A fractal polymer with loops recapitulates key features of chromosome organization

Kirill Polovnikov[1,2,*], Sergey Belan[3,4,*], Maxim Imakaev[2], Hugo B. Brandão[6,†], and Leonid A. Mirny[2,5,†]

[1]Skolkovo Institute of Science and Technology, 121205 Moscow, Russia
[2]Institute of Medical Engineering and Science, MIT, Cambridge
[3]Landau Institute for Theoretical Physics, Russian Academy of Sciences, 1-A Akademika Semenova av., 142432 Chernogolovka, Russia
[4]National Research University Higher School of Economics, Faculty of Physics, Myasnitskaya 20, 101000 Moscow, Russia
[5]Department of Physics, MIT, Cambridge
[6]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, 02142
[*]These authors contributed equally
[†]Corresponding authors: Leonid A. Mirny (mirny@mit.edu); Hugo B. Brandão (brandao@hugoresearch.com)

February 1, 2022

Abstract

Chromosomes are exceedingly long polymers compacted in a cell nucleus. While it was long suggested that mammalian chromosomes are folded into loops, experimental detection of such loops has remained a daunting task. The most comprehensive experimental information about chromosome spatial organization is provided by Hi-C experiments that measure the frequency of contacts between all chromosomal positions. The lack of a tractable physical model of a polymer folded into loops limits our ability to interpret experimental data. It thus remains unknown how to obtain accurate and quantitative information about the nature of chromosomal looping from Hi-C. Here, we introduce a model of a polymer with random loops, solve it analytically and extend it by simulations for real chains. Remarkably, our model faithfully reproduces complex shapes of experimental contact probability curves universal among mammalian cells. This agreement suggests that chromosomes are indeed organized into a sequence of randomly positioned loops and allows to estimate loop sizes. We further show that excluded volume in real chains can induce osmotic and topological repulsion between loops. Thus, our new framework allows interpretation of experimental data and suggests that interphase chromosomes are crumpled polymers further folded into a sequence of randomly positioned loops.

Chromosomes of high eukaryotes during interphase, i.e. when they are not compacted for cell division, have long been conjectured to be folded into loops [1, 2]. Most recently it was proposed that such loops are formed by an active (energy-dependent) process of loop extrusion, where loop-extruding motors associate to chromosomes, extrude loops, and dissociate, thus maintaining the chromosome polymer in the steady state with

random loops (see [3] for review). Detecting extruded loops from microscopy is difficult due to their transient nature (≈ 10 min life-time [4]) and high volume-fraction of chromatin. Hi-C experiments provide information about contacts, but an extruded loop cannot be told apart from any other occasional contact. To see whether characteristics of Hi-C data can reveal the presence of loops we develop a theoretical framework of a polymer folded into a series of random loops.

A Hi-C experiment produces a map of contact frequency $P(i,j)$ between all pairs of genomic positions i and j [5]. Beyond a variety of features visible in $P(i,j)$ maps, these data can be characterized by the dependence of the contact frequency $P(s)$ on the genomic distance between positions $s = |i-j|$, averaged over all positions. The scaling of $P(s) \sim s^{-1}$, which is different from $\sim s^{-3/2}$ expected for a random walk, was used to argue that chromosomes are folded into the crumpled (fractal) states [5–7]. The full shape of the $P(s)$ curves exhibit *universal* features for interphase chromosomes of higher eukaryotes (see Fig. 1B), and change when the motor proteins performing loop extrusion are perturbed (reviewed in [3]). Here, we develop a theory of a polymer folded into random loops and show that it reproduces the nearly universal $P(s)$ seen in experiments across various animal cell types.

Polymer chain folded into loops

We start with a classical bead-spring model of a polymer chain [11–13] that corresponds to a three-dimensional random walk (i.e. ideal chain with fractal dimension $d_f = 2$). In general, a chain can be characterized by its fractal dimension, d_f, where the mean-squared spatial size of the polymer segment of s beads behaves as $r^2(s) \sim s^{2/d_f} b^2$ for beads of the length b. An important experimentally measured characteristic of such a chain is the contact probability between ends of the segment, $P_0(s)$, that according to the mean-field argument [5, 7, 14], is inversely proportional to the volume spanned by the segment

$$P_0(s) \propto \frac{1}{s^{3/d_f}} \tag{1}$$

Though (Eq.1) is not true in general, it holds for the class of fractional Brownian trajectories, for which the effective Hamiltonian has been recently developed [15–17].

Now, we consider this fractal chain folded into consecutive loops with the average contour length, λ, and separated by gaps that have an average contour length, g, both exponentially distributed (Fig. 1A). The effect of loops on $P(s)$ can be calculated under the assumption of fixed loops as follows. First, one calculates the contributions of different diagrams *at equilibrium*. Second, one averages these probabilities over all possible pairs of monomers i, j involving different diagrams, such that $|i-j| = s$ (Fig. 2A). Finally, the remaining averaging over the distribution of random loops and gaps is performed.

Exact diagrammatic expansion for ideal chains

For the ideal chain, $d_f = 2$, our approach gives a precise analytical expression for contact probability $P(s)$ in the presence of random loops (see Supplementary Information). The ultimate result is factorized into the unconstrained conditional probability $P_0(s)$ of a loops-free chain (Eq.1) and function $\mathcal{P}$ of the scaled genomic distance s/λ and parameter λ/g:

$$P(s) = P_0(s)\, \mathcal{P}(s/\lambda;\ \lambda/g). \tag{2}$$

The function $\mathcal{P}$ accurately accounts the contributions from four classes of diagrams (a)-(d) depicted in (Fig. 2A); it is expressed in the form of multiple integrals involving the Bessel functions, which are to be computed numerically.

Strikingly, despite the simplicity of the ideal chain model, the shape of the $P(s)$ curve, best represented by its log-derivative, $\frac{d\log P(s)}{d\log(s)}$, qualitatively matches the corresponding curves computed from experimental Hi-C data for mammalian cells, see Fig. 2 B,C and Fig. 1B.

In accord with experiments (Fig. 2C), loops perturb the fractal shape of the $P(s)$ curve and result in the formation of a shoulder at $s \approx \lambda$, and a corresponding "peak" and "dip" on the log-derivative plot, see Fig. 1C. Moreover, in line with our theory, experiments where loops

