Formation of Chromosomal Domains by Loop Extrusion

Geoffrey Fudenberg[*,1], Maxim Imakaev[*,2], Carolyn Lu[3],
Anton Goloborodko[2], Nezar Abdennur[4], Leonid A. Mirny[1,2]

* Joint first authors listed alphabetically
1. Graduate Program in Biophysics, Harvard University, Cambridge, Massachusetts, USA.
2. Department of Physics, Massachusetts Institute of Technology (MIT), Cambridge, Massachusetts, USA.
3. Program for Research in Mathematics, Engineering and Science for High School Students (PRIMES) and Undergraduate Research Opportunities Program (UROP), MIT, Cambridge Massachusetts, USA.
4. Program in Computational and Systems Biology, MIT, Cambridge, Massachusetts, USA.

Abstract

Characterizing how the three-dimensional organization of eukaryotic interphase chromosomes modulates regulatory interactions is an important contemporary challenge. Here we propose an active process underlying the formation of chromosomal domains observed in Hi-C experiments. In this process, cis-acting factors extrude progressively larger loops, but stall at domain boundaries; this dynamically forms loops of various sizes within but not between domains. We studied this mechanism using a polymer model of the chromatin fiber subject to loop extrusion dynamics. We find that systems of dynamically extruded loops can produce domains as observed in Hi-C experiments. Our results demonstrate the plausibility of the loop extrusion mechanism, and posit potential roles of cohesin complexes as a loop-extruding factor, and CTCF as an impediment to loop extrusion at domain boundaries.

Main Text

Recently, chromosome conformation capture experiments have revealed contiguous domains of enriched interactions in eukaryotic interphase chromosomes at the sub-megabase scale, referred to either as topologically-associated domains (*1*, *2*) or, more simply, domains (*3*, *4*). Functional studies have provided evidence for the role of these domains for gene expression and development (*5*–*7*). Identifying mechanisms of domain formation remains an important open question.

Domains are contiguous regions of enriched contact frequency that appear as squares in a Hi-C map (**Fig 1A**) (*1*–*4*), and are relatively insulated from neighboring regions. Domains differ from larger-scale compartments (*8*, *9*) in that they do not necessarily form an alternating ‘checkerboard’ pattern of enriched contact frequencies. Domains often have relatively well-defined boundaries (*1*–*3*). Many domains have homogeneous interiors, while others have complex and hierarchical structures (**fig. S1**). More recently, peaks of interactions were observed between loci at the boundaries of domains (“peak-loci”(*3*)).

Previous polymer models of interphase organization have focused primarily on characterizing chromosome structure rather than mechanisms of folding (*10*–*12*). Others (*13*, *14*) used hard-wired interaction preferences to model alternating patterns that are characteristic of compartments, rather than domains. One proposed mechanism giving good agreement to the observed domain organization relies on supercoiling (*15*), though the role of supercoiling in eukaryotes remains unclear.

We find several distinct observations supporting that domains are organized by a 1D process that operates in *cis*, along the chromatin fiber, rather than by hard-wired interaction preferences. First, we find that interactions between pairs of peak-loci at large separations on the same chromosome, or from different chromosomes, display no enrichment (**fig. S2**). This shows that the ability to mediate an interaction peak is not an intrinsic property of peak-loci, but instead suggests that peaks are realized by a mechanism that acts along the chromosome. Second, upon deletion of a domain boundary, the domain spreads to the next boundary (*2*, *7*); this shows that preferential interactions between loci in a domain are not hard-wired and that boundary elements play a crucial role. Third, enrichment of inwards-facing CTCF binding sites was observed at peak-loci (*3*, *16*). Importantly, if proper CTCF orientation underlies loop formation, this can be naturally realized by a cis-acting translocation mechanism (see below). Together these favor a process that operates along the chromatin fiber and is limited by boundary elements (BEs). We demonstrate how this can be accomplished by loop extrusion, extending previous 1D models (*17*).

Here we examine a mechanism whereby interphase chromosomal domains are actively compacted by cis-acting loop-extruding factors (LEFs). In this process, LEFs associate and dissociate from the chromatin fiber, and translocate along it (**Fig. 1B-C**). When a LEF binds the chromatin fiber, it first holds together two directly adjacent regions. The LEF then translocates along the chromatin fiber in both directions, holding together progressively distant regions of a chromosome, i.e. extruding a loop (**Movie-M1, Movie-M2**). Translocation stops when the LEF encounters an obstacle, either another LEF, or a BE; if halted only on one side, LEFs continue to extrude on the other. Throughout this process, LEFs can stochastically dissociate, releasing the extruded loop.

