Analysis of Week 12's Paper

Chromosal Organization through Loop Extrusion

Minghang Li

Abstract

Week 12's lecture discussed the paper by Fudenberg et. al., titled by *"Formation of Chromosomal Domains by Loop Extrusion"*. It investigated the underlying mechanism of the formation of one of the most common and important chromosomal structure – Topologically Associating Domains (TADs). It proposed the hypothesis that TADs are formed by loop extrusion regulated by Boundary Elements (BEs) and Loop Elongation Factors (LEFs) and tested the hypothesis through rigorous modelling and simulation.

1 Background Introduction

1.1 Chromosomal Conformation Capture (3C) Technologies

The genetic information stored in DNA sequences is not presented in a linear manner, but is rather folded and organized into three-dimensional (3D) structures, thereby allowing remotely located genomic elements interact with each other. To understand the 3D spatial genome organization better, various techniques have been developed over the years. Beyond the conventional microscopy observation that uncovered some basic principles about genome microscopy, the development of chromosome conformation capture (3C) technologies has also produced many important insights into the chromatin structure [1].

The 3C technology was first introduced in 2002 by Dekker et. al. originally in the hope of studying the folding of yeast chromosome [2]. It quickly inspired the emergence of an abundance of 3C-derived genomics methods. Fig. 1 shows an overview of the 3C-derived technologies, which all start with the same steps. First, the DNA is crosslinked using a fixative agent, most often formaldehyde, to preserve the 3D representation of the genome. Second, the crosslinked DNA is digested with restriction enzymes to generate DNA fragments of a certain size. The DNA fragments are then ligated to specific adapters and amplified by PCR. The amplified DNA fragments are then sequenced and the sequence information is used to infer the spatial relationship between the captured DNA fragments [3]. Amongst all the 3C-derived technologies, Hi-C is probably the most widely used one. In Hi-C, the restriction ends of DNA fragments are filled in with biotin labeled nucleotides before the ligation step. Following a blunt ligation (usually by T4 DNA ligase), the biotin-labeled DNA fragments are purified and then captured by streptavidin-coated magnetic beads. After biotin-pulldown, the fragments are sheared into reads and then mapped back to the reference genome to generate a contact map, which is then used to score and infer the 3D structure of the genome [3]. Biotin pulldown and paired-end sequencing enables the detection of ligation junctions, and Next Generation Sequencing (NGS) methods makes such "all versus all" detection at a genome-wide scale possible.

1.2 Two Major Classes of Chromosomal Patterns

Hi-C is capable of capturing the 3D structure of the genome at a resolution of kb level. In the Hi-C contact maps, two major classes of patterns are most evident. The first one is the A/B compartmentalization which shows checkerboard-like patterns [5]. The A compartment is usually
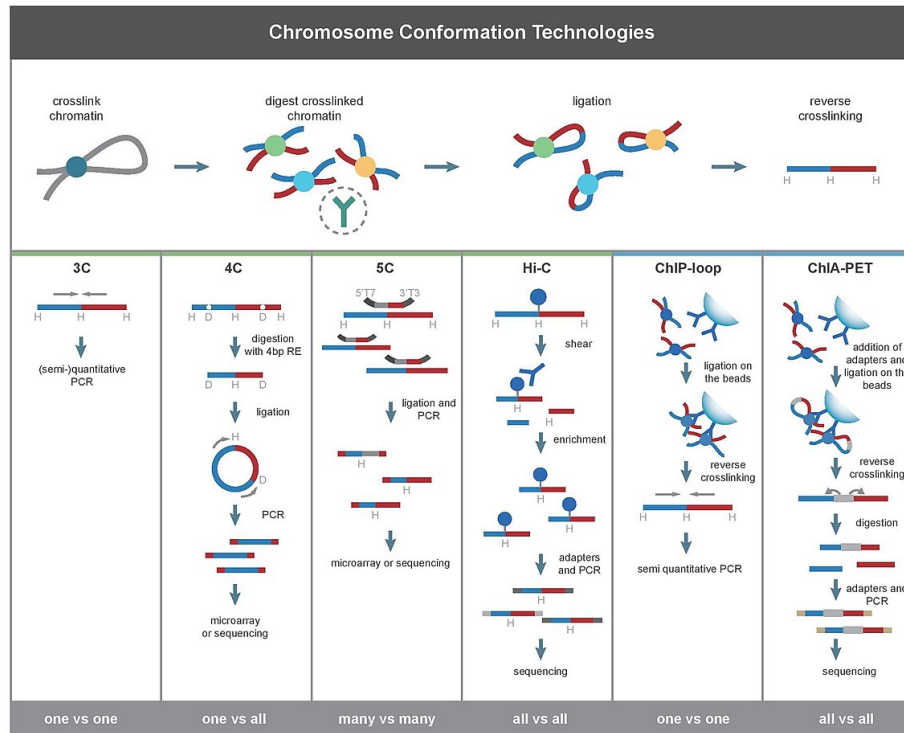
Figure 1: **An Overview of the 3C Technologies.**This figure comes from [4], which was adapted from [3]

