

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

Ignacio Madrid

May 7, 2018

## 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 of 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 a model of randomly cross-linked (RCL) polymers. A 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

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

and, analogically

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

The model simulated is described in figure 1.

## 2.2 Langevin dynamics of RCL polymers

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

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

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:

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

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. 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:

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

In the following, we implement the Euler's scheme of Eq. 3, 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. 3 will be specified.  $\Delta t = 0.005 \text{ sec}$  will be the usual value. Anyhow, to assure numerical stability, condition in Eq. 4 must always be verified.

$$\sqrt{2D\Delta t} \leq \tau \Delta r^* \quad (4)$$

In Eq. 4,  $\Delta r^*$  stands for the smallest spatial scale used (it will normally be the monomers encounter distance as it is defined later on) and  $\tau$  is a confidence factor that will be set at  $\tau = 0.2$  in the simulations.

### 2.3 Simulation of the induction of Double Strand Breaks (DSBs)

In particular, we are interested on the probability of repair after two DSBs. One DSB (DSB «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 (DSB «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.  $g = B_1 - A_2$ .

### 2.4 Definition of encounter times and probabilities

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

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

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

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

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

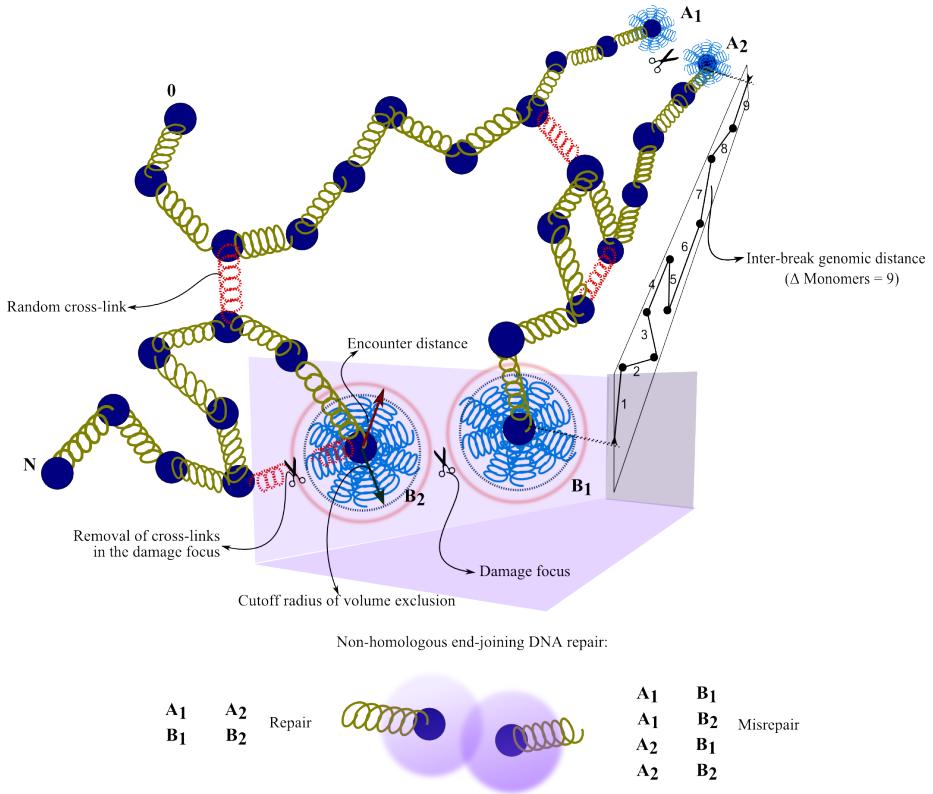


Figure 1: Model of an RCL polymer for chromatin repair simulations. Two random breaks are induced with a deterministic inter-break genomic distance between them. Random cross-links are built uniformly over all combinations of non-neighbor monomers. In the break moment, the cross-links in damage foci (e.g. the CL in monomer  $B_2$ ) might be removed along with the broken bonds. Excluded volume interactions can also be added in the form of a repulsive harmonic potential in the disconnected monomers, simulating the reported chromatin expansion induced just after the DSBs. Finally, repair occurs when two separated neighbors meet at a distance inferior to an encounter distance  $\varepsilon$ .

### 3 Results

#### 3.1 First Encounter Time (FET) distribution

Since an important peak was observed at instant 0 in the first simulations, we decided to wait some time after the induction of double strand breaks (DSBs), before measuring if any pair of monomers have encountered. Concerning the encounters events, if two or more pairs of monomers encounter at the same simulated instant, we toss a coin to choose uniformly over those pairs to be the first encounter. Since the FET distribution can be approximated by a single exponential [?], an exponential function with parameters  $a, \lambda, t \mapsto a \exp(-\lambda t)$  is fitted to the frequency histograms, aiming to compare the  $\lambda$  fitted in each one of the experiments sets.

