A chromatin extension model for insulator function based on comparison of high-resolution chromatin conformation capture and polytene banding maps

Michael R. Stadler[1] and Michael B. Eisen[1,2,3]

1. Department of Molecular and Cell Biology, University of California, Berkeley, CA
2. Department of Integrative Biology, University of California, Berkeley, CA
3. Howard Hughes Medical Institute, University of California, Berkeley, CA

Abstract

Insulator proteins bind to specific genomic loci and have been shown to play a role in partitioning genomes into independent domains of gene expression and chromatin structure. Despite decades of study, the mechanism by which insulators establish these domains remains elusive. Here, we use genome-wide chromatin conformation capture (Hi-C) to generate a high-resolution map of spatial interactions of chromatin from *Drosophila melanogaster* embryos. We show that from the earliest stages of development the genome is divided into distinct topologically associated domains (TADs), that we can map the boundaries between TADs to sub-kilobase resolution, and that these boundaries correspond to 500-2000 bp insulator elements. Comparing this map with a detailed assessment of the banding pattern of a region of a polytene chromosome, we show that these insulator elements correspond to low density polytene interbands that divide compacted bands, which correspond to TADs. It has been previously shown that polytene interbands have low packing ratios allowing the conversion of small genomic distances (in base pairs) into a large physical distances. We therefore suggest a simple mechanism for insulator function whereby insulators increase the physical space between adjacent domains via the unpacking and extension of intervening chromatin. This model provides an intuitive explanation for known features of insulators, including the ability to block enhancer-promoter interactions, limit the spread of heterochromatin, and organize the structural features of interphase chromosomes.

Introduction

Beginning in the late 19th century, cytological investigations of the polytene chromosomes of insect salivary glands implicated the physical structure of interphase chromosomes in their metabolic functions (1–5). Progressively more detailed optical and electron microscopic analysis of polytene chromosomes in *Drosophila melanogaster* have identified a stereotyped banding pattern with compacted, DNA-rich "bands" alternating with extended, DNA-poor "interband" regions(6–10). While much is now known about the molecular properties of these two types of chromatin (11–20), the precise molecular nature of the banding structure remains elusive.

Several methods have been developed in the past decade for the isolation and high-throughput characterization of chromosomal regions that are colocalized within nuclei (21–24), yielding genome-wide maps of chromatin structure in a number of organisms and tissues. These experiments revealed that the interphase chromatin fiber is organized into topologically-associated domains (TADs), contiguous regions of the genome that exhibit enriched three-dimensional contacts between loci within the TAD, and depleted contacts between loci in different TADs. Eagen et al. recently showed that TADs identified from genome-wide chromatin conformation capture (Hi-C) at approximately 15 kb resolution from polytene nuclei of *Drosophila melanogaster* largely correspond to the bands of polytene chromosomes, and suggested that inter-TADs are regions of decompacted chromatin containing the promoters and regulatory elements of housekeeping genes (25).

We have been interested in the establishment of interphase chromatin structure during early *Drosophila* development because of its relationship to transcriptional activation (26–28),

and have used deep sequencing of Hi-C libraries prepared with a 4-cutter restriction enzyme to generate a map of chromosome structure in cellular blastoderm embryos at sub-kilobase resolution. These data have many interesting features that we will present elsewhere, some of which have recently been discussed by (29,30). Here we focus exclusively on the relationship between our high-resolution Hi-C data, high-resolution polytene band patterns described by Vatolina et al. (31), and additional genomic data from our lab (26) and others (32–35) that shed light the molecular origins of chromosome organization.

Results and Discussion

We performed *in situ* Hi-C (22) in the early *Drosophila melanogaster* embryo using the MboI restriction enzyme (^GATC) and sequenced ~276 million productive linkages, a combination that yielded sub-kilobase resolution of chromatin architecture. From these data, we used methods similar to those employed by other groups (21,22,29,36,37) to identify regions of self-association (TADs) as well as boundary elements between those regions. Crucially, our data allowed the identification of TADs as small as a few kilobases. The boundary elements we identified between these TADs appear to correspond to classical insulator elements: regions of 500-2000 bp that are sensitive to DNAse digestion and are strongly bound by known insulator proteins. Noting that interband regions are also characterized by DNase hypersensitivity and insulator protein association, we wondered about the detailed relationship between our Hi-C TADs and boundaries and the banding structure observed in polytene chromosomes.