Boundary elements (BEs) are fixed genomic loci that stall LEF translocation, ensuring that extruded loops do not cross domain boundaries. This leads to enrichment of interactions within domains and effective insulation between domains. This insulation does not arise from an intrinsic ability of BEs to physically block interactions between distal genomic regions, but relies on their ability to regulate the translocation of LEFs. Note that extruded loops differ from loops that might be formed by proteins which simply bridge two genomic elements when they come into spatial proximity (*14*, *18*, *19*), as the latter mechanism has no way of distinguishing between distant or proximal chromosomal regions.

To consider how loop-extrusion dynamics can spatially organize a chromosome, we model a 10Mb region of the chromatin fiber as a polymer subject to the activity of associating and dissociating LEFs (**Fig 1C**). We model the chromatin fiber as a series of 10nm monomers, each representing roughly three nucleosomes, ~600bp (*20*). As previously (*20*), the polymer has excluded volume interactions, has no topological constraints and is simulated by Langevin dynamics using OpenMM (*21*). LEFs impose a system of bonds on the polymer: a bound LEF forms a bond between the two ends of an extruded loop, and the bond is re-assigned to increasingly separated pairs of monomers as a LEF translocates along the chromosome; when a LEF unbinds, this bond is removed. BEs, which halt LEFs translocation, were placed in fixed positions with sequential separations of 180kb, 360kb, and 720kb, through the 10Mb region.

The dynamics of loop extrusion are determined by two independent parameters (**Fig 2B, fig. S3**): the average linear *separation* between LEFs, and the LEF *processivity*, i.e. the average size of a loop extruded by an unobstructed LEF over its lifetime. Our model is additionally characterized by parameters governing the diffusivity of chromatin, the polymer stiffness and density, and the Hi-C capture radius (**Methods**). For each set of parameter values, we ran polymer simulations long enough to allow many association/dissociation events per LEF (10-160 events, **Movie-M1, Movie-M2**). From simulations, we obtain an ensemble of chromosome conformations (**Fig 1D**) and compute contact frequency maps (“simulated Hi-C”, **Fig 1E**) that can be compared with experimental Hi-C data.

For a range of LEF processivities and separations we observe the formation of domains on a simulated Hi-C map (**Fig 2C-E**), with many features observed in experimental Hi-C maps. For some parameter values we observe formation of homogenous domains; other simulated parameter sets lead to formation of peaks at corners of domains, or enrichment of contacts at the boundary of domains (**Database D1**). We observed neither domains nor peaks in the limit of short processivity and large separation that corresponds to a free polymer with excluded volume (**fig. S4**).

We next tested the ability of our model to reproduce an important quantitative characteristic of domain organization observed in Hi-C data: the dependence of Hi-C contact frequency *P(s)* with distance *s*, used previously (*14*, *15*, *20*, *22*, *23*). We aim to reproduce both *P(s)* within domains (separately for domain sizes 180kb, 360kb and 720kb) and *P(s)* between domains, which is ~1.5 fold smaller (**Fig 2A, fig. S5**). For each of 6912 parameter-sets, we determined the goodness-of-fit as the geometric standard deviation between the four experimental and simulated *P(s)* curves (**Methods**). Since many different parameter-sets reproduce experimental *P(s)* (**Fig 2A**), we consider how frequently each pair of values for LEF processivity and LEF separation are found in the top-100 parameter-sets (**Fig 2B**).

The best agreement with Hi-C data is achieved when separation between LEFs (~120Kb) is approximately equal their processivity (120-240kb, **Fig 2A**). In this regime, LEFs extrude loops relatively independently, as there are substantial gaps between LEFs (30-72% coverage of domains by loops). Due to the LEF-stalling function of BEs, loops of various sizes are dynamically formed within, but not between domains. Each loop, in turn, leads to enrichment of intra-domain interactions by a direct contact between its bases. Collectively, these effects lead to ~2-fold enrichment of contact probability within a domain (**fig. S6**). Notably, while adjacent domains display depletion of contact probability, polymer conformations display high spatial overlap of adjacent domains rather than appearing as segregated globules (**fig. S7C**).