gene-rich, transcriptionally active and accessible (characterized by DNase I sensitivity), and the B compartment is usually the heterochromatin region that is low in expression and has repressive histone marks [1][3]. The second one is the Topologically Associating Domains (TADs) which are characterized as continuous regions of mildly (2-4 folds) higher contact frequency than between loci in neighboring TADs [5]. They usually reside within a single contiguous compartment and don't necessarily exhibit the characteristic "checkerboard" pattern as A/B compartment does [6]. 50% of TADs have peaks of interactions at their boundaries, and they are capable of forming dynamic boundaries [7]. Up until the publishing of this paper, the formation of TADs was still not well understood. Previous works have proposed TADs forming mechanism based on similar monomer attraction [8] and supercoiling [9], yet they fail to explain the formation of TADs in a satisfactory manner.

2   Brief Summary of the Model

To give out a satisfactory explanation for TAD formation, the authors proposed a loop-extrusion based mechanism. In this process, TADs are formed through the dynamic introduction of chromatin loop by *cis*-acting loop-extruding factors (LEFs) and are bounded by boundary elements (BEs). The loop extrusion process was modelled by coupling the 1D dynamics of LEFs with 3D polymer dynamics. The following subsections will give a concise summary of the details of the proposed mechanism.

2.1   Basic Settings of the 1-D Model

The authors first defined the dynamics of LEFs with BEs. The translocation of LEFs along the chromatin fiber was simulated on a 1D lattice, where each position was characterized by association (birth) probability, dissociation (death) probability and BE occupancy (stalling prob-
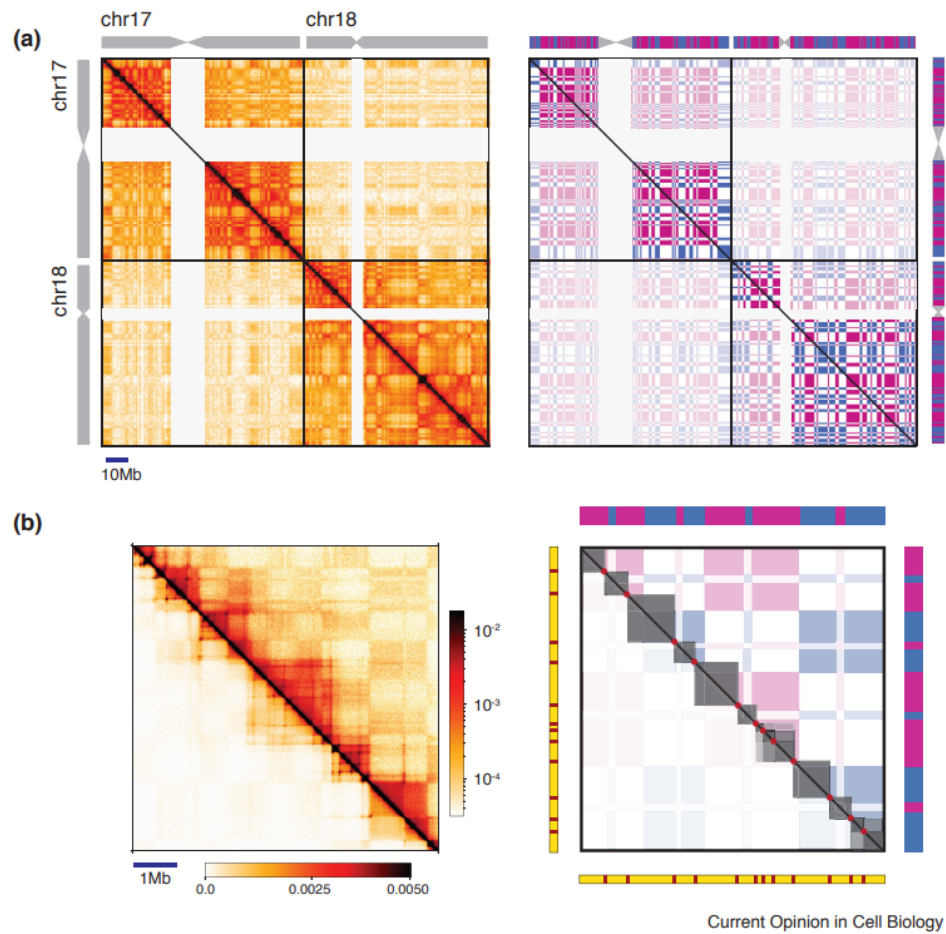
Figure 2: **Hallmark Patterns in Hi-C Maps: Compartmentalization and TADs.** This figure comes from [5]. Fig (a) shows the A/B compartmentalization pattern, and Fig (b) shows the TADs pattern.

