

## Topological and nontopological mechanisms of loop formation in chromosomes: Effects on the contact probability

Kirill Polovnikov  and Bogdan Slavov *Skolkovo Institute of Science and Technology, 121205 Moscow, Russia*

(Received 20 March 2023; revised 26 April 2023; accepted 5 May 2023; published 30 May 2023)

Chromosomes are crumpled polymer chains further folded into a sequence of stochastic loops via loop extrusion. While extrusion has been verified experimentally, the particular means by which the extruding complexes bind DNA polymer remains controversial. Here we analyze the behavior of the contact probability function for a crumpled polymer with loops for the two possible modes of cohesin binding, topological and nontopological mechanisms. As we show, in the nontopological model the chain with loops resembles a comb-like polymer that can be solved analytically using the quenched disorder approach. In contrast, in the topological binding case the loop constraints are statistically coupled due to long-range correlations present in a nonideal chain, which can be described by the perturbation theory in the limit of small loop densities. As we show, the quantitative effect of loops on a crumpled chain in the case of topological binding should be stronger, which is translated into a larger amplitude of the log-derivative of the contact probability. Our results highlight a physically different organization of a crumpled chain with loops by the two mechanisms of loop formation.

DOI: [10.1103/PhysRevE.107.054135](https://doi.org/10.1103/PhysRevE.107.054135)

### I. INTRODUCTION

Chromosomal DNA polymers are organized into loops which serve both structural and gene regulatory roles inside the cell nucleus [1,2]. Folding into chromatin loops provides an effective means of compaction of 10-cm-long human chromosomes inside a micron-sized nucleus. In particular, it was proposed that upon the entry to mitosis a fluffy interphase chromosome is arranged into a dense array of loops via the mechanism of lengthwise compaction [3], which results in the bottlebrush structure long observed in mitosis [4,5].

Interphase chromosomes are also organized into loops, however, of significantly lower linear density compared to the mitotic chains [1,6,7]. The loops are formed by SMC proteins (structural maintenance of chromosomes, e.g., cohesin or condensin motors), and their extruding ability has been independently demonstrated in single-molecule imaging by two groups [8,9]. In human cells cohesin provides the main source of short-scale loops on chromosomes. However, the mechanism of cohesin binding to chromosomes is not known. It has been proposed that cohesin can topologically embrace two strands of DNA, forming a loop of progressively increasing size in time [2,10]. In this topological binding model cohesin complex resembles a physical ring that can entrap two DNA strands without chemically binding them. Alternatively, in the nontopological model, cohesin operates as a cross-linker: in the base of the loop it creates a bond between two sites of the chain, which is moving in the course of extrusion [2,10]. Which of the models is realized in the cell is still under debate. As we show here, such a subtle difference in the microscopic structure of cohesin complex can affect chromosome organization at all scales.

A crucial difficulty towards the analytical description of a chromosome as a chain with loops is nonideal statistics of

chromosomal polymer. The physical state of a chromosome can be characterized by the scaling of the average contact probability function  $P(s)$  with the contour distance  $s = |i - j|$  (an analog of the return probability of a random walk to the origin). For chromosomes this function can be extracted from Hi-C experiments [11]. As various data suggest, at large scales, in the range of  $s = 1\text{--}5$  Mb, the scaling of  $P(s) \sim s^{-1}$  is different from  $\sim s^{-3/2}$ , as expected for a three-dimensional random walk, indicating that chromosomes are folded into nonideal states [11–15]. This  $\sim -1$  exponent of the contact probability corresponds to the fractal dimension  $d_f \approx 3$ , which is an asymptotic property of topologically crumpled chains. At shorter scales  $s < 1$  Mb the scaling of  $P(s)$  has a characteristic shoulder, which can be recapitulated by the simplest model of a fractal chain with cohesin-mediated loops breaking the scale invariance [7]. Importantly, upon elimination of cohesin loops in experiments [16,17], the  $\sim s^{-1}$  scaling stretches to shorter distances up to  $s \approx 40$  kb, suggesting that short-scale cohesin loops are formed on top of a crumpled polymer with strong (power-law) correlations along the genomic contour [7].

Here we demonstrate that due to the intrinsic long-range correlations in folding of crumpled chromosomes, the two models of cohesin binding produce drastically different relaxation behavior of the chain with loops. In the nontopological model a chromosome resembles a comb-like polymer with a backbone and loopy side chains. Conformation properties of the chain at large scales are controlled by the backbone (the shortest path) in this model. In contrast, in the topological binding model the backbone is not formed and the polymer remains linear, while folded into loops. Importantly, in this case the chain length is not shortened by the addition of the loops. Furthermore, all the loops are statistically coupled with each other in the topological model, being the segments of a

single crumpled polymer. Note that this physical difference between the models vanishes in the case of ideal chain statistics,  $d_f = 2$ , due to absence of long-range correlations in the chain.

We investigate the effect of the specific binding mechanism on the resulting contact probability function  $P(s)$  for the two models. We treat the loops as frozen disorder on the fractal polymer [7] and assume formation of a well-defined sequence of loops and gaps on the chain with exponential distribution of their lengths (no nested or overlapping loops are allowed) fixed on each conformation. In order to take into account intrinsic correlations in the crumpled chains, we make use of the Gaussian measure of fractional Brownian paths (fBm) as recently proposed [18]. We show that while the nontopological model under these assumptions can be computed analytically [7], the topological model is not fully analytically tractable. For the latter case it is possible to construct a perturbation by a single loop (the one-loop approximation), and analytically compute the first-order correction to the power-law scaling of  $P(s)$ . Despite the lack of a full analytical solution for the topological model, our results show how the two models generate different statistics of contacts across scales.

