PNAS
RESEARCH ARTICLE | BIOPHYSICS AND COMPUTATIONAL BIOLOGY

Theory of chromatin organization maintained by active loop extrusion

Brian Chan[a] and Michael Rubinstein[a,b,c,d,e,1]

Edited by Andrew J. Spakowitz, Stanford University, Stanford, CA; received January 2, 2023; accepted April 13, 2023 by Editorial Board Member Monica Olvera de la Cruz

The active loop extrusion hypothesis proposes that chromatin threads through the cohesin protein complex into progressively larger loops until reaching specific boundary elements. We build upon this hypothesis and develop an analytical theory for active loop extrusion which predicts that loop formation probability is a nonmonotonic function of loop length and describes chromatin contact probabilities. We validate our model with Monte Carlo and hybrid Molecular Dynamics–Monte Carlo simulations and demonstrate that our theory recapitulates experimental chromatin conformation capture data. Our results support active loop extrusion as a mechanism for chromatin organization and provide an analytical description of chromatin organization that may be used to specifically modify chromatin contact probabilities.

Significance

The loop extrusion model describes organization of chromatin into dynamic loops by protein complexes such as cohesin. We developed a theoretical model that quantifies how key parameters, including cohesin residence time on chromatin, extrusion velocity, and the number density of chromatin-bound cohesins, regulate genomic contacts. This work informs strategies for epigenetic regulation to engineer topologically associated domains and modify cohesin properties for amplifying or suppressing the frequency of contacts between loci of interest, such as enhancers and promoters. Additional details like cohesin binding locations can supplement our model to fine-tune predictions for specific genomic regions.

loop extrusion | chromatin organization | polymer physics

Motivation

One way to regulate gene expression is to modify chromatin contact probabilities between specific genomic sites, for example, an enhancer and promoter. An analytical description of chromatin organization would provide crucial information in deciding how to selectively modify chromatin contacts and thus gene expression. This work provides analytical equations with biologically interpretable parameters that describe how active loop extrusion regulates chromatin during interphase. Our model quantifies the degree to which these parameters should be modulated to produce predictable changes in chromatin contacts.

Chromatin Organization

To provide context for our work, we review basic levels of chromatin organization as well as the experimental techniques used and models proposed to explain it. In human cells, DNA is organized into 46 complexes called chromosomes that range from 50 megabase pairs (Mbp) to 250 Mbp in genomic length (1). Chromosomes segregate into distinct territories in the nucleus (Fig. 1*A*) (2, 3). Each chromosome contains one long complex known as chromatin, consisting of double-stranded DNA and proteins (3, 4). Chromatin is generally classified as either euchromatin or heterochromatin. Euchromatin has loosely packed DNA and proteins and is more accessible to protein binding. Heterochromatin has a tight packing of DNA and proteins and is less accessible (5). Chromatin organization and its regulation between the scales of nucleosomes and chromosomes (between 200 bp and 100 Mbp) are not well understood.

Experimental Techniques to Probe Chromatin Conformations

Fluorescence in situ hybridization (FISH) experiments have measured rms spatial distances $r(s)$ between chromatin sections as a function of genomic distance in terms of base pairs s, for s between 100 kilobase pairs (kbp) and 100 Mbp and spatial distances between 500 nm and 10 μm (6, 9, 10). FISH data are often used to infer a genomic region's fractal dimension D, which is the scaling exponent that describes how the mass m of a part of size r of the object varies with r (*i.e.*, $m \propto r^D$ or $r \propto m^{\frac{1}{D}}$). Although FISH does not directly probe chromatin mass, studies assume that the genomic distance s is directly proportional to mass. Under this approximation, early FISH data on human chromosome 4 suggest that for 100 kbp $< s <$ 1 Mbp, chromatin has on average a fractal dimension of $D \approx 2$ with $< r^2 > \approx 3 \frac{\mu\mathrm{m}^2}{\mathrm{Mbp}} s$, while for larger s, a more compact packing with $D \approx 3$ is observed (see blue lines in Fig. 1*B*) (7).

In addition to imaging, chromatin organization is probed by chromosome conformation capture techniques such as chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) (11) and Hi-C (8, 12), which probe physical contacts between chromatin

Author contributions: B.C. and M.R. designed research; B.C. performed research; B.C. and M.R. analyzed data; and B.C. and M.R. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. A.J.S. is a guest editor invited by the Editorial Board.

Copyright © 2023 the Author(s). Published by PNAS. This article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

[1] To whom correspondence may be addressed. Email: michael.rubinstein@duke.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2222078120/-/DCSupplemental.

Published May 30, 2023.

Downloaded from https://www.pnas.org by DUKE UNIVERSITY LIBRARIES ACQUISITIONS-PERIODICAL on June 13, 2023 from IP address 152.16.191.125.

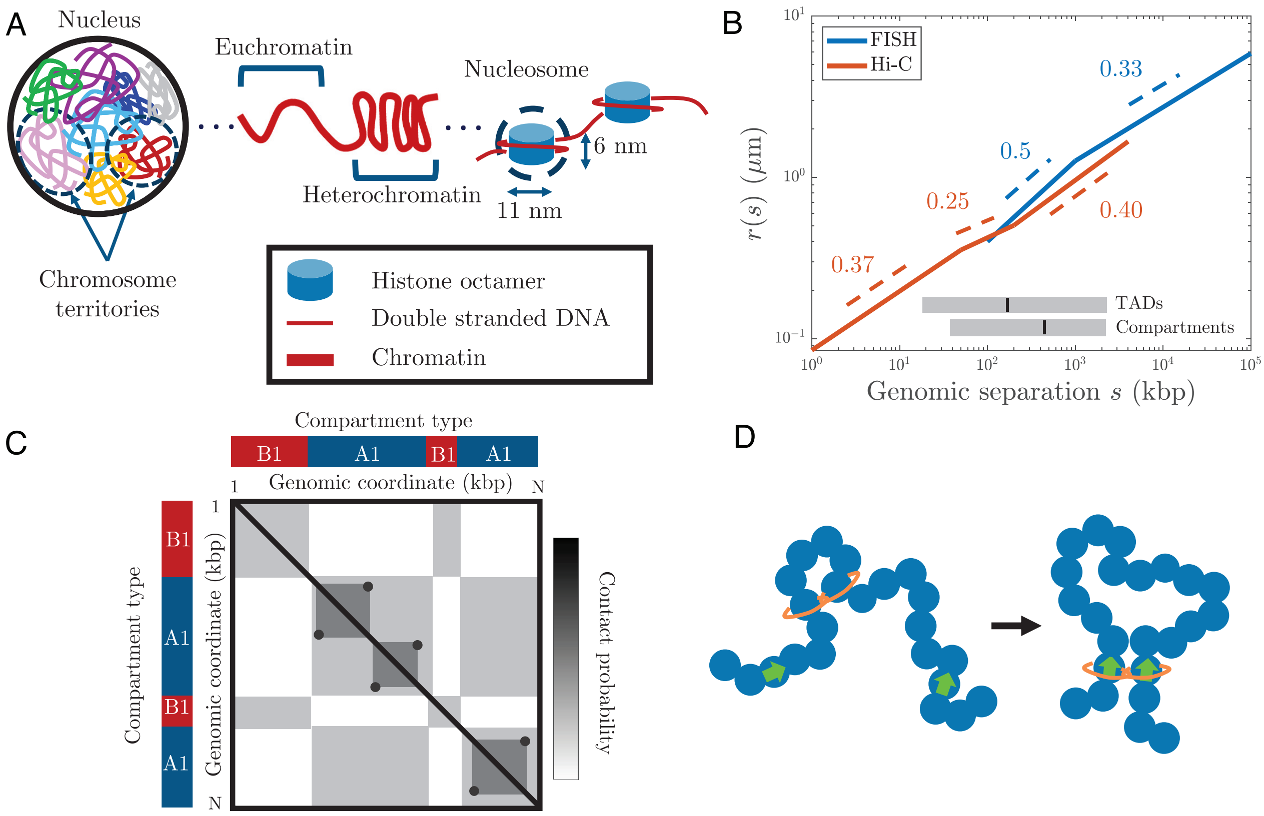

Fig. 1. Overview of chromatin organization. (*A*) Chromatin is made of nucleosome arrays, which are complexes of ~150 bp of double-stranded DNA wrapped 1.65 times around an octamer of histone proteins. Nucleosomes are about 11 nm in diameter and 6 nm in height and are connected by about ~50-bp DNA spacers (3, 4). (*B*) Schematic of chromatin rms internal distances *r(s)* measured by FISH (blue lines) (6, 7) and suggested by Hi-C (red lines) (8). Absolute values for Hi-C are scaled to match FISH data at 0.5 μm. Numbers next to dashed lines indicate slopes on a log-log scale (reciprocal of fractal dimensions *D*). (*C*) Schematic of TADs and compartmentalization in a Hi-C contact map for an *N* kbp chromatin segment. (*D*) Schematic of loop extrusion. Cohesin (orange rings) extrudes chromatin (blue circles) until reaching CTCF anchors (green arrows), creating anchored loops.

Downloaded from https://www.pnas.org by DUKE UNIVERSITY LIBRARIES ACQUISITIONS-PERIODICAL on June 13, 2023 from IP address 152.16.191.125.

fragments with resolution down to 1 kbp. ChIA-PET detects contacts between chromatin fragments bound by specific proteins by using antibodies to target a protein of interest (11). Hi-C produces contact maps, or heat maps, that indicate the probability that two chromatin fragments are in spatial proximity (8, 12).

