



## Nonequilibrium Chromosome Looping via Molecular Slip Links

C. A. Brackley,<sup>1</sup> J. Johnson,<sup>1</sup> D. Michieletto,<sup>1</sup> A. N. Morozov,<sup>1</sup> M. Nicodemi,<sup>2</sup> P. R. Cook,<sup>3</sup> and D. Marenduzzo<sup>1</sup>

<sup>1</sup>*SUPA, School of Physics and Astronomy, University of Edinburgh, Peter Guthrie Tait Road, Edinburgh EH9 3FD, United Kingdom*

<sup>2</sup>*Dipartimento di Fisica, Università di Napoli Federico II, INFN Napoli, CNR, SPIN,*

*Complesso Universitario di Monte Sant'Angelo, 80126 Naples, Italy*

<sup>3</sup>*Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, United Kingdom*

(Received 11 April 2017; published 26 September 2017)

We propose a model for the formation of chromatin loops based on the diffusive sliding of molecular slip links. These mimic the behavior of molecules like cohesin, which, along with the CTCF protein, stabilize loops which contribute to organizing the genome. By combining 3D Brownian dynamics simulations and 1D exactly solvable nonequilibrium models, we show that diffusive sliding is sufficient to account for the strong bias in favor of convergent CTCF-mediated chromosome loops observed experimentally. We also find that the diffusive motion of multiple slip links along chromatin is rectified by an intriguing ratchet effect that arises if slip links bind to the chromatin at a preferred “loading site.” This emergent collective behavior favors the extrusion of loops which are much larger than the ones formed by single slip links.

DOI: 10.1103/PhysRevLett.119.138101

**Introduction.**—The formation of long-range contacts, or loops, within DNA and chromosomes critically affects gene expression [1,2]. For instance, looping between specific regulatory elements, such as enhancers and promoters, can strongly increase transcription rates in eukaryotes [1]. The formation of these loops can often be successfully predicted by equilibrium polymer physics models, which balance the energetic gain of protein-mediated interactions with the entropic loss of looping [3–5].

However, recent high-throughput chromosome conformation capture (“Hi-C”) experiments [6,7] have fundamentally challenged the view that equilibrium physics is sufficient to model chromosome looping. These experiments showed that the genomes of most eukaryotic organisms are partitioned into domains, many of which are enclosed within a chromosome loop, 100–1000 kilobasepairs (kbp) in size. The bases of these loops tend to be enriched in binding sites for the CCCTC-binding factor (CTCF) [7,8]. The DNA-binding motif of CTCF is not palindromic, so it has a specific direction along DNA. Surprisingly, Hi-C analyses revealed that most of the CTCF binding sequences only form a loop if they are in a “convergent” orientation [Fig. 1(a)] [7,9]. Very few contacting CTCFs have a “parallel” orientation, and virtually none a “divergent” one. This strong bias is puzzling, because, if we imagine drawing arrows on the chromatin fiber (corresponding to CTCF directionality), then two loops with a pair of convergent or divergent arrows at their base have the same 3D structure [7,10]; hence, they would be equally likely according to equilibrium polymer models. Here we propose a nonequilibrium model that can account for this bias.

In most cases CTCF-mediated loops are associated with cohesin [13], a ringlike protein complex thought to bind DNA by topologically embracing it [14]. There are two

models for how cohesin might achieve this—as a dimer acting as a pair of molecular “handcuffs” in which each ring embraces one DNA duplex [Fig. 1(a)], or as a single ring that embraces two duplexes [15]. In both cases, the dimer or ring acts as a sliding bridge or molecular slip link [16,17]. Experiments show that cohesin topologically links to DNA (with binding mediated by “loader proteins” [13,18]), can slide along DNA or chromatin diffusively, and remains bound for  $\tau \sim 20$  min before dissociating (a process mediated by “unloader proteins”) [18–23].

One recent attempt to address the mechanism underlying CTCF-mediated looping is the “loop extrusion model” which argues that cohesin (or other “loop extruding factors”) can create loops of 100–1000 kbp by actively traveling in opposite directions along the chromosome [24–26]. This model is appealing as it naturally explains the bias in favor of convergent loops, if cohesin gets stuck when it finds a CTCF binding site pointing towards it (but passes over CTCF otherwise). However, the model is based on some assumptions lacking experimental evidence: it requires (i) that each cohesin can determine and maintain the correct direction in order to extrude (rather than shrink) a loop, and (ii) that cohesin must extrude loops at a speed of  $v \sim 5$  kbp/min, which is faster than that of RNA polymerase. While cohesin is known to have ATPase activity, this is not thought to be involved in directional motion; rather, it drives the gate-opening mechanism needed to link to DNA [13].

Here, we propose an alternative model for the formation of CTCF-mediated loops, which does not require unidirectional motion, nor any energetically costly explicit bias favoring loop extrusion. We start from the observation that the molecular topology of cohesin—that of a slip link—is compatible with diffusive sliding along chromatin [19]. From this premise, we formulate a nonequilibrium model

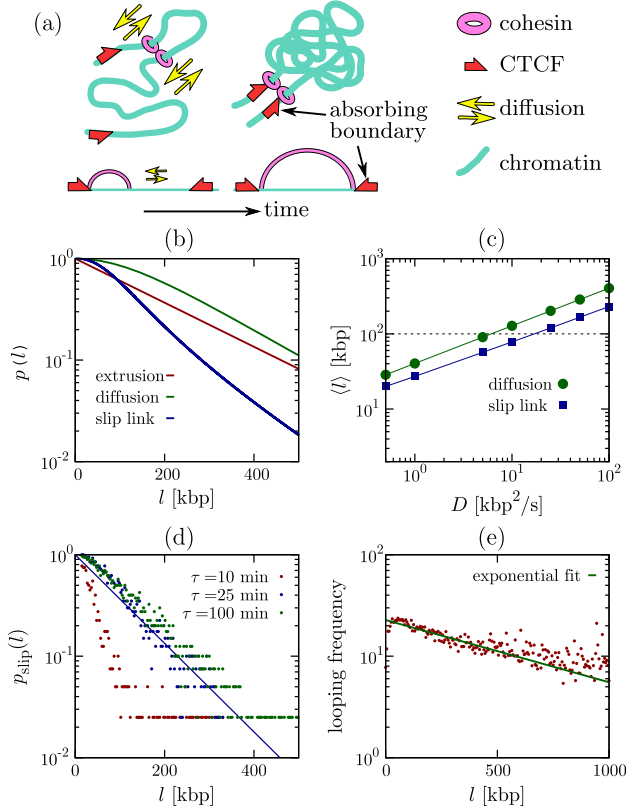


FIG. 1. Nonequilibrium chromosome looping. (a) Schematic of our model of diffusing slip links. (b) Probability of nonequilibrium loop formation in exactly solvable 1D models as a function of loop size  $l$ . Curves correspond to models involving (i) extrusion, (ii) diffusion, and (iii) slip links. Parameters are  $k_{\text{off}}^{-1} = 20$  min and (i)  $v = 10$  kbp/min; (ii), (iii)  $D = 25$  kbp<sup>2</sup>/s; (iii)  $\sigma_{\text{sl}} = 1$  kbp, and  $c = 1$  [11]. (c) Average loop size for models involving diffusion and slip links. Parameters are as in (b), apart from  $D$  which is varied. (d) Nonequilibrium looping probability for a slip link, computed from BD simulations, with different  $k_{\text{off}}^{-1} = \tau$ . The blue line shows an exponential fit for  $k_{\text{off}}^{-1} = 25$  min. (e) Analysis of ChIA-PET experiments for CTCF contacts within a Mbp [12] (log-linear plot).

where the binding and unbinding kinetics of cohesin violates detailed balance, modeling the fact that both its loading and unloading onto chromatin requires ATP [13,27]. We show that within this nonequilibrium context passive sliding is sufficient to account for both the creation of loops of hundreds of kbp before dissociation, and the bias favoring convergent CTCF binding. We further show that many-body interactions between diffusing slip links which bind close to a preferred “loading site” lead to the emergence of an “osmotic ratchet” promoting loop extrusion over shrinking, again in the absence of any bias in the microscopic molecular diffusion.

**Single slip links.**—We begin by discussing an exactly solvable 1D model where a slip link consisting of two cohesin rings in a dimer slides along the chromatin fiber. We assume that this binds with the cohesin rings at adjacent

positions on the fiber (as in [25]), and that there is a constant detachment rate  $k_{\text{off}} = \tau^{-1}$ . Data suggest that cohesin interacts with CTCF in a directional manner [10,25,28], so we assume that slip links bind to CTCF sites which face them, and are reflected off those which do not; further we assume that when the slip link reaches the two convergent CTCF sites it undergoes a conformational change decreasing  $k_{\text{off}}$ . We consider two CTCF proteins bound to the fiber at a separation  $l$  in a convergent orientation. [The case of a divergent pair is treated in [29], and as expected leads to no stable looping (Fig. S1 [29]).] For simplicity, we allow the rings forming one cohesin to diffuse until their separation reaches  $l$ , or until the dimer spontaneously unbinds, and consider both to be absorbing states. This is a nonequilibrium model as the binding-unbinding kinetics violate detailed balance, in line with experimental evidence that ATP is required for both [13,27].

At time  $t$ , the slip link holds together a chromatin loop of size  $x(t)$ . In order to take into account the entropic loss associated with this loop, we include an effective thermodynamic potential  $V(x)$  (detailed below). The probability that the cohesin holds a loop of size  $x$  at time  $t$ , obeys the following generalized Fokker-Plank equation:

$$\frac{\partial p(x,t)}{\partial t} = -k_{\text{off}} p(x,t) + \frac{\partial}{\partial x} \left( \frac{1}{\gamma} \frac{dV}{dx} p(x,t) \right) + D \frac{\partial^2}{\partial x^2} p(x,t), \quad (1)$$

where  $D$  and  $\gamma$  are the effective diffusion and drag coefficients describing the relative motion between chromatin and cohesin. The fluctuation-dissipation theorem implies  $D = k_B T / \gamma$ . The initial condition for Eq. (1) is  $p(x,0) = \delta(x - \sigma_{\text{sl}})$ , where  $\sigma_{\text{sl}}$  is the size of the slip link. Boundary conditions are reflecting at  $x = \sigma_{\text{sl}}$  and, for simplicity, absorbing at  $x = l$ —replacing the latter with an attractive interaction between CTCF and cohesin does not affect our results (Fig. S1 [29]).

We consider three possible cases. First, we model “loop extrusion” as in [24,25,40] by setting  $D = 0$  and  $(1/\gamma)(dV/dx) = -v$ , with  $v$  the extrusion speed. Second, we consider a “diffusion” model where cohesin diffuses in the absence of a potential,  $V = 0$ . Third, we consider a cohesin dimer diffusing in a potential  $V(x) = ck_B T \log(x)$ , which models the entropic cost of looping via the known contact probability  $p_{\text{eq}}(x) \sim x^{-c}$ . Here  $c$  is a universal exponent: in 3D,  $c = 1.5$  for random walk loops [41],  $c \sim 2.1$  for internal looping within self-avoiding chains [41,42], and  $c = 1$  for contacts within a “fractal globule” [11]. We refer to the case with a logarithmic potential as the “slip link” model, as it most closely resembles the dynamics of slip links on polymers [4,16,17].

As detailed in [29], we can analytically find the probability that a cohesin dimer binding at  $t = 0$  will, at some point, form a CTCF-mediated loop before detaching.

Denoting this probability by  $p(l)$ , the three models predict the following dependence on loop size  $l$  [Fig. 1(b)]:

$$p_{\text{extr}}(l) = e^{-k_{\text{off}}l/v}; \quad p_{\text{diff}}(l) = \frac{1}{\cosh(\alpha l)},$$

$$p_{\text{slip}}(l) = \left(\frac{l}{\sigma_{\text{sl}}}\right)^n \frac{I_{m-1}(\alpha l)K_m(\alpha l) + I_m(\alpha l)K_{m-1}(\alpha l)}{I_{m-1}(\alpha \sigma_{\text{sl}})K_m(\alpha l) + I_m(\alpha l)K_{m-1}(\alpha \sigma_{\text{sl}})}, \quad (2)$$

where  $\alpha = \sqrt{k_{\text{off}}/D}$ ,  $n = (1 - c)/2$ , and  $m = (1 + c)/2$ ;  $I$  and  $K$  denote the modified Bessel functions of the first and second kind, respectively. Note that we have taken the  $\sigma_{\text{sl}} \rightarrow 0$  limit for the loop extrusion [ $p_{\text{extr}}(l)$ ] and diffusion [ $p_{\text{diff}}(l)$ ] cases.

For large  $l$ , Eqs. (2) predict exponential decay of CTCF-mediated looping probabilities for all cases [Fig. 1(b)], with a power law correction for slip links,  $p_{\text{slip}}(l) \sim e^{-\alpha l} l^{-c/2}$ . This is markedly different from the power laws determining the looping probability of an equilibrium polymer [16,17]. The decay length is  $v/k_{\text{off}}$  for the loop extrusion model [25], and  $\alpha^{-1} = \sqrt{D/k_{\text{off}}}$  for the diffusion and slip link models; these are therefore the typical looping lengths formed before cohesin detaches. CTCF-mediated loop lengths *in vivo* are typically  $\sim 100$  kbp [7,10]; taking  $\tau = 20$  min means loop extrusion is viable if  $v > 5$  kbp/min, whereas the diffusion or slip link models require  $D > 10$  kbp<sup>2</sup>/s [Fig. 1(c); see Conclusions and [29] for a discussion of the likely *in vivo* value of  $D$ ].

