Heterogeneous cross-linked polymers to reconstruct chromatin reorganization during cell differentiation

O. Shukron and D. Holcman

Chromatin of mammalian nucleus folds into discrete contact enriched regions such as Topologically Associating domains (TADs). The folding hierarchy of the TADs and their internal organization is highly dynamic through cellular differentiation, where structural changes within and between TADs are correlated with gene activation and silencing. How these changes in chromatin conformation is correlated with gene regulation can be revealed by polymer models. Here we introduce a heterogeneous randomly cross-linked (RCL) polymer model that accounts for multiple interacting TADs. We study how connectivity within and between TADs affects chromatin organization and dynamics. Closed formula for the steady-state encounter probability within and between TADs indicate how cross-links across TADs affect TAD size and the dynamics of the chromatin. We apply the present polymer model to characterize the degree of compaction and show how secondary structures such as meta-TADs can be detected in Hi-C data. Finally, we apply the present approach to reconstruct high-order TAD organization due to inter-TAD connectivity for three TADs obtained from chromosomal capture data for the mammalian X chromosome during three successive stages of differentiation.

INTRODUCTION

Mammalian chromosome folds into discrete mega-bps (mbp) contact enriched regions termed Topologically associating domain (TADs). Although the precise role of TADs remain unclear, they participate in gene regulation [1, 2] and allows gene to activate synchronously, a concept known as replication timing units [3]. Gene regulation within TAD is modulated by transient loops at sub-TAD scale [1, 4, 5], formed by regulatory elements such as enhancers and promoters [6]. However, sparse connectors between TADs (at a scale >mbps) can significantly influence chromatin dynamics and gene regulation by generating bridges within TADs [7, 8]. We present here a polymer model that accounts for interactions between TADs to study chromatin dynamic at long scales, an area that remains largely unexplored.

Genome organization can be probed by chromosome conformation capture (CC) methods [2, 4, 9], which report simultaneous genomic contacts (loops) at scales of kilo bps (kbps) to mbps. TADs appear as matrix block in the contact maps by averaging encounter events over an ensemble of millions of cells [1, 9]. CC contact maps provide a statistical summary of steady-state looping frequencies, but do not contain neither direct information about the size of the folded genomic section nor any transient genomic encounter times [10]. Throughout cell differentiation stages, the boundaries between TADs remain stable, [1, 11] but their internal looping pattern is highly variable. Moreover, TADs can form hierarchy structures into meta TADs, formed by inter TAD connectivity, correlated with transcription state of the chromatin [11].

Coarse-grained polymer models are used to represent chromatin in steady state and in transient regimes, but also to study the statistical properties of the chromatin at a given scale. Starting with the Rouse polymer [12], composed of N monomers connected sequentially by harmonic springs, the linear connectivity in the Rouse model does not account for the complexity of molecular interactions and therefore cannot account for contact enriched regions such as TADs. To include short and long-range looping, other polymer models have been developed [10, 13–18], such as self avoiding interactions, random loops [8, 13, 19, 20] and epigenomic state [21, 22] demonstrating that TADs formation results from the intrinsic polymer properties.

Here, to account for interactions between multiple TAD, we develop a parsimonious polymer model of the chromatin. We extend the construction of the randomly cross-linked (RCL) polymer model, introduced in [7, 10], to account for multiple connected TADs of various sizes and connectivity within and between TADs. RCL can be generated by binding molecules such as CTCF [1] or by loop extrusion mechanism [23], although the exact mechanism by which they form is not the focus of the present work. We randomize the positions of the cross-links to capture the heterogeneity in chromatin structure in an ensemble of cells. To study the polymer dynamics, we obtain closed formulas for the encounter probability (EP), variance and the radius of gyration of the heterogeneous RCL polymer that reveals the role of parameters. We show that TADs and higher order genome organization can be modulated by external connectivity between TADs. This external connectivity also affect steady-state and time-dependent statistical properties of monomers within each TADs. To demonstrate the impact of the present model, we study the reorganization of three neighboring TADs on the mammalian X chromosome throughout successive stages of cellular differentiation. By computing the average number of connectors within and between TADs from fitting the analytical expression for the EP to the empirical 5C data [1], we evaluate for the first time the rate of compaction and decompaction of TADs, correlated to gene silencing and

activation, respectively. To conclude, the present framework allows to study chromatin dynamics from 5C data, to derive its statistical properties and reveal the influence of multiple connected TADs on each other and genome reorganization throughout cell differentiation.

CONSTRUCTING A RCL POLYMER FOR A SINGLE ASSOCIATING DOMAIN (AD)

The RCL polymer for a single AD [10] in dimension $d = 3$ is composed of N monomers at positions $\boldsymbol{R} = [r_1, r_2, ..., r_N]^T$, connected sequentially by harmonic springs [12] and N_c spring connectors are added between random non-nearest neighboring (NN) monomer pairs (Fig. 1A) leading to a realization $\mathcal{G}$. The added connectors describe the cross-linking by binding molecules such as CTCF, and their random position captures the heterogeneity over cell population. The energy of the RCL polymer is the sum of the spring potential of the linear backbone and that of added random connectors (called realization $\mathcal{G}$) [15, 20] :

$$\phi_\xi^{\mathcal{G}}(\boldsymbol{R}) = \frac{\kappa}{2}\boldsymbol{R}^T(\boldsymbol{M} + B^{\mathcal{G}}(\xi))\boldsymbol{R}, \tag{1}$$

where $\kappa = dk_BT/b^2$ is the spring constant, b the standard-deviation of the spring connector between monomers, k_B is the Boltzmann's constant and T the temperature. The matrix $\boldsymbol{M}$ is the Rouse matrix for the linear backbone [12], defining NN monomers connectivity

$$M_{m,n} = \begin{cases} -1, & |m-n| = 1; \\ -\sum_{j\neq m}^{N} M_{m,j}, & m = n; \\ 0, & \text{otherwise}, \end{cases} \tag{2}$$

and $B^{\mathcal{G}}(\xi)$ is a square symmetric connectivity matrix, representing connections between N_c randomly chosen non NN monomer pairs, and $0 \leq \xi \leq 1$ is the fraction of maximal possible non NN monomer pairs, and related to N_c by

$$N_c(\xi) = \lfloor \frac{\xi(N-1)(N-2)}{2} \rfloor. \tag{3}$$

We define the added connectivity matrix as

$$B_{mn}^{\mathcal{G}}(\xi) = \begin{cases} -1, & |m-n| > 1, r_m, r_n \text{ connected}; \\ -\sum_{i\neq j}^{N} B_{mj}^{\mathcal{G}}(\xi), & m = n; \\ 0, & \text{otherwise}. \end{cases}$$

For each polymer realization, we randomize the choice of N_c monomer pairs to connect out of the possible $(N-1)(N-2)/2$ pairs, i.e., we change the matrix $B^{\mathcal{G}}(\xi)$. The dynamics of the resulting monomers (vector $\boldsymbol{R}$) is driven by Brownian motion and the field of forces derived from the energy 1,

$$\begin{aligned} \frac{\mathrm{d}\boldsymbol{R}}{\mathrm{d}t} &= -\frac{d}{b^2}D\nabla\phi_\xi^{\mathcal{G}}(\boldsymbol{R}) + \sqrt{2D}\frac{\mathrm{d}\boldsymbol{\omega}}{\mathrm{d}t} \\ &= -\frac{d}{b^2}D\left(\boldsymbol{M} + B^{\mathcal{G}}(\xi)\right)\boldsymbol{R} + \sqrt{2D}\frac{\mathrm{d}\boldsymbol{\omega}}{\mathrm{d}t}, \end{aligned} \tag{4}$$

where $D = \frac{k_BT}{\zeta}$ is the diffusion constant, ζ is the friction coefficient, $\boldsymbol{\omega}$ are standard d-dimensional Brownian motions with mean 0 and variance 1. To study the statistical properties of 4, a mean-field approach [10] is used, where we replaced $B^{\mathcal{G}}(\xi)$ by its average $\langle B^{\mathcal{G}}(\xi)\rangle$ (average over all realization $\mathcal{G}$ for a given ξ [10, 15]). To recover the connectivity fraction ξ and the mean number of cross-links (N_c), we have used the formula for the steady-state variance, the EP between monomers m, n of a single AD [10], and the expression for the EP fitted to empirical 5C data of a single TAD. We shall now explain how this procedure can be extended to many TADs, so that the entire properties of the chromatin can be accounted for.

