Published in final edited form as:

Trends Biochem Sci. 2018 June ; 43(6): 469–478. doi:10.1016/j.tibs.2018.03.006.

Principles of Chromosome Architecture Revealed by Hi-C

Kyle P. Eagen^{1,*}

¹Department of Biochemistry and Molecular Genetics, Northwestern University Feinberg School of Medicine, Chicago, IL 60611, USA

Abstract

Chromosomes are folded and compacted within interphase nuclei, but the molecular basis of this folding is poorly understood. Chromosome conformation capture methods, such as Hi-C, combine chemical cross-linking of chromatin followed by fragmentation, DNA ligation, and high-throughput DNA sequencing to detect neighboring loci genome-wide. Hi-C has revealed the segregation of chromatin into active and inactive compartments, and the folding of DNA into self-associating domains and loops. Depleting CTCF, cohesin, or cohesin-associated proteins was recently shown to affect the majority of domains and loops in a manner that is consistent with a model of DNA folding through extrusion of chromatin loops. Compartmentation was not dependent on CTCF or cohesin. Hi-C contact maps represent the superimposition of CTCF/cohesin-dependent and -independent folding states.

Keywords

Chromosome structure; chromosome conformation capture; Hi-C; nuclear compartments; topologically associating domains (TADs); DNA loops

Chromatin Folding Paradigms Revealed by Chromosome Conformation Capture

The folding and compaction of DNA within nuclei and chromosomes has fascinated biologists for over a century. At the molecular level, wrapping of DNA around an octameric core of histones to form a nucleosome [1,2] results in so-called “beads-on-a-string”, fully extended, 11 nm chromatin fibers. At a cytological level, chromosomes are contained within interphase nuclei 5–10 μ m in diameter, or individualized into chromatids hundreds of nanometers in diameter during mitosis. For several decades, bridging the three orders of magnitude gap in folding between the molecular and cytological levels remained a mystery.

A combination of molecular biology with advances in DNA sequence analysis led to the development of the “chromosome conformation capture” [3] family of technologies that

*Correspondence: eagen@northwestern.edu (K.P. Eagen).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

assess DNA folding. Hi-C [4], which combines chemical cross-linking of chromatin, followed by restriction enzyme digestion, proximity ligation, and next-generation, high-throughput DNA sequencing has revealed three chromatin folding paradigms (Figure 1). At a scale of entire nuclei and whole chromosomes, loci segregate into two main **compartments** (see Glossary) [4] and at least six **subcompartments** [5]. From lengths of several to hundreds of kilobases, two features are observed, **loops** [5] and **topologically associating domains** or **TADs** [6–8].

Chromosome Compartmentation

A plaid or checkboard pattern in mammalian Hi-C contact maps was the first characterized folding paradigm and was observed for both intra- and interchromosomal contact maps [4]. This pattern revealed a set of loci that are more likely to interact with one another than expected for the random polymer conformation of a chromosome. This set of loci has a long-range contact pattern that is distinct from a second set of loci that also interact among themselves. Computational correlation and principal component analysis classified each genomic locus into either the A or B compartment at one megabase resolution. Comparing the compartment assignment with other biochemical features of chromatin identified the A compartment as more accessible to DNase I, more gene rich, and more transcriptionally active than the B compartment. Compartment A was also more enriched for the active histone H3K36me3 modification [4]. On this basis, the A and B compartments detected by Hi-C reflect the folding of chromosomes into euchromatin and heterochromatin, respectively.

Improvements in Hi-C contact map resolution, through more efficient sample preparation and more extensive DNA sequencing, revealed that correlation and principal component analysis could not completely explain the long-range contact pattern of high resolution HiC contact maps [5]. Instead of two sets, five computationally identified plus one manually annotated set of loci with shared interaction patterns were observed [5]. Two of these six sets had interaction patterns that correlated with each other and were enriched for histone modifications that are characteristic of euchromatin. On this basis, these two sets corresponded to the initially identified A compartment and were renamed subcompartments A1 and A2. The other four sets had interaction patterns that correlated more strongly with each other than with A1 or A2 and were enriched for features characteristic of heterochromatin. These corresponded to the initially identified B compartment and consequently reclassified as subcompartments B1, B2, B3, and B4 [5].

The six identified subcompartments likely represent a lower bound on the true number of long-range contact patterns. Computational identification of subcompartments relied on analysis of interchromosomal contact maps [5], which have far fewer contacts than intrachromosomal maps and therefore need to be analyzed at lower resolution for accurate annotation. Since the number of contact patterns increased from two to six as resolution increased from 1 Mb to 100 kb, it may be possible to identify even more patterns at resolutions greater than 100 kb. This was recently suggested to be the case when the same data used to identify subcompartments at 100 kb resolution was reanalyzed at 5 kb resolution [9]. Changes in the contact pattern could be observed within subcompartments

[9], but whether these changes reflect the presence of additional patterns or a better ability to define boundaries at higher resolutions remains to be determined.

