Combining CRISPR-Cas with RecET towards efficient genetic manipulations on Bacteria

Alicia Pereira e Calvo-Villamañán

Abstract: In the past decade, the CRISPR-Cas9 technology has greatly changed and improved the way scientists perform genetic manipulations. Nonetheless, the impact of this technology in the manipulation of Bacterial genomes has been low, mostly due to Bacteria's poor capacity to survive the double strand breaks created by Cas9.

The aim of this thesis was to develop genetic manipulation tools for Bacteria, by coupling CRISPR- Cas9 with the *Escherichia coli*'s Rac prophage's system, RecET. We hypothesized that RecET's ability to perform recombination between linear molecules would help overcome the death problem caused by Cas9, when trying to use this technology to self-target the Bacterial chromosome in order to perform genetic manipulations.

We developed three different techniques: an *in vivo,* direct cloning technique using RecET, and both a gene deletion technique, and a genome engineering technique, coupling CRISPR-Cas9 with RecET. We concluded that RecET has great potential to be used for genetic manipulations. It can perform *in vivo*, direct cloning of plasmids, with an efficiency that can be improved by performing it in deletion mutants. Combined activity of RecET and CRISPR-Cas9 can also improve the survival rate after targeting the bacterial chromosome, and can also help reducing the deletion size that emerges from this survival, hence emerging as a new gene deletion technique. Coupling of the two systems for genome editing, on the other hand, did not prove successful with the proposed strategy, and further setups will be tested in the future.

Keywords: CRISPR-Cas9, RecET, genome editing, *Escherichia coli*

Introduction: The first mention of the CRISPR array (Clustered Regularly Interspaced Short Palindromic Repeats array) in the literature dates back to 1987, when a group in the Osaka University described an unusual structure on the *Escherichia coli* chromosome consisting of five homologous sequences arranged as direct repeats and separated by flanks of 32 nucleotides[1].

Over the years that followed, the system was slowly characterised. Especially relevant discoveries include when in 2002 the first Cas proteins (CRISPR associated) were described[2, 3], when in 2005 their function was unveiled, with the arise of the hypothesis that CRISPR-Cas elements could provide for a genetic memory of previous infections and function as an adaptive immune system that could be passed on to the progeny[4, 5], and in 2009 when it was made evident that these elements were not randomly selected from the donor molecule, but in fact were selected on the base of their sequence in the donor being followed by a conserved di- or trinucleotide sequence called the PAM sequence (proto-spacer adjacent motifs)[6].

Nonetheless, the key moment for the arise of this system as a molecular biology tool happened in 2012, when the protein Cas9 from the CRISPR type II system was showed to be easily programmable[7]. The study showed that Cas9 proteins require only a dual-RNA structure, formed between two key molecules, the *trans*-activating tracr-RNA and the targeting crRNA (CRISPR-RNA), to be able to cleave the target DNA. Furthermost, they showed that this targeting could be easily reprogrammed by simply designing an crRNA molecule to cleave specific DNA sites[7].

Non-surprisingly, this study opened an unprecedented plethora of possibilities for genome targeting and genome editing, and began what is now known as the CRISPR-Cas9 *craze*.

The CRISPR-Cas9 mechanism is divided into three phases. The 1st phase is called the adaption phase. In this phase the genetic memory is created, by inserting between two direct repeats from the CRISPR-array new short sequences belonging to the invader, called spacers[8]. Spacer selection within the invader's chromosome is guided by the presence of a PAM sequence next to the target sequence[9]. The 2nd phase is called the expression phase. In this phase the entire CRISPR *locus* is transcribed into a pre-crRNA, which is later processed into individual crRNAs[10]. During this phase the tracrRNA molecules are also transcribed. The 3rd and last phase is the interference phase. In this phase the crRNA-tracrRNA duplex binds to the Cas9 protein and guides it to the target to trigger its degradation. This interference requires the presence of the PAM sequence in the target[11], which

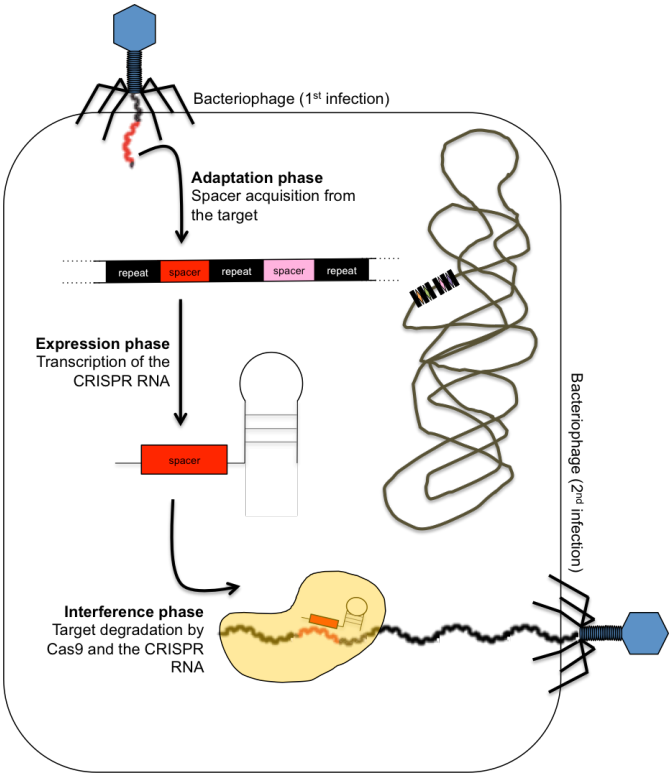

Figure 1 – Key steps of CRISPR-Cas9 immunity. 1) Adaptation phase: insertion of new spacers into the CRISPR *locus*. 2) Expression phase: transcription of the CRISPR *locus* and processing of the CRISPR RNA. 3) Interference phase: detection and degradation of mobile genetic elements by CRISPR RNA and Cas9

avoids the attack to the cell's own CRISPR *locus* by triggering what is called "non-self activation". Once the cr-RNA-tracrRNA-Cas9 complex binds to a target, Cas9 cleaves it creating a double strand break (DSB) thanks to its RuvC and HNH actives sites, which cleave the DNA (+) strand and the (-)strand respectively[12].