RCL POLYMER MODEL FOR MULTIPLE INTERACTING ADS

To construct multiple interacting TADs, we extend equations 1-4 to include $N_T > 1$ ADs of variable sizes, inter and intra AD-connectivity. The heterogeneous polymer consists of connecting sequentially N_T RCL polymers of $N = [N_1, N_2, \ldots, N_{N_T}]$ monomers, respectively (Fig. 1B). The position of monomers $\boldsymbol{R}$ consists of N_T vectors in the block matrix form

$$\boldsymbol{R} = \left[[R^{(1)}], [R^{(2)}], .., [R^{(N_T)}]\right]^T, \tag{5}$$

where the superscript indicates the AD $1, \ldots, N_T$ it belongs to and square brackets indicate a block matrix. The polymer's linear backbone is a chain of N_T Rouse matrices, defined in the matrix block form by

$$\begin{aligned} \boldsymbol{M} &= diag([M^{(1)}], [M^{(2)}], \ldots, [M^{(N_T)}]) \\ &= \begin{bmatrix} [M^{(1)}] & 0 & 0 & \ldots & 0 \\ 0 & [M^{(2)}] & 0 & \ldots & 0 \\ 0 & 0 & [M^{(3)}] & \ldots & 0 \\ . & . & 0 & \ldots & . \\ . & . & & \ldots & . \\ 0 & . & . & \ldots & [M^{(N_T)}] \end{bmatrix}, \end{aligned} \tag{6}$$

where each $[M^{(j)}]$ is a Rouse matrix of the form 2 with N_j monomers. The square-symmetric connectivity fraction matrix $\Xi = \{\xi_{ij}\}$, $1 \leq i, j \leq N_T$ accounts for connectors within and between ADs, and the mean-field average added block connectivity matrix is $\langle B^{\mathcal{G}}(\Xi)\rangle$ (Fig 1C) given by

$$\langle B^{\mathcal{G}}(\Xi)\rangle_{mn} = \begin{cases} \beta_m(\Xi), & m = n; \\ 0, & |m-n| = 1; \\ -\xi_{ij}, & m \in A_i, n \in A_j, \end{cases} \tag{7}$$

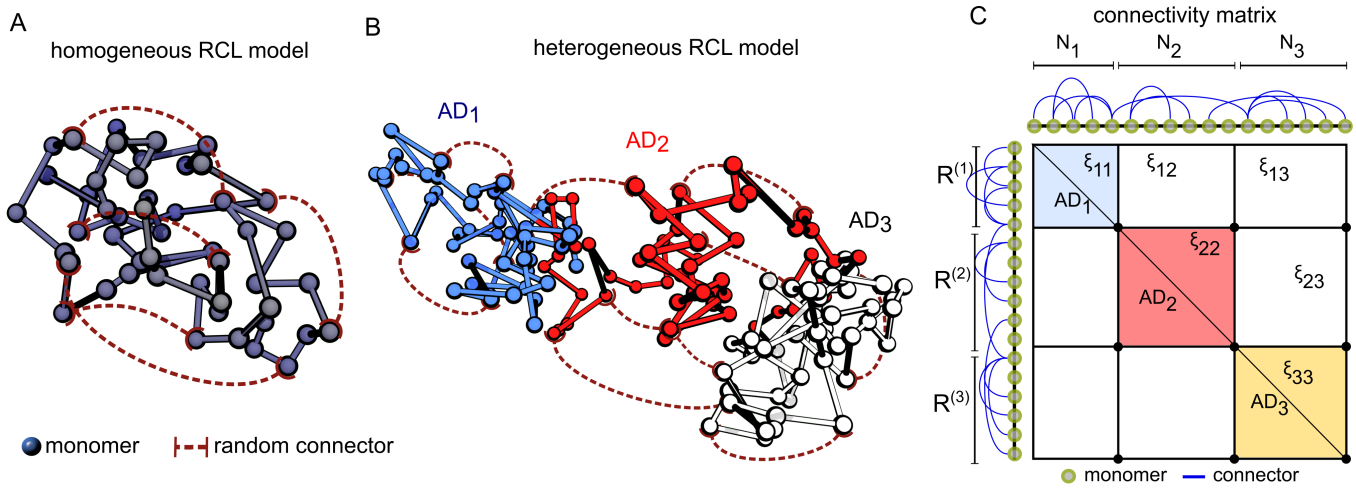

FIG. 1. **RCL polymer model for single and multiple interacting associating domains. A.**Schematic representation of the RCL polymer for a single associating domain (AD). Monomers (circles) are connected sequentially by harmonic springs to form the polymer's backbone (gray). Spring connectors (dashed red) are added between randomly chosen non nearest-neighbor monomers. The choice of monomer pairs to connect is randomized in each realization. All monomers share similar level of connectivity (homogeneous RCL model) **B.** Schematic representation of the RCL polymer model for three connected ADs (blue, red, white), where monomers (circles) are randomly connected (dashed red) within and between ADs by springs. Monomer connectivity within and between ADs can vary (heterogeneous RCL model). **C.** schematic representation of the symmetric connectivity matrix $\langle B^{\mathcal{G}}(\Xi)\rangle$, corresponding to the connected monomers over the three ADs in panel B (diagonal bocks), where each AD is of variable number of monomer N and connectivity fraction ξ_{ij}.

where

$$\beta_m(\Xi) = \begin{cases} -3\xi_{ii} + \sum_{j=1}^{N_T} N_j \xi_{ij}, & m \in A_i; \\ -2\xi_{ii} + \sum_{j=1}^{N_T} N_j \xi_{ij}, & m = \partial A_i, \end{cases} \tag{8}$$

and ∂A_i are the indices of the boundaries of A_i. The derivation of matrix $\langle B^{\mathcal{G}}(\Xi)\rangle$ is described in the appendix. The energy of the heterogeneous RCL polymer is computed from Eq. 1, with the connectivity fraction Ξ, the block matrix 6, and average random connectivity matrix $\langle B^{\mathcal{G}}(\Xi)\rangle$. The mean-field stochastic equations describing the dynamics of monomers $\boldsymbol{R}$ is

$$\frac{\mathrm{d}\boldsymbol{R}}{\mathrm{d}t} = -d\frac{D}{b^2}\left(\boldsymbol{M} + \langle B^{\mathcal{G}}(\Xi)\rangle\right)\boldsymbol{R} + \sqrt{2D}\frac{\mathrm{d}\boldsymbol{\omega}}{\mathrm{d}t}, \tag{9}$$

where $\boldsymbol{\omega}$ are Brownian motions with mean 0 and standard-deviation 1.

Statistical properties of heterogeneous RCL and numerical validation

We summarize here the main statistical properties of the heterogeneous RCL representing TADs. We first present the mean square radius of gyration (MSRG), second the encounter probability (EP) of the heterogeneous RCL polymer in two cases: encounters between monomers of the same AD (intra AD) and monomers of different ADs (inter-AD). Third, we discuss the mean square displacement of monomers of the heterogeneous RCL polymer in each AD. The computations are detailed in the Appendix.

The mean square radius of gyration (MSRG) characterizes the folding of each AD inside a ball of radius R_g. When the condition of dominant intra-AD connectivity (assumption H1, appendix 36) is satisfied, we obtain the MSRG for A_i (see Appendix)

$$\langle R_g^2\rangle^{(i)} \approx \frac{b^2}{(1-\xi_{ii})(\zeta_0^{(i)}(\Xi) - \zeta_1^{(i)}(\Xi))}, \tag{10}$$

where

$$\zeta_0^{(i)}(\Xi) = y^{(i)}(\Xi) + \sqrt{y^{(i)}(\Xi)^2 - 1},$$
$$\zeta_1^{(i)}(\Xi) = y^{(i)}(\Xi) - \sqrt{y^{(i)}(\Xi)^2 - 1}. \tag{11}$$

and

$$y^{(i)}(\Xi) = 1 + \frac{\sum_{k=1}^{N_T} \xi_{ik} N_k}{2(1-\xi_{ii})}, \tag{12}$$

Second, under the condition of non-vanishing connectivities (H_n, see condition 46), the EP between monomer m and n inside A_i is given by

$$P(m \in A_i, n \in A_i) \propto \left(\frac{d}{2\pi\sigma_{m,n}^2(\Xi)}\right)^{d/2}. \tag{13}$$

where

$$\sigma^2_{m,n}(\Xi) = \begin{cases} \langle R_g^2\rangle^{(i)}\big(\frac{(\zeta_0^{(i)}(\Xi)^{m-n}-1)^2 - 2\zeta_0^{(i)}(\Xi)^{m+n-1}}{\zeta_0^{(i)}(\Xi)^{2m-1}} + 2\big), m \geq n; \\ \langle R_g^2\rangle^{(i)}\big(\frac{(\zeta_0^{(i)}(\Xi)^{n-m}-1)^2 - 2\zeta_0^{(i)}(\Xi)^{m+n-1}}{\zeta_0^{(i)}(\Xi)^{2n-1}} + 2\big), m < n, \end{cases} \tag{14}$$

and $\langle R_g^2\rangle^{(i)}$ is the MSRG of A_i, defined in relation 10. When each monomer belongs to two distinct ADs i and j ($i \neq j$), the EP formula is modified to (see Appendix)

$$P(m \in A_i, n \in A_j)(\Xi) \propto \left(\frac{d}{2\pi\sigma^2_{m^{(i)}n^{(j)}}(\Xi)}\right)^{d/2}, \tag{15}$$

where

$$\sigma^2_{m^{(i)}n^{(j)}}(\Xi) = \\ = \langle R_g^2\rangle^{(i)}(1+\zeta_0^{(i)}(\Xi)^{1-2m}) + \langle R_g^2\rangle^{(j)}(1+\zeta_0^{(i)}(\Xi)^{1-2n}) \\ + b^2\left(\frac{1}{N_i\sum_{k\neq i}^{N_T} N_k\xi_{ik}} + \frac{1}{N_j\sum_{k\neq j}^{N_T} N_k\xi_{jk}}\right). \tag{16}$$

Finally, the Mean-Square Displacement of a monomer of the averaged heterogeneous RCL polymer inside A_i for intermediate times (see appendix) is given by

$$\langle\langle r_{m\in A_i}(t)^2\rangle\rangle \approx 2dD_{cm}t + \frac{db^2 Erf\Big[\sqrt{2dDt\sum_{k=1}^{N_T}\frac{N_k\xi_{ik}}{b^2}}\Big]}{2\sqrt{(1-\xi_{ii})\sum_{k=1}^{N_T} N_k\xi_{ik}}}, \tag{17}$$

where $D_{cm} = \frac{D}{\sum_{k=1}^{N_T} N_k}$ and Erf is the classical Error function. We now investigate the steady-state EP within ADs (Eq. 13) and between ADs (Eq. 15) by comparing Brownian simulations and formulas: for that purpose, we constructed a RCL polymer containing three ADs with $N_1 = 50, N_2 = 40, N_3 = 60$ total monomers, so that condition H1 (appendix Eq. 36) about dominant intra-connectivity is satisfied. We impose the number of connectors in each AD to be at least twice compared to the one between ADs (see Fig. 2A).