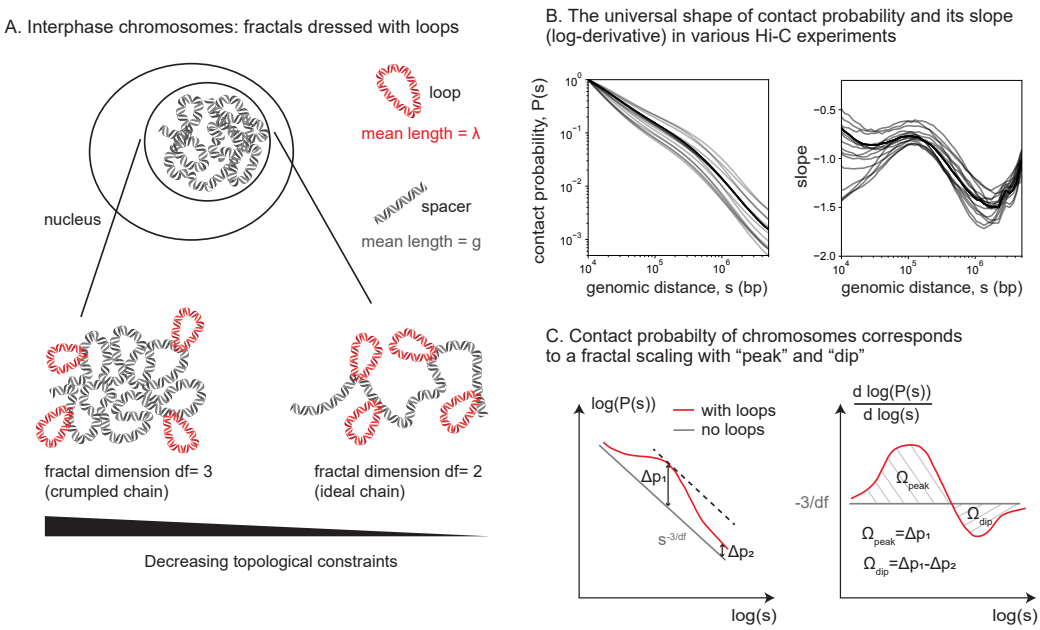

Figure 1: Folding of a fractal polymer into random loops recapitulates universal $P(s)$ profiles in Hi-C experiments. **A**: a sketch of fractal polymers with different fractal dimension d_f folded into loops. The mean size of the loops is λ, the mean size of the gaps (spacers) is g. **B**: a collection of 24 contact probability curves $P(s)$ (left) and its log-derivatives (right) from experimental Hi-C data for different human cell types [8–10]. The shape of the profiles is universal across experimental conditions and cell types. **C**: a sketch of the typical behaviour of the contact probability $P(s)$ with and without loops (left) along with its logarithmic derivative (right) up to several megabases. A geometrical argument explains that formation of the dip is due to difference in elevation of $P(s)$ at small and large scales.

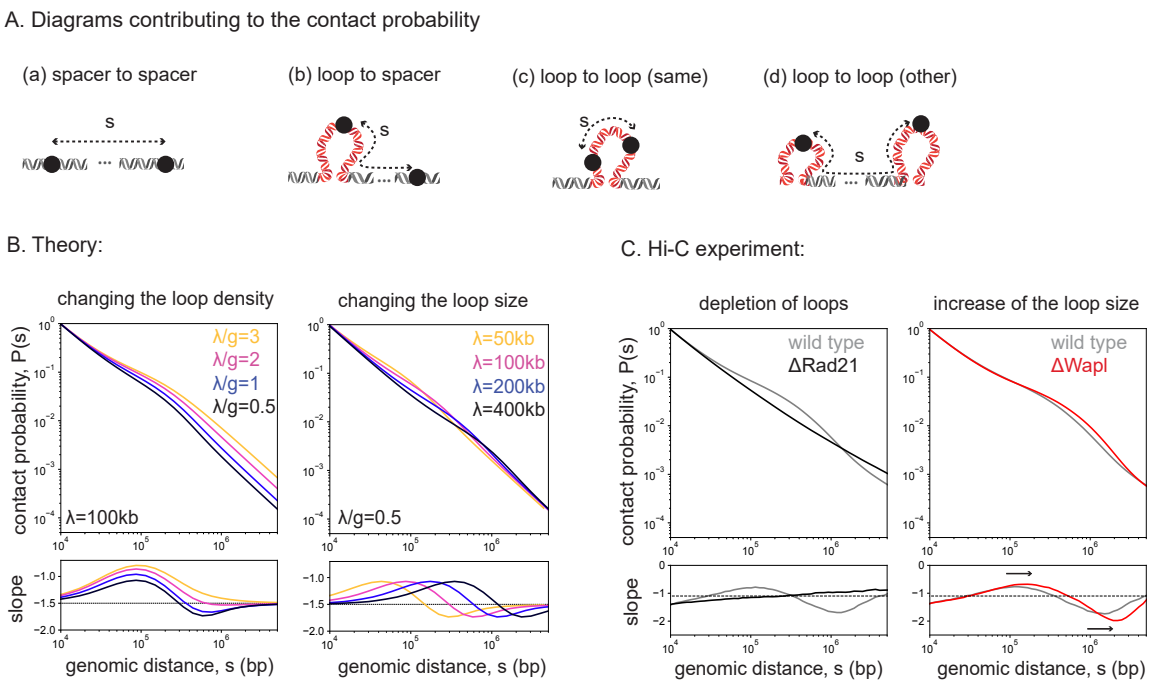

Figure 2: Model of the contact probability for the ideal chain ($d_f = 2$) folded into an array of random loops. **A**: Illustration of the possible diagrams contributing to the contact probability. The three dots (...) denote a chromosome segment that may contain an arbitrary number of loops. **B**: Evolution of the $P(s)$ and its log-derivative upon the change of the loop density, λ/g (left) and the mean loop length, λ (right), where the value of the other parameter is fixed as indicated. **C**: Behavior of the $P(s)$ and its log-derivative in a Hi-C (specifically, Micro-C) experiment upon (left) disruption of cohesin complexes by Rad21 depletion, thus eliminating loops; (right) Wapl depletion, thus increasing of the loops size. Rad21 is a subunit of cohesin, thus, its near-complete depletion results in disruption of cohesin-mediated loops on chromosomes. Wapl protein on the contrary unloads cohesin from chromosomes, thus its depletion increases cohesin residence time and, as a result, the average loop length λ. The cell type is mouse embryonic stem cell, same for the two datasets [18].

are eliminated by depletion of Rad21 (one of the subunits of the loop-extruding motor, cohesin) [18, 19] give $P(s)$ with an almost constant slope (i.e fractal) from 10 kb to $\approx 5,000$ kb, yielding a near constant in the log-derivative without the characteristic peak and the dip (see Fig. 2C, left). Furthermore, experiments where loop sizes were increased by depleting the protein Wapl (see Fig. 2C, right) demonstrate displacement of the peak to the right. Consistently, we see (Fig. 2B) that the shoulder on $P(s)$ and the peak on the log-derivative curve, travel to the right upon the increase of the mean loop length, λ, at fixed loop density λ/g.

It is straightforward to understand how loops perturb the contact probability $P(s)$, generating the curves we observe in the theory and experiments. Indeed, the loops impose two effects on $P(s)$ at different scales. At small scales, the dominant contribution to the elevation of $P(s)$ comes mainly from contacts between two loci within the same loop (e.g. see Fig. S6, for loop densities $\lambda/g \geq 1/5$). Within a loop, the contact probability is increased due to the smaller physical size of a loop compared to an open chain of the same contour length, and due to a shorter effective contour length between two points near the loop base. Thus,

$$\mathcal{P}(s \leq \lambda) \approx 1 + p(s/\lambda)\,\frac{\lambda}{\lambda + g}, \qquad (3)$$

where $p(s/\lambda)$ is some function describing the shape of the shoulder, which is weighted by the fraction of the chain within the loops.

