École polytechnique

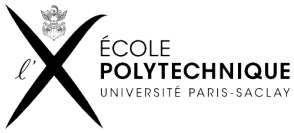
Rapport de stage de recherche

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

Rapport non confidentiel

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



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, July 2, 2018
Date

Ignacio Madrid Canales
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 lorsque les DSBs sont regroupes en clusters. Des polymeres grossieres a reticulations aleatoires (Random Cross-Linked, RCL) ont ete utilises comme modeles pour 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. 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 lorsque 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

Introduction

Chapter 2

Biological background

2.1 Chromatin: the complex, compact and mobile wrapping of DNA

In cells DNA is never alone. Since Walther Flemming's observations in 1878 it is known that DNA organizes in a more complex, heterogeneous structure: the chromatin. This macromolecular complex formed by DNA, RNA and proteins [] folds in a reticular multiscale organization (Fig.) along which DNA is wrapped ranging from the 2 nm dyameter duplex DNA strands to the micrometric chromosome.

However, it was not until 1997 when it was finally proved that chromatin is indeed not static and that undergoes constrained diffusion (subdiffusion) []. The biological function of chromatin mobility, and particularly its implications and changes during DNA damage response (DDR) are still poorly understood and have been discussed ever since [].

2.2 Double-strand breaks (DSB): induction and repair pathways

During cell cycle the genetic material is susceptible to suffer from various types of damage. Double Strand Breaks (DSB) are one of the most cytotoxic ones [], occurring when both strands of duplex DNA break. For instance, DSB happen after ionizing radiation (IR) exposure and cancer chemotherapy []. Spontaneously, mammalian cells can experiment around 50 DSBs per cell cycle [].

DSB repair is a crucial task for cells, and two main pathways of DSB repair have evolved: non-homologous end joining (NHEJ) and homologous recombination (HR) (Fig.). NHEJ is the major pathway of DSB repair and consists simply on the religation of the broken ends. NHEJ is fast but error-prone, inducing random insertions or deletions (indels) in the repaired region. On the other hand, HR is usually error-free, but needs an homologous donor sequence to use as template (so it requires to wait until DNA replication) [shielding]. Moreover, the risks of failed repair via NHEJ increase as DSB cluster in a genomic neighborhood. Misrepaired two-DSB clusters may result in a variety of serious chromosomic aberrations (Fig.).

Furthermore, both HR and NHEJ pathways requires a complex and yet not fully understood assembly of proteins and biochemical changes around the damage foci (DF). In particular, during NHEJ the repair machinery has been proven to favor decondensation of the chromatin around DF

[] (Fig.). At the same time, the combined action of 53BP1 protein, the LINC nuclear envelope complex and microtubules in the periphery of the nucleus enhance the chromatin mobility around DSB DF [Lottersberg15]. However, the relationship between enhanced diffusion and repair is not evident, and even counterintuitive: why and how repair could be favored by the broken DNA ends mobility? Contrasting results arguing for the detrimental or enhancing effect of DSB mobility on faithful repair have been showed, as some mechanisms have been proposed to fit them [ibid]. Here we rejoin the debate, aiming to characterize the factors that contribute to alter the chromatin diffusion, interfering or improving the repair probability of a two-DSB cluster, namely: the distance between the DSBs, the degree of folding and the dispersion effect induced by the DNA repair machinery.

Chapter 3

Methods

3.1 The RCL polymer

To simulate the chromatin we consider the model of a randomly cross-linked (RCL) polymer [ofir]. An RCL polymer is formed by a chain of N monomers connected by harmonic springs of constant κ as backbone (also known as Rouse polymer, or Gaussian chain [doi]), and N_c extra cross-links (CLs) which connect non-neighbor monomers with springs of constant κ'. 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 ξ 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

$$M_{m,n} = (D_{\text{backbone}} - A_{\text{backbone}})_{m,n} = \begin{cases} -1 & \text{if } |m-n| = 1 \\ -\sum_{i=1, i\neq n}^{N} M_{i,n} & \text{if } m = n \\ 0 & \text{otherwise} \end{cases} \tag{3.1}$$

and, analogically

$$B_{m,n} = (D_{\text{CLs}} - A_{\text{CLs}})_{m,n} = \begin{cases} -1 & \text{if } m \text{ and } n \neq m \text{ are connected by a CL} \\ -\sum_{i=1, i\neq n}^{N} B_{i,n} & \text{if } m = n \\ 0 & \text{otherwise} \end{cases} \tag{3.2}$$