## II. A FRACTAL POLYMER FOLDED INTO RANDOM LOOPS

We exploit a conventional bead-spring model of a flexible polymer chain, generalized for the arbitrary fractal dimension  $d_f$ , i.e., the average square of the segment size of length  $s$  is

$$r_0^2(s) = b^2 s^{2/d_f}, \quad (1)$$

where  $b$  is a scale of a single bead. In the case of absence of any interactions between the beads other than harmonic polymer bonds, the statistics of chain is ideal with  $d_f = 2$  [19]. A swollen chain in a good solvent in three dimensions (3D) has fractal dimension  $d_f \approx 1.7$  [20], however, excluded volume interactions are likely screened at the relevant scales for chromatin [13]. Topological constraints (in 3D) in the fractal globule model force the chain to fold with the asymptotic fractal dimension  $d_f = 3$  [14,21]. Without excluded volume  $d_f$  can be larger than the dimension of the space: e.g., a phantom randomly branching polymer has  $d_f = 4$  [22]. In general, in order to describe a polymer with any fractal dimension  $d_f \geq 2$ , an effective quadratic Hamiltonian was proposed in Ref. [18]. It generates Gaussian polymer conformations, which map onto trajectories of a fBm.

Next, we consider an ensemble of fractal chains folded into sequences of loops and gaps, see Fig. 1(b). The loops on each chain in the ensemble are randomly positioned, have a random contour length drawn from exponential distribution with the average  $\lambda$ , and are separated by exponentially distributed gaps (spacers) with the average length  $g$ . The loops are fixed for each chain. Similarly to our earlier work [7], we follow the frozen disorder approach, i.e., each chain in the ensemble is *thermally equilibrated* together with its set of the loops. Importantly, the fractal statistics [Eq. (1)] is considered as the inherent property of the chain, due to specific interactions in the polymer (e.g., large-scale topological constraints) which are not perturbed by the addition of loops.

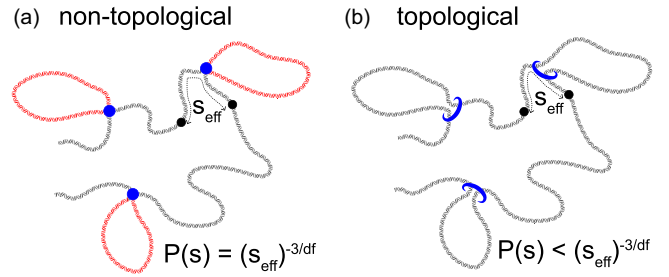


FIG. 1. Two possible binding mechanisms of cohesin to DNA: (a) nontopological model and (b) topological model.

In this paper we focus on the difference between the two models of the loop formation, depicted in Fig. 1. In the nontopological binding model cohesin complex forms a covalent bond in the loop base. Therefore, a chain with loops can be partitioned into a fBm backbone (main chain) and a set of fBm bridges (loops). Notably, the main chain and the loops are fractal (generally nonideal) chains both having fractal dimension  $d_f$  and statistically independent from each other. Relaxation of the polymer at large scales takes place along the shortest path, i.e., along the backbone [ $s_{\text{eff}}$ , see Fig. 1(a)]. In contrast, in the case when the ring-shaped binding protein can embrace two strands of chromatin [topological model of binding, Fig. 1(b)], the chromosome backbone is not formed and stress in the chain propagates along the whole polymer contour. A spacer and its neighboring loop belong to the same fBm polymer, which implies strong (power-law in the contour distance) tangent correlations between them for  $d_f > 2$ . Thus, these correlations are not destroyed by addition of the loops, which physically determines the difference in the contact statistics of a chain with loops at small and large scales.

### A. Nontopological model

Decoupling of correlations between the backbone and the loops in the nontopological model allows for the analytical expression for the  $P(s)$ , which was derived recently by us in Ref. [7]. The same model was exploited in Ref. [23] to quantify the influence of random loops on the intrachain distances in the ideal chain. We briefly describe the main steps and results below.

First, one needs to compute the equilibrium contributions of different diagrams, which classify relative positions of the points  $i, j$  ( $s = |i - j|$ ) with respect to the loop bases: (a) spacer-to-spacer, (b) loop-to-spacer, (c) loop-loop (same), (d) loop-loop (different). For each diagram one computes the variance of the spatial distance  $\sigma^2(i, j) = \langle r^2(i, j) \rangle$  between the points  $i, j$  of interest and makes use of the Gaussian relation  $P(i, j) \sim \sigma^{-3/2}(i, j)$  for the corresponding equilibrium contact probability. For that the vector  $\vec{r}$  is decomposed into the sum of independent vectors (see Fig. 2)  $\vec{r}_1$  (loop),  $\vec{r}_2$  (backbone) for diagram (b); and  $\vec{r}_1$  (loop),  $\vec{r}_2$  (backbone),  $\vec{r}_3$  (loop) for diagram (d). In diagrams (b)–(d) a loop resembles a fBm bridge of the same dimension  $d_f$ . The analytical tractability of this model is due to the statistical independence of the vectors  $\vec{r}_i$ ,  $i = 1, 2, 3$ , as they correspond to separate branches (side loops and the backbone) in the comb-like polymer. Second, one averages the resulting contact probabilities  $P(i, j)$  over

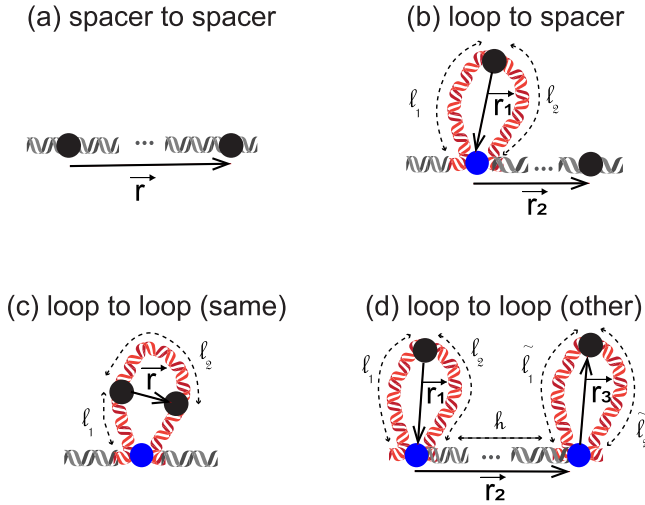


FIG. 2. Nontopological model. Diagrams contributing to the contact probability  $P(s)$ : (a) spacer-spacer, (b) loop-spacer, and (c) loop-loop (same), (d) loop-loop (other).

all possible pairs of monomers  $i, j$  belonging to different diagrams, such that  $|i - j| = s$ , using the appropriate weights of the diagrams. The exponential distribution of loops and gap sizes allows to make use of the well-known result for the propagators of the two-state Markov process and properly weigh contributions of different diagrams [24]. Finally, the remaining averaging over the distribution of random loops and gaps is performed, which results in the sum of multiple integrals involving Bessel functions [7].

