# Repair dynamics of clustered DSB damaged chromatin trough coarse-grained polymer models

*Auteur :*
Ignacio Madrid
*Promotion :*
X2015
*Option :*
Département de
Mathématiques appliquées
*Champ :*
MAP594 Modélisation
probabiliste et statistique
*Enseignant référent :*
Pr. Vincent Bansaye

*Tuteur de stage :*
Pr. David Holcman
*Dates du stage :*
26 mars - 31 aout 2018
*Adresse :*
École normale supérieure
Institut de Biologie
46 rue d'Ulm
75005 Paris, France

ÉCOLE **POLYTECHNIQUE**
UNIVERSITÉ PARIS-SACLAY

ENS | PSL★

# Déclaration d'intégrité relative au plagiat

Je soussigné MADRID CANALES Ignacio certifie sur l'honneur :

1. Que les résultats décrits dans ce rapport sont l'aboutissement de mon travail.

2. Que je suis l'auteur de ce rapport.

3. Que je n'ai pas utilisé des sources ou résultats tiers sans clairement les citer et les référencer selon les règles bibliographiques préconisées.

*Je déclare que ce travail ne peut être suspecté de plagiat.*

à Paris, June 27, 2018                    Ignacio MADRID CANALES
**Date**                                  **Signature**

**Résumé**

 Les cassures double-brin (DSBs de par ses sigles en anglais) sont un type d'endommagement de l'ADN particulierement cytotoxique, pour lequel la jonction d'extrémités non homologues (Non Homologous End Joining ou NHEJ) est le mecanisme de reparation prefere en grande partie du cycle cellulaire. Neanmois, ce dernier peut induire d'importantes reorganisations chromosomiques quand les DSBs sont regroupes en clusters. Des polymeres grossieres a reticulations aleatoires (Random Cross-Linked, RCL) ont ete utilises comme modeles de la chromatine, et simulations stochastiques de dynamique moleculaire ont ete performees avec l'objet de characteriser les facteurs qui affectent la probabilite de reparation d'un cluster de deux DSBs, a savoir : la distance genomique entre les DSBs, le degre de repliement du polymere (nombre de reticulations) et la dispersion induite par la machinerie de reparation (simulee par forces d'exclusion et par la suppression des reticulations autour des focus de dommage). Les resultats montrent que... TODO : Add results

**Abstract**

 Les cassures double-brin (DSBs de par ses sigles en anglais) sont un type de dommage d'ADN particulierement cytotoxique, pour lequel la jonction d'extrémités non homologues (Non Homologous End Joining ou NHEJ) est le mecanisme de reparation prefere en grande partie du cycle cellulaire. Neanmois, il peut induire d'importantes reorganisations chromosomiques quand les DSBs sont regroupes en clusters. Des polymeres grossieres a reticulations aleatoires (Random Cross-Linked, RCL) ont ete utilises comme modeles de la chromatine, et simulations stochastiques de dynamique moleculaire ont ete performees avec l'objet de characteriser les facteurs qui affectent la probabilite de reparation d'un cluster de deux DSBs, a savoir: la distance genomique entre les DSBs, le degre de repliement du polymere (nombre de reticulations) et la dispersion induite par la machinerie de reparation (simulee par forces d'exclusion et par la suppression des reticulations autour des focus de dommage). Les resultats montrent que... TODO: Add results

# Contents

# Chapter 1

# Motivation

Double Strand Breaks (DSB) are a highly cytotoxic type of DNA damage that occur when both strands of duplex DNA break, for instance, after ionizing radiation (IR) and cancer chemotherapy. Two principal pathways of DSB repair have evolved: non-homologous end joining (NHEJ) and homologous recombination (HR). NHEJ is the major pathway of DSB repair and consists on the synapsis and ligation of the broken ends. Here, we aim to simulate and study the repair and misrepair rates after two DSBs, via the NHEJ pathway. Polymer models (and in particular, a generalization of the Gaussian chain, the Randomly Cross-Linked (RCL) Polymer) will be used as a model of the chromatin. Stochastic simulations of molecular dynamics will be performed with the goal of characterize the factors that contribute to interfere or improve the repair probability, namely: the distance between the DSBs, the degree of folding of the polymer (simulated by the number of cross-links connecting non-neighbor monomers) and the dispersion effect induced by the DNA repair machinery (simulated by exclusion forces and the local removal of cross-links in the damage foci).