An schematic illustration of an RCL polymer is described in Fig. 3.1-A along with its respective Laplacian matrices (Fig.3.1-B).

3.2 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. (3.3):

$$dR_t = -\frac{1}{\zeta}\nabla\phi(R_t)dt + \sqrt{2D}dW_t \tag{3.3}$$

where ζ is the friction coefficient, ϕ 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:

$$\begin{aligned}\phi(R) &= \frac{\kappa}{2}\sum_{n=1}^{N-1}||R_{n+1} - R_n||^2 + \frac{\kappa'}{2}\sum_{i,j\in\mathcal{R}}||R_i - R_j||^2 \\ &= \frac{\kappa}{2}\mathrm{tr}(R^T M R) + \frac{\kappa'}{2}\mathrm{tr}(R^T B R)\end{aligned} \tag{3.4}$$

where κ and κ' 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. 3.1 and Eq. 3.2. Since $\frac{\kappa}{\gamma} = \frac{3D}{b^2}$ where b is the standard deviation of the bonds length, we'll consider the matrix equation:

$$dR_t = -\frac{3D}{b^2}(M + B)R_t dt + \sqrt{2D}dW_t \tag{3.5}$$

In the following, we implement the Euler's scheme of Eq. 3.5, setting $D = 0.008\ \mu^2$m/s and $b = 0.2\ \mu$m. When it is important, the Δt used in the discretization of Eq. 3.5 will be specified. $\Delta t = 0.005$ sec will be the usual value. Anyhow, to assure numerical stability, condition in Eq. 3.6 must always be verified [Ref].

$$\sqrt{2D\Delta t} \leq c\Delta r^\star \tag{3.6}$$

In Eq. 3.6, $\Delta r^\star$ 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.

3.3 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. 3.7).

$$\mathrm{MRG}(R) = \sqrt{\frac{1}{N}\sum_{m=1}^{N}(R_m - \bar{R})^2} \tag{3.7}$$

It has been proven [**?**] that for a given connectivity fraction ξ 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. 3.8.

$$\text{MSRG}(\xi) \approx \frac{3b^2}{4(1-\xi)\sqrt{N\xi}} \tag{3.8}$$

3.4 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. 3.1-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).

3.5 Definition of encounter times and probabilities

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

$$T^{\epsilon}_{m,n} := \inf\{t \geq 0 : ||(R_m)_t - (R_n)_t|| < \epsilon\} \tag{3.9}$$

where ϵ is the encounter distance. Then, the repair probability is defined as

$$\mathbb{P}(\text{Repair}) = \mathbb{P}(\inf\{T^{\epsilon}_{A_1,A_2}, T^{\epsilon}_{B_1,B_2}\} \leq \inf\{T^{\epsilon}_{A_1,B_1}, T^{\epsilon}_{A_1,B_2}, T^{\epsilon}_{A_2,B_1}, T^{\epsilon}_{A_2,B_2}\})$$

and the First Encounter Time (FET) as:

$$\text{FET} = \inf\{T^{\epsilon}_{A_1,A_2}, T^{\epsilon}_{B_1,B_2}, T^{\epsilon}_{A_1,B_1}, T^{\epsilon}_{A_1,B_2}, T^{\epsilon}_{A_2,B_1}, T^{\epsilon}_{A_2,B_2}\}$$

