



A data-driven chromatin model reveals spatial and dynamic features of genome organization

Antonio B. Oliveira Junior^{a,1} , Matheus F. Mello^a, Ronaldo J. Oliveira^{a,b} , Esteban Doder-Rojas^a , Sumitabha Brahmachari^a , Vinicius G. Contessoto^{a,c} , and José N. Onuchic^{a,c,d,e,1}

Affiliations are included on p. 8.

Contributed by José N. Onuchic; received October 24, 2025; accepted December 20, 2025; reviewed by Jin Wang and Bin Zhang

Compacting chromatin within the cellular nucleus presents a significant challenge for biology. Chromosomes must be both condensed and spatially organized to enable essential processes such as transcription and replication. Chromosome conformation capture experiments (e.g., Hi-C) provide valuable information about the spatial organization and, therefore, the connectivity between different genomic regions. These experiments inspired polymer models that describe the physical mechanism of the chromosomal energy landscape. The Full-Inversion Chromatin model (FI-Chrom), a data-driven approach for modeling genome organization, uses Hi-C contact maps to infer pairwise interaction potentials between all chromosomal loci. It combines Graphics Processing Unit (GPU)-accelerated simulations with efficient training of tens of millions of parameters derived from the maximum-entropy principle to determine 3D structures of chromosomes that accurately reproduce Hi-C-like data. FI-Chrom does not make any a priori assumptions regarding chromosome architecture, making it applicable to any chromosome conformation capture experiment. Its derived structural ensembles capture all essential features from the short- and long-range interactions of typical chromosome organization, such as segregated compartments, chromosome territories, and fully or partially formed loops. Although Hi-C contains only structural information, FI-Chrom extends these data by revealing an emergent dynamical mechanism encoded in the inferred energy landscape. For example, simulations show that chromatin loops are not static architectural features but rather transient structural elements. Statistical analyses further indicate that loops confined within a single compartment occur more frequently than those spanning multiple compartments, highlighting the dynamic and compartment-dependent nature of chromatin organization.

genome architecture | chromatin dynamics | polymer physics

The three-dimensional chromatin organization of eukaryotes within the nucleus is required for regulating genome functions such as gene expression (1–3). Experimental advancements, like chromosome conformation capture (Hi-C), provide valuable information about the spatial contacts between different genomic regions (4–8). Hi-C maps provide a two-dimensional contact map, creating a need for theoretical models to extract the genome's three-dimensional organization and to understand the associated functional mechanism.

Computational models for chromosome architecture can be broadly categorized as bottom-up or top-down approaches (9, 10). Bottom-up models attempt to predict chromosome structures from first principles, using detailed information about the physical and chemical interactions between DNA, histones, and other associated molecules (11–18). Top-down models, on the other hand, take advantage of experimental data such as Hi-C ligation maps to infer effective energy functions and their corresponding conformational ensembles (19). In both cases, however, it is important to highlight that chromosomes are fluid-like and therefore must be represented as a structural ensemble rather than a well-defined structure (20–22).

Data-driven models often integrate polymer physics and experimental constraints with the maximum entropy principle (MaxEnt) (23) to ensure minimal bias in the structures ensemble (24). Early work demonstrated that chromosome organization could be understood using energy landscapes learned from Hi-C contact maps. (25, 26). Other studies incorporated epigenetic features to predict chromosome organization across cell types (27–30) and explored large-scale genome architecture through a balance of compaction, phase separation, and lamina adhesion (31, 32). Recent frameworks have

Significance

Chromatin must fold into organized three-dimensional structures within the cell nucleus to enable gene regulation and replication. Chromosome conformation capture (Hi-C) experiments have provided insight into understanding this spatial organization. However, translating contact maps into polymer models remains a challenge. Here, we present the Full-Inversion Chromatin (FI-Chrom) model, a data-driven framework that infers effective interactions from Hi-C data using maximum-entropy principles. FI-Chrom uses no a priori constraints. Instead, it extracts interaction landscapes directly from Hi-C experiments. FI-Chrom reproduces key features of genome folding, such as A/B compartmentalization to transient loop extrusion. By bridging data-driven inference with polymer physics, FI-Chrom provides a versatile computational framework for exploring chromosomal organization across cell types, organisms, and experimental conditions.

Reviewers: J.W., Stony Brook University; and B.Z., Massachusetts Institute of Technology.

The authors declare no competing interest.

Copyright © 2026 the Author(s). Published by PNAS. This article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

¹To whom correspondence may be addressed. Email: antonio.oliveira@rice.edu or jonuchic@rice.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2530583123/-DCSupplemental>.

Published January 23, 2026.

further advanced this field by combining multiomics and imaging datasets, highlighting their functional connections to gene regulation and cellular phenotypes (33).

Here, we introduce the Full-Inversion Chromatin model (FI-Chrom), a data-driven MaxEnt approach to modeling chromosomes. FI-Chrom builds upon the strengths of the OpenMiChroM toolkit (34). Unlike phenomenological models, such as MiChroM, which rely on predefined and interpretable energy terms, FI-Chrom adopts a full-inversion strategy in which all pairwise interactions are directly inferred from the experimental contact map. It incorporates all the contact frequencies from Hi-C maps as bias in the energy function, and the pairwise interactions are optimized using the Adam algorithm (35). This data-driven inference approach enables FI-Chrom to capture local structural variability and transient features, including short-lived loop configurations, that are difficult to encode explicitly in coarse-grained or phenomenological energy landscapes. This approach generates an effective force field for nonbonded interactions, allowing the polymer model to converge toward 3D structures consistent with Hi-C probabilities.

The FI-Chrom contact maps generated from ensembles of 3D structures quantitatively reproduce experimental Hi-C maps for different chromosome sizes. These maps capture important genomic organization features, such as euchromatin-heterochromatin (A/B type) phase separation and the contact polymer scaling factor $P(s)$. Importantly, despite the FI-Chrom method training each pairwise interaction independently without explicitly encoding A/B compartment annotations, it accurately recapitulates the differences between these compartments. This differential interaction strength suggests a natural tendency for B-type regions to self-organize into compact, spatially segregated clusters, consistent with of phase separation mechanism.

We also demonstrate that loop regions within topologically associating domains (TADs) (36) are transiently formed, as revealed by the inspection of 3D structures. Our analyses indicate that loop formation is characterized by the progressive extension of loops until they are stabilized by architectural proteins such as CTCF (37). The formation of loops is dynamically regulated, with configurations fluctuating between fully formed loops and partially formed or open structures. These observations are obtained directly from the simulations of our model, without the inclusion of any prior information about chromosome dynamics. Notably, FI-Chrom accurately reproduces contact frequencies between loop boundaries and captures the characteristic “stripe” patterns likely associated with actively extruded loops (38). The ability to capture these transient and locally variable structures further illustrates the advantage of the full-inversion framework over models that rely on fixed, phenomenological energy terms. FI-Chrom thus represents a powerful and versatile computational framework that can be readily applied to a wide range of experimental datasets.