Let us see what  $P(s)$  for different diagrams look like in this model. As noted above, the variance of the vector connecting any two points at distance  $s$  in the chain can be expressed through the variances for a segment free of loops,  $\sigma_{\text{free}}^2(s)$  (the backbone), and for a fBm bridge of size  $L$ ,  $\sigma_{\text{bridge}}^2(s, L)$ . For a free segment along the backbone the result is given by Eq. (1), i.e.,

$$\sigma_{\text{free}}^2(s) = b^2 s^{2/d_f}, \quad (2)$$

as the backbone constitutes a fBm trajectory, which is statistically independent of the intervening of loops. The bridge of size  $L$  with the fractal dimension  $d_f$  can be described as a fBm polymer with the same dimension conditionally constrained by the loop. Making use of the Gaussian measure of the fBm paths, one can derive the following result for the fBm bridge (see Ref. [7] and derivation of the conditional propagators for the topological model below):

$$\begin{aligned} \sigma_{\text{bridge}}^2(s, L) &= \sigma_{\text{free}}^2(s) \left( 1 - \frac{[\sigma_{\text{free}}^2(L) + \sigma_{\text{free}}^2(s) - \sigma_{\text{free}}^2(L-s)]^2}{4\sigma_{\text{free}}^2(s)\sigma_{\text{free}}^2(L)} \right). \end{aligned} \quad (3)$$

Note that for the ideal chain  $d_f = 2$ , one obtains the well-known expression for the ideal bridge

$$\sigma_{\text{bridge}}^2(s, L) = \sigma_{\text{free}}^2(s) \frac{L-s}{L} = \frac{b^2 s(L-s)}{L}. \quad (4)$$

Now let us consider diagram (a). In the presence of an arbitrary number of nontopological loops between two points of interest the contact probability reads

$$P_{(a)}(s, x) = \frac{1}{(2\pi\sigma_{\text{free}}^2[(1-x)s])^{3/2}}, \quad (5)$$

where  $x$  ( $0 \leq x < 1$ ) denotes the fraction of the subchain length occupied by the loops. In other words, the intervening loops lead to a reduction of the effective contour distance between points of interest.

Next, let us consider a subchain of length  $s$  with one end belonging to the gap region and another end belonging to the loop. The loop containing one of the two sites of interest is parametrized by the lengths  $l_1$  and  $l_2$  as shown in Fig. 2(b). Clearly, the separation vector  $\vec{r}$  between two sites can be represented as a sum of mutually independent zero mean Gaussian random vectors  $\vec{r}_1$  and  $\vec{r}_2$ . Thus, we have

$$\begin{aligned} P_{(b)}(s, l_1, l_2, x) &= \frac{1}{(2\pi(\sigma_{\text{bridge}}^2[l_2, l_1 + l_2] + \sigma_{\text{free}}^2[(1-x)(s-l_2)]))^{3/2}}, \end{aligned} \quad (6)$$

where  $0 \leq x < 1$ ,  $l_1 \geq 0$ , and  $0 \leq l_2 \leq s$ . As before,  $x$  is the fraction of contour length along the backbone occupied by the loops.

Now let us consider a subchain located inside the loop, see Fig. 2(c). This situation corresponds to a pure bridge, therefore

$$P_{(c)}(s, l_1, l_2) = \frac{1}{(2\pi\sigma_{\text{bridge}}^2[s, l_1 + l_2])^{3/2}}, \quad (7)$$

with the parameters belonging to the range  $l_1 \geq 0$  and  $l_2 \geq s$ .

Finally, for diagram (d) the vector  $\vec{r}$  is decomposed into the sum of independent Gaussian vectors  $\vec{r}_1$  in the first bridge,  $\vec{r}_2$  along the backbone, and  $\vec{r}_3$  in the second bridge. Thus, the contact probability for this diagram reads [see Fig. 2(d)]

$$P_{(d)}(s, l_1, l_2, h, \tilde{l}_1, \tilde{l}_2, x) = \frac{1}{(2\pi\sigma_d^2)^{3/2}}, \quad (8)$$

with the following variance:

$$\sigma_d^2 = \sigma_{\text{bridge}}^2[l_2, l_1 + l_2] + \sigma_{\text{free}}^2[(1-x)h] \quad (9)$$

$$+ \sigma_{\text{bridge}}^2[\tilde{l}_2, \tilde{l}_1 + \tilde{l}_2]. \quad (10)$$

Expressions (5)–(8) determine the contact probabilities for all four classes of diagrams for any loop density  $\lambda/g$  and any fractal dimension  $d_f$  of the chain. The final result for the contact probability in the nontopological model is given by the weighted sum

$$P^{\text{nontopo}}(s) = \sum_{i=a,b,c,d} \langle p_{(i)}(s | \{A\}_i) \rangle, \quad (11)$$

where  $\langle p_{(i)}(s | \{A\}_i) \rangle$  are the particular contributions to the contact probability from each diagram  $i$ , integrated over the possible values of diagram-specific parameters  $\{A\}_i$  and properly weighted. In particular, the exponential distribution of loops and gap sizes allows to use the probabilities for the

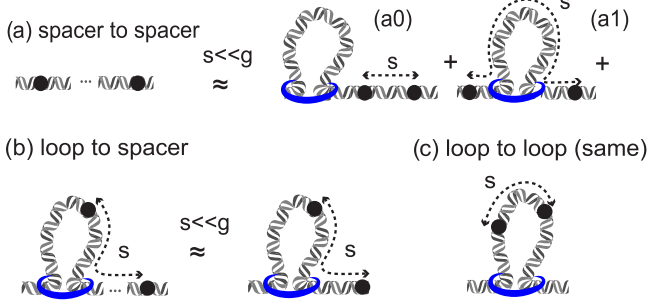


FIG. 3. One-loop approximation for the topological model. Diagrams (a) spacer-spacer, (b) loop-spacer and (c) loop-loop (same).

two-state Markov process for calculation of the weights (see Refs. [7,23] for details).

