NOVA SCHOOL OF SCIENCE & TECHNOLOGY

DEPARTAMENT OF LIFE SCIENCES

MARIA FILIPE GONÇALVES

BSc in Cell and Molecular Biology

ENGINEERING A CAS9-MONOMERIC STREPTAVIDIN FUSION TO INCREASE CRISPR KNOCK-IN EFFICIENCY *IN VITRO*

MASTER IN MOLECULAR GENETICS AND BIOMEDICINE

NOVA University Lisbon

November, 2021

DEPARTAMENT OF
LIFE SCIENCES

ENGINEERING A CAS9-MONOMERIC STREPTAVIDIN FUSION TO INCREASE CRISPR KNOCK-IN EFFICIENCY *IN VITRO*

MARIA FILIPE GONÇALVES

BSc in Cell and Molecular Biology

Adviser: Vasco M. Barreto, PhD
Principal Investigator, CEDOC – NOVA Medical School

Co-adviser: Rita Sobral, PhD
Assistant Professor, NOVA School of Science and Technology

Examination Committee:

Chair: Margarida Casal Ribeiro Castro Caldas Braga, PhD
Assistant Professor, NOVA School of Science and Technology

Rapporteur: Alisson Marques de Miranda Cabral Gontijo, PhD
Principal Investigator, CEDOC – NOVA Medical School

Adviser: Vasco Temudo e Melo Cabral Barreto, PhD
Principal Investigator, CEDOC – NOVA Medical School

MASTER IN MOLECULAR GENETICS AND BIOMEDICINE

NOVA University Lisbon
November, 2021

Engineering a Cas9-monomeric Streptavidin Fusion to Increase CRISPR Knock-in Efficiency *In Vitro*

ACKNOWLEDGMENTS

Gostaria de utilizar esta oportunidade para agradecer às pessoas que tornaram a elaboração desta tese possível. Primeiramente, agradeço ao meu orientador Vasco Barreto pela oportunidade de realizar este projeto, bem como pela disponibilidade e confiança que sempre demonstrou no meu trabalho. Como co-orientadora interna da faculdade, agradeço também à professora Rita Sobral pelo seu interesse e contribuição.

A todas as pessoas que fizeram parte do serviço de CRISPR, atual Genetagus, e do DNA Breaks Lab ao longo da minha estadia no CEDOC, Ângela, Catarina, João Proença, João Jorge, Miguel, Pedro, Polina, Nadiya, Gabriela e Rabia, que me aconselharam e ajudaram em tudo o que puderam, e também proporcionaram muitos momentos divertidos os quais guardo com carinho. Um agradecimento especial à Catarina que me acompanhou de perto nos primeiros meses do meu trabalho (e à distância nos últimos meses), que viveu comigo os falhanços e vitórias das experiências, sempre com a sua imensa paciência e apoio. Não poderia deixar de agradecer ao Manuel por toda a disponibilidade que teve para responder às minhas dúvidas e dar novas ideias.

Quero deixar um agradecimento muito especial aos meus pais e restante família - avós, tios e priminhos - que sempre apoiaram as minhas decisões, deram-me força, apoio emocional, muitos momentos felizes e proporcionaram todas as condições para que pudesse concluir este curso. Também à Cindy por ter estado ao meu lado durante toda a escrita desta tese e por sempre alegrar os meus dias.

Finalmente, aos meus amigos, especialmente às minhas girls, Cat, Rita, Pipa e Carol, expresso também o meu agradecimento por colocarem sempre um sorriso na minha cara, ouvirem as minhas preocupações e lutarem juntamente comigo para conseguirmos finalizar esta etapa das nossas vidas com sucesso.

ABSTRACT

The discovery of the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR), a bacterial adaptative defence system, was one of the major recent biotechnological breakthroughs, which has revolutionized gene editing in several fields. This system can be artificially manipulated to guide a Cas9 endonuclease with a single guide RNA (sgRNA) to regions of interest in the genome of a given cell. When a target sequence is recognized, Cas9 inserts a double-strand break that triggers the DNA repair of that region by either the error-prone non-homologous end joining (NHEJ) or the error-free, template-dependent, homology-directed repair (HDR) pathway. To increase the local concentration of the template DNA, skewing the repair to the HDR pathway, a Cas9 fused to a monomeric streptavidin (MSA) has been used, which was tethered to a biotinylated template DNA, thus increasing the efficiency of repair. In this thesis, two-point mutations - S14R and T39W, hypothesized to increase streptavidin's affinity to biotin - were inserted into the *MSA* gene of the Cas9-MSA-encoding plasmid. The engineered Cas9-MSA** plasmid, as well as the original Cas9-MSA and Cas9-wild-type (WT) plasmids, were used to transfect DR-GFP cells and were shown to have comparable HDR activities. Furthermore, with the addition of an exogenous template to the transfection conditions, the mutated Cas9-MSA produced the highest number of HDR-caused events with the biotinylated template, with a 1.3-fold increase of efficiency when compared to the original Cas9-MSA, and a 2.1-fold increase when compared to the traditional Cas9-WT system. These results are encouraging to explore this field of research further and improve the CRISPR-Cas9 technique to reach the required efficacy and safety for gene therapy applications.

Keywords: CRISPR-Cas9, Cas9-MSA, biotinylated DNA template, HDR, DR-GFP

RESUMO

A descoberta em bactérias do sistema de defesa adaptativo *Clustered Regularly Interspaced Short Palindromic Repeats* (CRISPR) foi um dos principais recentes avanços biotecnológicos, pois revolucionou a edição genética em diferentes campos. Este sistema pode ser artificialmente manipulado para direcionar uma endonuclease Cas9 com um RNA guia (sgRNA) para regiões de interesse no genoma de uma determinada célula. Quando uma sequência-alvo é reconhecida, a Cas9 insere uma quebra de cadeia dupla que desencadeia a reparação do DNA dessa região, através de *non-homologous end joining* (NHEJ) - propensa a erros - ou de *homology-directed repair* (HDR) - sem erros e dependente de um molde. Para aumentar a concentração local do molde de DNA, melhorarando a eficiência de reparação por HDR, foi utilizada uma Cas9 fundida a uma streptavidina monomérica (MSA), a qual foi ligada a um molde de DNA biotinilado, deste modo aumentando a eficiência da reparação. Nesta tese, duas mutações pontuais - S14R e T39W - foram inseridas no gene *MSA* do plasmídeo que codifica para a Cas9-MSA, com a hipótese de aumentar a afinidade da streptavidina para a biotina. O plasmídeo modificado Cas9-MSA**, assim como os plasmídeos de Cas9-MSA original e de Cas9-wild-type (WT), foram utilizados para transfetar células DR-GFP, e mostraram atividades de HDR semelhantes. Com a adição de um molde exógeno às condições de transfeção, a Cas9-MSA mutada produziu o maior número de eventos causados por HDR com o molde de DNA biotinilado, com um aumento de eficiência de 1.3 vezes quando comparado com a Cas9-MSA original, e de 2.1 vezes quando comparado com o sistema tradicional de Cas9-WT. Estes resultados são encorajadores para explorar esta área de investigação e melhorar a técnica de CRISPR-Cas9, para que possa possuir a eficácia e segurança necessárias para aplicações em terapia genética.

Palavras chave: CRISPR-Cas9, Cas9-MSA, molde de DNA biotinilado, HDR, DR-GFP

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1

INTRODUCTION

1.1 DNA Damage - Causes and Consequences

DNA damage is a variation in the structure of the genetic material - triggered by alterations in nucleotide base components or breaks in one or both DNA strands - and prevents the basic genetic processes from functioning correctly. If unrepaired, DNA damage can persist and consequently result in changes in nucleotide sequences within the genome, also known as mutations, deletions or rearrangements.

Several types of DNA damage may occur under the influence of either exogenous or endogenous sources. For example, in addition to intrinsic DNA instability caused by spontaneous base deamination [1], endogenous causes of mutations include replication errors that arise when the DNA polymerase's proofreading activity is not activated or because of replication slippage at repetitive sequences [2]. Moreover, endogenous sources of DNA damage include subproducts of cell metabolism such as reactive oxygen species (ROS) - which at high quantities cause oxidative stress in the cell leading to potential base oxidation - as well as single-strand breaks (SSBs) or double-strand breaks (DSBs) [1]. Notably, upon the spontaneous hydrolysis or DNA glycosylase cleavage of the *N*-glycosyl bond, abasic sites are originated, which, being unstable, are frequently removed by endonucleases, creating a gap in the double-stranded DNA (dsDNA) molecule that could produce DSBs.

Exogenous DNA damage sources include physical agents and genotoxic chemicals. The most notable physical mutagens are ultraviolet radiation (UV) - which mostly gives rise to pyrimidine dimers and commonly results in the distortion of the double helix - and ionizing radiation - which can directly affect the DNA by inducing DSBs or indirectly through the formation of reactive free radicals. Some chemical agents responsible for creating mutations are base analogs, such as 5-bromouracil, which has a high probability of being found in a form (enol) that pairs with G rather than A, resulting in point

mutations. Additionally, alkylating agents - which have a high affinity for nucleophilic base ring nitrogens, leading to direct base mispairings - and intercalating agents - flat molecules that can slot in between base pairs in the double helix, causing single-nucleotide deletions or insertions [3].

The consequences of these nucleotide sequence alterations differ greatly depending on the type of mutation and the genome region where it occurs. Many mutations located in intergenic DNA and non-coding gene sequences turn out to be silent as they do not significantly affect the functioning of the genome. With respect to mutations in coding regions, these can be synonymous when the new nucleotide produces a codon which translates into the same amino acid, non-synonymous (or missense) when there is an amino acid change in the protein sequence or nonsense if the mutation originates a termination codon, creating a premature STOP that leads to a truncated protein or activates the nonsense-mediated mRNA decay [4]. Insertions or deletions of nucleotides that are not three or a multiple of three result in a frameshift where all the codons downstream of the mutation change, generating different amino acids and consequently a new polypeptide sequence that is often also truncated [5].

1.2 DNA Damage Repair

Cells have evolved mechanisms to combat DNA damage resulting from endogenous or exogenous stressors, collectively designated as the DNA damage response (DDR). This process consists of a coordinated program that involves a network of interrelating pathways responsible for identifying lesions signalled via sensor proteins and countering particular types of damage through transducer and effector elements. These cause several changes in cell behaviour and trigger specific cascades facilitating DNA repair, cell cycle regulation, chromatin remodelling, immune-mediated elimination of damaged cells and apoptosis [6].

Upon detecting DNA damage, certain checkpoints that implement cell cycle arrest are activated to provide the required time that allows DNA repair to occur prior to cell division. If this damage is irreversible, then programmed death machineries associated with the DDR are triggered, enabling the elimination of these irreparably damaged cells. This way, DDR pathways ultimately decide cell fate, where cell survival or death is determined at the molecular level based on a threshold of pro-survival versus pro-death factors [7] [8]. Since 70 000 damage events are estimated to occur every day [9], if these DNA repair mechanisms did not exist, such alterations would disturb replication and transcription processes, causing failure in essential cell functions and, with the build-up of these mutations, the genetic information would be compromised after a small number of cell cycles.

It has been described that eukaryotic DNA polymerases exhibit error rates of approximately 10^{-4} – 10^{-5} and, when these rare replication errors occur [10], proofreading repair is the first used mechanism, reducing error rates to 10^{-6} – 10^{-7} [2]. If the wrong nucleotide is added to the 3'-end of the newly synthesized DNA strand, the exonuclease sub-unit of the polymerase removes it and proceeds with the

replication process [11]. As to post-replication errors, mismatched bases not repaired via proofreading are mended by mismatch repair (MMR) [12], as well as mutations caused by base analogs. In mammals, MutS α and MutL α recognize the mismatch, and exonuclease I (ExoI) excises the wrong nucleotides in the damaged DNA strand [13], increasing the proliferating cell nuclear antigen (PCNA) levels [14]. With the help of ssDNA-binding replication protein A (RPA) [15], replication factor C (RFC) [16], and polymerase δ [17], DNA is resynthesized and ligated with DNA ligase I.

While proofreading and MMR repair replication errors, chemical damages to the nucleotide bases are repaired through homology-dependent excision mechanisms such as base excision repair (BER) and nucleotide excision repair (NER). BER is mostly applied to non-helix-distorting lesions caused by deaminations, oxidations and alkylations. The BER pathway is initiated by recognizing the altered DNA base by a DNA glycosylase, cleaving the *N*-glycosylic bond between the base and the sugar-phosphate backbone and consequently giving rise to an abasic site, which is processed by an AP endonuclease that cleaves the phosphodiester bond, thus generating an SSB [18]. Afterward, DNA polymerase β removes the 5′-terminal deoxyribose phosphate residue [19] and adds the correct nucleotide to the nick, and XRCC1–DNA ligase III α attaches the DNA ends. On the contrary, NER is utilized in major DNA lesions, such as those induced by UV radiation and intercalating agents, which induce insertions or deletions that cause a disparity in the reading frame [20]. In the global genomic repair NER present in mammalians, the XPC - HR23B - centrin 2 heterodimer recognizes and binds DNA sites where the usual double-helix structure is damaged, followed by damage verification by XPA [21] [22]. The TFIIH complex is responsible for DNA unwinding, and endonucleases XPG and XPF-ERCC1 place incisions at the denaturation bubble's 3' and 5' ends, respectively [23] [24]. The formed gap is filled by DNA polymerase δ or ε, with the help of PCNA, and the DNA ligase seals the last nick [25].

Direct repair (DR) is the pathway activated to eliminate DNA defects induced by UV light and alkylating agents, and it does not require nucleotide excision, resynthesis, or ligation. Two of the main DR mechanisms include the reversal of pyrimidine dimers by photolyase and alkylation damage by alkyltransferase and AlkB family dioxygenases [26].

DSBs are a complex and dangerous type of DNA lesions as they can result in loss and rearrangement of genomic sequences and cell death. Nevertheless, some DSBs are produced in a programmed manner for a defined biological purpose; examples of phenomena relying on programmed DSBs include the mating-type in yeast [27], the V(D)J and class-switch recombination reactions that shape the antigen receptor repertoires [28], and meiosis in eukaryotes [29]. Furthermore, as a response to DSBs, two specific pathways can be recruited to the site of the damage: homology-directed repair (HDR) and non-homologous end-joining (NHEJ).

1.3 Homology-Directed Repair

The HDR pathway mechanism involves the exchange of identical sequences between the DNA molecule with the DSB and another undamaged homologous one, the latter being used as a repair template (**Figure 1.1**). Therefore, this is a slow high-fidelity template-dependent process, notably error-free, and is mainly activated after DNA duplication occurs in late S and G2 phases. This pathway is divided into three main stages: presynaptic, synapsis, and postsynaptic.

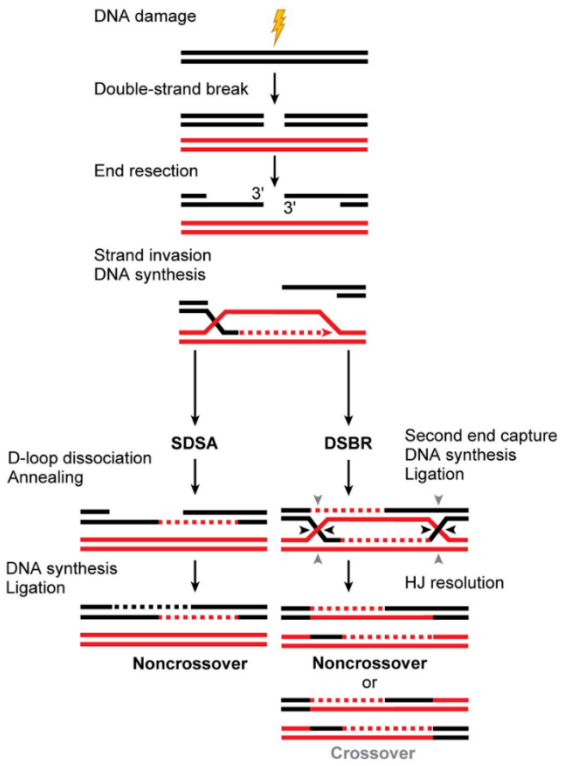

Figure 1.1 - Homology-directed repair, adapted from San Filippo *et al.* (2008) [30].

First, in the presynaptic stage, the MRN mammalian complex – consisting of meiotic recombination 11 (MRE11), radiation-sensitive 50 (RAD50), and Nijmegen breakage syndrome 1 (NBS1) proteins – binds the DNA ends in the DSB and recruits ataxia telangiectasia mutated (ATM) to the break site [31], where it becomes activated after autophosphorylation [32]. Consequently, ATM phosphorylates several DSB response proteins such as the H2AX chromatin remodelling histone and DNA nucleases C-terminal binding protein interacting protein (CtIP), Bloom syndrome protein (BLM) [33], ExoI, and MRE11 [34]. The MRN complex and CtIP are responsible for an initial resection of 50–100 nucleotides from 5' DNA ends, creating an early intermediate which is then processed by ExoI and DNA2, with the help of the BLM helicase, to form extensive regions of ssDNA [35]. These DNA tails are in turn

coated with RPA [36], and later breast cancer type 2 susceptibility protein (BRCA2) and RAD52 promote the binding of RAD51 to the 3' overhangs, displacing RPA [30].

In the synapsis stage, the RAD51 + ssDNA complex searches for homologous sequences in the double-stranded repair template and also promotes the invasion process, in which the ssDNA invades the dsDNA donor to form a joint molecule with a displaced strand – the D-loop. RAD54 stimulates homologous pairing since it stabilizes the nucleoprotein filament and later catalyses the removal of RAD51 from dsDNA, aiding in uncovering the 3' end, thus providing DNA polymerase access to initiate the repair DNA synthesis reaction [37].

In the postsynaptic stage, after RAD51 turnover, DNA polymerases δ or η elongate the strand using the donor strand as a template and replacing the nucleotides lost during end resection, with the help of the PCNA sliding clamp [38]. To resolve the recombination intermediate, several sub-pathways can be identified. In synthesis-dependent strand annealing (SDSA), after DNA synthesis, the invading strand is displaced from the homologous donor and anneals to complementary sequences in the initial DSB, exposed by cleavage of the other ssDNA overhang, hence forming noncrossover products. In the canonical DSB repair model (DSBR), a second end capture occurs as the 3′ end of the other side of the break also primes DNA synthesis, forming two Holliday junctions (dHJ). From here, the dHJ can either undergo dissolution or resolution to generate distinct intact duplex molecules. Alternatively, an early D-loop cleavage can take place, in which the extended D-loop structure is degraded by the MUS81-EME1 nuclease complex in nicked Holliday junction intermediates, promoting crossovers through the direct cleavage of the strand invasion intermediate [39].

For the dissolution of the dHJ, particularly observed in somatic cells, a joint activity of BLM helicase and a complex containing topoisomerase III alpha (TopoIII α), RecQ-mediated genome instability protein 1 (Rmi1) and 2 (Rmi2) is required, eventually leading to noncrossover products [40]. As for the resolution branch of DSBR, this process is carried out by resolvases through the nucleolytic cleavage of the dHJ and can yield a crossover - if the inner strands of one HJ and the outer strands of the other are cut - or noncrossover products – if both junctions are cut in the same plane. It is unclear which enzymes are implicated in the resolution stage and the factors involved in the choice of resolvase, particularly in eukaryotes, in which the whole HDR system is very complex [41]. After the Holliday junction's cleavage, the nicked strands are ligated to complete the process.

1.4 Non-Homologous End Joining

The NHEJ pathway ligates any DSB ends without exchanging genetic material with a homologous DNA molecule. Hence, although this is a rapid, highly efficient process, it is also highly error-prone because insertions and/or deletions can result from its action. Furthermore, unlike HDR, the

specific genes for this repair mechanism are expressed throughout all cell cycle phases, leading to a continuously activated state.

