Biophysical Journal
Article

Biophysical Society

Leveraging polymer modeling to reconstruct chromatin connectivity from live images

Sayantan Dutta,[1] Ashesh Ghosh,[1] Alistair N. Boettiger,[2] and Andrew J. Spakowitz[1,3,4,*]
[1]Department of Chemical Engineering, Stanford University, Stanford, California; [2]Department of Developmental Biology, Stanford University, Stanford, California; [3]Department of Materials Science and Engineering, Stanford University, Stanford, California; and [4]Program in Biophysics, Stanford University, Stanford, California

ABSTRACT Chromosomal dynamics plays a central role in a number of critical biological processes, such as transcriptional regulation, genetic recombination, and DNA replication. However, visualization of chromatin is generally limited to live imaging of a few fluorescently labeled chromosomal loci or high-resolution reconstruction of multiple loci from a single time frame. To aid in mapping the underlying chromosomal structure based on parsimonious experimental measurements, we present an exact analytical expression for the evolution of the polymer configuration based on a flexible-polymer model, and we propose an algorithm that tracks the polymer configuration from live images of chromatin marked with several fluorescent marks. Our theory identifies the resolution of microscopy needed to achieve high-accuracy tracking for a given spacing of markers, establishing the statistical confidence in the assignment of genome identity to the visualized marks. We then leverage experimental data of locus-tracking measurements to demonstrate the validity of our modeling approach and to establish a basis for the design of experiments with a desired resolution. Altogether, this work provides a computational approach founded on polymer physics that vastly improves the interpretation of in vivo measurements of biopolymer dynamics.

SIGNIFICANCE The dynamics of chromatin structure is important for understanding a number of in vivo biological processes, such as gene regulation and replication of the genetic material. In this article, we develop a polymer-physics-based method of reconstructing chromatin architecture by finding the most probable connectivity between indistinguishable spots from live imaging. Combining theory, simulation, and live imaging data, we identify the boundaries in length scale and timescale where the algorithm functions with high accuracy. This framework, coupled with live imaging of chromosomes fluorescently labeled at multiple points, establishes a basis for determining the interplay of the polymer structure, biological processes, and the surrounding medium impacting the dynamics of chromatin.

INTRODUCTION

Recent advances in imaging, gene editing, and fluorescent marking of chromosomal loci enable the study of chromosomal organization and dynamics at exquisite detail (1–8). However, live imaging of chromosomes with different colored marks remains experimentally challenging due to spectral overlap between fluorophores, limitations in orthogonal attachment chemistries, and restrictions in spatial and temporal resolution of imaging. Fully exploiting live imaging of chromosomes requires a reliable approach to reconstruct the dynamic chromosomal architecture from live images of multiple marks with single colors, which would enable us to interpret how the observed motion relates to the governing chromosomal physics and active biological processes.

Polymer modeling has emerged as an essential tool for interpreting experimental measurements of chromosomal organization and dynamics. Tracking of individual chromosomal loci results in complex behavior such as subdiffusive motion and temporal memory (9–13), which are well described using classical polymer physics models (14) adapted to include environmental viscoelasticity (15), active biological processes (16–21), and architectural looping (22–24). Extending these experimental methods to the imaging (4–6,25) and tracking (26,27) of multiple genomic loci poses additional challenges for interpreting the motion and predicting the underlying structure. However, polymer modeling is well suited to reveal the physical effects driving the interlocus coordination within the observed behavior

Submitted March 31, 2023, and accepted for publication August 1, 2023.
*Correspondence: ajspakow@stanford.edu
Editor: Alexandra Zidovska.
https://doi.org/10.1016/j.bpj.2023.08.001

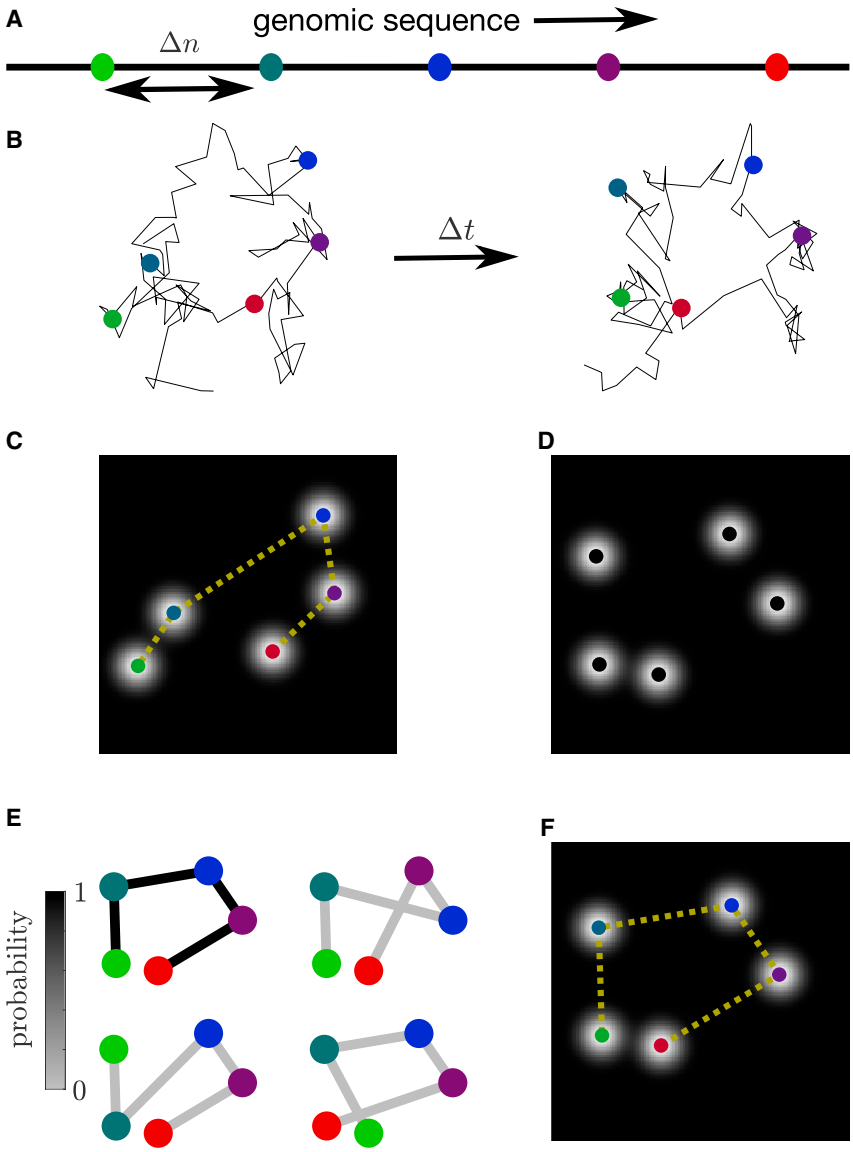

FIGURE 1 A demonstration of the algorithm of reconstructing chromatin connectivity. (*A*) Schematic of five different fiducial marks spaced with spacing Δn along a genomic sequence. (*B*) Chromosomal (polymer) configurations at two consecutive time points separated by time Δt from Rouse polymer simulations. (*C* and *D*) Synthetic images demonstrating live imaging of fiducial marks tagged with fluorescent label of the same color. The colored circles in (*C*) denote the genomic identity of spots shown in (*A*), whereas identity of spots is unknown at the later timepoint (*D*). (*E*) Four assignments of spots to different permutations of genomic positions in order of probability. The darker connecting lines indicate higher probability. (*F*) True assignment of the polymer positions from the simulated model. The most probable assignment is the same as the ground truth representing an ideal scenario of tracking.