ability). The LEF molecule was modelled by "two heads" connected with a linker, analogous to SMC protein complexes. Each head of the LEF occupied one lattice position and cannot be overlapped except for at birth events. At each time-step, each LEF head translocates to the neighboring lattics position in opposite direction until encountering another LEF head or a boundary element (BE). The stalling of one LEF head does not affect the translocation of the other LEF head. At each discrete unit of time, a LEF undergoes dissociation with a probability that is equal to the maximum value of the death probability. For the occurrence of birth events, the two heads of the LEF are either initiated on the same lattice location or on neighboring sites (when the site immediately to the right of the selected site is unoccupied) with a probability of 0.5. The dynamics of LEFs were depicted in Fig. 3.
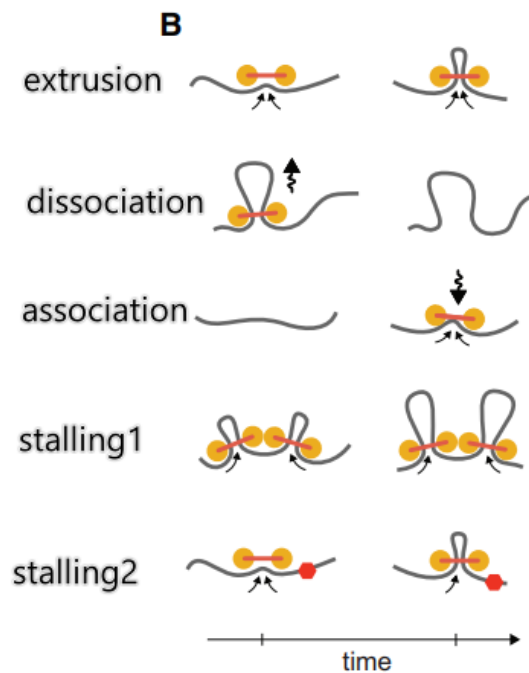

Figure 3: **The Dynamics of LEFs.** This figure comes from [6].

2.2   Basic Settings of the 3-D Polymer Model

As is later shown in the paper, the 1-D model can very well depict the TAD as an average of experiments from many cells, yet it is difficult to model insulation solely by the 1-D model. For example, an enhancer and promoter separated by 100kb can still come in contact in the 3D space. To account for this, the authors proposed a 3D polymer model to simulate the 3D structure of the chromatin fiber.

The 3D chromatin model is a polymer connected by 10-nm monomers (roughly the size of 3 nucleosomes or 600bp) subject to Langevin dynamics, which according to my understanding models the stochastic behavior of molecules in solvent. The monomers are further modelled by excluded volume interactions, which prohibits monomer overlapping, and are linked with harmonic bonds that restrict their relative position but allow for rotation in certain degrees (stiffness) and chain passing (which represents the activity of topoisomerase and enables loop twisting and loop inside loops).

2.3 Connecting the 1-D and 3-D Models

LEFs serve as the connection points between 1-D and 3-D models: they affects the monomers of the 3-D model by introducing additional harmonic bond between the monomers held by the two heads of the LEFs. The bond is constantly reassigned to increasingly separated pairs of monomers as the LEF heads move along the chromosome. Furthermore, before starting the 3D simulation, the 1D LEF dynamics was first run for 4 million steps to achieve convergence of LEF initial positions. At each 1-D simulation time step, according to the hyperparameter *3D-to-1D dynamics*, the corresponding 3D Langevin dynamics simulations were performed before starting the 1D-simulation time steps.

2.4 Important Parameters of the Model As A Whole

To simplify the simulation effort and to account for the still-unknown biological setup of exact LEF dynamics parameters, the authors chose to first use a simple setting with uniform birth probability, constant death probability, a fix number of LEFs and a discrete nubmer of impermeable BEs. The parameters have been relaxed in latter explorations such as the permeability of BEs, or even different types of BEs. Under such scenario, the LEF dynamics can be well described by only two metrics: the LEF processivity ($\frac{2}{\text{death rate}}$) and LEF separation ($\frac{\text{total number of lattice sites}}{\text{number of LEFs}}$). The processivity accounts for the average size of the loop extruded by an LEF if not stopped early; and the separation represents the distance between two pairs of LEFs on average.

