Universal contact statistics of looped polymers resolve cohesin density and stoichiometry in vivo

Kirill Polovnikov[1,2*] and Dmitry Starkov[1]
[1]*Skolkovo Institute of Science and Technology, 121205 Moscow, Russia and*
[2]*Institute for Physics & Astronomy, University of Potsdam, 14476 Potsdam-Golm, Germany*
(Dated: September 4, 2025)

Loop extrusion by cohesin complexes is central to genome organization, yet the stoichiometry of extrusion *in vivo* remains unresolved. Resolving this requires a quantitative link between microscopic loop density and macroscopic Hi-C contact statistics. We develop a perturbation theory for polymers decorated with quenched loops and measured under a finite Hi-C capture radius. The theory predicts a universal scaling law for the contact probability $P(s)$: its logarithmic derivative displays a distinct minimum at genomic distance $s_{\min} \sim \sqrt{T v_0^{\text{eff}}}$, determined solely by the loop–gap period T (inverse loop density) and the effective fragment length v_0^{eff}. This two-parameter reduction is validated by molecular-dynamics simulations of active loop extrusion and recapitulates short-scale features of Hi-C data. Applying the framework to mammalian datasets, we infer a conserved density of 6 ± 1 cohesin-mediated loops per megabase. Comparison with independent cohesin counts reveals near one-to-one correspondence between loops and complexes, providing quantitative evidence for monomeric loop extrusion. Our results establish a universal physical principle for short-scale chromatin folding and show how polymer theory bridges genome architecture with the molecular mechanisms that shape it.

INTRODUCTION

The folding of mammalian chromosomes is a paradigmatic example of multiscale organization in soft matter: polymer chains of gigabase length (tens of centimeters if stretched) compact into micron-sized chromosome territories, yet retain local structures such as topologically associating domains (TADs) and loops that shape gene regulation, replication, and repair [1–3]. A central mechanism underlying this organization is loop extrusion, whereby structural maintenance of chromosomes (SMC) complexes processively reel in DNA to form dynamic loops [4–6]. Direct *in vitro* experiments have visualized both condensin and cohesin extruding loops [7–9], but the physical implementation of this mechanism *in vivo*—and in particular the stoichiometry of extrusion—remains unresolved.

Polymer physics provides the natural framework to connect microscopic extrusion to macroscopic Hi-C contact maps. The contact probability $P(s)$ between loci separated by genomic distance s displays robust scaling behavior across species: in the absence of cohesin, $P(s) \sim s^{-\gamma_c}$ with $\gamma_c \approx 1$, consistent with crumpled globules of fractal dimension $d_f \approx 3$ [10–13]. The introduction of cohesin-mediated loops disrupts this scale invariance, producing shoulders in $P(s)$ and oscillatory features in its logarithmic derivative [6, 14]—universal signatures of extrusion.

A key unknown, however, is how many cohesins anchor each loop. Biochemically, cohesin is a tripartite ring [15], capable of extruding as a monomer [8, 9]. Yet microscopy has observed dimeric or oligomeric structures [16, 17], raising the possibility of paired motors. Absolute protein counts [16, 18] cannot resolve this ambiguity without knowledge of loop density, while Hi-C itself introduces a finite capture radius [19, 20] that obscures quantitative inference.

Here we address this challenge by developing a perturbation theory for polymers decorated with quenched loops under a finite capture radius. We show that, below the extrusion scale ($s \lesssim 100$ kb), chromosome contact statistics simplify dramatically: $P(s)$ depends only on two parameters—the loop density and the effective fragment length imposed by Hi-C. This reduction yields a universal scaling law in which the logarithmic derivative of $P(s)$ exhibits a characteristic dip at $s_{\min} \sim \sqrt{T v_0^{\text{eff}}}$. This dip serves as a direct readout of loop density, independent of processivity, loop size distribution, or large-scale topology.

Applying the framework to over twenty mammalian Hi-C datasets, we infer a conserved density of 6 ± 1 cohesin-mediated loops per megabase. Independent protein counts align with this density in a near one-to-one ratio, strongly supporting monomeric extrusion. By extracting molecular stoichiometry from universal scaling features, our work demonstrates how polymer theory can bridge collective genome folding to the microscopic mechanics of its organizing motors.

CHROMOSOME FOLDING ACROSS SCALES

Chromosomes exemplify a multiscale polymer system, where distinct physical regimes emerge at different contour separations s (in base pairs). At the largest scales ($s \gtrsim 10$ Mb), chromosomes segregate into territories of a few microns, reflecting confinement of gigabase-long polymers inside the nucleus [10, 21]. At intermediate

