

Electronic Supplementary Information: Osmotic mechanism of loop extrusion process

Tetsuya Yamamoto and Helmut Schiessel

September 12, 2017

S1 Derivation of the fluxes of cohesin rings

Here we summarize the derivation of eqs. (1) and (4) in the main article. We treat a stretch of chromatin fiber, which is flanked by two BEs (such as CTCF proteins in converging orientations). This chromatin region has a loader in the middle ($z = 0$) and an unloader at each end ($z = \pm M$). The region has $2M$ binding sites, which cohesin monomers and dimers can occupy. For simplicity, we assume that one ring of a dimer is at the positive half and the other ring is at the negative half. With this approximation, the number of microstates has the form

$$W = \left(\frac{M!}{N_d!(M - N_d)!} \right)^2 \frac{(M - N_d)!}{N_+!(M - N_d - N_+)!} \frac{(M - N_d)!}{N_-!(M - N_d - N_-)!}, \quad (\text{S1})$$

where N_d is the number of dimers, N_+ is the number of monomers in the positive half ($z > 0$), and N_- is the number of monomers in the negative half ($z < 0$). We derived eq. (S1) by assuming that the two rings of a dimer move independently. The factorials $N_+!$ and $N_-!$ in the denominator ensure that cohesin monomers cannot penetrate through other cohesin monomers. In the same vein, the factorial $N_d!$ in the denominator ensures that the rings of cohesin dimers cannot penetrate through the rings of other cohesin dimers. However, our mean field theory may not fully capture the fact that the rings of dimers cannot penetrate through monomers and vice versa. The free energy of the system thus has the form

$$\begin{aligned} F &= -k_B T \log W \\ &\simeq M k_B T [2\psi_d \log \psi_d + \psi_+ \log \psi_+ + \psi_- \log \psi_- \\ &\quad + (1 - \psi_d - \psi_+) \log(1 - \psi_d - \psi_+) + (1 - \psi_d - \psi_-) \log(1 - \psi_d - \psi_-)], \end{aligned} \quad (\text{S2})$$

where we used the Stirling's approximation to derive the last step. ψ_+ ($= N_+/M$) and ψ_- ($= N_-/M$) are the densities of monomers in the positive and negative halves. ψ_d ($= N_d/M$) is the density of dimers.

We generalize eq. (S2) to cases in which the probability ψ_+ and ψ_- of finding monomers and the probability ψ_d of finding dimers are not uniform along the chromatin fiber in the spirit of ref. [1] (see sec. 2.3). With this treatment, the free energy of the system has the form

$$F = k_B T \int_{-M}^M dz [\psi_d(z) \log \psi_d(z) + \psi_m(z) \log \psi_m(z) + (1 - \psi_d(z) - \psi_m(z)) \log(1 - \psi_d(z) - \psi_m(z))], \quad (\text{S3})$$

where $\psi_m(z)$ is the probability of finding monomers at z . Both $\psi_m(z)$ and $\psi_d(z)$ are even functions because of the symmetry of the system. This equation returns to eq. (S2) for cases in which the distribution functions, $\psi_m(z)$ and $\psi_d(z)$, do not depend on the position z . The chemical potential of cohesin monomers has the form

$$\begin{aligned} \mu_m(z) &= \frac{\delta F}{\delta \psi_m(z)} \\ &= k_B T [\log \psi_m(z) - \log(1 - \psi_d(z) - \psi_m(z))]. \end{aligned} \quad (\text{S4})$$