Regarding the canonical NHEJ (c-NHEJ) system (**Figure 1.2**), once a DSB occurs, the Ku heterodimer – composed of Ku70 and Ku80 proteins – binds to both DSB ends and acts as a "recruiting hub," so other NHEJ proteins can be located to the damage site to promote the joining of DNA ends [42]. DNA-dependent protein kinase catalytic subunit (DNA-PKcs) binds Ku-attached DNA ends, forming a complex that prevents end resection since it inhibits ExoI and DNA2 processing [43], and aids in bringing the broken ends together [44]. Subsequently, Ku slides internally, leading to the contact of DNA-PKcs with DNA [45], and its serine/threonine kinase activity is activated [46], phosphorylating itself as well as ARTEMIS. Once activated, ARTEMIS's endo [47] and exonucleolytic [48] activities are utilized to process the DNA ends, followed by strand synthesis by DNA polymerase μ or λ [49]. Finally, to seal the broken processed ends, ligation happens through Ligase IV [50]. It is known that X-ray repair cross-complementing protein 4 (XRCC4) promotes adenylation and helps DNA Ligase IV to interact with DNA [51], with its action being further promoted by XLF (XRCC4-like factor) [52].

Even when c-NHEJ is disabled, residual levels of NHEJ are still observed. This is due to alternative NHEJ (alt-NHEJ) action, which is initiated in the absence of the Ku complex by binding the MRN complex to the broken DNA ends. After 5' to 3' end resection, ssDNA overhangs are generated, with micro-homologous sequences between them that can then anneal to each other. The sequences flanking the original DSB that have not been aligned are removed, often resulting in dsDNA repair products containing a micro-homology region and deletions. Little is known about alt-NHEJ involved proteins and their detailed mechanism, but Poly-ADP-ribose polymerase 1 (PARP1) has been thought to have a role reminiscent of the one from DNA-PKcs in c-NHEJ [53]. Since alt-NHEJ is DNA ligase IV-independent, the enzymes that catalyse strand ligation in mammalian cells are DNA ligase I and III [54].

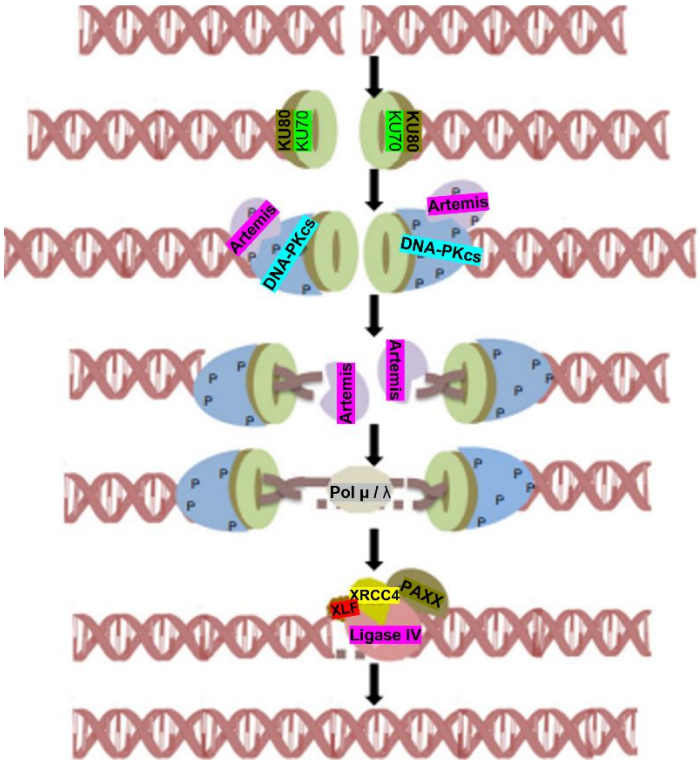

Figure 1.2 - Canonical non-homologous end joining, adapted from Sharma *et al.* (2016) [55].

1.5 CRISPR-Cas System

The Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) – CRISPR associated (Cas) system is a prokaryote adaptative immune system that allows the acquisition of resistance to external genetic elements obtained from invading foreign DNA of phages and plasmids. Most Eubacteria and most characterized Archaea capture snippets of the invading nucleic acids and use them to create DNA sequences named CRISPR arrays that allow storing memory in host chromosomes. The CRISPR locus is essentially a region of the prokaryote's DNA with palindromic repeats, intercalated with these small captured DNA sequences (spacers) followed by *Cas* genes. Thus, in bacteria, under a new attack by the same phage, short RNA sequences are produced from the CRISPR arrays, which target the foreign genetic material, and a Cas or similar enzyme is sent to degrade the hostile genetic material, disabling the virus.

With the insight into how the CRISPR-Cas system and its components function, this process could be artificially manipulated to guide endonucleases with short RNA-designed sequences to target regions of interest in a genome. Hence, there has been a revolution in genomic editing in numerous fields, from therapeutics and infectious agents to agriculture and food industries.

1.5.1 The History of CRISPR's Discovery

The first observation of the CRISPR locus was made in 1987 by Ishino *et al.* when, upon the sequencing of the *Escherichia coli iap* gene, five highly-homologous sequences of 29 nucleotides arranged as direct repeats intercalated with nonrepeating spacer 32 nucleotide sequences were discovered in the 3'-end flanking region of the gene [56]. In 1993, repetitions of 30-34 nucleotide sequences with short inverted repeats, spaced by 35–39 nucleotide sequences, were also observed in *Haloferax mediterranei* [57]. Subsequently, in 2000, Mojica *et al.* found these repeated spaced elements in 20 different microorganisms and referred to them as "short regularly interspaced repeats" (SRSRs) in the first report showing that these sequences are widespread among different phylogenetic groups [58]. Finally, at Mojica's suggestion, the term CRISPRs was coined by Jansen *et al.* in 2002, who identified several CRISPR-associated (*Cas*) genes upstream of the CRISPR loci – *Cas1*, *Cas2*, *Cas3* (with motifs characteristic of helicases of the superfamily 2) and *Cas4* (with homologies to the RecB exonucleases family) [59]. These findings led three groups to research and postulate the function of CRISPRs independently, uncovering that CRISPR loci have a role in adaptative immunity, since they noticed that spacer sequences in CRISPRs matched those of bacteriophage genomes and conjugative plasmids [60] [61] [62].

Pourcel *et al.* analysed the CRISPR locus of *Yersinia pestis* strains and discovered that the most recently acquired spacers had a homologous sequence at another locus in the genome, inside an inactive prophage. They postulated that CRISPRs could take up pieces of foreign DNA as part of a defence mechanism, thus representing a memory of past infections [60]. Bolotin *et al.* found that the phage sensitivity of *Streptococcus thermophilus* strains was negatively correlated with the number of spacers in the CRISPR arrays carried by the strain, and they also speculated that CRISPRs provided the cell immunity against phage infection by spacer-encoded anti-sense RNA [61]. In 2005, Mojica *et al.* identified 88 spacers in different microorganisms and concluded that spacer CRISPR sequences derived from preexisting bacteriophages and conjugative plasmids. Moreover, they discovered that these extrachromosomal elements could not infect the respective spacer-carrier strain, implying the existence of CRISPR-based immunity [62], which was first experimentally confirmed by Barrangou *et al.* in 2007. They performed infection experiments in a phage-sensitive wild-type *S. thermophilus* strain using virulent bacteriophages and analysed the CRISPR loci of the generated phage-resistant mutants, finding that additional spacer sequences had been inserted. It was further verified that, after the deletion of these spacer sequences, the previously phage-resistant strains became phage-sensitive. Additionally, they proved that the inactivation of the *Cas9* gene (referred to as *Cas5*, at the time) led to a loss of phage resistance, proving that this protein, which was already thought to be a nuclease because of its HNH-type nuclease motif, is necessary for immunity. In contrast, the inactivation of *Cas7* did not alter phage resistance, and this protein was shown to be involved in the insertion of new spacers, as the Cas7 knockout strain was unable to generate phage-resistant bacteria [63].

In 2008, Brouns *et al.* noted that, after CRISPR transcription, a set of Cas proteins were responsible for cleaving precursor CRISPR RNAs (pre-crRNAs), originating mature crRNAs that then act as small guide RNAs required for interfering with virus proliferation [64]. In the same year, a strain of *Staphylococcus epidermidis* was found to contain a CRISPR spacer homologous to the nickase gene present in staphylococcal conjugative plasmids, by Marrafini *et al.* They showed that CRISPR interference prevented plasmid transformation and that this interference was extremely specific, as disrupting the nickase sequence with silent mutations originated transformants and, on the contrary, the introduction of the CRISPR repeats and specific spacer in a strain lacking it, restored the interference (the bacteria became resistant to transformation by that plasmid). Moreover, by inserting a self-splicing intron sequence into the plasmid's nickase gene, they discovered that CRISPR-interference was blocked, which proved that the interference machinery targets DNA directly and not mRNA [65].

Comparative analysis of acquired spacers post-infection of different phages in *S. thermophilus* strains with the genomic sequences of the wild-type phages showed that the newly added spacer is homologous to a region in the phage genome, referred to as "proto-spacer"; moreover, a nucleotide motif was identified downstream of this region that is important for the resistance phenotype – the proto-spacer adjacent motif (PAM) [66]. Furthermore, it was clarified that there is a specific cleavage of the bacteriophage or plasmid dsDNA within the proto-spacer, three nucleotides upstream of the PAM [67].

Briefly, the organization and adaptative immunity related to CRISPR systems was understood: there is an initial stage of adaptation, where the insertion of a short sequence from the viral DNA in the CRISPR locus becomes a spacer; then, during a second stage of processing, transcription of the pre-crRNA and crRNA maturation occur; in the final stage of Cas and crRNA-mediated interference, there is the specific cleavage of the proto-spacer of the target sequence.

In 2011, Deltcheva *et al.* performed differential RNA sequencing of *Streptococcus pyogenes*, discovering trans-activating crRNAs (tracrRNAs). These tracrRNAs showed 24-nucleotide complementarity to the repeat regions of crRNA precursor transcripts and were proved to be necessary for crRNA maturation. Moreover, it was postulated that the resulting RNA duplex was recognized and site-specifically diced by RNase III in the presence of Csn1, also referred to as Cas9 [68]. Sapranauskas *et al.* were able to transfer the *S. thermophilus* CRISPR/Cas system into an *E. coli* strain, providing heterologous protection against plasmid transformation and phage infection, and establishing that Cas9 is the sole required protein for CRISPR-encoded interference due to its nuclease activity [69].

One year later, Jinek *et al.* further elucidated the mechanism for Cas9 cleavage. It was concluded that the mature crRNA is annealed to the tracrRNA forming a duplex RNA structure that directs the Cas9 protein to sites complementary to the crRNA-guide sequence, where the Cas9 HNH nuclease domain cleaves the complementary strand and the Cas9 RuvC-like domain cleaves the non-complementary strand, creating a DSB. In addition, they managed to design a single guide RNA (sgRNA) chimera

which mimicked tracrRNA:crRNA and verified that this molecule succeeded in directing sequence-specific Cas9 dsDNA cleavage [70]. These findings highlighted the potential of using this system with only two elements - sgRNA and Cas9 - for RNA-programmable gene targeting and editing, which will be further explored in section 1.6.5.

1.5.2 Classification of CRISPR-Cas Systems

Prior to 2011, the nomenclature associated with CRISPR was fragmented and did not reflect the evolutionary relationship of the systems. Only then a new classification system was proposed, based on the presence of certain Cas proteins on each of the three CRISPR types [71].

The typical type I loci contain the *Cas3* gene, which encodes a protein with distinct helicase and DNase domains for target sequence degradation, and other genes that encode proteins creating Cascade-like complexes with different compositions. Within the Cascade complex, Cas6 is the main enzyme that catalyses the processing of the long pre-crRNA transcript into a mature crRNA in type I CRISPR [72].

In type II systems, in addition to the ubiquitous Cas1 and Cas2, the large Cas9 protein alone seems sufficient for generating crRNA and cleavage of target DNA. The processing of pre-crRNA in these systems is achieved through a mechanism that includes a duplex formation between tracrRNA and part of the repeat in crRNA, and the double-stranded RNA-specific RNase III catalyses the cleavage in the presence of Cas9 [68].

Concerning type III CRISPR systems, Cas6 protein aids in the processing of pre-crRNA, which is then incorporated into Csm2/Cas10 (sub-type III-A) or Cmr5/Cas10 (sub-type III-B) complexes. These protein complexes are responsible for further crRNA maturation and later DNA or RNA cleavage.

Recently, two extra CRISPR system types have been described. Type IV, the most similar to type III-B loci, lacks *Cas1* and *Cas2* genes, and is often not closely located to a CRISPR array. This system type encodes a multisubunit crRNA–effector complex comprised of a large and sometimes a small subunit composed of the proteins Csf1, Cas5, and Cas7. Type V systems are characterized by the Cpf1 interference and adaptor modes and contain the *Cas1* and *Cas2* genes. This system is located adjacently to CRISPR arrays and is most comparable to type II systems since Cpf1 is similar to Cas9 as it contains a RuvC nuclease domain but, unlike Cas9, Cpf1 has a smaller size, produces a staggered DSB [73], and lacks an HNH nuclease domain [74].

Because Cas9 is the sole effector protein in type II CRISPR systems, these systems are ideal for performing genetic manipulation experiments, which is why type II systems will be emphasized in this thesis.

1.5.3 Type II CRISPR Molecular Mechanism in Microbial Defense

As mentioned before, the molecular process behind CRISPR-mediated immunity can be divided into three different stages: adaptation, processing, and interference (**Figure 1.3**).

During the adaptation phase, post phage infection, new spacers derived from viral DNA are acquired and incorporated into the host's genomic CRISPR locus, which allows the cell to adapt and store this sequence as "memory" to facilitate the inactivation of the invader if a new infection occurs. For this to take place, first, a proto-spacer is selected, and spacer material is generated from it (pre-spacer), followed by the new spacer's insertion in the CRISPR arrays preceded by an AT-rich leader sequence and flanked by diverse *Cas* genes. Although the full process of integrating spacers has not been completely described, Cas1 and Cas2 proteins are necessary to carry out spacer acquisition [75] since they are from an integrase complex, consisting of two distal Cas1 dimers bridged by a Cas2 dimer. Next, the pre-spacer is held above the protein complex, making contact with DNA on the opposite side. Finally, integration is accomplished at each end of the CRISPR repeat, though a nucleophilic attack by the proto-spacer 3'-OH ends, with the aid of the integration host factor (IHF) [76].

In the processing phase, expression is initiated, and the pre-crRNA is transcribed as a long transcript, containing the repeats and spacer sequences, which are then cleaved into short crRNAs. For this effect, tracrRNA, another element expressed from the CRISPR operon, pairs with pre-crRNA, and the two are processed by Cas9 and RNase III to produce smaller tracrRNA:crRNA complexes [68].

During the interference phase, the invading genetic material is targeted and cleaved. In type II CRISPR systems, Cas9 binds to tracrRNA:crRNA complexes, causing a conformational change in the protein's structure that brings its two nuclease domains together and creates a groove for DNA interaction. If a sequence complementary to the crRNA sequence is found in the foreign DNA, Cas9's HNH domain cleaves it while the RuvC domain cleaves the other non-complementary strand simultaneously, with a blunt DSB being created three nucleotides upstream of the PAM site [70].

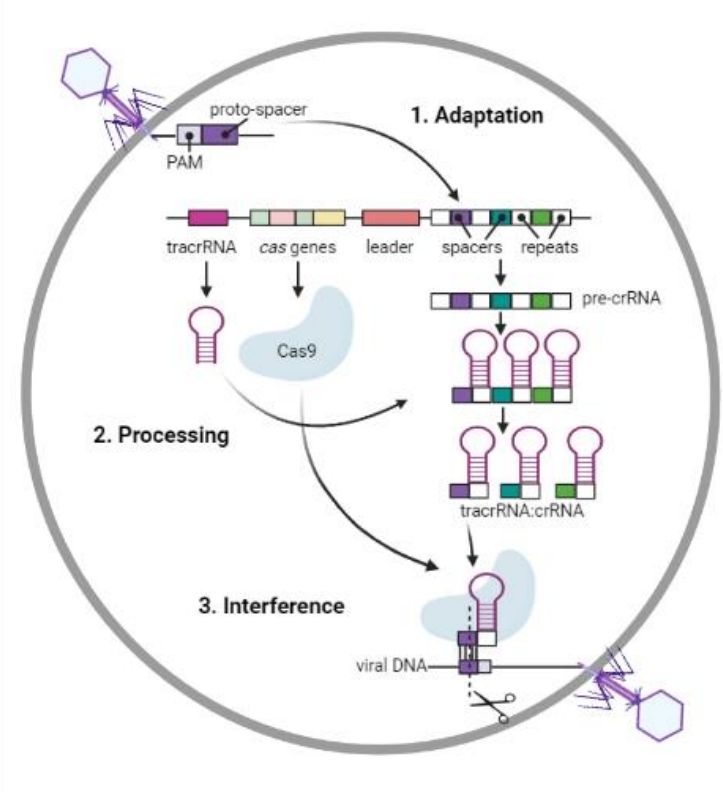

Figure 1.3 - Molecular mechanisms behind the main phases of the CRISPR type II microbial system.

1.6 Genome Editing

1.6.1 Principles of Genome Editing

As defined by Robb *et al.*, "gene editing and genome editing are techniques for genome engineering that involve DNA repair mechanisms for incorporating site-specific modifications into genomic DNA" [77]. Genome editing of eukaryotic cells is not a new field; instead, it has a history that goes back a few decades.

The first work in this field was conducted when researchers established that a DNA sequence with long regions of homology at both ends, when introduced into the cell, can be directed and integrated into a particular region in the host's genome through homologous recombination (HR) [78]. However, this method of utilizing only HR to achieve genetic modifications presented several weaknesses, such as inefficient integration of the external DNA and random DNA integration in undesired genomic locations, causing the number of effectively modified cells to be very low in each experiment. Given these limitations, a breakthrough happened when it was understood that gene targeting could be promoted by inducing DSBs at a specific genomic target in eukaryotic cells. Rouet *et al.* showed that

the ectopic expression of I-SceI yeast meganuclease in mouse embryonic stem cells triggered DSBs repaired by either NHEJ or HDR. Furthermore, it was observed that a DSB induced a 2 to 3 orders of magnitude increase in the efficiency of gene targeting, and DSBs that had been repaired without a template showed several small deletion mutations, some suggestive of repair based on regions or microhomology [79]. These scientific developments paved the way for the nuclease-assisted genome editing advances that soon followed.

Different endonuclease-based systems have been used to generate these DSBs with high-precision gene-editing techniques. All of them are based on the site-specific cleavage of DNA through nuclease activity and consequent triggering of the cell's DDR, particularly NHEJ and HDR.

As discussed previously in section 1.4., repair of DSBs via NHEJ holds a higher probability of causing insertions or deletions in the original DNA sequence when compared to HDR. Essentially, gene-editing tools are used to create programmed insertions, deletions, or replace a DNA sequence in the cell's genome with another particular one. The NHEJ pathway is often used to mediate gene knockouts by causing frameshift mutations and premature stop codons, whereas HDR introduces specific donor sequences using a donor DNA template. These options carry the possibility of generating different outcomes in gene expression, such as gene disruption, gene overexpression, repair of genetic mutations, and replacement of a missing gene.

There are four main techniques of site-specific genome editing in the subject of targeted nucleases with potential applications in model and non-model organisms, specifically the already mentioned meganucleases (MegNs), zinc finger nucleases (ZFNs), transcription activator-like effector nuclease (TALENs), and CRISPR-Cas9 (**Figure 1.4**).