* kipolovnikov@gmail.com

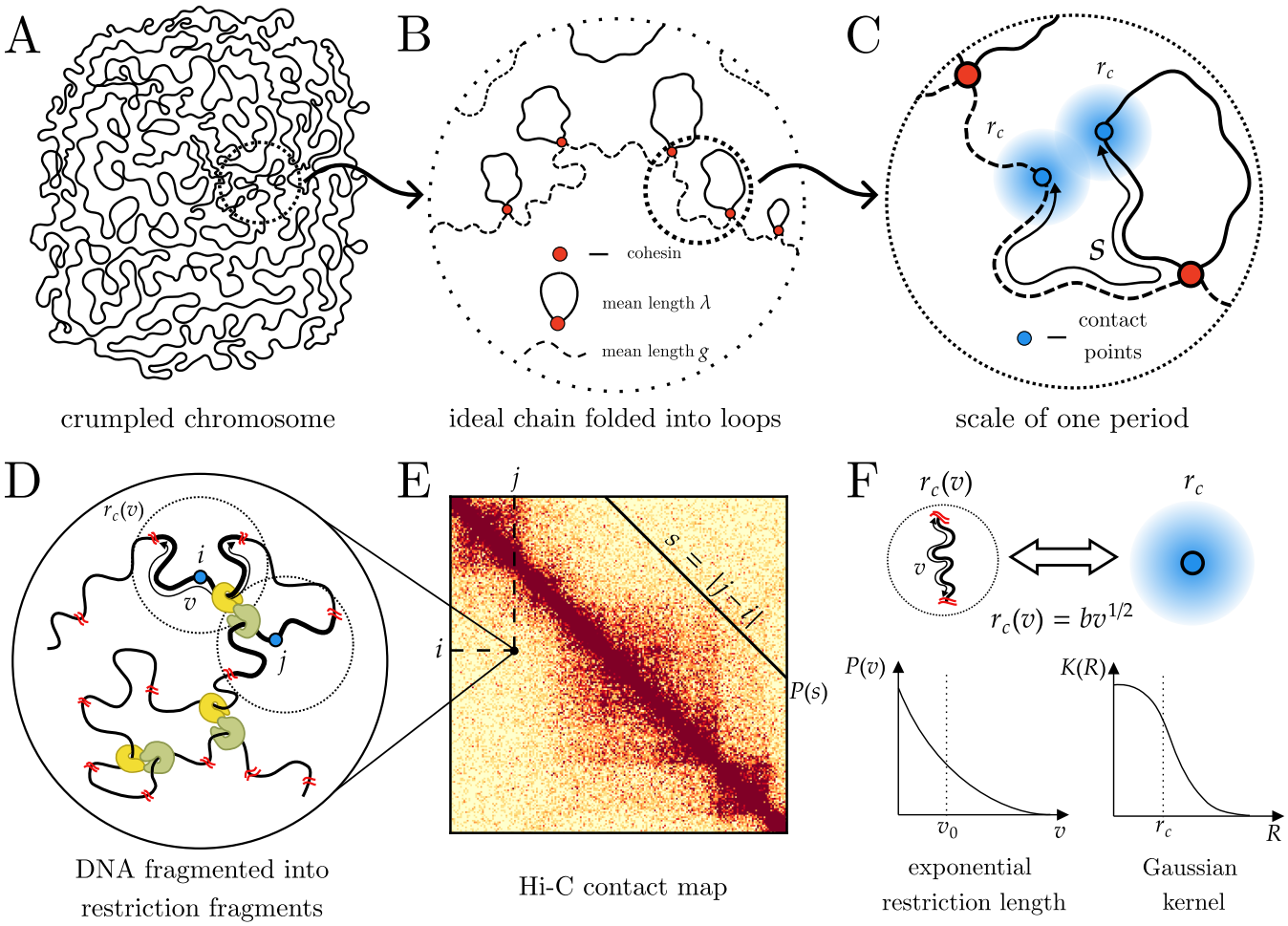

FIG. 1. From crumpled chromosomes to a minimal loop-and-cutoff model of Hi-C contacts. **(A)** A compact interphase chromosome is represented as a crumpled polymer globule. **(B)** Zooming in, the polymer backbone behaves as an ideal chain intermittently bridged by cohesin complexes (red). Loop lengths (solid) and gap lengths (dashed) are independently and exponentially distributed with mean values λ and g, respectively. **(C)** At the scale of a loop–gap period $T = \lambda + g$, Hi-C detects only pairs of loci that fall within the capture radius r_c (blue halo), so that contacts depend jointly on contour separation s and r_c. **(D)** In Hi-C, cross-linkers (yellow and green clamps) first stabilize spatially proximate segments, after which the fiber is digested at restriction sites (red wavy marks), producing fragments of random length v. The capture radius thus becomes fragment dependent, $r_c(v)$. Two example fragments at contour positions i and j are shown. **(E)** After ligation and sequencing, each captured pair (i, j) contributes to a Hi-C map. Grouping all pairs by genomic separation $s = |j - i|$ yields the contact probability $P(s)$. **(F)** Two equivalent formalisms for contact detection: (left) an exponential distribution of fragment lengths $p(v)$ combined with fragment-dependent capture radius $r_c(v)$; (right) a Gaussian detection kernel $K(R)$ with typical capture radius r_c. The two descriptions coincide when $r_c(v) \propto v^{1/2}$.

separations ($\sim$ 0.5–10 Mb), Hi-C and imaging consistently reveal scale-free statistics: the spatial size of a segment grows as $r(s) \sim s^{1/d_f}$ with fractal dimension $d_f \approx 3$, characteristic of a crumpled globule [10, 12, 22]. This crumpled regime arises from topological constraints in dense chromatin, analogous to nonconcatenated ring polymers in polymer melts.

Below the entanglement length ($N_e \approx$ 50–150 kb in dense chromatin [10, 23, 24]), topological constraints are no longer effective, and the chain recovers ideal statistics with $d_f = 2$. This "ideal-chain regime" corresponds to segments that relax freely within topological cages, in agreement with classical Rouse dynamics [25]. Thus, chromosome folding exhibits a crossover from ideal to crumpled behavior as s increases beyond N_e, paralleling well-known transitions in polymer melts and gels.

Loop extrusion introduces a new physical ingredient: a dynamically generated array of loops with mean size λ and gap g, producing a characteristic loop–gap period $T = \lambda + g$. Decorating a crumpled chain with such loops effectively dilutes entanglements, increasing the apparent entanglement length by an order of magnitude ($N_e \sim$ 1–2 Mb) [6, 14]. At the same time, loop–loop and loop–gap interactions create distinctive features in contact statis-

tics, including a robust oscillatory structure in the logarithmic derivative of the contact probability $P(s)$.

Taken together, chromosome folding can be organized into four physical regimes:

(I) **Crumpled regime** ($s > N_e$): compact statistics with $d_f \approx 3$, shaped by entanglements.

(II) **Ideal-with-loops regime** ($\lambda \lesssim s \lesssim N_e$): effectively ideal chains folded by many loops, with enlarged N_e.

(III) **Loop–gap resonance** ($s \approx \lambda, g$): oscillatory features in $P(s)$ reflecting periodic loop–gap architecture.

(IV) **Short-scale density-controlled regime** ($l_K < s < \lambda, g$): entanglements are irrelevant; folding depends only on loop density T^{-1} and the Hi-C capture radius.

In this work, we focus on Regime IV. We show that in this short-scale window, the complexity of chromosome folding reduces to just two effective parameters—loop density and capture radius—allowing us to develop a perturbation theory that quantitatively connects Hi-C contact statistics to cohesin stoichiometry.

CHROMATIN FOLDED BY COHESIN MOTORS: A FRACTAL POLYMER WITH QUENCHED LOOP DISORDER

Chromatin in the absence of cohesin can be viewed as a polymer with fractal statistics, interpolating between ideal and crumpled regimes. A polymer of fractal dimension d_f is defined by the scaling of its end-to-end distance with contour length s:

$$r(s) \sim b\, s^{1/d_f},$$

where b is the bead size. The case $d_f = 2$ corresponds to an ideal Gaussian chain, while $d_f \approx 3$ describes compact crumpled conformations stabilized by topological constraints [10, 12, 13]. Chromatin contact maps in cohesin-depleted cells are consistent with the latter [6], motivating fractal models as the baseline for genome folding.

To describe folding with cohesin, we introduce *quenched loop disorder*: additional harmonic bonds that connect non-adjacent loci and mimic cohesin clamps [6, 26]. Loops and gaps are drawn independently from exponential distributions with mean sizes λ and g, producing a disordered array of non-overlapping loops (Fig. 1B). Because cohesin grips rather than threads the fiber, loops are topologically neutral and statistically independent of the backbone [26]. This independence permits a diagrammatic decomposition of contact probabilities into gap–gap, loop–gap, and loop–loop contributions (Fig. S1), averaged over loop realizations.

At first sight, the active nature of loop extrusion might suggest that such quenched models are inappropriate. However, polymer relaxation timescales provide a natural separation. The Rouse relaxation time of a segment of length s, $\tau_R(s) \sim \tau_0(s/l_K)^2$ (with Kuhn length $l_K \sim 3$–4 kb and $\tau_0 \sim 10^{-2}$ s [6, 23, 24]), grows only quadratically with s, whereas the extrusion time $\tau_{\rm ex}(s) = s/r$ grows linearly with s (extrusion speed $r \sim 1$–2 kb/s [8, 9]). Equating the two yields a crossover scale $s^* \approx 0.5$–1 Mb: below this, relaxation is faster than extrusion, so loops can be treated as effectively equilibrated disorder.

Thus, on the short genomic scales central to Regime IV ($s < \lambda, g \sim 50$–100 kb), chromatin can be modeled as an ideal polymer ($d_f = 2$) decorated with quenched random loops. This equilibrium reduction enables an analytical perturbation theory that directly links Hi-C contact statistics to loop density.

