

Supplemental Information

Formation of Chromosomal Domains by Loop Extrusion

Geoffrey Fudenberg, Maxim Imakaev, Carolyn Lu, Anton Goloborodko, Nezar Abdennur, and Leonid A. Mirny

Supplemental Information for Formation of Chromosomal Domains by Loop Extrusion

Supplemental Notes

Experimental evidence for differences between TADs and compartments

TADs differ from larger-scale A/B compartments in they do not necessarily form an alternating ‘checkerboard’ pattern of enriched contact frequencies (Lajoie et al., 2014), and several TADs often reside within a single contiguous compartment (Gibcus and Dekker, 2013; Gorkin et al., 2014). TADs are largely stable across cell-types, while compartments are not (Dekker and Heard, 2015). TADs and compartments display distinct responses to biological perturbations, including cohesin knockdown (Seitan et al., 2013), and distinct relationships with other nuclear landmarks, including the lamina (Kind et al., 2015). Although correlated with replication timing, available evidence indicates that TADs are independent of DNA replication itself, they: appear early in G1, before replication has a chance to influence chromosome organization (Dileep et al., 2015; Naumova et al., 2013); are maintained in cells where replication has been halted by thymidine block (Naumova et al., 2013); and remain present, though are on-average weaker, in senescent cells that do not replicate their DNA (Chandra et al., 2015; Criscione et al., 2016).

Insulation in 1D vs 3D

Insulation can be easily represented in 1D when the chromosome is drawn as a straight line with the insulating element between the enhancer and promoter (**Fig. 5A**). This creates an impression that insulation is relatively easy to achieve. However, intuition developed in 1D is misleading when applied to long flexible chromatin fibers in 3D, particularly because the physical size of an insulating element (3-50nm) is much less than the size of a TAD (300-1000nm). For example, an enhancer and promoter separated by ~100kb (i.e. much greater than the persistence length of the chromatin fiber) can come in contact while still being spatially far from the insulating element in 3D. Our simulations of BEs as bulky objects or locally very stiff regions argue that it is very difficult to insulate large genomic regions from each other by small factors that act locally. Indeed, our simulations suggest that a crucial mechanistic insight is often overlooked in illustrations of insulating elements.

In this work, we propose that insulation at the scale of TADs can be achieved by loop extrusion limited by BEs; crucially, LEFs convert the difficult prospect of insulating in 3D into the more intuitive 1D case and allow insulation to be mediated over spatial and genomic distances much larger than the size of bound proteins forming a BE.

Extended Experimental Procedures:

Model Overview

To simulate how the mechanism of loop extrusion can actively compact TADs, we modeled the process by coupling the 1D dynamics of loop extrusion by LEFs with 3D polymer dynamics. We use a discrete model for 1D LEF dynamics, which imposes a system of bonds on the simulated 3D polymer dynamics. This allows us to generate average simulated contact maps for different parameter values in our model, which can in turn be compared with experimental Hi-C contact maps.

LEF dynamics with boundary elements

LEF translocation along a chromatin fiber was simulated on a 1D lattice, where each position was characterized by the following parameters: association (birth) probability, dissociation (death) probability, and BE occupancy (stalling probability).

Since the molecular mechanism of loop extrusion is not known, we model a LEF very generally as having two “heads” connected by a linker, analogously to SMC protein complexes. Each head of a LEF occupied one lattice position at a time, and no two heads could occupy the same lattice position except at birth events. At each time-step, each LEF head translocates to the neighboring lattice site, if that site is not occupied or is not a boundary element (BE). If one head cannot translocate, the other head is

unaffected. At each time-step, a LEF dissociates with a probability equal to the maximum of the death probabilities at the positions of the two heads of a LEF. For birth events, LEF heads were either started on the same lattice site, or on adjacent lattice sites (if the site immediately to the right of a chosen site was unoccupied) with 50% probability; this was necessary to avoid producing a checkerboard pattern on the contact map at the monomer-scale, since the two heads of a LEF that has never been stalled would otherwise always occupy either (even, odd), or (odd, even) positions.

In principle, this formulation of LEF dynamics allows for a very complex implementation where each parameter is variable throughout the genome. However, since it remains unknown how the values might vary across the genome, we first considered a simpler system with uniform birth probability, constant death probability, a fixed number of LEFs, and a discrete number of completely impermeable BEs. In this formulation, LEF dynamics are well described by the LEF **processivity** (2/death rate, since each head of a non-stalled LEF translocates by 1 lattice-site per time-step), and **separation** (total number of lattice sites / number of LEFs). Processivity is additionally interpretable as the average size of a loop extruded by an unobstructed LEF over its lifetime. Small elaborations of this basic 1D model can allow for LEF pausing, permanent stalling at particular genomic elements, active loading and unloading, or even multiple classes of LEFs with different processivity. Note that while 1D dynamics of LEFs are taken as independent from the 3D dynamics of the polymer chain, the interactions induced by LEFs cannot simply be described as an effective pairwise potential. This is because the potential of interactions created by a loop-extruding factor is time-dependent, and LEFs are not independent as they can block each other while on DNA.

3D simulations

To perform Langevin dynamics polymer simulations we used OpenMM, a high-performance GPU-assisted molecular dynamics API (Eastman and Pande, 2010; Eastman et al., 2013). To represent chromatin fibers as polymers, we used a sequence of spherical monomers of 1 unit of length in diameter. Here and below all distances are measured in monomer sizes (~3 nucleosomes, ~10nm), density is measured in particles per cubic unit, and energies are measured in kT.

Neighboring monomers are connected by harmonic bonds, with a potential $U = 100(r - 1)^2$ (here and below in units of kT). Polymer **stiffness** is modeled with a three point interaction term, with the potential $U = S(1 - \cos(\alpha))$, where alpha is the angle between neighboring bonds, and S is a stiffness parameter.

To allow chain passing, which represents activity of topoisomerase II, we used a soft-core potential for interactions between monomers, similar to (Le et al., 2013; Naumova et al., 2013). All monomers interacted via a repulsive potential

$$U = 1.5 \left(-1 + \left(\frac{1.05 \cdot r}{\sqrt{6/7}} \right)^{12} \cdot \left(\left(\frac{1.05 \cdot r}{\sqrt{6/7}} \right)^2 - 1 \right) \cdot \frac{823543}{46656} \right)$$

This is a fast and efficient potential designed to be a constant of 1.5 kT up to $r=0.7-0.8$, and then quickly go to zero at $r=1.05$. Simulations were performed in periodic boundary conditions, with the size of the cubic box set to achieve spatial **density** of 0.05 or 0.2.

To connect 1D LEF simulations with 3D polymer simulations, we first run 1D LEF dynamics for 4 million time steps to achieve thorough equilibration of LEF positions even in the most dense regimes. We then assigned bonds to the monomers in polymer simulations according to the current position of LEFs. The two monomers held by the two heads of each LEF were connected by a harmonic bond with the potential $U = 4(r - 1)^2$. We then performed the number of steps of Langevin dynamics, defined by a parameter setting the number of 3D-simulation time-steps per 1D-simulation time-steps (**3D-to-1D**

dynamics parameter). After that, we advanced LEF dynamics and restarted polymer simulations from the final position of the previous run, but with new position of LEFs (i.e. bonds). Since shifting bond positions in OpenMM is impossible without a computationally expensive re-initialization procedure, for computational efficiency, we advanced LEF dynamics by four steps at a time.

We note that this coarse-grained description of extrusion does not necessarily assume the chromatin fiber is threaded through the LEF, and the processes we describe could be realized by anything that acts processively along the chromosome to generate progressively larger loops.

Minimal 3D polymer model

We performed simulations of a polymer chain consisting of 8 groups of three TADs of 300, 600, and 1200 monomers each, arranged sequentially (300, 600, 1200, 300, 600, 1200...), totaling (300 + 600 + 1200) * 8 monomers. Impermeable BEs were placed between each pair of neighboring TADs. Simulations were started from a compact polymer conformation, as described in (Imakaev et al., 2015), created on a cubic lattice a box of the size (PBC box – 2). Simulations were then advanced for 100 blocks of simulations (400 steps of LEF dynamics). After that, 500 blocks of simulations were performed and their conformations were recorded. After that, LEF dynamics was advanced by 4 million steps, and the process was repeated (100 blocks of simulations, then 500 recorded blocks of simulations) three more times. The latter was done to achieve better averaging. Overall, this yielded 2000 conformations for each parameter set.

To achieve better averaging of contact maps (which are more difficult to average than $P(s)$ curves (Naumova et al., 2013)) for all main and supplemental figures, the above numbers were increased as follows: 200 non-recorded blocks of simulations, 1000 blocks of simulations recorded, repeated 10 times to yield 10000 conformations. This changed the goodness-of-fit by no more than 0.03, showing that our parameter sweep produced sufficient number of conformations to evaluate $P(s)$ curves.

Generating simulated contact maps, $P(s)$, and FISH CDFs

Generating contact frequency maps for comparison to Hi-C data relies on a final parameter, the **Hi-C capture radius**, which is the distance at which two monomers are determined to be in contact. Contact maps were generated using different values of contact radius: 2, 3, 4, 5, 6, 7, 8, 10. To generate a contact map, each conformation was separated into 7 blocks of two TADs (TADs 0,1; 1,2; 2,3... 6,7). Then for each block of two TADs, a contact map was calculated. Resulting 7*2000 contact maps from each of the 2000 conformation were summed together and saved.

To calculate $P(s)$ plots from simulated data, a contact map with 1 monomer resolution was used. The contact map was not normalized, since the final $P(s)$ curves were allowed to be vertically shifted on a log-log plot (i.e. multiplied by a constant) for calculating goodness-of-fit.

For display, simulated contact maps were first binned using 4 monomer (2.4 kb) bins. Then, the map was normalized such that an average sum over all rows equals one. The map was then clipped at 0.0035; values of zero were shown as white, values of 0.0035 or larger were shown as black.