After the CRISPR-Cas9 mechanism was understood, and its programmable nature was described, many studies arose on its use for genome editing. Many different organisms' genomes have been edited since 2012, by coupling CRISPR-Cas9 with diverse technologies. Nonetheless, the impact of CRISPR-Cas9 on the engineering of bacterial genomes has been overwhelmingly low in comparison.

One crucial step in the use of CRISPR-Cas9 for genome editing is the capacity to rectify the DSBs created by Cas9[13]. Most bacteria rely only on homologous recombination (HR) to do so[14], and this system is not efficient enough at repairing the DSBs. This is a factor that greatly limits the use of this technology for modifying bacteria.

A few reports exist already on the use of CRISPR-Cas9 for genome editing in bacteria. Nonetheless, none of them were efficient enough for this technology to replace the traditional methods of genome editing in bacteria. This thesis aims to develop a novel genome engineering technique by coupling CRISPR-Cas9 with RecET. RecET is a recombination system from the endogenous *E. coli* Rac prophage. Although originally used in place of the Redαβ system for conventional recombineering, new interest has been placed on this viral system due to its recently described ability to perform recombination between linear DNA molecules.

By coupling CRISPR-Cas9 with RecET we aim to overcome the DSB repair problem that depending only on HR rises, increasing the bacteria's ability to survive Cas9 targeting, and therefore enhancing the efficiency of CRISPR-Cas9 as a tool for bacterial engineering.

Material and methods: Strains: During this project only *Escherichia coli* strains were used. MG1655 was used as a base for the several mutation strains. MG1656 was used as a control for the *lacZ* targeting experiments. DH5 α was used for plasmid maintenance. Mutation strains library: a library of knock-out mutants was built in this project for enhancing RecET activity. Single KO mutants were developed for the following genes: *recD, recQ, recF, recG, recJ, recX, ruvAB, ruvC, ruvABC, mutS, sfiA, xseA, xthA* and *sbcB.* Double KO mutants were developed for *recDxseA* and for *recDsbcB.* All mutants were constructed using the λ-red recombineering system. Plasmid construction *in vitro:* Plasmids throughout this thesis were constructed using Golden Gate, on a 15μL total volume assembly-reaction mix, with final working concentrations of 1X NEB T4 buffer, 1X BSA, 20U *Bsa*I and 2000U NEB T4 ligase, and 100ng of the linearised vector backbone and equimolar amounts of the other assembly pieces. The assembly reactions were performed in a thermocycler as follows: 25 cycles of 3min at 37ºC and 4min at 16ºC, one cycle of 5 min at 50ºC and one cycle of 5 min at 80ºC. Plasmids were also constructed using Gibson Assembly, in a 15μL reaction containing 1XISO Buffer, 0.64μL 10U/μL T5 exonuclease, 20μL 2U/μL Phusion polymerase, 160μL 40U/μL Taq ligase and 100ng of the linearised vector backbone and equimolar amounts of the other assembly pieces The

assembly mix was incubated in a thermocycler at 50ºC for 60min. Plasmid construction *in vivo* (developed in this thesis): For plasmid construction *in vivo,* strains carrying the RecET were cultivated until OD600nm reached 0.4-0.6. At that point, RecET expression was induced for 15-40min, depending on the number of copies of the system per cell and of the promoter controlling its expression. After induction, an electrocompetent protocol was followed, consisting of two washes with water, one wash with glycerol 10%, and cells were stored in glycerol 10% at -80ºC. The linearised fragments of the vector to construct. A 2µL linear DNA mix was electroporated into this cells, carrying between 100ng and 200ng of total DNA, and a molar ratio of 1:5 backbone:insert. The linear fragments were flanked by shared 40-50bp long homology regions, in order to be able to circularise the vector.

Results: The global goal of this thesis is to couple CRISPR-Cas9 with RecET to develop reliable and easy-to-use genome editing techniques in bacteria. In order to do so, the work was divided into three separate parts. First step was the development and enhancing of *E. coli* strains able to perform RecET driven recombination for *in vivo* cloning. Secondly, the study of the effect of RecET recombination on the size of chromosomal deletions triggered by Cas9. And third, the coupling of both systems to perform genome editing in bacteria.

Development of an *in vivo* cloning technique in *E. coli* using RecET: The first step was to obtain strains able to perform RecET recombination. We constructed the pACV2 plasmid, which carries the RecET proteins under the pTet promoter, a good choice of promoter since it is very tight and allows for a strong repression of expression unless induced. The following step was the enhancing of RecET activity. In order to do so, a small library of mutation strains was built, and to test and compare the performance of each of the mutation strains, three transformations were carried out: 1) pUC19 (10ng) was transformed in all strains for normalising all the transformations, 2) pUC19 (10ng) in a linear form was transformed to test the performance of construction of plasmids *in vivo* using RecET recombination (each of the extremities of the linear fragment shared a 50bp homology as depicted in figure 2), 3) psgRNAc (100ng) was transformed in two fragments sharing 50bp homology region, as already explained before.

