Chromatin reconstruction and dynamics using random Loops accounting for Chromosome Capture data

O. Shukron, L. Giogetti, E. Heard and D. Holcman[1] *

October 20, 2014

Abstract

Keywords: Modeling, Inverse problem, First passage times,

1 Introduction

[Unfinished]
The spatio-temporal organization of the chromatin plays an essential role in the regulation of sub-cellular activity such as gene expression[1].

3d structure of the Igh locus was suggested by [5] after tagging several position along the loci.

2 Experimental data and Methods

2.1 The experimental data

We used the experimental 5C data generated by Nora et al.[6] for the chromosome contact frequencies of the X chromosome in a 4.5 Gb region encompassing the X inactivation center. In our work we focused on a subset of

*[1]Ecole Normale Supérieure, IBENS, 46 rue d'Ulm 75005 Paris, France.

the data, including a 94,082 bp region, termed TAD D and E (see [6]). Two replicates of the experiments were conducted. In our analysis we use the average of these two experimental replicates.

The encounter histograms generated by the 5C experiments describes the contact between genomic loci of variable sizes over millions of nuclei. To map these contact into uniform sized segments, we followed data coarse-graining as described in [4], to map segment encounter frequencies to that of evenly spaced, equal size beads. A bead size of 3000 bp was chosen according to the mean size of restriction segments resulted by HindIII enzyme digestion, used in the process of the 5C experiments [6] [4] (Supplementary Materials). This choice of bead size resulted in coarse-grained polymer of 307 beads.

The coarse-grained bead encounter frequencies includes 14,509 data points and was used to calculate the bead encounter probability as a function of the distance in beads units. For each bead, equidistant encounter frequencies were averaged, and the resulting encounter frequency signal was divided by the total number of encounters, to get the encounter probability as a function of bead distance.

For each bead $n = 1..307$ we fit the experimental encounter probability signal with a function of the form

$$p_n(d) = \alpha_n d^{-\beta_n} \tag{1}$$

with p_n the encounter probability of bead n, $\alpha_n = \frac{1}{\sum_{j=1}^{k} j^{-\beta_n}}$, d is the distance in bead units, and β_n is a parameter to be determined by the fitting procedure.

2.2 The polymer model

To represent the chromosome polymer and to explore the different architectures that can explain the appearance of the TADs, we chose to use the Rouse chain. The Rouse chain describes the dynamics of a linear polymer as a collection of massless beads connected by harmonic springs and driven by the thermal forces of diffusion. The corresponding system of stochastic differential equations describing the time progression of a chain of N beads is given, in the 3-dimensional case, by

$$\frac{dR}{dt} = -\frac{3D}{b^2} KR + \sqrt{2D}\frac{dW}{dt} \tag{2}$$

where, $R(t) = [R_1(t), R_2(t), .., R_N(t)]^T$ describes the 3D coordinates of N beads at time t, D is the diffusion constant, b is the standard-deviation of the distance between adjacent beads of the chain, W is an independent $N \times 3$ Brownian motion with mean 0 and variance 1 in each component, and K is the Kirchhoff bead connectivity matrix, which reflects different chain connectivities.

We have constructed our polymer model to have L loops of random sizes. To form each loops, we have randomly chosen 2 non-neighboring beads and altered the connectivity in the Kirchhoff matrix, with the condition that no bead can be connected to form more than one loop.

2.3 Simulations

Throughout simulations, for each fixed number of loops, the chosen beads to connect varied randomly. Such a choice was made to refer to the heterogeneity if the spatial organization inside TAD between cells, even in the same cell phase [6].

Simulations were always carried out until the chain's relaxation time, in which point any two beads were determined to have encountered if their distance satisfied $|R_j - R_k| < \epsilon < b, \quad (j \neq k)$. The chain's relaxation time is given by the slowest mode of the linear chain

$$\tau = \frac{b^2}{12D \sin(\frac{\pi}{2N})}$$

for which the number of simulation steps performed is $\frac{\tau}{\Delta t}$. The time step, Δt was set so to prevent simulation 'blow-ups' by demanding that the quotient of the norms of beads position at two subsequent time steps would be smaller that unity, which resulted in $\Delta t < \frac{b^2}{12D}$.

