Cohesin distribution alone predicts chromatin organization in yeast via conserved-current loop extrusion

Tianyu Yuan[1,2], Hao Yan[1,2], Kevin C. Li[3], Ivan Surovtsev[2,3], Megan C. King[1,3,4,*], and Simon G. J. Mochrie[1,2,5,*]

[1]Integrated Graduate Program in Physical and Engineering Biology, Yale University, New Haven, Connecticut 06520, USA

[2]Department of Physics, Yale University, New Haven, Connecticut 06520, USA

[3]Department of Cell Biology, Yale School of Medicine, New Haven, Connecticut 06520, USA

[4]Department of Molecular, Cell and Developmental Biology, Yale University, New Haven, Connecticut 06511, USA

[5]Department of Applied Physics, Yale University, New Haven, Connecticut 06520, USA

[*]correspondence to simon.mochrie@yale.edu or megan.king@yale.edu

October 31, 2023

Abstract

Inhomogeneous patterns of enhanced chromatin-chromatin contacts within 10-100 kb-sized regions of the genome are a generic feature of chromatin spatial organization. These features, termed topologically associating domains (TADs), have led to the loop extrusion factor (LEF) model, where TADs arise from loop extrusion by cohesin complexes. Currently, our ability to model TADs relies on the observation that in vertebrates TAD boundaries are correlated with DNA sequences that bind CTCF, which therefore is inferred to block loop extrusion. However, although TADs feature prominently in their Hi-C maps, non-vertebrate organisms either do not express CTCF or show few TAD boundaries that correlate with CTCF sites. In all of these organisms, the counterparts of CTCF remain unknown, frustrating comparisons between Hi-C data and simulations. To extend the LEF model across the tree of life, here, we propose the *conserved-current loop extrusion (CCLE) model* that interprets loop-extruding cohesin as a nearly-conserved probability current. Thus, we derive a position-dependent loop extrusion velocity from cohesin

ChIP-seq data alone. To demonstrate its utility in organisms lacking CTCF, we apply the CCLE model to the Hi-C maps of interphase *Schizosaccharomyces pombe* and meiotic *Saccharomyces cerevisiae*. In both cases, the model accurately predicts the TAD-scale Hi-C maps. It follows that loop extrusion by cohesin is indeed the primary mechanism underlying TADs in these systems. The model also gives new values for loop extrusion parameters such as the LEF density and processivity, which compare well to independent estimates, and provides new insights into *in vivo* LEF composition and function.

Introduction

Our knowledge of chromatin architecture has been transformed by sequencing-based chromatin-capture (Hi-C) techniques, which provide quantitative metrics of relative population-averaged contact probability between all pairs of genomic loci [1–10]. Hi-C data, typically presented as a contact map, together with theory and modeling have led to a new understanding of chromatin organization, based on two coexisting mechanisms that operate on largely different length scales: (1) On several-megabase scales, "checkerboard patterns" in Hi-C maps, encompassing distant contacts on the same chromosome and contacts on different chromosomes, have led to a block co-polymer-inspired picture of chromatin compartments, comprised of several types of epigenetically-distinguished heterochromatin and euchromatin, which each exhibits preferential affinity for like-regions [11–24]. (2) At smaller length scales – tens of kilobases – overlapping squares of high contact probability within definite regions of the same chromosome, termed topologically associating domains (TADs), have led to the loop extrusion factor (LEF) model [25–33], in which LEFs first bind to the chromatin polymer, and then initiate ATP-dependent loop extrusion by moving their two anchor points away from each other. Loop extrusion at an anchor stalls when the anchor encounters either another anchor or a so-called boundary element (BE). LEFs can also dissociate from chromatin. These processes collectively give rise to a dynamic steady-state of chromatin loops, in turn leading to TADs in Hi-C maps [26, 27, 30, 34].

Depletion of cohesin, a member of the structural maintenance of chromosomes (SMC) complex family, leads to the disappearance both of TADs and of their accompanying enhanced contact probabilities [35, 36], thus identifying cohesin as the predominant LEF. Bolstering cohesin's LEF identity, single-molecule experiments show that cohesin possesses ATP-dependent loop extrusion activity on DNA *in vitro* [37–39]. In vertebrates, the locations of TAD boundaries show a strong correlation with binding sites of the DNA-binding protein, CTCF [40, 41], while depletion of CTCF, causes loss of most TAD boundaries [36, 40]. These observations suggest that CTCF is the most important BE in vertebrates. Indeed, loop extrusion simulations using boundary elements, whose locations are defined by peaks in the CTCF chromatin immunoprecipitation sequencing (ChIP-seq) signal, are able to recapitulate many aspects of experimental vertebrate Hi-C maps [26, 27, 30, 34].

However, although TADs feature prominently in Hi-C maps across the tree of life, many non-vertebrate organisms either do not express CTCF orthologs, including yeasts (*Schizosaccharomyces pombe* and *Saccharomyces cerevisiae* [42]), plants (*Arabidopsis thaliana* [43] and *Oryza sativa* [44]), and *Caenorhabditis elegans* [45], or show only a limited number of TAD boundaries that correlate with CTCF binding sites, as in the case of *Drosophila melanogaster* [46]. In all of these organisms, even if the LEF model is applicable, which remains uncertain, the identities of the boundary elements are unknown, frustrating quantitative comparisons between Hi-C data and simulations. Even in vertebrates a number of important TAD boundaries are located away from CTCF binding sites.

With the goal of modeling TADs across the tree of life beyond vertebrates, here, we introduce a novel, physics-based version of the LEF model, named the conserved-current loop extrusion (CCLE) model, that should be applicable in any organism, in which loop extrusion is important. Specifically, by interpreting loop-extruding LEFs as a probability current, that is approximately conserved at steady-state, we derive a position-dependent loop extrusion velocity, using cohesin ChIP-seq data as input, which we then incorporate into loop extrusion simulations without boundary elements. This model has intuitive appeal in that loop extrusion velocities are small at genomic locations with high cohesin ChIP-seq signal, as if the LEFs are blocked there, while the velocities are high at positions with low cohesin ChIP-seq signal, because LEFs spend little time in locations where they are not blocked. By design, CCLE is agnostic concerning the identities of BEs and other proteins that interact with cohesin.

A key inspiration for the development of CCLE was Ref. [47], which successfully simulates chromatin organization in meiotic *S. cerevisiae*, using a version of the previous vertebrate-focused LEF models, except with cohesin binding sites replacing CTCF binding sites. CCLE improves on this approach in principle, by eliminating both the need to specify how to pick out cohesin binding sites from ChIP-seq data and the need to specify how such binding sites affect loop extrusion, both of which are accomplished in an *ad hoc* fashion in Ref. [47] (and its CTCF-based precursors [26, 27, 30, 34]).

To focus on the role of loop extrusion and avoid possible ambiguities associated with chromatin compartments, in this paper, we apply our model to meiotic budding yeast, following Ref. [47], and interphase fission yeast. Neither of the corresponding Hi-C maps shows a checkerboard pattern, characteristic of chromatin compartments. Using cohesin ChIP-seq data as input, we are able to match the measured Hi-C maps of *S. pombe* quantitatively, with just four fitting parameters, namely the LEF processivity in the absence of obstructions, the chromatin persistence length, the linear density of loop-extruding cohesins (ρ), and the linear density of cohesive cohesins (ρ_c). Both ρ and ρ_c are involved in transforming the cohesin ChIP-seq signal to LEF occupancy probability. For meiotic *S. cerevisiae*, we use a fifth parameter to empirically account for increased polymer volume exclusion in the more-compact meiotic chromosome.

Because the essentially simple CCLE model is capable of accurately converting cohesin ChIP-seq data

into Hi-C maps for both interphase fission yeast and meiotic budding yeast, it seems inescapable that TADs in these systems predominantly occur as a result of cohesin-based loop extrusion, and not via an alternative-to-loop-extrusion mechanism. CCLE transforms cohesin ChIP-seq data into an ensemble of fluctuating loop configurations, that define three-dimensional chromosomal organization. The excellent experiment-simulation agreement achieved by the conserved-current approach, across the entire *S. pombe* genome, sets a benchmark for LEF-model-based simulations, that seek to describe experimental Hi-C maps, and strongly suggests that the loop configurations generated, as well as the best-fit values of the density and processivity of loop-extruding cohesin, are realistic.

Results

Conserved-Current Loop Extrusion (CCLE) Quantitatively Describes TADs and Loop Configurations in Interphase *S. pombe*

CCLE model

In order to develop our approach, we envision loop extrusion as giving rise to probability currents of LEF anchors, flowing through chromatin lattice sites. Assuming two-sided loop extrusion, we are led to the following coarse-grained master equations for the probabilities, R_n and L_n, that chromatin site n is occupied by the right-moving or left-moving anchor of a LEF:

$$\frac{dR_n}{dt} = V_{n-1}R_{n-1}(1-P_n) - V_nR_n(1-P_{n+1}) + A_n - D_n, \tag{1}$$

and

$$\frac{dL_n}{dt} = U_{n+1}L_{n+1}(1-P_n) - U_nL_n(1-P_{n-1}) + a_n - d_n, \tag{2}$$

where V_n is the rate at which right-moving LEF anchors step from site n to site $n+1$, U_n is the rate at which left-moving LEF anchors step from site n to site $n-1$, $P_n = R_n + L_n$ is the probability that site n is occupied by either a left- or right-moving LEF anchor. The first and second terms on the right hand sides of each of EQ. 1 and EQ. 2 correspond to the current – *i.e.* the number per second – of LEF anchors to and from, respectively, site n by loop extrusion along the chromatin, while A_n and D_n (a_n and d_n) are the association and dissociation currents of right-moving (left-moving) LEF anchors at site n, respectively.

