Topological Spectra and Entropy of Chromatin Loop Networks

Andrea Bonato,[1] Michael Chiang,[2] Dom Corbett,[2] Sergey Kitaev,[3] Davide Marenduzzo,[1] Alexander Morozov,[1] and Enzo Orlandini[4]

[1]*Department of Physics, University of Strathclyde, Glasgow G4 0NG, Scotland, United Kingdom*
[2]*SUPA, School of Physics and Astronomy, The University of Edinburgh, Edinburgh EH9 3FD, Scotland, United Kingdom*
[3]*Department of Mathematics and Statistics, University of Strathclyde, Glasgow G1 1XH, United Kingdom*
[4]*Department of Physics and Astronomy, University of Padova and INFN, Sezione Padova, Via Marzolo 8, I-35131 Padova, Italy*

(Received 29 August 2023; accepted 19 April 2024; published 14 June 2024)

The 3D folding of a mammalian gene can be studied by a polymer model, where the chromatin fiber is represented by a semiflexible polymer which interacts with multivalent proteins, representing complexes of DNA-binding transcription factors and RNA polymerases. This physical model leads to the natural emergence of clusters of proteins and binding sites, accompanied by the folding of chromatin into a set of topologies, each associated with a different network of loops. Here, we combine numerics and analytics to first classify these networks and then find their relative importance or statistical weight, when the properties of the underlying polymer are those relevant to chromatin. Unlike polymer networks previously studied, our chromatin networks have finite average distances between successive binding sites, and this leads to giant differences between the weights of topologies with the same number of edges and nodes but different wiring. These weights strongly favor rosettelike structures with a local cloud of loops with respect to more complicated nonlocal topologies. Our results suggest that genes should overwhelmingly fold into a small fraction of all possible 3D topologies, which can be robustly characterized by the framework we propose here.

DOI: 10.1103/PhysRevLett.132.248403

Within mammalian cells, DNA interacts with proteins called histones to form a composite polymeric material known as chromatin, which is the building block of chromosomes and provides the genomic substrate for cellular processes, such as transcription—the copying of DNA into RNA [1,2]. Understanding the mechanisms underlying and regulating chromatin transcription is important, as these determine the pattern of active and inactive genes in a cell [3]. An important factor linked to transcription is 3D chromatin structure, as DNA elements such as promoters and enhancers often need to come together forming a loop to trigger transcription [2–5]. Within this context, polymer models have provided key insights into chromatin structure and loop formation, and into their link to transcription, concomitantly showing that physical principles may have far-reaching consequences in biology [6–15].

As an example, a simple model for chromatin organization is shown in Fig. 1(a). Here, chromatin is viewed as a semiflexible polymer which interacts with chromatin-binding proteins associated with transcription—such as RNA polymerases and transcription factors. There is a set of binding sites on the chromatin fiber, which have a high affinity for proteins, or protein complexes; the rest of the fiber has a weaker attraction for them, for instance, due to nonspecific or electrostatic interactions. When proteins can bind multivalently, which is generally the case for protein complexes, microphase separation of proteins and binding sites spontaneously emerges through a thermodynamic positive feedback loop, known as the "bridging-induced attraction" [7,8], which works as follows [Fig. 1(a)(ii)]. First, possibly through a fluctuation, the local density of binding sites in 3D may locally increase. This recruits chromatin-binding proteins which, if multivalent, further enhances the concentration of binding sites, in turn increasing protein concentration, and ultimately triggering a positive feedback resulting in the self-assembly of clusters of protein complexes and binding sites. Such clusters are accompanied by the formation of chromatin loops, which incurs an entropic cost growing nonlinearly with the number of loops [16] so that clusters do not coarsen past a typical size, given by the competition between gain in binding energy and loss in entropy. This type of clustering leads to structures very much like the transcription factories —clusters of RNA polymerases and gene promoters or enhancers—observed experimentally in living cells [3,17].