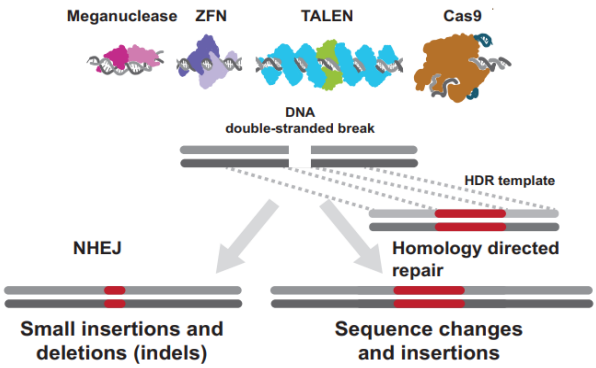

Figure 1.4 - Nucleases used in genome editing and the two possible DDR mechanisms activated along with associated repair products, adapted from Robb *et al.* (2019) [77].

1.6.2 Meganucleases

MegNs are naturally occurring endodeoxyribonucleases that can be divided into at least five families, from which the LAGLIDADG protein family is the most well-described [80]. Proteins from this family are generally encoded within self-splicing elements and show one of two main activities – they start as invasive endonucleases with the ability to mobilize their coding sequence, and, upon invasion, they obtain an RNA maturase activity to help ensure proper splicing of their intron [81].

These enzymes can cut dsDNA at specific recognition sites; hence, through the modification of their recognition sequence by protein engineering, its targeted sequence can also be altered, which implicates the modification of a significant portion of its DNA-interacting residues as well as the optimization of adjacently positioned residues on the protein scaffold [82].

Because of their very long recognition sites, MegNs are highly specific and easily delivered due to their relatively small size and low cytotoxicity. However, there is a limited number of available natural occurring MegNs, and each of them is only able to accommodate minor variations in the recognition site, which is a major disadvantage because the probability of finding a MegN with the ability to cleave a given gene at a precise location becomes tremendously low [83].

1.6.3 Zinc Finger Nucleases

ZFNs, first described by Kim *et al.*, consist of synthetic site-specific endonucleases created by linking zinc finger domains to the cleavage domain of an endonuclease, often FokI [84]. Thus, a highly specific pair of "genomic scissors" is produced by fusing the DNA-binding and DNA-cleaving domains.

The DNA-binding region contains a tandem array of Cys_2-His_2 fingers, each consisting of 30 amino acids in a conserved configuration, in which the helix binds a 3 bp DNA sequence (**Figure 1.5**) [85] [86]. In early studies, three zinc fingers were used to bind a 9 bp target. Because the nuclease domain to which they are attached functions as a dimer, a pair of ZNFs is needed to target a specific locus - one recognizes the sequence upstream, the other recognizes the sequence downstream of the targeted site [87]. Hence, ZFN dimers (the active species) can specify 18 bp of DNA per cleavage site [88]. Furthermore, more recent studies have described the addition of more zinc fingers, up to six, to specify longer cleavage DNA targets with improved specificity and efficiency.

By changing the recognition sequences of the DNA-binding domain of the ZFN, this mechanism has been used for several genome editing applications, including gene disruption – when the DSB is repaired through NHEJ – and gene correction or addition – through HDR repair with a DNA homologous template.

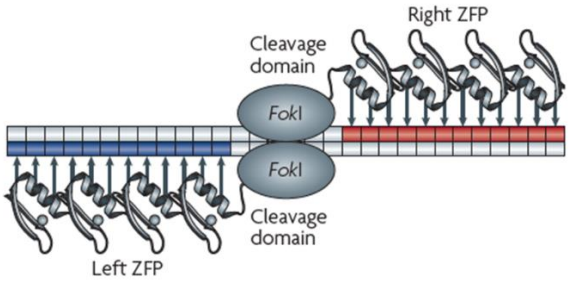

Figure 1.5 – Schematic representation of a ZFN dimer binding its target sequence, adapted from Urnov *et al*. (2010) [89].

In 2002, to disrupt a gene in *Drosophila melanogaster*, mRNAs encoding ZFNs designed for targets within that gene were injected into an early fly embryo, and up to 10% of the progeny produced by the resulting adult flies showed the desired mutation in the gene of interest [90]. ZFN driven gene disruption has also been applied in mammalian somatic cells; for example, to knockout the dihydrofolate reductase (dhfr) gene in Chinese hamster ovary (CHo) [91]. Homology-based ZFN editing entails the simultaneous supply of an appropriately designed homologous DNA molecule along with the site-specific ZFNs. For example, Urnov *et al*. inserted an extrachromosomal DNA donor with zinc-finger nucleases designed against an X-linked severe combined immune deficiency mutation in the *IL2R* γ gene, which yielded approximately 18% gene-modified human cells [92]. Furthermore, ZFNs delivered as plasmid DNA have been used in human embryonic stem (ES) and induced pluripotent stem cells (iPS) cells along with a given donor DNA molecule to efficiently target a drug resistance marker to a specific gene [93].

One of the major disadvantages of the ZFN is that the target site choice is restricted, because openly available ZFN elements can only target one site in every 200 bps of a random DNA sequence, which can be challenging in knock-in experiments [94]. In addition, ZFNs can be toxic to the cells because of the occurring off-target cleavages, which have been observed in other genomic sequences with only single nucleotide changes compared to the target sequence.

1.6.4 Transcription Activator-like Effector Nucleases

TALENs are similar to ZFNs since they include a non-specific FokI nuclease domain fused to a customizable DNA-binding domain. However, instead of zinc finger proteins, the TALENs' DNA-binding domain contains extremely conserved repeats from transcription activator-like effectors (TALEs) - proteins secreted by Xanthomonas bacteria and injected into host plant cells to bind to the genomic DNA and alter transcription in these cells, thereby facilitating the pathogenic bacterial colonization [95]. DNA binding by TALEs is mediated by a domain of arrays of highly conserved 33–35 amino acid repeats; in each repeat, two hypervariable amino acid residues – found at positions 12 and 13 – recognize one base pair in the target DNA (**Figure 1.6**).

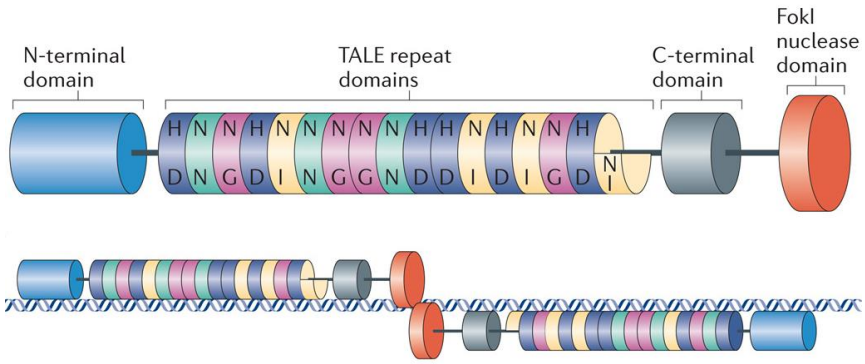

Figure 1.6 - Schematic representation of a TALEN - in which letters inside TALE repeats denote the two hypervariable residues - and how DNA target binding occurs, adapted from Joung *et al.* (2013) [96].

Through NHEJ repair of the generated DSB, Carlson *et al.* performed cytoplasmic injections of TALEN mRNAs into livestock zygotes capable of inducing LDL receptor gene knockout in up to 75% of embryos analysed, thereby generating a model for familial hypercholesterolemia [97]. TALENs have also been used to introduce specific insertions in human ES and IPS cells by Hockemeyer *et al.*, who engineered TALENs for five distinct genomic loci and used double-stranded donor templates for HDR repair, achieving high targeting efficiencies [98].

A major advantage in this genomic editing method is that, while each zinc finger recognizes three nucleotides, each TALE motif recognizes a single nucleotide, thus allowing TALENs to target specific sites with fewer constraints [99]. Moreover, unlike zinc fingers, DNA recognition by individual TALE modules appears to be mostly independent of neighboring modules but, when compared to ZFNs, one of the most noticeable limitations is their large size and consequent impending delivery [100].

1.6.5 CRISPR-Cas9

After CRISPR was discovered as an adaptative immune system in prokaryotes and its main components and molecular mechanisms were described, as explained in sections 1.5.1 and 1.5.3 of this introduction, Jinek *et al.* engineered a single guide RNA (sgRNA) which mimicked tracrRNA:crRNA and observed that this molecule directed sequence-specific Cas9 dsDNA cleavage, demonstrating that gene targeting can be manipulated *in vitro* [70]. This was a major step that facilitated the use of CRISPR-Cas9 as a genome editing technique, and soon after, several other groups managed to engineer the Type II bacterial CRISPR-Cas9 system to function with custom sgRNAs in human cells conducting site-specific cleavage [101] [102] [103]. Compared to the targeting efficiencies observed with TALENs or ZFNs targeting the same loci in human cells, the efficiencies achieved with CRISPR-Cas9 were comparable or

slightly higher. It was also reported that, with the simultaneous introduction of multiple sgRNAs into human cells, multiplex gene editing of targeted loci of the genome could be attained, and a Cas9-mutant nuclease termed nickase was generated to only cause SSBs, promoting HDR while minimizing NHEJ-mediated repair [104]. In terms of off-target activity, a TALEN approach to disrupt a marker gene showed minimal off-target effects in comparison to the corresponding ZFN [105]. Moreover, compared to TALENs, CRISPR-Cas9 has revealed a higher possibility of generating off-target effects for the same knockout editing project [106].

Contrasting with the ZFN and TALEN technologies, in which DNA binding is based on recognition by engineered proteins involving protein-DNA interactions, CRISPR-Cas9 depends on RNA-DNA recognition for genome editing. This offers various advantages over ZFNs and TALENs since RNA design and production for genomic targets is much easier when compared to protein engineering of nucleases. Additionally, the ability to address numerous targets simultaneously with multiple sgRNAs, easy prediction regarding off-target sites, and the low expenses associated with CRISPR-Cas9 methodology have led to its worldwide implementation in research laboratories, and countless successful applications to genome engineering have rapidly ensued.

1.6.6 Design of Experiments Using the CRISPR-Cas9 System

A protocol for the CRISPR-Cas9 system-mediated genomic editing has been established by Ran *et al*. and is briefly described below [107].

The first step is the selection of an appropriate sgRNA specific to the desired target. For the *S. pyogenes* Cas9 (SpCas9) system, the target sequence must immediately precede an NGG PAM, given that the 20-nucleotide sgRNA sequence binds the target DNA strand mediating the cleavage by Cas9's RuvC domain on the strand where the PAM is denoted (i.e., the non-target strand) – 3 nucleotides upstream of the PAM sequence – and by the HNH domain on the strand that is hybridized to the sgRNA (i.e., the target strand). The sgRNA can be designed with the aid of online CRISPR design tools that take a genomic sequence of interest and identify suitable target sites within it while considering the efficiency of the target modification and assessing cleavage at potential off-target sites [108] [109]. Once the sgRNA sequences are selected, the sgRNA must be constructed and delivered into the cell of interest. For this purpose, plasmids containing most of the sgRNA sequence and a cloning site to insert the 20-nucleotide spacer, sometimes along with the *SpCas9* gene, are available.

For experiments that require homologous recombination between the cleaved genomic DNA and an exogenous template, template design must follow certain rules. For example, targeted DNA modifications require the use of plasmid-based donor repair templates containing homology arms flanking the site of modification that are usually longer than 500 bp. This method allows substantial alterations to be generated, such as the insertion of large reporter genes like fluorescent proteins or

antibiotic resistance markers. However, in recent times, for single nucleotide substitutions or small insertions, ssDNA donor oligonucleotides have been used to contain flanking homologous sequences of approximately 40 bp on each side.

Cas9 and sgRNA-encoding plasmids and optional HDR repair templates are then transfected into cells, followed by the isolation of clonal cell lines with specific modifications, which can be achieved by isolating single cells by either Fluorescence-Activated Cell Sorting (FACS) or serial dilutions. Once the clones have been isolated, an expansion period is necessary to propagate the isolated clonal cell line. The final step in this process is functional testing/screening, and it can be done through different methods, most commonly Polymerase Chain Reaction (PCR) amplification followed by sequencing of the modified region, amplification refractory mutation system (ARMS) PCR, or via an assay based on a specific nuclease that cleaves the mismatched region of the hybrid DNA resulting from the annealing of mutated and wild-type (WT) sequences [110].

1.6.7 Applications of CRISPR-Cas9 Genome Editing

The ability to target and manipulate the genome of living organisms through CRISPR-Cas9 has been appealing to countless researchers worldwide. Genome engineering technology is reasonably applicable and shows the potential to revolutionize genome studies and transform different fields such as basic scientific research, nutrition, agriculture, and medicine.

With the continuous growth of the world's population, the risk of lack of agricultural resources increases. Therefore, the need to improve natural food production arises, which can be accomplished efficiently and time-saving using CRISPR-Cas9. This system has been used in approximately 20 crop species for improved yields, the addition of valuable traits, for example, higher nutritional values, prolonged shelf life, pest resistance, and tolerance to stress [111]. Furthermore, it has been shown that stress due to pathogenic microorganisms contributes to 16% of potential crop losses [112], and the negative regulators of disease resistance and grain development can be knocked out to achieve larger crop yields and host resistance against specific pathogens [113].

Much interest in CRISPR and other gene-editing methods revolves around their potential clinical applications in human diseases. CRISPR-Cas9 has already been applied in several genetic diseases, both *in vitro* and *in vivo*. For example, three research groups have successfully removed the mutated exon in the gene responsible for the most common type of Duchenne's muscular dystrophy in a mouse disease model, which resulted in partial rescue of muscle function and enhanced muscle force [114] [115] [116]. Moreover, Liang *et al.* further studied CRISPR-Cas9-mediated gene editing in human cells as they managed to target the cleavage of the β-globin gene in human tripronuclear zygotes, with an efficiency of 54% [117].

The first-in-human phase I clinical trial to test the safety and practicability of multiplex CRISPR-Cas9 editing was conducted to engineer T cells in three patients with refractory cancer. Researchers first

extracted T-cells from the patient's blood and subsequently used the CRISPR-Cas9 technology to delete genes that would interfere with the fight against cancer cells. These modified T-cells were infused back into the patients, and the engineered T-cells were detected nine months post-infusion [118]. CRISPR/Cas-9 for targeted transfusion-dependent β-thalassemia (TDT) therapy and sickle cell disease (SCD) has also been applied in clinical trials. Hematopoietic stem and progenitor cells obtained from healthy patients underwent electroporation with Cas9 and a sgRNA targeting the *BCL11A* erythroid-specific enhancer. One patient with TDT and another with SCD received autologous cells edited with CRISPR-Cas9 targeting the same *BCL11A* enhancer and, more than a year later, showed high levels of allelic editing in the bone marrow and blood, and sustained increases in fetal hemoglobin levels [119].

The CRISPR technology has also been used in the development of vaccines. For example, Yuan *et al*. developed a marker-free system for editing vaccina virus (VACV) vectors – used in cancer immunotherapies, oncolytic therapies, and as a vector to develop vaccines that prevent infectious diseases – by utilizing the CRISPR-Cas9 system with two repair donor vectors that targeted two separate genes, promoting reduced neurovirulence of the VACV and enhancing its antitumor immunity [120].

1.6.8 Improving CRISPR-Cas9 Specificity

Along with all the benefits and numerous applications of the CRISPR-Cas9 mechanism, it is also relevant to point out its downsides, off-target activity being a very important one. Different studies have shown that some mismatches between the guide sequence and the complementary target DNA can be tolerated – since specificity is mostly determined by the first 10-12 bp seed sequence adjacent to the PAM [121], leading to potential off-target cleavages and indel mutations across the genome [122] [123]. This can limit the utility of CRISPR-Cas9 for applications that require highly specific genome editing, particularly for therapeutic purposes in which even a low frequency of unintentional mutations may have detrimental consequences.

1.6.8.1 Engineered sgRNAs

Many efforts have been made towards finding approaches to manage better off-target activity, one of them being based on the modification of sgRNAs. Truncated sgRNAs, with shorter regions of target complementarity, have been used since they increase the Cas9-sgRNA complex binding sensitivity to mismatches, hence reducing undesired mutagenesis (by 5000-fold at some off-target sites) without sacrificing on-target genome editing efficiencies [124]. In addition, a chemical modification at specific sites of the ribose-phosphate backbone of sgRNAs has been described, which reduced off-target cleavage while preserving high on-target activity [125]. Furthermore, Kocak *et al.* have recently demonstrated that adding a secondary hairpin structure onto the 5' end of the sgRNA can improve the specificity of the Cas9-sgRNA complex by several orders of magnitude [126].

1.6.8.2 Engineered Cas9 Proteins

As an alternative to sgRNA engineering, modifications in the Cas9 protein have been performed. As mentioned, the first of them has been the creation of a Cas9 with nickase activity which, combined with a pair of sgRNAs that targeted opposing strands of the cleavage site, introducing a DSB, while extending the length of base pairings between the sgRNA and its target DNA [127]. The nuclease FokI domain has also been fused to the amino-terminal of a catalytically inactive Cas9 so that the FokI endonuclease is responsible for cleaving the DNA assisted by a pair of sgRNAs with closely located target sites [128]. Slaymaker *et al.* developed enhanced specificity SpCas9 (eSpCas9) [129], and Kleinstiver *et al*. created High Fidelity(HF)-SpCas9 [130] through substitutions in the residues of the parent proteins that established non-specific DNA contacts with the target, thus reducing off-target effects; however, it was shown that the on-target performance was also reduced in some target genes. Other engineered Cas9 variants were recently developed, which revealed improved specificity without compromising on-target activity [131] [132] [133].

1.6.8.3 Streptavidin-Cas9 Fusion

For HDR to be activated in knock-in experiments, the synchronized assembly of Cas9, a sgRNA, and a DNA donor at the target site is required. Nevertheless, the recruitment of the DNA template to a homologous sequence is a key rate-limiting step, which contributes to the low HDR efficiencies in genome editing. In 2017, Carlson-Stevermer *et al.* used Cas9 and a sgRNA containing an aptamer to bind streptavidin. In this system, a biotinylated ssDNA template – which induces a three-nucleotide switch from BFP to GFP – is recruited through biotin binding to streptavidin. Once all the components were transfected into BFP expressing cells, GFP expression was measured, and the rate of switch from BFP to GFP was 18 fold higher when compared to unmodified sgRNAs [134]. However, the introduction of an aptamer needs to be extremely specific and adjusted to each sgRNA. This means that, for each new target, an optimization step is required. In the same year, a different system was applied, which consisted of Cas9 fused to avidin, a sgRNA, and a biotin-modified ssDNA template. Using this method, the ratio of precise knock-in was shown to increase up to 3-fold compared to the traditional CRISPR-Cas9 system in mouse zygotes, with ~20% of the manipulated embryos being precisely edited via HDR [135].

Gu *et al.* in 2018 described the fusion of monomeric streptavidin (MSA) to the C-terminus of the *Cas9*-encoding region with an optimized linker [136]. Since MSA is a monomer that does not go through oligomerization, fusing the two proteins provides an efficient method to recruit the biotinylated DNA to the cleavage site while avoiding aggregation of the fusion molecule. Biotinylated dsDNA templates were designed to introduce fluorescent reporter genes at 20 different genes in mouse embryos. Despite being long inserts with up to over 4 kb, these reporter constructs were precise and efficiently inserted into the target DNA sites, although a variety of HDR editing rates were detected, depending on the gene location and accessibility [136].

