

# Progress report: Repair probabilities for the RCL polymer with two random breaks

Ignacio Madrid

April 27, 2018

## 1 Introduction

The model simulated is described in figure 1. We supposed a random cross-linked polymer (RCL polymer) subdued to Langevin dynamics, i.e., if  $(R_t)_{t \geq 0} = (R_1, \dots, R_N)_{t \geq 0}$  are the monomer positions of the polymer in time, 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 an RCL polymer is given by the classical Rouse potential and by the potential induced by the random cross-links:

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

where  $\kappa$  and  $\kappa'$  are spring constants we will suppose equal,  $M$  is the Rouse matrix (degree - connectivity),  $B$  is the degree - connectivity matrix with the added random connectors, and  $\mathcal{R}$  is the set of monomers which have been randomly connected. Since  $\frac{\kappa}{\gamma} = \frac{3D}{b^2}$  where  $b$  is the standard deviation of the bonds length, we'll consider the equation :

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

In the following, we implement the Euler's scheme of the previous equation setting  $D = 0.008 \mu^2 \text{m/s}$  and  $b = 0.2 \mu \text{m}$ . In particular, we are interested on the probability of repair after two random breaks (in monomers  $A_1 - A_2, B_1 - B_2$ , where  $A_1 \sim \mathcal{U}[1, N - g - 2]$ ,  $1 \leq A_1 < A_2 = A_1 + 1 < B_1 < B_2 = B_1 + 1 < N$ ) with a deterministic inter-break genomic distance  $g$  between them ( $B_1 - A_2 = g$ ).

So, if we define the first encounter time as

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

where  $R_m$  and  $R_n$  are the positions of monomers  $m$  and  $n$ , and  $\epsilon$  is the encounter distance, then the repair probability is

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

NB: By pure chance, the expected repair probability is  $2/6 = 1/3$ .

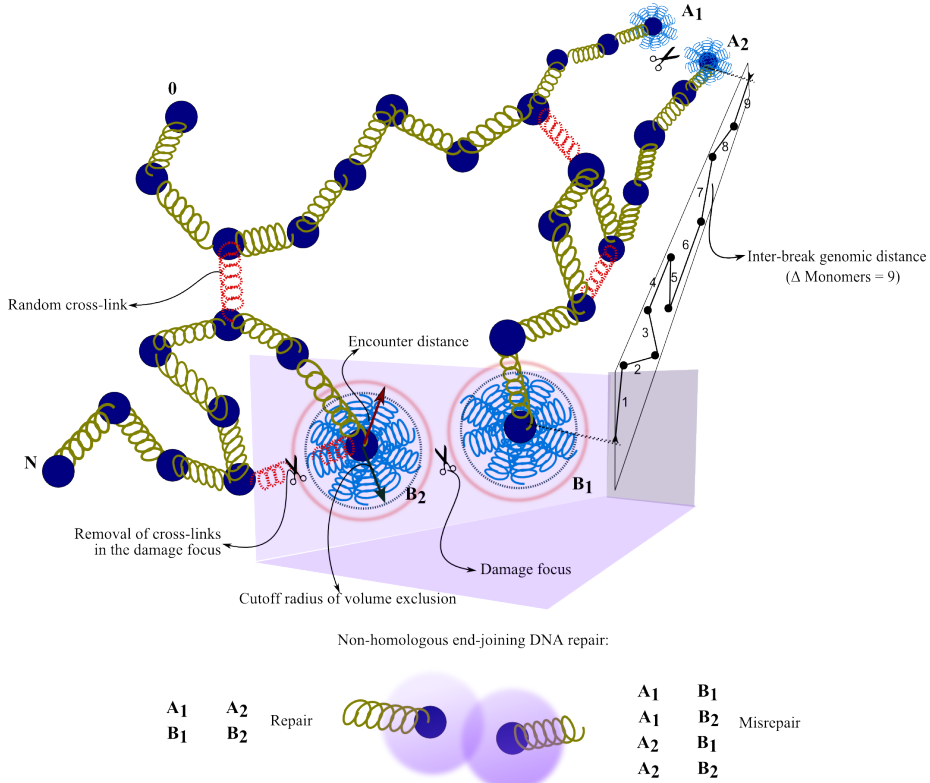


Figure 1: Model of an RCL polymer for chromatin repair simulations. Two random breaks are induced with a deterministic inter-break genomic distance between them. Random cross-links are built uniformly over all combinations of non-neighbor monomers. In the break moment, the cross-links in damage foci (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  $\epsilon$ .

