Statistics of randomly cross-linked polymer models to interpret chromatin conformation capture data

O. Shukron and D. Holcman

Group of Applied Mathematics and Computational Biology, IBENS, Ecole Normale Supérieure, 46 rue d'Ulm 75005 Paris, France

(Received 30 April 2017; published 27 July 2017)

Polymer models are used to describe chromatin, which can be folded at different spatial scales by binding molecules. By folding, chromatin generates loops of various sizes. We present here a statistical analysis of the randomly cross-linked (RCL) polymer model, where monomer pairs are connected randomly, generating a heterogeneous ensemble of chromatin conformations. We obtain asymptotic formulas for the steady-state variance, encounter probability, the radius of gyration, instantaneous displacement, and the mean first encounter time between any two monomers. The analytical results are confirmed by Brownian simulations. Finally, the present results are used to extract the mean number of cross links in a chromatin region from conformation capture data.

DOI: 10.1103/PhysRevE.96.012503

DNA in the nucleus is constantly remodeled by regulatory factors, and compacted genomic regions form transient and stable loops [1,2]. Looping is thus a key event in chromatin regulation: Although rare for a single polymer, it is frequent in a population of hierarchy folded genomes. Genome organization is probed by chromatin conformation capture (CC) techniques [3–5], which give access to simultaneous looping events in an ensemble of millions of chromatin segments. This experimental approach provides contact frequency matrices at various scales from few-kilo- to mega-base pairs (mbp). Analysis of these matrices remains difficult, but already has revealed that mammalian genomes contain mbp blocks of enriched connectivity, called topologically associating domains (TADs) [1,6]. The role of TADs and their organization remains unclear, although they are involved in gene regulation [1,4] and replication. TADs appear by averaging encounters over an ensemble of millions of samples [5] and represent steady-state looping frequencies, but do not contain either information about the size of the folded genomic section or any transient genomic encounter times.

To reconstruct chromatin at a given scale and explore its transient properties, polymer models are used as a coarse-grained representation. The Rouse model [7], characterized by nearest neighbor interactions, predicts an encounter probability (EP) that decays with $|m-n|^{-3/2}$ between monomer m and n, but cannot account for long-range interactions observed inside TADs [1,8]. Other polymer models include attractive and repulsive forces between monomers [9–14] to account for long-range interactions and have been used to probe the heterogeneous steady-state organization of the chromatin [15,16].

We study here a randomly cross-linked (RCL) polymer model used in Ref. [8] to describe the ensemble of steady-state chromatin conformations, present in CC data [1,3,6]. Cross links could be generated by either binding molecules (CTCF [1]) or by a hypothetical loop extrusion mechanism, but this exact formation mechanism is not the focus of the present model. Randomly cross-linked polymers were previously studied on a scale of a single protein molecule [17] and for cross-linked networks [18], where monomer connectivity modulates the energy landscape. The steady-state statistical properties of cross-linked polymers are similar to other physical areas, such as resistor networks, where analytical formulas were derived for the mean-square distance between resistor with prescribed connectivity, such as rings and stars [19,20]. Other applications came from the dynamic of random loop models in polymer physics [10,21,22] or fractal networks [23,24].

The RCL polymer configuration space was so far mostly explored numerically [12,16,22,25]. However, computing the encounter probability and the mean first encounter time, which are key quantities of interest to extract chromatin dynamics from the CC, was left open. We derive here formulas for the EP, the variance, and the radius of gyration of the RCL polymer that we use in a key step of chromatin reconstruction using polymer models. The present model can be used to determine from CC empirical EP the average number of cross links, a quantity inaccessible from CC experiments. We further derive an asymptotic formula for the mean first encounter time between any two monomers, which plays a key role in gene regulation [26]. Our asymptotic derivations are further confirmed by Brownian simulations.

A general procedure to extract the average number of cross links in a genomic section based on the EP decay of the 5C data [8] is available but it requires performing heavy iterative simulations. Using the present analysis, we derive an analytical formula that allows us to determine the number of loops or connectors directly from the EP of CC data.

I. THE RCL POLYMER MODEL

A linear polymer in dimension d ($d=3$) consists of N monomers with positions $\boldsymbol{R}=[r_1,r_2,\ldots,r_N]^T$, connected sequentially by harmonic springs [7], and we added spring connectors between random non-nearest-neighboring (non-NN) monomer pairs [Fig. 1(a)]. The energy of the RCL polymer introduced in Refs. [10,21] is the sum of the spring potential of linear backbone and that of random connectors

$$\phi_{\mathcal{G}}(\boldsymbol{R})=\frac{\kappa}{2}\sum_{n=2}^{N}(r_n-r_{n-1})^2+\frac{\kappa}{2}\sum_{\mathcal{G}}(r_m-r_n)^2, \qquad (1)$$