(28,29). Genome-scale mapping of genomic contacts using conformation capture techniques, such as Hi-C (1–3,30), has provided invaluable insight into the multiscale organization of chromosomal DNA, and physical insight into the driving forces behind compartmentalization and local associations is captured by polymer models that incorporate organizational proteins and epigenetic marks (23,31–42).

Here, we propose an algorithm based on the principles of polymer physics to assign genomic identity to fluorescent markers within live images of chromosomes, where the spots are identically labeled and therefore indistinguishable (Fig. 1 *A–F*). This approach leverages the statistical behavior of a flexible polymer for the analysis, as this has been shown to capture in vivo chromosomal behavior (9–11,24,43). We present the analytical solution for the governing statistics that maps the genomic identity of spots between subsequent images, providing a basis for spot identification. We demonstrate the quantitative reliability of this algorithm by comparison with two-spot experimental data. This work provides a basis for establishing confidence in spot identification as a function of number of spots, temporal and spatial resolution, and inter-genomic spacing of labeled marks, opening the door to fully exploit experimental measurements of chromosomal organization and dynamics in living cells.

MATERIALS AND METHODS

Probabilistic description of polymer configuration

We begin our analysis by defining the Rouse polymer model (14) that describes the physics of a flexible polymer chain subject to Brownian fluctuations in a viscous medium. This polymer model often acts as a basis for describing the dynamics of chromosome loci (9,15,24,44) at a length scale significantly higher than the persistence length of chromatin. We define a

polymer chain as a spatial curve with monomer position n that runs from 0 at one end of the chain to the total chain length N, defined as the number of Kuhn lengths b in the polymer. In this model, the position of the nth segment of the polymer $\vec{r}(n,t)$ evolves according to the Langevin equation of motion:

$$\xi \frac{\partial \vec{r}(n,t)}{\partial t} = \frac{3k_BT}{b^2}\frac{\partial^2 \vec{r}(n,t)}{\partial n^2} + \vec{f}_B(n,t), \quad (1)$$

where ξ is the coefficient of viscous drag. The Brownian force $\vec{f}_B$ arises from thermal fluctuations at temperature T and is governed by the fluctuation dissipation theorem, written as

$$\langle \vec{f}_B(n,t)\vec{f}_B(n',t')\rangle = 2k_BT\xi\delta(n-n')\delta(t-t')\boldsymbol{I}, \quad (2)$$

where $\boldsymbol{I}$ is the identity matrix.

This polymer model is often studied as a linear superposition of orthonormal modes, which are eigenfunctions associated with Eq. (1) and are defined as

$$\varphi_p(n) = \begin{cases} 1 & p = 0 \\ \sqrt{2}\cos\left(\frac{pn\pi}{N}\right) & p > 0. \end{cases} \quad (3)$$

The time-dependent amplitude of the pth mode can be calculated as $\vec{X}_p(t) = N^{-1}\int_0^N \vec{r}(n,t)\varphi_p(n)\mathrm{d}n$. We derive (17,45) the expression for $P_p(\vec{X}_p|\vec{X}_p^{(0)};t)$—the probability density of a normal mode being $\vec{X}_p$ at time t given the amplitude of the same mode is $\vec{X}_p^{(0)}$ at time zero (Supporting material, section 1). For $p \neq 0$, this is given by

$$P_p\left(\vec{X}_p\middle|\vec{X}_p^{(0)};t\right) = \frac{1}{\mathcal{N}_p}\exp\left[-\frac{1}{2}\frac{C_p(0)\left(\vec{X}_p^2+\vec{X}_p^{(0)^2}\right) - 2C_p(t)\vec{X}_p\cdot\vec{X}_p^{(0)}}{C_p^2(0)-C_p^2(t)}\right]. \quad (4)$$

$C_p(t) = \langle \vec{X}_p^{(0)}\cdot\vec{X}_p\rangle = (Nb^2/p^2\pi^2)\exp(-p^2t/\tau_R)$, where $\tau_R = \xi N^2b^2/(3\pi^2k_BT)$ is the characteristic time of relaxation of the longest wavelength mode (i.e., $p = 1$), often referred to as the Rouse time. $\mathcal{N}_p$ are normalization constants corresponding to individual modes. For center-of-mass diffusion (i.e., $p = 0$), the probability P_0 obeys Gaussian diffusion with a center-of-mass diffusivity $D_{COM} = k_BT/(N\xi)$.

We now consider a set of marked sites on the polymer chain. At time zero, there are $M^{(0)}$ marked sites located at monomer positions $n_i^{(0)}$, where $i = 1,2,\ldots,M^{(0)}$ and spatial positions $\vec{r}_i^{(0)} = \vec{r}(n_i^{(0)},0)$. At time t, there are M marked sites located at monomer positions n_i, where $i = 1,2,\ldots,M$ and spatial positions $\vec{r}_i = \vec{r}(n_i,t)$. Frequently, the marked sites remain the same across time frames (i.e., $\{n\} = \{n^{(0)}\}$), but our theory accommodates instances where the mark identities change from one frame to another, such as marks leaving the frame of view, and coupling of asynchronous imaging of two different type of marks. The spatial positions are centered by subtracting the center of mass of the marked sites, such that $\Delta\vec{r}_i = \vec{r}_i - M^{-1}\sum_{j=1}^M \vec{r}_j$. By construction, $\sum_{i=1}^M \Delta\vec{r}_i = 0$, leaving $M-1$ independent vectors $\Delta\vec{r}_i$, for $i = 1,2,\ldots,M-1$, and $\Delta\vec{r}_M = -\sum_{i=1}^{M-1}\Delta\vec{r}_i$ (similar for $\Delta\vec{r}_i^{(0)}$).