At steady state, assuming the difference between LEF binding and unbinding terms is small compared

to the loop-extruding terms, and that the mean probabilities of right-moving and left-moving LEF anchors being at site n are equal (i.e., $\langle R_n \rangle = \langle L_n \rangle = \frac{1}{2} \langle P_n \rangle$), and neglecting correlations among the anchor probabilities, EQ. 1 and 2 lead to (Supplementary Methods)

$$V_n = \frac{\rho L/\tau}{\langle P_n \rangle \left(1 - \langle P_{n+1} \rangle\right)} \tag{3}$$

and

$$U_n = \frac{\rho L/\tau}{\langle P_n \rangle \left(1 - \langle P_{n-1} \rangle\right)}, \tag{4}$$

where ρ is the mean LEF density in units of kb^{-1}, L is the mean LEF processivity in units of kb, and τ is the mean LEF lifetime, estimated to be of order 10^3 s in Ref. [31]. The ratio, L/τ, can be interpreted as the loop extrusion velocity of an isolated LEF along the chromatin polymer, *i.e.* twice the velocity of an isolated LEF anchor.

The utility of Eq. 3 and 4 becomes apparent, when we realize that cohesin ChIP-seq data specifies the n-dependence of $\langle P_n \rangle$. Thus, we can use cohesin ChIP-seq to calculate the position-dependent loop extrusion velocities. These velocities are subsequently used as input for Gillespie-type loop extrusion simulations (Methods), that implement loop extrusion, LEF association and dissociation, and mutual LEF blocking, as for the existing LEF models, but for which there are no explicit boundary elements. These simulations generate time-dependent loop configurations, that we average over time and over multiple independent simulations to calculate contact probability maps (Methods). Our approach neatly sidesteps needing to know the identities and positions of boundary elements (BEs) by exploiting the fact that the effect of BEs is encoded in $\langle P_n \rangle$ and, therefore, is incorporated into the LEF velocities via EQ. 3 and 4.

Functionally, there are two populations of cohesin complexes, namely trans-acting, cohesive cohesins, which give rise to cohesion between sister chromatids, and cis-acting, loop-extruding cohesins [32]. For *S. pombe* both types are present in interphase, and contribute to cohesin ChIP-seq. To determine P_n, which is the probability that chromatin site n is occupied by loop-extruding cohesin, we assume that cohesive cohesin is randomly-loaded along chromatin, giving rise to an n-independent contribution to the cohesin ChIP-seq, which we describe with a fitting parameter, ρ_c, that represents the mean density of cohesive cohesin.

Finally, to compare experimental and simulated Hi-C maps, we incorporate the polymer physics of self-contacts, within the confined volume of the nucleus, into simulated loop configurations, using a simple, albeit approximate, analytic approach, described in Methods.

CCLE simulations accurately reproduce experimental fission yeast Hi-C maps

As a first application of CCLE, we sought to describe the fission yeast Hi-C map from Ref. [6], on the basis of ChIP-seq data of the protein, Psc3, which is a component of the core cohesin complex, also from Ref. [6]. The right-hand side of Fig. 1*A* depicts a 1 Mb portion of the Hi-C map of *S. pombe*'s Chr 2 starting 0.3 Mb from the end of the left arm and extending to 1.3 Mb from the end of the left arm, using a logarithmic false-color intensity scale. Each pixel in the experimental Hi-C maps corresponds to 10 kb. The maximum interaction distance shown is 120 kb. In comparison, the left-hand side of Fig. 1*A* presents the corresponding conserved-current simulated Hi-C map using the best-fit parameters (See Table. 3 and Supplementary Methods for parameter values and optimization). The original resolution of the simulation is 1 kb, which is binned to 10 kb to match the experimental resolution. Fig. 1*B* magnifies the experiment-simulation comparison for three representative sub-regions, each 150 kb in size. In both Fig. 1*A* and *B*, visual inspection immediately reveals significant concurrence between the simulated and experimental Hi-C maps, including a high-degree of left-right symmetry, corresponding to excellent agreement between the experimental and simulated Hi-C maps. Clearly, our simulations well reproduce the inhomogeneous pattern of overlapping squares, found experimentally. In comparison to other published comparisons between Hi-C experiments and simulations at the scale of TADs [15, 21, 23, 26, 27, 35, 47], by eye, we judge the agreement displayed in Fig. 1*A* and *B* to be comparable or superior. By contrast, although the left- and right-hand sides of Fig. 1*C* and *D* look generally similar, clearly there is no left-right symmetry in these figures, which compare the experimental Hi-C maps of two different regions of Chr 2 with each other, and which therefore are expected to be dissimilar. Fig. 1*E* and *F* present maps of the ratio of the left and right sides of Fig. 1*A* and of the left and right sides of Fig. 1*D*, respectively, in logarithmic scale. Each pixel in the ratio maps shown in Fig. 1*E* and *F* is the ratio between the corresponding pixels, E_n and S_n, whichever is larger, from the two compared Hi-C maps, respectively, *i.e.* $e^{|\log(E_n/S_n)|}$. Therefore, all pixels in the ratio maps are always greater or equal to 1. Overall, Fig. 1*E* is lightly shaded, indicating relatively small discrepancies overall between the simulation and experiment of Fig. 1*A* and *B*. By contrast, Fig. 1*F* contains many more darker pixels, indicating relatively large differences between the experimental Hi-C maps of different genomic regions.

Overall, we have simulated Hi-C maps for a total of five 1.2-Mb-sized genomic regions (only the middle 1 Mb regions are displayed to avoid any possible distortions from the simulation boundaries), one from each chromosome arm that exceeds 1.2 Mb in length. (The left arm of chromosome 3 was omitted, since it is 1.1 Mb long.) These five regions together span more than one-third of the entire fission yeast genome. In every case, we achieve good agreement between the experimental and simulated Hi-C maps (Fig. 1 and Supplementary Figs. 1–4).

An additional, commonly-applied way to compare Hi-C experiments and simulations is to examine the mean contact probability, $P(s)$, for loci with genomic separation, s. Fig. 2 displays five $P(s)$-versus-s

curves, each corresponding to one of the five regions simulated. Experimental and simulated mean contact probabilities are shown as the open circles and solid lines, respectively. Evidently, the experimental mean contact probabilities are very similar for different genomic regions, except for large separation contacts in the 1.2-2.4 Mb region of Chr 3, whose probability exceeds that of the other regions by about 15%. For every region, the simulated mean contact probability agrees well with its experimental counterpart, throughout the range of genomic separations studied (20 kb-500 kb), bolstering *a posteriori* the simple, analytic approach to polymer self contacts, described in Methods.

To achieve the good agreement between simulation and experiment, evident in Fig. 1*A*, Fig. 1*B* and Fig. 2, we chose to quantify and minimize discrepancies between experimental and simulated Hi-C maps, similarly to Ref. [27, 47], using an objective function, that can be interpreted as the *mean pairwise ratio* (MPR) of experimental and simulated pixels, whichever is larger, averaged over all pixels:

$$\mathrm{MPR} = \frac{1}{N}\sum_{n=1}^{N} e^{|\log(\frac{E_n}{S_n})|}, \tag{5}$$

where E_n and S_n are the experimental and simulated contact probabilities, respectively, for pixel n, and N is the total number of pixels in the map of Fig. 1*A*. Note that the MPR value of the two compared maps is exactly the mean of all pixel values of their ratio map. A value of MPR closer to unity indicates better agreement. The MPR between the simulated and experimental Hi-C maps of Fig. 1*A* is 1.0911, while for the experimental Hi-C maps from two different regions (Fig. 1*D*), it is 1.1965, showing a deviation from unity more than twice as large (Table. 1 and 2).

In addition to the MPR, we also calculated the Pearson correlation coefficient (PCC) for pairs of Hi-C maps (Table. 1 and 2). Even though the experimental Hi-C maps shown in Fig. 1*D* correspond to non-overlapping genomic regions, their PCC is 0.9288, surprisingly close to unity. Nevertheless, the PCC for the simulated and experimental Hi-C maps of Fig. 1*A*, is 0.9801, more than three times closer to 1 than the former case. For all five regions studied, the MPR scores and PCCs correspond to distinctly smaller differences overall and significantly greater correlation, respectively, between experimental and simulated Hi-C maps than between different experimental Hi-C maps (Table. 1 and 2), demonstrating that the CCLE model is able to predict and reproduce TAD patterns in *S. pombe*, based solely on cohesin ChIP-seq data.

To obtain the best agreement between simulation and experiment, we optimized the model parameters by minimizing EQ. 5, as described in more detail in Supplementary Methods. The resultant best-fit model parameters are displayed in Table. 3. For all regions studied, the optimized values of the LEF density and the LEF processivity are similar and have small errors, suggesting that the best-fit values of these parameters are robust and that cohesin-driven loop extrusion processes are essentially uniform across the *S. pombe* genome, at least at a resolution of 10 kb. The best-fit LEF density of 0.033 kb^{-1}

Table 1: Mean pairwise ratio (MPR) score and Pearson correlation coefficient (PCC) between model-generated and experimental Hi-C after model parameter optimization

Region	MPR	Pearson
S. pombe Chr1: 0.6–1.6 Mb	1.0984	0.9780
S. pombe Chr1: 4.3–5.3 Mb	1.1102	0.9738
S. pombe Chr2: 0.3–1.3 Mb	1.0911	0.9801
S. pombe Chr2: 1.9–2.9 Mb	1.0877	0.9812
S. pombe Chr3: 1.3–2.3 Mb	1.0902	0.9810
S. cerevisiae Chr13: 290–790 kb	1.2499	0.8925
S. pombe Chr2: 0.3–1.3 Mb (condensin)	1.1793	0.9351

Table 2: MPR score (and PCC) between experimental Hi-C maps from different regions in *S. pombe*