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

3.6 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 ϕ_{ev} that is added to the potential ϕ defined in Eq. 3.4:

$$\phi_{ev}(R) = -\frac{\kappa_{ev}}{2} \sum_{i,j\ :\ i\neq j} ||R_i - R_j||^2 1_{||R_i - R_j|| < \sigma} \tag{3.10}$$

where σ 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 $\kappa_{ev} = \kappa$:

$$\phi_{local-ev}(R) = -\frac{\kappa_{ev}}{2} \sum_{i\in \mathrm{DF}} \sum_{j\neq i} ||R_i - R_j||^2 1_{||R_i - R_j|| < \sigma} \tag{3.11}$$

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 $\underset{sphere}{\phi_{repair}}$ is defined in Eq. 3.12.

$$\underset{sphere}{\phi_{repair}}(R) = -\frac{\kappa_{ev}}{2} \sum_{i\in \mathrm{DF}} \sum_{j\notin \mathrm{DF}} ||R_i - R_j||^2 1_{||R_i - R_j|| < \sigma} \tag{3.12}$$

Thereby, volume exclusion potentials originate a new Laplacian E as defined in Eq. 3.13, that will be added to the Laplacian matrices considered up to now (Eq. 3.5):

$$E_{m,n} = \begin{cases} -1 & \text{if } m \in \mathrm{DF}\ ,\ n \notin \mathrm{DF} \text{ and } ||R_m - R_n|| < \sigma \\ -1 & \text{if } n \in \mathrm{DF}\ ,\ m \notin \mathrm{DF} \text{ and } ||R_m - R_n|| < \sigma \\ -\sum_{i=1,\ i\neq n}^{N} E_{m,i} & \text{if } m = n \\ 0 & \text{otherwise} \end{cases} \tag{3.13}$$

So when exclusion forces are considered the simulated system will subdue the dynamic described by Eq. 3.14.

$$dR_t = -\frac{3D}{b^2}(M + B + E)R_t dt + \sqrt{2D}dW_t \tag{3.14}$$

We call the total $M + B + E$ the total Laplacian matrix of the system. The two decondensation effects are summarized in Fig. 3.1-D.

3.7 Scaling distances under condensation: ξ-adaptive encounter distances and exclusion radii

As indicates Eq. 3.8 (cf. Fig. **??** for an empirical example), as N_c (equivalently ξ) 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 ε, 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 ε and σ for different degrees of compaction will allow to measure the repair rates conditionally to the compaction, invariantly to the polymer size. So ε and σ will be redefined to be ξ-adaptive and preserve the spatial scale. Thus, ε and σ will be proportional to the polymer mean size, and in particular, to the MRG. Starting from Eq. 3.8 we will define the adaptive encounter distance $\tilde{\varepsilon}(\xi)$ as in Eq. 3.15, and the adaptive cutoff radius