The flux J_m of cohesin monomers is thus derived as

$$\begin{aligned} J_m &= -\frac{1}{\zeta_c} \psi_m(z) \frac{\partial}{\partial z} \mu_m(z) \\ &= -D_c \left[\frac{1 - \psi_d(z)}{1 - \psi_d(z) - \psi_m(z)} \frac{\partial}{\partial z} \psi_m(z) + \frac{\psi_m(z)}{1 - \psi_d(z) - \psi_m(z)} \frac{\partial}{\partial z} \psi_d(z) \right], \end{aligned} \quad (\text{S5})$$

where ζ_c denotes the friction constant of cohesin rings. In the last step we used the Einstein relationship $D_c = k_B T / \zeta_c$. For cases in which cohesin dimers are only rarely loaded onto the chromatin fiber ($\psi_d \ll 1$), eq. (S5) has the approximate form

$$J_m \simeq -D_c \frac{1}{1 - \psi_m(z)} \frac{\partial}{\partial z} \psi_m(z), \quad (\text{S6})$$

where we omitted higher order terms with respect to $\psi_d(z)$. Eq. (S6) can be rewritten in the form

$$J_m = -D_c \frac{\partial}{\partial z} \psi_m(z) - \frac{D_c}{k_B T} \psi_m(z) \frac{\partial}{\partial z} \Pi_{\text{osm}}(z), \quad (\text{S7})$$

where $\Pi_{\text{osm}}(z)$ is the osmotic pressure generated by cohesin monomers

$$\Pi_{\text{osm}}(z) = -k_B T \log(1 - \psi_m(z)). \quad (\text{S8})$$

Eq. (S8) returns to the familiar form of the osmotic pressure of an ideal gas for $\psi_m(z) < 1$.

The chemical potential of cohesin dimers has the form

$$\begin{aligned}\mu_d(z) &= \frac{\delta F}{\delta \psi_d(z)} \\ &= k_B T [\log \psi_d(z) - \log(1 - \psi_d(z) - \psi_m(z))].\end{aligned}\quad (\text{S9})$$

The flux J_d of cohesin dimers is thus given by

$$\begin{aligned}J_d &= -\frac{1}{\zeta_c} \psi_d(z) \frac{\partial}{\partial z} \mu_d(z) \\ &= -D_c \left[\frac{1 - \psi_m(z)}{1 - \psi_d(z) - \psi_m(z)} \frac{\partial}{\partial z} \psi_d(z) + \frac{\psi_d(z)}{1 - \psi_d(z) - \psi_m(z)} \frac{\partial}{\partial z} \psi_m(z) \right] \\ &\simeq -D_c \frac{\partial}{\partial z} \psi_d(z) - \frac{D_c}{k_B T} \psi_d(z) \frac{\partial}{\partial z} \Pi_{\text{osm}}(z),\end{aligned}\quad (\text{S10})$$

where we used the fact that cohesin dimers are only rarely loaded onto the chromatin fiber and omitted higher order terms with respect to $\psi_d(z)$ to derive the last form of eq. (S10).

We derived eq. (S1) by assuming that the two rings of a dimer diffuse independently. In a similar manner, one can show that for cases in which the two rings diffuse in registry, eq. (S10) should be replaced by

$$J_d = -\frac{1}{2} D_c \frac{\partial}{\partial z} \psi_d(z) - \frac{D_c}{k_B T} \psi_d(z) \frac{\partial}{\partial z} \Pi_{\text{osm}}(z),\quad (\text{S11})$$

where the factor 1/2 represents the fact that the two rings move together, analogous to the Flory-Huggins theory [1]. The solution of eq. (S11) changes to

$$\psi_d(z) = \frac{J_d}{J_m} \left(1 - \alpha e^{-2J_m(M-z)/D_c} \right),\quad (\text{S12})$$

where we used a factor $\alpha (= 1 - J_m/k_{\text{off}}^d)$. The broken curves in fig. 2 in the main article are predicted by using eq. (S12).