### B. Topological model

In contrast, in this case all loops and spacers between them are statistically coupled for  $d_f \neq 2$ , which does not allow to derive the full analytical expression accounting for all four diagrams. Indeed, in a chain with long-range correlations the conditional probability of contact between any points  $i, j$  depends on all the constraints in the chain. In the language of random walks, due to the intrinsic long-range memory in a fBm walk with the Hurst parameter  $H > 1/2$  ( $H = 1/d_f$ ), the conditional return probability at time  $t = T$  depends on all the intermediate returns (loops) at previous times  $t < T$ .

However, treating the loops as a perturbation to the fractal chain scaling, one can obtain a self-consistent linear order correction for  $\mathcal{P}^{\text{topo}}(s)$  using the Gibbs measure of fBm polymer conformations [18,25]. For that the non-Markovian propagators of the single loop diagrams (see Fig. 3) should be computed. Let us consider the looping probability  $P(i, j | k, n)$  of a fBm random walk at time points  $i, j > i$  conditioned on the loop at time points  $k, n$ , such as  $k < n$ . There are three one-loop diagrams corresponding to this situation that we aim to process (Fig. 3):

(a) spacer-spacer:

(a0)  $k < n < i < j$  points  $i, j$  both locate in the gap right next to the loop;

(a1)  $i < k < n < j$  points  $i, j$  belong to the consecutive gaps and there is one loop in between them;

(b) loop-spacer:  $k < i < n < j$  one point belongs to a loop while the other belongs to the gap next to it;

(c) loop-loop:  $k < i < j < n$  both points are inside the loop.

Introducing dimensionless variables  $x = 1 - l/s$  and  $y = (k - i)/s$ , where  $s = j - i$  and  $l = n - k$ , one can express the constrained propagators  $P(i, j | k, n)$  for each of the diagrams as follows (see Supplemental Material [26]):

$$P(i, j | k, n) = P(s, x, y) = \frac{1}{s^{3H}} \times I(x, y), \quad (12)$$

where the loops-induced correction function  $I(x, y)$  is

$$I(x, y) \propto \frac{1}{\left[1 - \frac{1}{4}(1-x)^{-2H} g^2(x, y)\right]^{3/2}}, \quad (13)$$

and the function  $g(x, y)$  has the following universal form:

$$g(x, y) = (1-y)^{2H} + |y+1-x|^{2H} - |x-y|^{2H} - |y|^{2H}. \quad (14)$$

The absolute values in Eq. (14) are to be expanded depending on relative positions of the monomers in each particular diagram. Clearly, the function  $g(x, y)$  reflects the additional damping factor to the unconstrained contact probability  $s^{-3H}$  in Eq. (12), which is due to the loop formation. Importantly, it contains information on the non-Markovian properties of the fractal polymer, specific to the topological binding model.

In the one-loop approximation, the contact probability  $\mathcal{P}^{\text{topo}}(s)$  for the topological model is determined by the sum of contributions coming from diagrams (a), (b), and (c), i.e.,

$$\mathcal{P}^{\text{topo}}(s) = P_{a_1}(s) + P_{a_0}(s) + P_b(s) + P_c(s). \quad (15)$$

The weight of each diagram for exponentially distributed loop and gap lengths can be expanded in a limit  $\lambda/g \ll 1, s/g \ll 1$  (see Supplemental Material [26]). The resulting expression for the averaged contact probability in the one-loop approximation can be written as follows:

$$\mathcal{P}^{\text{topo}}(s) \propto \left[1 + \frac{\lambda}{g} f(s)\right] \frac{1}{s^{3H}}, \quad (16)$$

where the first term reflects the contribution of the free propagator and the function  $f(s)$  in the brackets responds for the perturbative contribution of the one-loop diagrams

$$f(s) = -1 + \frac{s}{\lambda} \left\{ -1 + 2 \int_{-\infty}^0 s dx \rho[s(1-x)] \int_{-\infty}^{x-1} dy [I(x, y) - 1] + \int_0^1 s dx \rho[s(1-x)] \left[ \int_0^x dy I(x, y) + 2 \int_{x-1}^0 dy I(x, y) \right] + \int_{-\infty}^0 s dx \rho[s(1-x)] \left[ \int_x^0 dy I(x, y) + 2 \int_{x-1}^x dy I(x, y) \right] \right\}. \quad (17)$$

Taking a log-derivative of Eq. (16) and keeping only linear in  $\lambda/g$  terms, we arrive at the following perturbative expansion:

$$\frac{\partial \log \mathcal{P}^{\text{topo}}(s)}{\partial \log s} = -3H + \frac{\lambda}{g} \Delta(s), \quad (18)$$

where  $\Delta(s) = s \frac{\partial f(s)}{\partial s}$ . It is straightforward to show that  $\lim_{s \rightarrow 0} f(s) = 0$  so that the loop-free result gets recovered at small scales (short polymer subchains do not feel the loop constraints). In the opposite limit one obtains  $\lim_{s \rightarrow \infty} f(s) = 3H$ . A similar expression for the ideal chain case  $H = 1/2$  has been previously obtained in Ref. [27].