	Chr1: 0.6–1.6 Mb	Chr1: 4.3–5.3 Mb	Chr2: 0.3–1.3 Mb	Chr2: 1.9–2.9 Mb	Chr1: 1.3–2.3 Mb
Chr1: 0.6–1.6 Mb	1 (1)	-	-	-	-
Chr1: 4.3–5.3 Mb	1.1753 (0.9402)	1 (1)	-	-	-
Chr2: 0.3–1.3 Mb	1.1830 (0.9322)	1.1850 (0.9327)	1 (1)	-	-
Chr2: 1.9–2.9 Mb	1.1934 (0.9305)	1.1868 (0.9332)	1.1965 (0.9288)	1 (1)	-
Chr1: 1.3–2.3 Mb	1.2002 (0.9225)	1.1511 (0.9544)	1.1901 (0.9290)	1.1766 (0.9411)	1 (1)

Pearson correlation coefficients (PCCs) are listed in parentheses. Note that optimization based on MPR score is done on the 1 Mb (500 kb)-cropped regions from the regions in *S. pombe* (*S. cerevisiae*) given in Table. 3. See Supplementary Methods for details of the optimization process.

corresponds to a cellular copy number of about 400 loop-extruding cohesin complexes per cell. If we also take the best-fit "cohesive cohesin density" parameter at face value, then there are approximately an additional 400 cohesive cohesins per cell, that is, 800 cohesins in total, which may be compared to the copy numbers from Ref. [48] of the constituents of the core cohesin complex, namely Psc3, Psm1, Psm3, and Rad21, of 723, 664, 1280, 173, respectively (mean 710). The best-fit values of the chromatin persistence length are also similar for different genomic regions, reflecting the similar behavior of the mean contact probability, $P(s)$, versus genomic separation, s, for different genomic regions (Fig. 2). The best-fit values of the chromatin persistence length, which lie in the 75-95 nm range, may be compared to the value of 70 nm, reported in Ref. [49]. The relatively larger errors in the best-fit values of persistence length reflect the fact that our polymer model is only sensitive to the persistence length for large genomic separations (Eq. S37 and Supplementary Fig. 11). The best-fit values of the cohesive cohesin density show a greater variation for different genomic regions than the other parameters, hinting that the level of sister chromatid cohesion may vary across the genome. Alternatively, however, the apparent variation in this parameter could reflect different Psc3 ChIP-seq background levels or shifts for different regions (Methods).

Table 3: Optimized parameters of CCLE simulations for different genomic regions in interphase *S. pombe* and meiotic *S. cerevisiae*

S. pombe	LEF density, ρ (1/kb)	Mean processivity, L (kb)	Persistence length (nm)	Cohesive cohesin density, ρ_c (1/kb)
Chr1: 0.5–1.7 Mb	0.033 (± 0.001)	30.0 (± 0.7)	85 (± 19)	0.037 (± 0.003)
Chr1: 4.2–5.4 Mb	0.033 (± 0.001)	31.2 (± 0.8)	75 (± 23)	0.014 (± 0.003)
Chr2: 0.2–1.4 Mb	0.033 (± 0.001)	30.0 (± 0.8)	80 (± 21)	0.050 (± 0.004)
Chr1: 1.8–3.0 Mb	0.033 (± 0.001)	30.0 (± 0.6)	80 (± 13)	0.028 (± 0.005)
Chr1: 1.2–2.4 Mb	0.030 (± 0.001)	33.3 (± 1.0)	95 (± 17)	0.043 (± 0.004)

S. cerevisiae	LEF density, ρ (1/kb)	Mean processivity, L (kb)	Persistence length (nm)	Cohesive cohesin density, ρ_c (1/kb)	Gaussian σ (kb)
Chr13: 240–840 kb	0.058 (± 0.003)	38.4 (± 3.3)	160 (± 19)	0.018 (± 0.009)	90 (± 25)

LEF density is given in terms of number of LEFs per kilobase-pair; processivity is defined as the averaged LEF processivity in the corresponding region, in the absence of obstacles; cohesive cohesin content is also given in terms of number of cohesive cohesins per kilobase-pair. Parameters are optimized to minimize the mean pairwise ratio (MPR). See Supplementary Methods for details of statistical error calculation and the optimization process.

In addition to cohesin, condensin, which is also a member of the SMC complex family, has been found to extrude DNA loops *in vitro* [50]. However, when we carry out CCLE simulations using the ChIP-seq signal of interphase *S. pombe* condensin, the resultant best-fit simulated Hi-C map shows poor agreement with experiment (Supplementary Fig. 10 and Table. 1), reinforcing that cohesin alone predominantly determines TAD-scale chromatin organization in interphase *S. pombe*.

CCLE self-consistently reproduces cohesin ChIP-seq data

In addition to predicting chromatin organization as measured by Hi-C maps and mean contact probability curves, loop extrusion simulations simultaneously yield position-dependent LEF occupancy probabilities. For a typical 200 kb-sized region of *S. pombe*'s Chr 2, Fig. 3*A* compares the simulated time- and population-averaged probability that a chromatin lattice site is occupied by a LEF (red curve) to the corresponding experimental ChIP-seq data for Psc3 (blue curve), converted to occupancy probability, as described in Methods. Evidently, the simulated LEF occupancy probability matches the experimental Psc3 occupancy probability well. Indeed, as shown in Fig. 3*B*, the cross-correlation of experimental and simulated occupancy probability is nearly 0.7. In both cases, a number of peaks are apparent, extending above background to about twice, or less than twice, background. In the context of CCLE, this relatively weak contrast in occupancy probability gives rise to a corresponding relative lack of contrast in *S.*

pombe's interphase Hi-C pattern.

While unsurprising, given that we set the position-dependent LEF velocities using the Psc3 ChIP-seq data, the good agreement between experimental and simulated occupancy probabilities implies that our simulations are self-consistent, and that the assumptions, leading to EQ. 3 and 4, are valid for the parameters that are determined to best-describe the Hi-C contact map. The extent to which our best-fit simulations satisfy the assumptions, underlying CCLE, can be further assessed by examining the simulated distributions of the following dimensionless quantities, all of which should be small when CCLE is applicable:

$$\frac{R_n - L_n}{R_n + L_n}, \tag{6}$$

which specifies the fractional imbalance in the numbers of left- and right-moving LEF anchors at each lattice site;

$$\frac{A_n - D_n}{\frac{1}{2}(V_{n-1}R_{n-1}(1-P_n) + V_n R_n(1-P_{n+1}))}, \tag{7}$$

which is the ratio of the net current of right moving LEF anchors at each site, that violates current conservation, as a result of binding or unbinding, to the mean current of right-moving anchors, that satisfies current conservation; and

$$\frac{a_n - d_n}{\frac{1}{2}(U_{n+1}L_{n+1}(1-P_n) + U_n L_n(1-P_{n-1}))}, \tag{8}$$

which is the corresponding conserved-current violation ratio for left-moving LEF anchors. As shown in Supplementary Fig. 5, all three of these quantities are distributed around zero with only small excursions (standard deviations ~ 0.1), consistent with the assumptions, leading to EQ. 3 and 4.

Loop configurations in interphase *S. pombe*

The good agreement between our simulated Hi-C maps and experimental Hi-C maps suggests that the corresponding simulated loop configurations are realistic of loop configurations in live *S. pombe*. Fig. 4*A* shows three representative simulated loop configurations for a 1.2 Mb region of Chr 2, corresponding to the best-fit parameters. In this figure, as in Ref. [51], each loop is represented as a semicircle connecting the genomic locations of the two LEF anchor points. Because the model does not permit LEF anchors to pass each other, correspondingly semicircles never cross, although they frequently contact each other and nest, as is apparent for the configurations in Fig. 4*A*. The distributions of loop sizes for all five regions simulated in *S. pombe* are presented in Fig. 4*B*. Evidently, the loop size distributions are similar for all five regions with an overall mean and standard deviation of 22.1 kb and 19.5 kb, respectively. The mean loop size may be compared to the number of base pairs within the chromatin persistence length, estimated to comprise 3.5 kb by taking the chromatin linear density to be 50 bp/nm [49]. Thus, typical

loops contain several (~ 6) persistence lengths of the chromatin polymer. Fig. 4*C* shows the distributions of chromatin backbone segment lengths, *i.e.* the distributions of lengths of chromatin segments, that lie outside of loops. Again, these distributions are similar for all regions simulated with an overall mean and standard deviation of 28.9 kb and 26.8 kb, respectively, again corresponding to several (~ 8) persistence lengths between loops. Since a LEF's anchors bring the chromosomal loci, bound by the LEF, into spatial proximity, loops lead to a significant linear compaction of the chromatin polymer. Fig. 4*D* shows the distributions of chromatin compaction, defined as the fraction of the chromatin contour length within the backbone. These distributions too are similar for all five regions simulated, with overall mean and standard deviation of 0.4161 and 0.0663, implying that the chromatin polymer's contour length in fission yeast is effectively 2.5-times shorter with loops than without.

CCLE Describes TADs and Loop Configurations in Meiotic *S. cerevisiae*

CCLE simulations reproduce experimental meiotic budding yeast Hi-C maps

To further examine the ability of CCLE to describe TAD-scale chromatin organization, we next sought to describe the Hi-C map of meiotic *S. cerevisiae* from Ref. [47], using the ChIP-seq data of the meiotic cohesin core sub-unit, Rec8, from Ref. [52]. In contrast to the semi-dilute polymer solution envisioned to describe chromatin in interphase, in meiosis, the chromatin polymer is significantly compacted and is conspicuously organized about the chromosomal axis. Therefore, meiotic chromatin represents a very different polymer state than interphase chromatin, in which to test CCLE.