From each contact map, we calculated $P(s)$ as follows. For each diagonal of the contact matrix, we evaluated the average values of contact probability within 300-monomer TADs, 600-monomer TADs, 1200-monomer TADs and regions not belonging to any TAD (trans). We then averaged the values in logarithmically-spaced bins with a step of 1.2 starting at 7 (7,8,10,12,15,...,x,x*1.2, x*1.2^2,..., N) where N is the length of a given $P(s)$ curve (300, 600, 1200 for TADs, 1300 for between-TADs, in monomers).

For each parameter value, simulated FISH CDFs were calculated by aggregating distances for a pair of ‘within-TAD’ loci in the largest (1200 monomer) TAD and a pair of ‘between-TAD’ loci that span

the BE between the 600 and 1200 monomer TAD, both at a separation of 600 monomers (360kb) for the same sets of conformations used to build contact maps.

Experimental P(s) and Hi-C maps

To calculate experimental $P(s)$, we used publicly-available data from (Rao, 2014 (Rao et al., 2014)), processed in-house using ICE (Imakaev et al., 2012), for GM12878 inSitu protocol and MboI restriction enzyme. Data was binned at 2kb. Since a whole-chromosome 2kb resolution contact map is too large to fit in memory, we split the data in blocks of 20Mb, overlapping by 10Mb (starting at 0, 10Mb, 20Mb, 30Mb, etc.). We removed bins with <100 reads, and then iteratively corrected each block. We obtained GEO: GSE63525 GM12878 primary replicate TAD and loop annotations from (Rao et al., 2014). For all annotated TADs, we found the 20MB block most centered at this TAD, and evaluated $P(s)$ within this TAD. To calculate $P(s)$ between TADs, we defined a set of boundaries as the union of start and end-points for all TADs. $P(s)$ was then calculated between loci-pairs upstream and downstream of the boundary up Mb separation. To display Hi-C maps in **Figure 1**, the same dataset was used, binned at 5kb resolution to reduce sampling noise, and corrected in the same way as for $P(s)$ calculations.

Goodness-of-fit

To compare experimental and simulated $P(s)$, we first selected experimental TADs that are comparable in size to each simulated TAD. We selected TADs that are between 0.9 and 1.1 of the size of simulated TAD, assuming each monomer is 600kb (giving 180kb, 360kb and 720kb TADs). We then calculated median $P(s)$ for TADs in each category, as well as 10th and 90th percentiles. $P(s)$ was calculated in logarithmically spaced bins, as described above. The three curves were plotted for comparison with simulated scalings. Next, we evaluated experimental $P(s)$ at the same values of s as simulated $P(s)$ curves were calculated. Note that since these values are logarithmically spaced, each portion of the log-log plots shown contribute equally. We then define the goodness-of-fit as the *geometric standard deviation* of the ratio of simulated to experimental $P(s)$; this has an interpretation as a typical fold-deviation of $P(s)$.

A value of 1 indicates a perfect fit and larger values indicate worse fits. The best fit we observed was 1.103 (processivity 200, separation 200, 5000 steps, stiffness 2, density 0.2, cutoff 10). The second best fit was the same as first but with processivity of 400 (fit 1.110). The worst fit had a value of 4.46 for (processivity 100, separation 50, 1000 steps, stiffness 0, density 0.05, and cutoff 2). The poor fit in this case is likely because the small tight loops formed in this regime formed a very dense fiber that has many fewer long-range contacts than experimental data. Notably, the second worst fit was similar, while the third worst fit, 4.03, had completely different parameter values: (processivity 100, separation 1600, 5000 steps, stiffness 0, density 0.05, and cutoff 2), which lies in the free polymer regime. All parameter values and fit coefficients are summarized in **Data D1**.

Since the best fitting models had diverse parameter values, we took first 100 best fitting models (fit values 1.103 to 1.195) and assessed how frequently each pair of (processivity, separation) occurs in this list, and what the best fit was for each pair (**Fig 2**).

We note that any biological realization of active linear compaction by loop extrusion has many additional unknowns beyond the parameters we use to describe the process; this can include the size of the LEF and the interactions between LEFs and other DNA-bound proteins. The former may affect both the distance between LEF proteins at which they become stalled, as well as the distance between two regions of DNA bound by a LEF. Moreover, experimentally-mapped Hi-C TADs display a great deal of diversity, and would most likely be described by different sets of parameters. For these reasons, we focused on which combinations of parameters frequently gave a good fit, and the characteristics of TADs in this parameter regime, rather than relying on particular best-fitting parameter set. We found that while nearly any value of any parameter can be found in the top parameter-sets, we find that certain pairs of LEF processivity and LEF separation are found more often.

Modifications to the minimal model

To simulate a model of semi-permeable boundary (**fig. S1**), we introduced a genomic element in the middle of the longest TAD that permanently stalls LEF heads with a probability of 20%. We note that the effective stall rate is actually higher, as once one LEF becomes stalled at the boundary, another LEF is likely to stall against the original LEF, etc. Thus, only a small probability to stall LEFs is needed to “nucleate” a TAD boundary, even in a relatively dilute regime. In our case, a probability of 0.2 creates a strong visible TAD boundary.

To create a model with uneven loading of LEFs (**fig. S1**), we increased the birth rate 30-fold in the region of the length of (TAD length / 30), located starting at the 10th monomer of the TAD. This made it such that ½ of all LEFs are born next to a TAD boundary, while the other ½ are born randomly.

To simulate folding of the “vermicelli” chromosomes, we increased the processivity in the best-fitting model 10-fold and 20-fold, to 2000 and 4000 respectively. For each value, we observe prominent folding of the chromosome in a linearly organized “vermicelli” state, reminiscent of the first stage of mitotic compaction (Goloborodko et al., 2016; Naumova et al., 2013).

ChIP-seq peaks around oriented CTCF motifs

Motifs were assigned to narrow peak calls by interval intersection using bedtools (Quinlan and Hall, 2010). If more than one motif mapped to a single peak (0.5% of hits), only the first motif found was assigned. The CTCF peaks identified in GM12878 (GSM935611) that have an overlapping motif occurrence (PFM: CTCF_known1) from (Kheradpour and Kellis, 2014) were ordered by fold enrichment value. The directionality profiles are centered at the peaks, but oriented so that all motif instances “point” in the 3’ direction. We then selected the 4000 most and least enriched CTCF peaks from this set and produced histograms of the summit positions of all called ChIP peaks for CTCF (GSM935611), Smc3 (GSM935376) and Rad21 (GSM935332). Histograms are centered at the peaks, but oriented so that all motif instances “point” in the 3’ direction.

Model of directional BEs with varying permeability

To convert ChIP-seq peak strength to the occupancy of simulated BEs, we performed the following steps: first, we calculated the middle of each ChIP-seq narrow peak as the (start + end)/2. For peaks without a unique CTCF_known1 motif match, the orientation was assigned using the distance to the nearest rad21 peak. Separately for each orientation, we then calculated a binned profile of the sum of peak fold-change-over-input values for peaks whose centers fall in each 600bp bin. Bins with no peaks were assigned a value of -infinity. We then transformed these values with a logistic function $f(x)=1/(1 + \exp(-x/20 - \mu))$, where μ was selected to be 3, and changed to 2 or 4 for respectively higher or lower CTCF occupancy, and 20 was selected to allow for a wide range of BE occupancies. We used a logistic function, as it naturally way to convert from arbitrarily dispersed values to numbers between 0 and 1. The resulting profiles were used as BE permeability for each orientation.

We modelled a 15MB region of human chr14, 60,000,000 to 75,000,000, which was chosen as it has gene deserts, gene-rich regions, and is uninterrupted by poor coverage bins. Simulations were performed in two replicates for each region, and 50,000 conformations (steps of SMC dynamics) were obtained for each replicate. The resulting contact map was then calculated, and averaged in 20x20 blocks, corresponding to 12kb resolution, to obtain a 1250x1250 contact map.

To compare simulated contact map with the Hi-C contact map, we used GM12878 *in Situ* data from (Rao et al., 2014), mapped, filtered and binned using *hiclib* software for each restriction enzyme (MboI and DpnII) and then pooled together. The data was obtained at 10kb resolution, linearly interpolated to 6kb resolution, and block-averaged in 2x2 blocks to obtain a 12kb-resolution contact map (1250x1250 bins).

To calculate spearman correlations between experimental and simulated contact maps as a function of genomic position, we first removed the strong effect of genomic distance on interaction frequency. As previously (Naumova et al., 2013) the observed interaction matrices were divided by an expected interaction matrix, calculated as the mean number of interactions at a given distance, using a sliding window with a linearly increasing size. Spearman correlations were then calculated using sliding a genomic window of 3Mb along by 1Mb maximum separation.

Model of direct BE-to-BE interactions

For simulations with direct BE-to-BE interactions, the inter-monomer interaction energy was calculated for $r < \text{attractionRadius}$, and was zero otherwise. The expression below defines U , the energy of interaction between any two monomers. Sticky1 and Sticky2 are index variables that are equal 1 for sticky BEs, and zero for all other monomers. Function $\text{step}(x)$ is a step-function; equals 1 for $x>0$, and zero otherwise. attrE is attraction energy in kT , was selected to be 1.5, 3, and 5. r is distance between centers of the particles.

```
emin12 = 46656.0 / 823543.0; rmin12 = sqrt(6.0 / 7.0)); attractionRadius = 1.5; repulsionRadius = 1;
ATTRdelta = (attractionRadius - repulsionRadius) / 2.0; rshft = (r - repulsionRadius - ATTRdelta) /
ATTRdelta * rmin12; ATTRReTot = min(Sticky1, Sticky2) * attrE * kT;
Eattr = -rshft^12 * (rshft^2 - 1.0) * ATTRReTot / emin12 - ATTRReTot; rsc = r * rmin12;
Erep = (rsc^12 * (rsc^2 - 1.0) / emin12 + 1 ) *3*kT; U = step(1 - r) * Erep + step(r - 1) * Eattr;
```

This function is described in our bitbucket repository: <https://bitbucket.org/mirnylab/openmm-polymer>, revision 9b4303b from 2015-12-27, file openmmlib/openmmlib.py, line 1213. Simulations were performed in 10 independent runs, recording 1000 blocks of 10000 MD steps each after skipping the first non-recorded 100 blocks.