Transcription factories lead to the spontaneous emergence of a network of chromatin loops joining the binding

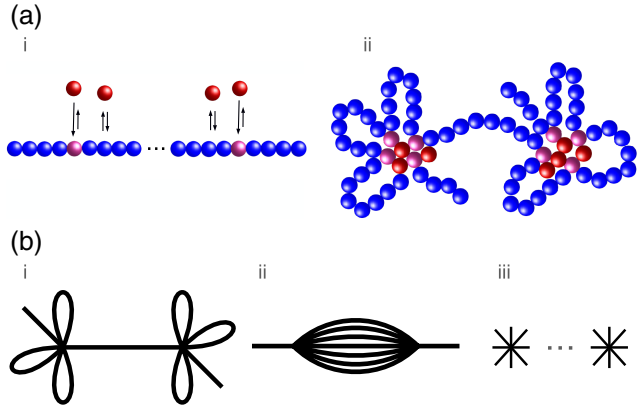

FIG. 1. (a)(i) Schematics of the problem. A chromatin fiber is modeled by a polymer, on which transcription units (pink spheres) are bound by multivalent proteins (red spheres), representing complexes of transcription factors and polymerases. (a)(ii) Emerging states, with clusters arising through the bridging-induced attraction. (b)(i) The chromatin loop network associated with the state in (a)(ii). This is a chain of two rosettes, each containing a cluster four binding sites in the fiber. (b)(ii) An alternative chromatin loop network, known as the watermelon topology. (b)(iii) The local polymer structure at the loop base is the same in (b)(i) and (b)(ii): this is what determines the entropy of the topology in the limit in which the distance between binding sites goes to infinity. In the case of chromatin, though, this limit is not relevant and the weights of the diagrams (b)(i) and (b)(ii) are in practice very different.

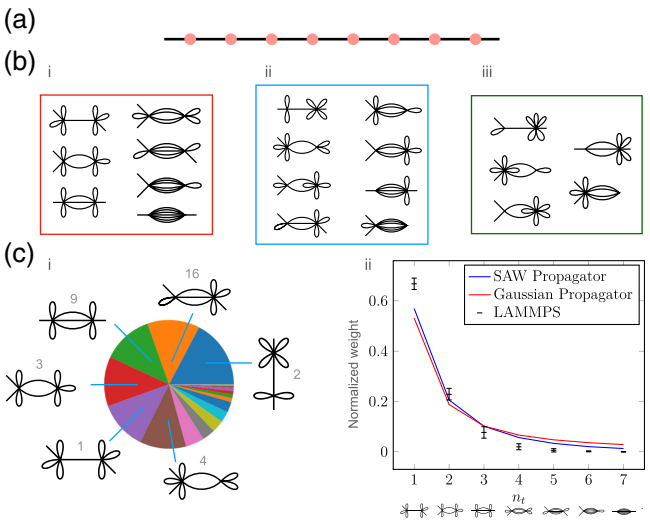

FIG. 2. (a) Setup of our explicit numerical calculation. A gene locus is modeled by a chromatin fiber with $n = 8$ equispaced TUs, with mutual distance $l = 30\sigma = 90$ kilobase pairs [24]. (b)(i)–(b)(iii) Sketch of all possible emerging topologies. There are 20 inequivalent topologies, which are here grouped into three classes (b)(i)–(b)(iii). Network topologies in each of the three classes have the same node degree distributions [Fig. 1(b)(iii)], and hence the same value of $\gamma_{\mathcal{G}}$ [20]. (c) Results of simulations. (c)(i) Pie chart showing the relative frequencies with which the 20 topologies in B are observed. The top six are shown; note the numbers labeling the topologies correspond to their combinatorial multiplicities given in Supplemental Material, Table S1. (c)(ii) Normalized frequency of occurrence of topologies in Bi as a function of n_t. The frequency of the i-th topology is normalized by its combinatorial multiplicity Ω_i (see Table S1).

