domains and looping is a hierarchical process, consistent with recent discoveries on chromosome metaTADs [12]. In addition, we found that Hi-C data from the Hox-B locus, a region associated with key events during embryonic development, can be reproduced with a 95% accuracy, supporting the view that our model captures some of its key folding mechanism.

Models similar to the SBS model have been proposed, such as the dynamic loop (DL) model [22], where diffusing binders are replaced by an effective interaction between the chain beads. Importantly, they were shown to have similar thermodynamics features, supporting the robustness of these approaches to describe the large-scale features of chromosome folding, beyond the specific details of the interaction potentials considered. More recently, extrusion loop (EL) models have been considered, focusing on the role of cohesin mediated interactions between CTFC binding sites [23,43,44]. They fall in the same class including the SBS [24] and DL [22] models, as they are made of a chain of interacting beads where, additionally, CTFC sites can be bridged by a specific type of binders (viewed as cohesin). Considering the simplicity of the ingredients, EL models can describe very well genomic regions where folding is mainly driven by CTCF. A difference from the approach discussed here is that here we make no *a priori* hypotheses on the nature of the folding factors and derive their different types and positions as an output of our method. Polymer models can help in understanding chromosome folding beyond Hi-C pairwise contact frequencies, as they return the entire 3D conformations and the relative positions of different regions (e.g., gene promoters and regulators), which could be tested, for example, by FISH experiments.

ACKNOWLEDGMENTS

This work was supported by CINECA ISCRA Grants No. HP10CYFPS5 and No. HP10CRTY8P, NIH Grant No. 1U54DK107977-01, and computer resources at INFN, CINECA, and Scope at the University of Naples.

C.A., A.M.C., and S.B. contributed equally to this work.

[1] T. Misteli, Beyond the sequence: Cellular organisation of genome function, Cell **128**, 787 (2007).

[2] A. Tanay and G. Cavalli, Chromosomal domains: Epigenetic contexts and functional implications of genomic compartmentalization, Curr. Opin. Genet. Dev. **23**, 197 (2013).

[3] W. Bickmore and B. van Steensel, Genome architecture: domain organisation of interphase chromosomes, Cell **152**, 1270 (2013).

[4] J. Dekker and T. Misteli, Long-range chromatin interactions, Cold Spring Harb. Perspect. Biol. **7**, a019356 (2015).

[5] J. Dekker and L. Mirny, The 3D genome as moderator of chromosomal communication, Cell **164**, 1110 (2016).

[6] M. Spielmann and S. Mundlos, Structural variations, the regulatory landscape of the genome and their alteration in human disease, Bioessays **35**, 533 (2013).

[7] D. Lupianez *et al.*, Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions, Cell **161**, 1012 (2015).

[8] D. Hnisz *et al.*, Activation of proto-oncogenes by disruption of chromosome neighborhoods, Science **351**, 1454 (2016).

[9] E. Lieberman-Aiden, N. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. Lajoie, P. Sabo, M. Dorschner *et al.*, Comprehensive mapping of long-range interactions reveals fold- ing principles of the human genome, Science **326**, 289 (2009).

[10] J. R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, and B. Ren, Topological domains in mammalian genomes identified by analysis of chromatin interactions, Nature (London) **485**, 376 (2012).

[11] N. Nora, B. Lajoie, E. Schulz, L. Giorgetti, I. Okamoto, N. Servant, T. Piolot, N. van Berkum, J. Meisig, J. Sedat *et al.*, Spatial partitioning of the regulatory landscape of the x-inactivation center, Nature (London) **485**, 381 (2012).

[12] J. Fraser *et al.*, Hierarchical folding and reorganisation of chromosomes are linked to transcriptional changes during cellular differentiation, Mol. Syst. Biol. **11**, 852 (2015).

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

[14] J. Phillips-Cremins, M. Sauria, A. Sanyal, T. Gerasimova, B. Lajoie, J. Bell, C. Ong, T. Hookway, C. Guo, Y. Sun, M. Bland, W. Wagstaff, S. Dalton, T. McDevitt, R. Sen, J. Dekker, J. Taylor, and V. Corces, Architectural protein subclasses shape 3D organisation of genomes during lineage commitment, Cell **153**, 1281 (2013).

[15] J. Langowski, Polymer chain models of dna and chromatin, Eur. Phys. J. E **19**, 241 (2006).

[16] D. Marenduzzo, C. Micheletti, and P. R. Cook, Entropy-driven genome organization, Biophys. J. **90**, 3712 (2006).