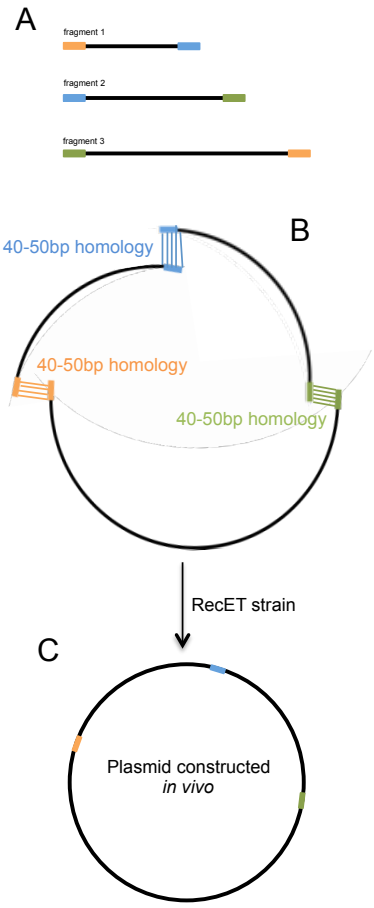

Figure 2 – RecET *in vivo* cloning strategy. A. The different fragments that will make the plasmid must be transformed into the cell in linear form. B. The fragments must share homology regions between them, so as to form a closed plasmid. C. The RecET system allows for the plasmid to be constructed *in vivo.*

A preliminary comparison was carried out between the several single KO mutation strains to determine the best performing ones. The KO mutants for *recD, sbcB, xseA, ruvABC* and *recJ* showed the best recET activity, and the experiments were repeated only on this second set of strains.

As it can be seen in figure 3, there is room for improvement of RecET activity for direct *in vivo* cloning. This conclusion was traced by comparing the wt and the control results on this figure (wt was transformed with 10ng of pUC19 in linear form, and control with 10ng of pCU10 in circular form). Since in both cases the same amount of the same plasmid was transformed, the differences in the efficiency of transformation allow us to trace the conclusion that not all the molecules of linear pUC19 that were transformed in the wt were re-circularised by RecET.

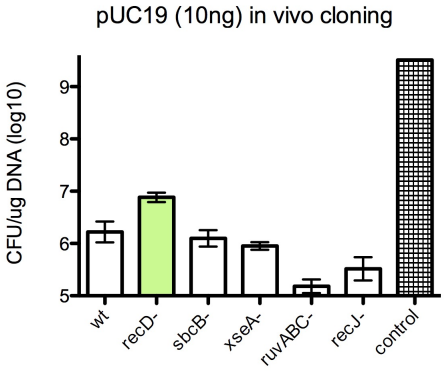

Figure 3 – Comparison between the RecET efficiency for *in vivo* cloning of pUC19 (one single fragment) in the best performing mutation strains. All strains carried the pACV2 plasmid, which has the RecET proteins under the pTet promoter. All strains were induced for 15min with 1nM aTc. Highlighted in green indicates positive significant difference with the wt. Error bars show standard deviation. Control is wt transformed with 10ng pUC10 in circular form.

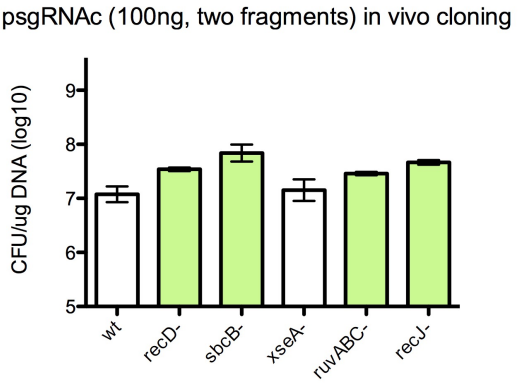

Figure 4 - Comparison between the RecET efficiency for *in vivo* cloning of psgRNAc (two fragments assembly) in the best performing mutation strains. All strains carried the pACV2 plasmid, which has the RecET proteins under the pTet promoter. All strains were induced for 15min with 1nM aTc. Highlighted in green indicates positive significant difference with the wt. Error bars show standard deviation.

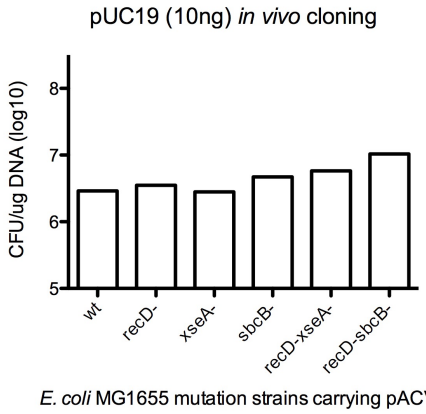

Figure 5 – Comparison between the RecET efficiency for *in vivo* cloning of pUC19 (one single fragment) in the single and double mutation strains. All strains carried the pACV2 plasmid, which has the RecET proteins under the pTet promoter. All strains were induced for 15min with 1nM aTc.

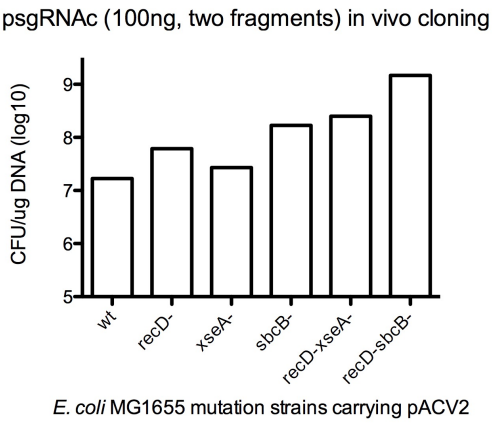

Figure 6 - Comparison between the RecET efficiency for *in vivo* cloning of psgRNAc (two fragments assembly) in the single and double performing mutation strains. All strains carried the pACV2 plasmid, which has the RecET proteins under the pTet promoter. All strains were induced for 15min with 1nM aTc.

