Random loop model describing the Topologically Associating Domains, application to chromatin structure

Ofir Shukron
David Holcman
Group of applied Mathematics and computational Biology
ENS

January 8, 2015

Agenda

1. Part I
   - Introduction
   - The experimental setting
     - Chromosome Conformation Capture Carbon Copy (5C)
     - The experimental Data
   - Analysis of the data
     - TAD D and E
   - Theoretical model
     - The encounter probability
     - Peaks of the encounter data
   - Simulation Results
     - Simulations with a simple rouse chain
     - Random loops model
   - Summary
2. Part II - Reconstruction
   - Introduction
     - The encounter probability

Introduction

1. The spatio-temporal organization of a chromosome has significant implication for nuclear activities, such as gene regulation.
2. However, the dynamic organization of the chromatin is not entirely understood.
3. Dynamic organization of the chromatin allows for long range gene regulation.
4. New methods to record chromatin conformation were recently developed.
5. Stable regions in the genome, termed TAD, were shown to exist in the population level.
6. Here we show that using random polymer looping model we can explain the appearance of TADs.
7. Given the static interaction maps, what can we say about the underlying dynamical polymer model?

Chromosome Conformation Capture Experiments

A set of methods to simultaneously record millions of looping events occurring within the genome. The general steps are:

1. Nuclei are extracted from millions of cells;
2. Formaldehyde induces protein-DNA and protein-protein cross-links;
3. Restriction enzymes digest the cross-linked DNA;
4. The cross-linked DNA is purified, diluted and ligated;
5. Cross-links are reversed;
6. PCR to amplify ligation junctions;
7. The histogram of segments' encounter is produced.

The experimental data

- Two replicate of the 5C experiments were conducted by Nora et. al 2012.
- Undifferentiated mouse embryonic stem cells.
- We focus on a 920,432 bp subset of the data, around the X inactivation center of the X chromosome.
- The region harbors the Xist enhancer and Tisx promoter.
- The subset of the data contains two adjacent Topologically Associating Domains (TADs).

Topologically Associating Domains (TADs)

Conserved structures showing chromosome interactions on the Mb scale, with higher intra than inter-region interactions. It is believed that the TAD forms a 'regulatory unit' for gene expression. However, its existence and role are not completely verified in a single cell.

Figure: The 4.5 Mb region (left), enlargement of TAD D and E (center). Displayed median count in a 30kb window every 6kb, gene expression Peasrsons correlation (right)

From restriction segments to beads

- Coarse-graining of the data was made by choosing a bead-size of 3000 bp, corresponding to the mean segment length resulted from the digestion of HindIII.

- The genomic section was evenly partitioned. Each bead receives a start and end index according to the segment it covers.
- for example:

bp range	start ind	end ind
500-3500	1	2
4000-4500	2	2
5000-12001	2	4

Bead encounter probability

TAD D and E

- We work with the average of the two experimental replicates.
- A total length of 920,432 bp - resulting in 307 beads (TAD D 107 beads, TAD E 200 beads).
- For each bead we calculate the encounter probability as a function of distance (bead units).
- TAD E has several strong specific interactions. TAD D has weak intra long range interactions. Strong inter-TAD specific interactions

Theoretical model

The Rouse model

To understand the dynamical process underlying the observed encounter probabilities we need a employ a polymer model. We start with the classical and most simple model, the Rouse chain.

- A Rouse chain describes polymer dynamics as a stochastic motion of a collection of microscopic "beads" connected by harmonic springs
- the 3D motion of bead n in the chain of N beads

$$\frac{dR_n}{dt} = -\frac{3D}{b^2}(2R_n(t) - R_{n+1}(t) - R_{n-1}(t)) + f_n(t)$$

- R_n- the position of bead n
  b- the standard deviation of the distance between adjacent beads
  D- the diffusion constant
  f_n- white Gaussian noise
- From the theory, $Pr(\|R_n - R_m\| < \epsilon) \sim |n-m|^{-1.5}$

The enconter probability

For the case of TAD D, TAD E, and the two together, we estimate the bead encounter probability, p, and fit it with a function of the form

$$p_n(d) = \alpha d^{-\beta}$$

where, d is the distance in [bead] units, n is the bead number, $\alpha = \frac{1}{\sum_{j=1}^{d_{max}} j^{-\beta}}$ and β is a parameter to be estimated. We report the values of β for each bead in each case.

TAD D
mean data fit. β =0.71409
Encounter Prob.
Distance [beads]
Fitted β for TAD D
Spline
β
Fitted β values
Bead number
TAD E
mean data fit. β =0.69223
Encounter Prob.
Distance [beads]
Fitted β for TAD E
Spline
β
Fitted β values
Bead number

Peaks of the encounter data

- A peak-calling procedure was implemented to find pairs of beads which interact frequently.
- Peak analysis was performed separately for TAD D, E and between TADs.
- About 90% of the peaks in the encounter data result from specific interactions between TADs and in TAD E.

Simulation with simple rouse chain

- A simple Rouse model cannot reproduce the TAD, as expected.
- Placing loops corresponding to the peaks of the encounter data does not reproduce the TADs.

