

Polymer Chain Dynamics, Analysis of
Topologically Associating Domains - Summary
of Methods and Findings

December 12, 2014

Contents

1	Theoretical Material	5
1.1	The Topologically associating domains (TADs)	5
1.2	A short review of Chromosome Capture methods	6
2	Analysis of the experimental data	9
2.1	Coarse-graining of the data	9
2.2	The encounter matrix	10
2.3	From encounter frequency to encounter probability	10
2.4	Showing the TADs	10
2.5	Symmetry of encounter data	11
2.6	Peaks of the encounter data	12
2.7	Interpretation of the peaks	14
2.8	Estimating b^2 from the data	14
2.9	Estimating $E[P(1)]$ from the data	15
2.10	Determining the spring constant for peaks	15
2.11	Fitting the encounter data	16
3	Simulations	19
3.1	The Simulation Framework	19
3.2	Initial simulation settings	21
3.3	Verification of the validity of simulation framework	21
3.3.1	The encounter frequency	23
3.4	A polymer with a single cross-link	24
3.4.1	Increasing the encounter distance	26
3.5	One, two and three loops at fixed positions	27
3.5.1	Internal loops	28
3.6	Random fixed loops	30
3.6.1	One TAD	30

3.6.2	Two TADS	31
3.7	Simulation of variable loop number	31
3.7.1	One TAD	31
3.7.2	Two TADs	35
3.8	Dynamic loop model	35
3.9	TAD D and E with cross-links corresponding to the peaks of the experimental data	37
3.10	Conditional encounter probability	38
3.10.1	Three beads in a linear chain	38
3.10.2	Two beads in the random fixed loops model	38
3.10.3	Two beads and a loop	39
4	Review of Literature	43
4.1	From Nora et.al 2012 [14]	43
4.2	From Dekker et al 2013 [2]	45
4.3	The Hierarchy of the 3D genome. Gibcus 2013 [7]	47

Chapter 1

Thoretical Material

1.1 The Topologically associating domains (TADs)

The Topologically Associating Domains (TADs) [14], which belong to the X inactivation center (Xic), are discrete adjacent chromosomal regions prominently interacting among themselves. By interaction we mean close physical proximity of one region to another part of the same chromosome, determined by the resolution of the 5C technique [4] [1], for which chromosomal regions are fixed in formaldehyde, and contact are defined as region close enough to be considered as having protein-protein interactions.

The Xic is a 4.5 Mb region located on embryonic stem cells' X chromosome and is believed to be conserved throughout the mammalian family. Nora et. al[14], recognized 9 TADs within the Xic region, and termed them in alphabetical order A to I (see Figure 1.1). These region span 200kb to 1Mb in size.

The Xic orchestrate the inactivation of the X chromosome by controlling the transcription of its coding sequences. The major player in the inactivation process is the Xist sequence.

The Xist sequence is located on the TAD termed E. The repressive anti-sense of Xist, termed Tsix, is located on the adjacent D TAD (Figure 1.2). In this report we focus only on TADs D and E. The TADs D and E are approximately of size 324 kb and 537 kb respectively (see Figure 1.1) and contain non-protein and protein coding sequences(For a graphical view see Figure 1.2), among which, the TAD D harbors the Xite regulator of Xist gene, and Tsix which is the repressive anti-sense transcript of Xist.

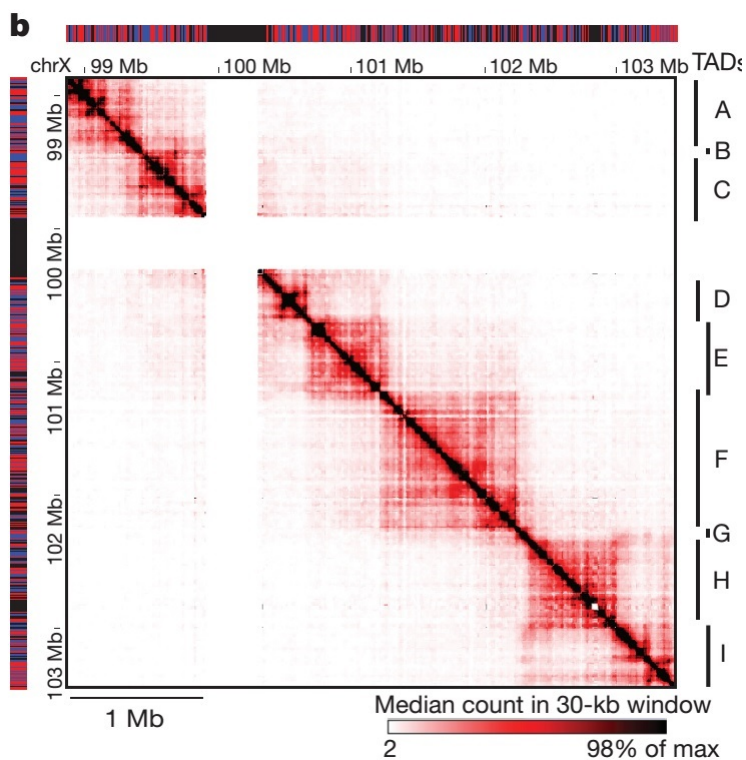


Figure 1.1: A display of the TADs in the region of the X chromosome. Taken from Nora et al. Nature. 2012 Apr 11;485(7398):381-5 [14]. The red-blue bar on top and left show the restriction segments cut by HINDIII (forward and reverse segments)



Figure 1.2: The genes and non protein coding regions of TAD D and E. Image taken from from Nora et al. Nature. 2012 Apr 11;485(7398):381-5 [14]

1.2 A short review of Chromosome Capture methods

The 5C method [4] relies on the first steps of the seminal 3C method proposed by Dekker et al. 2002 [3]. From the [supplementary material](#) of [3]. The steps of the 3C can be described as follows (see also Figure 1.4:

1.2. A SHORT REVIEW OF CHROMOSOME CAPTURE METHODS 7

1. Roughly 10^8 intact nuclei are isolated
2. Nuclei are **cross-linked** by using 1% **formaldehyde**, or its relative paraformaldehyde, for 10 minutes, which induces protein-protein and DNA-protein cross-linking. The cross-link is made by formaldehyde in the guanine nucleotide, between DNA-bound proteins and the DNA. The **lysine** amino acid of the protein is connected through its $-NH_2$ group to the **Methylene bridge** ($-CH_2-$)

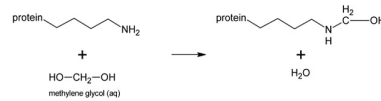


Figure 1.3: The chemical bond between the NH_2 group of formaldehyde and the lysine $-NH_2$ group of the protein

Formaldehyde can be polymerized if alcohol is not added to it. Therefore, it is relevant to ask how the procedure in Nora et al [14] was performed?

3. Cross-linking distance is between $10 - 100nm$ [2]
4. All non cross-linked proteins are removed. It is important to note that the DNA is purified from non cross-linked proteins only *after* the cross-links were made.
5. A site-specific restriction enzyme (**EcoRI** or **HindIII**, which are 6bp cutters) is used to digest the cross-linked DNA for 1 hour.
6. The restriction enzymes are inactivated using SDS and incubation for 20 minutes.
7. The reaction is 15 times diluted to favor relevant DNA-end ligation.
8. The free ends of the cleaved DNA are ligated (DNA is still cross-linked with proteins) for 45 minutes using **ligase**.
9. there are roughly 50 – 60% or erroneous junction created by ligation, those include circular ligation and ligation caused by incomplete digestion of the restriction enzyme
10. Reverse cross-linking by overnight incubation: the cross-links are destroyed, leaving segments of ligated DNA. The reverse cross-linking is done by incubation at 70^0 .
11. DNA is purified from non cross-linked proteins.
12. The products of the ligation are detected and quantified using PCR, by designing primers to specific ligation junctions.

13. a control template is created by using the same restriction enzyme on non cross-linked DNA and ligation of DNA fragments without dilution

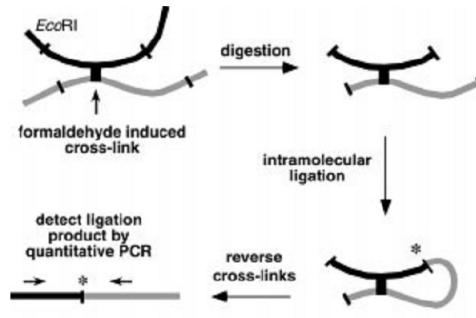


Figure 1.4: A schematic representation of the 3C method. Image taken from Dekker et al. 2002 [3]

Chapter 2

Analysis of the experimental data

Two replicates of 5C experiments were conducted on the chromosome region encompassing the Xic region. The data was coarse-grained as explained in [2.1](#), in accordance with the procedure by [\[8\]](#)[\[14\]](#). The encounter data includes 14,508 pairwise encounters of genomic loci. Each genomic segments is defined by a start and end-bead index.

2.1 Coarse-graining of the data

The following paragraphs are an adapted version of the text in from Giorgetti et al. [\[8\]](#) supplementary material.

In the genomic region of the TAD D and E, 124 forward (FOR) and 126 reverse (REV) primers anneal to alternate HindIII restriction fragments of variable size. Since the average length of FOR and REV fragments in this region is 3078 bp, a 3000-bp beads was chosen, resulting in polymers of $N=108$ beads in the case of TAD D alone (chrX:100378306-100699670, total length 321,364 bp) and $N = 307$ beads in the case of TAD D+E (total length 920,432 bp). Thirty five percent of restriction fragments were longer than 3000 bp and were mapped on multiple beads, while the remaining 65% contributed, together with their nearest neighbor, to define the expected contacts of single beads.

The uneven sampling of the genomic region, provided by restriction fragments of different length, must be mapped onto the even sampling provided

by equally spaced beads. This is performed as follows[8]:

1. The 5 end of the first bead in the chain is positioned at the 5 end of the first restriction fragment in the region of interest
2. Each restriction fragment in the region is assigned two indices i and j ($i, j = 1N$) corresponding to the beads to which its 5 and 3 ends overlap.
3. 5C counts corresponding to each pair of restriction fragments are assigned to the corresponding pairs of indices (i, j) and (h, k), for example:
4. If two or more consecutive FOR or REV restriction fragments map to the same bead, their contributions are summed ($< 10\%$ of all interactions).
5. 5C counts of two experimental replicates are averaged for each pair of restriction fragments and their standard deviation is taken as a measure of experimental uncertainty.

2.2 The encounter matrix

The data provided by Giorgetti et al.[8] contains the encounter frequency of beads (not segments!) constructed as explained in 2.1. Sorting the encounter data into a 307 by 307 matrix we can visualize the pair-wise encounter frequencies. Reminder: beads are numbered sequentially from the 5-end to the 3-end of the polymer.

2.3 From encounter frequency to encounter probability

We start by transforming the encounter frequency into encounter probability as follows. Each line of the encounter matrix is divided by the sum of all encounters of the line.

2.4 Showing the TADs

TADs are seen by displaying the normalized encounter data, after removing closest beads data from the matrix of encounters. The normalization is done by taking each row k of the encounter matrix, removing the data for the

closest neighbors of bead k and dividing by the sum of the row. A median filter of size 30 beads is then applied to the image (see methods in [14]).