There is surprisingly little data associating features of polytene structure to the genome at high resolution. Vatolina et al. used exquisitely careful electron microscopy to identify the fine banding pattern of the 65 kb region between polytene bands 10A1-2 and 10B1-2, revealing that this region, which appears as a single interband under a light microscope, actually contains six faint bands and seven interbands (31). Vatolina et al. then used available molecular genomic data to infer genomic locations for these bands and interbands, and validated four of their interband assignments by FISH. We compared these inferred mappings of polytene banding patterns to our Hi-C data.

Figure 1 shows the correspondence between Vatolina et al.'s polytene map from this region and our high-resolution Hi-C data, along with measures of early embryonic DNase hypersentivity from (26) and the binding of four insulator proteins (32). Strikingly, there is a near-perfect correspondence between the assignments of Vatolina et al. and our Hi-C data: bands correspond to TADs, and interbands correspond to the insulator elements that separate the TADs. Eagen et al. previously identified a broad correspondence between polytene interbands and inter-TAD regions from their Hi-C data (25). However, our higher resolution Hi-C data reveals fine structure within these inter-TAD regions, and as a result we can now see a precise correspondence between insulators (regions of 500-2000 bp that are bound by insulator proteins and divide topological domains) and polytene interbands that has not, to our knowledge, been described previously.

A key feature that distinguishes polytene interbands from bands is their low compaction ratio: they span a larger physical distance per base pair. Our observation that insulators correspond so precisely to interbands suggest a simple and intuitive mechanism for insulator activity whereby insulators function by increasing the physical space between adjacent domains via the unpacking and extension of intervening chromatin.

Insulators have been associated with various functional properties. Principal among these, in addition to their role in organizing chromatin into domains, insulators block interactions between enhancers and promoters exclusively when inserted between them (38–44) and protect transgenes from position effect variegation and block the spread of chromatin silencing states (45–50). Without invoking complex interactions or chromosomal geometries, this chromatin extension model for insulator function can explain these defining characteristics via simple physical separation.

Materials and Methods

Oregon R strain embryos were collected, aged, fixed with 5% PFA for 28 minutes, and hand sorted to isolate early cellularized (mid stage 5) embryos. In situ Hi-C was performed on 75-500 embryos as described in Rao et al. (22) with minor modifications, using the restriction enzyme MboI. Paired-end sequencing with 100 bp per side was performed on the Illumina Hi-Seq 2500. We prepared and sequenced libraries from seven independent biological and technical replicates. Combined data for all stage 5 samples is shown in this paper.

Sequenced reads were aligned to the flybase 5_22 genome (dm3) using bowtie (51), retaining only read pairs in which both ends mapped uniquely to the genome. Resulting reads were assessed and filtered for quality similar to (22). Linkage matrices at various resolutions were then generated by binning the genome into bins of different sizes. Local directionality scores for high-resolution (500 bp) bins were generated by comparing the number of upstream and downstream linkages within a 15 kb window for each bin. Boundary elements were identified heuristically by searching for sites of transitions between upstream-bias and downstream-bias and applying thresholds to identify boundaries with varying stringency.

ChIP data for insulator proteins was downloaded from the modEncode server and is described in (32,33,35). DNase-seq data was available at the UCSC genome browser (52,53)and was generated by (26).

References

[1]   Balbiani EG "Sur la structure du noyau des cellules salivares chez les larves de Chironomus." *Zoologischer Anzeiger* 4: 637–641, 662–666, 1881.

[2]   Balbiani EG "Sur la structure intime du noyau du Loxophyllum meleagris." *Zoologischer Anzeiger* 13: 110–115, 132–136, 1890.