sites in 3D (Fig. 1). These networks are important as they determine the 3D structure of genes, which as anticipated underlie the transcriptional state of a given cell [15,18]. Networks such as these depicted in Fig. 1(b) were previously considered in polymer physics, for instance, in [19,20], where it was found that their entropy mainly depends on the number of loops and the local number of legs associated with each cluster (Fig. 1(b)(iii)). The relevant thermodynamic ensemble is, however, fundamentally different for the chromatin networks we consider here, as in these cases the typical distance between binding sites l cannot be taken to be arbitrarily large as implicitly done in [19,20], but instead is a model parameter which remains finite and corresponds to a typical ∼50–100 kilobase pair (kbp) chromatin loop [21,22].

In this Letter, we show that the theory of partitions [23] provides a powerful framework to enumerate the topologies of the emerging chromatin loop networks. By performing numerical simulations of the folding of a typical gene locus with the model sketched in Fig. 1, we find that different topologies, with the same entropy in the limit of $l \to \infty$, appear with vastly different frequencies. A striking example of this is provided by the "rosette" and "watermelon" topologies in Fig. 1: despite having the same $l \to \infty$ behavior for their entropic weight, we show that the former is observed orders of magnitude more often than the latter. We resolve this apparent paradox by computing the amplitudes of the statistical weights associated with these diagrams: the ratio between amplitudes of different diagrams follows simple patterns which reflect the biases seen in simulations. We suggest that the topological weights we compute—i.e., the statistical weights of a diagram corresponding to a given topology—are important factors to understand the principles of loop network formation in chromatin, as well as in all polymer systems where binding sites have a typical finite separation between them. Our results can be applied in the future to describe actual 3D chromatin loop topologies observed in experiments or computer simulations.

Here, for concreteness, we focus on a stretch of chromatin with $n = 8$ binding sites, or transcription units [TUs, see Fig. 2(a)]. This case is relevant biologically, as when looking at the genomewide distribution of gene loci with n TUs, the most frequent cases are those with $n \sim 5$–10 [15]. Note we also expect similar results for generic values of n.

To enumerate all possible configurations of a gene locus with $n = 8$, we start by observing that what is important is the relative position of the TUs, which primarily determines the transcriptional activity of the promoter and hence of the gene [15,22]. Equivalently, we need to count all possible

partitions of n (initially distinguishable, or labeled) binding sites into different clusters, which equals the Bell numbers B_n [31]. B_n is a large number: for $n = 8$, $B_8 = 4140$, whereas $B_n \sim n^n$ for large n.

For simplicity, here we further focus on the possible gene locus topologies where the TUs are grouped into $k = 2$ clusters (each of size ≥ 1), without any singletons; different cases (with $k = 3, 4$) are discussed in the companion paper [32] and lead to qualitatively similar results. We call the number of possible configurations with n labeled (or distinguishable) TUs and k clusters $N(n, k)$. For $k = 2$, such a number can be found via the theory of partitions [23], and is given by $N(8, 2) = S(8, 2) - 8 = 119$. Here, $S(8, 2)$ denotes the Stirling number of the second kind [33], which counts the number of ways in which eight points can be partitioned into two nonempty clusters—the subtraction of eight is needed to remove singletons. This reasoning and formula can be generalized, so that the number of configurations of a gene locus with n distinguishable TUs and 2 clusters with ≥ 2 TUs in each is given by

$$N(n, 2) = S(n, 2) - n = 2^{n-1} - n - 1. \tag{1}$$

Additional properties of $N(n, k)$ are discussed in [32].