2.5 Symmetry of encounter data

We next examine whether the data can be regarded as symmetric in the sense of encounter probabilities for each bead. That is, whether bead k has the same probability to meet its closest neighbors from the right and left in the linear chain, or not.

The symmetry of encounter frequency data is shown by measuring the left and right encounter frequencies of each one of the beads comprising the chain. The left and right encounter data are defined in terms of bead index, where for the bead j the left encounter frequencies are defined as a vector of length $j - 1$ with the encounter data of bead j with $1, \dots, j - 1$, and $j + 1, \dots, N$ for the right encounter frequencies.

For bead index $j = 1, \dots, N$ we have different number of beads on the left and right. For comparison, we need to match the number of beads from the left and right for each bead. The number of beads on the left and right of bead j is then taken to be $n_j = \min(j - 1, N - j + 1)$. Let L_j be the left encounter data, R_j be the right encounter data vectors of lengths n_j

$$L_j = [e(j - 1), e(j - 2), \dots, e(j - n_j)]$$

and

$$R_j = [e(j + 1), e(j + 2), \dots, e(j + n_j)]$$

where $e(k)$ is the encounter frequency of the k^{th} closest bead to bead j . Note the reverse order of L_j .

We normalize each vector by dividing it by the sum of its encounters

$$\hat{L}_j = [e(j - 1), e(j - 2), \dots, e(j - n_j)] / \sum_{k=j-n_j}^{j-1} e(k)$$

and similarly for \hat{R}_j .

To test the difference between left and right frequencies, we calculate the mean difference by

$$d_j = \frac{1}{n_j} \sum_{k=1}^{n_j} (\hat{L}_j(k) - \hat{R}_j(k))$$

The Figure 2.1 below shows the results of plotting the mean difference for the two replicates of the 5C data

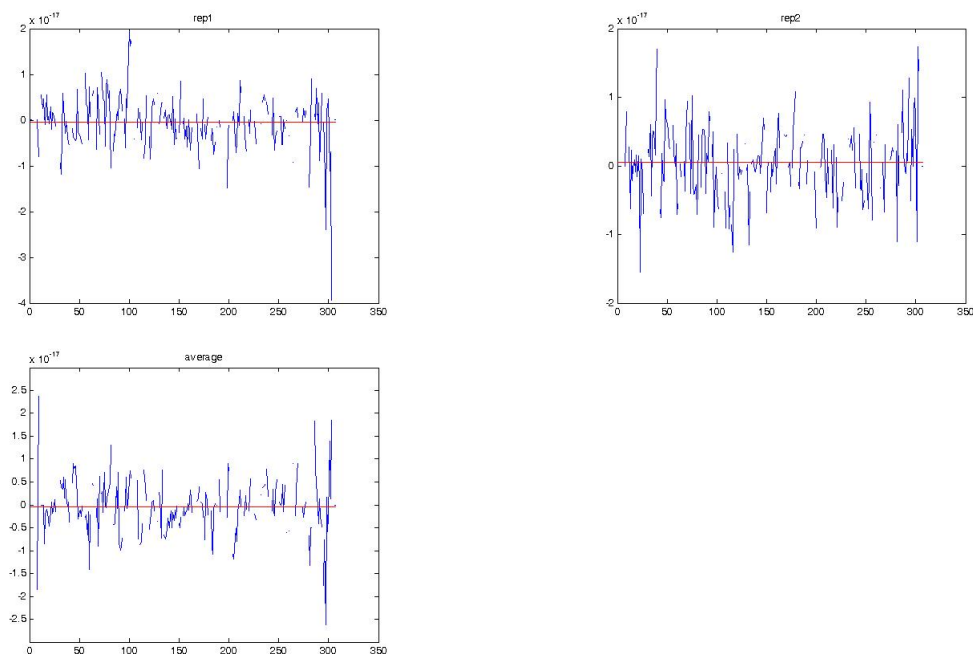


Figure 2.1: The mean difference (y axes) between left and right encounter frequencies for each bead number (x axes) is shown for Rep1 (top left) Rep2 (top right) and average (bottom)

The red lines in the Figure 2.1 signify the mean of the difference data. For the three figures it is of the order 10^{-18} . When points are missing, it means that the segment (restriction segment) occupies more than one bead, therefore the encounter data is zero there.

We can therefore treat the right and left data in the same manner and 'fold' the encounter data such that from now on we test the encounter data by bead distance. We will refer to the one-sided encounter as the encounter data from now on.

2.6 Peaks of the encounter data

For some beads the encounter data shows peaks at some distance from the origin (Figure 2.2). These peaks represent a frequent encounter with a distal part of the chain.

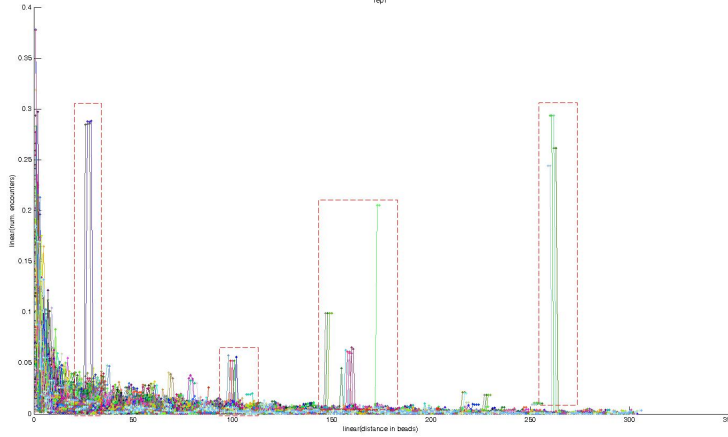


Figure 2.2: The encounter probability of each bead (Y axis) as a function of bead distance (X axis). The dashed red rectangles shows places of prominent non-nearest neighbors encounters

To accurately identify pair of beads that interact more frequently than expected, we employ a peak calling procedure on the 5C probability data. We analyze TAD D and the region between TADs separately from TAD E to avoid peak bias. We average all encounter frequency signals, taken from all genomic distances, and smooth the resulting mean signal using loess smoothing technique to represent the expected background encounter signal, $E(d)$, and calculate its standard deviation σ_E . The background Z-score distribution, $F_b(z)$, is calculated using all data by $z_B = \frac{O_{i(d)}(d) - E(d)}{\sigma_E}$, for each $i(d) = 1, \dots, N(d)$, with $N(d)$ the number of observations available for genomic distance d . A Weibull distribution is fitted to the background signal, where negative values Z_B are discarded, because we are only interested in peak values. We then estimate the threshold value, T_B , above which z-scores are considered as peaks of the data, as the 0.99 percent of this distribution. The value of T_B will be used for judging observations in a particular genomic distances for being peaks.

For each genomic distance d , we calculate a z-score for all available observations at that distance by $z(\{i\}_d) = \frac{O(\{i\}_d) - E(d)}{\sigma_O(d)}$ and fit a Weibull distribution to it, $F_d(z)$, discarding negative z-scores. For each d , we calculate a rejection value T_d of $F_d(z)$, above which observation at genomic distance d are considered peaks. We use the threshold values as our test statistics for multiple hypothesis tests.

Using the background distribution we calculate the p-values for T_d on the background distribution, namely, $p(d) = F_B(F_d^{-1}(0.99))$, calculate the positive False Discovery Rate[17] to get q -values, $q(p)$. As a new the global threshold of the data we choose $T_G = \max_d(p(d); q(p(d)) < 0.01)$.

2.7 Interpretation of the peaks

2.8 Estimating b^2 from the data

The bead distance pdf of the Gaussian chain has the distribution of

$$\Phi(|R_n - R_m|) = \left(\frac{3}{2\pi b^2 |n - m|} \right)^{1.5} \exp\left(-\frac{3|R_n - R_m|^2}{2|n - m|b^2}\right)$$

For an encounter between beads we set $R_n = R_m$, hence

$$\Phi(|R_n - R_m|) = \left(\frac{3}{2\pi b^2 |n - m|} \right)^{1.5}$$

To estimate b^2 , we calculate the encounter from nearest neighbors (plug $|n - m| = 1$ in the equation above)

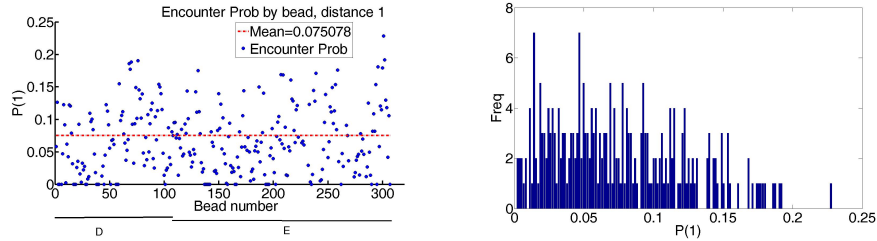


Figure 2.3: Encounter probability of nearest neighbors is used to estimate b^2 . Left- the nearest-neighbor encounter probability ($P(1)$) for all 307 beads. Right-the encounter probability histogram for nearest-neighbors, $P(1)$

Using the estimation of the expected probability of nearest neighbors $E[P(1)] = 0.075078$, which is the most reliable since it is least affected by structure, we can calculate

$$b^2 = \left(\frac{1.5}{\pi E[P(1)]} \right)^{1.5} = 16.0377$$

2.9 Estimating $E[P(1)]$ from the data

2.10 Determining the spring constant for peaks

For peaks which display encounter probability much larger than nearest neighbor probability, we adjust the spring constant to reflect the frequent encounter. The assumption made here is that all beads of the chain are connected with similar springs, however beads for which the encounter probability exceeds nearest-neighbor encounter probability confers a different dynamics, which is reflected in the spring constant.

In the Rouse model the relation between the spring constant and b^2 is given by

$$b^2 = \frac{3K_B T}{k}$$

with K_B the Boltzmann constant and T the absolute temperature. According to the estimation done previously, we have

$$k = \frac{3K_B T}{16.0377}$$

If we observe $\hat{P} > E[P(1)]$ for some beads, we want to estimate the new spring constant \hat{k} , and we can write

$$\hat{P} = \left(\frac{1.5}{\pi b^2} \right)^{1.5} = \left(\frac{k}{2\pi K_B T} \right)^{1.5}$$

therefore, the spring constant between those beads is given by

$$\hat{k} = \frac{2\pi K_B T}{\hat{P}^{1.5}}$$

the ratio between the spring constants k and \hat{k} is

$$\frac{k}{\hat{k}} = \frac{1.5\hat{P}^{1.5}}{\pi b^2} = \left(\frac{\hat{P}}{E[P(1)]} \right)^{1.5}$$

since $E[P(1)] < \hat{P}$ we get $k > \hat{k}$, so the spring corresponding to the observation \hat{P} is more restrictive.

The ratio above can also guide us in simulations when assigning spring constants \hat{k} for some observed probability $\hat{P} > E[P(1)]$, that can be written in terms of $P(1)$.

