Essay on Fudenberg et al.:

Formation of Chromosomal Domains by Loop Extrusion [1]

Jennifer Probst
ETH Zürich: 16-703-423
jennifer.probst@stud.unibas.ch

Submitted to Prof. Richard Neher
January 10th, 2021

1 Introduction

More and more studies in the recent years have emphasized the importance of spatial chromatin organization in the regulation of gene expression [2; 3; 4; 5]. Thereby, it has been shown, that chromatin conformation in vivo is not an equilibrium process and chromatin is constantly remodelled through various modifications of histones and DNA [6]. Two major mechanisms of chromatin organization have been reported [2]. First, an ATP-dependent organization of self-interacting regions (topologically associated domains, TADs) on megabase (Mb) scale [7; 8] and secondly the spatial compartmentalization on chromosomal scale (10s-100s Mb) into active (euchromatin) and inactive chromatin (heterochromatin) as a result of affinity based phase separation. The second mechanism has been widely studied [9], but until recently there was no agreement on the formation of TADs.

The paper by Fudenberg et al. [1] discusses as one of the first articles a mechanism called loop extrusion as an underlying mechanism for TAD formation. Loop extrusion has been mentioned before [11; 12], but never in connection with the formation of TADs, and effects of loop extrusion on 3D spatial organisation have never been examined in greater detail. Furthermore, the loops formed between pairs of boundary loci have been viewed as stable. In their paper, the authors present evidence from simulations alongside discussing experimental observations [10] that speak against this view and argue that emergence from multiple loops formed through dynamic loop extrusion is the true underlying mechanism of TAD formation. Main actors in the proposed mechanism are cis-acting loop-extrusion factors (LEFs) that form chromatin loops and thereby enable interactions between regions of the chromatin that are distant in 1D. The maximal loop size in the process is determined by boundary elements (BEs), that stall LEFs.

The authors tried to verify their hypothesis through polymer simulations, compared their results to experimental Hi-C data [10] and made specific predictions for future experiments.

2 Methods

2.1 Minimal 3D Model

Loop extrusion dynamics by LEFs were simulated in a 1D lattice where LEFs could associate or dissociate and each position was defined by an association (birth) probability, dissociation (death) probability, and BE occupancy (stalling probability). When LEFs bound the chromatin fiber in the model, they started extruding a loop by forming a bond between monomers at the two ends of an extruded loop, and re-assigned the bond to increasingly separated pairs of monomers as a LEF translocates along the chromosome in both directions. This translocation was stopped upon encountering a BE or another LEF, or through stochas-

tic dissociation. LEF dynamics were described by two parameters: LEF processivity, which determines the average time LEFs stay on the chromatin once bound and the average linear separation between two bound LEFs.

These simulations of 1D loop extrusion dynamics were coupled with 3D polymer dynamics, which allowed modelling 3D interactions between TADs and the rest of the loop, and reduced self-avoidance effects in the random walk simulation. A polymer of 10nm monomers was used to model the chromatin fiber and the model was run using OpenMM subject to Langevin dynamics with exclusion of volume interactions and topological constraints. The genome sequence used for simulations was about 1 Mb long and roughly 15 Mb for the experiments; which matches the genomic length scale of TADs. BEs were set at fixed positions (180kb, 360kb and 720kb) along the polymer to investigate patterns for TADs of different sizes.

The discrete model for 1D LEF dynamics thereby imposed a system of bonds on the simulated 3D polymer dynamics, which allowed to generate an ensemble of chromosome conformations and an average simulated contact map for each parameter set of LEF processivity and average linear separation values. This led to the opportunity to compare the simulated contact map with experimental Hi-C contact maps.

2.2 Hi-C

The Hi-C method [13] based on Chromosome Conformation Capture is used to detect 3D chromatin interactions inside the nucleus and thereby characterize large scale chromatin organization. Extending previous Chromosome Conformation Capture techniques through extensive parallel sequencing, this method enables unbiased identification of interactions between chromatin sites in the entire genome. Therefore, pieces of DNA being in close 3D proximity in the nucleus, are ligated together. Thereafter, they are pulled down and amplified for high throughput sequencing, which is then used to find the nucleotide sequences of the fragments. This gives us a conformation capture for the entire genome and allows to quantify the average number of interactions between genomic loci that were nearby in 3D space in the nucleus.