[3]   R. King and H. Beams, "Somatic synapsis in Chironomus, with special reference to the individuality of chromosomes." vol. 5, no. 3, pp. 577–591.

[4]   T. S. Painter, "The morphology of the third chromosome in the salivary gland of rosophila melanogaster and a new cytological map of this element." *Genetics*, 1935.

[5]   Heitz E and Bauer H, "Beweise fur die Chromosomennatur der Kernschleifen in den Knauelkernen von Bibio hortulanus." *Zeitschrift fur Zellforschung* 17: 67–82, 1933.

[6]   C. B. Bridges, "Salivary chromosome maps with a key to the banding of the chromosomes of Drosophila melanogaster," *J. Hered*., pp. 60–64, 1934.

[7]   G. Lefevre, "A photographic representation and interpretation of the polytene chromosomes of Drosophila melanogaster salivary glands.," *The Genetics and Biology of Drosophila*, vol. a. pp. 31–66, 1976.

[8]   C. Benyajati and A. Worcel, "Isolation, characterization, and structure of the folded interphase genome of Drosophila melanogaster," *Cell*, vol. 9, no. 3, pp. 393–407, 1976.

[9]   C. D. Laird and W. Y. Chooi, "Morphology of transcription units in Drosophila melanogaster," *Chromosoma*, vol. 58, no. 2, pp. 193–218, 1976.

[10]   M. Rabinowitz, "Studies on the cytology and early embryology of the egg of Drosophila melanogaster," *J. Morphol*., vol. 69, 1941.

[11]   E. S. Belyaevat and I. F. Zhimulevt, "Molecular and Cytogenetical Characterization of the 1OA1-2 Band and Adjoining," no. 1, 1994.

[12]   H. Eggert, A. Gortchakov, and H. Saumweber, "Identification of the Drosophila interband-specific protein Z4 as a DNA-binding zinc-finger protein determining chromosomal structure.," *J. Cell Sci*., vol. 117, no. Pt 18, pp. 4253–4264, 2004.

[13]   A. A. Gortchakov, H. Eggert, M. Gan, J. Mattow, I. F. Zhimulev, and H. Saumweber, "Chriz, a chromodomain protein specific for the interbands of Drosophila melanogaster

polytene chromosomes," *Chromosoma*, vol. 114, no. 1, pp. 54–66, 2005.

[14]   M. Berkaeva, S. Demakov, Y. B. Schwartz, and I. Zhimulev, "Functional analysis of Drosophila polytene chromosomes decompacted unit: The interband," *Chromosom. Res.*, vol. 17, no. 6, pp. 745–754, 2009.

[15]   S. A. Demakov, O. V. Andreenkov, M. B. Berkaeva, T. Y. Vatolina, E. I. Volkova, E. Z. Kvon, V. F. Semeshin, and I. F. Zhimulev, "Functional organization of interbands in Drosophila polytene chromosomes," *Russ. J. Genet.*, vol. 46, no. 10, pp. 1254–1256, 2010.

[16]   T. Y. Vatolina, S. A. Demakov, V. F. Semeshin, I. V. Makunin, V. N. Babenko, E. S. Belyaeva, and I. F. Zhimulev, "Identification and molecular genetic characterization of the polytene chromosome interbands in Drosophila melanogaster," *Russ. J. Genet.*, vol. 47, no. 5, pp. 521–532, 2011.

[17]   K. Zhao, C. M. Hart, and U. K. Laemmli, "Visualization of chromosomal domains with boundary element-associated factor BEAF-32," *Cell*, vol. 81, no. 6, pp. 879–889, 1995.

[18]   K. Byrd and V. G. Corces, "Visualization of chromatin domains created by the gypsy insulator of Drosophila.," *J. Cell Biol.*, vol. 162, no. 4, pp. 565–74, Aug. 2003.

[19]   C. Y. Pai, E. P. Lei, D. Ghosh, and V. G. Corces, "The centrosomal protein CP190 is a component of the gypsy chromatin insulator," *Mol. Cell*, vol. 16, no. 5, pp. 737–748, 2004.