In the simulations, the authors tested various combinations of processivity and separation and analyzed the effects of these parameters on the shape of the TADs.

3 Brief Summary of the Results

3.1 Evaluation of the Model

The authors used a measurement called $P(s)$ (a function of genomic distance s) to test the ability of the model to reproduce Hi-C contact map data. They plotted $s - P(s)$ curves for the model and the experimental data they obtained and confirmed that the shapes are similar. To measure the reproduction quantitatively, they also determined the *goodness of fit* using the geometric standard deviation of the ratios of the experimental data nad simulated $P(s)$ curves. Using such measures, the authors identified the best parameter settings (processivity and separation) to reproduce the experimental data.

3.2 Proposed Candidates for BEs and LEFs

It is natural to look for the possible molecular candidates for LEFs and BEs after building a successive model. As is summarized by the authors in the paper, it is highly likely that cohesin is a possible LEF in interphase. Cohesin is a Strutural Maintenance of Chromosomes (SMC) complex that have been proposed to be able to extrude loops of chromatin fiber and recently been proved to be able to move along the chromatin. They are also observed to be enriched at TAD boundaries and corner peaks, and removal of cohesin leads to the loss of TADs.

CCCTC-binding factor (CTCF) is a possible candidate for BEs. Similar to cohesin, CTCFs are also enriched at TAD boundaries and corner peaks, and the depletion of CTCF sleas to similar effect as cohesin depletion. Furthermore, CTCF is known to be able to interact with cohesin, especially in an orientation-dependent manner. More interestingly, inward-oriented CTCF sites are observed to occur at high frequency at TAD boundaries and corner-peaks, which is consistent with the model that CTCF is a BE.
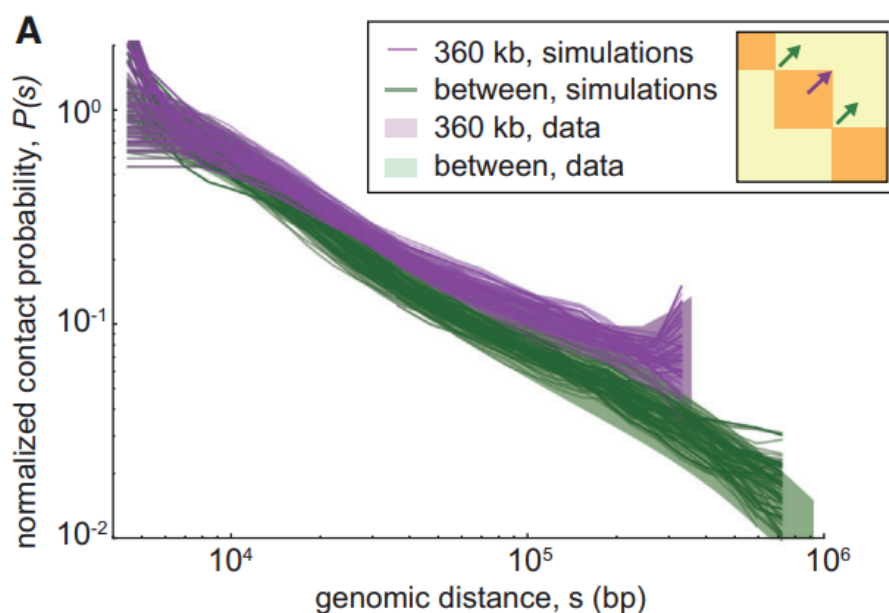
Figure 4: **Evaluation of Model.** This figure comes from [6]. Experimental $P(s)$ (shaded areas) versus simulated $P(s)$ for the 100 best fitting parameter sets (lines, one per parameter set) within TADs (purple) and between TADs (green).

3.3   The Effect of Processivity and Separation

As is shown in Fig. 5, the authors evaluated for the best parameters for the model as well as exploring the effect of changing the processivity and separation. The best agreement with Hi-C data is achieved for processivity of 120-240 kb and separation of 120 kb. It also shows that processivity affects the strength of the corner peak, while separation affects the strength of the TAD cliques as well as width of the TAD. The effect of processivity is especially natural since the larger the LEF processivity, the higher the probability of two LEFs to reach the BEs, hence the higher contact probability of the fragments at the end of TADs, resulting in the stronger corner peaks.