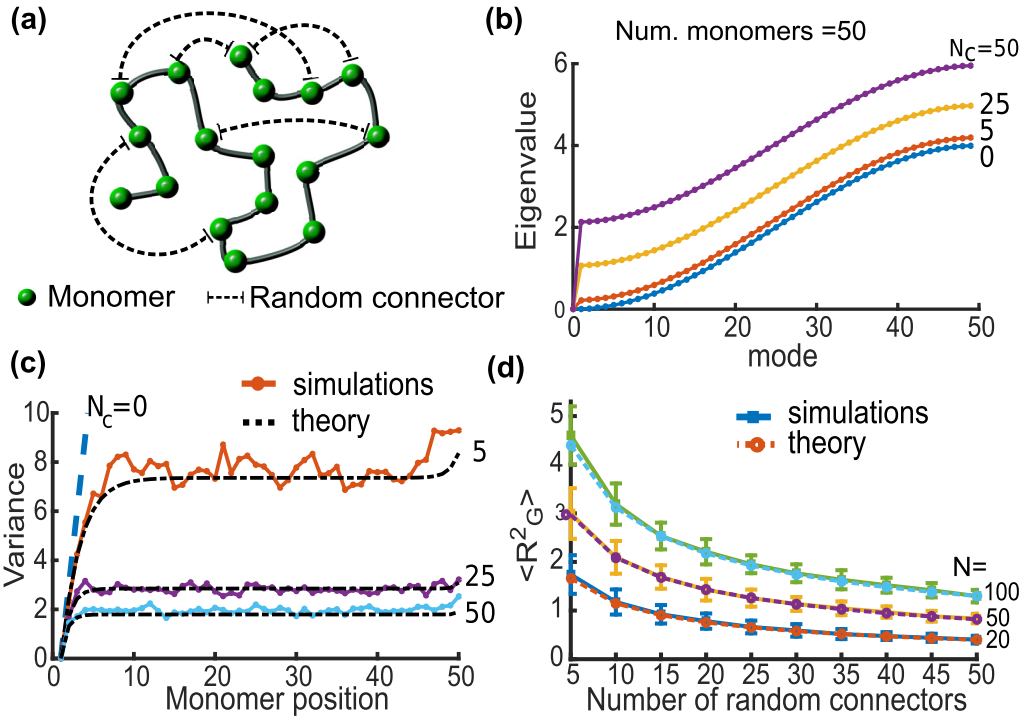

FIG. 1. Properties of a randomly cross-linked (RCL) polymer. (a) RCL polymers are composed of a linear backbone of N monomers (spheres), and $N_c(\xi)$ random connectors (dashed) between non-nearest-neighboring monomers pairs. (b) Eigenvalues of the RCL polymer [Eq. (16)] with $N = 50$ monomers, and $N_c(\xi) = 5$ (blue), 25 (red), and 50 (yellow) connectors. (c) Variance of the monomers distance: analytical [dashed, Eq. (28)] versus simulations [Eq. (13)] between monomer 1 and monomers 2–50 of the RCL polymer, 500 realizations with $N = 50, b = \sqrt{d}, D = 1, \Delta t = 0.01s$, for $N_c(\xi) = 5$ (blue), 25 (yellow), and 50 (green) added random connectors, computed after 10^4 steps, corresponding to the slowest relaxation time τ_0 [see Eq. (22)]. (d) Mean square radius of gyration $\langle R_G^2(\xi)\rangle$ with $N = 20$ (blue), 50 (yellow), and 100 (green) monomers: analytical [dashed, Eq. (32)], where $N_c(\xi) \in [5,50]$, vs stochastic simulations of 13 (continuous).

where $\kappa = dk_BT/b^2$ is the spring constant, b is the standard deviation of the connector between connected monomers, k_B is the Boltzmann's constant, and T is the temperature. The ensemble $\mathcal{G}$ is composed of N_c randomly chosen indices m,n among the non-NN monomers and this set is recomputed for each polymer realization to account for the large polymer conformational space. The connectivity fraction $0 \leqslant \xi \leqslant 1$ is the fraction of the total connector numbers $N_L = \frac{(N-1)(N-2)}{2}$, defined by

$$N_c(\xi) = \lfloor \xi N_L \rfloor. \tag{2}$$

For each polymer realization, we choose $N_c(\xi)$ pairs from N_L possible non-NN monomers, thus leading each time to a new ensemble of indices in $\mathcal{G}$ [Eq. (1)]. The dynamics of the resulting polymer model (vector $\boldsymbol{R}$) is given by Smoluchowski limit of the Langevin equation, which is the sum of Brownian motion and the gradient force induced by the potential energy [Eq. (1)],

$$\begin{aligned}\frac{d\boldsymbol{R}}{dt} &= -\frac{1}{\zeta}\nabla\phi_{\mathcal{G}}(\boldsymbol{R}) + \sqrt{2D}\frac{d\boldsymbol{\omega}}{dt} \\ &= -\frac{d}{b^2}D\left[M + B^{\mathcal{G}}(\xi)\right]\boldsymbol{R} + \sqrt{2D}\frac{d\boldsymbol{\omega}}{dt},\end{aligned} \tag{3}$$

where $D = \frac{k_BT}{\zeta}$ is the diffusion constant, ζ is the friction coefficient, $\boldsymbol{\omega}$ are independent white noises with mean 0 and variance 1, and M is the $N \times N$ Rouse matrix [7]:

$$M_{m,n} = \begin{cases} -\sum_{j\neq m} M_{m,j}, & m = n; \\ -1 & |m-n| = 1; \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

For a given connectivity fraction ξ, the square symmetric matrix $B^{\mathcal{G}}(\xi)$ with random connectivity is given by

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

The steady-state properties of an ensemble of RCL polymers are contained in the mean-field model, where we replace the matrix $B^{\mathcal{G}}(\xi)$ in Eq. (3) by its average $\langle B^{\mathcal{G}}(\xi)\rangle$ (averaging over all realizations $\mathcal{G}$ of non-NN monomer pairs when the number of connectors $N_c(\xi)$ is fixed). We construct $\langle B^{\mathcal{G}}(\xi)\rangle$ using the probability density of the monomer connectivity.