It is also useful to classify the distinct types of topologies which can be created, where we omit the labeling of the TUs, or equivalently consider them indistinguishable. We call the number of such inequivalent "unlabeled" topologies with n TUs and k clusters $N_u(n, k)$. This classification is relevant if we are interested in calculating the relative importance, or statistical weights, of different loop network topologies, without worrying about the detailed labeling. For two-cluster networks, we find that there are 20 such inequivalent unlabeled topologies [Fig. 2(b) and Supplemental Material, Table S1 [24]]. In general, $N_u(n, 2)$ can be found analytically and is given by [32]

$$N_u(n, 2) = \frac{n(n-1)}{2} - 2 - \frac{\lfloor \frac{n-1}{2} \rfloor \lfloor \frac{n+1}{2} \rfloor}{2}, \tag{2}$$

where $\lfloor x \rfloor$ denotes the largest integer which is smaller than x (the floor function of x). Clearly, $N_u(n, 1) = 1$, whereas it is very challenging to find $N_u(n, k)$ explicitly for $k > 2$. Asymptotically, $N_u(n, 2) \sim n^2$, therefore the growth rate is much smaller than that of the number of labeled networks, $N(n, 2)$, which grows as $\sim 2^n$ [Eq. (1)].

Each inequivalent topology i is associated with a combinatorial weight, or multiplicity, Ω_i, which counts the number of different labeled networks corresponding to it (see Fig. 2 in [32] for an example of two different labeled configurations corresponding to the same unlabeled topology). These multiplicities are listed in Supplemental Material Table S1 [24]; note that they satisfy the constraint $\sum_{i=1}^{N_u(n,2)} \Omega_i = N(n, 2)$. One way to characterize the different topologies is by counting the number of legs that loop directly back to their vertex (or cluster) of origin, and the number of legs that begin at one vertex and end at a distinct vertex (or cluster). We call these two quantities n_l and n_t, for the number of loops and the number of ties, respectively; we note that $n_t + n_l$ is an invariant (which equals $n - 1$, or 7 in our chosen example).

To quantify the statistical weights, or relative importance, of the different topologies in Fig. 2(b), we simulate, by using coarse-grained molecular dynamics run within the LAMMPS package [34], the behavior of a chromatin fiber with contour length L, monomer size σ and persistence length 3σ [8,22,35], with eight equally separated TUs whose mutual distance is $l = 30\sigma$, interacting with ten multivalent spherical complexes of TFs and polymerases [see Supplemental Material [24] for more details]. Focussing on configurations with two clusters as in the theoretical analysis above, we then compute the topological spectrum of our model chromatin fiber, by computing the frequency of each of these, which is an estimate of its statistical weight. Note that for simplicity in these simulations we do not include weak interactions between proteins and non-TU beads, although we do not expect this simplification to change the qualitative trends we observe.

Our simulations show that, remarkably, only a handful of topologies contribute significantly to the population of possible 3D structures of our model gene locus: accordingly, the top three topologies account for over 40% of the total structures, and the top six for just under 80% [Fig. 2(c)(i)]. Overall, we find that nonlocal topologies with a higher number of intercluster ties n_t are much less likely than rosettes, characterized by clouds of local loops and $n_t = 1$. Figure 2(c)(ii) shows the frequency of the symmetric diagrams in Fig. 2(b)(i) as a function of n_t, normalized with respect to their respective combinatorial multiplicities. These normalized frequencies give the topology-specific weight of each configuration, which we refer to as topological weights. Our numerical results show that such weights drop sharply with n_t, to the extent that the weight of the rosette structure [Fig. 1(b)(i), $n_t = 1$] is over 2 orders of magnitude larger than that of the watermelon configuration [Fig. 1(b)(i), $n_t = 7$].