From the modal decomposition, we find the probability P_M of a set of monomer segments being at $\{\Delta\vec{r}\}$ at time t given the initial monomer segments were at $\{\Delta\vec{r}^{(0)}\}$ at time zero, resulting in the expression:

$$P_M(\{\Delta\vec{r}\}|\{\Delta\vec{r}^{(0)}\};t) = \int\prod_{p=1}^{\infty}\mathrm{d}\vec{X}_p\mathrm{d}\vec{X}_p^{(0)}P_p\left(\vec{X}_p\middle|\vec{X}_p^{(0)};t\right) \times \prod_{i=1}^{M-1}\delta\left[\Delta\vec{r}_i - \sum_{p=1}^{\infty}\vec{X}_p\Delta\varphi_p(n_i)\right] \times \prod_{j=1}^{M^{(0)}-1}\delta\left[\Delta\vec{r}_j^{(0)} - \sum_{p=1}^{\infty}\vec{X}_p^{(0)}\Delta\varphi_p^{(0)}\left(n_j^{(0)}\right)\right]. \quad (5)$$

We define the centered normal modes $\Delta\varphi_p(n_i) = \varphi_p(n_i) - M^{-1}\sum_{j=1}^M\varphi_p(n_j)$ and $\Delta\varphi_p^{(0)}\left(n_i^{(0)}\right) = \varphi_p\left(n_i^{(0)}\right) - \left[M^{(0)}\right]^{-1}\sum_{j=1}^{M^{(0)}}\varphi_p\left(n_j^{(0)}\right)$. We note that converting to centered coordinates and normal modes eliminates the center-of-mass diffusion (i.e., $p = 0$) from the governing statistical distribution.

Utilizing properties of the Fourier transform and Gaussian integrals (see Supporting material, section 2), we evaluate the integral in Eq. 5 as

$$P_M(\{\Delta\vec{r}\}|\{\Delta\vec{r}^{(0)}\};t) = \frac{1}{\mathcal{N}}\exp\left[-\frac{1}{2}\left(\Delta r_{i,\alpha} - B_{ij}\Delta r_{j,\alpha}^{(0)}\right) \times \left(\boldsymbol{\sigma}^{-1}\right)_{ii'}\left(\Delta r_{i',\alpha} - B_{i'j'}\Delta r_{j',\alpha}^{(0)}\right)\right], \quad (6)$$

where $\mathcal{N}$ is a normalization constant, the repeated indices i,i' run from 1 to $M-1$, and j,j' run from 1 to $M^{(0)}-1$, and the index α sums over the imaging dimensions d (i.e., three-dimensional images with x,y,z data have $d = 3$). We define the matrices $\sigma_{ii'} = \mathcal{C}_{ii'}^{(t)} - \mathcal{C}_{ij}^{(\Delta)}(\mathcal{C}^{(0)-1})_{jj'}\mathcal{C}_{j'i'}^{(\Delta)}$, and $B_{ij} = \mathcal{C}_{ij'}^{(\Delta)}(\mathcal{C}^{(0)-1})_{j'j}$, with correlation matrices whose elements are given by

$$\mathcal{C}_{ii'}^{(t)} = \sum_{p=1}^{\infty}C_p(0)\Delta\varphi_p(n_i)\Delta\varphi_p(n_{i'}),$$
$$\mathcal{C}_{jj'}^{(0)} = \sum_{p=1}^{\infty}C_p(0)\Delta\varphi_p^{(0)}\left(n_j^{(0)}\right)\Delta\varphi_p^{(0)}\left(n_{j'}^{(0)}\right), \quad (7)$$
$$\mathcal{C}_{ij}^{(\Delta)} = \sum_{p=1}^{\infty}C_p(t)\Delta\varphi_p(n_i)\Delta\varphi_p^{(0)}\left(n_j^{(0)}\right).$$

The algorithm of spot assignment

The mark positions can be reordered within the set $\{\Delta\vec{r}\}$ to iterate through the genomic assignments of the M marks, and we define the set associated with the kth assignment as $\{\Delta\vec{r}\}_k$. If there are M spots imaged, there are $M!$ permutations of spots assignments to different genomic positions (i.e., $k = 1,2,\ldots,M!$) that can be considered. We define the probability of assignment k being the true configuration p_k, which is proportional to the local probability density of the corresponding configuration, computed as

$$p_k = \frac{P_M\left(\{\Delta\vec{r}\}_k\middle|\{\Delta\vec{r}^{(0)}\};t\right)}{\sum_{k'=1}^{M!}P_M\left(\{\Delta\vec{r}\}_{k'}\middle|\{\Delta\vec{r}^{(0)}\};t\right)}. \quad (8)$$

We demonstrate the assignment process by using trajectories from numerical simulations of the Rouse model. Given the connectivity of spots

leading to a specific configuration at $t = 0$ (Fig. 1 *C*) and positions of spots at time Δt without specific connectivity (Fig. 1 *D*), we aim to find the most probable connectivity at time Δt. Among all possible configurations, we show four configurations with highest probability (Fig. 1 *E*). The probability of the most probable configuration is significantly higher than all other configurations, and it corresponds to the ground truth from the dynamic simulation (Fig. 1 *F*).

Numerical simulation of Rouse polymer model

To simulate a Rouse polymer, we perform a numerical integration of a discrete form of the Langevin equation, given in Eq. (1). We nondimensionalize this equation by scaling the positions with respect to the root mean-squared end-to-end distance ($\vec{\rho} = \vec{r}/\sqrt{Nb^2}$), time with respect to Rouse time ($\tau = \pi^2 t/\tau_R$), and segment position with respect to total number of Kuhn segments ($\eta = n/N$). This results in the expression:

$$\frac{\partial \vec{\rho}(\eta,\tau)}{\partial \tau} = \frac{\partial^2 \vec{\rho}(\eta,\tau)}{\partial \eta^2} + \vec{\varphi}_B(\eta,\tau), \tag{9}$$

where the dimensionless Brownian force $\vec{\varphi}_B(\eta,\tau)$ has the property $\langle \vec{\varphi}_B(\eta,\tau)\cdot\vec{\varphi}_B(\eta',\tau')\rangle = 2\delta(\eta-\eta')\delta(\tau-\tau')\boldsymbol{I}$. For numerical integration, we discretize the polymer into N segments represented by $(N+1)$ beads. We simulate the time evolution of the coordinates of the ith bead in dimension α as

$$\begin{aligned}\rho_{i,\alpha}(\tau+\delta\tau) &= \rho_{i,\alpha}(\tau) \\ &+ \frac{\rho_{i-1,\alpha}(\tau) - 2\rho_{i,\alpha}(\tau) + \rho_{i+1,\alpha}(\tau)}{\delta\eta^2}\delta\tau \\ &+ \sqrt{\frac{2\delta\tau}{3\delta\eta}}\mathbb{N}(0,1),\end{aligned} \tag{10}$$

where $\delta\eta = 1/N$ is the normalized spacing between the beads, $\delta\tau$ is the normalized timestep for integration, and $\mathbb{N}(0,1)$ is taken from a Gaussian distribution of numbers with 0 mean and unit variance. We generate the initial configurations of the simulation from an equilibrium distribution of the Rouse polymer by setting $\rho_{i+1,\alpha}(0) = \rho_{i,\alpha}(0) + \sqrt{\frac{\delta\eta}{3}}\mathbb{N}(0,1)$. For the simulations presented in this manuscript, we choose $N = 100$ and $\delta\tau = 10^{-5}$.