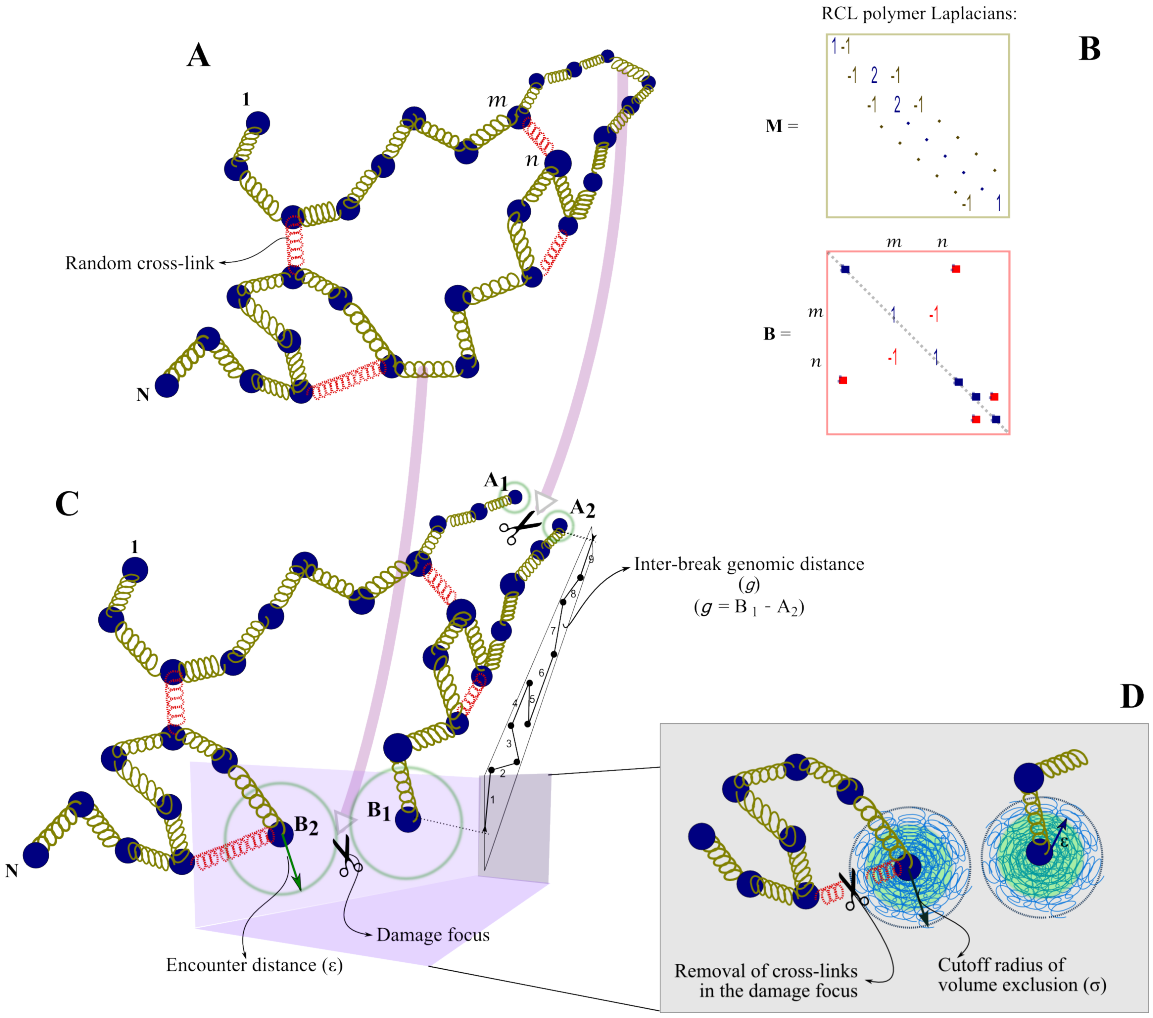


Figure 3.1: Model of an RCL polymer for chromatin repair simulations. **A**: Example of the Rouse backbone with some cross-links. **B**: Laplacian matrices of the RCL polymer represented in **A**: the Rouse matrix (M) and the CLs matrix (B). **C**: Scheme of the induction of two DSBs A and B at a distance 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.

of exclusion $\tilde{\sigma}(\xi)$ as in Eq. 3.16, where $v > 1$ is a factor which indicates how much bigger is the exclusion sphere with respect to the encounter distance.

$$\begin{aligned} \tilde{\varepsilon}(\xi) &= 2\sqrt{\frac{3b^2}{N(1-\xi)\sqrt{y^2-1}}} \\ y &= 1+\frac{N\xi}{2(1-\xi)} \end{aligned} \tag{3.15}$$

$$\tilde{\sigma}(\xi) = v \cdot \tilde{\varepsilon}(\xi) \tag{3.16}$$

3.8 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. 3.5) until relaxation time τ, which is computed analytically [Ref] as:
$$\tau = \frac{\cdots}{\cdots} \tag{3.17}$$

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. 3.14.