Models of bulky and stiff boundary elements

To simulate a of bulky BEs, we started with the same polymer chain of 8 groups of three TADs of length 300, 600 and, 1200 monomers as used in the models with loop extrusion. However, instead of introducing any LEFs, several polymer chains were connected at each BE (either 3 of length 10 attached 1-per-monomer to the monomers around the BE via harmonic bonds, or 5 chains of length 6). For stiff BEs, the 10 monomers around the BE had an increased stiffness of 6, while other monomers had a stiffness of 1, as defined above. As there were no LEF dynamics, conformations were recorded every 50000 MD steps, after an initial period of 100 non-recorded blocks for a total of 1000 blocks and contact maps were averaged over 5 independent runs.

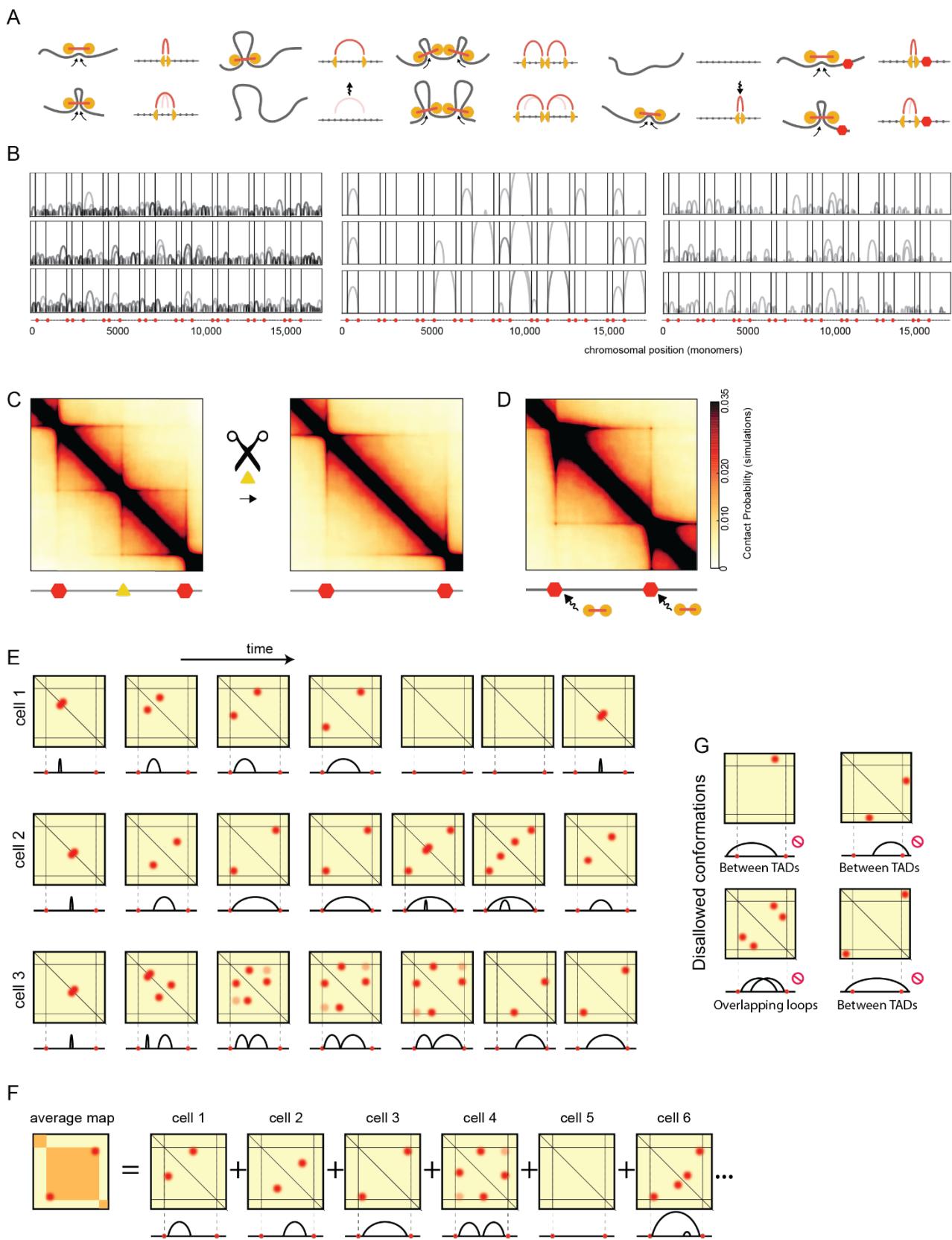


Figure S1

Figure S1. Model of the mechanism of loop extrusion, related to Figure 1.

A. Illustration of LEF update rules and lattice used for simulation of 1D LEF dynamics. As in the main figure, LEFs shown as linked pairs of yellow circles, chromatin fiber in grey. In the associated 1D lattices, LEF positions are shown by yellow triangles, and red arcs indicate the lattice sites (monomers) that are held by the LEF before and after the update. Left to right: extrusion, stalling at a LEF, dissociation, association, stalling at a BE (red hexagon); cartoons shown next to and their associated 1D lattice updates.

B. Three examples for each of three parameter values shown to highlight the diversity of looping interactions for three processivity-separation. Arcs indicate monomers held together by a given LEF. Arc height reflects loop size, and reinforced loops (i.e. those held by multiple LEFs) are in darker shades of grey. Positions of boundary elements (BEs) shown by vertical bars, illustrating how loops only form in-between neighboring BEs. Lattice with position of BEs as red hexagons shown schematically below. *Left:* processivity 720kb, separation 30kb. In the limit of small separation and large processivity, LEFs form a densely packed array of consecutive nested non-overlapping loops stretching between consecutive BEs. *Middle:* processivity 960kb, separation 960kb. In the limit of large processivity and large LEF separation, LEFs extrude the entire region between, and are stably bound at, consecutive BEs, forming a loop between the two boundary elements with high probability. *Right:* processivity 120kb, separation 120kb. In this regime of medium processivity and medium separations, both 2-4 times less than a TAD size, LEF dynamics encompass many events: a small fraction of LEFs happen to live long enough to extrude the entire region and transiently connect two BEs, although most LEFs dissociate before reaching the nearest boundary.

C. Modification of the minimal model: simulated deletion of a 20%-permeable BE halfway between two boundary elements separated by 720kb.

D. Modification of the minimal model: symmetric loading of LEFs. Half of the LEFs were loaded in a region directly adjacent to the start of each TAD, while the other half were loaded at random positions.

E. Loop extrusion is a stochastic process that follows a different trajectory in each cell. This figure outlines possible dynamics of LEFs using three examples. Each trajectory consists of 7 snapshots (frames), showing LEF positions as arches, and LEF-mediated loops (red dots) on the schematic single-cell Hi-C map. BE positions are indicated as red hexagons on the arch diagram, and as lines on the Hi-C map.

For cell 1, a LEF associates to the chromatin fiber (frame 1), and extrudes a loop until it meets a boundary (frame 3). It then continues to extrude a loop from the other side (frame 4), but soon dissociates (frame 5). After some time, a new LEF lands in a different location (frame 7).

For cell 2, a LEF associates to the chromatin fiber (frame 1) and continues to extrude a loop until both sides of a LEF are stalled against boundaries (frame 3). A LEF stays in the same stalled position (frame 4), while a new LEF associates to the chromatin within the loop created by the original LEF (frame 5). This new LEF continues to translocate, and the old LEF stays (frame 6), and finally dissociates (frame 7).

In cell 3, a LEF associates to the chromatin (frame 1), and another LEF associates nearby (frame 2). The two LEFs stall against each other (frame 3), which creates a weaker interaction between the beginning of the first LEF and the end of the second LEF as they are now nearby in space. The two LEFs continue to translocate until each of them hits a BE (frames 4 and 5). The first LEF then dissociates (frame 6), and the second LEF starts to translocate again (frame 7). This example

illustrates how corner-peaks may occur at distances longer than the LEF processivity, because they can be mediated by a chain of two or more LEFs stalled against each other.

F. Hi-C maps are formed as an average over many cells. In each of the cells, the configuration of LEFs is different. However, in each configuration, LEFs only form loops within TADs (i.e. loops do not cross BEs). Note that because LEFs stall at BEs, many configurations will have a LEF connecting two ends of a TAD. As a result, an average map may have a peak of interactions between a start and an end of a TAD. We also note that direct BE-to-BE loops are not the only contributors to the peak of interactions between neighboring BEs. As shown in cell 4, two LEFs stalled against each other and against neighboring BEs will bring the BEs close together in space, and thus contribute to the peak of interactions. In the more dense regimes of LEF dynamics (small separation, high processivity), even longer consecutive LEF arrays may contribute to the BE-to-BE interaction peak.

G. Certain configurations of loops are disallowed in the LEF dynamics described in this paper. In particular, a LEF will never connect two regions separated by a completely impermeable BE. Similarly, two loops will never cross (though they could be nested, as shown above).