[20]   M. K. Gilbert, Y. Y. Tan, and C. M. Hart, "The drosophila boundary element-associated factors BEAF-32A and BEAF-32B affect chromatin structure," *Genetics*, vol. 173, no. 3, pp. 1365–1375, 2006.

[21]   E. Lieberman-Aiden, N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. a Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. a Mirny, E. S. Lander, and J. Dekker, "Comprehensive mapping of long-range interactions reveals folding principles of the human genome.," *Science*, vol. 326, no. 5950, pp. 289–93, Oct. 2009.

[22]   S. S. P. Rao, M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov, J. T. Robinson, A. L. Sanborn, I. Machol, A. D. Omer, E. S. Lander, and E. L. Aiden, "A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping," *Cell*, pp. 1–16, Dec. 2014.

[23]   M. J. Fullwood, M. H. Liu, Y. F. Pan, J. Liu, H. Xu, Y. Bin Mohamed, Y. L. Orlov, S. Velkov, A. Ho, P. H. Mei, E. G. Y. Chew, P. Y. H. Huang, W.-J. Welboren, Y. Han, H. S. Ooi, P. N. Ariyaratne, V. B. Vega, Y. Luo, P. Y. Tan, P. Y. Choy, K. D. S. A. Wansa, B. Zhao, K. S. Lim, S. C. Leow, J. S. Yow, R. Joseph, H. Li, K. V Desai, J. S. Thomsen, Y. K. Lee, R. K. M. Karuturi, T. Herve, G. Bourque, H. G. Stunnenberg, X. Ruan, V. Cacheux-Rataboul, W.-K. Sung, E. T. Liu, C.-L. Wei, E. Cheung, and Y. Ruan, "An oestrogen-receptor-alpha-bound human chromatin interactome.," *Nature*, vol. 462, no. 7269, pp. 58–64, 2009.

[24]   R. A. Beagrie, A. Scialdone, M. Schueler, D. C. A. Kraemer, M. Chotalia, S. Q. Xie, M. Barbieri, I. de Santiago, L.-M. Lavitas, M. R. Branco, J. Fraser, J. Dostie, L. Game, N. Dillon, P. A. W. Edwards, M. Nicodemi, and A. Pombo, "Complex multi-enhancer contacts captured by genome architecture mapping," *Nature*, 2017.

[25]   K. P. Eagen, T. A. Hartl, and R. D. Kornberg, "Stable Chromosome Condensation Revealed by Chromosome Conformation Capture," *Cell*, vol. 163, no. 4, pp. 934–946, 2015.

[26]   X.-Y. Li, S. Thomas, P. J. Sabo, M. B. Eisen, J. a Stamatoyannopoulos, and M. D. Biggin, "The role of chromatin accessibility in directing the widespread, overlapping patterns of Drosophila transcription factor binding.," *Genome Biol.*, vol. 12, no. 4, p. R34, Jan. 2011.

