

Progress report:
Repair probabilities for the RCL polymer with
two DSBs

Ignacio Madrid

May 11, 2018

Contents

1 Motivation	2
2 Methods	2
2.1 The RCL polymer	2
2.2 Langevin dynamics of the RCL polymer	3
2.3 Useful statistical properties of the RCL polymer	4
2.4 Induction of Double Strand Breaks (DSBs)	4
2.5 Definition of encounter times and probabilities	4
2.6 Simulation of chromatin decondensation around DSBs: Removal of CLs and Excluded Volume Interactions	5
2.7 Scaling distances under condensation: ξ -adaptive encounter dis- tances and exclusion radii	6
2.8 Pipeline of the simulations	7
3 Results	9
3.1 First Encounter Time (FET) distribution	9
3.2 Changes on the repair probability conditionally to g , N_c and de- condensation	11
3.2.1 Effect of g on the repair probability	11
3.2.2 Effect of N_c on the repair probability	11
3.2.3 Effect of volume exclusion on the repair probability	12
3.2.4 Effect of volume exclusion with more restrictive encounter distances	18
3.3 Repair probability in different domains	20
3.3.1 TODO	26
4 Partial conclusions	27
5 Perspectives	27

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 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).

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 κ as backbone, 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

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

and, analogically

$$B_{m,n} = (D_{\text{CLS}} - A_{\text{CLS}})_{m,n} = \begin{cases} -1 & \text{if } m \text{ and } n \neq m \text{ are connected by a CL} \\ -\sum_{i=1}^N B_{i,n} & \text{if } m = n \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

An schematic illustration of an RCL polymer is described in Fig. 1-A along with its respective Laplacian matrices (Fig.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 $(R_t)_{t \geq 0} = (R_1, \dots, R_N)_{t \geq 0}$ with each $R_i \in \mathbb{R}^3, i = 1, \dots, N$, the dynamics are described by Eq. (3):

$$dR_t = -\frac{1}{\zeta} \nabla \phi(R_t) dt + \sqrt{2D} dW_t \quad (3)$$

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:

$$\begin{aligned} \phi(R) &= \frac{\kappa}{2} \sum_{n=1}^{N-1} \|R_{n+1} - R_n\|^2 + \frac{\kappa'}{2} \sum_{i,j \in \mathcal{R}} \|R_i - R_j\|^2 \\ &= \frac{\kappa}{2} \text{tr}(R^T M R) + \frac{\kappa'}{2} \text{tr}(R^T B R) \end{aligned} \quad (4)$$

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. 1 and Eq. 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:

$$dR_t = -\frac{3D}{b^2} (M + B) R_t dt + \sqrt{2D} dW_t \quad (5)$$

In the following, we implement the Euler's scheme of Eq. 5, setting $D = 0.008 \mu^2 \text{m/s}$ and $b = 0.2 \mu\text{m}$. When it is important, the Δt used in the discretization of Eq. 5 will be specified. $\Delta t = 0.005 \text{ sec}$ will be the usual value. Anyhow, to assure numerical stability, condition in Eq. 6 must always be verified [Ref].

$$\sqrt{2D\Delta t} \leq c\Delta r^* \quad (6)$$

In Eq. 6, Δr^* 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. 7).

$$\text{MRG}(R) = \sqrt{\frac{1}{N} \sum_{m=1}^N (R_m - \bar{R})^2} \quad (7)$$

It has been proven [2] 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 (MRG²) can be approximated by Eq. 8.

$$\text{MSRG}(\xi) \approx \frac{3b^2}{4(1-\xi)\sqrt{N\xi}} \quad (8)$$

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. $g = B_1 - A_2$. A scheme of two DSBs is represented in Fig. 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

$$T_{m,n}^\epsilon := \inf\{t \geq 0 : \|(R_m)_t - (R_n)_t\| < \epsilon\} \quad (9)$$

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

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

and the First Encounter Time (FET) as:

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

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 ϕ_{ev} that is added to the potential ϕ defined in Eq. 4:

$$\phi_{ev}(R) = -\frac{\kappa_{ev}}{2} \sum_{i,j : i \neq j} ||R_i - R_j||^2 1_{||R_i - R_j|| < \sigma} \quad (10)$$

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 induce volume exclusion for the DF monomers only and take $\kappa_{ev} = \kappa$:

$$\phi_{local-ev}(R) = -\frac{\kappa_{ev}}{2} \sum_{i \in DF} \sum_{j \neq i} ||R_i - R_j||^2 1_{||R_i - R_j|| < \sigma} \quad (11)$$

Finally, 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 is defined in Eq. 12.

$$\phi_{repair}^{sphere}(R) = -\frac{\kappa_{ev}}{2} \sum_{i \in DF} \sum_{j \notin DF} ||R_i - R_j||^2 1_{||R_i - R_j|| < \sigma} \quad (12)$$

Thereby, volume exclusion potentials originate a new Laplacian E as defined in Eq. 13, that will be added to the Laplacian matrices considered up to now (Eq. 5):

$$E_{m,n} = \begin{cases} -1 & \text{if } m \in DF, n \notin DF \text{ and } ||R_m - R_n|| < \sigma \\ -1 & \text{if } n \in DF, m \notin DF \text{ and } ||R_m - R_n|| < \sigma \\ -\sum_{i=1, i \neq n}^N E_{m,i} & \text{if } m = n \\ 0 & \text{otherwise} \end{cases} \quad (13)$$

So when exclusion forces are considered the simulated system will subdue the dynamic described by Eq. 14.