Excluded Volume and Topological Constraints

Fractal dimension $D \approx 2$ is predicted by the ideal linear chain model applicable for interpenetrating flexible linear polymers on length scales larger than correlation length (13). For this reason, many studies model chromatin as a polymer chain with fractal dimension $D = 2$ to understand higher-level chromatin organization (7, 10, 12, 14–16). The ideal linear chain model does not have excluded volume, while real polymers and chromatin clearly do. However, for polymer melts and solutions at sufficient concentrations, excluded volume interactions are screened at length scales larger than correlation length, allowing ideal chain statistics to be used at these length scales (13). Additionally, for real polymers, entanglements due to the uncrossability of chains may play a role on long length scales (13). However, the topoisomerase II enzyme can pass genomic regions through each other, thus removing some topological constraints (17), and recent studies suggest that chromosomes and chromatin domains are largely unentangled (18, 19). Note that the apparent fractal dimensions may vary for different cell lines and genomic regions, as other studies find that the apparent fractal dimension can be greater than 3, even on scales smaller than 500 kbp (20, 21). More detailed imaging experiments are needed to understand the chromatin structure of specific genomic regions.

Contact Probabilities

From Hi-C contact maps, one can calculate the average contact probability between chromatin sections as a function of genomic distance s characterized by a power law with exponent γ,

$$p(s) \approx As^{-\gamma}, \qquad [1]$$

where A is a constant. The mean-field prediction for the exponent γ for objects with fractal dimension D in spatial dimension d is $\gamma = d/D$, though this likely oversimplifies the relation between chromatin size and contact probability (14). On average, human chromatin can exhibit at least three contact probability scaling regimes with different γ exponents. Some data show $1 < \gamma_1 \leq 1.5$ on scales shorter than 100 kbp, corresponding to fractal dimension between $D \approx 2$ and $D \approx 3$ at a mean-field level (see *SI Appendix*, Fig. S14 for example) (22). On average, the second and third scaling regimes have $\gamma_2 \approx 0.75$ and $\gamma_3 \approx 1.3$ corresponding to fractal dimensions of $D \approx 4$ and $D \approx 2.3$ respectively, with the cross-over occurring at several hundred kbp (23, 24). In Fig. 1*B*, we compare FISH data from chromosome 4 with these three scaling regimes from the average contact probability in a specific 10 Mbp region of chromosome 8 with coordinates 130 Mbp—140 Mbp in GM12878 cells (*SI Appendix*, Fig. S14) using $\gamma_1 \approx 1.1, \gamma_2 \approx 0.76$, and $\gamma_3 \approx 1.2$, assuming the mean-field relation $\gamma = d/D$. Note that contact probability scaling exponents in specific chromatin regions may differ (14).

The γ_2 exponent has been observed to change depending on the presence of different proteins thought to regulate chromatin loops such as cohesin [a structural maintenance of chromosomes (SMC) protein complex] and the nipped-B-like protein NIPBL (a protein that facilitates cohesin–chromatin binding). Cells depleted of cohesin or NIPBL exhibit an increase in average γ_2 on scales smaller than 200 kbp (22, 24, 25). For example, a 50 Mbp region in chromosome 8 of HCT116 cells shows an increase from $\gamma_2 \approx 0.73$ to $\gamma_2 \approx 0.94$ corresponding to a decrease in fractal dimension from $D \approx 4.2$ to $D \approx 3.2$ (*SI Appendix*, Fig. S15) (22).

The three different scaling regimes of exponent γ suggest that chromatin organization changes on the corresponding length scales. As an analogy, consider the different fractal dimensions in a melt of densely packed nonconcatenated polymer rings. These polymers have fractal dimension $D = 2$ on small length scales between Kuhn length and entanglement length, and $D = 3$ on larger scales due to the nonconcatenation condition according to the fractal loopy globule model (26). If these polymer rings form double-folded conformations, the polymers would have $D = 4$ on scales larger than distance between branchings of these double-folds (14, 27, 28). Since the spatial dimension is $d = 3$, the fractal dimension should decrease on the largest length scales to maintain a polymer volume fraction below unity. While the different γ exponents observed in human chromatin are consistent with the fractal dimensions discussed above, they could be caused by the interplay of polymer relaxation and active processes perturbing chromatin regions facilitated by proteins such as cohesin, and possibly by the effect of nuclear confinement, rather than an equilibrium melt of rings with transient entanglements.

Topologically Associated Domains (TADs) and Compartmentalization

Hi-C also reveals two levels of chromatin organization known as TADs and compartmentalization (Fig. 1*C*). TADs appear as squares along the main diagonal of a contact map ranging from tens of kbp to several Mbp (24, 29). TADs are also marked by increased contacts between loci at TAD boundaries or anchors, which appear as dark corners of the TAD squares. This corresponds to local enhancement of contact probability within a continuous chromatin section. Human TADs range from 20 kbp to 3 Mbp in genomic length with a median of about 200 kbp (8) (see gray and black bars in Fig. 1*B*). Compartmentalization appears as a checkered pattern with widths ranging from 40 kbp to 3 Mbp and a median of approximately 500 kbp (8, 24) (see gray and black bars in Fig. 1*B*). Compartmentalization is interpreted as long-range associations between nonconsecutive chromatin sections. Six compartment types (A1, A2, B1, B2, B3, and B4) correlated with distinct epigenetic marks have been observed, and chromatin regions of one type preferentially associate with other regions of the same type (8, 12, 24). Below we focus on describing TAD regulation.

Regulation of TADs

TAD regulation models. The processes that regulate TADs are unclear. One model is loop extrusion, in which a protein extrusion complex binds or topologically embraces two neighboring loci and translocates along chromatin, forming larger and larger loops until reaching and stalling at protein boundary elements oriented in the correct directions (Fig. 1*D*) (16, 23, 30–36). Here, we define a loop as a section of chromatin with two loci bound to each other. The loop extrusion model ensures that genomic regions nearby along the chromatin contour colocalize, which may be important for bringing *cis*-regulatory elements in contact with gene promoters. Another model known as String and Binders Switch or transcription factor (TF) binding posits that multivalent TFs or protein complexes bind to specific genomic loci to form local loops (37–42). This model makes no guarantee that the smallest loops between binding sites are formed. The present paper focuses on loop extrusion and argues that simple TF-binding models are unlikely to regulate certain classes of chromatin loops.

Loop extruding proteins. Although the exact proteins critical to loop extrusion are unclear, both the cohesin and condensin SMC complexes exhibit extrusion ability in vitro, and the CCCTC-binding factor (CTCF) is the leading candidate for the boundary elements (43–48). Condensin organizes chromatin in mitotic cells, while cohesin is thought to organize chromatin in both mitotic and interphase cells (49). Therefore, for definiteness, for the remainder of this paper, we refer to "cohesin" and "CTCF anchors" as the extrusion complex and boundary elements respectively, though we note that the precise proteins involved in loop extrusion are not resolved.

It is unclear whether in vivo cohesin topologically embraces chromatin within a ring-like structure or directly binds chromatin and whether cohesin acts as a monomer with a single ring that can bind or embrace two loci at the ends of a loop or a dimer with two rings that each embraces or binds chromatin. For definiteness, we refer to cohesin as having two "domains", one positive (+) and one negative (–), that can each bind or embrace chromatin. Although the choice of (+) and (–) domains is arbitrary, in this work, we identify the (+) domain as translocating to larger genomic coordinates, while the (–) domain translocates to smaller genomic coordinates. We do not distinguish between binding and topologically embracing, nor between monomer and dimer, as our model does not depend on these details.

Loop extruding mechanisms. Precise details of loop extrusion related to the activity, symmetry, synchronicity, traversal, binding specificity, and binding activity of cohesin remain unresolved. In active models, cohesin domains consume energy and act as motors to translocate in biased directions along the chromatin fiber toward CTCF anchors (16, 23, 30, 31). In passive models, extrusion is thermally driven and does not have inherent directionality (32–35). Cohesin could extrude chromatin symmetrically, in which two domains of the extrusion complex simultaneously translocate in opposite directions with the same rates, or asymmetrically, in which the domains extrude with different rates.

The two cohesin domains could extrude synchronously, in which the domains always translocate at the same time, or asynchronously, in which the domains do not necessarily translocate together. Furthermore, cohesins may or may not be able to traverse each other during extrusion, as seen in other SMC complexes (50, 51), and recent experiments have shown that cohesin can pass over DNA-bound particles larger than its size in vitro (52). Traversal could be impeded, in which cohesin domains slow down when they encounter one another, or unimpeded, in which the domains do not change extrusion speed upon encounter.

Two issues arise due to binding: cohesin's binding specificity and activity. Cohesin could either bind to specific loci (i.e., one or a few sites per TAD) or nonspecifically, with the number of binding sites proportional to TAD genomic length (31). Note that while cohesin and NIPBL have been observed to colocalize and accumulate on chromatin, there is no direct evidence that cohesin binds to specific locations along the genome (31, 49, 53). Lastly, cohesin binding may or may not consume ATP. Studies have shown that chromatin-bound cohesin is insensitive to the presence of ATP and that cohesin can bind DNA without ATP (53, 54).

Downloaded from https://www.pnas.org by DUKE UNIVERSITY LIBRARIES ACQUISITIONS-PERIODICAL on June 13, 2023 from IP address 152.16.191.125.

Modeling Chromatin Organization Regulation

Theoretical and simulation models are used to study loop extrusion and TF binding as well as compartmentalization. Chromatin is often modeled as a coarse-grained bead-spring polymer due to its polymeric nature and apparent fractal dimension $D = 2$ at smaller length scale (Fig. 1*B*). Block heteropolymers are used to simulate compartmentalization, in which chromatin regions with similar histone marks have stronger attractive interactions than those with different histone marks (15, 25, 55). TFs are represented as i) additional beads that can bind to the polymer, ii) short-range attraction between specific beads, or iii) transient bonds between specific beads (38, 56–58). Loop extrusion is modeled either as an additional bond between beads that processes along the chromatin fiber or as rings that encircle chromatin (16, 23, 35, 59).

Previous studies have demonstrated that passive loop extrusion (32, 33), active loop extrusion (16, 23), and TF binding (38, 39) can all form TADs of physiological genomic lengths. Several theoretical and simulation studies suggest that passive loop extrusion with specific cohesin binding and no cohesin traversal could form TADs due to an osmotic pressure arising from neighboring cohesins (32–34). On the other hand, one study that compared active loop extrusion with an active–passive hybrid (in which one domain of cohesin was active and the other was passive) found that the hybrid model could not recover TAD corners (31).