This information can then be stored in a contact matrix, where each entry represents the count of interactions between two loci, or alternatively be represented as a heat map with DNA coordinates along both axis and color intensity representing the interaction counts (see Figure 1). To obtain simulated Hi-C maps the contact maps of each block of two simulated TADs were summed up for each conformation obtained and row-wise normalized.

3 Results

3.1 Quantitative Analysis of Loop Extrusion: Initial Model Validation

The authors demonstrated in their simulations that loop extrusion through combined action of LEFs and BEs can lead to the formation of TADs (or contact enrichment within TADs) and finer-scale features of Hi-C data. This TAD formation could be detected for a range of LEF processivity and separation on simulated Hi-C maps (Figure 1). Depending on the exact parameter values they observed homogeneous TADs, the formation of corner peaks for high processivity or enrichment of contacts at the boundary of TADs for low separation values.

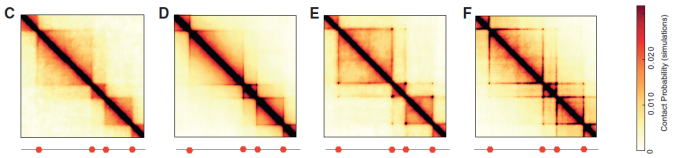

Figure 1: Simulated contact maps for some distinct processivity-separation values. [1]

In order to further validate the model, simulated Hi-C contact probability ($P(s)$) as a function of the linear genomic distance s was compared to experimental frequencies, which is a previously used measure for comparison and validation of polymer models [14; 15; 16]. The best agreement with experimental data was achieved for LEF separation of ∼120kb and LEF processivity of ∼120-240kb, determined by the goodness of fit as the ratio of simulated and experimental $P(s)$ for each parameter set. In Figure 2 the $P(s)$ for the 100 best fitting parameter sets is plotted alongside with experimental $P(s)$, separating within and between TAD regions for each $P(s)$. We observe that the simulated $P(s)$ match the experimental data within as well as between TAD regions.

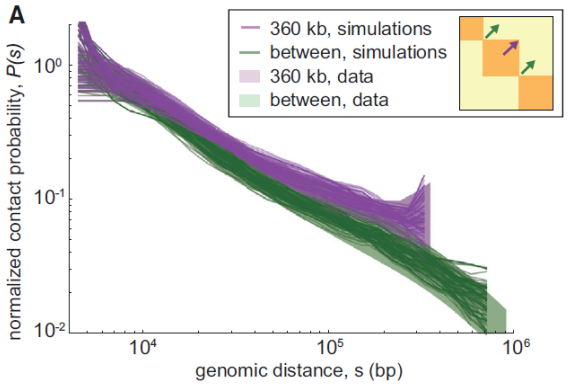

Figure 2: Experimental versus simulated P(s) for the 100 best fitting parameter sets calculated from Hi-C contact maps. [1]

Furthermore, the authors demonstrated that TADs arise as a feature of a population in averaged simulated Hi-C maps and are only visible in a fraction of contact maps directly. Additionally, upon deletion of a BE, TADs in their model spread until the next BE, a behaviour also observed in experiments. In summary, the model behaves as expected in terms of formation of TADs, fitting well to experimental data for a certain range of parameters.

3.2 Evidence for the Dynamic Loop Model

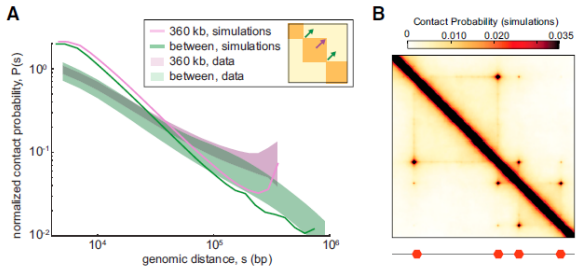

Figure 3: (A) $P(s)$ experimental vs. simulated for a strong loop parameter set. (B) Simulated contact map for a simple strong loop. [1]

As already mentioned in the introduction, TADs have popularly been viewed as stable loops formed between pairs of BEs. Figure 3 from the original publication [1] shows $P(s)$ of simulations generated with single strong loops between BEs. Interestingly, these simulations achieved some of the worst fits to Hi-C data besides lacking visible TADs. They therefore failed to generate TADs and explain the experimentally observed data. Less steep experimental curves indicate more evenly spread contacts along the chromosome in the TAD region and therefore argue against one single, static loop.