$$dR_t = -\frac{3D}{b^2} (M + B + E) R_t dt + \sqrt{2D} dW_t \quad (14)$$

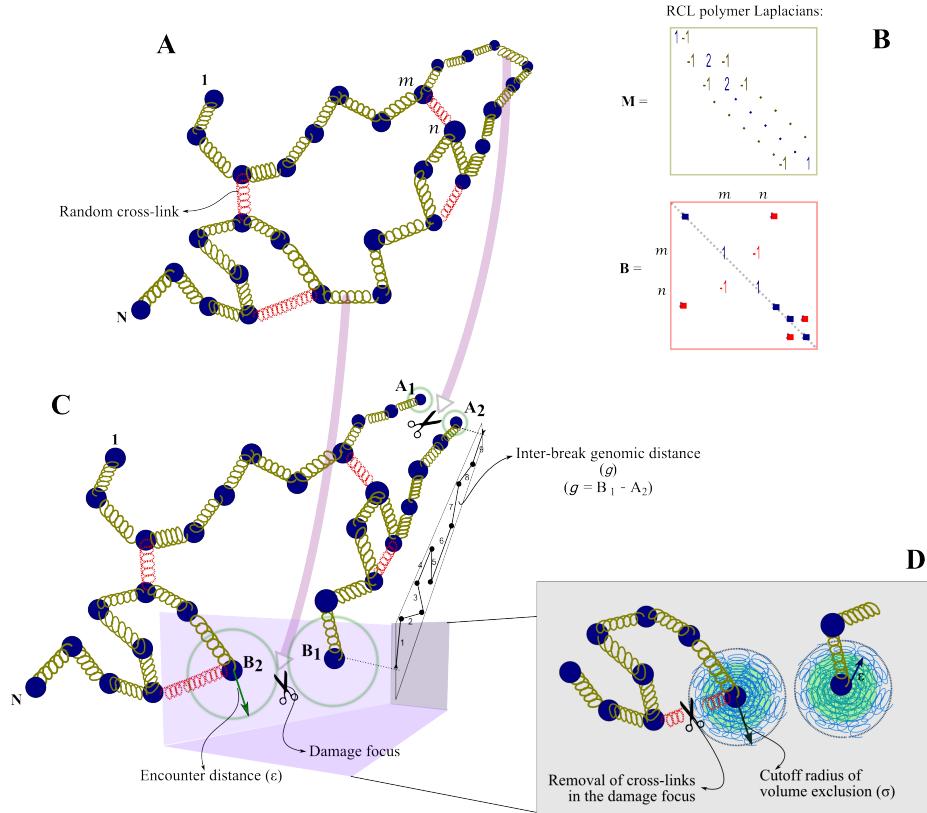


Figure 1: Model of an RCL polymer for chromatin repair simulations. **A:** Example of the Rouse backbone with some cross-links. **B:** Laplacian matrices of the RCL polymer represented in **A**: the Rouse matrix (M) and the CLs matrix (B). **C:** Scheme of the induction of two DSBs A and B at a distance 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.

We call the total $M + B + E$ the total Laplacian matrix of the system. The two decondensation effects are summarized in Fig. 1-D.

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

As indicated by Eq. 8 (cf. Fig. 6 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 decompact polymer, and thus for the same ε , an encounter which is rare enough for the latter 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. 8 we will define the adaptive encounter distance $\tilde{\varepsilon}(\xi)$ as in Eq. 15, and the adaptive cutoff radius of exclusion $\tilde{\sigma}(\xi)$ as in Eq. 16, where $v > 1$ is a factor which indicates how much bigger is the exclusion sphere with respect to the exclusion volume.

$$\begin{aligned}\tilde{\varepsilon}(\xi) &= 2\sqrt{\frac{3b^2}{N(1-\xi)\sqrt{y^2-1}}} \\ y &= 1 + \frac{N\xi}{2(1-\xi)}\end{aligned}\tag{15}$$

$$\tilde{\sigma}(\xi) = v \cdot \tilde{\varepsilon}(\xi)\tag{16}$$

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 $A_2 = A_1 + 1$, $B_1 = A_2 + g$, $B_2 = B_1 + 1$.
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. 5) until relaxation time τ , which is computed analytically [Ref] as:

$$\tau = \frac{\dots}{\dots}$$

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 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. 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. 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. 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,

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

3 Results

3.1 First Encounter Time (FET) distribution

Since the FET distribution can be approximated by a single exponential [1], an exponential function with parameters a, λ such that $t \mapsto a \exp(-\lambda t)$ is fitted to the frequency histograms, aiming to compare the λ fitted in each one of the experiments sets.

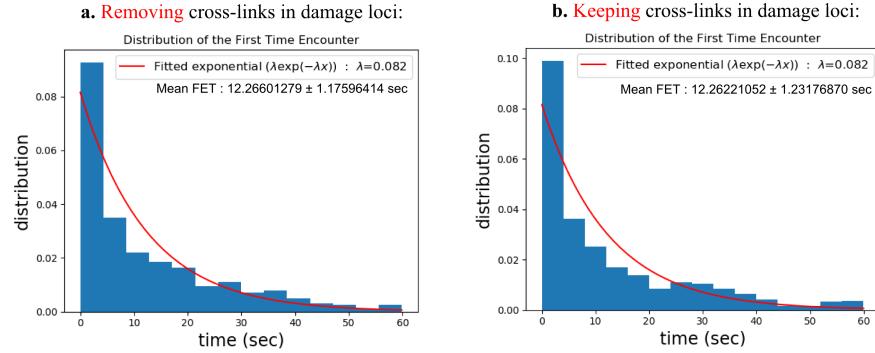
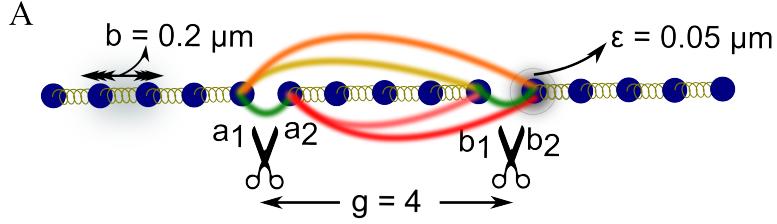


Figure 2: Distribution of the first encounter time keeping and removing the cross-links in the cleaved monomers. An exponential function has been fitted to each histogram. $N = 100$ monomers, $N_c = 5$, $g = 10$, $T_{wait} = 25\text{sec}$, and $T_{max} = 60\text{sec}$.

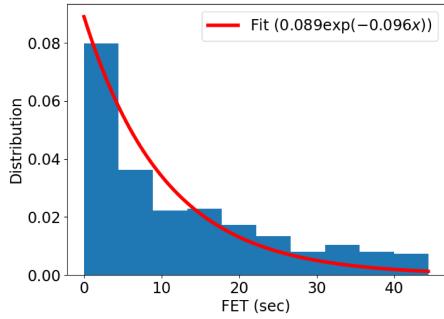
The results in Fig. 2 indicate that the removal of cross-links does not seem to affect the mean first encounter time, the estimated rate parameter (λ) is 0.08 for both cases. Nonetheless, it's no surprising considering that for $N = 100$ and $N_c = 5$ the mean number of connectors in the DF is about 0.4, so the removal of CLs is not actually occurring for the majority of cases. To overcome this situation and to properly study the effect of removing the CLs a new Laplacian matrix will be included, which will impose a certain number N_d of extra cross-links in the DF monomers. Fig. 3 shows the difference in the FET distributions when we impose $N_d = 2$ CLs in the DF. We can see that when we remove the 2 imposed CLs the mean FET increases in 2 seconds, and the repair rate increases as well. We also observe that when we remove the CLs the most common misrepair corresponds to the looping of the central segment of the broken polymer (A_2-B_1).

In the following, we will determine the effect of the removal of CLs with respect to the inter-break distance g and the number of random CLs N_c .



