Supplementary Information

On the theory of condensin mediated loop extrusion in genomes

Ryota Takaki

Physics Department, The university of Texas at Austin

Atreya Dey, Guang Shi, and D. Thirumalai

Chemistry Department, The university of Texas at Austin

(Dated: May 27, 2020)

CONTENTS

I. SIMULATIONS

The purpose of the simulations is to show that the extracted parameter values by fitting the theoretical extrusion rate or equivalently the velocity of loop extrusion (LE) to experiments are reasonable. In particular, we use simulations to argue that the value of $\Delta R \approx 26$ nm (see the main text for details) is consistent with the known condensin architecture. To this end, we imagine that during the ATPase cycle the SMC motor undergoes a conformational change from a "open" (top structure in Fig. S1) to a "closed" (bottom structure) state, which brings the motor domains close to the hinge region. This process is allosterically driven, in a manner similar to other cargo-carrying motors (myosins, kinesins and dynein), by binding and hydrolysis of ATP. We envision that in the SMC the allosteric transitions are effectuated through the movement of the flexible elbow region.

A. Model for Condensin

We modeled the two heads of condensin as spheres that are connected by finitely-extensible nonlinear elastic (FENE) potential [1] to the coiled coils (CCs) that connect the motor domains to the hinge (Fig.S1). The CC in the SMCs are reminiscent of the lever arm in Myosin V. The angle θ_1 and θ_2 are formed at the junctions connecting the motor heads to the first bead on the CCs (Fig.S1). As in molecular motors, a change in the conformation change initiated in the head domain is somewhat amplified over condensin through the CCs. We envision this process as the principle mechanism by which a spool (roughly $\frac{\Delta R}{0.34} = 76$ base pairs (bps) in a single step) of dsDNA could be extruded.

We used 19 and 18 beads for upper CC and lower CC, respectively. The diameter of each bead is 1 nm diameter each. We used 3 beads (diameter 0.4 nm each) in the middle of the CCs for the elbow region. The angle potential in the elbow region, marking the break in an otherwise stiff CC, is sufficiently weak to facilitate the allosteric propagation of conformational changes in the motor head. For the hinge and two motor heads we used 4 nm diameter beads.

All lengths are measured in units of $\sigma = 1$ nm corresponding to the diameter of the beads in the CC. We express energy in the unit of k_BT, where k_B is the Boltzmann constant and T is the temperature. The mass of all the particles were set to $m = 1$. We performed