## 2 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.

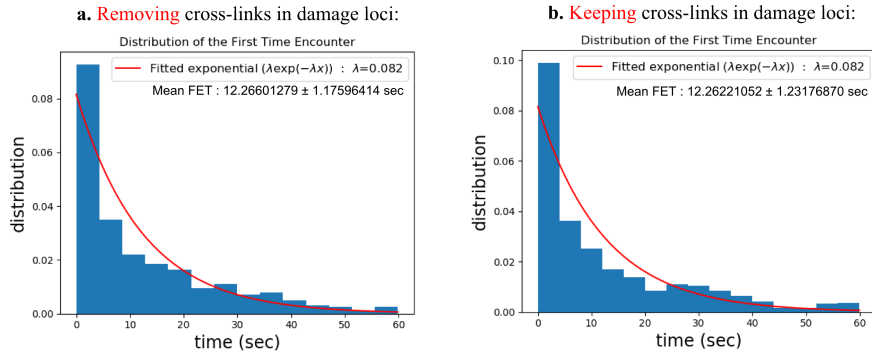


Figure 2: Distribution of the first encounter time keeping and removing the cross-links in the cleaved monomers. An exponential distribution has been fitted to each histogram. The results corresponds to a polymer of 100 monomers, 5 random cross-links, with two DSBs at a genomic distance of 10 and a waiting 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 the cross-links seems not affect the mean first encounter time, nor the ratio of fail/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 50 seconds the simulation is discarded) the distribution is biased and corresponds to  $\mathcal{L}(FET|FET < 50)$ . For instance, when the simulation maximum time has been increased to 200 seconds, the parameter is estimated at 0.04.

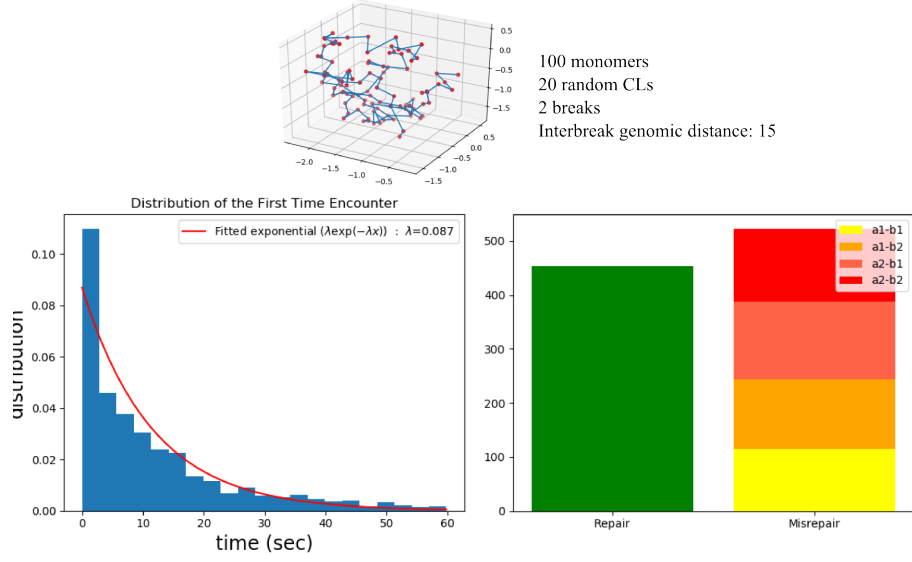


Figure 3: Distribution of the first encounter time keeping the CLs in the damage foci, and distribution of the repair and misrepair events (1000 realisations).

### 3 Repair probability

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

We see 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). Some hypothesis: the wrong combination  $A_2 - B_1$  is no possible anymore, so in fact, by pure chance, we have a  $2/5$  probability of repair (instead of  $2/6$ ), and the combinations  $A_2 - B_2$ ,  $A_1 - B_1$  could be very rare. Besides, 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 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

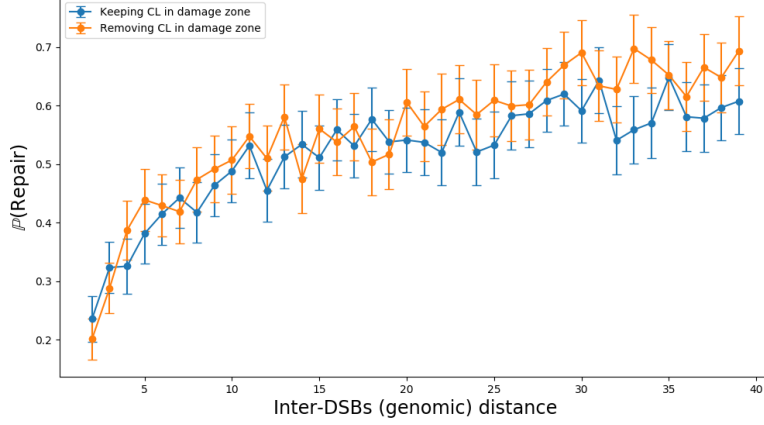


Figure 4: Repair probability against the genomic distance between the two DSBs. Polymer of 100 monomers with 15 random cross-links. Encounter distance of  $0.1 \mu\text{m}$  (and  $b = 0.2 \mu\text{m}$ ).