Our 1D theory does not account for the motion of the chromatin fiber, or for the coupling between instantaneous polymer conformation in 3D and slip link diffusivity; to account for these aspects, we also present results from 3D Brownian dynamics (BD) simulations [29]. We modeled a chromatin fiber as a bead-and-spring polymer with bead diameter  $\sigma = 30$  nm, compaction  $C = 100$  bp/nm, and persistence length  $l_p = 4\sigma$  [43]; cohesin slip links were modeled by two rigid rings (each with diameter  $2R \sim 3.5\sigma$ , and thickness  $\sigma_{\text{sl}} = \sigma$ , while  $D \sim 5$  kbp<sup>2</sup>/s [29]). Each ring embraces the fiber, and the two rings are linked via a semiflexible hinge, favoring a planar handcuff configuration with the center of the rings a distance  $2R$  apart (Fig. S5 [29]). Figure 1(d) shows the *nonequilibrium* looping probability  $p_{\text{slip}}(l)$  for different values of  $k_{\text{off}}$  found in these 3D simulations. The results confirm our 1D model predictions that large loops can form via diffusive sliding—e.g., a 100 kbp loop forms with probability  $\sim 0.3$  if  $k_{\text{off}}^{-1} = \tau = 25$  min (see also [29], Figs. S7, S8, and Movie 1). As in the 1D models, the decay of  $p_{\text{slip}}(l)$  is exponential [Fig. 1(d)].

Hi-C experiments measuring the frequency of contacts between all genomic loci largely support a power law decay of contact probability [44]. However, that analysis does not distinguish between CTCF-mediated loops and other contacts [43–46]. Chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) experiments [10] are able to single out contacts where both anchor points are bound to a

protein of interest. Intriguingly, in CTCF ChIA-PET data [12], fitting to an exponential leads to reasonable decay lengths (loop size) of  $\sim 500$ – $1000$  kbp [Fig. 1(e)], whereas fitting to a power law yields an effective exponent which is far from those expected from equilibrium polymer physics [Fig. S11(a) [29]].

**Multiple slip links and the osmotic ratchet.**—So far, we have considered a single slip link. When multiple slip links coexist on the same chromatin segment, they may interact either sterically or entropically. To quantify how this affects loop formation, we performed simple 1D simulations capturing the stochastic dynamics of each side (monomer) of  $N$  slip links which interact solely via excluded volume, and diffuse along a chromatin fiber of size  $L$  discretized into segments of length  $\sigma_{\text{sl}}$  [Figs. 2(a)–2(c), and [29]]. Each slip link can exist in an unbound or chromatin-bound state with binding and unbinding rates  $k_{\text{on}}$  and  $k_{\text{off}}$ , respectively. When binding, the two slip link monomers always occupy neighboring sites along the fiber. In [29], we present a model which also includes a “looping weight” (Figs. S3 and S4), accounting for the entropy of a loop network [16,17]. This effective potential has a quantitative effect but does not modify the qualitative trends; hence, we report here results from the simpler case without the weight.

We consider two cases: (i) with slip links binding at random (unoccupied) locations on the fiber, and (ii) with binding occurring at a preferred “loading site.” Figure 2(a) shows the time average of the maximal loop size  $\langle l_{\text{max}} \rangle$  in a steady state as a function of  $N$  for the first case. As the fiber gets more crowded, the slip links form consecutive loops [Fig. 2(a), inset, and Fig. S3(c)] competing with each other. Consequently,  $\langle l_{\text{max}} \rangle$  decreases steadily with  $N$  [Fig. 2(a)].

A strikingly different result is found when slip links always bind at the same location. This scenario mimics the experimental finding that linking of cohesin to DNA is facilitated by a loader protein (e.g., Scc2 or NIPBL), which has preferential binding sites within the genome [1,18,23]. In this case, we observe that the maximum loop size *increases* with  $N$  [Fig. 2(b)], favoring loop growth over shrinking. Therefore, the system now works as a ratchet, rectifying the diffusion of the two ends of the loop subtended by a slip link. The typical loop network found in a steady state is different from the case of random rebinding, and entails a significant proportion of nested loops [6 out of 11 in Fig. 2(b), inset, and Fig. S3(d)], which reinforce each other. Figure 2(c) shows the probability distribution of sizes for the largest loop and confirms the dramatic difference between the cases with and without loading. We also performed BD simulations of a chromatin fiber interacting with  $N$  slip links which can bind and unbind, with a loading site [29]. These 3D simulations confirm the ratchet effect, and show that the outer loops can easily span hundreds of kbp [Fig. 2(d); Fig. S8 [29]] even with as few as  $N = 3$  slip links. This ratchet effect may provide a microscopic basis for the loop extrusion model in



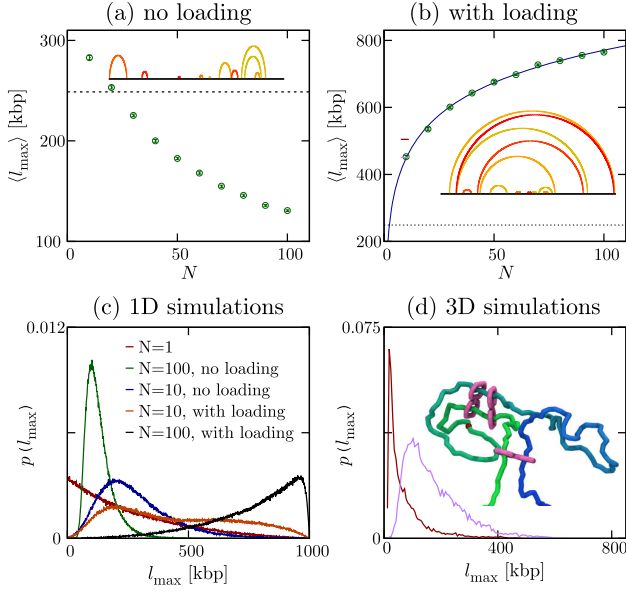


FIG. 2. Multiple slip links and the osmotic ratchet. (a),(b) Results from 1D simulations of diffusing slip links rebinding either (a) randomly, or (b) at a loading site. Plots show the time average of the largest loop (for the case with looping weight see [29], Fig. S4). Parameters are  $\sigma_{\text{sl}} = 1$  kbp,  $L = 1000$  kbp,  $k_{\text{on}}^{-1} = k_{\text{off}}^{-1} = 25$  min, the diffusion coefficient of a monomer is  $D \sim 33.35$  kbp<sup>2</sup>/s, while  $N$  is varied. There are reflecting boundary conditions at the two ends of the fiber. Typical configurations for  $N = 20$  are shown as insets, as “looping diagrams” showing the loop network [29]. The dotted line in (a),(b) denotes the average loop size with a single slip link; the solid line in (b) is a fit to  $a + b \log N$  (see text). (c) Probability distribution of the largest loop size for different  $N$  with or without loading (for the case with looping weight see [29], Fig. S4). (d) Results from 3D BD simulations of multiple slip links with loading, for  $L = 3000$  kbp,  $k_{\text{off}}^{-1} = 25$  min. The plots show the probability distribution of the size of the largest loop for  $N = 1$  (with  $k_{\text{on}} \rightarrow \infty$ ), and  $N = 3$  (with  $k_{\text{on}} = 10k_{\text{off}}$ , snapshot shown as an inset).

[24–26], valid under conditions where several cohesins (or other slip links) are bound to the same chromatin region.

While the simulations discussed thus far were done in dilute conditions, simulations of slip links moving on chromatin fibers at physiological concentrations reach qualitatively similar conclusions (Fig. S9 [29]). The chromatin contact patterns arising from 1D or 3D simulations are also reminiscent of intra-TAD contact maps observed by Hi-C (Figs. S9, S10 [29]), although *in vivo* we expect other mechanisms besides cohesin-mediated looping to contribute to contact formation [43,44,46].

To understand the emergence of a self-organized ratchet, we construct a simple theory by analyzing the 1D model without looping weight (see [29] for more details). The key factor is the existence of a nonuniform slip link density  $\rho(x)$ , and hence an osmotic pressure; the associated gradient creates a force that rectifies the motion of cohesin rings placed close to the loading site. If volume exclusion

does not significantly affect the density and pressure profiles (an assumption which holds in our 1D stochastic simulations, Fig. S2 [47]), we can write down the following phenomenological equation determining the size of a loop,  $l$ , subtended by a symmetrically progressing slip link starting from the loader:

$$\frac{dl}{dt} = -2D\sigma_{\text{sl}} \left[ \frac{\partial \rho}{\partial x} \right]_{x=l/2} = k_{\text{on}} N_{\text{off}} \sigma_{\text{sl}} e^{-al/2}, \quad (3)$$

where  $N_{\text{off}} = Nk_{\text{off}}/(k_{\text{on}} + k_{\text{off}})$  is the average number of unbound cohesins. The maximal speed of this “osmotic ratchet” is achieved for loops close to the loading site. Equation (3) predicts that at a given time,  $l$  should grow logarithmically with  $N$ , and our data are indeed fitted well by the functional form  $a + b \log N$  [Fig. 2(b)].

**Conclusions.**—In summary, we proposed a dynamical model through which molecular slip links might organize chromosomal loops. First, we showed that diffusive sliding of cohesin [18,48,49] naturally explains the experimentally observed bias favoring convergent over divergent CTCF loops. Second, the probability of formation of cohesin/CTCF-mediated loops does not obey a power law, in stark contrast with the case of polymer loops in thermodynamic equilibrium. Finally, we found that when multiple slip links bind to chromatin at a “loading site” rather than randomly, a ratchet effect arises, which favors the formation of much larger loops than are possible with single slip links. Each of these results critically depends on our assumption that the cohesin binding kinetics violate detailed balance, which is motivated by the fact that its loading and unloading requires ATP.

An important consequence of our work is that it predicts which values of the 1D diffusion coefficient,  $D_0$  (in  $\mu\text{m}^2/\text{s}$ ) and chromatin compaction,  $C$  (in bp/nm), are needed for slip links to form CTCF-mediated loops of hundreds of kbp, as found in mammalian genomes [7]. For a single slip link, we need  $D = D_0 C^2 > 10$  kbp<sup>2</sup>/s; for multiple slip links due to the ratchet effect a substantially smaller  $D$  suffices. The worst-case scenario occurs for the least compact structure, a 10-nm chromatin fiber with  $C \sim 20$  bp/nm: this requires  $D_0 > 0.025 \mu\text{m}^2/\text{s}$ . Recent experiments in *Xenopus* egg extracts [49] found  $D_0 = 0.2525 \pm 0.0031 \mu\text{m}^2/\text{s}$  for acetylated cohesin on chromatin [49], comfortably fulfilling the requirement (see [29] for a more quantitative discussion of these and other experiments). Those experiments were performed on a stretched fiber, whereas *in vivo* cohesin dimers are associated with folded chromatin and need to work in a crowded nucleoplasm. Our BD simulations suggest that long enough loops can still be formed when these aspects are taken into account. We hope that our work will prompt new studies to measure diffusion of multiple cohesins on reconstituted chromatin fibers. Particularly our model shows that an extrusion mechanism could arise without the previously

proposed motor activity, and it is important that future experiments are designed to discriminate between these models.

Research outputs generated through the EPSRC Grant No. EP/I004262/1 can be found at [50].

This work was supported by ERC (CoG 648050, THREEDECELLPHYSICS), by IS CRA Grants No. HP10CYFPS5 and HP10CRTY8P, and by the Einstein BIH Fellowship Award to M. N.