Additionally, similar $P(s)$ were observed for TADs with and without corner peaks, which are a feature of single, strong loops of constant size. It was furthermore observed, that no permanent contact between BEs was required to simulate TADs with visibly strong peaks and in their simulations the loop extrusion mechanism could produce TADs with and without corner peaks.

Therefore, the authors concluded that corner peaks seem to be an optional feature of TADs, resulting by collective activity of LEFs in the region between BEs.

3.3 The Role of Boundary Elements

3.3.1 Comparison of different BE Models

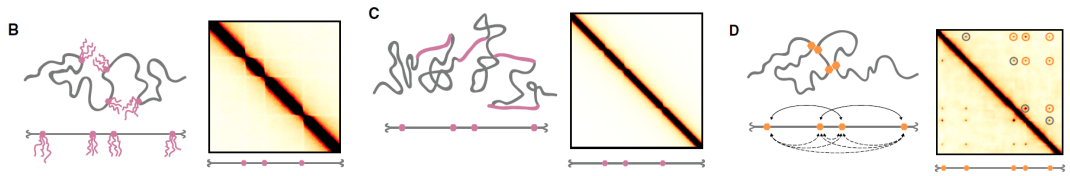

Figure 4: (a) Model of BEs as bulky object (b) Model of BEs as a stiff region of chromatin (c) Model with direct BE-to-BE attraction [1]

In the previously described model (chapter 2.1), insulation between neighboring TADs was assumed to emerge from LEF translocation regulation on a large spatial and genomic distance. Various simulations were performed to test further models for BEs (Figure 4). This included

large bulky objects acting as physical boundaries, BEs as a stiff region of chromatin and a model with direct BE-to-BE attraction. In scenario (a) and (b), insulation of directly neighboring regions was observed, but no long-range insulation and the models failed to obtain the TADs observed in experimental data. In the simulation of scenario (c) with directly interacting BEs, where any two BEs in close spatial proximity could interact, distinct peaks of contact probability were observed not only between adjacent BEs, but between all of them and insulation was negligible, which also did not fit the experimental data.

Together, these simulations highlight the role of LEFs stalled by BEs for imposing insulation at the scale of whole TADs and limiting interaction of loci to those within TADs.

3.3.2 Roles of Cohesin and CTCF

Having shown the importance of LEFs and BEs in TAD formation, the authors tried to identify possible molecular candidates for LEFs and BEs. Cohesin, which is a structural maintenance of chromosomes complex, is hypothesized to act as a LEF in interphase in vivo. It has been experimentally shown, that cohesins impact interphase TAD organization, chromatin looping and have motor protein activities. Moreover, these protein complexes have been observed to be enriched at interphase TAD boundaries [7] and corner peaks [10], and upon depletion TADs get less distinctive [17].

Fudenberg et al. argue that CTCF is an equivalently relevant candidate to be a boundary element. Similarly to cohesin, inward-oriented CTCF sites are also enriched at interphase TAD boundaries [18] and corner peaks [10]. Additionally, CTCF was observed to interact with cohesin in an orientation-dependent manner in immunoprecipitation experiments [19].

With these proposed roles of cohesin as a LEF and CTCF as a BE, predictions for outcomes of future experiments and simulations could be made.

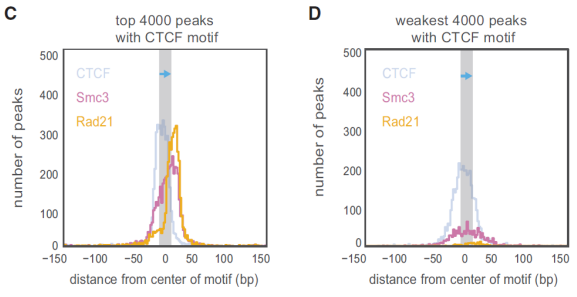

Figure 5: Distribution of CTCF ChIP-seq peaks around the 4000 strongest / weakest motif-associated CTCF binding peaks. The blue arrows indicate orientation of the CTCF sites. [1].

The authors predicted cohesin accumulation at CTCF-binding sites with bound CTCF, similar to LEF accumulation at BEs in their previous simulations. In ChIP-seq experiments (Figure 5) enrichment of cohesin peaks correlates with strongly bound CTCF peaks, in an orientation-dependent manner. Similarly, model predictions on cohesin and CTCF depletion and increased cohesin-binding time could be verified experimentally. Overall, the results lead to the conclusion that CTCF is a highly probable BE candidate with preferential orientation.