For a fixed number of connectors $N_c(\xi)$, the probability that monomer m has $k \leqslant (N-2)$ non-NN connections is obtained by choosing k position in row m of the matrix $B^{\mathcal{G}}(\xi)$ (excluding the super-, sub-, and diagonal), and the remaining $Nc - k$

connectors in any row or column $n \neq m$, thus

$$Pr_m(k) = \begin{cases} \frac{C^{N_c(\xi)-k}_{N_L-(N-3)} C^k_{N-3}}{C^{N_c(\xi)}_{N_L}}, & 1 < m < N; \\ \frac{C^{N_c(\xi)-k}_{N_L-(N-2)} C^k_{N-2}}{C^{N_c(\xi)}_{N_L}}, & m = 1, N, \end{cases} \tag{5}$$

where the binomial coefficient is $C^j_i = \frac{i!}{(i-j)!j!}$. This probability is the hypergeometric distribution for the number of connections for monomer m. The mean number of connectors for each monomer is therefore

$$\beta_m(\xi) = \begin{cases} \frac{(N-3)N_c(\xi)}{N_L} \approx (N-3)\xi, & 1 < m < N; \\ \frac{(N-2)N_c(\xi)}{N_L} \approx (N-2)\xi, & m = 1, N. \end{cases} \tag{6}$$

Using the mean values in Eq. (6), we obtain the expression for the matrix $\langle B^{\mathcal{G}}(\xi)\rangle$

$$\langle B^{\mathcal{G}}_{mn}(\xi)\rangle = \begin{cases} -\xi, & |m-n| > 1; \\ \beta_m(\xi), & m = n; \\ 0, & \text{otherwise}, \end{cases} \tag{7}$$

which can be decomposed as the sum

$$\langle B^{\mathcal{G}}(\xi)\rangle = \xi(N\boldsymbol{I_d} - \boldsymbol{M} - 1_N), \tag{8}$$

where $\boldsymbol{I_d}$ is the $N \times N$ identity matrix and 1_N is a $N \times N$ matrix of ones. To study the mean properties of the RCL polymer, we study the stochastic process (3) using the average matrix $\langle B^{\mathcal{G}}(\xi)\rangle$ in relation (8).

II. EIGENVALUES OF THE RCL POLYMER

To study the steady-state properties of system (3), we diagonalize the averaged connectivity matrix $M + \langle B(\xi)\rangle$ using Rouse normal coordinates $\boldsymbol{U} = [u_0, u_1, ..u_{N-1}]$ [7], defined as

$$\boldsymbol{U} = \boldsymbol{V}\boldsymbol{R}, \tag{9}$$

where

$$\boldsymbol{V} = (\alpha^n_p) = \begin{cases} \sqrt{\frac{1}{N}}, & p = 0; \\ \sqrt{\frac{2}{N}} \cos\left(\left(n - \frac{1}{2}\right)\frac{p\pi}{N}\right), & \text{otherwise}. \end{cases} \tag{10}$$

The Rouse orthonormal basis [7] diagonalizes M to

$$VMV^T = \Lambda = \text{diag}(\lambda_0, \lambda_1, \ldots, \lambda_{N-1}), \tag{11}$$

where

$$\lambda_p = 4\sin^2\left(\frac{p\pi}{2N}\right), \quad p = 0, \ldots, N-1, \tag{12}$$

are the eigenvalues of the Rouse matrix [relation (4)]. Substituting $\langle B^{\mathcal{G}}(\xi)\rangle$ for $B^{\mathcal{G}}(\xi)$ in system (3) and multiplying it from the left by $\boldsymbol{V}$ in (10), we obtain the mean-field equations

$$\frac{d\boldsymbol{U}}{dt} = -\frac{d}{b^2} D[\Lambda + \boldsymbol{V}\langle B^{\mathcal{G}}(\xi)\rangle \boldsymbol{V}^T]\boldsymbol{U} + \sqrt{2D}\frac{d\boldsymbol{\eta}}{dt}, \tag{13}$$

where $\boldsymbol{\eta} = \boldsymbol{V}\omega$ are independent white noises with mean 0 and variance 1. From identity (8), the matrix $\langle B^{\mathcal{G}}(\xi)\rangle$ commutes with M and therefore is diagonalizable using the same orthonormal basis $\boldsymbol{V}$:

$$\boldsymbol{V}\langle B^{\mathcal{G}}(\xi)\rangle \boldsymbol{V}^T = \text{diag}(\gamma_0(\xi), \ldots, \gamma_{N-1}(\xi)). \tag{14}$$

Using (8) and (14), we obtain the eigenvalues

$$\gamma_p(\xi) = \begin{cases} 0, & p = 0; \\ \xi(N - \lambda_p), & 1 \leqslant p \leqslant N-1. \end{cases} \tag{15}$$

To conclude, the eigenvalues of system (13) are the sum of eigenvalues of the Rouse matrix M and $\langle B^{\mathcal{G}}(\xi)\rangle$:

$$\chi_p(\xi) = \gamma_p(\xi) + \lambda_p = N\xi + 4(1-\xi)\sin^2\left(\frac{p\pi}{2N}\right). \tag{16}$$

The stochastic system (13) consists of N independent equations. For $\xi = 0$, we recover the Rouse polymer [7], whereas for $\xi = 1$, we obtain a fully connected polymer, with a circular matrix $\boldsymbol{M} + \langle B^{\mathcal{G}}(\xi)\rangle$, for which all eigenvalues is equal to N except for the first vanishing one. Using relations (9), (16), and (10) in (1), the energy of the RCL polymer is written as

$$\phi_{\mathcal{G}}(\boldsymbol{U}) = \frac{\kappa}{2}\sum_{p=1}^{N-1} \chi_p(\xi) u_p^2. \tag{17}$$

The statistics of the RCL system [relation (3)], can be recovered from (13) in the diagonalized form [expression (17)], by scaling ξ with the ratio of mean number of random connectors to the mean of total number of connectors:

$$\xi^* = \xi\frac{N_c(\xi)}{N + N_c(\xi)}. \tag{18}$$

The ensemble of eigenvalues (16) for RCL polymers, for $N = 50$ monomers, and $N_c(\xi) = 5$, 25, and 50 added random connectors is shown in Fig. 1(b).