C. A. B. and J. J. contributed equally to this work.

- 
- [1] B. Alberts, A. Johnson, J. Lewis, D. Morgan, and M. Raff, *Molecular Biology of the Cell* (Garland Science, New York, 2014).
- [2] S. Chambeyron and W. A. Bickmore, *Curr. Opin. Cell Biol.* **16**, 256 (2004).
- [3] D. Marenduzzo, C. Micheletti, and P. R. Cook, *Biophys. J.* **90**, 3712 (2006).
- [4] A. Hanke and R. Metzler, *Biophys. J.* **85**, 167 (2003).
- [5] J. M. G. Vilar and L. Saiz, *Phys. Rev. Lett.* **96**, 238103 (2006).
- [6] J. Dekker, K. Rippe, M. Dekker, and N. Kleckner, *Science* **295**, 1306 (2002).
- [7] S. S. P. Rao, M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov, J. T. Robinson, A. L. Sanborn, I. Machol, A. D. Omer, E. S. Lander *et al.*, *Cell* **159**, 1665 (2014).
- [8] J. R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, and B. Ren, *Nature (London)* **485**, 376 (2012).
- [9] E. de Wit, E. S. M. Vos, S. J. B. Holwerda, C. Valdes-Quezada, M. J. A. M. Versteegen, H. Teunissen, E. Splinter, P. J. Wijchers, P. H. L. Krijger, and W. de Laat, *Mol. Cell* **60**, 676 (2015).
- [10] M. Oti, J. Falck, M. A. Huynen, and H. Zhou, *BMC Genomics* **17**, 252 (2016).
- [11] L. A. Mirny, *Chromosome Res.* **19**, 37 (2011).
- [12] Z. Tang *et al.*, *Cell* **163**, 1611 (2015).
- [13] F. Uhlmann, *Nat. Rev. Mol. Cell Biol.* **17**, 399 (2016).
- [14] P. J. Huis in 't Veld, F. Herzog, R. Ladurner, I. F. Davidson, S. Piric, E. Kreidl, V. Bhaskara, R. Aebersold, and J. M. Peters, *Science* **346**, 968 (2014).
- [15] K. Nasmyth, *Nat. Cell Biol.* **13**, 1170 (2011).
- [16] R. Metzler, A. Hanke, P. G. Dommersnes, Y. Kantor, and M. Kardar, *Phys. Rev. E* **65**, 061103 (2002).
- [17] D. Michieletto, *Soft Matter* **12**, 9485 (2016).
- [18] J. Stigler, G. Çamdere, D. E. Koshland, and E. C. Greene, *Cell Rep.* **15**, 988 (2016).
- [19] M. T. Ocampo-Hafalla and F. Uhlmann, *J. Cell Sci.* **124**, 685 (2011).
- [20] D. Gerlich, B. Koch, F. Dupeux, J. M. Peters, and J. Ellenberg, *Curr. Biol.* **16**, 1571 (2006).
- [21] R. Ladurner, V. Bhaskara, P. J. H. in 't Veld, I. F. Davidson, E. Kreidl, G. Petzold, and J. M. Peters, *Curr. Biol.* **24**, 2228 (2014).
- [22] A. S. Hansen, I. Pustova, C. Cattoglio, R. Tjian, and X. Darzacq, *eLife* **6**, e25776 (2017).
- [23] G. A. Busslinger, R. A. Stocsits, P. van der Lelij, E. Axelsson, A. Tedeschi, N. Galjart, and J.-M. Peters, *Nature (London)* **544**, 503 (2017).
- [24] E. Alipor and J. F. Marko, *Nucleic Acids Res.* **40**, 11202 (2012).
- [25] G. Fudenberg, M. Imakaev, C. Lu, A. Goloborodko, N. Abdennur, and L. A. Mirny, *Cell Rep.* **15**, 2038 (2016).
- [26] A. L. Sanborn, S. S. P. Rao, S.-C. Huang, N. C. Durand, M. H. Huntley, A. I. Jewett, I. D. Bochkov, D. Chinnappan, A. Cutsosky, J. Li *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **112**, E6456 (2015).
- [27] Y. Murayama and F. Uhlmann, *Cell* **163**, 1628 (2015).
- [28] L. Uuskula-Reimand, H. Hou, P. Samavarchi-Tehrani, M. Vietri Rudan, M. Liang, A. Medina-Rivera, H. Mohammed, D. Schmidt, P. Schwalie, E. J. Young *et al.*, *Genome Biol.* **17**, 182 (2016).
- [29] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.119.138101> for more details on analytical calculations and simulations, and for additional simulation results, supplemental figures and movies. The Supplemental Material also includes Refs. [30–39].
- [30] M. R. Evans and S. N. Majumdar, *Phys. Rev. Lett.* **106**, 160601 (2011).
- [31] P. de los Rios and A. Barducci, *eLife* **3**, e02218 (2014).
- [32] S. Plimpton, *J. Comput. Phys.* **117**, 1 (1995).
- [33] J. Langowski, *Eur. Phys. J. E* **19**, 241 (2006).
- [34] K. Kremer and G. S. Grest, *J. Chem. Phys.* **92**, 5057 (1990).
- [35] H. Hajjoul *et al.*, *Genome Res.* **23**, 1829 (2013).
- [36] A. Rosa and R. Everaers, *PLoS Comput. Biol.* **4**, e1000153 (2008).
- [37] A. Rosa, N. B. Becker, and R. Everaers, *Biophys. J.* **98**, 2410 (2010).
- [38] J. Shimada and H. Yamakawa, *Macromolecules* **17**, 689 (1984).
- [39] X. Wang, H. B. Brandao, T. B. K. Le, M. T. Laub, and D. Z. Rudner, *Science* **355**, 524 (2017).
- [40] A. Goloborodko, M. V. Imakaev, J. F. Marko, and L. Mirny, *eLife* **5**, e14864 (2016).
- [41] B. Duplantier, *J. Stat. Phys.* **54**, 581 (1989).
- [42] E. Carlon, E. Orlandini, and A. L. Stella, *Phys. Rev. Lett.* **88**, 198101 (2002).
- [43] C. A. Brackley, J. Johnson, S. Kelly, P. R. Cook, and D. Marenduzzo, *Nucleic Acids Res.* **44**, 3503 (2016).
- [44] 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).
- [45] C. A. Brackley, S. Taylor, A. Papantonis, P. R. Cook, and D. Marenduzzo, *Proc. Natl. Acad. Sci. U.S.A.* **110**, E3605 (2013).
- [46] A. M. Chiariello, S. Bianco, C. Annunziatella, A. Esposito, and M. Nicodemi, *Sci. Rep.* **6**, 29775 (2016).
- [47] Volume exclusion can though still affect the dynamics, so our theory should not be seen as quantitative.
- [48] I. F. Davidson, D. Goetz, M. P. Zaczek, M. I. Molodtsov, P. J. Huis in 't Veld, F. Weissmann, G. Litos, D. A. Cisneros, M. Ocampo-Hafalla, R. Ladurner *et al.*, *EMBO J.* **35**, 2671 (2016).
- [49] M. Kanke, E. Tahara, P. J. Huis in 't Veld, and T. Nishiyama, *EMBO J.* **35**, 2686 (2016).
- [50] DOI: 10.7488/ds/2129.

# Non-equilibrium chromosome looping via molecular slip-links

## Supplementary Material

C. A. Brackley<sup>1,\*</sup>, J. Johnson<sup>1,\*</sup>, D. Michieletto<sup>1</sup>, A. N. Morozov<sup>1</sup>, M. Nicodemi<sup>2</sup>, P. R. Cook<sup>3</sup>, D. Marenduzzo<sup>1</sup>

<sup>1</sup> *SUPA, School of Physics and Astronomy, University of Edinburgh, Peter Guthrie Tait Road, Edinburgh, EH9 3FD, UK,*

<sup>2</sup> *Dipartimento di Fisica, Universita' di Napoli Federico II, INFN Napoli, CNR, SPIN, Complesso Universitario di Monte Sant'Angelo, 80126 Naples, Italy*

<sup>3</sup> *Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford, OX1 3RE, UK*

\* *Equal contributions*

### EXACTLY SOLVABLE NON-EQUILIBRIUM 1D MODELS

In this section we discuss the derivation of the solution of the 1D non-equilibrium model reported in the main text.

Let us denote by  $p(x, t)$  the probability that the two monomers (heads) of a slip-link dimer (cohesin) are at a separation  $x$ , at a time  $t$  after the slip-link binds to the chromatin fiber. The probability distribution  $p(x, t)$  obeys the following generalized Fokker-Planck equation,

$$\frac{\partial p(x, t)}{\partial t} = -k_{\text{off}} p(x, t) + \frac{\partial}{\partial x} \left[ \frac{1}{\gamma} \frac{dV}{dx} p(x, t) \right] + D \frac{\partial^2}{\partial x^2} p(x, t), \quad (\text{S1})$$

where  $k_{\text{off}}$  is the slip-link detachment rate,  $V(x)$  is the potential energy associated with the configuration in which the slip-link monomers hold a chromatin loop of size  $x$ , while  $D$  and  $\gamma$  are respectively the diffusion and drag coefficient for slip-links moving along the chromatin fiber. As usual,  $D$  and  $\gamma$  are related through the fluctuation-dissipation theorem (Stokes-Einstein formula),  $D = k_B T / \gamma$ . Eq. (S1) should be solved with the initial condition that  $p(x, t = 0) = \delta(x - \sigma_{\text{sl}})$ , as we assume the slip-link binds to two adjacent regions of the chromatin fiber. Also, there is a reflecting boundary at  $x = \sigma_{\text{sl}}$ , and an absorbing boundary at  $x = l$ : this is because once the slip-link binds to the convergent CTCF sites we assume that it “clicks” and sticks to them irreversibly (i.e. we assume that a pair of cohesin rings forming a bridge between two CTCF sites is a very stable complex).

The instantaneous probability at time  $t$  that a slip-link with separation  $\sigma_l < x < l$  unbinds from the chromatin fiber is

$$p_{\text{off}}(t) dt = k_{\text{off}} dt \int_{\sigma_{\text{sl}}}^l dx p(x, t). \quad (\text{S2})$$

In our simple analytical model, once the slip-link detaches, it cannot bind again; i.e., this is an absorbing state. Therefore the probability that the slip-link unbinds before reaching the  $x = l$  absorbing state can be found by integrating over all time,

$$P_{\text{off}} = k_{\text{off}} \int_0^\infty dt \int_{\sigma_{\text{sl}}}^l dx p(x, t). \quad (\text{S3})$$

As this is the probability that the slip-link unbinds while its separation is less than  $l$ , and as  $x = l$  is an absorbing state, the probability that the system reaches the  $x = l$  absorbing state is given by

$$p(l) = 1 - P_{\text{off}}. \quad (\text{S4})$$

In other words, the slip-link cannot diffuse indefinitely in a finite 1D segment without either unbinding or reaching the absorbing state at  $x = l$ .

In order to solve these equations it is useful to define the following quantity,

$$Q(x) = \int_0^\infty dt p(x, t). \quad (\text{S5})$$

Note that  $Q(x)$  may be viewed as the Laplace transform of  $p(x, t)$ ,

$$\hat{p}(x, s) = \int_0^\infty dt e^{-st} p(x, t) \quad (\text{S6})$$

computed at  $s = 0$ . By integrating Eq. (S1) over time from  $t = 0$  to  $\infty$ , we find that  $Q(x)$  obeys the following ordinary differential equation

$$-\delta(x - \sigma_{\text{sl}}) = -k_{\text{off}}Q(x) + D\frac{d^2Q}{dx^2} + \frac{d}{dx} \left[ \frac{1}{\gamma} \frac{dV}{dx} Q(x) \right], \quad (\text{S7})$$

where the Dirac-delta function comes from the  $t = 0$  boundary condition. Since  $P_{\text{off}}$  has an absorbing boundary at  $x = l$ , we also have the boundary condition that  $Q(l) = 0$ . The probability of eventually falling off the chromatin fiber is therefore  $P_{\text{off}} = k_{\text{off}} \int_{\sigma_{\text{sl}}}^l dx Q(x)$ . Therefore, the probability of forming a CTCF-mediated loop is equal to

$$p(l) = 1 - k_{\text{off}} \int_{\sigma_{\text{sl}}}^l dx Q(x). \quad (\text{S8})$$

Let us now compute  $p(l)$  for the three cases discussed in the main text. For simplicity, in the active loop extrusion and free-diffusing cohesin model, we will consider from the start the limit  $\sigma_{\text{sl}} \rightarrow 0$ , as a non-zero value of  $\sigma_{\text{sl}}$  is only required for the slip-link model with logarithmic potential.

### Active extrusion model

For the active extrusion model  $D = 0$  and  $(1/\gamma)(dV/dx) = -v$ , where  $v$  is the constant extrusion speed (a value of  $v > 0$  leads to an increase in the size of the loop,  $x$ ), so Eq. (S7) reduces to a first order differential equation in  $Q(x)$ ,

$$-\delta(x) = -k_{\text{off}}Q(x) + vQ'(x), \quad (\text{S9})$$

where  $'$  denotes derivative with respect to  $x$ . The solution is

$$Q(x) = \frac{1}{v} e^{-k_{\text{off}}x/v}, \quad (\text{S10})$$

so that

$$p(l) = e^{-k_{\text{off}}l/v}. \quad (\text{S11})$$

Note that in this case we cannot apply the boundary condition at  $x = l$  as the equation is first order.

### Free diffusing slip-link model

In the free diffusion model  $dV/dx = 0$ , and the equation for  $Q(x)$  is

$$-\delta(x) = -k_{\text{off}}Q(x) + DQ''(x). \quad (\text{S12})$$

The solution in this case is

$$Q(x) = \frac{1}{\sqrt{Dk_{\text{off}}}} \frac{e^{-\alpha x} - e^{-2\alpha l} e^{\alpha x}}{1 + e^{-2\alpha l}}, \quad (\text{S13})$$

where here and in what follows we have defined  $\alpha = \sqrt{k_{\text{off}}/D}$ , as in the main text. Consequently, the probability of forming a CTCF-mediated loop can be found to be

$$\begin{aligned} p(l) &= 1 - \frac{1 - 2e^{-\alpha l} + e^{-2\alpha l}}{1 + e^{-2\alpha l}}, \\ &= \frac{1}{\cosh(\alpha l)}. \end{aligned} \quad (\text{S14})$$

### Slip-link in a logarithmic potential

If the diffusing slip-link is subject to a logarithmic potential,  $V(x) = ck_B T \log x$ , which captures the entropic cost of looping, then the equation for  $Q(x)$  is

$$-\delta(x - \sigma_{\text{sl}}) = -k_{\text{off}}Q(x) + \left[ \frac{aQ(x)}{x} \right]' + DQ''(x), \quad (\text{S15})$$

where we have defined  $a = ck_B T / \gamma$ . The homogeneous version of Eq. (S15) can be written as

$$DQ''(x) + \frac{aQ'}{x} - \frac{aQ}{x^2} - k_{\text{off}}Q = 0. \quad (\text{S16})$$

If we write  $Q = x^n f$ , then  $Q$  is a solution to Eq. (S16) when the function  $f(x)$  solves the following differential equation:

$$x^2 f'' + \left( 2n + \frac{a}{D} \right) x f' - \left[ \frac{k_{\text{off}}}{D} x^2 - (n-1) \left( n + \frac{a}{D} \right) \right] f = 0. \quad (\text{S17})$$

Now, by setting  $n = (1/2) - (a/2D) = (1-c)/2$ , we note that Eq. (S17) can be written in the form,

$$x^2 y'' + xy' - (x^2 - A^2)y = 0, \quad (\text{S18})$$

where  $A$  is a constant: this is the modified Bessel equation. Therefore the general solution of Eq. (S17) can be written in terms of the modified Bessel functions of the first and second kind as follows,

$$f(x) = C_1 I_m(\alpha x) + C_2 K_m(\alpha x), \quad m = \frac{1}{2} + \frac{a}{2D}, \quad (\text{S19})$$

where  $C_1$  and  $C_2$  are constants to be determined from the boundary conditions,  $I_m$  and  $K_m$  respectively denote the order  $m$  modified Bessel functions of the first and second kind. Note that  $n + m = 1$ .