The difference in topological weight between rosettes and watermelon is at first sight surprising because the entropic exponent of the two graphs is the same [20]. However, this is not the whole contribution to the topological weight of a graph, $Z_{\mathcal{G}}$, which is the partition function of the graph and which, at least for $\sigma \ll l \ll L$, can be generically written as follows [20]:

$$Z_{\mathcal{G}} \sim A_{\mathcal{G}} \mu^{N_{\mathcal{G}}} l^{\gamma_{\mathcal{G}}-1}. \tag{3}$$

In Eq. (3), μ is the connective constant of the chain [36], which is the same for all networks, $\gamma_{\mathcal{G}}$ is a topology-dependent universal entropic exponent, which is the same

for each of the topologies in one of the three classes highlighted by boxes in Figs. 2(b)(i)–2(b)(iii), whereas $A_{\mathcal{G}}$ is an amplitude, which is in general nonuniversal. In our case, the distance between TU (or local loop size) l cannot become arbitrarily large but is instead a model parameter that is finite: for instance, for human chromosomes, it is ∼50–100 kbp [21,22]. Therefore, the entropic exponent need not be more important than the amplitude to determine the topological weight.

To get more insight into the amplitudes in Eq. (3) for different topologies, we compute the topological weights for the case of a network of phantom random walks, without volume exclusion. This Gaussian network case can be solved analytically (see Ref. [32]), and the weight of a generic graph $\mathcal{G}$ with n TUs is given by

$$Z_{\mathcal{G}} = \int d\mathbf{x}_0 \ldots d\mathbf{x}_{n+1} \delta(\mathcal{G}) \prod_{i=0}^{n} e^{-\frac{3(\mathbf{x}_{i+1}-\mathbf{x}_i)^2}{2l\sigma}}. \tag{4}$$

The term $\delta(\mathcal{G})$ is a product of Dirac δ functions which specifies the network topology [32]. Performing the Gaussian integrals in Eq. (4), we find the ratio of the weights for the rosette to the watermelon topologies to be $7^{3/2}$. More generally, the weights of the topologies in Fig. 2(b)(i), within the Gaussian approximation, are [24,32,37]

$$Z_{n_t} = \frac{Z_{\mathcal{R}}}{n_t^{3/2}}, \tag{5}$$

with $Z_{\mathcal{R}}$ the rosette weight. Therefore, the amplitudes decay with n_t in a power-law fashion inherited from the metric exponent of the random walk.

Equation (5) shows that topologies with nonlocal loops are statistically unlikely and that the rosette topology is more dominant than the watermelon one, exactly as found in our simulations, even though in the latter the difference in magnitude between their respective frequencies is even higher. Indeed, an inspection of the numerical frequencies found in Fig. 2(c)(ii) suggests that the decay is faster than the power-law fit predicted in Eq. (5). This behavior is likely due to the fact that the Gaussian theory outlined above neglects excluded volume. To see this, we note that for a set of n_t self-avoiding walks connecting the two clusters in Fig. 2(b), with position $\mathbf{0}$ and $\mathbf{x}$ the propagator, or probability distribution of having an end-to-end distance $\mathbf{x}$ with a contour length l, becomes

$$p(x \equiv |\mathbf{x}|; l) \sim \left(\frac{x}{R_F}\right)^{n_t g} e^{-n_t\left(\frac{x}{R_F}\right)^{\delta}};$$
$$g = \frac{\gamma - 1}{\nu}; \quad \delta = \frac{1}{1-\nu}, \tag{6}$$

where R_F (which depends on l) is the Flory radius of a self-avoiding walk, and $\gamma \simeq 7/6$ and $\nu \simeq 3/5$ are its entropic and metric exponents [36]. In our case, therefore, $g \simeq 5/18$ and $\delta \simeq 5/2$. By repeating the calculation for the ratio between the weights of the topologies in Fig. 2(b)(i), but now including self-avoidance by using the propagator in Eq. (6), after some algebra we obtain

$$\frac{Z^{\mathrm{SA}}_{n_t}}{Z^{\mathrm{SA}}_{\mathcal{R}}} = \frac{\Gamma\left(\frac{n_t g+3}{\delta}\right)}{n_t^{\frac{n_t g+3}{\delta}} \Gamma\left(\frac{g+3}{\delta}\right)}, \tag{7}$$