S2 Spontaneous unloading of cohesin rings

We here treat cases in which cohesin rings are unloaded before they reach the ends of the stretch of chromatin fiber, see fig. 1 in the main article. A cohesin loader is localized in the middle of the chromatin region. In steady state, the dynamics of cohesin monomers is represented by the form

$$-\frac{\partial}{\partial z} J_m(z) - k_{\text{off}}^m \psi_m(z) = 0.\quad (\text{S13})$$

The first term of eq. (S13) represents the accumulation of cohesin monomers by the gradient of the flux of monomers. The second term of eq. (S13) represents the spontaneous unloading of monomers and k_{off}^m is the rate constant that

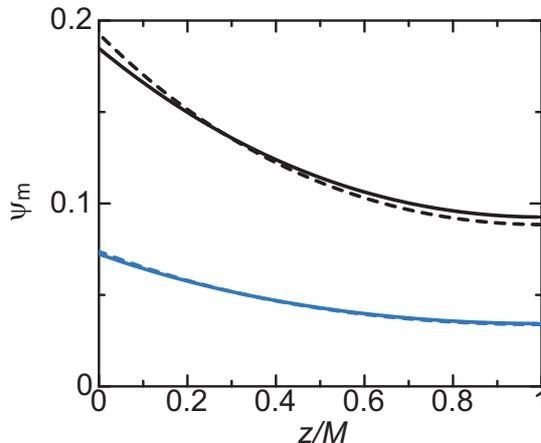


Figure S1: The probability $\psi_m(z)$ of finding a cohesin monomer at each binding site is shown as a function of the position z/M for $k_{\text{on}}^m c/D_c = 0.1$ (blue) and 0.3 (black). We used $k_{\text{off}}^m c/D_c = 2.0$ for the calculations. The solid curve is derived by numerically calculating eq. (S13) and the broken curve is derived by using an approximate form, eq. (S15).

accounts for this process. $\psi_m(z)$ is the probability of finding cohesin monomers at z . The flux $J_m(z)$ of cohesin monomers has the form

$$J_m(z) = -D_c \frac{\partial}{\partial z} \psi_m(z) - \frac{D_c}{k_B T} \psi_m(z) \frac{\partial}{\partial z} \Pi_{\text{osm}}(z). \quad (\text{S14})$$

The first term on the right-hand side of eq. (S14) is due to the thermal diffusion of monomers and the second term is due to the osmotic pressure $\Pi_{\text{osm}}(z)$ ($\equiv -k_B T \log(1 - \psi_m(z))$) generated by monomers. The boundary conditions to solve eq. (S13) are i) the flux of monomers is zero at $z = M$ due to the BEs, $J_m(M) = 0$, and ii) the flux of monomers is equal to the loading rate of the monomers, $J_m(0) = k_{\text{on}}^m c(1 - \psi_m(0))$ (k_{on}^m is the rate constant that accounts for the binding process and c is the concentration of cohesin rings in the solution). We numerically solve eq. (S13) by using the Runge-Kutta method, see fig. S1. This leads to the probability distribution $\psi_m(z)$ as a function of the dimensionless parameters $k_{\text{on}}^m cM/D_c$ and $k_{\text{off}}^m M^2/D_c$.

For cases in which the rescaled binding rate $k_{\text{on}}^m cM/D_c$ of monomers is relatively small, the solution of eq. (S13) has an approximate form

$$\psi_m(z) = \frac{\tilde{k}_{\text{on}}^m \tanh \sqrt{\tilde{k}_{\text{off}}^m} \cosh(\sqrt{\tilde{k}_{\text{off}}^m}(1 - z/M))}{\tilde{k}_{\text{on}}^m + \sqrt{\tilde{k}_{\text{off}}^m} \tanh \sqrt{\tilde{k}_{\text{off}}^m} \sinh \sqrt{\tilde{k}_{\text{off}}^m}}, \quad (\text{S15})$$

where we used the rescaled rate constants $\tilde{k}_{\text{on}}^m = k_{\text{on}}^m cM/D_c$ and $\tilde{k}_{\text{off}}^m = k_{\text{off}}^m M^2/D_c$, see the broken curve in fig. S1. Our numerical calculations of eq. (S13) agree