Another approach to modeling chromatin organization is to construct effective energy landscapes for chromatin folding using epigenetic data, for example, locations of CTCF-associated loops (60–62). Molecular dynamics simulations of polymers equilibrated in these energy landscapes recapitulate Hi-C contact maps, contact probability scaling, and compartmentalization of specific genomic regions. These models are agnostic to specific biochemical mechanisms but nevertheless provide important insight into chromatin organization.

Modeling and experiments produce wide predicted ranges of loop extrusion parameters known as "cohesin processivity" (the average loop length extruded by unimpeded cohesin) and "cohesin separation" (the average genomic spacing between bound cohesins) denoted λ and d, respectively. For example, active extrusion models suggest cohesin processivity to be about 200 kbp (16, 63). However, experimental studies of cohesin residence time in interphase (~20 to 25 min) (64, 65) and extrusion speed in interphase and ex cellula (~20 to 60 kbp/min) (25, 47, 48) suggest longer processivities between 400 and 1,500 kbp. Previous simulations predict a cohesin separation of about 120 kbp, while experimental estimates range from 50 kbp to 300 kbp (16, 65, 66). Additionally, studies that simulate TF binding demonstrate that bound TFs tend to cluster, stabilize loops, and form TADs (38, 59, 67, 68). Moreover, TF binding remains a possibility as extrusion in live cells has yet to be observed. Thus, despite the vast number of studies that model chromatin organization, there is no consensus on the processes that occur in vivo.

Chromatin Organization and Gene Regulation

Chromatin organization is important for regulating gene expression (37, 69). Certain chromatin compartments identified by Hi-C are gene rich and correlate with active histone marks, while others correlate with repressive histone marks (8, 12). Chromatin organization can impact gene expression by facilitating or insulating contacts between gene promoters and regulatory elements (such as enhancers and silencers), which can be separated by hundreds of kbp (37, 70–72). Additionally, TAD boundaries are enriched for gene transcription start sites and contacts between promoters and enhancers, and in some cases, genes within the same TAD are coregulated (8, 29, 73). Epigenetic tools have been developed that modulate gene expression by altering chromatin organization, such as modifying chromatin accessibility and engineering specific loci to colocalize and form loops (74–76). Thus, an analytical description of chromatin organization could help inform experiments to modify gene expression.

Overview of Results

Our work supports active loop extrusion with nonspecific, steady-state binding as a mechanism for chromatin organization. We develop a model of chromatin organization regulated by active loop extrusion with three main components: i) a binding and unbinding component, ii) a one-dimensional component of loop extrusion kinetics, and iii) a three-dimensional component for the spatial organization of chromatin. The first two components are similar to previous models (31, 77), though we focus our results on TAD formation and contact probabilities rather than chromosome compaction. Our model predicts chromatin contact probabilities, which we compare with Hi-C and ChIA-PET data targeted toward CTCF since these two experimental assays are commonly used to probe chromatin organization. We also use our model to describe the effects of cohesin traversal on chromatin loop distribution.

Results and Discussion

Model Description. Here, we outline our theoretical model for active loop extrusion. See *SI Appendix* for additional detail. In this work, active extrusion refers to cohesin domains consuming energy and translocating in biased directions toward CTCF anchors (16, 23, 30, 31), while passive extrusion refers to cohesin domains translocating purely based on thermal fluctuations (32–35). Consider a chromatin region discretized into loci of z kbp each, with a TAD containing N loci defined by CTCF anchors at loci i and $i+N$. The genomic distance between anchors is zN kbp. Cohesins initially bind to consecutive loci x and $x+1$. We assume that the number density of chromatin-bound cohesins during the G1 phase of the cell cycle is time independent corresponding to a steady state. For definiteness, in this paper, we describe passive cohesin binding that follows equilibrium detailed balance and does not consume energy, as in previous work (77). The free-energy change per pair of loci due to cohesin binding with two domains attached to different chromatin loci is $\mu k_B T$ (Fig. 2*B*). The average steady-state separation between neighboring bound cohesins is e^{μ} loci, corresponding to $d = z\mathrm{e}^{\mu}$ kbp. For example, with $z = 1$ (each locus representing 1 kbp), the estimated cohesin separation range from 50 kbp to 300 kbp (66) corresponds to a chemical potential range of $\mu k_B T \approx 4 k_B T$ to $\approx 6 k_B T$. Cohesins have an energy barrier for unbinding $h k_B T > 0$. We choose a single frequency ν to describe attempt rates of cohesin binding, unbinding, and translocation. Cohesin residence time is $\tau_r \approx \frac{1}{\nu} e^{h}$. In the example with $z = 1$, cohesin separation is $\mathrm{e}^{\mu} \approx 150$ kbp for $\mu k_B T = 5 k_B T$ and residence time is $\nu^{-1} e^{h} \approx 30$ min for $\nu = 0.1\ \mathrm{s}^{-1}$ and $h k_B T = 5 k_B T$.

The (–) and (+) cohesin domains translocate to neighboring loci with probabilities p_- and p_+ per timestep with duration ν^{-1} respectively, independent of one another (Fig. 2 *A*, *C*, and *D*). The (–) and (+) domains cannot translocate further than loci i or $i+N$, respectively. Although experimental and theoretical studies have shown that condensin loop extrusion velocity may be dependent on external load applied to DNA (78, 79), we coarse-grain the biophysical details of a loop extrusion step and assume a constant translocation probability on the scale of our

Downloaded from https://www.pnas.org by DUKE UNIVERSITY LIBRARIES ACQUISITIONS-PERIODICAL on June 13, 2023 from IP address 152.16.191.125.

Downloaded from https://www.pnas.org by DUKE UNIVERSITY LIBRARIES ACQUISITIONS-PERIODICAL on June 13, 2023 from IP address 152.16.191.125.

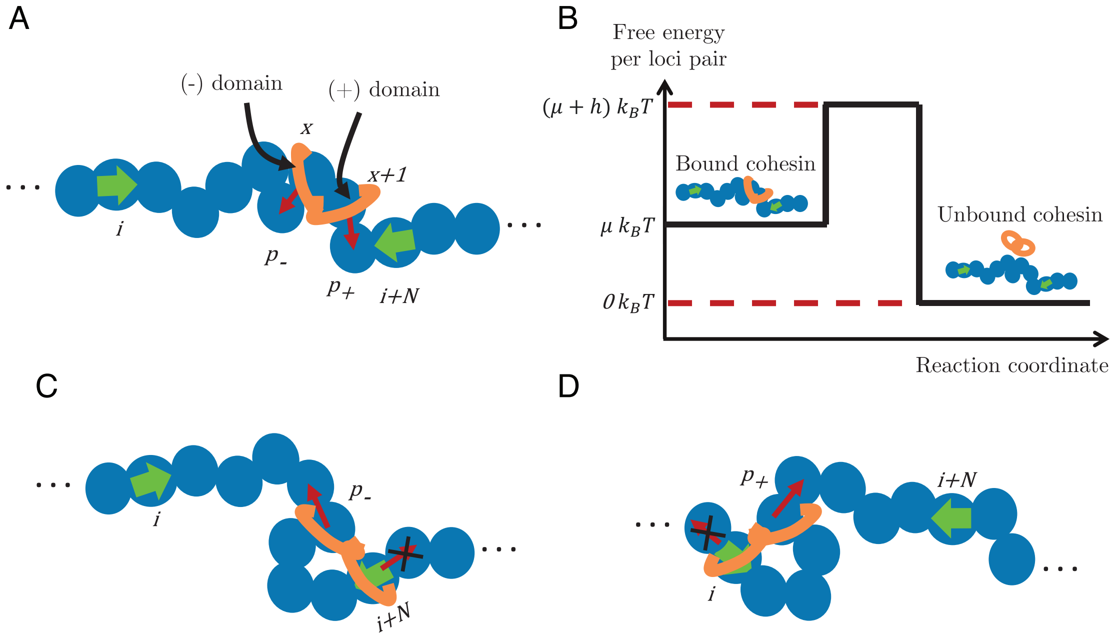

Fig. 2. Schematic description of active extrusion model. Blue circles indicate non-CTCF loci. Green arrows indicate CTCF anchors. The right-pointing arrow at locus *i* stops (−) cohesin domains, and the left-pointing arrow at locus *i*+*N* stops (+) cohesin domains. A pair of orange rings represents cohesin. (*A*) Cohesins bind to a pair of consecutive loci *x* and *x*+*1*. (*B*) Energy landscape that determines cohesin binding and unbinding kinetics. (*C*) The (+) domain does not translocate if bound to the anchor at locus *i*+*N*. (*D*) The (−) domain does not translocate if bound to the anchor at locus *i*.

loci discretization (1 kbp per locus). In the main text, we assume that cohesin traversal is unimpeded, meaning that cohesin domains do not slow down upon meeting neighboring cohesins. We explore impeded cohesin traversal in *SI Appendix*. For simplicity, the remainder of this paper assumes that $p_- = p_+ = p$. Unimpeded cohesin processivity in kbp is given by $\lambda = 2p\nu z\tau_r = 2pze^h$ (the factor of two comes from the translocation of both domains). Recall that z is the number of kbp per locus and ν is the binding, unbinding, and translocation attempt frequency. A residence time of 30 min and a probability $p = 0.5$ of translocating 1 kbp (one locus) per $\nu^{-1} = 10$ s results in a processivity of approximately 150 kbp.