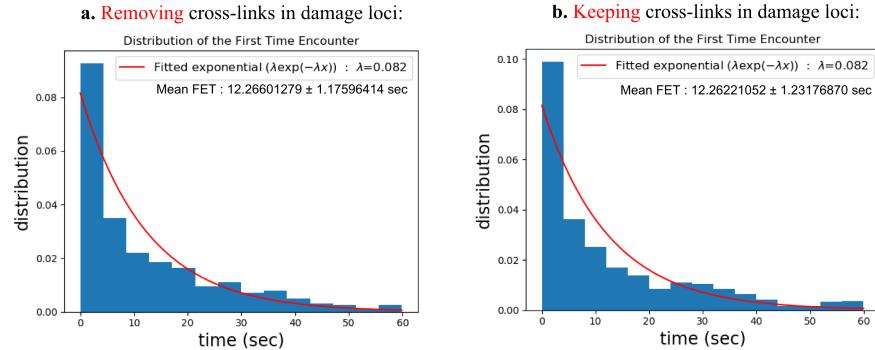


Figure 2: Distribution of the first encounter time keeping and removing the cross-links in the cleaved monomers. An exponential distribution has been fitted to each histogram. The results corresponds to a polymer of 100 monomers, 5 random cross-links, with two DSBs at a genomic distance of 10 and a waiting after-break time of 25 seconds. The events which take longer than 60 seconds are ignored.

The results of figure 2 indicate that the removal of cross-links does not seem to affect the mean first encounter time, nor the ratio of misrepair/repair events. The estimated rate parameter ( $\lambda$ ) is 0.08 for both cases. It should be noted that since the maximum time of simulation is set at 60 seconds (i.e. if nothing happens in 60 seconds the simulation is discarded) the distribution is biased and corresponds to  $\mathcal{L}(FET|FET < 60)$ . For instance, when the simulation maximum time has been increased to 200 seconds, the rate parameter is estimated at 0.04.

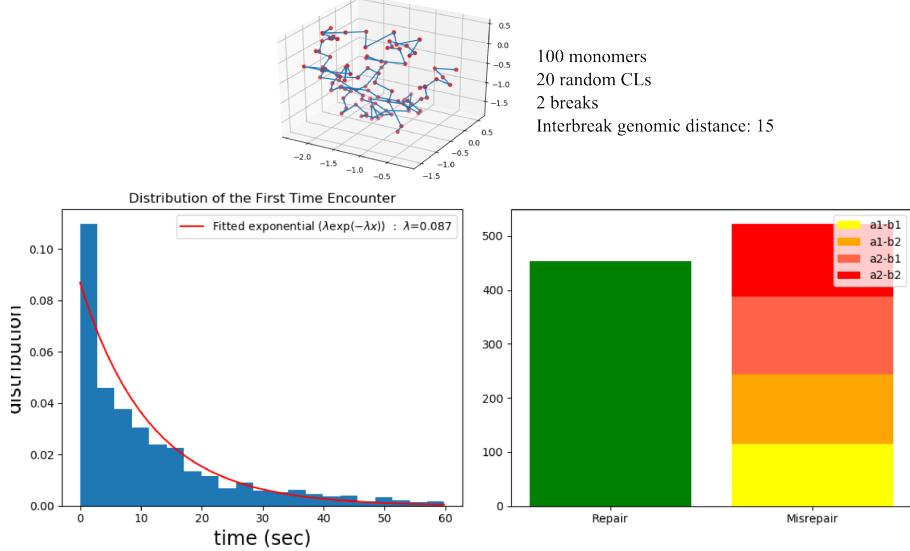


Figure 3: Distribution of the first encounter time keeping the CLs in the damage foci, and distribution of the repair and misrepair events (1000 realisations).

### 3.2 Repair probability

#### 3.2.1 Effect of the genomic distance between two DSBs

We see in figure 4 that regardless the removal of cross-links the repair probability increases in the same way as the inter-break genomic distance increases. So, removing the cross-links in the disconnected monomers does not seem to induce a better repair probability

The most interesting remark though is the gap between the probabilities at a genomic distance of 1 (i.e. the two DSBs are connected,  $A_2$  and  $B_1$  being neighbors) and the other genomic distances (figure 5). As only valid cuts are allowed (i.e. the polymer rests fully connected after the breaks) the segment  $A_2 - B_1$  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.

#### 3.2.2 Effect of the number of cross-links

Increasing the number of cross-links approaches the simulated probabilities to  $1/3$ , which is the result expected by pure chance. Indeed, increasing the number of CLs and therefore approaching all monomers not only increases the chances of a good matching, but also the bad match case. It is interesting that systems which had a repair probability  $\leq 1/3$  because had DSBs at a small genomic

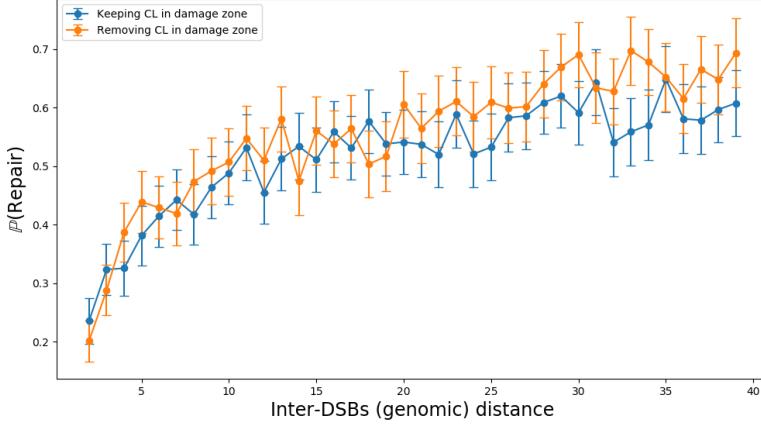


Figure 4: Repair probability against the genomic distance between the two DSBs. Polymer of 100 monomers with 15 random cross-links. Encounter distance of  $0.1 \mu\text{m}$  (and  $b = 0.2 \mu\text{m}$ ).

distance (so the chances of a bad matching where higher) improve their repair probabilities when we increase the number of cross-links. In other words, **the compactness may help to repair polymers cut in a small neighborhood.**

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 (cf. figure 7). Besides, we can see that the position of the breaks does not affect the ensemble mean of the gyration radius.