To solve Eq. (S15), we note that it is equivalent to Eq. (S16) with the following boundary conditions:

$$\begin{aligned} Q(l) &= 0 \\ \left[ DQ' + \frac{aQ}{x} \right]_{x=\sigma_{\text{sl}}} &= -1, \end{aligned} \quad (\text{S20})$$

where the second boundary condition comes from integrating Eq. (S15) over an infinitesimal interval containing  $x = \sigma_{\text{sl}}$ . Eqs. (S20) can be used to determine the two constants  $C_1$  and  $C_2$  in Eq. (S19), to obtain

$$C_1 = -\frac{1}{D\alpha\sigma_{\text{sl}}^n} \left[ \frac{K_m(\alpha l)}{I_{m-1}(\alpha\sigma_{\text{sl}})K_m(\alpha l) + K_{m-1}(\alpha\sigma_{\text{sl}})I_m(\alpha l)} \right], \quad (\text{S21})$$

$$C_2 = \frac{1}{D\alpha\sigma_{\text{sl}}^n} \left[ \frac{I_m(\alpha l)}{I_{m-1}(\alpha\sigma_{\text{sl}})K_m(\alpha l) + K_{m-1}(\alpha\sigma_{\text{sl}})I_m(\alpha l)} \right]. \quad (\text{S22})$$

Note that we have used the following identities:

$$\begin{aligned} I'_m(\alpha x) &= I_{m-1}(\alpha x) - \frac{m}{\alpha x} I_m(\alpha x), \\ K'_m(\alpha x) &= -K_{m-1}(\alpha x) - \frac{m}{\alpha x} K_m(\alpha x). \end{aligned} \quad (\text{S23})$$

The solution of Eq. (S15) which satisfies the relevant boundary conditions is then given by

$$Q(x) = \left( \frac{x}{\sigma_{\text{sl}}} \right)^n \frac{1}{D\alpha} \left[ \frac{I_m(\alpha l)K_m(\alpha x) - I_m(\alpha x)K_m(\alpha l)}{I_{m-1}(\alpha\sigma_{\text{sl}})K_m(\alpha l) + K_{m-1}(\alpha\sigma_{\text{sl}})I_m(\alpha l)} \right]. \quad (\text{S24})$$

From this we obtain

$$\begin{aligned} p(l) &= 1 - k_{\text{off}} \int_{\sigma_{\text{sl}}}^l dx Q(x) \\ &= \left( \frac{l}{\sigma_{\text{sl}}} \right)^n \frac{I_{m-1}(\alpha l)K_m(\alpha l) + I_m(\alpha l)K_{m-1}(\alpha l)}{I_{m-1}(\alpha\sigma_{\text{sl}})K_m(\alpha l) + I_m(\alpha l)K_{m-1}(\alpha\sigma_{\text{sl}})}, \end{aligned} \quad (\text{S25})$$



where we have used the following identities:

$$\begin{aligned} \int dx x^n I_m(\alpha x) &= \frac{x^n I_{m-1}(\alpha x)}{\alpha} \\ \int dx x^n K_m(\alpha x) &= -\frac{x^n K_{m-1}(\alpha x)}{\alpha}, \end{aligned} \quad (\text{S26})$$

which hold for indefinite integrals provided that  $n + m = 1$ .

## NON-EQUILIBRIUM 1D MODELS OF A SINGLE SLIP-LINK: STOCHASTIC SIMULATIONS

In this section we consider 1D stochastic simulations of a single slip-link diffusing in a logarithmic potential in the presence of two CTCF proteins, at mutual distance  $l$ , which act as barriers. For simplicity, as in the main text (Fig. 1) we only consider the relative distance between the slip-link monomers,  $x$ , and assume that it performs a random walk in an effective potential, whereas in reality both monomers diffuse and are subject to a potential dependent on the monomer-monomer separation – we expect the two situations to be qualitatively analogous. With respect to the case considered in the main text, we here assume that there are no absorbing states, but rather that the slip-link gains an energy  $\epsilon$  when it reaches a separation between the monomers  $x = l$  (i.e., sticking between CTCF and cohesin is not permanent here, so  $x$  can decrease later on). Correspondingly, the detachment rate will decrease at  $x = l$ : for concreteness, we assume  $k_{\text{off}}$  is constant, and equal to  $k_0$ , for  $x \neq l$ , while it is equal to  $k_{\text{off}} = k_0 e^{-\epsilon/(k_B T)}$  for  $x = l$ . The single cohesin we model, once detached, rebinds at rate  $k_{\text{on}} = k_0$ , and when it does the monomers always start close together, so  $x = \sigma_{\text{sl}}$  (which is equal to the lattice spacing in our simulations). The logarithmic potential is  $V(x) = ck_B T \log x$ , and we choose here  $c = 2.1$  which corresponds to the formation of internal loops in a self-avoiding walk (see discussion in the main text, different values of  $c$  lead to the same qualitative trends). The logarithmic potential and CTCF-cohesin interactions are incorporated in the algorithm via a standard Metropolis acceptance test.

Figure S1 shows a plot of the probability that the slip-link is attached to the chromatin and has  $x = l$  (i.e., the probability that a CTCF-mediated loop forms) once steady state is reached. As might be expected, we find that increasing  $\epsilon$  strongly favours the CTCF-mediated loops, with respect to other states where the slip-link subtends a smaller loop size. This case is instructive because it suggests that a thermodynamic directional attraction between CTCF and cohesin (here, the interaction parametrised by  $\epsilon$ ) is sufficient to favour the formation of CTCF-mediated loops. It should be noted that the model is still a non-equilibrium one, because  $k_{\text{off}}$  is constant for  $x \neq l$ , and, mainly, because upon rebinding the slip-link always returns to the case with  $x = \sigma_{\text{sl}}$ . This second feature renders our model (both here and in the main text) to some extent similar to the diffusion “with resetting” model considered in [1], although here the motion is further constrained by the logarithmic potential. There is also some similarity with the “ultra-affinity” mechanism discussed in Ref. [2], which is also a non-equilibrium process.

Based on our results, we therefore suggest that non-equilibrium (re)binding (i.e., the resetting) and thermodynamic directional attraction are enough to explain the bias favouring the formation of convergent CTCF loops ( $\epsilon \neq 0$ ) with respect to divergent ones (where there is no directional attraction, and hence  $\epsilon = 0$ ). Again, and as in the main text, because this is a non-equilibrium model, the probability of formation of CTCF-loop is not compatible with a power law: rather it decays approximately exponentially (see the log-linear plot in Fig. S1).

## 1D MODELS WITH MANY INTERACTING SLIP-LINKS, AND THE OSMOTIC RATCHET

### 1D model without looping weight

We now consider the case of multiple slip-links studied in the main text, and derive the formula for the density and effective extrusion force in the case where slip-links always rebind at the same “loading site”. This is the case which leads to the osmotic ratchet discussed in the main text. In this section, we consider 1D models (3D simulations are described separately below).

We first consider a simplified model without “looping weights”, where  $N$  slip-links simply diffuse on a chromatin fiber of length  $L$ : i.e., this model neglects the entropic cost associated with the formation of a given loop network.

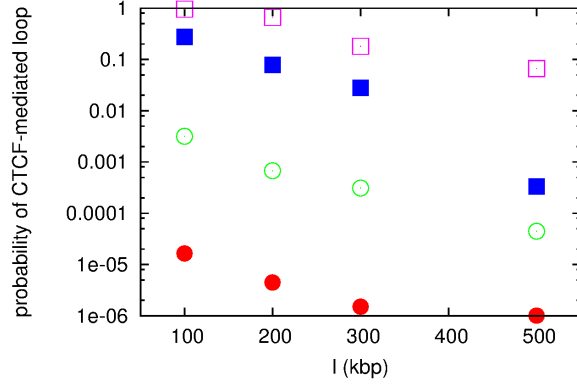


FIG. S1: Plot of the probability of formation of a CTCF-mediated loop as a function of loop size, for different values of  $\epsilon$ . From bottom to top, curves correspond to  $\epsilon = 0, 5k_B T, 10k_B T$  and  $15k_B T$  respectively. The case of  $\epsilon = 0$  models a divergent CTCF loop, the other cases with  $\epsilon \neq 0$  model convergent CTCF loops with different assumptions for the strength of the thermodynamic attraction between CTCF and the slip-link.

If we disregard excluded volume interactions we can write the following partial differential equation for the (average) density  $\rho(x, t)$  of slip-links bound to chromatin at position  $x$ , where the loading site is located at  $x = 0$ ,

$$\begin{aligned} \frac{\partial \rho(x, t)}{\partial t} &= k_{\text{on}} N_{\text{off}} \delta(x) - k_{\text{off}} \rho(x, t) + D \frac{\partial^2 \rho(x, t)}{\partial x^2} \\ &= \frac{k_{\text{on}} k_{\text{off}} N}{k_{\text{on}} + k_{\text{off}}} \delta(x) - k_{\text{off}} \rho(x, t) + D \frac{\partial^2 \rho(x, t)}{\partial x^2}, \end{aligned} \quad (\text{S27})$$

where  $N_{\text{off}} = k_{\text{off}} N / (k_{\text{on}} + k_{\text{off}})$  is the average number of unbound cohesins (which are available to bind at the loading site). The three terms on the right hand side of Eq. (S27) respectively denote binding at the loading site with rate  $k_{\text{on}}$ , unbinding with rate  $k_{\text{off}}$  from any site, and diffusion. Note that here  $D$  is the diffusion constant for a slip-link monomer moving along the chromatin fiber. This equation does not include noise, therefore it should be seen as a mean field theory, which predicts the average value of  $\rho(x, t)$ . The steady state solution of Eq. (S27) which decays for  $x \rightarrow \infty$  (relevant for  $L \rightarrow \infty$ ) is given by

$$\rho(x) = A e^{-\alpha |x|} \quad (\text{S28})$$

where  $A$  is a constant and in a similar way to before we define  $\alpha = \sqrt{k_{\text{off}}/D}$ . Similarly to what was previously shown in the section “Exactly solvable non-equilibrium models”, the constant  $A$  can be determined by integrating Eq. (S27) around 0, from  $x = -\epsilon$  to  $x = +\epsilon$ , and then sending  $\epsilon \rightarrow 0$ . This procedure leads to the requirement that

$$A = \frac{1}{2D\alpha} \frac{k_{\text{on}} k_{\text{off}} N}{k_{\text{on}} + k_{\text{off}}}, \quad (\text{S29})$$

and therefore  $\rho(x)$  in steady state is given by

$$\rho(x) = \frac{N}{2} \frac{k_{\text{on}} k_{\text{off}}}{k_{\text{on}} + k_{\text{off}}} \frac{1}{D\alpha} e^{-\alpha |x|}. \quad (\text{S30})$$

Stochastic 1D computer simulations of  $N$  slip-links diffusing *with* excluded volume interactions on a chromatin fiber of size  $L$  confirm that the average density profile of bound slip-links is an exponentially decaying function centred on the loading site, in good agreement with Eq. (S30) even for a large number of slip-links (Fig. S2).

The 1D pressure exerted by the slip-link gas is equal to  $p(x) = N k_B T \rho(x)$ ; for a given slip-link at position  $x$ , there will be a difference in the pressure on the inside and outside of each head of the link, resulting on an outward force. Since the size of one of the slip-link head is  $\sigma_{\text{sl}}$ , we can estimate the force acting on a head at position  $x$  as follows:

$$\begin{aligned} f(x) &\simeq -k_B T \sigma_{\text{sl}} \frac{\partial \rho}{\partial x} \\ &= \frac{k_B T \sigma_{\text{sl}} N k_{\text{on}} k_{\text{off}}}{2D(k_{\text{on}} + k_{\text{off}})} e^{-\alpha |x|} \end{aligned} \quad (\text{S31})$$

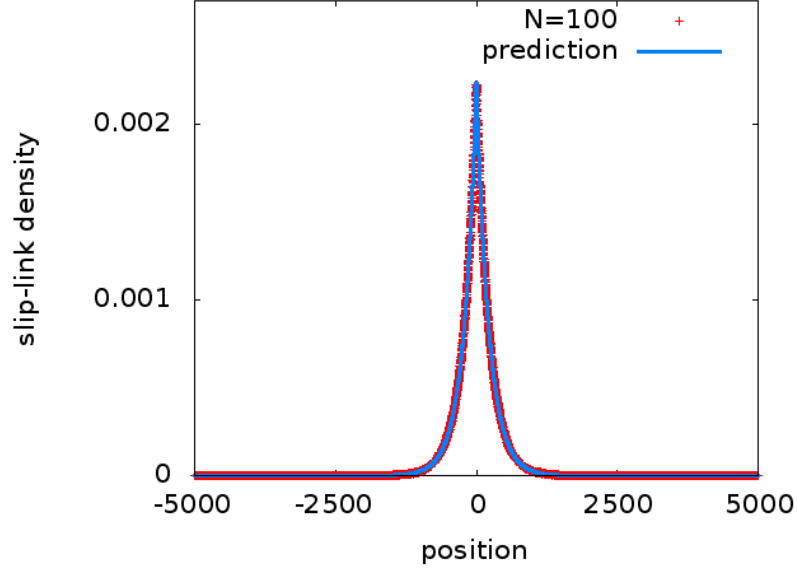


FIG. S2: Plot of the density of slip-link monomers as a function of position, for a chromatin fiber of length  $L = 10000$ , with a loading site in the middle, at  $x = 0$ , and  $N = 100$  slip-links which have  $k_{\text{on}} = k_{\text{off}} = 0.00001$ , and  $D = 0.5$  (all parameters in simulation units). Points show results from a 1D simulation, and the line is given by Eq. S30

If we now imagine a slip-link placed symmetrically around the loading site, so that its two heads are at positions  $\pm x$ , then the osmotic pressure will tend to increase the size of the loop  $l = 2x$ . If we assume for simplicity, that the loop will remain symmetrical with respect to the loading site, we can write down the following equation for the effective extrusion velocity of the loop,  $v = dl/dt$ ,

$$\gamma \frac{dl}{dt} = \frac{k_B T \sigma_{\text{sl}} N k_{\text{on}} k_{\text{off}}}{D(k_{\text{on}} + k_{\text{off}})} e^{-\alpha l/2}, \quad (\text{S32})$$

where  $\gamma$  is the slip-link's effective drag coefficient. Eq. (S32) predicts that the maximal extrusion speed is when the loop is close to the loading site, where it can be approximated as

$$v \sim v(l \rightarrow 0) = \frac{\sigma_{\text{sl}} k_{\text{on}} k_{\text{off}} N}{(k_{\text{on}} + k_{\text{off}})}, \quad (\text{S33})$$

where we have used the fluctuation-dissipation relation  $D = k_B T / \gamma$ . Note that the solution of Eq. (S32) is given by

$$l(t) = \frac{2}{\alpha} \log \left[ 1 + \frac{N k_{\text{on}} k_{\text{off}} \alpha \sigma_{\text{sl}}}{2(k_{\text{on}} + k_{\text{off}})} t \right], \quad (\text{S34})$$

so that this simple theory predicts that extrusion should slow down with loop size, which should only increase logarithmically at later times. Note that Eq. (S34) predicts the average evolution of the loop size for a slip-link binding at the loading site at  $t = 0$ , whereas in Figure 2 in the main text we plot the size and probability distribution of the *largest* loop at a given time. However, this simplified theory is useful as it clarifies that loops can be extruded provided the steady state slip-link density  $\rho(x)$  is not constant. Of course, if there is not a preferred loading site, the first term in Eq. (S27) becomes  $k_{\text{on}} N_{\text{off}} / L$ : in this case  $\rho(x)$  is constant in steady state, and there is no longer an osmotic pressure driving extrusion, in line with the results discussed in the main text for the case with random rebinding.