We define two dimensionless parameters

$$\alpha = \frac{Nz}{\lambda} = \frac{N}{2p\nu\tau_r}, \qquad [2]$$

and

$$\beta = \frac{\lambda}{d} = 2p\nu\tau_r e^{-\mu}, \qquad [3]$$

α is the ratio between TAD length and cohesin processivity. $\alpha \ll 1$ indicates that on average, most cohesins that bind within a TAD will extrude the entire TAD, while $\alpha \gg 1$ suggests that the average length of a cohesin-mediated loop is smaller than the TAD length. β provides a measure of loop nesting and is the ratio between cohesin processivity and separation (Fig. 3), also described in ref. 77. Note that the number of kbp per loci z cancels out in the definitions of both α and β.

Active Loop Extrusion with Nonspecific, Passive Binding and Unimpeded Cohesin Traversal Produces TAD Squares and Multiple Contact Probability Scaling Regimes. Hybrid MD–MC simulations of active extrusion with nonspecific, passive binding and unimpeded cohesin traversal produce squares along the main diagonal of contact maps like TADs in human Hi-C contact maps (Fig. 4*A*). The example in Fig. 4*A* shows the results of a hybrid MD–MC simulation of an ideal polymer chain with $\alpha \approx 2$ and $\beta \approx 0.5$ (see Eqs. **2** and **3** for definitions of α and β), corresponding to the case with limited nesting and the two anchors not held together very frequently. The parameters are chosen to produce biologically relevant cohesin processivity and separation of approximately 100 kbp and 200 kbp, respectively. In contrast, simulations of passive extrusion as well as simulations with specific, passive binding and completely impeded cohesin traversal do not produce this pattern (*SI Appendix*, Fig. S9). Furthermore, while simulations with specific binding produce TAD corners, they also produce "X" patterns, with secondary diagonals perpendicular to the main diagonal (*SI Appendix*, Fig. S9*B*). These secondary diagonals, sometimes called chromatin jets, are seen in Hi-C in certain species such as *Bacillus subtilis* bacteria and specific genomic sites in humans (31, 80, 81). Although we do not focus on this phenomenon, our theory can be extended to consider this case.

Active loop extrusion causes multiple contact probability scaling regimes. When cohesin separation is larger than half the cohesin processivity, there are three regimes: one on scales shorter

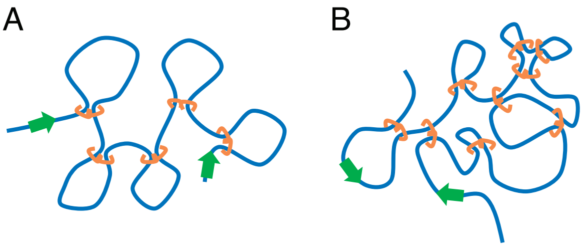

Fig. 3. Examples of the effect of cohesins (orange ovals) on chromatin (blue curves) conformations for (*A*) $\beta \ll 1$and (*B*) $\beta \gg 1$. Green arrows represent CTCF anchors. For small β, cohesin-bound loops do not overlap, forming an array of loops. For large β, cohesin-bound loops form nested structures. Both (*A*) and (*B*) correspond to $\alpha \gg 1$in which the TAD length is much longer than the average length of a cohesin-mediated loop.

Downloaded from https://www.pnas.org by DUKE UNIVERSITY LIBRARIES ACQUISITIONS-PERIODICAL on June 13, 2023 from IP address 152.16.191.125.

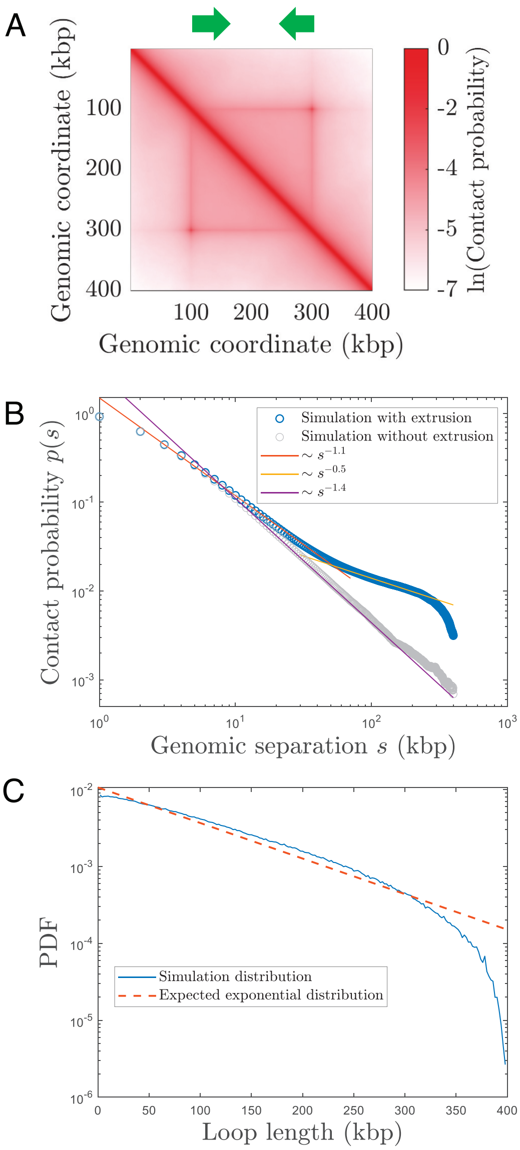

Fig. 4. Hi-C and contact probabilities of active loop extrusion from MD–MC simulations. (*A*) Simulated Hi-C contact map. Green arrows indicate two CTCF anchors at 100 kbp and 300 kbp, which stop (−) and (+) cohesin domains, respectively. (*B*) Average contact probability as a function of genomic separation with the same parameters as in (*A*) but without anchors (blue circles) and the MD simulation for the same chain without extrusion (gray circles). (*C*) Distribution of cohesin-mediated loop lengths for the MD–MC simulation with extrusion but without anchors.

than half of processivity, a second between half the processivity and separation, and a third on scales larger than separation (*SI Appendix*, Fig. S16*A*). The first cross-over occurs at approximately half the cohesin processivity because the contact probability of a typical cohesin-mediated loop is nonmonotonic (as cohesin holds two loci together), with the local minimum at approximately half the loop length. As the processivity and separation length scales approach each other, the middle regime collapses into a single cross-over (*SI Appendix*, Fig. S16*B* and Fig. 4*B*). A last regime may be present on even larger scales, to ensure a volume fraction less than unity, though it would not be present in our simulations of an ideal polymer without excluded volume. The contact probability behavior is more complex when cohesin separation is smaller than half the cohesin processivity and will be discussed in future work.

Fig. 4*B* shows the average contact probability of a simulation with the same cohesin processivity and separation as the simulation in Fig. 4*A* but without anchors. Since the cohesin separation is on the same order as half the cohesin processivity, there are two scaling regimes. In analogy with the transition from $1 < \gamma_1 \leq 1.5$ to $\gamma_2 \approx 0.75$ seen in human Hi-C, Fig. 4*B* exhibits a cross-over from $\gamma_1 \approx 1.1$ to $\gamma_2 \approx 0.5$ at approximately 50 kbp (see red and yellow lines). The final decay for $s > 300$ kbp is due to chain end effects. The initial decay of contact probability with $\gamma_1 \approx 1.1$ is slower than the decay of contact probability with $\gamma_1 \approx 1.4$ for a simulation of an ideal polymer without loop extrusion (see gray circles and purple line in Fig. 4*B*). This difference reflects the fact that the distribution of loop sizes (expected to be exponential with mean equal to processivity) formed in the process of active loop extrusion includes many loops that are much smaller than the processivity (Fig. 4*C*) that increase contact probability *p(s)* for smaller *s*.

Full Extrusion Probability. The one-dimensional component of our model predicts that the full extrusion probability *P(N)* with unimpeded cohesin traversal, or the probability that two CTCF anchors are held together by a cohesin, is a nonmonotonic function of the genomic separation between the anchors N given by

$$P(N) \approx 1 - \left(\frac{1 + e^{-\mu}\left(1 - \frac{1}{\alpha}\left(e^{-\alpha} - e^{-2\alpha}\right)\right)}{1 + e^{-\mu}} \right)^N, \qquad [4]$$

where $\mu k_B T > 0$ is the chemical potential per loci pair and α is the ratio between TAD length and cohesin processivity defined in Eq. **2**. See *SI Appendix* for the derivation of this result.

If N is much smaller than cohesin processivity ($\alpha \ll 1$), most bound cohesins will extrude the full TAD because the average cohesin residence time is much longer than the minimum time needed for full extrusion. In the small N limit and assuming cohesin separation much larger than TAD length $e^{\mu} \gg N$,

$$P(N) \approx Ne^{-\mu} = \alpha\beta, \qquad [5]$$

If N is much larger than processivity ($\alpha \gg 1$), the average time it takes for a single cohesin complex to extrude the full TAD is much longer than the average residence time and most cohesins unbind before reaching both TAD anchors. This results in full extrusion probability exponentially decreasing with N, with

$$P(N) \approx 2p\nu\tau_r e^{-\mu - \frac{N}{2p\nu\tau_r}} = \beta e^{-\alpha}, \qquad [6]$$

in the large N limit, expanding Eq. **4** to the first order in $e^{-\mu}$. With constant cohesin processivity, TADs with longer length N (and thus increasing α) have smaller full extrusion probabilities. With constant TAD length N, decreasing cohesin separation (increasing the nesting parameter β) increases full extrusion probability, just like the small N limit above.

Furthermore, our model suggests that the TAD length which maximizes the full extrusion probability is approximately 60% of the cohesin processivity. Although there is no simple analytical form of the maximum of Eq. **4**, we approximate its location by equating the two limiting behaviors and find that $\alpha_{opt} e^{\alpha_{opt}} = 1$. The α_{opt} that satisfies this condition is given by $\alpha_{opt} = W(1) \approx 0.6$, where *W()* the Lambert W function (82).

Eq. **4** agrees with full extrusion probabilities from MC simulations (*SI Appendix*, Fig. S7). Impeded cohesin traversal produces the same functional form for the full extrusion probability as Eq. **4**,

Downloaded from https://www.pnas.org by DUKE UNIVERSITY LIBRARIES ACQUISITIONS-PERIODICAL on June 13, 2023 from IP address 152.16.191.125.

with renormalized cohesin translocation probability p and processivity λ (*SI Appendix*). We expect the full extrusion probability when only one cohesin domain is allowed per locus to also follow the same functional form as Eq. **4** (*SI Appendix*).