Overview of the FI-Chrom Workflow. FI-Chrom infers an effective pairwise interaction field directly from Hi-C contact probabilities and uses it to generate 3D structural ensembles that are consistent with experimental data. At its core, the model combines a homopolymer backbone with a data-driven nonbonded term derived using the MaxEnt approach, which biases configurations toward the experimental ligation landscape. The total potential is

$$U(\vec{r}) = U_{HP}(\vec{r}) + \sum_{ij} \lambda_{ij} f(r_{ij}),$$

where U_{HP} enforces chain connectivity and excluded volume, λ_{ij} are effective pairwise parameters learned from Hi-C, and $f(r_{ij})$ is a smooth contact switch that approximates the probability of loci i and j being in contact (See *Materials and Methods* for detailed definition and parametrization of each term).

Fig. 1 summarizes the training procedure. We initialize the training with a symmetric force-field $\{\lambda_{ij}\}$ with small attraction in all nonannotated regions, such as centromeres/telomeres. Next, we perform GPU-accelerated Langevin simulations from diverse starting conformations to sample in silico contact frequencies. To update $\{\lambda_{ij}\}$, we use the Adam algorithm (35) with the gradient with respect to the difference between simulated and experimental contact probabilities. We iterate simulation $\rightarrow$ update $\rightarrow$ simulation until convergence of the map-level error metric, while monitoring for overfitting signatures (e.g., emergence of anomalously large positive entries in the effective field).

Practically, each training iteration aggregates many short trajectories across replicates to ensure robust statistics of ϕ_{ij}^{sim} over a wide range of genomic distances. Because FI-Chrom does not impose a priori compartment labels, the learned $\{\lambda_{ij}\}$ captures emergent mesoscale organization (A/B segregation, subcompartments) as well as local features (TADs, loop-associated stripes) directly from the data. The result is an effective landscape that quantitatively reproduces Hi-C maps across chromosomes while yielding physically interpretable 3D ensembles. Full mathematical definitions, optimizer hyperparameters, convergence criteria, and simulation details are provided in *Materials and Methods*.

The Full-Inversion Chromatin Model Recovers Highly Detailed Contact Maps. To evaluate the effectiveness and reliability of the FI-Chrom method, we used the Hi-C experiments from the GM12878 human cell line (5). This choice was guided by the high quality of these maps and the range of experimental data that they encompass (39). We select three chromosomes of varying lengths for this study: chromosome 2 (243 Mb), chromosome 10 (135 Mb), and chromosome 19 (59 Mb). The resolution of the polymer model is 50 kb of DNA per bead. This choice aligns with previous studies (26, 34, 40). However, it is important to highlight that the resolution in FI-Chrom modeling is inherently

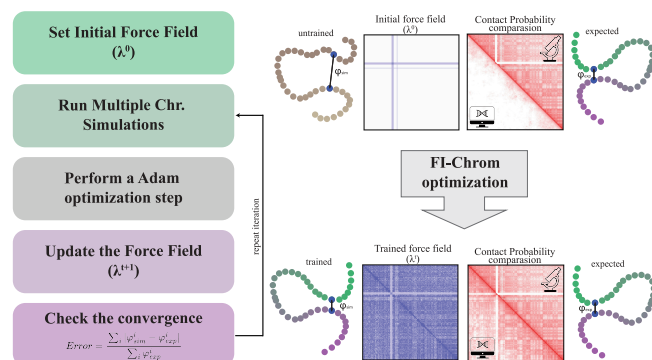


Fig. 1. Schematic representation of the iterative FI-Chrom workflow. Experimental Hi-C contact probabilities are used to infer an effective pairwise interaction through iterative optimization. The resulting force field is then used in polymer simulations to generate structural ensembles, whose simulated contact probabilities are compared with experiments until convergence is achieved.

contingent on the quality of the Hi-C data used as input for its training.

In Fig. 2, we present the model for chromosome 19. The corresponding analyses for chromosomes 2 and 10 are shown in *SI Appendix*, Figs. S1 and S2. Fig. 2*A* compares the Hi-C map (*Upper triangle*) and the simulated contact map (*Lower triangle*). The optimized energy function for chromosome 19 generates an ensemble of structures that accurately reproduces the checkerboard-like pattern. This pattern correlates with the euchromatin-heterochromatin phase separation observed in the experimental Hi-C maps. We perform quantitative comparisons between these contact maps by showing the correlation matrices comparison (Fig. 2*D*); the Pearson's correlation coefficient between them is $R = 0.98$. From these correlation matrices, we can extract the A/B compartmentalization (Fig. 2*H*) by performing a Principal Component Analysis (PCA) and extracting its first principal component (PC1) as performed in Aiden et al. (4). Fig. 2*E* shows the agreement of the contact probability as a function of the genomic distance between simulation and experiment, even for large genomic distances where modeling usually diverges from experiments (26). This can be further verified through Pearson's Correlation by genomic distance (Fig. 2*F*) between the Hi-C and the FI-Chrom model.

Fig. 2*C* shows the matrix representation of the pairwise parameters in the force field when the training pipeline converges. Each element in this matrix represents the interaction intensity (λ_{ij}), where i and j denote the indices of the beads in the chromosome

(see *Materials and Methods* for the full definition of the potential). Notably, the optimization convergence (Fig. 2*I*) is reached after 60 iterations for chromosome 19. For larger chromosomes, more iterations are typically required. For instance, chromosome 2 required approximately 100 iterations to achieve convergence (*SI Appendix*, Fig. S1). Importantly, inspecting the distribution of interaction parameters can also reveal overfitting. When the learned force field begins to exhibit anomalously strong or widespread interactions that can be understood as nonphysical. This may indicate that the model is adapting too closely to the noisy features of the Hi-C data, suggesting overfitting.

To understand the global features of the force field parameters, we use the previously reported subcompartment annotations (5) and average the interactions among the annotations for the trained force field. In Fig. 2*G*, we show the average interaction $\langle \lambda \rangle$ for all pairs of subcompartments. The result is consistent with previous models, in which the chromosome interaction is governed by a copolymer potential (26). Additionally, the relative strength of interaction between different subcompartment annotations aligns with an alternative computational approach (41). It shows a higher interaction between compartment B (and its subcompartments B1, B2, B3, and B4) compared to compartment A (and its subcompartments A1 and A2). This stronger attraction among B-type regions suggests a tendency of these loci to form more compact and spatially segregated clusters, indicative of phase separation, a phenomenon that becomes visually apparent in the ensemble of 3D structures discussed in the next section.

