UNIVERSITÀ
DI SIENA
1240

DIPARTIMENTO DI BIOTECNOLOGIE, CHIMICA E FARMACIA

DEPARTMENT OF BIOTECHNOLOGY, CHEMISTRY AND PHARMACY

DOCTORAL COURSE IN BIOCHEMISTRY AND MOLECULAR BIOLOGY – BIBIM 2.0

BIO/11

Cycle XXXIII

Coordinator: Prof. Annalisa Santucci

BUILDING A PLATFORM FOR FLEXIBLE AND SCALABLE TESTING OF GENETIC EDITORS

Tutors:

Prof. Annalisa Santucci

Prof. Andrea Crisanti

Dr. Roberto Galizi

TERESA PETRUCCI

Firmato il 30/04/2021

PhD candidate: Teresa Petrucci

Academic year: 2020/2021

“But don’t hurry the journey at all.
Better if it lasts for years,
so you’re old by the time you reach the island,
wealthy with all you’ve gained on the way”.

Ithaca - Konstantinos P. Kavafis

A mamma e a papa’,
per avermi sempre lasciato libera di andare, di vivere e di cadere.
Ma stando sempre li’ dietro, pronti ad esserci,
ogni volta che ne avessi bisogno.

Table of contents

Abstract

Cell-free systems allow to perform *in-vitro* transcription-translation reactions without requiring living organisms, revolutionising scientific research over the last decade.

This allows to easily synthesise a variety of molecular components for genetic editing applications without requiring expensive and time-consuming procedures such as cell culture, animal maintenance etc.

In this work, I aimed to develop a high-throughput platform for the rapid, flexible and scalable *in-vitro* testing of various genetic editors, such as those part of the CRISPR/Cas repertoire. I used the commercially available *E. coli* cell extract (MyTXTL) in combination with a fully customisable design to generate fluorescent reporters, that allow standardised testing of various CRISPR components against any predesigned target or protospacer adjacent motif (PAM) sequences.

In order to increase the scalability of this screening platform, I utilised automated liquid handling technologies (Echo 525) and explored the possibility to introduce a high throughput cloning method (BASIC assembly).

I believe that this approach will be highly valuable for the screening of CRISPR components prior than their final application in *in-vivo* systems, such as humans or animals.

These genetic editors could then be used in many biological and artificial systems, such as gene editing, metabolomics and genetic engineering.

1. Introduction

1.1 CRISPR system

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) is a powerful tool which can be used to edit the genome in a targeted manner. It is already widely used for scientific research, allowing to easily alter DNA sequences and modify gene function.

It takes origins from an adaptive immune system of prokaryotic organisms, a natural defence mechanism that bacteria and archaea have developed to foil attacks by viruses and other invaders (Barrangou et al., 2007; Barrangou et al., 2014; Fineran et al., 2012; Marraffini et al., 2010). CRISPR sequences were initially discovered in *E. coli* genome and derived from repetitive DNA fragments of bacteriophages that had previously attacked the prokaryote. Using CRISPR-associated (Cas) genes, they could store a record of this invasion, which allowed to destroy the DNA of the invader upon further exposure.

CRISPRs are diverse families of DNA repeats which all share a common architecture. Each CRISPR locus includes conserved short repeat sequences of approximately 20 base pairs (bp) length separated by unique spacer sequences.

In details, the mechanism of protection in each CRISPR-Cas system could be divided into three stages (figure 1.1): in the first one, called adaptation or spacer acquisition, specific Cas proteins cut the phage DNA in fragments of approximately 20 base pairs (bp) length, which constitute the conserved repeats. This foreign DNA selected for integration is called protospacer. Then, the repeat fragments are associated with adjacent unique spacer sequence, constituting the CRISPR array. The integration of the protospacer begins with a concerted cleavage–ligation reaction that occurs at the end of the first repeat, via nucleophilic attack of the protopacer to the self-DNA. Consequently, the repeat is bent, leading to a second cleavage–ligation reaction at the spacer side of the sequence, that allows the insertion of the protospacer into the CRISPR locus (McGinn et al., 2019). Downstream the target site is located a conserved protospacer-adjacent motif (PAM) that mediates the recognition of the invader DNA and its distruption. This sequence allows for distintion beteween bacterial self from non-self-DNA, because it is only found next to complementary protospacer targets in foreign DNA, and is absent from repeat in the host CRISPR locus. The PAM sequence and length depends on the CRISPR-Cas type and the mechanism of PAM recognition likely occurs through protein-DNA interactions.

The second stage of the CRISPR-Cas adaptive immunity is the expression or production of CRISPR RNA: first, the CRISPR array is transcribed in long precursor crRNA (pre-crRNA); then specific enzymes process and mature it in individual CRISPR RNAs (crRNAs), which base pair with the trans-activating CRISPR RNA (tracrRNA) to form a functional guide RNA (gRNA). The tracrRNA harbours a sequence that is complementary to the repeat motif of the crRNA precursor. Through sequence homology, the gRNA guides another Cas protein with endonuclease function to a specific site in the genome (Barrangou, 2013; Marraffini, 2016). This protein uses the tracrRNA portion of the guide as a handle, while the crRNA one directs the CRISPR complex to a matching viral sequence, which constitutes the target site.