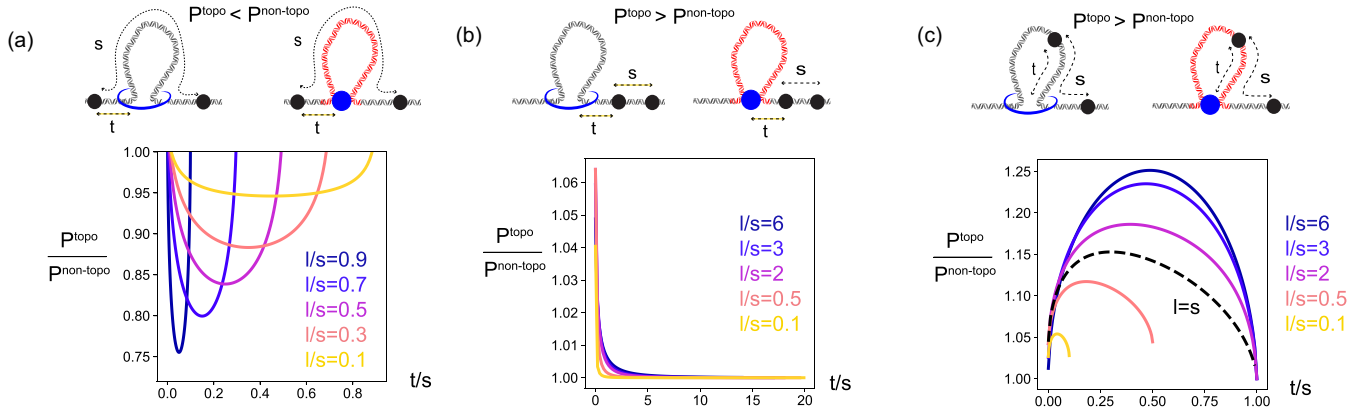


FIG. 4. Difference between the contact statistics of the two models for different diagrams and  $d_f = 3$ . On all panels  $t$  is the distance between the left point and the loop base. (a) For diagram (a1) with one loop in between the points the topological model yields a smaller contact probability than the nontopological. (b) For diagram (a0) without loops in between the topological model has a slightly larger  $P^{\text{topo}}$  than  $P^{\text{nontopo}}$  for the points located close to the loop base. (c) For diagram (b) the topological model yields a larger contact probability, especially when the loop size is larger than the contour distance between the points,  $s < l$ .

### C. Difference in the shape of $P(s)$ for the two models

Having the results for the two models, it is important to see how the residual correlations in the topological binding mechanism change the contact probability across scales. To understand it, in what follows we will analyze their impact for the main diagrams (a) and (b). Diagram (c) is not relevant for us here, since it provides equivalent results for both models. Diagram (d) is not included in the one-loop approximation, which is used to compute the topological model. However, at large contour distances  $s$  it can be approximated by the behavior of diagram (a). Indeed, contribution of the backbone segment in Eq. (10) becomes dominant at  $s \gg \lambda$ .

Diagram (a) consists of two diagrams: (a0) and (a1). Figure 4(a) shows that for the two points located on the backbone and one loop in between them [diagram (a1)] the contact probability is always larger for the nontopological binding of cohesin,  $P_{a1}^{\text{nontopo}} > P_{a1}^{\text{topo}}$ . Importantly, the difference gets stronger with an increase of the loop size  $l$  in between. The effect is the most pronounced when the two points are located close to the loop base,  $l/s \approx 1$ , i.e., when most of the contour distance is occupied by the loop. The source of this difference is clear. From the viewpoint of a random walk, the time elapsed between the points is smaller for the nontopological model as it effectively goes along the backbone ( $s_{\text{eff}}$ ). In other words, the effective contour distance between the points in diagram (a1) is shortened by formation of the loops in between. In the case of topological binding such shortening is absent for  $H < 1/2$  and the shortest contour distance equals to the original distance  $s > s_{\text{eff}}$ . Therefore, the equilibrium distance between the points is larger in the topological model and the contact probability is smaller.

However, for  $H = 1/2$ , the difference in  $P(s)$  for diagram (a1) between the models vanishes, despite the fact that the shortest path along the polymer in the topological model is larger than in the nontopological. One can demonstrate this as follows. As in the previous section, let us denote by  $i, j$  the two points of interest and by  $k, n$  the base of the loop. The vector  $\vec{r}_{ij} = \vec{r}_i - \vec{r}_j$  can be decomposed as

$$\vec{r}_{ij} = \vec{r}_{ik} + \vec{r}_{kn} + \vec{r}_{nj}, \quad (19)$$

where  $\vec{r}_{kn} = 0$  due to the loop. Since in the case of an ideal chain all the vectors in Eq. (19) are independent, the distance between the points  $i, j$  is controlled by the distances along the backbone

$$(\vec{r}_{ij})^2 = (\vec{r}_{ik})^2 + (\vec{r}_{nj})^2 = (k-i) + (j-n), \quad (20)$$

i.e., the loop is effectively eaten up. Thus, the shortest distance in the nontopological model becomes the effective contour distance in the topological for  $H = 1/2$  for diagram (a1).

Interestingly, for diagrams (b) and (a0) the situation is the opposite [see Figs. 4(b) and 4(c): in the topological model the contact probability is larger,  $P_{a0,b}^{\text{topo}} > P_{a0,b}^{\text{nontopo}}$ , in the case of the nonideal chain. Similarly to diagram (a1), upon increase of the loop size  $l$  the effect gets stronger. However, the physics of the difference between the models in this case is not related with the chain shortening, but with residual correlations that are not destroyed by the formation of a loop via the topological mechanism. Indeed, for example, for diagram (b) residual correlations between a point in the loop and a point in the spacer persist in the topological mode of loop formation. Since these correlations are negative (for  $H < 1/2$ ), the equilibrium distance between the points is shorter than in the nontopological case, where the points fluctuate independently. Accordingly, for the ideal chain statistics ( $H = 1/2$ ) the two points in either model fluctuate independently, and the contact probability is the same. Interestingly, for sufficiently large loops,  $l \gg s$ , the maximal fold difference between the models corresponds to the symmetric positions of the points relative to the loop base,  $t/s = 1/2$  [see Fig. 4(c)]. However, upon decrease of the loop size  $l$  the optimal relative distance to the base drops down to  $t/s = 1/4$  at  $l = s$ , meaning that the point in the spacer should be three times further from the loop base than the one in the loop. Such an asymmetry in the configuration is related with different statistical properties of spacer and loop regions, which becomes noticeable when their contour sizes get similar.