In simulation we can fix $b^2 = 3 = \dim$ (in 3D), $K_B T = 1$ hence we get $k = 3K_B T/b^2 = 1$. We can now estimate $E[P(1)] = (1.5/(\pi b^2))^{1.5} = (2\pi)^{-1.5}$ and plug it into the relation $\hat{k}/k = (\hat{P}/E[P(1)])^{1.5} = (\hat{P}/(2\pi)^{-1.5})^{1.5} = \frac{\hat{P}^{1.5}}{2\pi} = \hat{k}$. The probability $\hat{P} \sim E[P(1)]$ and therefore, $\hat{k} = \gamma E[P(1)]/(2\pi)$ $\gamma = \text{const}$

2.11 Fitting the encounter data

In this and subsequent sections we work with the first replicate of the data.

To examine the behavior of the polymer as a whole, we fit for each bead's encounter data a function of the form

$$e_j(d) = \frac{1}{\sum_{i=1}^{N_j} d_i^{-\beta}} d^{-\beta} \quad (2.1)$$

e_j is the normalized encounter frequency of bead j , d is the distance in bead units, and β is a parameter to be determined by the fitting process.

The fitted exponent (β) value obtained by minimizing the error in the model 2.1 are shown in Figure 2.4

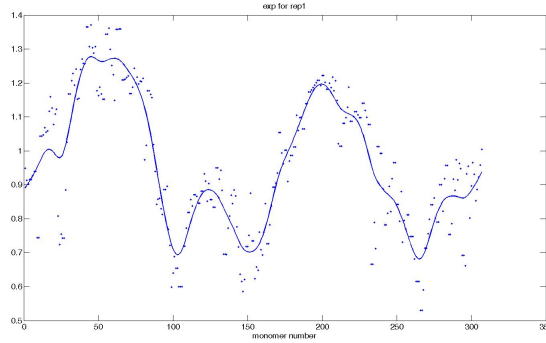


Figure 2.4: The fitted β value for each bead 1 to 307. The continuous curve represents a smoothing of the fitted exponent values using a smoothing spline with tolerance of 2 (see Matlab documentation on [spaps](#) function)

In Figure 2.4, we see that the fitted exponent curve is somewhat tri-modal. With one peak located in roughly at bead 54, one at 125, and the

left most at 200. The first and the last places correspond the mid points of the two TADs of the encounter.

Chapter 3

Simulations

3.1 The Simulation Framework

We have constructed a computational stochastic simulation framework, which allows us to examine the behavior of the polymer chain in various configurations. We use the [Rouse model](#) to mathematically describe the polymer chain as a series of beads connected by harmonic springs (see figure [3.1](#)). The stochastic differential equation describing the dynamics of bead n in a chain of N beads is

$$\dot{r}_n = -D\nabla_{r_n}\phi_{Rouse} + \sqrt{2D}\dot{w}_n$$

with the potential

$$\phi_{Rouse} = \frac{\kappa}{2} \sum_{n=1}^N (r_n - r_{n-1})^2$$

which in the linear chain case gives rise to the set of equations

$$\dot{r}_n = -\frac{d\kappa D}{b^2} (2r_n - r_{n+1} - r_{n-1}) + \sqrt{2D}\dot{w}_n$$

for the inner beads, and

$$\begin{aligned} \dot{r}_1 &= -\frac{d\kappa D}{b^2} (r_1 - r_2) + \sqrt{2D}\dot{w}_1 \\ \dot{r}_N &= -\frac{d\kappa D}{b^2} (r_N - r_{N-1}) + \sqrt{2D}\dot{w}_N \end{aligned}$$

for the end beads. Where r_n are the coordinates of the n^{th} bead, d is the dimension, κ is the spring constant, D is the diffusion constant, b is

the standard-deviation of the distance between adjacent beads, and w is N-dimensional white Gaussian noise with mean 0 and variance 1 in each component.

The flexibility of our framework enables us to simulate a single or an ensemble of chains in open space or in 3 types of domains: a sphere, a cylinder, and between two infinite plates. As an extension to the classical Rouse model, we can also simulate the chain with variable minimal distance, $l_0 > 0$, between adjacent beads. Hence, we modify the Rouse potential to read

$$\phi_{Rouse} = -\frac{\kappa}{2} \sum_{n=1}^N ((\|r_n - r_{n-1}\| - l_0)^2)$$

and simulate the coupled (coordinate-wise) system of equations it yields.

In addition, non-adjacent beads can be fixed to form stable loops throughout the simulation (see Figure 3.1 left panel). In the course of simulation (such as in subsection 3.6), we can randomize the location of these fixed loops, a model we name *random fixed loop model*. We extend this idea to allow the formation and dissociation of loops in time, a model termed dynamic loop model. In this model only when the affine beads are located within some minimal distance ϵ to one another they form a loop with association rate λ per unit of time. Looped affine beads can dissociate with dissociation rate $1 - \lambda$.

At the end of simulations, an analysis is performed to calculate the encounter probability between beads of the chain, and a report is generated.

We consider that two beads R_i and R_j had encountered if

$$dist_{ij} = \|r_i - r_j\| \leq \epsilon$$

with $0 < \epsilon < b$. The theoretical encounter probability in a linear Rouse chain is given by

$$Pr(dist_{ij} < \epsilon) = \alpha|i - j|^{-1.5} \quad (3.1)$$

with $\alpha > 0$.

At the end of simulations we extract the encounter histogram of beads and fit it with a function of the form

$$Pr(dist_{ij} \leq \epsilon) = \alpha_{ij}|i - j|^{-\beta}$$

for each pair of beads in the chain.

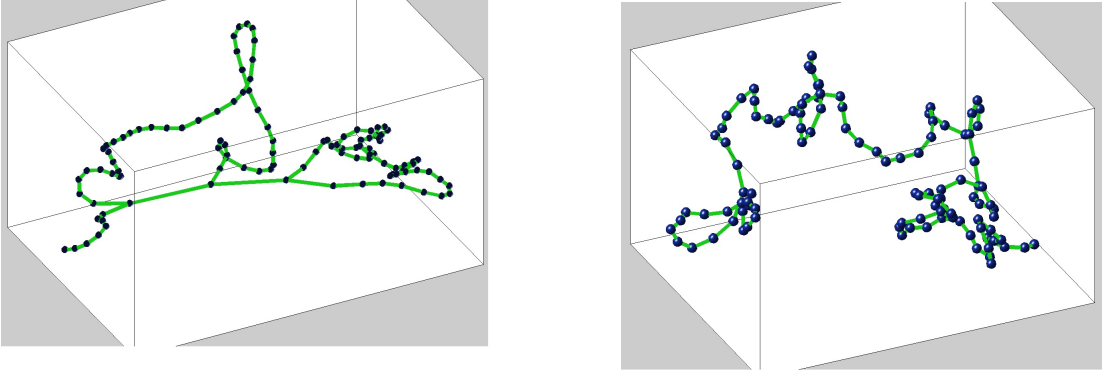


Figure 3.1: Two examples of the beads-springs model. A linear chain model (Right) and a chain with internal 2 loops (Left)

3.2 Initial simulation settings

For the subsequent simulations, unless stated otherwise, we determine the simulation time step, Δt and the simulation time as follows. the simulation step Δt was determined such that it will prevent simulation 'blow-ups'

$$\Delta t = \frac{b^2}{12D}$$

Which is the values that stabilizes the numerical solution to the ode with no noise. when noise is present, this values can be increased.

The *chain relaxation time* is given by

$$\tau_p = \frac{b^2 \Delta t}{12d^2 \sin^2\left(\frac{p\pi}{2N}\right)}$$

where p is the mode number. For simulations we use the relaxation time of the slowest mode $p = 1$.

3.3 Verification of the validity of simulation framework

Here we make sure that the output simulation framework obeys the expected behavior of the Rouse polymer. A polymer of 64 beads, with no cross-linking

is simulated for 40,000 steps, the STD of link length was set to $b = 0.1$. The result is presented in Figure 3.2

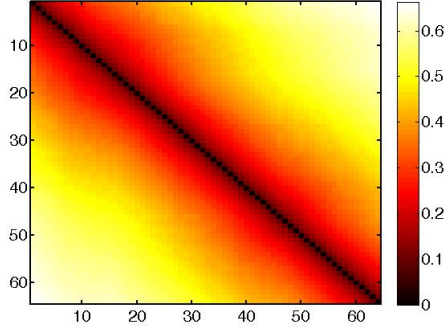


Figure 3.2: The mean bead distance matrix after 50,000 steps of the Rouse model

For an easier comparison of the simulation output with theoretical results, we work from now on with the Root-Mean-Square (RMS) distance between beads.

To see the expected increase in distance, we plot the cross-section of the RMS distance of the first bead to beads 2, 3, ..., 64. The distribution of distances between any two beads n and m of the Rouse polymer, have the following distribution

$$\Phi(R_n - R_m, n - m) = \left[\frac{3}{2\pi b^2 |n - m|} \right]^{3/2} \exp \left[-\frac{3(R_n - R_m)^2}{2|n - m|b^2} \right] \quad (3.2)$$

Where, R_n is the vector of coordinates of bead n . The equation above is the consequence of the fact that the sum of normally distributed random variables is again a normally distributed random variable. In addition, it can be shown that

$$\langle (R_n - R_m)^2 \rangle = |n - m|b^2 \quad (3.3)$$

Using the result above we compare the RMS distance between bead 1 and beads 2, 3, ..., 64 with the function $|n - m|b^2$, appearing in equation 3.3, with $b = 0.1$. As can be seen from Figure 3.3, there is an excellent agreement between the output of the simulations and the theoretical results.

3.3. VERIFICATION OF THE VALIDITY OF SIMULATION FRAMEWORK23

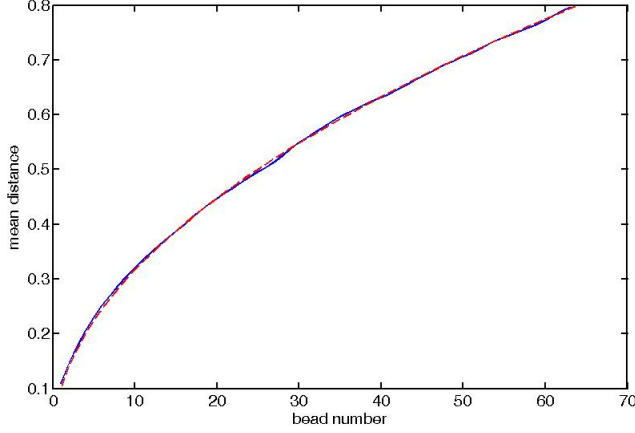


Figure 3.3: The RMS distance of bead 1 to all other 63 beads is plotted (blue curve) vs. the theoretical result (dashed red curve) $dist(1, n) = nb^2$. As can be seen, there is an excellent agreement between the curves, indicating that the simulation framework operates as expected

The RMS distance between adjacent bead was found to be 0.105, in agreement with the bond length $b = 0.1$. The mean square end-to-end distance was found to be 0.6319, in agreement with the expected 0.64. Overall we can conclude that the system operates as expected.