At large scales, when $s \gg \lambda + g$, i.e. two monomers are separated by sufficiently many loops[1], the contact probability becomes elevated due to a shorter effective contour length between the two points, since the shortest path between monomers skips all intervening loops. Thus,

$$\mathcal{P}(s \gg \lambda + g) \propto \left(\frac{\lambda}{g} + 1\right)^{3/d_f}. \qquad (4)$$

The interplay between the shoulder at small scales and the elevation of $P(s)$ at large scales determines the shape of $P(s)$ as well as the "peak" and the "dip" of the log-derivative. Fig. 1C displays a simple geometric argument for the dip: the shoulder, (Eq.3), at $s \approx \lambda$ must be higher than the elevation at $s \gg \lambda$, (Eq.4). As the spacer, g, shrinks (loop density λ/g increases), the elevation of $P(s)$ (Eq.4) grows, while the height of the shoulder (Eq.3) quickly saturates. Therefore, upon a gradual compaction of the chain into a dense array of loops, the dip becomes more shallow. As the analytical solution for $d_f = 2$ suggests (Fig. 2B, left), there is a critical loop density $(\lambda/g)^* \approx 2$, at which the dip completely disappears. Physically, this marks the transition to a dense array of loops where effective shortening of the polymer at large scales reaches the level of compaction at scales of a single loop.

One-loop approximation for non-ideal chains

As we see, the ideal chain model recapitulates general features of experimental data. However, it does not capture an important aspect of chromosome organization such as the value of the $P(s)$ slope closer to -1, rather than $-3/2$ in most mammalian cell types [5–7, 20–22]. Indeed, the ≈ -1 scaling stretches for two orders of magnitude in genomic scale s, when cohesin-mediated loops are eliminated [18, 19, 23] (Fig. 2C). According to (Eq.1), it corresponds to the almost compact fractal folding with $d_f \approx 3$. Such characteristics reflect the fractal (crumpled) globule folding of a polymer, e.g. in a equilibrium melt of non-concatenated unknotted rings [6, 7, 14, 20, 22] or as a long-lived non-equilibrium state of a collapsed linear chain [21].

The case of $d_f > 2$ is not tractable analytically because of statistical cross-coupling between *all* loops and gaps. In fact, it is the consequence of the central limit theorem: $d_f > 2$ imposes strong non-local correlations between distal chain segments [15, 16]. However, in the limit of sufficiently sparse loops $\lambda/g \ll 1$ one can obtain a perturbative expression for $P(s)$ of a non-

[1] The limit corresponds to the amount of loops $n \gg \lambda/g$, which is equivalent to the contour length stored in the gaps being much greater than the loop size λ.

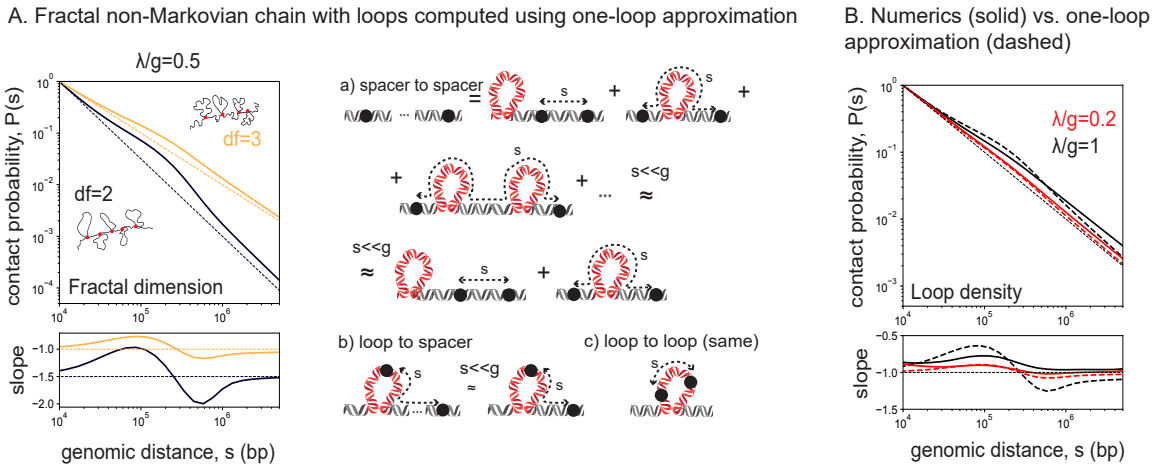

Figure 3: Beyond the ideal chain: fractal polymer of $d_f > 2$ with loops **A**: (left) One-loop approximation for the fractal polymer. Effect of the fractal dimension on $P(s)$ and its log derivative ($\lambda/g = 0.5, \lambda = 100$kb). (right) Expansion to the diagrams with weights linear in λ/g. **B**: Comparison of one-loop approximation and numerical scheme for different loop densities ($d_f = 3, \lambda = 100$kb).

Markovian fractal chain with an arbitrary $d_f \geq 2$:

$$\mathcal{P}(s) = \frac{P(s)}{P_0(s)} \approx 1 + \frac{\lambda}{g} f\left(\frac{s}{\lambda}\right). \qquad (5)$$

Using the effective Hamiltonian for the fractal chain with statistics of the fractional Brownian motion, as suggested recently [15], it is possible to obtain *an exact* linear-order expression for the shape function $f(s/\lambda)$ (see Supplementary Information). The intuition behind the structure of this result (Eq.5) is transparent: in the case of weak disorder of loops, the leading contribution comes from the loop-free fractal scaling $P_0(s)$, while the single-loop diagrams (see Fig. 3A) produce a correction linear in the small parameter $\lambda/g \ll 1$.

Consequently, in the linear order in λ/g the logarithmic derivative of the contact probability (Eq.5) reads

$$\frac{d \log P(s)}{d \log s} \approx -3/d_f + \frac{\lambda}{g}\Delta(s), \qquad (6)$$

where $\Delta(s) = sf'(s)$ encodes for the universal shape of the log-derivative function.

By construction, the baseline (loops-free) slope of $P(s)$ in (Eq.6) is its asymptotic value $\frac{\log P_0(s)}{\log s} = -3/d_f$. In particular, the loopy chain folded with the fractal dimension $d_f = 3$ (Fig. 3A) has a slower decaying $P(s)$ than the ideal chain with $d_f = 2$ at all scales, the former being consistent with Hi-C data for mammalian somatic cells (compare, e.g., with Fig.1B, 2C). Interestingly, upon the increase of the fractal dimension d_f, the amplitudes of peak and dip notably diminish (Fig. 3A and Fig. S3): additional compactness due to loops has a weaker effect on already crumpled chains with $d_f > 2$.

Though the one-loop approximation allows us to understand the non-Markovian effect of loops on a fractal chain, it yields the quantitatively correct result only at very small density of loops ($\lambda/g \leq 0.1$, Fig. S2). An alternative way to look into the effect of $d_f > 2$ on the $P(s)$ is to sample positions of loci at given contour distance s on the numerically generated sequences of loops and gaps. In this numerical approach (first suggested in [24] for the ideal chain) the spatial sizes of chain segments are considered statistically independent, which breaks down at $d_f > 2$. Nev-

ertheless, this approach has an important advantage of virtually taking into account all the diagrams (a)-(d), Fig. 2A, with their proper sample weights. As we demonstrate in the Fig. S4, the numerical approach perfectly agrees with the ideal chain theory ($d_f = 2$), irrespective of the value of λ/g.