Focusing on figures 3 and 4, testing of the mutation strains allowed us to conclude that the only mutant that consistently shows improvement of RecET activity is *ΔrecD.* Nonetheless, since there are biological differences in the way a linearised plasmid is re-circularised, and in the way two linear fragments are joined together to make a circular plasmid, we can't rule out the possibility that the other mutants might improve the RecET activity as well, depending on the circumstances. Experiments were repeated as well on the double knock-outs for *recDxseA* and *recDsbcB,* which showed a stronger improvement of the RecET system.

Development of a gene deletion technique in *E. coli* coupling CRISPR-Cas9 with RecET: The next step on the thesis workflow was to study whether RecET could be coupled with CRISPR-Cas9 to create controlled deletions in the *E. coli* chromosome. As it was already explained in the introduction, one crucial step in the use of CRISPR-Cas9 for modifying the genome is the host's capacity to rectify the DSBs that are created by Cas9. Most bacteria rely only on HR for this repair, and this factor greatly limits the use of CRISPR' Cas9 for genome editing, due to the requirement of providing a homology DNA template for the HR to occur. Due to this limitation, creating deletions on most bacterial genomes using CRISPR-Cas9 targeting is quite difficult. HR requires big homologies for its machinery to work, and since these homologies are scarce throughout the genome, most Cas9 targeting leads to cell death.

Nonetheless, it was observed that the consequences of Cas9 cleavage are not alike throughout the entire *E. coli* chromosome. A study was published on 2016 on the fact that some regions of its chromosome allow for better dealing with the DSBs created by Cas9 than others. When targeting *lacZ*, the authors observed that some cells survived the cutting by creating large deletions on the *lacZ* region. These deletions ranged from 12.9kb to 35kb. Two aspects about *lacZ* turned out to be of paramount importance for this findings; (i) first, the fact that there are no essential genes in the periphery of *lacZ* which allowed for deletions to occur in the region and (ii) second, the fact that the *lacZ* region is flanked by repetitive extragenic palindromic (REP) sequences. REP regions are found all throughout the *E. coli* chromosome and they have been implicated in a variety of functions, ranging from gene regulation to control of the DNA structure. The study showed that recombination events between REP elements were on the base of most of the large deletions observed (others were due to micro-homologies). Our aim was to couple RecET to CRISPR-Cas9 to create a system that would not depend on such big homology regions, but rather smaller ones. By doing so, we would expect the recombination events to be more common and the deletion sizes to be smaller[15, 16].

The first step was to re-program Cas9 to target the *lacZ* gene. We began by testing the efficiency of Cas9 targeting by transforming the pCas9sg::*lacZ2* (vector targeting *lacZ*) plasmid into wt *E. coli* MG1655 and MG1656, as well as the following RecET strains, wt, Δ*recD, ΔxseA*, *ΔsbcB*, *ΔrecJ*, *ΔruvABC*, *ΔrecD*Δ*xseA* and *ΔrecDΔsbcB*. The pCas9sg (empty vector) plasmid carries a *Bsa*I site for cloning spacers upstream the tracrRNA, and also Cas9, both of them constitutively expressed. pCas9sg::*lacZ2* has the *lacZ2* sgRNA cloned upstream the tracrRNA sequence.

We can observe in table 1 that, as expected, in the case of MG1655 with no induction of RecET, Cas9 targeting proves to be lethal to most of the population. In the case of MG1656 with no induction of RecET no difference was observed between the transformation with the empty vector and the vector targeting *lacZ*, results to be expected due to the fact that *E. coli* MG1656 has the *lacZ* region in the chromosome deleted. In the rest of the cases, Cas9 targeting has a clear killer effect on the population, but a higher percentage of cells survive than in the non-RecET-induced control, except on the case of the *ruvABC* mutant, in which no cells were able to survive the Cas9 targeting.

Table 1 – Rate of survival after Cas9 killing using the pCas9sg::*lacZ2* plasmid. Values in % by dividing the efficiency of transformation of pCas9sg::*lacZ2* between the efficiency of transformation of pCas9sg for each strain.

MG1656 pACV2	**MG1655 pACV2 (%)**						
wt	**wt**	**wt**	***ΔrecD***	***ΔxseA***	***ΔsbcB***	***ΔruvABC***	***ΔrecJ***
N.A.	0,211	6,000	6,667	1,714	3,182	0,000	1,600
RecET not induced	**RecET induced for 20 min**						

After understanding whether RecET helped the cells surviving Cas9 targeting, the next step on this part of the project was to study the size of the chromosomal deletion that was created throughout this survival. As already explained before, the 2016 study revealed that targeting *lacZ* in the lacZ2 position lead to deletions that ranged from 12.9kb to 35kb. We wanted to see whether RecET rescuing could lead to smaller deletions. For that, we designed pairs of primers that allowed us to cover the *lacZ* region to see the size of the potential deletion.

We screened 10 colonies for each of the mutation strains (wt, *ΔrecD, ΔxseA, ΔsbcB, ΔruvABC, ΔrecJ, ΔrecDΔxseA* and *ΔrecDΔsbcB*). This first PCR allowed us to know approximately the range of the deletion size for each of the colonies tested. After all the colonies were screened, we focused on covering examples of the different deletions present, amplified across the deletion with the appropriate pair of primers and sent the results for sequencing. Results are depicted in figure 7, which shows that the deletion size after Cas9 targeting ranges from 10.9kb to 17.5kb, a significantly improvement on the size of deletion relative to the previously mentioned 2016 study.