1.6.9 Objectives of This Thesis

Building upon these recent discoveries and described systems, following the hypothesis that the binding of the biotinylated DNA to the Cas9-fused MSA would lead to an enrichment of donor DNA for HDR at the desired editing site, in this work, two-point mutations were introduced in the *MSA* gene that are hypothesized to potentially increase this protein's affinity to biotin (Sheldon Park [137]). This plasmid encodes the protein fusion with MSA's Serine-14 mutated to an Arginine (S14R) and Threonine-39 mutated to a Tryptophan (T39W) – Cas9-MSA** plasmid – and was tested in a reporter cell line.

The reporter termed DR-GFP (**Figure 1.7**) comprehends a GFP cassette adjacent to the promoter interrupted by an I-SceI restriction site giving rise to two in-frame premature termination codons (*SceGFP*), preventing GFP expression. Approximately 3.7 kb downstream of SceGFP is located a GFP cassette 5' and 3' truncated, which has the correct GFP sequence without the I-SceI restriction site (*iGFP*) [138]. If a sgRNA is used to guide Cas9 to the restriction site in the *SceGFP* gene, and a DNA template is provided, the DSB originated by Cas9 can be repaired by HDR through either the endogenous intact GFP cassette – acting as a donor of wild-type (WT) genetic information to the broken *SceGFP* gene – or the exogenously provided template. This editing converts the cells into GFP positive ones, with the quantity of GFP positive events observed in a flow cytometer being a measure of HDR.

An ssDNA template (GFPtemp) was designed to correct the GFP gene and contain a restriction site for HphI. This would be of interest since it would allow to digest DNA from cells transfected with different Cas9 plasmids with this restriction enzyme, as a way of discriminating the percentage of cells edited using the exogenous template DNA (cleavage by HphI would take place), relatively to those edited with the endogenous template from the *iGFP* cassette on the cell line's genome (cleavage by HphI would not be possible), which would serve as an internal control in the context of the competition between the two HDR templates.

Cas9-MSA fusions were used with and without point mutations to transfect DR-GFP U2OS cells, along with a sgRNA and biotinylated ssDNA template (GFPtemp), and HDR was measured through flow cytometry GFP-positive events and compared between conditions. Due to shipping delays, the HphI enzyme did not arrive in the time frame of this practical work; therefore, sequencing of the targeted editing region was performed for each condition as an alternative way to evaluate the favoured use of either the exogenous or endogenous DNA template.

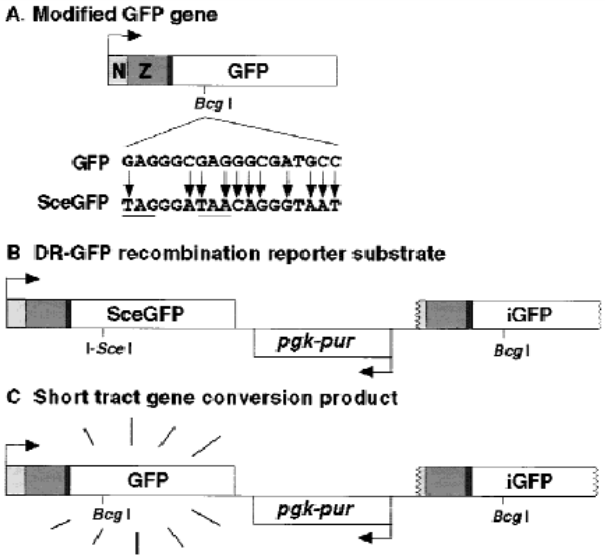

Figure 1.7 - DR-GFP reporter schematic representation, adapted from Pierce *et al.* (1999) [138].

2

MATERIALS AND METHODS

2.1 DNA Sequences and Primer Design

All primers used were ordered from IDT and designed to introduce the two amino acid-changing mutations in the *MSA* gene (S14R and T39W). Sequencing reactions were performed by Stabvida using the Sanger method, and the results were analysed with SnapGene editor software (Insightful Science; https://www.snapgene.com/). The primers encoding the sgRNA were ordered from Sigma and designed through the CCTop tool [109] to contain overhangs that allowed direct cloning in pX330 (#48137; http://n2t.net/addgene:48137 ; RRID:Addgene_48137) [107], which was done prior to this work.

The sequences of all the used primers, sgRNA, and template DNAs are presented in **Table 2.1.**

Table 2.1 - Primers, sgRNA and DNA template used, along with respective sequences. Nucleotides in bold originate the desired mutations in the MSA gene. Nucleotides in italics on the sgRNA sequence allow its direct cloning into pX330.

	Name	**Sequence**
Primers	REVRosa26-GFPinsert	TTTCCTCATTTTATTAGGAAAGG
	mStreptav-S14R-F	ACGTGGTACAACCAG**CGG**GGTTCTACCTTCACC
	mStreptav-S14R-R	GGTGAAGGTAGAACC**CCG**CTGGTTGTACCACGT
	mStreptav-T39W-F	AACCGTGCGCAGGGC**TGG**GGTTGCCAGAACTCT
	mStreptav-T39W-R	AGAGTTCTGGCAACC**CCA**GCCCTGCGCACGGTT
	Cas9-seq8	ACAACAAGCACCGGGATAAGC
	GFP5-FWext	ATGGTGAGCAAGGGCGAGG
	GFP3-RVext	TTGAAGTTCACCTTGATGCCGTTC
	Fw_Cas9mSAtmpl	CAGATGAACTTCAGG
	Rv_Cas9mSAtmpl	GTAAACGGCCACAAG
sgRNA	sgRNA11	*CACC*GGATAACAGGGTAATACCTA

DNA template	GFPtemp	CAGATGAACTTCAGGGTCAGCTTGCCA-TAGGTGGCATCGCCCTCACCCTCGCCGGACAC-GCTGAACTTGTGGCCGTTTAC

2.2 Construction of Cas9-MSA** Plasmid with S14R and T39W Mutations

2.2.1 Individual and Fusion PCRs

Three individual fragments (A, B, and C) with overlapping regions (**Figure 2.1**) were amplified from the described plasmid pX330 spCas9-MSA (#113096; http://n2t.net/addgene:113096 ; RRID:Addgene_113096) [136] by PCR. According to the manufacturer's conditions, reactions were optimized for each amplicon using the Hifi DNA polymerase (VWR, #733-2618). As annealing temperatures (Ta) differ according to the primers used in each reaction, **Table 2.2** shows the primers used for each amplicon, along with optimized Ta and respective PCR product length.

Each reaction contained 1 ng of DNA template, 0.5 µM of each primer (IDT), 0.2 mM of each dNTP (Thermo Fisher Scientific), 1X Hifi buffer (VWR), 1 U of Hifi DNA Polymerase (VWR), and Milli-Q H2O (up to 50 µL). Samples were placed in a C1000 Thermal Cycler (Bio-Rad) and subjected to the following temperature conditions: 95ºC for 2 min; 30 cycles of 95ºC for 30 s, Ta for 40 s, 72ºC for 1 min/kb; 72ºC for 5 min, 4ºC forever. Amplicons were electrophoretically separated in a 2% agarose gel and recovered using the QIAquick Gel Extraction Protocol (Qiagen) with some modifications. Briefly, after excising and weighting the DNA band from the agarose gel, three gel volumes of ADB buffer (5.5 M GuSCN, 20 mM Tris-HCl pH 6.8) were added to the gel slice. This mixture was incubated at 50ºC for 10 min. After the gel slice dissolved completely, one gel volume of isopropanol was added to the sample and mixed. To isolate the DNA, the sample was applied to the QIAquick column and centrifuged for 1 min at 16,000 rcf. The flow-through was discarded, and 750 µL of DNA Wash Buffer (80% EtOH, 10 mM NaCl, 1 mM Tris-HCl, pH 7.5) was added to the column and incubated at room temperature for 5 min, and centrifuged for 1 min at 16,000 g. Again, the flow-through was discarded, and the column was spun at 16,000 g for 3 min to dry the matrix. Next, DNA was eluted in 20 µl of Elution Buffer (pre-warmed at 60ºC, Grisp) through centrifuging for 2 min at 16,000 rcf and subsequently stored at -20ºC.

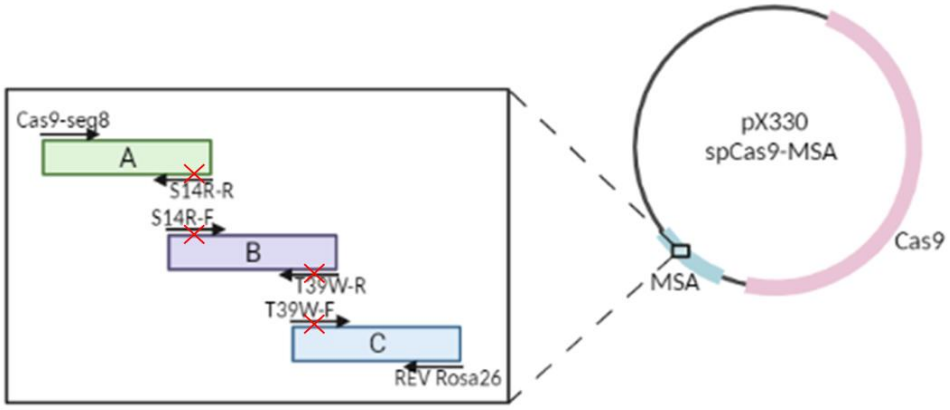

Figure 2.1 - Schematic representation of the spCas9-MSA plasmid and the fusion PCR-based method used to introduce the S14R and T39W mutations (represented as red crosses in the primers) into the MSA gene.

Table 2.2 - Primer sets used to insert the two-point mutations in the MSA gene for each targeted PCR product, optimal Ta used in the PCR program, and respective DNA fragment length originated after each PCR reaction.

PCR Product	Primer Name	Optimal Ta (ºC)	PCR Product Length (bp)
Amplicon A	Cas9-seq8	57.9	509
	mStreptav-S14R-R		
Amplicon B	mStreptav-S14R-F	65.3	108
	mStreptav-T39W-R		
Amplicon C	mStreptav-T39W-F	54.6	406
	REV Rosa26-GFPinsert		

The three purified fragments were fused using a fusion PCR with Cas9-seq8 and REV Rosa26-GFP insert primers. The PCR reaction contained the DNA volume of each amplicon according to **Table 2.3** (and Milli-Q H2O up to 15 µL), 0.2 mM of each dNTP (Thermo Fisher Scientific), 1X Hifi buffer (VWR), 1 U of Hifi DNA Polymerase (VWR) and Milli-Q H2O (up to 50 µL). Samples were placed in the C1000 Thermal Cycler (Bio-Rad) and subjected to the following temperature conditions: 95ºC for 2 min; 10 cycles of 95ºC for 30 s, 55ºC for 40 s, 72ºC for 30 s. Afterward, 0.5 µM of each primer was added, and the samples were then subjected to the following temperature conditions, in the same Thermal Cycler: 95ºC for 2 min; 25 cycles of 95ºC for 30 s, 54.6ºC for 40 s, 72ºC for 1 min; 72ºC for 5 min, 4ºC forever. Fusion PCR products were electrophoretically separated in a 1% agarose gel and recovered using the QIAquick Gel Extraction Protocol (Qiagen) as previously described for individual amplicons.

Table 2.3 - Volume of each amplicon in fusion PCR (1 and 2) and control tubes (t1, t2, t3, H2O).

Tube 1	Tube 2	Control t1	Control t2	Control t3	H2O
Amplicon A (5 µL) Amplicon B (5 µL) Amplicon C (5 µL)	Amplicon A (2 µL) Amplicon B (2 µL) Amplicon C (2 µL)	Amplicon A (5 µL)	Amplicon B (5 µL)	Amplicon C (5 µL)	No amplicon

2.2.2 Generating Cas9-MSA *DCM-* Plasmid

The purified fusion PCR product (insert) and pX330 spCas9-MSA (vector) were digested with BamHI and StuI restriction enzymes. Because StuI is blocked by overlapping DCM methylation, prior to its digestion, the vector was used to transform *E. coli* GT115 competent cells (#gt115-11, InvivoGen), a DCM-deficient strain. Briefly, 50 µL of chemically competent cells, prepared according to a previous study[139], were added to 10 µL of the Cas9-MSA plasmid and incubated for 20 minutes on ice. Then, the GT115 bacteria + DNA mixture was heat-shocked at 42ºC for 80 seconds and chilled on ice for 3 minutes. Afterward, 1 mL of SOC medium [139] was added, and the mixture was incubated for 1 hour at 37ºC with 150 rpm rotation. The samples were then centrifuged at 4ºC for 5 minutes at 1,500 rcf. Finally, the supernatant was discarded, leaving only about 50 µL, in which the pellet was resuspended. Next, the mixture was plated on LB agar with ampicillin at 100 mg/mL (Sigma) and incubated overnight at 37ºC.

2.2.3 DNA Isolation of Cas9-MSA *DCM-* Plasmid

Several colonies from *E. coli* GT115 transformation were picked to perform DNA isolation (miniprep) by isopropanol precipitation. These were grown separately in 5 mL of LB broth with ampicillin overnight at 37ºC. Afterward, cell suspensions were centrifuged at 1,910 rcf, 4ºC for 15 min, and supernatants were poured off. The pellets were then resuspended in 200 µL of P1 resuspension buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 100 µg/mL RNase) and transferred to a new Eppendorf tube. 200 µL of P2 lysis buffer (200 mM NaOH, 1% SDS) was then added, and the tube was inverted three times very carefully and left incubating for 1 min at room temperature, before adding 300 µL of P3 neutralization buffer (3 M potassium acetate pH 5.5), inverting the tubes three times very carefully and incubating at room temperature for 2 min. Finally, these mixtures were spun at 16,100 rcf for 10 min, and supernatants were transferred to new Eppendorf tubes. Subsequently, 420 µL of isopropanol was added, the tubes were inverted and centrifuged at 16,100 rcf for 10 min. Again, supernatants were poured off, and pellets were dried at room temperature and later resuspended in 50 µL of Milli-Q H2O.

2.2.4 Insert and Vector Digestion with StuI and BamHI

Single and double digestion mixtures for the insert were prepared as follows: 0.5 µg of purified fusion PCR product, 1X 3.1 buffer (NEB), 5 U of BamHI (R6021, Promega) and/or 5 U of StuI (R0187S, NEB), and Milli-Q H2O (up to 25 µL). The mixtures were incubated for 1 h at 37ºC.

Similarly, single and double digestion mixtures were also prepared for the vector: 1 µg of Cas9-MSA *DCM-* plasmid, 1X 3.1 buffer (NEB), 10 U of BamHI (Promega) and/or 10 U of StuI (NEB), and Milli-Q H2O (up to 50 µL). Due to the high molecular weight of the plasmid, in the double digestion case, StuI was added first, and mixtures were incubated for 2 h at 37ºC, and BamHI was only then sequentially added with a further 2 h incubation at 37ºC.

Along with single and double digestions, uncut controls for both the vector and insert were prepared, using the digestion mixtures described previously but without adding either restriction enzymes. Digestion products were analysed by gel electrophoresis in a 1% agarose gel and purified with the QIAquick Gel Extraction Protocol (Qiagen).

2.2.5 Ligation of Digested Insert and *DCM-* Vector

To carry out the vector-insert ligation, T4 ligase (EL0014, Thermo Scientific) was used, and the ligation mixture was prepared in the following way, after optimization: 27.2 ng of digested pX330 spCas9-MSA *DCM-*, 196.2 ng of digested fusion PCR product, 1X T4 Buffer (Thermo), 1 U of T4 ligase and Milli-Q H2O (up to 20 µL). In addition, a negative control was also included, using the same ligation mixture but without the insert DNA. Each mixture was divided into two Eppendorf tubes – 10 µL were incubated for 2 hours at room temperature, and the other 10 µL were incubated overnight at 4ºC, followed by heat inactivation at 65ºC for 10 min. Ligation mixtures – vector + insert, negative control (vector only), and positive control (Cas9-MSA original plasmid) – were then used to transform *E. coli* DH5 α competent cells, using the method described in section 2.2.2.

2.2.6 Colony PCR to Detect Positive Ligation Colonies

All colonies were picked and added to a colony PCR reaction, each containing: 0.5 µM of each primer – Cas9-seq8 and REVRosa26-GFPinsert (IDT) –, 0.2 mM of each dNTP (Thermo Fisher Scientific), 1X GoTaq buffer (Promega), 1 U of GoTaq DNA Polymerase (Promega) and Milli-Q H2O (up to 10 µL). The PCR protocol used was: 95ºC for 5 min; 95ºC for 30 s, 54.6ºC for 30 s and 72ºC for 1 min per kb (since the insert had 956 bp, 1 min was used in this step), for 30 cycles; a final extension of 72ºC for 7 minutes. The PCR products were electrophoretically separated in a 2% agarose gel and visualized in ChemiDoc Touch Imaging System. Four colonies were chosen amongst the positive ones – 7, 9, 19, and 22 – from which plasmid DNA was extracted through a miniprep, as described in section 2.2.3., and then sent for

Sanger sequencing. Because it was verified that the desired mutations were present in all DNA samples, a glycerol stock of the Cas9-MSA** plasmid was created and kept at -80ºC.

2.3 Asymmetric PCR for ssDNA Template Production

A protocol was developed and optimized based on the principle of asymmetric PCR to facilitate the generation of ssDNA template GFPtemp without experiments being dependent on the repeated ordering of oligo sequences. PCR mixtures contained: 190 ng of DNA from the previously ordered GFPtemp, Fw_Cas9mSAtmpl (forward), and Rv_Cas9mSAtmpl (reverse) primers in concentrations shown in **Table 2.4**, 0.2 mM of each dNTP (Thermo Fisher Scientific), 2 mM of MgCl2, 1X GoTaq buffer (Promega), 1 U of GoTaq DNA Polymerase (Promega) and Milli-Q H2O (up to 50 µL). Samples were placed in the C1000 Thermal Cycler (Bio-Rad) and subjected to the following program: 94ºC for 10 s; 200 cycles of 94ºC for 10 s, 42ºC for 10 s, 72ºC for 10 s; 72ºC for 1 min, 4ºC forever. Amplicons were analysed by gel electrophoresis in a 4% agarose gel.

Table 2.4 - Used primer concentrations in each PCR reaction. While maintaining the same forward primer concentration, several dilutions of the reverse primer were tested – from 1:10 to 1:2,000 – along with control tubes with only reverse (+) and only forward (-) primers.

Primers	-	+	1:10	1:100	1:200	1:500	1:1,000	1:2,000
Forward (µM)	2	0	2	2	2	2	2	2
Reverse (µM)	0	2	0.2	0.02	0.01	0.004	0.002	0.001

2.4 Cell Culture

DR-GFP U2OS cells were cultured in 100 x 20 mm cell culture petri dishes, with 10 mL of Dulbecco's Modified Eagle Medium (DMEM; Gibco), completed with 10% heat-inactivated Fetal Bovine Serum (Biowest) and 1% Penicillin/Streptomycin (Gibco), at 37ºC and 5% CO2.