TWO APPROACHES FOR MODELING CONTACT CAPTURE IN HI-C PROTOCOLS

To connect polymer models with Hi-C data, one must account not only for polymer geometry but also for the detection process. Hi-C does not register all close encounters between loci: it captures only those fragments cross-linked, digested, and ligated into chimeric molecules. The resulting contact map A_{ij} (Fig. 1E) is then summarized by the contact probability

$$P(s) = \frac{1}{N-s} \sum_{i=1}^{N-s} A_{i,i+s},$$

but its short-scale features are inevitably shaped by the experimental capture radius.

Capture Model A: Fragment-based description. In a mechanistic view, restriction enzymes cut DNA into fragments of random length v, drawn from an exponential distribution with mean v_0 [19]. A pair of loci contributes to $P(s)$ only if their fragments overlap. The corresponding capture radius is stochastic, with variance

$$\tilde{r}_c^2(v_i, v_j) = r^2(v_i) + r^2(v_j), \qquad r(v) = bv^{1/2},$$

where $r(v)$ is the spatial extent of a fragment. Averaging over the distribution of fragment lengths produces the observed capture statistics.

Capture Model B: Gaussian kernel. For analytical progress, it is convenient to coarse-grain the detection process into a smooth kernel. We assume that two loci separated by spatial distance R are registered with probability

$$K(R) = \exp\left(-\frac{3R^2}{2r_c^2}\right), \tag{1}$$

where r_c is an effective capture radius. This Gaussian description averages over fragment heterogeneity, cross-linking variability, and ligation efficiency, and avoids the sharp cutoffs of step-function models [27–29].

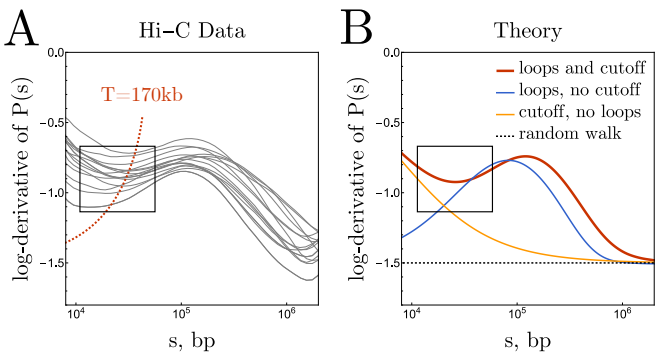

FIG. 2. Loops and capture radius are both required to reproduce the universal dip in short-scale Hi-C contact statistics. **(A)** Logarithmic derivative of the contact probability $P(s)$ for 16 mammalian Hi-C datasets [3, 19, 30]. All maps display a characteristic dip (black rectangle) at separations of tens of kilobases. The red dotted line indicates dip positions predicted for fixed loop–gap period $T = 170$ kb and varying fragment length v_0. **(B)** Polymer-model predictions. Only the combined model with quenched loops and finite capture radius (red) reproduces the observed dip. Models lacking either loops (orange) or capture radius (blue) fail to capture this feature. The dashed line marks the $-3/2$ slope of an unrestricted ideal chain. Model parameters: bead size = 1 kb, loop–gap period $T = 170$ kb, fragment length $v_0 = 5$ kb.

Equivalence and parameter mapping. The two formalisms are mathematically equivalent: averaging Model A yields a kernel of Gaussian form, with the correspondence

$$r_c^2 = \tfrac{3}{2}\, b^2 v_0. \tag{2}$$

Thus, the mean fragment length v_0 and the effective capture radius r_c are interchangeable control parameters. In practice, we adopt the Gaussian kernel (Model B) for tractability, treating v_0 as a phenomenological parameter of the Hi-C protocol.

Effect on scaling of $P(s)$. Consider first a loop-free chain. At genomic separations s much larger than the capture scale $r_c^2/b^2 \sim v_0$, the Gaussian kernel is effectively a step function, and the asymptotic scaling recovers the familiar ideal-chain law $P_0(s) \sim s^{-3/2}$. At very short separations $s \ll v_0$, however, entire segments fall within the capture radius, leading to a plateau in $P(s)$ and a vanishing log-derivative. The exact Gaussian result, valid at all $s > l_K$, is

$$P_0(s) = \big(\gamma(s) + 1\big)^{-3/2}, \qquad \gamma(s) = \tfrac{b^2 s}{r_c^2}. \tag{3}$$

This expression highlights the crossover controlled by v_0: finite capture radii strongly affect the short-scale statistics of Hi-C but leave the large-scale scaling of chromatin essentially unchanged.

COMBINED EFFECT OF LOOPS AND CAPTURE RADIUS: PERTURBATION THEORY OF THE SHORT-SCALE DIP

To analyze how quenched cohesin loops and finite Hi-C capture radius jointly shape the contact probability $P(s)$, we combine two ingredients: (i) the quenched disorder formalism for loops and (ii) the Gaussian capture kernel (Model B). The general expression reads

$$P(s) = 4\pi \int_0^\infty K(R)\, \mathcal{P}(R \mid s)\, R^2\, dR, \tag{4}$$

where $\mathcal{P}(R \mid s)$ is the disorder-averaged distribution of end-to-end distances for a segment of contour length s. This "full theory" yields analytical formulas for $P(s)$ across genomic scales, qualitatively reproducing log-derivative profiles seen in Hi-C (Fig. 2A,B). However, because many parameters (r_c, λ, g, N_e) enter simultaneously, robust inference requires a reduction of variables.

Perturbative expansion at short scales. We therefore focus on Regime IV ($s < \lambda, g$), where s/λ, $s/g \ll 1$ and experimental curves display a characteristic V-shaped dip (Fig. 2A). At very short s the polymer behaves as a free random walk, $P(s) \approx P_0(s)$ from (3), giving a monotonic slope with no dip. The dip must therefore arise from loop corrections.

A simple scaling argument clarifies why the short-scale correction depends only on the loop–gap period $T = \lambda + g$, rather than on λ or g separately. Consider placing an interval of length s randomly along the alternating loop–gap sequence. Three cases can occur:

(a) **Gap–gap:** both loci fall in the same gap. This configuration is indistinguishable from a free polymer segment and therefore contributes $P^{(a)}(s) \simeq P_0(s)$. Its statistical weight is $W^{(a)} \approx g/T$, independent of s to leading order.

(b) **Loop–gap:** one locus lies in a loop and the other in an adjacent gap. The weight of such configurations scales as $W^{(b)} \sim s/T$, since the probability of straddling a loop–gap boundary is proportional to the interval length. The corresponding contact probability is again close to the free case, $P^{(b)}(s) \simeq P_0(s)$, so the total contribution is of order $(s/T)P_0(s)$.

(c) **Loop–loop:** both loci are inside the same loop. This has constant weight $W^{(c)} \approx \lambda/T$, independent of s, but the associated contact probability is modified because both ends are constrained within a loop. This produces a correction of order $(s/\lambda)P_0(s)$.

Crucially, the λ appearing in the loop–loop correction cancels against the factor λ/T in its weight. As a result, the total correction is proportional to $(s/T)P_0(s)$, with no explicit dependence on λ or g. This reasoning shows that what matters is not the absolute loop size or

gap length, but only their sum T, i.e. the mean spacing between consecutive loop anchors. Thus the loop density T^{-1} emerges as the sole control parameter governing short-scale corrections to $P(s)$.