# Chapter 2

# Methods

## 2.1 The RCL polymer

To simulate the chromatin we consider the model of a randomly cross-linked (RCL) polymer. An RCL polymer is formed by a chain of $N$ monomers connected by harmonic springs of constant $\kappa$ as backbone, and $N_c$ extra cross-links (CLs) which connect non-neighbor monomers with springs of constant $\kappa'$. CLs are added uniformly over all $\frac{(N-1)(N-2)}{2}$ combinations of non-neighbor monomer pairs. We also define the RCL polymer connectivity $\xi$ as the fraction of added cross-links with respect to the total number of possible pairings, i.e. $\xi = \frac{2N_c}{(N-1)(N-2)}$.

We characterize the RCL polymer by two $N \times N$ admittance (Laplacian) matrices: the Rouse matrix $M$ which describes the backbone, and $B$ which describes the added connectors. A Laplacian matrix results from the difference between the diagonal degree matrix $D$ that accounts for the total number of connections each monomer has, and the adjacency matrix $A$ that accounts for the connectivity ($A_{m,n} = 1$ if monomers $m$ and $n$ are connected, 0 otherwise). So

$$M_{m,n} = (D_{\text{backbone}} - A_{\text{backbone}})_{m,n} = \begin{cases} -1 & \text{if } |m - n| = 1 \\ -\sum_{i=1, i \neq n}^{N} M_{i,n} & \text{if } m = n \\ 0 & \text{otherwise} \end{cases}$$

(2.1)

and, analogically

$$B_{m,n} = (D_{\text{CLs}} - A_{\text{CLs}})_{m,n} = \begin{cases} -1 & \text{if } m \text{ and } n \neq m \text{ are connected by a CL} \\ -\sum_{i=1, i \neq n}^{N} B_{i,n} & \text{if } m = n \\ 0 & \text{otherwise} \end{cases}$$

(2.2)

An schematic illustration of an RCL polymer is described in Fig. 2.1-A along with its respective Laplacian matrices (Fig.2.1-B).

## 2.2 Langevin dynamics of the RCL polymer

We supposed the RCL polymer subdued to Langevin dynamics, i.e., if monomer positions of the polymer in time $t$ are represented by the $N \times 3$ matrix $(R_t)_{t \geq 0} = (R_1, ..., R_N)_{t \geq 0}$ with each $R_i \in \mathbb{R}^3, i = 1, ..., N$, the dynamics are described by Eq. (2.3):

$$dR_t = -\frac{1}{\zeta} \nabla \phi(R_t) dt + \sqrt{2D} dW_t \tag{2.3}$$

where $\zeta$ is the friction coefficient, $\phi$ is an harmonic potential, $D$ is the diffusion coefficient and $(W_t)_{t \geq 0}$ is a $N \times 3$-dimensional Brownian motion.

The harmonic potential of a RCL polymer is given by the classical Rouse potential and by the potential induced by the random cross-links:

$$\phi(R) = \frac{\kappa}{2} \sum_{n=1}^{N-1} ||R_{n+1} - R_n||^2 + \frac{\kappa'}{2} \sum_{i,j \in \mathcal{R}} ||R_i - R_j||^2$$

$$= \frac{\kappa}{2} \text{tr}(R^T M R) + \frac{\kappa'}{2} \text{tr}(R^T B R) \tag{2.4}$$

where $\kappa$ and $\kappa'$ are spring constants we will suppose equal, $\mathcal{R}$ is the set of monomers which have been randomly connected, and $M$ and $B$ are the Laplacian matrices introduced in Eq. 2.1 and Eq. 2.2. Since $\frac{\kappa}{\gamma} = \frac{3D}{b^2}$ where $b$ is the standard deviation of the bonds length, we'll consider the matrix equation:

$$dR_t = -\frac{3D}{b^2}(M + B)R_t dt + \sqrt{2D} dW_t \tag{2.5}$$

In the following, we implement the Euler's scheme of Eq. 2.5, setting $D = 0.008 \ \mu^2 \text{m/s}$ and $b = 0.2 \ \mu\text{m}$. When it is important, the $\Delta t$ used in the discretization of Eq. 2.5 will be specified. $\Delta t = 0.005$ sec will be the usual value. Anyhow, to assure numerical stability, condition in Eq. 2.6 must always be verified [Ref].

$$\sqrt{2D\Delta t} \leq c\Delta r^\star \tag{2.6}$$

In Eq. 2.6, $\Delta r^\star$ stands for the smallest spatial scale used (it will normally be the monomers encounter distance as it is defined later on) and $c$ is a confidence factor that will be set at $c = 0.2$ in the simulations.