III. ENCOUNTER PROBABILITY (EP) BETWEEN MONOMERS OF THE RCL POLYMER

The RCL polymer belongs to the class of generalized Gaussian chain models studied in Refs. [9,19,20,22], for which the EP between any two monomers m and n at equilibrium is given by

$$P_{m,n}(\xi) = \left(\frac{d}{2\pi\sigma^2_{m,n}(\xi)}\right)^{\frac{d}{2}}. \tag{19}$$

To compute expression (19) explicitly, we estimate now the variance $\sigma^2_{m,n}(\xi) = \langle(r_m - r_n)^2\rangle$ in normal coordinates [Eq. (9)]:

$$\sigma^2_{m,n}(\xi) = \sum_{p=0}^{N-1}\left(\alpha^m_p - \alpha^n_p\right)^2\langle u^2_p(\xi)\rangle. \tag{20}$$

Although computational methods to study the steady-sate variance of Gaussian models were introduced already in Ref. [20], we provide here a computation of the variance using

the normal coordinates [Eq. (9)] and the eigenvalues [Eq. (16)], which we will use below to compute time-dependent polymer properties.

The time-dependent variance is computed from the decoupled Ornstein-Uhlenbeck equations (13) [27], and we obtain

$$\left\langle u_p^2(\xi)\right\rangle = \frac{b^2}{\chi_p(\xi)}\left[1-\exp\left(-\frac{2D\chi_p(\xi)t}{b^2}\right)\right]. \tag{21}$$

The relaxation times $\tau_0 \geqslant \tau_1(\xi) \geqslant \cdots \tau_{N-1}(\xi)$ are

$$\tau_p(\xi) = \frac{b^2}{2D\chi_p(\xi)}, \tag{22}$$

and the slowest $\tau_0(\xi)$ corresponds to the diffusion of the center of mass. At steady state,

$$\left\langle u_p^2(\xi)\right\rangle = \frac{b^2}{2(1-\xi)\left[y(N,\xi)-\cos\left(\frac{p\pi}{N}\right)\right]}, \tag{23}$$

where

$$y(N,\xi) = 1+\frac{N\xi}{2(1-\xi)}. \tag{24}$$

Substituting relations (10) and (23) into (20), we get

$$\sigma_{m,n}^2(\xi) = \sum_{p=0}^{N-1}\frac{b^2\left[\cos\left(\frac{p(m-\frac{1}{2})\pi}{N}\right)-\cos\left(\frac{p(n-\frac{1}{2})\pi}{N}\right)\right]^2}{N(1-\xi)\left[y(N,\xi)-\cos\left(\frac{p\pi}{N}\right)\right]}. \tag{25}$$

For $N \gg 1$, the sum (25) is approximated by an integral (Euler Mac-Laurin formula),

$$\sigma_{m,n}^2(\xi) = \int_{-\pi}^{\pi}\frac{b^2\left[\cos\left(x\left(m-\frac{1}{2}\right)\right)-\cos\left(x\left(n-\frac{1}{2}\right)\right)\right]^2 dx}{2\pi(1-\xi)[y(N,\xi)-\cos(x)]}$$

$$= \oint_{|z|=1}\frac{-b^2(z-z^{m+n})^2(z^m-z^n)^2 dz}{4\pi i(1-\xi)[z-\zeta_0(N,\xi)][z-\zeta_1(N,\xi)]z^{2(m+n)+1}}, \tag{26}$$

where the boundaries of integration $[0,\pi]$ is transformed in the complex plane using the contour of the unit disk parameterized by $z = e^{ix}$ and we define

$$\zeta_0(N,\xi) = y(N,\xi)+\sqrt{y^2(N,\xi)-1},$$

$$\zeta_1(N,\xi) = y(N,\xi)-\sqrt{y^2(N,\xi)-1}. \tag{27}$$

When $\zeta_0(N,0)=1$, we recover from expression (25) the variance $\sigma_{m,n}^2(0) = b^2|m-n|$ of the Rouse chain [$N_c(\xi)=0$] [7]. The integrand in (26) is symmetric in m and n and has a pole of order $2(m+n)+1$ at $z=0$ and simple poles at $z=\zeta_0(N,\xi), z=\zeta_1(N,\xi)$. Because $y(N,\xi)\geqslant 1$, we have $\zeta_0(N,\xi)\geqslant 1$, which is outside the contour $|z|=1$, and $\zeta_1(N,\xi)\leqslant 1$, for all $N,\xi\geqslant 0$. The pole $\zeta_0(N,\xi)$ is not in the disk and does not contribute to the residues of (26). For $\xi > 0$, we solve the integral (26) to obtain an exact expression for the variance. With the notations $\zeta_0=\zeta_0(N,\xi)$, $\zeta_1=\zeta_1(N,\xi)$, we have

$$\sigma_{m,n}^2(\xi) = \begin{cases}\dfrac{b^2\left\{\left[\zeta_0^{m-n}-1\right]^2-2\zeta_0^{m+n-1}+2\zeta_0^{2m-1}\right\}}{(1-\xi)(\zeta_0-\zeta_1)\zeta_0^{2m-1}}, m\geqslant n;\\ \dfrac{b^2\left[\left(\zeta_0^{n-m}-1\right)^2-2\zeta_0^{m+n-1}+2\zeta_0^{2n-1}\right]}{(1-\xi)(\zeta_0-\zeta_1)\zeta_0^{2n-1}}, m<n.\end{cases} \tag{28}$$