Many Hi-C domains have peaks of interactions at their corners (~50%, (*3*)). Domains with and without peaks have similar *P(s)*, suggesting a similar underlying organizational mechanism, independent of the corner-peak (**fig. S5**). Our model can produce both types of domains, as increasing LEF processivity naturally strengthens peaks at domain corners while LEF dynamics still provide within-domain enrichment (**Fig 2E-F, fig. S6, fig. S8**). Still, these visibly strong peaks are not in permanent contact. Similarly, in Hi-C data, loci at corner-peaks do not appear to be in permanent contact (**fig. S8**). In fact, we find that strong loops between BEs provides among the worse fits (**fig. S9B**), with exceedingly strong corner-peaks and a lack of visible domains. While domains and TADs have been portrayed as loops, our results indicate that that a single stable loop does not describe domains as observed in Hi-C (*15*, *24*). Together, our model suggests that domains result from the dynamic activity of LEFs in the region between BEs, whereas corner-peaks emerge when LEFs transiently form BE-to-BE loops.

Several extensions of our basic model can reproduce additional features observed in Hi-C data. Introducing BEs that are present in a faction of cells or are semi-permeable can lead to nested domain-in-domain organization (**Fig 3A**). Similarly, deletion of a BE would lead to spreading of a domain until the next BE, similar to observations in genome-engineering experiments (*2*, *7*) (**fig. S10**). Finally, uneven loading of LEFs leads to asymmetric domains (**Fig 3B**). We note that *in-vivo* LEF dynamics may have many additional subtleties. For example, the lifetimes of LEFs stalled at BEs may be different from moving LEFs; LEFs may backtrack or pause; there may be several types of LEFs with different processivities; and LEFs may be actively loaded or unloaded at particular elements.

Certain architectural proteins are attractive molecular candidates for LEFs and BEs. Both cohesin and condensin complexes have been hypothesized to have the ability to extrude chromatin loops (*17*, *25*), and may serve as LEFs. In interphase, cohesins have been implicated in domain organization (*26*–*28*) and chromatin loops (*29*) beyond their role in sister chromatid cohesion, and have been observed to dynamically bind to chromatin, even before replication (*30*). Additionally, cohesin is enriched at interphase domain boundaries (*1*) and loops (*3*), and its depletion decreases domain strength (*26*, *27*). Finally, increasing cohesin binding time by depleting the cohesin unloader Wapl (*31*) condenses interphase chromosomes into a prophase-like ‘vermicelli’ state; a similar change occurs in our model if LEF processivity is greatly increased (**fig. S11**).

Boundary-elements in our model correspond to any impediment to loop extrusion; CTCF is a particularly relevant molecular candidate (*32*, *33*). CTCF is enriched at domain boundaries (*1*, *3*, *16*, *27*), its depletion decreases domain strength (*27*), and it has a relatively long residence time on chromatin (*34*). In addition, analogous to LEF accumulation at BEs in our model (**Fig 3D**), cohesin accumulates at CTCF binding sites, but only when CTCF is bound at these sites (*35*).

Finally, if CTCFs halt loop 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 sites at domain boundaries and loop bases (*3*, *16*), even at very large genomic separations (**Fig 3C, fig. S12**). Indeed, CTCF binding sites are oriented such that the C-terminus of bound CTCF (*34*), known to interact with cohesin (*36*), faces the interior of domains. The interaction of CTCF and cohesin may stabilize cohesin-mediated loops either directly or by shielding cohesins from the unloading action of SA2-interacting Wapl (*37*), similar to shugoshin (*38*) and sororin (*39*).

With these molecular roles, we predict that depletion of either CTCF or cohesin from chromosomes would disrupt domains, but would differentially affect spatial distances. Depletion of CTCF would not affect distances between loci within domains, but would decrease distances between neighboring domains to the within-domain level. In contrast, depletion of cohesin would increase distances for loci within domains. Available imaging data supports decompaction following cohesin depletion (*26*, *27*, *40*), and lack of decompaction following CTCF depletion (*40*). Full validation would require targeting specific regions and using more efficient methods for architectural protein removal.

The mechanism of loop extrusion studied here is similar to the proposed mechanism of mitotic chromosome condensation (*17*, *20*), but with the addition of BEs and many fewer, less processive, LEFs. Accordingly, increasing the number and processivity of LEFs and removing BEs could underlie the transition from interphase to mitotic chromosome organization. Conversely, upon exit from mitosis, interphase 3D chromosome organization can be re-established simply by restoring BE positions, which could potentially be epigenetically inherited bookmarks (*41*).