The right-hand side of Fig. 5*A* depicts a 500 kb portion of the experimental Hi-C map of meiotic *S. cerevisiae*'s Chr 13 from 290 kb to 790 kb, using a logarithmic false-color intensity scale. In comparison, the left-hand side of Fig. 5*A* presents the corresponding best-fit simulated Hi-C map. The resolution of both maps shown is 2 kb. The original resolution of the simulated Hi-C map is 500 bp and is later binned to 2 kb to match the experimental resolution. Fig. 5*B* magnifies the experiment-simulation comparison for three representative 80 kb sub-regions. Both Fig. 5*A* and *B* reveal similar patterns of high-probability contacts for the experimental and simulated Hi-C maps, manifested in a high-degree of left-right symmetry. However, in contrast to the patch-like TAD patterns, consisting of overlapping squares of more-or-less evenly-distributed enhanced contact probability, that feature in the experimental and simulated Hi-C maps of interphase *S. pombe*, the TADs of meiotic *S. cerevisiae* are dominated by discrete lines of high contact probability and their intersection points. Nevertheless, Fig. 5*A* and *B* make it clear that in spite of their strikingly different appearance to the TAD patterns of interphase *S. pombe*, the grid-like patterns of TADs in meiotic *S. cerevisiae* are also well reproduced by CCLE model. Indeed, meiotic *S. cerevisiae*'s CCLE-simulated Hi-C map shows a compelling resemblance to experiment. In addition, the three quantities, given by EQ. 6, EQ. 7 and EQ. 8, are distributed around zero with relatively small fluctuations (Supplementary Fig. 6), indicating that CCLE model is self-consistent in

this case also.

The very different appearance of the Hi-C maps of interphase *S. pombe* and meiotic *S. cerevisiae* directly follow from the much greater contrast of meiotic *S. cerevisiae*'s cohesin ChIP-seq profile, which consists of a dense pattern of strong, narrow peaks, which extend above the background to reach 4 or 5 times background (Fig. 5*D*). Also different is the best-fit value of the LEF density, which is about twice as large in meiotic *S. cerevisiae* as in interphase *S. pombe* (Table. 3). However, the best-fit value for the LEF processivity of meiotic *S. cerevisiae* is similar to that of interphase *S. pombe*, suggesting that the underlying mechanism that governs loop extrusion by cohesin is identical in both species. Values for the LEF density (0.058 kb^{-1}) and processivity (38.4 kb) may be compared to the best-fit values given in Ref. [47] of 0.031 kb^{-1} and 76 kb, respectively, in the case that their "barrier parameter" is 0.95, or 0.04 kb^{-1} and 64 kb, respectively, for a "barrier parameter" of 0.90. The best-fit value of the "cohesive cohesin density" is noticeably smaller for meiotic *S. cerevisiae* than for interphase *S. pombe*. It is also significantly smaller than the best-fit density of loop-extruding cohesins, suggesting that the preponderance of cellular cohesins are involved in loop extrusion in meiotic *S. cerevisiae*.

In spite of the agreement between simulation and experiment, evident in Fig. 5*A* and *B*, the experiment-simulation comparison for meiotic *S. cerevisiae* shows a higher MPR and smaller PCC than for interphase *S. pombe* (Table. 1), indicating a poorer fit. However, we ascribe this poorer fit, at least in part, to the larger experimental errors of the former data set. These larger experimental errors are apparent when we calculate the MPR and the PCC for the two duplicate Hi-C data sets available in Ref. [47], which take values of 1.3353 and 0.8986, respectively, indicating larger discrepancies between the two nominally identical meiotic *S. cerevisiae* data sets than between experiments and simulations for interphase *S. pombe*.

CCLE self-consistently reproduces Rec8 ChIP-seq data in *S. cerevisiae*

For a typical 200-kb-sized region of Chr 13 of *S. cerevisiae*, Fig. 5*D* compares the simulated time- and population-averaged probability that a chromatin lattice site is occupied by a LEF (red curve) to the corresponding experimental ChIP-seq data for Rec8 (blue curve), converted to occupancy probability, as described in Methods. Clearly, the simulated LEF occupancy probability matches the experimental Rec8 occupancy probability well, with a cross correlation value of almost 0.8 (Fig. 5*E*).

Loop configurations in meiotic *S. cerevisiae*

Fig. 5*G* shows three representative simulated loop configurations for the 240 – 840 kb region of Chr 13, corresponding to the best-fit parameters. In comparison to the loop configurations in interphase *S. pombe*, the loops in meiotic *S. cerevisiae* appear more regularly spaced, corresponding to the more regularly-distributed peaks of meiotic *S. cerevisiae*'s cohesin ChIP-seq data. Fig. S8*A* and *B* present

the distributions of loop sizes and backbone segment lengths, respectively, for the same region in *S. cerevisiae*. The mean and standard deviation of these quantities are: 19.82 and 16.06 kb (mean), and 15.97 and 14.67 kb (SD), respectively. Fig. S8 *C* shows the distributions of chromatin relative compaction, whose mean and standard deviation are 0.2129 and 0.0702, *i.e.* the chromatin polymer's contour length in meiotic *S. cerevisiae* is effectively 5-times shorter with loops than without, twice as compact as the interphase *S. pombe*'s chromatin.

Discussion

By examining an approximate steady-state solution of master equations describing the motions of chromatin loop extruding complexes, we have been led to a new version of the loop extrusion factor model – the conserved-current loop extrusion (CCLE) model – that uses cohesin ChIP-seq data as the sole input. To demonstrate its utility, we applied the CCLE model to accurately reproduce the TAD-scale ($\sim 10 - 100$ kb) Hi-C maps for both of interphase *S. pombe*, for the first time, and of meiotic *S. cerevisiae*, in effect recapitulating the results of Ref. [47] but using a model with an improved physical basis. Overall, our results are compelling evidence that chromatin organization in interphase *S. pombe* and meiotic *S. cerevisiae* on the scale of TADs is principally the result of loop extrusion, and that the cohesin complex is the dominant loop extrusion factor for TAD-scale loops in both systems, marking the base of all or the overwhelming majority of TAD-scale loops. Importantly, the fact that the CCLE model can convert cohesin ChIP-seq data into an accurate Hi-C map, highlights that essential aspects of the three-dimensional chromatin configuration are encoded in the one-dimensional cohesin distribution.

In our current CCLE implementation, cohesin binds to chromatin at random locations. However, the fission yeast protein, Mis4, has been previously identified as a cohesin loading factor [53, 54], prompting us to envision a modification of the model to incorporate position-dependent cohesin binding with a binding rate proportional to the Mis4 ChIP-seq signal. However, for *S. pombe*, the effect of such a modification must necessarily be small, because there is no correlation between the Mis4 and Psc3 ChIP-seq signals (Supplementary Fig. 7). Therefore, the overall distribution of cohesin along the genome (given by the Psc3 ChIP-seq) is independent of where the cohesin was putatively loaded (specified by the Mis4 ChIP-seq). This observation suggests that, in *S. pombe* at least, the collective spreading of cohesins following association to chromatin is sufficiently large, so as to obscure their initial positions.

The fact that the cohesin distribution along the chromatin is strongly linked to chromatin looping, as evident by the success of the CCLE model, allows for new insights into *in vivo* LEF composition and function. For example, recently, two single-molecule studies [37, 38] independently found that Nipbl, which is the mammalian analogue of Mis4, is an obligate component of the loop-extruding human cohesin complex. Ref. [37] also found that cohesin complexes containing Pds5, instead of Nipbl, are unable to

extrude loops. On this basis, Ref. [32] proposed that, while Nipbl-containing cohesin is responsible for loop extrusion, Pds5-containing cohesin is responsible for sister chromatid cohesion, neatly separating cohesin's two functions according to composition. However, the success of CCLE in interphase *S. pombe*, together with the observation that the Mis4 ChIP-seq signal is uncorrelated with the Psc3 ChIP-seq signal (Supplementary Fig. 7) allows us to infer that Mis4 cannot be a component of loop-extruding cohesin in *S. pombe*. On the other hand, Pds5 is correlated with Psc3 in *S. pombe* (Supplementary Fig. 7) suggesting that both proteins are involved in loop-extruding cohesin, contradicting a hypothesis that Pds5 is a marker for cohesive cohesin in *S. pombe*.

In contrast to the absence of Mis4-Psc3 correlation in *S. pombe*, in humans, Nipbl ChIP-seq and Smc1 ChIP-seq are correlated (Supplementary Fig. 7), consistent with Ref. [32]'s hypothesis that Nipbl can be involved in loop-extruding cohesin in humans. However, Ref. [55] showed that human Hi-C contact maps in the absence of Nipbl's binding partner, Mau2 (Ssl3 in *S. pombe* [56]) show clear TADs, consistent with loop extrusion, albeit with reduced long-range contacts in comparison to wild-type maps, indicating that significant loop extrusion continues in live human cells in the absence of Nipbl-Mau2 complexes. These collected observations suggest the existence of two populations of loop-extruding cohesin complexes *in vivo*, one that involves Nipbl-Mau2 and one that does not. Both types are present in mammals, but only Mis4-Ssl3-independent loop-extruding cohesin is present in *S. pombe*.

Beyond yeast, because the CCLE model is agnostic to the identity of any particular boundary element, in principle it extends the LEF model to organisms across the tree of life, including organisms that do not express CTCF. By contrast, prior LEF models have been limited to vertebrates, which (1) express CTCF and where (2) CTCF is the principal boundary element. In future work, it will be interesting to explore the applicability of CCLE model to other model organisms, that exhibit TADs, including, for example, *Drosophila melanogaster*, *Arabidopsis thaliana*, *Oryza sativa*, *Caenorhabditis elegans*, and *Caulobacter crescentus*, as well as to vertebrates.

Methods

Gillespie-type loop extrusion simulations

For interphase *S. pombe* (meiotic *S. cerevisiae*), we represent 1.2 Mb (600 kb) regions of the chromatin polymer as an array of 1200 discrete lattice sites, each comprising 1000 bp (500 bp), and simulate chromatin loop configurations using a Gillespie-type algorithm applied to the LEFs, populating these lattice sites, similarly to Refs. [27, 29]. Because the simulations are of Gillespie-type, the time separations between successive reactions are exponentially-distributed, while which reaction is realized occurs with a probability that is proportional to the rate of that reaction. LEFs are modeled as objects with two anchors, which initially bind to empty, adjacent chromatin lattice sites with a uniformly-distributed

binding probability. Upon binding, a LEF's anchors translocate stochastically and independently in opposite directions at rates specified by position-dependent LEF velocities, determined from experimental cohesin ChIP-seq data. Only outward steps, that grow loops, are permitted. Translocation of a LEF anchor into a lattice site, that is already occupied by another LEF's anchor, is forbidden. LEFs dissociate at a constant rate, dissipating the corresponding loop. In our simulations, after unbinding, a LEF immediately rebinds to an empty pair of neighboring sites, maintaining a constant number of bound LEFs. Neither simulated contact probabilities nor simulated mean occupancy probabilities depend on the time scale of loop extrusion. Therefore, the results of our simulations for chromatin-chromatin contacts and LEF occupancies depend on the mean LEF processivity, which is the ratio of the mean LEF velocity and the dissociation rate, and which therefore is the appropriate fitting parameter.