Despite the fact that the topological model can be solved only in the one-loop approximation (i.e., for  $\lambda/g \ll 1$ ), one can understand the qualitative difference between the models for all  $\lambda/g$ . Indeed, the conformational statistics of a loopy

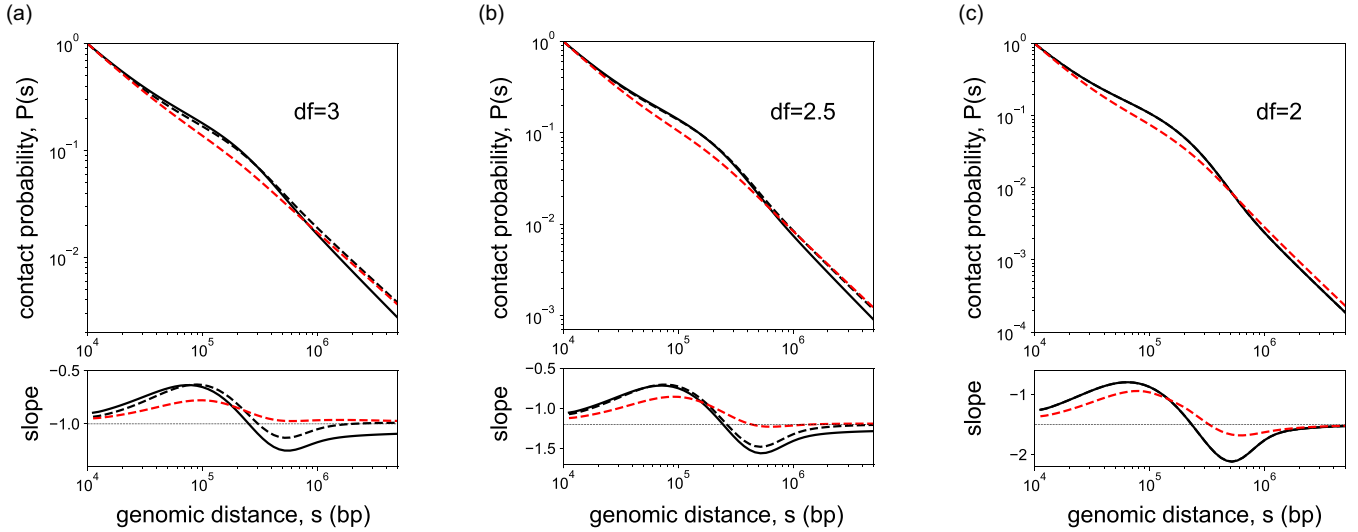


FIG. 5. The normalized contact probability and its log-derivative for the binding model  $P^{\text{nontopo}}(s)$  (red, dashed), for the nontopological in the one-loop approximation  $\mathcal{P}^{\text{nontopo}}(s)$  (black, dashed), and for the topological in the one-loop approximation  $\mathcal{P}^{\text{topo}}(s)$  (black, solid). Loop density is  $\lambda/g = 1$ . Different fractal dimensions of the chain: (a)  $d_f = 3$ , (b)  $d_f = 2.5$ , (c)  $d_f = 2$ . The difference between the topological and nontopological models vanishes as the chain is becoming more ideal. Conversion between the monomers and base pairs can be done by assuming that one monomer contains 1 kb of chromatin length. Accordingly,  $\lambda = g = 100$  kb here. The particular choice of the coarse graining, however, is marginally important, since the amplitudes of the peak and dip, as well as the normalized  $P(s)$ , do not depend on the absolute values of  $\lambda$  and  $g$  (only on their ratio,  $\lambda/g$  and  $s/\lambda$ ) in both models.

chain at  $s \gg \lambda$  is determined by the distance between the points located on the main chain, i.e., by diagram (a), see Fig. 2(a). This distance is effectively shortened in the case of the nontopological binding model, thus leading to a higher contact probability, similarly to the one-loop diagram (a1), see Fig. 4(a). With an increase of the separation  $s$  between the points the fraction of the chain within the backbone (along the shortest path) increases to the bulk value  $\frac{g}{\lambda+g}$ . For the reference loop density  $\lambda/g = 1$ , this corresponds to half of the contour distance effectively shortened. For the case of a single-loop diagram (a1), this yields  $\sim 15\%$  for the difference in the absolute values of  $P(s)$  at large scales [Fig. 4(a)]. Therefore, with an increase of loop density  $\lambda/g$  the drop of  $P^{\text{topo}}$  compared to  $P^{\text{nontopo}}$  at large scales increases, in particular, resulting in a deeper dip of the log-derivative.

On the other hand, at short scales  $s \approx \lambda$  all diagrams depicted in Fig. 4 contribute similarly to  $P(s)$ . While diagram (a1) with one loop in between yields a larger value of  $P(s)$  in the case of nontopological binding, this is not the case for diagrams (b) and (a0). Thus, we expect the difference between the models at short scales to be less pronounced.

This analysis is confirmed by Fig. 5, where for the sake of accuracy we compare the two models in the one-loop approximation  $\mathcal{P}^{\text{topo}}(s)$  and  $\mathcal{P}^{\text{nontopo}}(s)$ , as well as the full theory for the nontopological model,  $P^{\text{nontopo}}(s)$ . For that we use expressions (5)–(7) for the single-loop diagrams in the nontopological case and insert them as the particular contributions for calculation of  $\mathcal{P}^{\text{nontopo}}(s)$ . The averaging of the single-loop diagrams, as for the topological model, is conducted using the linearized weights [Eqs. (S17)–(S20)]. First we note that at intermediate loop densities ( $\lambda/g = 1$ ) the one-loop approximation produces larger amplitudes of the peak and the dip on the log-derivative, for all fractal dimensions  $d_f$ . Interestingly,

the effect of the topological binding at the loop base is qualitatively similar. Indeed, in full agreement with our analysis above, if one compares  $\mathcal{P}^{\text{topo}}(s)$  and  $\mathcal{P}^{\text{nontopo}}(s)$  (black lines in Fig. 5), it is evident that the topological model yields a smaller contact probability at larger scales and, thus, larger amplitude of the dip. The corresponding difference at short scales  $s \leq \lambda$  is much less pronounced than at large scales. Still, one sees that the topological model at short scales produces somewhat larger and slower decaying  $\mathcal{P}^{\text{topo}}$ , especially for larger values of the fractal dimension. Upon decrease of the fractal dimension from  $d_f = 3$  to  $d_f = 2$  the two models in the one-loop approximation give more similar results, consistent with our analysis.

All together, our analysis shows that the main difference between the binding models concerns large scales. With an increase of the loop density, the topological model produces smaller  $P^{\text{topo}}(s)$  than the nontopological  $P^{\text{nontopo}}(s)$  at  $s \gg \lambda$ , resulting in a deeper dip of the log-derivative. Physically, this difference arises from the effective chain shortening that is established by the loops in the nontopological binding model. We note that the predicted deepening of the dip (located at  $s \approx 0.5$ – $1$  Mb) in the topological model should be potentially detectable in Hi-C and Micro-C experiments, where the resolution is limited mostly by the respective capture radius ( $\sim 50$ – $150$  nm for Hi-C [28];  $< 20$  nm for Micro-C [29]).