Rather than representing hierarchical structures, as the names compartments and subcompartments imply, the ability to classify loci based on their long-range contact pattern represents the principle of chromosome compartmentation: that sets of loci interact with themselves more frequently than expected and that these loci segregate from other sets of loci that also interact among themselves. The capability to more finely classify loci simply reflects a better ability to annotate the long-range contact pattern with increased resolution. What emerges is not a new or hierarchical folding paradigm, but an improved annotation of contact patterns.

An important correlate of annotating long-range contact patterns is that by their very nature compartments and subcompartments are defined in genomic coordinates. Transferring these genomic coordinates to a spatial understanding of compartmentation remains challenging. *Drosophila melanogaster* **polytene chromosomes** were shown to lack compartments [10], consistent with imaging data that these chromosomes do not regularly fold back on themselves [11]. This implies that compartmentation requires chromosomes to be flexible enough to make regular interactions [10], but does not indicate how distant loci with increased frequency of interactions are spatially arranged. Imaging many genomic loci along individual human chromosomes suggested a polarized arrangement of A and B compartments within chromosomes [12], but the fluorescence *in situ* hybridization (FISH) probes used for this analysis relied on Hi-C data that could not discern the subcompartment contact pattern. Compartmentation is also reflected in interchromosomal contact patterns [4,5], and it remains to be determined if a polarized arrangement of compartments applies to interactions between chromosomes.

Self-associating Chromatin Domains

Compartmentation refers to the off-diagonal boxes in Hi-C contact maps, but there are also on-diagonal boxes of enriched contact frequency. These are often referred to as topologically associating domains or TADs. On-diagonal boxes have two-fold significance. First, the elevated contact frequency implies that a locus within a TAD is likely to contact any other locus within the same TAD, or that chromatin within a TAD is self-associating. Second, this elevated contact frequency is not continuous, but instead is divided into noticeable blocks, indicating that self-association is also not continuous but is broken into distinct domains. Therefore, TADs have boundaries and chromatin within one TAD is insulated from, or less likely to interact with, chromatin from another TAD. The significance of TADs was pointed out nearly simultaneously in flies, mice, and humans [6–8], suggesting that self-association is an evolutionarily conserved and general principle of chromosome folding.

Compartment off-diagonal boxes often align with TADs indicating that the two are related. The boundaries of TADs frequently define the rows and columns of the plaid pattern of Hi-C contact maps. Yet, the two have been observed to be separable. Both *Drosophila* polytene chromosomes [10] and mouse maternal zygotic chromatin [13,14] have TADs, but lack

compartments. This supports the concept that compartmentation arises from TAD-TAD interactions [10].

Like compartments and subcompartments, the ability to annotate TADs depends on sample quality and sequencing depth. As the resolution of the contact map improves, additional TAD boundaries can be recognized, and, therefore, more and apparently smaller TADs are identified [5,15,16]. Analogous to the relationship between compartments and subcompartments, the ability to better detect boundaries with more data does not necessarily indicate hierarchy in chromosome organization or that newly identified boundaries point to new concepts in chromatin folding. More simply, TADs are better annotated as sample quality and sequencing depth increases.

Another analogy between compartmentation and self-association is that like compartments and subcompartments, TADs are defined in genomic coordinates. A physical understanding of TADs came from Hi-C analysis of *Drosophila* polytene chromosomes [10]. Polytene chromosome bands are equivalent to TADs, indicating that TADs represent a folded state so dense they can be seen by light microscopy. Furthermore, the location of TADs is conserved between polytene tissue, cultured Kc167 cells, and *Drosophila* embryos indicating that the polytene organization generally reflects interphase chromosome organization.

TADs are observed across metazoans [6–8] and their connection to polytene chromosome bands [10,17,18] indicates that one principle of interphase chromosome organization is an evolutionarily conserved axial organization. A synonym for polytene bands is **chromomeres**, which are serially aligned beads or granules of chromatin resulting from local coiling of DNA [19,20]. Super-resolution microscopy revealed that a series of TADs reflects a succession of compact, physical structures in single *Drosophila* cells [21]. Therefore, TADs are also synonymous with chromomeres. Most frequently studied in model systems, such as lampbrush and polytene chromosomes, Hi-C analysis has now extended the chromomere concept to most metazoan genomes, including humans, and allows for the sequence within each chromomere to be determined.

Chromatin Loops

In addition to boxes of enriched contact frequency, focal peaks of contact enrichment can also be observed in Hi-C contact maps [5]. These focal peaks are often radially symmetric, indicating that they are distinct from TADs or the off-diagonal boxes of compartmentation. A peak represents two loci that are in closer proximity to one another than neighboring loci and therefore represent chromatin loops. There are ~10,000 loops in the human genome with some loop boundaries, also referred to as loop anchors, shared between two or more loops. FISH studies indicated that loops are apparently transient as loop anchors are greater than 250 nm apart in three-quarters of a cell population [5].