ChIP-seq to occupancy probability

The *S. pombe* (*S. cerevisiae*) experimental cohesin ChIP-seq data is averaged over each 1000-bp-sized (500-bp-sized) lattice site to yield the ChIP-seq signal, C_n, for each chromatin lattice site, n. To convert Psc3 ChIP-seq data at lattice site n, C_n, to the occupancy probability, P_n, we write

$$P_n = 2N_{LEF}\frac{C_n}{\sum_{n=1}^{N} C_n} - \rho_c \tag{9}$$

where ρ_c is the cohesive cohesin density (number per chromatin lattice site), N is the total number of lattice sites, and N_{LEF} is the number of LEF simulated in the region represented by N lattice sites. The factor of 2 is because one LEF has two anchors and occupies two lattice sites. Although we call ρ_c "cohesive cohesin density", it is important to note that any experimental backgrounds or shifts in the published ChIP-seq data are also subsumed into this parameter. The excellent experiment-simulation agreement justifies the assumption of a uniform distribution of cohesive cohesins *a posteriori*.

Modeling the chromatin polymer inside the nucleus

To model chromatin-chromatin contacts, we treat chromatin inside the cell nucleus as a Gaussian polymer in spherical confinement. For a Gaussian polymer in the continuous limit, the probability density, $p(r, \theta, \phi, n)$, that a location n along the polymer is positioned at coordinates (r, θ, ϕ) follows the diffusion equation, $\partial p/\partial n = (l^2/6)\nabla^2 p$, where l is the Kuhn length. Spherical confinement within a nuclear radius a can be enforced by finding solutions to the diffusion equation either with reflecting or absorbing boundary conditions at $r = a$. As discussed in Refs. [57–59], reflecting boundary conditions correspond to an attractive surface-polymer interaction given by $\epsilon = -k_B T \log(6/5)$, while absorbing boundary conditions correspond to zero polymer-surface interaction. Comparison of chromatin mean contact probability, $P(s)$, using reflecting and absorbing boundary conditions is presented in Supplementary Fig. 11. To crudely

account for "Rabl configurations", in which chromatin is known to be attached to the nuclear envelope [60, 61], we chose to use the solution for reflecting boundary conditions to calculate the self-contact probability between any two points along the chromatin polymer. Calculation of the corresponding self-contact probability for two locations with genomic separation, n, is presented in detail in Supplementary Methods. To then incorporate loops we replace the actual genomic separation (n) between the two loci of interest by their *effective* genomic separation (n_{eff}), which is also discussed in detail in Supplementary Methods. In brief, n_{eff} is the backbone length between the two locations of interest modified by the contribution of any loops that contain the two locations. Our approximate, analytic approach to polymer self-contacts may be justified *a posteriori* on the basis of the excellent data-simulation agreement apparent in Figs. 1 and 2.

Because of the increased compaction of meiotic chromosomes, compared to interphase chromosomes, we can expect volume exclusion to play a more prominent role in meiotic nuclei than in interphase nuclei. In order to account for volume exclusion in meiotic *S. cerevisiae*, we introduce an additional empirical factor, $e^{-s^2/2\sigma^2}$, which reduces the probability of contacts with large genomic separations. This factor is characterized by an additional fitting parameter, namely σ, which represents a genomic distance scale, beyond which chromatin-chromatin contacts are reduced because of volume exclusion. Inclusion of this factor leads to improved agreement between experimental and simulated $P(s)$ curves at longer length scales ($\sim$64 kb to $\sim$120 kb) (Fig. 5F), but has little effect on simulated Hi-C maps at shorter length scales ($\lesssim$ 64 kb).

Simulated Hi-C contact map generation

To generate simulated Hi-C contact maps, for each genomic region and each set of parameter values, we performed 200 independent loop extrusion simulations, each comprised of 100000 time steps. In our simulations, loops achieve a dynamic steady state well within 15000 time steps, as gauged by convergence of the chromatin backbone length to its steady-state value. To ensure that only steady-state loop configurations are included in our simulated Hi-C maps, we discard the first 15000 time steps. Then, for each independent loop extrusion simulation, we create a time-averaged Hi-C map by averaging together 100 simulated Hi-C contact maps, each corresponding to a loop configuration generated by CCLE model, that are evenly-distributed across the simulation time period after the initial 15000 time steps of non-steady-state period. To generate a simulated Hi-C contact map for a given loop configuration, we calculate the contact probability for each pair of lattice sites in the given genomic region, using their effective genomic distance, n_{eff}, and the functional form of self-contact probability for a polymer inside a sphere (Supplementary Methods). Next, we average together the time-averaged Hi-C maps from all 200 simulations to produce an ensemble-averaged Hi-C map. Because the CCLE simulations are performed using a finer resolution than the resolution of experimental Hi-C maps, our ensemble-averaged Hi-C

contact maps are block-averaged to match the resolution of the corresponding experimental map, and scaled so that the mean contact probability of each simulated Hi-C map along its second diagonal is equal to the mean contact probability along the second diagonal of the corresponding experimental Hi-C map, that is, we require $P_{\rm sim}(20\ kb) = P_{\rm exp}(20\ kb)$ for *S. pombe* and $P_{\rm sim}(4\ kb) = P_{\rm exp}(4\ kb)$ for *S. cerevisiae*. The simulated Hi-C maps so-obtained are compared to experimental maps in Fig. 1, 5 and Supplementary Figs. 1 – 4,

To assess the noise within our simulated Hi-C maps, we also calculated the MPR and the PCC between the average of two sets of 200 independent, time-averaged Hi-C simulations, giving a MPR of 1.0174 and a PCC of 0.9992 between these two replicates. Evidently, these values are close to unity, indicting that our final simulated maps accurately represent the CCLE model with little noise.

Code availability

Codes used for CCLE simulation and data analysis of ChIP-seq and Hi-C maps are freely and publicly available at https://github.com/bigpaul97/CCLE. All studied experimental data are included and cited in the article and/or Supplementary Information.

References

[1] Lieberman-Aiden, E., Berkum, N. L. van, Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., Sandstrom, R., Bernstein, B., Bender, M. A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L. A., Lander, E. S., and Dekker, J. "Comprehensive mapping of long-range interactions reveals folding principles of the human genome". In: *Science* 326.5950 (2009), pp. 289–293.

[2] Berkum, N. L. van, Lieberman-Aiden, E., Williams, L., Imakaev, M., Gnirke, A., Mirny, L. A., Dekker, J., and Lander, E. S. "Hi-C: A Method to Study the Three-dimensional Architecture of Genomes". In: *Journal of Visualized Experiments : JoVE* (2010).

[3] Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S., and Ren, B. "Topological domains in mammalian genomes identified by analysis of chromatin interactions". In: *Nature* 485.7398 (2012), pp. 376–380.

[4] Dixon, J. R., Gorkin, D. U., and Ren, B. "Chromatin domains: the unit of chromosome organization". In: *Molecular Cell* 62.5 (2016), pp. 668–680.

[5] Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. "Three-dimensional folding and functional organization principles of the Drosophila genome". In: *Cell* 148.3 (2012), pp. 458–472.

[6] Mizuguchi, T., Fudenberg, G., Mehta, S., Belton, J.-M., Taneja, N., Folco, H. D., FitzGerald, P., Dekker, J., Mirny, L., Barrowman, J., and Grewal, S. I. S. "Cohesin-dependent globules and heterochromatin shape 3D genome architecture in S. pombe". In: *Nature* 516.7531 (2014), pp. 432–435.

[7] Dekker, J. "Two ways to fold the genome during the cell cycle: insights obtained with chromosome conformation capture". In: *Epigenetics & Chromatin* 7.1 (2014), pp. 1–12.

[8] Dekker, J. and Heard, E. "Structural and functional diversity of topologically associating domains". In: *FEBS Letters* 589.20 (2015), pp. 2877–2884.

[9] Pollard, T. D., Earnshaw, W. C., Lippincott-Schwartz, J., and Johnson, G. *Cell biology E-book*. Elsevier Health Sciences, 2016.

[10] Jerkovic, I. and Cavalli, G. "Understanding 3D genome organization by multidisciplinary methods". In: *Nature Reviews Molecular Cell Biology* 22.8 (2021), pp. 511–528.

[11] Barbieri, M., Chotalia, M., Fraser, J., Lavitas, L.-M., Dostie, J., Pombo, A., and Nicodemi, M. "Complexity of chromatin folding is captured by the strings and binders switch model". In: *Proceedings of the National Academy of Sciences* 109.40 (2012), pp. 16173–16178.

[12] Giorgetti, L., Galupa, R., Nora, E. P., Piolot, T., Lam, F., Dekker, J., Tiana, G., and Heard, E. "Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription". In: *Cell* 157.4 (2014), pp. 950–963.

[13] Barbieri, M., Xie, S. Q., Torlai Triglia, E., Chiariello, A. M., Bianco, S., Santiago, I. de, Branco, M. R., Rueda, D., Nicodemi, M., and Pombo, A. "Active and poised promoter states drive folding of the extended HoxB locus in mouse embryonic stem cells". In: *Nature Structural & Molecular Biology* 24.6 (2017), pp. 515–524.