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 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.

### 3.3 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:

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

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

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

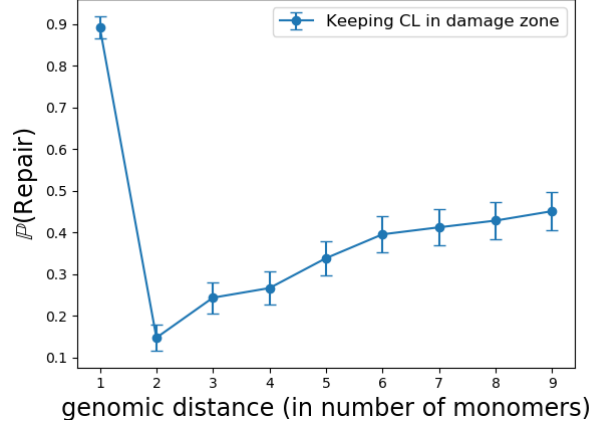


Figure 5: Repair probability against the genomic distance between the two DSBs. Polymer of 100 monomers with 10 random cross-links. Encounter distance of  $0.1 \mu\text{m}$  (and  $b = 0.2 \mu\text{m}$ ).

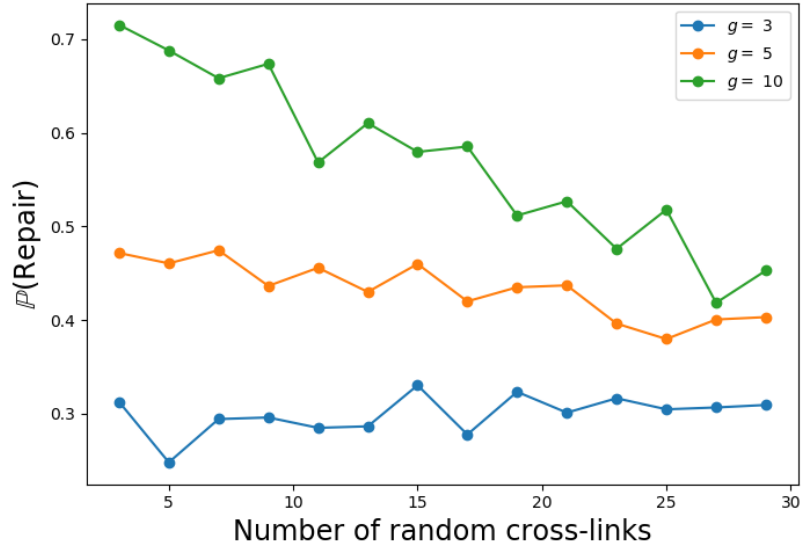


Figure 6: Polymer of 100 monomers. Repair probability against the number of random cross-links when removing the random CLs in damage foci.

$\sigma$  is a cutoff radius for the excluded volume interactions. Larger values

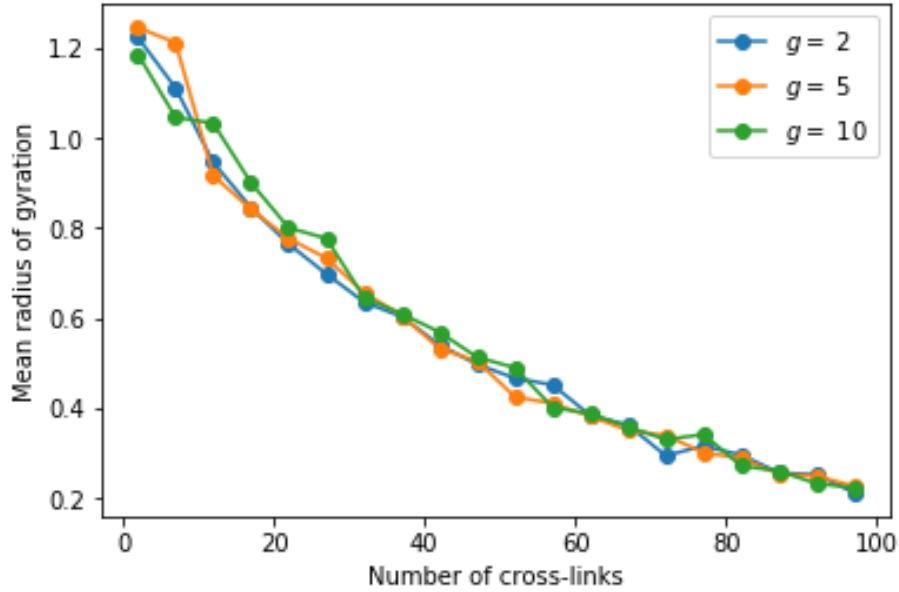


Figure 7: Polymer of 100 monomers. Mean radius of gyration in function of the number of cross-links.