FIGURE 1

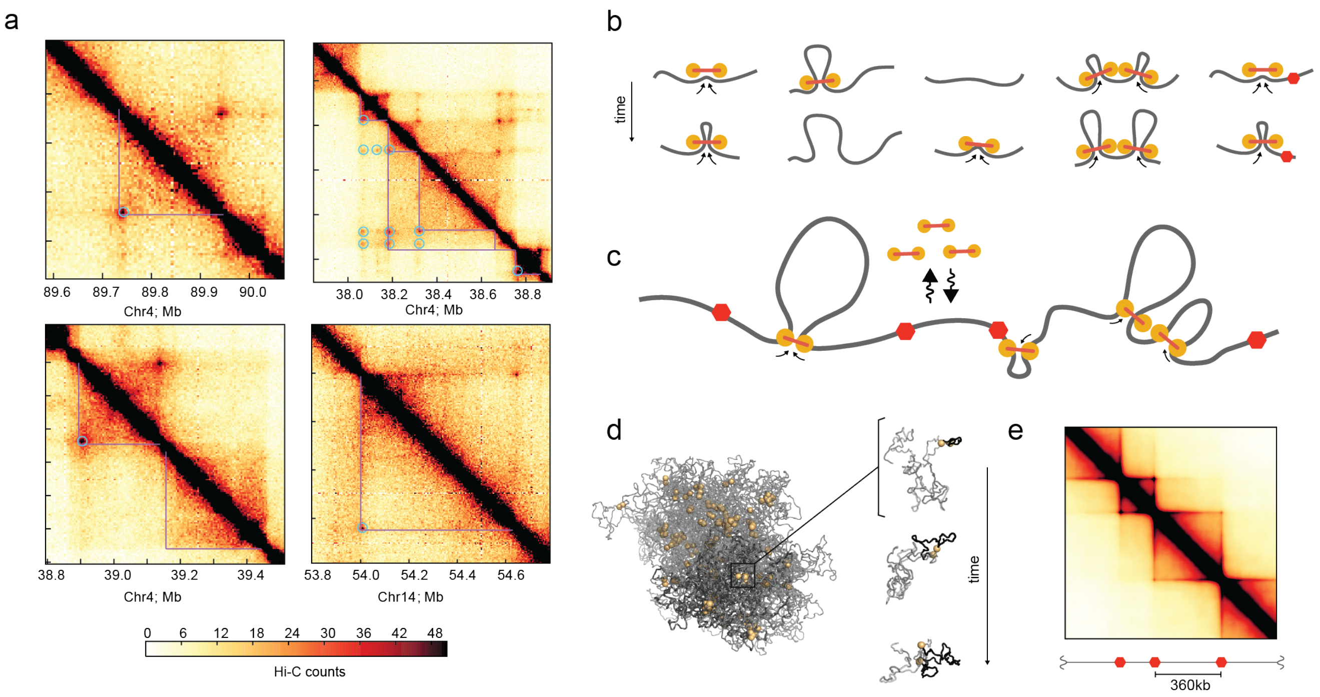

Fig 1. Loop extrusion as a mechanism domain formation.

a. Examples of Hi-C contact maps at 5kb resolution showing domains from four chromosomal regions (GM12878 in-situ MboI (*3*)), highlighting domains (purple lines) and interaction peaks (blue circles).

b. Model of LEF dynamics, LEFs shown as linked pairs of yellow circles, chromatin fiber in grey. From left to right: extrusion, dissociation, association, stalling upon encountering a neighboring LEF, stalling at a BE (red hexagon).

c. Schematic of LEF dynamics (**Movie-M1, Movie-M2**).

d. Conformations of a polymer subject to LEF dynamics, with processivity 120kb, separation 120kb. *Left:* shows LEFs (yellow), and chromatin (grey), for one conformation, where darker grey highlights the combined extent of three regions of sizes (180kb, 360kb, 720kb) separated by BEs. *Right:* shows the progressive extrusion of a loop (black) within a 180kb region.

e. Simulated contact map for processivity 120kb, separation 120kb.