Together, this analysis shows that folding into loops perturbs fractal non-Markovian polymer states with $d_f > 2$ in a way similar to ideal chains, producing universal log-derivative curves, though above a different baseline $-3/d_f$. Importantly, such an extension of the ideal chain theory allows us to rationalize slopes of $P(s)$ observed in Hi-C experiments.

Real loopy chains: repulsion between the loops

We next consider the extension of the ideal chain theory to the real chains with loops (Fig. 4A). We perform equilibrium simulations of a loopy polymer ($N = 25,000$ beads, $\approx$ 25Mb chromosomal region) with excluded volume and topological constraints (see Methods).

We observe that shapes of $P(s)$ curves for real chains deviate at moderate loop densities $\lambda/g > 1$ from the predictions of the ideal chain theory, exhibiting a pronounced dip at $s > \lambda$ on the log-derivative[2] (Fig. 4B and Fig. S7). Such pronounced dip, however, is also evident in the log-derivatives from Hi-C data for a broad range of cells and conditions.

Real chains can induce repulsion between the loops, that in turn could stiffen the polymer backbone. To quantify this effect we measure the emerging persistence length l_p of the backbone and find that indeed it increases with the loop density, $l_p \propto \lambda/g$ (Fig. 4C and Fig. S7E). Thus, we start observing another phenomenon in the loopy polymer – osmotic (excluded volume) and topological repulsion between the loops [7, 25], which leaves a distinct signature in the $P(s)$ curve (Fig. 4B).

This real-chain effect on $P(s)$ can be understood by extending our theoretical framework to a chain with phantom loops but stiff spacers. The shape of the $P(s)$ at scales $s \approx \lambda$ (Eq.3) is determined largely by intra-loop contacts and is not sensitive to stiffening of spacers. However, at large scales $s \gg \lambda + g$ the contact probability is decreased, because loops become more segregated from each other due to the stiffer backbone. Thus,

$$\mathcal{P}\left(s \gg \lambda + g\right) \propto \left(\frac{\lambda}{g} + 1\right)^{3/d_f} \frac{1}{l_p^{3(1-d_f^{-1})}}, \quad (7)$$

and the mean squared size of the backbone reads $r^2(s) = (s/l_p)^{2/d_f} l_p^2\, b^2$ at $s > l_p$. In other words, loop-loop repulsion drives stronger screening of contacts between the loops. Therefore, the increase in contact frequency at large scales due to chain shortening (elevation of Δp_2) is counteracted by spatial separation of the loops (Fig. 4A). As a result, the dip on the log-derivative becomes deeper.

In order to develop an *effective* numerical description of the real chains, suitable for interpretation of Hi-C data, we extended our model of the ideal chain by introducing the phenomenological persistence length of the spacers, l_p, while keeping the loops unchanged. In full agreement with our theoretical argument (Eq.7) above, incorporation of effective stiffness of spacers to the phantom chain model allows to recapitulate the observed log-derivatives from simulations of real chains, Fig. 4D. A gradual increase of effective l_p at fixed loop density λ/g on an ideal chain results in deepening of the dip (Fig. S7D). Achieving best agreement with simulations of topologically constrained loops, however, requires the values of l_p somewhat larger than the l_p measured in simulations (Fig. S7E). This argues that repulsion between the loops in real chain leads not only to stiffening of the backbone but also to additional screening of contacts between more distant loops. Both effects are captured by the theory with the *effective* persistence of the backbone.

The observed stiffening is reminiscent to osmotic stretching of spacers of low-density bottle-

[2] At scales of a single loop the log-derivatives of ideal and real chains are similar. This is a consequence of general screening of the excluded volume interactions in a highly concentrated system (the concentration blob is approximated by several kb), [11, 13].

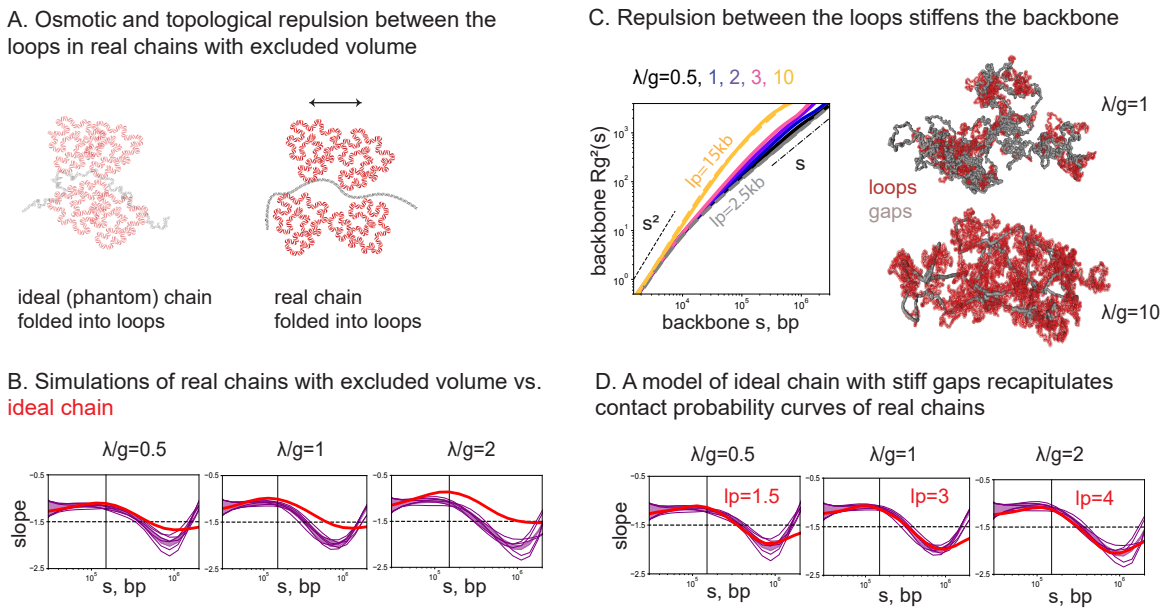

Figure 4: Simulations of real loopy chain. **A**: A sketch of phantom and real chains with excluded volume and topological constraints. Loops in real chains are segregated by means of osmotic and topological repulsion between the loops. **B**: (purple) Log-derivatives of $P(s)$ from polymer simulations with real chain for various loop densities. Three simulation replicates at two cutoffs=3,5 are shown at each plot. The filled area is restricted by the two curves averaged over the replicates. (red) Curves for the ideal chain are computed using the numerical scheme. **C**: (left) Behaviour of the gyration radius of the backbone $R_g^2(s)$ as a function of the backbone contour distance s for various loop density for the real chain. Dashed curves correspond to the $R_g^2(s)$ with particular l_p from the worm-like chain (WLC) model. (right) Snapshots of conformations from simulations for two values of loop density. **C**: (purple) Same as on B. (red) Curves for the ideal chain with the effective persistence of spacers, l_p, computed using the numerical scheme.

brushes in bad solvent [26], driven by the lack of space between the *linear* neighboring side chains. However, we note that the osmotic stretching between the respective *loopy* side chains is expected to kick in only at $\lambda/g \geq 4$, when two neighbouring loops start to overlap[3]. Therefore, we attribute the observed stiffening at lower loop densities $\lambda/g < 4$ to topological repulsion between all the loops, which collectively swell the chain backbone.