2.5 Transfection of DR-GFP Cells

U2OS cells were transfected through electroporation. To detach cells from culture plates, the medium was removed, cells were washed with PBS 1X (50% of the initial medium volume) and incubated with trypsin 1X (Gibco) for 5 minutes at 37ºC. Complete DMEM was added to stop trypsin's action and resuspend the detached cells, followed by cell counting and centrifugation at 300 rcf for 4 minutes. Supernatants were discarded, and cell pellets were washed with PBS 1X and centrifuged again in the same conditions. Cells were resuspended in 1 M Amaxa buffer (5 mM KCl, 15 mM MgCl2, 120 mM

Na2HPO4/NaH2PO4 Ph 7.2, 50 mM Mannitol, 0.05% PEG) to a known concentration of 1 million cells per 100 µL of cell suspension. Tubes were prepared to contain 1.5 µg of Cas9-encoding plasmid (Cas9-WT, Cas9-MSA or Cas9-MSA**), 0.4 µg of RFP-encoding plasmid, 1.5 µg of sgRNA-encoding plasmid (sgRNA11) and, when applicable, 0.2 nmol of template ssDNA (GFPtemp either with or without biotinylation in the 5′ end) – **Table 2.5**; to which 100 µL of cell suspension was added and readily transferred to electroporation cuvettes. The program X-001 – with expected efficiency of 97-99% and expected viability of 80-96%, in this cell line – was used for electroporation in the Nucleofector II device (Amaxa Biosystems). After transfection, approximately 1 mL of completed DMEM (50%) + FBS (50%) was added to resuspend the cells, which were cultured in a well of a 6-well plate or divided by three wells of 24-well plates.

Table 2.5 - Plasmid composition of each condition tube to which the cell suspension was added for electroporation.

A0	**A1**	**A2**	**A3**
Cas9-WT RFP	Cas9-WT sgRNA11 RFP	Cas9-WT sgRNA11 GFPtemp RFP	Cas9-WT sgRNA11 Bio-GFPtemp RFP
B0	**B1**	**B2**	**B3**
Cas9-MSA RFP	Cas9-MSA sgRNA11 RFP	Cas9-MSA sgRNA11 GFPtemp RFP	Cas9-MSA sgRNA11 Bio-GFPtemp RFP
C0	**C1**	**C2**	**C3**
Cas9-MSA** RFP	Cas9-MSA** sgRNA11 RFP	Cas9-MSA** sgRNA11 GFPtemp RFP	Cas9-MSA** sgRNA11 Bio-GFPtemp RFP

2.6 Flow Cytometry and Genotyping Analysis

Upon 72 hours post-electroporation, cells were prepared for flow cytometry. Briefly, cells were washed with PBS 1X, trypsinized, and centrifuged at 300 rcf for 5 minutes. The supernatants were discarded, and cells were resuspended in FACS buffer (PBS with 2% FBS) – in the proportion of 300 µL of buffer to 1 million cells. In FACS Aria II, transfected cells were interrogated for RFP expression by a laser of 561 nm with fluorescence measuring through a 610/20 band-pass filter, and for GFP expression

by a laser of 488 nm with fluorescence measuring through a 530/30 band-pass filter. Data from flow cytometry was examined in the FlowJo software (Tree Satr, USA).

After sorting, cells from each transfection condition were put back in culture in 48-well plates and, when 90% confluency was reached, genomic DNA was extracted. For this, cells were washed with PBS 1X, trypsinized, and centrifuged at 300 rcf for 5 minutes. Next, cell pellets were washed with 500 µL of PBS 1X and centrifuged using the same conditions. After discarding the supernatants, pellets were resuspended in lysis Quantum Buffer (1,44x10^{-4}% SDS, 10mM Tris-HCl pH 8.3, 0.1 mg/mL Proteinase K) in the ratio of 30 µL of buffer per 10 thousand cells, and subjected to the following temperatures in the C1000 Thermal Cycler (Bio-Rad): 50ºC for 90 min; 95ºC for 15 min. DNA from cells of one triplicate of each transfection condition was used as template for a PCR which amplified the edited portion of the *SceGFP* gene, in which reactions mixtures consisted of 100 ng of DNA template, 0.5 µM of each primer - GFP5-FWext and GFP3-RVext – (IDT), 0.2 mM of each dNTP (Thermo Fisher Scientific), 1X GoTaq buffer (Promega), 1 U of GoTaq DNA Polymerase (Promega) and Milli-Q H2O (up to 50 µL). The PCR temperature program in C1000 Thermal Cycler (Bio-Rad) was: 95 ºC for 2 min; 25 cycles of 95ºC for 20 s, 53.5ºC for 20 s, 72ºC for 30 s; 72ºC for 2 min, 4ºC forever. PCR products were electrophoretically separated in a 2% agarose gel, purified using the QIAquick Gel Extraction Protocol (Qiagen), and sent for sequencing.

2.7 Statistical Analysis

Averages and standard deviation values were calculated for triplicates of each transfection condition, and two-way ANOVA tests followed by Turkey's multiple comparisons tests were performed through GraphPad Prism version 9.2.0 (GraphPad Software, San Diego, California USA, www.graphpad.com).

3

RESULTS

3.1 Insertion of Point Mutations in the *MSA* Gene

To potentially increase MSA's affinity to biotin, S14R and T39W mutations were introduced in the *MSA* gene. Fragments A, B, and C were amplified through PCR from the Cas9-MSA plasmid with primers that inserted the desired mutations, and PCR products were electrophoretically separated in an agarose gel (**Figure 3.1A**).

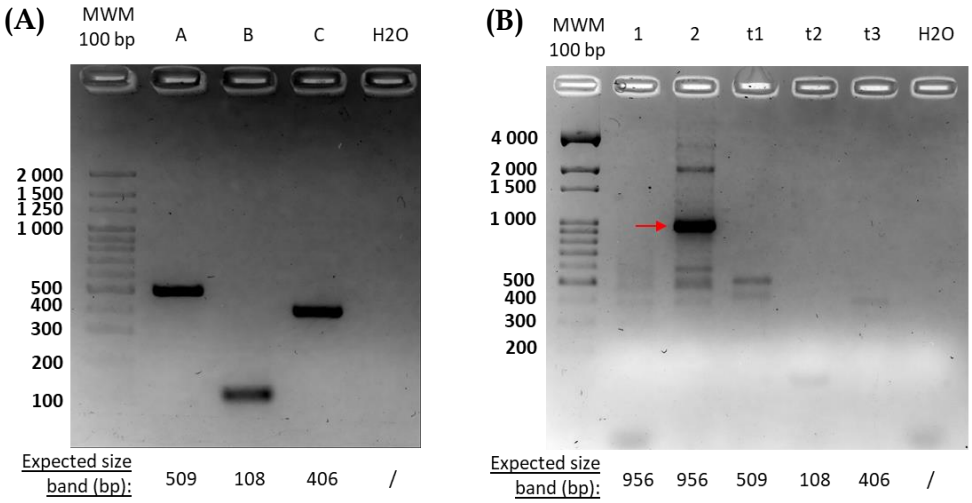

Figure 3.1- Agarose gel images of **(A)**: individual PCR products (A, B and C amplicons) and **(B)**: fusion PCR products using the most external primers (1 – 5 µL of each amplicon; 2 – 2 µL of each amplicon; t1 – A amplicon only; t2 – B amplicon only; t3 – C amplicon only; H2O – negative control, according to Table 3). 2% and 1% agarose gels were used, respectively. Gels were visualized in ChemiDoc Touch Imaging System (Bio-Rad) with GreenSafe used as nucleic acid stain (Nzytech).

Bands with expected molecular weight for each amplicon were observed and DNA from each band was extracted to perform fusion PCR using the most external primers. Fusion PCR products and controls were separated by electrophoresis (**Figure 3.1B**). No fusion PCR product was detected when 5 µL of amplicons A, B and C were used (lane 1), being only detectable when a volume of 2 µL of each amplicon was applied (lane 2). Furthermore, all control reactions originated the expected size bands (**Figure 3.1B**).

3.2 Digestion of the Fusion PCR Product and the Cas9-MSA *DCM*- Plasmid

After incubating single and double digestion mixtures of the insert (purified fusion PCR product) and the vector (Cas9-MSA *DCM*- plasmid) with BamHI and/or StuI, along with uncut controls for each, all digestion products were separated through electrophoresis as shown in **Figure 3.2**.

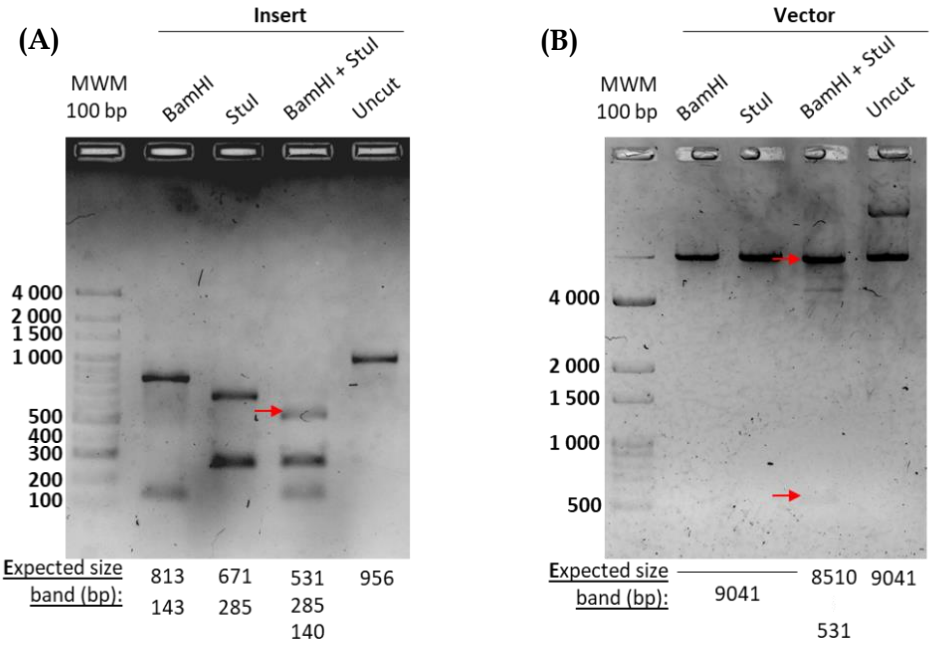

Figure 3.2 - Agarose gel images of single and double digestion products, as well as uncut controls of **(A)**: fusion PCR insert and **(B)**: Cas9-MSA plasmid. 1% agarose gels were and visualized in ChemiDoc Touch Imaging System (Bio-Rad) with GreenSafe used as nucleic acid stain (Nzytech).

Insert digestion products displayed the expected molecular weights and the double digestion 531 bp band, which corresponds to the fragment that contains S14R and T39W mutations, was purified

for subsequent ligation (**Figure 3.2A**). With respect to the vector, the double digestion resulted in the release of a 531 bp DNA fragment only when a sequential addition of StuI followed by BamHI was performed (**Figure 3.2B**). DNA from the double digestion product corresponding to the 8,510 bp band was purified for later ligation.

3.3 Ligation Colony PCR and Sequencing

Upon ligation of digested insert and vector with T4 ligase overnight at 4ºC, followed by transformation of E. coli DH5 α competent cells with ligation mixtures, all colonies grown in LB agar + ampicillin were picked and used to execute a colony PCR. Primers Cas9-seq8 and REVRosa26-GFPinsert - previously used to fuse A B and C fragments that constitute the insert – were used in the PCR mixture to evaluate the incorporation of the fusion PCR product in each colony. Colony PCR products were separated by electrophoresis (Figure 3.3) and positive colonies – the ones from which a band corresponding to the fusion product appeared – were selected.

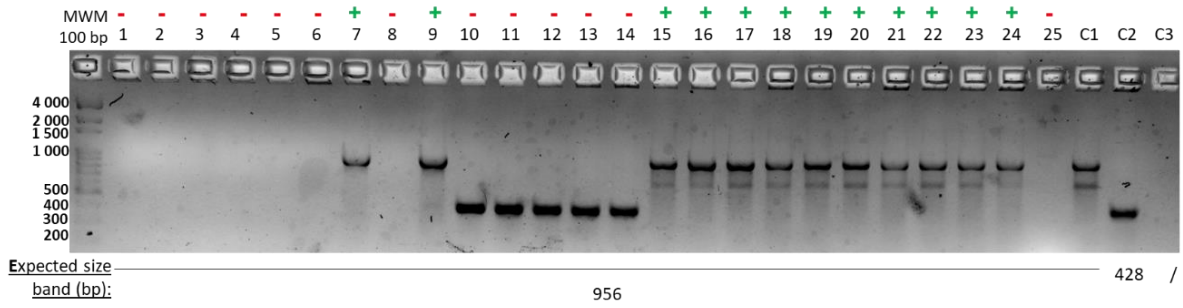

Figure 3.3- Image of 2% agarose gel from colony PCR products: 1-25 – DNA from each ligation colony; C1 – Positive control (Cas9-MSA plasmid); C2 – Vector only; C3 – Negative control. Gel visualized in ChemiDoc Touch Imaging System (Bio-Rad) with GreenSafe used as nucleic acid stain (Nzytech).

Plasmid DNA was extracted from colonies 7, 9, 19 and 22 and sent to Sanger sequencing with primer Cas9-seq8 to verify if the two desired mutations were present in the Cas9-MSA plasmid. Sequencing results aligned with the original plasmid sequence are shown in **Figure 3.4**, from which we can infer that all four colonies acquired the mutated Cas9-MSA** plasmid.

Figure 3.4 - Sequencing results from Cas9-MSA** extracted from colonies 7, 9, 19 and 22, aligned with the Cas9-MSA plasmid in the S14R and T39W mutations site.

3.4 Production of Single-stranded Donor DNA

An asymmetric PCR method was optimized to generate an ssDNA template for cell transfection. The electrophoretic separation of the PCR products on an agarose gel is shown in **Figure 3.5**. Different dilutions of the reverse primer were tested with the same concentration of forward primer since the forward GFPtemp strand was the desired one. Of the analysed dilutions, the PCR mixture with 1:500 (0.004 µM) of the Rv_Cas9mSAtmpl primer and 2 µM of Fw_Cas9mSAtmpl appeared to be the one with the best yield of the expected 80-nucleotide ssDNA template (lane 6), which corresponds to the band observed in the PCR product amplified with only the reverse primer (lane 2), with the expected molecular weight.

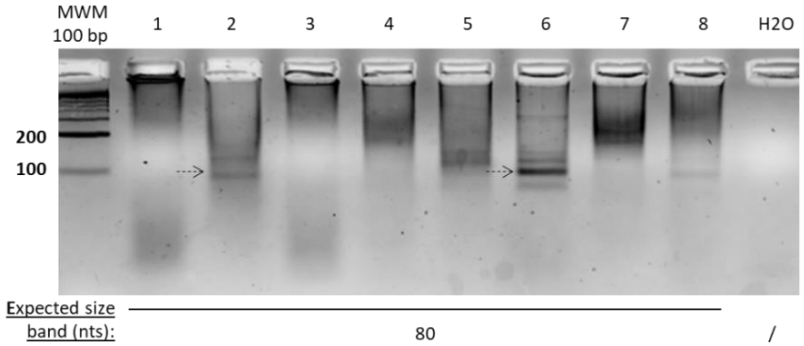

Figure 3.5 - Image of 4% agarose gel after electrophoresis of asymmetric PCR products using different concentrations of primers for production of ssDNA GFPtemp, according to Table 6: 1: 2 µM of the forward primer only; 2: 2 µM of the reverse primer only; 3-8: 2 µM of forward primer and concentration of reverse primer varying from 0.2

µM, 0.02 µM, 0.01 µM, 0.004 µM, 0.002 µM and 0.001 µM, respectively; H20 - negative control. Gel visualized in ChemiDoc Touch Imaging System (Bio-Rad) with GreenSafe used as nucleic acid stain (Nzytech).

3.5 Testing HDR Activities of Cas9-MSA and Cas9-MSA**

Cas9 fusions' ability to induce DNA cleavage and HDR-based repair was assessed in the described DR-GFP reporter. After transfection with the different Cas9 plasmids with/without the addition of sgRNA, GFP positive events were quantified through flow cytometry. A negative control of cells that underwent the same electroporation conditions in the absence of either Cas9 or sgRNA-encoding plasmids and cells transfected with only either GFP or RFP-encoding plasmids were used to define different cell populations and establish the correct gates for further experiments (**Figure 3.6**).

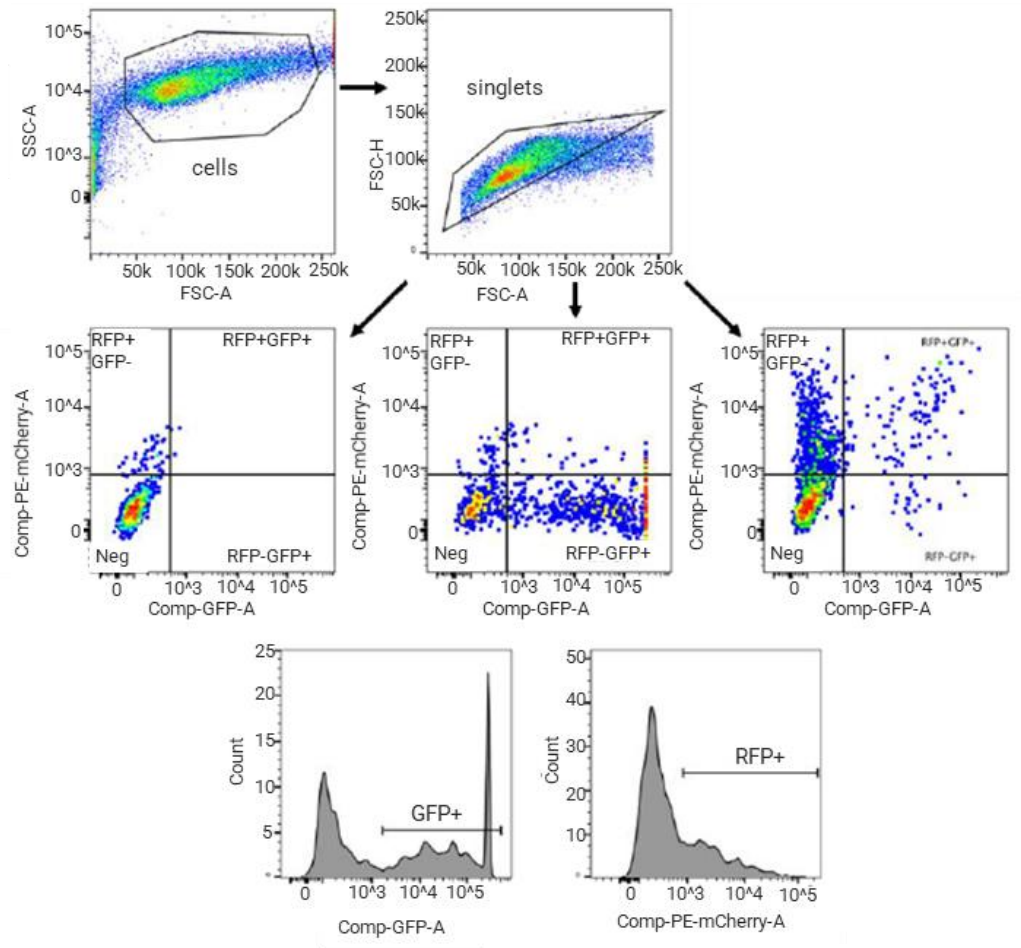

Figure 3.6 - Flow cytometry gating strategy to identify single DR-GFP cells and their GFP and/or RFP positive populations.

Because all cells, not including the controls described previously, were co-transfected with RFP, only percentages of GFP-positive events within all RFP positive cells were considered. Both Cas9-MSA and Cas9-MSA** fusions were shown to be active with no significant difference between each other's HDR activity. However, the Cas9-MSA fusion demonstrated a slight decrease in the ability to cause GFP-positive events compared to Cas9-WT (**Figure 3.7A**). Furthermore, the difference between the number of GFP-positive events when transfecting with Cas9 plasmids + sgRNA and transfecting with Cas9 plasmids only, was calculated (**Figure 3.7B**), validating the previous result.