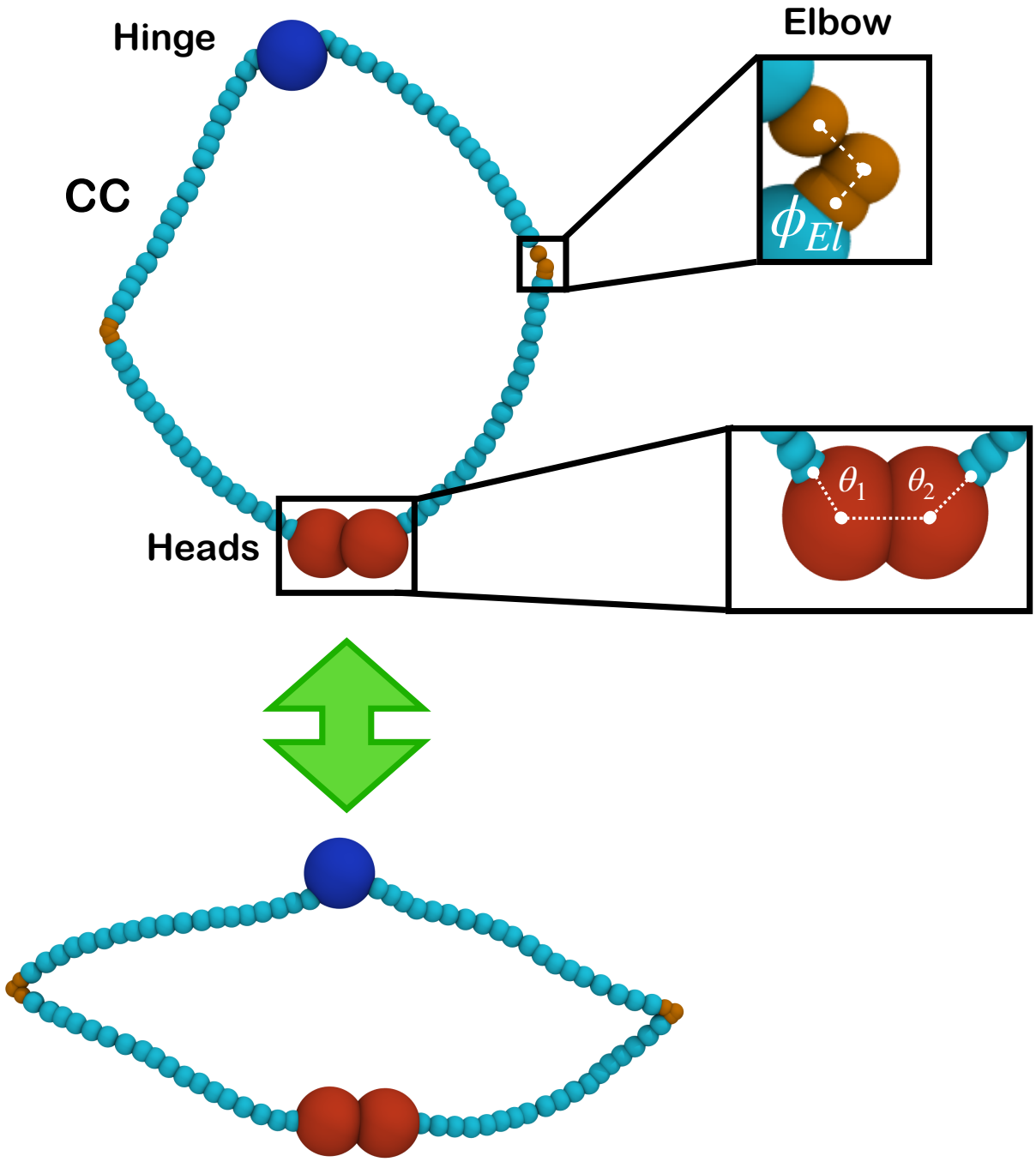

FIG. S1. Caricature of the condensin motor, an example of the Structural Maintenance of Chromosomes (SMC), used in the simulations. The two heads are shown as red spheres. A magnified image of the angle between the motor heads at the junctions to the two arms of the SMC are shown in the lower box. The angle at the elbow is depicted in the upper box. The coiled coils (CCs) connecting the motor to the hinge (purple sphere) are treated as a semi-flexible polymers that are kinked at the flexible elbow region. We envision that the allosteric transition between the open and the closed states (shown as by the green arrow) is driven by ATP binding and hydrolysis to the motor domains.

low-friction Langevin dynamics simulations using OpenMM [2] software using a time-step of $\Delta t_L = 0.01\tau_L$, where $\tau_L = 0.4\sqrt{m\sigma^2/k_BT}$. The value of the friction coefficient is $0.01/\tau_L$. The friction coefficient was chosen to be as low as possible for a stable simulation.

B. Energy function

Because the goal is to merely illustrate that the hypothesized allosteric mechanism for SMC-mediated LE is plausible, we chose a simple energy function to monitor the conformational changes in condensin. The explicit form of the energy function is,

$$E(\vec{r}_1, \vec{r}_2, ..., \vec{r}_N, \vec{\phi}, \vec{\theta}) = \sum_{i=1}^{N-1} U_{FENE}(r_{i,i+1}) + \sum_{i\neq j}^{N} U_N(r_{i,j}) + \Big(\sum_{i\in CC\neq El} U^{CC}_{ANG}(\phi_i) + \sum_{i\in El} U^{El}_{ANG}(\phi_i) \Big) + \sum_{i\in Head} U_{CNF}(\theta_i). \tag{S1}$$

The first term in Eq. S1 enforces the connectivity of the beads and is given by,

$$U_{FENE}(r_{i,i+1}) = -\frac{1}{2} k_F R_F^2 \log\Big[1 - \frac{(r_{i,i+1} - r^0_{i,i+1})^2}{R_F^2}\Big], \tag{S2}$$

where k_F is the stiffness of the potential, R_F is the upper bound for the displacement, and $r^0_{i,i+1}$ is the equilibrium distance between the beads, i and $i+1$. The second term in Eq.S1, accounting for excluded volume interactions, is given by, $\mathrm{U}_N(r_{i,j}) = \epsilon_N \Big(\frac{\sigma}{r_{i,j}}\Big)^{12}$, where ϵ_N and σ are the strength and range of the interaction, respectively. We used additive interactions, which means that σ is the sum of the radii of the two interacting beads. The third and the fourth terms in Eq.S1 are the two angle potentials that control the bending stiffness of the CCs. The potential $U_{ANG}(\phi_i)$ is taken to be,