of  $\sigma$  make the waiting for looping times last longer (figure 8) and thus make the encounter events rarer, however they do not affect dramatically the repair 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)

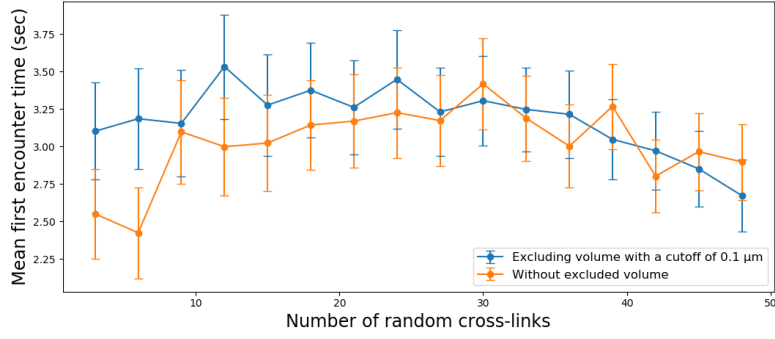


Figure 8: Mean first encounter times in function of the number of cross-links, with and without a radius of 0.1  $\mu\text{m}$  of excluded volume in the damage zone.

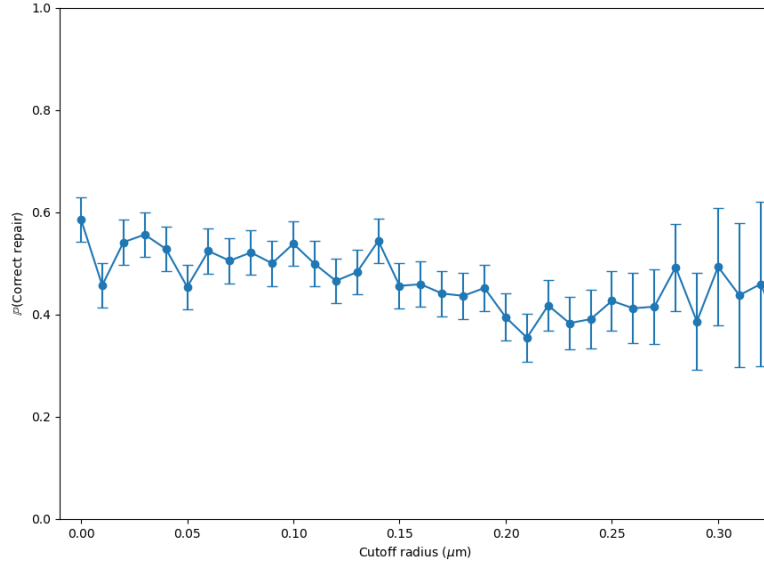


Figure 9: Probability and 95% confidence interval of repair, in function of the cutoff radius of volume exclusion.