[27]   M. M. Harrison, X.-Y. Li, T. Kaplan, M. R. Botchan, and M. B. Eisen, "Zelda binding in the early Drosophila melanogaster embryo marks regions subsequently activated at the
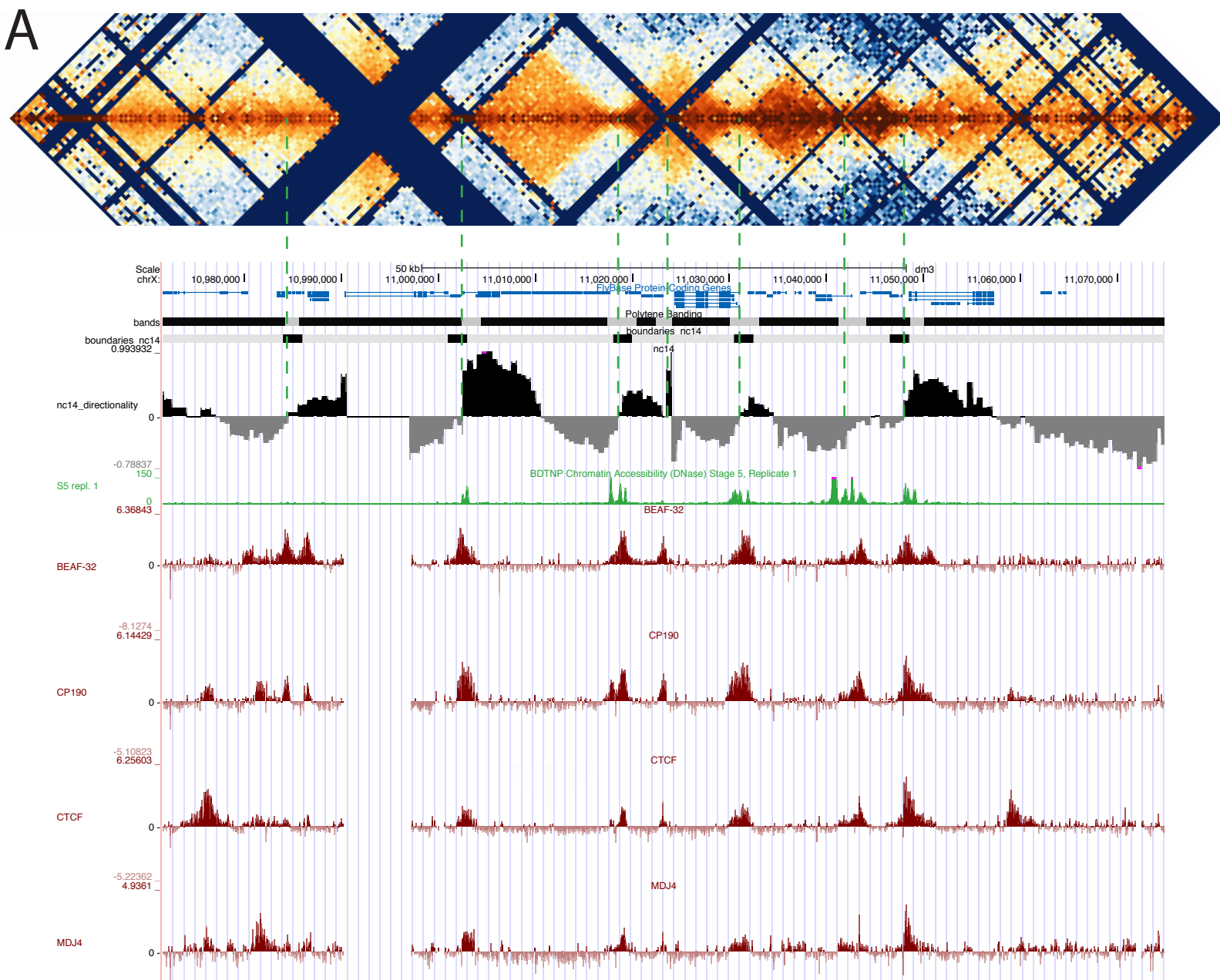
maternal-to-zygotic transition.," *PLoS Genet.*, vol. 7, no. 10, p. e1002266, Oct. 2011.

[28]   X.-Y. Li, M. M. Harrison, J. E. Villalta, T. Kaplan, and M. B. Eisen, "Establishment of regions of genomic activity during the Drosophila maternal to zygotic transition.," *Elife*, vol. 3, p. e03737, 2014.

[29]   T. Sexton, E. Yaffe, E. Kenigsberg, F. Bantignies, B. Leblanc, M. Hoichman, H. Parrinello, A. Tanay, and G. Cavalli, "Three-dimensional folding and functional organization principles of the Drosophila genome.," *Cell*, vol. 148, no. 3, pp. 458–72, Feb. 2012.

[30]   C. B. Hug, A. G. Grimaldi, K. Kruse, and J. M. Vaquerizas, "Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription," *Cell*, vol. 169, no. 2, p. 216–228.e19, 2017.