### 1D Model with looping weight, and looping diagrams

The model discussed above corresponds to the case “without looping weight”. The case “with looping weight” discussed in the main manuscript can be considered in 1D simulations by introducing an entropic potential which

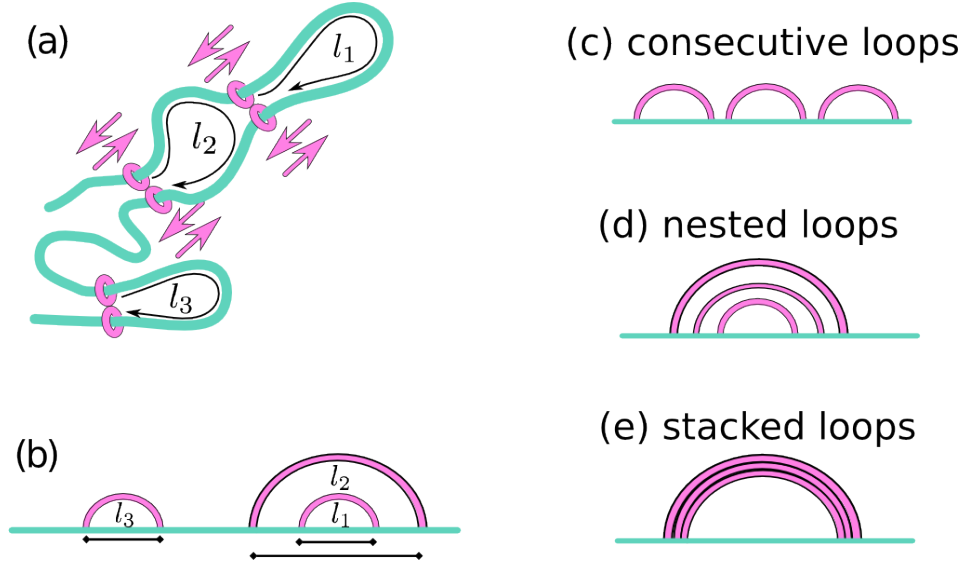


FIG. S3: (a,b) Cartoon of a network of slip-links, and associated loops, viewed either in 3D (a), or 1D (b). In this network, there are three loops:  $l_1$ ,  $l_2$  and  $l_3$ . Of these,  $l_1$  and  $l_3$  are “simple loops”. (c-e) Cartoons of looping diagrams which typically arise in our simulations. The arrangement of slip-links in (c) is referred to as “consecutive loops”, while that in (d) is referred to as “nested loops”. For nested loops, if arcs are close together, this corresponds to a 3D structure where slip-links are closely stacked onto one another: we referred to this pattern as “stacked loops”. If different arcs are coloured differently, stacked arcs appear as rainbow patterns (Suppl. Movies 2 and 3).

affects the motion of slip-link monomers (in practice, this is done through a standard Metropolis test). For simplicity, we assume that all loops are Gaussian, i.e., we disregard self-avoidance effects in this calculation. To compute the looping of a given configuration of slip-link heads (e.g., that in Figs. S3a,b), we first identify all loops. The number of loops,  $n$ , is equal to the number of bound slip-links,  $N_b$ , and we label their sizes  $l_1, \dots, l_n$  (see Figs. S3a,b). We then identify the number of “simple loops”, which do not contain another loop inside. In general, there will be a number  $n_s \leq n$  of simple loops. The probability of formation of each loop is  $\sim l^{-3/2}$ , and this is weighted by another factor  $e^{-\tilde{\kappa}/l^2}$  for simple loops to model the energetic cost of bending;  $\tilde{\kappa}$  is a constant associated with the persistence length of the chromatin fiber. The looping weight is then

$$w_{\text{looping}} = \left( \prod_{i=1, \dots, n} \frac{1}{l_i^{3/2}} \right) \left( \prod_{i=1, \dots, n_s} e^{-\frac{\tilde{\kappa}}{l_i^2}} \right). \quad (\text{S35})$$

This looping weight is defined up to a multiplicative constant, and, in turn, it defines the potential in which the slip-links move (up to an irrelevant additive constant) via

$$V_{\text{looping}} = -k_B T \log w_{\text{looping}}. \quad (\text{S36})$$

In Figure 2 in the main manuscript we present results without looping weight; Figure S4 shows the results of similar simulations but including the looping weight, with  $\kappa \approx 8$  (in units of  $\Delta x^{-2}$ , where  $\Delta x = \sigma_{\text{sm}}/2$ ), a choice corresponding to a rather flexible polymer. The results show that the looping weight makes a notable quantitative change, but the qualitative trends are very similar to those in Figure 2 in the main manuscript, with the model with loading leading to the ratchet effect discussed above and in the main manuscript.

Diagrams such as that in Figure S3 are useful to determine visually the looping topology without the need to show the 3D configuration. Such diagrams are used extensively for RNA secondary structure representations; we refer to these in our context as “looping diagrams”. In the main text and below we refer to some specific loop configurations which are more easily described by these diagrams: these are the “consecutive loop” arrangement in Figure S3c, the “nested loop” one in Figure S3d, and “the stacked loops” of Figure S3e, where some of the loops in a nested loop arrangement are packed close to each other. As discussed below, stacked loops are entropically favoured, hence they appear often in our simulations, particularly if the number of slip-links is large: in Supplementary Movies 2 and 3, where each arc is colored differently, stacked loops appear as rainbow patterns.

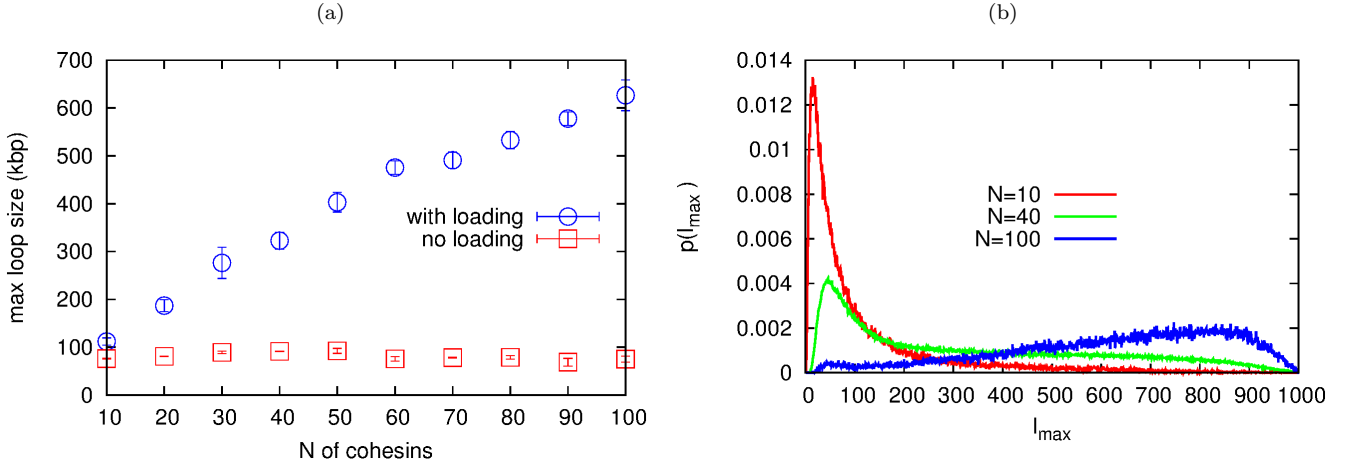


FIG. S4: The osmotic ratchet in 1D simulations with looping weight. (a) Plot of the time average of the largest loop,  $\langle l_{\max} \rangle$ , as a function of  $N$  for the model with and without loading (see legend). (b) Distributions of the largest loop size for different  $N$ , for the model with loading.

### 3D BROWNIAN DYNAMICS OF A CHROMATIN FIBER WITH MOLECULAR SLIP-LINKS

In this section we give details and additional results for the three-dimensional Brownian dynamics (BD) simulations of a slip-link sliding diffusively on a chromatin fiber, which are discussed in the main text.

#### Brownian dynamics: force field and other simulation details

In our BD simulations we follow the evolution of a chromatin fiber and of a slip-link which is topologically bound to the fiber. The dynamics are evolved using a velocity-Verlet integration scheme within the LAMMPS software [3] in Brownian dynamics mode (NVT ensemble).

The chromatin fiber is modelled, as in Ref. [4], as a bead-spring self-avoiding and semi-flexible polymer; each of its beads have size  $\sigma$ . If we denote the position of the centre of the  $i$ -th chromatin bead by  $\mathbf{r}_i$ , the separation vector between beads  $i$  and  $j$  by  $\mathbf{d}_{i,j} = \mathbf{r}_i - \mathbf{r}_j$ , and its modulus by  $d_{i,j} = |\mathbf{r}_i - \mathbf{r}_j|$ , then we can express the finitely-extensible non-linear (FENE) spring potential modelling the connectivity of the chain as follows:

$$U_{\text{FENE}}(i, i+1) = -\frac{k}{2} R_0^2 \ln \left[ 1 - \left( \frac{d_{i,i+1}}{R_0} \right)^2 \right], \quad (\text{S37})$$

for  $d_{i,i+1} < R_0$  and  $U_{\text{FENE}}(i, i+1) = \infty$ , otherwise; here we chose  $R_0 = 1.6 \sigma$  and  $k = 30 \epsilon / \sigma^2$ .

The semi-flexibility (bending rigidity) of the chain is described through a standard Kratky-Porod potential, defined in terms of the positions of a triplet of neighbouring beads along the polymer as follows:

$$U_{\text{B}}(i, i+1, i+2) = \frac{k_B T l_p}{\sigma} \left[ 1 - \frac{\mathbf{d}_{i,i+1} \cdot \mathbf{d}_{i+1,i+2}}{d_{i,i+1} d_{i+1,i+2}} \right], \quad (\text{S38})$$

where in the main text we set the persistence length  $l_p = 4\sigma$  (which maps to  $\simeq 120$  nm – see below; this is reasonable for chromatin [5]).

Self-avoidance is ensured by introducing a repulsive Weeks-Chandler-Anderson (WCA) potential between every chromatin bead as follows:

$$U_{\text{LJ}}(i, j) = 4\epsilon \left[ \left( \frac{\sigma_c}{d_{i,j}} \right)^{12} - \left( \frac{\sigma_c}{d_{i,j}} \right)^6 \right] + \epsilon, \quad (\text{S39})$$

for  $d_{i,j} < 2^{1/6} \sigma$ , and  $U_{\text{LJ}}(i, j) = 0$  otherwise. In Eq. (S39) we set  $\epsilon = k_B T$ .



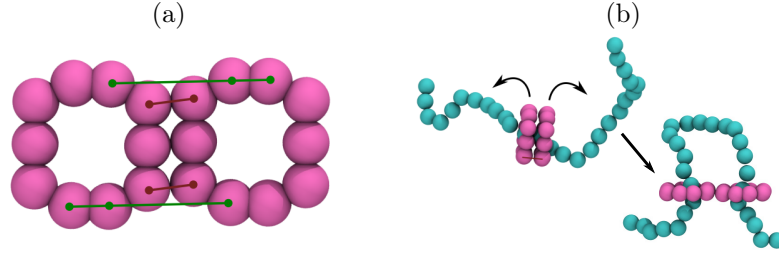


FIG. S5: Snapshots showing the slip-link model used in the Brownian dynamics simulations. (a) Each slip-link consists of a pair of rings, each composed from 10 beads of diameter  $\sigma$ . Each ring moves as a rigid body, and the pair is held together by two FENE bonds, indicated by brown lines. To keep the rings in an open handcuff configuration two Kratky-Porod bending interactions are added between beads as indicated by green lines. The separation between the centre point of each ring is  $2R = 3.5\sigma$ . (b) When a slip-link is added to the chromatin fiber, it is first arranged as a folded handcuff, with each ring encircling an adjacent bead of the polymer. As the dynamics evolve the bending potential acts to quickly unfold the handcuff and generate a loop.