For $0<\xi\ll 1, k>1$, we approximate the terms (27) by

$$\zeta_0^k(N,\xi)\approx\exp(k\sqrt{N\xi});$$
$$\zeta_1^k(N,\xi)\approx\exp(-k\sqrt{N\xi}); \tag{29}$$

and use (29) in expression (28) to obtain the asymptotic expression for the variance

$$\sigma_{m,n}^2(\xi)\approx\frac{b^2}{\sqrt{N\xi}}[1-\exp(-|m-n|\sqrt{N\xi})]. \tag{30}$$

To check the range of validity of formula (28), we use Brownian simulations [Fig. 1(c)], computed after a relaxation time τ_0 (10^4 numerical steps) for $N=50, N_c=5,25$, and 50. Substituting relation (28) in (19), we obtain an expression for the steady-state EP $P_{m,n}(\xi)$ between any two monomers. We then compare the EP obtained from Brownian simulations of RCL polymer for $N=20,50$ with the analytical formula 19 for $N_c(\xi)=25,50$ connectors [Fig. 2(a)], which shows very good agreement.

IV. MEAN SQUARE RADIUS OF GYRATION (MSRG) OF THE RCL POLYMER

The mean square radius of gyration (MSRG) $\langle R_G^2(\xi)\rangle$ characterizes the size of the RCL polymer and can be computed from the expression of variance (28) by the following formula [7]:

$$\left\langle R_G^2(\xi)\right\rangle = \frac{1}{N^2}\sum_{m=1}^{N}\sum_{n=1}^{m}\sigma_{m,n}^2(\xi). \tag{31}$$

To compute the sum (31), we use elementary formula for the sum of geometric series. Using relation (28) in (31) and with the notations $\zeta_0=\zeta_0(N,\xi)$, $\zeta_1=\zeta_1(N,\xi)$, we obtain

$$\left\langle R_G^2(\xi)\right\rangle = \frac{b^2}{N^2(1-\xi)(\zeta_0-\zeta_1)}\left[\frac{(1+2\zeta_0)N(1+N)}{2\zeta_0}+\frac{N\left[2(1+\zeta_0)^2-\zeta_0^3\right]}{1-\zeta_0^2}-\frac{\zeta_0^3\left(1-\frac{1}{\zeta_0^{2N}}\right)}{\left(1-\zeta_0^2\right)^2}+\frac{2(1+\zeta_0)\left(1-\frac{1}{\zeta_0^N}\right)}{(1-\zeta_0)^2}\right]. \tag{32}$$

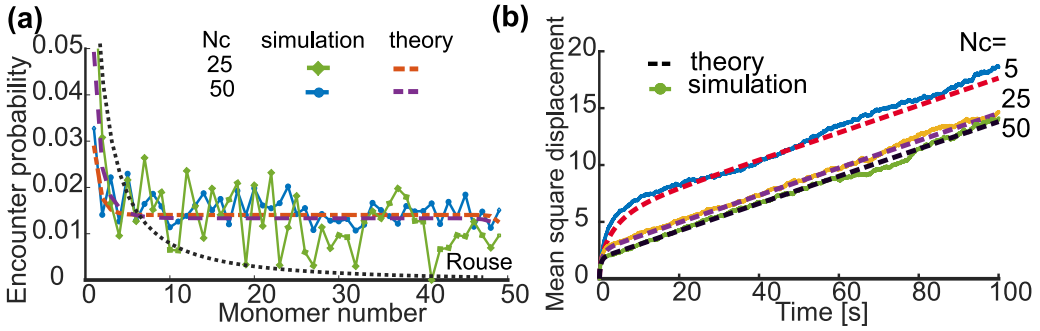

FIG. 2. Properties of the RCL polymer: (a) Encounter probability between monomers 1 and 2–50, simulated from Eq. (3). The statistics of the simulations is recorded after the slowest relaxation time [Eq. (22)]. Parameters are $N = 50$ monomers; $N_c(\xi) = 25$ (green diamonds) and 50 (blue circles) connectors. We average over 500 realizations (changing each time the ensemble $\mathcal{G}$) and compare with the analytical formula, Eq. (19) (dashed curves), with $D = 1, d = 3, b = \sqrt{d}, \epsilon = b/10, \Delta t = 0.01s$. The encounter probability of the Rouse polymer where $N_c(\xi) = 0$ (dotted black), which cannot account for long-range connectivity. (b) Mean square displacement simulations for a RCL polymer, where $N = 50$ monomers and $N_c(\xi) = 5$ (blue), 25 (yellow), and 50 (green) added connectors. The formulas [Eq. (35)] are shown in dashed lines.

In the low connectivity case $N_c(\xi) \ll \frac{N^2}{2}$, we use (29) in (32) and by discarding terms of higher order in $O(N^{-1})$, we obtain the asymptotic expansion

$$\left\langle R_G^2(\xi)\right\rangle \approx \frac{3b^2}{4(1-\xi)\sqrt{N\xi}}. \tag{33}$$

In Fig. 1(d), we compare formula (32) with $\langle R_G^2(\xi)\rangle$ computed from Brownian simulations for $N = 20{,}50$, and 100 monomers and $N_c(\xi) \in [5{,}50]$ added random connectors, and both agree.