For each tested polymer connectivity we constructed the bead encounter frequencies histogram and derived the bead encounter probability from it. The bead encounter probability was then fitted similarly to the fitting in eq. 1.

For a linear Rouse chain, the expected value of β is 1.5 [3]. We interpret $\beta < 1.5$ as long range interaction resulting from non nearest-neighbor bead interactions. Because the addition of non-neighboring connections to the linear chain can only increase the long range encounter probability, we have focused on interpreting the fitted values in the range $\beta < 1.5$.

3 Results

3.1 Analysis of the experimental data

To evaluate the mean encounter probability in the experimental data, we have calculated the β value for the 3 cases of TAD D, TAD E and TAD D+E for each of the beads in those genomic region. This provides us with a basis for comparison of the results of simulations with the experimental data and to the inference on the spatial organization of the chromosome.

The calculation of β for each bead in the case of TAD D+E resulted in a pattern which was aligned with the significant long range interactions (Figure 1 b), represented by the peaks the encounter probability graph (Figure **??**). Indeed, the mean β value was 0.729, which is well below the expected value for the linear Rouse chain (Figure 1 c).

We then turned to examine whether long range interactions stem from inter or intra-TAD polymer looping.TAD D has almost no significant long range interactions (supplementary material), although the mean fitted β value was 0.71, which indicates either packed organization of TAD D or heterogeneity of the location of loops within the cell population examined in the 5C experiments. Intra-TAD long range interaction within TAD E contribute about half of the significant long range encounter peaks in the encounter probability graph (supplementary material), whereas the other half stem from inter-TAD long range interactions.

Given the calculation of the β values from the experimental data we now turn to explore which polymer architecture give rise to the observations.

3.2 Random fixed loops simulations

To examine if fixed loops in the polymer can recreate the TADs, we have placed connection between beads corresponding to the peaks of the encounter probability and simulated our model until relaxation time. In a 307 beads polymer, these fixed loops were insufficient to recreate a TAD-like structure. Only localized nearest neighbors interactions emerged by this model(Figure 2) which cannot account for the observed long range interaction map.

Although these fixed loops are insufficient by themselves to create the encounter maps expected, we noticed peaks on the boundaries of TADs. We have postulated that these peaks, which connects the two boundaries of a TAD, are of significance to the spatial organization and the functionality

of regulatory elements within the TAD. We therefore set to examine the encounter probability of a polymer having a large fixed loop between two predefined ends.

Furthermore, to reflect the heterogeneity in the spatial organization of the chromatin of cells in the HiC experiment [2] [6], we have added loops between randomly chosen beads on the linear chain between the two boundaries we have determined for the big loop (Figure 3 a).

Increasing the number of internal random loops from 1 to 10, we see an encounter pattern which resembles that of a TAD (Figure 3).

Next, we added a second, adjacent region, to form a loop, and sequentially added 1 to 10 internal random fixed loops in each. (Figure 4)

4 Discussion

5 Figure

References

[1] Thomas Cremer and Christoph Cremer. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nature reviews genetics*, 2(4):292–301, 2001.

[2] Job Dekker, Marc A Marti-Renom, and Leonid A Mirny. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nature Reviews Genetics*, 14(6):390–403, 2013.

[3] M. Doi and S.F. Edwards. *The theory of polymer dynamics*. International series of monographs on physics. Clarendon Press, 1986.

[4] Luca Giorgetti, Rafael Galupa, Elphège P. Nora, Tristan Piolot, France Lam, Job Dekker, Guido Tiana, and Edith Heard. Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell*, 157(4):950–963, May 2014.

[5] Suchit Jhunjhunwala, Menno C van Zelm, Mandy M Peak, Steve Cutchin, Roy Riblet, Jacques JM van Dongen, Frank G Grosveld, Tobias A Knoch, and Cornelis Murre. The 3d structure of the immunoglobulin heavy-chain

locus: implications for long-range genomic interactions. *Cell*, 133(2):265–279, 2008.

[6] Elphège P Nora, Bryan R Lajoie, Edda G Schulz, Luca Giorgetti, Ikuhiro Okamoto, Nicolas Servant, Tristan Piolot, Nynke L van Berkum, Johannes Meisig, John Sedat, Joost Gribnau, Emmanuel Barillot, Nils Blüthgen, Job Dekker, and Edith Heard. Spatial partitioning of the regulatory landscape of the x-inactivation centre. *Nature*, 485:381–5, 2012 May 17 2012.