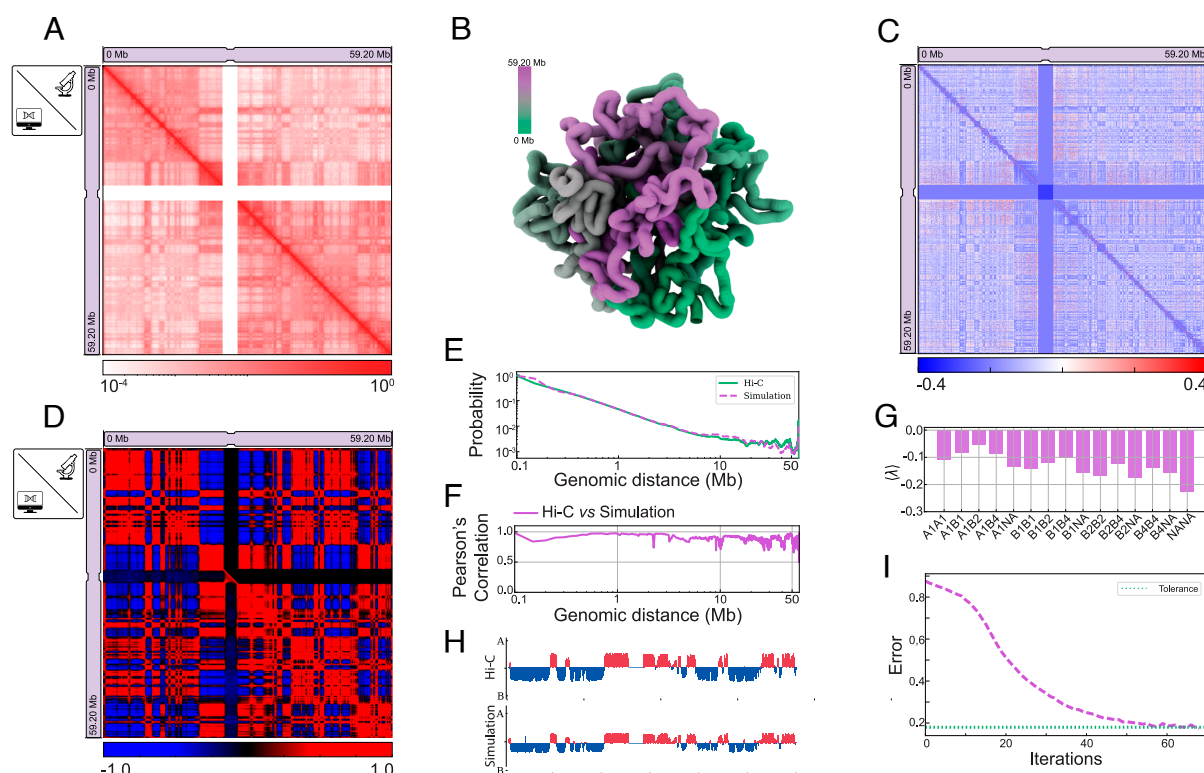


Fig. 2. Full-Inversion Chromatin model for chromosome 19 in the cell line GM12878. (A) Comparison between experimental (*Upper triangle*) and simulated (*Lower triangle*) contact maps. The centromere probability regions were removed due to the nonexistence of data. (B) An example of a 3D structure colored by the genomic index. (C) The force-field matrix obtained after training. These values represent the pairwise parameters (λ_{ij}). (D) A comparison between the experimental (*Top Right*) and simulated (*Bottom Left*) correlation matrix was obtained for the contact maps. The Pearson's R is 0.989. (E) Contact probability scaling comparison. (F) Pearson's correlation by genomic distance. (G) The mean λ for types. The subcompartment annotation labels are obtained in Rao et al. (5). (H) The first eigenvector of the PCA obtained from the correlation matrix (B). (I) The convergence of the training pipeline as a function of the number of iterations. The green (pointed) line represents the threshold for stopping the training pipeline.

Structural Characterization of the 3D Ensemble of Structures.

To generate the *in silico* Hi-C, as shown in Fig. 2, we performed multiple chromosome simulations, in which every replicate starts from a random initial configuration. This helps to ensure the phase space is properly explored (see *Materials and Methods* for details). In this process, 200,000 structures were generated for each chromosome using their optimized force field. Since the *in silico* Hi-C presents high-fidelity contact probabilities, this section analyzes how the 3D structures are affected by the addition of optimized nonbonded interactions.

In Fig. 3 *A* and *B*, we show the radial distribution profile (RDP) for compartments and subcompartments, respectively. The behavior of the structural ensembles presents features inherent to chromosome architecture, such as the phase separation between compartments (and subcompartments). Inside the chromosome territory, the compartments show relative spatial preferences, where the heterochromatin (here represented by B compartments) is frequently found in the interior of the chromosome territory, consistent with microscopy data (42), and has more compact clusters (this detail is shown in the structures inside the RPD plots, separated by compartments). On the other hand, the A compartments show smaller clusters and are more likely to be found in the periphery of the chromosome, as it has been observed in previous computational models of single chromosomes (25, 26, 43).

To characterize the topological states of the chromosome configurations, we applied the minimal rope length (44–46) to find topological invariants over the ensemble of structures. Fig. 3*C* shows the probability distributions of this knot invariant for the trained chromosome 10 (in purple) and the homopolymer (in green). The homopolymer model is highly knotted, as expected from entropic considerations for such long, locally flexible chains (46, 47). In contrast, the optimized chromosome ensembles have topological distributions peaked near 0, meaning

the 3D configurations are mostly unknotted at the megabase scale. In such a scenario, forming compartments [or topologically associating domains (TADs)] helps to keep the chromosomes untangled.

Additionally, to understand how the additional energy (from the optimized model) affects the topology, we started multiple simulations using as initial configurations structures from the equilibrium ensemble of the homopolymer model. These initial structures contain a large number of knots, with an average minimal rope length of 140. Fig. 3*D* presents the average minimal rope length relaxation of the trained chromosome model. This result shows a reduction of the minimal rope length over time, showing that the trained potential is sufficient to untangle the chromosome structural ensemble. Untangling is possible due to the soft-core nature of the self-avoidance potential in the homopolymer model, which allows chain crossing and mimics the activity of topoisomerases, enzymes that transiently cut and rejoin DNA strands to resolve topological constraints (48). It also demonstrated that the most likely structural ensemble for chromosomes has a minimal amount of knotting.