4 Discussion

The results of this paper can be summarized into two main findings namely the construction of the loop extrusion model (chapter 3.1, 3.2) and the identification as well as confirmation of prime candidates for LEFs and BEs (chapter 3.3). The authors showed compelling evidence that their model is able to form TADs agreeing to Hi-C data on a number of experimentally observed features such as enriched contact domains, corner peaks, finer structures like nested TADs, lines at TAD edges or grids of peaks. Furthermore, it succeeded in explaining the absence of interaction peaks of distant BEs or BEs on different chromosomes and demonstrated the merging of TADs observed in boundary deletion experiments. This model was the first of its kind modelling interaction peaks between BEs as dynamic loops instead of simple static loops, which showed much better agreement to experimental Hi-C and microscopy data. In the paper's discussion the authors compare their model results to results of a similar model published in a paper by Sanborn et al. [20] in the same year, which observed a different behaviour in terms of robustness of contact probabilities $P(s)$ to changes in simulation parameters. Fudenberg et al. explain these differences in that they considered a much wider range of LEF processivities and separations, which were not tested in the model of Sanborn et al.

This wide range of simulation parameters tested allowed the authors to eliminate different BE models and predict outcomes of future experiments. They collected strong evidence from consistent simulations and experiments for the probable roles of Cohesin as LEF and CTCF as BE candidate, that acts directional. Finally, they propose further experiments to examine alternative factors like additional protein complexes involved via mechanisms distinct from loop extrusion in megabase scale chromatin modulation and argue that these mechanisms have far-ranging consequences for many processes in the nucleus.

A classic limitation to all simulations is that every model only describes reality at a simplified level and that considered model parameters are only a subset of the actual parameters influencing the process of TAD formation. The effects of other parameters such as for example, the size of LEFs, could be included. Additionally, the length of genome sequences that were used for simulations and in experiments differed and BEs were only set up in the same three positions on the simulated polymer. These are all factors that could have an impact on the outcome of the simulations. Lastly, I would suggest simulations and experiments with chromatin sequences differing from human chromosome 14, to examine the generalizability of the loop extrusion model to the whole human genome.

In conclusion, the paper represents a landmark in the discovery process of chromatin organization mechanisms. Fudenberg et al. built a theoretical approach to experimentally interpret observed patterns of contact probabilities generated by Hi-C methods on an abstracted level. With verified model predictions in cohesin / CTCF enrichment and depletion experiments,

this model triggered a large number of follow-up experiments. Further research was conducted to characterize the molecular underpinnings of the loop extrusion mechanism.

5 Developments

Many of Fudenberg et al.'s findings [1] have been confirmed in recent years either through further simulations or experimentally. In the following paragraphs some selected findings are presented.

In a paper published in 2017 from Rao et al. [21], the loop extrusion model was utilized for further molecular dynamic simulations where the behaviour of a polymer was explored in a solvent containing extrusion complexes. Rao et al. could corroborate the observation that cohesin loss eliminates all loop domains. This was also validated experimentally in work from Nora et al. [22] in 2017 on mouse embryonic stem cells, where they found that CTCF depletion disrupted loops between TAD boundaries and disturbed the insulation of neighboring TADs. They further concluded from their experiments, that CTCF is a major blocking factor in the extrusion, which agrees with the loop extrusion model of Fudenberg et al. Additionally, they found chromatin structuring on chromosomal scale (above Mb scale) to be preserved and therefore to behave quite differently to the Mb level organization upon CTCF depletion. The authors concluded that insulation of chromosomes on a local scale becomes decoupled from the global compartmentalization when CTCF is degraded.

Identical observations could be made in cohesin removal experiments by Schwarzer et al. [23] in 2017, who observed two different reactions in chromatin organization upon cohesin removal. Similar to the targeted CTCF degradation by Nora et al. [22], upon depletion of cohesin-loading factor in mouse livers, TADs and associated Hi-C peaks disappeared globally. The compartmental segregation was observed to be remarkably resistant to CTCF depletion. They confirmed that genome compartmentalisation determined by chromatin state is cohesin-independent and formation of TADs by loop extrusion is cohesin-dependent.