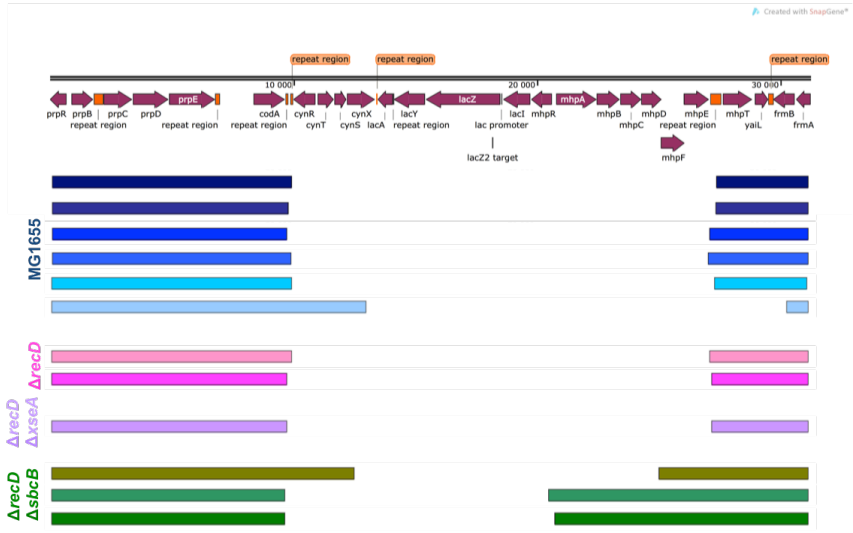

Figure 7 – Representation of the sequencing results for the analysis of the deletion size coupling RecET with Cas9 targeting the *E. coli* chromosome on *lacZ2*. In purple arrows, the different genes present in the region, with their transcription direction indicated. In orange, the various repetitive regions present in the region. *lacZ2* target position displayed in light green. The horizontal bars represent the sequencing results for strains carrying the RecET system in the pACV2 plasmid, induced for 20min.

Development of a genome editing technique in *E. coli* coupling CRISPR-Cas9 with RecET: The final and most important aim of this thesis was to develop a genome editing technique by coupling CRISPR-Cas9 with RecET. Results obtained in the direct cloning and gene deletion parts of this thesis allowed us to know that we had a good system in which cells expressing RecET were able to perform recombination between two linear DNA molecules sharing short homology regions, and also that RecET offered an advantage for cell survival after Cas9 cuts in the chromosome. Our final question was whether RecET could repair the bacterial chromosome after Cas9 targeting, by promoting recombination between the cleaved chromosome and a homology cassette. If this were the case, coupling of these two systems could result in a better and faster genome editing technique than those already available for most species of bacteria.

In order to answer that question, we designed a series of experiments that consisted on simultaneously inducing the expression of Cas9 (programming it to target the *lacZ1* or *lacZ2* position) and transforming a cassette carrying the kanamycin resistance gene flanked by 50pb-long homology regions to the immediate sites of the Cas9 cut *locus*. These experiments were carried in the RecET strains developed throughout this thesis.

First, we carried out the electrocompetence inducing protocol from the “Construction of plasmids by *in vivo* cloning using RecET” on cells carrying both pACV2 (RecET under pTet) and either pACV10 (Cas9 under pLacO-1, with a constitutive expression of the *lacZ1* sgRNA) or pACV17 (Cas9 under pLacO-1, with a constitutive expression of the *lacZ2* sgRNA).

Then, for induction of Cas9, we tested three different setups, (i) a first one in which we induced Cas9 in the last 5min before the electrocompetence inducing protocol, during the last wash with glycerol and during the recovery phase, (ii) a second one in which we induced Cas9 during the last wash of the electrocompetence inducing protocol and the recovery phase, and (iii) a last one in which we only induced its expression during the recovery phase. In all cases Cas9 was induced during the plating.

Lastly, we tested transforming the cells with different amounts of the cassette; 1000ng, 500ng, 200ng, 100ng, 50ng and 20ng.

We performed several controls, a negative control in which nor RecET or Cas9 were induced, to see if the repression of the expression of both the systems was tight; a negative control for RecET, in which RecET was not induced, to see whether RecET had any effect on the setup; and a negative control for Cas9, in which Cas9 was not induced.

In all cases, regardless of the combination of time of induction for Cas9 and amount of cassette transformed, the amount of colonies recovered in the plating was very low. More importantly, the amount of colonies recovered for each of the cassettes’ concentration was the same for the negative control of Cas9 and the tests. The cells were being able to incorporate the cassette into their genome, in the desired *locus*, but Cas9 was not playing a role on the process.

Discussion: Since the naturally existing DNA repairing mechanisms present in most bacteria are not enough to rectify the DSBs created by Cas9, the aim of this thesis was to couple RecET with the CRISPR-Cas9 system to provide an alternative and efficient DNA repair mechanism that would make it possible for bacterial cells to survive the Cas9 targeting, hence making the CRISPR-Cas9 tools that already work efficiently in other organisms, available for bacteria too.

As explained before, RecET is a recombination system from the endogenous *E. coli* Rac prophage. It was originally used in place of the Redαβ system for conventional recombineering. Nonetheless, back in 2012 a renovated interest was placed on the RecET system with the discovery that the RecE/RecT proteins were able to mediate highly efficient homologous recombination between two linear DNA molecules (linear plus linear homologous recombination, LLHR)[17]. In the same article that showed that RecET could be efficiently used for LLHR, prospects of using this system for direct, *in vivo*, cloning were immediately shown. We thought that these same prospects could be traced for using the system for repairing the Cas9 induced DSBs.