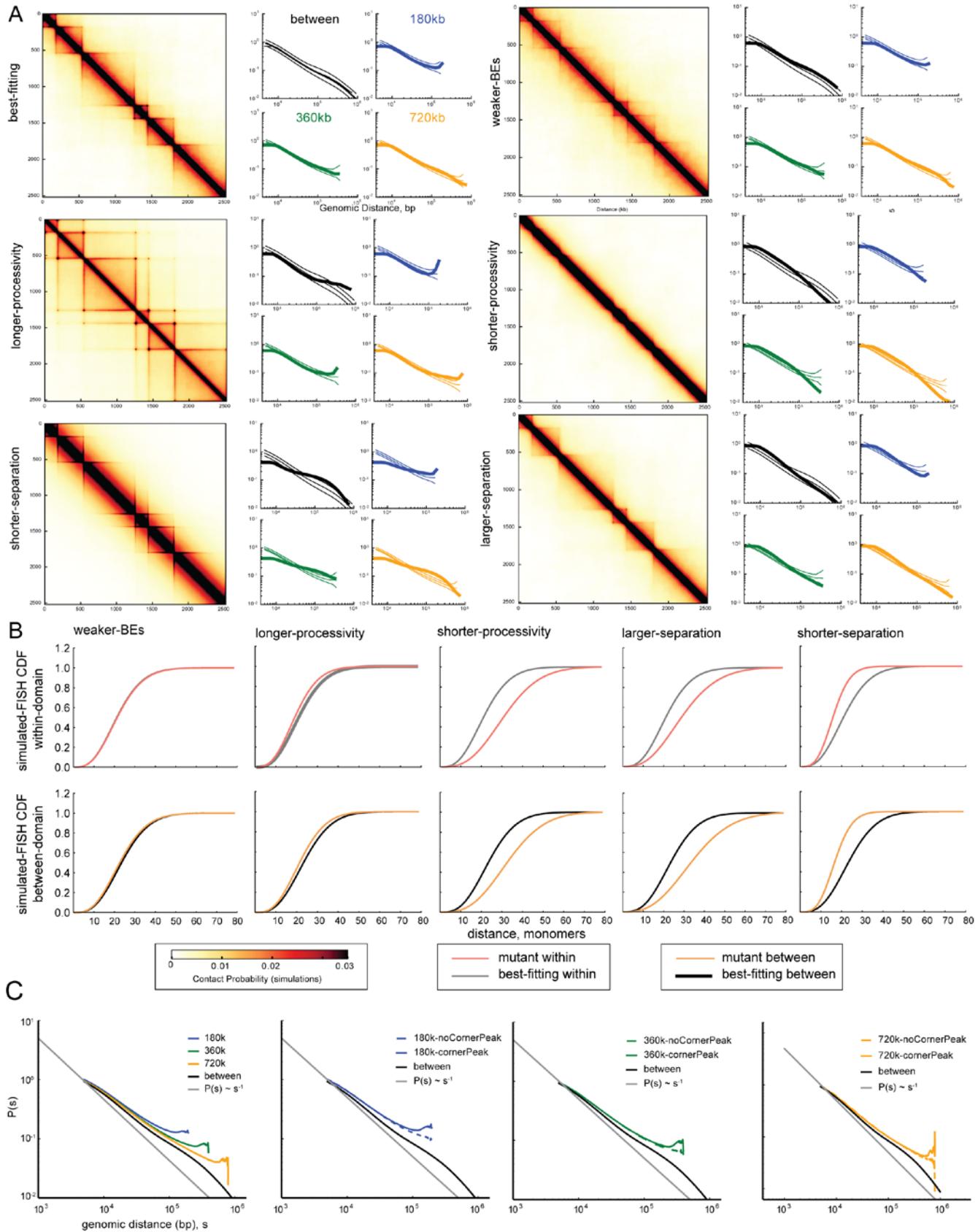


Figure S2

Figure S2. Experimentally testable predictions from perturbations of LEF dynamics, related to Figure 2.

- A.** Contact maps and $P(s)$ for perturbations of the best-fitting model (processivity 120kb, separation 120kb, 5000 steps, stiffness 2, density 0.2, cutoff 10.)
- B.** Simulated FISH CDF functions for pairs of region within the TAD, and in two neighboring TADs, both separated by 600 monomers (360 kb). Grey or black line corresponds to the best-fitting model, while orange line corresponds to the perturbation of the model.

Perturbations of the model are as follows:

For weaker-BEs, all BEs were made semi-permeable, with 50% permeability. For *longer-processivity, shorter-processivity, shorter-separation*, and *larger-separation* models, the processivity or separation were respectively increased or decreased by a factor of 4 compared to the best-fitting model.

The models correspond to possible experimental perturbations:

Weaker-BEs can correspond to CTCF depletion. *Longer-processivity* may correspond to the decreased unloading of LEFs, which can be achieved experimentally by depleting the LEF unloader (e.g. for cohesin, Wapl); *shorter-processivity* may come from over-expressing the unloader. *Shorter-separation* may be achieved by increasing the number of chromatin-bound active LEFs, either by increasing loading of LEFs (e.g. for cohesin, Nipbl), or by overexpressing LEFs; *larger-separation* can come from the opposite perturbations.

Results of the perturbations are the following:

Weaker-BEs (CTCF depletion) shows blurred TAD boundaries on the contact map. The FISH CDF doesn't change much for within-TAD probes, but the between-TAD CDF shifts to be closer to the within-TAD CDF.

Higher-processivity displays very strong corner peaks. Spatial distances decrease slightly, but not as much as in other processivity or separation changes.

Shorter-processivity model displays an almost complete loss of TADs, has increased distances in FISH for both within- and between-TAD, and makes the two curves more alike.

Shorter-separation tilts $P(s)$ curves to be less steep and decreases FISH distances. It also leads to sharper TAD boundaries.

Larger-separation makes TADs less pronounced and leads to an increase in FISH distances within and between TADs.

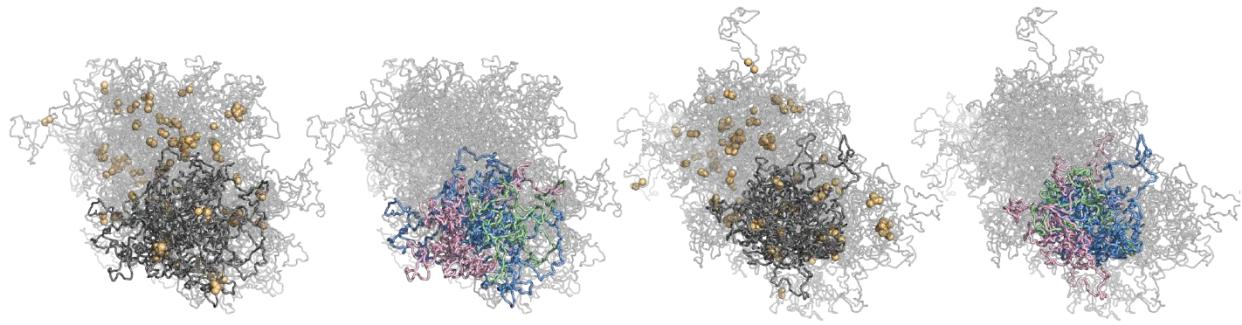
C. Average contact probability, $P(s)$, vs. genomic distance, s . Calculated from 2kb resolution GM12878 in-situ DpnII maps (Rao et al., 2014).

Left subplot: $P(s)$ for TADs of 180kb, 360kb, and 720kb, and between TADs using published TADs from GEO: GSE63525_GM12878_primary+replicate_Arrowhead_domainlist.txt.gz (Rao et al., 2014).

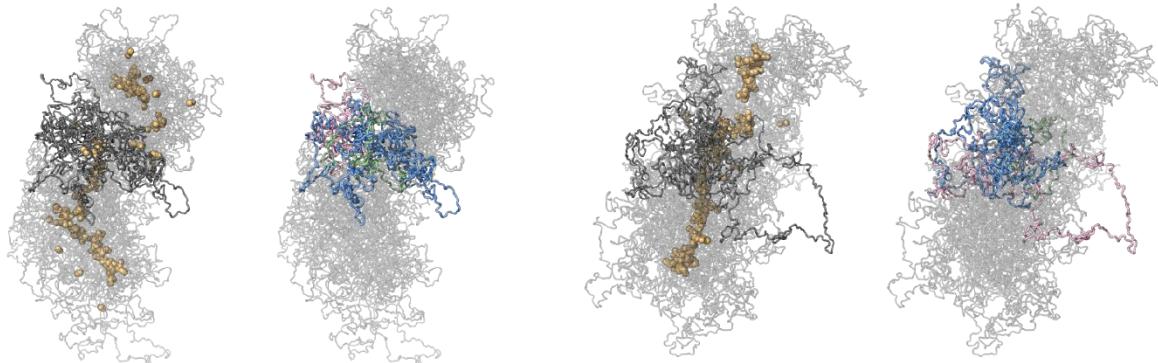
Other subplots: show TADs of the indicated lengths with and without corner peaks, as determined by having a published peak less than 50kb from the TAD corner (sum of distances between starts and between ends of peak/TAD),

Published peaks from GEO: GSE63525_GM12878_primary+replicate_HiCCUPS_looplist.txt.gz (Rao et al., 2014).

A



B



C

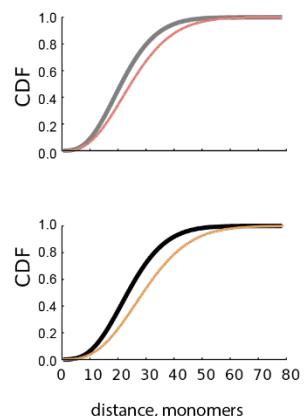
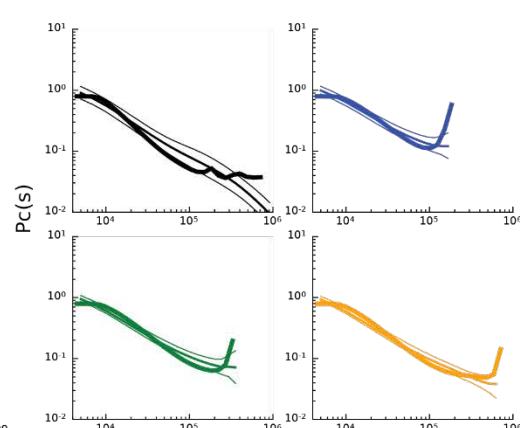
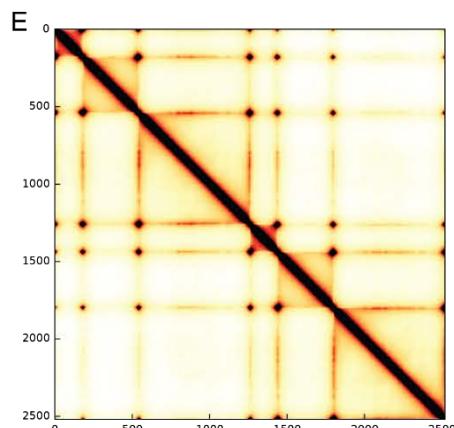
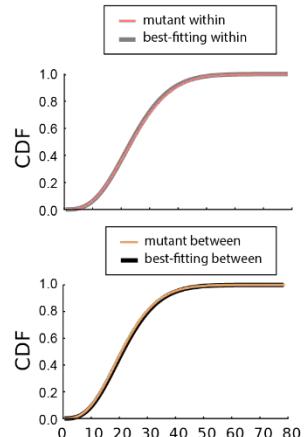
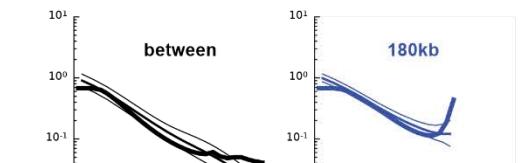
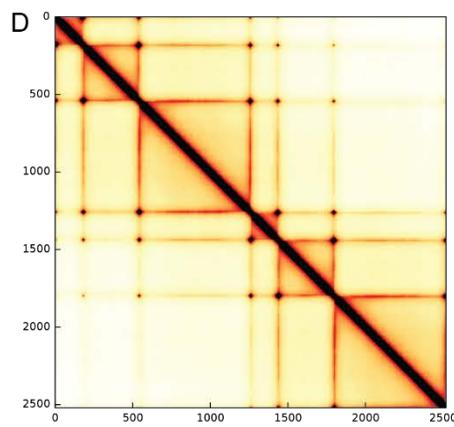
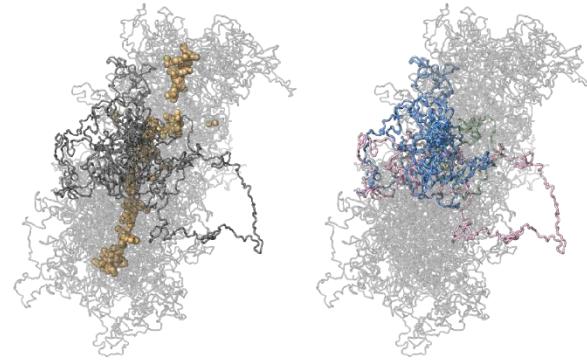