### 3.2.3 Effect of Excluded Volume

Excluded volume interactions to simulate self-avoidance are added to the polymer dynamics through a new harmonic potential that is added to the potential considered in the Langevin equation:

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

In practice, to simulate the dispersion of chromatin in the break region, we induce volume exclusion to the damage foci monomers only and take  $\kappa_{ev} = \kappa$ :

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

$\sigma$  is a cutoff radius for the excluded volume interactions. Larger values of  $\sigma$  make the waiting for looping times last longer (figure 8) and thus make

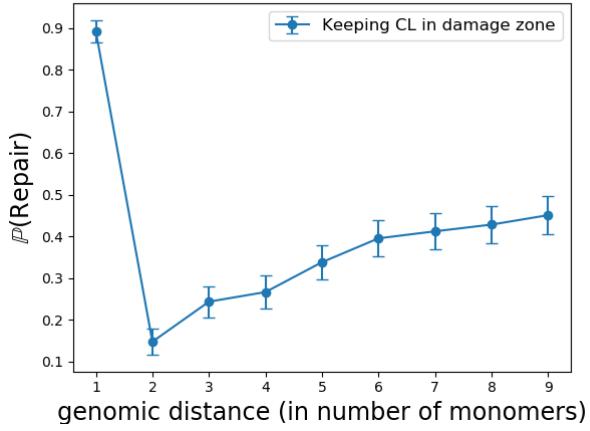


Figure 5: Repair probability against the genomic distance between the two DSBs. Polymer of 100 monomers with 10 random cross-links. Encounter distance of  $0.1 \mu\text{m}$  (and  $b = 0.2 \mu\text{m}$ ).

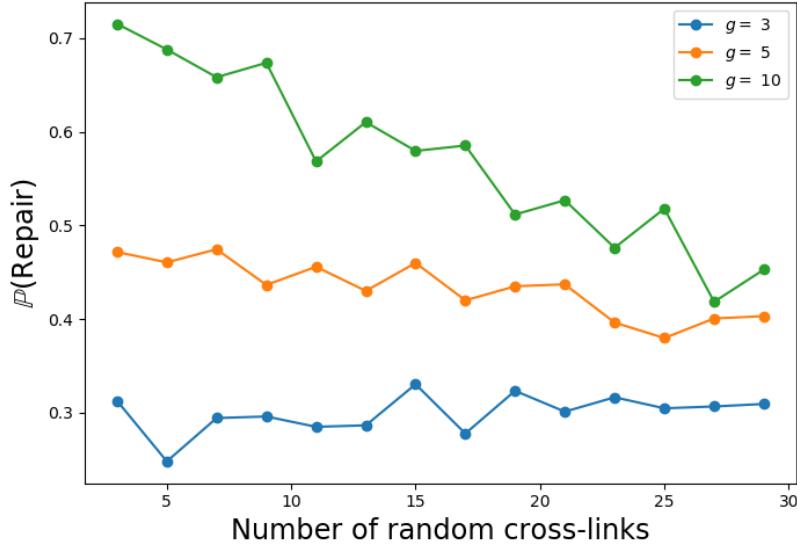


Figure 6: Polymer of 100 monomers. Repair probability against the number of random cross-links when removing the random CLs in damage foci.

the encounter events rarer, however they do not affect dramatically the repair

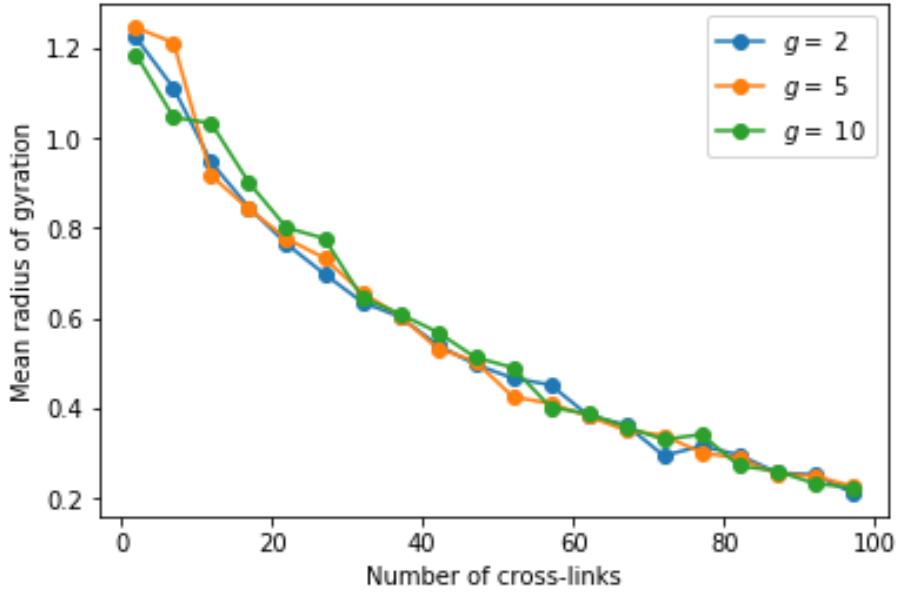


Figure 7: Polymer of 100 monomers. Mean radius of gyration in function of the number of cross-links.

probability overall (figure 9).