Carrying out the perturbative expansion gives

$$P(s) \approx P_0(s)\left(1 + \frac{s}{T} f(\gamma(s))\right), \tag{5}$$

with

$$f(\gamma) = \frac{3\gamma}{2(\gamma+1)}, \qquad \gamma(s) = \frac{b^2 s}{r_c^2}. \tag{6}$$

Origin and scaling of the dip. Taking the logarithmic derivative of (5) yields

$$y(s) = \frac{d\log P(s)}{d\log s} = -\frac{3\gamma}{2(\gamma+1)} + \frac{s}{T}\,\frac{3\gamma(\gamma+2)}{2(\gamma+1)^2}. \tag{7}$$

The first term reflects the finite capture radius (monotonic decay), while the second encodes loop corrections (growing with s/T). Their competition produces a local minimum—a universal dip—absent in models with only one ingredient (Fig. 2B).

These results allow us to characterize quantitatively the short-scale behaviour of $P(s)$. Figure 2B shows that only the combined action of loops and a finite capture radius produces a distinct local minimum in the log-derivative (red curve, highlighted region); models with only loops or only a cutoff do not. This dip is therefore a unique signature of their interplay, with its genomic position set by both T and v_0.

Comparison with numerical calculations confirms that perturbation theory (PT, (7)) captures the full model with high accuracy (Fig. 3). Both PT and full theory predict that increasing v_0 shifts the minimum upward and to the right in the (s, y) plane (Fig. 3A). Remarkably, minima from experimental Hi-C curves (Fig. 2A, dotted line) fall along a similar diagonal band, suggesting a conserved loop density $1/T$ across cell types.

Scaling arguments clarify these observations. The minimum position $s_{\min}$ lies between the two *different* characteristic scales T and $v_0 \ll T$ and, therefore, scales as their geometric mean:

$$s_{\min} \sim \sqrt{T \cdot v_0}. \tag{8}$$

The depth of the dip, $y_{\min}$, being dimensionless, must depend only on their ratio:

$$y_{\min} = y_{\min}(\alpha), \quad \alpha = v_0/T. \tag{9}$$

Since increasing v_0 flattens $P(s)$ at small scales, $y_{\min}(\alpha)$ grows monotonically with α. This behavior is confirmed in Fig. 3B, where PT and full theory for fixed $\alpha = 1/40$ show invariant y-values of the dip. Furthermore, all curves collapse under the rescaling $s \to s/v_0$ (Fig. 3B, inset), demonstrating that $s_{\min} \sim \sqrt{T v_0}$. Analytical calculations in SI Appendix confirm these scaling results for the log-derivative minimum.

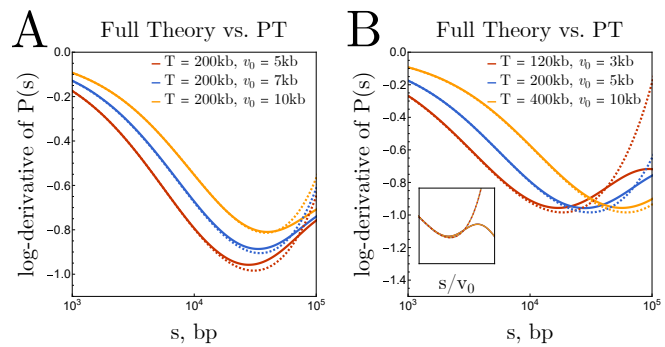

FIG. 3. Comparison of full-theory (solid lines) and perturbation-theory (dashed lines) predictions. **(A)** Logarithmic derivative of $P(s)$ for fixed loop–gap period T and varying restriction fragment lengths v_0. **(B)** Logarithmic derivative of $P(s)$ for constant ratio $T/v_0 = 40$. Inset: rescaling s by v_0 leads to a collapse.

Implication. Perturbation theory thus reduces the apparent complexity of chromosome folding to just two effective parameters, T and v_0. This explains the ubiquity of the short-scale dip in Hi-C profiles and enables robust inference of cohesin density directly from experimental contact statistics.

ACTIVE EXTRUSION SIMULATIONS VALIDATE PERTURBATION THEORY

To test the analytical predictions, we performed molecular dynamics simulations of two-sided *active loop extrusion* and compared steady-state contact probabilities against perturbation theory (PT). Cohesins were simulated with the standard extrusion kinetics [5], parameterized by the mean spacing d (inverse density), processivity l, and extrusion rate r (see Methods). Blocking factors such as CTCF or RNA polymerase were omitted [32, 33], allowing us to isolate the fundamental contribution of cohesin motors.

Phantom chains: ideal benchmark

As a baseline, we simulated phantom chains with random-walk statistics (Fig. 4A,B, Fig. S3). After equilibration, the log-derivative of $P(s)$ agreed quantitatively with PT for quenched loops, with d directly playing the role of the loop–gap period T. In this regime, both PT and simulations depend only on loop density and capture length v_0, while processivity l is irrelevant (Fig. 4B). Only at larger scales, near $s \approx \lambda$, does l influence the contact profiles. Varying d (160 and 320 kb; Fig. S3) confirmed the one-to-one correspondence between PT and active simulations, validating the quenched-loop picture at short scales.