Figure S3

Figure S3. Spatial properties of TADs organized by LEFs; Increased LEF lifetime creates a prophase-like vermicelli state, related to Figure 3.

- A.** Conformations of a polymer subject to LEF dynamics for two time-points, for the best-fitting model. For each, left image shows LEFs (yellow spheres), and chromatin (grey); right image shows three neighboring regions between BEs of sizes (180kb, 360kb, 720kb) in (green, pink, blue) for the same conformations.
- B.** Conformations of a polymer subject to LEF dynamics for very high LEF processivity (1200kb), ten-times the processivity in **Fig 1E**; such high processivity might occur following depletion of LEF-unloading factors (e.g. Wapl). *Left*: shows LEFs (yellow spheres), and chromatin (grey), *right*: shows three neighboring regions between BEs of sizes (180kb, 360kb, 720kb) in (green, pink, blue) for the same conformation (as colored in **Fig 1D**). Interestingly, LEFs form a continuous backbone, despite the presence of boundaries.
- C.** As previously, but for even larger processivity (6000kb).
- D.** Processivity, separation = 1200kb, 120kb (as for **B**). *Left*: contact map, *Middle*: associated $P(s)$ for three TAD sizes and between TADs. *Right*: CDF. 3D-to-1D ratio 5000/4, stiffness 2, density .2, cutoff 8.
- E.** as for **D**, but with larger processivity (6000kb).

Notably, in addition to the emergence of a backbone in the spatial conformations, simulated contact maps display particularly bright peaks between BEs for highly processive LEFs. In contrast, Hi-C for mitotic chromosomes displayed relatively homogeneous contact maps (Naumova et al., 2013). This is congruent with a picture where condensins, which operate during mitosis, are not subject to stalling at BEs, unlike cohesins in interphase.

Interestingly, the results in (D,E) contrast with the result (**Figure S2**) that increasing the processivity to 480kb (4-fold) from the best-fitting model decreases FISH distances. Indeed, despite large changes on the contact map, the processivity 1200kb (10-fold increased) simulated FISH CDF are almost unchanged from the best-fitting model with 120kb, and the processivity 6000kb CDF actually shifts to larger distances. This behavior arises because the simulated FISH CDF measured at 360kb reflects both the overall compaction, as well as the average loop size. While the overall compaction of the chromosome is certainly higher for 1200kb and 6000kb than for 480kb, the average loop size is also higher due to the increased formation of nested loops (Goloborodko et al., 2015), increasing the loop size in this very dense regime increases the average distance of non-BE monomers to the LEF backbone, which in turn increases distances between any two non-BE monomers. We note this is analogous to how over-condensed mitotic chromosomes are thicker and shorter than normal mitotic chromosomes. Together, these results indicate the importance of measuring compaction at multiple scales for testing predictions of the loop extrusion mechanism.

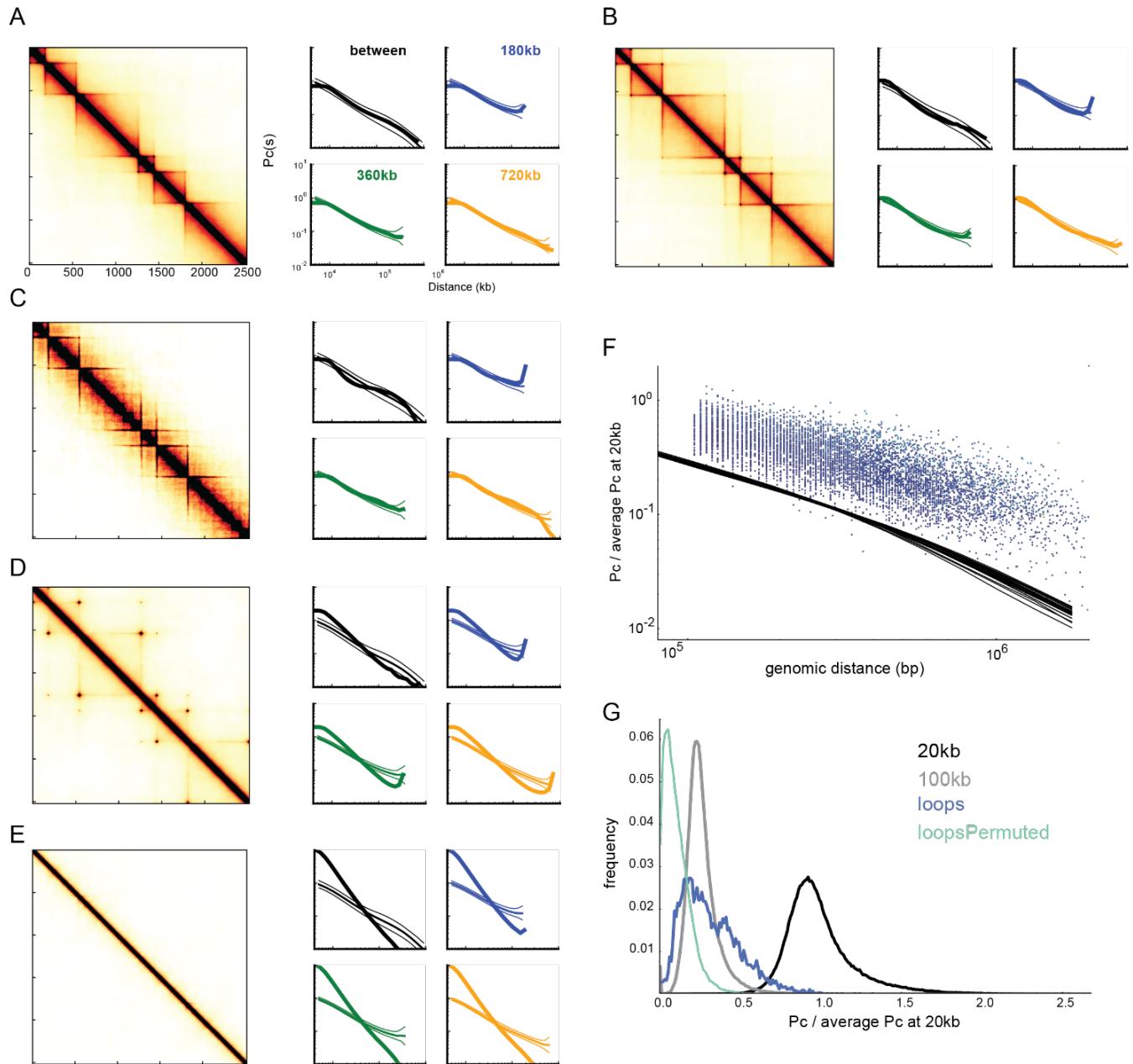


Figure S4

Figure S4. Corner-peaks are transient and naturally emerge at higher LEF lifetimes, related to Figure 4.

A. *Left:* Contact map showing TADs, for processivity 120kb, separation 120kb, as in **Fig 1E**. *Right:* contact probability vs. distance for TADs of three sizes, and between-TADs (trans). Corner-peaks are connected by a LEF (9%, 0.7%, <0.01%) of the time for TADs of (180kb, 360kb, 720kb).

B. As above, for processivity 240kb, separation 120kb. Corner-peaks are connected by a LEF (25%, 7%, 0.2%) of the time for TADs of (180kb, 360kb, 720kb).

Together, **a** and **b** show increasing LEF processivity naturally strengthens peaks at TAD corners. Note that for the increased processivity, the corner-peak for the largest 720kb TAD is clearly visible on the contact map and gives a slight increase of $P(s)$ at the end of the plot. Yet, the start and the end of this TAD are connected by a LEF only 0.2% of the time. This emphasizes that corner-peaks are not permanent loops; permanent loop would have much more drastic effect on the contact map.

C. An example of a dense regime that results in a contact map with very strong domain borders, yet fits the scaling quite well and is found within the 100 best-fitting models: goodness-of-fit of 1.1828 (rank 65). Parameters of the model are: processivity 720kb, separation 30kb, 3D-to-1D ratio 300/4, stiffness 0, density 0.05, cutoff 8.

This example illustrates that fitting multiple quantitative features of TADs, beyond simply $P(s)$, may be valuable for future studies.

D. An example of a “simple loop” regime with poor fit of 1.4137 (rank 2208 out of 6912). Processivity 960kb, separation 960kb, 3D-to-1D ratio 1000/4, stiffness 2, density 0.2, cutoff 8. This regime display strong loops between neighboring BEs.