RESULTS

Dynamics of chromosomal loci is consistent with Rouse model

We utilize the trajectories of two markers of chromatin separated by a distance of ~120 kb attached to a mouse chromosome of length ~160 Mb imaged in two different channels (46) (Fig. 2 *A*) to check if the Rouse polymer model (14) realistically represent the dynamics of chromosome loci. We use the quantity mean-squared change in distance (MSCD), defined as $\langle(\Delta\vec{r}(t) - \Delta\vec{r}(0))^2\rangle$, where $\Delta\vec{r}(t)$ is the displacement between the two spots at time t. This quantity is specifically powerful for tracking experimental dynamics as it only considers the relative displacement between the spots and eliminates any experimental drift or large-scale motion. For a long flexible polymer, MSCD of two spots in the middle of the chain separated by a distance of Δn Kuhn units can be calculated as follows (17):

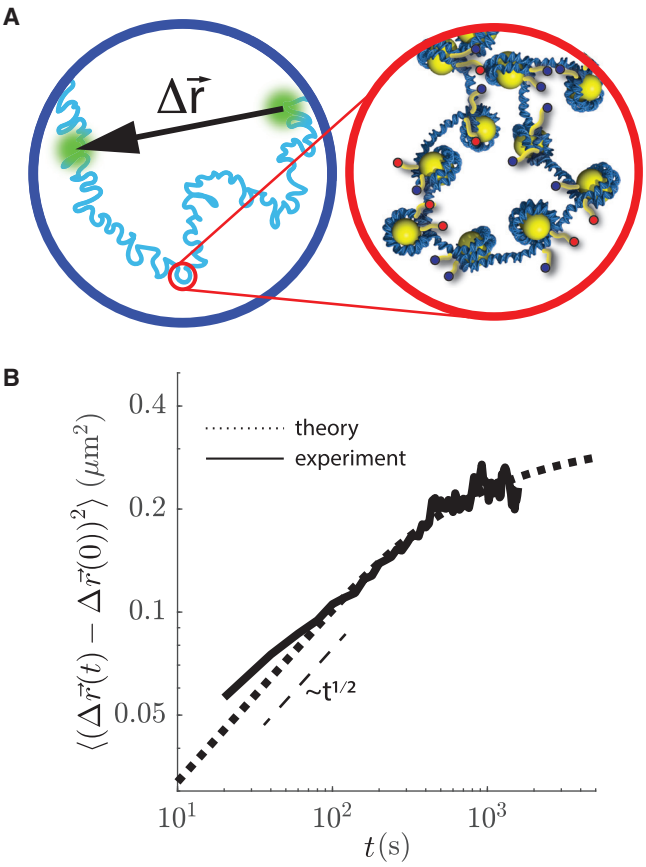

FIGURE 2 Polymer model for chromosome dynamics. (*A*) A schematic of two loci marked on a chromosome (represented as a flexible polymer) and the vector denoting their relative displacement (*left*), and the right image shows a magnified version of the same with elaborated microscopic features such as helical coiling of chromatin around nucleosomes (*right*). (*B*) Mean-squared change in displacement (MSCD) of two such loci as a function of time from live imaging data (46) (*solid line*). The dotted line represents theoretical predictions of the MSCD using the Rouse model. The dashed line represents a slope of 0.5 in the logarithmic scale.

$$\langle(\Delta\vec{r}(t) - \Delta\vec{r}(0))^2\rangle = 16\sum_{p=0}^{\infty}\Delta C_{2p+1}(t)\sin^2\left[\frac{(2p+1)\pi\Delta n}{2N}\right], \tag{11}$$

where $\Delta C_p(t) = (Nb^2/p^2\pi^2)[1 - \exp(-p^2 t/\tau_R)]$, N is the total number of Kuhn segments in the polymer chain, b is the Kuhn length, and τ_R is the Rouse time of the polymer.

The MSCD calculated from live imaging data is consistent with Rouse polymer dynamics with a $t^{0.5}$ scaling in the early time and saturation in long time (17,29,47) (Fig. 2 *B*). Next, to estimate the length and timescale of the chromosomal dynamics, we fit Eq. (11) to the experimental data with the root mean-squared end-to-end distance $b\sqrt{N}$ and the Rouse time τ_R as free parameters (Fig. 2 *B*). The analysis reveals the value of these two parameters to be 14.80 μm and 2.4×10^8 s, respectively.

We note that the genomic distance between the two spots imaged here is significantly higher ($\sim 10^3$ times) than the persistence length of DNA and consists of a large number ($\sim 10^3$) of nucleosomes, the units of microscopic models of chromatin (41,42). Our result suggests that the chromosomal dynamics at large length scale (relevant for live imaging) can be captured by a flexible polymer model (Fig. 2 *A*, left), whereas features such as semiflexible nature of the chain and excluded volume effect play a more important role in the mechanics at smaller length scales (Fig. 2 *A*, right).

Dynamics of memory of polymer configurations

The probability distribution derived in Eq. (6) takes the form of a multivariate Gaussian distribution (materials and methods). The time-dependent matrix $B_{ij}(\{n\},\{n^{(0)}\},t)$ linearly transforms the configuration at time 0 to the mean configuration at time t as $\langle \Delta r_{i,\alpha}\rangle = \sum_{j=1}^{M^{(0)}-1} B_{ij}\Delta r_{j,\alpha}^{(0)}$ for each segment i. The individual elements of the matrix $\sigma_{ij}(\{n\},\{n^{(0)}\},t)$ denote covariance between the positions of the polymer segments i and j. The diagonal elements of the matrix σ_{ii} represent the variance of individual polymer segments at t (see Supporting material, section). The mean and variance of the position of the Mth spot is set by the other $M-1$ spots as $\langle \Delta\vec{r}_M\rangle = -\sum_{i=1}^{M-1}\langle \Delta\vec{r}_i\rangle$, and $\langle(\Delta\vec{r}_M - \langle\Delta\vec{r}_M\rangle)^2\rangle = \sum_{i=1}^{M-1}\sum_{i'=1}^{M-1}\sigma_{ii'}$, respectively.

We show the progression of the mean and variance of a polymer conformation over time in Fig. 3 *A*. At time $t = 0$, we generate a polymer configuration from the equilibrium distribution of a two-dimensional Rouse polymer described by 100 Kuhn segments, shown in Fig. 3 *A*. The mean conformation over time is shown as the solid black lines, and the standard deviation (i.e., square root of the variance) at each monomer position is indicated by the gray circles. The analysis shows that the polymer loses the features of the initial structure with time. With increasing time, only the large-scale features are retained, and eventually, the mean positions of all of the segments coincide with the center of mass. The variance of the positions rises and at $t > \tau_R$ approaches a steady state (Fig. 3 *A* and *B*). In all time, the variances of the bead positions decrease from the free ends to the middle of the chain due to dangling chain ends (Fig. 3 *B*).