3.3.1 The encounter frequency

According to the theory, the encounter frequency of a bead m in a Rouse chain as a function of the distance of beads from it on the linear chain, should be proportional to $1/|m - n|$, where $n \neq m$ is the distance in bead units. To see that the simulation framework obeys this rule, the encounter frequency matrix was calculated. The encounter distance was set to 0.1, the matrix is presented in Figure 3.4. To see the decay of the encounter matrix, we plot the normalized encounter events of bead 1, and present the fit of the function

$$y = ax^{-b} \quad (3.4)$$

The encounter data was normalized prior to fitting by dividing it by the sum of encounter of bead 1. The parameters fitted were found to be $a = 0.2436$ and $b = 1.036$, in agreement with the expected $e(1) \propto 1/n$

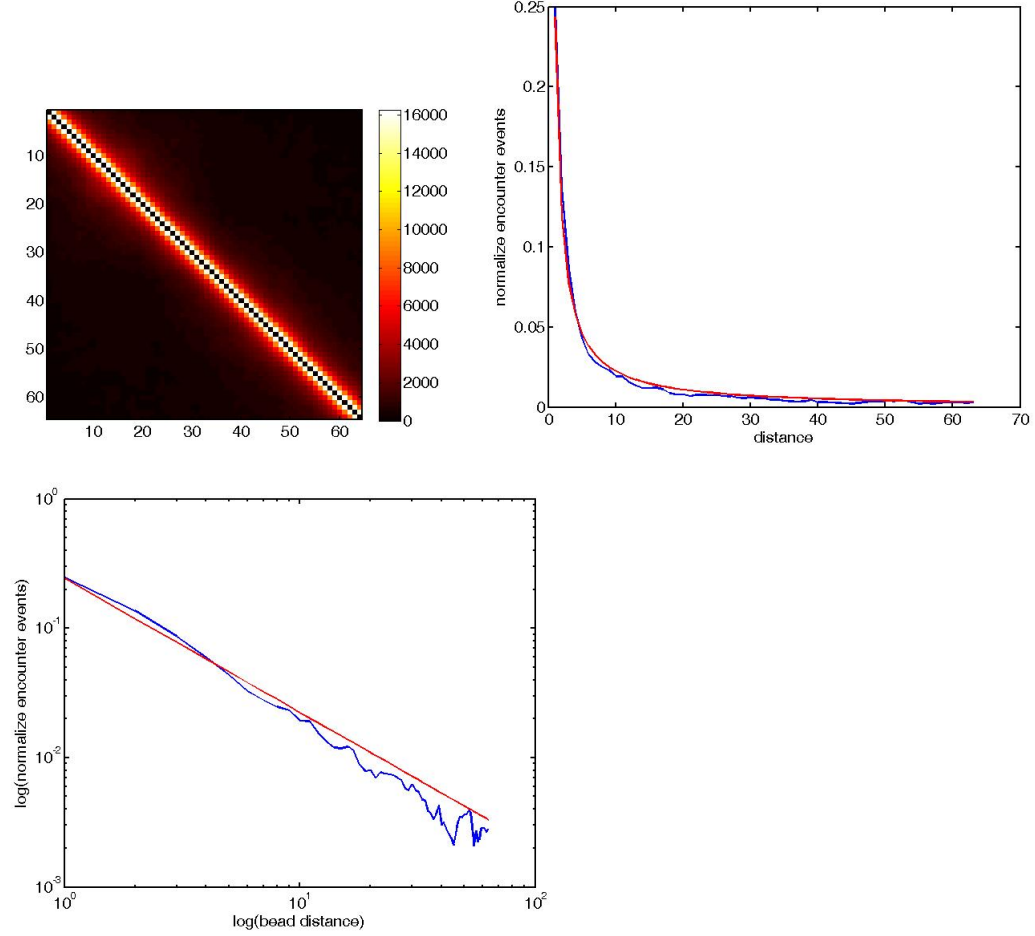


Figure 3.4: The encounter frequency and verification of theoretical results. The encounter matrix (top left) for a 64 beads Rouse chain. The decay of the rows of the encounter matrix is presented for the first bead (top right) for which we have the most number of beads (blue line) and a fit to the encounter curve (red line). the encounter frequency of bead 1 in a log-log scale (bottom left, blue curve) and a fit according to the theory (red line).

3.4 A polymer with a single cross-link

Next, we calculate the distance matrix for a 64 bead polymer for which bead 24 and 44 are cross-linked, creating a loop of size 20 [beads]. The simulation is run from relaxation time, a total of 50,000 steps. The relaxation time is

determined according to the formula

$$\tau_p = \frac{b^2 \Delta t}{12d^2 \sin^2\left(\frac{p\pi}{2N}\right)} \quad (3.5)$$

For $\Delta t = 10^{-4}$, $p = 1$, $d = \sqrt{2D\Delta t}$, $N = 64$, we get

$$\tau_1 = 57.68 \quad (3.6)$$

minutes, which is the relaxation time of the first Rouse mode derived from the cross-correlation function.

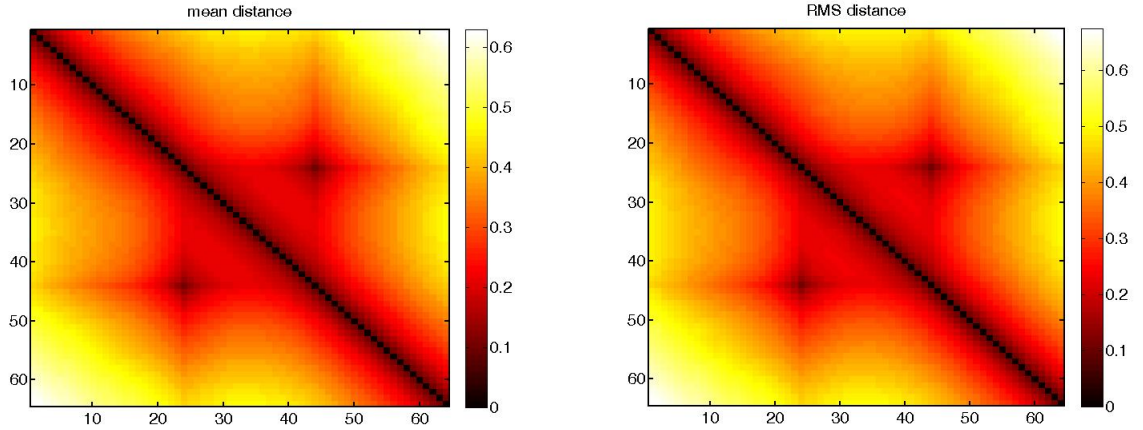


Figure 3.5: mean Distance matrix (left) and RMS distance matrix (right) over 50,000 steps of the simulation. Beads 24 and 44 are connected in a polymer of length 64 beads. As expected, the two matrices display similar qualitative behavior but with slightly different values

To see the decay of bead distance, we plot the cross-sections of the bead-distance matrix presented in Figure 3.5 at the rows 24 and 44. It

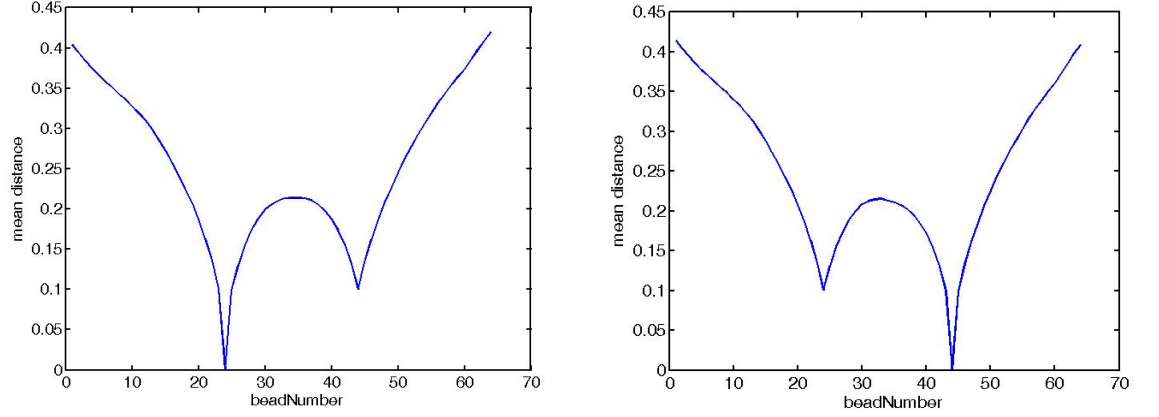


Figure 3.6: The cross section of the mean bead distance matrix presented in Figure 3.5 for bead 24 (left) and bead 44 (right). The height at point 44 (left panel) and point 24 (right panel) is the bond length $b = 0.1$ as expected.

3.4.1 Increasing the encounter distance

Next we examined whether changing the encounter length in the cross-linked chain can explain the appearance of TADs. The encounter frequency between any two beads of the chain is shown in Figure 3.7. The encounter length varied between half and twice of the connector length (0.1).

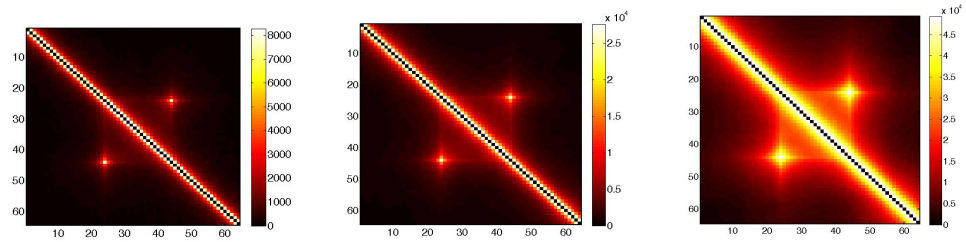


Figure 3.7: Encounter frequency matrices of a Rouse polymer with 64 beads with bead 24 and 44 connected by harmonic potential when varying the encounter distance from 0.05 (left), 0.1 (middle), and 0.2 (right). The bond length was set to $b = 0.1$, the encounter distance was set to 0.1. Simulation ran over 50,000 steps of 10^{-4} seconds. Each pixel represent bead pair (i, j) of the polymer.

The encounter histogram of bead 24 is displayed in Figure 3.8.

