Modeling the Effect of Histone Methylation on Chromosomal Organization in Colon Cancer Cells

Neil Chowdhury
Phillips Exeter Academy
Exeter, NH, USA

Mentor: Maxim Imakaev
Massachusetts Institute of Technology
Cambridge, MA, USA

Biology Award

Modeling the Effect of Histone Methylation on Chromosomal Organization in Colon Cancer Cells

Author: Neil Chowdhury, **Mentor**: Maxim Imakaev

Abstract

Loop extrusion and compartmentalization are the two most important processes regulating the high-level organization of DNA in the cell nucleus. These processes are largely believed to be independent and competing (Sanborn et al. 2015; Nuebler et al. 2018). Chromatin consists of nucleosomes, which contain coils of DNA wrapped around histone proteins. Besides packing DNA, nucleosomes contain an “epigenetic code” - tails of histone proteins are chemically modified at certain positions to leave certain “histone marks” on the chromatin fiber. This paper explores the effect of the H3K9me3 histone modification, which typically corresponds to inactive and repressed chromatin, on genome structure. Interestingly, in H3K9me3 domains, there are much fewer topologically associating domains (TADs) than in other domains, and there is a unique compartmentalization pattern. A high-resolution polymer model simulating both loop extrusion and compartmentalization is created to explore these differences.

Keywords: Hi-C, genome architecture, histone modification, molecular dynamics simulations

Table of Contents

Introduction

Chromatin, the combination of DNA and DNA-associated proteins in the cell nucleus, is composed of chromatin fiber, nucleosomes, and histones. Figure 1 shows that nucleosomes consist of a histone octamer made of four types of histone proteins (H2A, H2B, H3, and H4), as well as a short length of DNA. Chromatin consists of a string of nucleosomes in chromatin fiber.

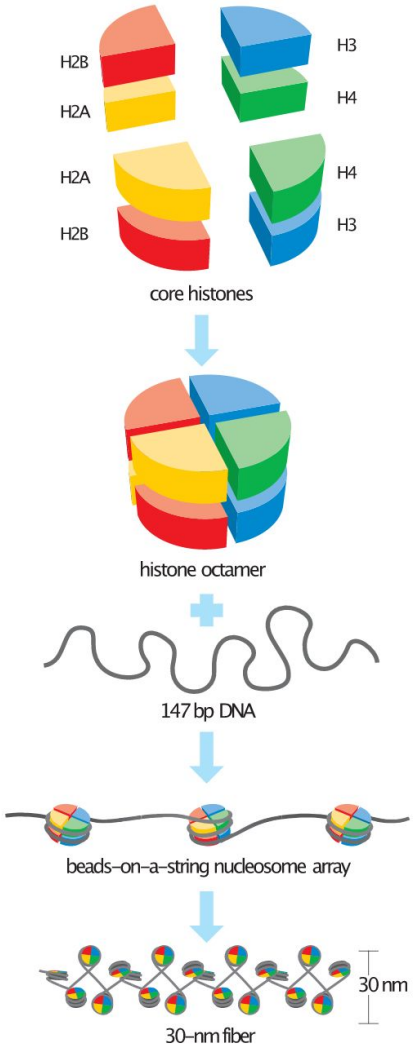

Figure 1. Structure of chromatin.[1]

[1] By David O Morgan - The Cell Cycle. Principles of Control., Attribution, https://commons.wikimedia.org/w/index.php?curid=89674546

The large-scale organization of chromatin is a topic of major interest. Hi-C is a genome-wide technique to probe the 3D structure of genomes by recording the events when two chromosomal regions are in direct proximity. It converts these events into unique molecular products that can be read out using high-throughput sequencing (Lieberman-Aiden et al. 2009). The result of Hi-C is a set of *contact probabilities* between pairs of chromosomal regions. Hi-C is represented as a *contactmap*, a square matrix where cells represent the contact probability between the DNA region indicated by its row and column (Figure 2); darker cells in the figure correspond to a higher contact probability. As reflecting across the main diagonal is equivalent to switching the order of the row and column, the Hi-C matrix is symmetric about the main diagonal, the line going from the top left to the bottom right. Details of the Hi-C protocol are reviewed in (Lajoie, Dekker, and Kaplan 2015).

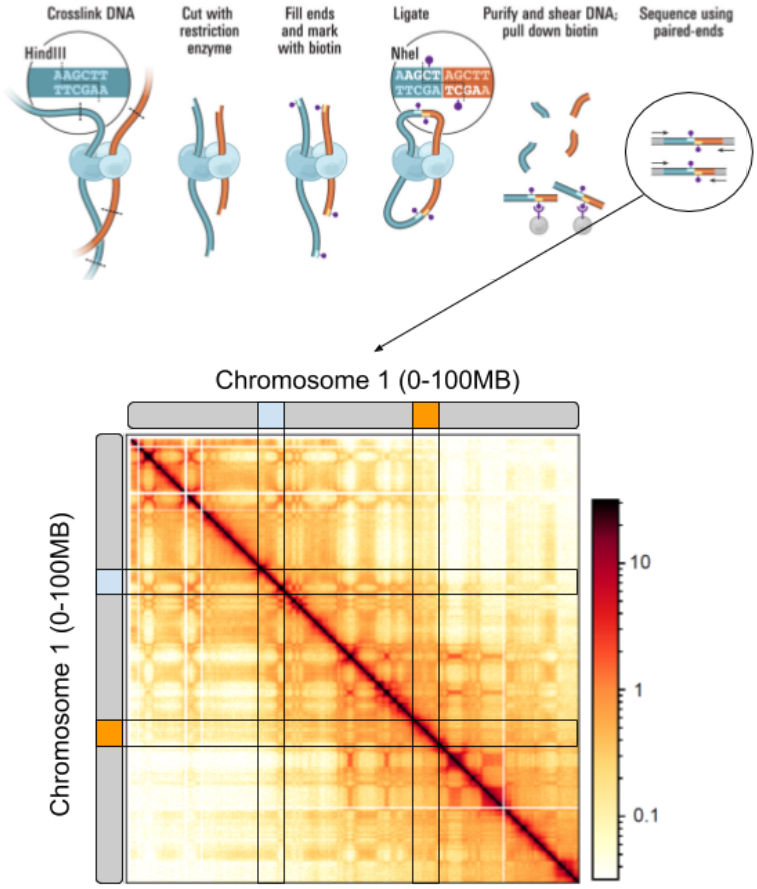

Figure 2. Summary of the Hi-C process. Hi-C process (top image) is taken from (Lieberman-Aiden et al. 2009).

One immediate feature of the Hi-C matrix is the differentiation between *cis* and *trans* interactions (Figure 3) (Lieberman-Aiden et al. 2009). Cis interactions represent the contacts between regions on the same chromosome, while trans interactions represent the contacts

between regions on different chromosomes. Cis interactions are generally stronger than trans interactions because regions on the same chromosome are physically linked together.

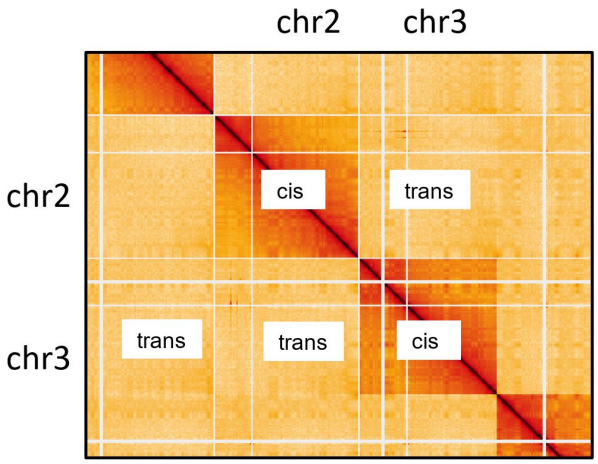

Figure 3. Cis and trans interactions in a Hi-C matrix. Uses Hi-C data from (Schwarzer et al. 2017). Taken from (Chowdhury and Abraham 2019).