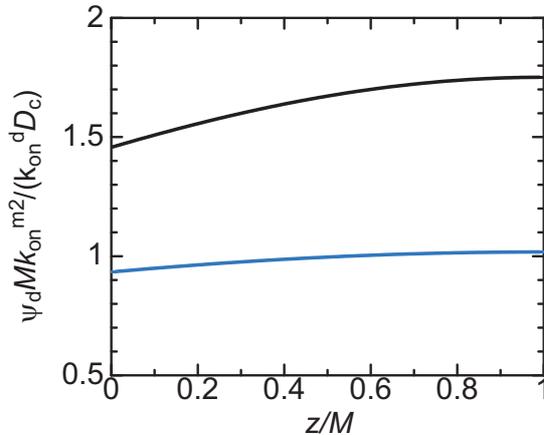


Figure S2: The probability $\psi_d(z)$ of finding a cohesin dimer at each binding site (rescaled by $D_c k_{\text{on}}^d / (M k_{\text{on}}^{m2})$) is shown as a function of the position z for $k_{\text{off}}^d M^2 / D_c = 0.3$ (blue) and 0.5 (black). We used $k_{\text{on}}^m c M / D_c = 3.0$ and $k_{\text{off}}^m M^2 / D_c = 1.0$ for the calculations.

with the approximate form eq. (S15) for small values of \tilde{k}_{on}^m (or large values of \tilde{k}_{off}^m).

In the steady state, the dynamics of cohesin dimers is represented by the form

$$-\frac{\partial}{\partial z} J_d(z) - k_{\text{off}}^d \psi_d(z) = 0 \quad (\text{S16})$$

The first term on the right-hand side of eq. (S16) represents the accumulation of dimers due to the gradient of the flux of the dimers. The second term represents the spontaneous unloading of dimers and k_{off}^d is the rate constant that accounts for the unloading process. The flux $J_d(z)$ of dimers has the form

$$J_d(z) = -D_c \frac{\partial}{\partial z} \psi_d(z) - \frac{D_c}{k_B T} \psi_d(z) \frac{\partial}{\partial z} \Pi_{\text{osm}}(z). \quad (\text{S17})$$

The first term of eq. (S17) is due to the thermal diffusion and the second term is due to the osmotic pressure of cohesin monomers. We assumed that the cohesin dimers are only rarely loaded onto the chromatin region, see also sec. S1. The boundary conditions to solve eq. (S16) are i) the flux of dimers is zero at $z = M$ due to the BEs, $J_d(M) = 0$, and ii) the flux of dimers is equal to the loading rate of the dimers, $J_d(0) = k_{\text{on}}^d c^2 (1 - \psi_m(0))^2$.

We derive the rescaled probability distribution $\tilde{\psi}_d(z)$ ($\equiv M k_{\text{on}}^{m2} \psi_d(z) / (D_c k_{\text{on}}^d)$) of dimers as functions of the dimensionless parameters $k_{\text{on}}^m c M / D_c$, $k_{\text{off}}^m M^2 / D_c$, and $k_{\text{off}}^d M^2 / D_c$ by numerically solving eqs. (S13) and (S16). Our theory predicts that the probability distribution function $\psi_d(z)$ is a monotonically increasing function of the position z for cases in which the unloading rate constant

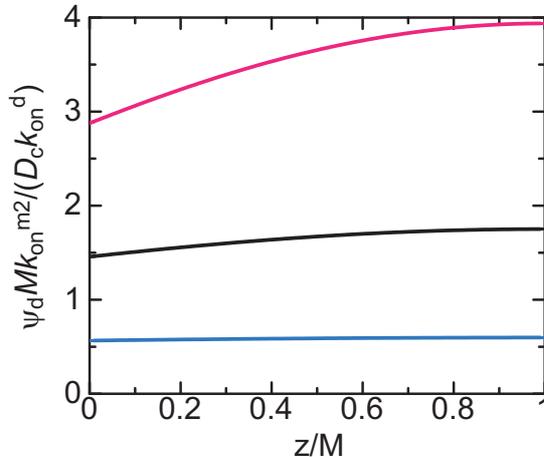


Figure S3: The probability $\psi_d(z)$ of finding a cohesin dimer at each binding site (rescaled by $D_c k_{on}^d / (M k_{on}^m)$) is shown as a function of the position z for $k_{off}^m M^2 / D_c = 0.5$ (blue), 1.0 (black), and 2.0 (magenta). We used $k_{on}^m cM / D_c = 3.0$ and $k_{off}^d M^2 / D_c = 0.3$ for the calculations.