This example also shows that simply bringing neighboring BEs together is insufficient to create TAD observed in Hi-C.

E. In the ‘free-polymer limit’ of short processivity and large separation, the effect of LEFs on the polymer is minimal. *Left:* contact map showing neither TADs nor peaks. *Right:* contact probability vs. distance for TADs of three sizes, and between-TADs (trans). Processivity 60kb, separation 960kb, 3D-to-1D ratio 5000/4, stiffness 0, density 0.05, cutoff 5.

F. contact probability $P(s)$ vs. genomic distance for corner-peaks (blue circles) using published locations from GM12878 and the associated Hi-C contact maps at 10kb resolution, normalized to 1 at 20kb. Black lines show average chromosome-wide $P(s)$ (one line per chromosome).

G. histogram of contact probability for all published peak-pairs (blue) versus all loci separated by 20kb (black) or by 100kb (grey), and permuted peak-pairs (green, with shuffling along diagonals of the Hi-C map to preserve the distribution of peak-pair genomic distances).

Together, **F** and **G** show that peak-loci do not appear to be in permanent contact in Hi-C data, and corner-peaks are best considered as transient loops.

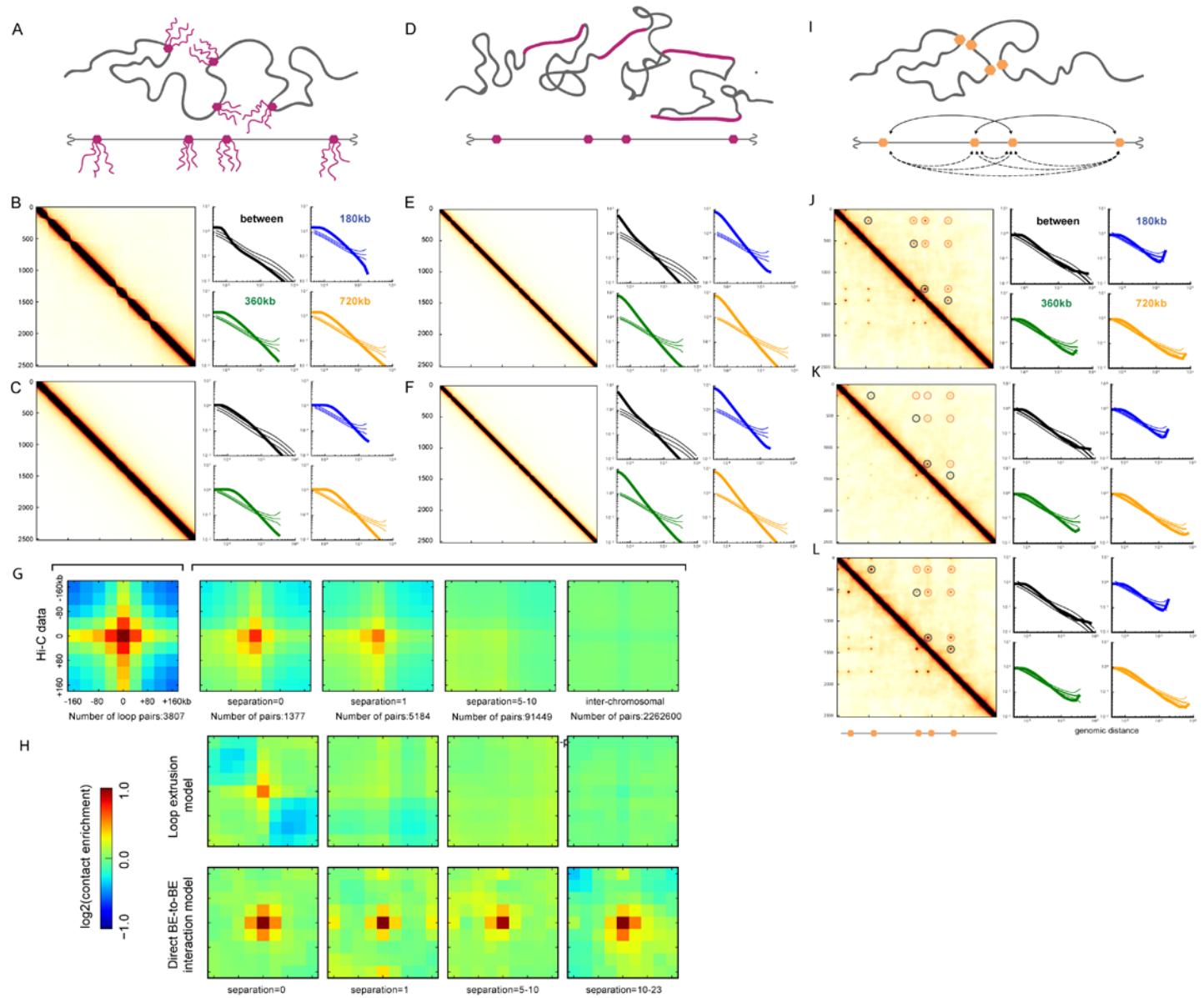


Figure S5

Figure S5. Difficulty of insulating large domains with small elements, related to Figure 5.

A. A schematic illustration of a model that explores whether BEs modeled as large bulky objects (e.g. bound by proteins or RNA) can insulate neighboring TADs. In these simulations, each BE is bound by several polymer chains (3 of length 10 in B; 5 of length 6 in C). Note that since our monomer represents 3 nucleosomes (~600 kDa), the bound mass would be 18 mDa in each case, which is more than 100 CTCF molecules (82-130 kDa) or 250 Rad21 molecules (72 kDa).

This simulation shows that direct physical blocking of interactions by BEs would not lead to insulation between neighboring TADs. While the bulky BE does create partial insulation between regions directly adjacent to the BE (a “break” in the diagonal of the contact map), this insulation does not extend much further and does not reach the length scale of a full TAD. Moreover, this bulky BE model never leads to peaks of contact probability between proximal BEs.

B-C. Contact maps and $P(s)$ for a bulky BE model (a) with 3 bristles of length 10 (**B**) or 5 bristles of length 6 (**C**).

D. A schematic illustration of a model in which we explore whether BEs may correspond to stiff regions of chromatin, thus physically separating two TADs. In this simulation, the majority of the chromatin fiber is flexible (stiffness 1), interspersed with stiff BEs (10 monomers with stiffness 6). While this model also creates insulation between regions directly adjacent to the BE, it does not extend much further and does not reach the length scale of a full TAD.

E,F. Contact maps and $P(s)$ for the stiff BE model; contact radius 10 for **E**, and 5 for **F**.

G. Enrichment plots around peak loci at 40kb resolution for observed-over-expected GM12878 in-situ (Rao et al., 2014) contact maps (i.e. after removing distance-dependence).

Left: enrichment only around peak pairs called for this dataset (i.e pairs of peak loci, GEO: GSE63525_GM12878_primary+replicate_HiCCUPS_looplist.txt.gz, (Rao et al., 2014)). Peak pairs show a pronounced peak in average enrichment.

Right: For this plot, we split the list of all peak pairs into a list of individual unique peak loci. We then calculate enrichment around pairs of peak loci separated by a given number of loci in between. We note that separation of zero is less enriched than that of peak-pairs, as separation zero does not imply that the two peak loci form a corner-peak; they could be the end of one corner peak and the beginning of the next peak. We also calculated enrichments for all between-chromosomal pairs of peak loci.

We find that peak loci at further separations display weaker peak in average enrichment. Notably, the peak completely disappears on the between-chromosomal map. This shows that the ability of peak-loci to mediate an interaction peak is not an inherent property, and suggests peaks are realized by a mechanism that acts along the chromosome. We note that a very small (~1%) depletion of interactions between peak loci between chromosomes, which is barely visible on the interaction map as a bluer cross, is consistent with the idea that loci involved in the formation of peaks are sterically excluded from interactions with other genomic regions (Doyle et al., 2014).

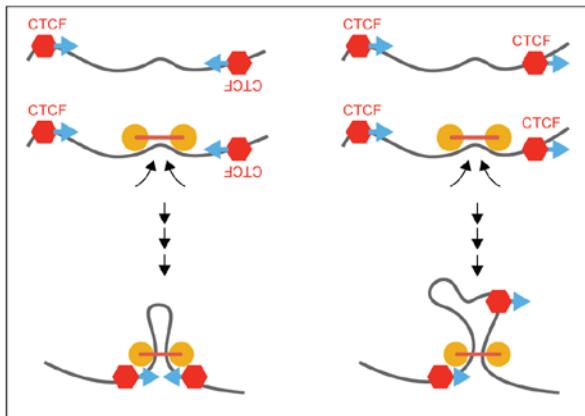
H. As in (*G, right*), but calculated for simulated datasets. Since we do not simulate different chromosomes, the last column corresponds to BEs separated by more than 10 other BEs. *Top:* Best-fitting loop extrusion model as in **1E**. *Bottom:* a model with direct BE-to-BE interactions (attraction strength 3, see below). Note that in the BE-to-BE interaction model strength of enrichment does not depend on the separation.

Importantly, in contrast with peaks formed by loop extrusion, which are limited to increased interactions between proximal BEs, peaks from the model with direct BE-to-BE interactions have increased interactions at all separations.