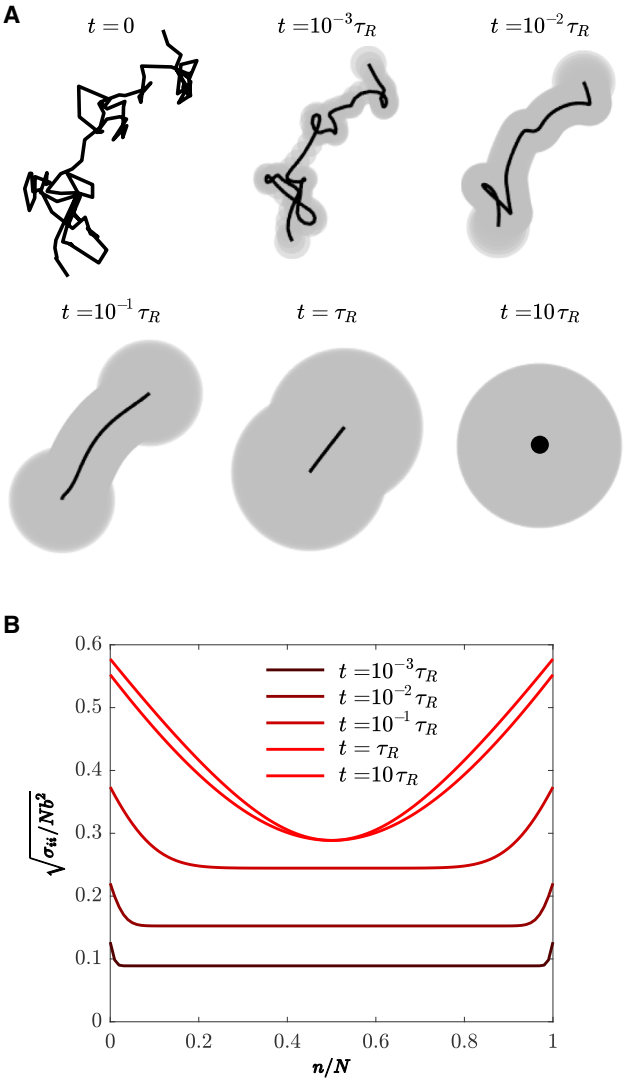

FIGURE 3 Theoretical prediction from our polymer dynamics model. (*A*) Evolution of the mean and standard deviations of the polymer bead positions given a structure at $t = 0$ as a function of time. The dark line represents the mean position of polymer segments. The gray circles represent the standard deviation of each bead position. (*B*) Standard deviations of bead positions as a function of genomic position (n) at different times.

Characterizing the performance of spot assignments

Our theoretical framework provides the necessary input to assign the imaged marks to specific genomic positions (materials and methods; Fig. 1). We evaluate the performance of assignment for different number of spots M, genomic distance between spots Δn, and time lag Δt utilizing trajectories from the numerical simulations. For high genomic distance and short time lag, the standard deviation of the individual spots is significantly smaller than the interspot distances (Fig. 4 *A*). Furthermore, the ground truth has significantly higher probability than all other possible assignments. As the time lag increases (Fig. 4 *B*) or the genomic distance between the spots decreases (Fig. 4 *C*), the standard deviation associated with the individual spot position approaches and eventually surpasses the interspot distances. In those cases, a number of assignments have similar probability (Fig. 4 *B* and *C*), and often the assignment corresponding to the ground truth does not have the maximum probability. For tracking the polymer with high confidence and accuracy, we need to be in a regime where the relative probabilities among the assignments coincide with conditions demonstrated in Fig. 4 *A*.

To quantitatively assess the confidence of the assignment, we define the entropy of assignment (48) as $S = -\sum_k p_k \log p_k$. For assignments with high confidence, a single

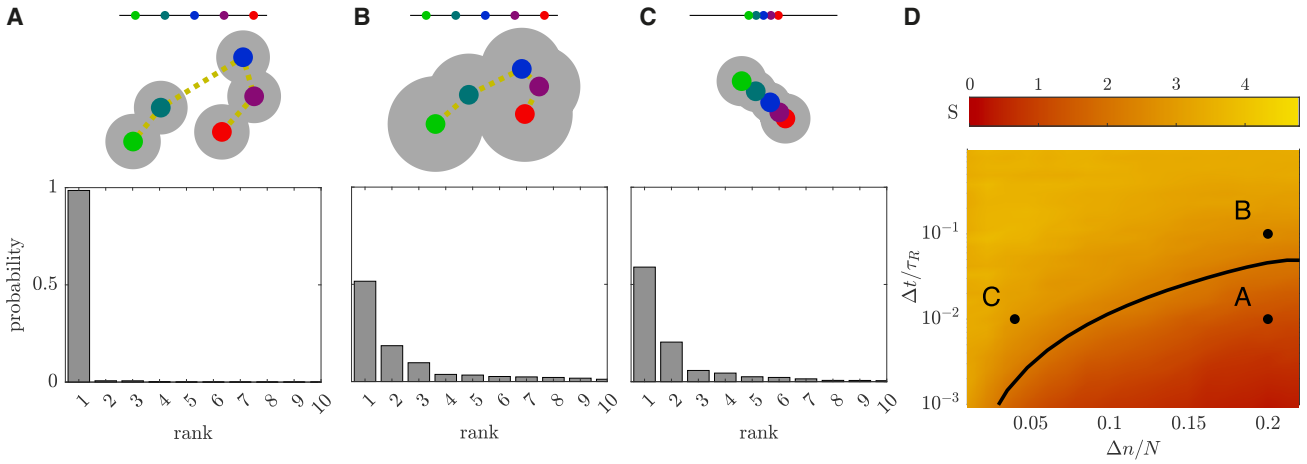

FIGURE 4 Performance of spot assignment as a function of physical parameters. (*A–C*) Genomic positions (*top*), mean position and standard deviation of the marks (*middle*), and relative probabilities of spot assignment (*bottom*) for different genomic spacing Δn and time lag Δt. (*D*) Entropy of the spot assignment as a function of Δn and Δt for five spots. The specific parameters shown in (*A*)–(*C*) are marked with solid circles. The solid line represents the timescale, when the standard deviation of spot position is comparable to the interspot distances.

assignment has probability close to 1, and the entropy S approaches 0. On the contrary, conditions where multiple assignments have significant nonzero probability leads to larger entropy S associated with the uncertainty in spot assignment. We calculate the entropy of the assignment for a range of genomic distance and time lag and average it over 100 different equilibrium simulations (Fig. 4 *D*). The entropy increases as the marks become genomically closer and the time lag between the frame increases. In all the cases, the entropy S remains small for a significant range of parameters. We identify the boundary for the low entropy region in the parameter space as conditions where the positional variance of individual segments approaches the average intergenome distance between spots. Specifically, we calculate the time lag Δt for which the mean standard deviation of the spot is equal to half of the root mean-squared distance between the spots from the equilibrium statistics of Rouse polymer, evaluated as

$$\frac{1}{M-1}\sum_{i=1}^{M-1}\sigma_{ii}(\{n\},\{n\},\Delta t) = \frac{\Delta n b^2}{4}. \qquad (12)$$

The corresponding curve closely represents the boundary between the high and low entropy regions. Fig. 4 *D* shows this boundary for five spots, providing an estimate of the genomic distances that can be resolved for a given time resolution for tracking the polymer. We obtain the same timescale for $M = 4$ and $M = 6$ spots as well. We compare the boundary location for $M = 4$, 5, and 6 in Fig. 5. The timescale is indistinguishable for different numbers of spots for small spot distances. However, it reduces with number of spots for higher spot distances.