Compartmentalization is responsible for the checkerboarding pattern observed in both cis and trans interactions in Hi-C (Lieberman-Aiden et al. 2009). It is believed to be formed by monomers of the chain having different patterns of attraction to one another (Figure 4) (Falk et al. 2019; Nuebler et al. 2018). It is commonly understood that there are two main compartments, called A and B, respectively consisting of active (euchromatic) and inactive (heterochromatic) DNA. The checkerboarding pattern most likely results from heterochromatic B-B attraction, causing B domains to phase separate from A domains (Falk et al. 2019).

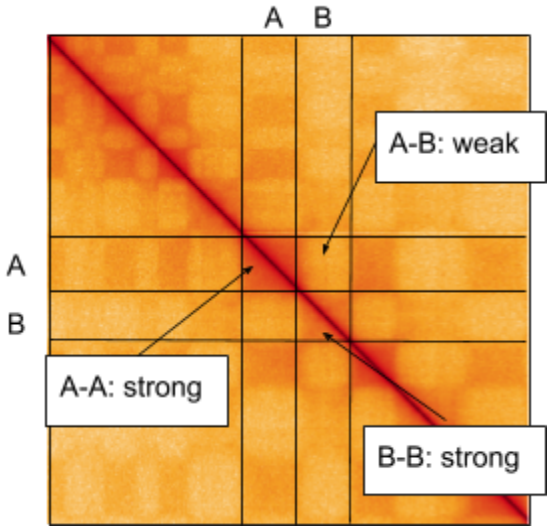

Figure 4. A/B compartmentalization in part of a Hi-C matrix.

In addition, several low-level organizational features formed by loop extrusion are visible in high-resolution Hi-C, including TADs, loops, and stripes (Figure 5). Cohesin is responsible for extruding loops, and it functions by attaching onto the genome and connecting two regions with its legs, progressively increasing the size of the loop (Fudenberg et al. 2016; Sanborn et al. 2015). Cohesins can be stopped from extruding in one direction of the chain if they hit a CCCTC-binding factor (CTCF) protein loaded at a CTCF site (Fudenberg et al. 2016; Phillips and Corces 2009).

TADs are compact domains with a high mean contact probability in their interior compared to interactions with their exterior (Reviewed in Szabo, Bantignies, and Cavalli 2019). They appear as darker squares on the main diagonal of the Hi-C matrix and form when cohesins extrude loops within the TAD boundaries, as shown in the blue square in Figure 5. TADs contain CTCF sites at their boundaries which constrain cohesins within the TAD.

"Loops" (also known as "dots" or "peaks") are enriched regions of contact probability between TAD boundaries, circled in green in Figure 5 (Fudenberg et al. 2016; Rao et al. 2014). Stripes (also known as “flames”) are caused by loop anchors interacting with the entire body of a TAD, outlined in pink in Figure 5 (Vian et al. 2018).

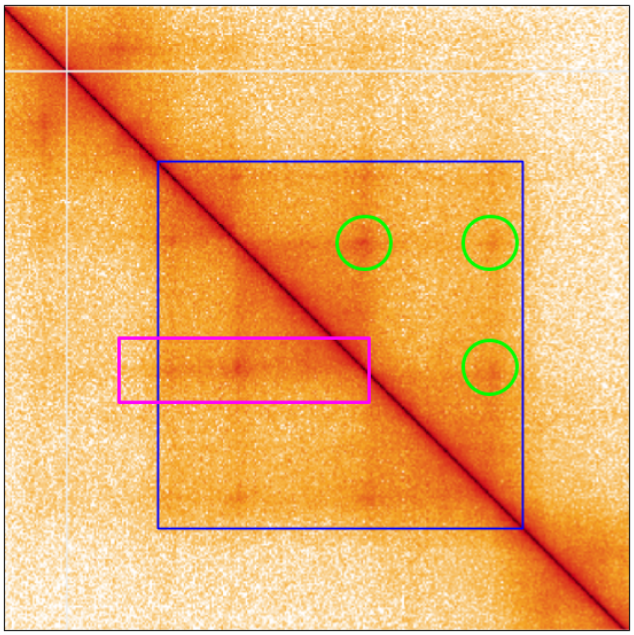

Figure 5. Hi-C map with a TAD outlined in blue, a stripe outlined in pink, and multiple loops circled in green.

The formation of each of these loop extrusion features is described in Figure 6.

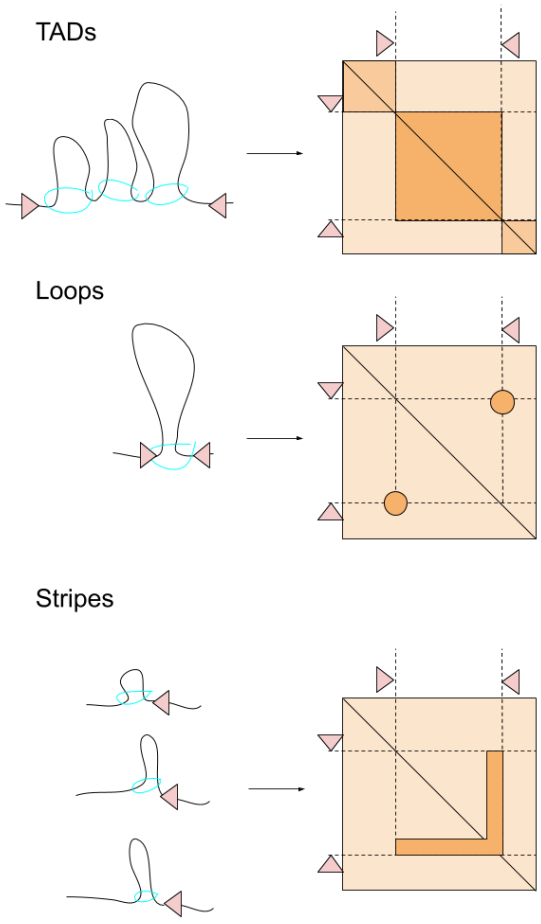

Figure 6. TADs form by loop extrusion by cohesins (blue rings) between the TAD boundaries (red triangles). Loop extrusion within the domain causes increased interactions between regions within TAD, which are visible in Hi-C. Loops form when regions at CTCF sites are moved close to each other by a cohesin. Stripes form when a cohesin is fixed at one side, strengthening interactions between the region at that side and the rest of the TAD.

The goal of running simulations is to demonstrate that a set of polymer forces can produce a contactmap that replicates patterns in Hi-C. This lets us make conjectures about the internal processes regulating genome structure. Modeling loop extrusion in simulations has been explored by Fudenberg et al., and A/B compartmentalization has been explored by Falk et al. (Fudenberg et al. 2016; Falk et al. 2019). Nuebler et al. created a model combining the compartmentalization and loop extrusion models and found that compartmentalization and loop extrusion are competing processes in megabase-scale chromosome organization (Nuebler et al. 2018).

This paper focuses on the effect of the histone modification H3K9me3, specifically in the HCT116 (human colon cancer) cell line. The analysis will additionally use experimental data

from HCT116 cells in which two DNA methyltransferases, DNMT1 and DNMT3b, have been disrupted. These are known as DKO cells, and they are deficient in histone modification (Jacinto et al. 2007). RAD21-depleted cells, which are missing a subunit of the cohesin protein complex, do not undergo loop extrusion and will also be analyzed (Seitan et al. 2013; Rao et al. 2017). Hi-C and chromatin immunoprecipitation-sequencing (ChIP-seq) data for DKO cells were produced by George Spracklin in the Job Dekker lab at UMass Medical Center as part of a manuscript in preparation (Spracklin et al. 2020); data for wild type HTC116 cells were obtained from (Rao et al. 2017).

Although other studies of simulations have explored loop extrusion and A/B compartmentalization separately and together, this work is the first high-resolution modeling of Hi-C with more than two compartments in combination with loop extrusion.