3.4   Exploration of Alternative Hypotheses

As is mentioned in the introduction, there are several other plausible hypotheses of interest. To prove the superiority and generality of the model, the authors investigated a few of alternative hypotheses and showed that they more or less failed to faithfully reproduce TADs in the simulation.

Investigate TAD formation Via Single Stable Loop

As is shown in Fig. 6, it is clearly shown that single stable loops are incapable of producing TADs. Conspicuously, the single table loops produce exceedingly strong corner peaks and no intermediate interactions that gives the characteristic cliques in the Hi-C contact map.

Investigate TAD formation Via Solely Boundary Element Dynamics

The model proposed by the authors do not rely on the direct physical blockade of LEFs by BEs to enable insulation between TADs, instead, it relies on the BEs to regulate the translocation of LEFs. However, it is possible that direct physical insulation of BEs can lead to the formation of
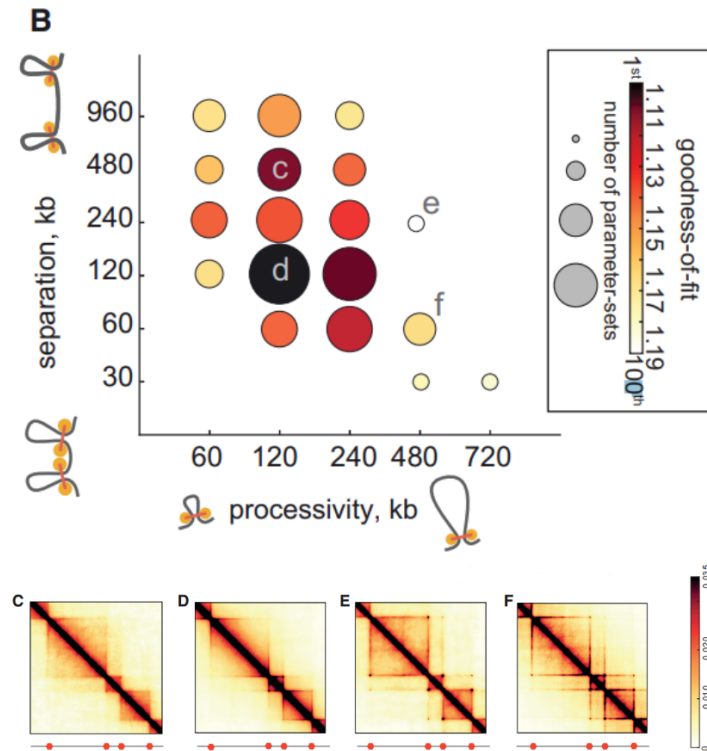
Figure 5: **The Effect of Processivity and Separation.** This figure comes from [6]. Figure C, D, E, F corresponds to the parameter set c, d, e, f in figure B, respectively.
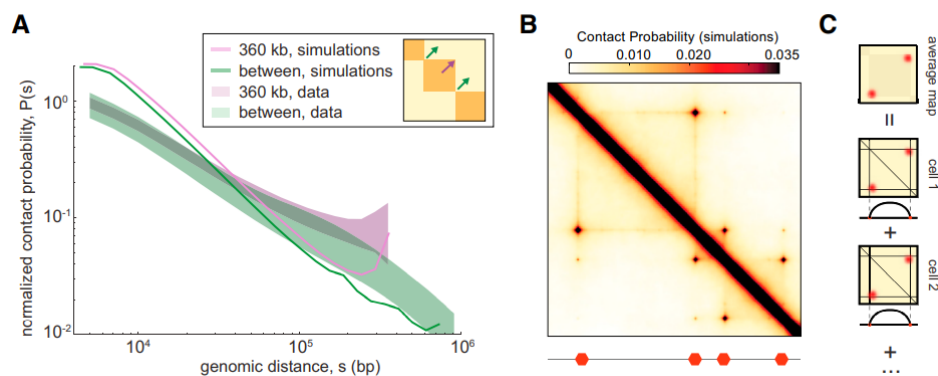

Figure 6: **Investigate TAD formation Via Single Stable Loop.** This figure is exactly Figure 4 in [6]. It shows that single stable loops are incapable of producing TADs.

TADs. To rule out the possibility that TADs simply result from the existence of bulky objects at boundaries, the authors tested such hypothesis in the simulation. They also investigated the effect of placing stiff chromatin fiber at BE location, which has similar effect as bulky BEs. The dimerization of BEs can also potentially lead to the formation of TADs – this represents the scenario that proteins are capable of selectively bind specific genomic elements.