To track how configurations change over the training steps, we used a collective variable Q (see *SI Appendix* for the mathematical definition) to assess the similarity between structures at each training step. Q ranges from 0 (completely different) to 1 (completely similar). Fig. 3*E* shows the distribution profile for each 10 step of FI-Chrom training. Initially consistent with homopolymer simulations, low values of Q indicate uncorrelated structures. However, as training progresses, Q converges to 0.30, indicating the emergence of correlations within the structure ensemble. While $Q = 0.30$ is low compared to highly structured systems, such as proteins, it suggests partial organization typical of heteropolymers, where interactions vary with sequence (25, 40). Notably, chromosomes exhibit high structural similarity within short genomic regions of approximately 1 Mb, where the

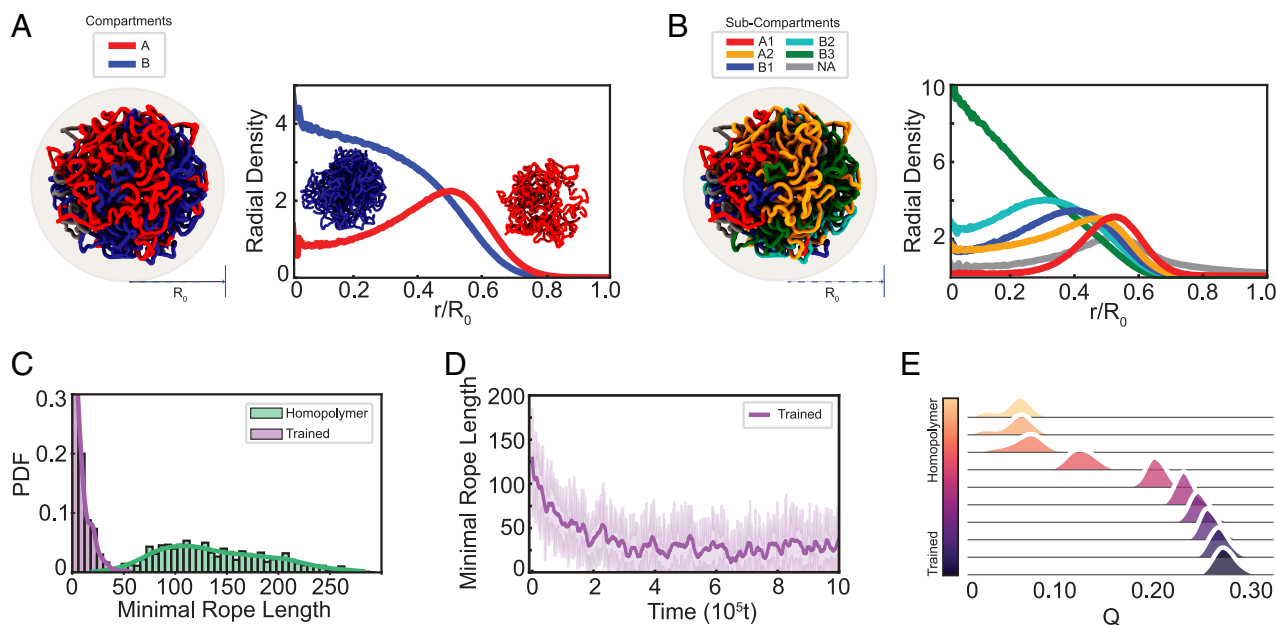


Fig. 3. Structural characterization of chromosome 10 simulations. (A) (Right) Radial density profiles (RDP) for chromosome compartment types A (red) and B (blue) were calculated over the ensemble of structures. $\rho(r)/N_{type}$ is the normalized radial density per type; r/R_0 is the normalized radius. A representative structure is shown colored by compartment types (Left). (B) (Right) RDP for chromosome subcompartment annotations by Rao et al. (5). (Left) representative structure colored by the subcompartments. (C) Probability distributions of the knot invariant measured as the minimal rope length for the homopolymer conformations (green) and the trained chromosome ensemble (purple). (D) Relaxation of trained chromosome topology measured with the average knot invariant as a function of time. (E) Distribution of the structural similarity over training steps. Each line shows the distribution of Q over an ensemble of structures for each training step. Q is a collective variable that measures similarities between pairs of structures (see *SI Appendix* for mathematical definition).

formation of topologically associating domains occurs. A detailed analysis of these regions is presented in the following section.

Structure and Dynamics Toward Function. A significant improvement in the FI-Chrom modeling is the high accuracy of the contact frequencies when compared to the experimental Hi-C data. Fig. 4A displays four comparisons between theoretical

and experimental contact maps, each with varying magnification levels, focusing on specific regions of interest. These zoomed-in views offer a detailed examination of the near-diagonal features, even at a few megabase resolutions, such as TADs and loops.

The compartmentalization patterns are highly correlated with the sequence annotations for long-range genomic distances