$k_{off}^d M^2 / D_c$ is relatively small, see fig. S2, and $\psi_d(z)$ is a monotonically decreasing function of the position z for cases in which the unloading rate constant $k_{off}^d M^2 / D_c$ is relatively large (not shown).

Substituting eq. (S17) into eq. (S16) leads to the form

$$\frac{\partial^2}{\partial \tilde{z}^2} \tilde{\psi}_d + \frac{1}{1 - \psi_m} \frac{\partial \psi_m}{\partial \tilde{z}} \frac{\partial \tilde{\psi}_d}{\partial \tilde{z}} + \left[\frac{M^2 k_{off}^m}{D_c} \psi_m - \frac{M^2 k_{off}^d}{D_c} \right] \tilde{\psi}_d(z) = 0, \quad (\text{S18})$$

where we used the rescaled coordinate $\tilde{z} (\equiv z/M)$ and rescaled probability distribution function $\tilde{\psi}_d(z) (\equiv M k_{on}^m \psi_d(z) / (D_c k_{on}^d))$ of dimers. The boundary condition i) (see below eq. (S17)) leads to the fact that the gradient $\partial \psi_d(z) / \partial z$ is zero at $z = M$. Eq. (S18) implies that $\psi_m(M) > k_{off}^d / k_{off}^m$ is the necessary condition for the probability distribution $\psi_d(z)$ to be a monotonically increasing function of the position. Indeed, it is also the sufficient condition. The probability distribution $\psi_d(z)$ of dimers is a monotonically decreasing function when the gradient $\partial \psi_d(z) / \partial z$ is negative at $z = 0$, see the broken curve in fig. S4. The probability distribution $\psi_d(z)$ is a non-monotonic function of the position z in a parameter region, which is delineated by the two conditions, see fig. S4.

To summarize, the physics does not change with the cases in which cohesin monomers and dimers are unloaded spontaneously on the track as long as the detailed balance is broken at the cohesin loading site $z = 0$. There are a couple of differences in details: i) there is a window of parameters at which the distribution function of cohesin dimers is a non-monotonic function of the position z and ii) the gradient of the distribution function of cohesin monomers and dimers

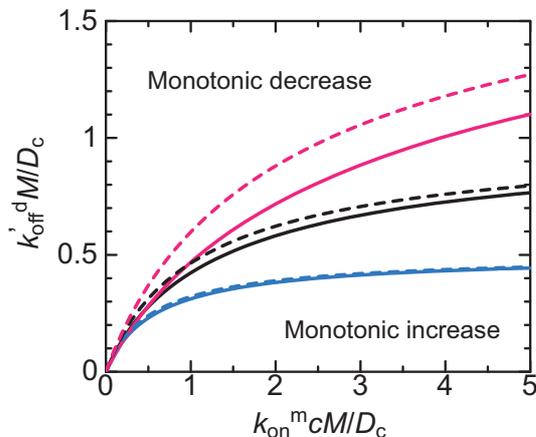


Figure S4: The state diagram of the distribution of cohesin dimers is shown as a function of the rescaled unloading rate $k'_{\text{off}}{}^{\text{d}}M^2/D_c$ of dimers and the rescaled loading rate $k_{\text{on}}{}^{\text{m}}cM/D_c$ of monomers. The probability distribution function $\psi_{\text{d}}(z)$ is a monotonically increasing function when the rescaled unloading rate $k'_{\text{off}}{}^{\text{d}}M^2/D_c$ of dimers is smaller than the solid curve and is a monotonically increasing function when the rescaled unloading rate $k'_{\text{off}}{}^{\text{d}}M^2/D_c$ of dimers is smaller than the broken curve. The probability distribution function $\psi_{\text{d}}(z)$ is a non-monotonic function in the region between the broken and solid curves. We used $k'_{\text{off}}{}^{\text{d}}M^2/D_c = 0.5$ (blue), 1.0 (black), 2.0 (magenta) for the calculations.

with respect to the position z is zero at $z = M$ because BEs (such as CTCFs with converging orientation) stop the motion of cohesin monomers and dimers.