[14] Brackley, C. A., Brown, J. M., Waithe, D., Babbs, C., Davies, J., Hughes, J. R., Buckle, V. J., and Marenduzzo, D. "Predicting the three-dimensional folding of cis-regulatory regions in mammalian genomes using bioinformatic data and polymer models". In: *Genome Biology* 17.1 (2016), p. 59.

[15] Brackley, C. A., Johnson, J., Kelly, S., Cook, P. R., and Marenduzzo, D. "Simulated binding of transcription factors to active and inactive regions folds human chromosomes into loops, rosettes and topological domains". In: *Nucleic Acids Research* 44.8 (2016), pp. 3503–3512.

[16] Di Pierro, M., Zhang, B., Aiden, E. L., Wolynes, P. G., and Onuchic, J. N. "Transferable model for chromosome architecture". In: *Proceedings of the National Academy of Sciences* 113.43 (2016), pp. 12168–12173.

[17] Chiariello, A. M., Annunziatella, C., Bianco, S., Esposito, A., and Nicodemi, M. "Polymer physics of chromosome large-scale 3D organisation". In: *Scientific Reports* 6.1 (2016), p. 29775.

[18] Di Pierro, M., Cheng, R. R., Lieberman Aiden, E., Wolynes, P. G., and Onuchic, J. N. "De novo prediction of human chromosome structures: Epigenetic marking patterns encode genome architecture". In: *Proceedings of the National Academy of Sciences* 114.46 (2017), pp. 12126–12131.

[19] Brackley, C., Johnson, J., Michieletto, D., Morozov, A., Nicodemi, M., Cook, P., and Marenduzzo, D. "Extrusion without a motor: a new take on the loop extrusion model of genome organization". In: *Nucleus* 9.1 (2018), pp. 95–103.

[20] Bianco, S., Lupiáñez, D. G., Chiariello, A. M., Annunziatella, C., Kraft, K., Schöpflin, R., Wittler, L., Andrey, G., Vingron, M., Pombo, A., Mundlos, S., and Nicodemi, M. "Polymer physics predicts the effects of structural variants on chromatin architecture". In: *Nature Genetics* 50.5 (2018), pp. 662–667.

[21] Buckle, A., Brackley, C. A., Boyle, S., Marenduzzo, D., and Gilbert, N. "Polymer simulations of heteromorphic chromatin predict the 3D folding of complex genomic loci". In: *Molecular Cell* 72.4 (2018), pp. 786–797.

[22] Conte, M., Fiorillo, L., Bianco, S., Chiariello, A. M., Esposito, A., and Nicodemi, M. "Polymer physics indicates chromatin folding variability across single-cells results from state degeneracy in phase separation". In: *Nature Communications* 11.1 (2020), p. 3289.

[23] Gerguri, T., Fu, X., Kakui, Y., Khatri, B. S., Barrington, C., Bates, P. A., and Uhlmann, F. "Comparison of loop extrusion and diffusion capture as mitotic chromosome formation pathways in fission yeast". In: *Nucleic Acids Research* 49.3 (2021), pp. 1294–1312.

[24] Conte, M., Irani, E., Chiariello, A. M., Abraham, A., Bianco, S., Esposito, A., and Nicodemi, M. "Loop-extrusion and polymer phase-separation can co-exist at the single-molecule level to shape chromatin folding". In: *Nature Communications* 13.1 (2022), p. 4070.

[25] Alipour, E. and Marko, J. F. "Self-organization of domain structures by DNA-loop-extruding enzymes". In: *Nucleic Acids Research* 40.22 (2012), pp. 11202–11212.

[26] Sanborn, A. L., Rao, S. S. P., Huang, S.-C., Durand, N. C., Huntley, M. H., Jewett, A. I., Bochkov, I. D., Chinnappan, D., Cutkosky, A., Li, J., Geeting, K. P., Gnirke, A., Melnikov, A., McKenna, D., Stamenova, E. K., Lander, E. S., and Aiden, E. L. "Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes". In: *Proceedings of the National Academy of Sciences* 112.47 (2015), E6456–E6465.

[27] Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., and Mirny, L. A. "Formation of chromosomal domains by loop extrusion". In: *Cell Reports* 15.9 (2016), pp. 2038–2049.

[28] Goloborodko, A., Imakaev, M. V., Marko, J. F., and Mirny, L. "Compaction and segregation of sister chromatids via active loop extrusion". In: *Elife* 5 (2016), e14864.

[29] Goloborodko, A., Marko, J. F., and Mirny, L. A. "Chromosome compaction by active loop extrusion". In: *Biophysical Journal* 110.10 (2016), pp. 2162–2168.

[30] Nuebler, J., Fudenberg, G., Imakaev, M., Abdennur, N., and Mirny, L. A. "Chromatin organization by an interplay of loop extrusion and compartmental segregation". In: *Proceedings of the National Academy of Sciences* 115.29 (2018), E6697–E6706.

[31] Banigan, E. J. and Mirny, L. A. "Loop extrusion: theory meets single-molecule experiments". In: *Current Opinion in Cell Biology* 64 (2020), pp. 124–138.

[32] Davidson, I. F. and Peters, J.-M. "Genome folding through loop extrusion by SMC complexes". In: *Nature Reviews Molecular Cell Biology* 22.7 (2021), pp. 445–464.

[33] Oudelaar, A. M. and Higgs, D. R. "The relationship between genome structure and function". In: *Nature Reviews Genetics* 22.3 (2021), pp. 154–168.

[34] Fudenberg, G., Abdennur, N., Imakaev, M., Goloborodko, A., and Mirny, L. A. "Emerging evidence of chromosome folding by loop extrusion". In: *Cold Spring Harbor symposia on quantitative biology.* Vol. 82. Cold Spring Harbor Laboratory Press. 2017, pp. 45–55.

[35] Rao, S. S., Huang, S.-C., Glenn St Hilaire, B., Engreitz, J. M., Perez, E. M., Kieffer-Kwon, K.-R., Sanborn, A. L., Johnstone, S. E., Bascom, G. D., Bochkov, I. D., Huang, X., Shamim, M. S., Shin, J., Turner, D., Ye, Z., Omer, A. D., Robinson, J. T., Schlick, T., Bernstein, B. E., Casellas, R., Lander, E. S., and Aiden, E. L. "Cohesin loss eliminates all loop domains". In: *Cell* 171.2 (2017), pp. 305–320.

[36] Wutz, G., Várnai, C., Nagasaka, K., Cisneros, D. A., Stocsits, R. R., Tang, W., Schoenfelder, S., Jessberger, G., Muhar, M., Hossain, M. J., Walther, N., Koch, B., Kueblbeck, M., Ellenberg, J., Zuber, J., Fraser, P., and Peters, J.-M. "Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins". In: *The EMBO Journal* 36.24 (2017), pp. 3573–3599.

[37] Davidson, I. F., Bauer, B., Goetz, D., Tang, W., Wutz, G., and Peters, J.-M. "DNA loop extrusion by human cohesin". In: *Science* 366.6471 (2019), pp. 1338–1345.

[38] Kim, Y., Shi, Z., Zhang, H., Finkelstein, I. J., and Yu, H. "Human cohesin compacts DNA by loop extrusion". In: *Science* 366.6471 (2019), pp. 1345–1349.

[39] Golfier, S., Quail, T., Kimura, H., and Brugués, J. "Cohesin and condensin extrude DNA loops in a cell cycle-dependent manner". In: *Elife* 9 (2020), e53885.

[40] Nora, E. P., Goloborodko, A., Valton, A.-L., Gibcus, J. H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L. A., and Bruneau, B. G. "Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization". In: *Cell* 169.5 (2017), pp. 930–944.

[41] Kentepozidou, E., Aitken, S. J., Feig, C., Stefflova, K., Ibarra-Soria, X., Odom, D. T., Roller, M., and Flicek, P. "Clustered CTCF binding is an evolutionary mechanism to maintain topologically associating domains". In: *Genome Biology* 21 (2020), pp. 1–19.

[42] Ong, C.-T. and Corces, V. G. "CTCF: an architectural protein bridging genome topology and function". In: *Nature Reviews Genetics* 15.4 (2014), pp. 234–246.

[43] Wang, C., Liu, C., Roqueiro, D., Grimm, D., Schwab, R., Becker, C., Lanz, C., and Weigel, D. "Genome-wide analysis of local chromatin packing in Arabidopsis thaliana". In: *Genome Research* 25.2 (2015), pp. 246–256.

[44] Dong, Q., Li, N., Li, X., Yuan, Z., Xie, D., Wang, X., Li, J., Yu, Y., Wang, J., Ding, B., Zhang, Z., Li, C., Bian, Y., Zhang, A., Wu, Y., Liu, B., and Gong, L. "Genome-wide Hi-C analysis reveals extensive hierarchical chromatin interactions in rice". In: *The Plant Journal* 94.6 (2018), pp. 1141–1156.

[45] Heger, P., Marin, B., and Schierenberg, E. "Loss of the insulator protein CTCF during nematode evolution". In: *BMC Molecular Biology* 10 (2009), pp. 1–14.

[46] Kaushal, A., Mohana, G., Dorier, J., Özdemir, I., Omer, A., Cousin, P., Semenova, A., Taschner, M., Dergai, O., Marzetta, F., Iseli, C., Eliaz, Y., Weisz, D., Shamim, M. S., Guex, N., Lieberman Aiden, E., and Gambetta, M. C. "CTCF loss has limited effects on global genome architecture in Drosophila despite critical regulatory functions". In: *Nature Communications* 12.1 (2021), p. 1011.

[47] Schalbetter, S. A., Fudenberg, G., Baxter, J., Pollard, K. S., and Neale, M. J. "Principles of meiotic chromosome assembly revealed in S. cerevisiae". In: *Nature Communications* 10.1 (2019), p. 4795.

[48] Marguerat, S., Schmidt, A., Codlin, S., Chen, W., Aebersold, R., and Bähler, J. "Quantitative analysis of fission yeast transcriptomes and proteomes in proliferating and quiescent cells". In: *Cell* 151.3 (2012), pp. 671–683.