To construct the encounter frequency matrix, we use equation 9 in dimension $d = 3$, with $b = \sqrt{3}\mu m$ and diffusion coefficient $D = 1\mu m^2/s$ starting with a random walk initial polymer configuration. The connectors are placed uniformly in each A_i and in between ADs according to the numbers indicated in Fig. 2A. We ran 5000 simulations until polymer relaxation time, defined as

$$\tau(\Xi) = \max_{1\leq i\leq N_T}\left(\frac{b^2}{2dD\sum_{k=1}^{N_T} N_k\xi_{ik} + 4(1-\xi_{ii})\sin^2(\frac{\pi}{2N_i})}\right),$$

This is the longest relaxation time of RCL chains among N_T [10] ADs of the order of tens of thousands of steps.

We then constructed the simulation encounter frequency matrix, where the encounter distance is $\epsilon = b/10$. The encounter frequency matrix shows three distinct diagonal blocks (Fig. 2A) resulting from high intra AD connectivity, and further reveals a high-order organization (cyan blue in Fig. 2A), which resembles the meta-ADs discussed in [11]. We thus propose here that hierarchical AD organization is a consequence of weak inter connectivity properties.

Second, we computed using numerical simulations, the steady-state EP (Eqs. 13 and 15) from the encounter frequency matrix (Fig. 2A) by dividing each row with its sum. We then compared the results of the simulations with the theoretical EP (Eqs. 13 and 15) in Fig. 2B: we show three sample curves for monomer 20 (upper left), 70 (upper right) and 120 (lower) located in the middle of each AD, which are in good agreement with the theory. Finally, the MSRG $\langle R_g^2\rangle^{(i)}$ for A_1, A_2 and A_3 is $2.42, 1.5, 2.15\mu m$ (simulations), compared to $2.38, 1.31, 2.09$, respectively, obtained from expression 10, which are in good agreement.

To validate the MSD expression (Eq. 17), we simulated eq. 9 for 500 steps with a time step $\Delta t = 0.05$s, past the relaxation time $\tau(\Xi)$ (Eq. 18) and computed the MSD for each monomers in each AD $i = 1, 2, 3$. In Fig. 2C we plotted the mean MSD in each AD against the theoretical expression Eq. 17 (dashed), which are in good agreement. The overshoot of the MSD of A_1, results from the fact that center of masses of ADs remain weakly coupled (see Eq. 35). The height of the MSD curve is inversely proportional to the total connectivity of each AD, with A_1 (Fig. 2C, orange, 26 connectors), A_2 (blue, 46 connectors) and A_3 (red, 37 connectors).

We conclude that various statistical properties can be extracted from the present heterogenous RLC polymer model that we shall use now to reconstruct chromatin in different phases of nucleus cell differentiation.

Genome reorganization with multiple TADs from the 5C data reconstructed from the heterogeneous RCL polymer during cellular differentiation

We now show how the heterogeneous RCL model can be used to reveal the chromatin properties based on 5C data of the X chromosome. We focus on the chromatin reorganization during three successive stages of differ-

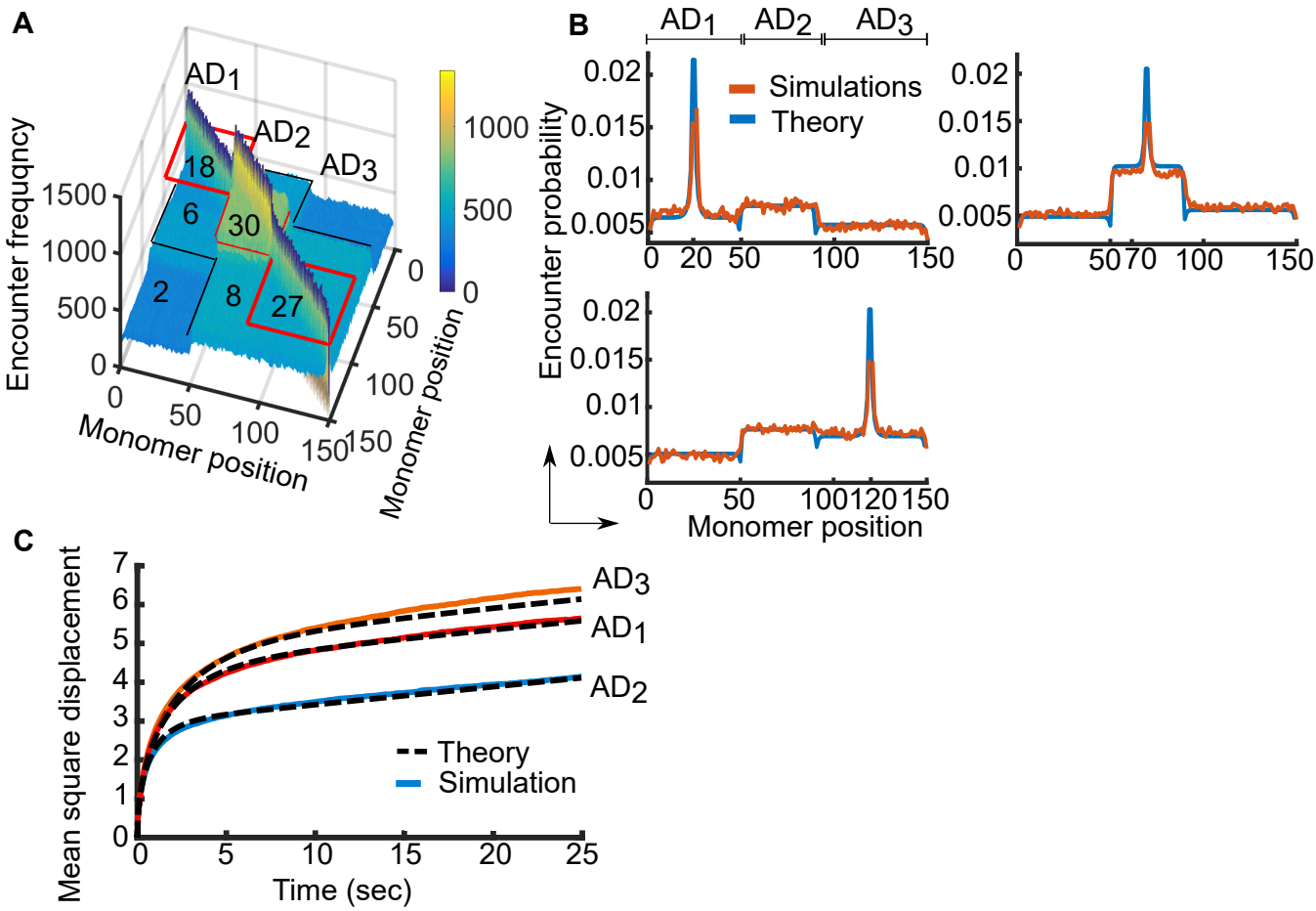

FIG. 2. **Statistical properties of the heterogeneous RCL polymer. A.** Encounter frequency matrix of a polymer with 3 AD block (AD_1, AD_2, AD_3) of $N_1 = 50, N_2 = 40, N_3 = 60$ monomers each, result of 5000 simulations of the system 29 with $\Delta t = 0.1s, D = 1, d = 3, b^2 = \sqrt{d}$. The number of added connectors appears in each block. Three distinct diagonal ADs are visible (red boxes) where secondary structure appears (black lines) due to weak inter-AD connectivity. **B.** Encounter probability (EP) of the heterogeneous RCL described in panel A, where the simulation EP (orange) is in agreement with theoretical EP (blue, Eqs. 13, 15), plotted for the middle monomer in each AD: monomer 20 (top left), monomer 70 (top right) and monomer 120 (bottom left). **C.** The mean square displacement of monomers in each AD of the heterogeneous RCL polymer, simulated as described in panel A, simulation (full) versus theory (dashed Eq. 17) are in good agreement for time up to 25 s. The MSD for AD 1 overshots at $t > 15$ because the center of masses of the three AD cannot be fully decoupled (see Eq. 35).

entiation in cell lineage commitment: undifferentiated mouse embryonic stem cells (MESC), Neuronal precursor cells (NPC) and mouse Embryonic fibroblasts (MEF). We used the average of two replica of a subset of the 5C data generated in [1], harboring 3 topologically associating domains (TADs): TAD D, E, and F, which span a genomic section of about 1.9 Mbp. We coarse-grained the 5C encounter frequency data at a scale of 6 kb, which is twice the median length of the restriction segment of the HindII enzyme used in producing the 5C data [7, 8]. At this scale we empirically found that long-range persistent peaks of the 5C encounter data are smoothed out sufficiently to enable us to use expressions 13 and 15 for fitting the 5C EP. The resulting coarse-grained encounter frequency matrix includes pair-wise encounter data of 302 equally-sized genomic segments. To determine the position of TAD boundaries, we mapped the TAD boundaries reported in bp (see [1] Supplementary Information, 'Analysis of 5C data, section) to genomic segments after coarse-graining. We constructed a heterogeneous RCL polymer with $N_D = 62, N_E = 88, N_F = 152$ monomers for TAD D, E and F, respectively. To compute the minimum number of connectors within and between TADs, we fitted the EP of each monomer in the coarse-grained empirical EP matrix using formulas 13 and 15. In Fig. 3A, we present the fitted EP matrices for MESC (left) NPC (middle) and MEF (right).