The nonmonotonic behavior of full extrusion probabilities is consistent with ChIA-PET data. Since ChIA-PET measures the frequency of contacts between loci bound by specific proteins (11), ChIA-PET targeted toward CTCF can serve as a proxy for full extrusion probabilities $P(N)$ (33, 83). We fit the prediction of our model for full extrusion probability (Eq. **4**) to ChIA-PET counts targeted toward CTCF in GM12878 cells genome-wide, which we denote by I (84). We assume that experimental noise of ChIA-PET counts between CTCF anchors is larger than background contacts and can be simply modeled by an additive constant (*SI Appendix*, Fig. S12). We stress that this assumption is made only when considering contacts between two CTCF anchors; these contacts are dominated by conformations in which cohesin holds the two anchors together, compared to background contacts when the anchors are not directly held together (*SI Appendix*, Fig. S12). In the following section, we do not make this assumption when considering nonanchor loci.

This fit yields an effective cohesin processivity of (190 ± 10) kbp and a chemical potential of about (5.1 ± 0.3) k_BT per pair of 1-kbp loci, which translates to a cohesin separation of approximately (160 ± 40) kbp (Fig. 5*A*). Predictions for cohesin processivity and separation are consistent with previous experimental observations and modeling predictions (16, 25, 47, 48, 63–66). The fitted values of cohesin processivity and separation suggest that full extrusion probabilities for TADs longer than 400 kbp are approximately 10% or smaller; for example, the full extrusion probability for a 505-kbp TAD is approximately 7.5%, which is consistent with recent experimental studies (85). Furthermore, we simulate a chromatin region in chromosome 8 in GM12878 cells with two consecutive TADs, using processivity and separation within one standard deviation of the fitted values (185 kbp and 150 kbp, respectively). The simulated processivity and separation vary from the predicted values because the simulated cohesin bonds break more frequently than e^{-b} per τ_{MC} and form less frequently than $e^{-(b+\mu)}$ per τ_{MC} in MD–MC simulations if the bonds are overstretched (*SI Appendix*). We find good agreement between the simulated and experimental Hi-C maps (8), with a stratum-adjusted correlation coefficient (86) of 0.87 (Fig. 5*B*).

In contrast, a simple TF binding model cannot explain the nonmonotonic behavior of full extrusion probability, as attraction between CTCF anchors does not increase with the genomic length for shorter TADs and the full extrusion probability monotonically decreases with genomic length (*SI Appendix*). While a complicated potential between CTCF anchors could in principle create a nonmonotonic dependence, such interaction potential is hard to justify. For completeness, we fit the decreasing portion of the ChIA-PET intensity to a TF binding model (*SI Appendix*, Fig. S13).

TAD Edge Contact Probabilities. The three-dimensional component of our model is needed to calculate contact probabilities other than those between two anchors. To demonstrate this, we calculate TAD edge contact probabilities (Fig. 6). TAD edge contacts represent contacts between a CTCF anchor and other loci within the same TAD and are generalizations of TAD "stripes", which have been associated with gene regulation via superenhancers (53). Consider a TAD of length N with anchors at loci i and $i+N$. Edge contacts are those between the TAD anchor locus i and a locus $i+s$, where s is the separation between i and $i+s$ for $0<s<N$. We classify chromatin conformations into four cases (Fig. 7), whose relative contributions are dictated by cohesin processivity and separation:

a) Open contribution: The two anchors i and $i+N$ are not bound by the same cohesin.
b) Closed contribution: The two anchors i and $i+N$ are bound by the same cohesin.
c) Intermediate contribution 1: One cohesin is bound to locus i and locus $i+w$, where $0<w\leq s<N$. There is no cohesin bound to both TAD anchor loci i and $i+N$.
d) Intermediate contribution 2: One cohesin is bound to locus i and locus $i+w$, where $0<s<w<N$. There is no cohesin bound to both TAD anchor loci i and $i+N$.

We assume that cohesin binding-unbinding and one-dimensional active loop extrusion kinetics are independent of chromatin conformation. The open contribution accounts for the loopy structure created throughout the entire genomic region (not just the TAD of interest) by loop extrusion of many cohesins. As long as the average time a cohesin domain spends on a TAD anchor is longer than the average time spent on a nonanchor locus ($\tau_r = \frac{1}{\nu}e^b > \frac{1}{p\nu}$), we distinguish the closed and intermediate contributions because

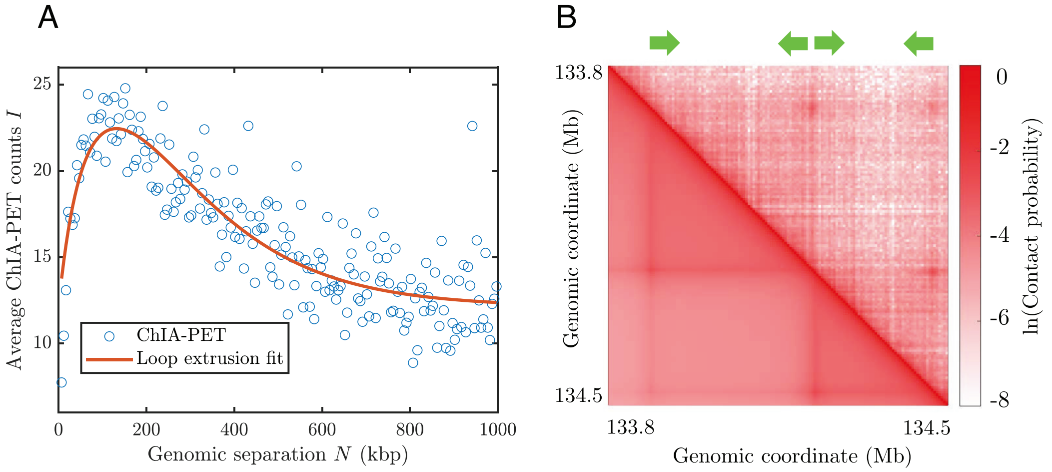

Fig. 5. Comparison of the active extrusion model with experimental ChIA-PET and Hi-C data. (*A*) ChIA-PET data [blue circles, (84)] are fit by the active loop extrusion model represented by Eq. **4** with additional multiplicative and additive constants (red solid line, see *SI Appendix* for fitting details) (RMSE = 2.0). (*B*) Comparison of Hi-C contact maps between simulation (lower triangular portion) and experimental data (upper triangular portion) (8). Experimental contact probabilities were normalized such that the average contact probability at 50 kbp was the same as in simulation.

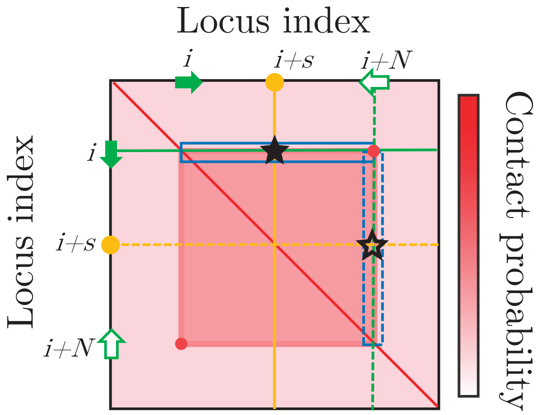

Fig. 6. Example TAD edges on a schematic Hi-C map. The TAD is between loci *i* and *i*+*N* (filled and unfilled green arrows respectively). For clarity, edges are shown only on the upper triangular portion of the map, as the upper and lower triangular portions are equivalent. The first TAD edge is boxed in solid blue lines and represents contacts between locus *i* and *i*+*s* for 0<*s*<*N*. An example *s* is shown with a yellow circle. The contact between *i* and *i*+*s* is the intersection of solid green and solid yellow lines (filled black star). The second TAD edge between locus *i*+*N* and *i*+*s* for 0<*s*<*N* is boxed in dashed blue lines. The contact between *i*+*N* and *i*+*s* is the intersection of the dashed green and dashed yellow lines (unfilled black star).

they involve cohesin bound to the TAD anchor locus i and another locus within the TAD. Additionally, we assume that on average, only one cohesin domain is bound to the TAD anchor locus i in intermediate cases, which is reasonable for the fitted cohesin processivity and separation above (*SI Appendix*, Fig. S5).

Recall that loop extrusion causes at least two contact probability scaling regimes with exponents γ_1 and γ_2 (*Introduction* and Fig. 4). Open conformation contact probabilities (Fig. 7*A*) are approximated by two scaling expressions with exponents γ_1 and γ_2 and with the cross-over at half the cohesin processivity:

$$p_{open}(s) = \begin{cases} As^{-\gamma_1}, & s \leq \frac{\lambda}{2} \\ A\left(\frac{\lambda}{2}\right)^{\gamma_2-\gamma_1} s^{-\gamma_2}, & s > \frac{\lambda}{2} \end{cases}. \qquad [7]$$

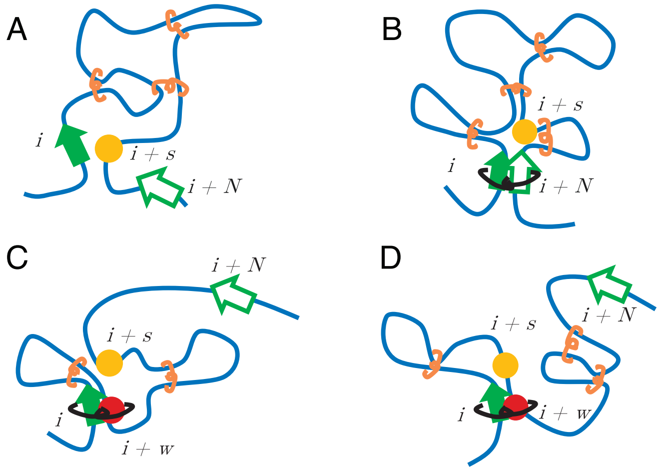

Fig. 7. Examples of conformations corresponding to different contributions to TAD edge contact probability between CTCF anchor locus *i* (filled green arrow) and locus *i* + *s* (yellow circle). Orange ovals correspond to cohesins that contribute to the fractal structure of chromatin. Unfilled green arrows correspond to the CTCF anchor at locus *i*+*N*. (*A*) Open contribution. (*B*) Closed contribution. (*C*) Intermediate contribution 1. (*D*) Intermediate contribution 2.