Interference or targeting is the final stage, when the prokaryote, invaded by the virus, recognises it; the CRISPR complex binds the foreign DNA and finally the Cas endonuclease cleaves the invader sequence, destroying the bacteriophage (Hille et al., 2018; Sorek et al., 2013; Van der Oost et al., 2014).

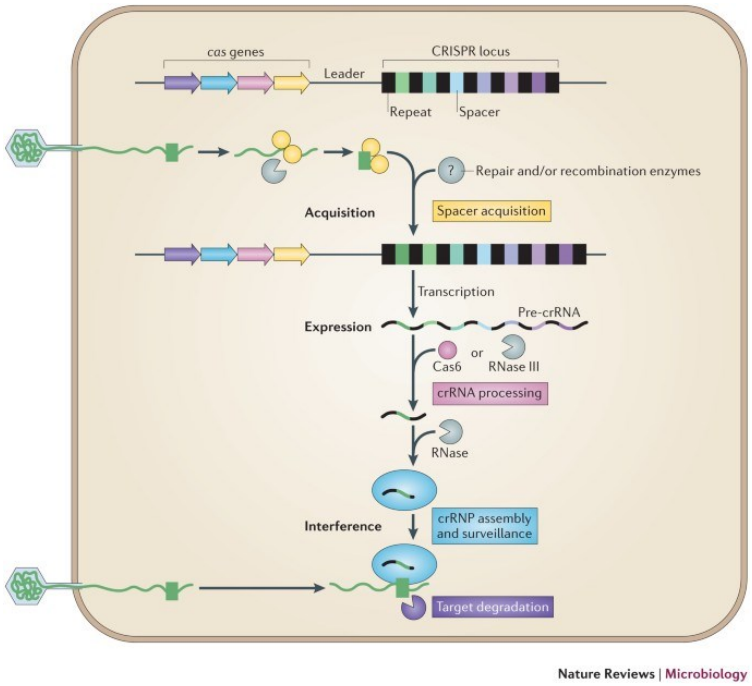


Figure 1.1: Stages of the CRISPR-Cas adaptive immunity in prokaryotes

In the first stage, adaptation, the genome derived from the bacteriophage DNA is cut in fragments; these foreign repeats are then assembled with unique spacers into the bacterial genome, constituting the CRISPR array. In the second stage, production of CRISPR RNA, the CRISPR array is matured and processed into single CRISPR RNAs. These crRNAs base pair with the tracrRNA to form a functional gRNA. In the third stage, targeting, the prokaryote recognises the bacteriophage thanks to the repeat sequences previously integrated into the genome; therefore, the CRISPR machinery is activated, and the viral genome is destroyed.

Source: Van der Oost et al., Unravelling the structural and mechanistic basis of CRISPR–Cas systems, *Nature Reviews Microbiology*, 2014

1.1.1 CRISPR-Cas classification

The diversity of CRISPR-Cas systems is classified in two main classes (figure 1.2), further divided into types and subtypes, based on criteria such as evolution, distribution, signature proteins, function (Sorek et al., 2013; Shabbir et al., Wang et al., 2020; Makarova et al., 2015). The main difference between Classes 1 and 2 is the complexity of their enzymes: Class 1 is based on multiple sub-unit effectors, while Class 2 contains single Cas proteins.

Focusing on Class 2 CRISPR-Cas systems, they only represent 10% of the CRISPR loci and, unlike Class 1, they are only found in bacteria and can target both DNA and RNA, depending on the type. Class 2 CRISPR-Cas systems are classified in two types, II, V and VI and each of them is characterised by a signature gene encoding for a specific Cas endonuclease responsible for cleavage and mechanism

of action. In details, type II has *cas9* gene, targets the DNA and requires a tracrRNA for function; type V has *cas12*, targets the DNA and also requires a tracrRNA; type VI has *cas13* and is the only class II system that targets the RNA.

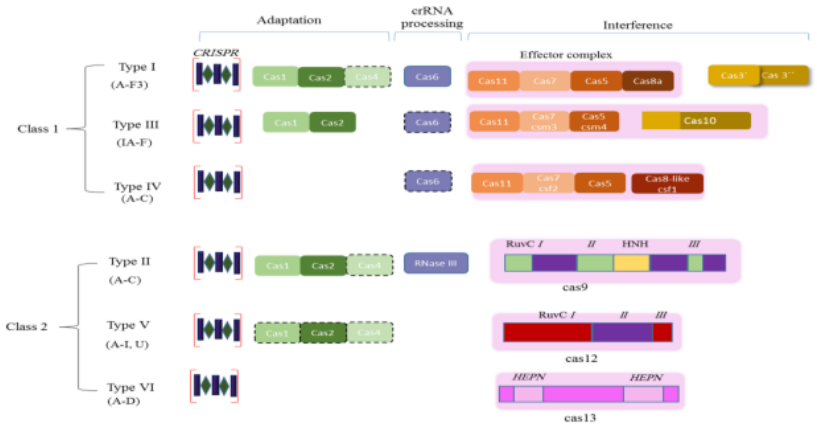


Figure 1.2: Class 1 and Class 2 CRISPR systems

Characteristics, types and subtypes of Class 1 and Class 2 CRISPR systems. The Class 1 is composed of multiple sub-unit Cas proteins, while the Class 2 is based on a single, large, multidomain enzyme. Each type is characterised by a signature endonuclease protein, such as Cas9 for type II of Class II and Cas10 for type III of Class I. Other differences are the target of these enzymes, which could be DNA or RNA, based on the Cas protein, and if they require a tracrRNA for function.