In humans, greater than 85% of loop anchors are bound by cohesin or CTCF [5]. Cohesin is thought to form a tri-partite ring [22,23], suggesting that it could establish or maintain chromatin loops by entrapping two segments of chromatin in *cis*. CTCF was identified for its ability to bind insulators and prevent enhancers from communicating with promoters [24].

CTCF has a characteristic binding motif that is nonpalindromic [25]. Of the loops where both anchors can be assigned to a unique CTCF motif, approximately 90% of the motifs point towards each other [5]. This has been referred to as the convergent rule [26].

A predominance of convergent CTCF motif orientations at loop anchors is not expected by chance. If CTCF is involved in looping independent of the orientation of its motif, a quarter of the motif pairs would be expected to be convergently oriented, a quarter divergently oriented, and half tandemly oriented. The enrichment for convergently oriented motifs suggests a biochemical mechanism for loop formation in contrast to looping via diffusion.

Loop Extrusion Hypothesis

A distinct, but related problem in chromosome biology is how chromosomes establish their thread-like structure during cell division. Structural maintenance of chromosome (SMC) complexes, including cohesin and condensin, are a family of multiprotein enzymes identified for their role in shaping chromosome architecture during mitosis and meiosis [27]. Cohesin was initially discovered for its role in holding together sister chromatids, a process known as sister chromatid cohesion [28–30]. The condensin complex was identified due to its requirement for proper chromosome morphology during mitotic chromosome condensation in *Xenopus laevis* egg extracts [31].

Related to how thread-like chromatids form during cell division is how chromatids remain individualized and maintain their identity [32]. Nasmyth proposed that condensin might help resolve sister chromatids if it acted as a processive enzyme that enlarges loops of DNA [32]. This process would separate one DNA molecule from another, and, as neighboring loops converge, explain why condensin is found along the axis of mitotic chromatids. This conceptual model could be recapitulated mathematically and is referred to as loop extrusion [33].

Polymer simulations have applied the concept of loop extrusion to the loops and TADs observed by Hi-C [26,34]. If cohesin or cohesin-containing complexes stochastically bound to a stretch of DNA and extruded chromatin until the complex reached a CTCF protein, two CTCFs flanking the complex could each act as boundary elements to stop or stall the complex, thereby forming a loop (Figure 2). Simulated contact maps from these models could recapitulate the box and peak signals of TADs and loops, respectively, from experimental Hi-C contact maps. If an additional constraint is added such that the complex stops or stalls when it encounters CTCF bound to its motif in an orientation that points towards the complex, then the locations of computationally predicted loops agree with those observed experimentally. This provides a potential explanation for the enrichment of convergently oriented CTCF motifs at loop anchors. The convergent rule was tested by CRISPR/Cas9-mediated editing of CTCF motifs at loop anchors, and in each case where the motif was mutated the loop was no longer observed [26,35,36].

Editing CTCF motifs tests the convergent rule and loop extrusion hypothesis with regard to DNA sequence at a select number of loci, but not the proteins or macromolecular complexes that act on chromatin genome-wide. Tests of the loop extrusion hypothesis in this regard

came from Hi-C analysis of cells deficient for CTCF or cohesin. Auxin-mediated, acute protein degradation of CTCF [37,38] or cohesin [38,39] altered chromatin looping. Degradation of CTCF led to a decrease in aggregate Hi-C contact intensity in the vicinity of CTCF- and cohesin-bound loops identified in wild-type cells [37]. A similar result was observed after genetic removal of cohesin followed by aggregation of single-nucleus Hi-C contact intensity in the vicinity of loops identified in bulk Hi-C experiments [13]. Degradation of CTCF or cohesin resulted in a genome-wide loss of loops at individual loci [38,39]. Removal of CTCF resulted in a substantial loss of insulation between many neighboring TADs [37,38]. Many TADs were also lost upon removal of cohesin [38,39]. Loops and TADs were reestablished after restoration of CTCF [37] or cohesin [39]. These results contrasted with earlier reports on the roles of CTCF and cohesin, likely due to differences in the degree of cohesin or CTCF depletion or due to differences in the analysis, such as resolution, that may impact the ability to discern TAD boundaries [37,39–42]. Together, these results identify important roles for cohesin and CTCF in determining the location of chromatin loops and many TADs.

Removal or restoration of CTCF and cohesin identifies key proteins that organize chromosomes into TADs and loops, but the loop extrusion model proposes an active mechanism of chromatin folding that is not fully tested by protein removal or restoration. Since cohesin is proposed to be a minimal subunit of the loop extrusion complex [26,34], altering how long cohesin remains on chromatin is one way to further test the loop extrusion model. Cohesin's residency time on chromatin is affected by proteins that promote loading or unloading of cohesin. The SCC2/SCC4 complex (also known as NIPBL/MAU2) loads cohesin onto chromatin [43], whereas WAPL promotes its release [44,45].