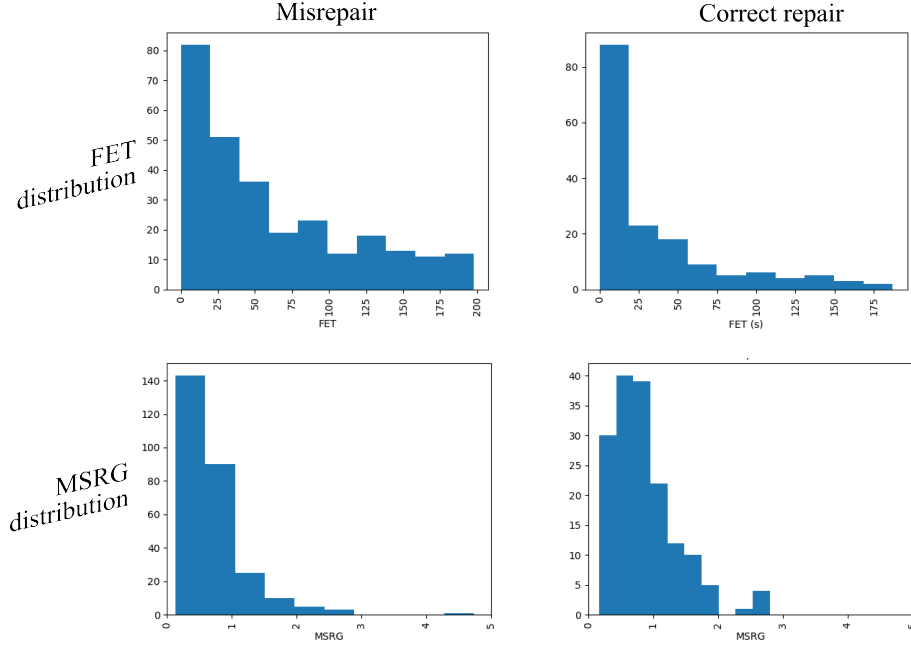


Figure 10: Distribution of the first encounter times and the mean square radius of gyration in repair and misrepair events considering local excluded volume interactions.

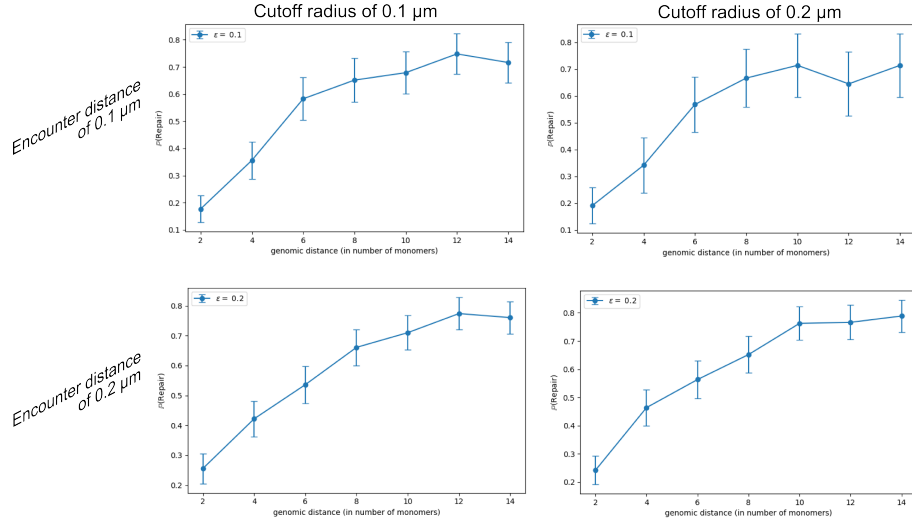


Figure 11: Probability and 95% confidence interval of repair, in function of interbreak genomic distance for exclusion cutoffs  $b/2$  and  $b$  ( $b = 0.1 \mu\text{m}$ ).

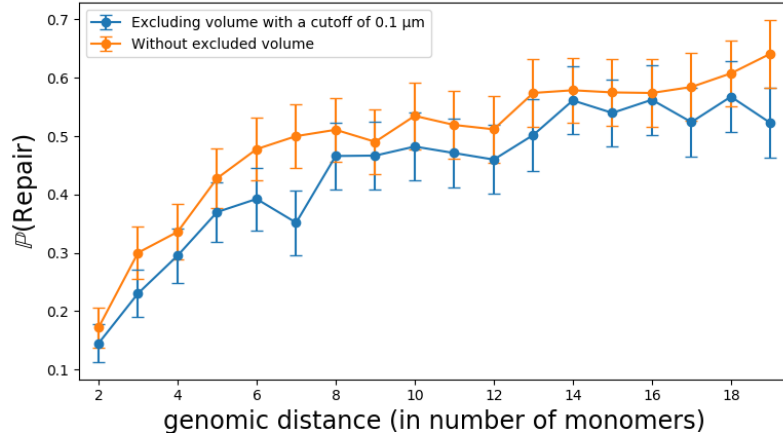


Figure 12: Probability of repair in function of the genomic separation between the breaks.

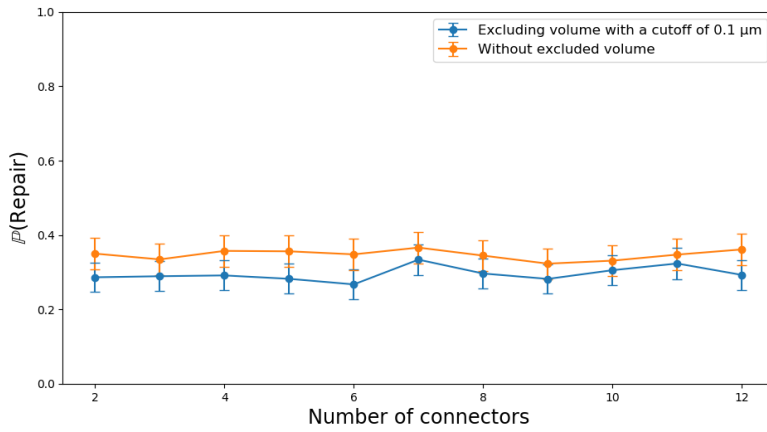


Figure 13: Probability of repair in function of the number of random cross-links, for two breaks separated by 4 monomers ( $g=5$ ).