[49] Arbona, J.-M., Herbert, S., Fabre, E., and Zimmer, C. "Inferring the physical properties of yeast chromatin through Bayesian analysis of whole nucleus simulations". In: *Genome Biology* 18 (2017), pp. 1–15.

[50] Ganji, M., Shaltiel, I. A., Bisht, S., Kim, E., Kalichava, A., Haering, C. H., and Dekker, C. "Real-time imaging of DNA loop extrusion by condensin". In: *Science* 360.6384 (2018), pp. 102–105.

[51] Bundschuh, R. and Hwa, T. "Statistical mechanics of secondary structures formed by random RNA sequences". In: *Physical Review E* 65.3 (2002), p. 031903.

[52] Ito, M., Kugou, K., Fawcett, J. A., Mura, S., Ikeda, S., Innan, H., and Ohta, K. "Meiotic recombination cold spots in chromosomal cohesion sites". In: *Genes to Cells* 19.5 (2014), pp. 359–373.

[53] Ciosk, R., Shirayama, M., Shevchenko, A., Tanaka, T., Toth, A., Shevchenko, A., and Nasmyth, K. "Cohesin's binding to chromosomes depends on a separate complex consisting of Scc2 and Scc4 proteins". In: *Molecular Cell* 5.2 (2000), pp. 243–254.

[54] Takahashi, T. S., Yiu, P., Chou, M. F., Gygi, S., and Walter, J. C. "Recruitment of Xenopus Scc2 and cohesin to chromatin requires the pre-replication complex". In: *Nature Cell Biology* 6.10 (2004), pp. 991–996.

[55] Haarhuis, J. H., van der Weide, R. H., Blomen, V. A., Yáñez-Cuna, J. O., Amendola, M., van Ruiten, M. S., Krijger, P. H., Teunissen, H., Medema, R. H., van Steensel, B., Brummelkamp, T. R., de Wit, E., and Rowland, B. D. "The cohesin release factor WAPL restricts chromatin loop extension". In: *Cell* 169.4 (2017), pp. 693–707.

[56] Makrantoni, V. and Marston, A. L. "Cohesin and chromosome segregation". In: *Current Biology* 28.12 (2018), R688–R693.

[57] DiMarzio, E. A. "Proper accounting of conformations of a polymer near a surface". In: *The Journal of Chemical Physics* 42.6 (1965), pp. 2101–2106.

[58] Casassa, E. F. "Equilibrium distribution of flexible polymer chains between a macroscopic solution phase and small voids". In: *Journal of Polymer Science Part B: Polymer Letters* 5.9 (1967), pp. 773–778.

[59] Allegra, G. and Colombo, E. "Polymer chain between attractive walls: A second-order transition". In: *The Journal of Chemical Physics* 98.9 (1993), pp. 7398–7404.

[60] Mizuguchi, T., Barrowman, J., and Grewal, S. I. "Chromosome domain architecture and dynamic organization of the fission yeast genome". In: *FEBS Letters* 589.20 (2015), pp. 2975–2986.

[61] Schreiner, S. M., Koo, P. K., Zhao, Y., Mochrie, S. G., and King, M. C. "The tethering of chromatin to the nuclear envelope supports nuclear mechanics". In: *Nature Communications* 6.1 (2015), p. 7159.

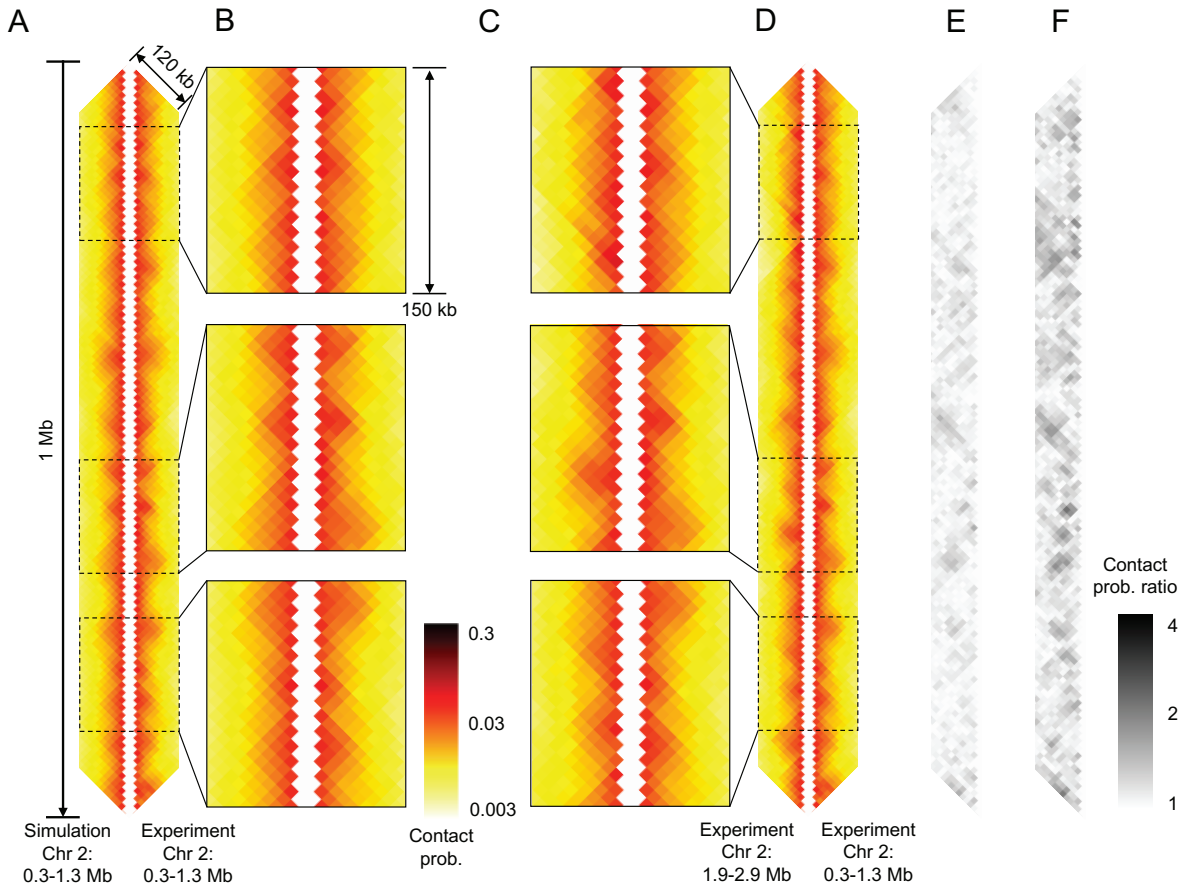

Figure 1: Conserved-current loop extrusion (CCLE) model recapitulates TAD-scale chromatin organization in interphase *S. pombe* using cohesin ChIP-seq data. (A) Comparison between Hi-C map of 1 Mb region generated by the CCLE model (using cohesin ChIP-seq data from Ref. [6]) and the experimental Hi-C map [6] of the same region. (B) Magnified comparisons between experiment and simulation in (A) for three representative regions. (C) Magnified comparisons between experiment and simulation in (D) for three representative regions. (D) Comparison between two experimental Hi-C maps [6] of two different regions. All Hi-C maps show interactions up to genomic separations of 120 kb. (E) Contact probability ratio maps between simulated and experimental Hi-C maps shown in (A). Each pixel represents the ratio of contact probabilities of the corresponding pixels of simulation and experiment in (A). (F) Contact probability ratio maps between two experimental Hi-C maps shown in (D). Pixels with darker shades indicate a poorer agreement than pixels with lighter shades. All maps are displayed in log-scale. Simulated Hi-C map is generated by CCLE model with best-fit parameters given in Table. 3 The optimization process is discussed in Supplementary Methods.

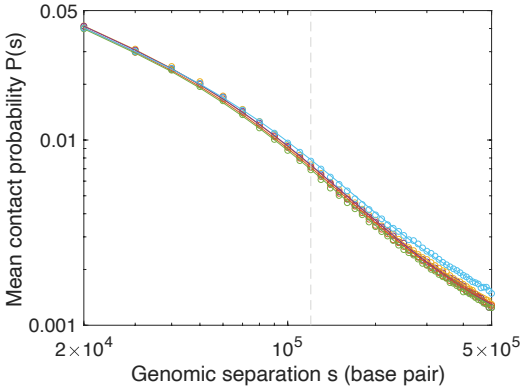

Figure 2: CCLE reproduces $P(s)$ curves of experimental Hi-C contact maps across the entire *S. pombe* genome. Chromatin mean contact probabilities, $P(s)$, as a function of genomic separation, s, are plotted in circles for five 1.2 Mb genomic regions, using experimental Hi-C contact map data from Ref. [6]. The corresponding $P(s)$ curves of CCLE-simulated Hi-C contact maps are plotted as lines. Different colors represent different regions: yellow, 0.5-1.7 Mb region of Chr 1; purple, 4.2-5.4 Mb region of Chr 1; red, 0.2-1.4 Mb region of Chr 2; green, 1.8-3.0 Mb region of Chr 2; cyan, 1.2-2.4 Mb region of Chr 3. The dashed, vertical line indicates the maximum genomic separation (120 kb) presented in the Hi-C comparison maps and ratio maps shown in Fig. 1.