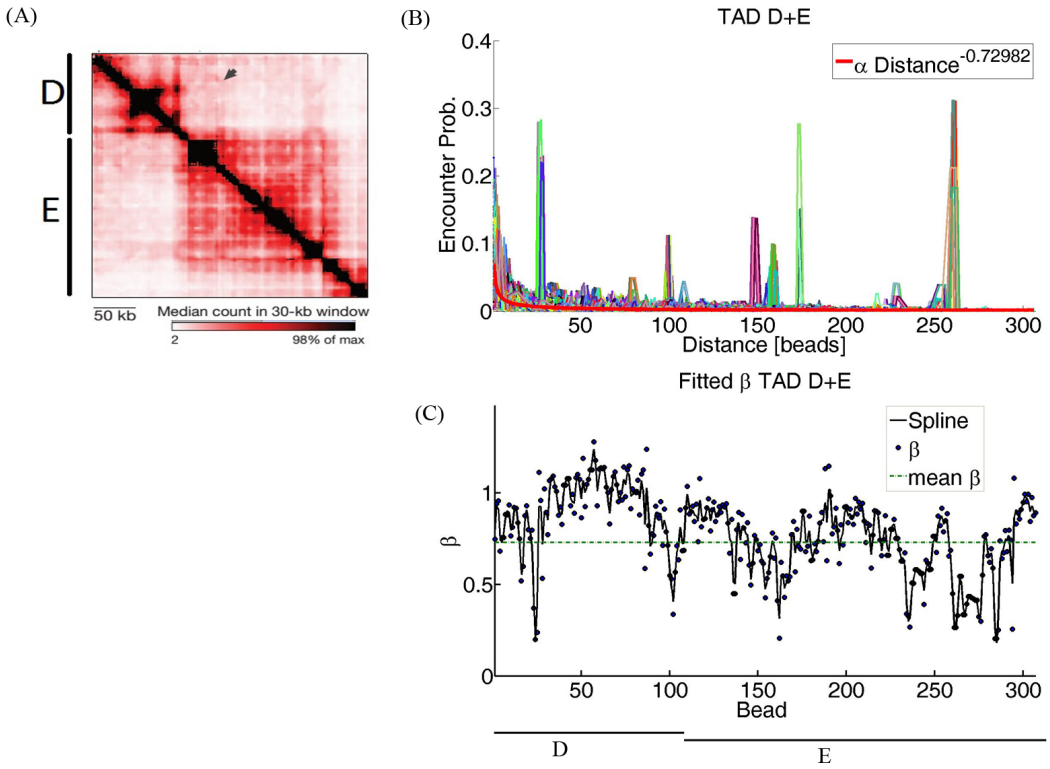

Figure 1: **Analysis of the experimental data** (A) Average contact histogram of the 2 replicates of the 5C experiments in 2 discrete genomic regions of high self interactions termed TAD D and E (Nora et al [6]) (B) The encounter probability graphs for each of the 307 beads show long range interactions between beads in TAD E and between TAD D and E, a fit of the form $\alpha Distance^{-\beta}$ to the mean encounter of each distance (red curve), was found to have $\beta = 0.729$, well below the expected $\beta = 1.5$ for a linear Rouse polymer, implying compact configuration of the polymer and looping (C) The calculated β value for each of the 307 beads (circles) was fitted with a smoothing spline (black curve) to show a fluctuating pattern around the mean ($\beta = 0.729$, green dashed line) with sharp decrease in values for beads having long range interactions, e.g beads 25, 102, 165, 285.