Results

H3K9 domains induce a distinct compartmentalization pattern separate from A/B

A technique based in eigendecomposition of cis interactions can be used to annotate the partitioning of the genome into the broad A/B classes. The cis interactions are first divided by the means on their diagonals, null rows and columns are truncated, and then it is eigendecomposed into its eigenvectors and eigenvalues. Since the outer product of the eigenvector with the largest value (the first eigenvector) makes up the greatest part of the matrix in the eigendecomposition, it can be used to assign A/B compartments.

A indicates where the first eigenvector (E1) is positive, and *B* indicates where the first eigenvector is negative. In the outer product of the eigenvector, A-A and B-B interactions multiply to get a positive value, indicating a stronger interaction, and A-B interactions multiply to a negative value, indicating weaker interaction (Figure 7).

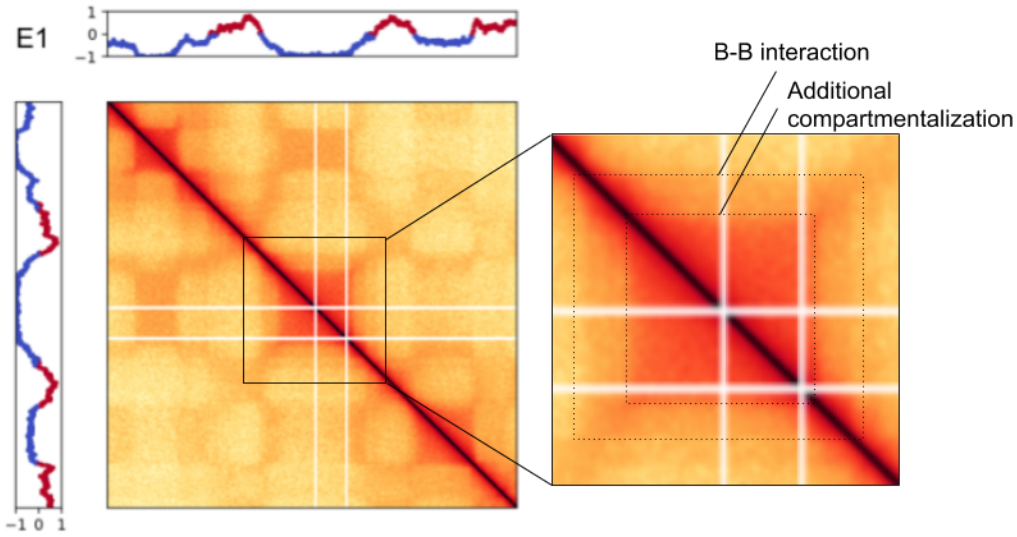

Figure 7. E1 plotted with a section of a Hi-C map (HCT116/Rad21 depletion, chromosome 5, 152-160Mb). The eigenvector is colored by compartment (red: positive, A; blue: negative, B). The right image is a zoomed-in portion of the image on the left, showing additional, stronger compartmentalization within a B-B interaction.

However, Figure 7 shows a B-B domain interaction with an additional layer of compartmentalization inside it, suggesting that there are more than two compartmental states. To account for this, H3K9me3 ChIP-seq can be used to annotate the *H3K9* compartment, a subset of the B compartment (Figure 8); the H3K9 compartment includes loci where the H3K9me3 histone modification is present. It appears that this H3K9 compartment corresponds to the increased layer of compartmentalization.

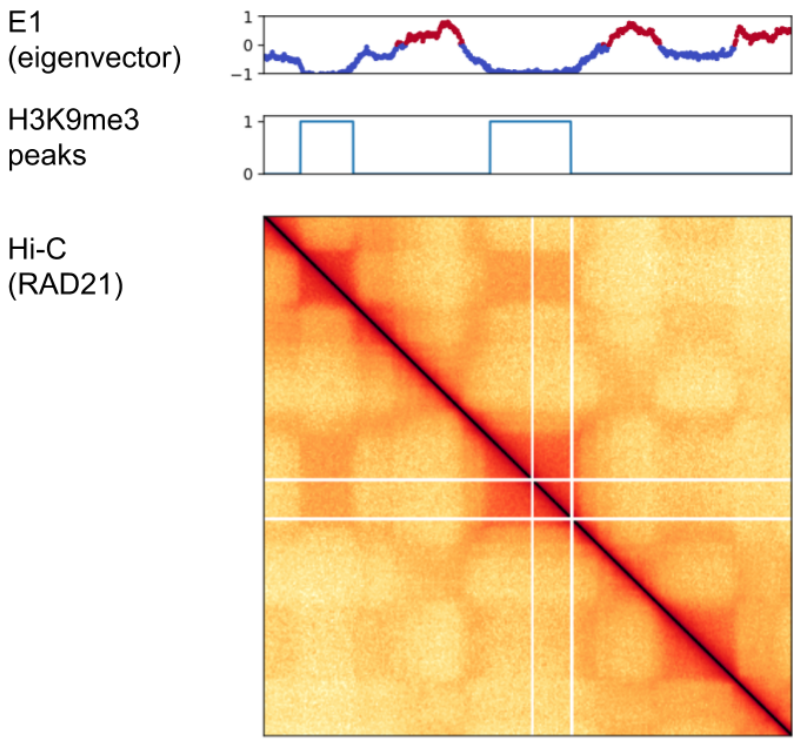

Figure 8. H3K9me3 ChIP-seq plotted with E1.

As seen in Figure 8, regions with higher levels of H3K9me3 appear to compartmentalize within B regions, as H3K9-H3K9 interactions appear stronger than B-B or B-H3K9 interactions. H3K9 domains form a “subcompartment” within the B compartment (Spracklin et al. 2020). There are other subcompartments within A and potentially outside both A and B; for simplicity’s sake, we explore only the weak A (normal A), polycomb B (normal B), and heterochromatic B (H3K9) subcompartments in this paper, as labeled by (Spracklin et al. 2020).

To quantitatively describe interactions between A, B, and H3K9 compartments, I measured average contact frequencies between A, B, and H3K9 compartments in HCT116 and RAD21 Hi-C (Figure 9). I found that all three compartments most frequently interact with themselves; two loci with the same compartment label are more likely to interact than two loci with different compartment labels. The mean contact probabilities between different compartments vary: A-H3K9 contact probability is the lowest (0.57), whereas A-B and B-H3K9 are intermediate (0.87 and 0.92).

A physical process behind compartment formation still remains unclear. A recent publication showed that A-B compartmentalization is likely driven by B-B attraction, and not by A-A attraction or by attraction to the nuclear periphery (Falk et al. 2019). I hypothesized that causing the pairs B-B, H3K9-H3K9 and B-H3K9 to have different affinities could drive the formation of such a complex 3-compartment structure. Polymer models were used to check if varying attraction between A, B, and H3K9 compartments could reconstruct the average

inter-compartment contact probabilities inferred from the data; this is explored in the next section.

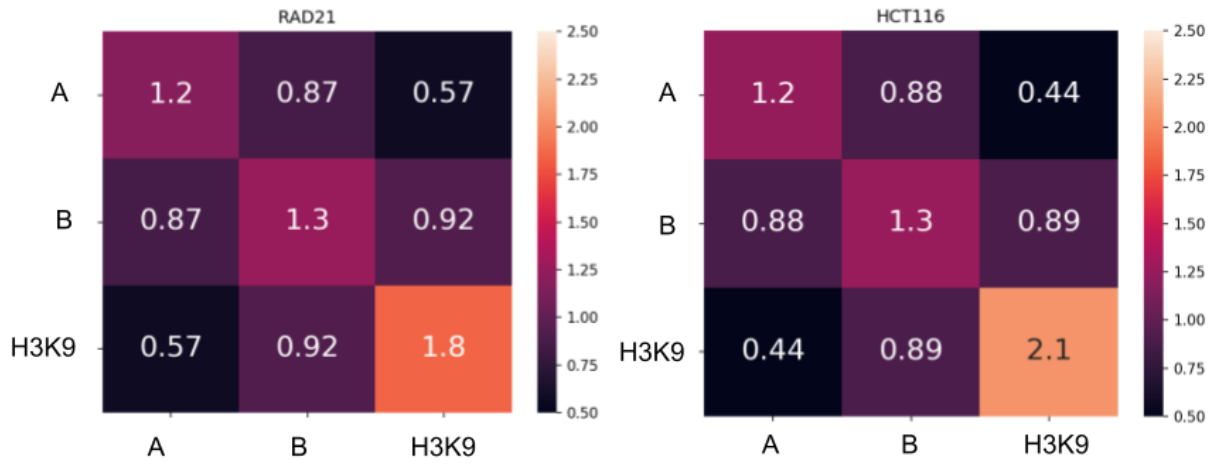

Figure 9. Left: RAD21-depletion interaction frequency heatmap. Right: HCT116 interaction frequency heatmap. Cells indicate the mean interaction frequency between the compartment in the row and column.