[31]   T. Y. Vatolina, L. V. Boldyreva, O. V. Demakova, S. A. Demakov, E. B. Kokoza, V. F. Semeshin, V. N. Babenko, F. P. Goncharov, E. S. Belyaeva, and I. F. Zhimulev, "Identical functional organization of nonpolytene and polytene chromosomes in Drosophila melanogaster," *PLoS One*, vol. 6, no. 10, 2011.

[32]   N. Nègre, C. D. Brown, P. K. Shah, P. Kheradpour, C. a Morrison, J. G. Henikoff, X. Feng, K. Ahmad, S. Russell, R. a H. White, L. Stein, S. Henikoff, M. Kellis, and K. P. White, "A comprehensive map of insulator elements for the Drosophila genome.," *PLoS Genet.*, vol. 6, no. 1, p. e1000814, Jan. 2010.

[33]   S. E. Celniker, L. A. L Dillon, M. B. Gerstein, K. C. Gunsalus, S. Henikoff, G. H. Karpen, M. Kellis, E. C. Lai, and J. D. Lieb, "Unlocking the secrets of the genome," *Nature. June*, vol. 18, no. 4597249, pp. 927–930, 2009.

[34]   S. M. Gallo, D. T. Gerrard, D. Miner, M. Simich, B. Des Soye, C. M. Bergman, and M. S. Halfon, "REDfly v3.0: toward a comprehensive database of transcriptional regulatory elements in Drosophila.," *Nucleic Acids Res.*, vol. 39, no. Database issue, pp. D118-23, Jan. 2011.

[35]   S. Contrino, R. N. Smith, D. Butano, A. Carr, F. Hu, R. Lyne, K. Rutherford, A. Kalderimis, J. Sullivan, S. Carbon, E. T. Kephart, P. Lloyd, E. O. Stinson, N. L. Washington, M. D. Perry, P. Ruzanov, Z. Zha, S. E. Lewis, L. D. Stein, and G. Micklem, "modMine: Flexible access to modENCODE data," *Nucleic Acids Res.*, vol. 40, no. D1, pp. 1082–1088, 2012.

[36]   M. Lupien, J. Eeckhoute, C. a Meyer, Q. Wang, Y. Zhang, W. Li, J. S. Carroll, X. S. Liu, and M. Brown, "FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription.," *Cell*, vol. 132, no. 6, pp. 958–70, Mar. 2008.

[37]   E. Crane, Q. Bian, R. P. McCord, B. R. Lajoie, B. S. Wheeler, E. J. Ralston, S. Uzawa, J. Dekker, and B. J. Meyer, "Condensin-driven remodelling of X chromosome topology during dosage compensation," *Nature*, vol. 523, no. 7559, pp. 240–244, 2015.

[38]   C. Holdridge and D. Dorsett, "Repression of hsp70 heat shock gene transcription by the suppressor of hairy-wing protein of Drosophila melanogaster.," *Mol. Cell. Biol.*, vol. 11, no. 4, pp. 1894–900, 1991.

[39]   M. Peifer and W. Bender, "Sequences the Drosophila effects adjacent," vol. 85, no. December, pp. 9650–9654, 1988.

[40]   P. A. Smith and V. G. Corces, "The suppressor of Hairy-wing binding region is required for gypsy mutagenesis," *MGG Mol. Gen. Genet.*, vol. 233, no. 1–2, pp. 65–70, 1992.

[41]   P. K. Geyer, M. M. Green, and V. G. Corces, "Mutant gene phenotypes mediated by a Drosophila melanogaster retrotransposon require sequences homologous to mammalian enhancers," *Proc Natl Acad Sci U S A*, vol. 85, no. 22, pp. 8593–8597, 1988.

[42]   P. K. Geyer and V. G. Corces, "DNA position-specific repression of transcription by a Drosophila zinc finger protein," *Genes Dev.*, vol. 6, no. 10, pp. 1865–1873, 1992.

[43]   R. Kellum and P. Schedl, "A Group of scs elements function as domain boundaries in an enhancer-blocking assay." vol. 12, no. 5, pp. 2424–2431, 1992.