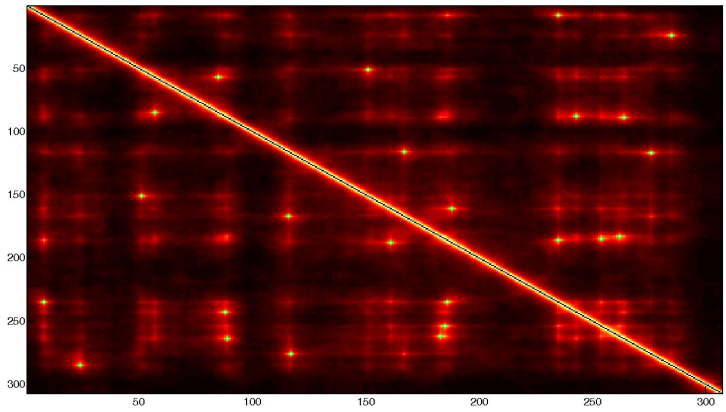

Figure 2: **Simulation of a chain with loops at positions corresponding to the peaks of the encounter data** was found insufficient to create an encounter histogram resembling the appearance of two TADs as in the results of the 5C experiment (Figure 1.A)

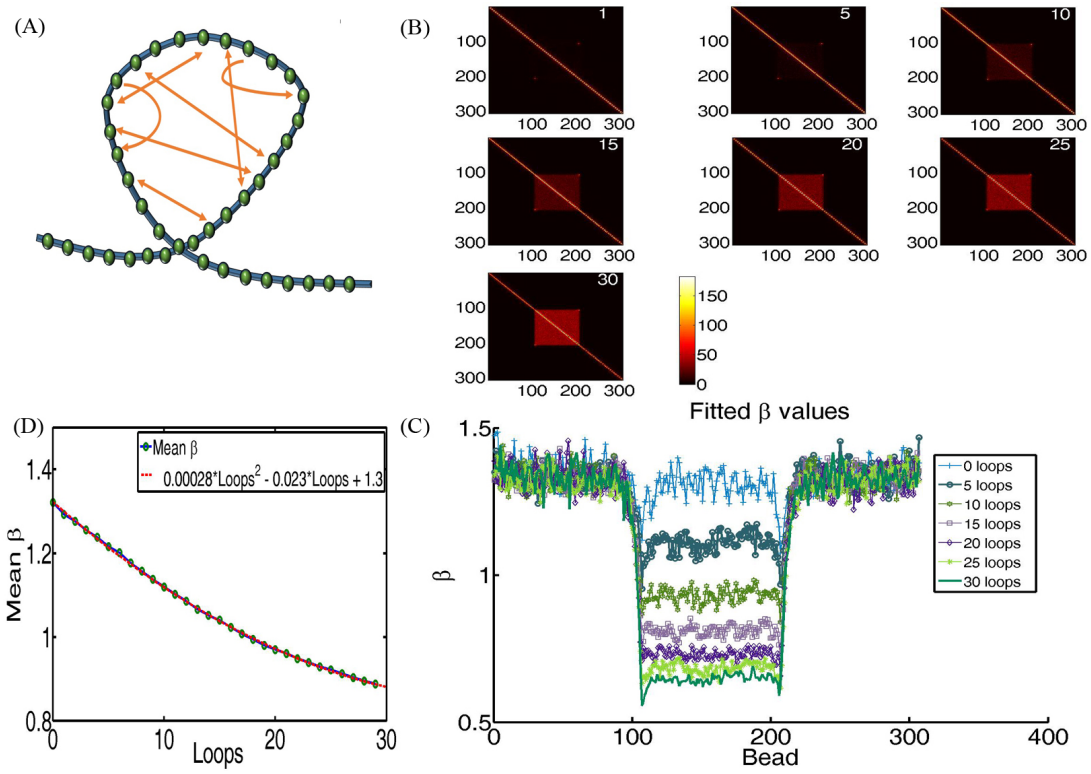

Figure 3: **Encounter profile for a polymer model with one fixed loop and 1-30 random internal loops.** (A) A sketch of the polymer model used in simulation. Beads 107 and 207 were connected to form a 100 beads fixed loop throughout simulations, 1 to 30 internal loops were sequentially added. Each loop was formed by randomly picking two beads in the range 108-206, as an example shown by the orange arrows (B) The encounter histograms for 7 cases of adding internal random loops (number indicated in white in each box) resembles the TAD region in the central part of the polymer corresponding to the position of the fixed loop.(C) For each number of random internal loops, a model of the form $\alpha Distance^{-\beta}$ was fitted to the encounter probability of each bead, the resulting β values show a sharp decrease for beads in the loop range (107-207) with increased number of internal loops, whereas, as expected, beads outside the tail show similar behavior independent of the number of loops (D) The mean β values as a function of number of internal loops decreases quadratically and allows to infer from the measured mean encounter probability the number of loops in the polymer.