We computed the average number of connectors N_c, within and between each TAD by averaging the connectivity values obtained from the fitting the EP over all monomers to obtain the connectivity matrix Ξ and used relations 19 and 20.

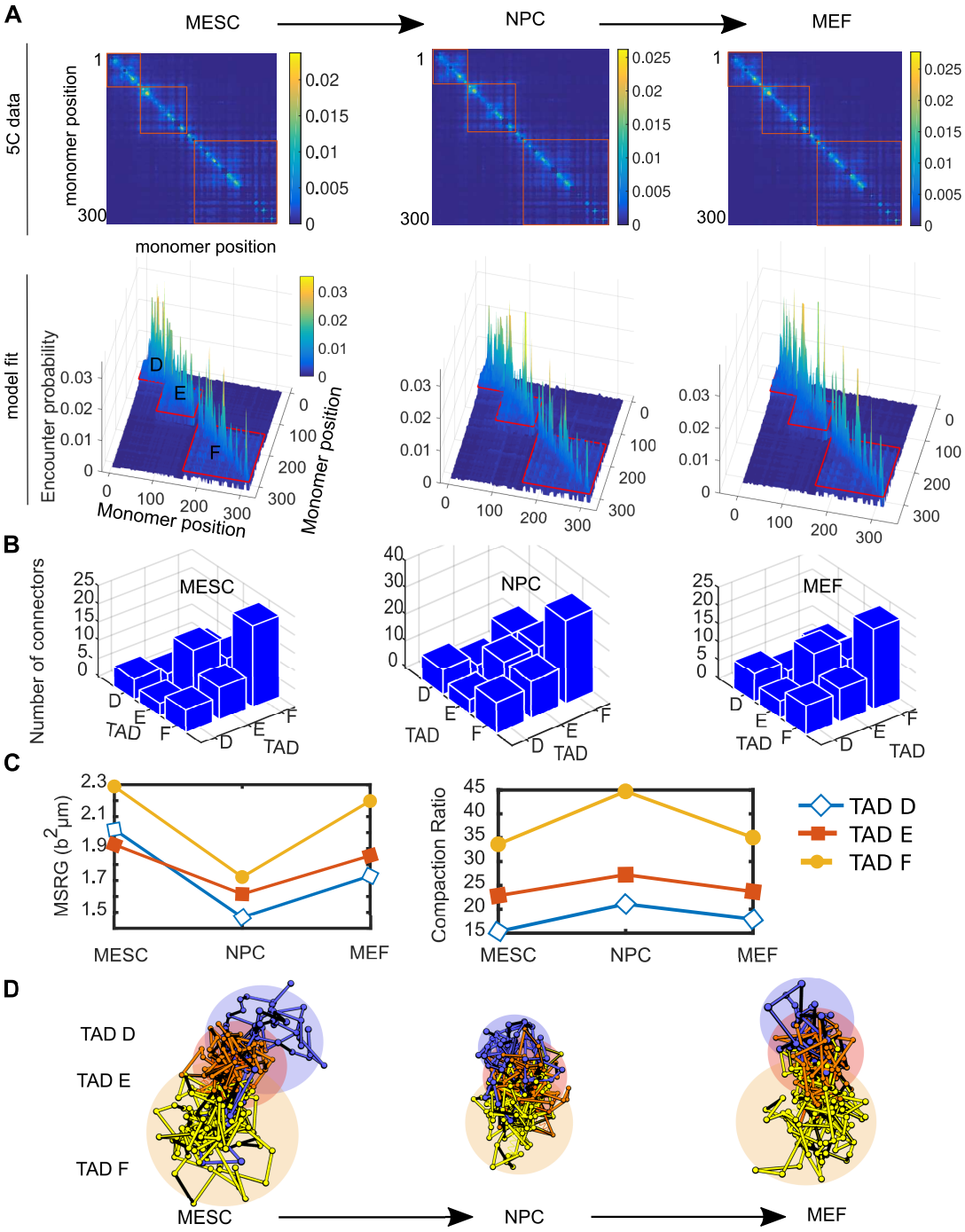

FIG. 3. **Reorganization of the X Chromosome during differentiation. A.** Hi-C maps and result of fitting expressions 13 and 15 to the empirical 5C encounter probability data [1] at a scale of 6kb for 3 TADs (red rectangles) at 3 successive stages of cell differentiation: Embryonic stem cell (MESC, left), Neuronal precursor cells (NPC, middle) and Embryonic Fibroblasts (MEF, right). **B.** Mean number of connectors in each AD obtained by fitting the EP as described in panel A, within and between TAD D, E and F for MESC (left), NPC (middle) and MEF (right). The number of connectors within TAD F grows from 22 for MSEC to 32 for NPC and drops back to 22 for MEF cells. The inter-TAD F and D connectivity increases from 9 for MSEC to 14 for NPC and decreases to 9 for MEF. The number of connectors in TAD E remains 14 throughout the 3 stages, whereas the number of connectors for TAD D increases from 6 for MSEC to 9 for NPC and decreases to 7 for MEF. **C.** Mean square radius of gyration (left) in units of b^2. The compaction ratio (right) C_r is computed as the ratio of $Nb^2/6$(MSRG of the Rouse chain) to the MSRG of each TAD with the same number of monomers, shows that the compaction in TAD E is higher than in D at NPC stage despite having a higher MSRG (1.6 and 1.3 for TAD D and E, respectively). **D.** Three realizations of the RCL polymer showing the compaction of TAD D (blue) , E (red) and F (orange) during the transition from MESC to NPC and decompaction in the transition to MEF. Shaded areas represent ball with of gyration radius, similar to in panel C for each TAD in each stage.

The mean number of connectors in the differentiation from MESC to NPC showed an increase by 145-200% (Fig. 3B) within and between TADs. The number of connectors within TAD F increases by 145% from 22 for MESC to 32 in NPC, the inter-connectivity between TAD F and E increases by 150% from 9 at MESC to 14 for NPC and the connectivity between TAD F and D doubled from 6 connectors for MESC to 12 for NPC, whereas the number of connectors within TAD E remained constant of 14. In the differentiation stage between NPC to MEF the number of connectors within TADs D, E and F returned to values comparable to MESC, whereas the inter-TAD connectivity between TAD F and E decreased from 14 for NPC to 9 for MEF.

To evaluate the size of the folded TADs, we computed the mean square radius of gyration of the three TADs throughout differentiation stages. We found that the MSRG increases and decreases in correlation with the acquisition and lose of connectors in all TADs. We Computed the MSRG (Eq. 10) for each TAD using the calibrated connectivity matrix Ξ, obtained from fitting the experimental 5C EP (Fig. 3C, left) in the units of b^2. The MSRG of all TADs decreased from average of 2 $b^2\mu m$ at MESC stage to 1.7 $b^2\mu m$ for NPC and increased back to 2 $b^2\mu m$ for MEF cells. In the transition from MESC to NPC, the MSRG of TAD D surpassed that of TAD E (Fig. 3C, red squares) due to the increased in inter-TAD connectivity between TAD E and F at NPC stage.

DISCUSSION

In this work, we constructed a general RCL model [7, 10] for multiple interacting ADs of variable size, intra and inter-AD connectivities. We provide a framework for studying a complete representation of conformation capture (3C,5C, and Hi-C) data, where we account for both intra and inter-TAD connectivity. The RCL polymer model allows estimating the average number of crosslinks within and between each TAD, length-scales such as the mean square radius of gyration to characterize the size of the folded AD, and the mean square displacement in multiple ADs. These quantities cannot be derived from the empirical conformation capture data and requires complementary method, such as single particle tracking [24, 25]. The present analysis reveals also the statistics of complex matter at the transition point between liquid and gel [26, 27] and also how transcriptional bursting dynamics is modulated by the random architecture of chromatin sub-organization.

To validate the heterogeneous RCL model, we reconstructed multiple TAD reorganization during three successive stages of cell differentiation: mouse embryonic stem cell (ESC), neuronal precursor cells (NPC) and mouse embryonic fibroblasts (MEF). Indeed, we fitted expressions 13 and 15 to the empirical 5C encounter matrices of three differentiation stages (Fig. 3A) and showed that the RCL model produced contact enriched TADs with a variability in monomer connectivity within each TAD (Fig. 3A, super and sub diagonals of EP matrices). We use the average connectivity we obtained from the fitting procedure to derive the average number of connectors between and within TADs (Fig. 3B). In the coarse-graining scale of 6 kb, our results show that the X chromosome acquires connectors within and between TADs D E and F [1] in the transition between MESC to NPC and later returns to values comparable to those of MESC at MEF stage. The increasing connectivity for NPC cells is correlated with the acquisition of LaminB1 (see [1], Figure 3). The increase in the number of connectors in each TAD at NPC stage correlated with TAD compaction. Indeed, the MSRG curves (Fig. 3C left) decrease at NPC stage, indicating a higher chromatin compaction for all TADs. The compaction ratio (in relation to linear Rouse polymer) showed high compaction at NPC stages (Fig. 3C, right) for all TADs, which can be associated with heterochromatin state, suppressed gene expression and lamina associating domains [28]. Inter-TAD connectivity is quite stable as the number of connectors for TAD E remains constant at 14 however, but the MSRG and compaction ratio decreased at NPC stage. This result indicates that the a full description of the 5C data by polymer models must take into account inter-TAD connectivity. Overall, the RCL polymer captures the correlated reorganization of TAD D, E and F during differentiation. TAD reorganization we found is also compatible with transcription co-regulation in these TADs [1].