FIGURE 2

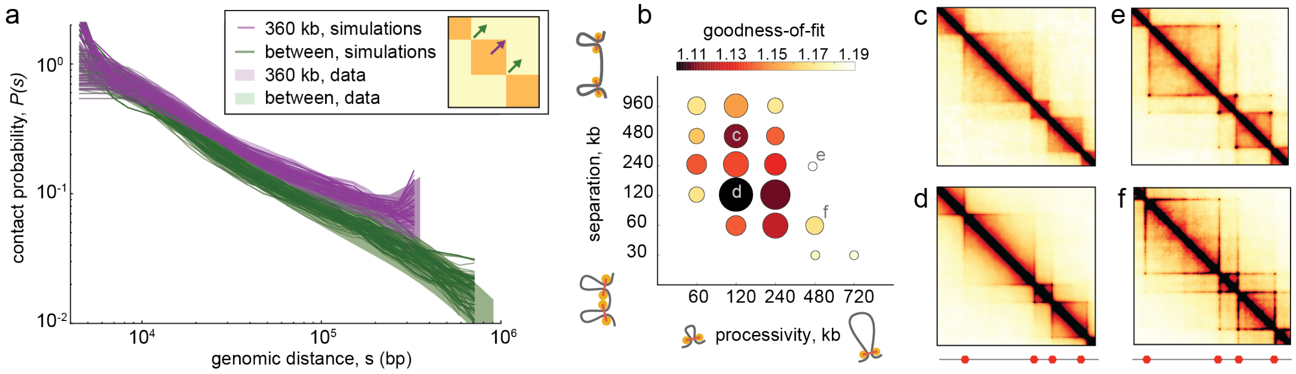

Fig 2. Quantitative analysis of loop extrusion.

a. Experimental *P(s)* (shaded areas) versus simulated *P(s)* for the 100 best-fitting parameter-sets (lines, one per parameter-set) within domains (purple) and between domains (green). Experimental *P(s)* calculated from 2kb contact maps and normalized to one at 4kb; shaded area shows 10th and 90th percentiles at each genomic distance. Simulated *P(s)* shown with vertical offsets obtained from fitting procedure (**Methods**).

b. Goodness-of-fit versus LEF processivity and separation for the 100 best-fitting parameter-sets (from 6912 total parameters-sets). Circle areas represent the number of parameter-sets among the top-100, while color quantifies the best-fit at each *processivity-separation* pair; a value of 1 indicates a perfect fit.

c-f. Simulated contact maps for the indicated *processivity-separation* pairs.

FIGURE 3

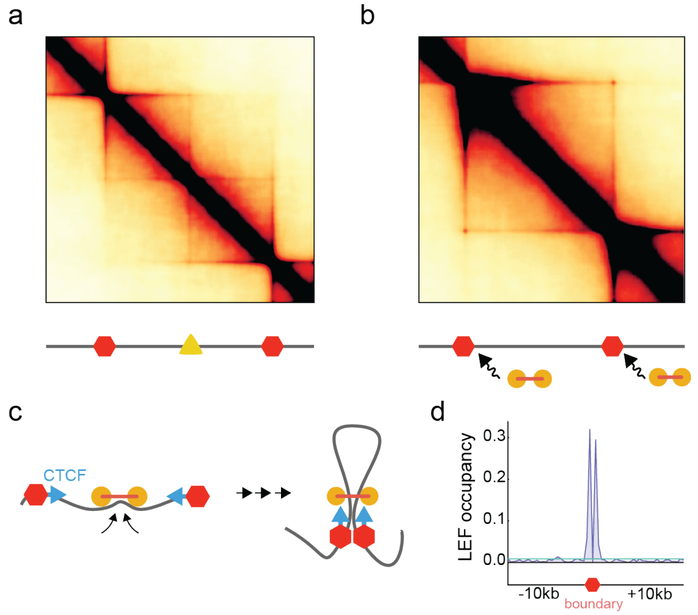

Fig 3. Additional consequences of loop extrusion.
a. Simulated contact map for a model with a semi-permeable BE (yellow triangle), displaying nested domains.
b. Simulated contact map for a model with asymmetric loading of LEFs.
c. Loop extrusion plus an orientation-dependent boundary function of CTCF can lead to enrichment of inward-oriented CTCF sites at domain boundaries, even over large genomic distances (**fig. S12**).
d. Accumulation of LEFs at BEs for processivity 120kb, separation 120kb.