Random Fixed Loop Model

- The enhancer-transcription factor elements' encounter motivates the simulation of a chain with randomly placed loops.
- These encounters might not be as frequent as 'stable' loops in the chromatin and therefore not shown significantly in the encounter maps.
- Simulate a chain of 307 beads, having random loops in a bounded region.
- Increasing number of loops at random positions.

Stable loop with random loops within

one TAD

1. Following the peaks observation, we form a mix of 'stable' big loop with internal random loops.
2. In a 307 beads chain, beads 107 and 207 were connected.
3. We iteratively add up to 30 connectors within the stable big loop.
4. We can start seeing the emergence of β curve pattern as in the experimental data.

Stable loop with random loops within

Two TADs

1. In a 307 beads chain, we create two stable loops by connecting bead 1 and 107, and bead 108 and 307
2. We iteratively add 30 connectors within each TAD.
3. We can start seeing the emergence of β curve pattern as in the experimental data.

Summary

- We presented a plausible and simple model to explain the appearance of the conserved TAD domains.
- The model represent the results obtained in the population level.
- A relationship between the mean number of loops and the encounter probability was shown by simulations.
- The parameter β used for our analysis can easily be retrieved from the experimental data.

Work in progress

1. Calculating the conditional probability of bead A meeting bead B before it mets bead C in the resulting polymer configuration.
2. Calculation of the mean first encounter time between beads in random loop model
3. Showing analytically the relationship between number of loops and decrease in β
4. Put in more rigorous form the relationship between the pattern of β and the mean observed chromatin structure.
5. constructing a heterogeneous random loop model, taking into account the peaks and varying spring constants.
6. Simulating a model with random dynamic loops, in which loops can form and dissociate.
7. simulating the beta polymer with random loops.
8. Experiment with other polymer models.

Reconstruction

Introduction

1. How do we go from encounter frequencies to structure?
2.

The encounter probability in the Rouse chain

We begin by employing a Rouse chain of N_c beads to represent the chromatin.
The pdf of a vector, R_{mn} between bead m and n in equilibrium is normal

$$P(R_{mn}) = \left(\frac{3}{2\pi b^2|m-n|}\right)^{3/2} \exp\left(-\frac{3R_{mn}R_{mn}^T}{2b^2|m-n|}\right)$$

Setting $R_{mn} = \vec{0}$ we get the encounter probability $\left(\frac{3}{2\pi b^2|m-n|}\right)^{3/2}$.
To match the HiC signals, we normalize by dividing by

$$\sum_{t=1}^{N_C-1}\left(\frac{3}{2\pi b^2|N-t|}\right)^{3/2} =$$

Therefore, the encounter probability does not depend on b^2

$$P(R_{mn} = \vec{0}) = \frac{|m-n|^{-3/2}}{\sum_{j=1}^{N}|N_r - j|^{-3/2}}$$

The encounter probability in a ring

Forcing the two ends of a Rouse chain to meet creates a "Rouse ring". A rouse ring can be represented by a Brownian bridge. The pdf of a vector r_m of the ring of N_r beads is given by

$$P(r_m) = \left(\frac{3N}{2\pi b^2 m(N_r - m)}\right)^{3/2} \exp\left(-\frac{3N_r r_m r_m^T}{2b^2 m(N_r - m)}\right)$$

Setting $r_m = \vec{0}$ and normalizing, we get

$$P(r_m = \vec{0}) = \frac{m(N_r - m)^{-3/2}}{\sum_{t=1}^{N_r - 1}(t(N_r - t))^{-3/2}}$$

which also does not depend on b^2

The encoutner probability in composite structures

The pdf of the vector between bead n of the chain and bead m of the ring is calculated as the sum of the two random vectors R_{nN_c} and r_m.
The calculation can be done using the convolution of the two pdfs, to give

$$P(r_m + R_{nN_c} = 0) = \left(\frac{3}{2\pi b^2(|N_c - n| + m(N_r - m))} \right)^{3/2} / S$$

$1 \leq n \leq N_c \qquad 0 \leq m \leq N_r$ and S the normalization constant

We can separate the the structures. The encounter probability behaves as that of the chain for $1 \leq n \leq N_c$ and as a ring for $0 \leq m \leq N_r$

From encounter probabilities to distances

Beads with equal encounter probability are located equidistant along the chain from a reference bead. The distance is the shortest one along the graph representing the polymer.
By inverting the encounter probability equation for each structure we can transform probabilities to distances.
In the case of the Rouse chain

$$D_1(P) = \left(P \sum_{t=1}^{N_c-1} |N_c - t|^{-3/2} \right)^{-2/3} \qquad (linear)$$

For the Rouse ring

$$D_2(P) = \frac{N_r + \sqrt{N_r^2 - 4N_r \left(P \sum_{m=1}^{N_r-1} (m(N_r - m)/N_r)^{-3/2} \right)^{-2/3}}}{2}$$

Using these two structure we'll invert the encounter probability signals from the HiC experiments.

Interpretation of the HiC using chains and rings

- In theory, we can decompose the HiC signal into regions where the polymer takes a ring or linear chain structure
- The question is how to distinguish between them?

As an example, we have simulated a complex structure of the following form