The total potential energy experienced by chromatin bead  $i$  is given by

$$U_i = \sum_j U_{\text{FENE}}(i, j) \delta_{j, i+1} + \sum_j \sum_k U_{\text{B}}(i, j, k) \delta_{j, i+1} \delta_{k, i+2} + \sum_j U_{\text{LJ}}(i, j), \quad (\text{S40})$$

and its dynamics can be described by the Langevin equation

$$m\ddot{\mathbf{r}}_i = -\xi\dot{\mathbf{r}}_i - \nabla U_i + \boldsymbol{\eta}_i, \quad (\text{S41})$$

where  $m$  is the bead mass,  $\xi$  is the friction coefficient, and  $\boldsymbol{\eta}_i$  is a stochastic delta-correlated noise. The variance of each Cartesian component of the noise,  $\sigma_{\eta}^2$ , satisfies the usual fluctuation dissipation relation  $\sigma_{\eta}^2 = 2\xi k_B T$ .

In order to model slip-links we build a pair of rings out of beads also of diameter  $\sigma$ , and allow each ring to move as a rigid body. The translational motion of the centre of mass of the ring is described by a Langevin equation as in Eq. (S41), while rotation is described by a similar equation where the force term is replaced by the torque on the centre of mass, calculated from the forces experienced by the component beads of the ring. Each ring is composed of 10 beads arranged so that it is large enough to encircle the chromatin fiber. The two rings are held together by a pair of FENE bonds (as in Eq. (S37)), and they are kept in an open “handcuff” arrangement via two bending interactions (as in Eq. (S38), but with  $l_p = 100\sigma$ ). The slip-link beads interact with each other, and with chromatin beads with the WCA potential described above. Figure S5A shows the arrangement of a pair of rings and indicates the interactions between them. Slip-links are attached to a chromatin fiber by first positioning them in a folded handcuff arrangement such that each ring encircles an adjacent polymer bead; the bending interactions between the two rings then act to open the the handcuff, and bend the polymer (see Fig. S5B). After this the slip-link is free to diffuse in 3D and along the polymer.

Below we discuss two types of simulations: equilibrium simulations, where slip-links are attached to the chromatin fibre at the start of the simulation, and their dynamics are followed until the end of the simulation; and non-equilibrium simulations, where slip-links from an unbound pool are attached to the chromatin fibre at a rate  $k_{\text{on}}$ , and bound slip-links are removed from the fibre at a rate  $k_{\text{off}}$ . In the latter case the binding-unbinding kinetics break detailed balance, since slip-links are always added at adjacent beads, but can be removed when they are in any configuration (e.g., when they are associated with a large chromatin loop). In practice, to simulate this system we have coupled LAMMPS to an in-house code modelling stochastic detachment and binding. Unbound slip-rings are not simulated directly via Brownian dynamics, but are taken into account to determine which slip-links rebind.

As is customary [6], we use simulation units where the mass of a polymer bead  $m = 1$ , and the distance and energy units  $\sigma = 1$  and  $\epsilon = k_B T$  respectively; the simulation time unit is given by  $\tau_{\text{LJ}} = \sigma \sqrt{m/\epsilon}$ . There are two other time scales in the system, the velocity decorrelation time  $\tau_{\text{in}} = m/\xi$  and the Brownian time  $\tau_{\text{B}} = \sigma/D_b$ ; we set the friction  $\xi = 1$  meaning that  $\tau_{\text{in}} = \tau_{\text{LJ}} = \tau_{\text{B}}$ . Here  $D_b = k_B T/\xi$  is the diffusion coefficient of a bead of size  $\sigma$ . From the Stokes’ friction coefficient for spherical beads of diameter  $\sigma$  we have that  $\xi = 3\pi\eta_{\text{sol}}\sigma$  where  $\eta_{\text{sol}}$  is the solution viscosity. For the slip-link rings we set a total mass of each ring of  $m_r = 2.75m$ ; keeping  $\tau_{\text{in}} = \tau_{\text{LJ}}$  ensures a suitably larger friction  $\xi_r$  for these larger proteins (this means we approximate that each ring diffuses like a sphere of diameter  $2.75\sigma$ ). The numerical integration of Eq. (S41) uses a time step  $\Delta t = 0.01\tau_{\text{LJ}}$ .

Most of the simulations presented below, or in the main text, consider a short chromatin fragment (4.5–6 Mbp) in dilute conditions (this renders the computations simpler as using a realistic chromatin density would require a very small system size leading to severe tangles in the chromatin fibre which would not be resolved during the simulation). In order to map simulation time to real time scales, we require that the mean square displacement of a polymer bead matches that of a chromatin loci measured experimentally in Ref. [7] (this is the same procedure used in e.g. Ref. [8]). This means that we also match the effective *in vivo* viscosity due to macromolecular crowding etc. For this dilute system we obtain that  $\tau_B = \tau_j = 0.1$  s. In order to assess how the system is affected by the dynamics of denser polymer solutions, we also perform some simulations of larger chromatin fragments (15 Mbp) at concentrations similar to that found *in vivo*; we again map simulation times to the mean square displacement of a chromatin loci measured experimentally, and in this case obtain that  $\tau_B = \tau_j = 0.02$  s. We note that the different mapping for the time unit in the dilute and dense chromatin simulations means that the non-equilibrium slip-link binding rates  $k_{\text{on}}$  and  $k_{\text{off}}$  also have different mappings in those cases. For the long 15 Mbp fibre simulations, tangles in the chromatin might impede slip-link diffusion, and would still take longer than the simulation time to resolve, so we choose an initial untangled spiral configuration resembling a mitotic chromosome [8] and the system is then equilibrated for at least  $2 \times 10^6$  time steps before slip-links are added. In reality, enzymes such as topoisomerase would allow resolution of tangles.

### Brownian dynamics: results for the equilibrium case

In our 3D simulations the rate at which slip-links diffuse along the chromatin is determined by factors such as the interaction between the slip-link and the chromatin (specifically the relative size of the rings and the polymer beads) as well as the 3D diffusion constants of the rings and polymer beads. As detailed above we have chosen physically realistic parameters for the 3D diffusion of the individual components, and we do not try to control the 1D diffusion further. Instead we measure the 1D diffusion directly from the simulation. By tracking the position on the chromatin of one side of a slip-link in a simulation with a single slip-link and no unbinding, the diffusion constant can be found from the mean square displacement (MSD). Figure S6 shows plots of the MSD for both the dilute and dense chromatin density cases. We find that the 1D diffusion constant is approximately the same in both cases when expressed in simulation units ( $D_0 = 0.051 \sigma \tau_{\text{LJ}}^{-1}$ ). However, since these two cases have different mappings to real units, this maps to two different values ( $D_{0,\text{dilute}} = 5 \times 10^{-4} \mu\text{m}^2\text{s}^{-1}$ , and  $D_{0,\text{dense}} = 2.3 \times 10^{-3} \mu\text{m}^2\text{s}^{-1}$ ). To cast these in  $\text{kbp}^2/\text{s}$ , we recall the assumed compaction in our BD simulations is  $C = 100$  bp/nm, so that these values correspond to  $D_{\text{dilute}} = 5$   $\text{kbp}^2/\text{s}$ , and  $D_{\text{dense}} = 23$   $\text{kbp}^2/\text{s}$ . Interestingly, both estimates are well below the values measured *in vitro*: e.g., in Ref. [9],  $D_0 \sim 0.2525 \mu\text{m}^2/\text{s}$  was found, assuming an underlying 10-nm fibre,  $C = 20$  bp/nm, this leads to  $D \sim 100$   $\text{kbp}^2/\text{s}$ . We note that our BD measurements of 1D diffusivity should only be viewed as an approximate estimate for the values appearing in our theory, since the latter considers the diffusion constant for the separation of the two sides of the slip-link – this quantity is linked to our BD estimates by a numerical factor (equal to 2 if the two heads of a slip-link can diffuse independently).

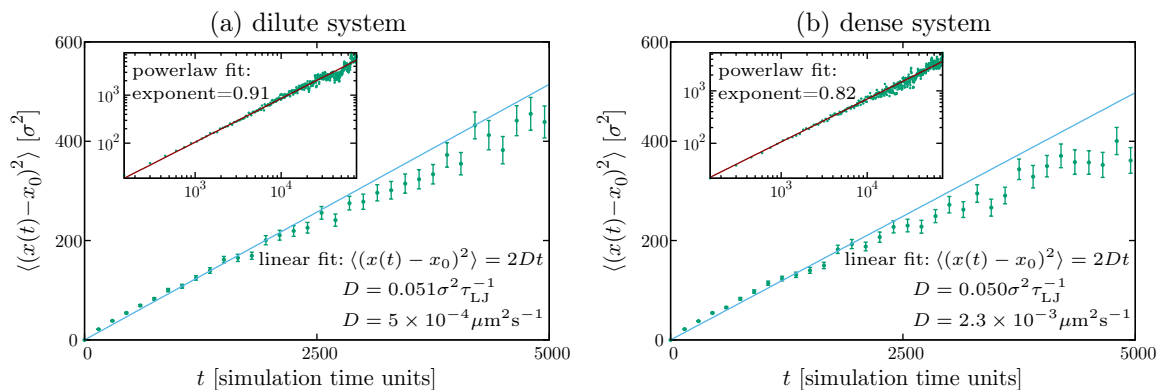


FIG. S6: Plots showing the mean squared displacement of the position on the chromatin of one side of a slip-link  $x(t)$ , from Brownian dynamics simulations of a single slip-link with no unbinding. Panel (a) shows the case of a dilute system (4.5 Mbp fibre in a periodic cubic system of size  $3 \mu\text{m}$ ), and panel (b) shows the case of a dense system (15 Mbp fibre confined in a sphere of diameter  $1.32 \mu\text{m}$ ). Insets show log-log plots with a fit to a power law function; the msd is approximately linear in time at short times. Lines in the main plots show a linear fit from which the diffusion constant is obtained. The mapping to real units uses  $\tau_{\text{LJ}} = 0.1$  s for the dilute system, and  $\tau_{\text{LJ}} = 0.02$  s for the dense system, as detailed in the text.

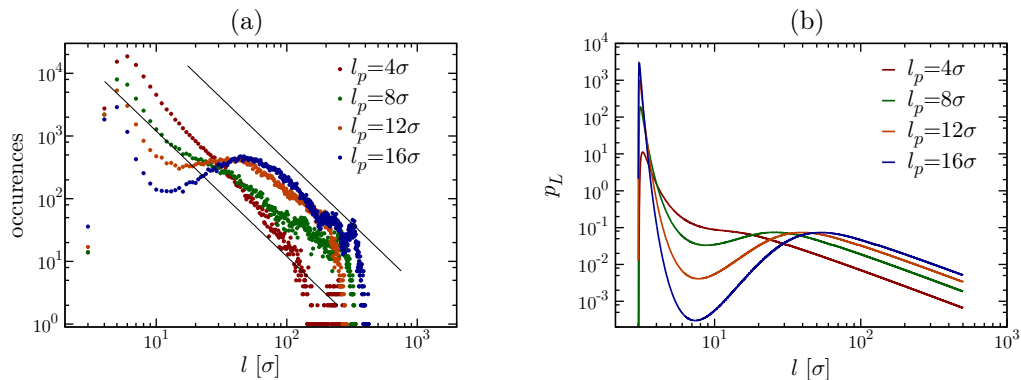


FIG. S7: (a) Log-log plot of the frequency of looping for slip-links diffusing on chromatin fibers with different flexibility (i.e., different values of the persistence length,  $l_p$ ). These results are from BD simulations for the equilibrium case of a single slip-link which does not unbind. (b) Probability of loop formation for an ideal semi-flexible polymer with different values of  $l_p$  as a function of loop size, according to the analytical approximation in Ref. [13], see the formula in the text. We assumed that the two ends of the loop need to lie at a 3D distance of  $r = 3\sigma$  from each other, and scaled the probability distributions by  $8l_p^3$  to show them more clearly on the same plot.

We now study the probability of formation of cohesin-mediated chromatin loops in a 3D simulation where a single slip-link is loaded on the fibre and cannot unbind. We consider chromatin fibres of different persistence length,  $l_p$ , and analyse the effect of this parameter on looping probability. As the unbinding rate is 0, the slip-link is eventually in thermal equilibrium, hence we expect that the distribution probability of loop lengths should be compatible with that arising in thermodynamic equilibrium.

Figure S7a shows the frequency of looping as a function of loop size, as found with our Brownian dynamics simulations. In line with our expectation, we find that the curves show power law decay for large  $l$ , as appropriate for looping probability in equilibrium. The exponent  $c \sim 2$ , is the one corresponding to internal looping in a self-avoiding walk [10–12]. Each of the curves in Figure S7a shows a peak, at small  $l$ . This peak remains in a similar position for all values of  $l_p$ . It can also be seen that, for stiffer polymers, there is a second shoulder, or smaller peak, for larger values of the loop length.

Figure S7b shows a prediction of looping probabilities obtained by using the analytical estimate in Ref. [13]. These results show that both peaks can be explained by considering a simple theory for semi-flexible polymers which neglects excluded volume interactions, by assuming that looping via a slip-link is equivalent to the constraint that the two ends of a loop are separated in 3D by a distance  $r$ . Physically, the first peak arises because very small loops cannot form, as a loop must at least span a distance  $r$ . The second peak is related to the well known optimal size of a loop in a semi-flexible polymer, which comes about due to the competition between the entropic cost, which shorter loops, and bending penalties, which favours longer loops [14].