We see in figure 10 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 11 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 12 and 13)

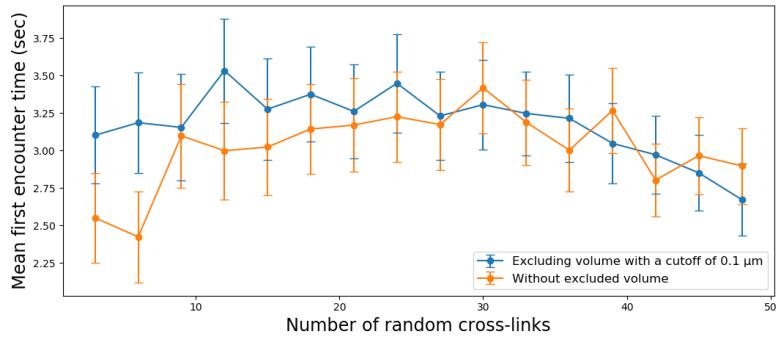


Figure 8: Mean first encounter times in function of the number of cross-links, with and without a radius of  $0.1 \mu\text{m}$  of excluded volume in the damage zone.

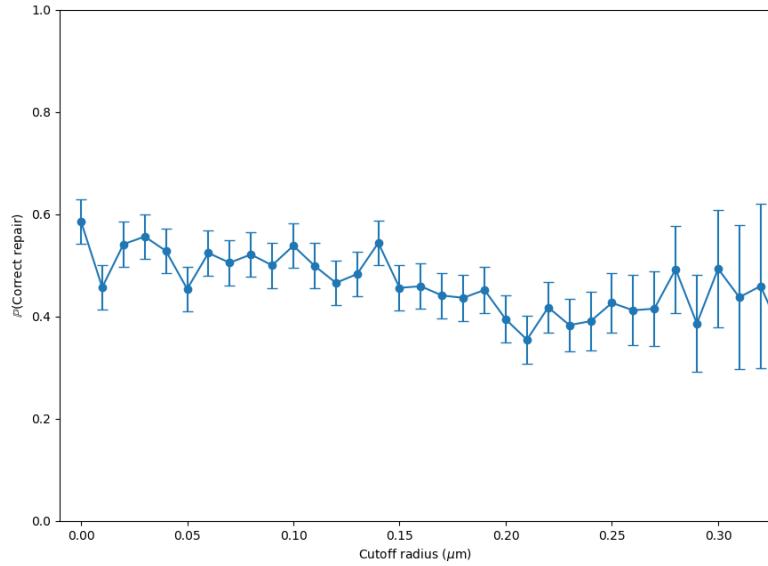


Figure 9: Probability and 95% confidence interval of repair, in function of the cutoff radius of volume exclusion.

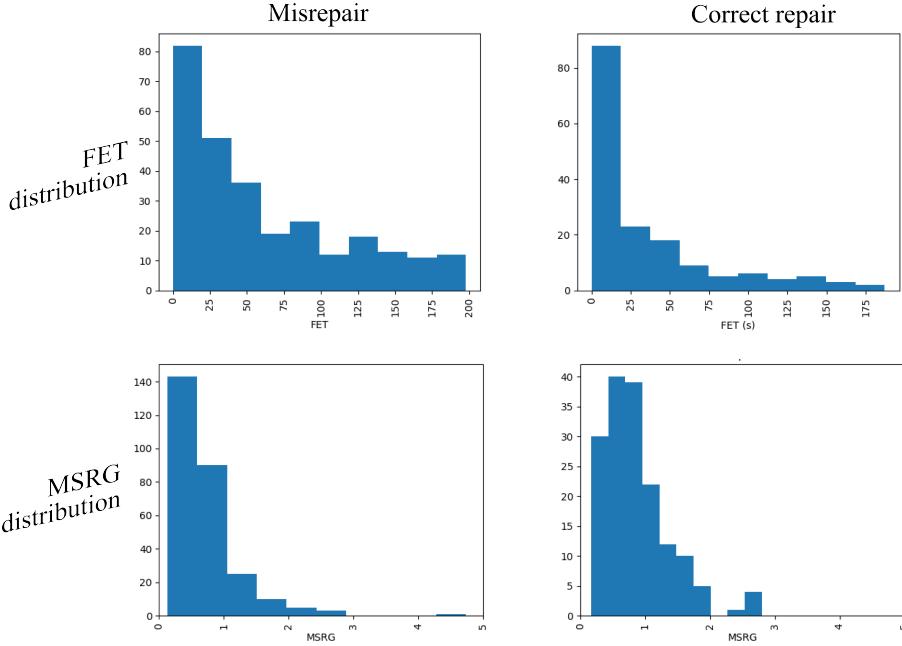


Figure 10: Distribution of the first encounter times and the mean square radius of gyration in repair and misrepair events considering local excluded volume interactions.