Finally, we note that this phenomenon of topologically induced backbone swelling is remarkably similar to the emergent self-avoidance of the primitive path in annealed lattice animals, used to describe crumpling of unknotted rings [22]. Together this analysis shows, that (i) excluded volume and topological repulsion lead to increased spatial separation of loops that (ii) can be effectively captured by our theory with effective stiffness of the spacers.

Discussion

While it has been long suggested that interphase chromosomes have loops, direct observation of loops inside living cells by microscopy or Hi-C is obstructed by the high volume density of chromatin and inability to resolve loops from just other folds of the chain or random contacts. We demonstrate that the model of a polymer chain folded into loops can quantitatively recapitulate the $P(s)$ curves observed in Hi-C experiments. Our theory allows to detect and characterize chromosome loops from chromosome conformation experimental data, and agrees with experiments where loops are perturbed experimentally [18, 19].

A polymer folded into random loops is a new and exciting physical system. Surprisingly, loops affect not only local folding of the polymer, but also long-range contacts between monomers far along the chain. As we show, the interplay between the loops-induced compaction at small scales and an arrangement of an array of loops at large scales results in a characteristic shape of the $P(s)$ and its log-derivative. We further demonstrate that repulsion between the loops in real chains affects long-range interactions as evident from the amplitude of the dip. Altogether, our theory sheds light on the spatial folding of chromosomes as loopy fractal polymer chains and allows to infer parameters of such organization.

Consistently with [27] we observe that the peak on the log-derivative, referred as λ_{inf}, corresponds to the mean loop length λ. For an ideal chain, the correction factor λ_{inf}/λ turns out to be rather close to unity ($\approx 0.8 - 0.9$) and this value does not change much with the loop density λ/g or d_f (Fig. 5A and Fig. S11). In the real chains, loops stiffen the backbone and shift the position of the peak from the true λ further to the left, with the correction factor $\lambda_{inf}/\lambda \approx 0.6 - 0.8$ (see Fig. 5A, bottom right panel). These observations allow to estimate typical sizes of chromosome loops from *in vivo* Hi-C data, yielding a range of $\lambda = 100 - 150$ kb (see Figs. S12-S22). Furthermore, in the sparse loops regime, it is possible to infer the loop density, λ/g, from $P(s)$ (see Methods). However, in the dense loop regime $\lambda/g > 0.5$ such inference of the loop density is not possible since reals chains with different λ/g yield indistinguishable log-derivative curves (see Fig. 5C and Methods)

By measuring positions and amplitudes of the peak and the dip of log-derivatives one can put theoretical predictions and experimental measurements on the same diagram (see "fountain" diagram, Fig. 5B). A cloud of Hi-C data points overlays the theoretical curves on this diagram, allowing to reveal important general character of interphase chromosomes in human cells [8–10].

First, we observe that the chromosomes resemble loopy chains with $\lambda/g \approx 1 - 3$, with the effective stiffness parameter of backbone $l_p \approx 1.5 - 2.5$ and, hence, the loops experience topological repulsion. Similar conclusions are drawn from the direct fitting of various datasets with our theory (see Figs. S12-S22 and a table with the inferred parameters for the data analysed in Fig. S23). Importantly, the repulsion between loops alters $P(s)$ and its derivative at scales much larger than the scale of a single loop (i.e. at

[3]Approximating a loop by a Brownian bridge, it has the mean squared size 4 times smaller than the linear counterpart.

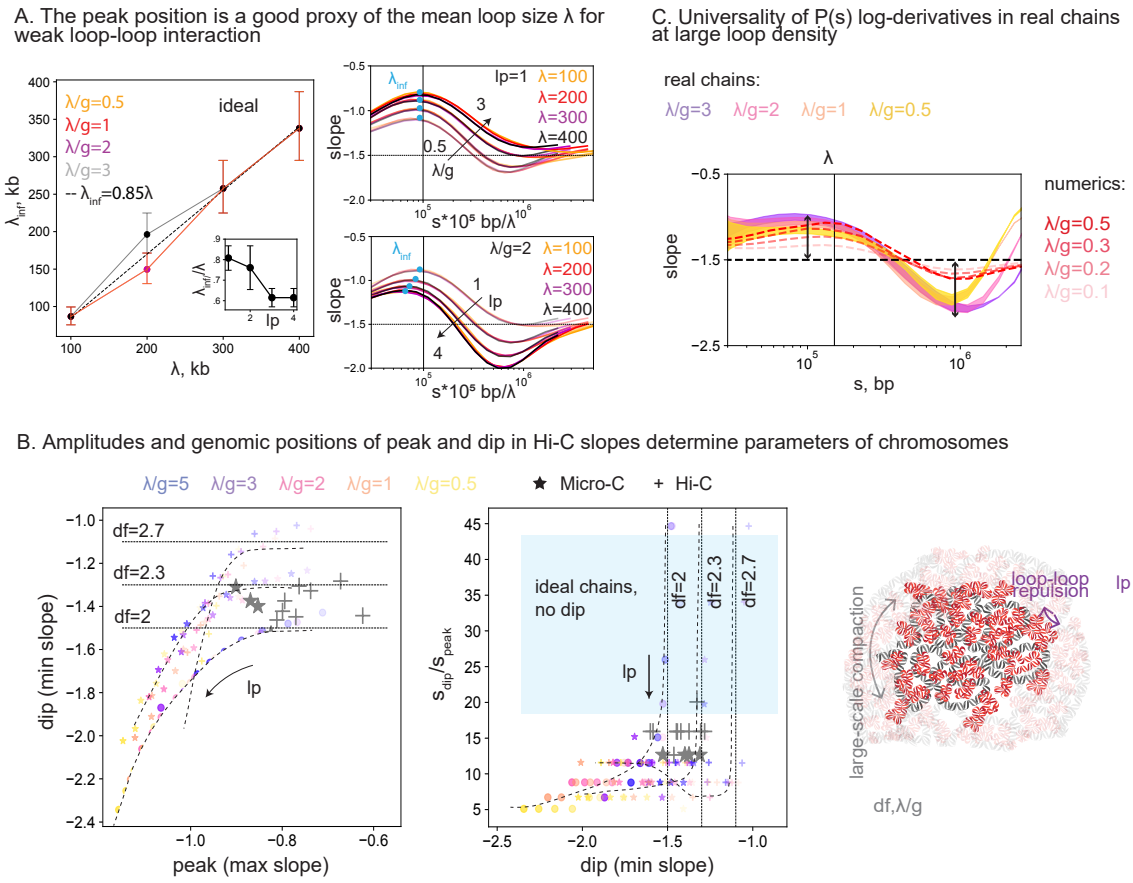

Figure 5: Estimation of parameters from the Hi-C data. **A**: The $\lambda_{inf} - \lambda$ diagram for various loop densities and ideal phantom chains, $d_f = 2, l_p = 1$. The dashed line is the best-fit. The inset shows the dependency of λ_{inf}/λ as a function of l_p. Collapse of the $P(s)$ log-derivatives scaled by λ for different λ is shown for varying loop density λ/g, fixed $l_p = 1$ (right top) and varying l_p, fixed $\lambda/g = 2$ (right bottom). **B**: Numerically-generated diagrams of polymer states in peak-dip local log-derivatives (left) and s_{dip}/s_{peak}-dip (right) for different $\lambda/g, d_f$ and l_p. The black dashed lines designate the level lines with fixed d_f. The horizontal (left) and vertical (right) black lines mark the asymptotic values of the minimal slope (i.e. when the dip is absent) for each value of the fractal dimension. The gray symbols (stars for Micro-C and crosses for Hi-C) correspond to the human differentiated cells based on experimental data from [8] and [9, 10]. (right) A sketch of a typical chromosome surrounded by other chromosomes. **C**: Log-derivative curves for real chains from polymer simulations are almost universal at large density of loops, $\lambda/g \geq 1$. Some deviations from the universal curve can be detected upon slight decrease of loop density below $\lambda/g = 1$ and most notable for much lower loop densities ($\lambda/g = 0.5, 0.3, 0.2, 0.1$), provided by numerics. Each strip corresponds to two values of cutoff=3 (upper line) and cutoff=5 (bottom line), measured in monomers.