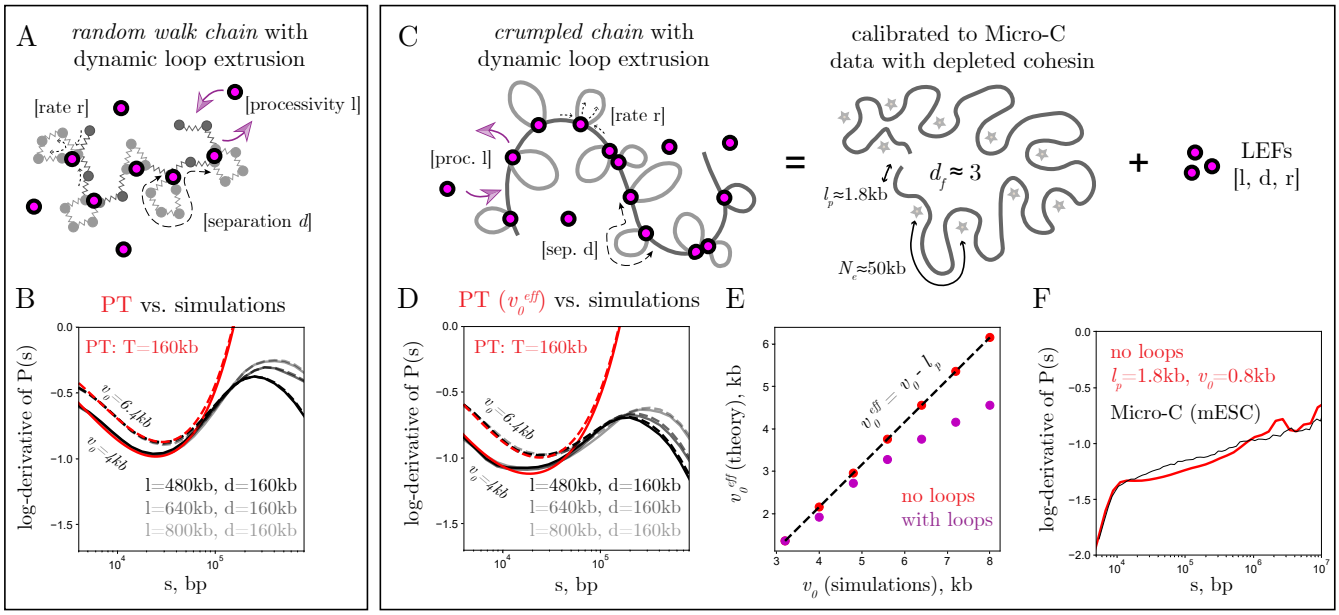

FIG. 4. Validation of perturbation theory (PT) predictions through molecular dynamics simulations of active loop extrusion with three parameters d ([kb], separation), l ([kb], processivity) and r ([kb/s], extrusion rate) on two polymer systems: (left panel) phantom chains with random-walk statistics and (right panel) crumpled ring polymers with a finite persistence. **(A)** Schematic representation of dynamic loop extrusion on a random-walk chain. **(B)** Comparison between theoretical predictions and simulation results for random-walk chains, showing two restriction fragment lengths: $v_0 = 4$ kb (solid line) and $v_0 = 6.4$ kb (dashed line). **(C)** Schematic of a crumpled polymer chain, calibrated against Micro-C data, with active loop extruders. **(D)** Theory-simulation comparison for crumpled chains using effective restriction lengths v_0^{eff}, with $v_0 = 4$ kb (solid) and $v_0 = 6.4$ kb (dashed). **(E)** Relationship between actual v_0 in simulations and effective v_0^{eff} values in PT for both looped and unlooped systems. **(F)** Calibration of loops-free crumpled chain simulation against Micro-C data from [31] for cells without cohesin. Curve matching at $s < 10$ kb enables unit conversion between simulations and biological chromatin (800bp per bead, the compaction ratio assumed for both phantom and crumpled chain simulations).

Crumpled chains: realistic chromatin background

We next examined chromatin-like polymers with crumpled statistics, modeled as equilibrated ring polymers in a melt (Methods). In the absence of cohesin, simulations reproduced $P(s) \sim s^{-\gamma_c}$ with $\gamma_c \approx 1$, consistent with the crumpled globule observed experimentally [6, 10, 26]. Calibration against Micro-C data yielded a mapping of ∼800 bp per bead, a persistence length $l_p \approx 1.8$ kb, and entanglement length $N_e \approx 50$ kb (Fig. S4), in line with microscopy and prior polymer studies [23, 24].

Interestingly, Micro-C data from cohesin-depleted mESCs [31] show a wedge-shaped log-derivative profile ($s \sim 1$–10^4 kb; Fig. 4F), rather than the plateau seen in Hi-C under similar conditions. Simulations revealed that this wedge arises when the capture length v_0 is smaller than l_p, explaining the Micro-C versus Hi-C contrast (Fig. S4B). Thus, a single calibrated crumpled-chain model reconciles both protocols, establishing a universal loops-free baseline for chromosome folding.

Active extrusion in crumpled chains

Introducing active loop extrusion on this background, we again compared simulations with PT. Remarkably, agreement was recovered provided that the capture length v_0 was replaced by an *effective* length

$$v_0^{\text{eff}} = v_0 - l_p, \quad v_0 \geq l_p, \tag{10}$$

which accounts for the excluded-volume effect of persistence-length blobs. At $v_0 \approx l_p$ few contacts are registered, while for $v_0 \gg l_p$ the flexible-chain PT applies directly (Fig. 4D,E).

Equation (10) was first validated in crumpled chains without loops (Fig. S5A): simulated log-derivatives were monotonic and matched perturbation theory with large T, confirming that the persistence correction alone captures the short-scale behavior. When loops were introduced, PT and simulations continued to agree provided $v_0 \lesssim 5$–6 kb (Fig. S5B). At larger fragment lengths, however, a modest additional reduction of v_0^{eff} by 10–20% was required. This effect can be traced to the geometry of loop anchors: cohesin bases act as bulky steric objects that locally thicken the chromatin fiber, displacing nearby segments and effectively screening short-range

contacts. When several loops happen to be anchored in close proximity, these bases form *clustered* steric hubs that enhance the screening, further reducing the effective capture of nearby fragments. As a result, restriction fragments overlapping single or clustered loop bases underreport interactions compared to fragments of the same genomic size placed in the bulk polymer. Within the PT framework this screening manifests as an effective shortening of the fragment length, i.e. a downward shift of v_0^{eff}. Importantly, no analogous correction was needed for the loop–gap period $T = d$, indicating that loop density remains the sole structural control parameter, while the renormalization of v_0^{eff} reflects local steric and clustering effects at loop bases.

Summary

Together, these simulations demonstrate that the perturbation theory remains valid under non-equilibrium active extrusion and realistic chromatin compaction, provided persistence is incorporated through the effective fragment length v_0^{eff} (10). This unifying description explains Hi-C and Micro-C profiles across protocols and enables robust inference of cohesin loop density directly from short-scale features of $P(s)$.

INFERENCE OF *IN VIVO* LOOP DENSITY

We next applied perturbation theory to infer cohesin loop density directly from Hi-C contact statistics. Figure 5A compares theoretical predictions with Hi-C data from Oksuz *et al.* [19] in human ESCs, generated with two crosslinking protocols (FA versus FA+DSG). The agreement is striking: additional crosslinking reduces the effective fragment length v_0^{eff} nearly threefold, while the loop periodicity T remains unchanged at ∼110–130 kb.

More generally, theoretical curves with fixed $T = 150$ kb and varying v_0^{eff} (2–11 kb) recapitulate the short-scale features of diverse Hi-C datasets (Fig. 5B). These include the initial decay of $P(s)$ and the local minimum in its log-derivative around $s \approx 20$–40 kb (highlighted in Fig. 2A). Variation across cell types and protocols is largely absorbed by changes in v_0^{eff}, while the inferred loop density remains stable at ∼6–7 loops/Mb. As expected, PT does not capture the saturation at $s \approx \lambda \sim 100$ kb, which reflects the finite processivity of extrusion.

To formalize the inference, we focused on the minima of the log-derivative. Each minimum is characterized by two observables: its genomic position $s_{\min}$ and depth $y_{\min}$, see (8),(9), where v_0 should be substituted with v_0^{eff}. This separation of dependencies allows the two fundamental parameters, T and v_0^{eff}, to be extracted directly from experimental minima.