### III. DISCUSSION

Chromosomes are polymers organized into loops by SMC complexes (e.g., condensins, cohesins) [1]. These loops are transient, as they are produced in a process of loop extrusion of a SMC complex that stochastically binds and unbinds DNA [6]. Therefore, these random loops are not directly visible

in the data (Hi-C or microscopy). Here and in our previous paper [7] we have suggested an analytically tractable polymer model, in which the presence of random loops, as well as their statistical characteristics, are encoded in the specific shape of the average contact probability curve. At the same time, this  $P(s)$  function can also be computed from the experimental Hi-C contact maps [11], providing a robust framework to characterize the loopy state of a chromosome.

Despite the fact that the loop extrusion by cohesin (and, generally, by SMC complexes) has been confirmed in *in vitro* experiments [8,9,30–32], the microscopic mechanism of binding and reeling on DNA remains contradictory. Like other SMC complexes, cohesin has a ring-shaped structure, but whether this ring is actually needed to topologically embrace two strands of DNA in the course of extrusion is a subject of ongoing experimental and computational research [2,10]. Three possible binding mechanisms have been proposed [10]: (i) topological, i.e., the cohesin is bound to one string chemically, while the ring opens and closes when cohesin needs to encircle the second strand of DNA; (ii) pseudo-topological, when both strands are embraced by the ring without chemical binding in the loop base, so that cohesin can be easily pulled away from the chromosome; and (iii) nontopological, when the two DNA strands are chemically bound to the ring, which implies the cohesin ring does not play an important role in binding the DNA. In particular, in mechanisms (i) and (ii) the chromatin strand is reeled into the cohesin ring in the course of extrusion together with the proteins residing on the chain. Clearly, translocation of obstacles larger than the ring size ( $\sim 35$  nm) into the loop would not be possible. However, a recent single-molecule study has shown that nanoparticles as large as  $\sim 200$  nm, as well as big protein machines (e.g., RNA polymerase), can be easily translocated into the loop by cohesin [10]. This experimentally observed ability of cohesin to avoid large obstacles on DNA points to the nontopological binding mechanism. However, other authors have previously noted that the topological binding could explain atypically long residence times of cohesin on chromatin ( $\sim 10$  min), as compared to other molecules without the ring structure [2]. Furthermore, cohesin was previously shown to topologically entrap two sister chromatids [33]. Thus, the role of the cohesin ring structure in DNA binding and extrusion of loops is not yet fully established.

In this paper we underscore the importance of a particular microscopic structure of the cohesin loop base on the three-dimensional organization of a chromosome across scales. Standing on the experimental observation that without cohesin loops chromosomes follow the statistics of a crumpled chain with fractal dimension  $d_f = 3$ , and not of an ideal chain with  $d_f = 2$  [7,16,17], we demonstrate that the conformational properties of a crumpled chain with loops would be different for the topological and nontopological structure of the loop base. While for the nontopological binding the backbone of

a loopy polymer forms with the loops emanating out of it (as in a comb-like polymer), in the topological case the polymer folds into loops, but its organization remains linear. This leads to the two main peculiarities of the two models: (i) effective shortening of the chromosome at large scales in the nontopological model, and (ii) negatively correlated fluctuations of adjacent loops and spacers in the topological model. Notably, for the ideal chain statistics ( $d_f = 2$ ) the correlations along the chain are absent, resulting in equivalent organization of the loopy chain in either binding model. However, for any  $d_f > 2$ , due to strong (power-law) negative correlations, the mode of cohesin binding has a quantitative effect on the three-dimensional organization of the chains into loops.

In order to quantify this effect we used an analytically tractable polymer model [7] and compared the behavior of the contact probability  $P(s)$  for topological and nontopological binding models. We based our theoretical argument on the quenched disorder approach and made several assumptions that allowed us to solve the nontopological model analytically for any fractal dimension  $d_f$  and any density of loops  $\lambda/g$  (see Ref. [7]). Under the same set of assumptions, the topological model is not analytically solvable, so here we employed the one-loop approximation which treats the loops as a perturbation to the fractal scaling of the chain. We obtained the perturbative form of  $P(s)$  in the loop density  $\lambda/g$  and analytically computed the linear order correction using the Gaussian measure of fBm trajectories as suggested recently. This result allowed us to quantitatively compare the contact probability between the two models for the single-loop diagrams [(a)–(c)]. We found that the binding mechanism in the loop base has indeed an impact on the  $P(s)$ , which is qualitatively similar to the difference between the one-loop approximation and the full theory within the single binding mechanism.

The revealed quantitative difference for the two binding models allows to rationalize the organization of chromosomes beyond the minimal analytically tractable models. We note the central role of the nonideal (crumpled) statistics of the chain in the observed difference. Thus, for rare species where the emergent statistics of the chromosomes is closer to the ideal chain (e.g., yeast [34]), our analysis implies universal contact probability behavior independent of the microscopic details imposed by the SMC machinery.

## ACKNOWLEDGMENTS

We thank Leonid Mirny, Sergei Nechaev, Mehran Kardar, Hugo Brandao, Vittore Scolari, and Sergei Belan for valuable discussions on the subject of the paper. The work is supported by the Russian Science Foundation (Grant No. 21-73-00176). K.P. acknowledges hospitality of LPTMS (CNRS-Université Paris-Saclay) and Institute Curie (Paris) where a part of the work has been done.

- [1] L. Mirny and I. Solovei, Keeping chromatin in the loop(s), *Nat. Rev. Mol. Cell Biol.* **22**, 439 (2021).  
 [2] E. J. Banigan and L. Mirny, Loop extrusion: Theory meets single-molecule experiments, *Curr. Opin. Cell Biol.* **64**, 124 (2020).

- [3] A. Goloborodko, J. F. Marko, and L. A. Mirny, Chromosome compaction by active loop extrusion, *Biophys. J.* **110**, 2162 (2016).  
 [4] J. Paulson and U. Laemmli, The structure of histone-depleted metaphase chromosomes, *Cell* **12**, 817 (1977).