For completeness, we report here the form of the analytical approximation for the distribution probability of the end-to-end distance  $r$  of a semiflexible polymer of size  $L$ ,  $p_L(r)$ , used in Ref. [13]:

$$p_L(r) = J(L) \left( \frac{1 - c\rho^2}{1 - \rho^2} \right)^{5/2} \exp \left( \frac{\sum_{i=-1}^0 \sum_{j=0}^3 c_{ij} \lambda^i \rho^{2j}}{1 - \rho^2} \right) \exp \left( - \frac{d\lambda ab(1+b)\rho^2}{1 - b^2\rho^2} \right) I_0 \left( - \frac{d\lambda a(1+b)\rho^2}{1 - b^2\rho^2} \right), \quad (\text{S42})$$

where

$$\begin{aligned}
 \rho &= \frac{r}{L} \\
 \lambda &= \frac{l_p}{L} \\
 a &= 14.054 \\
 b &= 0.473 \\
 c &= 1 - \left[ 1 + (0.38\lambda^{-0.95})^{-5} \right]^{-0.2} \\
 c_{ij} &= \begin{pmatrix} -3/4 & 23/64 & -7/64 \\ -1/2 & 17/16 & -9/16 \end{pmatrix} \\
 1 - d &= \begin{cases} 0 & \text{if } \lambda < 1/8 \\ \frac{1}{\frac{0.177}{\lambda - 0.111} + 6.4(\lambda - 0.111)^{0.783}} & \text{if } \lambda \geq 1/8 \end{cases} \\
 J(L) &= \frac{1}{8l_p^3} \begin{cases} \left(\frac{3\lambda}{\pi}\right)^{3/2} \left[ 1 - \frac{5\lambda}{4} - \frac{79\lambda^2}{160} \right] & \text{if } \lambda < 1/8 \\ 896.32\lambda^5 \exp\left(-14.054\lambda + \frac{0.246}{\lambda}\right) & \text{if } \lambda \geq 1/8 \end{cases}.
 \end{aligned} \tag{S43}$$

In Eq. (S43),  $I_0$  is a modified Bessel function of the first kind, and  $J(L)$  is the so-called Shimada and Yamakawa  $J$ -factor, measuring the ring closure probability for a wormlike chain (see [14] for details and the derivation of this factor).

### Brownian dynamics: non-equilibrium simulations, with loading sites and CTCF

In this section, we present results of non-equilibrium Brownian dynamics simulations of longer chromatin fibres, where we take into account loading and unloading, and we model the presence of CTCF.

First, Figure S8 shows results from simulations of a model chromatin fiber of size  $L = 2000\sigma$ , split into sections, in each of which we place a slip-link. At the end of each section we locate beads with high affinity for the slip-link – modelling convergent CTCF sites (the end “CTCF bead” of a given section is  $5\sigma$  away from the start “CTCF bead” of the next section). The results show the fraction of slip-links which reach the sticky CTCF sites for different values of the persistence length,  $l_p$ , and of the loop, or section, size (see also Suppl. Movie 1, valid for  $l_p = 8\sigma$  and a loop size of  $90\sigma$ ). The simulations “with ratchet effect” consider three slip-links per section, with the topology arranged so as to give three loops where the largest loop contains the middle loop which in turn contains the smallest loop (see also Suppl. Movie 2, and the corresponding nested rainbow rings determining looping topology). The “ratchet” effect, which is discussed in more detail below, leads to a dramatic increase in the fraction of largest loops which reach the CTCF sites. Our results also show that the stiffer the fiber, the more likely is the formation of a CTCF-mediated loops (Fig. S8B).

Our 3D simulations show that, especially for large  $N$ , nested loops are often formed by closely “stacked” slip-links (see inset of Fig. 2d, main text and Fig. S8). Stacking is favoured by entropic forces tending to diminish the total number of loops [15], thereby clustering slip-links, and can be recognised in 1D looping diagrams (SI, Fig. S3) as “rainbow” patterns (Suppl. Movies 2 and 3). Stacking is not, however, essential for the ratchet effect – what is essential is the set-up of a non-trivial osmotic pressure in 1D, as discussed in the section “1D models with many interacting slip-links, and the osmotic ratchet”.

Figure S9 shows results from larger simulations (a 15 Mbp chromatin fibre) with chromatin concentration similar to that found *in vivo*. To achieve this concentration, the fibre was confined in a sphere (diameter  $1.32 \mu\text{m}$  – we have checked that similar results are obtained if we consider periodic boundary conditions and a cubic box of size such that the chromatin density is the same). Twenty CTCF sites were positioned at regular intervals (they are separated by either 750 kbp or 1500 kbp) in various orientations; 16 cohesin loader sites were positioned at 750 kbp intervals between the CTCFs. Ten independent simulations were run, and in each simulation the CTCF sites were occupied with a probability 0.8, so that the occupancies could vary between simulations; results were then averaged over all simulations. There were 32 slip-links which were loaded stochastically at a randomly selected loader site at a rate where  $k_{\text{on}}^{-1} = 2.5 \text{ min}$ , and were unloaded at a rate where  $k_{\text{off}}^{-1} = 25 \text{ min}$ . Slip-links stick at CTCF sites which are orientated towards them, but reflect from CTCFs orientated away from them. The unbinding rate of a slip-link which is bound to two CTCF sites is reduced by multiplying by a factor  $f = 0.1$ .

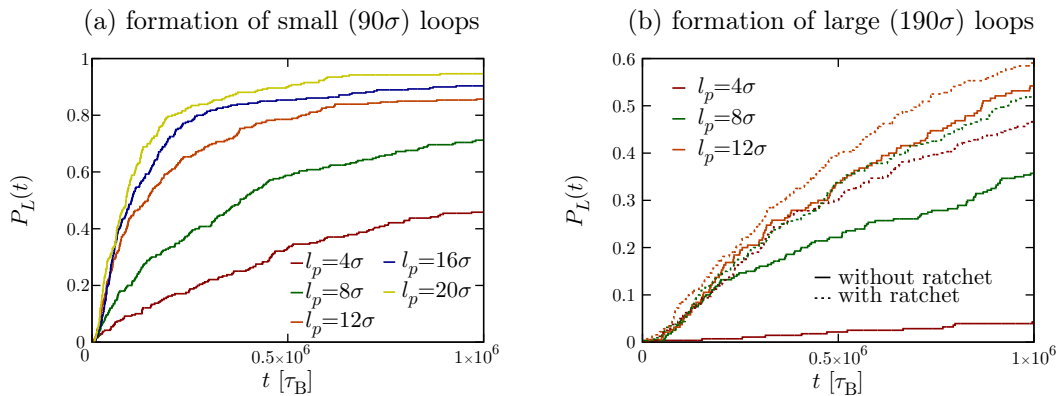


FIG. S8: The osmotic ratchet can substantially increase the probability of CTCF-mediated loop formation. A chromatin fiber of size  $L = 2000\sigma$  is divided up into sections of (a)  $90\sigma$  (b) or  $190\sigma$ , with a convergent pair of CTCF sites placed at either end of each section. Curves show the probability that the largest loop in each section has reached the CTCF sites, as a function of time (measured here in Brownian times, each corresponding to 0.1 s). The simulations “with ratchet effect” contain three slip-links per CTCF loop (see text), the others one slip-link per CTCF loop. Using parameters relevant for chromatin ( $\sigma = 30$  nm with a compaction of 100 bp/nm – i.e., 3 kbp per bead), these results correspond to loop sizes of 270 kbp and 570 kbp.

The conformation of the simulated chromosome can be visualized as a contact map [Fig. S9(c)] similar to the interaction maps generated from HiC data. This shows the frequency at which interactions occur between different loci; adjacent pairs of CTCF sites give rise to triangles, reminiscent of the topological domains observed in HiC [see also the blue segment in Fig. S9(b)]. Stable loops formed between convergent pairs of CTCF sites give rise to dark spots at the corner of some domains. A dark region is also observed close to loader sites (these would be lines perpendicular to the main diagonal in a square map, and would become stronger if more slip-links were added) – this is not usually observed in HiC maps, and highlights the fact that this is a somewhat artificial example. For instance, the loaders have been placed equidistant from the adjacent CTCFs, and the model does not include other chromatin-chromatin interactions (e.g. mediated by complexes of transcription factors and polymerase) or the fact that transcribing polymerases are thought to be able to push cohesins short distances excluding them from active genes (see [16], and see discussion in the main text).

For completeness, we present here some typical contact maps which can be obtained with our simpler 1D simulations, by assuming that all contacts are cohesin-mediated loops. We again note that we should only expect partial agreement with HiC maps, as we do not consider other mechanisms which lead to chromatin interactions, such as contacts mediated by bridging proteins [4, 17, 18]. Nevertheless, the qualitative agreement with TAD patterns in Hi-C is encouraging. The contact maps show that by introducing an interaction between CTCF and cohesin ( $1 k_B T$  is sufficient) it is possible to obtain a spot at the corner of a domain [i.e., between convergent CTCF sites, Figs. S10(a,d), as in Fig. S1]. By increasing  $N$ , a spot appears this time even in the *absence* of an interaction [Figs. S10(c,f,g,j)], as in the loop extrusion model: this is because the ratchet leads to effective extrusion. The simulations in Figs. S10(a-f,g,j) have a loading site in the middle: increasing the number of cohesins in the ratchet also leads to the appearance of a fainter line perpendicular to the main diagonal [e.g., Fig. S10(g)]. While this is not normally seen in mammalian TADs, Hi-C maps from some bacteria *do* show this feature, and it has been attributed to loop extrusion by condensin-like SMC-complexes [20] – our results suggest that passive sliding together with the ratchet effect could also account for this phenomenon. Placing the loading site close to the boundary leads to additional features in the contact maps: in particular in Figs. S10(h,i,k,l) for large enough  $N$  a line appears, approximately parallel to the border of the domain. Similar straight lines are often seen at the boundary of Hi-C contact maps in Ref. [19]. While the agreement is only qualitative, this is notable given the simplicity of the model. It would be of interest to compare more in details experimental contact maps with those predicted by our 1D or 3D models – to do so, we would need more information about the location of loading sites. It would also be important in the future to analyse contact maps obtained from simulating a combination of non-equilibrium slip-links and other kinds of chromatin bridging proteins such as those considered in the simulations of Refs. [4, 17, 18].



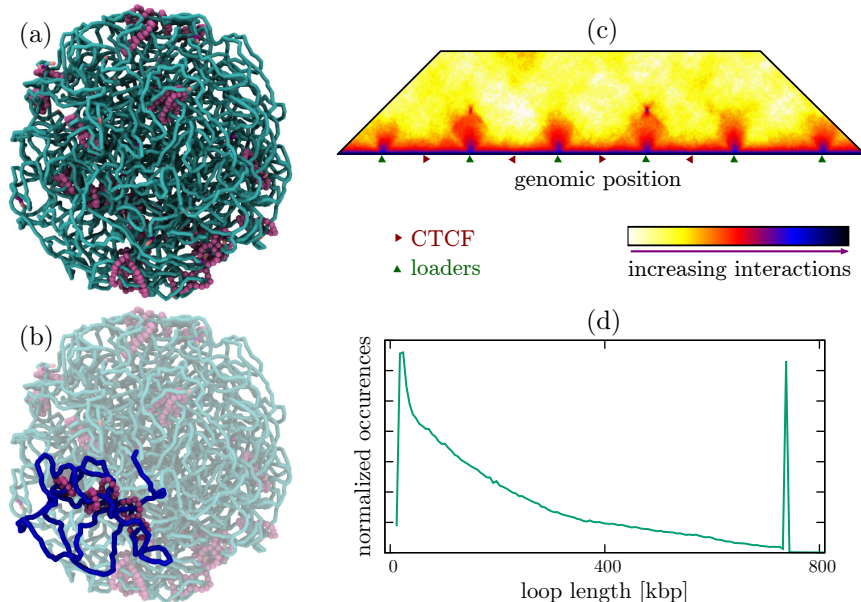


FIG. S9: Simulations of a 15 Mbp chromatin fibre, at a chromatin density similar to that found *in vivo*, and confined in a sphere. (a) A typical snapshot of the simulation. (b) The same snapshot is shown, but now a region of chromatin between two convergent CTCFs is highlighted in blue. (c) A HiC like interaction map is shown for a 300 kbp region of the simulated fibre. The colour at each point in the map indicates the frequency of interaction between the chromatin positions connected by a triangle with its apex at that point. Positions and orientations of CTCF sites, and positions of the loader sites are indicated. (d) Plot showing the probability distribution of loop lengths. The loop length is defined as the genomic distance between the chromatin beads encircled by each side of each slip-link, and all bound slip-links are included in the measurement. The peak at 750 kbp is due to slip-links sticking at convergent CTCF pairs; the height of this peak depends on the factor by which  $k_{\text{off}}$  is multiplied for slip-links forming CTCF loops, but we expect it to still be visible even for  $f = 1$ .

## ANALYSIS OF CHIA-PET DATA

To estimate the *in vivo* probability of finding cohesin and CTCF-mediated loops of a given length, we analysed ChIA-PET (Chromatin Interaction Analysis by Paired-End Tag Sequencing) data from Ref. [21] (data publicly available from the Gene Expression Omnibus (GEO), accession number GSE72816). In these experiments chromatin-chromatin interactions mediated by a specific proteins are identified using immunoprecipitation; in this way pairs of interacting chromatin regions which are both bound by CTCF are identified. In particular, Figure 1e of the main manuscript shows the contact probability between CTCF-bound sequences in GM12878 cells (data set from GEO accession number GSM1872886). Data were sorted into bins of size 5 kbp according to loop size. This contact probability, in the range 100 – 1000 kbp, is compatible with an exponential decay (see Fig. 1e), where the decay length of the exponential is of the order of hundreds of kbp, within the typical range of CTCF-mediated loops [22]. These data can also be fitted by a power law, but the fit is quite poor (Fig. S11a). The exponent resulting from the fit is  $\sim -0.35$ , which is lower than and far from those which can be explained by a polymer physics model (e.g.,  $\sim -2.1$  for an ideal self-avoiding walk,  $\sim -1.5$  for an equilibrium globule or  $\sim -1$  for a fractal globule). It is interesting to note that ChIA-PET contacts between sequences bound to RNA polymerase II appear to be better fitted by a power law albeit with a similarly low exponent (Fig. S11b; data set from GEO accession number GSM1872887 [21]). These results suggest that CTCF-mediated contacts (as well as RNA PolII-mediated ones) appear to obey contact decay laws which are incompatible with the laws which would be predicted on the basis of equilibrium polymer models. Notably, such polymer models can however account very well for the contact decay laws typically found in Hi-C experiments [23] (which probe chromatin-chromatin interactions genome wide without selecting for specific proteins). A possible explanation is that Hi-C contacts encompass many interactions arising randomly from spatial proximity in 3D (i.e. there does not have to be a protein mediated interaction, since any chromatin regions with close proximity can be captured), which can be explained by polymer physics assuming randomly diffusing polymers. Taken together these observations support our hypothesis that cohesin/CTCF mediated loops form through a non-equilibrium mechanism *in vivo*, and are distinct from other chromatin diffusion mediated loops.