Source: CRISPR-Based Diagnosis of Infectious and Non-infectious Diseases, Vangah et al., *Biological Procedures Online*, 2020

1.1.2 DNA repair pathways in eukaryotic cells

When a nuclease cleaves in correspondence with a specific site in the genome, a double-strand break (DSB) occurs; in eukaryotic cells two pathways efficiently could repair the break: homology-directed repair (HDR) or non-homologous end joining (NHEJ). Taking advantage of these natural mechanisms, their association with engineered nucleases could be used to alter targeted genome.

If the break is resolved via NHEJ repair, the two broken ends are ligated together, and occasional gain or loss of genetic information could occur at the cleavage site. Consequently, NHEJ could be used to disrupt or abolish the function of a gene via introduction of small insertions or deletions at the DSB.

Alternatively, a gene function could be corrected, or a new gene could be inserted at the site of break when the DSB is repaired via HDR. In this case, a designed donor DNA template is required, which has to anneal with homology-based process along with the locus. The information encoded on the donor template could then be used to repair the break, adding specific nucleotides at that specific site in the genome (figure 1.3). To result in a gene correction, required nucleotides, derived from the donor template, could be transferred at the site of break; it could be a single change, or a short heterologous sequence could be inserted. Furthermore, also gene addition could occur via HDR: in this case, the donor has to carry an open reading frame (ORF) or a transgene at the break position. Then, that specific nucleotide sequence will be inserted in the chromosome, resulting in the addition of a new gene in that specific location of the genome (Urnov et al., 2010).

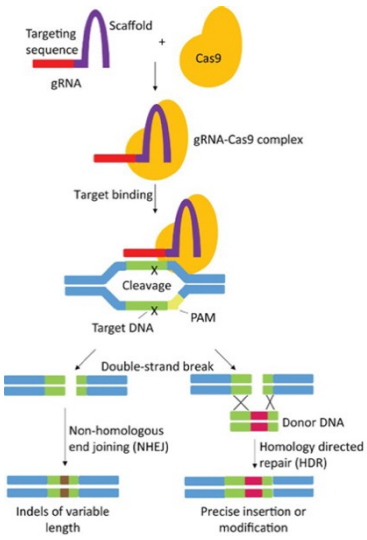


Figure 1.3: Repair mechanisms which could be associated with CRISPR-Cas systems

NHEJ and HDR are two cellular pathways which could be used to insert a specific mutation at the DSB. They occur after the nuclease cleaves the DNA and drive the alteration to a specific site. Several scenarios could happen, based on the kind of mechanism which is induced.

1.2 Nuclease proteins

Numerous nucleases have been engineered so far, in order to induce genetic modifications in various organisms. Taking advantage of their ability to bind and cleave the DNA, these proteins are revolutioning the genome editing in the last decade.

1.2.1 ZFNs

Zinc Finger Nucleases (ZFNs) were the first endonucleases to be used in genetic engineering.
They are artificial DNA-binding proteins that have been developed to edit the genome by inducing double-strand break in the DNA at a selected location. ZFNs have two functional domains: a DNA-binding and a DNA-cleaving domain (Urnov et al., 2010; Gaj et al., 2013). The DNA-binding one consists of a tandem array of Cysteine-Histidine (Cys_2-His_2) fingers, each recognising approximately a unique triplet (3 bp sequence) of DNA, with varying levels of selectivity. Therefore, each zinc-finger domain contacts 3 bp in the major groove of DNA. To generate a ZFN with specificity for a determined sequence, fingers binding the triplets which compose that specific motif are linked together.
The DNA-cleaving domain includes the nuclease domain of *FokI* restriction enzyme. *FokI* must dimerise to cleave the DNA, therefore the binding with the target site have to occur in the correct orientation and with appropriate spacing, in order to allow dimer formation, otherwise the DNA cleavage is not permitted (Urnov et al., 2010; Gaj et al., 2013).
Based on this modular design, ZFNs could be engineered to target a variety of chosen genomic sequence, inserting fingers specific for recognising required triplets (Gupta et al., 2012).
Because the nucleases to which they are attached only function as dimers, pairs of ZFNs are required to target any specific locus: one that recognizes the sequence upstream and the other that recognizes the sequence downstream of the site to be modified.
The association of engineered nucleases, inducing a double-strand break (DSB), such as ZFNs, with natural mechanisms that occur in eukaryotic cells and are able to repair induced by the nuclease (figure 1.4).