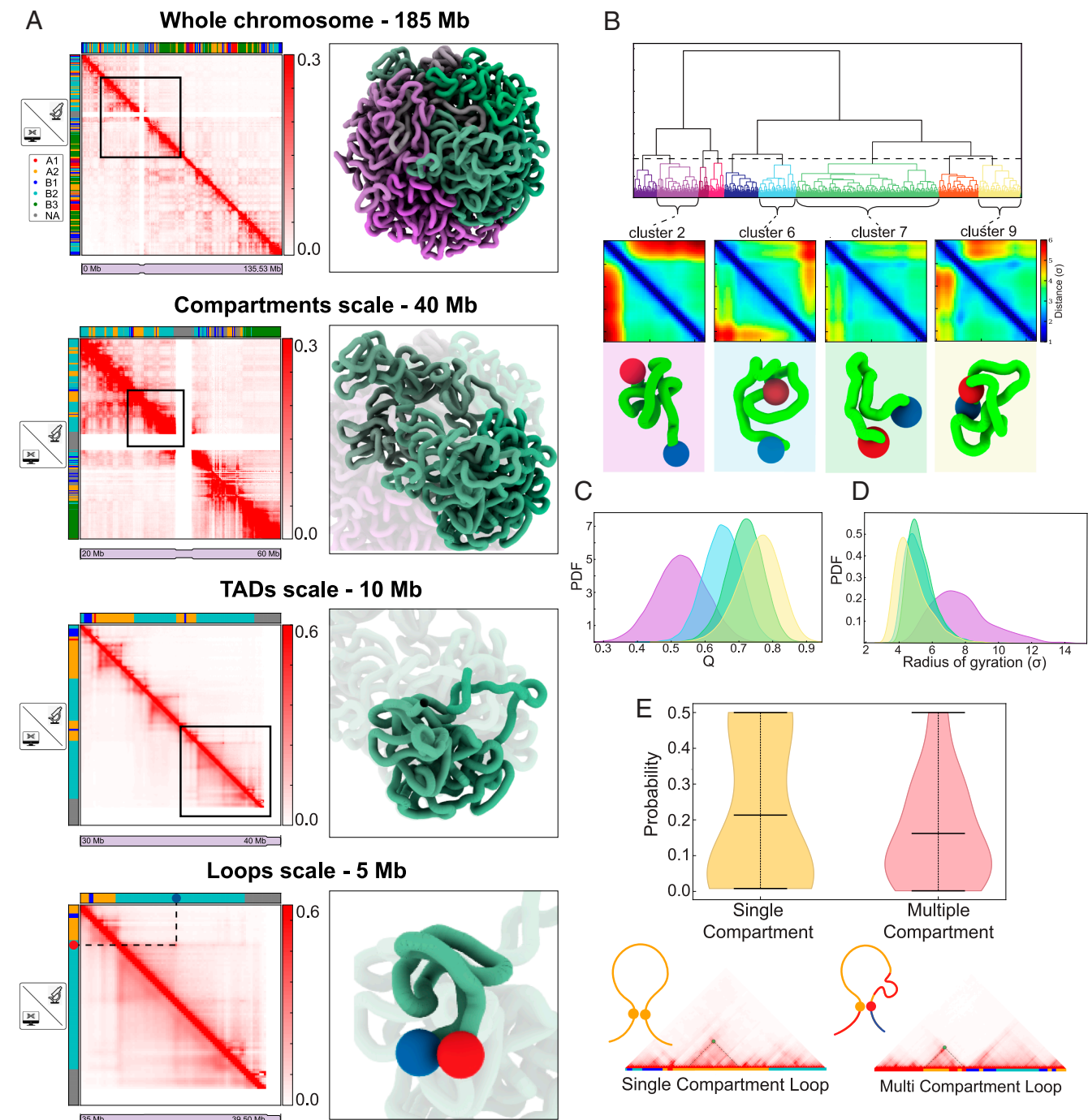


Fig. 4. Structural Characterization of Loop Formation in Human Chromosome 10. (A) Comparison between experimental (Top Right) and simulated (Bottom Left) contact maps for four different zoom levels, starting from the whole chromosome map (Whole chromosome) to the zoomed-in region between 35 and 39.5 Mb, where loop formation features become visible (Loops scale). The Bottom label indicates the genomic range for each zoom level, and the border labels display the sequence annotations by Rao et al. (5). Note that telomeric bins are omitted because Hi-C cannot reliably capture their repetitive sequences, and regions labeled as NA correspond to pericentromeric domains whose annotations differ slightly from intrachromosomal Hi-C signals; FI-Chrom infers their interactions directly from the experimental map. Aside from each zoom, a representative 3D structure colored by genomic distances is shown, with each row corresponding to the same zoom level in panel (A). (B) Hierarchical clustering analysis of 20,000 structures from the loop scale region. (C) Probability distribution function (PDF) of the similarity reaction coordinate Q . (D) PDF as a function of the radius of gyration (σ). (E) Violin plots comparing the probability distribution of fully formed loops in single-compartment (yellow) and multiple-compartment (pink) contexts.

(illustrated in the Whole chromosome and Compartments scale). On the other hand, we observe emergent features that go beyond the subcompartment annotations for shorter genomic distances (illustrated in the TADs and Loops scales). Particularly in the Loops scale (zoom in on a ~5 Mb length region), the formation of a “stripe” followed by a high-frequency contact is noticeable. This pattern may suggest topologically associating domains (TADs) emerging through the loop extrusion mechanism (49–51), with the “corner dot” being anchored by motifs bound to the architectural protein CTCF (highlighted by blue and red points in the last row in Fig. 4A). On the *Right* side, a representative 3D structure, focusing on the same regions, is shown in a zoomed view. The loop marked by CTCF boundaries (represented by red and blue beads) is illustrated in the Loop scale. This loop spans a genomic length of 1.25 Mb, corresponding to 25 beads in our model at a 50 kb resolution for chromosome 10.

Next, we will further explore the conformational aspects of this loop formation. Fig. 4B illustrates the hierarchical clustering analysis based on internal bead distances across 20,000 structures, with nine distinct clusters being identified. From the clustering analysis, the top four larger clusters are highlighted, along with their mean distance matrix and a representative structure. Cluster 2 (purple) represents open conformations, comprising 15.49% of the ensemble and exhibiting the greatest structural variability. Cluster 6 (blue) accounts for 12.55% of the ensemble and shows compacted structures but without the specific loop contacts. Cluster 7 is the largest cluster, representing 38.96% of the ensemble. To investigate full loop formation, we highlight Cluster 9 (yellow), which is primarily characterized by fully formed loops, mainly defined by the CTCF–CTCF contact, representing 14.17%.

We also explored the probability distributions using two structural reaction coordinates. Fig. 4C shows the structural similarity using the collective variable Q (see *SI Appendix* for more details), while Fig. 4D examines structural compaction through the radius of gyration. Cluster 9 (fully formed loops) exhibits the highest structural similarity ($Q \approx 0.75$) and shares a comparable radius of gyration with Clusters 6 and 7, suggesting a predominantly compact structure across these clusters. However, fully formed loops are relatively rare, constituting only 14.17% of the ensemble.

Finally, we investigated how loop formation correlates with chromosomal compartmentalization. Since compartmentalization emerges from underlying genomic interactions, it is plausible that the loop extrusion process would exhibit distinct behaviors depending on the compartmental context. To explore this idea, we analyzed loops annotated by Rao et al. (5) for chromosome 10, specifically selecting loops longer than 750 Kb (equivalent to more than 15 beads in our polymer model). Fig. 4E compares the probability of fully formed loops across 200,000 3D structures, categorizing loops into two groups: those contained entirely within a single compartment and those spanning multiple compartments. Loops within a single compartment were fully formed in approximately 21% of cases, slightly higher than the 17% observed for loops spanning multiple compartments. Notably, single-compartment loops exhibited a broader probability distribution, indicating greater variability. This finding suggests that the loop formation mechanism operates more efficiently and with a longer residence time within single compartments compared to loops crossing compartment boundaries.

This loop analysis can be extended to distinguish extrusion-dependent from extrusion-independent interactions. This was

achieved by performing additional simulations for chromosomes in mouse embryonic stem cells under three perturbative conditions: untreated, RAD21 depletion, and CTCF depletion (*SI Appendix*, Fig. S5). FI-Chrom accurately recapitulates the loss of extrusion-driven loops following RAD21 removal and the weakening of CTCF-anchored boundaries after CTCF depletion. The altered conformational landscapes revealed by hierarchical clustering further demonstrate that FI-Chrom captures not only steady-state structures but also the redistribution of structural ensembles induced by protein depletion, reinforcing its potential for modeling chromatin folding across diverse regulatory environments. (51, 52)

Discussion

Our work introduces the Full-Inversion Chromatin (FI-Chrom) model, a data-driven polymer designed to study 3D genome organization. Using inverse statistical mechanics and the maximum-entropy principle, FI-Chrom infers effective pairwise interactions directly from Hi-C data. This approach bridges experimental observations and theoretical modeling, providing a physically grounded route for reconstructing genome architecture *in silico*.

FI-Chrom explicitly incorporates all pairwise interactions into its calculations. To refine the optimization process, we employed a first-order gradient-based optimization method known as Adam (35). This approach enables us to identify interaction parameters that accurately reproduce the contact frequencies observed in Hi-C experiments. As an agnostic effective energy model, FI-Chrom offers a valuable alternative to traditional energy landscapes derived from physicochemical interactions based on first principles (13–15, 17, 18).