To prove this hypothesis, first we needed to gain some knowledge on the use of RecET for performing LLHR. The easiest way to do so was to replicate the already existing studies on the use of the system for direct *in vivo* cloning of plasmids. We created strains carrying the RecET system, and to improve its activity, we decided to create a library of strains carrying both RecET and certain mutations that had been associated previously with enhancing the activity of the system, or some that we thought were good candidates for improving it. Results of the *in vivo* direct cloning experiments are portrayed in figures 3 and 4.

There are many surprising things about the results we obtained. The first, most striking aspect, is the fact that the strains do not behave similarly in the two cases we tested.

Looking individually at each of the strains, we can see that deleting the exonuclease VII gene (*xseA*) had no effect on RecET's activity, and that only deleting the helicase RecD lead to enhancing of the system's activity on both types of transformation. In all the other cases, the strains behaved quite differently depending on the plasmid that was transformed. Deleting the gene for exonuclease I (*sbcB*) had little effect on RecET in the case of the linear pUC19 transformation, but enhanced almost ten times its activity in the case of psgRNAc. And on an even more despair case, deleting the genes *ruvABC* and *recJ* lead to diminishing of the system's activity in the case of the pUC19 transformation, whilst leading to a slight enhancement in the case of the psgRNAc transformation.

There does not seam to be an easy explanation for this disparity of results. The difference between the two experiments is the number of fragments that were transformed in each case. In the case of the pUC19 transformation the plasmid was linearised and transformed as a single linear DNA molecule with its extremities sharing 50bp of homology. In the case of the psgRNAc transformation, the plasmid was divided into two linear DNA fragments, and both were transformed. In this second case the extremities also shared 50bp of homology between them, but in this case RecET should recognise the existence of two individual homology regions.

In the case of pCU19, the homology regions to be recognised by RecET were placed within the same individual molecule, which was not the case for psgRNAc. This could greatly affect the probability of the homology regions being close enough from each other for RecET to match them and the recombination to occur. If this were the case, this could explain the differences observed in figures 3 and 4. Deleting the exonucleases' genes did not result in improvement in the pUC19 transformation, but always resulted on it in the psgRNAc transformation. A smaller chance of co-existing in the same place would mean a longer time before the homology regions met in space, which would mean a higher chance for exonucleases to digest the fragments that make the plasmid. Therefore, deleting exonucleases genes would improve the RecET activity by protecting the fragments from degradation before the system could match the homologies and promote recombination.

Another striking result from the data shown in figures 3 and 14 is the behaviour of the *recJ* mutant strain. RecJ had been shown to be essential for RecET's activity, and therefore we would expect this deletion mutant to show a diminished RecET activity[18]. Nonetheless, this is not what we observed. In the case of the pUC19 transformation, the mutant behaved quite closely from the wt, whilst in the case of the psgRNAc, it showed a slight enhancing.

Analysing the literature more closely, we realised that the relationship between RecET and RecJ had been traced in conditions quite different from ours. In the study in case, RecET was used to mediate recombination after UV-induced damage was inflicted to DNA. DNA broken ends from UV-damage often give rise to overhangs. For RecET to work, overhangs need to be digested and blunt ends created. Therefore, the role of the 5'-3' RecJ exonuclease was in this case essential, since although RecE is also a 5'-3' exonuclease, it only recognises blunt ends. RecJ was responsible for digesting the overhands and creating those blunt ends. Hence the diminished RecET activity when *recJ* was deleted. In our case, however, the DNA molecules that were transformed already presented blunt ends, and therefore deleting *recJ* should not have had an effect on the RecET activity. We observed, nonetheless, a slight enhancing of RecET's activity when transforming psgRNAc, which we cannot explain

Further analysis of RecET's activity on the different mutation strains needs to be performed for these questions to be answered. Special attention should be place on the double-deletion library that is currently being constructed. Preliminary results on this new library look promising (figures 5 and 6), with the activity having improved approximately 100 times on one of the double-mutants tested (the one with deletion of both *recD* and *sbcB*). Nonetheless, these results correspond to a single experiment with no replicates, and have therefore to be approached with caution.

Once we had a better grasp on how the RecET system works and we had the set of conditions in which our system worked the best, we moved forwards to the next part of the project, the development of a gene deletion technique by coupling RecET with CRISPR-Cas9.

We based the setup of our experiments on a study published in 2016 on the different consequences of Cas9 cleavage in the *E. coli* chomosome[16]. The authors reported that by targeting *lacZ* they observed the emergence of cells that carried a deletion on the *lacZ* region, with sizes ranging from 12.9kb to 35kb. They also reported that REP sequences were in the basis of the creation of most of these deletions, since recombination between these elements, that are distributed throughout all the chromosome and happen to flank the *lacZ* region, allowed for closing of the deadly DSB created by Cas9.

The results of this study suggested that Cas9 could be an interesting tool for creating deletions on the *E. coli* chromosome. Nonetheless, the system was too unreliable because it created deletions with sizes amongst too big a range. Our aim was to make this system more reliably by coupling RecET with Cas9 targeting. We wanted to know whether the deletion size would decrease when expressing RecET, whether we could narrow the range of sizes of the deletions and whether the entire setup would be as dependent on REP sequences as before, or if smaller micro-homologies would be enough for RecET rescue to work.

We tested the several RecET mutation strains on their ability to rescue the cells from Cas9 targeting of the *lacZ2* position, using the same workflow from the 2016 study, plus the previous induction of RecET in the cultures used for the experiment.

We can observe in table 1 that the different mutation strains presented different rates of killing by Cas9 targeting. In the case of MG1655 pACV2 not induced, only 0.2% of the cells survived the Cas9 targeting, proving that this is a rare event. The percentage of survival was greatly improved by expressing RecET.