$\approx 10\lambda \approx 1$ Mb), indicating that it affects large-scale chromosome organization. Such repulsion between loops can swell individual chromosomes and increase repulsion between individual chromosomes, strengthening "chromosomal territoriality". Furthermore, within a chromosome, topological repulsion could potentially make individual chromosomal domains, formed my loop extrusion, become more isolated from each other.

Second, we see that Hi-C data for a variety of cell types correspond to "crumpled polymer" models that have the fractal dimension larger than $d_f > 2$. Also removal of loops (cohesin depletion) reveals the plateau of the log-derivative at ≈ -1, indicating $d_f \approx 3$ at scales $s \approx 0.5 - 10$ Mb. These observations strongly suggest that chromosomes constitute topologically constrained polymers in the crumpled state $d_f \approx 3$, [7, 20, 28] that are further folded into random loops, similarly to the "loopy globule" concept [29]. Together, both observations suggest that topological interactions play important roles in chromosome organization at different scales: 100kb-1Mb for interactions between loops, and 1-10 Mb for folding into the crumpled state.

While powerful in reproducing experimental data and in its simplicity, our theory has several limitations. First, we consider only equilibrium loops, while the loops actively extruded by motors proteins (e.g. cohesins, condensins) constitute an active/non-equilibrium one. The fixed loops approach relies on the assumption that the time to actively extrude the loop is larger than the time of its passive relaxation. Thus, there exists a critical length scale, s^*, for which relaxation is still faster than extrusion. Quantitative estimates (see Methods) indicate that for the typical extrusion speed of ≈ 1 kb/s [24] the fixed loops assumption is justified for loops of size up to $s^* \approx 1\text{Mb} \gg \lambda$. Second, we do not consider very dense arrays of loops formed during mitosis [30] and meiosis [31], where bottle-brush rigidity of the chain can become dominant.

In summary, we have developed an analytical theory of a fractal polymer folded into stochastic loops, demonstrating that information about loops is contained in the shape of experimentally measurable contact probability curve, $P(s)$, and primarily in its log-derivative. Extended to non-ideal polymers with excluded volume interactions, our theory faithfully reproduces experimental curves that have a universal shape in a broad range of cell types. This agreement indicates that interphase chromosomes are indeed formed by chromatin fiber folded into stochastic loops, that are not directly visible in experimental data. Our analysis also suggests that chromatin fiber is a non-ideal chain, and that loops are interacting with each other through osmotic and topological repulsion, stiffening the backbone. Finally, our theory provides a powerful tool for the inference of loop sizes and characterization of large-scale chromosome folding. Broadly, the polymer model introduced in the paper represents a novel class of polymers folded into loops that can be further extended into non-equilibrium, heteropolymeric, or gel-like systems with loops.

Acknowledgements

We thank Mehran Kardar, Alexander Grosberg, Mikhail Tamm, Vittore Scolari, Sameer Abraham and Anton Goloborodko for valuable discussions on the subject of the paper. LM acknowledges support from the National Institutes of Health Common Fund 4D Nucleome Program HG011536 and a grant R01GM114190. The work of KP is supported by the Russian Science Foundation (Grant No. 21-73-00176). The work of SB on ideal chain model has been supported by Russian Science Foundation (Grant No. 20-72-00170).

Code availability

The code available at https://github.com/kipolovnikov/fractal-loops contains the following: (i) numerical evaluation of the analytical solution for the all-diagrams ideal chain model; (ii) numerical evaluation of the analytical solution for the one-loop approximation; (iii) numerical generation of the contact probability curves (numerical scheme); (iv) the fitting scheme. The folder also contains

a collection of numerical curves for various combinations of parameters, as well as a set of calculated logarithmic derivatives for Hi-C data analysed in this paper.

Methods

Estimation of the critical equilibrium loop extrusion scale

The fixed loops model is a reasonable approximation of the active loop extrusion process at scales, as the extrusion rate $r_{coh} \approx 1$ kb/s is smaller than the thermal rate of stress relaxation. Importantly, the thermal relaxation rate of a polymer depends on the contour (genomic) length scale, s; in the framework of the Rouse model, it can be estimated as $r_{th} = s/\tau_R(s) = l_K^2/(\tau_0 s)$, where l_K is the Kuhn length ($s \gg l_K$) and τ_0 is the microscopic Rouse time (the diffusion time of one monomer at the scale of order of the Kuhn length, [11, 32, 33]). Note that extrusion time is linear in genomic distance s, while thermal relaxation time is quadratic in s. Thus, at small length scales $s < s^*$ thermal relaxation occurs relatively fast, while at larger extrusion scales $s > s^*$ the chains would be at non-equilibrium. The critical equilibrium scale is

$$s^* = l_k^2 (r_{coh}\tau_0)^{-1} \tag{8}$$

The reported microscopic parameters for chromatin in the literature vary greatly and are quite ambiguous [34, 35]. For the coarse-grained modelling one typically takes $l_k = 3$ kb, [36, 37]. As a reference value of the Rouse diffusion coefficient one assumes $D = 10^{-2}\mu^2 s^{-1/2}$, [4, 36]. This yields $\tau_0 = r_k^4 D^{-2} \approx 10^{-2}$s for the Rouse microscopic time, where we have used the conversion ratio $c = 60$ bp/nm, thus, the length of the Kuhn segment is $r_k = 50$ nm. Plugging these values to (Eq.8) results in the following estimate of the equilibrium genomic length scale

$$s^* \approx 1\text{Mb} \tag{9}$$

Since the characteristic scale of loop extrusion, i.e. the mean loop size, is $\lambda \approx 100$ kb $< s^*$, these experimental numbers suggest that at the spatial scales of extrusion activity chromosomes equilibrate relatively fast. However, given wide variations in experimental data for l_k and τ_0, understanding the non-equilibrium effects of extrusion is of great importance (see [38] for recent developments in this direction).