Additionally, the FI-Chrom model generates ensembles of 3D structures that faithfully preserve essential conformational features of chromosomes, including the A/B compartmentalization (Figs. 2H and 3A) and subcompartment structures (Fig. 3B). The spatial organization of these (sub)compartments aligns well with both experimental results (4, 42, 53) and previous simulation models (see *SI Appendix*, Fig. S3 for comparison) (25, 26). Notably, our trained force field avoids generating highly knotted topologies (Fig. 3C and D), which is crucial since highly knotted chromosome configurations might impede DNA segregation (25).

Despite achieving higher correlation with Hi-C contact frequencies, FI-Chrom’s overall 3D structures closely resemble those generated by our previous copolymer model, MiChroM (26). The similarity distribution between FI-Chrom and MiChroM, illustrated in *SI Appendix*, Fig. S3, is largely indistinguishable. Therefore, we conclude that FI-Chrom primarily enhances local-scale agreements with experimental data without significantly altering the large-scale structural organization defined by compartmentalization.

At shorter-range scales (segments from 1 to 5 Mb), FI-Chrom effectively captures emergent features associated with loop extrusion mechanisms (49, 54). By analyzing ensembles of 3D structures, we can statistically characterize the formation of loops. Fig. 4C shows that loops frequently adopt partially formed configurations (clusters 6 and 7), which correlate with stripes commonly observed in high-resolution Hi-C experiments (55). Structures containing fully formed loops (defined by anchored contacts) were observed in about 14.17% of simulated structures. This result is consistent with our finding that loops confined within single compartments show higher formation efficiency and greater variability compared to loops spanning multiple

compartments (Fig. 4E). Although FI-Chrom is trained assuming thermodynamic equilibrium and not explicitly calibrated for dynamic behavior, the generated ensemble samples the phase space extensively enough to yield density-of-state profiles consistent with experimental loop-formation data.

Beyond characterizing chromosome-scale dynamics, the FI-Chrom model allows exploration of energy landscapes within individual genomic domains independently of specific biochemical mechanisms. Its straightforward training and implementation, enabled by the Open-MiChroM toolkit (as detailed in *Materials and Methods*), suggest FI-Chrom's applicability even to genomes lacking obvious compartmentalization, such as bacterial chromosomes (56) and yeast genomes (8, 57), or investigate structural changes during the cell cycle (22).

Summarizing, FI-Chrom is a robust and versatile computational framework capable of generating chromosome models that accurately reproduce experimental Hi-C interactions. By combining statistical mechanics, machine learning optimization, and polymer physics, this approach significantly enhances our understanding of chromosome architecture. FI-Chrom enables the exploration of genome dynamics across a broader range of organisms and contexts, further bridging experimental observations with theoretical models.

Materials and Methods

Full-Inversion Chromatin Model Energy Function. The FI-Chrom energy function, denoted as $U(\vec{r})$, is described as

$$U(\vec{r}) = U_{HP}(\vec{r}) + \sum_{ij} \lambda_{ij} f(r_{ij}), \quad [1]$$

where $U_{HP}(\vec{r})$ is the homopolymeric backbone potential that includes nearest-neighbor bonds, angles, soft-core repulsion, and hard-core repulsion potentials, described in detail in *SI Appendix, Theory*. The coefficients λ_{ij} are Lagrange multipliers introduced by the maximum entropy approach (see *SI Appendix, Theory* for details). The λ_{ij} values are determined using iterative optimization to ensure that the simulated contact frequencies match the experimental measurements. $f(r_{ij})$ is the contact probability between genomic loci i and j , defined as

$$f(r_{ij}) = \frac{1}{2} \left(1 + \tanh \left[\mu(r_c - r_{ij}) \right] \right), \quad [2]$$

where the parameters μ and r_c are calibrated based on experimentally determined contact probability maps. The first constraint ensures that $f(r_{ij})$ returns 1 when loci are in close contact (i.e., $f(1) = 1$). The second constraint adjusts $f(r_{ij})$ to match the experimental probability for next nearest neighbors (i.e., $f(2) = \min\{P_{i,i+2}^{\text{exp}}\}$). In this work, values $\mu = 3.22$ and $r_c = 1.78$ were obtained from the GM12878 human cell line (5, 26).

Parameter Optimization. The optimization procedure identifies suitable values for the Lagrange multipliers λ_{ij} , which represent the interaction strengths between locus pairs i and j . For a polymer of length N beads, the number of parameters to optimize is $S = (N - 3)(N - 2)/2$. This formula accounts for all pairwise contacts, excluding first and second neighbors, already represented by bond and angle potentials. For instance, human chromosome 1 at 50 kb resolution has $N = 4,960$ beads, resulting in $S = 12,288,403$ interaction parameters. This number increases drastically with higher-resolution models, as parameters scale quadratically with the number of beads.

To manage these millions of parameters, we used the first-order optimizer Adam (35), which adaptively estimates the first and second moments of gradients to adjust learning rates individually. Specifically, Adam computes weighted moving averages of gradients (m_t) and squared gradients (v_t), using hyperparameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$. It employs bias corrections to counteract initial biases, applying updates based on corrected gradient

estimations. This adaptive method facilitates efficient convergence, especially with complex models and extensive datasets.

In the FI-Chrom model, the gradient is given by $g_t = \beta(\varphi_{ij}^{\text{exp}} - \varphi_{ij}^{\text{sim}})$, detailed further in *SI Appendix*, with β defined as $k_b T$, combining the Boltzmann constant (k_b) and information-theoretic temperature (T). Here, $\varphi_{ij}^{\text{sim}}$ and $\varphi_{ij}^{\text{exp}}$ represent simulation-derived and experimentally measured contact probabilities, respectively.

Training Protocol. The first step is processing Hi-C data to extract experimental contact probabilities. Starting with raw ligation counts, the matrix is balanced to achieve double stochasticity (constant row and column sums) via the Knight and Ruiz (KR) algorithm (58). The experimental contact probability matrix ($\varphi_{ij}^{\text{exp}}$) is obtained by normalizing each row by its maximal element, usually $C_{i,i+1}$, aligning with the model's assumption that adjacent beads are always in contact.

Next, an initial symmetric force field matrix (λ_{ij}^0) is created, initially set to zeros or initial guesses. Nonannotated regions (e.g., centromeres, telomeres) receive a baseline attraction (around $-0.2e$), averaged from all interactions, as illustrated in Fig. 1.

Chromosome simulations using the current force field iteration maximize sampling accuracy across multiple replicates, starting from varied initial configurations. Averaged simulation data are then used to update parameters via the Adam optimizer. Convergence is assessed through the error metric $\Delta = \sum_{ij} |\varphi_{ij}^{\text{sim}} - \varphi_{ij}^{\text{exp}}| / \sum_{ij} \varphi_{ij}^{\text{exp}}$. An error below the threshold (typically between 0.10 ~ 0.15) defines that the training is complete; otherwise, the process iterates further (Fig. 1).