Polymer simulations reproduce the compartmentalization pattern in HCT116 cells

To reproduce features observed in HCT116 cells, a polymer model was designed that would account for interactions between all 3 types of subcompartments. The model also included the process of loop extrusion, which is necessary for the formation of TADs, and also is known to interfere with A-B compartmentalization (Nuebler et al. 2018).

This model is flexible and allows us to predict how perturbations to loop extrusion and compartmentalization are reflected in the contactmap. Several perturbations of the model were performed to verify that it accurately reflects Hi-C.

The first perturbation is running with a two-state A/B compartmentalization model (without H3K9 compartmentalization). Simulations of a region in HCT116 were run with a three-state heteropolymer model and a two-state model, with and without loop extrusion. The three-state model is much better at recapitulating the Hi-C interaction frequencies than the two-state model (Figure 10). This suggests that H3K9-H3K9 attraction is indeed higher in HCT116.

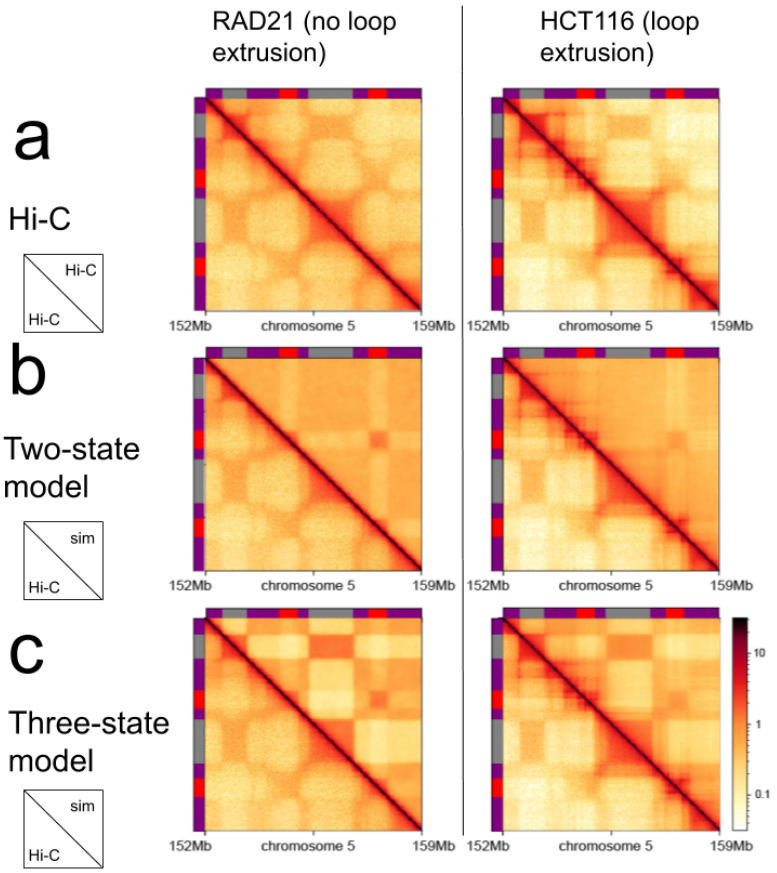

Figure 10. Two-state and three-state models. The left column represents RAD21-depletion conditions, where loop extrusion is prevented, and the right column represents normal HCT116 cells. Row (a) is the original cell Hi-C, row (b) shows Hi-C and the two-state model contactmap split across the main diagonal, and row (c) shows Hi-C and the three-state model contactmap split across the main diagonal. The colored bars to the left and right indicate the domain type: red (A), purple (B), and gray (H3K9).

Loss of methylation corresponds with a loss of H3K9 compartmentalization and re-establishment of TADs in H3K9 domains

The data analyses performed by (Spracklin et al. 2020) indicate that DKO cells have deficiencies of H3K9me3 in some heterochromatic domains. The domains where H3K9me3 is depleted are called disrupted domains, and the domains where H3K9me3 remains are called persistent domains.

H3K9 domains in HCT116 are generally devoid of loop extrusion features like stripes, loops, and TADs. However, these features appear in several disrupted domains, as seen in DKO Hi-C. Figure 11 shows one example of a disrupted domain that contains features of loop extrusion.

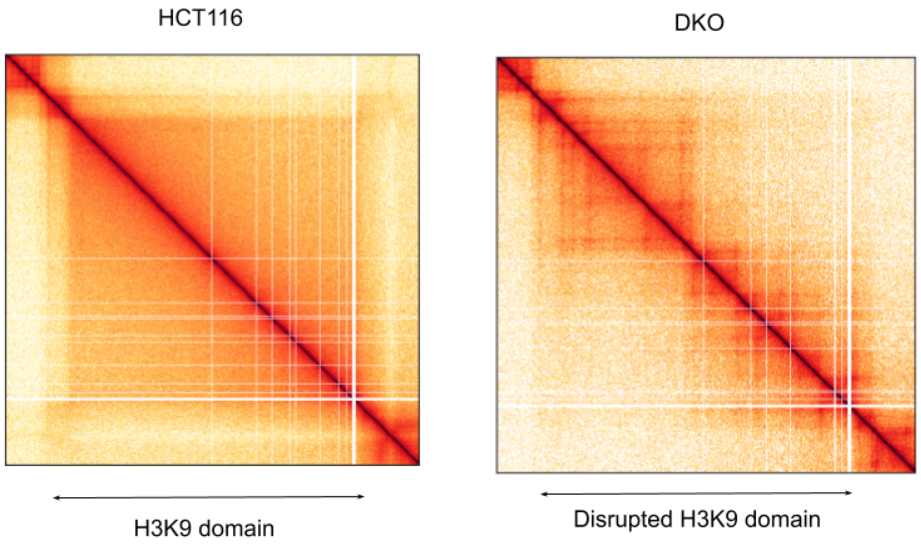

Figure 11. A disrupted H3K9 domain (chromosome 6, 90-96Mb) in HCT116 and DKO. In DKO, TADs, loops, and stripes reappear. This is one example of TAD reappearance in disrupted domains; this pattern is observed throughout the genome.

To probe the differences between disrupted and persistent domains, three perturbations to the polymer model were made, all based on the HCT116 “control” with CTCFs from HCT116 ChIP-seq. The RAD21 variation has no loop extrusion, the DKO (persistent) variation has CTCFs from DKO ChIP-seq, and the DKO (disrupted) variation uses DKO CTCFs and further removes H3K9 compartmentalization by treating H3K9 domains as normal B regions.

As seen in Figure 12, the disrupted DKO model (d), without H3K9 compartmentalization, is a good match for the disrupted domain, while the persistent DKO model (c), with H3K9 compartmentalization, is a good match for the persistent domain. Therefore, disrupted HP1 domains behave like the normal B compartment in DKO conditions. This lets us conclude that H3K9 compartmentalization and TAD disappearance is correlated with the H3K9me3 histone modification.