where Γ denotes the gamma function, and the label SA stands for self-avoidance. Importantly, Eq. (7) predicts a different functional dependence on n_t with respect to that of the Gaussian network, as $(Z^{\mathrm{SA}}_{n_t}/Z^{\mathrm{SA}}_{\mathcal{R}}) \sim (e^{-(n_t g/\delta)}/\sqrt{n_t})$ for large n_t. By setting $n_t = 7$, we obtain that $Z^{\mathrm{SA}}_{\mathcal{R}}/Z^{\mathrm{SA}}_{\mathcal{W}} \simeq 42.4$, as opposed to $7^{3/2} \simeq 18.5$ for the Gaussian network. In other words, excluded volume interactions within the chromatin fiber sharpen the difference between the topological weights, rendering predictions closer to values observed in simulations. The remaining discrepancy may be due to the fact that Eq. (7) still disregards mutual avoidance between the n_t chains connecting the two clusters.

Do the trends found in the simple model in Fig. 2 transfer to more realistic models for chromatin organization? To address this question, we analyze HiP-HoP simulations [24,38] of 3D structures of human chromatin in lymphoblastoid cells (Fig. 3). The HiP-HoP model accounts for

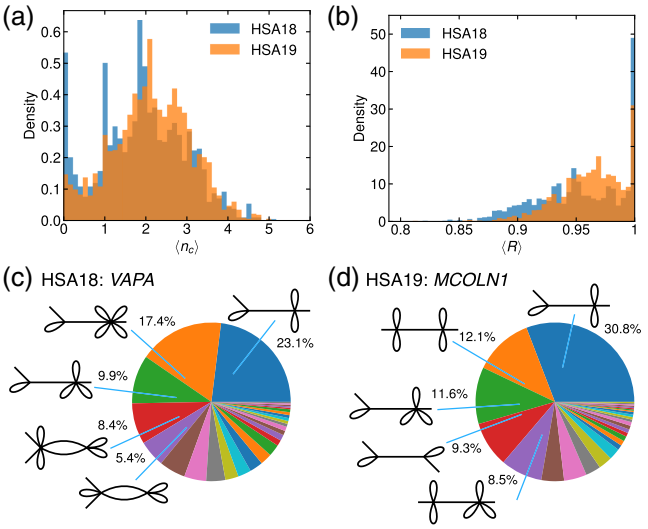

FIG. 3. (a),(b) Probability density of (a) the average number of clusters n_c and (b) rosette score $R = 1 - [(n_t - n_c + 1)/(n_{\mathrm{ns}} - n_c)]$ observed for each TU in HiP-HoP simulations for human chromosomes (HSA) 18 and 19. Here, n_{ns} is the number of nonsingleton TUs, and the average is taken over 600 simulated structures for each TU. A higher R indicates the topology is more dominated by rosettes. (c),(d) Pie charts showing the relative frequencies of observing different looping topologies with two clusters for two gene loci, (c) *VAPA* and (d) *MCOLN1*. Both genes have eight highly interacting TUs (i.e., those that interact with the promoter in > 10% of all sampled structures, including the promoter itself). Note that TUs can be singletons in some of these topologies.

loop extrusion [39], chromatin heteromorphicity [40] and folding by inactive, as well as active, chromatin-binding proteins, and it has been shown to recapitulate well Hi-C contact maps [38]. Each TU—typically a promoter or enhancer—is associated with a chromatin locus, which contains other TUs contacting it frequently. These loci are the analog of the chromatin segment considered in Fig. 2. The 3D folding of loci genomewide shows that (i) typical structures contain $n_c > 1$ clusters of TUs [Fig. 3(a), note the distribution peaks around $n_c \simeq 2$ as in Fig. 2], and (ii) rosettelike structures with local loops are dominant genomewide [Figs. 3(b)–3(d), and Supplemental Material, Figs. S1 and S2]. Interestingly, the predominance of rosettes is more pronounced in gene-rich chromosomes (HSA 19) than in gene-poor ones [HSA 18, Fig. 3(b)]. ATP-dependent loop extrusion subtly affects results, as the rosette score in Fig. 3 slightly increases without it (Supplemental Material, Fig. S1).