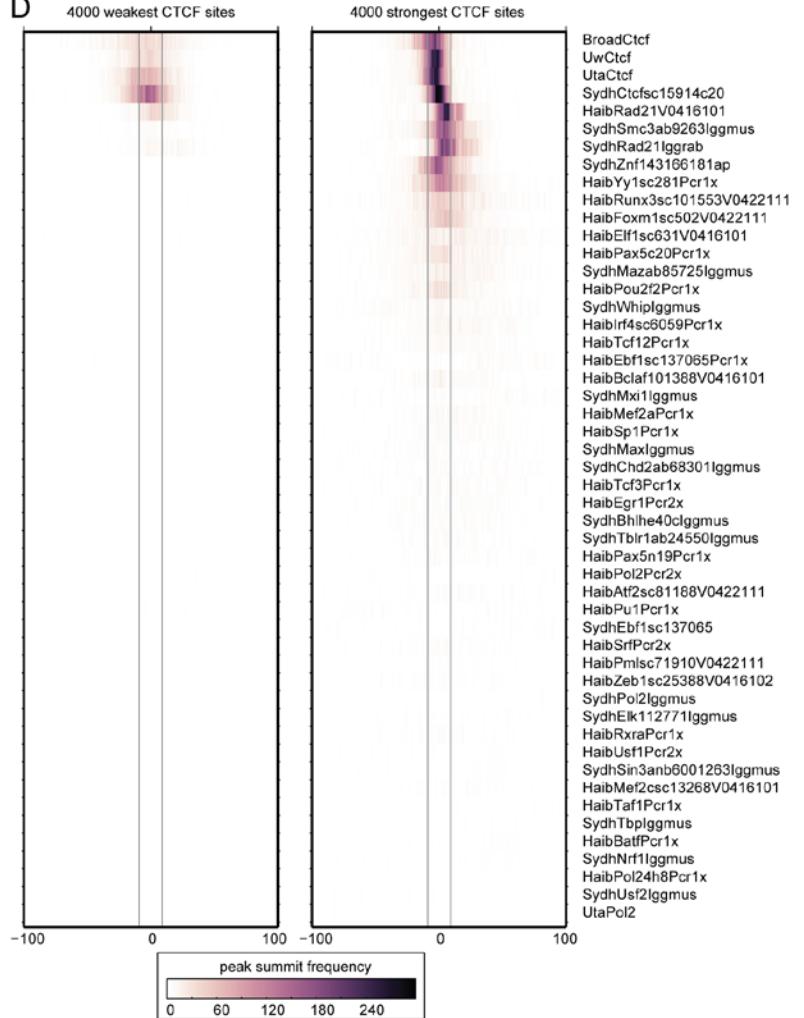
I-L. *models with direct BE-to-BE interactions form peaks, but not TADs.* In this alternate model, all pairs of BEs experience a short-range attractive force (with strength $1.5kT$, $3kT$, and $5kT$ in **J,K,L**). When two BEs are far away from each other, they do not interact and move randomly due to the diffusive motion of chromatin. When two BEs come in close proximity (<25 nm), they experience an attractive force. This would happen, for example, under the assumption that CTCF molecules bound at BEs have an intrinsic affinity for each other.

Note that, in contrast with experimental data, strong peaks are observed between all possible pairs of BEs in this model, and not preferentially between BEs at short separations along the linear genome. The outcome of this simulation suggests something more than long-range looping-mediated interaction is required to form TADs and peaks; in particular, an additional mechanism is required to restrict interactions to occur between proximal elements along the linear genome.

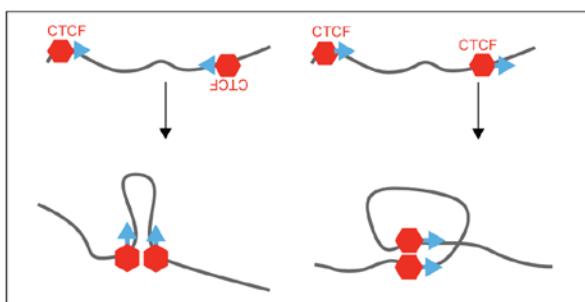
A



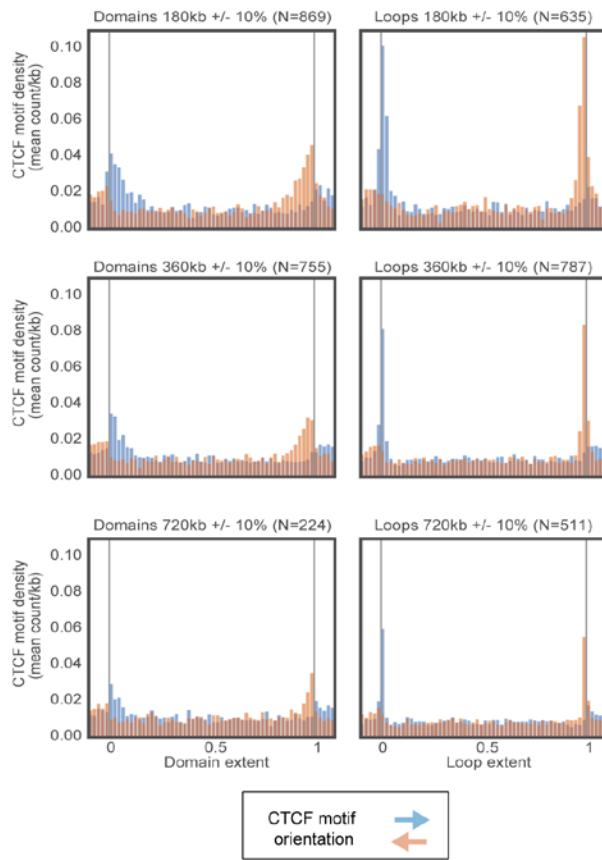
D



B



C



E

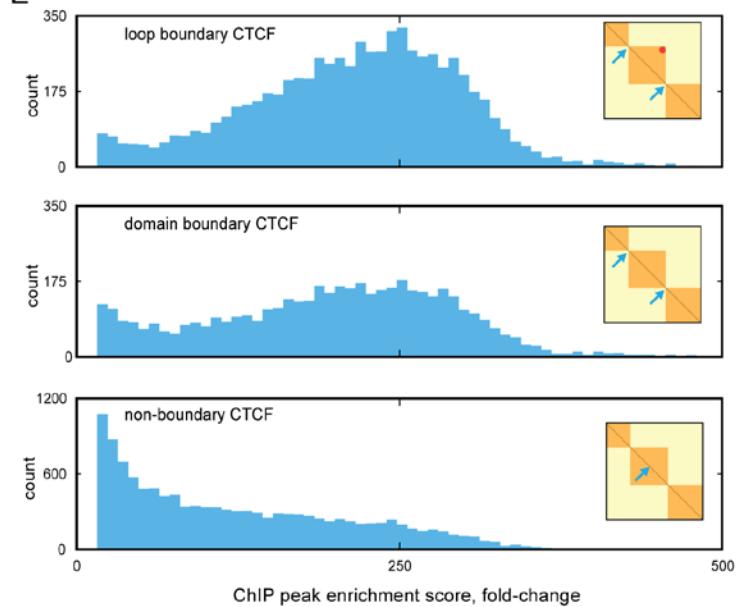
**Figure S6**

Figure S6. CTCF as a directional BE, related to Figure 6.

Genomics studies (Rao et al., 2014; Rudan et al., 2015) have revealed preferential enrichment of inwardly oriented CTCF motifs at TAD boundaries and between peak locus pairs; this has been taken to suggest that CTCF can stabilize chromatin loops in an orientation-dependent manner. We find that preferential enrichment persists over a range of TAD sizes and peak locus pair genomic distances (see below). Here we illustrate how two mechanisms of loop formation would affect relative to motif orientation in 3D. The first mechanism relies on loop extrusion, and in the second, two looping elements interact after a stochastic encounter in 3D. We illustrate how loop extrusion can lead to the observed orientation-specific enrichment, while 3D encounters cannot.

A. loop extrusion maintains the orientation of motifs over large separation, whether for convergent motif-pairs (*left*), or for same-orientation motif-pairs (*right*). If CTCFs halt loop LEF extrusion and stabilize loops in an orientation-dependent manner, then the mechanism of loop extrusion explored here can explain the observed enrichment in convergent CTCF motifs at TAD boundaries and loop bases, even at very large genomic separations.

B. encounters in 3D at genomic distances much larger than the persistence length of the chromatin fiber maintain no preferential orientation of motif-pairs.

C. *Left:* enrichment plot of CTCF motif orientation (similar to (Rudan et al., 2015)) over TADs called in GM12878 (Rao et al., 2014), for TADs of three sizes: 180kb, 360kb, and 720kb, with a +/-10% size range. Each TAD in a given size range was divided into 50 bins, and then extended by 5 bins to the left and right. The total number of + or - motif occurrences in each bin was aggregated, and divided by the bin size to give a motif "density" per kb. The bar heights give the average motif density in each bin. Motif occurrences and orientations were obtained from a publicly-available resource of regulatory motifs derived from a combination of known literature motifs and motifs discovered from ENCODE TF ChIP-seq data (Kheradpour and Kellis, 2014). We selected matches corresponding to motif "CTCF_known1" (Jaspar ID: MA0139.1, consensus 5'-TGGCCACCAGGGGGCGCTA-3') passing a position weight-matrix based p-value threshold. *Right:* similarly, but for peak loci (also called loops, in (Rao et al., 2014)). Note that motif orientation enrichment is similar over each range of distances, with TADs displaying a ~2-fold and loops ~3- to 5-fold enrichment of inward-facing motifs at each range of distances.

D. histograms of GM12878 ChIP-seq peak summits of TFs (ENCODE consortium, 2012) around the 4000 weakest (left) and 4000 strongest (right) CTCF core motif instances. The strength of a motif instance was determined by the maximal fold change vs control of the CTCF ChIP-seq peak overlapping the motif. The histograms span a 200-bp window and are aligned along the center of the core motif, whose width is depicted with vertical gray lines. The histograms are oriented along the reference strand of the corresponding motif instance. Shown are the 50 most enriched factors; the first five unique factors are CTCF, RAD21, SMC3, ZNF143 and YY1. Note that the cohesin subunit (SMC3 and RAD21) densities are shifted downstream relative to those of the other factors.

E. histograms of GM12878 ChIP-seq peak fold change values for CTCF peaks with a motif match that is proximal to annotated loop bases, domain boundaries and the remaining peaks not associated with either. The ChIP-seq data used is the uniformly processed peaks from the Sydh ENCODE production center. Motifs were assigned to peaks as in B; peaks were assigned to annotated loop base regions and to 10-kb windows around the annotated domain boundaries from (Rao et al., 2014). These histograms demonstrate that CTCF is more enriched at loop bases and domain boundaries relative to other regions.