Keeping CLs in DF

B.1

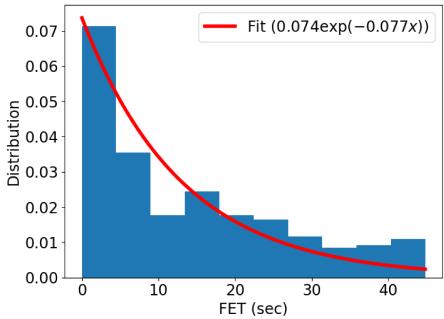


B.2



Removing CLs in DF

C.1



C.2

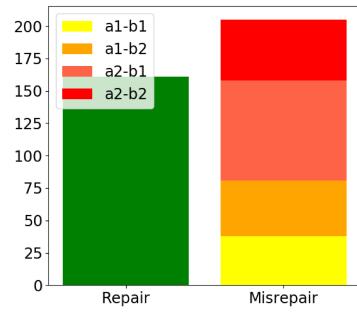


Figure 3: Distribution of the first encounter times and events, keeping and removing the cross-links when imposing CLs in the DF. Both results corresponds to $I = 500$ simulations, with $N = 100$, $N_c = 12$, $N_d = 2$, $g = 10$, $\varepsilon = 0.05\mu\text{m}$, $T_{max} = 45$ sec, $T_{wait} = 2.5$ sec and $\sigma = 0$ (no volume exclusion). **A:** Schematic representation of the experimental set and possible encounters. **B:** Results keeping CLs in DF. **1.** Histogram of relative frequencies and fitted exponential function, with an estimated rate of 0.096 and a mean FET of 12.600 ± 1.032 sec. **2.** Encounter events distribution. **C:** Results removing CLs in DF. **1.** Histogram of relative frequencies and fitted exponential function, with an estimated rate of 0.077 and a mean FET of 14.171 ± 1.117 sec. **2.** Encounter events distribution.

3.2 Changes on the repair probability conditionally to g , N_c and decondensation

3.2.1 Effect of g on the repair probability

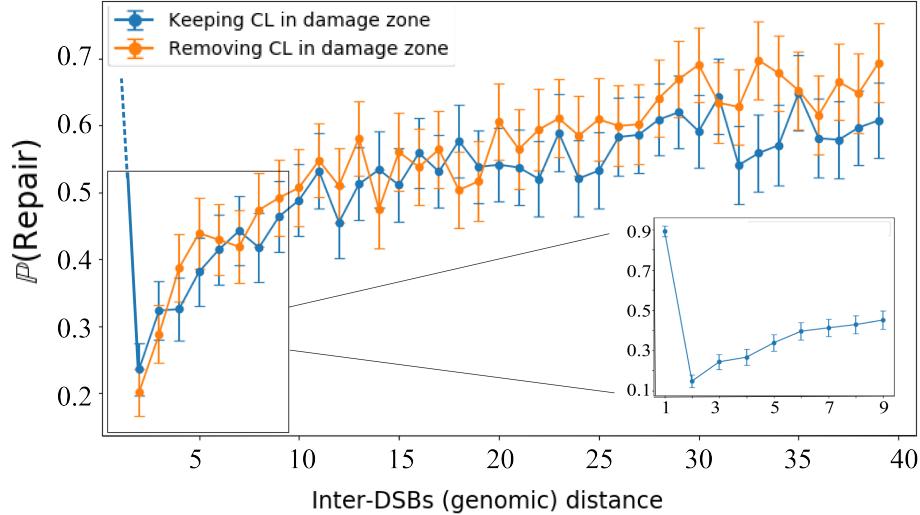


Figure 4: Repair probability as function of the genomic distance g between the two DSBs. $N = 100$, $N_c = 15$, $\varepsilon = 0.1\mu\text{m}$ and $N_d = 0$ (no CLs imposed in DF).

Fig. 4 shows that regardless the removal of cross-links, repair probability increases as the inter-break genomic distance increases.

Besides, when we keep CLs in the DF, we can remark an important gap between the probabilities at $g = 1$ (i.e., the two DSBs are connected, $A2$ and $B1$ being neighbors) and the other genomic distances (Fig. 4 Zoom). As only valid cuts are allowed (the polymer rests fully connected after the breaks) the central segment is indeed connected to the upstream and/or the downstream fragment. The probability of being connected to only one of them is higher than the probability of being connected to both, so the segment could be pushed towards one of the free monomers rapidly and then induce a good ratio of repair. This case will be treated as an out-layer from now on, and thus the following comparative analysis will be performed for $g \geq 2$.

3.2.2 Effect of N_c on the repair probability

Increasing the number of cross-links approaches the simulated probabilities to $1/3$, which is the result expected in the well mixed case, as we observe in Fig. 5. Indeed, increasing the number of CLs and therefore approaching all monomers not only increases the chances of repair, but also of mismatch. It is interesting that systems which had a repair probability $\leq 1/3$ because had DSBs at a very

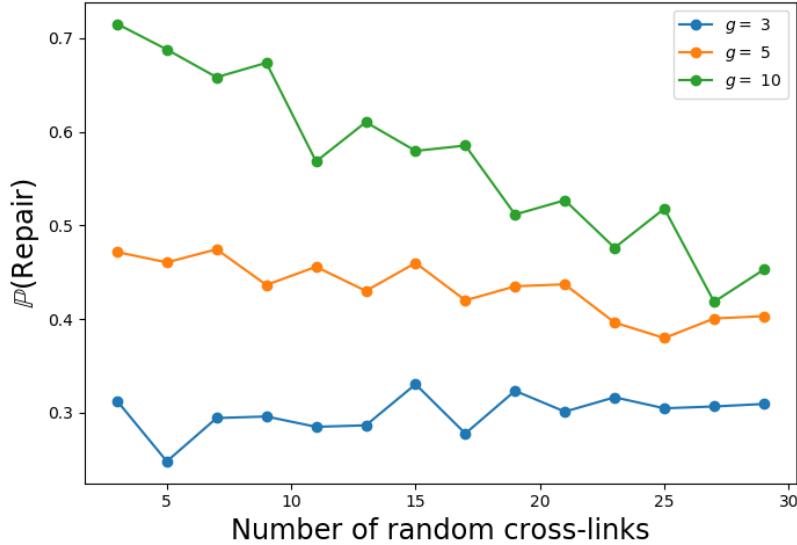


Figure 5: Repair probability as a function of N_c for $N = 100$, $\varepsilon = 0.1\mu\text{m}$ and $N_d = 0$.

small genomic distance ($g \leq 3$) (so the chances of a bad matching were higher) improve their repair probabilities when we increase the number of cross-links.

In effect, increasing the number of random connectors increases as well the compactness of the polymer, revealed as a decrease of the mean radius of gyration (Fig. 6). Besides, we can see that the position of the breaks does not affect the ensemble mean of the gyration radius. In conclusion, **the compactness help to repair polymers cut in a small neighborhood**.

3.2.3 Effect of volume exclusion on the repair probability

Larger values of σ make the waiting for looping times last longer (figure 7) and thus make the encounter events rarer, however they do not affect dramatically the repair probability overall (figure 8).