Moreover, our construction of the RCL polymer accounts for multiple connected regions with weak inter-TAD connectivity, as suggested by experimental CC data [1, 4, 5, 11]. Despite such weak inter-AD connectivity, multiple ADs can affect a locus dynamics inside a TAD. The expression we obtained for the steady-state encounter probability (Eqs. 13 and 15) can be used to describe the 5C encounter probability matrices. We validated these expressions by Brownian simulations and showed that we can capture both TADs and higher order structures (metaTADs [11]) resulting from inter-TAD connectivity (Fig. 2A and B). We further demonstrated how the dynamics of monomers is affected by the local connectivity between and within ADs, using the MSD curves (Fig. 2C). This result explains the large variability in MSD behavior seen experimentally in single particle tracking experiments [7, 24, 29] based on the variability in local chromatin organization in cell population, represented here by the random connector localisation.

To conclude, the present framework is new tool to reconstruct systematically chromatin structural reorganization from HiC matrices and can be used to interpret chromatin capture data. In particular, using any experimental data of ligation proximity experiments (Hi-C, 5C

4C), the model reveals statistical properties beyond CC data.

APPENDIX

From HiC statistical maps to the average connectivity matrix $\langle B^{\mathcal{G}}(\Xi)\rangle$

To construct $\langle B^{\mathcal{G}}(\Xi)\rangle$, we first define the indices b_i that represent the boundary between successive blocks i and $i+1$, $1 \le i \le N_T$ (Fig. 4A):

$$b_i = \sum_{m=1}^{i} N_m, \tag{18}$$

with $b_0 = 0$. Indices b_i, b_j define the grid $T_{ij} = (b_i, b_j)$, and represent the bottom right corner of each block of $\langle B^{\mathcal{G}}(\xi)\rangle$ (Fig. 4A). The block i, j is in between $T_{i-1,j-1} + (1,1)$ to T_{ij} and it is represented by $[T_{ij}]$.
For a fixed connectivity Ξ_{ij}, the number of connectors in block $[T_{ij}]$ is

$$N_c(\xi_{ij}) = \lfloor \xi_{ij} N_L \rfloor, \tag{19}$$

where N_L is the maximal possible number of non nearest neighbor (NN) connected pairs in block $[T_{ij}]$, given by

$$N_L = \begin{cases} (N_i-1)(N_i-2)/2, & |i-j| = 0; \\ N_i N_j - 1, & |i-j| = 1; \\ N_i N_j, & \text{otherwise.} \end{cases} \tag{20}$$

We make the choice that the number $N_c(\xi_{ij})$ connected pairs is independent between blocks $[T_{ij}], (i \neq j)$, and all monomer pairs are equally likely to be chosen within each block. Due to symmetry, we shall only consider choices of monomer pairs in the upper triangular part of $\langle B^{\mathcal{G}}(\xi)\rangle$. The probability,$Pr_{ii}(k)$, of monomer $b_i < m < b_{i+1}$ to have $k \ge 0$ connectors in $[T_{ii}]$, is computed by choosing k positions in row m and the remaining $N_L - k$ connectors in any row and column $n \neq m$, leading to the hypergeometric probability

$$Pr_{ii}(k) = \begin{cases} \frac{C_k^k C_{N_L-(N_i-3)}^{N_c(\xi_{ii})-k}}{C_{N_L}^{N_c(\xi_{ij})}}, & b_i < m < b_{i+1}; \\ \frac{C_k^{N_c(\xi_{ii})} C_{N_L-(N_i-2)}^{N_c(\xi_{ii})-k}}{C_{N_L}^{N_c(\xi_{ii})}}, & m = b_i, b_{i+1}, \end{cases} \tag{21}$$

and for off-diagonal blocks $[T_{ij}]$, $j > i$

$$Pr_{ij}(k) = \frac{C_k^{N_j} C_{N_L-N_j}^{N_c(\xi_{ij})-k}}{C_{N_L}^{N_c(\xi_{ij})}}. \tag{22}$$

where $C_i^j = \frac{i!}{(i-j)!j!}$ is the binomial coefficient. The expected number of connectors $\bar{\beta}_m^{(i,j)}(\xi)$, for monomer $b_i < m < b_{i+1}$ in each block $[T_{ij}]$ is

$$\bar{\beta}_m^{(i,j)} = \begin{cases} \xi_{ij} N_j, & |i-j| > 1; \\ \xi_{ii}(N_i - 2), & |i-j| = 0, m = b_i, m = b_{i+1}; \\ \xi_{ii}(N_i - 3), & |i-j| = 0, b_i < m < b_{i+1}. \end{cases} \tag{23}$$

Here we consider that the mean number of non NN connectors $\beta_m(\xi)$ for monomer $b_i \le m \le b_{i+1}$ is the sum of the mean number of connectors in each block $[T_{ij}]$, $j = 1, .., N_T$

$$\begin{aligned} \beta_m(\Xi) &= \sum_{k=1}^{N_T} \bar{\beta}_m^{(i,k)} = \langle B(\Xi)\rangle_{mm} \\ &= \begin{cases} -3\xi_{ii} + \sum_{j=1}^{N_T} N_j \xi_{ij}, & b_i \le m \le b_{i+1}; \\ -2\xi_{ii} + \sum_{j=1}^{N_T} N_j \xi_{ij}, & m = b_i, b_{i+1}. \end{cases} \end{aligned} \tag{24}$$

The matrix $\langle B^{\mathcal{G}}(\Xi)\rangle$ can then be defined for each block $[T_{ij}]$ by

$$[T_{ij}] = \begin{cases} -\xi_{ii}\left(\mathbf{1}_{ii} + M^{(i)}\right) + I_i \sum_{k=1}^{N_T} N_k \xi_{ik}, & i = j; \\ -\xi_{ij}\mathbf{1}_{ij}, & i \neq j, \end{cases} \tag{25}$$

where $\mathbf{1}_{ij}$ is a $N_i \times N_j$ matrix of ones, $M^{(i)}$ is a Rouse matrix of $N = N_i$ monomers (Eq. 2), and I_i is an $N_i \times N_i$ identity matrix.

Diagonalization of the average connectivity matrix

To study the steady-state properties of equation 9, we decouple the system 9 into independent coordinates by diagonalizing it in the normal coordinates

$$\boldsymbol{U} = \boldsymbol{V}\boldsymbol{R} = [u_0, u_1, ...], \tag{26}$$

where

$$\boldsymbol{V} = diag([V^{(1)}], [V^{(2)}], \ldots, [V^{(N_T)}]) \tag{27}$$

is a diagonal block matrix where each block $[V^{(i)}]$ is a $N_i \times N_i$ matrix of the eigenvectors of a Rouse polymer of N_i monomers [12], given by

$$[V^{(i)}]_{m,n} = \begin{cases} \sqrt{\frac{1}{N_i}}, & m = 1; \\ \sqrt{\frac{2}{N_i}} \cos\left(\frac{m\pi}{N_i}(n - \frac{1}{2})\right), & 1 < m \le N_i. \end{cases} \tag{28}$$

We multiply 9 from the left by 27 to obtain the system of equations in the normal form

$$\frac{d\boldsymbol{U}}{dt} = -d\frac{D}{b^2}\left(\Lambda + \boldsymbol{V}\langle B^{\mathcal{G}}(\Xi)\rangle \boldsymbol{V}^T\right)\boldsymbol{U} + \sqrt{2D}\frac{d\boldsymbol{\eta}}{dt}, \tag{29}$$

where $\boldsymbol{\eta} = \boldsymbol{V}\boldsymbol{\omega}$ are Brownian motion with mean 0 and standard-deviation 1, and

$$\Lambda = diag([\Lambda^{(1)}], [\Lambda^{(2)}], \ldots, [\Lambda^{(N_T)}]) \tag{30}$$