Polymer simulations

Setup

Polymer simulations are done using polychrom (available at https://github.com/open2c/polychrom), a wrapper around the open source GPU-assisted molecular dynamics package OpenMM [39]. In simulations of long linear polymers with loops the polymer is represented as a chain of $N \approx 20000$ monomers with one monomer corresponding to 1 kb, so that the total simulated genomic length is 20Mb. This is sufficient for the simulation of polymer effects at the intermediate scales, i.e. <10 Mb. The chains are equipped with harmonic bonds U_{bond} and excluded volume U_{ev} interactions. Simulations are conducted in the box with periodic boundary conditions (PBC) at volume density $\rho\sigma^3 = 0.3$. The energy of harmonic bonds is as follows

$$U_{bond} = \frac{3}{2a^2} \sum_{i=1}^{N-1} \left(r_{i,i+1} - l_b\right)^2, \tag{10}$$

where a is the standard deviation of the monomer-to-monomer distance $r_{i,i+1} = |\mathbf{r_{i+1}} - \mathbf{r_i}|$ from the equilibrium bond length l_b. We preset $a \approx 0.06$ (by default in polychrom one defines the "wiggle distance", which is by $\sqrt{3/2}$ smaller than a).

The excluded volume potential U_{ev} is introduced via the auxiliary Weeks-Chandler-Anderson (WCA) potential $U(r_i, r_j)$ [40, 41], which is a lifted Lennard-Jones repulsive branch

$$U(r_{ij} = |\mathbf{r_i} - \mathbf{r_j}|) = \begin{cases} 4\varepsilon\left((\sigma/r_{ij})^{12} - (\sigma/r_{ij})^6\right) + \varepsilon, & r_{ij} \le 2^{1/6}\sigma \\ 0, & r_{ij} > 2^{1/6}\sigma \end{cases} \tag{11}$$

where σ is the characteristic scale of the excluded volume repulsion and $\varepsilon = 1$. In order to avoid

strong repulsive forces close to the singularity of (Eq.11) in simulations, the WCA potential is further smoothly truncated

$$U_{ev}(r_{ij}) = \mathcal{H}(U(r_{ij}) - \varepsilon_{tr})\varepsilon_{tr} \times \left(1 + \tanh\left[\frac{U(r_{ij})}{\varepsilon_{tr}} - 1\right]\right) + \mathcal{H}(\varepsilon_{tr} - U(r_{ij}))U(r_{ij}), \quad (12)$$

at the prescribed truncation value $\varepsilon_{tr} = 10$, corresponding to a strong mutual volume exclusion of the beads. In (Eq.12) $\mathcal{H}$ is the step function. The potential (Eq.12) acts between every pair of beads, except for neighboring ones. The value of the equilibrium bond length $l_b = 0.6\sigma$ was chosen smaller than the excluded volume scale specifically to prohibit strand passage for simulations of real chains.

Simulation stages

First the linear chain without loops is equilibrated in the PBC box starting from a random walk configuration. After the equilibration (during the time, roughly corresponding to the Rouse relaxation time of $\approx$ 1 Mb chain segment), the averaged contact probability resemble the three-dimensional fractal with $d_f = 2$ ($\frac{d \log P(s)}{dlogs} \approx -3/2$).

At the second step we gradually extrude the loops on the chain and then fix the loops positions. Loops are simulated as additional harmonic bonds between non-neighboring beads of the same energy as the polymer bonds (Eq.10). The loop lengths $l_1, l_2, ..., l_k$, $k = 100$ are drawn randomly from the exponential distribution with the mean $\lambda = 100$. The starting positions of the loop extruders $x_1, x_2, ..., x_k$ are drawn randomly, such that the gap lengths $x_{i+1} - x_i - l_i$ for $i = 1, 2, ..., k-1$ are distributed exponentially with the mean g (we vary g from 20 to 200). We extrude the loops up to the pre-calculated lengths and then equilibrate the chain with fixed extruders at their final positions (another $\approx$1Mb Rouse time). For each set of parameters of loop extrusion we perform three independent runs (replicates) and the computed observables are averaged.

The contact probability is calculated with cutoff=3, 5 and 7 monomers, emulating the crosslinking radius in the actual contact capture experiment. Note that the change in the $P(s)$ logarithmic slopes under different cutoff is observed at small scales, while at large scales it is almost negligible, see Fig. S7C. Since we do not aim at description of the statistics of contacts at scales <40 kb, the chosen coarse-graining model (1 kb per monomer) is adequate for the purposes of the work.

Quantitative analyses of the chains from simulations (contact probability, $P(s)$, and gyration radius, $R_g^2(s)$) are conducted using the module for the analyses of polymer conformations in polychrom (polymer_analyses).

Fitting procedure

First we have generated the data required to calculate $P(s)$ numerically, using the statistical sampling of contacts (numerical scheme). We have swept the following parameters: loop density $\lambda/g = 0.5, 1, 2, 3, 5$, persistence length of the backbone $l_p = 1, 1.5, 2, 2.5, 3, 3.5, 4$, fractal dimension $d_f = 2, 2.1, 2.3, 2.5, 2.7, 3$. The mean loop size is fixed and equal to $\lambda = 100$. The total length of the chain is $N = 10000$. Simulating the contact probability $P(s = |i - j|)$ in numerics, $n_{samples} = 100$ points i, j on the chain are chosen randomly for each configuration of the random loops. The conditional contact probabilities for each configuration of loops and gaps are then averaged; the total number of configurations is $n_{replica} = 1.5 * 10^5$.

Then, we fit the numerically-generated collection of the $P(s)$ log-derivatives to experimental data. For each value of d_f the best-fit values of parameters: mean loop size λ, loop density λ/g and persistence of the backbone lp – are found as the ones that minimize the error within the genomic region $s \in [40\text{kb}, 4\text{Mb}]$. The error is measured as the root-mean-squared error (rMSE) between the experimental and numerical curves, thus, it has the same "units" as the log-derivatives. In Figs. S12-S22 (first row) we display three curves with the minimal error for each fractal dimension d_f (columns) in the space

of outlined above parameters.

Inference of loop size and loop density

Consistently with [27] we observe that the peak on the log-derivative corresponds to the mean loop size, λ. For an ideal chain, the correction factor λ_{inf}/λ turns out to be very close to unity (≈ 0.96) and does not change much with the loop density λ/g or d_f (Fig. 5A and Fig. S11). In the non-ideal case with topological repulsion, loops stiffen the main chain and shift the position of the peak λ_{inf} from the true λ further to the left, (see Fig. 5A, bottom right panel). This effect can be captured by the system of ideal loops on a stiff backbone (e.g. $l_p \approx 4$ for $\lambda/g \approx 2-3$) yielding the correction factor λ_{inf}/λ of ≈ 0.6 (Fig.4B and Fig.S7). Thus, a more accurate estimate of λ would require fitting of the whole experimental curve by the model. The straightforward fitting of the collection of experimental datasets (Rao et al, [8], Krietenstein et al. [9] and Oksuz et al. [10] for Hi-C datasets, see Figs. S12-S22) suggests that the mean loop size is consistent across the cell types, yielding a range of $\lambda = 100 - 150$ kb.

At the same time, accurate estimation of the loop density, λ/g, from $P(s)$ is possible only for sparse loop arrays. The non-uniqueness of determination of λ/g and l_p is manifested in the diagonal form of the heatmaps for the error of fit in these two variables, see Figs. S12-S22 (second row). Indeed, simulations of real chains with sufficiently dense loops ($\lambda/g > 1$) show that systems with constant λ and different g yield very similar $P(s)$ and their log-derivatives (Fig. 5C, Figs. S8-S10); e.g., peak and dip of the log derivative are similar across different loop densities suggesting that without further assumptions (e.g. assuming some degree of loop repulsion) one cannot confidently elucidate g. Physically, this results from two opposite effects in real chains canceling each other: (i) shortening of gaps increases the contact frequency between distant monomers, (ii) while stronger repulsion between the loops lead to screening of distal contacts. For sparse loops, however, ($\lambda/g < 1$), the effect (ii) is weak and loops slightly perturb the $P(s)$ curve. As a consequence, the log-derivative becomes sensitive to g in the sparse loops regime, allowing its inference from the data. E.g. the fitting scheme applied to the mouse zygotic cells data from [27] identifies the value of loop density $\lambda/g \approx 0.5$ (see Fig. S22).