Expression of RecET alone increased the chances of surviving the Cas9 targeting to 6%, and this result was enhanced by deleting the *recD* gene, which resulted in a 6.7% chance of surviving the chromosomal attack. Deleting *xseA* or *sbcB,* both exonucleases' genes, led to a decrease in the survival rate, when compared with the wt RecET expressing strain. This can be due to the fact that those two exonucleases play a role in the degradation of the chromosome, which is essential for finding the homology regions for recombination, and therefore closing of the chromosome. Deletion of *ruvABC* led to no rate of survival. As explained before, the RuvABC complex plays a role in solving the Holiday Junction, which should not affect RecET activity since there is no formation of this structure during RecET recombination. Nonetheless, RuvABC had already been shown to have an effect of unknown cause on RecET, and once again that effect is highlighted with these results. Further studies will have to be carried to deeper understand the relationship between the two systems, but until then,

the cause for this effect will remain unknown. Lastly, deletion of *recJ* resulted as well in a decrease of the rate of survival. Again, this could be explained since during the degradation of the chromosome that follows the creation of a DSB sticky ends are formed, which need to be turned into blunt ends by RecJ for RecET to work.

Summing up, regardless of the differences between the mutations strains, RecET clearly is enhancing the ability of the cells to survive Cas9 targeting. Next, we needed to understand whether the fact that the recombination effects were happening more often was due to a more efficient recombination between REP sequences, or due to the recombination between micro-homologies.

In regards to the size of the deletion, results from figure 7 are highly interesting and allowed us to trace several conclusions.

The first one is that coupling RecET to the Cas9 targeting system allows for smaller deletions than would have otherwise taken place. Deletions as small as 10.9kb were observed on the RecET strain with both *recD* and *sbcB* deleted. Going back to the results obtained in the *in vivo* direct cloning part of this thesis, figure 5 and 6 show that the double mutant for *recD* and *sbcB* had a much higher efficiency of transformation than the other strains. This means that the RecET activity was much higher in this mutant strain than in the others. Pairing that result with the results shown in figure 7 we can say that the key to getting smaller deletions lays in greatly enhancing the RecET system's activity. Further studies with the on-construction double mutant library are of great importance to confirm this conclusion, and may also be the solution for creating an effective and reliable gene knock-out system in bacteria using Cas9.

The second conclusion was that the range of the deletion size was greatly narrowed down by expressing RecET. It was narrowed down from 12.9-35kb to 10.9-17.6kb, once again proving that using RecET makes the setup more reliable. Looking at all the results except the double mutant *ΔrecDΔsbcB*'s (since it behaved better than the other strains), we can see that the size of the deletion is quite stable, between 17.0 and 17.7kb. Looking at the double mutant *ΔrecDΔsbcB* results individually, the deletion size range drops to between 10.9kb and 12.4kb. These results show that RecET can indeed make the system more reliable, and that by greatly enhancing the activity of RecET we could have a useful tool for creating deletions in the chromosome of *E. coli*.

Lastly, the last conclusion that the data shown in figure 7 allows us to draw is that almost all recombination events were due to recombination between REP elements. Only the double mutant *ΔrecDΔsbcB* showed deletions in which REP recombination did not take place. Once again, this shows the potential of using RecET for creating a gene deletion tool, since a greater activity of this system allowed for micro-homologies to be sufficient for repairing the DSB.

The gene-deletion experiments allowed us to have a better grasp on the combined activity of the CRISPR-Cas9 system and the RecET system. Nonetheless, in no moment of these experiments did the two systems have to act simultaneously, competing for access to the same region of DNA.

This, however, is the case of the experiments for the last part of this thesis, for the development of a genome editing technique by coupling CRISPR-Cas9 to RecET. The final question that we wanted to answer was whether RecET would be able to repair the bacterial chromosome after Cas9 cutting, given the providing of a recombination cassette with homologies of 50bp.

In this case, there could be a problem with the coupling of the two systems. Cas9 was reported to stay bound to the DNA after cleavage of the target[19]. This fact was then described as potentially resulting in a negative effect on DNA repair. In the case of the gene deletions experiments, this longer binding of Cas9 was not problematic, since Cas9 eventually releases itself from the chromosome and the exonucleases resume their activity. But it could pose a problem in the coupling of RecET and Cas9 for genome editing.

Since this was going to be the most sensitive part of the thesis, we designed our experiments trying to cover as many potential problems as possible. We tested it with different targets, different amounts of expression of RecET and Cas9, and with different amounts of cassette. Nonetheless, the results obtained were not satisfactory.

The lack of results made us revisit the planning, and we decided to focus on the possible incompatibility between the two systems. Going back to the study that had reported that Cas9 stays