The prefactor for $s > \frac{\lambda}{2}$ assures the continuity of the two regimes. The contact probabilities of chromatin in the closed conformations (Fig. 7*B*) can be approximated by the mean-field contact probability between two loci s monomers apart with contact probability scaling exponent γ_1, within a section of N monomers with ends bound to each other:

$$p_{closed}(s) = A\left(s^{-\frac{\gamma_1}{3-\gamma_1}} + (N-s)^{-\frac{\gamma_1}{3-\gamma_1}}\right)^{3-\gamma_1}, \qquad [\mathbf{8}]$$

where A is the same constant as in Eq. **7** (13). If cohesin processivity, separation, and N are all on the same order (as is the case for a typical TAD), we use γ_1 to describe the closed conformations because the TAD is not long enough to observe the γ_2 regime in the closed conformations. Since the TAD anchors are held together in the closed conformations, the contact probability is nonmonotonic and the effective scaling exponent extends for only half of the TAD length (the first term in Eq. **8**). The second half of the TAD length corresponds to the increasing part of the contact probability curve, represented by the second term in Eq. **8**. This limits the applicability of the power law decay to smaller values of s and eliminates the second scaling regime. Note that TADs much longer than processivity and separation may exhibit multiple γ exponents even in closed conformation (*SI Appendix*).

The contact probabilities from intermediate conformation 1 (Fig. 7*C*) are approximated by Eq. **7** with s replaced by $(s - w + 1)$:

$$p_{int,1}(s-w+1) = \begin{cases} A(s-w+1)^{-\gamma_1}, & s-w+1 \leq \frac{\lambda}{2} \\ A\left(\frac{\lambda}{2}\right)^{\gamma_2-\gamma_1}(s-w+1)^{-\gamma_2}, & s-w+1 > \frac{\lambda}{2} \end{cases}. \qquad [\mathbf{9}]$$

Contact probabilities from intermediate conformation 2 (Fig. 7*D*) are approximated by Eq. **8** with N replaced by w:

$$p_{int,2}(s) = A\left(s^{-\frac{\gamma_1}{3-\gamma_1}} + (w-s)^{-\frac{\gamma_1}{3-\gamma_1}}\right)^{3-\gamma_1}. \qquad [\mathbf{10}]$$

We use γ_1 only to describe intermediate conformation 2 if cohesin processivity, separation, and N are all on the same order because loci i and s are within a loop as with the closed contribution (in this case, bounded by loci i and i+w rather than by i and i+N). At small s, all contributions have the same behavior, so the TAD edge contact probability p_{edge} decays as a power law $p_{edge} \sim s^{-\gamma_1}$. For large s, the closed contribution (Fig. 7*D*) dominates and causes p_{edge} to increase with s as $p_{edge} \sim (N-s)^{-\gamma_1}$. Our theory agrees with MD–MC simulations (*SI Appendix*, Eqs. **S10–S18** and Fig. S10). These MD–MC simulations are used to validate our theory rather than explicitly model a specific chromatin region. Our theory applies to any γ exponents that could be due to extrusion, even with the effect of nuclear confinement (*Introduction*).

Using the cohesin processivity and separation fitted from genome-wide ChIA-PET data (190 kbp and 160 kbp respectively), we predict the shape of TAD edge contact probabilities for four TADs in GM12878 cells (Fig. 8) (8). We assume that the average processivity and separation fitted across the genome are representative of any TAD in GM12878 cells, though we recognize that slight variations may occur. We average the two edges per TAD (see solid and dashed blue boxes in Fig. 6), assuming them to be approximately equal. To calculate the TAD edge contact probabilities, we use $\gamma_1 = 1.1$ and $\gamma_2 = 0.76$ for genomic separations shorter and longer than 95 kbp, respectively, as fitted from Hi-C

Downloaded from https://www.pnas.org by DUKE UNIVERSITY LIBRARIES ACQUISITIONS-PERIODICAL on June 13, 2023 from IP address 152.16.191.125.

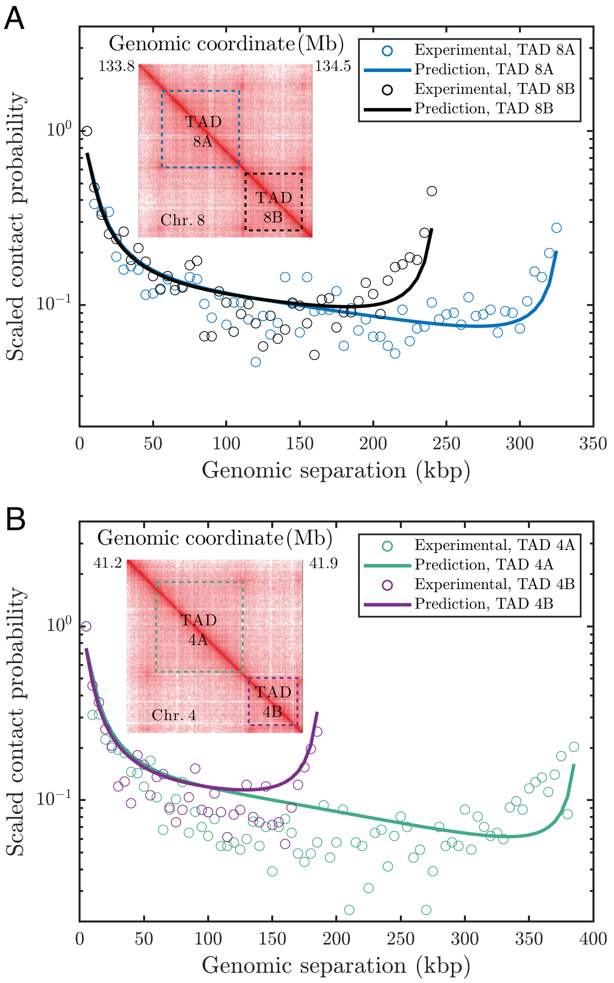

Fig. 8. TAD edge contact probabilities. Comparison of edge contact probabilities from GM12878 Hi-C (*Insets*, edges highlighted by dashed lines) between experimental Hi-C (open circles) (8) and theoretical predictions (solid lines) for (*A*) two TADs in chromosome 8 (blue and black, same region as in Fig. 5*B*) and (*B*) two TADs in chromosome 4 (green and purple). Experimental data were scaled such that the contact probability at 5 kbp equals unity. All four TAD edge contact probabilities were fit simultaneously to the theory in *SI Appendix*, Eq. **S18**. The only fitting parameter was the coefficient *A* from Eqs. **7–10**, with $A = 6.5 \pm 0.2$ used for all curves (RMSE = 0.05).

data averaged over a 10-Mbp interval of chromosome 8 with a genomic coordinate interval of 130 Mbp–140 Mbp (*Introduction*, Fig. 1*B*, and *SI Appendix*, Fig. S14). Our theory shows very good agreement with experimental data.

Qualitative Comparisons with Hi-C Experiments. Our analytical theory is also consistent with experimental studies that degrade or knock down the wings apart-like protein homolog (WAPL). WAPL is known to aid in cohesin unbinding from chromatin; cohesin can still bind to chromatin in WAPL-depleted cells, but the residence time increases (87, 88). Our theory predicts that full extrusion probability for long TAD lengths increases as cohesin residence time increases. For TAD lengths N much longer than cohesin processivity, doubling the cohesin residence time increases the full extrusion probability by $\approx 2e^{\frac{Nz}{2\lambda}} = 2e^{\frac{N}{4p\nu\tau_r}}$ where Nz is the TAD length in kbp and τ_r is the original cohesin residence time (see Eq. **6**). Longer residence times slow the decay of full extrusion probability for long TAD lengths, as cohesin complexes are less likely to unbind from chromatin before reaching the CTCF anchors. Hi-C experiments after WAPL degradation or knockdown show higher intensities of TAD corners associated with CTCF anchors for long TAD lengths compared to wild-type cells, suggesting that longer cohesin residence times are associated with increased frequencies of CTCF anchor contacts (89–93). Several of these studies also suggest that multiple cohesin complexes can accumulate and stack at CTCF anchors to stabilize loops (92, 93). This is consistent with our model, which allows multiple cohesins to bind to the same pair of anchors. For example, consider two cohesins bound to the same anchors. The two cohesins unbind independently. When one cohesin unbinds, the other cohesin may still hold the anchors together, thus stabilizing the loop. Increasing the cohesin residence time (by knocking down WAPL, for example) would enhance this effect.

Additionally, our theory predicts that increasing cohesin separation (or decreasing the number of bound cohesins per TAD) should decrease full extrusion probabilities. Eq. **4** suggests that cohesin separation and full extrusion probability are inversely proportional. Biologically, decreasing the number of bound cohesins per TAD could be achieved by knocking down SCC4, a protein that facilitates cohesin–chromatin binding (94). This agrees with the decrease in TAD corner intensities seen in Hi-C maps of cells deficient in SCC4 (89). In these experiments, SCC4-deficiency caused fewer cohesins to bind to chromatin but did not affect cohesin residence time.