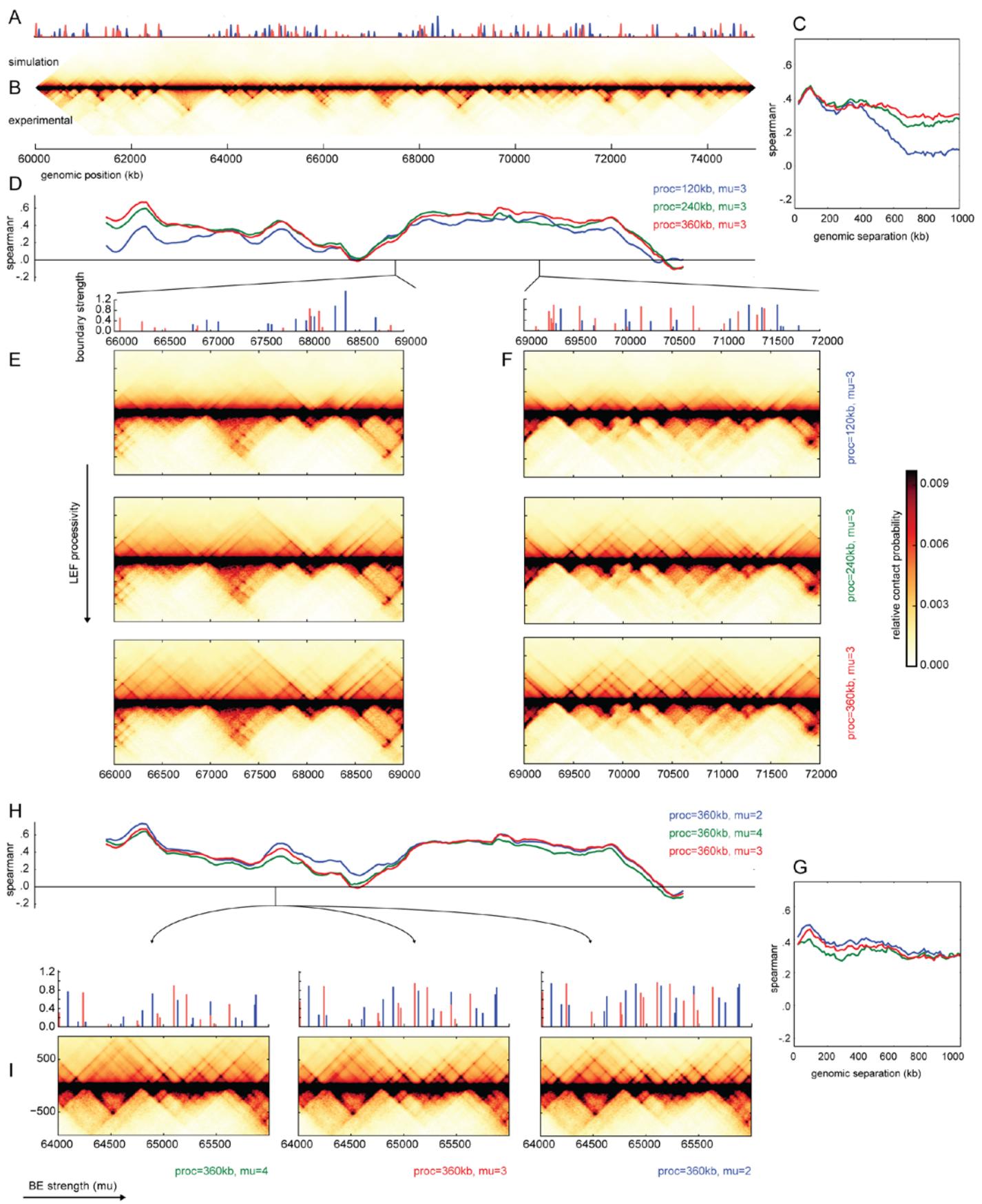


Figure S7

Figure S7. Model with orientation-specific BEs of varying permeability, related to Figure 7.

- A.** Orientation-specific BE strength profile across the simulated region of human chromosome 14, cell type GM12878, 12kb resolution. BE strength reflects the average number of BEs bound within a 12kb (20 monomer) region.
- B.** Simulated contact map of the full region of human chr14, GM12878 cell type. Maps are compared with experimental maps for the same regions at the same 12kb resolution.
- C.** Spearman correlation between experimental and simulated observed/expected contact maps as a function of genomic distance for simulations with BE parameter $\mu=3$, and LEF processivity 120kb (blue), 240kb (green), and 360kb (red). Note this corresponds to correlation between the maps at increasing distances from the center line of the maps.
- D.** Correlation profile between experimental and simulated observed/expected along the chromosome for sliding genomic windows of 3Mb along the chromosome by 1Mb maximum separation.
- E, F.** Simulated contact maps for two indicated regions with BE parameter $\mu=3$, and LEF processivity 120kb, 240kb, and 360kb.
- G.** Spearman correlation between experimental and simulated observed/expected contact maps as a function of genomic distance for simulations with LEF processivity 360kb for BE parameter $\mu= 2$ (blue), 4 (green), 3 (red).
- H.** Correlation profile between experimental and simulated observed/expected (as in D).
- I.** Simulated contact maps for two indicated regions with LEF processivity 360kb for BE parameter $\mu= 2, 3, 4$.

Movie and Database Captions

Movie-M1, related to Figure 1.

This movie illustrates dynamics of a 120-kb TAD in a best-fitting model (**Figure 1E**). 320 consecutive simulation blocks (i.e. rounds of 3D-simulation time-steps followed by 1D-simulation time-steps) are shown, with 6 linearly interpolated conformations between any two blocks. Loop bases are shown as spheres, and regions within loops are shown in black.

Movie-M2, related to Figure 1.

Modification of the **Movie M1** for the model with twice-longer processivity, 240 kb compared to 120kb in the **Movie M1**. Monomers in loops are highlighted by black. For this value of processivity, loops-inside-loops occurred frequently. Monomers in the inside of multiple loops are highlighted by purple. Note that when the outer loop in a loop-in-loop conformation disappears, purple changes to black.

Data D1, related to Figure 2 and Experimental Procedures *goodness-of-fit*.

The database contains goodness-of-fit for all 6912 parameter-sets.

Supplemental References

- Chandra, T., Ewels, P.A., Schoenfelder, S., Furlan-Magaril, M., Wingett, S.W., Kirschner, K., Thuret, J.Y., Andrews, S., Fraser, P., and Reik, W. (2015). Global reorganization of the nuclear landscape in senescent cells. *Cell Rep.* 10, 471–484.
- Criscione, S.W., De Cecco, M., Siranosian, B., Zhang, Y., Kreiling, J.A., Sedivy, J.M., and Neretti, N. (2016). Reorganization of chromosome architecture in replicative cellular senescence. *Sci. Adv.* 2, e1500882.
- Dekker, J., and Heard, E. (2015). Structural and functional diversity of Topologically Associating Domains. *FEBS Lett.* 589, 2877–2884.
- Dileep, V., Ay, F., Sima, J., Vera, D.L., Noble, W.S., and Gilbert, D.M. (2015). Topologically-associating domains and their long-range contacts are established during early G1 coincident with the establishment of the replication timing program. *Genome Res.* gr – 183699.
- Eastman, P., and Pande, V. (2010). OpenMM: A Hardware-Independent Framework for Molecular Simulations. *Comput. Sci. Eng.* 12.
- Eastman, P., Friedrichs, M.S., Chodera, J.D., Radmer, R.J., Bruns, C.M., Ku, J.P., Beauchamp, K.A., Lane, T.J., Wang, L.P., Shukla, D., et al. (2013). OpenMM 4: A Reusable, Extensible, Hardware Independent Library for High Performance Molecular Simulation. *J Chem Theory Comput* 9, 461–469.
- Gibcus, J.H., and Dekker, J. (2013). The Hierarchy of the 3D Genome. *Mol. Cell* 49, 773–782.
- Goloborodko, A.A., Imakaev, M., Marko, J., and Mirny, L. (2016). Compaction and segregation of sister chromatids via active loop extrusion. *bioRxiv*.
- Gorkin, D.U., Leung, D., and Ren, B. (2014). The 3D Genome in Transcriptional Regulation and Pluripotency. *Cell Stem Cell* 14, 762–775.
- Imakaev, M., Fudenberg, G., McCord, R.P., Naumova, N., Goloborodko, A., Lajoie, B.R., Dekker, J., and Mirny, L.A. (2012). Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods* 9, 999–1003.

- Imakaev, M. V., Tchourine, K.M., Nechaev, S.K., and Mirny, L. a (2015). Effects of topological constraints on globular polymers. *Soft Matter* 11, 665–671.
- Kheradpour, P., and Kellis, M. (2014). Systematic discovery and characterization of regulatory motifs in ENCODE TF binding experiments. *Nucleic Acids Res.* 42, 2976–2987.
- Kind, J., Pagie, L., de Vries, S.S., Nahidazar, L., Dey, S.S., Bienko, M., Zhan, Y., Lajoie, B., de Graaf, C.A., Amendola, M., et al. (2015). Genome-wide maps of nuclear lamina interactions in single human cells. *Cell* 163, 134–147.
- Lajoie, B.R., Dekker, J., and Kaplan, N. (2014). The Hitchhiker's Guide to Hi-C Analysis: Practical guidelines. *Methods* 72, 65–75.
- Le, T.B.K., Imakaev, M. V., Mirny, L.A., and Laub, M.T. (2013). High-Resolution Mapping of the Spatial Organization of a Bacterial Chromosome. *Science* (80-.). 342, 731–734.
- Naumova, N., Imakaev, M., Fudenberg, G., Zhan, Y., Lajoie, B.R., Mirny, L.A., and Dekker, J. (2013). Organization of the mitotic chromosome. *Science* 342, 948–953.
- Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842.
- Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al. (2014). A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* 159, 1665–1680.
- Seitan, V.C., Faure, A.J., Zhan, Y., McCord, R.P., Lajoie, B.R., Ing-Simmons, E., Lenhard, B., Giorgetti, L., Heard, E., Fisher, A.G., et al. (2013). Cohesin-Based chromatin interactions enable regulated gene expression within preexisting architectural compartments. *Genome Res.* 23, 2066–2077.