Computer Simulations. Simulations were performed using the Open-MiChroM toolkit (34). Initially, polymer chains were equilibrated in open configurations (straight or spiral shapes), employing optional spherical flat-bottom harmonic constraints to generate collapsed structures. Following the collapse, this constraint was removed for a secondary equilibration step, which randomized configurations prior to sampling.

Simulations presented in this work involved 20 replicates, each consisting of 1×10^7 steps, with structures saved every 1×10^3 steps. This yielded 200,000 structures per training iteration. Langevin dynamics maintained a constant theoretical temperature ($T = 1.0$) with a damping coefficient of 10τ and a timestep $\Delta t = 0.01\tau$.

Data, Materials, and Software Availability. All computational methods described in this study are implemented in the Open-MiChroM toolkit (v1.1.1), available at <https://github.com/junioreif/OpenMiChroM> (59) and archived at Zenodo (DOI: 10.5281/zenodo.18087962) (60). The full set of scripts used for force-field optimization, training of interaction parameters, chromosome simulations, and downstream analyses are provided in the same repository under the folder 'Tutorials/'. Processed and raw Hi-C datasets used for training and validation were obtained from The Gene Expression Omnibus (GEO) accession number GSE63525 (61). The final trained interaction matrices and simulation input files supporting the findings of this work are deposited in Zenodo under DOI: <https://doi.org/10.5281/zenodo.18088647> (62). All other data are included in the manuscript and/or *SI Appendix*.

ACKNOWLEDGMENTS. We thank Peter Wolynes for important discussions and suggestions during the development of this work. This research was supported by the Center for Theoretical Biological Physics, sponsored by the NSF (Grants PHY-2019745, PHY-2014141, and PHY-2210291) and by the Welch Foundation (Grant C-1792). J.N.O. is a Cancer Prevention and Research Institute of Texas Scholar in Cancer Research. A.B.O.J. acknowledges the Robert A. Welch Postdoctoral Fellow Program. Financial support was also provided in part by the Brazilian agencies Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG, APQ-02303-21) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, 402495/2022-5 and 307992/2022-5). We would like to thank Advanced Micro Devices, Inc. for donating critical hardware and

support resources from its High-Performance Computing Fund, which made this work possible.

Author affiliations: ^aCenter for Theoretical Biological Physics, Rice University, Houston, TX 77005; ^bDepartamento de Física, Laboratório de Biofísica Teórica, Instituto de Ciências Exatas, Naturais e Educação, Universidade Federal do Triângulo Mineiro, Uberaba, MG

38025-180, Brazil; ^cDepartment of Physics and Astronomy, Rice University, Houston, TX 77005; ^dDepartment of Chemistry, Rice University, Houston, TX 77005; and ^eDepartment of Biosciences, Rice University, Houston, TX 77005

Author contributions: A.B.O.J., V.G.C., and J.N.O. designed research; A.B.O.J. and V.G.C. performed research; A.B.O.J. contributed new reagents/analytic tools; A.B.O.J., M.F.M., R.J.O., E.D.-R., S.B., V.G.C., and J.N.O. analyzed data; and A.B.O.J., V.G.C., and J.N.O. wrote the paper.