Figure 6A plots the theoretical trajectories of $(s_{\min}, y_{\min})$ for $T = 100$–300 kb, parameterized by v_0^{eff}. Experimental minima from 28 Hi-C datasets fall cleanly

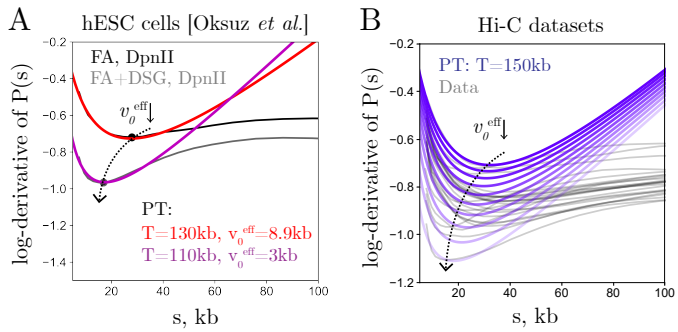

FIG. 5. Parameter inference from chromatin contact data. **(A)** Comparison of Hi-C protocols in hESCs using data from [19]: standard Hi-C (FA crosslinking, DpnII digestion) versus enhanced Hi-C (FA+DSG crosslinking, DpnII digestion). Red/magenta curves: perturbation theory (PT) fits with optimized periodicity T and effective fragment length v_0^{eff}. **(B)** PT predictions (blue curves) for fixed $T = 150$ kb and varying v_0^{eff} (2–11 kb), overlaid with Hi-C data from multiple studies.

onto these trajectories. Across five independent studies [3, 19, 30, 34, 35], encompassing multiple mammalian cell types and protocols, the inferred periodicity converges to $T = 174 \pm 38$ kb (mean ± s.d.), corresponding to a conserved loop density of 6.0 ± 1.3 loops per megabase. The robustness of this value across species, laboratories, and experimental conditions points to a fundamental organizing principle of mammalian chromatin.

COHESIN STOICHIOMETRY ANALYSIS

We then compared the loop density inferred from Hi-C with independent measurements of cohesin abundance. Table 1 compiles RAD21 copy numbers from fluorescence microscopy and spectroscopy, as well as from mass spectrometry, in mouse ESCs and HeLa cells [16, 18]. Normalized to genomic length (cohesins/Mb), these counts align closely with the inferred density of 6.0 ± 1.3/Mb (Fig. 6B).

Dividing cohesin density by loop density yields 1.02 ± 0.15 cohesins per loop (mean ± s.d.), providing quantitative evidence for a monomeric extrusion mechanism. While uncertainties in protein counts are substantial (up to ∼40%), leaving room for a minor dimeric fraction (∼2–17%), the dominant contribution is consistent with single cohesin complexes anchoring loops.

Thus, by combining polymer theory, Hi-C statistics, and protein stoichiometry, we provide direct evidence that loop extrusion in vivo is carried out predominantly by monomeric cohesin motors.

DISCUSSION

Physics message. We show that short-scale chromosome folding admits a *two-parameter* description: in the

TABLE I. Estimates of cohesin abundance across different cell types and measurement methods. Data are derived from Cattoglio *et al.* [16] and Holzmann *et al.* [18]. Bold values indicate the average across the two measurement methods used in each study. Linear cohesin density was calculated from the absolute number of dynamically chromatin-bound proteins.

Ref.	Method	Protein	Cell phase	Cell line	# of dynamically bound	Genome size, Mb	Linear density, per Mb
C. Cattoglio et al, 2019	FCM	RAD21	G1/G2	mESC	52456 ± 5015	3 × 2716	6.44 ± 0.62
C. Cattoglio et al, 2019	"in-gel" FCS	RAD21	G1/G2	mESC	34586 ± 14169	3 × 2716	4.24 ± 1.74
C. Cattoglio et al, 2019	**average**	**RAD21**	**G1/G2**	**mESC**	**43521 ± 7515**	**3 × 2716**	**5.34 ± 0.92**
J. Holzmann et al, 2019	LC-MS	RAD21	G1	HeLa	69275 ± 14698	7900	8.77 ± 1.86
J. Holzmann et al, 2019	LC-MS	RAD21	G2	HeLa	74750 ± 11064	2 × 7900	4.73 ± 0.70
J. Holzmann et al, 2019	FCS/iFRAP	RAD21	G2	HeLa	104952 ± 26360	2 × 7900	6.64 ± 1.67
J. Holzmann et al, 2019	**average**	**RAD21**	**G1/G2**	**HeLa**	**82992 ± 10715**	**3/2 × 7900**	**6.71 ± 0.87**

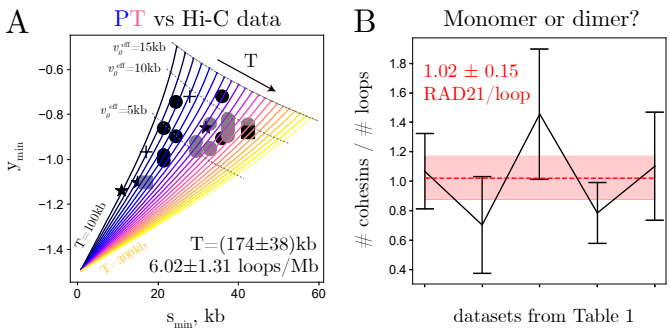

FIG. 6. Quantitative inference of cohesin loop properties from Hi-C datasets [3, 19, 30, 34, 35]. **(A)** Log-derivative minima positions ($s_{\min}$) and slopes ($y_{\min}$) for 28 Hi-C datasets (symbols), overlaid on PT predictions for loop periodicities $T = 100$–300 kb (colored curves). Increasing v_0^{eff} moves points along each curve. Inferred mean periodicity $T = 174 \pm 38$ kb corresponds to 6.02 ± 1.31 loops/Mb. **(B)** Cohesin counts from Table 1 normalized by inferred loop number, yielding virtually a single cohesin complex per loop. Shaded region: uncertainty range; dashed line: monomeric expectation.

window $l_K \lesssim s \lesssim 50$–$100$ kb, the contact statistics are governed solely by (i) the mean loop–gap period T (the inverse loop density) and (ii) an effective Hi-C fragment length v_0^{eff} that encapsulates protocol- and fiber-level physics. A perturbation theory around the loop-free ideal chain, with quenched loop disorder [6, 26] and a finite capture radius [19, 20], yields a universal form for the logarithmic derivative of $P(s)$ and predicts a distinct minimum at

$$s_{\min} \sim \sqrt{T\, v_0^{\text{eff}}},$$

with depth $y_{\min} = y_{\min}(v_0^{\text{eff}}/T)$. This reduction constitutes an identifiable inverse problem: two directly measurable observables $(s_{\min}, y_{\min})$ determine the two physical parameters (T, v_0^{eff}).

From scaling to inference. The minimum arises from a competition between detection (finite capture radius) and sparse quenched links (loops). A simple diagrammatic expansion shows that, at short contour separations, only three local configurations contribute at leading order (gap–gap, loop–gap, loop–loop), and their statistical weights sum to a correction proportional to s/T, independent of λ and g separately. This explains why *density*, not individual loop sizes, controls the short-scale response. The perturbation theory quantitatively matches both the full analytical calculation and molecular-dynamics simulations of *active* two-sided extrusion (Fig. 3–4), where extruders were implemented with standard stochastic kinetics [5, 33]. This agreement establishes that non-equilibrium motor action leaves a quasi-equilibrium fingerprint at short s, since local relaxation is faster than extrusion dynamics [6].

Biological implication as a corollary. Applying this scaling framework across diverse mammalian Hi-C datasets [3, 19, 30, 34, 35] yields a conserved density of 6 ± 1 cohesin-mediated loops per megabase. Independent RAD21 counts from fluorescence microscopy and mass spectrometry [16, 18] align with this value, implying approximately one cohesin per loop base. Thus, the polymer-physics inference points to *monomeric* loop extrusion as the prevailing in vivo mechanism; widespread dimerization [17, 36] would require cohesin abundances inconsistent with measurements. Reports of dimer-like foci are naturally reconciled as collision-induced co-localization in a crowded loop array [36].