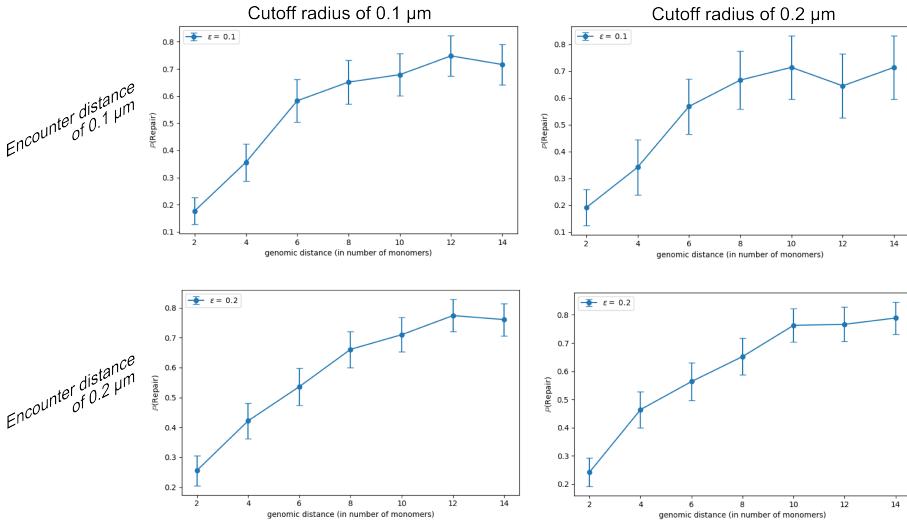


Figure 11: Probability and 95% confidence interval of repair, in function of interbreak genomic distance for exclusion cutoffs  $b/2$  and  $b$  ( $b = 0.1 \mu\text{m}$ ).

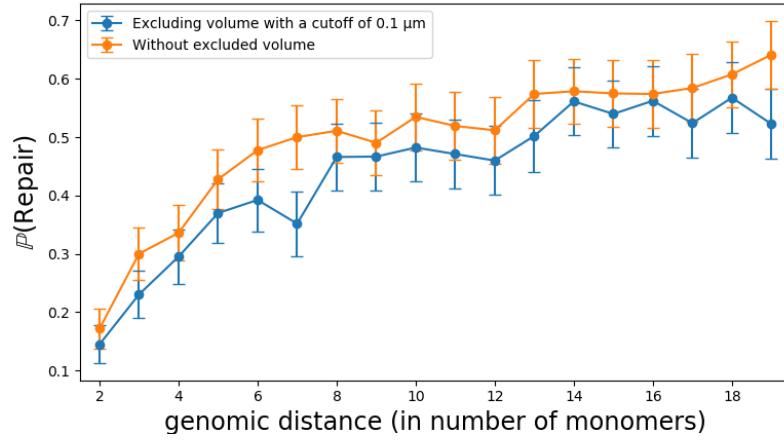


Figure 12: Probability of repair in function of the genomic separation between the breaks.

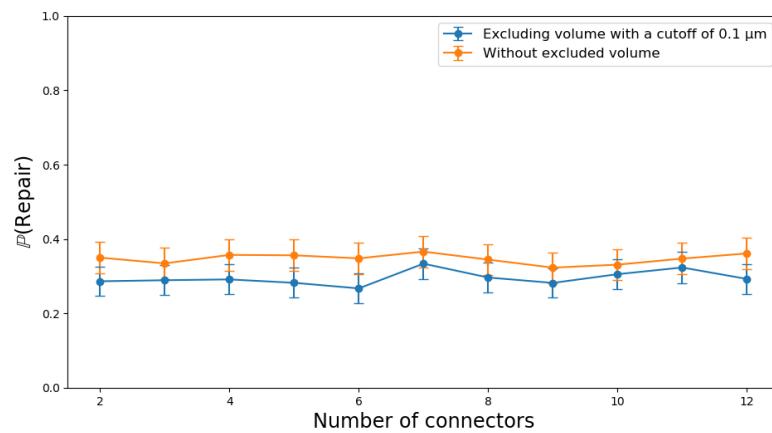


Figure 13: Probability of repair in function of the number of random cross-links, for two breaks separated by 4 monomers ( $g=5$ ).

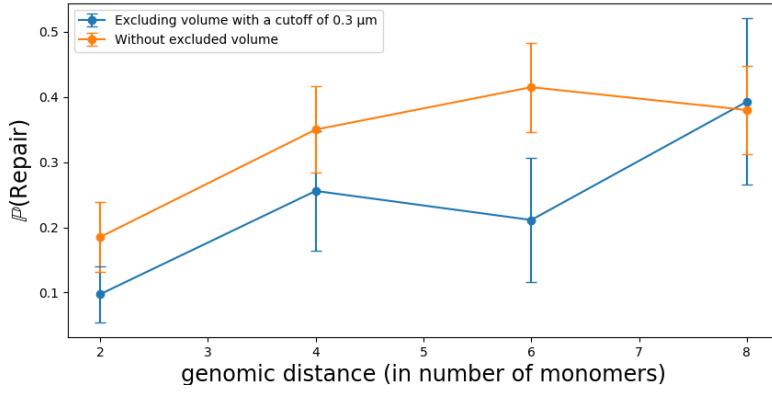


Figure 14: Probability of repair in function of the genomic separation between the breaks. Even for small genomic distances and high connectivity (20 random cross-links) the local volume exclusion seems to interfere with the correct repair, rather than helping it.

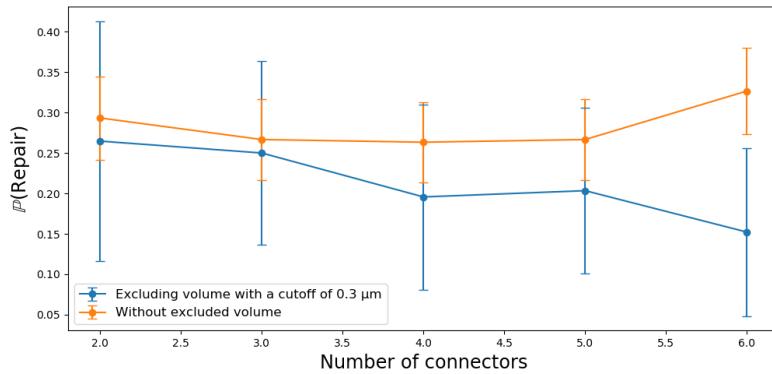


Figure 15: Probability of repair in function of the total number of cross-links, with an inter-break genomic distance  $g = 4$ .