We see in figure 9 that, as expected, the misrepair scenarios are associated with smaller radius of gyration, and thus with more compact polymers. This could be an indication that relaxation of the polymer chain in the damage loci could lead to better repair probabilities.

We can see in figure 10 that the repair probability in function of the genomic separation between the breaks behaves similarly that previously (cf. figure 4).

Surprisingly however the repair probability function considering excluded volume interactions is always below the repair probability in the non-excluded case (figures 11 and 12)

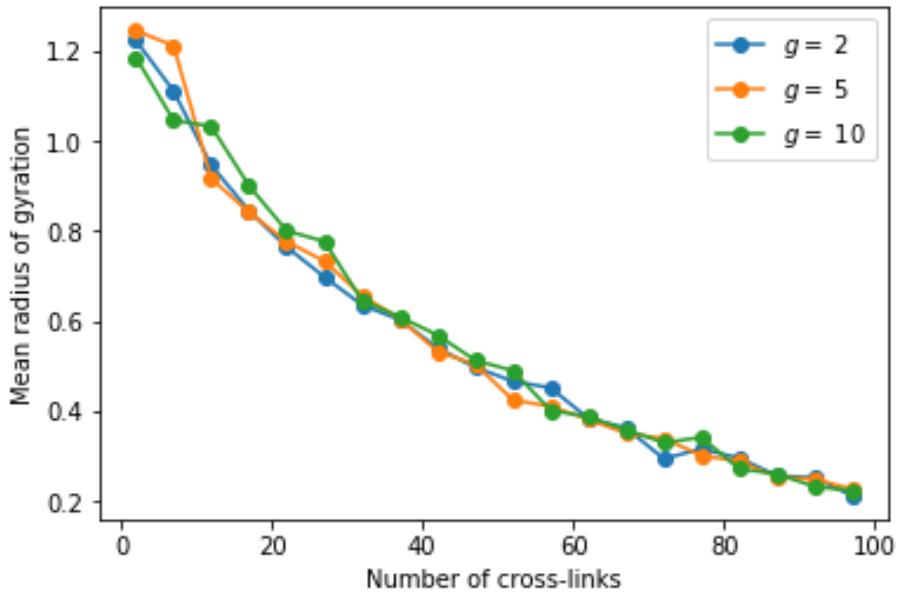


Figure 6: Polymer of 100 monomers. Mean radius of gyration in function of the number of cross-links.

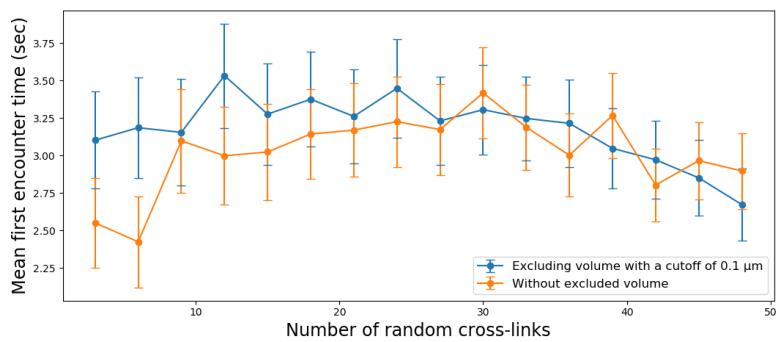


Figure 7: Mean first encounter times in function of the number of cross-links, with and without a radius of 0.1 μm of excluded volume in the damage zone.

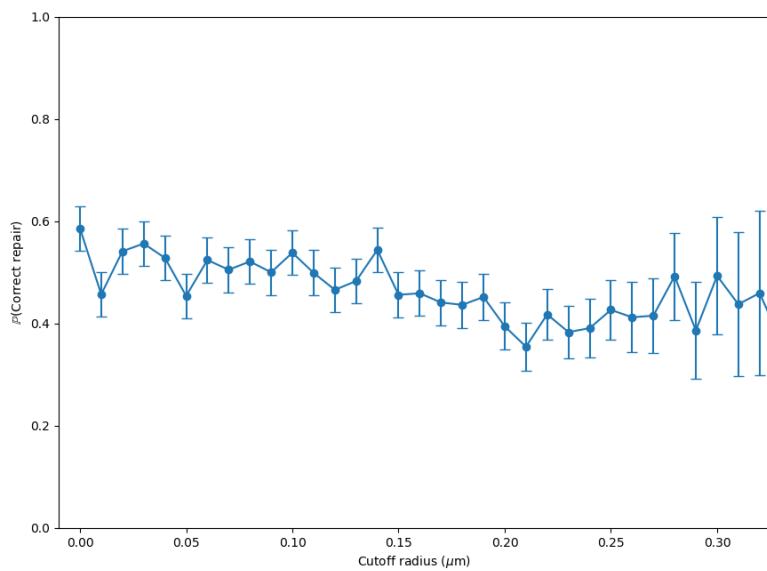


Figure 8: Probability and 95% confidence interval of repair, in function of the cutoff radius of volume exclusion.

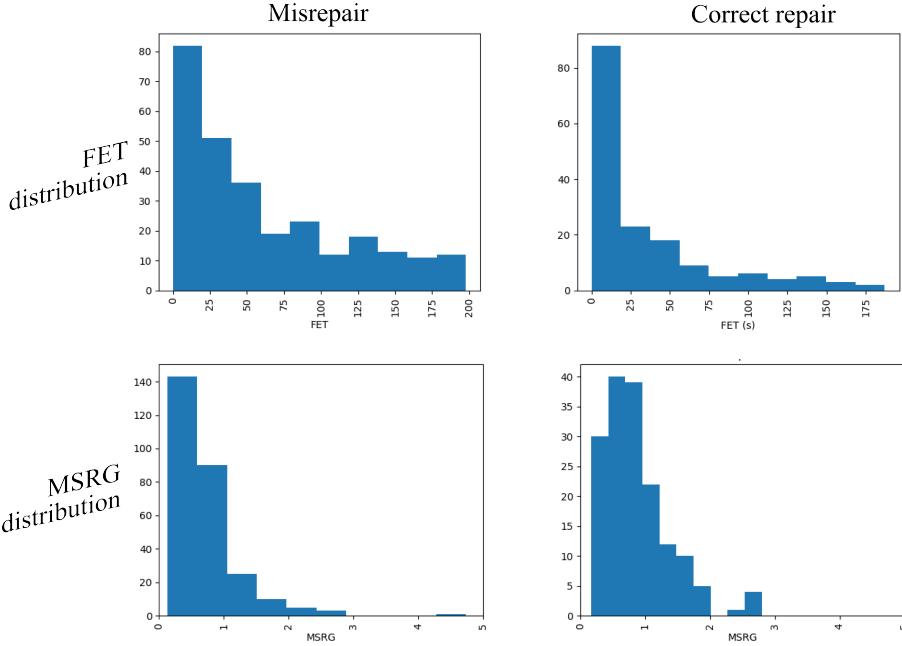


Figure 9: Distribution of the first encounter times and the mean square radius of gyration in repair and misrepair events considering local excluded volume interactions.