Although it is possible to evaluate the relative probabilities of all possible configurations for a small number of spots, the number of possible assignments increases exponentially with the number of spots, and algorithmic steps are necessary to render the approach computationally tractable. For example, an increase in the number of spots from 5 to 10 results in the number of possible assignments increasing from $5! = 120$ to $10! > 3 \times 10^6$. For high genomic distance and low time lag, σ_{ij} is effectively a diagonal matrix, which makes the probability of individual spot position independent of others and allows the algorithm to scale polynomially (instead of exponentially) with the number of spots to find the most probable assignment of spots (49). We show that the spot identification is of similar accuracy in the corresponding parameter sets (Supporting material, section 3; Fig. S1).

Tracking trajectories from live imaging

We demonstrate that our polymer-based spot assignment realistically represents the dynamics of a chromosome (Fig. 2). Next, we identify the time resolution of the microscopy in

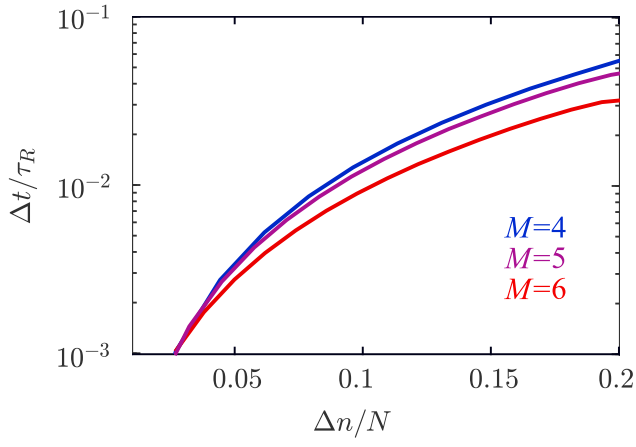

FIGURE 5 The timescale where the standard deviation of the spot position is comparable to the interspot separation as a function of genomic distance between spots for $M = 4$, 5, and 6 spots.

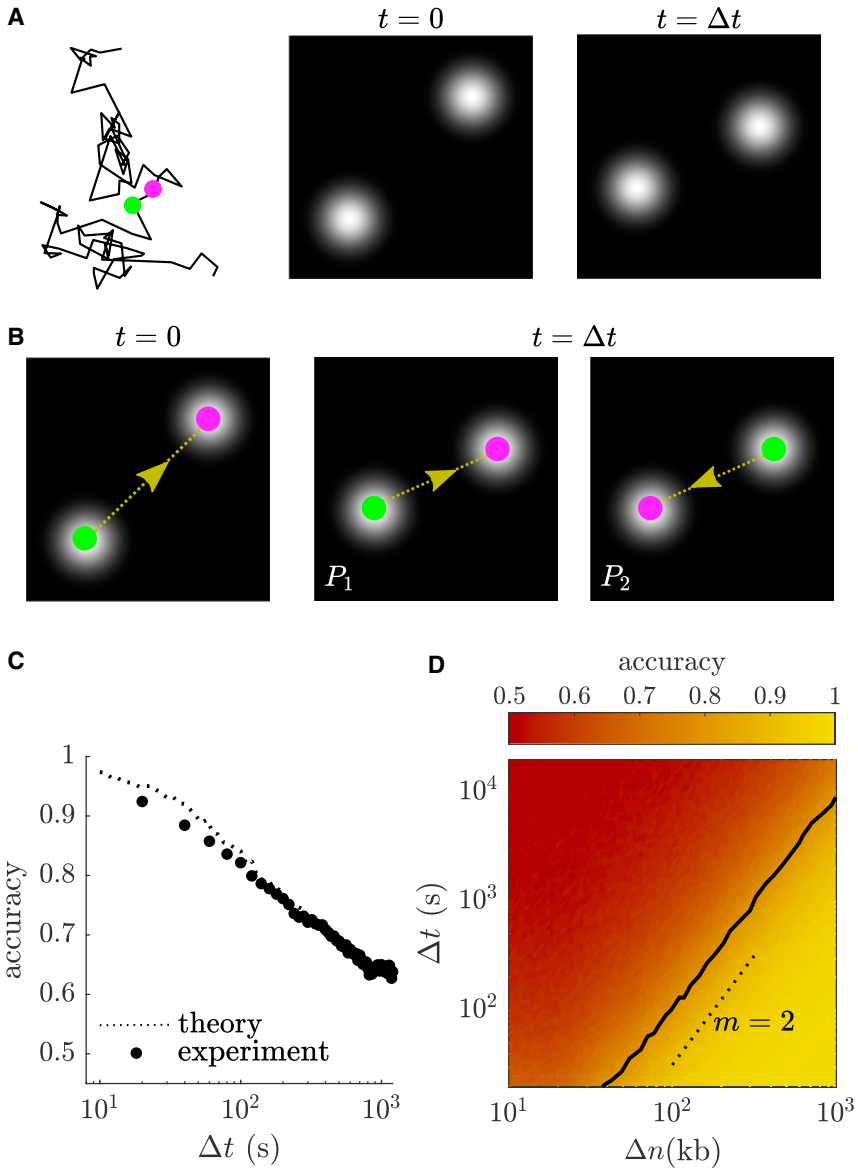

FIGURE 6 Accuracy of tracking in live imaging data. (*A*) A schematic of live imaging of two spots of a polymer (chromosome) with two different markers and the corresponding images at two time points separated by δt. (*B*) A schematic of the spot identification problem: known spot identity at $t = 0$, two spots with unknown identities at $t = \Delta t$, true assignment (denoted by probability P_1), and false assignment (denoted by probability P_2) (*from left to right*). (*C*) For a specific pair of markers, accuracy of spot assignment as a function of time resolution. The overlaid dotted lines represent the same for configurations taken from equilibrium distribution of the Rouse model. (*D*) Accuracy as a function of genomic distance and time resolution for configurations from equilibrium distribution of the Rouse model. The solid line represents a contour map for 80% accuracy, and the dotted line represents a line with slope $m = 2$.

real units for accurate tracking. We utilize the same live imaging data shown in Fig. 2 with two marks separated by 120 kb labeled with two different channels situated on a $\sim$ 160-Mb-long mouse chromosome for this purpose. We show that when the distance between spots is significantly smaller than the chain length (i.e., $\Delta n/N = 7.5 \times 10^{-4} \ll 1$), the expression of probability density becomes only a function of interspot distances and is insensitive to exact spot position on the genome (Supporting material, section 4; Fig. S2). We utilize this approximation and the physical parameters estimated in Fig. 2 *B* to evaluate the performance of our algorithm on the experimental data, and we calculate the relative probabilities of the two possible assignments given the spot positions at an earlier time point (Fig. 6 *A*).