Summary. We develop analytical equations that describe active loop extrusion with nonspecific binding and cohesin traversal. We derive analytical expressions for contact probabilities and full extrusion probabilities. The analytical predictions of our model agree with experimental ChIA-PET data for full extrusion probabilities and Hi-C data for contact probabilities. Our equations also agree with MD and hybrid MC-MD simulations of the full active extrusion model. We expect extrusion with impeded cohesin traversal to follow the same analytical equations but with a renormalized cohesin processivity (*SI Appendix*). Our theory suggests that full extrusion probabilities should increase with cohesin residence time but decrease with cohesin separation. These predictions are consistent with Hi-C experiments performed on WAPL- and SCC4-deficient cells.

We show that active loop extrusion with nonspecific, passive binding and cohesin traversal produces at least two contact probability scaling regimes, with the cross-over between regimes occurring on the order of half the cohesin processivity. We also predict a nonmonotonic dependence of full extrusion probability on TAD length with a maximum at a TAD length at about 60% the cohesin processivity, which is consistent with experimental ChIA-PET data targeted to CTCF in GM12878 cells (Fig. 5*A*). Nonmonotonicity arises because our model assumes many cohesin-binding sites per TAD and that the number of cohesin-binding sites increases with TAD length. Our analysis of ChIA-PET data from GM12878 cells predicts that average cohesin processivity is (190 ± 10) kbp and separation is (160 ± 40) kbp, both of which are within ranges of previous modeling and experimental studies (16, 25, 47, 48, 63–66). These parameters suggest that in GM12878 cells, the optimum TAD length to maximize full extrusion probability is approximately 110 kbp. Furthermore, we use our analytical equations and the parameters obtained from the analysis of ChIA-PET data to calculate the expected edge contact probabilities and found our predictions in good agreement with Hi-C data of edge probabilities for four TADs in GM12878 cells (Fig. 8). Our findings may be used to predict and modify the probability of contacts between specific chromatin loci of interest based on cohesin characteristics.

Downloaded from https://www.pnas.org by DUKE UNIVERSITY LIBRARIES ACQUISITIONS-PERIODICAL on June 13, 2023 from IP address 152.16.191.125.

Materials and Methods

Simulations. MC simulations were used to study cohesin binding, unbinding, and translocation in loci index space. Each MC simulation considered one genomic region of length $2N$ with one pair of CTCF anchors located at loci $i=N/2$ and $i+N=3N/2$, with TAD length N ranging from 100 to 2,000 loci. MC simulations were performed in MATLAB. Hybrid MD–MC simulations were used to explore the three-dimensional component of our model and study contact probabilities produced by loop extrusion. MD was performed using the Large-scale Atomic/Molecular Massively Parallel Simulator package (95). We use the Kremer–Grest bead-spring model to simulate ideal linear polymer chains without excluded volume where bonded beads are connected by a finite extensible nonlinear elastic (FENE) potential. Nonbonded beads do not interact unless held together by a cohesin. Each MD–MC simulation contains a single chain of $N_{beads} = 400$ beads. Chains are initiated as random walks and equilibrated for approximately 10 times their relaxation times using MD before starting the MD–MC simulations (*SI Appendix*). The MC procedure described in *SI Appendix* was implemented in C and coupled to MD to simulate loop extrusion. MD was used to follow the spatial trajectories of each bead, and MC was used to determine extrusion complex binding, unbinding, and translocation. Cohesin is represented by an additional FENE bond potential between a pair of beads. See *SI Appendix* for extended methods. Note that the ideal linear chain MD–MC simulations are used primarily to validate our derived analytical equations on a system with fractal dimension $D = 2$. Our theory would apply to other polymer models of chromatin with different fractal dimensions as well.

Data, Materials, and Software Availability. All study data are included in the article and/or *SI Appendix*.

ACKNOWLEDGMENTS. M.R. acknowledges funding support by the NSF under grant number EFMA-1830957 and the NIH under grant numbers P01-HL164320 and 5P01-HL108808.

Author affiliations: [a]Department of Biomedical Engineering, Duke University, Durham, NC 27708; [b]Department of Mechanical Engineering and Materials Science, Duke University, Durham, NC 27708; [c]Department of Chemistry, Duke University, Durham, NC 27708; [d]Department of Physics, Duke University, Durham, NC 27708; and [e]Institute for Chemical Reaction Design and Discovery (World Premier International Research Center Initiative-ICReDD), Hokkaido University, Sapporo 001-0021, Japan

1. A. Piovesan, M. Caracausi, F. Antonaros, M. C. Pelleri, L. Vitale, GeneBase 1.1: A tool to summarize data from NCBI gene datasets and its application to an update of human gene statistics. *Database 2016, baw153 (2016).*
2. T. Cremer, C. Cremer, Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat. Rev. Genet.* **2**, 292–301 (2001).
3. R. Phillips *et al.*, *Physical Biology of the Cell* (Garland Science, 2012), 10.1201/9781134111589.
4. K. Luger, J. C. Hansen, Nucleosome and chromatin fiber dynamics. *Curr. Opin. Struct. Biol.* **15**, 188–196 (2005).
5. J. C. Hansen *et al.*, The 10-nm chromatin fiber and its relationship to interphase chromosome organization. *Biochem. Soc. Trans.* **46**, 67–76 (2018).
6. R. K. Sachs, G. Van Den Engh, B. Trask, H. Yokota, J. E. Hearst, A random-walk/giant-loop model for interphase chromosomes. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 2710–2714 (1995).
7. A. Rosa, R. Everaers, Structure and dynamics of interphase chromosomes. *PLoS Comput. Biol.* **4**, 1000153 (2008).
8. S. S. P. Rao *et al.*, A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–80 (2014).
9. K. Bystricky, P. Heun, L. Gehlen, J. Langowski, S. M. Gasser, Long-range compaction and flexibility of interphase chromatin in budding yeast analyzed by high-resolution imaging techniques. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 16495–16500 (2004).
10. J. Mateos-Langerak *et al.*, Spatially confined folding of chromatin in the interphase nucleus. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 3812–3817 (2009).
11. M. J. Fullwood, Y. Ruan, ChIP-based methods for the identification of long-range chromatin interactions. *J. Cell. Biochem.* **107**, 30–39 (2009).
12. E. Lieberman-Aiden *et al.*, Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
13. M. Rubinstein, R. Colby, *Polymer Physics* (Oxford University Press, ed. 1, 2003).
14. J. D. Halverson, J. Smrek, K. Kremer, A. Y. Grosberg, From a melt of rings to chromosome territories: The role of topological constraints in genome folding. *Rep. Prog. Phys.* **77**, 022601 (2014).
15. D. Jost, P. Carrivain, G. Cavalli, C. Vaillant, Modeling epigenome folding: Formation and dynamics of topologically associated chromatin domains. *Nucleic Acids Res.* **42**, 9553–9561 (2014).
16. G. Fudenberg *et al.*, Formation of chromosomal domains by loop extrusion. *Cell Rep.* **15**, 2038–2049 (2016).
17. W. Antonin, H. Neumann, Chromosome condensation and decondensation during mitosis. *Curr. Opin. Cell Biol.* **40**, 15–22 (2016).
18. F. Tavares-Cadete, D. Norouzi, B. Dekker, Y. Liu, J. Dekker, Multi-contact 3C reveals that the human genome during interphase is largely not entangled. *Nat. Struct. Mol. Biol.* **27**, 1105–1114 (2020).
19. D. Goundaroulis, E. Lieberman Aiden, A. Stasiak, Chromatin is frequently unknotted at the megabase scale. *Biophys. J.* **118**, 2268–2279 (2020).
20. S. Wang *et al.*, Spatial organization of chromatin domains and compartments in single chromosomes. *Science* **353**, 598–602 (2016).
21. Y. Takei *et al.*, Integrated spatial genomics reveals global architecture of single nuclei. *Nature* **590**, 344–350 (2021).
22. M. Cremer *et al.*, Cohesin depleted cells rebuild functional nuclear compartments after endomitosis. *Nat. Commun.* **11**, 1–16 (2020).
23. 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).
24. W. Schwarzer *et al.*, Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51–56 (2017).
25. S. S. P. Rao *et al.*, Cohesin loss eliminates all loop domains. *Cell* **171**, 305–320.e24 (2017).
26. T. Ge, S. Panyukov, M. Rubinstein, Self-similar conformations and dynamics in entangled melts and solutions of nonconcatenated ring polymers. *Macromolecules* **49**, 708–722 (2016).
27. A. Rosa, R. Everaers, Conformational statistics of randomly branching double-folded ring polymers. *Eur. Phys. J. E Soft Matter* **42**, 7 (2019).
28. P. G. de Gennes, Statistics of branching and hairpin helices for the dAT copolymer. *Biopolymers* **6**, 715–729 (1968).
29. J. R. Dixon *et al.*, Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
30. E. Alipour, J. F. Marko, Self-organization of domain structures by DNA-loop-extruding enzymes. *Nucleic Acids Res.* **40**, 11202–11212 (2012).
31. E. J. Banigan, A. A. van den Berg, H. B. Brandão, J. F. Marko, L. A. Mirny, Chromosome organization by one-sided and two-sided loop extrusion. *Elife* **9**, e53558 (2020).
32. T. Yamamoto, H. Schiessel, Osmotic mechanism of the loop extrusion process. *Phys. Rev. E* **96**, 03042 (2017).
33. C. A. Brackley *et al.*, Nonequilibrium chromosome looping via molecular slip links. *Phys. Rev. Lett.* **119**, 138101 (2017).
34. A. Bonato, C. A. Brackley, J. Johnson, D. Michieletto, D. Marenduzzo, Chromosome compaction and chromatin stiffness enhance diffusive loop extrusion by slip-link proteins. *Soft Matter* **16**, 2406–2414 (2020).
35. C. A. Brackley *et al.*, Extrusion without a motor: A new take on the loop extrusion model of genome organization. *Nucleus* **9**, 95–103 (2018).
36. K. Nasmyth, Disseminating the genome: Joining, resolving, and separating sister chromatids during mitosis and meiosis. *Annu. Rev. Genet.* **35**, 673–745 (2001).
37. S. Schoenfelder, P. Fraser, Long-range enhancer–promoter contacts in gene expression control. *Nat. Rev. Genet.* **20**, 437–455 (2019).
38. C. A. Brackley *et al.*, Simulated binding of transcription factors to active and inactive regions folds human chromosomes into loops, rosettes and topological domains. *Nucleic Acids Res.* **44**, 3503–3512 (2016).
39. C. Annunziatella *et al.*, Molecular dynamics simulations of the strings and binders switch model of chromatin. *Methods* **142**, 81–88 (2018).
40. A. S. Weintraub *et al.*, YY1 is a structural regulator of enhancer-promoter loops. *Cell* **171**, 1573–1588.e28 (2017).
41. G. Li *et al.*, Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148**, 84–98 (2012).
42. M. Barbieri *et al.*, Complexity of chromatin folding is captured by the strings and binders switch model. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 16173–16178 (2012).
43. N. Mayerova, L. Cipak, J. Gregan, Cohesin biology: From passive rings to molecular motors. *Trends Genet.* **36**, 387–389 (2020).
44. F. Uhlmann, SMC complexes: From DNA to chromosomes. *Nat. Rev. Mol. Cell Biol.* **17**, 399–412 (2016).
45. A. S. Hansen, CTCF as a boundary factor for cohesin-mediated loop extrusion: Evidence for a multi-step mechanism. *Nucleus.* **11**, 132–148 (2020).
46. E. P. Nora *et al.*, Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* **169**, 930–944.e22 (2017).
47. Y. Kim, Z. Shi, H. Zhang, I. J. Finkelstein, H. Yu, Human cohesin compacts DNA by loop extrusion. *Science* **366**, 1345–1349 (2019).
48. I. F. Davidson *et al.*, DNA loop extrusion by human cohesin. *Science* **366**, 1338–1345 (2019).
49. I. F. Davidson, J.-M. Peters, Genome folding through loop extrusion by SMC complexes. *Nat. Rev. Mol. Cell Biol.* **22**, 445 (2021).
50. E. Kim, J. Kerssemakers, I. A. Shaltiel, C. H. Haering, C. Dekker, DNA-loop extruding condensin complexes can traverse one another. *Nature* **579**, 438–442 (2020).
51. H. B. Brandão, Z. Ren, X. Karaboja, L. A. Mirny, X. Wang, DNA-loop-extruding SMC complexes can traverse one another in vivo. *Nat. Struct. Mol. Biol.* **28**, 642–651 (2021).
52. B. Pradhan *et al.*, SMC complexes can traverse physical roadblocks bigger than their ring size. *Cell Rep.* **41**, 111491 (2022).
53. L. Vian *et al.*, The energetics and physiological impact of cohesin extrusion. *Cell* **173**, 1165–1178.e20 (2018).
54. B. W. Bauer *et al.*, Cohesin mediates DNA loop extrusion by a "swing and clamp" mechanism. *Cell* **184**, 5448–5464.e22 (2021).
55. 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).
56. K. Kumari, B. Duenweg, R. Padinhateeri, J. R. Prakash, Computing 3D chromatin configurations from contact probability maps by inverse brownian dynamics. *Biophys. J.* **118**, 2193–2208 (2020).
57. D. Meluzzi, G. Arya, Recovering ensembles of chromatin conformations from contact probabilities. *Nucleic Acids Res.* **41**, 63–75 (2013).
58. M. Socol *et al.*, Rouse model with transient intramolecular contacts on a timescale of seconds recapitulates folding and fluctuation of yeast chromosomes. *Nucleic Acids Res.* **47**, 6195–6207 (2019).