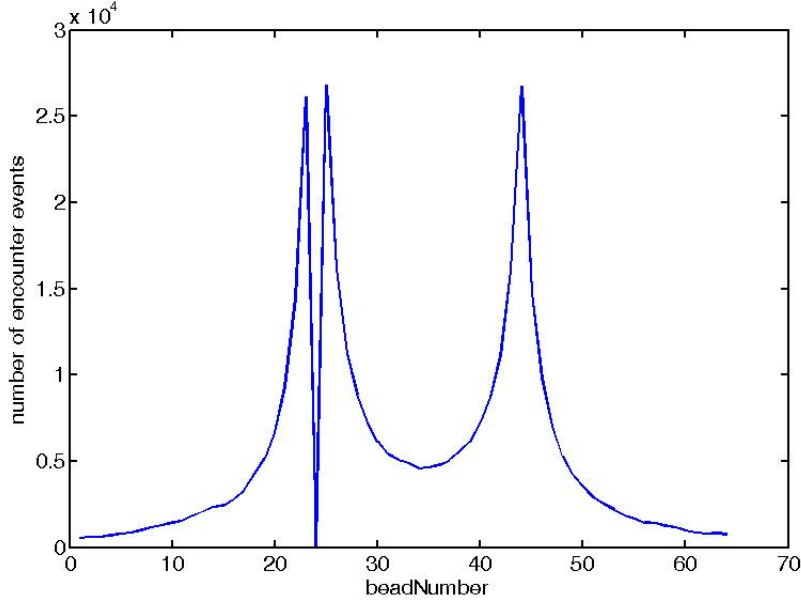


Figure 3.8: the encounter histogram of bead 24 as a function of bead number. The polymer included 64 beads with Kuhn length of 0.1, the encounter distance was set to be 0.1. the self encounters were trivially zeroed out

3.5 One, two and three loops at fixed positions

Here we present *in-silico* experimental results conducted for the case of fixed 1, 2, and 3 loops, i.e. beads connected to form loops remain the same throughout the simulations. We use a polymer composed of 64 beads and sequentially simulate the model with loops between bead [10 19], [28 37] and [46 55] of the linear chain. For each case 10,000 simulations were performed.

The encounter histograms are presented in Figure 3.9. The mean encounter probability calculated for each case is shown in Figure 3.9. The encounter data was fitted with a function of the form $\alpha x^{-\beta}$. The beta values are reported in the legend of Figure 3.9. Since the loops length was set to 9, the encounter probability for the three cases, as a function of distance, shows an increase around distance 9. The values of the probability at that distance are 0.03, 0.034, and 0.035, for 1, 2 and 3 fixed loops respectively.

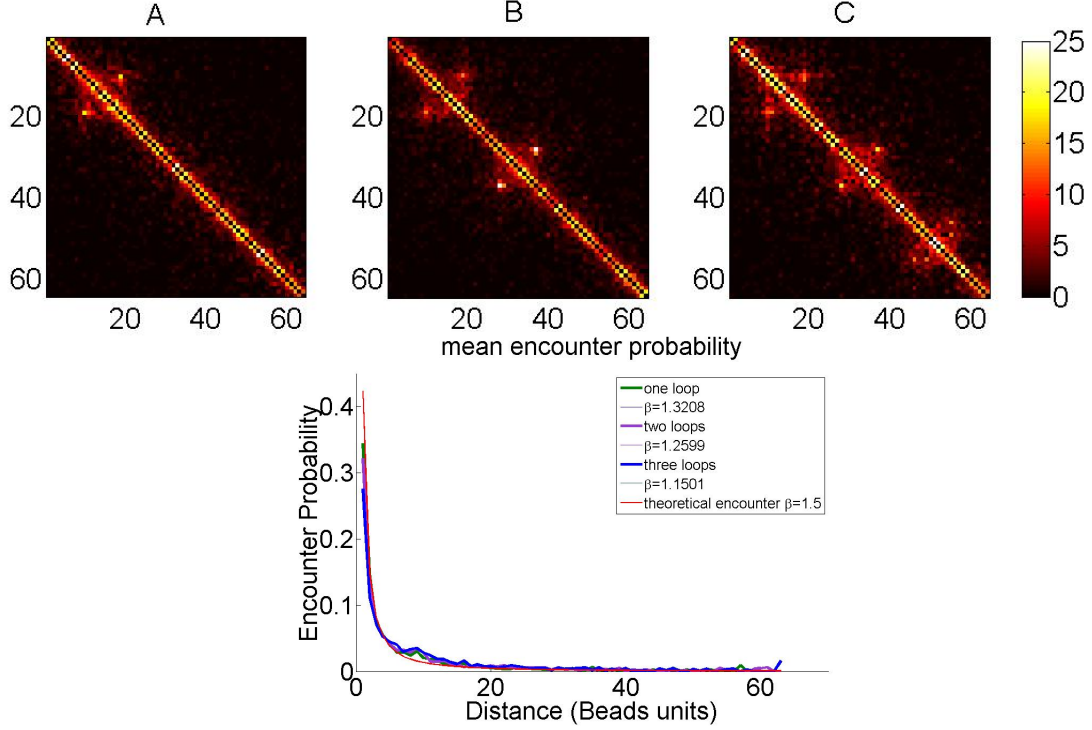


Figure 3.9: The encounter histograms (upper panel) of the 3 fixed position loops cases. (A) a single loop between beads [10 19]. (B) two loops in [10 19], [28 37]. (C) three loops in [10 19], [28 37], and [46 55]. Each point in the 64 by 64 matrices above represent number of encounters between bead i and j . The mean encounter probability (lower panel) for the 3 fixed loops cases: one fixed loop (green), 2 fixed loops (purple), and 3 fixed loops (blue). The fitted β values for the 3 cases are reported in the legend. The theoretical encounter probability curve ($\beta = 1.5$) is shown in red. The bump around distance 9 indicates shows the higher encounter probability for the three cases, since the length of each loop was set to 9.

3.5.1 Internal loops

We now turn to examine the encounter probability of a model containing one 'big' loop when internal loops are sequentially added in it (see Figure 3.10 lower right panel). That is, we connect bead i and j ($j \gg i$) and add n loops at of random lengths between bead $i + 1$ and bead $j - 1$, with the constraint that any bead cannot be connected to form more than one loop.

For the results presented here, we chose the beads 5 and 60 to form the big loop, we perform 10 simulation rounds, in each we increase the number of internal loops from 1 to 10. The connected beads forming the internal loops

are chosen at random. Encounter histogram for the 10 cases are shown in Figure 3.10 (upper panel). As the number of internal loops is increased, the encounter probability between beads in the loop increases.

The calculated mean encounter probabilities for the 10 cases are presented in Figure 3.10 (lower left panel)

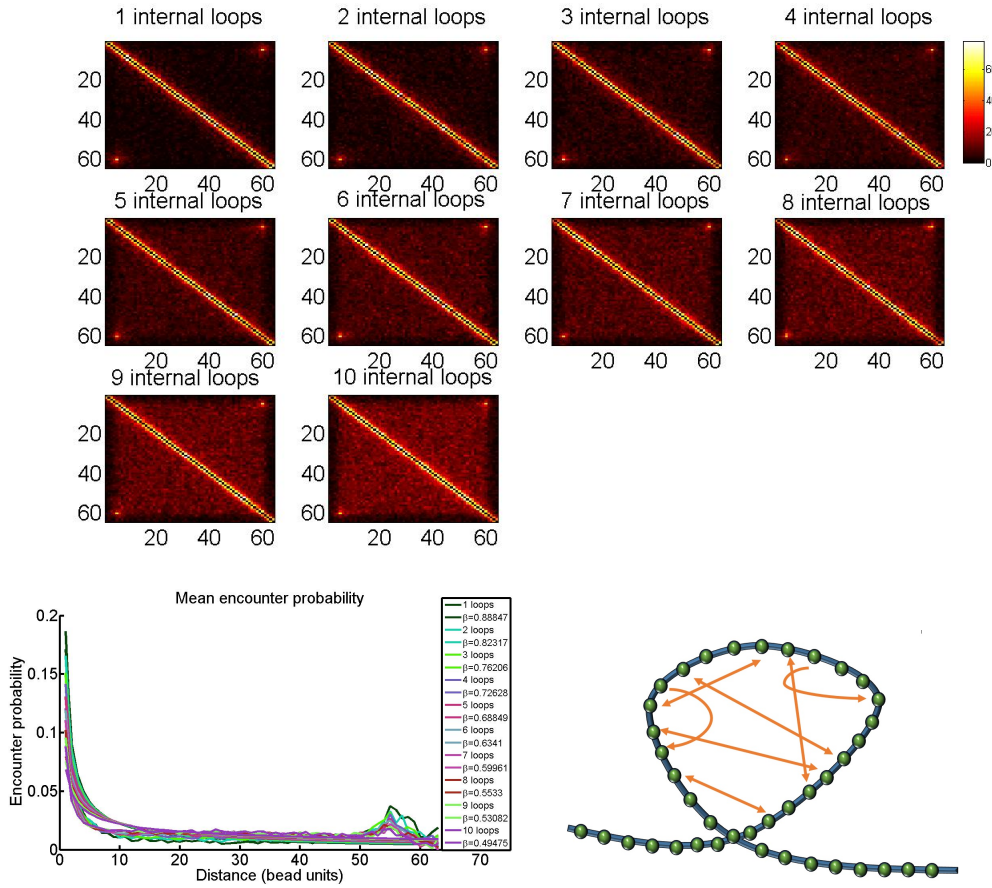


Figure 3.10: The encounter histogram (upper panel) of the model with one big loop between beads 5 and 60, when 1 to 10 internal loops are added to it. The mean encounter probabilities curves (lower left panel) as the number of internal loops is increased from 1 to 10. The simulated encounter data was fitted with a function of the form $\alpha x^{-\beta}$, where $x = 1.63$ is the distance in beads units. The bump in distance 55 represents the connection between bead 5 and 59, which constitute the 'big' loop in our simulations and is seen as a consistent bright spot in the encounter histogram (upper panel). A cartoon of the looped model (lower right panel) with internal loops, arrows indicate the beads to be connected. In the simulation in 3.5.1 we simulate the polymer in the loop region, without the tails

3.6 Random fixed loops

In this experiment we add random length loops between bead i and j . We term the range of beads between i and j as a TAD.

3.6.1 One TAD

To simulate a TAD, we employ the random fixed loop model and restrict the bead pairs forming the loops to take indices between beads 1 to 32, the second half of the chain remains a linear Rouse chain. No bead can participate in forming more than one loop. We sequentially add 1 to 8 random fixed loops in the region of the TAD, matching the case of 2 to 16 beads participating in loop formation.

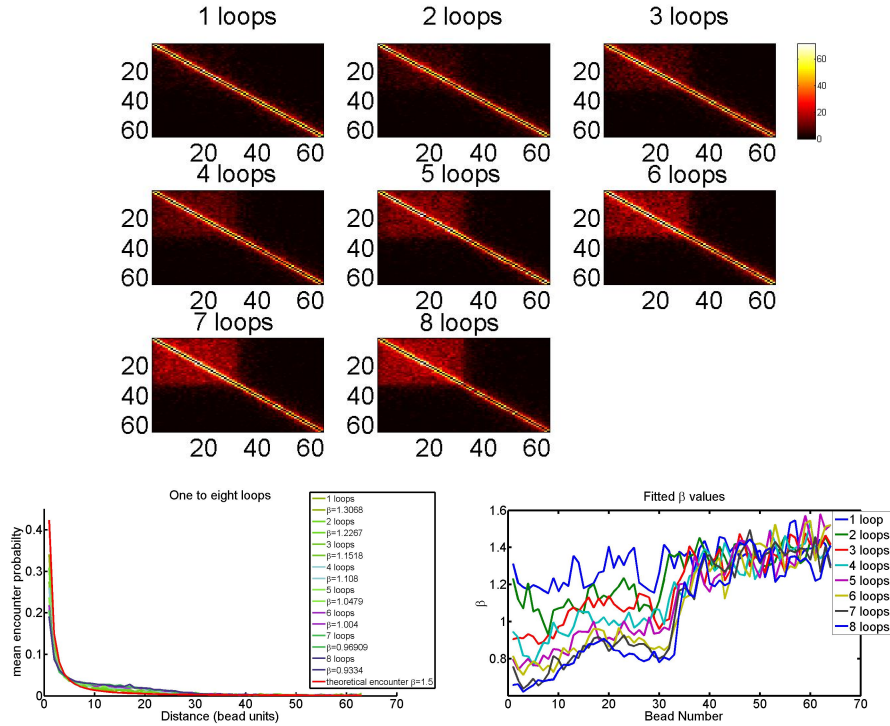


Figure 3.11: The encounter histogram (upper panel) of the case of random fixed loops in the region 1 to 32. As the number of random loops increases from 1 to 8 we see the emergence of a TAD region. The mean encounter probability (lower left panel) of the eight cases. The decrease in β as the number of loops increases is attributed to the increased encounter between the first 32 beads. The fitted β values (lower right panel) for the encounter data of each bead is displayed for the 8 cases.