Specifically, we quantify the accuracy by calculating the fraction of cases where the ground truth (i.e., the correct assignment of spots marked by green and purple) has greater probability than the second possible assignment (i.e., $P_1 > P_2$) (Fig. 6 *B*). For time resolution less than 100 s, we achieve accuracy more than 80% (Fig. 6 *C*). For the same basepair distances, we calculate the theoretical accuracy by considering the initial points from an equilibrium distribution of the Rouse model and the final point from the probability distribution from the long-chain approximation of Eq. (6) for the same configuration (Fig. 6 *C*). We found that the accuracy for the experimental trajectories closely matches the theoretical accuracy, suggesting again that our model results are consistent with the dynamics of a chromosome in a living cell. Using this theoretical framework, we determine the variation of the accuracy with both the genomic distance between the spots and the time resolution of the

microscopy. The analysis reveals that to achieve a constant accuracy, the required time resolution scales with the square of genomic distance (Fig. 6 *D*).

DISCUSSION

In this article, we show that the physical behavior of a flexible polymer is consistent with the chromosomal dynamics observed in live imaging data (Fig. 2). Next, we derive a probabilistic description of temporal memory of the configuration of a flexible polymer (Fig. 3). We propose an algorithm for finding the most probable connectivity of fiducial marks on chromosomes based on their relative position at another time point. This algorithm leads to the reconstruction of the dynamics of chromatin architecture from live-cell imaging data (Fig. 1). Using numerical simulations, we show that the algorithm is capable of tracking configurations of polymers with high confidence in time steps for which the displacement due to fluctuation is significantly less than the interspot distances (Figs. 4 and 5). Finally, analyzing the live trajectories of fluorescent spots on chromosomes, we find the appropriate time resolution in real units for mammalian chromosomes for assignment of spots separated by a given genomic distance (Fig. 6).

We demonstrate and test the proposed algorithm for a flexible free polymer subjected to Brownian fluctuations. We would like to point out, however, that the theoretical framework is also designed to take into account several biologically relevant events connected to chromosome dynamics. For example, active forces by molecular motors can be included in this model (17). On the other hand, looping of chromosomes mediated by cohesin (22–24) can be taken into account by adding constraints on the distances between two polymer segments. Summarily this model is adaptable to incorporate existing experimental knowledge into tracking. Future work will examine at what time and length scale it is necessary to incorporate these specific physical factors for accurate tracking.

We test our algorithm with live imaging data of two fluorescent spots and identify the time resolution where we can track the spots accurately. Given the quantitative accuracy in spot assignment of two spots, our algorithm forms the basis for systematic assignment of multiple fluorescent spots. This will enable us to reconstruct the live dynamics of chromatin architecture with more detail, leading us to gain a deeper understanding of chromosomal physics and the associated biological processes that drive chromosomal motion.

SUPPORTING MATERIAL

Supporting material can be found online at https://doi.org/10.1016/j.bpj.2023.08.001.

AUTHOR CONTRIBUTIONS

A.J.S., and A.N.B. formulated the problem. A.J.S., S.D., and A.G. developed the theory. S.D. developed the numerical simulations and the code for tracking. S.D. analyzed the experimental trajectories. All authors contributed in writing the manuscript.

ACKNOWLEDGMENTS

We are grateful to Jude Lee, Liang-Fu Chen, Timothy Downing, and Elizabeth Read for valuable discussions. Funding for this work is provided by the National Science Foundation, Understanding the Rules of Life (Award Number 2022182).

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Kempfer, R., and A. Pombo. 2020. Methods for mapping 3D chromosome architecture. *Nat. Rev. Genet.* 21:207–226.
2. Fraser, J., I. Williamson, …, J. Dostie. 2015. An overview of genome organization and how we got there: from FISH to Hi-C. *Microbiol. Mol. Biol. Rev.* 79:347–372.
3. Le, T. B. K., M. V. Imakaev, …, M. T. Laub. 2013. High-resolution mapping of the spatial organization of a bacterial chromosome. *Science.* 342:731–734.
4. Habermann, F. A., M. Cremer, …, I. Solovei. 2001. Arrangements of macro-and microchromosomes in chicken cells. *Chromosome Res.* 9:569–584.
5. Boettiger, A., and S. Murphy. 2020. Advances in chromatin imaging at kilobase-scale resolution. *Trends Genet.* 36:273–287.
6. Mateo, L. J., N. Sinnott-Armstrong, and A. N. Boettiger. 2021. Tracing DNA paths and RNA profiles in cultured cells and tissues with ORCA. *Nat. Protoc.* 16:1647–1713.
7. Chen, B., L. A. Gilbert, …, B. Huang. 2013. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell.* 155:1479–1491.
8. Rodriguez, J., and D. R. Larson. 2020. Transcription in living cells: molecular mechanisms of bursting. *Annu. Rev. Biochem.* 89:189–212.
9. Weber, S. C., A. J. Spakowitz, and J. A. Theriot. 2010. Bacterial chromosomal loci move subdiffusively through a viscoelastic cytoplasm. *Phys. Rev. Lett.* 104, 238102.
10. Weber, S. C., A. J. Spakowitz, and J. A. Theriot. 2012. Nonthermal ATP-dependent fluctuations contribute to the in vivo motion of chromosomal loci. *Proc. Natl. Acad. Sci. USA.* 109:7338–7343.
11. Weber, S. C., M. A. Thompson, …, J. A. Theriot. 2012. Analytical tools to distinguish the effects of localization error, confinement, and medium elasticity on the velocity autocorrelation function. *Biophys. J.* 102:2443–2450.
12. Polovnikov, K. E., M. Gherardi, …, M. V. Tamm. 2018. Fractal folding and medium viscoelasticity contribute jointly to chromosome dynamics. *Phys. Rev. Lett.* 120, 088101.
13. Di Pierro, M., D. A. Potoyan, …, J. N. Onuchic. 2018. Anomalous diffusion, spatial coherence, and viscoelasticity from the energy landscape of human chromosomes. *Proc. Natl. Acad. Sci. USA.* 115:7753–7758.
14. Doi, M., and S. F. Edwards. 1988. The Theory of Polymer Dynamics, 73. oxford university press.

15. Weber, S. C., J. A. Theriot, and A. J. Spakowitz. 2010. Subdiffusive motion of a polymer composed of subdiffusive monomers. *Phys. Rev.* 82, 011913.