$$U_{ANG}(\phi_i) = \epsilon_b(1 + \cos\phi_i), \tag{S3}$$

where ϵ_b, the energy scale for bending, is related to the persistence length of the semi-flexible CC (see Section.II). We used a different value of ϵ_b for $U^{CC}_{ANG}(\epsilon^{CC}_b)$ and $U^{El}_{ANG}(\epsilon^{El}_b)$ to realize the difference in the stiffness between the elbow region, and the rest of the CC. Because the persistence of the CC, l^{CC}_p, is not known we varied ϵ^{CC}_b to cover a range of plausible values

Parameter	Value
k_F	$50(k_BT/\text{nm}^2)$
R_F	$1.5(\text{nm})$
ϵ_b^{CC}	$4, 15, 30, 45, 60, 75, 90,$ $105, 120, 135, 150(k_BT)$
ϵ_b^{El}	$4(k_BT)$
ϵ_N	$5(k_BT)$
k_C	$100(k_BT/\text{rad}^2)$

TABLE S1. Parameters for the molecular dynamics simulation.

of l_p^{CC} (see Table S1).

The last term in Eq.S1 models the conformation change in the motor head of condensin due to ATP binding, and is taken as,

$$U_{CNF}(\theta_i) = k_C\big(\theta_i^0 - \theta_i\big)^2, \tag{S4}$$

where k_C is the spring constant for the potential, and θ_i^0 is the equilibrium angle for the angle potential. Before the conformational change we set $\theta_i^0 = 2.4$ (radian) in the open state, which is roughly the angle calculated by ATP engaged state of prokaryotic SMC [3]. Because the structure for closed state is unavailable we chose $\theta_i^0 = 4.0$ (radian) for closed state, which leads to $\theta_i \sim \pi$ (radian) in equilibrium configuration for condensin (see Fig.S5), in order to obtain a large conformational change. The transition between the open and closed states results in the scrunching of the DNA and extrusion of the loop. The parameter values in the energy function used in the simulations are in Table S1.

II. PERSISTENCE LENGTH OF THE COILED-COIL

We calculated the persistence length (l_p^{CC}) for coiled-coil that suffices to amplify using simulations for a single isolated semi-flexible polymer with the elbow. The single polymer has 40 beads in total with 3 flexible beads, which provides a coarse-grained descrition of the CC for condensin (Sec.I A). We set $\epsilon_b^{El} = 4(k_BT)$ and varied ϵ_b^{CC} (see Table.S1). Since the CC in our model has flexible kink at the elbow, we computed the effective persistence length using end-to-end distance of the semi-flexible polymer. Namely, we obtained contour

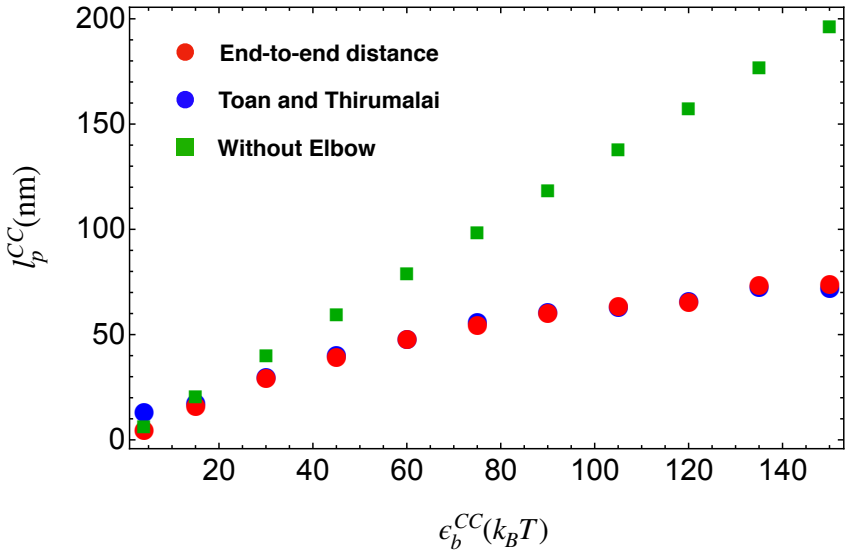

FIG. S2. Persistence length (l_p^{CC}) as a function of bending stiffness (ϵ_b^{CC}). Red circles; l_p^{CC} calculated using end-to-end distance. Blue circles; l_p^{CC} calculated using the correlation of tangent angle used in Toan and Thirumalai [4]. Green squares are for a semi-flexible polymer *without* flexible elbow.

length (L) and end-to-end distance (R) from simulations then numerically solved $< R^2 >= 2l_pL\big(1 - \frac{l_p}{L}(1 - e^{-l/l_p})\big)$ [5] for l_p. We also computed the persistence length from the decay of tangent angle correlation used previously in simulating a model for DNA [4]. We found that the two methods gives consistent value of effective persistence length (see Fig.S2).