Scope and limits. Our theory purposefully targets the regime where it is predictive and falsifiable: $l_K \lesssim s \ll \lambda, g \lesssim 10^2$ kb, with $s_{\min} \gtrsim l_K$ and $\sqrt{T v_0^{\text{eff}}} < g$. In this window, topology is effectively screened and the equilibrium assumption holds despite active extrusion. The parameter v_0^{eff} is phenomenological: it aggregates fragment flexibility (with a persistence correction $v_0^{\text{eff}} \approx v_0 - l_p$ for $l_p \sim 1$–2 kb [23, 24]), crosslinking strength, multi-ligation and nuclear crowding [37, 38], hence it is non-universal and protocol dependent. Outside the short-scale window—e.g. near $s \sim \lambda$ or into the crumpled regime $s \gtrsim N_e$—additional parameters (processivity, entanglement length N_e [10, 22]) enter and identifiability is lost, by design of the framework.

Relation to prior modeling. The inferred density refines and unifies earlier simulation- and theory-based estimates of 4–8 loops/Mb from polymer fits [5, 39–42], live-cell imaging [43], and analytical models [6, 44]. Our approach is orthogonal: it exploits a *local* scaling feature (the dip) whose position and depth provide a parameter-complete readout at short scales, immune to the degeneracies that limit long-scale fits.

Outlook. Conceptually, the work illustrates how minimal quenched-disorder plus detection physics yields *uni-*

versal short-scale statistics and turns population Hi-C maps into a quantitative probe of molecular stoichiometry. The same strategy—identify a separation of scales, derive a low-parameter perturbation theory for a protocol-weighted observable, and fit identifiable features—should generalize to (i) multi-motor mixtures (e.g. STAG1/2-specific or mitotic condensin-I/II densities [45]), (ii) time-resolved maps during cell-cycle transitions [46, 47], and (iii) other capture modalities such as Micro-C or SPRITE with distinct kernels. More broadly, it shows how polymer theory can elevate genome-wide contacts from descriptive statistics to *physics-calibrated meters* of the underlying molecular machinery.

METHODS

Molecular dynamics simulations of loop extrusion

All simulations were performed with the `polychrom` framework (https://github.com/mirnylab/polychrom), which implements Langevin dynamics for bead–spring polymers. Chromatin fibers were modeled as chains of $N = 40{,}000$ beads connected by harmonic springs. One bead was mapped to $\sim$ 800 bp by comparison of the crumpled chain simulation with Micro-C data (see Results). After equilibration, simulation snapshots were collected and contact probabilities $P(s)$ were computed using contact-capture Model A. In this model, bead–bead contacts are registered at capture distances $r_c(v)$ corresponding to the typical spatial size of polymer segments of length v, measured independently in simulations. The raw contact counts $\mathcal{N}(r_c(v))$ were then weighted by the fragment-length distribution $\rho(v|v_0) = v^{-1}\exp(-v/v_0)$, simulating the fragmentation protocol in Hi-C. This yields the final contact probability

$$P(s|v_0) = \sum_{i=1}^{k} \mathcal{N}(r_c(v_i))\, \rho(v_i|v_0)\Delta v,$$

where $k = 20$ fragment sizes were used, equally spaced between $v_1 = 1$ kb and $v_k = 4v_0$, with step size Δv.

a. Phantom and crumpled chain ensembles. To benchmark perturbation theory in the simplest case, we first simulated phantom chains with random-walk statistics, i.e. with excluded-volume interactions switched off. To model more realistic chromatin folding, we simulated a single unknotted ring polymer in periodic boundary conditions. The chain was equipped with excluded-volume interactions using a Weeks–Chandler–Andersen potential of scale $\sigma \approx 15$ nm [48], and harmonic bonds of mean length $l_b = 0.8\sigma$ [6]. Simulations were performed at volume density $\rho\sigma^3 = 0.4$, which best reproduced Micro-C log-derivatives in the absence of cohesin. Fiber stiffness was introduced via a soft harmonic angular potential $\frac{1}{2}k_{\rm ang}(\theta_{i,i+1,i+2} - \pi)^2$ with $k_{\rm ang} = 1\,k_BT$. Mapping against Micro-C data yielded a compaction ratio of $\sim$ 800 bp per bead and, by fitting to a worm-like chain model (Fig. S4A), a persistence length $l_p \approx 1.8$ kb. The entanglement length was $N_e \approx 50$ kb, consistent with previous studies [6, 23, 24]. Importantly, the same ring-melt system also matched three different cohesin-depleted Hi-C datasets (Wutz, Rao, Schwarzer) when varying only the average fragment length v_0 (Fig. S4B). This calibrated system therefore served as our baseline model of chromatin without cohesin.

b. Active loop extrusion model. Two-sided loop extrusion was implemented by coupling 3D polymer dynamics to a 1D stochastic model of cohesin kinetics [5, 33]. Each cohesin was represented as an additional harmonic bond between two beads. At initialization, $N_0 = N/d$ cohesins were placed randomly along the chain, defining the mean spacing d. During simulations, each bound cohesin unbound with rate $k_{\rm off} = 2r/l$, where l is the processivity length and r the extrusion speed, and the same number of cohesins were reloaded at random unoccupied sites to maintain N_0 constant. Upon binding, cohesins extruded symmetrically at speed $r \approx 1$ kb/s until unbinding, collision with another cohesin, or reaching the chain ends. No extrusion barriers (e.g. CTCF, RNA polymerase) were included, to isolate the fundamental effect of cohesin motors. Simulation time was mapped to real time by calibrating the Rouse diffusion coefficient of individual beads, yielding $D_R \approx 10^{-2}\,\mu\mathrm{m}^2\,\mathrm{s}^{-1/2}$, consistent with experimental measurements (see [6] for details).

c. Simulation protocol. For each parameter set, three independent replicates were run for $\sim 10^4$ extrusion steps (corresponding to $\sim 10^4$ s of real time). After equilibration, 10^3 conformations were sampled for analysis in each replicate. Contact probability curves $P(s)$ and their log-derivatives were computed at fragment lengths v_0 in the kb range, matching experimental protocols, and directly compared with perturbation theory predictions and Hi-C data.

ACKNOWLEDGMENTS

We acknowledge funding from the Russian Science Foundation (Grant No. 25-13-00277). K.P. was additionally supported by the Alexander von Humboldt Foundation. We thank members of the Mirny and Metzler laboratories for helpful discussions and valuable feedback.

[1] J. Dekker and L. Mirny, The chromosome folding problem and how cells solve it, Cell **187**, 6424 (2024).