16. Lampo, T. J., N. J. Kuwada, …, A. J. Spakowitz. 2015. Physical modeling of chromosome segregation in Escherichia coli reveals impact of force and DNA relaxation. *Biophys. J.* 108:146–153.

17. Ghosh, A., and A. J. Spakowitz. 2022. Active and thermal fluctuations in multi-scale polymer structure and dynamics. *Soft Matter.* 18:6629–6637.

18. Mahajan, A., W. Yan, …, M. J. Shelley. 2022. Euchromatin Activity Enhances Segregation and Compaction of Heterochromatin in the Cell Nucleus. *Phys. Rev. X.* 12, 041033. https://doi.org/10.1103/PhysRevX.12.041033.

19. Saintillan, D., M. J. Shelley, and A. Zidovska. 2018. Extensile motor activity drives coherent motions in a model of interphase chromatin. *Proc. Natl. Acad. Sci. USA.* 115:11442–11447.

20. Goychuk, A., D. Kannan, …, M. Kardar. 2023. Polymer folding through active processes recreates features of genome organization. *Proc. Natl. Acad. Sci. USA.* 120, e2221726120.

21. Shin, S., H. Woo Cho, …, D. Thirumalai. 2023. Transcription-induced active forces suppress chromatin motion by inducing a transient disorder-to-order transition. *Biophys. J.* 122:19a.

22. Goloborodko, A., M. V. Imakaev, …, L. Mirny. 2016. Compaction and Segregation of Sister Chromatids via Active Loop Extrusion. *Elife.* 5:148644–e14916.

23. Nuebler, J., G. Fudenberg, …, L. A. Mirny. 2018. Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proc. Natl. Acad. Sci. USA.* 115:E6697–E6706.

24. Gabriele, M., H. B. Brandão, …, A. S. Hansen. 2022. Dynamics of CTCF-and cohesin-mediated chromatin looping revealed by live-cell imaging. *Science.* 376:496–501.

25. Lichter, P., C. J. Tang, …, D. C. Ward. 1990. High-resolution mapping of human chromosome 11 by in situ hybridization with cosmid clones. *Science.* 247:64–69.

26. Zidovska, A., D. A. Weitz, and T. J. Mitchison. 2013. Micron-scale coherence in interphase chromatin dynamics. *Proc. Natl. Acad. Sci. USA.* 110:15555–15560.

27. Ghosh, R. P., J. M. Franklin, …, J. T. Liphardt. 2019. A fluorogenic array for temporally unlimited single-molecule tracking. *Nat. Chem. Biol.* 15:401–409.

28. Lampo, T. J., A. S. Kennard, and A. J. Spakowitz. 2016. Physical modeling of dynamic coupling between chromosomal loci. *Biophys. J.* 110:338–347.

29. Newman, T. A. C., B. Beltran, …, S. M. Burgess. 2022. Diffusion and distal linkages govern interchromosomal dynamics during meiotic prophase. *Proc. Natl. Acad. Sci. USA.* 119, e2115883119.

30. Lieberman-Aiden, E., N. L. Van Berkum, …, J. Dekker. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science.* 326:289–293.

31. Rosa, A., and R. Everaers. 2008. Structure and Dynamics of Interphase Chromosomes. *PLoS Comput. Biol.* 4:e1000153.

32. Strickfaden, H., T. Cremer, and K. Rippe. 2012. Higher Order Chromatin Organization and Dynamics. *In* Genome Organization and Function in the Cell Nucleus Wiley Online Library, pp. 417–447.

33. Benedetti, F., J. Dorier, …, A. Stasiak. 2014. Models That Include Supercoiling of Topological Domains Reproduce Several Known Features of Interphase Chromosomes. *Nucleic Acids Res.* 42:2848–2855.

34. Paulsen, J., M. Sekelja, …, P. Collas. 2017. Chrom3D: Three-Dimensional Genome Modeling from Hi-C and Nuclear Lamin-Genome Contacts. *Genome Biol.* 18:21.

35. Chiariello, A. M., C. Annunziatella, …, M. Nicodemi. 2016. Polymer Physics of Chromosome Large-Scale 3D Organisation. *Sci. Rep.* 6:29775–29778.

36. Michieletto, D., E. Orlandini, and D. Marenduzzo. 2016. Polymer Model with Epigenetic Recoloring Reveals a Pathway for the de Novo Establishment and 3D Organization of Chromatin Domains. *Phys. Rev. X.* 6, 041047.

37. Jost, D., P. Carrivain, …, C. Vaillant. 2014. Modeling Epigenome Folding: Formation and Dynamics of Topologically Associated Chromatin Domains. *Nucleic Acids Res.* 42:9553–9561.

38. Di Pierro, M., B. Zhang, …, J. N. Onuchic. 2016. Transferable Model for Chromosome Architecture. *Proc. Natl. Acad. Sci. USA.* 113:12168–12173.

39. Di Pierro, M., R. R. Cheng, …, T. Schlick. 2017. De Novo Prediction of Human Chromosome Structures: Epigenetic Marking Patterns Encode Genome Architecture. *Proc. Natl. Acad. Sci. USA.* 114:12126–12131.

40. MacPherson, Q., B. Beltran, and A. J. Spakowitz. 2018. Bottom–up modeling of chromatin segregation due to epigenetic modifications. *Proc. Natl. Acad. Sci. USA.* 115:12739–12744.

41. Sandholtz, S. H., Q. MacPherson, and A. J. Spakowitz. 2020. Physical modeling of the heritability and maintenance of epigenetic modifications. *Proc. Natl. Acad. Sci. USA.* 117:20423–20429.

42. MacPherson, Q., B. Beltran, and A. J. Spakowitz. 2020. Chromatin compaction leads to a preference for peripheral heterochromatin. *Biophys. J.* 118:1479–1488.

43. Shi, G., L. Liu, …, D. Thirumalai. 2018. Interphase human chromosome exhibits out of equilibrium glassy dynamics. *Nat. Commun.* 9:3161.

44. Liu, L., G. Shi, …, C. Hyeon. 2018. Chain organization of human interphase chromosome determines the spatiotemporal dynamics of chromatin loci. *PLoS Comput. Biol.* 14, e1006617.

45. Ghosh, A., and A. J. Spakowitz. 2022. Statistical behavior of nonequilibrium and living biological systems subjected to active and thermal fluctuations. *Phys. Rev. E.* 105, 014415.

46. Alexander, J. M., J. Guan, …, O. D. Weiner. 2019. Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. *Elife.* 8, e41769.

47. Miné-Hattab, J., and R. Rothstein. 2012. Increased chromosome mobility facilitates homology search during recombination. *Nat. Cell Biol.* 14:510–517.

48. Shannon, C. E. 1948. A mathematical theory of communication. *The Bell system technical journal.* 27:379–423.

49. Munkres, J. 1957. Algorithms for the assignment and transportation problems. *J. Soc. Ind. Appl. Math.* 5:32–38.