Fig.S2 shows the persistence length of the CC as a function of ϵ_b^{CC}. Without the flexible elbow, the persistence length is well approximated by $l_p = l_b\epsilon_b/(k_BT)$, where $l_b \sim 1.3$ nm is mean bond length, as described elsewhere [6–8]. In the presence of the kink at the elbow the effective persistence length becomes shorter.

III. DISTRIBUTION OF HEAD-HINGE DISTANCE

In Fig.5 in the main text we showed the distribution of head-hinge distance for $l_p^{CC} \sim 70$ nm. Here we list distributions for different values of l_p^{CC} (Fig.S3). We also plot the mean value of the change of head-hinge distance between open and closed state ($\Delta R_s = R_1 - R_2$) as a function of l_p^{CC} (Fig.S4). For $l_p^{CC} \sim 5$ nm (Fig.S3(a)), we find the two distributions for open state and closed state overlaps. This is because the persistence length of CC is too small to propagate the conformational change initiated at the head domain. As l_p^{CC} increases the

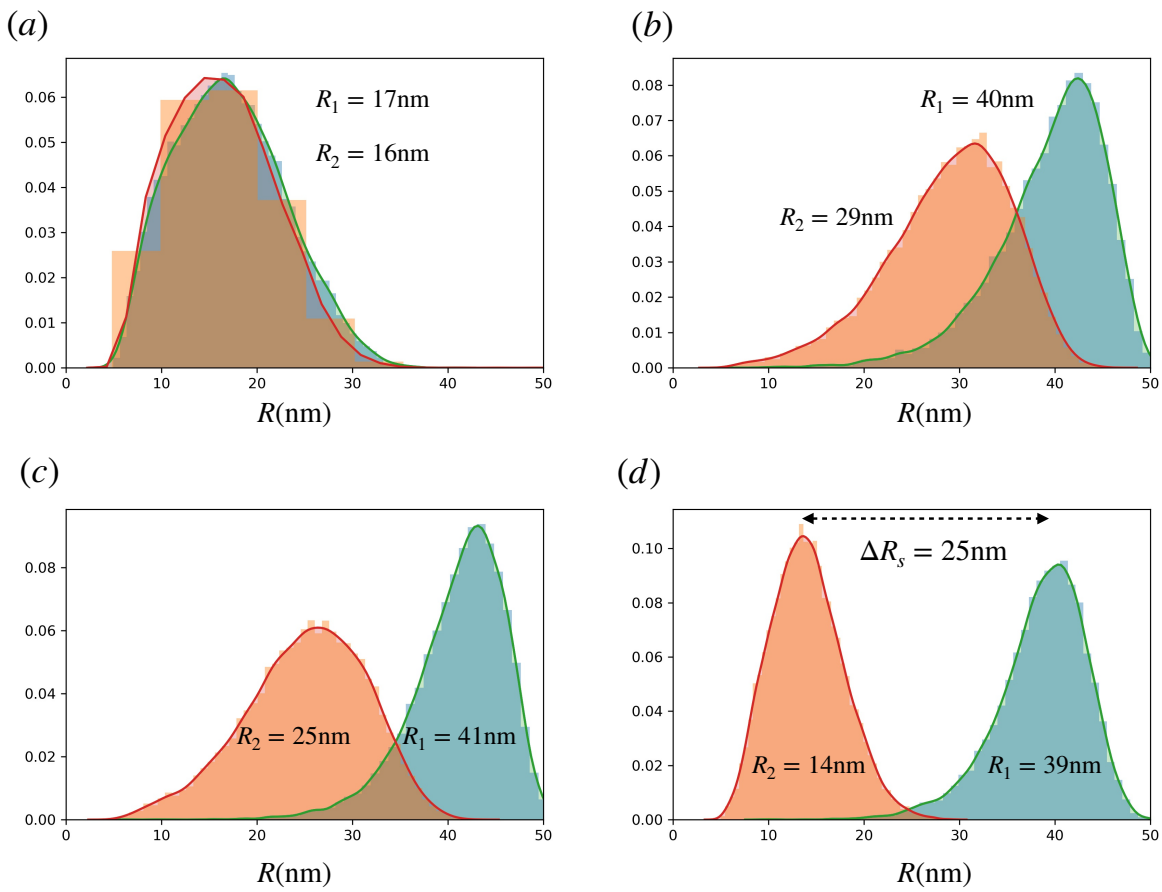

FIG. S3. Distributions of R for various l_p^{CC}. R_1 and R_2 are mean value of head-hinge distance for open state and closed state, respectively. The distributions are calculated from 50 trajectories, 40000 time points. (a) For $l_p^{CC} \sim 5$ nm, $\Delta R_s = 1 \pm 8$ nm, where error is calculated using standard deviation of the distributions. (b) For $l_p^{CC} \sim 40$ nm, $\Delta R_s = 11 \pm 9$ nm. (c) For $l_p^{CC} \sim 60$ nm, $\Delta R_s = 16 \pm 8$ nm. (d) For $l_p^{CC} \sim 90$ nm, $\Delta R_s = 25 \pm 6$ nm.