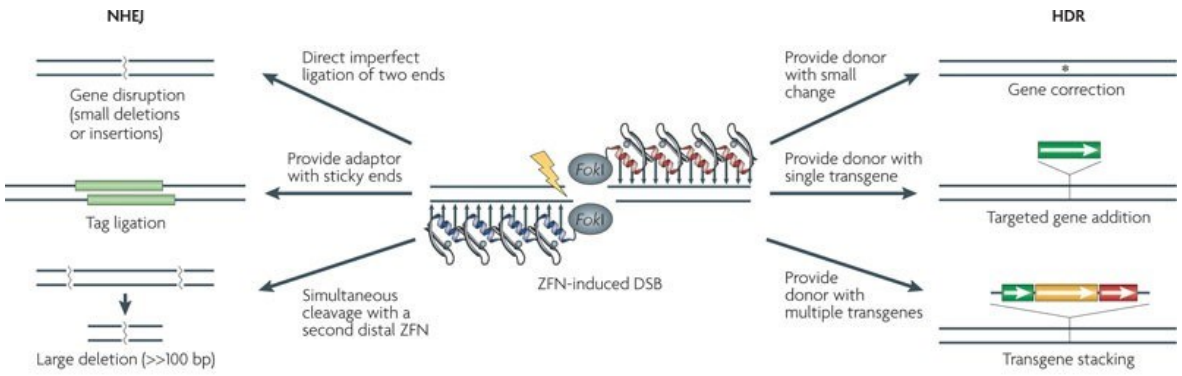


Figure 1.4: Genome editing by using ZFNs associated with via NHEJ and HDR repairs

When ZFNs are combined with natural repair mechanisms in eukaryotic cells, this association leads to events which could edit the genome at a designed location. ZFNs could be engineered to induce DSB at a specific site in the genome. Then, the association of these endonucleases with endogenous repair mechanisms, such as NHEJ and HDR, determines various scenarios: NHEJ repair could lead to gene disruption or abolishment via introduction of insertions at the DSB. On the other side, HDR repair could correct a gene function or add a new gene at the site of break by inserting a donor DNA template in that designed location of the chromosome.

Source: Urnov et al., Genome editing with engineered zinc finger nucleases, Nature Reviews, 2010

However, despite its application in the past, widespread use of ZFNs has significantly decreased. ZNFs have been designed to recognize all of the 64 possible codons of the human genome, but its application is still time consuming and has many limitations. For example, to recognize any defined sequence of DNA triplets, the assembly of libraries of different zinc finger moieties is required; pairs of ZFNs are mandatory to target any specific locus and these nucleases necessitate published finger archives to evaluate their affinity for DNA sequences. Moreover, these nucleases are often characterised by low efficiency and off-target activity. Finally, ZFNs only allow to target one site each time that should be at least 18 bp long, considering each individual zinc finger recognises three base pairs and a zinc-finger DNA-binding domain recognises at least nine base pairs. Consequently, this is a further disadvantage that limits the use of ZFNs to target any potential site in the genome.

1.2.2 TALENs

Transcription activator-like effector nucleases (TALENs) represent another tool used for gene editing. Similar to ZFNs, also TALENs induce DNA double-strand breaks which, stimulating NHEJ and

HDR repairs (figure 1.5), induces genetic modifications at a precise genomic location (Urnov et al., 2010).

TALENs are programmable DNA-binding proteins produced by plant pathogens in the genus *Xanthomonas*; once in the nucleus, TALENs bind specific DNA sequences and activate the gene expression in order to increase plant susceptibility o pathogens.

These proteins bind a specific sequence in the genome which contains tandem repeats of a 30-35 amino-acid motif; each repeat is highly conserved, except than two adjacent amino-acids, which constitute the repeat-variable di-residue (RVDs) (Gaj et al., 2013; Millet et al., 2010). These amino-acids specify the target; moreover, naturally occurring targets are preceded by a T, that is required for TALEN effector activity (Christian et al., 2010; Cermak et al., 2011). Other similarities with ZFNs are that also TALENs have a domain structure composed of amino-acid repeats, each of them recognising a single base pair, and include the same enzyme recognition site: *FokI*. Since this enzyme cleaves as a dimer, two nucleases have to bind the target site to enable *FokI* to dimerise and induce a DSB.

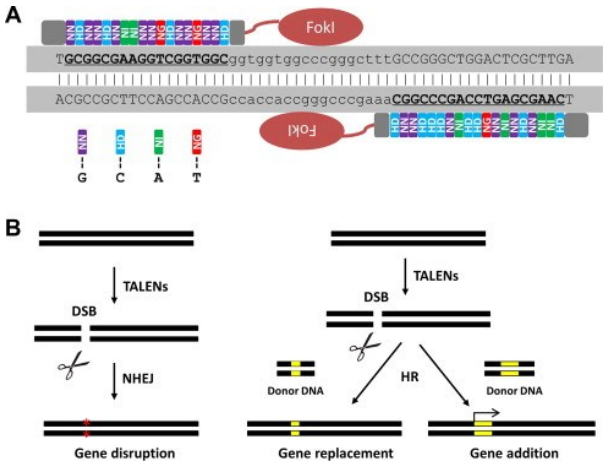


Figure 1.5: TALEN structure

A) TALEN repeats contain 30-35 amino acids that recognise a single base pair via the RVD sequence, which is related to a specific target site. Two monomeric TALENs are required to bind the target site to enable FokI to dimerize and cleave DNA; the only targeting limitation is that the binding site has to begin with a T. B) In common with ZFNs, TALENs share the modular structure, the restriction enzyme (*FokI*) and the association with endogenous repair mechanisms, such as NHEJ and HDR, to induce genetic modifications.