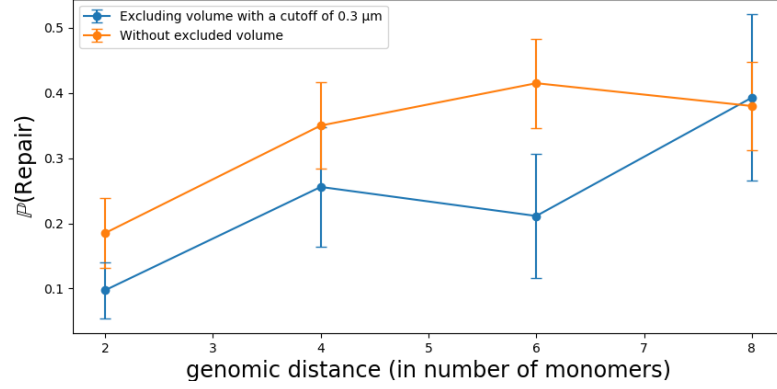


Figure 14: Probability of repair in function of the genomic separation between the breaks. Even for small genomic distances and high connectivity (20 random cross-links) the local volume exclusion seems to interfere with the correct repair, rather than helping it.

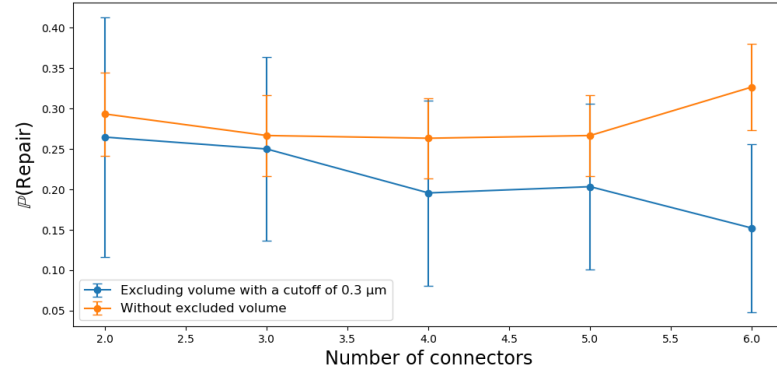


Figure 15: Probability of repair in function of the total number of cross-links, with an inter-break genomic distance  $g = 4$ .

### 3.4 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.

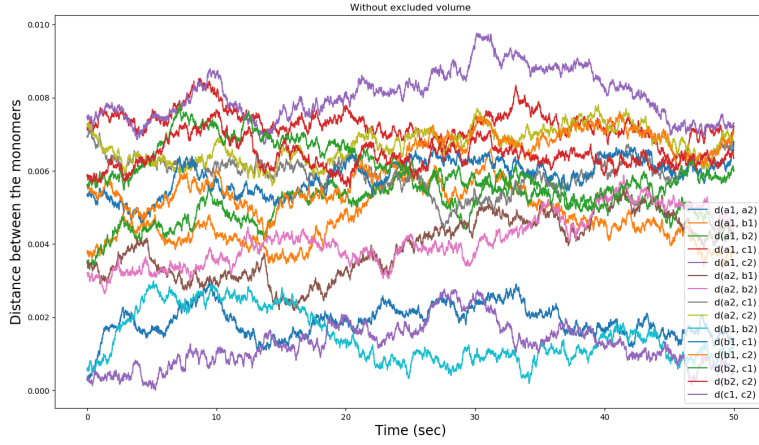


Figure 16: Track of the distances between the ends of three DSBs just before the breaks are induced at relaxation time. The ends of the same break tend to be nearer in comparison to the other distances.