### 3.2.4 Effect of volume exclusion with more restrictive encounter distances

After tracking the inter-break physical distances (in  $\mu\text{m}$ ) as it is presented in figure 16, 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.

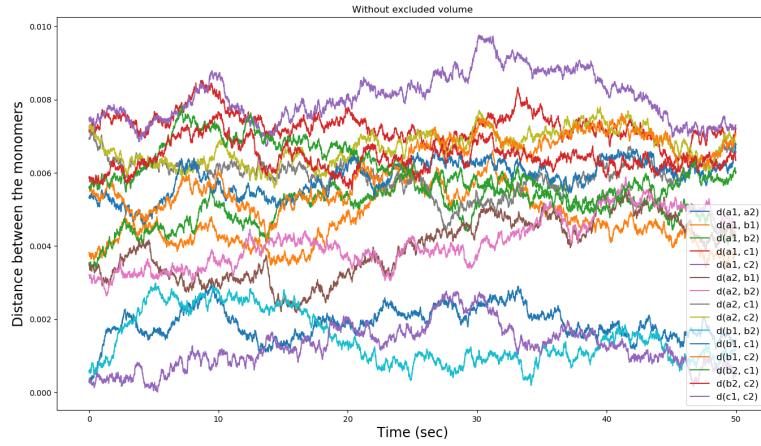


Figure 16: Track of the distances between the ends of three DSBs just before the breaks are induced at relaxation time. The ends of the same break tend to be nearer in comparison to the other distances.

Figure 17 shows the results with and without volume exclusion and with a encounter distance of  $\varepsilon = b \cdot 10^{-1} = 0.02 \mu\text{m}$ .

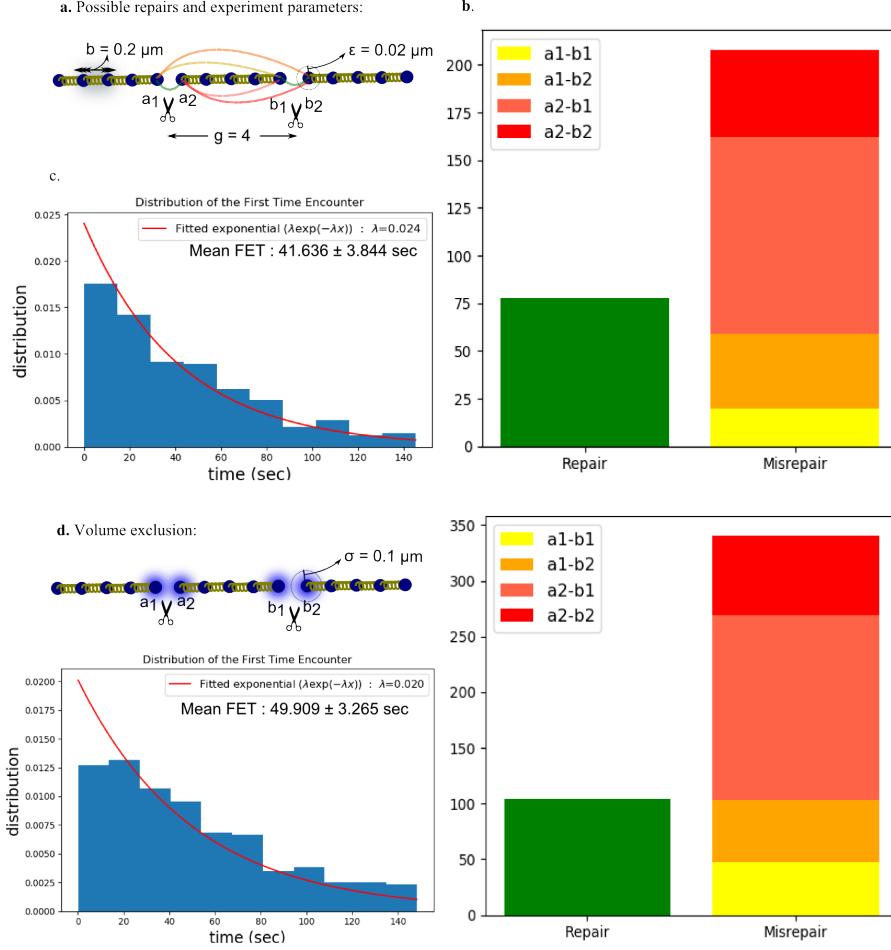


Figure 17: Encounter results with and without volume exclusion for a two DSBs at a genomic distance of 4 and considering 20 random cross-links. **a.** Experiment set. **a.** Encounter events distribution. The most frequent mismatch corresponds to the looping of the central fragment of the broken polymer. **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 18).