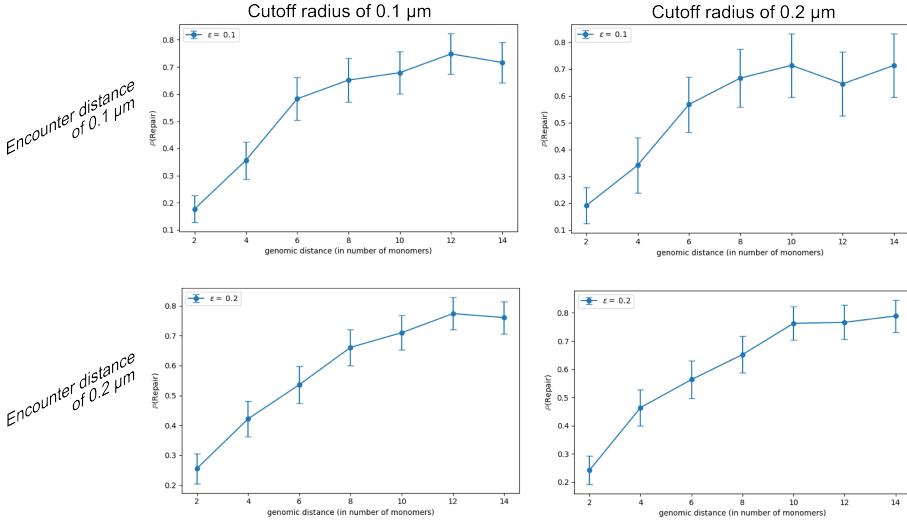


Figure 10: Probability and 95% confidence interval of repair, in function of interbreak genomic distance for exclusion cutoffs $b/2$ and b ($b = 0.1 \mu\text{m}$).

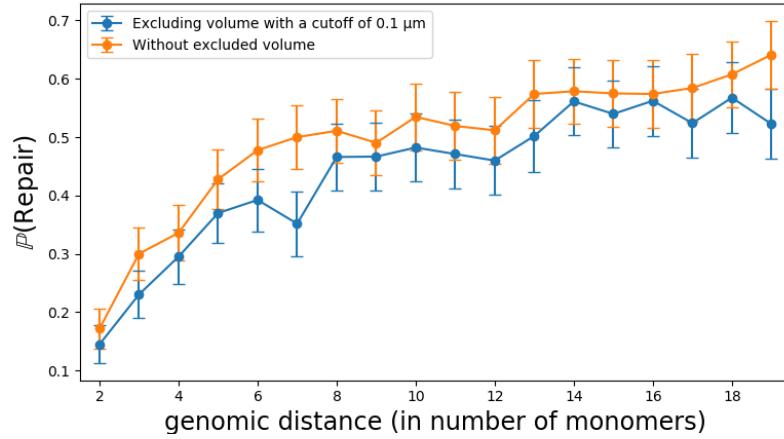


Figure 11: Probability of repair in function of the genomic separation between the breaks.

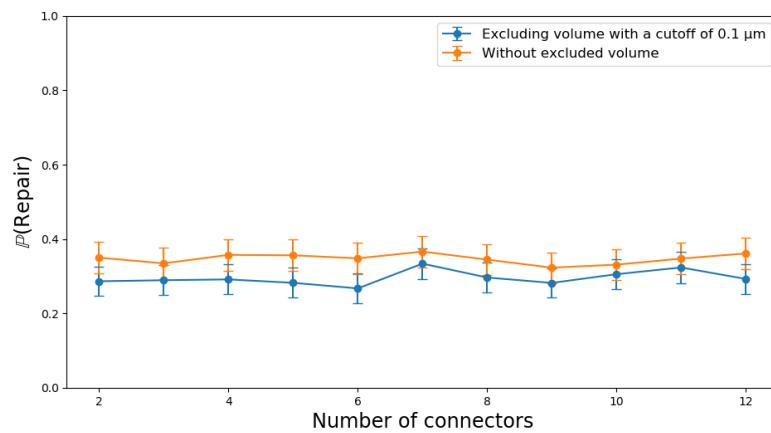


Figure 12: Probability of repair in function of the number of random cross-links, for two breaks separated by 4 monomers ($g=5$).

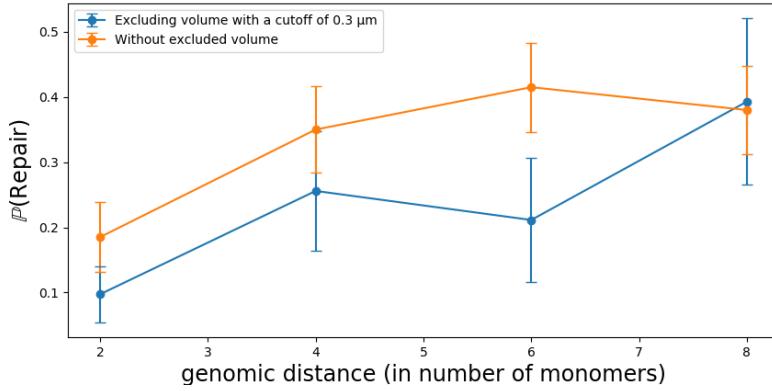


Figure 13: Probability of repair in function of the genomic separation between the breaks. Even for small genomic distances and high connectivity (20 random cross-links) the local volume exclusion seems to interfere with the correct repair, rather than helping it.

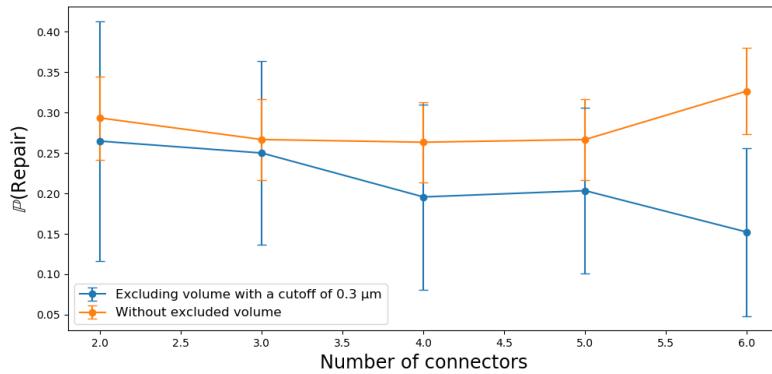


Figure 14: Probability of repair in function of the total number of cross-links, with an inter-break genomic distance $g = 4$.