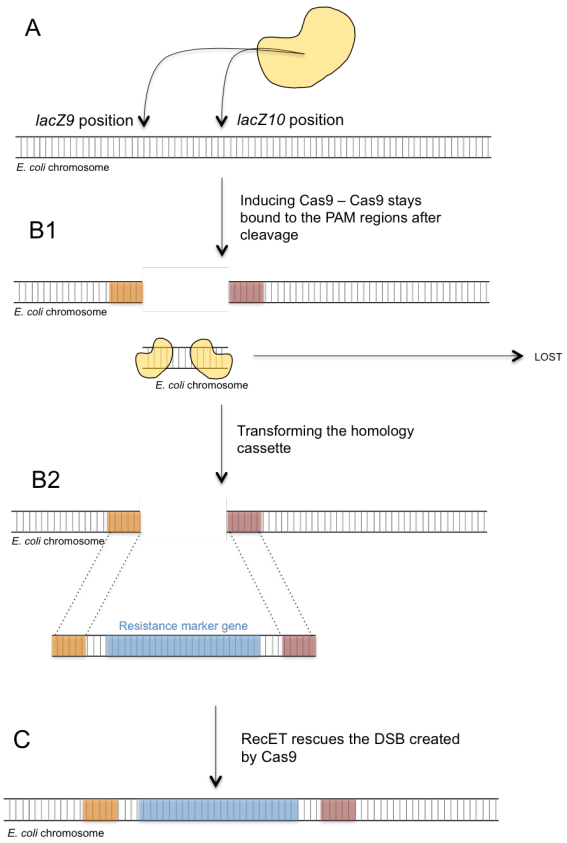

Figure 8 – Alternative scheme for the RecET rescue of the *E. coli* chromosome by promoting recombination with a cassette. A. First, Cas9 has to be programmed to target the *E. coli* chromosome twice. Keep in mind that whatever is between there two positions will be lost from the chromosome. Design of the targets should be such so that the PAM sequence of both the targets is position in the fragment that is lost from the chromosome. (B1 and B2 happen simultaneously in time, they have been divided in this figure for easier comprehension of the mechanism). B1. Cas9 induction leads to cleavage on the sites and Cas9 stays bound to the PAM regions of each of the targets. B2. The cassette carries two 50bp-long homology regions to the immediate sites surrounding the piece of the chromosome that is lost (orange and red squares). C. RecET promotes recombination between the homology regions of the cassette and the chromosome, effectively editing the genome.

bound to the chromosome after cleavage, it was reported that it has a tendency to protect the PAM regions. This means that after the cut, Cas9 is more likely to stay bound to the DNA extremity that has the PAM regions present on it.

We therefore designed a new system in which Cas9 cuts twice in the chromosome (figure 8). This double cleavage makes the setup more complicated, and hence maybe less attractive, but could solve the problem of the incompatibility between Cas9 and RecET. By cutting twice in the chromosome we end up with a piece of chromosome that will be lost after incorporation of the editing cassette. Choosing the spacers carefully so that the PAM regions are always present in the chromosome fragment that is lost, the fact that Cas9 remains attached to the chromosome after cleavage stops being a problem, hence eliminating the competition between Cas9 and RecET for accessing the homology regions.

This new system could lead to efficient genome editing of the bacterial chromosome using CRISPR-Cas9, and will be tested in the future in the continuation of this study.

Summing up everything that has been reported in this thesis, RecET emerges as an interesting mechanism to combine with the CRISPR-Cas9 system to create a versatile toolset for molecular biology.

Here we prove that strains carrying the RecET system can perform *in vivo* direct cloning. Using a strain that is able to correctively construct and produce plasmids in one single step greatly shortens the time that is needed to get a functional plasmid, and therefore is a great tool to have in a molecular biology lab.

We also prove that the RecET system is able to rescue the cells after Cas9 targeting, both by making recombination events more common and also by reducing the size of the deletion created after the rescuing. A highly enhanced activity of the RecET system is required for this setup to work efficiently.

Lastly, we open the door to a new genome editing technique in *E. coli,* by designing a system that would make it possible for Cas9 and RecET to lead to the incorporation of desired DNA sequences in the desired *locus* of the chromosome. Should this system work, it would become a tool that allows for scarless genome editing, without the need for resistance marker cassettes and double recombination events. This tool would greatly improve the efficiency of genome editing in bacteria and speed up the process as well.

Future perspectives include, first of all, the preparation of the double mutation library for further study of RecET. This library is currently under construction. Results of the already tested double mutants are promising, specially the double mutant for *recD* and *sbcB*, which makes the construction of this library of paramount importance.

Future work also includes the testing of the new system designed for genome editing, to confirm whether it would eliminate the incompatibility between Cas9 and RecET. If this system were to work and were efficient, it would be a great tool for the editing of bacterial genomes.

1. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakata, A. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *J Bacteriol* **169**, 5429-5433 (1987).
2. Jansen, R., Embden, J.D., Gaastra, W. & Schouls, L.M. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* **43**, 1565-1575 (2002).
3. Jansen, R., van Embden, J.D., Gaastra, W. & Schouls, L.M. Identification of a novel family of sequence repeats among prokaryotes. *OMICS* **6**, 23-33 (2002).
4. Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S.D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**, 2551-2561 (2005).
5. Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J. & Soria, E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* **60**, 174-182 (2005).
6. Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733-740 (2009).
7. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821 (2012).
8. Barrangou, R. et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709-1712 (2007).
9. Swarts, D.C., Mosterd, C., van Passel, M.W. & Brouns, S.J. CRISPR interference directs strand specific spacer acquisition. *PLoS One* **7**, e35888 (2012).
10. Deltcheva, E. et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602-607 (2011).
11. Semenova, E. et al. Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc Natl Acad Sci U S A* **108**, 10098-10103 (2011).
12. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A* **109**, E2579-2586 (2012).
13. Selle, K. & Barrangou, R. Harnessing CRISPR-Cas systems for bacterial genome editing. *Trends Microbiol* **23**, 225-232 (2015).
14. Shuman, S. & Glickman, M.S. Bacterial DNA repair by non-homologous end joining. *Nat Rev Microbiol* **5**, 852-861 (2007).
15. Ton-Hoang, B. et al. Structuring the bacterial genome: Y1-transposases associated with REP-BIME sequences. *Nucleic Acids Res* **40**, 3596-3609 (2012).
16. Cui, L. & Bikard, D. Consequences of Cas9 cleavage in the chromosome of Escherichia coli. *Nucleic Acids Res* **44**, 4243-4251 (2016).
17. Fu, J. et al. Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nat Biotechnol* **30**, 440-446 (2012).
18. Shiraishi, K., Hanada, K., Iwakura, Y. & Ikeda, H. Roles of RecJ, RecO, and RecR in RecET-mediated illegitimate recombination in Escherichia coli. *J Bacteriol* **184**, 4715-4721 (2002).
19. Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C. & Doudna, J.A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62-67 (2014).