The loop extrusion model predicts that if cohesin isn't efficiently loaded onto chromatin, loops should be less pronounced. Conditional genetic deletion of *Scs2/Nipbl* in mouse liver followed by aggregation of Hi-C contact intensity in the vicinity of loops identified in bulk Hi-C experiments from wild-type cells indicated a loss of loops and fewer TADs [46]. *Scs4/Mau2* was genetically deleted in haploid cultured human cells and examination of the Hi-C contact maps at individual loci instead of in aggregate revealed more subtle changes [47]. TADs and loops did not completely disappear, but became less apparent. Large loops were affected to a greater extent than smaller loops. Differences between deletion of *Scs2* versus *Scs4* could reflect the different experimental systems or differences in the computational analysis. *Scs2* and *Scs4* are essential for cohesin loading and viability [29,43] in budding yeast, but not in haploid cultured human cells [47]. *In vitro*, fission yeast cohesin can inefficiently be loaded onto chromatin in the absence of its loader complex, Mis4^{Scs2}-Ssl3^{Scs4} [48]. These results suggest that if less cohesin is loaded onto chromatin then small, but not large, loops can form by extrusion, and it is harder to detect TADs in Hi-C contact maps.

The loop extrusion model also predicts that blocking cohesin's release from chromatin should stabilize loops. Conditional genetic deletion of WAPL in mouse zygotes [13] followed by aggregation of single-nucleus Hi-C contact intensity in the vicinity of loops identified in bulk Hi-C experiments from wild-type zygotes indicated, on average, stronger signal at loops and TADs, with longer loops being more affected than shorter loops [13]. In

human cells, WAPL levels were altered by RNAi depletion [38] or genetic deletion [47] followed by Hi-C analysis both in aggregate and at individual loci. After removing WAPL, individual loops observed in wild-type cells increased in size [38,47]. Genetic deletion of WAPL led to loops being more apparent between the boundaries of large TADs and a mild decrease in TAD boundary strength [47]. RNAi depletion of WAPL resulted in larger TADs, with a more significant decrease in TAD boundary strength compared to WAPL deletion [38]. Differences with respect to TADs may be due to differences in the degree of WAPL depletion, differences in the computational analysis [38], or differences in Hi-C contact map resolution that may impact the ability to discern TAD boundaries. With respect to loop extrusion, if cohesin is less likely to be released from chromatin, smaller loops as well as many TADs become stronger and longer loops can form.

Increasing cohesin's residency time on chromatin by depleting WAPL would also decrease the likelihood that cohesin would stop or stall at a properly oriented CTCF motif. On this basis, another test of the loop extrusion hypothesis is to observe the orientation of CTCF motifs at loop anchors after WAPL depletion. When WAPL was removed from cells, fewer convergent CTCF sites were observed between loop anchors [38,47]. In the absence of WAPL, when cohesin reaches CTCF bound to its motif in the proper orientation, cohesin may bypass this CTCF, creating a longer loop and eventually stopping, stalling, or disengaging from chromatin at an improperly oriented CTCF motif further along the chromatin fiber [38,47].

Besides WAPL, PDS5A and PDS5B also interact with and inhibit cohesin's release from chromatin, yet simultaneous RNAi depletion of PDS5A and PDS5B affected Hi-C results differently than depletion of WAPL [38]. RNAi depletion of both PDS5A and PDS5B resulted in fewer loops than wild-type cells and fewer convergent CTCF sites between loop anchors [38]. Understanding these differences will come from further elucidating the functions of PDS5A and PDS5B. In the context of loop extrusion, PDS5A and PDS5B might promote release or stalling of cohesin when it encounters CTCF bound to its motif in the proper orientation. Fewer loops observed after co-depletion of PDS5A and PDS5B may reflect a blurring of the cohesin-dependent Hi-C signal and apparently fewer loops [38].

Many lines of evidence from a number of independent groups support the loop extrusion model and indicate that CTCF and cohesin are important factors for folding many TADs and loops. These experiments also revealed aspects of chromosome organization that do not depend on CTCF or cohesin.

Cohesin-independent Chromatin Folding

In each case when CTCF or cohesin was removed or depleted from chromatin, compartmentation could still be detected in Hi-C contact maps [13,37–39,46,47]. Compared to wild-type Hi-C contact maps, compartmentation was more apparent after cohesin removal [13,38,39,46,47]. Compartments, therefore, are independent of CTCF and cohesin. If cohesin-dependent chromatin folding arises because of loop extrusion, then compartmentalization must reflect a cohesin-independent mechanism of chromatin folding [39,46].

Off-diagonal boxes characteristic of compartmentation imply the presence of on-diagonal boxes. Off-diagonal boxes have boundaries, and, because chromosomes are polymers, these off-diagonal boundaries must align with on-diagonal boundaries. On-diagonal boxes are typically obscured when viewing whole-chromosome Hi-C contact maps because the color scale required to show weak, long-range contacts results in an essentially saturated signal close to the diagonal. Since TADs are on-diagonal boxes, a consequence of compartmentation in the absence of CTCF and cohesin is that some TADs are independent of CTCF and cohesin. Close inspection of contact maps reveals that many TADs can still be detected, by eye or computationally, after removal of CTCF or cohesin from chromatin [37,39,46]. Importantly, the presence of on-diagonal boxes does not imply off-diagonal boxes, as indicated by polytene chromosomes that have TADs but not compartments[10].