As is shown in Fig. 7, it is very obvious that no TADs are formed at all when BEs are modelled by either bulky objects or stiff chromatin fibers. The direct BE-to-BE mechanism, despite showing some interaction between distal loci, cannot distinguish between chromosomal regions, and result in uniform peaks of contact probability accross all the BE pairs. It also have zero separating effect between TADs, resulting in no characteristic clique shape of TADs at all.
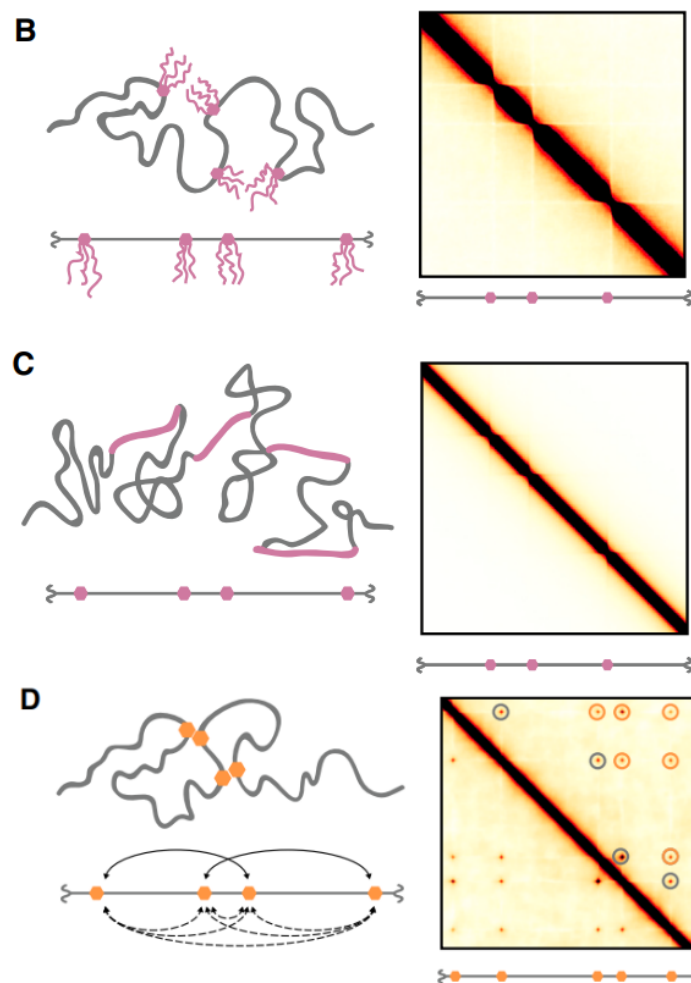

Figure 7: **Investigate TAD formation Via Solely Boundary Element Dynamics.** This figure is part of Figure 5 in [6]. Fig B models BE as bulky objects, Fig C models BE as stiff chromatin fibers, Fig D models the BE-to-BE TAD formation mechanism.

The simulation of the alternative hypotheses stress on the importance of the *cis*-acting manner of LEFs along the chromatin as well as sophisticated interaction between BEs and LEFs.

3.5   Exploration of Boundary Elements Properties

It has been shown in literature that flipping solely the orientation of BEs (to be exact, the proposed molecular BE candidate, CTCFs) can lead to the merging of two TADs.

The authors analyzed cohesin ChIP-seq peaks and found that CTCFs were enriched in exact orientation dependent manner around strongly bound CTCF peaks, suggesting CTCF indeed acts as a BE that impedes loop extrusion in an orientation-dependent manner – at strong CTCF binding sites, cohesins are enriched only in one direction but not the other. In weak CTCF binding sites, cohesins are enriched in both directions.
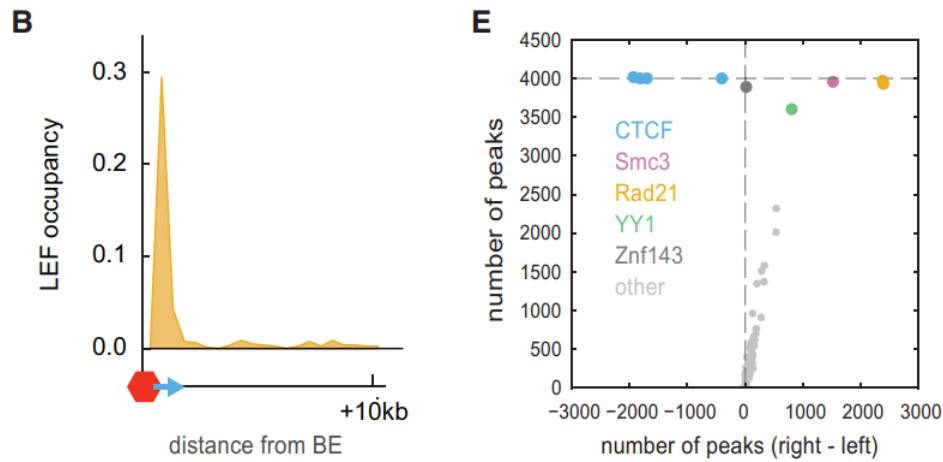