3.6.2 Two TADS

We follow similar procedure as with the one TAD simulation, only now we create random loops in two regions to form two TADs. The first regions is defined between bead 1 to 32, the second region between bead 33 to 64. In each simulation round we sequentially increase the number of random fixed loops from 1 to 10 (Figure 3.12).

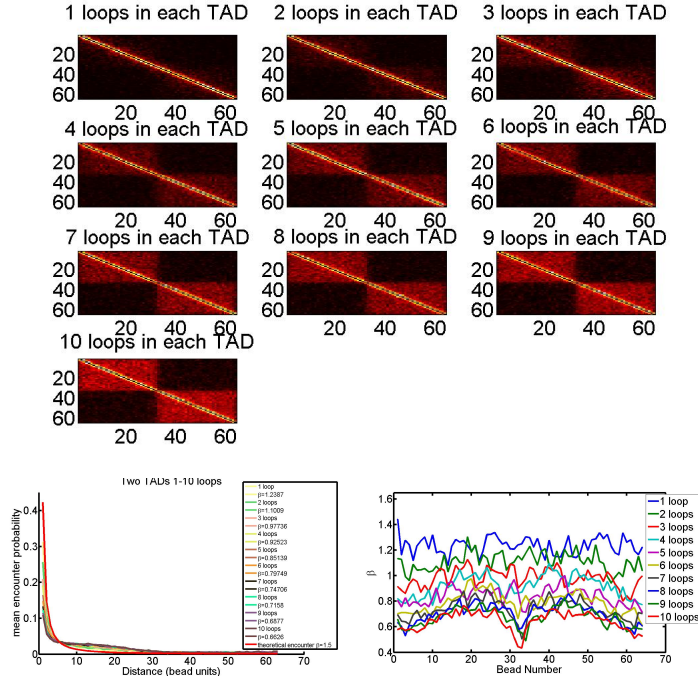


Figure 3.12: The encounter histograms (upper panel) of the case with two TADs. As the number of loops in each TAD increases from 1 to 10, the two TADs emerges. The mean encounter probability (lower left panel) of the 10 cases resembles that of the one TAD case due to 32 beads size of the TAD. The fitted β values for the encounter probability of each bead in the 10 cases (lower right panel).

3.7 Simulation of variable loop number

3.7.1 One TAD

To examine the effect of increasing the number of stable loops on the encounter probability, a chain of 32 beads was constructed. The number of stable loops varied from 0 to 16. the bond length $b = 0.1$, the encounter

distance was set to $b/2 = 0.05$, the time step was set to prevent simulation blow ups by the formula above, $\Delta t = 8.3 \times 10^{-4}$. For each number of loops, random pairs of beads were chosen to be connected.

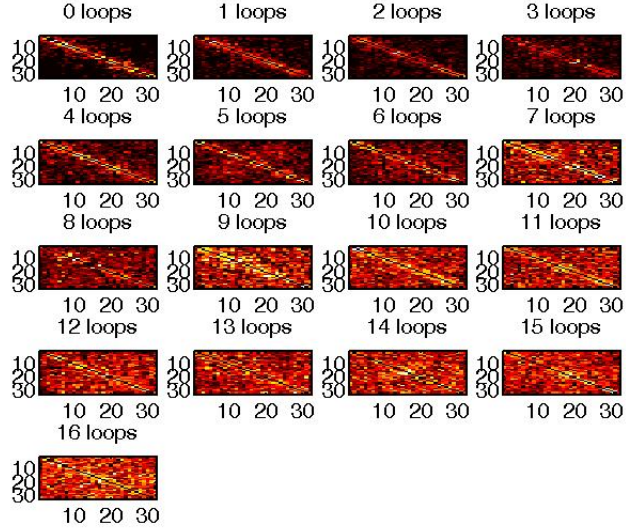
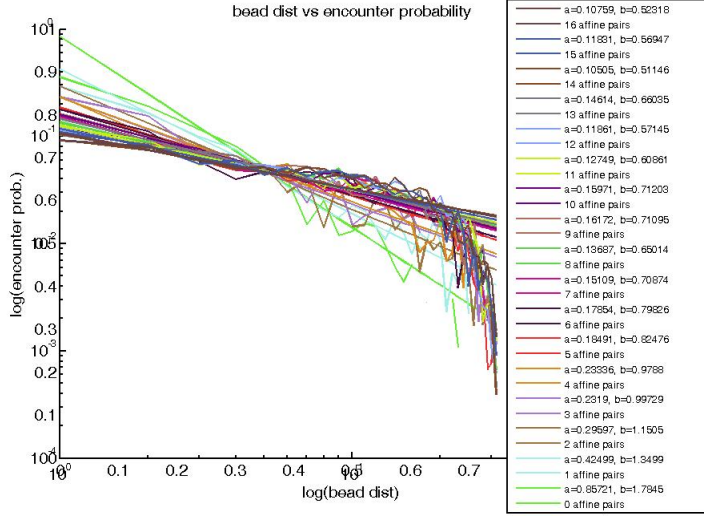
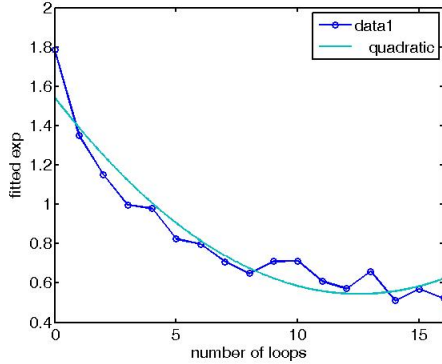


Figure 3.13: The encounter histogram of a chain of 32 beads. When the number of loops varies from 0 to 16 an area resembling a TAD emerges

for each choice of loop number, we fit a curve to the encounter probability as a function of distance [beads].



The fitted exponent values are decreasing roughly quadratically.



One TAD with a 'tail'

In this simulation we test the affect of the increase in the number of stable loops on the bead encounter probability. An ensemble of chains comprised of 64 beads were simulated to 1.2 times the relaxation time of the first Rouse mode, resulting in about 16,000 steps for each chain. For each loop number 300 simulation were carried out. The bond length was set to be $b = 0.1$, the encounter distance $\epsilon = b/2$, time step was set to be $\delta t = 10^{-4}$. Stable loops were defined between bead 1 and 32 which we expect to resemble a TAD. At each simulation round we increase the number of loops L by choosing at

random $2L$ bead indexes of bead to be connected, L is varied from 0 to 15. the encounter histogram shows an increasing encounter pattern in the region the loops are defined in (see Figure 3.14)

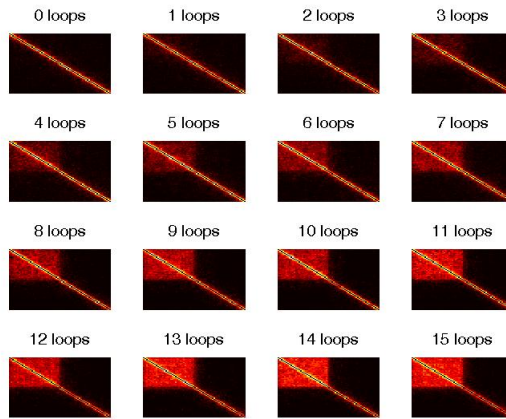


Figure 3.14: The affect of increasing the stable loops in a polymer of 64 beads. Loops were created at random between beads located in the first half of the polymer. Increasing the number of loops from 0 to 15 reveals a region resembling a TAD as seen in the data

The encounter probabilities for the chains when increasing the loops from 0 to 15, can be seen in Figure 3.15 For each encounter probability curve we fit a function of the form ad^{-b} . The change in the exponent b as a function of the number of loops is seen in Figure 3.16

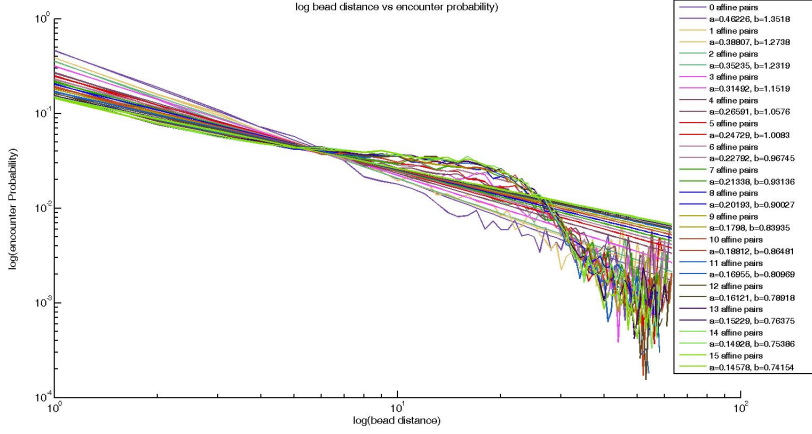


Figure 3.15: Encounter probability vs bead distance for the 16 models representing variable stable loops number if the first half of the chain of 64 beads (top)

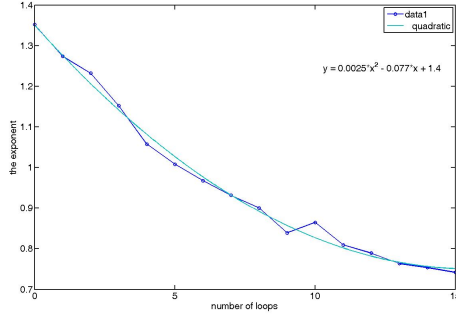


Figure 3.16: The change in the value of the fitted exponent to the encounter probability curves of models with increasing number of stable loops. The decrease in the values of b quadratic as can be seen by the cyan line fitted to the data.

3.7.2 Two TADs

3.8 Dynamic loop model

It can be assumed that stiff fixed loops in the chromatin are not always stable, but can open and close transiently. We turn to examine whether a model with loops that can be formed, held fixed, for some period, and be released, can explain the appearance of the TADS. The rational behind is

that the variability of cells used in the experiments some in different states, i.e. are found with different initial chromatin loops. When these cells are fixed and the chromatin is cross-linked, the cross-links that will be formed will depend on the proximity of the beads conditional on the loops present at the moment the experiment began.