TADs, initially described as a single feature – on-diagonal boxes of enriched contact frequency – may reflect two different phenomena: cohesin-dependent and cohesin-independent chromatin folding. In the context of loop extrusion, cohesin-dependent or loop TADs don't reflect self-association, but rather the average result of performing Hi-C on millions of cells. Loop TADs reflect many loops forming and breaking across a multitude of cells, with focal peaks representing more stable, yet still transient loops.

What are cohesin-independent or compartment TADs? In mammalian cells it is presently unclear, but studies from other organisms shed light on their physical nature. In *Drosophila*, TADs strongly correlate with the same histone modification and non-histone protein localization patterns that correlate with compartments and subcompartments [8,9,17,21,49]. For example, many TADs are enriched in their entirety for histone H3K27me3 and bound by Polycomb group proteins [8,9]. Polycomb-repressed TADs also form condensed structures in single *Drosophila* cells [21]. In the mouse, the formation of developmentally-regulated TADs depends on polycomb repressive complex 1, but does not correlate with CTCF or cohesin occupancy [50]. This suggests that polycomb-repressed chromatin may reflect one class of evolutionarily conserved compartment TADs.

Sub-kilobase resolution Hi-C analysis in *Drosophila* has also provided insights about the properties of compartment TADs. Although these studies reported focal peaks indicative of looping in the fly [15,16], these loops are qualitatively different from mammalian loops. Loops in the fly do not typically form at the corners of TADs, but rather within TADs [15,16] suggesting that *Drosophila* TADs are unrelated to looping. In flies and plants [51], all TADs are compartment TADs. Computational modeling of Hi-C data from *Drosophila*, plants, and mammals suggested that compartmentation is evolutionarily conserved, but mammalian Hi-C maps are best described by compartmentation plus CTCF/cohesin-dependent looping [9]. That *Drosophila* polytene chromosome bands are stable in the absence of cohesin [52], that polytene TADs are equivalent to polytene bands and diploid TADs [10], and that TADs in single *Drosophila* cells are physical, compact structures [21] suggest that mammalian compartment TADs reflect chromomeres.

Concluding Remarks: A Superimposition of Folding States

Hi-C contact maps from populations of wild-type mammalian cells reflect the superimposition of chromatin looping, self-association, and compartmentation. Looping during interphase may be due to chromatin extrusion, but evidence for a cohesin-associated complex with this biochemical activity is lacking (see Outstanding Questions). Single-molecule assays [53–58] will be essential in this regard. For example, condensin's ability to translocate along [54] and extrude loops of DNA [58] was recently determined using single-molecule imaging. Combined with Hi-C, microscopic, and computational analysis of chromosome structure as cells enter mitosis [59], these results agree with Nasmyth's model of mitotic chromatid formation by condensin-dependent loop extrusion [32]. Identifying if cohesin has this capability, or if condensin has the same activity outside of mitosis is important for substantiating the loop extrusion model during interphase.

Outstanding Questions Box

- Can a cohesin-associated complex with loop extrusion activity similar to that of condensin be biochemically isolated *in vitro*?
- What molecular interactions are needed for CTCF, only when bound to its motif in a “proper” orientation, to stabilize loops?
- What is the origin of the superimposition of cohesin-dependent and -independent folding states observed in Hi-C contact maps?
- What are the biochemical forces that segregate genomic intervals into different compartments and subcompartments?
- How are the boundaries of compartment TADs formed?
- What is the physical nature of intra-domain contacts within compartment TADs?
- How does looping, self-association, and compartmentation affect biological function?

If population-based Hi-C data reflects the superimposition of folding states, what is the cause of this convolution? That bulk Hi-C cannot distinguish between single-cells may be an oversimplified answer. Chromosomes undergo large conformational changes as cells divide [59,60]. Keeping the cell cycle in mind will be essential for interpreting Hi-C data [61]. For instance, what is the consequence of loop extrusion in G2, where sister chromatid cohesion must be maintained while intra-chromatid loops are forming and breaking? Single-cell or single-nucleus Hi-C assays [14,61] will be beneficial as their resolution improves, keeping in mind that datasets from single-cell genomics methods are inherently sparse.

Model systems, such as *Drosophila*, where Hi-C–detected loops are qualitatively different from mammalian loops will be important for elucidating the molecular basis of self-associating compartment TADs and compartmentation. This will first require cohesin removal experiments analogous to those in mammals to firmly establish that the majority of

chromatin folding in these organisms is cohesin-independent. The mechanism of defining compartment TAD boundaries can then be investigated as well as the enigmatic physical nature of contacts within these domains.