separation of the distributions become distinct. We note that for $l_p \sim 90$ nm, ΔR_s reaches 25 nm, which is comparable to the conformation change estimated in our theory.

Fig.S4 shows ΔR_S as a function of l_p^{CC}. ΔR_s linearly increases with l_p^{CC} reaching ~ 25 nm at $l_p^{CC} \sim 90$ nm.

IV. CONDENSIN POWER STROKE

A salient feature of molecular motors is the conversion of the chemical energy released due to ATP hydrolysis to mechanical work, which is often accompanied by a power stroke involving conformational changes. Processive motors, such as kinesin, myosin or dynein,

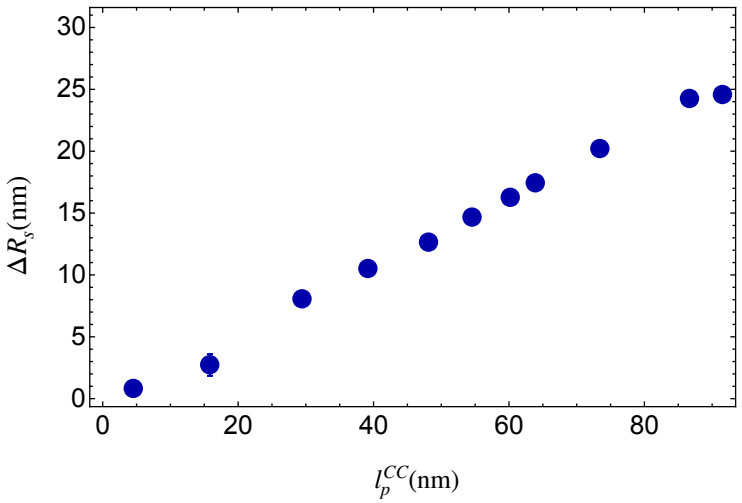

FIG. S4. Difference of head-hinge distance between open and closed state ($\Delta R_s = R_1 - R_2$) as a function of l_p^{CC}. Each points are calculated from 5 trajectories (4000 time points). Error bars are mostly smaller than the plot marker.

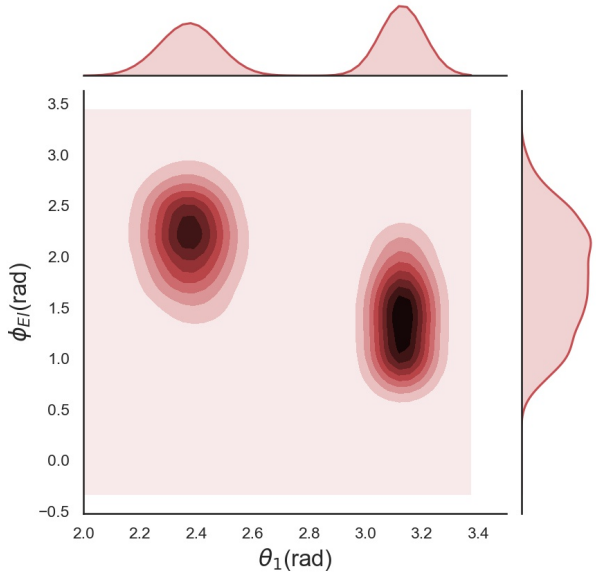

FIG. S5. Density distribution for the angles for head (θ_1) and for elbow (ϕ_{El}) from simulation. We plot the two dimensional distribution of angle from 50 trajectories, 40000 sample points, for $\epsilon_b^{CC} = 150(k_BT)$ corresponding $l_p^{CC} \sim 70$ nm. The two maximum are $(\theta_1, \phi_{El}) = (2.4, 2.1)$ and $(\theta_1, \phi_{El}) = (3.1, 1.4)$. On top and the right side we show the distribution of θ_1 and ϕ_{El}, respectively.