References

[1] Riggs, A. DNA methylation and late replication probably aid cell memory, and type I DNA reeling could aid chromosome folding and enhancer function. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* **326**, 285–297 (1990).

[2] Münkel, C. *et al.* Compartmentalization of interphase chromosomes observed in simulation and experiment. *Journal of molecular biology* **285**, 1053–1065 (1999).

[3] Fudenberg, G., Abdennur, N., Imakaev, M., Goloborodko, A. & Mirny, L. Emerging evidence of chromosome folding by loop extrusion. *Cold Spring Harbor symposia on quantitative biology* 45–55 (2017).

[4] Gabriele, M. *et al.* Dynamics of CTCF and cohesin mediated chromatin looping revealed by live-cell imaging. *bioRxiv* (2021).

[5] Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).

[6] Mirny, L. The fractal globule as a model of chromatin architecture in the cell. *Chromosome Res.* **19**, 37–51 (2011).

[7] Halverson, J. D., Smrek, J., Kremer, K. & Grosberg, A. Y. From a melt of rings to chromosome territories: the role of topological constraints in genome folding. *Reports on Progress in Physics* **77**, 022601 (2014).

[8] Rao, S. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).

[9] Krietenstein, N. *et al.* Ultrastructural details of mammalian chromosome architecture. *Molecular cell* **78**, 554–565 (2020).

[10] Akgol Oksuz, B. *et al.* Systematic evaluation

of chromosome conformation capture assays. *Nature Methods* **18**, 1046–1055 (2021).

[11] Grosberg, A. Y. & Khokhlov, A. R. *Statistical Mechanics of Macromolecules* (AIP: Woodbury, NY., 1994).

[12] Rubinstein, M. & Colby, R. Polymer Physics. *Oxford University Press, Oxford* (2003).

[13] De Gennes, P. G. Scaling Concepts in Polymer Physics. *Cornell university press* (1979).

[14] Halverson, J. D., Lee, W. B., Grest, G. S., Grosberg, A. Y. & Kremer, K. Molecular dynamics simulation study of nonconcatenated ring polymers in a melt. I. statics. *The Journal of Chemical Physics* **134**, 204904 (2011).

[15] Polovnikov, K., Nechaev, S. & Tamm, M. V. Effective Hamiltonian of topologically stabilized polymer states. *Soft Matter* **14**, 6561–6570 (2018).

[16] Polovnikov, K., Nechaev, S. & Tamm, M. V. Many-body contacts in fractal polymer chains and fractional Brownian trajectories. *Physical Review E* **99**, 032501 (2019).

[17] Polovnikov, K., Gherardi, M., Cosentino-Lagomarsino, M. & Tamm, M. Fractal folding and medium viscoelasticity contribute jointly to chromosome dynamics. *Physical Review Letters* **120**, 088101 (2018).

[18] Hsieh, T. *et al.* Enhancer-promoter interactions and transcription are maintained upon acute loss of CTCF, cohesin, WAPL, and YY1. *BioRxiv* (2021).

[19] Rao, S. *et al.* Cohesin loss eliminates all loop domains. *Cell* **171**, 305–320 (2017).

[20] Grosberg, A., Rabin, Y., Havlin, S. & Neer, A. Crumpled globule model of the three-dimensional structure of DNA. *Europhysics Letters* **23**, 373 (1993).

[21] Grosberg, A., Nechaev, S. & Shakhnovich, E. The role of topological constraints in the kinetics of collapse of macromolecules. *Journal de physique* **49**, 2095–2100 (1988).

[22] Grosberg, A. Annealed lattice animal model and Flory theory for the melt of non-concatenated rings: Towards the physics of crumpling. *Soft Matter* **10**, 560–565 (2014).

[23] Schwarzer, W. *et al.* Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51–56 (2017).

[24] Banigan, E. J. & Mirny, L. Loop extrusion: theory meets single-molecule experiments. *Current opinion in cell biology* **64**, 124–138 (2020).

[25] Marenduzzo, D. & Orlandini, E. Topological and entropic repulsion in biopolymers. *Journal of statistical mechanics: theory and experiment* **9**, L09002 (2009).

[26] Paturej, J., Sheiko, S. S., Panyukov, S. & Rubinstein, M. Molecular structure of bottle-brush polymers in melts. *Science advances* **2**, e1601478 (2016).

[27] Gassler, J. *et al.* A mechanism of cohesin-dependent loop extrusion organizes zygotic genome architecture. *The EMBO journal* **36**, 3600–3618 (2017).

[28] Imakaev, M. *et al.* Effects of topological constraints on globular polymers. *Soft Matter* **4**, 665–671 (2015).

[29] Grosberg, A. Extruding loops to make loopy globules? *Soft Matter* **110**, 2133–2135 (2016).

[30] Gibcus, J. *et al.* A pathway for mitotic chromosome formation. *Science* **359**, 2133–2135 (2018).

[31] Schalbetter, S. A., Fudenberg, G., Baxter, J., Pollard, K. S. & Neale, M. J. Principles of meiotic chromosome assembly revealed in s. cerevisiae. *Nature Communications* **10**, 1–12 (2019).

[32] Doi, M. & Edwards, S. The theory of polymer dynamics. *Oxford University Press, Oxford* (1988).

[33] Tamm, M. V. & Polovnikov, K. Dynamics of polymers: classic results and recent developments. Order, Disorder and Criticality: Advanced Problems of Phase Transition Theory. *World Scientific* (2018).

[34] Dekker, J., Rippe, K., Dekker, M. & N., K. Capturing chromosome conformation. *Science* **295**, 1306–1311 (2002).

[35] Bystricky, K. *et al.* Long-range compaction and flexibility of interphase chromatin in budding yeast analyzed by high-resolution imaging techniques. *PNAS* **101**, 16495–16500 (2004).

[36] Nuebler, J. *et al.* Chromatin organization by an interplay of loop extrusion and compartmental segregation. *PNAS* **115**, E6697–E6706 (2018).

[37] Tortora, M. M., Salari, H. & Jost, D. Chromosome dynamics during interphase: a biophysical perspective. *Current opinion in genetics & development* **61**, 37–43 (2020).

[38] Starkov, D., Parfenyev, V. & Belan, S. Conformational statistics of non-equilibrium polymer loops in Rouse model with active loop extrusion. *The Journal of Chemical Physics* **154**, 164106 (2021).

[39] Eastman, P. & Pande, V. Openmm: A hardware-independent framework for molecular simulations. *Computing in science & engineering* **12**, 34–39 (2010).

[40] Weeks, J., Chandler, D. & Anderson, H. Role of repulsive forces in determining the equilibrium structure of simple liquids. *The Journal of Chemical Physics* **54**, 5237 (1971).

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