Source: Chen et al., An efficient TALEN mutagenesis system in rice, Methods, 2014

However, despite their ability to disrupt, restore or insert a gene function, TALEN application has decreased in the last few years. These nucleases are expensive, difficult to handle and time consuming, limiting their widespread. Moreover, TALENs are limited to simple mutations, they could only target a single site at one time, their application often results in low efficiency and mosaicism and these nucleases have a large size, making it harder to deliver and express them into cells, which is another clear disadvantage for their practical application (Gupta et al., 2014).
Therefore, in the field of genetic engineering, new endonucleases are currently being using more than ZFNs and TALENs.

1.2.3 Cas9s

Taking advantage of their ability to cleave and modify the DNA, the application of CRISPR-Cas system in genome editing research is rapidly rising. When this natural tool is transferred into more complex organisms than bacteria and archaea, such as humans, animal cells or embryos, it could have important functions beyond immunity: gene regulation or DNA repair, gene therapy, drug delivery, control of vector-borne diseases.
CRISPR endonucleases could be used to introduce gene insertions or gene deletions through targeted DNA cleavage followed by NHEJ repair, could allow replacement of targeted sequences via HDR repair, could interfere with chromatin regulation, could be used to generate genetically modified organisms or, one of the most recent challenges, the use of CRISPR to adapt genome editing for treatment of human diseases (Pickar-Oliver et al., 2019; Barrangou, 2013; Wang et al., 2020).
In order to expand the applications of CRISPR system, the search for more Cas proteins is needed. The wild-type endonuclease derived from *Streptococcus pyogenes* (SpCas9) has been well characterised and, based on its mechanism of action, it cannot be used for many applications. In fact, this protein always requires a specific PAM motif (5'-NGG-3') located downstream its target site. Therefore, all those positions in the genome that are not associated with this PAM cannot be cleaved by the SpCas9, limiting significantly the range of applications of this nuclease. Consequently, the search for alternative Cas proteins that are able to target a broader range of sequences, ideally PAM-less, is required.

However, prior to any application, the specific CRISPR-Cas locus needs to be characterised and the first step includes the identification of the *cas* genes and the CRISPR array.

Nowadays, several computational pipelines have been developed to search for short palindromic repeat sequences which could represent a putative CRISPR locus. Most of these tools are based on sequence alignments and search for similarity, conservation and lengths of the repeats.

Additional tools allow also to characterise the crRNA:tracrRNA duplex (Hidalgo-Cantabrana et al., 2019), first deriving the sequences and the structures of both RNAs and then predicting their interaction. The main approaches to derive the gRNA sequences are: the pattern recognition genre, that searches for candidates comprising a PAM and around 20 bp that are coplemetary via base-pairing to the target sequence; the feature rule genre, which integrate key features that influence the target efficiency and could be used to screen among different gRNA candidates. Analysed features could be G/C content (it should not be too high or too low), poly-T sequences (typical terminator for gRNAs), composition of bases involved in Cas binding, exon position (gRNA targeting earlier exons is preferred). Another approach is the machine learning genre, that integrates algorithms for further refining more precise gRNA candidates (figure 1.6) (Zhang et al., 2020).