A polymer with 128 beads was simulated. We define two region with high affinity of their member beads to a specific point in them. These regions can be considered as the TADs. Each region is of size 25 beads. The first region is defined to be from bead 1 to 25, the second from bead 77 to 102.

To simulate the state of experiment initiation, we define two beads in each TAD, which have high probability to be found together, and run simulations until steady state. Only one pair from each TAD was examined in each simulation, which we vary from simulation to simulation. At the end of each simulation we retrieve the encounters frequency for all the beads.

Beads that come in close proximity to one another, are considered attached, and can be released and some rate k_{off} . The attachment distance was defined as half of the connector length between adjacent beads.

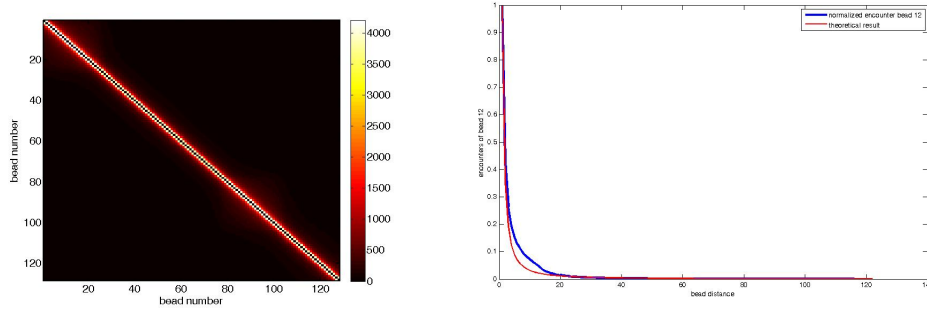


Figure 3.17: The mean encounter matrix for the transient loop model. the chain included 128 beads, with connector length of 0.1. Two regions were defined as the TADs: Bead 1 to bead 25, and bead 77 to 102. at each simulation, a pair of beads from each region were considered as having some affinity to one another. The model was simulated for each choice of affine pair until steady-state, 30000 steps. The mean encounter of all simulation in presented.

For beads that are not affine to any other, like the beads 26 to 77, the encounter frequency retrieved matches the theoretical result.

3.9 TAD D and E with cross-links corresponding to the peaks of the experimental data

We next wanted to verify whether using a Rouse polymer and connecting beads for which the peaks of the experimental data, would allow us to reproduce the TADs.

A polymer with 307 beads was created, corresponding to the TAD D (107 beads) and E (200 beads). The cross-links were made at the mid point of bead numbers displayed in Table ?? . Thirty simulations were performed and their average distance matrix after relaxation time is presented in Figure 3.18. As can be seen the location of the cross-links create 'hot spots' in the mean distance matrix. In comparison to Figure ??, we see that there is little resemblance. Thus concluding that creating cross-links according to the encounter matrix, is not sufficient to reproduce the TADs.

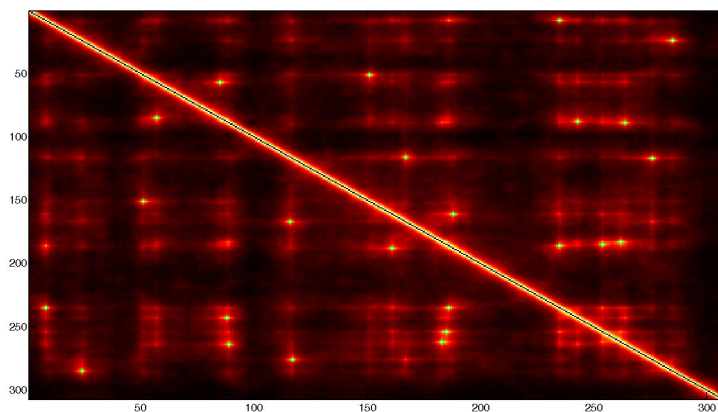


Figure 3.18: The mean of 30 simulation of the polymer with 307 beads, connected corresponding to the peaks of the encounter data (Luca et al). The small green circles represent locations of connected bead pairs. These circles overlap with peaks of the simulated encounter matrix. It is evident that the TAD structure was not reproduced with this approach. The x and y axes represent bead numbers. Each pixel represents the encounter number of bead i and j

3.10 Conditional encounter probability

3.10.1 Three beads in a linear chain

In the analysis of the experimental data as performed by Giorgetti et al[8], 3 loci were found to be significant to the internal organization of TAD D. These loci overlap the position of the highly interacting Xite/Tsix, Chic1, and Linx, which correspond to beads in the range 25-27, 33-35, and 86-89 respectively.

We therefore simulate a 108 beads chain and assess the conditional probability that bead 34 (Chic1) encounter bead 26 (Xite/Tsix) before bead 87 (Linx). All simulations were ran until an encounter occurred. Time step was set to 8.3×10^{-5} [sec]. The encounter distance was set to $b/5 = 0.02$, where b is the standard-deviation of the distance between adjacent beads.

The conditional encounter probability is summarized in the following table

prob loci	bead num.	prob	MFET [sec]
Chic1 \rightarrow Xite/Tsix	34 \rightarrow 26	0.9378	0.2273
Chic1 \rightarrow Linx	34 \rightarrow 87	0.0622	0.1903

Not that we work in units of $k_B T$, hence the times should be multiplied by this factor.

3.10.2 Two beads in the random fixed loops model

Next, we examine the conditional encounter probability of the 3 loci (beads 26, 34, and 87) in the context of the random fixed loops model as described in subsection 3.6. To simulate the D TAD, we use a chain of 108 beads, and run 10,000 simulations with a time step of 8.3×10^{-5} [sec], and an encounter distance of $b/5 = 0.02$. We set 10 randomly placed fixed loops for each simulation, where the beads 26, 34, and 87 are prohibited from forming loops. The conditional probability of bead 34 meeting bead 26 before it meets bead 87 is summarized Table 3.1 along with the mean first encounter time.

prob loci	bead num.	prob	MFET [sec]
Chic1 \rightarrow Xite/Tsix	34 \rightarrow 26	0.725	0.1154
Chic1 \rightarrow Linx	34 \rightarrow 87	0.275	0.1422

Table 3.1:

3.10.3 Two beads and a loop

We now turn to explore the conditional probability that a bead B encounters bead A before it encounter bead C in a chain containing one loop. There are 27 configurations for the position of the beads in relation to the loop. Here we address the case of a chain with two equally sized tails. Thus, due to symmetry we reduce the possible configurations to 14. We term the beads contained in the loop as *in the loop* and beads outside as beads *on the tail*. The possible configurations are:

1. A in the loop, B and C on the same tail
2. A in the loop, B and C on different tails
3. B in the loop, A and C on the same tail
4. B in the loop, A and C on different tails
5. C in the loop, A and B on the same tail
6. C in the loop, A and B on different tails
7. A and B in the loop, C on the tail
8. A and C in the loop, B on the tail
9. B and C in the loop, A on the tail
10. A , B and C in the loop
11. A , B and C are on the tail
12. A , B on one tail, C on the other tail
13. A , C on one tail, B on the other tail
14. B , C on one tail, A on the other tail

The different cases are summarized graphically in Figure 3.19

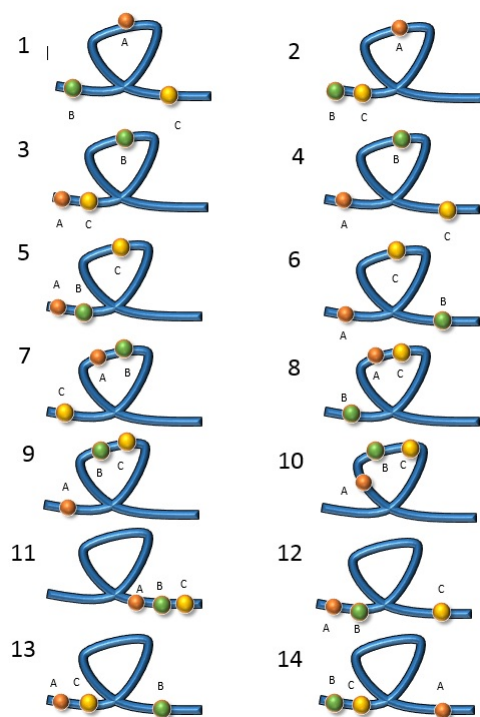


Figure 3.19: The possible configuration in the case of three beads and a loop, leaving tails of equal length

The conditional probability encounters and the mean first encounter time, are summarized in the Table 3.2 according to the cases specified above

case	Bead A-B	MFET A-B (sec)	Bead B-C	MFET B-C (sec)
1	0.15	0.34	0.85	0.23
2	0.76	3.97	0.23	3.11
3	0.37	1.47	0.63	1.46
4	0.53	1.79	0.47	2.25
5	0.86	0.29	0.16	0.15
6	0.29	2.57	0.71	4.14
7	0.79	0.55	0.21	0.45
8	0.63	1.83	0.37	2.20
9	0.26	0.46	0.74	0.81
10	0.52	0.13	0.48	0.15
11	0.54	0.14	0.46	0.11
12	0.97	0.29	0.03	0.13
13	0.56	8.39	0.44	6.19
14	0.01	0.02	0.99	0.3

Table 3.2: Summary of the results of simulating the conditional probability that bead B meets A before C (Bead A-B), and probability that bead B meets C before A (Bead B-C), with the mean first encounter time (MFET) given in units of seconds. For each case 10 simulation were performed

Chapter 4

Review of Literature

Here we give several key points from articles bearing important principals from the understanding of chromosome internal interaction and organization

4.1 From Nora et.al 2012 [[14](#)]

- The X inactivation center (Xic) orchestrate the initiation of the X chromosome inactivation by controlling the expression of non-protein-coding Xist transcript.
- 5C techniques were used to analyze the *spatial* organization of the region including the Xist. The region analyzed is 4.5 Mb in size.
- 200 kb to 1 Mb regions were discovered and named TAD.
- TADs remain consistent between male and female Embryonic stem cells (ESC) and between different cell lines.
- TADs remain persistent before and after cell differentiation and before and after X inactivation. Internal TAD connections remain high even after X inactivation. However, the intra-TAD connections are lower than those of the same TAD on the active X.
- on the inactive X, the long range interaction(> 50kb) inside the TAD are lost.
- disrupting the regions between TADs caused an ectopic chromosomal contacts
- 5C-Forward and 5C-Reverse oligonucleotides were designed to conduct test on undifferentiated mouse embryonic stem cells (ESCs)

- it was found that long range contacts ($>50\text{kb}$) happen within a series of discrete genomic blocks, roughly 0.2-1 Mb in size
- The chromosomal domains are self-associating.
- FISH experiment on 3D chromosomal conformation were conducted to verify the results.
- in the FISH, the distance between probes lying in the same 5C domain were significantly shorter than of probes lying in different domains.
- a strong correlation was found between 5C count and 3D distance.
- bacterial artificial chromosome probes showed in experiments that large DNA segments belonging to the same 5C domain co-localize much more than DNA segment on adjacent domains *throughout cell cycle*.
- a correlation was shown between the existence of H3K27me and the TADs. However, for a mouse with a knockout of these proteins, a significant change was *not* observed in the composition and size of the TADs.
- it was shown that H3K27me - an antibody signifying the presence of Histone 3- is not related to the folding of the genomic regions into TADs.
- when the boundary between TAD E and D (which include the Xist and Tsix), ectopic folding of TAD E was observed. It was assumed that boundary elements have an influence on the conformation of the TADs.
- after removing boundary elements between TADs, the TADs did not merge completely, hence concluded that there are other, intra-TADs element that control conformation.
- TAD did not change during cell differentiation (from embryonic stem cell to neuronal progenitor and embryonic fibroblast), but the *internal* connections inside each TAD did change.
- for the TADs that did change, some have become [Lamina Associated Domains](#) (LADs)
- the transcription dynamics showed positive correlation for TADs for which the promoter of the genes was located inside the TAD.
- the correlation between gene expression of genes in the same TAD did not depend on the distance between genes in the TAD.