3.2.4 Effect of volume exclusion with more restrictive encounter distances

After tracking the inter-break physical distances (in μm) as it is presented in figure 15, it seems that the encounter distance of $0.1 \mu\text{m}$ is rather permissive. So, smaller encounter radius will be considered now to see if excluded volume could help the correct repairs.

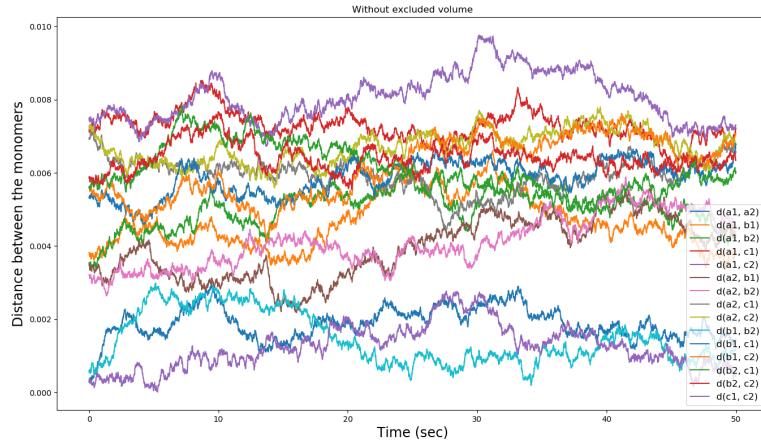


Figure 15: Track of the distances between the ends of three DSBs just before the breaks are induced at relaxation time. The ends of the same break tend to be nearer in comparison to the other distances.

Figure 16 shows the results with and without volume exclusion and with a encounter distance of $\varepsilon = b \cdot 10^{-1} = 0.02 \mu\text{m}$.

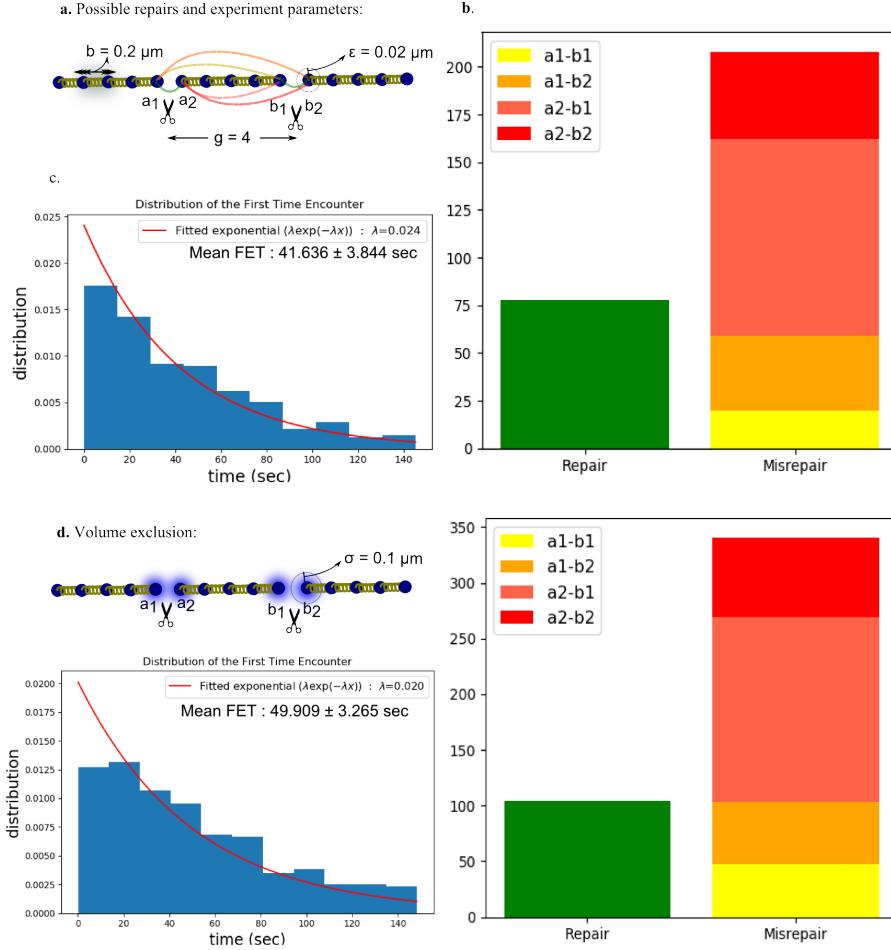


Figure 16: Encounter results with and without volume exclusion for a two DSBs at a genomic distance of 4 and considering 20 random cross-links. **a.** Experiment set. **a.** Encounter events distribution. The most frequent mismatch corresponds to the looping of the central fragment of the broken polymer. **c.** First Encounter Time distribution of all the events. **d.** Same results considering volume exclusion with a cutoff radius of $0.1 \mu\text{m}$.

3.3 Repair probability in different domains

We'll now consider a polymer replied in two connected sub-domains (TADs) (figure 17).

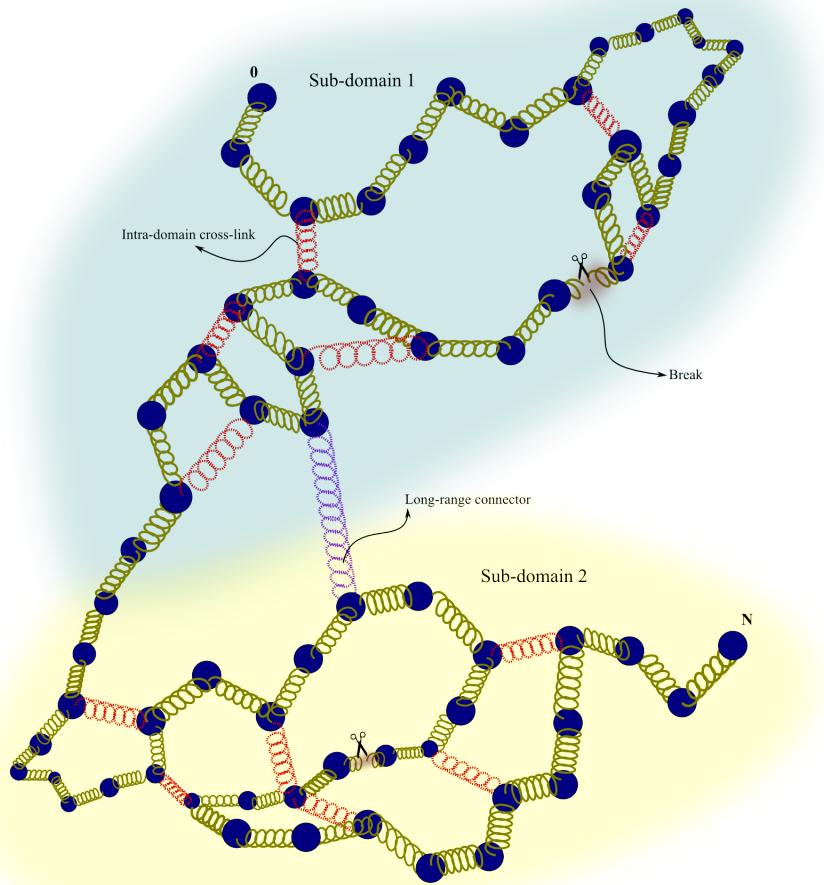


Figure 17: Model of the RCL polymer replied in two connected subdomains.