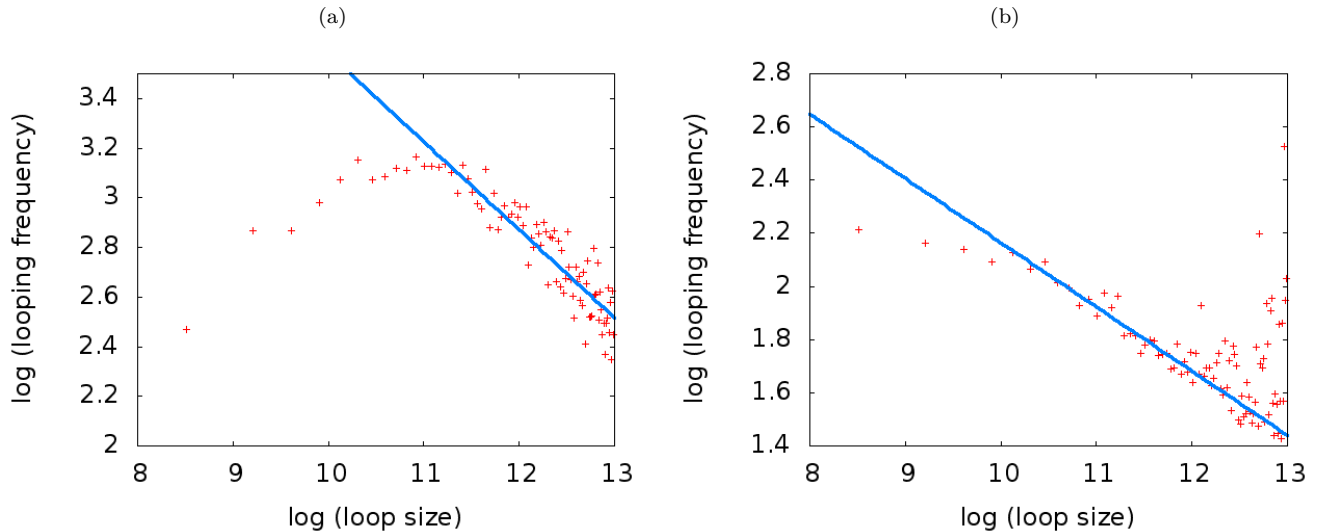


FIG. S11: (a) Log-log plot of the ChIA-PET data for CTCF-mediated loops. Solid lines show a fit to a power law for values of the abscissa (natural log of loop size) between 11 and 13, and the resulting exponent was  $\sim -0.35$ . (b) Log-log plot of the ChIA-PET data for contacts associated with RNA Polymerase II; the fit was performed for values of the abscissa (natural log of loop size) between 10 and 12, and the resulting exponent was  $\sim -0.24$ .

which studied cohesin diffusion on naked DNA and in the presence of nucleosomes; Ref. [9] also studied the case of cohesin diffusing on chromatin formed in *Xenopus* egg extracts. It is important to discuss these experiments as they measure diffusion coefficients which have a significant bearing on the conclusions we can draw from our analysis.

Recall that a key consequence of our work is that it predicts the minimal diffusion coefficient,  $D$ , required for slip-links to be able to organise chromosome loops of hundreds of kbp (the typical size of CTCF loops in mammalian genomes). We found that this minimal value is  $D = 10 \text{ kbp}^2/\text{s}$ . To compare with *in vitro* experiments, which normally measure diffusion on a quasi-1D substrate (DNA or chromatin), we note that  $D = D_0 C^2$ , where  $D_0$  (measured in  $\mu\text{m}^2/\text{s}$ ) is the 1D diffusion accessible experimentally, while  $C$  (measured in  $\text{bp}/\text{nm}$ ) is the compaction of DNA base pairs within the chromatin fibre. A value of  $C$  equal to  $20 \text{ bp}/\text{nm}$  corresponds to an open 10-nm chromatin fiber (with 200 bp per histone octamer). At the other end of the possible range of values, in a 30-nm fiber there can be 3 kbp of DNA per 30 nm, so  $C = 100 \text{ bp}/\text{nm}$ . The worst possible case for our theory occurs if  $C$  is small: for  $C = 20 \text{ bp}/\text{nm}$  we obtain the 1D diffusion coefficient needs to be at least  $D_0 = 0.025 \mu\text{m}^2/\text{s}$  for sufficiently large loops to form before cohesin detaches.

So the key question is, what diffusion rates for cohesin can be expected *in vivo*? The experiments in Ref. [9] have shown that cohesin diffusion on DNA or chromatin is dependent on its ATPase activity, salt concentration, and how it was loaded onto the substrate, and is greatly enhanced by cohesin acetylation. A value of  $D_0 = 0.2525 \pm 0.0031 \mu\text{m}^2/\text{s}$  was measured for acetylated cohesin on chromatin [9]. If similar rates are realized *in vivo* then a single cohesin (i.e. without the ratchet effect) can indeed extrude 100 kbp loops within 20 mins, even for the lower bound  $C$  value.

On the other hand, Refs. [24, 25] measured a smaller  $D_0$  – however these measurements were performed in the absence of acetylation. The results of Refs. [24, 25] also showed that non-acetylated cohesin can diffuse over single nucleosomes, but not over obstacles of diameter larger than 21 nm. By extrapolating those results to a nucleosomal fibre, Ref. [24] suggested that cohesin might only cover around 7 kbp in an hour by diffusion, which is too small to drive CTCF-mediated looping. Together with Ref. [9], these works suggest that the pore size, and therefore diffusivity of cohesin, may be regulated by ATP hydrolysis and acetylation. If the *in vivo* diffusion rate is closer to the slower estimate of [24], then our mechanism would require either a more compact fibre, or the additional ratchet effect to operate. We note though that the extrapolation of 7 kbp explored in an hour refers to a stretched array of nucleosomes, meaning a very low compaction – we propose that this is therefore not a good prediction for the *in vivo* case, so diffusion is likely more efficient at exploring chromatin there. To quantify the enhancement that can be expected from the ratchet effect, Figure S8 shows the probability of formation of a long 570 kbp convergent CTCF-mediated loop over time, for chromatin fibres with different numbers of cohesins. As detailed above, the results show that the osmotic ratchet is at work with as few as 3 bound cohesins per loop, and dramatically enhances looping probability.

We hope that our work will prompt new studies to accurately measure cohesin diffusion on reconstituted chromatin

fibers, and as a function of the number of cohesins, thereby allowing a test of our osmotic ratchet in the lab.

### CAPTIONS FOR SUPPLEMENTARY MOVIES

**Suppl. Movie 1:** This movie shows the looping diagrams associated with the dynamics corresponding to Suppl. Fig. 8a, where a chromatin fiber of length  $L = 2000\sigma$  and persistence length  $l_p = 8\sigma$  is divided into sections of size  $90\sigma$ ; we assume that each of the sections contains a single slip-link (modelling cohesin) at all times, and that it is delimited by a bead at each boundary which is sticky for the slip-link, to model the presence of CTCF convergent sites. Each of the arcs shown in the movie tracks the positions of the two ends of each slip-link along the chromatin fiber. The interaction between CTCF and cohesin is large enough to ensure virtually irreversible binding on the timescale of our simulations.

**Suppl. Movie 2:** As Suppl. Movie 1, but now with  $l_p = 4\sigma$ , and with sections of size  $190\sigma$ , with three slip-links per section. It can be seen that the simultaneous presence of the three slip-links leads to a ratcheting effect which favours loop formation.

**Suppl. Movie 3:** This movie shows the self-organization of the osmotic ratchet; as in Suppl. Movies 1 and 2 the looping diagrams are shown for each configuration. The dynamics are slightly different from that shown in Fig. 2 of the main manuscript: now there is not a fixed number of slip-links, but slip-links bind (i.e., are created, when the loading site is unoccupied) at rate  $k_{\text{on}} = 10^{-3} \text{ s}^{-1}$  and detach (i.e., are destroyed) at rate  $k_{\text{off}} = 10^{-4} \text{ s}^{-1}$ . The formation of “rainbow patterns” with arcs tightly stacked against each other is due to entropic forces which favour the presence of a single loop, kept together by several clustered slip-links, over that of many loops, where slip-links are homogeneously distributed.

- 
- [1] M. R. Evans and S. N. Majumdar. Diffusion with Stochastic Resetting, *Phys. Rev. Lett.* **106**, 160601 (2011).
  - [2] P. de los Rios and A. Barducci. Hsp70 chaperones are non-equilibrium machines that achieve ultra-affinity by energy consumption, *eLife* **3**, e02218 (2014).
  - [3] S. Plimpton. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **117**, 1 (1995).
  - [4] C. A. Brackley, J. Johnson, S. Kelly, P. R. Cook and D. Marenduzzo. Simulated binding of transcription factors to active and inactive regions folds human chromosomes into loops, rosettes and topological domains. *Nucleic Acids Res.* **8**, 3503 (2016).
  - [5] J. Langowski. Polymer models of DNA and chromatin. *Eur. Phys. J. E* **19**, 241 (2006).
  - [6] K. Kremer and G. S. Grest. Dynamics of entangled linear polymer melts: A molecular-dynamics simulation. *J. Chem. Phys.* **92**, 5057 (1990).
  - [7] H. Hajjoul, J. Mathon, H. Ranchon, I. Goiffon, J. Mozziconacci, B. Albert, P. Carrivain, J. M. Victor, O. Gadad, K. Bystrycky and A. Bancaud. High-throughput chromatin motion tracking in living yeast reveals the flexibility of the fiber throughout the genome. *Genome Res.* **23**, 1829 (2013).
  - [8] A. Rosa and R. Everaers. Structure and Dynamics of Interphase Chromosomes. *PLOS Comp. Biol.*, e1000153 (2008).
  - [9] M. Kanke, E. Tahara, P. J. Huis in’t Veld, and T. Nishiyama. Cohesin acetylation and Wapl-Pds5 oppositely regulate translocation of cohesin along DNA. *EMBO J.* **35**, 2686 (2016).
  - [10] E. Carlon, E. Orlandini and A. L. Stella. Roles of stiffness and excluded volume in DNA denaturation. *Phys. Rev. Lett.* **88**, 198101 (2002).
  - [11] B. Duplantier. Statistical mechanics of polymer networks of any topology. *J. Stat. Phys.* **54**, 581 (1989).
  - [12] A. Hanke and R. Metzler. Entropy loss in long-distance DNA looping. *Biophys. J.* **85**, 167 (2003).
  - [13] A. Rosa, N. B. Becker and R. Everaers. Looping Probabilities in Model Interphase Chromosomes. *Biophys. J.* **98**, 2410 (2010).
  - [14] J. Shimada and H. Yamakawa. Ring-closure probabilities for twisted wormlike chains. Application to DNA. *Macromolecules* **17**, 689 (1984).
  - [15] R. Metzler, A. Hanke, P. G. Dommersnes, Y. Kantor and M. Kardar. Tightness of slip-linked polymer chains. *Phys. Rev. E* **65**, 061103 (2002).
  - [16] G. A. Busslinger, R. A. Stocsits, P. van der Lelij, E. Axelsson, A. Tedeschi, N. Galjart and J. M. Peters. Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. *Nature* **544**, 503 (2017).
  - [17] M. Barbieri, M. Chotalia, J. Fraser, L.-M. Lavitas, J. Dostie, A. Pombo and M. Nicodemi. Complexity of chromatin folding is captured by the strings and binders switch model. *Proc. Natl. Acad. Sci. USA* **109**, 16173 (2012).
  - [18] M. Chiariello, S. Bianco, C. Annunziatella, A. Esposito and M. Nicodemi. Polymer physics of chromosome large-scale 3D organisation. *Scientific Reports* **6**, 29775 (2016).
  - [19] S. S. P. Rao, M. H. and Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov, J. T. Robinson, A. L. Sanborn, I. Machol, A. D. Omer, E. S. Lander, and E. Lieberman Aiden. A 3D map of the human genome at kilobase resolution

- reveals principles of chromatin looping. *Cell* **159**, 1665 (2014).
- [20] X. Wang, H. B. Brandao, T. B. K. Le, M. T. Laub and D. Z. Rudner. Bacillus Subtilis SMC Complexes Juxtapose Chromosome Arms as They Travel From Origin to Terminus. *Science* **335**, 524-527 (2017).
  - [21] Z. Tang *et al.*. CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription, *Cell* **163**, 1611 (2015).
  - [22] M. Oti, J. Falck, M. A. Huynen and H. Zhou. CTCF-mediated chromatin loops enclose inducible gene regulatory domains. *BMC Genomics* **17**, 252 (2016).
  - [23] A. L. Sanborn *et al.* Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci. USA* **112**, E6456 (2015).
  - [24] J. Stigler, G. Camdere, D. E. Koshland and E. C. Greene. Single-Molecule Imaging Reveals a Collapsed Conformational State for DNA-Bound Cohesin. *Cell Rep.* **15**, 988 (2016).
  - [25] I. F. Davidson *et al.* Rapid movement and transcriptional re-localization of human cohesin on DNA. *EMBO J.* **35**, 2671 (2016).