V. MEAN SQUARE DISPLACEMENT (MSD) OF A SINGLE MONOMER OF THE RCL POLYMER

Using the normal coordinate system (9) in dimension d, the mean square displacement (MSD) of monomers in the RCL polymer is given by

$$\begin{aligned}\left\langle r_m^2(t)\right\rangle &= \left\langle \left(\sum_{p=0}^{N-1} \alpha_p^m u_p(t)\right)^2 \right\rangle \\ &= \frac{2dDt}{N} + \sum_{p=1}^{N-1} \left(\alpha_p^m\right)^2 \left\langle u_p^2\right\rangle = 2dD_{cm}t \\ &\quad + \frac{2db^2}{N} \sum_{p=1}^{N-1} \frac{\cos^2\left(\frac{p\pi(m-\frac{1}{2})}{N}\right)\left[1-\exp\left(-\frac{2D\chi_p(\xi)t}{b^2}\right)\right]}{\chi_p(\xi)},\end{aligned} \tag{34}$$

where we used $\langle u_p, u_q\rangle = 0, \forall p \neq q$, and $D_{cm} = \frac{D}{N}$. Averaging over all monomers and approximating the sum in (34) by an integral (Eurler Mac-Laurin formula) for $N \gg 1$, we obtain

$$\begin{aligned}\left\langle\left\langle r_m^2(t)\right\rangle\right\rangle &= 2dD_{cm}t + \frac{2db^2}{N^2} \sum_{p=1}^{N-1} \frac{\left[1-\exp\left(-\frac{2D\chi_p(\xi)t}{b^2}\right)\right]}{\chi_p(\xi)} \sum_{m=1}^{N} \cos^2\left[\frac{p\pi(m-1/2)}{N}\right] = 2dD_{cm}t + \frac{db^2}{\pi}\int_0^\pi dx \frac{1-e^{\left(-\frac{2dD\chi_x(\xi)t}{b^2}\right)}}{\chi_x(\xi)} \\ &= 2dD_{cm}t + \frac{db^2}{\sqrt{\pi N\xi(1-\xi)}}\int_0^{\sqrt{2dDN\xi t/b^2}} \exp(-g^2)dg = 2dD_{cm}t + \frac{db^2\mathrm{Erf}[\sqrt{2DN\xi t/b^2}]}{2\sqrt{N\xi(1-\xi)}},\end{aligned} \tag{35}$$

where $\mathrm{Erf}[t]$ is the error function. Equation (35) characterizes the MSD for intermediate time scale $\tau_{N-1}(\xi) \ll t \ll \tau_1(\xi)$. For short time scale $t \ll \tau_{N-1}(\xi)$, the MSD is approximated by

$$\begin{aligned}\left\langle\left\langle r_m^2(t)\right\rangle\right\rangle &= \frac{b^2\int_0^{\sqrt{2dDN\xi t/b^2}} \exp(-g^2)dg}{\sqrt{\pi N\xi(1-\xi)}} \\ &\approx \frac{b\sqrt{2dDt}}{\sqrt{\pi(1-\xi)}}\left[1-\frac{\exp(-2dDN\xi t/b^2)}{2}\right].\end{aligned}$$

Thus, for $N\xi \gg 1$, the MSD behaves like

$$\left\langle\left\langle r_m^2(t)\right\rangle\right\rangle \propto \frac{db\sqrt{2dDt}}{\sqrt{\pi(1-\xi)}}. \tag{36}$$

We conclude that the homogeneous behavior of MSD for the RCL polymer model gives an anomalous exponent $\alpha = 0.5$, similar to the Rouse model due to the mean-field approximation. This result is in contrast with the spectrum of anomalous exponents obtained for each configuration [8]. Finally, for long time scales [$t \gg \tau_1(\xi)$] (slow diffusion of the polymer's center

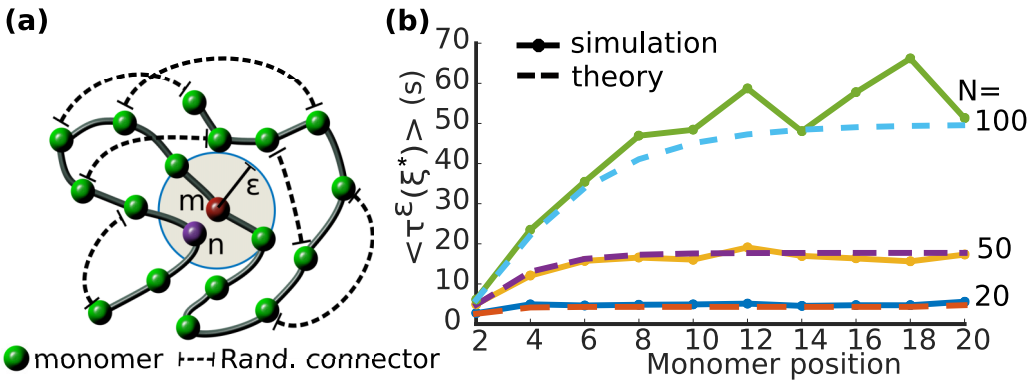

FIG. 3. Transient RCL polymer properties: (a) Two monomers m (red) and n (purple) meet when they enter a ball of radius ϵ. Random connectors (dashed arrows) are added to a linear Rouse chain. (b) Stochastic simulations (dots) of the MFET between monomer 1 and monomers 2–20 of RCL polymers with $N = 20$ (blue), 50 (yellow), and 100 (green) monomers, with $N_c(\xi) = 25$ random connectors, agree with the formula in Eq. (42) (dashed). Parameters: $\epsilon = b/10, D = 1, b = \sqrt{3}, \Delta t = 0.01s$, the RCL system is 3 [we used Eq. (42) with ξ^*, Eq. (18)].

of mass), the error function in (35) is almost constant and therefore

$$\langle\langle r_m(t)^2\rangle\rangle \approx 2d D_{cm} t + \frac{db^2}{2\sqrt{N\xi(1-\xi)}}. \tag{37}$$