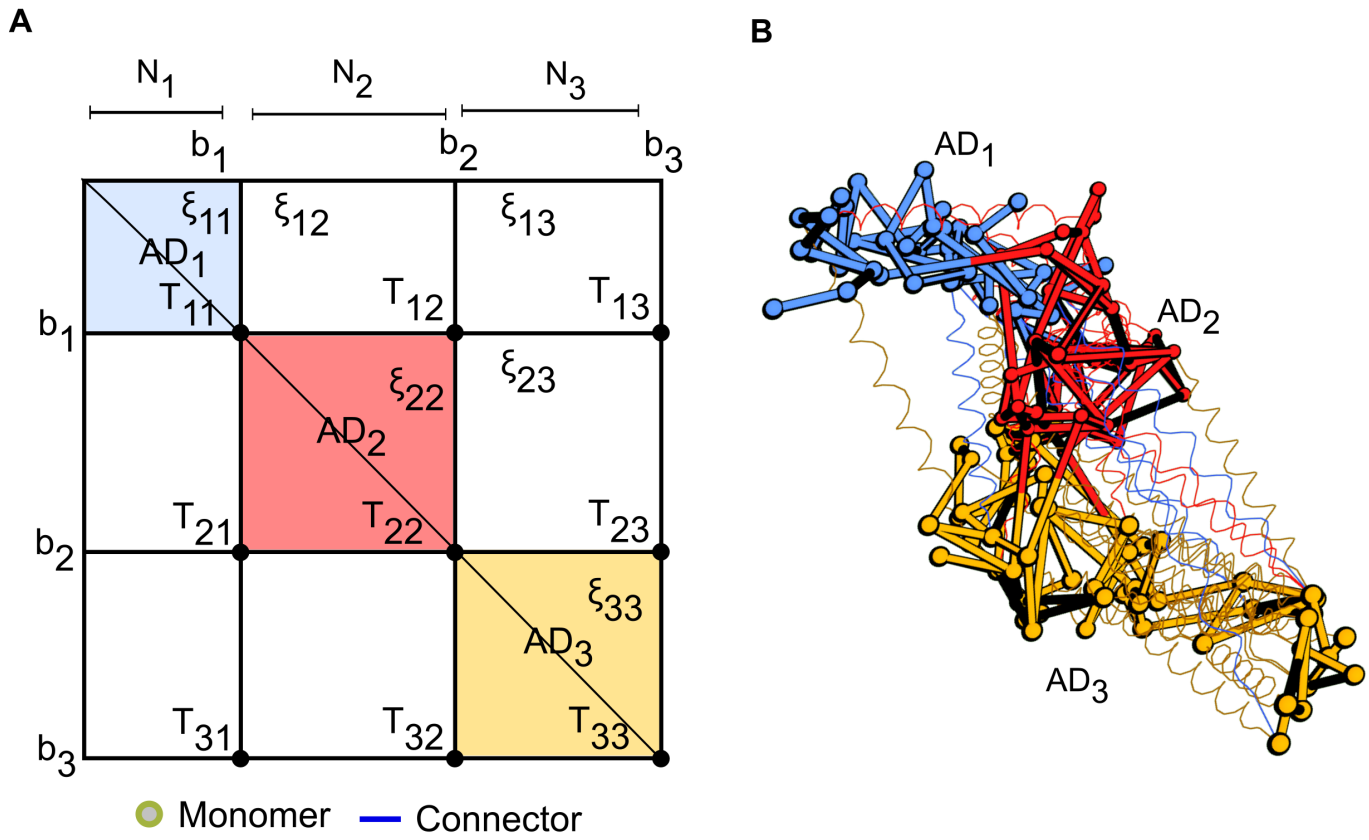

FIG. 4. **Constructing the heterogeneous RCL polymer from a matrixA.** Schematic description of 3 ADs: A_1 (blue) A_2 (red) and A_3 (orange) composed of N_1, N_2 and N_3 monomers connected linearly, containing additional random connectors within and between ADs, determined by the connectivity fraction ξ_{ij}(Eq. 19). The indices b_i represent the bottom right end of each block associated to the grid points $T_{ij} = (b_i, b_j)$, which define blocks $[T_{ij}]$. **(B)** Sample configuration of a heterogeneous RCL polymer with 3 ADs corresponding to the construction in panel A. Monomers (spheres) are connected linearly by harmonic springs to form the backbone and additional connectors (springs) are added randomly between non-nearest neighboring monomers within and between ADs.

is a $N_T \times N_T$ block diagonal matrix of the eigenvalues of Rouse chains [12] of $N_1, N_2, \ldots, N_T$ monomers defined by

$$[\Lambda^{(i)}] = diag(\lambda_0^{(i)}, \lambda_1^{(i)}, ..\lambda_{N_i-1}^{(i)}), \tag{31}$$

where

$$\lambda_p^{(i)} = 4\sin^2\left(\frac{p\pi}{2N_i}\right)), \quad p = 0, \ldots, N_i - 1 \tag{32}$$

are the Rouse eigenvalues.
From relation 25, each $[T_{ii}]$ commutes with $[M^{(i)}]$ (Eq. 2) and therefore the multiplication $\boldsymbol{V}\langle B(\Xi)\rangle\boldsymbol{V}^T$ in the right-hand side of 29, can be carried out for each block separately to obtain

$$[V^{(i)}][T_{ij}][V^{(j)}]^T = \begin{cases} -\xi_{ii}\left(N_i G_{ii} + [\Lambda^{(i)}]\right) + I_{N_i}\sum_{k=1}^{N_T} N_k \xi_{ik}, & i = j; \\ -\xi_{ij}\sqrt{N_i N_j} G_{ij}, & i \neq j, \end{cases} \tag{33}$$

where G_{ij} is a $N_i \times N_j$ matrix of zeros with 1 in its the top left cell. The stochastic equations that describes the dynamic of the normal coordinate $u_m^{(i)}$, $1 < m \leq N_i - 1$ in chain i is

$$\frac{\mathrm{d}u_m^{(i)}}{\mathrm{d}t} = -\frac{dD}{b^2}\left(\lambda_m^{(i)}(1-\xi_{ii}) + \sum_{k=1}^{N_T} N_k \xi_{ik}\right) u_m^{(i)} + \frac{\mathrm{d}\eta_m^{(i)}}{\mathrm{d}t}, \tag{34}$$

and the centers of masses $u_0^{(i)}$ is driven by

$$\frac{\mathrm{d}u_0^{(i)}}{\mathrm{d}t} = -\frac{dD}{b^2}\left(\sum_{k=1}^{N_T} N_k \xi_{ik} u_0^{(i)} - \sum_{k=1}^{N_T} \xi_{ik}\sqrt{N_i N_k} u_0^{(k)}\right) + \frac{\mathrm{d}\eta_0^{(i)}}{\mathrm{d}t}, \tag{35}$$

where $\eta_m^{(i)}$ are d-dimensional Gaussian noise with mean 0 and standard-deviation $\sqrt{2D}$.
In the following analysis, we consider the case of a dominant intra-AD connectivity, i.e

$$(H1): \quad N_i \xi_{ii} \gg \sum_{k \neq i}^{N_T} \xi_{ik}\sqrt{N_i N_k}, \tag{36}$$

which is well verified for the properties TADs of CC maps [1, 4, 5].

Mean square radius of gyration for each AD

The mean square radius of gyration (MSRG) characterizes the folded size of each AD. We derive here an expression for the distribution of the square radius of gyration of each AD under the assumption of dominant intra-AD connectivity 36. We start with the classical formula for the distribution of the square radius of gyration $P(R_g^2)$ is given by [30]

$$P(R_g^2) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(\mathtt{i_c}\beta R_g^2)\det(\mathbf{1}_{N_i-1} + \frac{\mathtt{i_c} d\beta b^2}{2N_i}[\Gamma^{(i)}]^{-1})^{-d/2} d\beta, \tag{37}$$

where $\mathtt{i_c}$ is the complex unit, $[\Gamma^{(i)}]$ is a $(N_i-1)\times(N_i-1)$ diagonal matrix of eigenvalues for the block connectivity matrix of AD i excluding the first vanishing one, and det is the determinant operator. We evaluate now the determinant in integral 37 [30]

$$\det(\mathbf{1}_{N_i-1} + \frac{\mathtt{i_c} d\beta b^2[\Gamma^{(i)}]^{-1}}{2N_i})^{-d/2} = \exp\left(\frac{d}{2}\sum_{p=1}^{\infty}\left(-\frac{2\mathtt{i_c}\beta b^2}{dN_i}\right)^p \frac{Tr([\Gamma^{(i)}]^{-p})}{p}\right), \tag{38}$$

where Tr is the trace operator. We truncate the series in 38 at order $p=2$ and substitute the resulting expressing in 37 to obtain

$$\begin{aligned} P(R_g^2) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp\left(-\frac{b^4\beta^2}{dN_i^2}Tr([\Gamma^{(i)}]^{-2}) + \mathtt{i_c}\beta(R_g^2 - \frac{b^2}{N_i}Tr([\Gamma^{(i)}]^{-1}))\right)\mathrm{d}\beta, \\ &= \sqrt{\frac{dN_i^2}{4\pi b^4 Tr([\Gamma^{(i)}]^{-2})}}\exp\left(-\frac{\left(R_g^2 - \frac{b^2 Tr([\Gamma^{(i)}]^{-1})}{N_i}\right)^2}{4b^4 Tr([\Gamma^{(i)}]^{-2})/dN_i^2}\right). \end{aligned} \tag{39}$$