- [5] J. H. Gibcus, K. Samejima, A. Goloborodko, I. Samejima, N. Naumova, J. Nuebler, M. Kanemaki, L. Xie, J. R. Paulson, W. C. Earnshaw *et al.*, A pathway for mitotic chromosome formation, *Science* **359**, 2133 (2018).
- [6] G. Fudenberg, N. Abdennur, M. Imakaev, A. Goloborodko, and L. Mirny, Emerging evidence of chromosome folding by loop extrusion, *Cold Spring Harb. Symp. Quant. Biol.* **82**, 45 (2017).
- [7] K. E. Polovnikov, B. Slavov, S. Belan, M. Imakaev, H. B. Brandao, and L. A. Mirny, Crumpled polymer with loops recapitulates key features of chromosome organization, bioRxiv: 10.1101/2022.02.01.478588v3 (2023).
- [8] T. Terakawa, S. Bisht, J. M. Eeftens, C. Dekker, C. H. Haering, and E. C. Greene, The condensin complex is a mechanochemical motor that translocates along DNA, *Science* **358**, 672 (2017).
- [9] M. Ganji, I. A. Shaltiel, S. Bisht, E. Kim, A. Kalichava, C. H. Haering, and C. Dekker, Real-time imaging of dna loop extrusion by condensin, *Science* **360**, 102 (2018).
- [10] B. Pradhan, R. Barth, E. Kim, I. F. Davidson, B. Bauer, van Laar, W. Yang, J. Ryu, van der Torre, J. Peters *et al.*, SMC complexes can traverse physical roadblocks bigger than their ring size, *Cell Rep.* **41**, 111491 (2022).
- [11] E. Lieberman-Aiden, N. L. van Berkum, L. Williams, M. Imakaev, T. Ragozcy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner *et al.*, Comprehensive mapping of long-range interactions reveals folding principles of the human genome, *Science* **326**, 289 (2009).
- [12] L. Mirny, The fractal globule as a model of chromatin architecture in the cell, *Chromosome Res.* **19**, 37 (2011).
- [13] 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, *Rep. Prog. Phys.* **77**, 022601 (2014).
- [14] A. Grosberg, Y. Rabin, S. Havlin, and A. Neer, Crumpled globule model of the three-dimensional structure of DNA, *Europhys. Lett.* **23**, 373 (1993).
- [15] A. Rosa and R. Everaers, Structure and dynamics of interphase chromosomes, *PLoS Comput. Biol.* **4**, e1000153 (2008).
- [16] T. S. Hsieh, C. Cattoglio, E. Slobodyanyuk, A. S. Hansen, X. Darzacq, and R. Tjian, Enhancer-promoter interactions and transcription are maintained upon acute loss of CTCF, cohesin, WAPL, and YY1, *BioRxiv* (2021).
- [17] S. S. P. Rao, S. Huang, B. G. St Hilaire, J. M. Engreitz, E. M. Perez, K. Kieffer-Kwon, A. L. Sanborn, S. E. Johnstone, G. D. Bascom, I. D. Bochkov *et al.*, Cohesin loss eliminates all loop domains, *Cell* **171**, 305 (2017).
- [18] K. Polovnikov, S. Nechaev, and M. V. Tamm, Effective Hamiltonian of topologically stabilized polymer states, *Soft Matter* **14**, 6561 (2018).
- [19] A. Y. Grosberg and A. R. Khokhlov, *Statistical Mechanics of Macromolecules* (AIP, Woodbury, 1994), p. 350.
- [20] P. G. De Gennes, *Scaling Concepts in Polymer Physics* (Cornell University Press, Ithaca, 1979).
- [21] A. Grosberg, S. Nechaev, and E. Shakhnovich, The role of topological constraints in the kinetics of collapse of macromolecules, *J. Phys. France* **49**, 2095 (1988).
- [22] S. Nechaev, A. Semenov, and M. Koleva, Dynamics of a polymer chain in an array of obstacles, *Physica A* **140**, 506 (1987).
- [23] S. Belan and V. Parfenyev, Footprints of loop extrusion in statistics of intra-chromosomal distances: An analytically solvable model, *arXiv:2301.03856*.
- [24] P. Pedler, Occupation times for two-state Markov chains, *J. Appl. Probab.* **8**, 381 (1971).
- [25] K. E. Polovnikov, S. Nechaev, and M. V. Tamm, Many-body contacts in fractal polymer chains and fractional Brownian trajectories, *Phys. Rev. E* **99**, 032501 (2019).
- [26] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevE.107.054135> for the details of computations for the topological model (propagators, linearized weights of the diagrams).
- [27] S. A. Belan and D. E. Starkov, Influence of active loop extrusion on the statistics of triple contacts in the model of interphase chromosomes, *JETP Lett.* **115**, 763 (2022).
- [28] G. Fudenberg and M. Imakaev, FISH-ing for captured contacts: Towards reconciling FISH and 3C, *Nat. Methods* **14**, 673 (2017).
- [29] N. Krietenstein, S. Abraham, S. V. Venev, N. Abdennur, J. Gibcus, T. S. Hsieh, K. M. Parsi, L. Yang, R. Maehr, L. A. Mirny *et al.*, Ultrastructural details of mammalian chromosome architecture, *Mol. Cell* **78**, 554 (2020).
- [30] Y. Kim, Z. Shi, H. Zhang, I. J. Finkelstein, and H. Yu, Human cohesin compacts DNA by loop extrusion, *Science* **366**, 1345 (2019).
- [31] S. Golfier, T. Quail, H. Kimura, and J. Brugués, Cohesin and condensin extrude DNA loops in a cell cycle-dependent manner, *Elife* **9**, e53885 (2020).
- [32] 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).
- [33] D. Ivanov and K. Nasmyth, A topological interaction between cohesin rings and a circular minichromosome, *Cell* **122**, 849 (2005).
- [34] H. Wong, H. Marie-Nelly, S. Herbert, P. Carrivain, H. Blanc, R. Koszul, E. Fabre, and C. Zimmer, A predictive computational model of the dynamic 3D interphase yeast nucleus, *Curr. Biol.* **22**, 1881 (2012).