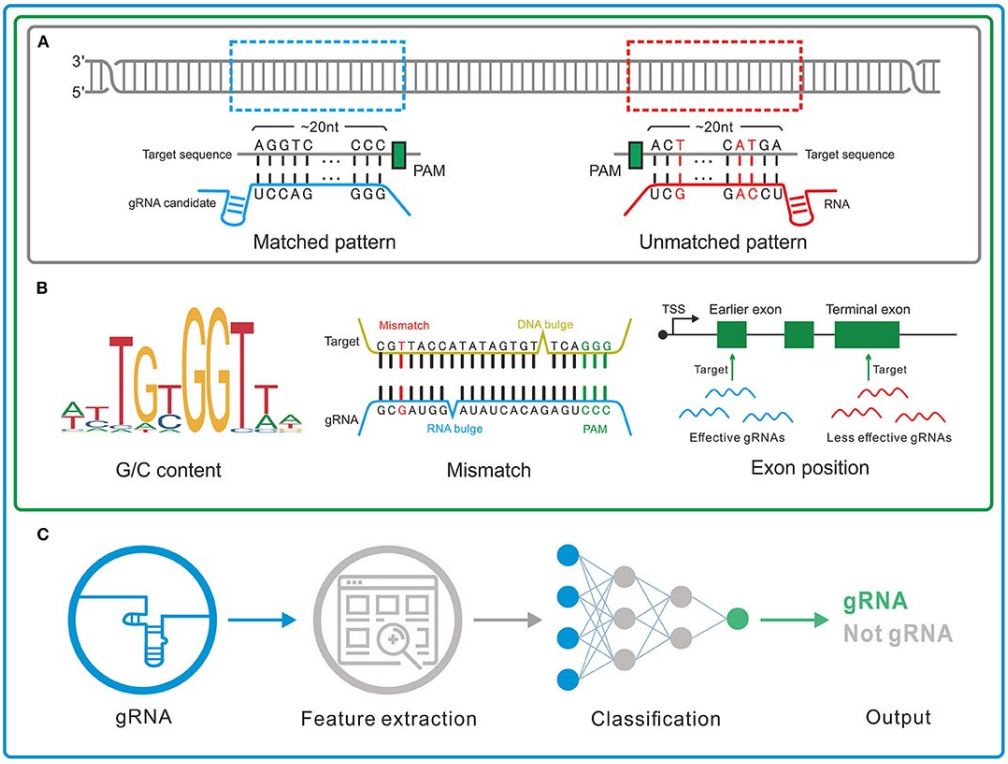


Figure 1.6: Different gRNA designers

A) Pattern recognition genre, based on base-pair complementarity between the gRNA candidate and the target sequence; B) feature rule gentre: it analyses further information related to the gRNA, such as G/C content, mismatches, exon position, base composition; C) machine learning genre, which is based on algorithms that, combined with the other features, define more precise gRNA candidates.

Source: Zhang et al., *In silico* Method in CRISPR/Cas System: An Expedite and Powerful Booster, *Frontiers in Oncology*, 2020

The discovery of new PAMs via computational tools is fundamental for expanding CRISPR applications, because it allows to overcome the requirement of the canonical NGG sequence and, therefore, target a wider range of sites. The PAM is a limiting step in cutting unique loci.

The spacers included in the CRISPR locus of bacteria are used for PAM prediction based on spacer-protospacer *in-silico* identification and the recognition of the conserved sequence flanking the protospacer. In details, there is a search for matches between these spacers and sequence databases of bacteriophages or plasmids. Via alignment, the putative PAMs could then be identified, using algorithms which represent the conservation of the nucleotide at each position, such as WebLogo. This tool generates a logo reporting a stack of letters for each position in the sequence and the height of each letter reflects the conservation of that base in that position (Karvelis et al., 2017).

Thanks to the joined collaboration with bioinformatical analysis, the CRISPR-Cas field is significantly expanding, especially regarding the engineering of new Cas proteins and the discovery of new PAM variants.

1.2.3.1 Cas9 structure and function

In Class II CRISPR-Cas systems, the endonuclease Cas9, driven by a gRNA duplex including crRNAs and tracrRNAs, induces a DSB in correspondence of a specific target site. However, prior to the interaction with the DNA and the cleavage event, some mandatory stages are required. This is necessary to convert the protein from its inactive status into the active one.

One of the first characterised Cas9 was the one derived from *Streptococcus pyogenes* (SpCas9).

Its structure consists of two lobes: recognition lobe (REC) and nuclease lobe (NUC). The REC lobe includes three regions: a α helix (referred also as the bridge helix), the REC1, which is responsible for binding the DNA, and the REC2 domains; the NUC lobe can be divided into RuvC and HNH, which are needed for the cleavage, and PAM interacting (PI) domain, that interacts with the PAM sequence of the target DNA (Nishimasu et al., 2014; Jiang et al., 2016). As for the SpCas9, the PAM is always the canonical 5'-NGG-3' motif.

Crystallography studies showed as the gRNA:target DNA duplex, negatively charged, interacts with the positively charged groove at the interface between REC and NUC lobes (Nishimasu et al., 2014). Therefore, that is the junction where all the components of the CRISPR complex, the Cas9, the gRNA and the target site, converge (figure 1.7). The interaction occurs at the genome location where the DNA sequence shows complementarity with the gRNA and the recognition of the PI leads to a DSB 3 bp upstream the PAM motif.