A study from Bintu et al. [24] in 2018 used some high-throughput oligopaint labeling and imaging techniques to look into single-cell level chromatin dynamics inside the nuclei of several different mammalian cell lines. Matching the predictions of Fudenberg et al., they could observe TAD formation with domain boundaries preferentially at CTCF and cohesin-binding sites. At a population level, the disappearance of TADs upon cohesin depletion could also be observed. Interestingly, at the single-cell level the abundance of TAD-like structures remained constant. Instead, the cells were missing preferential positioning which explained the loss of TADs at population level. The authors therefore concluded that one doesn't require cohesin for domain structure formation or maintenance on the single cell level, but rather for the preferential positions of BEs.

Having been cited 1145 times up to the current day, this paper is currently one of the most influential in its field and will hopefully further inspire future experiments to deepen our understanding about spatial chromatin organization.

References

[1] Fudenberg, Geoffrey, et al. "Formation of chromosomal domains by loop extrusion." Cell reports 15.9 (2016): 2038-2049.

[2] Mirny, Leonid A., Maxim Imakaev, and Nezar Abdennur. "Two major mechanisms of chromosome organization." Current opinion in cell biology 58 (2019): 142-152.

[3] Lanctôt, Christian, et al. "Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions." Nature Reviews Genetics 8.2 (2007): 104-115.

[4] Schuster-Böckler, Benjamin, and Ben Lehner. "Chromatin organization is a major influence on regional mutation rates in human cancer cells." nature 488.7412 (2012): 504-507.

[5] Fraser, James, et al. "Chromatin conformation signatures of cellular differentiation." Genome biology 10.4 (2009): 1-18.

[6] "Chromatin organization and structure" `https://neherlab.org/20181213_theoretical_biophysics.html` (accessed: 03/01/22)

[7] Dixon, Jesse R., et al. "Topological domains in mammalian genomes identified by analysis of chromatin interactions." Nature 485.7398 (2012): 376-380.

[8] Nora, Elphège P., et al. "Spatial partitioning of the regulatory landscape of the X-inactivation centre." Nature 485.7398 (2012): 381-385.

[9] Richards, Eric J., and Sarah CR Elgin. "Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects." Cell 108.4 (2002): 489-500.

[10] Rao, Suhas SP, et al. "A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping." Cell 159.7 (2014): 1665-1680.

[11] Nasmyth, Kim. "Disseminating the genome: joining, resolving, and separating sister chromatids during mitosis and meiosis." Annual review of genetics 35.1 (2001): 673-745.

[12] Alipour, Elnaz, and John F. Marko. "Self-organization of domain structures by DNA-loop-extruding enzymes." Nucleic acids research 40.22 (2012): 11202-11212.

[13] Lieberman-Aiden, Erez, et al. "Comprehensive mapping of long-range interactions reveals folding principles of the human genome." science 326.5950 (2009): 289-293.

[14] Barbieri, Mariano, et al. "Complexity of chromatin folding is captured by the strings and binders switch model." Proceedings of the National Academy of Sciences 109.40 (2012): 16173-16178.

[15] Benedetti, Fabrizio, et al. "Models that include supercoiling of topological domains reproduce several known features of interphase chromosomes." Nucleic acids research 42.5 (2014): 2848-2855.

[16] Rosa, Angelo, Nils B. Becker, and Ralf Everaers. "Looping probabilities in model interphase chromosomes." Biophysical journal 98.11 (2010): 2410-2419.

[17] Sofueva, Sevil, et al. "Cohesin-mediated interactions organize chromosomal domain architecture." The EMBO journal 32.24 (2013): 3119-3129.

[18] Rudan, Matteo Vietri, et al. "Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture." Cell reports 10.8 (2015): 1297-1309.

[19] Xiao, Tiaojiang, Julie Wallace, and Gary Felsenfeld. "Specific sites in the C terminus of CTCF interact with the SA2 subunit of the cohesin complex and are required for cohesin-dependent insulation activity." Molecular and cellular biology 31.11 (2011): 2174-2183.

[20] Sanborn, Adrian L., 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 112.47 (2015): E6456-E6465.

[21] Rao, Suhas SP, et al. "Cohesin loss eliminates all loop domains." Cell 171.2 (2017): 305-320.

[22] Nora, Elphège P., et al. "Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization." Cell 169.5 (2017): 930-944.

[23] Schwarzer, Wibke, et al. "Two independent modes of chromatin organization revealed by cohesin removal." Nature 551.7678 (2017): 51-56.

[24] Bintu, Bogdan, et al. "Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells." Science 362.6413 (2018).