What are the functional consequences of looping, self-association, and compartmentation? It is already unclear how loop extrusion during interphase results in a biological outcome. Acute removal of cohesin led to essentially no changes in gene expression [39], whereas genetic deletion [46,62] led to stronger, but possibly indirect, changes in a greater number of genes. Studying human disease [63–66] may aid in revealing the underlying basic biology, but resolving these differences and uncovering the function of each chromatin folding paradigm may require viewing chromatin beyond its role in gene regulation. Faithful segregation of chromosomes, in addition to repair of damaged DNA, is essential for properly transmitting genes from mother to daughter cell. Studying chromatin through the lens of cellular heredity may reveal important structure-function relationships.

Acknowledgments

K.P.E. is grateful to Haneen Ammouri and Celeste Rosencrance for helpful reading of this manuscript and apologizes to those whose work could not be cited due to space limitations. Research in the Eagen Lab is supported by an NIH Director's Early Independence Award DP5OD024587.

Glossary

Chromomere

Serial aligned beads or granules of chromosomes resulting from local coiling of the DNA. Separated from each other by interchromomeres, which are less dense. Chromomeres and interchromomeres are irregularly spaced along a chromosome.

Compartments or subcompartments

Sets of chromosomal regions with similar, long-range Hi-C contact patterns that occur more frequently than expected based upon the random polymer conformation of the chromatin fiber.

Compartmentation

The segregation of chromatin into two compartments or at least six subcompartments that is reflected by the long-range contact pattern of off-diagonal boxes of alternating enriched or depleted contacts in Hi-C contact maps. Often referred to as a “checkerboard” or “plaid” pattern.

Loop

A pair of loci in close physical proximity. For pairs that interact more strongly than their local neighbors, this is indicated by a radially symmetric focal peak of contact enrichment in Hi-C contact maps.

Polytene chromosomes

Giant chromosomes that result from many rounds of DNA replication without cell division while retaining pairing of homologous chromosomes and alignment of sister chromatids. This results in individual chromosomes that can easily be seen by light microscopy and

display an alternating pattern of dense bands, which represent condensed chromatin, separated by less dense, loosely folded interbands. Most often studied from the salivary glands of *Drosophila melanogaster*.

Topologically associating domains (TADs)

Boxes of enriched contact frequency that tile the diagonal in Hi-C contact maps. Also referred to as A/B domains [67], physical domains [8], topological domains [6], or contact domains [5].

References

1. Kornberg RD. Chromatin structure: a repeating unit of histones and DNA. *Science*. 1974; 184:868–871. [PubMed: 4825889]
2. Luger K, et al. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*. 1997; 389:251–260. [PubMed: 9305837]
3. Dekker J, et al. Capturing chromosome conformation. *Science*. 2002; 295:1306–1311. [PubMed: 11847345]
4. Lieberman-Aiden E, et al. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science*. 2009; 326:289–293. [PubMed: 19815776]
5. Rao SSP, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*. 2014; 159:1665–1680. [PubMed: 25497547]
6. Dixon JR, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*. 2012; 485:376–380. [PubMed: 22495300]
7. Nora EP, et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*. 2012; 485:381–385. [PubMed: 22495304]
8. Sexton T, et al. Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell*. 2012; 148:458–472. [PubMed: 22265598]
9. Rowley MJ, et al. Evolutionarily Conserved Principles Predict 3D Chromatin Organization. *Mol Cell*. 2017; 67:837–852. [PubMed: 28826674]
10. Eagen KP, et al. Stable Chromosome Condensation Revealed by Chromosome Conformation Capture. *Cell*. 2015; 163:934–946. [PubMed: 26544940]
11. Hochstrasser M, et al. Spatial organization of chromosomes in the salivary gland nuclei of *Drosophila melanogaster*. *J Cell Biol*. 1986; 102:112–123. [PubMed: 3079766]
12. Wang S, et al. Spatial organization of chromatin domains and compartments in single chromosomes. *Science*. 2016; 353:598–602. [PubMed: 27445307]
13. Gassler J, et al. A mechanism of cohesin-dependent loop extrusion organizes zygotic genome architecture. *EMBO J*. 2017; doi: 10.15252/embj.201798083
14. Flyamer IM, et al. Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature*. 2017; 544:110–114. [PubMed: 28355183]
15. Eagen KP, et al. Polycomb-mediated chromatin loops revealed by a subkilobase-resolution chromatin interaction map. *Proceedings of the National Academy of Sciences*. 2017; 114:8764–8769.
16. Cubeñas-Potts C, et al. Different enhancer classes in *Drosophila* bind distinct architectural proteins and mediate unique chromatin interactions and 3D architecture. *Nucleic Acids Res*. 2017; 45:1714–1730. [PubMed: 27899590]
17. Ulianov SV, et al. Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. *Genome Res*. 2016; 26:70–84. [PubMed: 26518482]
18. Stadler MR, et al. Convergence of topological domain boundaries, insulators, and polytene interbands revealed by high-resolution mapping of chromatin contacts in the early *Drosophila melanogaster* embryo. *elife*. 2017; 6:637662.
19. Wilson, EB. *The cell in development and inheritance*. New York: Macmillan; 1896.