Figure 18 shows the repair probabilities of two DSBs located at independent connected sub-domains. We see that the correct repair rate is visible higher. The probabilities are observed for two sub-domains where the genomic distance that separate them is increased. The distance between the two sub-domains does not seem to affect the (already good) probability of correct repair.

The following figures illustrates the following experiment: two DSBs are

induced **in the same sub-domain** with a genomic distance between the breaks as in the previous section. The repair probability is measured again to see if the presence of a juxtaposed sub-domain without breaks affects the repair in the first sub-domain.

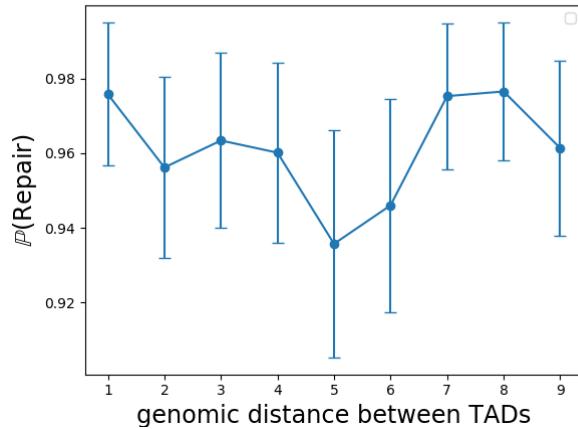


Figure 18: One DSB in each subdomain (two subdomains of size 50). Repair probability vs. inter-TAD separation distance.

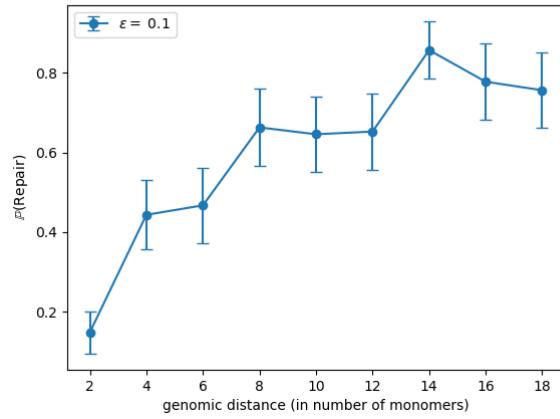


Figure 19: Probability and 95% confidence interval of repair, in function of the interbreak distance for two DSBs induced in one sub-domain of size 100 on a polymer formed by two sub-domains (100 and 50 monomers). The subdomain had 5 RCLs.

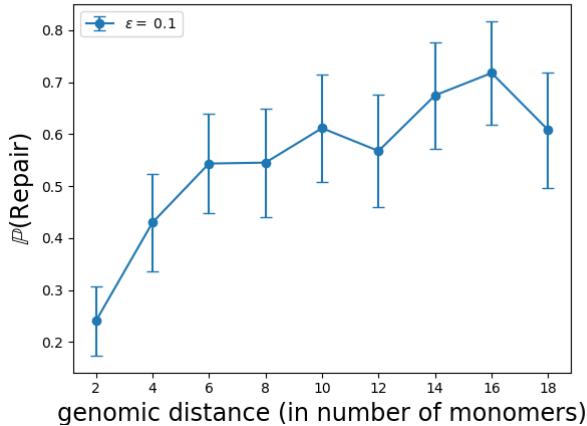


Figure 20: Probability and 95% confidence interval of repair, in function of the interbreak distance for two DSBs induced in one sub-domain of size 100 on a polymer formed by two sub-domains (100 and 50 monomers). The subdomain had 10 RCLs.

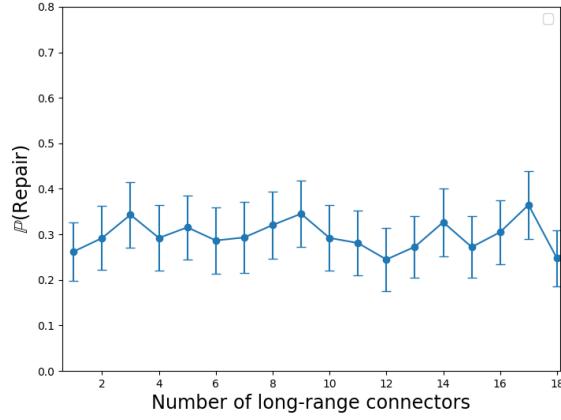


Figure 21: Probability of repair in function of the number of long-range connectors. Two DSBs were induced at a genomic distance of 5 in the first sub-domain of a polymer conformed by 2 sub-domains of 100 monomers with 10 inner cross-links.

Figures 22 and 24 suggest that the presence of a juxtaposed cross-linked polymer chain of any size does not affect the repair performance for two DSBs at a given genomic distance when the number of long-range connectors remains little. Figure 23 suggest that a higher number of long-range connectors increases the randomness of the system, making the probabilities go to $1/3$. If the correct repair probability was already about $1/3$ without long-range connections (for example, thanks to a high number of intra-domain cross-links), adding the long-range links does not change the repair rates.

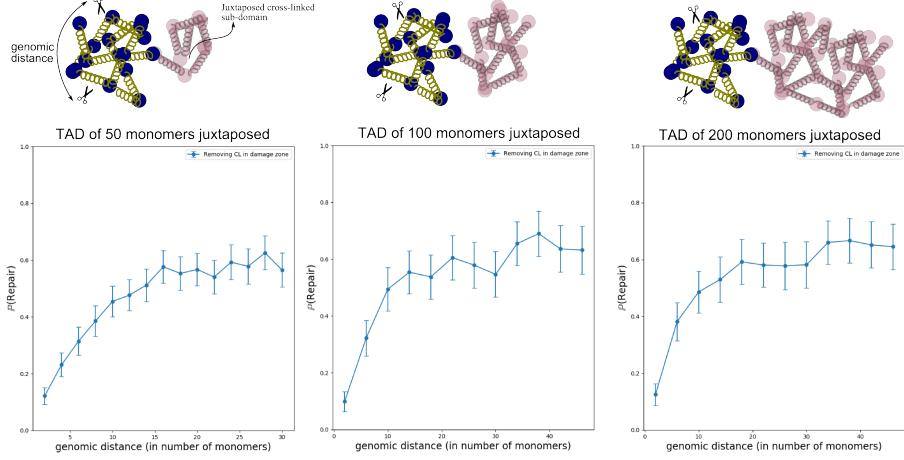


Figure 22: Repair probability of a 100-monomers sub-domain with two DSBs, in function of the genomic inter-break distance between the DSBs. Juxtaposed to the broken sub-domain, there is another sub-domain of size 50, 100 and 200 without breaks. The first sub-domain has 10 inner cross-links and the juxtaposed one has a connectivity of 0.2%. There are no long-range connectors between the two sub-domains.