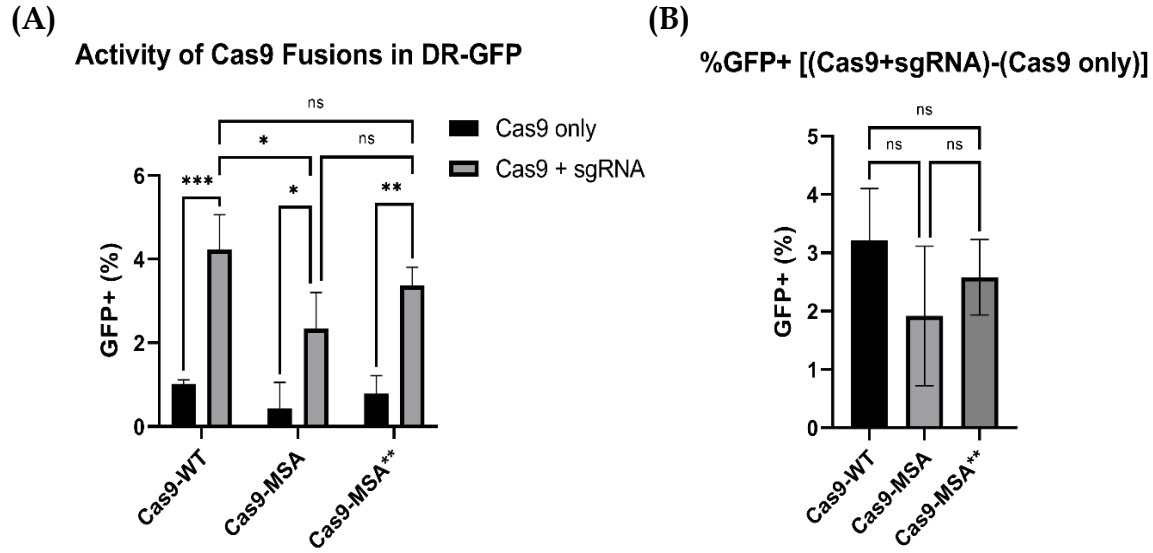

Figure 3.7- Analysis of HDR activities of Cas9-MSA fusions. **(A)**: Percentage of GFP+ events within transfected DR-GFP cells. **(B)**: Difference between GFP+ events with and without adding sgRNA. ns $p > 0.05$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$. SD is represented by the error bars (n=3).

3.6 Optimization of DNA Template Dosage in Transfection Experiments

Before adding the exogenous ssDNA templates to the conditions of the DR-GFP transfection experiments, the quantity of the used template had to be optimized. With this purpose, DR-GFP cells were transfected with Cas9-WT and sgRNA-encoding plasmids along with different amounts of GFPtemp donor DNA - 0.2 nmol, 0.4 nmol and 0.8 nmol.

Electroporated cells were analysed by flow cytometry and, although differences between closer amounts of DNA were not statistically significant, GFP-positive events were the highest in cells transfected with 0.2 nmol of GFPtemp and displayed an apparent tendency to decrease with the increase of ssDNA template used (**Figure 3.8**). Therefore, 0.2 nmol was the chosen quantity of ssDNA template used in further electroporations.

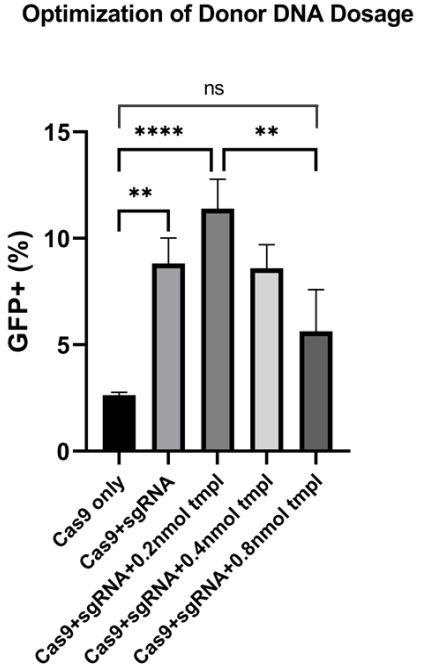

Figure 3.8 - Optimization of GFPtemp donor DNA dosage used in co-transfection of DR-GFP cells with Cas9 and sgRNA-encoding plasmids. ns $p > 0.05$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$. SD is represented by the error bars (n=3).

3.7 Assessment of Cas9 Fusions' Editing Efficiencies with Exogenous Templates

A transfection experiment was conducted in DR-GFP cells with Cas9 and sgRNA-encoding plasmids and adding the optimal amount of either GFPtemp or 5′-biotinylated GFPtemp (represented in **Figure 3.9A**). Electroporated cells (all RFP-positive cells) were sorted through flow cytometry and put back into culture, and GFP percentages were evaluated as a measure of HDR efficiency (**Figure 3.9B**).

With the addition of the non-biotinylated ssDNA template, the number of GFP-positive events upon transfection with the Cas9-MSA fusion showed no significant differences from Cas9-MSA**, but was increased when compared to Cas9-WT. Moreover, in transfections with the biotinylated ssDNA

template, Cas9-MSA** caused the biggest yield of GFP-positive events, with a 2.1-fold increase of HDR efficiency when compared to the commonly used Cas9-WT and non-biotinylated donor DNA (**Figure 3.9B)**.

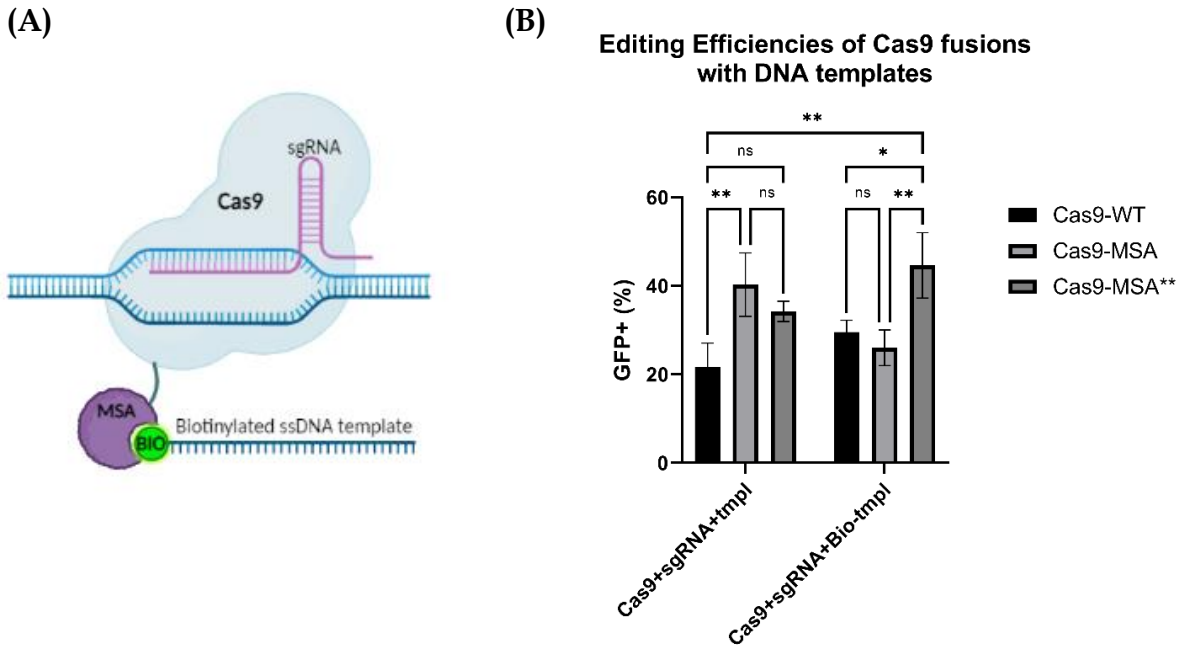

Figure 3.9 - (A): Schematic representation of the Cas9-MSA fusion's action mechanism, through the binding of MSA to the biotin molecule present in the exogenously provided ssDNA template. **(B)**: Evaluation of HDR activities of Cas9-MSA fusions in co-transfections with exogenous ssDNA non-biotinylated (tmpl) or 5'-biotinylated (Bio-tmpl) templates. ns $p > 0.05$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$. SD is represented by the error bars (n=3).

Figure 3.10 is a summary of the quantification results of GFP-positive events in all transfection conditions, and significant differences between Cas9 only, Cas9 + sgRNA, Cas9 + sgRNA + template and Cas9 + sgRNA + biotinylated template, for each Cas9-encoding plasmid. No significant differences between the co-transfection of Cas9-WT with either GFPtemp or the 5'-biotinylated GFPtemp were observed. Concerning Cas9-MSA, the largest number of HDR-caused GFP-positive events was achieved upon co-transfection with the sgRNA and the non-biotinylated GFPtemp, displaying a reduction of GFP percentages in the condition in which the biotinylated template was added. In addition, the highest average editing efficiency was achieved in DR-GFP cells transfected with the Cas9-MSA** fusion and the biotinylated GFPtemp ssDNA template.

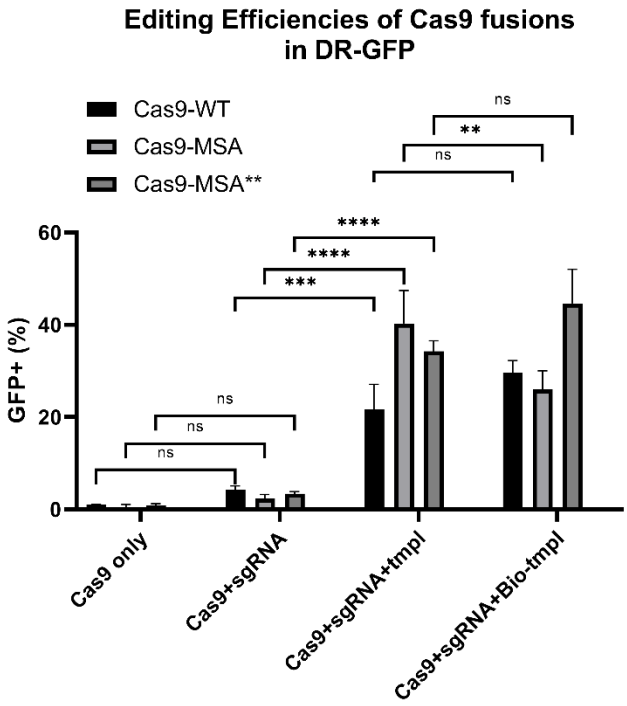

Figure 3.10 - Comparative results of GFP-positive events in DR-GFP cells transfected with different Cas9 plasmids, in the presence or not of the sgRNA, ssDNA template (GFPtemp) or biotinylated ssDNA template (Bio-GFPtemp). ns $p > 0.05$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$. SD is represented by the error bars (n=3).

3.8 Sequencing of Targeted Regions in Sorted DR-GFP Transfected Cells

Because the HphI enzyme did not arrive in the time frame of this practical work, sequencing of the targeted editing site was performed to evaluate the presence of superimposed sequence variants. Thus, after sorted DR-GFP cells from all transfection conditions were cultured until near-confluency, genomic DNA was extracted and used to amplify the sgRNA's target region in the *SceGFP* gene. PCR products were separated in an agarose gel by electrophoresis (**Figure 3.11**), and the 500 bp desired bands were purified and sent for Sanger sequencing. It should be noted that the primers used in this PCR amplify both *SceGFP* and *iGFP*.

Sanger sequencing results of DNA extracted from one triplicate of each condition are observed in **Figure 3.12**, as well as the sequences of *SceGFP*, *iGFP* and the exogenous GFPtemp ssDNA donor. Sequencing of the GFP sequence in cells transfected with Cas9-MSA** and sgRNA-encoding plasmids (condition C1) failed multiple times for all triplicates, therefore no results for this condition are presented in **Figure 3.12**.

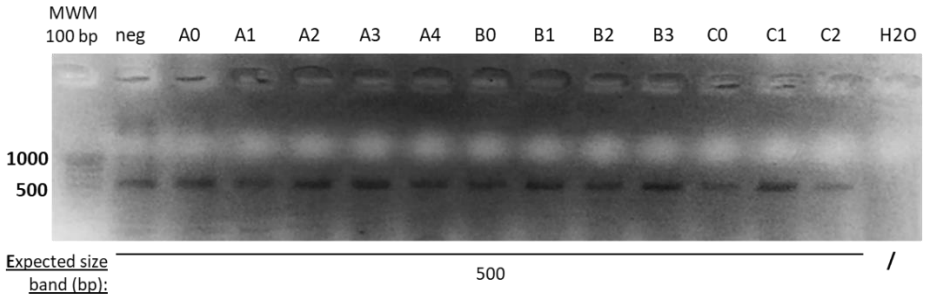

Figure 3.11 - Image of 2% agarose gel visualized under UV transilluminator after electrophoretic separation of PCR products, which were amplified from genomic DNA of transfected DR-GFP cells. neg – negative transfection control; A – Cas9-WT, B – Cas9-MSA, C – Cas9-MSA**, 0 – Cas9 only, 1 – Cas9 + sgRNA, 2 – Cas9 + sgRNA + GFPtemp, 3 – Cas9 + sgRNA + Bio-GFPtemp; H2O – negative PCR control.

Sanger sequencing chromatograms were analysed and, for the nucleotides that differ between SceGFP, iGFP, and the exogenous template (GFPtemp) sequences, the relative abundances of each distinctive nucleotide were calculated. Since the nucleotides that differentiate between the exogenous GFPtemp and the endogenous GFP sequences are in positions 6 and 24, peak heights of overlapping nucleotides in these positions were measured using BioEdit software [140], and relative average contributions of the exogenous ssDNA template are represented in **Table 3.1**. It was confirmed that the exogenous template only made contributions in conditions where it was used in cell transfection (except for condition A0, in which the high background noise levels originate a slight T detection in nucleotide 6). When the biotinylated template was used in electroporation, in Cas9-WT transfections, the contribution of this template's sequence in HDR editing appears to slightly decrease when compared to the one observed in the non-biotinylated donor condition. In contrast, in transfections with Cas9-MSA and Cas9-MSA**, GFPtemp contribution showed an increase of 1.31 and 1.27-fold in relation to the non-biotinylated template, respectively, demonstrating that the MSA-biotin interaction is improving Cas9 targeting and editing with the exogenous template.

(A)

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

TAG GGA TAA CAG GGT AAT ACC TAC – SceIGFP

GAG GGC GAG GGC GAT GCC ACC TAC – iGFP

GAG GGT GAG GGC GAT GCC ACC TAT – GFPtemp

(B)

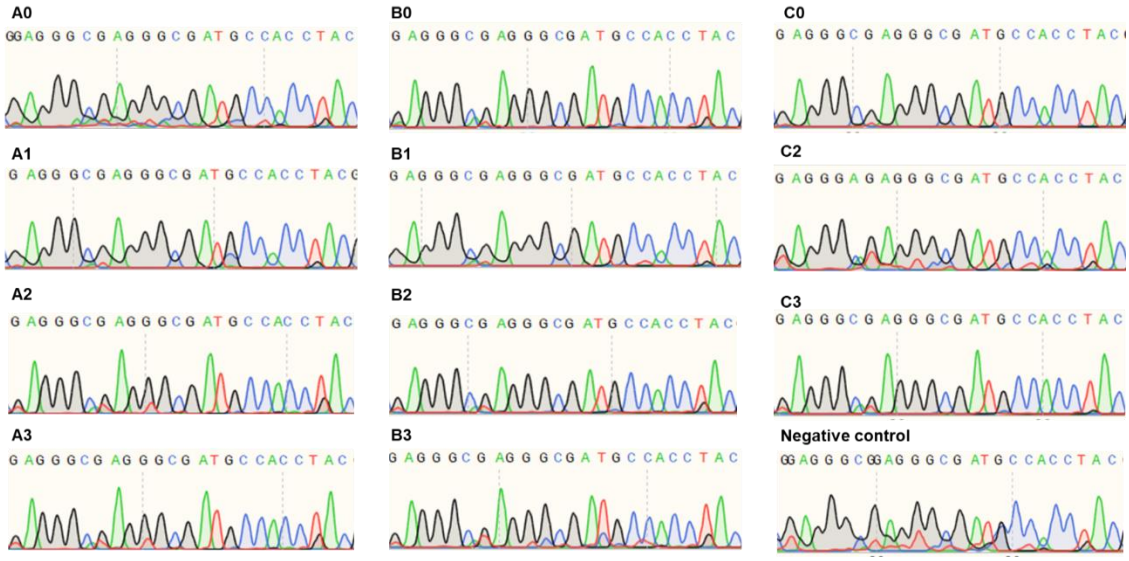

Figure 3.12 - **(A)**: 5′-3′ sequence of the targeted editing site in *SceGFP* and the corresponding one in the *iGFP* gene, as well as the GFPtemp template used in transfections. **(B):** Sequencing results of DNA from one triplicate of each condition, represented in the following nomenclature. Conditions beginning with: A – Cas9-WT, B – Cas9-MSA, C – Cas9-MSA**; conditions ending with: 0 – Cas9 only, 1 – Cas9 + sgRNA, 2 – Cas9 + sgRNA + GFPtemp, 3 – Cas9 + sgRNA + Bio-GFPtemp.

Table 3.1 - GFPtemp relative contributions in varying nucleotides 6 and 24 of DNA amplified and sequenced from sorted transfected DR-GFP cells, based on peak heights of each nucleotide for each position. Conditions beginning with: A – Cas9-WT, B – Cas9-MSA, C – Cas9-MSA**; conditions ending with: 0 – Cas9 only, 1 – Cas9 + sgRNA, 2 – Cas9 + sgRNA + GFPtemp, 3 – Cas9 + sgRNA + Bio-GFPtemp.

position	**A0**	**Average**	**A1**	**Average**	**A2**	**Average**	**A3**	**Average**
6	2.15%	1.08%	0%	0%	0%	**0.53%**	0.49%	**0.50%**
24	0%		0%		1.06%		0.50%	
position	**B0**	**Average**	**B1**	**Average**	**B2**	**Average**	**B3**	**Average**
6	0%	0%	0%	0%	2.15%	**1.75%**	2.92%	**2.30%**
24	0%		0%		1.35%		1.68%	
position		**C0**		**Average**	**C2**	**Average**	**C3**	**Average**
6		0%		0%	3.93%	**1.96%**	3.11%	**2.48%**
24		0%			0%		1.85%	

4

DISCUSSION AND CONCLUSIONS

Throughout this thesis, an approach based on the tethering of a Cas9-fused MSA to a biotinylated DNA template was presented to improve the enrichment of donor DNA for HDR at the desired editing site. Because HDR only takes place in specific cell cycle stages and HDR activity is comparatively weak in mammalian cells, designing strategies to increase editing efficiencies in knock-in experiments is essential.

In 2018, Gu *et al.* described the fusion of monomeric streptavidin (MSA) to the C-terminus of the *Cas9* encoding region, and used this fusion to precise and efficiently insert reporter genes in mouse embryos [136]. In this thesis, two-point mutations in the *MSA* gene were hypothesized to increase this protein's affinity to biotin (Sheldon Park, unpublished data), thus potentially enhancing the template DNA concentration near the cleavage site and further improving HDR-based repair of the DSB originated by Cas9.

With this objective, S14R and T39W mutations – with the exchange of MSA's Serine-14 to an Arginine and Threonine-39 to a Tryptophan – were inserted in the Cas9-MSA-encoding plasmid. The newly engineered fusion protein-encoding plasmid (Cas9-mSA**), as well as the original Cas-MSA and Cas9-WT plasmids, were used to transfect DR-GFP cells. This cell line has an integrated reporter that consists of a GFP cassette adjacent to the promoter interrupted by an I-SceI restriction site, giving rise to two in-frame premature termination codons, preventing GFP expression, and a downstream truncated GFP cassette that has the correct GFP sequence without the I-SceI restriction site [138]. In transfection experiments, with the aid of a sgRNA that guides Cas9 to the restriction site in the *SceGFP* gene, along with a donor DNA template, the produced DSB can be repaired by HDR through either the endogenous intact GFP cassette or the exogenously provided template, converting the cells into GFP-positive events that can be quantified through flow cytometry as a measure of HDR. Different conditions were tested - Cas9-WT, Cas9-MSA and Cas9-MSA** encoding plasmids, either alone or with the addition of the sgRNA, non-biotinylated or biotinylated ssDNA templates - and, after optimization, transfected cells were sorted through flow cytometry and their genomic DNA was extracted for sequencing

of the target cleavage site in *SceGFP*. Below, the main conclusions of the different experimental results are discussed.