[44] R. Kellum and P. Schedl, "A position-effect assay for boundaries of higher order chromosomal domains," *Cell*, vol. 64, no. 5, pp. 941–950, 1991.

[45] F. Recillas-Targa, M. J. Pikaart, B. Burgess-Beusse, A. C. Bell, M. D. Litt, A. G. West, M. Gaszner, and G. Felsenfeld, "Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 99, no. 10, pp. 6883–6888, 2002.

[46] R. R. Roseman, V. Pirrotta, and P. K. Geyer, "The su(Hw) protein insulates expression of the Drosophila melanogaster white gene from chromosomal position-effects.," *EMBO J.*, vol. 12, no. 2, pp. 435–42, Feb. 1993.

[47] T. G. Kahn, Y. B. Schwartz, G. I. Dellino, and V. Pirrotta, "Polycomb complexes and the propagation of the methylation mark at the drosophila Ubx gene," *J. Biol. Chem.*, vol. 281, no. 39, pp. 29064–29075, 2006.

[48] D. R. Mallin, J. S. Myung, J. S. Patton, and P. K. Geyer, "Polycomb group repression is blocked by the Drosophila suppressor of Hairy-wing [SU(HW)] insulator," *Genetics*, vol. 148, no. 1, pp. 331–339, 1998.

[49] C. J. A. Sigrist and V. Pirrotta, "Chromatin insulator elements block the silencing of a target gene by the Drosophila polycomb response element (PRE) but allow trans interactions between PREs on different chromosomes," *Genetics*, vol. 147, no. 1, pp. 209–221, 1997.

[50] S. Cuddapah, R. Jothi, D. E. Schones, T. Y. Roh, K. Cui, and K. Zhao, "Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains," *Genome Res.*, vol. 19, no. 1, pp. 24–32, 2009.

[51] B. Langmead, C. Trapnell, M. Pop, and S. Salzberg, "Ultrafast and memory-efficient alignment of short DNA sequences to the human genome," *Genome Biol.*, vol. 10, no. 3, p. R25, 2009.

[52] W. J. Kent, C. W. Sugnet, T. S. Furey, and K. M. Roskin, "The Human Genome Browser at UCSC W.," *J. Med. Chem.*, vol. 19, no. 10, pp. 1228–31, 1976.

[53] K. R. Rosenbloom, J. Armstrong, G. P. Barber, J. Casper, H. Clawson, M. Diekhans, T. R. Dreszer, P. A. Fujita, L. Guruvadoo, M. Haeussler, R. A. Harte, S. Heitner, G. Hickey, A. S. Hinrichs, R. Hubley, D. Karolchik, K. Learned, B. T. Lee, C. H. Li, K. H. Miga, N. Nguyen, B. Paten, B. J. Raney, A. F. A. Smit, M. L. Speir, A. S. Zweig, D. Haussler, R. M. Kuhn, and W. J. Kent, "The UCSC Genome Browser database: 2015 update," *Nucleic Acids Res.*, vol. 43, no. D1, pp. D670–D681, 2015.

Figure 1: (A) Combined view from the UCSC genome browser of polytene banding structure resolved by (31), Hi-C interaction heatmap, Hi-C directionality index, DNase accessibility, ChIP-chip data for the insulator proteins CP190, BEAF-32, dCTCF, and Mod(mdg4) for the region covering chrX:10,971,551-11,074,797. Dashed green lines indicate the locations of TAD boundaries identified by Hi-C. Hi-C blocks lacking data result from genomic regions that lack MboI cut sites (GATC) and are uninformative. (B) Graphical representation of the chromatin expansion model of insulator function. The bands shown are proportionally sized according to genomic size (bp, top) and physical distance in polytene chromosomes (bottom) as measured by (31). Insulator proteins are represented to indicate that they drive the extension of these regions, not to imply that they coat them entirely. The presence of active promoters within the borders of a subset of these interbands suggests that these regions are not entirely occluded by insulator proteins. The mechanism(s) by which insulator proteins decompact chromatin are unclear.

Figure 1