[2] E. Lieberman-Aiden *et al.*, Comprehensive mapping of long-range interactions reveals folding principles of the

human genome, Science **326**, 289 (2009).
[3] S. Rao *et al.*, A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping, Cell **159**, 1665 (2014).
[4] K. Nasmyth, Disseminating the genome: joining, resolving, and separating sister chromatids during mitosis and meiosis, Annual Review of Genetics **35**, 673 (2001).
[5] G. Fudenberg, N. Abdennur, M. Imakaev, A. Goloborodko, and L. Mirny, Emerging evidence of chromosome folding by loop extrusion, Cold Spring Harbor symposia on quantitative biology , 45 (2017).
[6] K. E. Polovnikov *et al.*, Crumpled polymer with loops recapitulates key features of chromosome organization, Physical Review X **13**, 041029 (2023).
[7] M. Ganji *et al.*, Real-time imaging of dna loop extrusion by condensin, Science **360**, 102 (2018).
[8] I. F. Davidson, B. Bauer, D. Goetz, W. Tang, G. Wutz, and J. M. Peters, DNA loop extrusion by human cohesin, Science **366**, 1338 (2019).
[9] Y. Kim, Z. Shi, H. Zhang, I. J. Finkelstein, and H. Yu, Human cohesin compacts DNA by loop extrusion, Science **366**, 1345 (2019).
[10] J. D. Halverson, J. Smrek, K. Kremer, and A. Y. Grosberg, From a melt of rings to chromosome territories: the role of topological constraints in genome folding, Reports on Progress in Physics **77**, 022601 (2014).
[11] L. Mirny, The fractal globule as a model of chromatin architecture in the cell, Chromosome Res. **19**, 37 (2011).
[12] A. Grosberg, Y. Rabin, S. Havlin, and A. Neer, Crumpled globule model of the three-dimensional structure of DNA, Europhysics Letters **23**, 373 (1993).
[13] S. Nechaev, A. Semenov, and M. Koleva, Dynamics of a polymer chain in an array of obstacles, Physica A: Statistical Mechanics and its Applications **140**, 506 (1987).
[14] B. Chan and M. Rubinstein, Activity-driven chromatin organization during interphase: Compaction, segregation, and entanglement suppression, Proceedings of the National Academy of Sciences **121**, e2401494121 (2024).
[15] K. Nasmyth, Cohesin: a catenase with separate entry and exit gates?, Nature Cell Biology **13**, 1170 (2011).
[16] C. Cattoglio *et al.*, Determining cellular CTCF and cohesin abundances to constrain 3D genome models, Elife **8**, e40164 (2019).
[17] F. Ochs *et al.*, Sister chromatid cohesion is mediated by individual cohesin complexes, Science **383**, 1122 (2024).
[18] J. Holzmann *et al.*, Absolute quantification of cohesin, CTCF and their regulators in human cells, Elife **8**, e46269 (2019).
[19] B. Akgol Oksuz *et al.*, Systematic evaluation of chromosome conformation capture assays, Nature Methods **18**, 1046 (2021).
[20] B. R. Lajoie, J. Dekker, and N. Kaplan, The Hitchhiker's guide to Hi-C analysis: practical guidelines, Methods **72**, 65 (2015).
[21] T. Cremer and C. Cremer, Chromosome territories, nuclear architecture and gene regulation in mammalian cells, Nature Review Genetics **2**, 292 (2001).
[22] A. Rosa and R. Everaers, Ring polymers in the melt state: the physics of crumpling, Physical Review Letters **112**, 118302 (2014).
[23] J. M. Arbona, S. Herbert, E. Fabre, and C. Zimmer, Inferring the physical properties of yeast chromatin through bayesian analysis of whole nucleus simulations, Genome Biology **18**, 1 (2017).
[24] M. M. Tortora, H. Salari, and D. Jost, Chromosome dynamics during interphase: a biophysical perspective, Current Opinion in Genetics and Development **61**, 37 (2020).
[25] M. Rubinstein, Dynamics of ring polymers in the presence of fixed obstacles, Physical Review Letters **57**, 3023 (1986).
[26] K. E. Polovnikov and B. Slavov, Topological and non-topological mechanisms of loop formation in chromosomes: Effects on the contact probability, Physical Review E **107**, 054135 (2023).
[27] A. Gavrilov *et al.*, Disclosure of a structural milieu for the proximity ligation reveals the elusive nature of an active chromatin hub, Nucleic Acids Research **41**, 3563 (2013).
[28] T. Nagano *et al.*, Comparison of Hi-C results using in-solution versus in-nucleus ligation, Genome Biology **16**, 175 (2015).
[29] G. Fudenberg and M. Imakaev, FISH-ing for captured contacts: towards reconciling FISH and 3C, Nature Methods **14**, 673 (2017).
[30] B. Bonev *et al.*, Multiscale 3D genome rewiring during mouse neural development, Cell **171**, 557 (2017).
[31] T. Hsieh *et al.*, Enhancer-promoter interactions and transcription are maintained upon acute loss of CTCF, cohesin, WAPL, and YY1, BioRxiv (2021).
[32] A. Monteagudo-Sánchez, D. Noordermeer, and M. V. Greenberg, The impact of DNA methylation on CTCF-mediated 3D genome organization, Nature Structural & Molecular Biology **31**, 404 (2024).
[33] H. B. Brandão *et al.*, RNA polymerases as moving barriers to condensin loop extrusion, Proceedings of the National Academy of Sciences **116**, 20489 (2019).
[34] G. Wutz *et al.*, Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins, The EMBO journal **36**, 3573 (2017).
[35] W. Schwarzer *et al.*, Two independent modes of chromatin organization revealed by cohesin removal, Nature **551**, 51 (2017).
[36] A. Brunner *et al.*, Quantitative imaging of loop extruders rebuilding interphase genome architecture after mitosis, Journal of Cell Biology **224**, e202405169 (2025).
[37] F. Tavares-Cadete, D. Norouzi, B. Dekker, Y. Liu, and J. Dekker, Multi-contact 3C reveals that the human genome during interphase is largely not entangled, Nature Structural & Molecular Biology **27**, 1105 (2020).
[38] P. Olivares-Chauvet *et al.*, Capturing pairwise and multi-way chromosomal conformations using chromosomal walks, Nature **540**, 296 (2016).
[39] A. L. Sanborn *et al.*, Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes, Proceedings of the National Academy of Sciences **112**, E6456 (2015).
[40] E. J. Banigan and L. A. Mirny, The interplay between asymmetric and symmetric DNA loop extrusion, Elife **9**, e63528 (2020).
[41] S. Ghosh and D. Jost, How epigenome drives chromatin folding and dynamics, insights from efficient coarse-grained models of chromosomes, PLoS computational biology **14**, e1006159 (2018).
[42] S. K. Ghosh and D. Jost, Genome organization via loop extrusion, insights from polymer physics models, Briefings in Functional Genomics **19**, 119 (2020).

[43] M. Gabriele *et al.*, Dynamics of CTCF-and cohesin-mediated chromatin looping revealed by live-cell imaging, Science **376**, 496 (2022).

[44] B. Chan and M. Rubinstein, Theory of chromatin organization maintained by active loop extrusion, Proceedings of the National Academy of Sciences **120**, e2222078120 (2023).

[45] J. Gibcus *et al.*, A pathway for mitotic chromosome formation, Science **359**, 2133 (2018).

[46] K. Abramo *et al.*, A chromosome folding intermediate at the condensin-to-cohesin transition during telophase, Nature cell biology **21**, 1393 (2019).

[47] H. Zhang *et al.*, Chromatin structure dynamics during the mitosis-to-G1 phase transition, Nature **576**, 158 (2019).

[48] K. Kremer and G. Grest, Dynamics of entangled linear polymer melts: A molecular-dynamics simulation, The Journal of Chemical Physics **92**, 5057 (1990).