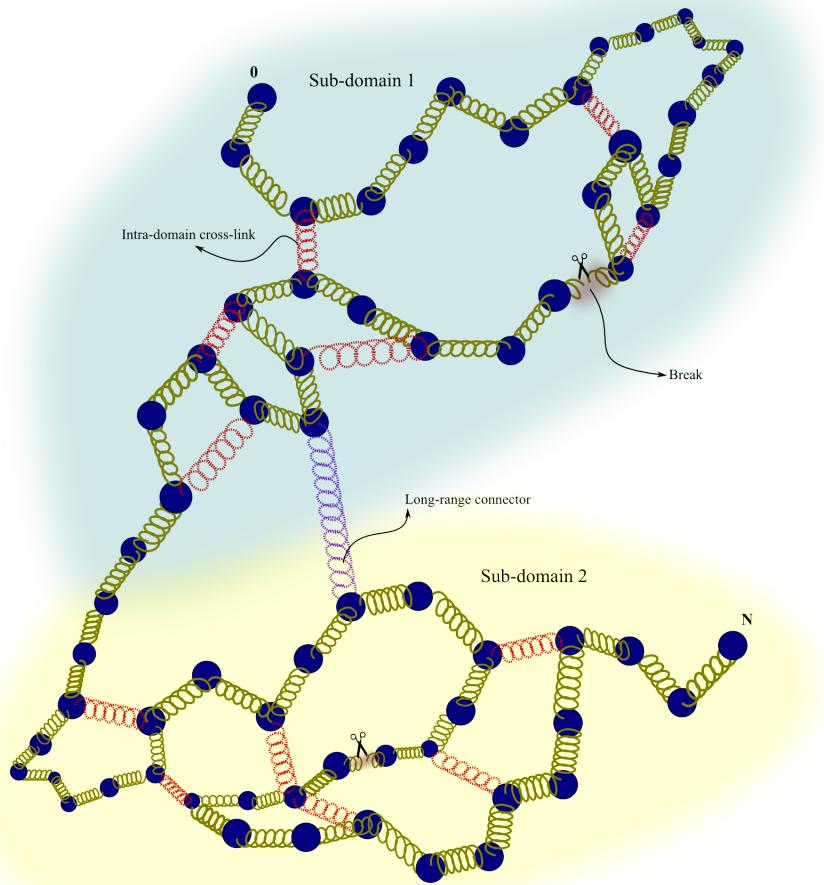


Figure 18: Model of the RCL polymer replied in two connected subdomains.

Figure 19 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.

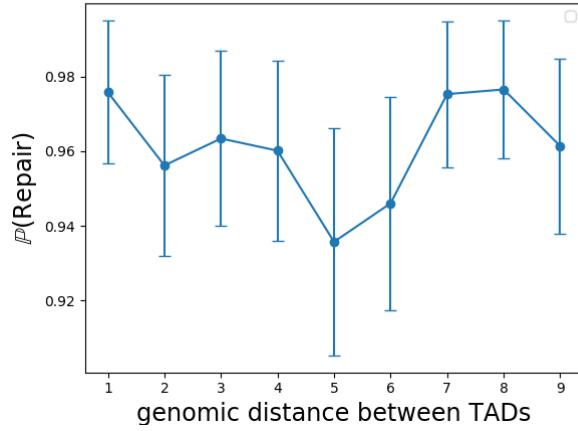


Figure 19: One DSB in each subdomain (two subdomains of size 50). Repair probability vs. inter-TAD separation distance.

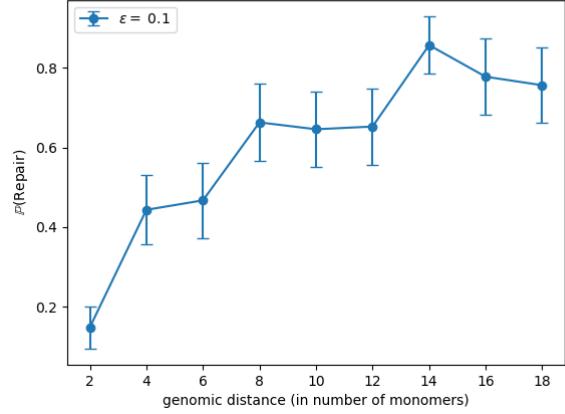


Figure 20: Probability and 95% confidence interval of repair, in function of the interbreak distance for two DSBs induced in one sub-domain of size 100 on a polymer formed by two sub-domains (100 and 50 monomers). The subdomain had 5 RCLs.

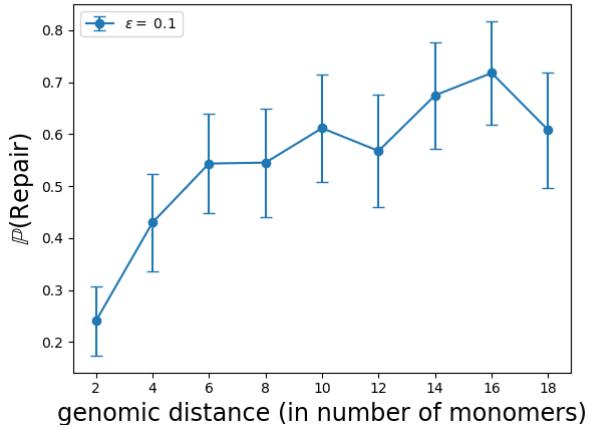


Figure 21: Probability and 95% confidence interval of repair, in function of the interbreak distance for two DSBs induced in one sub-domain of size 100 on a polymer formed by two sub-domains (100 and 50 monomers). The subdomain had 10 RCLs.

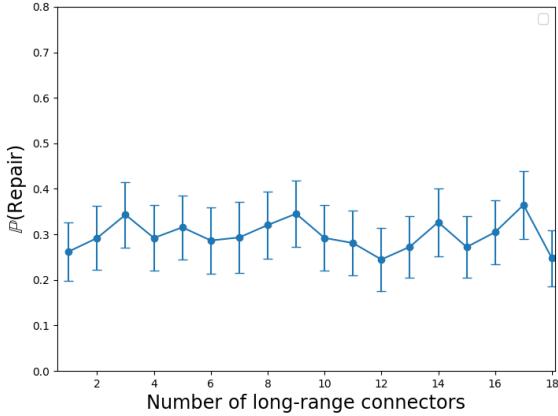


Figure 22: Probability of repair in function of the number of long-range connectors. Two DSBs were induced at a genomic distance of 5 in the first sub-domain of a polymer conformed by 2 sub-domains of 100 monomers with 10 inner cross-links.