S3 Conformational entropy of chromatin

We here take into account the conformational entropy of the chromatin fiber in an extension of the model presented in the main article. We assume that not more than one cohesin dimer is loaded on the stretch of chromatin fiber at any moment. The loaded cohesin dimer produces a loop of length $z_+ + z_-$ and the two free arms have length, N_+ ($= M - z_+$) and N_- ($= M - z_-$), see fig. 1 in the main article. Here and after, we use the subscripts $+$ and $-$ to represent the quantities of the free arms at the positive ($z > 0$) and negative ($z < 0$) halves, respectively. These lengths are represented by the number of chain segments. For simplicity, we assume that the size a of a binding site (defined in the main article) is equal to the length of a chain segment. We neglect the excluded volume interactions between chain segments.

With this treatment, the number of configurations of the free arms has the

forms

$$W_+ = \omega^{N_+} \quad (\text{S19})$$

$$W_- = \omega^{N_-}, \quad (\text{S20})$$

where ω is the number of the nearest neighbor sites in a lattice model. The number of configurations of the loop has the form

$$W_{\text{lp}} = c \left(\frac{l_a}{R} \right)^3 \omega^{z_+ + z_-}, \quad (\text{S21})$$

where l_a is the lattice constant (or Kuhn length) and $R (= \sqrt{z_+ + z_-} l_a)$ is the end-to-end distance of a free polymer of length $z_+ + z_-$. The factor $(l_a/R)^3$ accounts for the fact that the two ends of a loop are at the same position. c is a constant of the order of unity (which accounts for the fact that the number of configurations depends on the positions of the end points). With eq. (S19) - (S21), the free energy contributions of the chromatin fiber conformation has the form

$$F_{\text{con}} = -2Mk_B T \log \omega + \frac{3}{2} k_B T \log(z_+ + z_-). \quad (\text{S22})$$

This leads to the force applied to each ring of cohesin dimer in the form

$$\begin{aligned} f_{\text{con}} &= -\frac{\partial}{\partial z_+} F_{\text{con}} \\ &= -\frac{3}{2} \frac{k_B T}{z_+ + z_-}. \end{aligned} \quad (\text{S23})$$

The conformational entropy of the chromatin fiber thus tends to decrease the size of chromatin loops by counteracting the osmotic pressure generated by cohesin monomers. The force f_{con} is not applied to cohesin monomers and thus eqs. (1) - (3) in the main article are still effective.

Eqs. (S22) and (S23) are effective for cases in which the loop length $z_+ + z_-$ is larger than the order of unity. For cases in which the loop length is smaller, the bending elastic energy dominates the free energy contributions of the conformational entropy, see refs. [2] and [3]. The bending elastic energy has an approximate form

$$\frac{F_{\text{ela}}}{k_B T} \simeq \frac{\pi^2}{z_+ + z_-}, \quad (\text{S24})$$

which is derived by assuming that the loop is circular (the numerical factor π^2 in eq. (S24) is smaller than a more exact calculation [2, 3] by the factor of about 0.7) and using the fact that the Kuhn length l_a is twice the persistence length. It is important to note that the cohesin dimer starts with a loop of finite size and thus finite energy. By using eq. (S24), the energy barrier to initiate a loop is estimated to the order of $\frac{\pi^2}{2} k_B T$. Because microscopic details of loop

initiation by a cohesin dimer are unknown, we do not discuss the small loop regime any further.