## 2.3 Useful statistical properties of the RCL polymer

Some useful statistical properties will be measured to study the polymer repair. For instance, the Mean Radius of Gyration (MRG) will be computed as a proper indicator of the polymer compaction. The MRG is defined as the root mean

square distance from each monomer to the center of mass of the polymer (Eq. 2.7).

$$\mathrm{MRG}(R) = \sqrt{\frac{1}{N}\sum_{m=1}^{N}(R_m - \bar{R})^2} \qquad (2.7)$$

It has been proven [?] that for a given connectivity fraction $\xi$ and when the connectivity is low ($N_c \ll \frac{1}{2}N^2$) the mean square radius of gyration MSRG (MRG$^2$) can be approximated by Eq. 2.8.

$$\mathrm{MSRG}(\xi) \approx \frac{3b^2}{4(1-\xi)\sqrt{N\xi}} \qquad (2.8)$$

## 2.4 Induction of Double Strand Breaks (DSBs)

In particular, we are interested on the probability of repair after two DSBs. One DSB (damage focus «A») will be randomly induced between neighbor monomers $A_1$ and $A_2 = A_1 + 1$, where $A_1 \sim \mathcal{U}[1, N - g - 2]$. The other DSB (damage focus «B») will be induced in neighbor monomers $B_1$ and $B_2$. A deterministic inter-break genomic distance $g$ will be imposed between A and B, defined as the shortest distance between the DSBs along the backbone, i.e. $g = B_1 - A_2$. An scheme of two DSBs is represented in Fig. 2.1-C. The two breaks define thus three interconnected fragments: the upstream fragment (formed by monomers 1 to $A_1$), the central fragment (monomers $A_2$ to $B_1$) and the downstream fragment (monomers $B_2$ to $N$).

## 2.5 Definition of encounter times and probabilities

We define the first encounter time between monomers $m$ and $n$ as

$$T_{m,n}^{\epsilon} := \inf\{t \geq 0 : ||(R_m)_t - (R_n)_t|| < \epsilon\} \qquad (2.9)$$

where $\epsilon$ is the encounter distance. Then, the repair probability is defined as

$$\mathbb{P}(\mathrm{Repair}) = \mathbb{P}(\inf\{T_{A_1,A_2}^{\epsilon}, T_{B_1,B_2}^{\epsilon}\} \leq \inf\{T_{A_1,B_1}^{\epsilon}, T_{A_1,B_2}^{\epsilon}, T_{A_2,B_1}^{\epsilon}, T_{A_2,B_2}^{\epsilon}\})$$

and the First Encounter Time (FET) as:

$$\mathrm{FET} = \inf\{T_{A_1,A_2}^{\epsilon}, T_{B_1,B_2}^{\epsilon}, T_{A_1,B_1}^{\epsilon}, T_{A_1,B_2}^{\epsilon}, T_{A_2,B_1}^{\epsilon}, T_{A_2,B_2}^{\epsilon}\}$$

We remark that when monomers are well mixed, the expected repair probability is $2/6 = 1/3$ (favorable encounters / total possible encounters).

## 2.6 Simulation of chromatin decondensation around DSBs: Removal of CLs and Excluded Volume Interactions

It has been reported that the DNA repair machinery induces a relaxation of the chromatin around damage foci (DF) [Ref]. To simulate this dispersion two effects are considered independently: the removal of CLs in the damage foci to induce decompactation of the polymer around the breaks; and the addition of an exclusion potential around the damaged monomers, to simulate self-avoidance towards the cut ends.

In general, excluded volume forces are added to the polymer dynamics through a new harmonic potential $\phi_{ev}$ that is added to the potential $\phi$ defined in Eq. 2.4:

$$\phi_{ev}(R) = -\frac{\kappa_{ev}}{2} \sum_{i,j \ : \ i \neq j} ||R_i - R_j||^2 1_{||R_i - R_j|| < \sigma} \tag{2.10}$$

where $\sigma$ is a cutoff radius of exclusion and $1_{||R_i - R_j|| < \sigma} = 1$ if $||R_i - R_j|| < \sigma$ and 0 otherwise.