Figure 8: **The Orientation Dependent Accumulation of Cohesin.** This figure is part of Figure 6 in [6]. Fig B shows the theoretical explanation of the orientation-dependent enrichment of cohesin. Fig E shows the experimental data of cohesin ChIP-seq peaks at strong CTCF binding sites.

Furthermore, with the model at hand, the authors modify the BE permeability and directionality to try to recapitulate the effect of CTCF on TAD boundary. To correctly model and reproduce the experimental data, the authors converted ChIP-seq data for CTCF into BE permeability and directionality parameters with the help of cohesin peak and logistic transformation. According to the simulation and the analysis by the authors, the model to some extent reproduce and predicts the effect of CTCF on TAD boundary, yet agreement along the chromosome was non-uniform. However, the results do reflect that the increased permeability of BEs can lead to the emergence of large, homogeneous TADs, suggesting the importance of BEs in the formation of TADs and further supporting the model.

4   Strengths and Weaknesses of the Paper

Strengths

The paper proposed and examined a model for an important 3D chromosomal structure and showed that the model is capable of reproducing the formation of TADs in the simulation. The model is also capable of predicting the effect of BEs on TADs, which is a very important step towards understanding the mechanism of TAD formation.

To the best of my knowledge, this is the first paper that proposed a model for TAD formation with detailed and rigorous simulation as well as predictive power. The model is also very simple and easy to understand, which makes it a very good starting point for further research on TAD formation.

Weaknesses

Since this is one of the pioneering work, parts of the model still needs refinement and describe. For example, it is likely that the association possibility and dissociation possibility are not
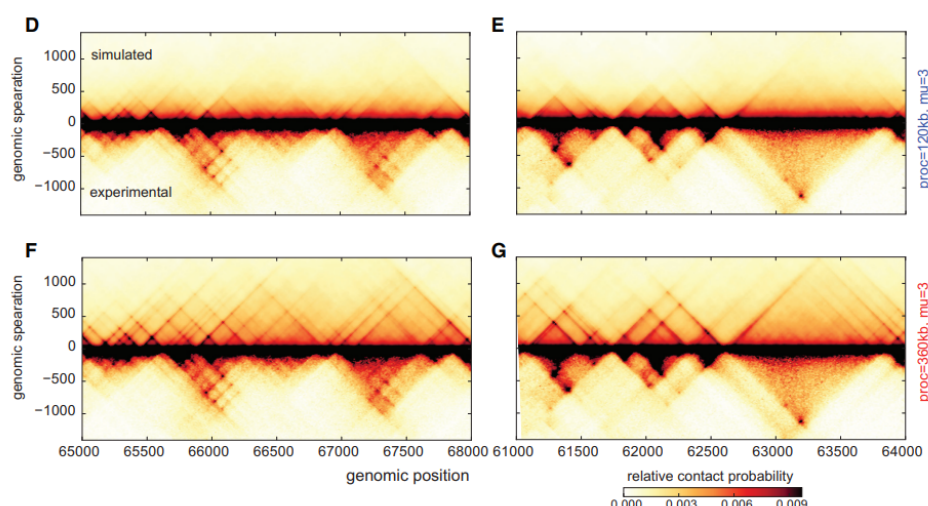
Figure 9: **Simulation of Changing TAD Boundaries with Different BE Permeability.** This figure is part of Figure 7 in [6]. LEF processivity is 120 kb (D and E) and 360 kb (F and G).

homogeneous across the genome. The ChIP-seq data of CTCF enrichment peaks were also not well modelled and reproduced in the simulation.

5 Future Prospectives

As is mentioned in the weaknesses, the models can be further refined to incorporate more subtleness in the chromatin conformational change. But the most important future work is to test the model in real biological systems, such as testing the proposed molecular candidate of BEs and LEFs. And in a paper published in 2019 [10], the loop extruding effect of cohesin has been proved, strongly supporting the model proposed in this paper.