With eq. (S23), the flux J_d of a cohesin dimer on the fiber has the form

$$J_d = -D_c \frac{\partial}{\partial z} \psi_d(z) - \frac{D_c}{k_B T} \psi_d(z) \frac{\partial}{\partial z} \Pi_{\text{osm}}(z) - \frac{3}{4} D_c \frac{\psi_d(z)}{z}, \quad (\text{S25})$$

where we assumed that the two rings of the dimer diffuse randomly and the positive and negative halves of the chromatin stretch are symmetric, $z_+ = z_- = z$. The boundary conditions to solve eq. (S25) have the forms

$$J_d = k_{\text{on}}^d c (1 - \psi_m(0))^2 \quad (\text{S26})$$

$$= k_{\text{off}}^d \psi_d(M), \quad (\text{S27})$$

which account for the fact that the binding and unbinding rates of cohesin dimers are both equal to the flux J_d in the steady state. With these boundary conditions, the probability $\psi_d(z)$ has the form

$$\psi_d(z) = \frac{J_d}{k_{\text{off}}^d} \left(\frac{z}{M} \right)^{-3/4} e^{J_m(z-M)/D_c} (1 - g(z)) \quad (\text{S28})$$

with

$$J_d = \frac{k_{\text{on}}^d J_m^2}{k_{\text{on}}^{m2}} \quad (\text{S29})$$

$$g(z) = \frac{M k_{\text{off}}^d}{D_c} \int_1^{z/M} du u^{3/4} e^{-J_m M(u-1)/D_c}. \quad (\text{S30})$$

The probability distribution $\psi_d(z)$ corresponds to the average frequency of contact between the segments at $-z$ and z . Our theory predicts that the probability distribution $\psi_d(z)$ shows a divergence at $z \rightarrow 0$, see fig. S5. This is an artefact due to our approximation, which treats the positions z of cohesin binding sites as a continuous variable. The free energy of producing a small loop is dominated by the bending elastic energy, see also eq. (S24) and discussion below. We thus expect that the probability $\psi_d(z)$ shows a maximum of finite height at the crossover between the energetic and entropic regime. The conformational entropy does not change the height of the peak at $z = M$, because it is determined by the boundary condition, but makes the peak less steep, again due to the entropic force f_{con} . These results are effective for cases in which 1) the statistics of the chain conformation is given by the Gaussian distribution, 2) the conformational dynamics of the chromatin fiber is faster than the diffusion of cohesin dimers, and 3) not more than one cohesin dimer is loaded on the stretch of chromatin fiber.

References

- [1] S. A. Safran, *Statistical thermodynamics of surfaces, interfaces, and membranes*, Westview Press, Boulder, CO, 2003.

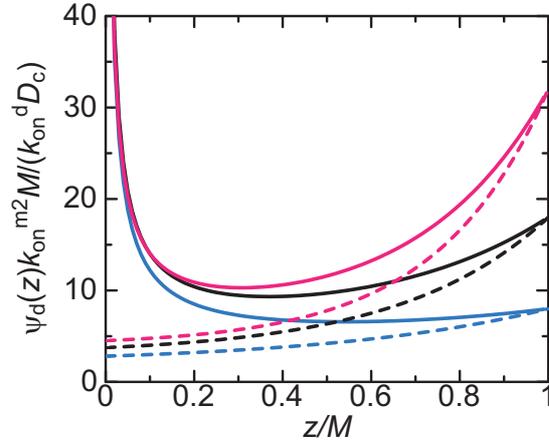


Figure S5: The probability $\psi_d(z)$ of finding a cohesin dimer (rescaled by $k_{\text{on}}^{m^2}M/(k_{\text{on}}^d D_c)$) is shown as a function of the rescaled position z/M . The broken and solid curves are predicted by using eq. (5) in the main article and eq. (S28), respectively. The rescaled rate constant $Mk_{\text{on}}^m c/D_c$ is set to 3.4 (cyan), 150.6 (black), 1092.0 (magenta). We used $Mk_{\text{off}}^d/D_c = 0.5$ and $Mk_{\text{off}}^m/D_c = 5.0$.

[2] H. Schiessel, *J. Phys.: Condens. Matter*, 2003, **15**, R699.

[3] H. Yamakawa and W. H. Stockmayer, *J. Chem. Phys.*, 1972, **57**, 2843.