1. T. Cremer, C. Cremer, Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat. Rev. Genet.* **2**, 292–301 (2001).
2. T. Sexton, H. Schober, P. Fraser, S. M. Gasser, Gene regulation through nuclear organization. *Nat. Struct. Mol. Biol.* **14**, 1049–1055 (2007).
3. W. A. Bickmore, The spatial organization of the human genome. *Annu. Rev. Genomics Hum. Genet.* **14**, 67–84 (2013).
4. E. Lieberman-Aiden *et al.*, Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
5. S. S. P. Rao *et al.*, A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
6. A. L. Sanborn *et al.*, Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E6456–E6465 (2015).
7. T. H. S. Hsieh *et al.*, Mapping nucleosome resolution chromosome folding in yeast by micro-C. *Cell* **162**, 108–119 (2015).
8. T. H. S. Hsieh, G. Fudenberg, A. Goloborodko, O. J. Rando, Micro-C XL: Assaying chromosome conformation from the nucleosome to the entire genome. *Nat. Methods* **13**, 1009–1011 (2016).
9. G. Guroy, J. Liang, Three-dimensional chromosome structures from energy landscape. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 11991–11993 (2016).
10. M. Di Pietro, Inner workings of gene folding. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 4774–4775 (2019).
11. X. Lin, B. Zhang, Explicit ion modeling predicts physicochemical interactions for chromatin organization. *eLife* **12**, RP90073 (2024).
12. J. Moller, J. Lequieu, J. J. de Pablo, The free energy landscape of internucleosome interactions and its relation to chromatin fiber structure. *ACS Cent. Sci.* **5**, 341–348 (2019).
13. S. Tripathi, S. Brahmachari, J. N. Onuchic, H. Levine, DNA supercoiling-mediated collective behavior of co-transcribing RNA polymerases. *Nucleic Acids Res.* **50**, 1269–1279 (2022).
14. R. Collepardo-Guevara *et al.*, Chromatin unfolding by epigenetic modifications explained by dramatic impairment of internucleosome interactions: A multiscale computational study. *J. Am. Chem. Soc.* **137**, 10205–10215 (2015).
15. J. Nuebler, G. Fudenberg, M. Imakaev, N. Abdennur, L. A. Mirny, Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E6697–E6706 (2018).
16. S. Watanabe, Y. Mishima, M. Shimizu, I. Suetake, S. Takada, Interactions of HP1 bound to H3K9me3 dinucleosome by molecular simulations and biochemical assays. *Biophys. J.* **114**, 2336–2351 (2018).
17. G. Arya, T. Schlick, Role of histone tails in chromatin folding revealed by a mesoscopic oligonucleosome model. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 16236–16241 (2006).
18. Q. MacPherson, B. Beltran, A. J. Spakowitz, Bottom-up modeling of chromatin segregation due to epigenetic modifications. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 12739–12744 (2018).
19. V. G. Contessoto, R. R. Cheng, J. N. Onuchic, Uncovering the statistical physics of 3D chromosomal organization using data-driven modeling. *Curr. Opin. Struct. Biol.* **75**, 102418 (2022).
20. W. T. Chu, X. Chu, J. Wang, Uncovering the quantitative relationships among chromosome fluctuations, epigenetics, and gene expressions of transdifferentiation on waddington landscape. *Adv. Sci.* **9**, 2103617 (2022).
21. X. Chu, J. Wang, Conformational state switching and pathways of chromosome dynamics in cell cycle. *Appl. Phys. Rev.* **7**, 031403 (2020).
22. V. G. Contessoto *et al.*, Energy landscape analysis of the development of the chromosome structure across the cell cycle. *Proc. Natl. Acad. Sci. U.S.A.* **122**, e2425225122 (2025).
23. E. T. Jaynes, Information theory and statistical mechanics. *Phys. Rev.* **106**, 620 (1957).
24. J. W. Pitera, J. D. Chodera, On the use of experimental observations to bias simulated ensembles. *J. Chem. Theory Comput.* **8**, 3445–3451 (2012).
25. B. Zhang, P. G. Wolynes, Topology, structures, and energy landscapes of human chromosomes. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 6062–6067 (2015).
26. M. Di Pietro, B. Zhang, E. L. Aiden, P. G. Wolynes, J. N. Onuchic, Transferable model for chromosome architecture. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 12168–12173 (2016).
27. M. Di Pietro, R. R. Cheng, E. L. Aiden, P. G. Wolynes, J. N. Onuchic, De novo prediction of human chromosome structures: Epigenetic marking patterns encode genome architecture. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 12126–12131 (2017).
28. Y. Qi, B. Zhang, Predicting three-dimensional genome organization with chromatin states. *PLoS Comput. Biol.* **15**, e1007024 (2019).
29. D. Michieletto, E. Orlandini, D. Marenduzzo, Polymer model with epigenetic recoloring reveals a pathway for the de novo establishment and 3D organization of chromatin domains. *Phys. Rev. X* **6**, 041047 (2016).
30. E. Dodero-Rojas *et al.*, Pymegabase: Predicting cell-type-specific structural annotations of chromosomes using the epigenome. *J. Mol. Biol.* **435**, 168180 (2023).
31. S. Brahmachari, V. G. Contessoto, M. Di Pietro, J. N. Onuchic, Shaping the genome via lengthwise compaction, phase separation, and lamina adhesion. *Nucleic Acids Res.* **50**, 4258–4271 (2022).
32. V. G. Contessoto *et al.*, Interphase chromosomes of the *Aedes aegypti* mosquito are liquid crystalline and can sense mechanical cues. *Nat. Commun.* **14**, 1–12 (2023).
33. Y. Zhang *et al.*, Computational methods for analysing multiscale 3d genome organization. *Nat. Rev. Genet.* **25**, 123–141 (2024).
34. A. B. Oliveira Junior, V. G. Contessoto, M. R. Mello, J. N. Onuchic, A scalable computational approach for simulating complexes of multiple chromosomes. *J. Mol. Biol.* **433**, 166700 (2020).
35. D. P. Kingma, J. Ba, Adam: A method for stochastic optimization. *arXiv [Preprint]* (2014). <https://doi.org/10.48550/arXiv.1412.6980> (Accessed 15 October 2025).
36. J. R. Dixon *et al.*, Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
37. T. M. Yusufzai, H. Tagami, Y. Nakatani, G. Felsenfeld, CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. *Mol. Cell* **13**, 291–298 (2004).
38. G. Fudenberg *et al.*, Formation of chromosomal domains by loop extrusion. *Cell Rep.* **15**, 2038–2049 (2016).
39. ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57 (2012).
40. R. R. Cheng *et al.*, Exploring chromosomal structural heterogeneity across multiple cell lines. *eLife* **9**, e60312 (2020).
41. M. Falk *et al.*, Heterochromatin drives compartmentalization of inverted and conventional nuclei. *Nature* **570**, 395–399 (2019).
42. J. H. Su, P. Zheng, S. S. Kinrot, B. Bintu, X. Zhuang, Genome-scale imaging of the 3D organization and transcriptional activity of chromatin. *Cell* **182**, 1641–1659.e26 (2020).
43. L. A. Mirny, The fractal globule as a model of chromatin architecture in the cell. *Chromosom. Res.* **19**, 37–51 (2011).
44. P. Pietrot, IN SEARCH OF IDEAL KNOTS*. *CMST* **4**, 9–23 (1998).
45. O. Gonzalez, J. H. Maddocks, Global curvature, thickness, and the ideal shapes of knots. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 4769–4773 (1999).
46. A. Y. Grosberg, Y. Rabin, Metastable tight knots in a wormlike polymer. *Phys. Rev. Lett.* **99**, 217801 (2007).
47. A. R. Khokhlov, A. Y. Grosberg, V. S. Pande, *Statistical Physics of Macromolecules* (American Inst. of Physics, 1989).
48. J. C. Wang, Cellular roles of DNA topoisomerases: A molecular perspective. *Nat. Rev. Mol. Cell Biol.* **3**, 430–440 (2002).
49. G. Fudenberg *et al.*, Formation of chromosomal domains by loop extrusion. *Cell Rep.* **15**, 2038–2049 (2016).
50. J. E. Phillips, V. G. Corces, CTCF: Master weaver of the genome. *Cell* **137**, 1194 (2009).
51. M. Gabriele *et al.*, Dynamics of CTCF- and cohesin-mediated chromatin looping revealed by live-cell imaging. *Science* **376**, 496–501 (2022).
52. T. H. S. Hsieh *et al.*, Enhancer-promoter interactions and transcription are largely maintained upon acute loss of CTCF, cohesin, WAPL or YY1. *Nat. Genet.* **54**, 1919–1932 (2022).
53. Y. Takei *et al.*, Integrated spatial genomics reveals global architecture of single nuclei. *Nature* **590**, 344–350 (2021).
54. M. Ganji *et al.*, Real-time imaging of DNA loop extrusion by condensin. *Science* **360**, 102–105 (2018).
55. S. S. Rao *et al.*, Cohesin loss eliminates all loop domains. *Cell* **171**, 305–320 (2017).
56. X. Wang *et al.*, Condensin promotes the juxtaposition of DNA flanking its loading site in bacillus subtilis. *Genes Dev.* **29**, 1661–1675 (2015).
57. L. Lazar-Stefanita *et al.*, Cohesins and condensins orchestrate the 4D dynamics of yeast chromosomes during the cell cycle. *EMBO J.* **36**, 2684–2697 (2017).
58. P. A. Knight, D. Ruiz, A fast algorithm for matrix balancing. *IMA J. Numer. Anal.* **33**, 1029–1047 (2013).
59. A. B. Oliveira Junior, OpenMiChroM GitHub repository. GitHub. <https://github.com/junioreif/OpenMiChroM>. Accessed 20 December 2025.
60. A. B. Oliveira Junior, OpenMiChroM: A toolkit for chromatin polymer simulations (version 1.1.1). Zenodo. <https://doi.org/10.5281/zenodo.18087962>. Deposited 29 December 2025.
61. S. Rao, M. Huntley, E. Lieberman Aiden, A three-dimensional map of the human genome at kilobase resolution reveals principles of chromatin looping. *Gene Expression Omnibus*. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE63525>. Accessed 15 October 2025.
62. A. B. Oliveira Junior, FI-Chrom trained interaction matrices and simulation input files. Zenodo. <https://doi.org/10.5281/zenodo.18088647>. Deposited 29 December 2025.