In conclusion, we have studied the topological spectrum of chromatin loop networks emerging upon the 3D folding of a gene in a simple polymer model. We have shown that the theory of partitions provides a useful and general way to enumerate the topologies of labeled chromatin fibers, whereas counting unlabeled inequivalent topologies is much more challenging. Our main finding is that, out of all such possible topologies, only a small fraction is in practice observed. Specifically, we predict that gene loci should be overwhelmingly organized in rosettelike structures with a predominant number of local loops between neighboring transcription units [8]. Notably, the dominance of rosettes is also found in more realistic models of human chromatin organization, accounting for adenosine triphosphate (ATP)-dependent processes such as loop extrusion, and for the presence of inactive, as well as active, chromatin.

In the future, it would be of interest to combine our framework with experimental methods, such as SPRITE [41], to classify the networks of chromatin loops formed in different types of cells. From a theoretical viewpoint, the graphs associated with the loop networks we discuss can be mapped into words [42], and it would be fascinating to explore this mapping more fully, which would allow us to view our 3D loop networks as a topological alphabet of chromatin folding.

This work was supported by the Wellcome Trust (223097/Z/21/Z). E. O. acknowledges support from Grant No. PRIN 2022R8YXMR funded by the Italian Ministry of University and Research. The work of S. K. was supported by Leverhulme Research Fellowship (No. RF-2023-065\9).

[1] C. R. Calladine and H. Drew, *Understanding DNA: The Molecule and How It Works* (Academic Press, New York, 1997).
[2] B. Alberts, A. Johnson, J. Lewis, D. Morgan, and M. Raff, *Molecular Biology of the Cell* (Taylor & Francis, London, 2014).
[3] P. R. Cook and D. Marenduzzo, Nucleic Acids Res. **46**, 9895 (2018).
[4] V. Y. Goel, M. K. Huseyin, and A. S. Hansen, Nat. Genet. **55**, 1048 (2023).
[5] Note these enhancer-promoter loops are different from cohesin-dependent loops which have modest effect on transcription; see, e.g., T.-H. S. Hsieh, C. Cattoglio, E. Slobodyanyuk, A. S. Hansen, X. Darzacq, and R. Tjian, Nat. Genet. **54**, 1919 (2022).
[6] M. Barbieri, M. Chotalia, J. Fraser, L.-M. Lavitas, J. Dostie, A. Pombo, and M. Nicodemi, Proc. Natl. Acad. Sci. U.S.A. **109**, 16173 (2012).
[7] C. A. Brackley, S. Taylor, A. Papantonis, P. R. Cook, and D. Marenduzzo, Proc. Natl. Acad. Sci. U.S.A. **110**, E3605 (2013).
[8] C. A. Brackley, J. Johnson, S. Kelly, P. R. Cook, and D. Marenduzzo, Nucleic Acids Res. **44**, 3503 (2016).
[9] L. Giorgetti, R. Galupa, E. P. Nora, T. Piolot, F. Lam, J. Dekker, G. Tiana, and E. Heard, Cell **157**, 950 (2014).
[10] D. Jost, P. Carrivain, G. Cavalli, and C. Vaillant, Nucleic Acids Res. **42**, 9553 (2014).
[11] M. Di Pierro, B. Zhang, E. L. Aiden, P. G. Wolynes, and J. N. Onuchic, Proc. Natl. Acad. Sci. U.S.A. **113**, 12168 (2016).
[12] S. Bianco *et al.*, Nat. Genet. **50**, 662 (2018).
[13] G. Shi and D. Thirumalai, Nat. Commun. **10**, 3894 (2019).
[14] S. Kadam, K. Kumari, V. Manivannan, S. Dutta, M. K. Mitra, and R. Padinhateeri, Nat. Commun. **14**, 4108 (2023).
[15] M. Chiang, C. A. Brackley, C. Naughton, R.-S. Nozawa, C. Battaglia, D. Marenduzzo, and N. Gilbert, bioRxiv https://www.biorxiv.org/content/10.1101/2022.06.09.495447v1 (2022).
[16] D. Marenduzzo and E. Orlandini, J. Stat. Mech. (2009) L09002.
[17] A. Papantonis and P. R. Cook, Chem. Rev. **113**, 8683 (2013).
[18] W. Winick-Ng *et al.*, Nature (London) **599**, 684 (2021).
[19] B. Duplantier, Phys. Rev. Lett. **57**, 941 (1986).
[20] B. Duplantier, J. Stat. Phys. **54**, 581 (1989).
[21] P. R. Cook, *Principles of Nuclear Structure and Function* (Wiley, New York, 2001).
[22] C. Brackley, N. Gilbert, D. Michieletto, A. Papantonis, M. Pereira, P. Cook, and D. Marenduzzo, Nat. Commun. **12**, 5756 (2021).
[23] G. E. Andrews, *The Theory of Partitions* (Cambridge University Press, Cambridge, 1998).
[24] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.132.248403, which includes Refs. [25–30] for details on simulations and example analytical calculations.
[25] A. Rosa and R. Everaers, PLoS Comput. Biol. **4**, e1000153 (2008).
[26] K. Kremer and G. S. Grest, J. Chem. Phys. **92**, 5057 (1990).
[27] C. A. Brackley, B. Liebchen, D. Michieletto, F. Mouvet, P. R. Cook, and D. Marenduzzo, Biophys. J. **112**, 1085 (2017).