20. Beermann W. Chromomeres and genes. *Results Probl Cell Differ.* 1972; 4:1–33. [PubMed: 4198831]
21. Szabo Q, et al. TADs are 3D structural units of higher-order chromosome organization in *Drosophila*. *Sci Adv.* 2018; 4:eaar8082. [PubMed: 29503869]
22. Gligoris TG, et al. Closing the cohesin ring: structure and function of its Smc3-kleisin interface. *Science.* 2014; 346:963–967. [PubMed: 25414305]
23. Huis in 't Veld PJ, et al. Characterization of a DNA exit gate in the human cohesin ring. *Science.* 2014; 346:968–972. [PubMed: 25414306]
24. Bell AC, et al. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell.* 1999; 98:387–396. [PubMed: 10458613]
25. Kim TH, et al. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell.* 2007; 128:1231–1245. [PubMed: 17382889]
26. Sanborn AL, et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proceedings of the National Academy of Sciences.* 2015; 112:E6456–65.
27. Uhlmann F. SMC complexes: from DNA to chromosomes. *Nat Rev Mol Cell Biol.* 2016; doi: 10.1038/nrm.2016.30
28. Losada A, et al. Identification of *Xenopus* SMC protein complexes required for sister chromatid cohesion. *Genes Dev.* 1998; 12:1986–1997. [PubMed: 9649503]
29. Michaelis C, et al. Cohesins: chromosomal proteins that prevent premature separation of sister chromatids. *Cell.* 1997; 91:35–45. [PubMed: 9335333]
30. Guacci V, et al. A direct link between sister chromatid cohesion and chromosome condensation revealed through the analysis of MCD1 in *S. cerevisiae*. *Cell.* 1997; 91:47–57. [PubMed: 9335334]
31. Hirano T, Mitchison TJ. A heterodimeric coiled-coil protein required for mitotic chromosome condensation in vitro. *Cell.* 1994; 79:449–458. [PubMed: 7954811]
32. Nasmyth K. Disseminating the genome: joining, resolving, and separating sister chromatids during mitosis and meiosis. *Annu Rev Genet.* 2001; 35:673–745. [PubMed: 11700297]
33. Alipour E, Marko JF. Self-organization of domain structures by DNA-loop-extruding enzymes. *Nucleic Acids Res.* 2012; 40:11202–11212. [PubMed: 23074191]
34. Fudenberg G, et al. Formation of Chromosomal Domains by Loop Extrusion. *Cell Reports.* 2016; 15:2038–2049. [PubMed: 27210764]
35. de Wit E, et al. CTCF Binding Polarity Determines Chromatin Looping. *Mol Cell.* 2015; doi: 10.1016/j.molcel.2015.09.023
36. Guo Y, et al. CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell.* 2015; 162:900–910. [PubMed: 26276636]
37. Nora EP, et al. Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell.* 2017; 169:930–944e22. [PubMed: 28525758]
38. Wutz G, et al. Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *EMBO J.* 2017; doi: 10.15252/embj.201798004
39. Rao SSP, et al. Cohesin Loss Eliminates All Loop Domains. *Cell.* 2017; 171:305–320e24. [PubMed: 28985562]
40. Sofueva S, et al. Cohesin-mediated interactions organize chromosomal domain architecture. *EMBO J.* 2013; 32:3119–3129. [PubMed: 24185899]
41. Seitan VC, et al. Cohesin-based chromatin interactions enable regulated gene expression within preexisting architectural compartments. *Genome Res.* 2013; 23:2066–2077. [PubMed: 24002784]
42. Zuin J, et al. Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proceedings of the National Academy of Sciences.* 2014; 111:996–1001.
43. Ciosk R, et al. Cohesin's binding to chromosomes depends on a separate complex consisting of Scc2 and Scc4 proteins. *Mol Cell.* 2000; 5:243–254. [PubMed: 10882066]
44. Gandhi R, et al. Human Wapl is a cohesin-binding protein that promotes sister-chromatid resolution in mitotic prophase. *Curr Biol.* 2006; 16:2406–2417. [PubMed: 17112726]