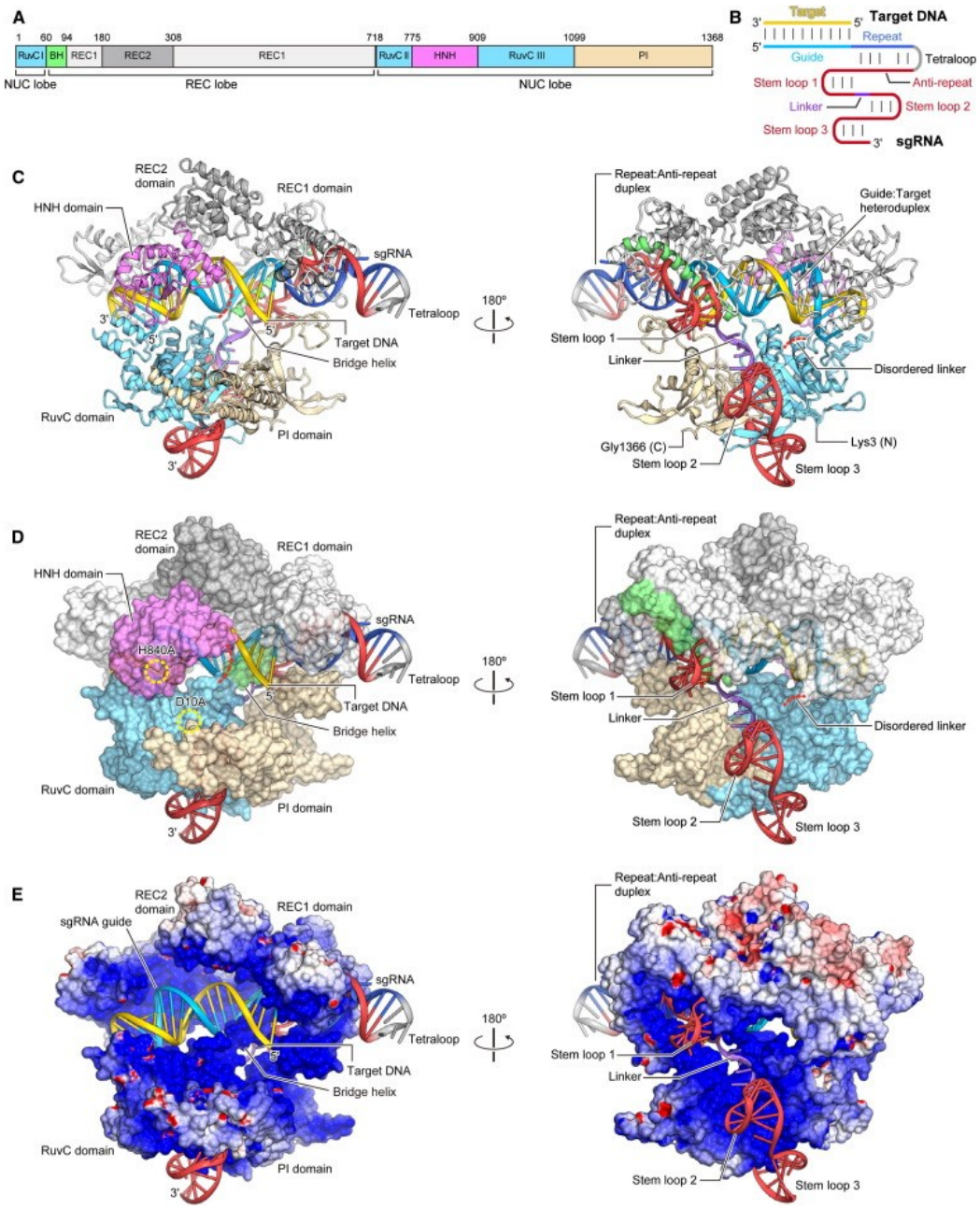


Figure 1.7: Crystal structure of the SpCas9 in complex with gRNA and target DNA

A) Domains constituting the SpCas9 protein; B) Representation of the gRNA:target DNA complex; C)Representation of the Cas9:gRNA:target DNA complex; D) Surface representation of the Cas9:gRNA:target DNA complex; E) Electrostatic surface potential of Cas9. Molecular graphic images were prepared using CueMol.

Source: Depardieu et al., Gene silencing with CRISPRi in bacteria and optimization of dCas9 expression level, Methods, 2020

The interaction of the Cas9 with the gRNA induces a conformational rearrangement of the protein structure, fundamental to transform the endonuclease into its active form, that is required to cleave the DNA. Only after these events the Cas9 could interact with the target site. In particular, the

rearrangement is induced by the interaction of REC1 with the guide:target heteroduplex; therefore, this domain is critical for the Cas9 function.

Analysing the whole process, the gRNA recognition by the Cas9 leads to the formation of the Cas9:gRNA binary complex; this interaction induces the conformational change of the endonuclease into its active form (Nishimasu et al, 2014). Afterwards, the PI domain recognises the PAM sequence on the target DNA: this is fundamental for the enzyme to avoid attempts with many incorrect sequences. If the Cas9 finds the appropriate PAM, it will interact with the remaining DNA on the target, in order to assess whether the residual motif is complementary to its guide sequence. Upon target binding, the Cas9 undergoes a second conformational change, which leads to a R-loop formation. In this position the nuclease is finally positioned to cleave the opposite strand of the target DNA (Lim et al., 2016; Jinek et al., 2014, Jiang et al., 2016).

Some amino acids in the PI domain are crucial for PAM recognition and interaction. In particular, two arginine residues (R1333 and R1335) interact with the major groove of the guanine bases of the PAM (NGG motif for SpCas9); moreover, and a lysine (K1107) and a serine (S1109) residues in the phosphate lock loop (figure 1.8), which is also included in the PI domain, interact with the minor groove. At this point, the phosphate lock loop could form hydrogen bonds with the phosphate in position +1 of the PAM, stabilising the DNA. Therefore, the target sequence rotates upwards the gRNA (Anders et al., 2014).