Downloaded from https://www.pnas.org by DUKE UNIVERSITY LIBRARIES ACQUISITIONS-PERIODICAL on June 13, 2023 from IP address 152.16.191.125.

59. A. Buckle, C. A. Brackley, S. Boyle, D. Marenduzzo, N. Gilbert, Polymer simulations of heteromorphic chromatin predict the 3D folding of complex genomic loci. *Mol. Cell* **72**, 786–797.e11 (2018).
60. M. Di Pierro, 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).
61. M. Di Pierro, 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).
62. B. Zhang, P. G. Wolynes, Topology, structures, and energy landscapes of human chromosomes. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 6062–6067 (2015).
63. J. Gassler *et al.*, A mechanism of cohesin-dependent loop extrusion organizes zygotic genome architecture. *EMBO J.* **36**, 3600–3618 (2017).
64. A. S. Hansen, I. Pustova, C. Cattoglio, R. Tjian, X. Darzacq, CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *Elife* **6**, e25776 (2017).
65. C. Cattoglio *et al.*, Determining cellular CTCF and cohesin abundances to constrain 3D genome models. *Elife* **8**, e40164 (2019).
66. J. Holzmann *et al.*, Absolute quantification of cohesin, CTCF and their regulators in human cells. *Elife* **8**, e46269 (2019).
67. C. A. Brackley, S. Taylor, A. Papantonis, P. R. Cook, D. Marenduzzo, Nonspecific bridging-induced attraction drives clustering of DNA-binding proteins and genome organization. *Proc. Natl. Acad. Sci. U.S.A.* **110**, E3605–E3611 (2013).
68. C. A. Brackley, Polymer compaction and bridging-induced clustering of protein-inspired patchy particles. *J. Phys. Condens. Matter* **32**, 11 (2020).
69. J. Fraser, I. Williamson, W. A. Bickmore, J. Dostie, An overview of genome organization and how we got there: From FISH to Hi-C. *Microbiol. Mol. Biol. Rev.* **79**, 347–372 (2015).
70. N. Hao, K. E. Shearwin, I. B. Dodd, Positive and negative control of enhancer-promoter interactions by other DNA loops generates specificity and tunability. *Cell Rep.* **26**, 2419–2433.e3 (2019).
71. B. Doyle, G. Fudenberg, M. Imakaev, L. A. Mirny, Chromatin loops as allosteric modulators of enhancer-promoter interactions. *PLoS Comput. Biol.* **10**, 1003867 (2014).
72. S. S. Gisselbrecht *et al.*, Transcriptional silencers in Drosophila serve a dual role as transcriptional enhancers in alternate cellular contexts. *Mol. Cell* **77**, 324–337.e8 (2020).
73. Y. Zhan *et al.*, Reciprocal insulation analysis of Hi-C data shows that TADs represent a functionally but not structurally privileged scale in the hierarchical folding of chromosomes. *Genome Res.* **27**, 479–490 (2017).
74. L. R. Polstein *et al.*, Genome-wide specificity of DNA binding, gene regulation, and chromatin remodeling by TALE- and CRISPR/Cas9-based transcriptional activators. *Genome Res.* **25**, 1158–1169 (2015).
75. S. L. Morgan *et al.*, Manipulation of nuclear architecture through CRISPR-mediated chromosomal looping. *Nat. Commun.* **8**, 15993 (2017).
76. A. M. D'Ippolito *et al.*, Pre-established chromatin interactions mediate the genomic response to glucocorticoids. *Cell Syst.* **7**, 146–160.e7 (2018).
77. E. J. Banigan, L. A. Mirny, Limits of chromosome compaction by loop-extruding motors. *Phys. Rev. X* **9**, 031007 (2019).
78. R. Takaki, A. Dey, G. Shi, D. Thirumalai, Theory and simulations of condensin mediated loop extrusion in DNA. *Nat. Commun.* **12**, 1–10 (2021).
79. M. Ganji *et al.*, Real-time imaging of DNA loop extrusion by condensin. *Science* **360**, 102–105 (2018).
80. H. B. Brandão *et al.*, RNA polymerases as moving barriers to condensin loop extrusion. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 20489–20499 (2019).
81. Y. Guo *et al.*, Chromatin jets define the properties of cohesin-driven in vivo loop extrusion. *Mol. Cell* **82**, 3769–3780.e5 (2022).
82. R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, D. E. Knuth, On the lambert W function. *Adv. Comput. Math.* **5**, 329–359 (1996).
83. W. Xi, M. A. Beer, Loop competition and extrusion model predicts CTCF interaction specificity. *Nat. Commun.* **12**, 1046 (2021).
84. Z. Tang *et al.*, CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription. *Cell* **163**, 1611–1627 (2015).
85. M. Gabriele *et al.*, Dynamics of CTCF- and cohesin-mediated chromatin looping revealed by live-cell imaging. *Science* **376**, 476–501 (2022).
86. T. Yang *et al.*, HiCRep: Assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. *Genome Res.* **27**, 1939–1949 (2017).
87. S. Kueng *et al.*, Wapl controls the dynamic association of cohesin with chromatin. *Cell* **127**, 955–967 (2006).
88. A. Tedeschi *et al.*, Wapl is an essential regulator of chromatin structure and chromosome segregation. *Nature* **501**, 564–568 (2013).
89. J. H. I. Haarhuis *et al.*, The cohesin release factor WAPL restricts chromatin loop extension. *Cell* **169**, 693–707.e14 (2017).
90. G. Wutz *et al.*, Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *EMBO J.* **36**, 3573–3599 (2017).
91. L. Costantino, T. H. S. Hsieh, R. Lamothe, X. Darzacq, D. Koshland, Cohesin residency determines chromatin loop patterns. *Elife* **9**, 1–31 (2020).
92. E. J. Banigan *et al.*, Transcription shapes 3D chromatin organization by interacting with loop-extruding cohesin complexes. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2210480120 (2023).
93. N. Q. Liu *et al.*, WAPL maintains a cohesin loading cycle to preserve cell-type-specific distal gene regulation. *Nat. Genet.* **53**, 100–109 (2021).
94. R. Ciosk *et al.*, Cohesin's binding to chromosomes depends on a separate complex consisting of Scc2 and Scc4 proteins. *Mol. Cell* **5**, 243–254 (2000).
95. S. Plimpton, Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **117**, 1–42 (1997).

Downloaded from https://www.pnas.org by DUKE UNIVERSITY LIBRARIES ACQUISITIONS-PERIODICAL on June 13, 2023 from IP address 152.16.191.125.