4.2 From Dekker et al 2013 [2]

1. This is a review of methods to analyze Chromosome Capture data and to infer mechanical properties of the DNA.
2. The 3C methods determines the encounter frequency of genomic segments, probably in the range of $10 - 100nm$
3. Chromosome are highly variable among cells [13]
4. There are some organizational principals at the scale of the whole nucleus
5. chromosomes occupy separate territories that do not usually mix[12]. When they do mix, they can potentially create functional interactions between loci on different chromosomes.
6. transcription does not occur diffusively throughout the nucleus, but happen at sub-nuclear sites enriched in components of the transcription and RNA processing machinery.
7. Therefore, actively transcribed genes tend to co-localize
8. transcriptionally inactive segments tend to associate with each other and can often be found localized at the nuclear periphery
9. sub-nuclear positioning of the loci is correlated with gene expression
10. imaging approaches do not allow analysis of the 3D folding of a complete genome at the kb resolution.
11. the 3C based methods do not distinguish functional from non-functional association of loci
12. the loci encounter frequency is affected by
 - (a) direct loci encounter
 - (b) encounters mediated by proteins
 - (c) indirect co-localization to the same sub-nuclear structure (e.g lamina)
 - (d) result of chromosome packing and folding
 - (e) random encounters
 - (f) interactions defined by the polymer physical characteristics
13. the 3D chromatin is highly variable even among identical cells[12].
14. analytical tools to interpret the 3c data focus first on specific point by-point looping interactions, e.g. between promoters and gene regulatory

elements

15. in previous article, thousands of long range ($> 4mb$) interactions were identified between promoters and enhancer-like regions [15]
16. one abundant class of long range interaction involves promoters looping to sites bound by the insulator protein CTCF. The hypothesis is that these looping events have architectural role.
17. genes are regulated by multiple distal elements [6]
18. average pattern of looping interactions around promoters is asymmetric. Looping interaction are most frequently observed $\sim 120kn$ upstream
19. human and mouse genome are each composed of over 2000 TADs covering 90% of the genome.
20. TADs are defined by genetically encoded boundary elements [14].
21. boundary elements are enriched with CTCF-bound loci
22. the mechanisms that establish TAD boundaries are still unclear.
23. analysis of mouse genome suggests tat enhancer-promoter interactions are particularly frequent within TADs [16]
24. genes within the same TAD tend to be coordinately expressed during cell differentiation [14]. Possibly because those genes share the same set of regulatory elements.
25. the 3D structure of the Igh locus was inferred by polymer modeling and genome tagging to give discrete areas of looping [10].
26. the 3D structure of bacterial genome was determined by a combination of 5C data, imaging and modeling [19]
27. in yeast it was demonstrated that a small set of spatial constraints is sufficient to yield a highly organized genome structure, i.e volume exclusion [18]
28. HiC data shows lack of specific interactions for loci $> 1Mb$ apart
29. HiC data for non-synchronized human cells show 3 regimes, each exhibiting a power law decline in the contact probability $P(s) \sim s^{-\alpha}$. for $s < 0.7Mb$ $\alpha \approx 0.7$, corresponding to the TADs. For $0.7Mb < s < 10Mb$, $\alpha \approx 1$. For $s > 10Mb$, α is not well measures due to poor statistics [11].
30. a polymer model with excluded volume can give rise to an encounter probability with $\alpha = 1$

31.

4.3 The Hierarchy of the 3D genome. Gibcus 2013 [7]

1. neighboring chromosomes can overlap considerably
2. loci on the same chromosome interact much more frequently than loci on different chromosomes.
3. the nuclear envelope (double lipid bilayer) restricts genomic DNA to confined 3D space and provides a solid anchor point that allows for specific chromatin interactions.
4. The nuclear lamina is the inner nucleus membrane is associated with inactive heterochromatic chromatin directly or indirectly via lamina associated proteins.
5. almost half of the genome in a given cell population is composed of lamina associated domains [9].
6. activated genes move to the nuclear interior and inactive genes are found in the lamina associated domains (how do they move there?)
7. loci gather near sub-nuclear structures such as the nucleoli, which are sub-nuclear structures dedicated to expression by RNA polymerase I.
8. it has been shown that open, gene rich, areas and closed, gene-poor chromosomal domain are generally located in separate sub-nuclear regions [5]
9. sub-chromosomal compartment were detected, named A and B, according to functional compartmentalization of gene activity [11].
10. this compartmentalization is related to the association of loci to transcription factories and the nuclear lamina.
11. loci found clustered in the A compartment are generally gene rich, transcriptionally active and DNase I hypersensitive
12. loci found clustered in the B compartment are generally gene poor, transcriptionally silent, and DNase I insensitive [11].
13. those compartments are cell type specific, and cannot be predicted according to gene density.
14. Chromatin domains preferentially associate with other domains of a comparable activity level.

15. The actively transcribed genes colocalize at foci called [transcription factories](#)
16. groups of active genes, sometimes from different chromosome are found in these foci, enriched in the activity of RNA polymerase and other transcription factors.
17. genes located within the same TAD tend to have coordinated dynamics of expression during differentiation
18. so far TADs have not been described in bacteria, yeast, or plants.
19. TADs are a conserved feature of cells in the mammalian cells
20. it is currently not clear what defines TAD boundary
21. TAD boundaries are enriched in [CTCF](#) sites
22. CTCF at the boundaries of the TAD could imply CTCF protein its role in mediating and blocking long range interactions
23. however, most CTCF binding sites are within TAD, so it is not sufficient to define TAD boundary [\[14\]](#)
24. it was suggested that CTCF recruits cohesin complexes.
25. the mechanism by which long range interactions are mediated is largely unknown
26. specific long range interactions recruit specific proteins to the gene promoter, and recruit RNA polymerase.
27. most specific long range interactions between promoters and enhancers can be found within the boundaries of the TADs.
28. there is a striking cell-to-cell diversity of subnuclear positioning of loci.
29. at the level of the whole nucleus, the spatial organization of the genome is rather static (chromosomes do not move a lot)
30. the positioning of the chromosome varies greatly between cells, but remains stable for any given cell.
31. at the scale of few hundreds of kb the chromatin is much more dynamic.
32. loci may interact with other loci within a radius of 1Mb or so.
- 33.

Bibliography

- [1] Elzo de Wit and Wouter de Laat. A decade of 3c technologies: insights into nuclear organization. *Genes & development*, 26(1):11–24, 2012.
- [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] Job Dekker, Karsten Rippe, Martijn Dekker, and Nancy Kleckner. Capturing chromosome conformation. *science*, 295(5558):1306–1311, 2002.
- [4] Josée Dostie, Todd A Richmond, Ramy A Arnaout, Rebecca R Selzer, William L Lee, Tracey A Honan, Eric D Rubio, Anton Krumm, Justin Lamb, Chad Nusbaum, et al. Chromosome conformation capture carbon copy (5c): a massively parallel solution for mapping interactions between genomic elements. *Genome research*, 16(10):1299–1309, 2006.
- [5] Peter Fraser and Wendy Bickmore. Nuclear organization of the genome and the potential for gene regulation. *Nature*, 447(7143):413–417, 2007.
- [6] Mark B Gerstein, Anshul Kundaje, Manoj Hariharan, Stephen G Landt, Koon-Kiu Yan, Chao Cheng, Xinmeng Jasmine Mu, Ekta Khurana, Joel Rozowsky, Roger Alexander, et al. Architecture of the human regulatory network derived from encode data. *Nature*, 489(7414):91–100, 2012.
- [7] Johan H Gibcus and Job Dekker. The hierarchy of the 3d genome. *Molecular cell*, 49(5):773–782, 2013.
- [8] Luca Giorgetti, Rafael Galupa, Elphège P Nora, Tristan Piolot, France Lam, Job Dekker, Guido Tian, and Edith Heard. Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell*, 157(4):950–963, 2014.

- [9] Lars Guelen, Ludo Pagie, Emilie Brasset, Wouter Meuleman, Marius B Faza, Wendy Talhout, Bert H Eussen, Annelies de Klein, Lodewyk Wessels, Wouter de Laat, et al. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature*, 453(7197):948–951, 2008.
- [10] 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.
- [11] Erez Lieberman-Aiden, Nynke L van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragoczy, Agnes Telling, Ido Amit, Bryan R Lajoie, Peter J Sabo, Michael O Dorschner, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *science*, 326(5950):289–293, 2009.
- [12] WF Marshall, A Straight, JF Marko, J Swedlow, A Dernburg, A Belmont, AW Murray, DA Agard, and JW Sedat. Interphase chromosomes undergo constrained diffusional motion in living cells. *Current Biology*, 7(12):930–939, 1997.
- [13] Iris Müller, Shelagh Boyle, Robert H Singer, Wendy A Bickmore, and Jonathan R Chubb. Stable morphology, but dynamic internal reorganisation, of interphase human chromosomes in living cells. *PloS one*, 5(7):e11560, 2010.
- [14] 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, et al. Spatial partitioning of the regulatory landscape of the x-inactivation centre. *Nature*, 485(7398):381–385, 2012.
- [15] Amartya Sanyal, Bryan R Lajoie, Gaurav Jain, and Job Dekker. The long-range interaction landscape of gene promoters. *Nature*, 489(7414):109–113, 2012.
- [16] Yin Shen, Feng Yue, David F McCleary, Zhen Ye, Lee Edsall, Samantha Kuan, Ulrich Wagner, Jesse Dixon, Leonard Lee, Victor V Lobanenko,

- et al. A map of the cis-regulatory sequences in the mouse genome. *Nature*, 488(7409):116–120, 2012.
- [17] John D Storey. A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(3):479–498, 2002.
- [18] Harianto Tjong, Ke Gong, Lin Chen, and Frank Alber. Physical tethering and volume exclusion determine higher-order genome organization in budding yeast. *Genome research*, 22(7):1295–1305, 2012.
- [19] Mark A Umbarger, Esteban Toro, Matthew A Wright, Gregory J Porreca, Davide Bau, Sun-Hae Hong, Michael J Fero, Lihua J Zhu, Marc A Marti-Renom, Harley H McAdams, et al. The three-dimensional architecture of a bacterial genome and its alteration by genetic perturbation. *Molecular cell*, 44(2):252–264, 2011.