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

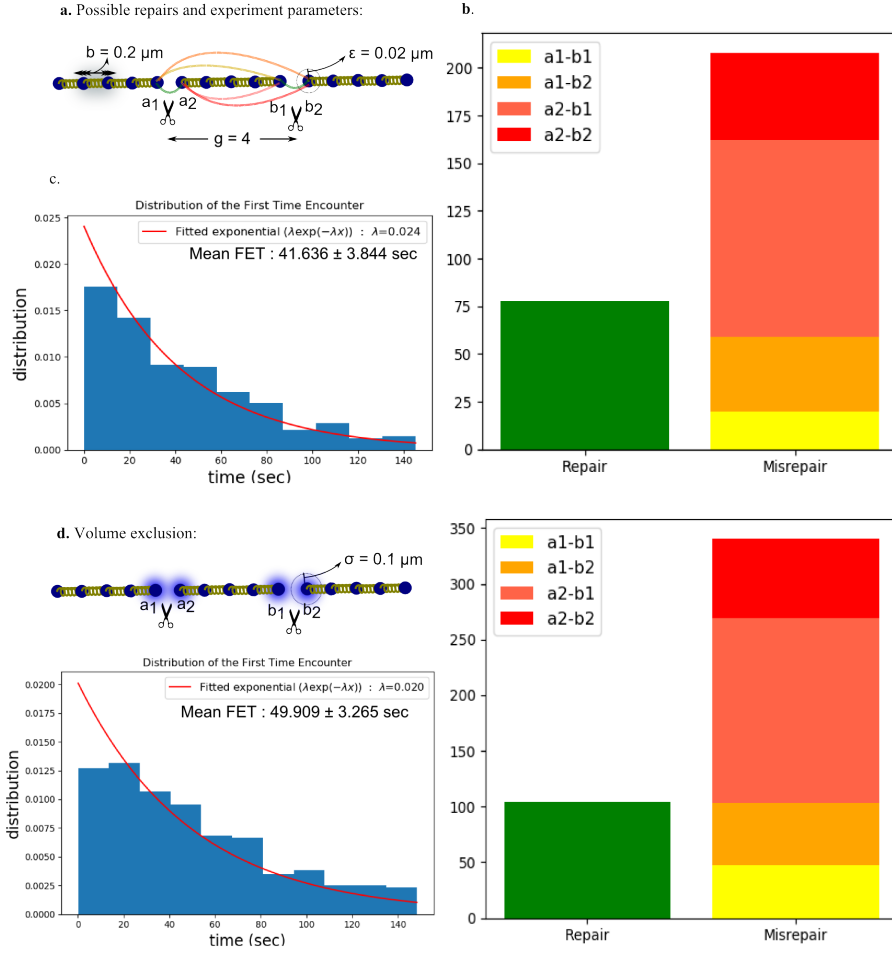


Figure 17: Encounter results with and without volume exclusion for a two DSBs at a genomic distance of 4 and considering 20 random cross-links. **a.** Experiment set. **a.** Encounter events distribution. The most frequent mismatch corresponds to the looping of the central fragment of the broken polymer. **c.** First Encounter Time distribution of all the events. **d.** Same results considering volume exclusion with a cutoff radius of  $0.1 \mu\text{m}$ .

## 4 Repair probability in different domains

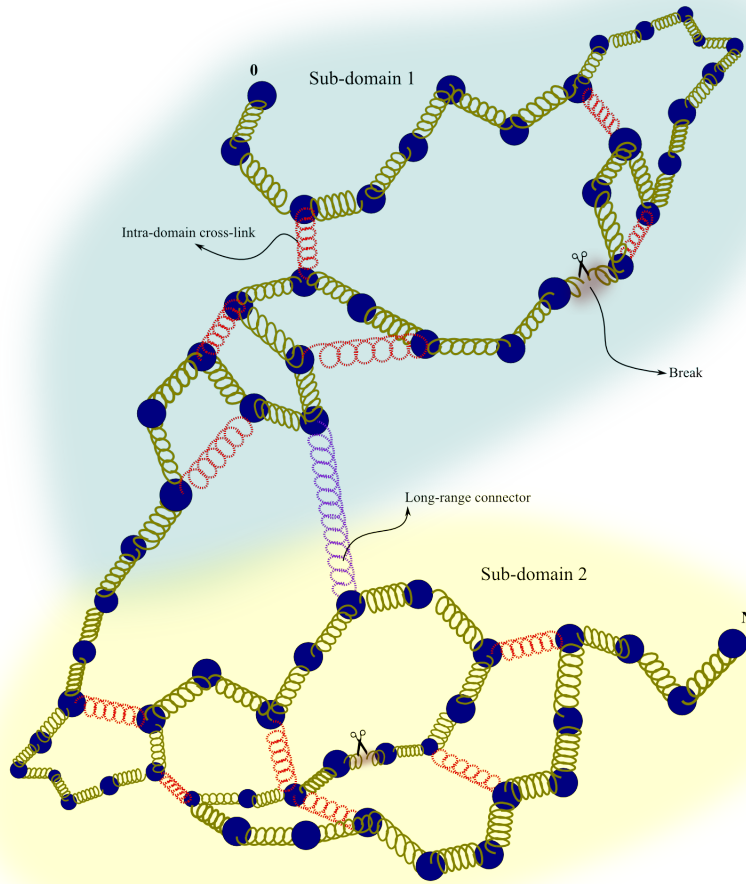


Figure 18: Model of the RCL polymer replied in two connected subdomains.