Under the present approximation, the distribution of the square radius of gyration of each AD i is Gaussian with mean and variance given respectively by

$$\langle R_g^2\rangle^{(i)} = \frac{b^2 Tr([\Gamma^{(i)}]^{-1})}{N_i}, \tag{40}$$

$$\sigma^2(R_g^2) = \frac{2b^4 Tr([\Gamma^{(i)}]^{-2})}{dN_i^2}. \tag{41}$$

Under the assumption of dominant intra-AD connectivity, we approximate $[\Gamma^{(i)}]$ by $[V^{(i)}]\left([M^{(i)}]+[T_{ii}]\right)[V^{(i)}]^T$ (Eqs. 6 and 33), excluding the first vanishing entry. For $N_i \gg 1$ the trace $Tr([\Gamma^{(i)}]^{-1})$ in 39 is computed by an integral (Euler-MacLaurin formula) and the Residue theorem

$$\begin{aligned} Tr([\Gamma^{(i)}]^{-1}) &\approx Tr\left([V^{(i)}]\left([M^{(i)}]+[T_{ii}]\right)[V^{(i)}]^T\right) \\ &= \sum_{p=1}^{N_i-1}\frac{1}{\lambda_p^{(i)}(1-\xi_{ii}) + \sum_{k=1}^{N_T} N_k\xi_{ik}} \\ &\approx \frac{N_i}{2\pi(1-\xi_{ii})}\int_{-\pi}^{\pi}\frac{\mathrm{d}x}{y^{(i)}(\Xi)-\cos(x)} \\ &= -\frac{N_i}{2\pi\mathtt{i_c}(1-\xi_{ii})}\oint_{|z|=1}\frac{\mathrm{d}z}{(z-\zeta_0^{(i)}(\Xi))(z-\zeta_1^{(i)}(\Xi))} \\ &= \frac{N_i}{(1-\xi_{ii})(\zeta_0^{(i)}(\Xi)-\zeta_1^{(i)}(\Xi))}, \end{aligned} \tag{42}$$

where

$$y^{(i)}(\Xi) = 1 + \frac{\sum_{k=1}^{N_T}\xi_{ik}N_k}{2(1-\xi_{ii})}, \tag{43}$$

which depends on connectivities of all ADs connected to A_i, $\lambda_p^{(i)}$ (Eq. 32) are the eigenvalues of the Rouse block matrix 2,

and

$$\zeta_0^{(i)}(\Xi) = y^{(i)}(\Xi) + \sqrt{y^{(i)}(\Xi)^2 - 1},$$
$$\zeta_1^{(i)}(\Xi) = y^{(i)}(\Xi) - \sqrt{y^{(i)}(\Xi)^2 - 1}. \quad (44)$$

Substituting 44 and 12 into 42 and then into 40, we obtain the expression for the MSRG of AD i

$$\langle R_g^2\rangle^{(i)} \approx \frac{b^2}{(1-\xi_{ii})(\zeta_0^{(i)}(\Xi) - \zeta_1^{(i)}(\Xi))}. \quad (45)$$

Encounter probability of monomers of the heterogeneous RCL polymer

To compute the encounter probability (EP) in the heterogeneous RCL polymer, we distinguish two cases: encounters between monomers of the same AD (intra AD), and monomers of different ADs (inter AD). We now derive formulas for the EP in these two cases under a non-vanishing connectivity assumption defined by

$$(H_n)\ \xi_{ij} > 0, 1 \le i,j \le N_T. \quad (46)$$

Intra-AD encounter probabilities

For monomers $1 \le m,n \le N_i$ of AD i, expressing the position in the normal coordinates 26, we get

$$\langle (r_m^{(i)} - r_n^{(i)})^2\rangle = \frac{2}{N_i}\langle\left(\sum_{p=1}^{N_i-1}\cos\left(\frac{p(m-\frac{1}{2})\pi}{N_i}\right)u_p^{(i)} - \cos\left(\frac{p(n-\frac{1}{2})\pi}{N_i}\right)u_p^{(i)}(\Xi)\right)^2\rangle = \frac{2}{N_i}\sum_{p=1}^{N_i-1}\left(\cos\left(\frac{p(m-\frac{1}{2})\pi}{N_i}\right) - \cos\left(\frac{p(n-\frac{1}{2})\pi}{N_i}\right)\right)^2\langle u_p^{(i)}(\Xi)^2\rangle. \quad (47)$$

We recall that the variance the Ornstein-Uhlenbeck system 34 is

$$\langle u_p^{(i)}(\Xi)^2\rangle = \frac{b^2\left(1 - \exp\left(-2\kappa(\lambda_p^{(i)}(1-\xi_{ii}) + \sum_{j=1}^{N_T} N_j\xi_{pj})t\right)\right)}{\lambda_p^{(i)}(1-\xi_{ii}) + \sum_{j=1}^{N_T} N_j\xi_{ij}}, \quad (48)$$

where $\langle u_p^{(i)}(\Xi)u_q^{(i)}(\Xi)\rangle = 0, p \ne q$. The steady-state limit of 48 gives

$$\langle u_p^{(i)}(\Xi)^2\rangle = \frac{b^2}{\lambda_p^{(i)}(1-\xi_{ii}) + \sum_{j=1}^{N_T} N_j\xi_{ij}}. \quad (49)$$

Substituting 49 in 47, we obtain

$$\langle (r_m^{(i)} - r_n^{(i)})^2\rangle = \frac{2b^2}{N_i}\sum_{p=1}^{N_i-1}\frac{\left(\cos\left(\frac{p(m-\frac{1}{2})\pi}{N_i}\right) - \cos\left(\frac{p(n-\frac{1}{2})\pi}{N_i}\right)\right)^2}{\lambda_p^{(i)}(1-\xi_{ii}) + \sum_{j=1}^{N_T} N_j\xi_{pj}}. \quad (50)$$

For $N_i \gg 1$, we approximate the sum 50 using the Euler-MacLaurin formula, making a change of variable $x = \frac{p\pi}{N_i}$, $\mathrm{d}x = \frac{N_i}{\pi}\mathrm{d}p$ to obtain

$$\langle (r_m^{(i)} - r_n^{(i)})^2\rangle = \frac{2b^2}{N_i}\sum_{p=1}^{N_i-1}\frac{\left(\cos(\frac{p\pi(m-\frac{1}{2})}{N_i}) - \cos(\frac{p\pi(n-\frac{1}{2})}{N_i})\right)^2}{\sum_{j=1}^{N_T}\xi_{ij}N_j + 4(1-\xi_{ii})\sin^2(p\pi/(2N_i)))} \approx \frac{b^2}{\pi(1-\xi_{ii})}\int_{-\pi}^{\pi}\frac{\left(\cos\left(x(m-\frac{1}{2})\right) - \cos\left(x(n-\frac{1}{2})\right)\right)^2}{y^{(i)}(\Xi) - \cos(x)}\mathrm{d}x, \quad (51)$$

where $y^{(i)}(\Xi)$ is defined in expression 43. The integral in 51 is computed in the complex plane along the contour of the unit disk [10] using the residue theorem and we get

$$\sigma_{m,n}^2(\Xi) = \begin{cases} \langle R_g^2\rangle^{(i)}\left(\frac{(\zeta_0^{(i)}(\Xi)^{m-n}-1)^2 - 2\zeta_0^{(i)}(\Xi)^{m+n-1}}{\zeta_0^{(i)}(\Xi)^{2m-1}} + 2\right), m \ge n; \\ \langle R_g^2\rangle^{(i)}\left(\frac{(\zeta_0^{(i)}(\Xi)^{n-m}-1)^2 - 2\zeta_0^{(i)}(\Xi)^{m+n-1}}{\zeta_0^{(i)}(\Xi)^{2n-1}} + 2\right), m < n, \end{cases} \quad (52)$$

where $\langle R_g^2\rangle^{(i)}$ is the MSRG of A_i defined by expression 45. Because the heterogeneous RCL model belongs to the class of generalized Gaussian models [31],the EP between monomer m and n of AD i is given by

$$P^{(i)}(m,n) \propto \left(\frac{d}{2\pi\sigma_{m,n}^2(\Xi)}\right)^{d/2}. \quad (53)$$

Inter-AD encounter probabilities

We now compute the EP between monomers of A_i and A_j $(i \ne j)$. We start by computing the variance of the vector between $r_m^{(i)}$ of A_i and $r_n^{(j)}$ of A_j, using the center of masses $r_{cm}^{(i)}, r_{cm}^{(j)}$ respectively.

$$\sigma_{m^{(i)}n^{(j)}}^2(\Xi) = \langle (r_m^{(i)}) - r_n^{(j)})^2\rangle = \langle (r_m^{(i)} - r_{cm}^{(i)} + r_{cm}^{(i)} - r_{cm}^{(j)} + r_{cm}^{(j)} - r_n^{(j)})^2\rangle = \langle (r_m^{(i)} - r_{cm}^{(i)})^2\rangle + \langle (r_n^{(j)} - r_{cm}^{(j)})^2\rangle + \langle (r_{cm}^{(i)} - r_{cm}^{(j)})^2\rangle. \quad (54)$$

The variance of the vector between a monomer m of A_i and its center of mass in normal coordinates (Eq. 26) at steady-state is

$$\langle (r_m^{(i)} - r_{cm}^{(i)})^2\rangle = \frac{b^2}{2N_i(1-\xi_{ii})}\sum_{p=1}^{N_i-1}\frac{\cos^2\left(\frac{p\pi}{N_i}(m-\frac{1}{2})\right)}{y^{(i)}(\Xi) - \cos(\frac{p\pi}{N_i})} \approx \frac{-b^2}{2\pi(1-\xi_{ii})\mathtt{i_c}}\oint_{|z|=1}\frac{(z^{2m-1}+1)^2}{z^{2m-1}(z-\zeta_0^{(i)}(\Xi))(z-\zeta_1^{(i)}(\Xi))}\mathrm{d}z = \frac{b^2(1+\zeta_0^{(i)}(\Xi)^{1-2m})}{(1-\xi_{ii})(\zeta_0^{(i)}(\Xi) - \zeta_1^{(i)}(\Xi))} = \langle R_g^2\rangle^{(i)}(1+\zeta_0^{(i)}(\Xi)^{1-2m})). \quad (55)$$