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. 3.5 for some waiting time T_{wait}

7. **Simulation until encounter**. Once T_{wait} is reached, simulation continues until an encounter, as defined in Eq. 3.9, 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 Θ summarizes the experiment parameters,

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

Chapter 4

Results

The repair dynamics of the simulated chromatin fiber are studied from two distinctive points of view. First, we ask how a DSB moves, and particularly how the presence of a DSB cluster changes the normal subdiffusive regime that characterizes unbroken Gaussian chains. To that extent, time-dependent statistical properties accounting for the polymer anomalous diffusion are measured after DSBs induction and after a fixed relaxation time (12000 steps), along a given fixed simulation time. We focus mainly in the Mean Square Displacement (MSD), defined in Eq. 3.7. Finally, we ask how DSBs are repaired. Which factors affect the misrepair rate? How are those changes explained by the mechanical alterations introduced before?

4.1 DSBs alter single-monomers anomalous diffusion around damage foci

MSD curves are obtained from monomer trajectories saved over different realizations. For intermediate time scales it is known that MSD fits a power function (Eq. 4.1) where A is constant and α is known as the anomalous diffusion exponent.

$$\mathrm{MSD}_m(t) := \left\langle ||R_m(t) - R_m(0)||^2 \right\rangle \sim At^{\alpha} \tag{4.1}$$

α characterizes the motion: $\alpha = 1$ indicates normal diffusion (Brownian motion); $\alpha > 1$, super-diffusion (directed motion): and, $\alpha < 1$ characterizes subdiffusion (confined motion). Chromatin motion is usually described as subdifussive [refs]. On its behalf, classical Rouse polymers exhibit $\alpha = 0.5$.

Here, MSD curves are obtained for each monomer of the chain and therefore an anomalous exponent α is fitted to each one of them. Fig. 4.2.

4.2 DF anomalous diffusion is altered in a N_c-dependent way

Results above also shows a clear trend in terms of N_c (see color distribution in Fig. - A).

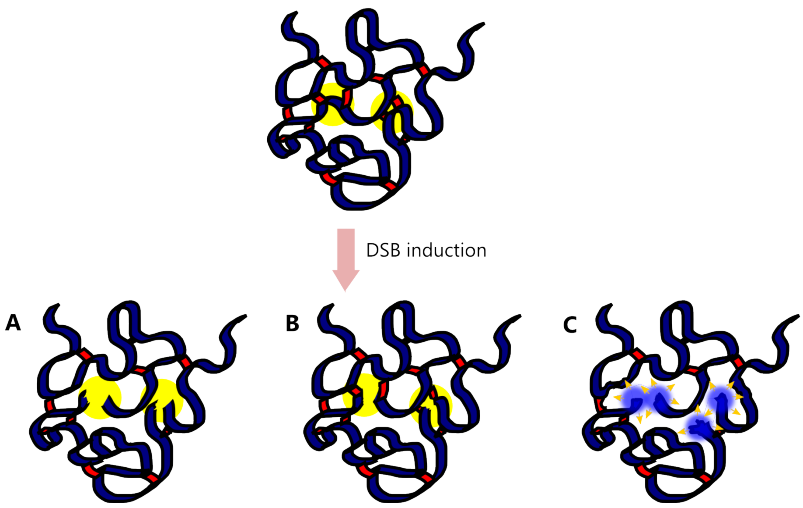


Figure 4.1: Hypothesis of possible DDR effects on the damage foci after DSB induction. **A.** Not only the backbone adjacency is lost after DSB; the CLs in the DF are also removed. **B.** Only the backbone adjacency is lost, the CLs in the DF are kept. **C.** Exclusion repulsive forces are induced around the DF. This scenario might occur keeping or removing CLs in DF.