45. Kueng S, et al. Wapl controls the dynamic association of cohesin with chromatin. *Cell*. 2006; 127:955–967. [PubMed: 17113138]
46. Schwarzer W, et al. Two independent modes of chromatin organization revealed by cohesin removal. *Nature*. 2017; 551:51–56. [PubMed: 29094699]
47. Haarhuis JHI, et al. The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. *Cell*. 2017; 169:693–707e14. [PubMed: 28475897]
48. Murayama Y, Uhlmann F. Biochemical reconstitution of topological DNA binding by the cohesin ring. *Nature*. 2014; 505:367–371. [PubMed: 24291789]
49. Hou C, et al. Gene density, transcription, and insulators contribute to the partition of the *Drosophila* genome into physical domains. *Mol Cell*. 2012; 48:471–484. [PubMed: 23041285]
50. Kundu S, et al. Polycomb Repressive Complex 1 Generates Discrete Compacted Domains that Change during Differentiation. *Mol Cell*. 2017; 65:432–446e5. [PubMed: 28157505]
51. Dong P, et al. 3D chromatin architecture of large plant genomes determined by local A/B compartments. *Molecular Plant*. 2017; 10:1497–1509. [PubMed: 29175436]
52. Eichinger CS, et al. Disengaging the Smc3/kleisin interface releases cohesin from *Drosophila* chromosomes during interphase and mitosis. *EMBO J*. 2013; 32:656–665. [PubMed: 23340528]
53. Stigler J, et al. Single-Molecule Imaging Reveals a Collapsed Conformational State for DNA-Bound Cohesin. *Cell Reports*. 2016; 15:988–998. [PubMed: 27117417]
54. Terakawa T, et al. The condensin complex is a mechanochemical motor that translocates along DNA. *Science*. 2017; 11:eaa6516.
55. Hansen AS, et al. CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *elife*. 2017; 6:e25776. [PubMed: 28467304]
56. Kanke M, et al. Cohesin acetylation and Wapl-Pds5 oppositely regulate translocation of cohesin along DNA. *EMBO J*. 2016; doi: 10.15252/embj.201695756
57. Davidson IF, et al. Rapid movement and transcriptional re-localization of human cohesin on DNA. *EMBO J*. 2016; doi: 10.15252/embj.201695402
58. Ganji M, et al. Real-time imaging of DNA loop extrusion by condensin. *Science*. 2018; doi: 10.1126/science.aar7831
59. Gibcus JH, et al. A pathway for mitotic chromosome formation. *Science*. 2018; 359:eaao6135. [PubMed: 29348367]
60. Naumova N, et al. Organization of the Mitotic Chromosome. *Science*. 2013; 342:948–953. [PubMed: 24200812]
61. Nagano T, et al. Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature*. 2017; 547:61–67. [PubMed: 28682332]
62. Viny AD, et al. Dose-dependent role of the cohesin complex in normal and malignant hematopoiesis. *J Exp Med*. 2015; 212:1819–1832. [PubMed: 26438361]
63. Lupiáñez DG, et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*. 2015; 161:1012–1025. [PubMed: 25959774]
64. Franke M, et al. Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature*. 2016; 538:265–269. [PubMed: 27706140]
65. Hnisz D, et al. Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science*. 2016; 351:1454–1458. [PubMed: 26940867]
66. Flavahan WA, et al. Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature*. 2016; 529:110–114. [PubMed: 26700815]
67. Zhang Y, et al. Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. *Cell*. 2012; 148:908–921. [PubMed: 22341456]

Trends Box

- Chromatin is folded by looping, self-association, and compartmentation.
- Many chromatin loops and TADs depend on CTCF and cohesin for their establishment and maintenance.
- A model of cohesin-dependent extrusion of DNA loops can explain many of the loops and TADs observed in Hi-C contact maps.
- Compartmentation does not depend on cohesin or CTCF.
- Hi-C contact maps from populations of cells reflect the superimposition of cohesin-dependent and -independent folding states.

Figure 1. Hi-C-Detected Chromatin Folding Paradigms

Left: Cartoon representation of Hi-C data, represented as a heatmap or Hi-C contact map.

The strength of each pixel indicates the relative, pair-wise contact probability of two loci.

TADs are on-diagonal boxes of contact enrichment. Loops are radially symmetric peaks of

contact intensity, often located at the corners of TADs in mammalian cells. Off-diagonal

boxes indicate interactions due to compartmentation. Right: Cartoon representation of DNA

folding. TADs are more tightly folded than regions between them. TADs and loops may be

either mostly transcriptionally active (grey) or inactive (black). Loops may also be more

tightly folded, but additionally have an increased likelihood of contact between their

boundaries or anchors. Compartmentation is indicated by homotypic (active-active or

inactive-inactive) TAD-TAD interactions. The *bona fide* pattern of chromatin folding is

unknown and only indicated schematically.

Figure 2. Cohesin-dependent Loop Extrusion Hypothesis

A loop extrusion complex containing cohesin (yellow ring) is loaded onto DNA and begins extruding a small loop of chromatin. Extrusion continues until cohesin dissociates from DNA or until it reaches CTCF protein bound to its motif in an orientation that points towards the complex. At this point, the complex stalls or stops on that side of the loop, but can continue extruding on the opposite side. If CTCF bound to an improperly oriented motif is encountered, the complex is likely to bypass that site. Once a second, properly oriented CTCF site is encountered, the complex stalls or stops at this second site. Stopping or stalling is likely transient, in which case loops are not stable structures. The continuous breaking and forming of loops may be apparent as loop TADs in Hi-C contact maps from populations of cells. The actual pattern of chromatin folding is unknown and only indicated schematically.