undergo dynamic allosteric transitions initiated by ATP binding to the motor head. We posit that a similar power stroke mechanism that produces allosteric transitions must also be operative in the SMC class of motors to translocate along the DNA, thus extruding loops. Based on AFM experiments [9, 10] and structural studies [3] we envision that open

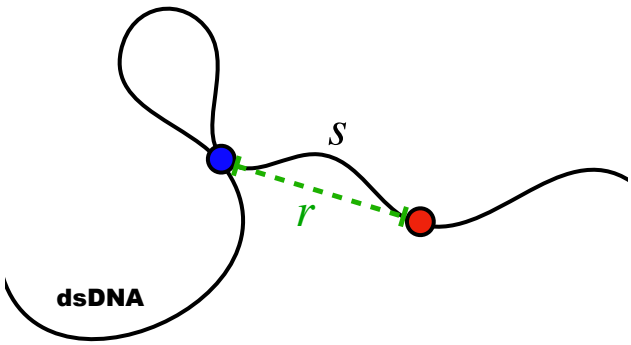

FIG. S6. A picture of a conformation of condensin bound to two loci separated by a genomic distance s (extruded loop length). The spatial distance between the attachment points in the DNA is r. For LE to occur condensin has to engage with at least two loci on the DNA.

to close transition condensin (scrunching process) corresponds to condensin power stroke. The scrunching is assumed to be initiated at heads (ATPase domain), which is amplified through CC.

In the main text, we obtained head-hinge distance as a measure of the power stroke. We illustrate in Fig.S5 the distribution for the angle for the head, θ_1, and for the elbow, ϕ_{El}, (see Fig.S1 for the definition of angles) calculated from the simulation trajectories. There is a clear separation in the two dimensional distribution of the angles θ_1 and ϕ_{El} (Fig.S1). The distribution of θ_1 in both the open (O shape displayed in top structure in Fig.S1) state and the closed state (B shape - a terminology used in [9] to describe the structure in the bottom of Fig.S1) is narrower than the fluctuations of ϕ_{El}. Thus, even using the simple model we find that conformations changes in the head is transmitted through the elbow leading to the open to closed transition (Fig. S1).

V. DERIVATION OF $P(\mathcal{L}|\mathcal{R})$

A major ingredient in the theory (see Eq.(1) in the main text) is the calculation of the contour length extruded loop as condensin is powered by ATP binding to the motor head and subsequent hydrolysis. To obtain Eq.(1) in the main text let us consider condensin separated by the spatial distance r that pinches a loop whose genomic length is s (Fig.S6). Given the distribution of the spatial distance r between two loci separated by a linear genomic distance s, $P(r|s)$, we would like to derive the distribution of s $P(s|r)$. Indeed, $P(s|r)$ is the probability of extruded length of dsDNA, s by condensin whose DNA binding domains

are separated by the distance, r. According to the Bayes's rule, we have $P(r|s)P(s) = P(s|r)P(r)$. The normalization dictates that $P(r) = \int_0^L P(r|s)P(s)\mathrm{d}s$. These two equations lead to,

$$P(s|r) = \frac{P(r|s)P(s)}{\int_0^L P(r|s)P(s)\mathrm{d}s}.$$

We assume that there is no preference for picking any specific genomic distance s on DNA (we do not account for sequence specific preference). Consequently, we take $P(s) = 1/L$. Therefore,

$$\begin{aligned} P(s|r) &= \frac{(1/L)P(r|s)}{(1/L)\int_0^L P(r|s)\mathrm{d}s} \\ &= \frac{P(r|s)}{\int_0^L P(r|s)\mathrm{d}s}. \end{aligned}$$

Thus, $P(s|r)$ and $P(r|s)$ differ only by a constant, $\int_0^L P(r|s)\mathrm{d}s$, if we consider a fixed r. It is clear that $P(r|s)$ is the radial probability density for the interior segments separated by a distance r for a semi-flexible polymer, which is derived elsewhere [11]. For the case $s = L$, $P(r|s)$ is the result for the end-to-end distribution for semi-flexible chains, $P(\mathcal{R}|\mathcal{L})$ [12]. It is known that the simple analytic result for $P(\mathcal{R}|\mathcal{L})$ [12] is accurate when compared to the exact result [13] or numerical simulations. Thus, we employ the simpler expression $P(\mathcal{R}|\mathcal{L})$ and assume that $P(\mathcal{L}|\mathcal{R})$ is equivalent to $P(\mathcal{R}|\mathcal{L})$ up to a normalization constant when expressed in terms of L with fixed R.