References

1. J. R. Dixon *et al.*, *Nature*. **485**, 376–380 (2012).
2. E. P. Nora *et al.*, *Nature*. **485**, 381–385 (2012).
3. S. S. P. Rao *et al.*, *Cell*. **159**, 1665–1680 (2014).
4. T. Sexton *et al.*, *Cell*. **148**, 458–72 (2012).
5. G. Andrey *et al.*, *Science*. **340**, 1234167 (2013).
6. O. Symmons *et al.*, *Genome Res.* **24**, 390–400 (2014).
7. D. G. Lupiáñez *et al.*, *Cell*. **161**, 1012–1025 (2015).
8. E. Lieberman-Aiden *et al.*, *Science*. **326**, 289–93 (2009).
9. M. Imakaev *et al.*, *Nat. Methods*. **9**, 999–1003 (2012).
10. E. Lieberman-Aiden *et al.*, *Science*. **326**, 289–293 (2009).
11. L. Giorgetti *et al.*, *Cell*. **157**, 950–963 (2014).
12. D. Bau *et al.*, *Nat. Struct. Mol. Biol.* **18**, 107–114 (2011).
13. D. Jost, P. Carrivain, G. Cavalli, C. Vaillant, *Nucleic Acids Res.*, 1–9 (2014).
14. M. Barbieri *et al.*, *Proc. Natl. Acad. Sci.* **109**, 16173–16178 (2012).
15. F. Benedetti, J. Dorier, Y. Burnier, A. Stasiak, *Nucleic Acids Res*. **42**, 2848–2855 (2013).
16. M. V. Rudan *et al.*, *Cell Rep.* **10**, 1297–1309 (2015).
17. E. Alipour, J. F. Marko, *Nucleic Acids Res.* **40**, 11202–12 (2012).
18. M. Bohn, D. W. Heermann, *PLoS One*. **5**, e12218 (2010).
19. V. F. Scolari, M. C. Lagomarsino, *Soft Matter*. **11**, 1677–1687 (2015).
20. N. Naumova *et al.*, *Science*. **342**, 948–53 (2013).
21. P. Eastman *et al.*, *J Chem Theory Comput*. **9**, 461–469 (2013).
22. T. B. K. Le, M. V Imakaev, L. a Mirny, M. T. Laub, *Science*. **342**, 731–4 (2013).
23. A. Rosa, N. B. Becker, R. Everaers, *Biophys. J.* **98**, 2410–2419 (2010).
24. B. Doyle, G. Fudenberg, M. Imakaev, L. A. Mirny, *PLoS Comput. Biol.* **10**, e1003867 (2014).
25. K. Nasmyth, *Annu. Rev. Genet.* **35**, 673–745 (2001).
26. S. Sofueva *et al.*, *EMBO J.* **32**, 3119–3129 (2013).
27. J. Zuin *et al.*, *Proc. Natl. Acad. Sci. U. S. A.* **111**, 996–1001 (2014).
28. T. Mizuguchi *et al.*, *Nature*. **516**, 432–435 (2014).
29. M. H. Kagey *et al.*, *Nature*. **467**, 430–435 (2010).
30. D. Gerlich, B. Koch, F. Dupeux, J.-M. Peters, J. Ellenberg, *Curr. Biol.* **16**, 1571–1578 (2006).
31. A. Tedeschi *et al.*, *Nature*. **501**, 564–8 (2013).
32. C.-T. Ong, V. G. Corces, *Nat. Rev. Genet.* **15**, 234–46 (2014).
33. C. Hou, H. Zhao, K. Tanimoto, A. Dean, *Proc. Natl. Acad. Sci. U. S. A.* **105**, 20398–20403 (2008).
34. H. Nakahashi *et al.*, *Cell Rep.* **3**, 1678–1689 (2013).
35. V. Parelho *et al.*, *Cell*. **132**, 422–433 (2008).
36. T. Xiao, J. Wallace, G. Felsenfeld, *Mol. Cell. Biol.* **31**, 2174–2183 (2011).
37. Z. Ouyang *et al.*, *Proc. Natl. Acad. Sci. U. S. A.* **110**, 11355–11360 (2013).
38. K. Hara *et al.*, *Nat. Struct. Mol. Biol.* **21**, 864–870 (2014).
39. T. Nishiyama *et al.*, *Cell*. **143**, 737–749 (2010).
40. L. D. Nolen, S. Boyle, M. Ansari, E. Pritchard, W. A. Bickmore, *Hum. Mol. Genet.* **22**, 4180–4193 (2013).
41. S. Kadauke, G. A. Blobel, *Epigenetics Chromatin*. **6**, 6 (2013).