Figures 23 and 25 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 24 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.

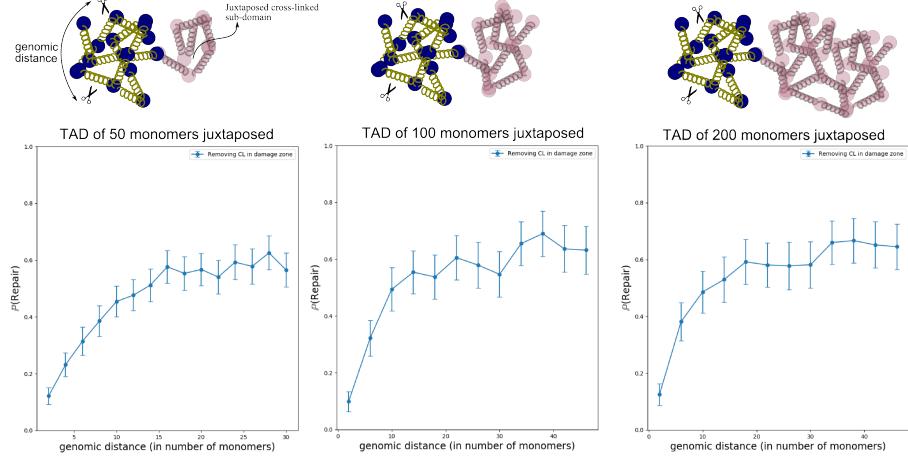


Figure 23: Repair probability of a 100-monomers sub-domain with two DSBs, in function of the genomic inter-break distance between the DSBs. Juxtaposed to the broken sub-domain, there is another sub-domain of size 50, 100 and 200 without breaks. The first sub-domain has 10 inner cross-links and the juxtaposed one has a connectivity of 0.2%. There are no long-range connectors between the two sub-domains.

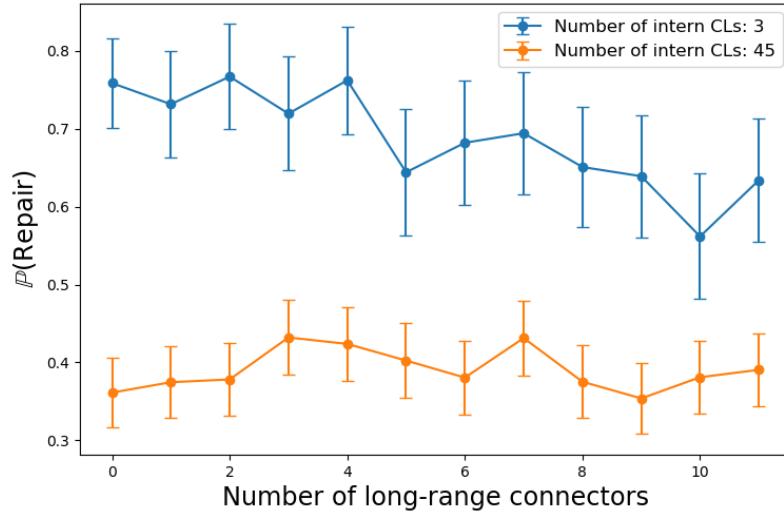


Figure 24: Probability of repair in function of the number of long-range connectors. Two DSBs were induced at a genomic distance of 5 in the first sub-domain of a polymer conformed by 2 sub-domains of 100 monomers with 3 and 45 inner cross-links.

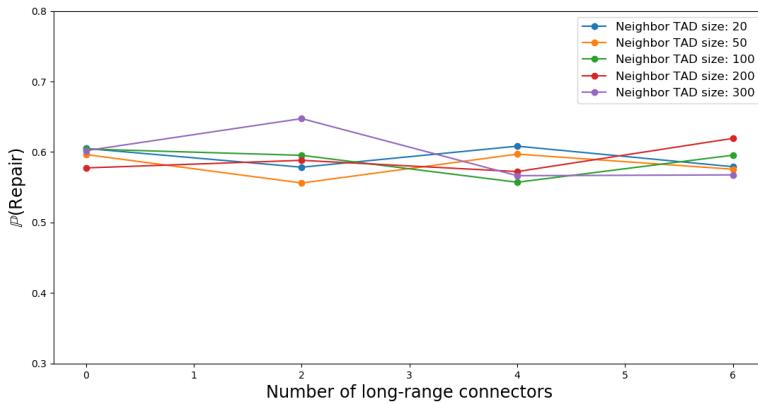


Figure 25: Probability of repair for 0,2,4 and 6 long-range connectors, and different sizes for the juxtaposed unbroken sub-domain. Two DSBs were induced at a genomic distance of 30 in the first sub-domain (100 monomers). The juxtaposed sub-domain has a constant connectivity fraction (0.2%).

### 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  $\xi$  AS A PROBABILITY SO, INSTEAD OF FIXING THE NUMBER OF CROSS-LINKS WE FIX  $\xi$  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 polymer<sup>1</sup>
- For the RCL polymer with subdomains : It has been reported that DSBs cluster. In particular, damaged active genes remain largely unrepaired and clustered to be repaired afterwards via homologous recombination in postreplicative cells<sup>2</sup>. 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.

---

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

<sup>2</sup>Aymard 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

amitai12 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.