VI. LOAD DEPENDENCE OF LE VELOCITY

In the experiment [14] the load acting on condensin is obtained indirectly using relative extension of dsDNA. However, in principle, it is possible to measure the direct load dependence of condensin by using a different experimental set up such as optical tweezers. We plotted using Eq.(4) in the main text, the LE velocity, directly as a function of external load f in Fig.S7. We also converted the experimental data from x to f using Eq.(5) in the main text. It is evident from Fig.S7 that condensin is a weak motor with a high response to the external load. Note that the rate of extrusion significantly drops by $f = 0.4$ pN, which is

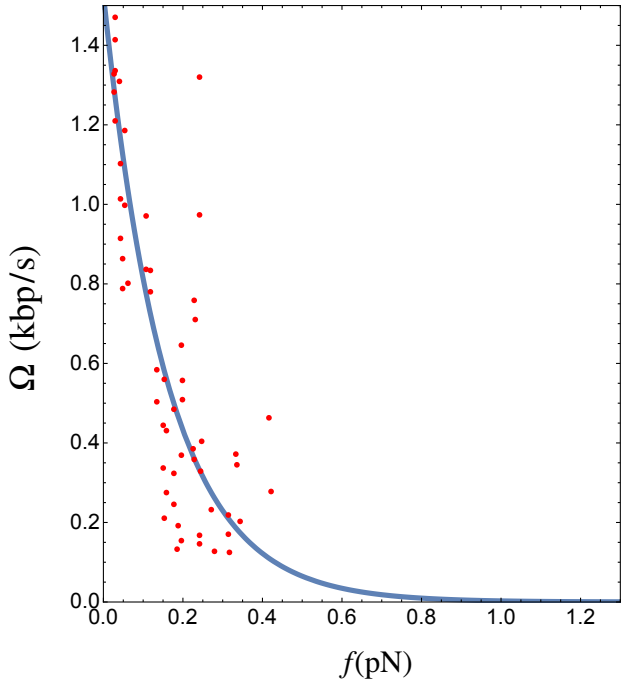

FIG. S7. Comparison of the velocity of LE predicted by theory (blue line) with the data (red dots) extracted from experiment [14].

consistent with previous experiments [15–17]. It appears that the stall force is no more than about $f_s = 1.0$ pN.

VII. EFFECT OF VARIABLE PERSISTENCE LENGTH FOR DNA

In the main text we used $l_p = 50$ nm as the persistence length of DNA, which is widely accepted value for dsDNA [18]. However, it could be interesting to explore l_p, which can be drastically altered in the presence of divalent cations, as a variable in our theory. In Fig.S8(a) we plotted $P(\mathcal{L}|R = 50 \text{ nm})$ using Eq.(1) in the main text for different l_p. As dsDNA become flexible the distribution of $P(\mathcal{L}|R = 50 \text{ nm})$ becomes wider, suggesting that most probable value of captured length of DNA by condensin would be larger with a large dispersion. Thus, in this situation our approximation, $\Delta l \approx \Delta R$, would become less accurate. Nevertheless, we can explore the velocity of extrusion for different l_p shown in Fig.S8(b) for a fixed $\Delta R = 26$ nm. As l_p decreases the velocity of extrusion becomes linear and slower because the load acting on DNA is higher for smaller l_p at the same extension.