References

[1] A. Denker and W. d. Laat, "The second decade of 3c technologies: Detailed insights into nuclear organization," *Genes Dev.*, vol. 30, no. 12, pp. 1357–1382, Jun. 15, 2016, Company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Institution: Cold Spring Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press Publisher: Cold Spring Harbor Lab, ISSN: 0890-9369, 1549-5477. DOI: 10. 1101/gad.281964.116. [Online]. Available: http://genesdev.cshlp.org/content/30/12/1357 (visited on 01/13/2023).

[2] J. Dekker, K. Rippe, M. Dekker, and N. Kleckner, "Capturing chromosome conformation," *Science*, vol. 295, no. 5558, pp. 1306–1311, Feb. 15, 2002, Publisher: American Association for the Advancement of Science. DOI: 10.1126/science.1067799. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.1067799 (visited on 01/13/2023).

[3] E. d. Wit and W. d. Laat, "A decade of 3c technologies: Insights into nuclear organization," *Genes Dev.*, vol. 26, no. 1, pp. 11–24, Jan. 1, 2012, Company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Institution: Cold Spring Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press Publisher: Cold Spring Harbor Lab, ISSN: 0890-9369, 1549-5477. DOI: 10.1101/gad.179804.111.

[Online]. Available: http://genesdev.cshlp.org/content/26/1/11 (visited on 01/13/2023).

[4] G. Li, L. Cai, H. Chang, *et al.*, "Chromatin interaction analysis with paired-end tag (ChIA-PET) sequencing technology and application," *BMC Genomics*, vol. 15, no. 12, S11, Dec. 19, 2014, ISSN: 1471-2164. DOI: 10.1186/1471-2164-15-S12-S11. [Online]. Available: https://doi.org/10.1186/1471-2164-15-S12-S11 (visited on 01/13/2023).

[5] L. A. Mirny, M. Imakaev, and N. Abdennur, "Two major mechanisms of chromosome organization," *Current Opinion in Cell Biology*, Cell Nucleus, vol. 58, pp. 142–152, Jun. 1, 2019, ISSN: 0955-0674. DOI: 10.1016/j.ceb.2019.05.001. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0955067418301601 (visited on 01/13/2023).

[6] G. Fudenberg, M. Imakaev, C. Lu, A. Goloborodko, N. Abdennur, and L. A. Mirny, "Formation of chromosomal domains by loop extrusion," *Cell Reports*, vol. 15, no. 9, pp. 2038–2049, May 31, 2016, ISSN: 2211-1247. DOI: 10.1016/j.celrep.2016.04.085. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2211124716305307 (visited on 01/13/2023).

[7] S. S. P. Rao, M. H. Huntley, N. C. Durand, *et al.*, "A 3d map of the human genome at kilobase resolution reveals principles of chromatin looping," *Cell*, vol. 159, no. 7, pp. 1665–1680, Dec. 18, 2014, Publisher: Elsevier, ISSN: 0092-8674, 1097-4172. DOI: 10.1016/j.cell.2014.11.021. [Online]. Available: https://www.cell.com/cell/abstract/S0092-8674(14)01497-4 (visited on 01/13/2023).

[8] D. Jost, P. Carrivain, G. Cavalli, and C. Vaillant, "Modeling epigenome folding: Formation and dynamics of topologically associated chromatin domains," *Nucleic Acids Research*, vol. 42, no. 15, pp. 9553–9561, Sep. 2, 2014, ISSN: 0305-1048. DOI: 10.1093/nar/gku698. [Online]. Available: https://doi.org/10.1093/nar/gku698 (visited on 01/13/2023).

[9] F. Benedetti, J. Dorier, Y. Burnier, and A. Stasiak, "Models that include supercoiling of topological domains reproduce several known features of interphase chromosomes," *Nucleic Acids Research*, vol. 42, no. 5, pp. 2848–2855, Mar. 1, 2014, ISSN: 0305-1048. DOI: 10.1093/nar/gkt1353. [Online]. Available: https://doi.org/10.1093/nar/gkt1353 (visited on 01/13/2023).

[10] Y. Kim, Z. Shi, H. Zhang, I. J. Finkelstein, and H. Yu, "Human cohesin compacts DNA by loop extrusion," *Science*, vol. 366, no. 6471, pp. 1345–1349, Dec. 13, 2019, Publisher: American Association for the Advancement of Science. DOI: 10.1126/science.aaz4475. [Online]. Available: https://www.science.org/doi/10.1126/science.aaz4475 (visited on 01/13/2023).