[28] A. L. Sanborn *et al.*, Proc. Natl. Acad. Sci. U.S.A. **112**, E6456 (2015).
[29] V. I. Risca, S. K. Denny, A. F. Straight, and W. J. Greenleaf, Nature (London) **541**, 237 (2017).
[30] R. Dreos, G. Ambrosini, R. Groux, R. C. Périer, and P. Bucher, Nucleic Acids Res. **45**, D51 (2017).
[31] M. Aigner, Discrete Math. **205**, 207 (1999).
[32] A. Bonato, M. Chiang, D. Corbett, S. Kitaev, D. Marenduzzo, A. Morozov, and E. Orlandini, companion paper, Phys. Rev. E **109**, 064405 (2024).
[33] B. C. Rennie and A. J. Dobson, J. Comb. Theory **7**, 116 (1969).
[34] S. Plimpton, J. Chem. Phys. **117**, 1 (1995).
[35] J. Langowski, Eur. Phys. J. E **19**, 241 (2006).
[36] P.-G. De Gennes, *Scaling Concepts in Polymer Physics* (Cornell University Press, Ithaca, 1979).
[37] A more general calculation of topological weights for any network, hence applicable to those in Figs. 2(b)(ii) and 2(b)(iii) as well, is given in the companion paper [32].
[38] M. Chiang, G. Forte, N. Gilbert, D. Marenduzzo, and C. A. Brackley, Methods Mol. Biol. **2301**, 267 (2022).
[39] G. Fudenberg, M. Imakaev, C. Lu, A. Goloborodko, N. Abdennur, and L. A. Mirny, Cell Rep. **15**, 2038 (2016).
[40] A. Buckle, C. A. Brackley, S. Boyle, D. Marenduzzo, and N. Gilbert, Mol. Cell **72**, 786 (2018).
[41] S. A. Quinodoz, P. Bhat, P. Chovanec, J. W. Jachowicz, N. Ollikainen, E. Detmar, E. Soehalim, and M. Guttman, Nat. Protoc. **17**, 36 (2022).
[42] S. Kitaev, *Patterns in Permutations and Words* (Springer, New York, 2011).