4.1 Construction of the Cas9-MSA** Plasmid

The two-point mutations – S14R and T39W – were inserted into the *MSA* gene through the described fusion PCR method. The fusion PCR product was only generated in reaction mixtures with a smaller amount of each of the three individual amplicons (**Figure 3.1B**), possibly because of the composition of the elution buffer used to elute the individual PCR products, or even because larger concentrations of template PCR DNA in the reaction's mixture might increase the binding of primers to non-specific sequences and lead to poor DNA synthesis. After optimising the primary individual and fusion PCR protocols, several repetitions were performed, and each product was purified to increase the PCR products' stock.

The insert, consisting of the fusion PCR product with the desired mutations, was readily digested with StuI and BamHI. The expected band pattern was observed when products were separated in the agarose gel (**Figure 3.2A**). Upon failure of the Cas9-MSA vector's digestion, it was noted that the activity of StuI was blocked by overlapping DCM methylation. To solve this problem, the vector was used to transform a DCM-deficient strain of *E. coli* competent cells, prior to its digestion. Single digestion products observed in **Figure 3.2B** correspond to the linear conformation of the plasmid, whereas the uncut plasmid shows different sized bands as it can remain in different isoforms, possibly open circular (moves slower in the agarose gel) and linear monomers resulting from possible nuclease contamination. Upon ligation of digested insert and vector with T4, followed by transformation of *E. coli* DH5α competent cells with ligation mixtures, and colony PCR using all colonies grown in LB agar + ampicillin, four positive colonies – that originated a colony PCR product corresponding to the fusion product's molecular weight (**Figure 3.3**) – were chosen, and their plasmid DNA was extracted and sent for sequencing, confirming that all four colonies acquired the Cas9-MSA** plasmid with the desired mutations (**Figure 3.4**).

4.2 Testing HDR Activities of Cas9 Fusions

Once the protocol for the generation of ssDNA template for transfection of DR-GFP cells was optimised (**Figure 3.5**), Cas9-MSA and Cas9-MSA** fusions' activities in the described DR-GFP reporter were evaluated. In an experience comparing transfection with Cas9 only and Cas9 + sgRNA-encoding plasmids, both fusions were active, causing DNA cleavage and triggering HDR-based repair but, contrary to Cas9-MSA**, Cas9-MSA displayed a minor reduction of originated GFP-positive events in comparison to Cas9-WT (**Figure 3.7A**). This must be further analysed as one of the potential explanations

could be that the mutations inserted in the *MSA* gene may lead to conformational changes in the protein's structure, perhaps attenuating MSA's negative interference in its fused Cas9 observed in Cas-MSA's HDR activity. Furthermore, since only the endogenous *iGFP* template is being used as a donor DNA for HDR in this transfection condition, without the positive compensation of the ligation of the biotinylated template to the Cas9-sgRNA complex, the effect that MSA has in Cas9 is shown to be ultimately detrimental for the nuclease activity.

To favor the HDR pathway instead of NHEJ, the donor DNA must be in the vicinity of the Cas9 cleavage site. Hence, using high concentrations of ssDNA template might appear as an option to improve HDR editing efficiencies; nevertheless, the template DNA itself has a toxic effect in cells. Using 0.2 nmol of GFPtemp, the highest number of GFP-positive events was detected in DR-GFP cells, displaying a declining tendency with the increase of ssDNA dosage (**Figure 3.8**). However, viability data could not be evaluated in this experiment since propidium iodide (PI), the available membrane impermeant stain that is excluded from viable cells, absorbs at a maximum wavelength of ~535 nm and emits at ~615 nm and, therefore, both absorption and emission spectra overlap with those of RFP - ~555 nm absorption maximum and ~585 emission maximum. Both markers would therefore need to be excited by the same 561 nm (yellow) laser in FACS Aria II and detected by the same filter, making it impossible to differentiate the different PI-positive and RFP-positive populations.

A transfection experiment using Cas9 and sgRNA-encoding plasmids, with the addition of either GFPtemp or 5'-biotinylated GFPtemp, was carried out (**Figure 3.9B**). Compared to what was observed when only Cas9 and sgRNA plasmids were used - hence when the endogenous iGFP template was acting as the sole HDR donor DNA - the HDR efficiency of all three Cas9-encoding plasmids increased. This variation may be justified by the presence of many copies of the exogenous template in the cell post-transfection, while only a single-copy of the *iGFP* gene exists in the cell's genome. These differences became more evident for Cas9-MSA** when the biotinylated GFPtemp was introduced since, among all conditions, cells that had been transfected with this donor ssDNA along with the Cas9-MSA** and sgRNA-encoding plasmids exhibited the biggest yield of GFP-positive events – with a 1.3-fold increase of HDR efficiency comparatively to the transfection with the same biotinylated template but with the unmutated Cas9-MSA plasmid (**Figure 3.10**), demonstrating that the S14R and T39W mutations can effectively contribute to better HDR efficiencies in this knock-in experiment. The fact that that the original Cas9-MSA showed an HDR decrease when the biotinylated exogenous template was provided, compared to the non-biotinylated one, is not in agreement with Gu *et al*. [136]. In their study, mouse embryos were cytoplasmically microinjected with Cas9-MSA mRNA and sgRNA, whereas, in this thesis, cells were transfected with plasmids encoding these CRISPR components. This may explain the difference in efficiency results since, upon transfection, the Cas9-MSA mRNA is readily translated into the protein on the cytoplasm, establishing the ribonucleoprotein complex early on, while, in transfections with plasmid DNA, CRISPR components need to enter the nucleus where transcription takes

place first, which possibly does not facilitate the formation of Cas9-MSA + sgRNA + biotinylated-template complexes. Moreover, Gu *et al.* used a dsDNA donor template symmetrically biotinylated in both 5' and 3' ends, which perhaps increases the probability of streptavidin-biotin binding, whereas an ssDNA template with only 5' biotinylation was used in this work.

The exogenous ssDNA template was designed to contain a silent mutation that inserts a HphI restriction site in the *SceGFP* gene when this template is used as donor in HDR-based repair. Unfortunately, due to shipping difficulties, this enzyme was not delivered in time for its experimental purpose. Thus, as an alternative, the GFP gene was amplified from the DNA of sorted RFP-positive cells of each transfection condition, to evaluate the competition between the usage of the genomic *iGFP* template and the exogenously provided ssDNA donor for *SceGFP* editing. Sanger sequencing chromatograms were analysed and, for the nucleotides that differ between SceGFP, iGFP and the exogenous template GFPtemp sequences, the relative abundances of each distinctive nucleotide were calculated. Low percentages were obtained because *SceGFP* and *iGFP* were both being amplified in the PCR that amplified the sequenced DNA, and the negative transfection control showed too much background noise to be used to normalize sequencing results of the other conditions. However, it was possible to observe that, when DR-GFP cells were transfected with Cas9-MSA and Cas9-MSA** plasmids, GFPtemp relative contribution showed an increase of 1.31 and 1.27-fold in relation to the non-biotinylated template, respectively. This demonstrates that, even in Cas9-MSA's case, where HDR efficiencies decreased when the biotinylated template was added, the percentage of used exogenous template for DSB repair still increases, indicating that the system of interaction between MSA and the biotin molecule in the ssDNA donor appears to be successfully working in concentrating DNA template molecules near the Cas9 cleavage site.

4.3 Future Prospects

Although the experiments performed in this thesis provided encouraging conclusions, additional experimental work and analysis must be conducted to achieve more informative and concusive results.

1 - First, since the sequencing of the DNA extracted from DR-GFP cells transfected with Cas9-MSA** and sgRNA-encoding plasmids failed several times, a transfection experiment in these conditions should be repeated, and new DNA must be extracted for further analysis and comparisons.

2 - Experiments in which different concentrations of the template DNA are co-transfected with Cas9 and sgRNA-encoding plasmids should be performed to evaluate the difference between Cas9-MSA and Cas9-MSA** HDR efficiencies when the ssDNA donor becomes limiting.

3 - Because the exogenously provided ssDNA template was designed to insert a HphI restriction site in the *SceGFP* edited region, HphI digestion of genomic DNA extracted from sorted cells of each transfection condition will be carried out once the enzyme arrives at the laboratory. This will provide an internal control and allow a better quantification of the percentage of cells edited with either the endogenous or the exogenous donor DNA.

4 - The tested biotinylated DNA template was only modified to have one biotin molecule at its 5'-end. However, other templates should be evaluated, with only 3' or both 5' and 3' biotinylation. This can be achieved through the developed asymmetric PCR method using biotinylated primers instead of the unmodified ones.

5 - Another strategy that is sought to be assessed is the design of an intermediate biotinylated ssDNA molecule that can hybridize to multiple copies of the same template oligo DNA, thus potentially increasing several-fold the local concentration of the template used for gene editing.

6 - Production and purification of the mutated fusion protein (Cas9-MSA**) needs to take place in order to describe the protein structure and function further, and establish possible correlations between the mutated amino acids in MSA and its biotin affinity. Transfection of cells with a Cas9-MSA**-biotinylated template complex assembled in vitro remains a priority.

7 - To evaluate differences in off-target activities between transfection conditions, some of the most significantly probable off-targets in DR-GFP cells' genome, for this sgRNA, must be amplified from DNA of transfected cells and sent for sequencing.

BIBLIOGRAPHY

[1] N. Chatterjee and G. C. Walker, "Mechanisms of DNA Damage, Repair, and Mutagenesis," *Environ. Mol. Mutagen.*, vol. 58, no. 5, pp. 235–263, 2017.

[2] T. A. Kunkel, "Evolving Views of DNA Replication (In) Fidelity OVERLAPPING FUNCTIONS," *cold Spring Harb. Lab. Press*, vol. 74, no. 0, pp. 91–101, 2009.

[3] A. J. Griffiths, W. M. Gelbart, J. H. Miller, and R. C. Lewontin., "The Molecular Basis of Mutation," in *Modern Genetic Analysis*, New York: W. H. Freeman, 1999.

[4] S. Brogna and J. Wen, "Nonsense-mediated mRNA decay (NMD) mechanisms," *Nat. Struct. Mol. Biol.*, vol. 16, no. 2, pp. 107–113, 2009.

[5] T. Brown, "Mutation, Repair and Recombination," in *Genomes. 2nd edition.*, Oxford: Wiley-Liss, 2002.

[6] J. H. J. Hoeijmakers, "Genome maintenance mechanisms for preventing cancer," *Nature*, vol. 411, no. 6835, pp. 366–374, 2001.

[7] S. Fulda, "Cell death and survival signaling in cancer therapy," *Signaling, Gene Regul. Cancer*, pp. 1–18, 2012.

[8] W. P. Roos, A. D. Thomas, and B. Kaina, "DNA damage and the balance between survival and death in cancer biology," *Nat. Rev. Cancer*, vol. 16, no. 1, pp. 20–33, 2016.

[9] A. Tubbs and A. Nussenzweig, "Endogenous DNA Damage as a Source of Genomic Instability in Cancer," *Cell*, vol. 168, no. 4, pp. 644–656, 2017.

[10] S. D. McCulloch and T. A. Kunkel, "The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases," *Cell Res.*, vol. 18, no. 1, pp. 148–161, 2008.

[11] B. D. Preston, T. M. Albertson, and A. J. Herr, "DNA replication fidelity and cancer," *Semin. Cancer Biol.*, vol. 20, no. 5, pp. 281–293, 2010.

[12] G. M. Li, "Mechanisms and functions of DNA mismatch repair," *Cell Res.*, vol. 18, no. 1, pp. 85–98, 2008.

[13] C. Schmutte, R. C. Marinescu, M. M. Sadoff, S. Guerrette, J. Overhauser, and R. Fishel, "Human exonuclease I interacts with the mismatch repair protein hMSH2," *Cancer Res.*, vol. 58, no. 20, pp. 4537–4442, 1998.

[14] A. Umar *et al.*, "Requirement for PCNA in DNA mismatch repair at a step preceding DNA resynthesis," *Cell*, vol. 87, no. 1, pp. 65–73, 1996.

[15] C. Ramilo *et al.*, "Partial Reconstitution of Human DNA Mismatch Repair In Vitro: Characterization of the Role of Human Replication Protein A," *Mol. Cell. Biol.*, vol. 22, no. 7, pp. 2037–2046, 2002.

[16] F. A. Kadyrov, L. Dzantiev, N. Constantin, and P. Modrich, "Endonucleolytic Function of MutLα in Human Mismatch Repair," *Cell*, vol. 126, no. 2, pp. 297–308, 2006.

[17] M. J. Longley, A. J. Pierce, and P. Modrich, "DNA polymerase δ is required for human mismatch repair in vitro," *J. Biol. Chem.*, vol. 272, no. 16, pp. 10917–10921, 1997.

[18] G. L. Dianov and U. Hübscher, "Mammalian base excision repair: The forgotten archangel," *Nucleic Acids Res.*, vol. 41, no. 6, pp. 3483–3490, 2013.

[19] Y. Matsumoto and K. Kim, "Excision of deoxyribose phosphate residues by DNA polymerase beta during DNA repair," *Science (80-.).*, vol. 269, no. 5224, pp. 699–702, 1995.

[20] T. Nouspikel, "Nucleotide excision repair: Variations on versatility," *Cell. Mol. Life Sci.*, vol. 66, no. 6, pp. 994–1009, 2009.

[21] K. Sugasawa *et al.*, "Xeroderma pigmentosum group C protein complex is the initiator of global genome nucleotide excision repair," *Mol. Cell*, vol. 2, no. 2, pp. 223–232, 1998.

[22] M. Araki *et al.*, "Centrosome Protein Centrin 2/Caltractin 1 Is Part of the Xeroderma Pigmentosum Group C Complex That Initiates Global Genome Nucleotide Excision Repair," *J. Biol. Chem.*, vol. 276, no. 22, pp. 18665–18672, 2001.
[23] A. O'Donovan, A. A. Davies, J. G. Moggs, S. C. West, and R. D. Wood, "XPG endonuclease makes the 3′ Incision in human DNA nucleotide excision repair," *Nature*, vol. 371, no. 6496. pp. 432–435, 1994.
[24] D. Mu, D. S. Hsu, and A. Sancar, "Reaction mechanism of human DNA repair excision nuclease," *J. Biol. Chem.*, vol. 271, no. 14, pp. 8285–8294, 1996.
[25] O. Popanda and H. W. Thielmann, "The function of DNA polymerases in DNA repair synthesis of ultraviolet-irradiated human fibroblasts," *BBA - Gene Struct. Expr.*, vol. 1129, no. 2, pp. 155–160, 1992.
[26] C. Yi and C. He, "DNA repair by reversal of DNA damage," *Cold Spring Harb. Perspect. Biol.*, vol. 5, no. 1, pp. 1–18, 2013.
[27] F. Pâques and J. E. Haber, " Multiple Pathways of Recombination Induced by Double-Strand Breaks in Saccharomyces cerevisiae ," *Microbiol. Mol. Biol. Rev.*, vol. 63, no. 2, pp. 349–404, 1999.
[28] P. Soulas-Sprauel *et al.*, "V(D)J and immunoglobulin class switch recombinations: A paradigm to study the regulation of DNA end-joining," *Oncogene*, vol. 26, no. 56, pp. 7780–7791, 2007.
[29] S. Keeney and M. J. Neale, "Initiation of meiotic recombination by formation of DNA double-strand breaks: Mechanism and regulation," *Biochem. Soc. Trans.*, vol. 34, no. 4, pp. 523–525, 2006.
[30] J. San Filippo, P. Sung, and H. Klein, "Mechanism of eukaryotic homologous recombination," *Annu. Rev. Biochem.*, vol. 77, pp. 229–257, 2008.
[31] M. F. Lavin, "ATM and the Mre11 complex combine to recognize and signal DNA double-strand breaks," *Oncogene*, vol. 26, no. 56, pp. 7749–7758, 2007.
[32] C. J. Bakkenist and M. B. Kastan, "DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation," *Nature*, vol. 421, no. 6922, pp. 499–506, 2003.
[33] Y. Shiloh, "ATM and related protein kinases: safeguarding genome integrity," *Nat. Rev. Cancer*, vol. 3, no. 3, pp. 155–168, 2003.
[34] D. Lim *et al.*, "ATM phosphorylates p95/nbs1 in an S-phase checkpoint pathway," *Nature*, vol. 404, pp. 613–617, 2000.
[35] E. P. Mimitou and L. S. Symington, "Nucleases and helicases take center stage in homologous recombination," *Trends Biochem. Sci.*, vol. 34, no. 5, pp. 264–272, 2009.
[36] M. S. Wold, "Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism," *Annu. Rev. Biochem.*, vol. 66, no. 1, pp. 61–92, 1997.
[37] W. D. Heyer, X. Li, M. Rolfsmeier, and X. P. Zhang, "Rad54: The Swiss Army knife of homologous recombination?," *Nucleic Acids Res.*, vol. 34, no. 15, pp. 4115–4125, 2006.
[38] X. Li, C. M. Stith, P. M. Burgers, and W. D. Heyer, "PCNA Is Required for Initiation of Recombination-Associated DNA Synthesis by DNA Polymerase δ," *Mol. Cell*, vol. 36, no. 4, pp. 704–713, 2009.
[39] G. Mazón, E. P. Mimitou, and L. S. Symington, "SnapShot: Homologous recombination in DNA double-strand break repair," *Cell*, vol. 142, no. 4, pp. 648.e1-648.e2, 2010.
[40] L. Wu and I. O. Hickson, "The Bloom's syndrome helicase suppresses crossing over during homologous recombination," *Nature*, vol. 426, no. 6968, pp. 870–874, 2003.
[41] J. M. Svendsen and J. W. Harper, "GEN1/Yen1 and the SLX4 complex: Solutions to the problem of Holliday junction resolution," *Genes Dev.*, vol. 24, no. 6, pp. 521–536, 2010.
[42] P. Frit, V. Ropars, M. Modesti, J. B. Charbonnier, and P. Calsou, "Plugged into the Ku-DNA hub: The NHEJ network," *Prog. Biophys. Mol. Biol.*, vol. 147, pp. 62–76, 2019.
[43] E. Y. Shim *et al.*, "Saccharomyces cerevisiae Mre11/Rad50/Xrs2 and Ku proteins regulate association of Exo1 and Dna2 with DNA breaks," *EMBO J.*, vol. 29, no. 19, pp. 3370–3380, 2010.
[44] D. A. Ramsden and M. Gellert, "Ku protein stimulates DNA end joining by mammalian DNA ligases: A direct role for Ku in repair of DNA double-strand breaks," *EMBO J.*, vol. 17, no. 2, pp. 609–614, 1998.
[45] S. Yoo and W. S. Dynan, "Geometry of a complex formed by double strand break repair proteins at a single DNA end: Recruitment of DNA-PKcs induces inward translocation of Ku protein," *Nucleic Acids Res.*, vol. 27, no. 24, pp. 4679–4686, 1999.