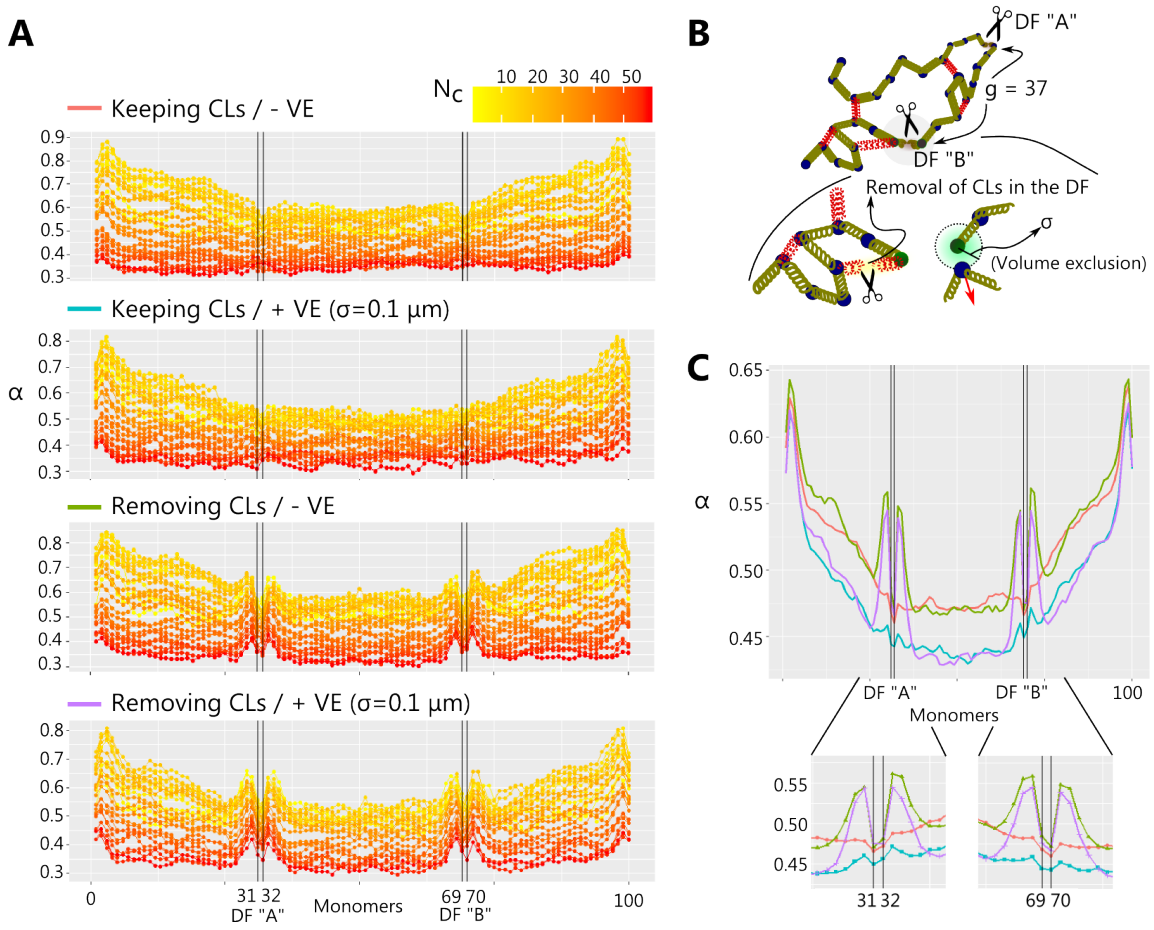


Figure 4.2: Anomalous diffusion exponent α measured under the four considered scenarios after the induction of two DSBs at fixed loci $A_1 = 31$ and $B_1 = 69$. **A.** α of each monomer from a 100-monomers RCL polymer obtained by fitting At^{α} to the MSD curves measured along 60 seconds ($\Delta t = 0.005$ sec) after have waited a relaxation time calculated by Eq. 3.17. The MSD is computed averaging over 600 simulations for different values of N_c ranging from 2 to 58, and for each of the four possible repair scenarios. **B.** Schematic representation of the simulated system illustrating the considered repair mechanisms. **C.** Averaged α of each monomer, where the average is taken over all the simulated connectivities (N_c) (i.e. averages of curves represented in the left panel). The colors are consistent with the labels titling the graphs at the left (**A**). Two windows zooming around the DF monomers appear at the bottom.