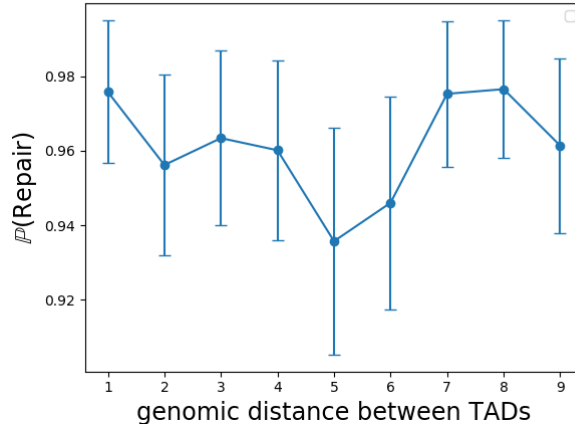


Figure 19: One DSB in each subdomain (two subdomains of size 50). Repair probability vs. inter-TAD separation distance.

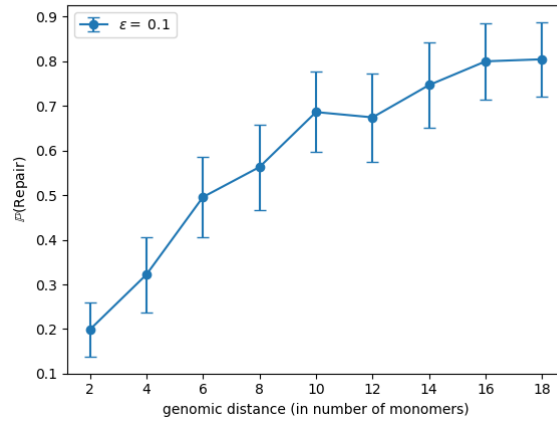


Figure 20: Control: Probability and 95% confidence interval of repair, in function of the interbreak distance. 5 RCLs over 100 monomers.

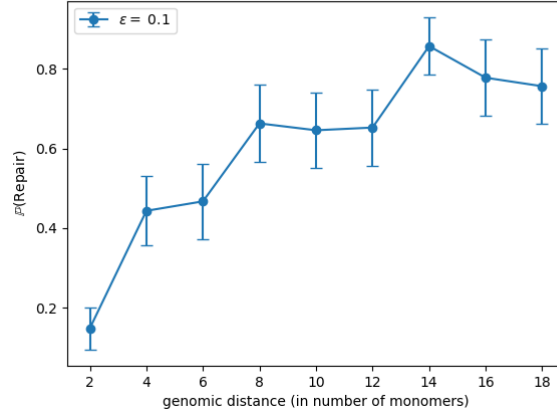


Figure 21: Probability and 95% confidence interval of repair, in function of the interbreak distance for two DSBs induced in one sub-domain of size 100 on a polymer formed by two sub-domains (100 and 50 monomers). The subdomain had 5 RCLs.

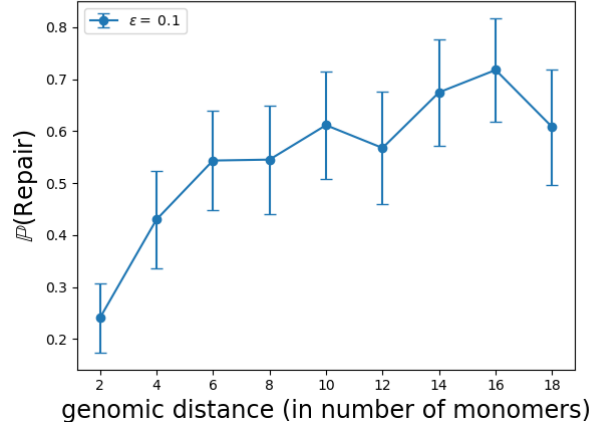


Figure 22: Probability and 95% confidence interval of repair, in function of the interbreak distance for two DSBs induced in one sub-domain of size 100 on a polymer formed by two sub-domains (100 and 50 monomers). The subdomain had 10 RCLs.



## 4.1 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?
- For two DSBs in a TAD: Simulate the repair probability vs. inter-break genomic distance (almost at the limit of the TAD). Does the probability decreases? Test with different number of random connectors (to self: simulate this first, could be interesting) (cf. figure with 2 domains and 10 random connectors per domain, the probability seems to decrease at the end, verify!)

## 5 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.

## 6 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