In our case, to simulate the dispersion of chromatin around DSBs only, we'll induce first volume exclusion for the DF monomers only and take $\kappa_{ev} = \kappa$:

$$\phi_{local-ev}(R) = -\frac{\kappa_{ev}}{2} \sum_{i \in \text{DF}} \sum_{j \neq i} ||R_i - R_j||^2 1_{||R_i - R_j|| < \sigma} \tag{2.11}$$

Later on, considering that in spite of the exclusion induced by repair proteins, the whole repair machinery looks for affine ends to join them, we'll consider repair exclusion spheres that repel all monomers except the other cut ends. So, if any two cut ends approach they won't feel exclusion forces. This new exclusion potential $\phi_{repair \atop sphere}$ is defined in Eq. 2.12.

$$\phi_{repair \atop sphere}(R) = -\frac{\kappa_{ev}}{2} \sum_{i \in \text{DF}} \sum_{j \notin \text{DF}} ||R_i - R_j||^2 1_{||R_i - R_j|| < \sigma} \tag{2.12}$$

Thereby, volume exclusion potentials originate a new Laplacian $E$ as defined in Eq. 2.13, that will be added to the Laplacian matrices considered up to now (Eq. 2.5):

$$E_{m,n} = \begin{cases} -1 & \text{if } m \in \text{DF} \text{ , } n \notin \text{DF and } ||R_m - R_n|| < \sigma \\ -1 & \text{if } n \in \text{DF} \text{ , } m \notin \text{DF and } ||R_m - R_n|| < \sigma \\ -\sum_{i=1, \ i \neq n}^{N} E_{m,i} & \text{if } m = n \\ 0 & \text{otherwise} \end{cases}$$

$$\tag{2.13}$$

So when exclusion forces are considered the simulated system will subdue the dynamic described by Eq. 2.14.
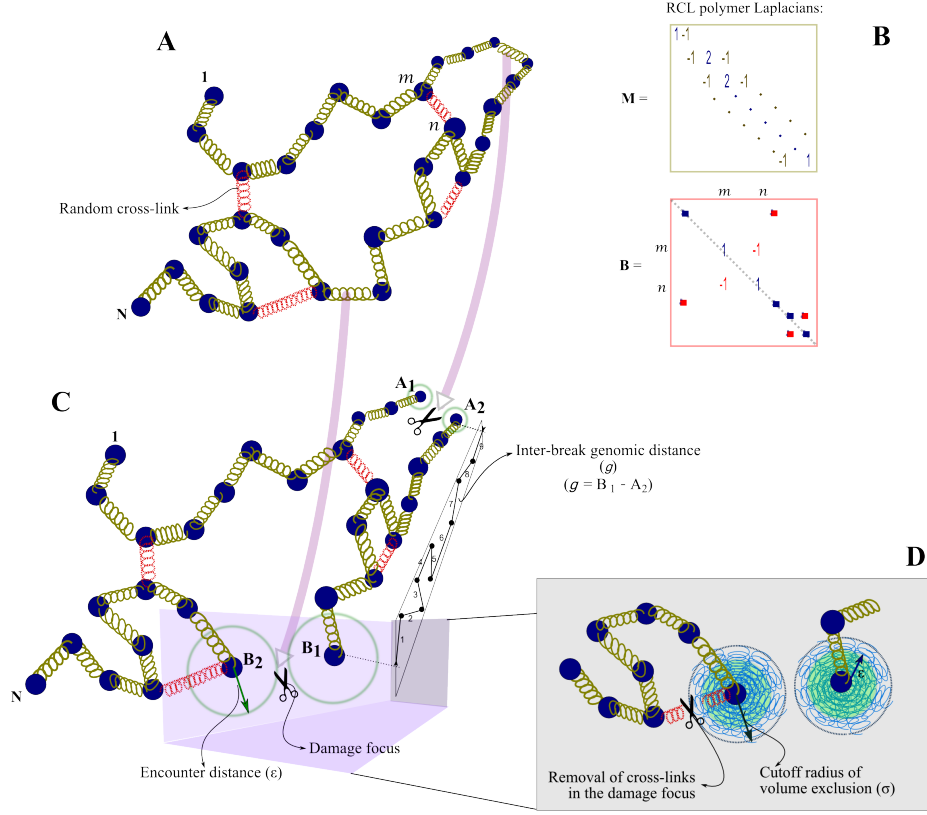
Figure 2.1: Model of an RCL polymer for chromatin repair simulations. **A**: Example of the Rouse backbone with some cross-links. **B**: Laplacian matrices of the RCL polymer represented in **A**: the Rouse matrix ($M$) and the CLs matrix ($B$). **C**: Scheme of the induction of two DSBs $A$ and $B$ at a distance $g$. **D**: To simulate the effect of the sparsity induced by the repair machinery, excluded volume forces are added in the damage foci and the concerned CLs are removed.

$$dR_t = -\frac{3D}{b^2}(M + B + E)R_t dt + \sqrt{2D}dW_t \qquad (2.14)$$

We call the total $M + B + E$ the total Laplacian matrix of the system. The two decondensation effects are summarized in Fig. 2.1-D.