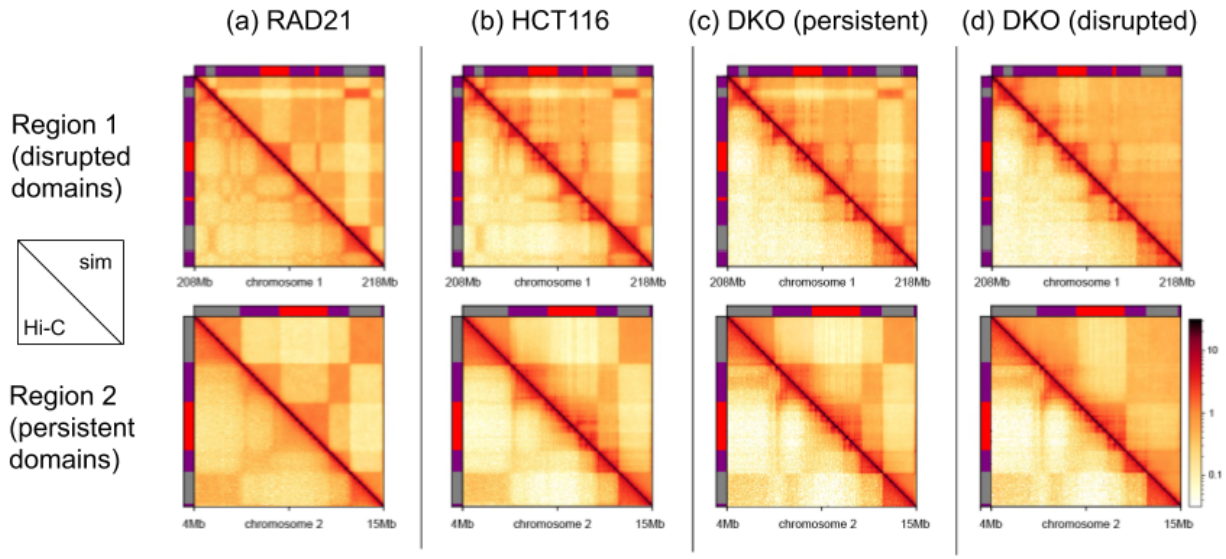

Figure 12. Contactmaps with Hi-C on the left half and a simulation contactmap on the right half, separated by the main diagonal. Region 1 contains two disrupted H3K9 domains and Region 2 contains two persistent domains. These occur in four different conditions: (a) RAD21: No extrusion, three-state model of compartmentalization; (b) HCT116: Extrusion, three-state model; (c) DKO (persistent): Extrusion, three-state model; (d) DKO (disrupted): Extrusion, two-state model. The colored bars to the left and right indicate the domain type: red (A), purple (B), and gray (H3K9).

The fact that TADs reappear in disrupted domains suggests that the loading of CTCFs to CTCF sites is inhibited by a factor correlating with H3K9me3 modification. It is also plausible that loop extrusion itself is prevented by the H3K9me3 modification; this is shown to not be true in the next section.

Loop extrusion occurs in H3K9 domains

Since TADs and loops, hallmarks of loop extrusion, are absent in H3K9 domains, one may think that loop extrusion itself is inhibited in H3K9 domains. However, I will show that these are not the result of inhibiting loop extrusion, but of just deactivating CTCFs. Loop extrusion still occurs normally in H3K9 domains.

One reason why extrusion must occur is that Hi-C shows that several H3K9 domains contain stripes traversing the length of the domain, which is only possible when cohesins are allowed to enter H3K9 domains. Simulations were run where loop extrusion was allowed or restricted, in an H3K9 domain containing stripes traversing the domain. Figure 13 shows the resulting contactmaps.

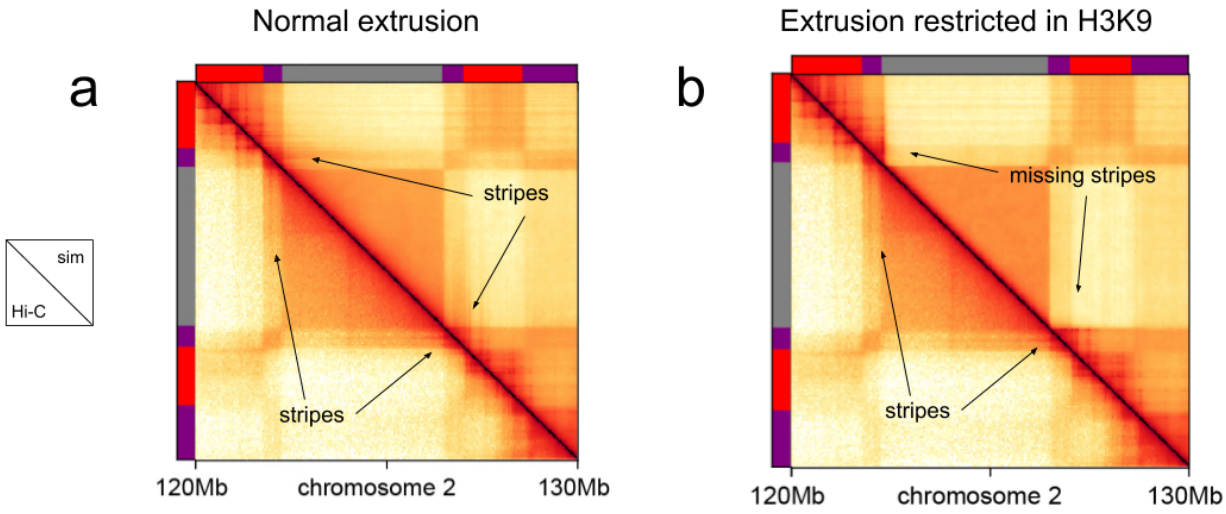

Figure 13. Simulations of an H3K9 domain with/without loop extrusion in the H3K9 domain. (a) Normal extrusion rules. (b) Restricted extrusion in H3K9. The colored bars to the left and right indicate the domain type: red (A), purple (B), and gray (H3K9).

The left contactmap shows the result of loop extrusion in normal conditions and the right contactmap shows the result when loop extrusion is prevented. Preventing extrusion results in the disappearance of stripes, as indicated in the figure, which is not observed in Hi-C. Therefore, it can be concluded that the extrusion of cohesins is not restricted in H3K9 domains. The only alteration to loop extrusion that H3K9 domains have is that they undergo depletion of CTCFs, as mentioned in the previous section.

Methods

Simulation model

The simulation is modeled as a system with five equivalent polymer chains, each with around 10,000 monomers. Each monomer represents one kilobase of the genome, so each chain represents 10 million base pairs of the genome. The model is run using the open-source GPU-assisted library OpenMM, which allows defining forces on the monomers and uses integration to determine positions and velocities for each timestep (Eastman et al. 2017). Bonds between monomers are harmonic bonds with a fixed equilibrium length and a force proportional to the change from rest length (i.e. $F = -k\Delta x$) (Figure 14).

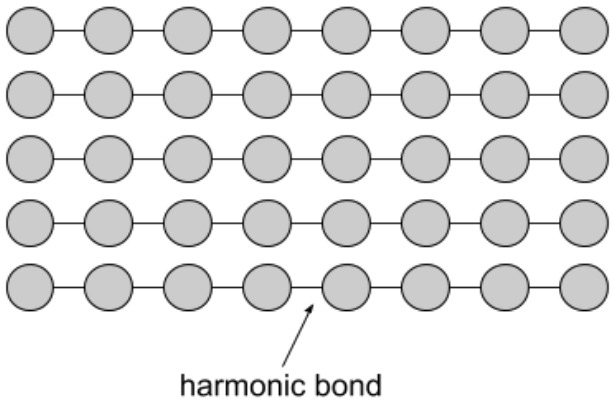

Figure 14. Simulation model. Consists of five polymer chains of monomers. Each monomer represents 1,000 base pairs. Harmonic bonds connect adjacent pairs of monomers on the same chain.