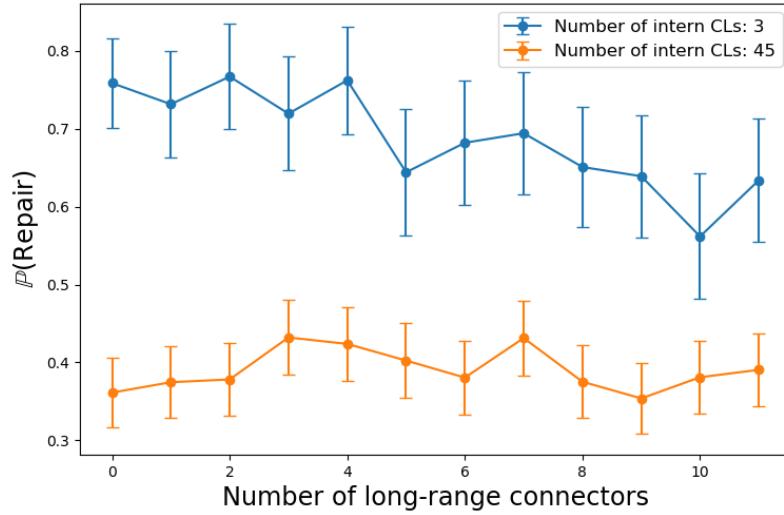


Figure 23: Probability of repair in function of the number of long-range connectors. Two DSBs were induced at a genomic distance of 5 in the first sub-domain of a polymer conformed by 2 sub-domains of 100 monomers with 3 and 45 inner cross-links.

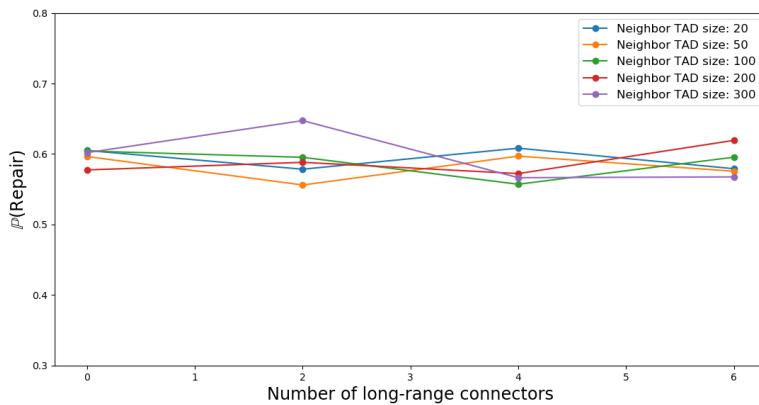


Figure 24: Probability of repair for 0,2,4 and 6 long-range connectors, and different sizes for the juxtaposed unbroken sub-domain. Two DSBs were induced at a genomic distance of 30 in the first sub-domain (100 monomers). The juxtaposed sub-domain has a constant connectivity fraction (0.2%).

3.3.1 TODO

- For one DSB in each TAD: Simulate the repair probability vs. intra-TAD connectors number
- For one DSB in each TAD: Simulate the repair probability vs. long-range connectors number
- For two DSB in a TAD: Compare repair probabilities with the results obtained in the previous section. Does a second TAD (with no breaks) affect the repair probability of the first one? How?
- THINK ABOUT STOCHASTICALLY GENERATING THE CROSS-LINKS, USING IN THE CONNECTIVITY ξ AS A PROBABILITY SO, INSTEAD OF FIXING THE NUMBER OF CROSS-LINKS WE FIX ξ AND HAVE $E(N_c) = \xi N_l$ INSTEAD OF A DETERMINISTIC $N_c = \lfloor \xi N_l \rfloor$. THE RCL POLYMER WILL BE THE AN ERDOS-RENYI GRAPH, THE SUB-DOMAINS POLYMER MATRIX WOULD CORRESPOND TO A WELL KNOWN STOCHASTIC BLOCK MODEL, ETC.

4 Partial conclusions

- When the mismatch rate is high (near DSBs and/or high connectivity) the most common misrepair seems to be the " $a_2 - b_1$ " one, which corresponds to the looping of the central segment of the cut RCL polymer.
- Keeping or removing the cross-links in the separated neighbors does not seem to particularly help the repair, at least without considering volume exclusion. The joint action of CLs removal and volume exclusion in the damage foci should be analyzed more carefully to conclude if this "relaxation" helps or not the repair.
- If there are enough cross-links, even for DSBs separated by a large genomic distance the repair probability goes down to $1/3$, i.e. the expected value by pure chance. The best correct repair probabilities occur, as one could expect, at large interbreak genomic distances with few random cross-links. However, if the breaks are close and the repair probability is less than $1/3$, the increment of cross-links improves the correct match rate.
- Local excluded volume in the damage foci seems to do not affect or even obstruct the repair, rather than improving its probability. Probability curves with local local exclusion volume are indeed down shifted in comparison with the non-excluded ones.

5 Perspectives

- Non-homologous end-joining repair : The repair probability curves in function of the inter-break genomic distance and the number of cross-links reveal a characteristic behavior (tendency to $1/3$ when the connectivity is high, a sort of logistic growth when the genomic distance between the breaks is increased) that could be verified analytically using the results available for the unbroken RCL polymer¹
- For the RCL polymer with subdomains : It has been reported that DSBs cluster. In particular, damaged active genes remain largely unrepairs and clustered to be repaired afterwards via homologous recombination in postreplicative cells². It could be interesting to study if DSB clustering is a property inherent to the polymer dynamics, and if it is facilitated by exclude volume interactions and the local removal of cross-links in the damage foci.

¹O. Shukron and D. Holcman. Statistics of randomly cross-linked polymer models to interpret chromatin conformation capture data. **Phys. Rev. E** 2017 Jul;96, 012503

²Aymard F. *et al.* Genome-wide mapping of long-range contacts unveils clustering of DNA double-strand breaks at damaged active genes. **Nat Struct Mol Biol.** 2017 Apr;24(4):353-361

References

- [1] Amitai A, Kupka I, Holcman D. Computation of the mean first-encounter time between the ends of a polymer chain. *Phys Rev Lett.* 2012 Sep 7;109(10):108302.
- [2] Shukron, O. and Holcman D. Statistics of randomly cross-linked polymer models to interpret chromatin conformation capture data. *Phys. Rev. E* 96, 012503 – Published 27 July 2017.