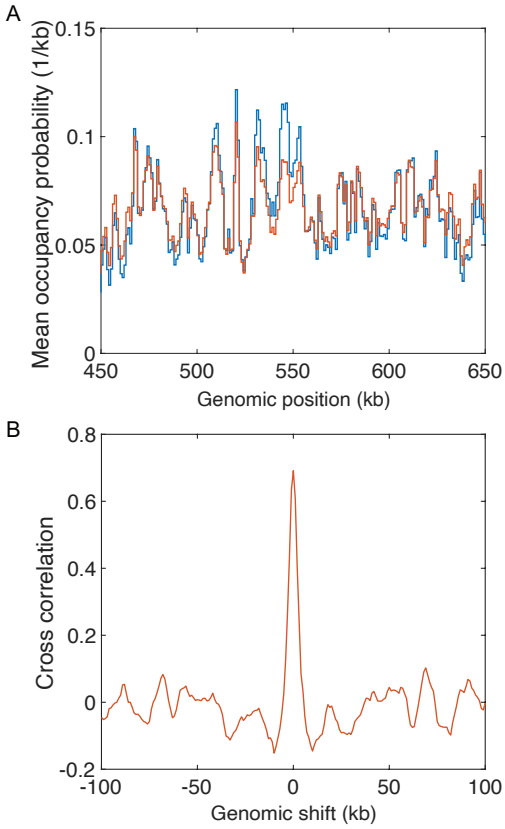

Figure 3: CCLE reproduces experimental cohesin occupancy landscape in *S. pombe*. (A) Comparison between experimental cohesin occupancy probability landscape (blue) and LEF occupancy probability landscape in CCLE simulation using interphase cohesin (Psc3) ChIP-seq data (red). The occupancy probability curves are normalized by the corresponding optimized number of LEFs of 40. The landscape is plotted for 450 to 650 kb region in Chr 2 of *S. pombe*. (B) Cross correlation between experimental cohesin occupancy probability landscape and LEF occupancy probability landscape in CCLE simulation, as a function of genomic shift.

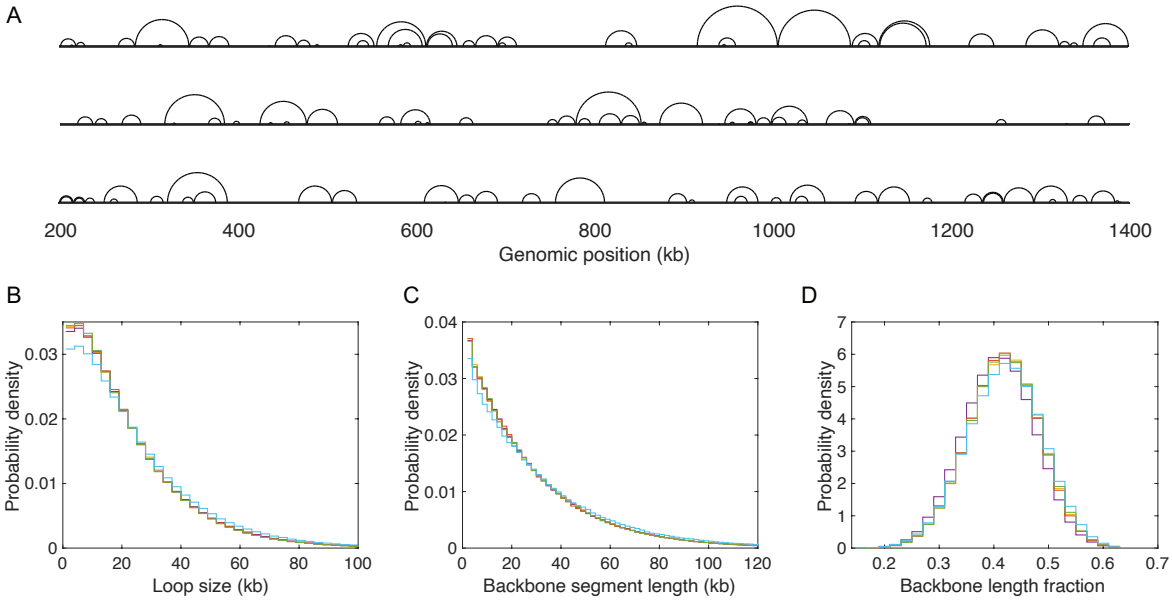

Figure 4: Loop configurations and properties. (A) Snapshots of three representative simulated loop configurations of interphase *S. pombe*'s Chr 2 between 200 kb and 1400 kb. In each case, the chromatin backbone is represented as a straight line, while loops are represented as semicircles connecting loop anchors, following Ref. [51]. Because LEF anchors block each other, loops can nest but they cannot cross. (B), (C), and (D) Distributions of loop size, backbone segment length, and chromatin compaction ratio (as measured by the fraction of the chromatin contour length within the backbone), respectively, for the five genomic regions in interphase *S. pombe*, listed in Table. 3: Yellow, Chr 1: 0.5 – 1.7 Mb; purple, Chr 1: 4.2 – 5.4 Mb; red, Chr 2: 0.2 – 1.4 Mb; green, Chr 2: 1.8 – 3.0 Mb; cyan, Chr 3: 1.2 – 2.4 Mb.

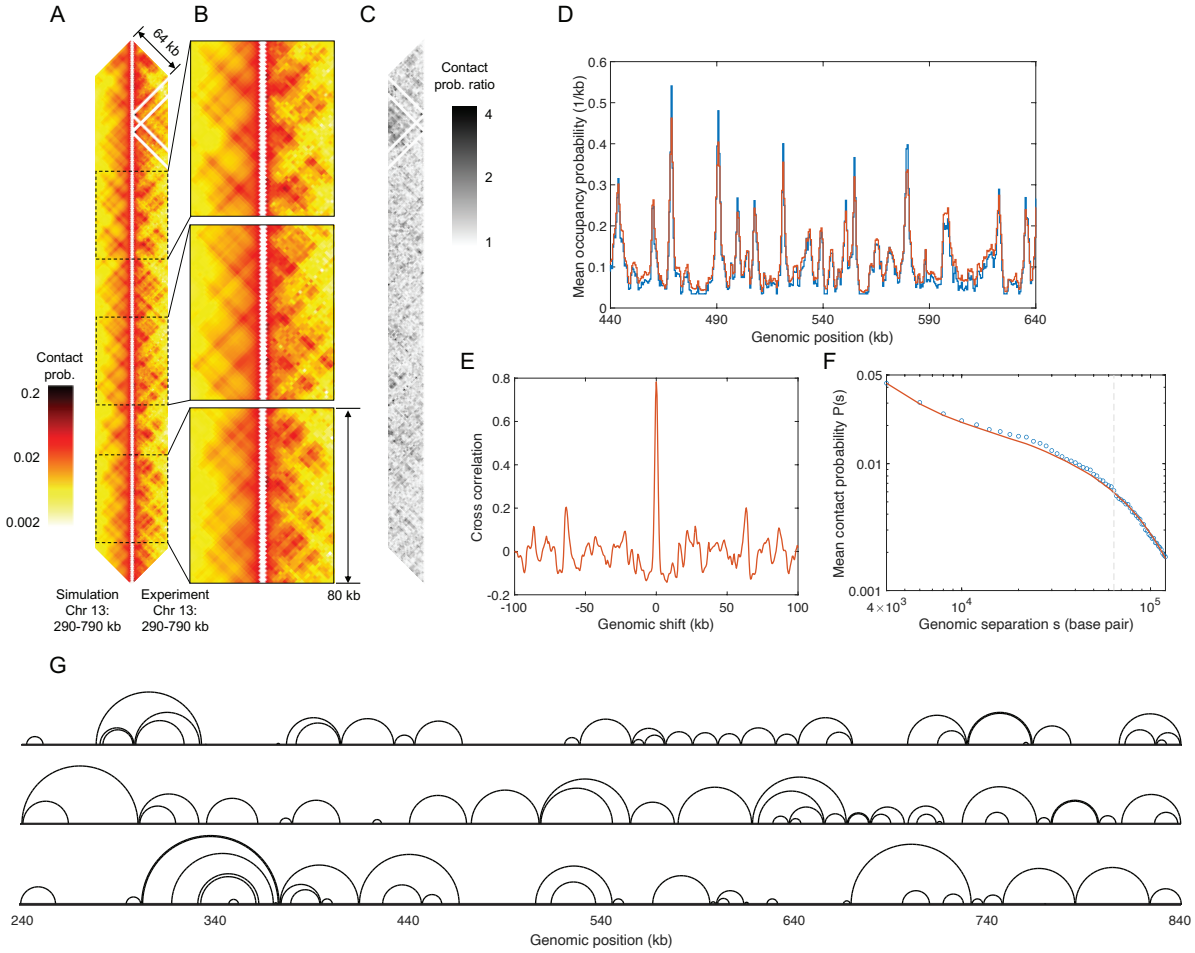

Figure 5: Conserved-current loop extrusion (CCLE) model also recapitulates TAD-scale chromatin organization in meiotic *S. cerevisiae* using cohesin ChIP-seq data. (A) Comparison between Hi-C map of 500 kb region generated by the CCLE model (using meiotic Rec8 ChIP-seq data from Ref. [52]) and the experimental Hi-C map [47] of the same region. Both Hi-C maps show interactions up to genomic separation of 64 kb. (B) Magnified experiment-simulation Hi-C comparison for three representative subregions, each 80 kb in size. (C) Contact probability ratio map between the experimental and simulated Hi-C maps in (A). (D) Normalized experimental meiotic cohesin (Rec8) occupancy probability (blue) and simulated LEF occupancy probability landscape (red). The occupancy probability curves are plotted for 440 to 640-kb region of Chr 13 and are normalized by the corresponding optimized number of LEFs of 13.3 in the shown region. (E) Cross correlation between experimental meiotic cohesin (Rec8) occupancy landscape and simulated LEF occupancy probability landscape. (F) Chromatin mean contact probability, $P(s)$, as a function of genomic distance, s, for experimental Hi-C (blue circles) and simulation (red line) with Gaussian correction factor described in Methods. (G) Snapshots of three representative simulated loop configurations of meiotic *S. cerevisiae*'s Chr 13 between 240 kb and 840 kb. In each case, the chromatin backbone is represented as a straight line, while loops are represented as semicircles connecting loop anchors, following Ref. [51]. Because LEF anchors block each other, loops can nest but they cannot cross, although they frequently come into contact. Simulation results shown in this figure are generated by CCLE using the best-fit parameters given in Table. 3. The optimization process is discussed in Supplementary Methods.