Similarly for A_j, we obtain

$$\langle (r_n^{(j)} - r_{cm}^{(j)})^2 \rangle = \langle R_g^2 \rangle^{(j)} (1 + \zeta_0^{(j)}(\Xi)^{1-2n})). \tag{56}$$

Under the assumption of weak inter-AD connectivity (Assumption H1, Eq. 36), the center of masses of ADs are independent. Thus the variance of the vectors between center of masses of A_i and A_j is

$$\langle (r_{cm}^{(i)} - r_{cm}^{(j)})^2 \rangle = \frac{\langle u_0^{(i)}(\Xi)^2 \rangle}{N_i} + \frac{\langle u_0^{(j)}(\Xi)^2 \rangle}{N_j}, \tag{57}$$

from the Ornstein-Uhlenbeck equation 35 at steady-state, the variance $\langle u_0^{(i)}(\Xi)^2 \rangle$ is

$$\langle u_0^{(i)}(\Xi)^2 \rangle = \frac{b^2}{\sum_{k \neq i}^{N_T} N_k \xi_{ik}}. \tag{58}$$

Substituting 58 for A_i and A_j into 57, we obtain

$$\langle (r_{cm}^{(i)} - r_{cm}^{(j)})^2 \rangle = \frac{b^2}{N_i \sum_{k \neq i}^{N_T} N_k \xi_{ik}} + \frac{b^2}{N_j \sum_{k \neq j}^{N_T} N_k \xi_{jk}}. \tag{59}$$

Substituting expressions 55, 56 and 59 into 54, we obtain the final expression for the inter-AD variance

$$\begin{aligned} \sigma^2_{m^{(i)} n^{(j)}}(\Xi) &= \frac{b^2(1 + \zeta_0^{(i)}(\Xi)^{1-2m})}{(\zeta_0^{(i)}(\Xi) - \zeta_1^{(i)}(\Xi))(1 - \xi_{ii})} + \frac{b^2(1 + \zeta_0^{(j)}(\Xi)^{1-2n})}{(\zeta_0^{(j)}(\Xi) - \zeta_1^{(j)}(\Xi))(1 - \xi_{jj})} + b^2 \left(\frac{1}{N_i \sum_{k \neq i}^{N_T} N_k \xi_{ik}} + \frac{1}{N_j \sum_{k \neq j}^{N_T} N_k \xi_{jk}} \right) \\ &= \langle R_g^2 \rangle^{(i)} (1 + \zeta_0^{(i)}(\Xi)^{1-2m}) + \langle R_g^2 \rangle^{(j)} (1 + \zeta_0^{(i)}(\Xi)^{1-2n}) + b^2 \left(\frac{1}{N_i \sum_{k \neq i}^{N_T} N_k \xi_{ik}} + \frac{1}{N_j \sum_{k \neq j}^{N_T} N_k \xi_{jk}} \right). \end{aligned} \tag{60}$$

The EP between monomer m of AD i and n of AD j is

$$P_{m^{(i)}, n^{(j)}}(\Xi) \propto \left(\frac{d}{2\pi \sigma^2_{m^{(i)} n^{(j)}}(\Xi)} \right)^{d/2}. \tag{61}$$

Mean-Square Displacement of monomers of the heterogeneous RCL polymer

We now study the dynamical properties using the mean square displacement (MSD) of monomers in A_i. The computational method is similar to [10], where here we replace the mean number of connectors of monomers $N\xi$ of a single AD by the sum $\sum_{k=1}^{N_T} N_k \xi_{ik}$, which is the mean number of connectors of monomers in A_i when there are multiple ADs.

When the dominant AD connectivity (assumption H1, Eq. 36) is satisfied, the relaxation times $\tau_p(\xi), p = 1, ..N^{(i)} - 1$ of the internal modes of the RCL polymer representing a single AD of $N^{(i)}$ monomers is [10]

$$\tau_p(\xi) = \frac{b^2}{2dD(N\xi + 4(1-\xi)\sin^2(p\pi/(2N)))}. \tag{62}$$

Thus, for intermediate time scales $\tau_{N-1}(\xi) \leq t \leq \tau_1(\xi)$ (Eq. 62)

$$\langle \langle r_m^{(i)}(t)^2 \rangle \rangle \approx 2dD_{cm}t + \frac{db^2 Erf[\sqrt{2dDt \sum_{k=1}^{N_T} N_k \xi_{ik}/b^2}]}{2\sqrt{(1-\xi_{ii}) \sum_{k=1}^{N_T} N_k \xi_{ik}}} \tag{63}$$

where $D_{cm} = \frac{D}{\sum_{k=1}^{N_T} N_k}$. Expression 63 can be further approximated as

$$\langle \langle r_m^{(i)}(t)^2 \rangle \rangle \approx 2dD_{cm}t + \frac{db\sqrt{dDt} \left(1 + \exp(-\frac{2dDt}{b^2} \sum_{k=1}^{N_T} N_k \xi_{ik}) \right)}{\sqrt{2\pi(1-\xi_{ii})}} \tag{64}$$

Expression 64 scales as $\sqrt{t}$ similar to Rouse. This effect is due to the averaging over all realization. We also highlight how the MSD curve depends on the connectivity between A_i and all other ADs.

[1] E. P. Nora, B. R. Lajoie, E. G. Schulz, L. Giorgetti, I. Okamoto, N. Servant, T. Piolot, N. L. van Berkum, J. Meisig, J. Sedat, *et al.*, Nature **485**, 381 (2012).

[2] M. Simonis, P. Klous, E. Splinter, Y. Moshkin, R. Willemsen, E. De Wit, B. Van Steensel, and

W. De Laat, Nature genetics **38**, 1348 (2006).
[3] B. D. Pope, T. Ryba, V. Dileep, F. Yue, W. Wu, O. Denas, D. L. Vera, Y. Wang, R. S. Hansen, T. K. Canfield, *et al.*, Nature **515**, 402 (2014).
[4] J. Dekker, K. Rippe, M. Dekker, and N. Kleckner, science **295**, 1306 (2002).
[5] J. R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, and B. Ren, Nature **485**, 376 (2012).
[6] M. Tark-Dame, H. Jerabek, E. M. Manders, D. W. Heermann, and R. van Driel, PLoS Computational Biology **10**, e1003877 (2014).
[7] O. Shukron and D. Holcman, PLOS Computational Biology **13**, e1005469 (2017).
[8] L. Giorgetti, R. Galupa, E. P. Nora, T. Piolot, F. Lam, J. Dekker, G. Tiana, and E. Heard, Cell **157**, 950 (2014).
[9] E. Lieberman-Aiden, N. L. Van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, *et al.*, science **326**, 289 (2009).
[10] O. Shukron and D. Holcman, Physical Review E **96**, 012503 (2017).
[11] J. Fraser, C. Ferrai, A. M. Chiariello, M. Schueler, T. Rito, G. Laudanno, M. Barbieri, B. L. Moore, D. C. Kraemer, S. Aitken, *et al.*, Molecular systems biology **11**, 852 (2015).
[12] M. Doi and S. Edwards, *The Theory of Polymer Dynamics Clarendon* (Oxford, 1986).
[13] M. Bohn and D. W. Heermann, PloS One **5**, e12218 (2010).
[14] I. Sokolov, Physical Review Letters **90**, 080601 (2003).
[15] M. Bohn, D. W. Heermann, and R. van Driel, Physical Review E **76**, 051805 (2007).
[16] D. W. Heermann, Current Opinion in Cell Biology **23**, 332 (2011).
[17] A. Amitai and D. Holcman, Physical Review E **88**, 052604 (2013).
[18] J. Langowski and D. W. Heermann, in *Seminars in cell & developmental biology*, Vol. 18(5) (Elsevier, 2007) pp. 659–667.
[19] S. Jespersen, I. Sokolov, and A. Blumen, The Journal of Chemical Physics **113**, 7652 (2000).
[20] J. Bryngelson and D. Thirumalai, Physical review letters **76**, 542 (1996).
[21] D. Jost, P. Carrivain, G. Cavalli, and C. Vaillant, Nucleic Acids Research , gku698 (2014).
[22] D. Michieletto, E. Orlandini, and D. Marenduzzo, Physical Review X **6**, 041047 (2016).
[23] G. Fudenberg, M. Imakaev, C. Lu, A. Goloborodko, N. Abdennur, and L. A. Mirny, Cell reports **15**, 2038 (2016).
[24] S. M. Gasser, Trends in Cell Biology (2016).
[25] A. Amitai, M. Toulouze, K. Dubrana, and D. Holcman, PLoS Computational Biology **11**, e1004433 (2015).
[26] T. S. Harmon, A. S. Holehouse, M. K. Rosen, and R. V. Pappu, eLife **6** (2017).
[27] D. Hnisz, K. Shrinivas, R. A. Young, A. K. Chakraborty, and P. A. Sharp, Cell **169**, 13 (2017).
[28] D. Peric-Hupkes, W. Meuleman, L. Pagie, S. W. Bruggeman, I. Solovei, W. Brugman, S. Gräf, P. Flicek, R. M. Kerkhoven, M. van Lohuizen, *et al.*, Molecular cell **38**, 603 (2010).
[29] V. Dion and S. M. Gasser, Cell **152**, 1355 (2013).
[30] B. Eichinger, Macromolecules **13**, 1 (1980).
[31] A. A. Gurtovenko and A. Blumen, in *Polymer Analysis Polymer Theory* (Springer, 2005) pp. 171–282.