## 2.7 Scaling distances under condensation: $\xi$-adaptive encounter distances and exclusion radii

As indicates Eq. 2.8 (cf. Fig. **??** for an empirical example), as $N_c$ (equivalently $\xi$) increases the polymer compaction increases too. Since the spatial scale of a compact polymer is different from the spatial scale used by a decompacted polymer, and thus for the same $\varepsilon$, an encounter which is rare enough for the later might be too common for the former, for a more compact polymer, smaller spatial scales should be used. In effect, scaling $\varepsilon$ and $\sigma$ for different degrees of compaction will allow to measure the repair rates conditionally to the compaction, invariantly to the polymer size. So $\varepsilon$ and $\sigma$ will be redefined to be $\xi$-adaptive and preserve the spatial scale. Thus, $\varepsilon$ and $\sigma$ will be proportional to the polymer mean size, and in particular, to the MRG. Starting from Eq. 2.8 we will define the adaptive encounter distance $\tilde{\varepsilon}(\xi)$ as in Eq. 2.15, and the adaptive cutoff radius of exclusion $\tilde{\sigma}(\xi)$ as in Eq. 2.16, where $v > 1$ is a factor which indicates how much bigger is the exclusion sphere with respect to the encounter distance.

$$\tilde{\varepsilon}(\xi) = 2\sqrt{\frac{3b^2}{N(1-\xi)\sqrt{y^2-1}}} \qquad (2.15)$$

$$y = 1 + \frac{N\xi}{2(1-\xi)}$$

$$\tilde{\sigma}(\xi) = v \cdot \tilde{\varepsilon}(\xi) \qquad (2.16)$$

## 2.8 Pipeline of the simulations

A number $I$ of independent polymer realizations will be simulated to extract mean statistical properties and the repair probabilities conditional to the parameters defined so far. The main pipeline of each realization goes as described below:

1. **Initialization of the polymer backbone** from a random walk in $\mathbb{R}^3$ where each step is i.i.d. $\mathcal{N}(0, b^2)$.

2. **Predefinition of DSBs $A$ and $B$.** We take $A_1 \sim \mathcal{U}[1, N-g-2]$ and then $A_2 = A_1 + 1$, $B_1 = A_2 + g$, $B_2 = B_2 + 1$.

3. **Induction of random cross-links**. $N_c$ cross-links are induced between non-neighbor monomers such that after breaks $A$ and $B$ the polymer rests fully connected. In practice, we make random connections until obtaining one which verifies the full connectivity condition.

4. **Relaxation**. Once the polymer is validly cross-linked, the polymer is subdued to Langevin dynamics (discretized as an Euler's scheme for Eq. 2.5) until relaxation time $\tau$, which is computed analytically [Ref] as:

$$\tau = \frac{...}{...}$$

5. **DSBs Induction and actualization of the total Laplacian matrix**. When the polymer reaches relaxation time, the two cuts are induced between the predefined monomers $A_1$, $A_2$ and $B_1$, $B_2$. Removed bonds are thus cleaned out from the Laplacian matrix. Besides, if CLs are removed from the separated monomers and if Volume Exclusion is included, those effects are also added to the Laplacian as in Eq. 2.14.

6. **Waiting**. As a way to exclude the repair events which happen immediately after the break, and since the DNA repair is not instantaneous (it requires the arrival of specific proteins to the cut ends), the polymer is let to evolve under Eq. 2.5 for some waiting time $T_{wait}$

7. **Simulation until encounter**. Once $T_{wait}$ is reached, simulation continues until an encounter, as defined in Eq. 2.9, happens between any of the cut ends. If two or more pairs of monomers encounter at the same simulated instant, we randomly choose over those pairs to be the first encounter. If no encounters occur before a threshold time $T_{max}$ the simulation is discarded.

8. Statistical properties and the encounter event that occurred (repair or misrepair) are saved and a new polymer is initialized to perform the same simulation.

Finally, the conditional probability of repair is estimated as

$$\mathbb{P}(\text{Repair}|\Theta) = \frac{\text{Number of } A_1\text{-}A_2 \text{ or } B_1\text{-}B_2 \text{ encounters}}{\text{Total number of encounters}}$$

where $\Theta$ summarizes the experiment parameters,

$$\Theta = (N, b, N_c, g, \varepsilon, \sigma, T_{wait}, T_{max}).$$