VI. MEAN FIRST ENCOUNTER TIME (MFET) $\langle\tau^\epsilon(\xi)\rangle$ BETWEEN MONOMERS OF THE RCL POLYMER

We compute here the mean time for two monomers of the RCL polymer to enter for the first time in a ball of radius $\epsilon > 0$, at which they can possibly interact [Fig. 3(a)]. The MFET for both the Rouse and β polymer [13,28] were computed [29] using the regular expansion with respect to $\epsilon > 0$ of the first eigenvalue λ_0^ϵ of the Fokker-Planck operator associated to the stochastic equation (13), so that

$$\langle\tau^\epsilon(\xi)\rangle \approx \frac{1}{D\lambda_0^\epsilon(\xi)}. \tag{38}$$

The first-order approximation in ϵ is given by [29]

$$\lambda_0^\epsilon(\xi) = \frac{4\pi\epsilon\int_{C-P} e^{-\phi_{\mathcal{G}}(U)}dU}{|\tilde{\Omega}(\xi)|} + O(\epsilon^2), \tag{39}$$

where $\phi_{\mathcal{G}}(U)$ is the diagonalized potential (17) and $|\tilde{\Omega}(\xi)|$ is the integral over the entire RCL configuration space, computed using Gaussian integrals

$$|\tilde{\Omega}(\xi)| = \int e^{-\phi_{\mathcal{G}}(U)}d\boldsymbol{U} = \int \prod_{p=1}^{N} e^{-\frac{\kappa}{2}\chi_p(\xi)u_p^2(\xi)}d\boldsymbol{U}$$
$$= \left[\frac{(2\pi)^{N-1}}{\prod_{p=1}^{N-1}\kappa\chi_p(\xi)}\right]^{\frac{d}{2}}. \tag{40}$$

The integral over the space $C - P$ of the closed RCL polymer ensemble with fixed connector between monomers m and n and additional $N_c(\xi)$ random connectors in relation (39) is computed directly and gives [30]

$$\int_{C-P} e^{-\phi_{\mathcal{G}}(U)}d\boldsymbol{U}$$
$$= \int e^{-\phi_{\mathcal{G}}(U)}\delta\left[\sum_{p=1}^{N}(\alpha_p^m - \alpha_p^n)u_p^2\right]d\boldsymbol{U}$$
$$= (2\pi)^{\frac{(N-2)d}{2}}\left[\frac{\kappa}{2}b^2\prod_{p=1}^{N-1}e^{-[\kappa\chi_p(\xi)]}\sigma_{m,n}^2(\xi)\right]^{\frac{d}{2}}, \tag{41}$$

where δ is the δ function. Using relations (40) and (41) in (38), we obtain the MFET between any two monomers m and n of the RCL polymer for a given connectivity fraction ξ in dimension $d = 3$:

$$\langle\tau_{m,n}^\epsilon(\xi)\rangle = \frac{1}{4\pi D\epsilon}\left[\frac{2\pi\sigma_{m,n}^2(\xi)}{\kappa b^2}\right]^{\frac{3}{2}}, \tag{42}$$

Using (30) into (42), we obtain the new looping formula

$$\left\langle\tau_{m,n}^\epsilon(\xi)\right\rangle \approx \frac{b^2[1-\exp(-|m-n|\sqrt{N\xi})]^{d/2}}{4\sqrt{N\xi}\pi D\epsilon(\kappa b^2)^{d/2}} + \mathcal{O}(N\xi),$$

where $|m-n| \ll N$ and $\xi \ll 1$. The analytical formula (42) agrees with Brownian simulations of the MFET for the RCL polymer [Eq. (3)] with $N = 20{,}50$, and 100 monomers, and $N_c(\xi) = 25$ added random connectors [Fig. 3(b)].

VII. APPLICATIONS OF THE RCL POLYMER MODEL TO CHROMATIN RECONSTRUCTION

We derived here several analytical formulas for the variance, encounter probability, radius of gyration, mean-square displacement, and the mean first encounter time of RCL polymer models. These formulas can be used to extract parameters from CC experiments [1,6]. Formula (19) can be used to fit the empirical encounter probability to extract the connectivity fraction ξ. This parameter has a direct interpretation and represents the mean number of cross links that can be mediated by CTCF molecules present in a genomic region.

The parameter ξ depends on the coarse-grained scale [8] and is used directly to estimate the radius of gyration [Eq. (32)] of any region of interest. This radius characterizes the size of the folded genomic region relative to other genomic segments. It also provides insight into the relative compaction of TADs and local organization of the chromatin in the cell nucleus.

To demonstrate the applicability of the present method, we coarse-grained the 5C data reported in Ref. [1] of male neuronal progenitors NPC-E14 cells, replicate 1. Coarse-graining was performed at a scale of 3 kbp according to the method presented in Ref. [31]. The assumptions of the RCL model require that monomers share similar average connectivity, and thus we took only a subset of the 5C data containing TAD H. The TAD did not contain any long-range persistent loops (peaks) and we decided to test the present model. The length of the genomic section in TAD H is 679 kbp, which after coarse graining resulted in a polymer of $N = 226$ monomers. We fit the EP [Eq. (19)] of each of the 226 monomers as explained in Ref. [8] and obtained an average connectivity of $\xi = 0.0022$, corresponding to $N_c(\xi) = 56$ added connectors, that could be interpreted as the number of binding molecules. Fitting the EP of TAD H with a power law $a|m-n|^{-\beta}$ lead to $\beta = 0.77$, showing that the Rouse ($\beta = 1.5$) model is inadequate to represent the empirical EP. With persistence length of $b = 0.05\ \mu$m and $N = 226$, $\xi = 0.0022$ in Eq. (32), we predicted that the radius of gyration is 43 nm for TAD H. Thus, the 679-kbp TAD H is compacted into a ball of volume 3.4×10^5 nm^3 (2 bp per nm^3).