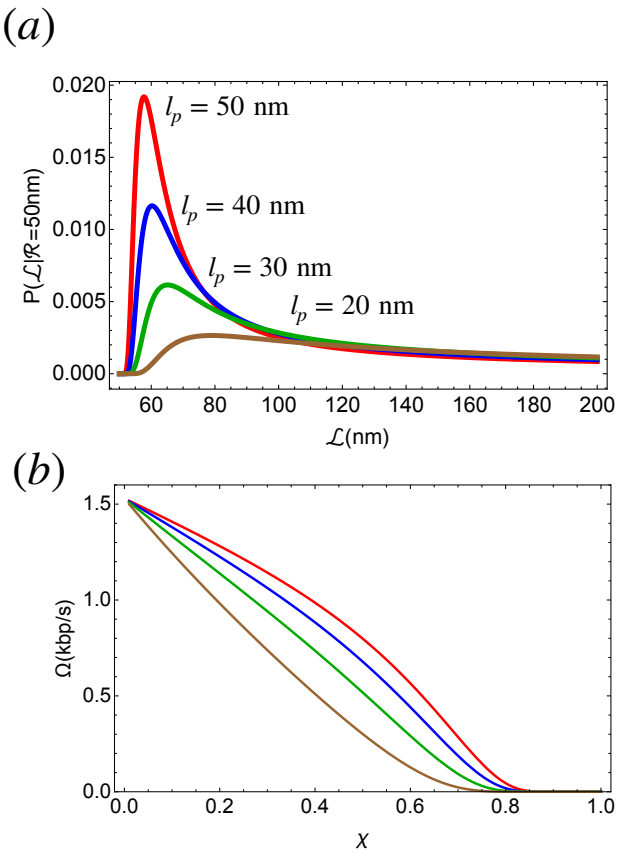

FIG. S8. Effect of variable persistence length for dsDNA. (a) Plot of $P(\mathcal{L}|R = 50\text{ nm})$. $l_p = 50$ nm (red), $l_p = 40$ nm (blue), $l_p = 30$ nm (green), and $l_p = 20$ nm (brown). (b) Extrusion rate of DNA for different l_p. $l_p = 50$ nm (red), $l_p = 40$ nm (blue), $l_p = 30$ nm (green), and $l_p = 20$ nm (brown). ΔR is fixed to be 26 nm $\sim$ 76 bps.

[1] K. Kremer and G. S. Grest, The Journal of Chemical Physics **94**, 4103 (1991).

[2] P. Eastman, J. Swails, J. D. Chodera, R. T. McGibbon, Y. Zhao, K. A. Beauchamp, L.-P. Wang, A. C. Simmonett, M. P. Harrigan, C. D. Stern, *et al.*, PLoS computational biology **13**, e1005659 (2017).

[3] M.-L. Diebold-Durand, H. Lee, L. B. R. Avila, H. Noh, H.-C. Shin, H. Im, F. P. Bock, F. Bürmann, A. Durand, A. Basfeld, *et al.*, Molecular cell **67**, 334 (2017).

[4] N. M. Toan and D. Thirumalai, The Journal of chemical physics **136**, 06B612 (2012).

[5] O. Kratky and G. Porod, Recueil des Travaux Chimiques des Pays-Bas **68**, 1106 (1949).

[6] Z. Benková, L. Rišpanová, and P. Cifra, Polymers **9**, 313 (2017).

[7] X. Li, C. M. Schroeder, and K. D. Dorfman, Soft Matter **11**, 5947 (2015).

[8] J. Midya, S. A. Egorov, K. Binder, and A. Nikoubashman, The Journal of chemical physics **151**, 034902 (2019).

[9] J. M. Eeftens, A. J. Katan, M. Kschonsak, M. Hassler, L. de Wilde, E. M. Dief, C. H. Haering, and C. Dekker, Cell reports **14**, 1813 (2016).

[10] J.-K. Ryu, A. J. Katan, E. O. van der Sluis, T. Wisse, R. de Groot, C. Hearing, and C. Dekker, bioRxiv (2019).

[11] C. Hyeon and D. Thirumalai, The Journal of chemical physics **124**, 104905 (2006).

[12] J. Bhattacharjee, D. Thirumalai, and J. Bryngelson, arXiv preprint cond-mat/9709345 (1997).

[13] J. Wilhelm and W. Frey, Phys. Rev. Lett. **77**, 2581 (1996).

[14] M. Ganji, I. A. Shaltiel, S. Bisht, E. Kim, A. Kalichava, C. H. Haering, and C. Dekker, Science **360**, 102 (2018).

[15] J. M. Eeftens, S. Bisht, J. Kerssemakers, M. Kschonsak, C. H. Haering, and C. Dekker, The EMBO journal **36**, 3448 (2017).

[16] J. Eeftens and C. Dekker, Nature structural & molecular biology **24**, 1012 (2017).

[17] T. R. Strick, T. Kawaguchi, and T. Hirano, Current biology **14**, 874 (2004).

[18] M. Rubinstein, R. H. Colby, *et al.*, *Polymer physics*, Vol. 23 (Oxford university press New York, 2003).