The model is also run with periodic boundary conditions (PBC) with a fixed density, shown in Figure 15. This means that monomers experience forces depending on their x-, y-, and z-coordinates modulo the PBC box width. Plots of the simulation conformations appear as cubes with sides the same length as the PBC width. This ensures that chains do not drift apart and that the simulation remains at a fixed density, similar to the confinement of the cell nucleus (Di Stefano et al. 2016).

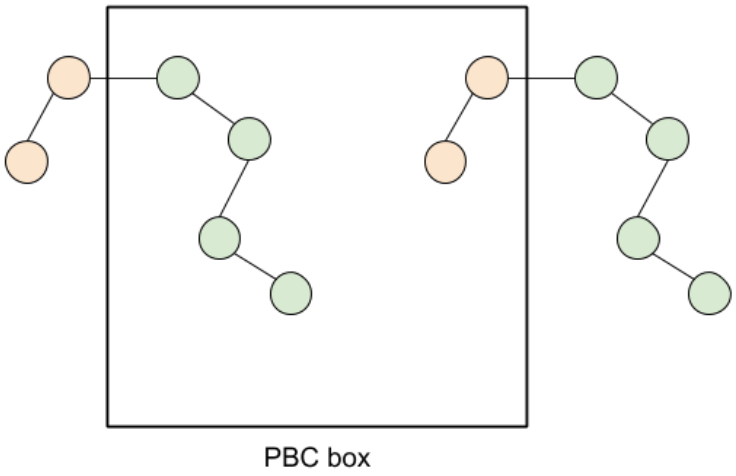

Figure 15. Periodic boundary conditions. All monomers remain inside the box, and as they pass through one side of the box, they appear on the other side but continue interacting with neighboring monomers.

The simulations developed in this paper were adapted from prior work in simulating Hi-C with only two compartments and simpler loop extrusion rules (Nuebler et al. 2018; Falk et al. 2019).

Modeling compartmentalization

It is believed that compartmentalization is a result of attractive forces between heterochromatic domains. This can be simulated by creating a variable attractive force between a pair of monomers that depends on only the monomer types of the pair; monomers in heterochromatic domains attract.

A force based on the smooth square well potential was used to simulate different types of interactions between monomers shown in Figure 16. The smooth square well force is a "smoothed" version of the square well force, with its potential energy V(x) linearly related to $x^{12}(x^2-1)$, where x is the distance between monomers. This heteropolymer force consists of two smooth square wells put together, one attractive and one repulsive (Nuebler et al. 2018). The y-axis represents potential energy and the x-axis represents the distance between a pair of monomers. When the monomers are too close to one another, they experience a large repulsion, but at a certain distance, they may experience a small attraction. The depth of this attraction well depends on the monomer types that interact.

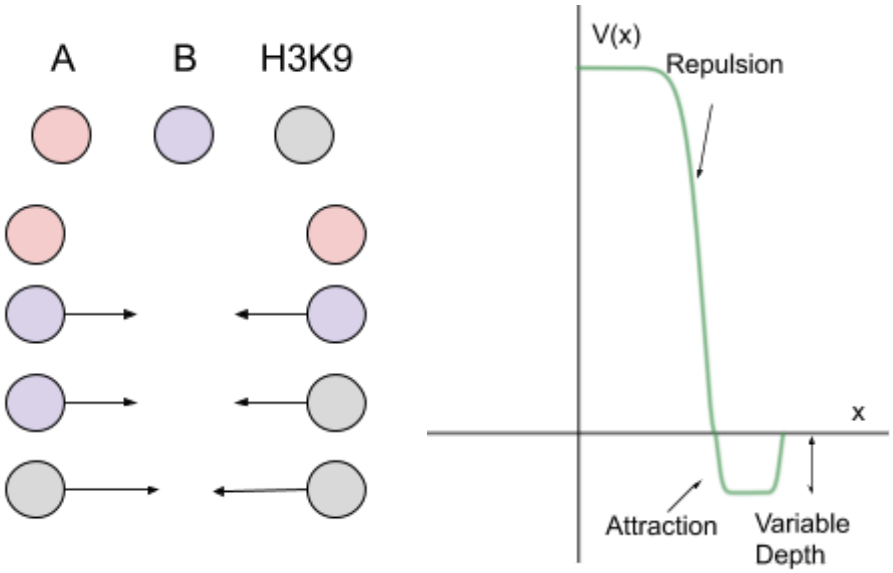

Figure 16. The Smooth Square Well potential. The depth of the attractive well depends on the monomers present. For example, two B monomers may interact more strongly than an A and a B monomer.

Simulations were run with a single monomer type and varying attraction energy to demonstrate the result of attraction on a polymer conformation. 3D maps of conformations taken after the simulations had equilibrated are shown in Figure 17. With low attraction energies, the polymer is mainly in an expanded state, shown in row 1, but higher attraction energies result in collapsed polymers, shown in row 2. The collapse is a result of the attraction of monomers outweighing the entropic effects that lead to expansion.

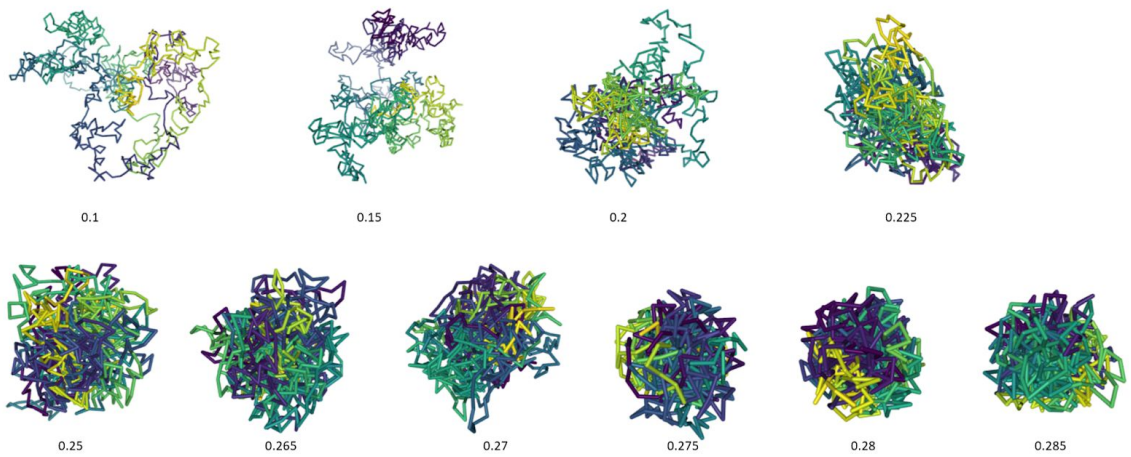

Figure 17. 3D maps of conformations with increasing attraction energy (k_BT), shown below the conformations.

Figure 18 shows the result of a two-state heteropolymer simulation, where monomers are assigned to either A or B. The attraction energy of BB was set to 0.25 k_BT, while the attraction energy of AA and AB was set to 0 k_BT, resulting in the cis contactmap shown below. It can be seen that the A domains interact more with each other, and the B domains interact more with each other, creating a checkerboarding pattern. The plot on the right shows a conformation in which A monomers have been colored red and B monomers have been colored blue. A and B are segregated into distinct locations in the conformation as the result of phase separation. The checkerboarding pattern produced appears similar to A/B compartmentalization pattern in Hi-C. Three-state models that account for the additional H3K9-H3K9 attraction are explored in the next sections.