[46] O. Hammarsten, L. G. DeFazio, and G. Chu, "Activation of DNA-dependent protein kinase by single-stranded DNA ends," *J. Biol. Chem.*, vol. 275, no. 3, pp. 1541–1550, 2000.
[47] Y. Ma, U. Pannicke, K. Schwarz, and M. R. Lieber, "Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination," *Cell*, vol. 108, no. 6, pp. 781–794, 2002.
[48] Y. Ma, K. Schwarz, and M. R. Lieber, "The Artemis:DNA-PKcs endonuclease cleaves DNA loops, flaps, and gaps," *DNA Repair (Amst).*, vol. 4, no. 7, pp. 845–851, 2005.
[49] B. Bertocci, A. De Smet, J. C. Weill, and C. A. Reynaud, "Nonoverlapping Functions of DNA Polymerases Mu, Lambda, and Terminal Deoxynucleotidyltransferase during Immunoglobulin V(D)J Recombination In Vivo," *Immunity*, vol. 25, no. 1, pp. 31–41, 2006.
[50] S. H. Teo and S. P. Jackson, "Identification of saccharomyces cerevisiae DNA ligase IV: Involvement in DNA double-strand break repair," *EMBO J.*, vol. 16, no. 15, pp. 4788–4795, 1997.
[51] U. Grawunder *et al.*, "Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells," *Nature*, vol. 388, no. 6641, pp. 492–495, 1997.
[52] P. Ahnesorg, P. Smith, and S. P. Jackson, "XLF interacts with the XRCC4-DNA Ligase IV complex to promote DNA nonhomologous end-joining," *Cell*, vol. 124, no. 2, pp. 301–313, 2006.
[53] W. Y. Mansour, T. Rhein, and J. Dahm-Daphi, "The alternative end-joining pathway for repair of DNA double-strand breaks requires PARP1 but is not dependent upon microhomologies," *Nucleic Acids Res.*, vol. 38, no. 18, pp. 6065–6077, 2010.
[54] L. Liang *et al.*, "Human DNA ligases I and III, but not ligase IV, are required for microhomology-mediated end joining of DNA double-strand breaks," *Nucleic Acids Res.*, vol. 36, no. 10, pp. 3297–3310, 2008.
[55] S. Sharma and S. C. Raghavan, *Nonhomologous DNA End Joining*, vol. 1. Elsevier Ltd., 2016.
[56] Y. Ishino, H. Shinagawa, K. Makino, M. Amemura, and A. Nakata, "Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isoenzyme conversion in Escherichia coli, and identification of the gene product," *J. Bacteriol.*, vol. 169, no. 12, pp. 5429–5433, 1987.
[57] F. J. M. Mojica, G. Juez, and F. Rodriguez-Valera, "Transcription at different salinities of Haloferax mediterranei sequences adjacent to partially modified PstI sites," *Mol. Microbiol.*, vol. 9, no. 3, pp. 613–621, 1993.
[58] F. J. M. Mojica, C. Díez-Villaseñor, E. Soria, and G. Juez, "Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria," *Mol. Microbiol.*, vol. 36, no. 1, pp. 244–246, 2000.
[59] R. Jansen, J. D. A. Van Embden, W. Gaastra, and L. M. Schouls, "Identification of genes that are associated with DNA repeats in prokaryotes," *Mol. Microbiol.*, vol. 43, no. 6, pp. 1565–1575, 2002.
[60] C. Pourcel, G. Salvignol, and G. Vergnaud, "CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies," *Microbiology*, vol. 151, no. 3, pp. 653–663, 2005.
[61] A. Bolotin, B. Quinquis, A. Sorokin, and S. Dusko Ehrlich, "Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin," *Microbiology*, vol. 151, no. 8, pp. 2551–2561, 2005.
[62] F. J. M. Mojica, C. Díez-Villaseñor, J. García-Martínez, and E. Soria, "Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements," *J. Mol. Evol.*, vol. 60, no. 2, pp. 174–182, 2005.
[63] R. Barrangou *et al.*, "CRISPR Provides Against Viruses in Prokaryotes," *Science (80-.).*, vol. 315, no. March, pp. 1709–1712, 2007.
[64] Brouns *et al.*, "Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes," *Science (80-.).*, vol. 321, no. 5891, pp. 960–964, 2008.
[65] L. A. Marraffini and E. J. Sontheimer, "CRISPR Interference Limits Horizontal Targeting DNA," *Science (80-.).*, vol. 322, no. 5909, pp. 1843–1845, 2008.
[66] H. Deveau *et al.*, "Phage response to CRISPR-encoded resistance in Streptococcus thermophilus," *J. Bacteriol.*, vol. 190, no. 4, pp. 1390–1400, 2008.
[67] J. E. Garneau *et al.*, "The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA," *Nature*, vol. 468, no. 7320, pp. 67–71, 2010.
[68] E. Deltcheva *et al.*, "CRISPR RNA maturation by trans-encoded small RNA and host factor

RNase III," *Nature*, vol. 471, no. 7340, pp. 602–607, 2011.
[69] R. Sapranauskas, G. Gasiunas, C. Fremaux, R. Barrangou, P. Horvath, and V. Siksnys, "The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli," *Nucleic Acids Res.*, vol. 39, no. 21, pp. 9275–9282, 2011.
[70] M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, and E. Charpentier, "A Programmable Dual-RNA – Guided DNA Endonuclease in Adaptive Bacterial Immunity," *Science (80-.).*, vol. 337, no. 6096, pp. 816–821, 2012.
[71] K. S. Makarova *et al.*, "Evolution and classification of the CRISPR-Cas systems," *Nat. Rev. Microbiol.*, vol. 9, no. 6, pp. 467–477, 2011.
[72] J. Carte, R. Wang, H. Li, R. M. Terns, and M. P. Terns, "Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes," *Genes Dev.*, vol. 22, no. 24, pp. 3489–3496, 2008.
[73] B. Zetsche *et al.*, "Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System," *Cell*, vol. 163, no. 3, pp. 759–771, 2015.
[74] K. S. Makarova *et al.*, "An updated evolutionary classification of CRISPR-Cas systems," *Nat. Rev. Microbiol.*, vol. 13, no. 11, pp. 722–736, 2015.
[75] I. Yosef, M. G. Goren, and U. Qimron, "Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli," *Nucleic Acids Res.*, vol. 40, no. 12, pp. 5569–5576, 2012.
[76] Y. Xiao, S. Ng, and K. H. Nam, "How Type II CRISPR-Cas establish immunity through Cas1-Cas2 mediated spacer integration," *Nature*, vol. 550, no. 7674, pp. 137–141, 2017.
[77] G. B. Robb, "Genome Editing with CRISPR-Cas: An Overview," *Curr. Protoc. Essent. Lab. Tech.*, vol. 19, no. 1, pp. 1–20, 2019.
[78] M. R. Capecchi, "Altering the Genome by Homologous Recombination," *Science (80-.).*, vol. 244, no. 4910, pp. 1288–1292, 1989.
[79] P. Rouet, F. Smih, and M. Jasin, "Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease," *Mol. Cell. Biol.*, vol. 14, no. 12, pp. 8096–8106, 1994.
[80] B. L. Stoddard, "Homing endonucleases: From microbial genetic invaders to reagents for targeted DNA modification," *Structure*, vol. 19, no. 1, pp. 7–15, 2011.
[81] J. P. Gogarten and E. Hilario, "Inteins, introns, and homing endonucleases: Recent revelations about the life cycle of parasitic genetic elements," *BMC Evol. Biol.*, vol. 6, pp. 1–5, 2006.
[82] B. L. Stoddard, "Homing endonucleases from mobile group i introns: Discovery to genome engineering," *Mob. DNA*, vol. 5, no. 1, 2014.
[83] A. M. Khalil, "The genome editing revolution: review," *J. Genet. Eng. Biotechnol.*, vol. 18, no. 1, 2020.
[84] Y. G. Kim, J. Cha, and S. Chandrasegaran, "Hybrid restriction enzymes: Zinc finger fusions to Fok I cleavage domain," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 93, no. 3, pp. 1156–1160, 1996.
[85] J. Miller, A. D. McLachlan, and A. Klug, "Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes.," *EMBO J.*, vol. 4, no. 6, pp. 1609–1614, 1985.
[86] S. A. Wolfe, L. Nekludova, and C. O. Pabo, "DNA RECOGNITION BY Cys2His2 ZINC FINGER PROTEINS," *Annu. Rev. Biophys. Biomol. Struct*, vol. 29, pp. 183–212, 2000.
[87] É. S. Vanamee, S. Santagata, and A. K. Aggarwal, "FokI requires two specific DNA sites for cleavage," *J. Mol. Biol.*, vol. 309, no. 1, pp. 69–78, 2001.
[88] Q. Liu, D. J. Segal, J. B. Ghiara, and C. F. Barbas, "Design of polydactyl zinc-finger proteins for unique addressing within complex genomes," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 94, no. 11, pp. 5525–5530, 1997.
[89] F. D. Urnov, E. J. Rebar, M. C. Holmes, H. S. Zhang, and P. D. Gregory, "Genome editing with engineered zinc finger nucleases," *Nat. Rev. Genet.*, vol. 11, no. 9, pp. 636–646, 2010.
[90] K. J. Beumer *et al.*, "Efficient gene targeting in Drosophila by direct embryo injection with zinc-finger nucleases," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 105, no. 50, pp. 19821–19826, 2008.
[91] Y. Santiago *et al.*, "Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 105, no. 15, pp. 5809–5814, 2008.
[92] F. D. Urnov *et al.*, "Highly efficient endogenous human gene correction using designed zinc-finger nucleases," *Nature*, vol. 435, no. 7042, pp. 646–651, 2005.

[93] J. Zou *et al.*, "Gene Targeting of a Disease-Related Gene in Human Induced Pluripotent Stem and Embryonic Stem Cells," *Cell Stem Cell*, vol. 5, no. 1, pp. 97–110, 2009.
[94] R. M. Gupta and K. Musunuru, "Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9," *J. Clin. Invest.*, vol. 124, no. 10, pp. 4154–4161, 2014.
[95] J. Boch and U. Bonas, "Xanthomonas AvrBs3 Family-Type III effectors: Discovery and Function," *Annu. Rev. Phytopathol.*, vol. 48, pp. 419–436, 2010.
[96] J. K. Joung and J. D. Sander, "TALENs: a widely applicable technology for targeted genome editing," *Nat Rev Mol Cell Biol*, vol. 14, no. 1, pp. 49–55, 2013.
[97] D. F. Carlson *et al.*, "Efficient TALEN-mediated gene knockout in livestock," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 109, no. 43, pp. 17382–17387, 2012.
[98] D. Hockemeyer *et al.*, "Genetic engineering of human ES and iPS cells using TALE nucleases," *Nat Biotechnol*, vol. 29, no. 8, pp. 731–734, 2012.
[99] S. Ramalingam *et al.*, "Generation and genetic engineering of human induced pluripotent stem cells using designed zinc finger nucleases," *Stem Cells Dev.*, vol. 22, no. 4, pp. 595–610, 2013.
[100] S. H. Khan, "Genome-Editing Technologies: Concept, Pros, and Cons of Various Genome-Editing Techniques and Bioethical Concerns for Clinical Application," *Mol. Ther. - Nucleic Acids*, vol. 16, no. June, pp. 326–334, 2019.
[101] P. Mali *et al.*, "RNA-guided human genome engineering via Cas9," *Science (80-.).*, vol. 339, no. 6121, pp. 823–826, 2013.
[102] S. W. Cho, S. Kim, J. M. Kim, and J. S. Kim, "Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease," *Nat. Biotechnol.*, vol. 31, no. 3, pp. 230–232, 2013.
[103] M. Jinek, A. East, A. Cheng, S. Lin, E. Ma, and J. Doudna, "RNA-programmed genome editing in human cells," *Elife*, vol. 2013, no. 2, pp. 1–9, 2013.
[104] L. Cong *et al.*, "Multiplex Genome Engineering Using CRISPR/Cas Systems," *Science*, vol. 339, no. 6121, pp. 819–823, 2013.
[105] C. Mussolino, R. Morbitzer, F. Lütge, N. Dannemann, T. Lahaye, and T. Cathomen, "A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity," *Nucleic Acids Res.*, vol. 39, no. 21, pp. 9283–9293, 2011.
[106] J. Zhang *et al.*, "Comparison of gene editing efficiencies of CRISPR/Cas9 and TALEN for generation of MSTN knock-out cashmere goats," *Theriogenology*, vol. 132, pp. 1–11, 2019.
[107] F. A. Ran, P. D. Hsu, J. Wright, V. Agarwala, D. A. Scott, and F. Zhang, "Genome engineering using the CRISPR-Cas9 system," *Nat. Protoc.*, vol. 8, no. 11, pp. 2281–2308, 2013.
[108] J. Aach, P. Mali, and G. M. Church, "CasFinder: Flexible algorithm for identifying specific Cas9 targets in genomes," *bioRxiv*, pp. 1–8, 2014.
[109] M. Stemmer, T. Thumberger, M. Del Sol Keyer, J. Wittbrodt, and J. L. Mateo, "CCTop: An intuitive, flexible and reliable CRISPR/Cas9 target prediction tool," *PLoS One*, vol. 10, no. 4, pp. 1–11, 2015.
[110] and E. J. R. Dmitry Y. Guschin, Adam J. Waite, George E. Katibah, Jeffrey C. Miller, Michael C. Holmes, "A Rapid and General Assay for Monitoring Endogenous Gene Modification Dmitry," *Methods Mol. Biol.*, vol. 649, pp. 247–256, 2010.
[111] A. Ricroch, P. Clairand, and W. Harwood, "Use of CRISPR systems in plant genome editing: Toward new opportunities in agriculture," *Emerg. Top. Life Sci.*, vol. 1, no. 2, pp. 169–182, 2017.
[112] E. C. Oerke, "Crop losses to pests," *J. Agric. Sci.*, vol. 144, no. 1, pp. 31–43, 2006.
[113] K. Chen, Y. Wang, R. Zhang, H. Zhang, and C. Gao, "CRISPR/Cas Genome Editing and Precision Plant Breeding in Agriculture," *Annu. Rev. Plant Biol.*, vol. 70, pp. 667–697, 2019.
[114] C. E. Nelson *et al.*, "In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy," *Science (80-.).*, vol. 351, no. 6271, pp. 403–407, 2015.
[115] C. Long *et al.*, "Postnatal genome editing partially restores dystrophin expression in a mouse model of muscular dystrophy," *Science (80-.).*, vol. 351, no. 6271, pp. 400–403, 2015.
[116] M. Tabebordbar *et al.*, "In vivo gene editing in dystrophic mouse muscle and muscle stem cells," *Science (80-.).*, vol. 351, no. 6271, pp. 407–411, 2015.
[117] P. Liang *et al.*, "CRISPR/Cas9-mediated gene editing in human tripronuclear zygotes," *Protein Cell*, vol. 6, no. 5, pp. 363–372, 2015.
[118] E. A. Stadtmauer *et al.*, "CRISPR-engineered T cells in patients with refractory cancer," *Science*

(80-.)., vol. 367, no. 6481, pp. 1–20, 2020.
[119] H. Frangoul *et al.*, "CRISPR-Cas9 Gene Editing for Sickle Cell Disease and β-Thalassemia," *N. Engl. J. Med.*, vol. 384, no. 3, pp. 252–260, 2021.
[120] M. Yuan *et al.*, "A marker-free system for highly efficient construction of vaccinia virus vectors using CRISPR Cas9," *Mol. Ther. - Methods Clin. Dev.*, vol. 2, no. August, p. 15035, 2015.
[121] F. Jiang and J. A. Doudna, "CRISPR-Cas9 Structures and Mechanisms," vol. 46, no. 1, pp. 505–529, 2017.
[122] Y. Fu *et al.*, "High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells," *Nat. Biotechnol.*, vol. 31, no. 9, pp. 822–826, 2013.
[123] P. D. Hsu *et al.*, "DNA targeting specificity of RNA-guided Cas9 nucleases," *Nat Biotechnol*, vol. 31, no. 9, pp. 827–832, 2013.
[124] Y. Fu, J. D. Sander, D. Reyon, V. M. Cascio, and J. K. Joung, "Improving CRISPR-Cas nuclease specificity using truncated guide RNAs," *Nat. Biotechnol.*, vol. 32, no. 3, pp. 279–284, 2014.
[125] D. E. Ryan *et al.*, "Improving CRISPR-Cas specificity with chemical modifications in single-guide RNAs," *Nucleic Acids Res.*, vol. 46, no. 2, pp. 792–803, 2018.
[126] D. D. Kocak, E. A. Josephs, V. Bhandarkar, S. S. Adkar, J. B. Kwon, and C. A. Gersbach, "Increasing the specificity of CRISPR systems with engineered RNA secondary structures," *Nat. Biotechnol.*, vol. 37, no. 6, pp. 657–666, 2019.
[127] F. A. Ran *et al.*, "Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity," *Cell*, vol. 154, no. 6, pp. 1–10, 2013.
[128] J. P. Guilinger, D. B. Thompson, and D. R. Liu, "Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification," *Nat. Biotechnol.*, vol. 32, no. 6, pp. 577–582, 2014.
[129] I. M. Slaymaker, L. Gao, B. Zetsche, D. A. Scott, W. X. Yan, and F. Zhang, "Rationally engineered Cas9 nucleases with improved specificity," *Science (80-.).*, vol. 351, no. 6268, pp. 84–88, 2016.
[130] B. P. Kleinstiver *et al.*, "High-fidelity CRISPR-Cas9 variants with undetectable genome- wide off-targets," *Nature*, vol. 529, no. 7587, pp. 490–495, 2016.
[131] J. S. Chen *et al.*, "Enhanced proofreading governs CRISPR-Cas9 targeting accuracy," *Nature*, vol. 550, no. 7676, pp. 407–410, 2017.
[132] J. K. Lee *et al.*, "Directed evolution of CRISPR-Cas9 to increase its specificity," *Nat. Commun.*, vol. 9, no. 1, pp. 1–10, 2018.
[133] C. A. Vakulskas *et al.*, "A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human hematopoietic stem and progenitor cells," *Nat. Med.*, vol. 24, no. 8, pp. 1216–1224, 2018.
[134] J. Carlson-Stevermer *et al.*, "Assembly of CRISPR ribonucleoproteins with biotinylated oligonucleotides via an RNA aptamer for precise gene editing," *Nat. Commun.*, vol. 8, no. 1, 2017.
[135] M. Ma *et al.*, "Efficient generation of mice carrying homozygous double-floxp alleles using the Cas9-Avidin/Biotin-donor DNA system," *Cell Res.*, vol. 27, no. 4, pp. 578–581, 2017.
[136] B. Gu, E. Posfai, and J. Rossant, "Efficient generation of targeted large insertions by microinjection into two-cell-stage mouse embryos," *Nat. Biotechnol.*, vol. 36, no. 7, pp. 632–637, 2018.
[137] S. Park, "Monomeric Streptavidin Mutants, Methods of Using The Same and Processes of Manufacturing Proteins," 2018.
[138] A. J. Pierce, R. D. Johnson, L. H. Thompson, and M. Jasin, "XRCC3 promotes homology-directed repair of DNA damage in mammalian cells," *Genes Dev.*, vol. 13, no. 20, pp. 2633–2638, 1999.
[139] H. Inoue, H. Nojima, and H. Okayama, "High efficiency transformation of Escherichia coli with plasmids," *Gene*, vol. 96, no. 1, pp. 23–28, 1990.
[140] T. Hall, "BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT," *Nucleic Acids Symp. Ser.*, vol. 41, pp. 95–98, 1999.

2021 MARIA FILIPE GONÇALVES ENGINEERING A CAS9-MONOMERIC STREPTAVIDIN FUSION TO INCREASE CRISPR KNOCK-IN EFFICIENCY *IN VITRO*

NOVA SCHOOL OF SCIENCE & TECHNOLOGY