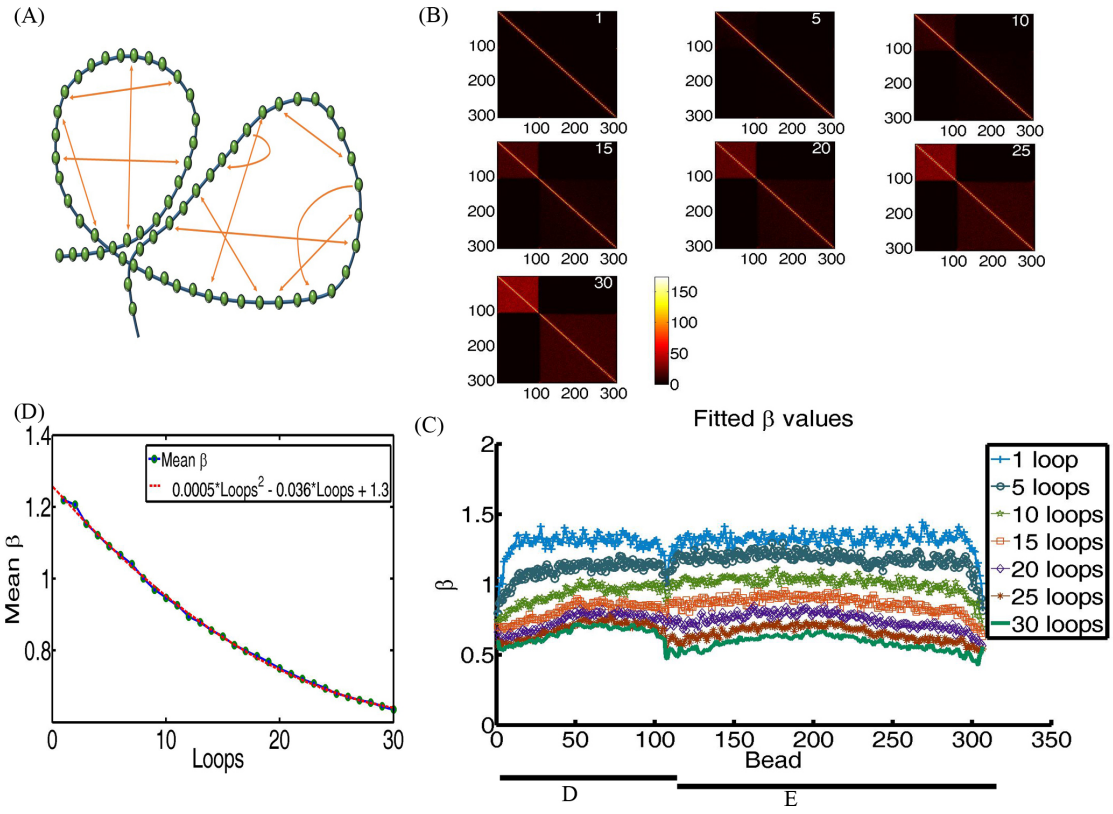

Figure 4: **Encounter profile for a polymer model with two fixed loops and 1-30 random internal loops in each** (A) A sketch of the polymer model used in simulations. Beads 1, 107 and beads 108, 307 were connected to form two large loops corresponding to the positions of TAD D and E in the experimental data. One to 30 internal loops were sequentially added in each large loops. Each loop was formed by randomly picking two beads in the range 2-106 and two in the range 108-306, as in the example shown by the orange arrows (B) The encounter histograms for 7 cases of adding internal random loops in each TAD (number indicated in white in each box) resembles the TADs of the experimental data (Figure 1.(A)) as the number of random loops is increased (C) For each number of random internal loop, a model of the form $\alpha Distance^{-\beta}$ was fitted to the encounter probability of each bead. Beads on the edge between the two fixed loops (e.g beads 107-8) show decrease in β due to the high, long range encounter probability with beads participating in each fixed loop. (D) A quadratic relationship was found empirically between the number of loops and the mean β.