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

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

[19] M. Tark-Damme, R. van Driel, and D. Heermann, Chromatin folding—From biology to polymer models and back, J. Cell Sci. **124**, 839 (2011).

[20] M. Barbieri, A. Scialdone, A. Gamba, M. Nicodemi, and A. Pombo, Polymer physics, scaling and heterogeneity in the spatial organisation of chromosomes in the cell nucleus, Soft Matter **9**, 8631 (2013).

[21] M. Nicodemi and A. Pombo, Models of chromosome structure, Curr. Opin. Cell Biol. **28**, 90 (2014).

[22] M. Bohn and D. Heermann, Diffusion-driven looping provides a consistent framework for chromatin organisation, PLoS ONE **5**, e12218 (2010).

[23] A. Sanborn, S. Rao, S. Huang, N. Durand, M. Huntley, A. Jewett, I. Bochkov, D. Chinnappan, A. Cutkosky, J. Li, K. Geeting, A. Gnirke, A. Melnikov, D. McKenna, E. Stamenova, E. Lander, and E. Aiden, Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes, Proc. Natl. Acad. Sci. USA **112**, E6456 (2015).

[24] M. Nicodemi and A. Prisco, Thermodynamic pathways to genome spatial organization in the cell nucleus, Biophys. J. **96**, 2168 (2009).

[25] M. Barbieri, M. Chotalia, J. Fraser, L. Lavitas, J. Dostie, A. Pombo, and M. Nicodemi, Complexity of chromatin folding is captured by the strings and binders switch model, Proc. Natl. Acad. Sci. USA **109**, 16173 (2012).

[26] M. Allen and D. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, Oxford, 1987).

[27] K. Kremer and G. Grest, Dynamics of entangled linear polymer melts: A molecular-dynamics simulation, J. Chem. Phys. **92**, 5057 (1990).

[28] C. Brackley, S. Taylor, A. Papantonis, P. Cook, and D. Marenduzzo, Nonspecific bridging-induced attraction drives clustering of dna-binding proteins and genome organization, Proc. Natl. Acad. Sci. USA **110**, E3605 (2013).

[29] S. Plimpton, Fast parallel algorithms for short-range molecular dynamics, J. Comput. Phys. **117**, 1 (1995).

[30] A. M. Chiariello *et al.*, Polymer physics of chromosome large-scale 3D organisation, Sci. Rep. **6**, 29775 (2016).

[31] M. Bishop and J. P. J. Michels, The shape of ring polymers, J. Chem. Phys. **82**, 1059 (1985).

[32] N. Sehgal, A. Fritz, K. Morris, I. Torres, Z. Chen, J. Xu, and R. Berezney, Gene density and chromosome territory shape, Chromosoma **123**, 499 (2014).

[33] M. Nicodemi and A. Coniglio, Macroscopic glassy relaxations and microscopic motions in a frustrated lattice gas, Phys. Rev. E **57**, R39(R) (1998).

[34] M. Nicodemi, Force Correlations and Arch Formation in Granular Assemblies, Phys. Rev. Lett. **80**, 1340 (1998).

[35] M. Tarzia *et al.*, Glass transition in granular media, Europhys. Lett. **66**, 531 (2004).

[36] A. Coniglio and M. Nicodemi, The jamming transition of granular media, J. Phys.: Condens. Matter **12**, 6601 (2000).

[37] A. Coniglio *et al.*, Segregation of granular mixtures in presence of compaction, Europhys. Lett. **43**, 591 (1998).

[38] M. Nicodemi and H. J. Jensen, Creep of Superconducting Vortices in the Limit of Vanishing Temperature: A Fingerprint of Off-Equilibrium Dynamics, Phys. Rev. Lett. **86**, 4378 (2001).

[39] D. S. Grebenkov, M. P. Ciamarra, M. Nicodemi, and A. Coniglio, Force Correlations and Arch Formation in Granular Assemblies, Phys. Rev. Lett. **100**, 078001 (2008).

[40] M. Nicodemi and A. Prisco, Symmetry-breaking model for X-chromosome inactivation, Phys. Rev. Lett. **98**, 108104 (2007).

[41] A. Scialdone *et al.*, Conformation regulation of the X chromosome inactivation center: A model, PLoS Comput. Biol. **7**, e1002229 (2011).

[42] M. Nicodemi, B. Panning, and A. Prisco, A thermodynamic switch for chromosome colocalization, Genetics **179**, 717 (2008).

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

[44] A. Goloborodko, J. Marko, and L. Mirny, Chromosome compaction by active loop extrusion, Biophys. J. **110**, 2162 (2016).