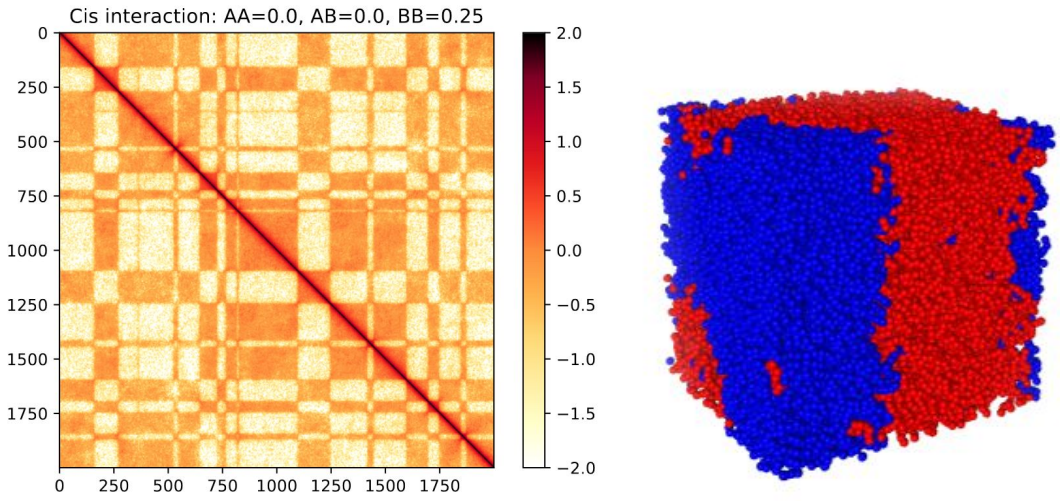

Figure 18. A two-compartment model. Left: Cis contactmap (colormap shows logged interaction frequencies) showing the checkerboarding pattern, Right: a conformation from the simulation (Red monomers: A, blue monomers: B)

Iteratively fitting compartments

Chromatin compartments have been simulated in (Nuebler et al. 2018; Falk et al. 2019). A simulation model lets one view the effect of compartmentalization using a given set of attraction energies, but this is not easily reversed; it's hard to find the attraction energies that lead to a compartmentalization pattern. In a two-state model, the strength of compartmentalization can be adjusted by primarily sweeping one parameter, the B-B attraction. However, for a three-state model, there are five adjustable parameters: the attraction energies of A-A, A-B, A-H3K9, B-B, B-H3K9, and H3K9-H3K9. To find an optimal set of attraction energies, I designed an iterative process that allows us to converge on the correct parameters (Figure 19).

Let E_i be the attraction energy matrix at iteration i, and $f(E)$ be the interaction matrix produced by the attraction energy matrix E. The iterative process works by first running simulations to compute $\frac{f(E_i+dE)}{|dE|}$, where dE is a small adjustment to the attraction energy between every pair of monomers (A-A, A-B, A-H3K9, B-B, B-H3K9, H3K9-H3K9). These changes are used in a linear-regression fashion to predict the resulting mean interactions for a given attraction energy matrix. For the next iteration, the attraction energy matrix E_{i+1} is chosen that minimizes the squared difference between the predicted mean interactions and the interaction frequencies observed from Hi-C is chosen. Finally, the attraction energies are all multiplied by a constant factor determined in a linear sweep to find the optimal interaction frequency compared to the Hi-C interaction frequencies.

The end result of iteration is shown in Figure 19. At the end of the 5th iteration, it was found that the simulated interaction frequencies are very close to the interaction frequencies observed in Hi-C. The optimal attraction energies show that the H3K9-H3K9 energy is higher than B-B, which is higher than A-A. The resulting interaction frequencies produced contain the relative observations observed from the Hi-C frequencies: loci in the same compartment attract more frequently than loci in different compartments, and the A-B and B-H3K9 attractions are more intermediate in strength. The attraction energies found at the end of iteration are used as the parameter for compartmentalization in the simulations.

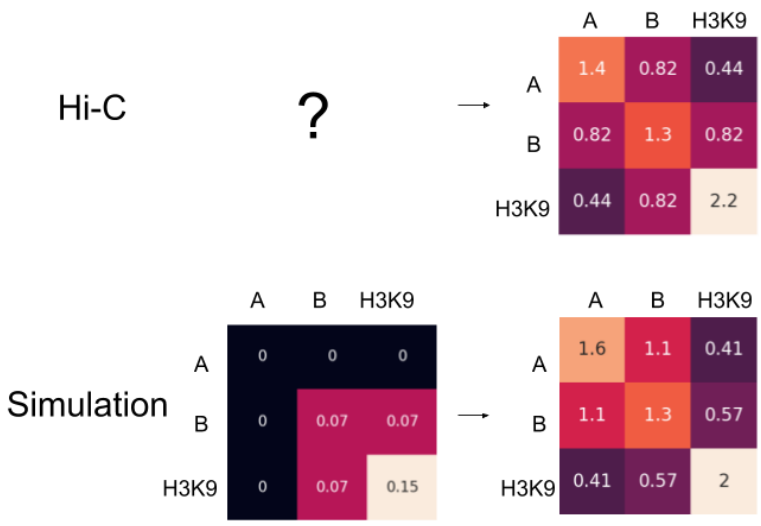

Figure 19. Iteratively fitting compartments. The first row shows an unknown set of attraction energies creating the desired interaction frequencies from Hi-C, which were derived from the chromosome on which the simulation was run. The next row shows the optimal attractive energies and simulated interaction frequencies generated by the iterative process.

Modeling loop extrusion

In this paper's loop extrusion model, cohesins are protein complexes that bind to the chromatin fiber at random positions and facilitate loop extrusion. Over time, the cohesin expands; one of its legs moves downstream, while the other moves upstream. For example, a cohesin connecting monomers i and j will connect $i-1$ and $j+1$ in the next time step. The cohesin exerts a harmonic force pulling the regions connected by its legs together (Figure 20). Cohesins also have a lifetime; they unload from the genome randomly with a probability of 1/lifetime. Cohesin legs can be in a stalled state, for example, when they collide with one another on the strand. Stalled cohesin legs do not move, but the other cohesin leg may continue to extrude unidirectionally.

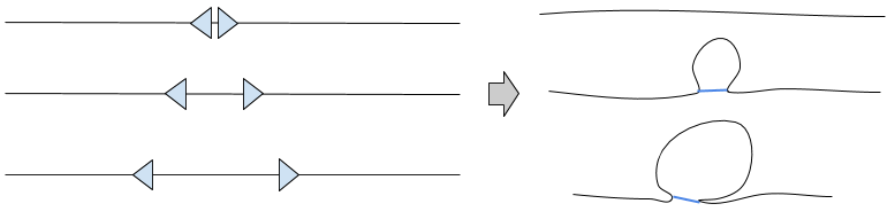

Figure 20. Left: The cohesin legs and the direction they travel in. Right: The resulting polymer chain conformation, with the cohesin bond indicated in blue.

CTCF is a transcription factor that regulates loop extrusion. CTCFs are probabilistically loaded onto specific sites on chromatin and stall cohesin legs, preventing them from extruding either downstream or upstream, depending on the CTCF's direction (Figure 21). In this paper's cohesin model, cohesins are also prevented from unloading while stalled at a CTCF.

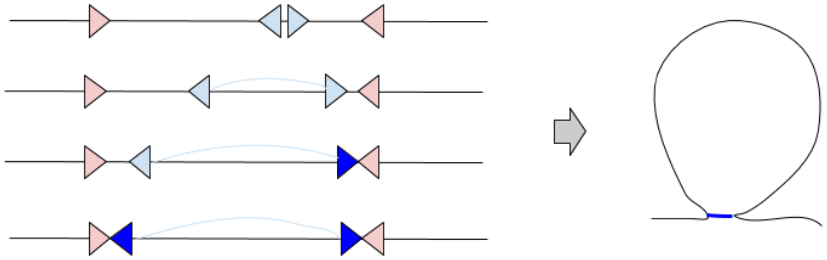

Figure 21. Left: Cohesins (blue) stalling on CTCF boundaries (red). Stalled legs are depicted in dark blue. Right: the resulting loop with a maximum size.

The cohesin trajectory accounting for CTCFs is applied to simulations. This results in visible TADs, stripes, and loops in the contactmap (Figure 22). This shows that the loop extrusion model does a good job of replicating features seen in Hi-C.