Finally, a possible test for the robustness of the RCL to coarse graining at any scale is that the value of the MSRG should persist. By coarse graining, we change the number N of monomers and the variance b, which should be known experimentally for each scale. In the absence of such knowledge, from Eq. (30) or (33), we see that to keep the MSRG constant for all scales, b^2 needs to be proportional to $\sqrt{N\xi}$, the square root of the mean number of connectors. Here, the coarse graining is imposed by the 5C protocol at resolution 3 kb. We change the coarse graining from 3- to 10-kb resolution of TAD H and we find that the number of connectors decreases from 56 at 3 kb to 7 at 10 kb and thus b^2 should increase from 0.025 to 0.075 μm at 10-kb resolution.

Another application of the present analysis is the fitting of the MSD [Eq. (35)] to single-particle trajectory data: By fitting the experimental MSD curves using Eqs. (35)–(37), we obtain the degree of connectivity ξ. We can then interpret the mean deviation of loci dynamic from pure diffusion as the confinement due to cross-linked genomic environment [8,32,33].

To conclude, the main goal of this paper was to derive asymptotic formulas to extract the connectivity ξ or the mean number of connectors of a polymer model to account for the block matrices (topological associated domains) present in CC data (3C, 5C, and Hi-C). The procedure consists in fitting the EP of the RCL model [Eq. (19)] to CC data to extract the value of connectivity that can later be used in formula (42) to compute the mean first encounter time between any two monomers and thus for any two genes of interest. Encounter times are key for understanding processes, such as mammalian X chromosome inactivation [1] or non-homologous-end joining after DNA double-strand break [29,34].

[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 (London) **485**, 381 (2012).

[2] M. Tark-Dame, H. Jerabek, E. M. M. Manders, D. W. Heermann, and R. van Driel, PLoS Comput. Biol. **10**, e1003877 (2014).

[3] J. Dekker, K. Rippe, M. Dekker, and N. Kleckner, Science **295**, 1306 (2002).

[4] M. Simonis, P. Klous, E. Splinter, Y. Moshkin, R. Willemsen, E. De Wit, B. V. Steensel, and W. De Laat, Nat. Genet. **38**, 1348 (2006).

[5] E. Lieberman-Aiden, N. L. V. 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).

[6] J. R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, and B. Ren, Nature (London) **485**, 376 (2012).

[7] M. Doi and S. Edwards, *The Theory of Polymer Dynamics Clarendon* (Oxford University Press, Oxford, UK, 1986).

[8] O. Shukron and D. Holcman, PLoS Comput. Biol. **13**, e1005469 (2017).

[9] I. M. Sokolov, Phys. Rev. Lett. **90**, 080601 (2003).

[10] M. Bohn, D. W. Heermann, and R. van Driel, Phys. Rev. E **76**, 051805 (2007).

[11] M. Bohn and D. W. Heermann, PloS ONE **5**, e12218 (2010).

[12] D. W. Heermann, Curr. Opin. Cell Biol. **23**, 332 (2011).

[13] A. Amitai and D. Holcman, Phys. Rev. E **88**, 052604 (2013).

[14] J. Langowski and D. W. Heermann, Seminars in Cell & Developmental Biology **18**, 659 (2007).

[15] D. Jost, P. Carrivain, G. Cavalli, and C. Vaillant, Nucl. Acids Res. **42**, 9553 (2014).

[16] M. Barbieri, M. Chotalia, J. Fraser, L.-M. Lavitas, J. Dostie, A. Pombo, and M. Nicodemi, Proc. Natl. Acad. Sci. USA **109**, 16173 (2012).

[17] Y. Suzuki, J. K. Noel, and J. N. Onuchic, J. Chem. Phys. **134**, 245101 (2011).

[18] C. Roos, A. Zippelius, and P. M. Goldbart, J. Phys. A **30**, 1967 (1997).

[19] A. A. Gurtovenko and A. Blumen, in *Polymer Analysis Polymer Theory* (Springer, Berlin, 2005), pp. 171–282.

[20] B. Eichinger, Macromolecules **13**, 1 (1980).

[21] J. D. Bryngelson and D. Thirumalai, Phys. Rev. Lett. **76**, 542 (1996).

[22] S. Jespersen, I. M. Sokolov, and A. Blumen, J. Chem. Phys. **113**, 7652 (2000).

[23] S. Alexander and R. Orbach, J. Phys. Lett. **43**, 625 (1982).

[24] T. Nakayama, K. Yakubo, and R. L. Orbach, Rev. Mod. Phys. **66**, 381 (1994).

[25] E. R. Duering, K. Kremer, and G. S. Grest, Phys. Rev. Lett. **67**, 3531 (1991).

[26] S. Kadauke and G. A. Blobel, Biochim. Biophys. Acta **1789**, 17 (2009).

[27] Z. Schuss, *Theory and Applications of Stochastic Processes: An Analytical Approach*, Applied Mathematics, Vol. 170 (Springer, Berlin, 2009).
[28] R. W. Pastor, R. Zwanzig, and A. Szabo, J. Chem. Phys. **105**, 3878 (1996).
[29] A. Amitai, I. Kupka, and D. Holcman, Phys. Rev. Lett. **109**, 108302 (2012).
[30] A. Amitai, A. Seeber, S. M. Gasser, and D. Holcman, Cell Rep. **18**, 1200 (2017).
[31] L. Giorgetti, R. Galupa, E. P. Nora, T. Piolot, F. Lam, J. Dekker, G. Tiana, and E. Heard, Cell **157**, 950 (2014).
[32] S. C. Weber, J. A. Theriot, and A. J. Spakowitz, Phys. Rev. E **82**, 011913 (2010).
[33] S. C. Weber, A. J. Spakowitz, and J. A. Theriot, Phys. Rev. Lett. **104**, 238102 (2010).
[34] A. Amitai and D. Holcman, Phys. Rep. **678**, 1 (2017).