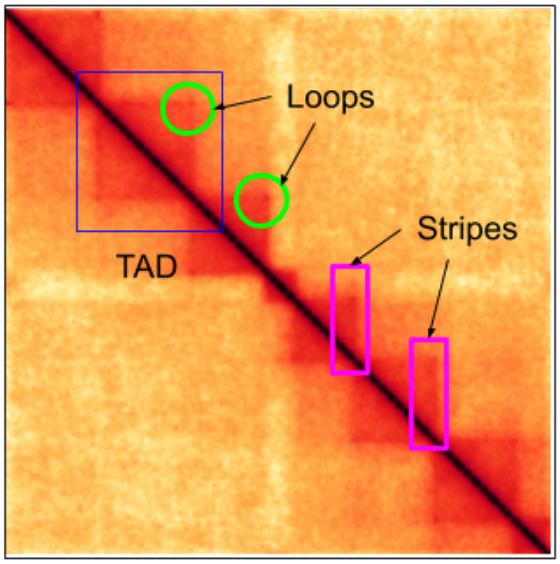

Figure 22. TADs formation in a simulated contactmap. The TAD boundaries prevent extrusion in either direction and probabilistically contain loaded CTCFs.

Conclusion

I have demonstrated that the H3K9me3 histone modification is linked to CTCF deactivation and stronger compartmentalization and that the effect of H3K9me3 can be modeled in high-resolution simulations. The H3K9me3 modification correlates with the deactivation of CTCF

sites and changes to compartmentalization patterns, which are thought to be separate processes.

Furthermore, this paper advances the study of genome structure by improving the ability of researchers to replicate Hi-C maps in silico. For example, the method to iteratively fit a compartment model to a Hi-C map allows for the complete replication of compartmentalization structure in Hi-C. This enables the exploration of more advanced loop extrusion and compartmentalization rules, which may help us better understand the molecular processes that regulate genome structure.

Acknowledgments

I would like to thank my mentor, Maxim Imakaev, for teaching me about molecular dynamics simulations and Hi-C and guiding me throughout this project. Additionally, I would like to thank George Spracklin and Nezar Abdennur, who suggested this project and gave me advice throughout. I am working with them on preparing a research paper of which this work will be a part. Sameer Abraham mentored me last year and taught me about compartmentalization in the genome, which was the main building block for this project. Finally, I thank Professor Leonid Mirny at MIT for providing me with computational resources and guidance, and the MIT PRIMES program for enabling me to do computational biology research.

Bibliography

Chowdhury, Neil, and Sameer Abraham. 2019. “A Method to Recognize Universal Patterns in Genome Structure Using Hi-C.” *MIT PRIMES Research Papers*. https://math.mit.edu/research/highschool/primes/materials/2019/Chowdhury.pdf.

Di Stefano, Marco, Jonas Paulsen, Tonje G. Lien, Eivind Hovig, and Cristian Micheletti. 2016. “Hi-C-Constrained Physical Models of Human Chromosomes Recover Functionally-Related Properties of Genome Organization.” *Scientific Reports* 6 (October): 35985.

Eastman, Peter, Jason Swails, John D. Chodera, Robert T. McGibbon, Yutong Zhao, Kyle A. Beauchamp, Lee-Ping Wang, et al. 2017. “OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics.” *PLoS Computational Biology* 13 (7): e1005659.

Falk, Martin, Yana Feodorova, Natalia Naumova, Maxim Imakaev, Bryan R. Lajoie, Heinrich Leonhardt, Boris Joffe, et al. 2019. “Heterochromatin Drives Compartmentalization of Inverted and Conventional Nuclei.” *Nature* 570 (7761): 395–99.

Fudenberg, Geoffrey, Maxim Imakaev, Carolyn Lu, Anton Goloborodko, Nezar Abdennur, and Leonid A. Mirny. 2016. “Formation of Chromosomal Domains by Loop Extrusion.” *Cell Reports* 15 (9): 2038–49.

Jacinto, Filipe V., Esteban Ballestar, Santiago Ropero, and Manel Esteller. 2007. “Discovery of Epigenetically Silenced Genes by Methylated DNA Immunoprecipitation in Colon Cancer Cells.” *Cancer Research* 67 (24): 11481–86.

Lajoie, Bryan R., Job Dekker, and Noam Kaplan. 2015. “The Hitchhiker’s Guide to Hi-C Analysis: Practical Guidelines.” *Methods* 72 (January): 65.

Lieberman-Aiden, Erez, Nynke L. van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragoczy, Agnes Telling, Ido Amit, et al. 2009. “Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome.” *Science* 326 (5950): 289–93.

Nuebler, Johannes, Geoffrey Fudenberg, Maxim Imakaev, Nezar Abdennur, and Leonid A. Mirny. 2018. “Chromatin Organization by an Interplay of Loop Extrusion and Compartmental Segregation.” *Proceedings of the National Academy of Sciences of the United States of America* 115 (29): E6697–6706.

Phillips, Jennifer E., and Victor G. Corces. 2009. “CTCF: Master Weaver of the Genome.” *Cell* 137 (7): 1194–1211.

Rao, Suhas S. P., Su-Chen Huang, Brian Glenn St Hilaire, Jesse M. Engreitz, Elizabeth M. Perez, Kyong-Rim Kieffer-Kwon, Adrian L. Sanborn, et al. 2017. “Cohesin Loss Eliminates All Loop Domains.” *Cell* 171 (2): 305–20.e24.

Rao, Suhas S. P., Miriam H. Huntley, Neva C. Durand, Elena K. Stamenova, Ivan D. Bochkov, James T. Robinson, Adrian L. Sanborn, et al. 2014. “A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping.” *Cell* 159 (7): 1665–80.

Sanborn, Adrian L., Suhas S. P. Rao, Su-Chen Huang, Neva C. Durand, Miriam H. Huntley, Andrew I. Jewett, Ivan D. Bochkov, et al. 2015. “Chromatin Extrusion Explains Key Features of Loop and Domain Formation in Wild-Type and Engineered Genomes.” *Proceedings of the National Academy of Sciences of the United States of America* 112 (47): E6456–65.

Schwarzer, Wibke, Nezar Abdennur, Anton Goloborodko, Aleksandra Pekowska, Geoffrey Fudenberg, Yann Loe-Mie, Nuno A. Fonseca, et al. 2017. “Two Independent Modes of

Chromatin Organization Revealed by Cohesin Removal." *Nature* 551 (7678): 51–56.
Seitan, Vlad C., Andre J. Faure, Ye Zhan, Rachel Patton McCord, Bryan R. Lajoie, Elizabeth Ing-Simmons, Boris Lenhard, et al. 2013. "Cohesin-Based Chromatin Interactions Enable Regulated Gene Expression within Preexisting Architectural Compartments." *Genome Research* 23 (12): 2066.
Spracklin, George, Nezar Abdennur, Maxim Imakaev, Neil Chowdhury, Sriharsa Pradhan, Leonid Mirny, and Job Dekker. 2020. "DNA Methylation Is Required to Maintain Heterochromatin Architecture in Colon Cancer Cells." Manuscript in preparation.
Szabo, Quentin, Frédéric Bantignies, and Giacomo Cavalli. 2019. "Principles of Genome Folding into Topologically Associating Domains." *Science Advances* 5 (4): eaaw1668.
Vian, Laura, Aleksandra Pękowska, Suhas S. P. Rao, Kyong-Rim Kieffer-Kwon, Seolkyoung Jung, Laura Baranello, Su-Chen Huang, et al. 2018. "The Energetics and Physiological Impact of Cohesin Extrusion." *Cell* 173 (5): 1165–78.e20.

本参赛团队声明所提交的论文是在指导老师指导下进行的研究工作和取得的研究成果。尽本团队所知，除了文中特别加以标注和致谢中所罗列的内容以外，论文中不包含其他人已经发表或撰写过的研究成果。若有不实之处，本人愿意承担一切相关责任。

参赛队员：Neil Chowmy　　指导老师：

2020/9/13

年　　月　　日