Once the target site is stabilised in this position, the Cas9 HNH and RuvC domains could cleave the DNA template. Specifically, HNH cleaves the complementary strand, whereas RuvC the non-complementary one, leading to a DSB at that specific site (∼ 3-4 nucleotides upstream the PAM sequence) (Jiang et al., 2017).

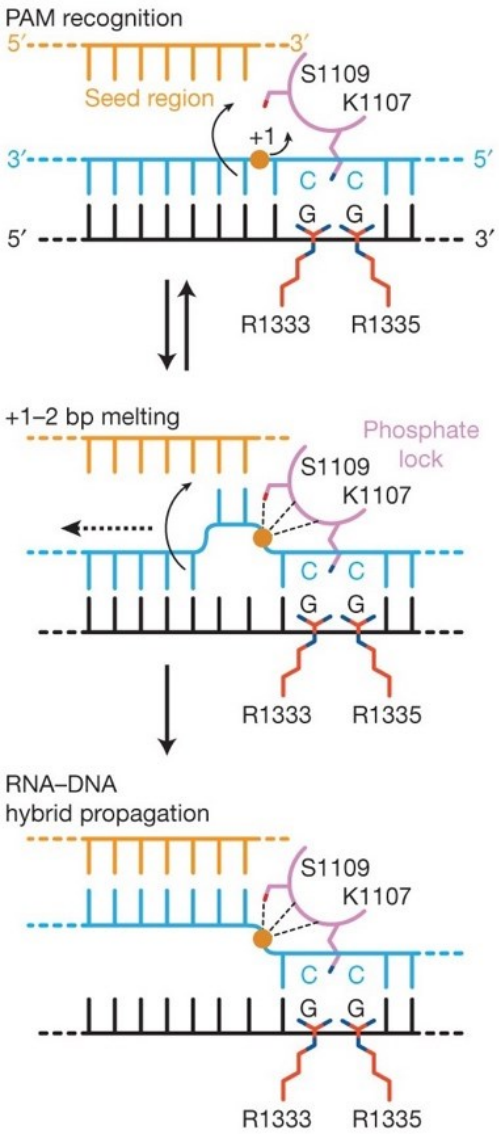


Figure 1.8: PAM recognition and interaction

Interaction between the Cas9 and the gRNA at PAM binding site. R1333 and R1335 of the PI domain interact with the G sites of the PAM; S1109 and K1107 in the phosphate look (also in the PI domain) interact with the minor groove, leading to hydrogen bonds which stabilise the DNA template. At this point, the Cas9:target:gRNA complex has the correct conformation for the nuclease stage. Source: bibliography n.25

After the nuclease has cleaved the DNA, cellular repair mechanisms, which naturally happen, could be used to introduce a mutation at that location in the genome. Like ZFN and TALEN nucleases (above sections 1.1.2.1 and 1.1.2.2), also the Cas9 activity could be associated to HDR and NHEJ pathways (figure 1.3).

1.2.3.2 Engineered Cas9 in order to broaden PAM specificity

In the last few years CRISP-Cas9 application is significantly increasing in different fields.

Theoretically, this tool could be designed to cleave any target sequence, provided that a specific PAM is located downstream the target site.

In SpCas9 the canonical PAM is 5'-NGG-3' and this represents the only constraint about CRISPR-Cas9 widespread, considering this sequence is mandatory for the target site recognition by the endonuclease (Siksnys et al., 2016).

The determination of alternative PAMs begins with *in-silico* sequence extrapolations via bioinformatic analysis (Mojica et al., 2009; Karvelis et al., 2017), that develop nucleotide prediction at each position of the sequence landscape. Finally, methods such as WebLogo (Crooks et al., 2004) or PAM Weels could be used to represent derived PAM variants. Then, in order to assess the nuclease activity and PAM preferences for selected Cas9, *in-vivo* or *in-vitro* cleavage assays are performed. Generally, in addition to the screening of previously suggested PAMs, the analysis could include also randomised libraries of DNA sequence (Chatterjee et al., 2004; Karvelis et al., 2015), beneficial to increase the possibility to determine new potential motives.

In order to broaden the PAM specificity, various Cas9 mutants have been engineered so far, introducing specific mutations in the protein sequence. However, none of these have shown absence of PAM specificity yet; the only improvement was the expansion of the canonical PAM motif.

Example of SpCas9 orthologues engineered with point mutations which expanded PAM specificity are VQR (D1135V/R1335Q/T1337R) and EQR (D1135E/R1335Q/T1337R) variants, which recognise 5'-NGAN-3' and 5'-NGNG-3' PAMs, while the VRER variant (D1135V/G1218R/R1335E/T1337R) is specific for the 5'-NGCG-3' motif (Anders et al., 2016; Kleinstiver et al., 2015). All these Cas9s have in common the T1337R substitution. Other engineered variants were the high-fidelity SpCas9-HF1 (N497A/R661A/Q695A/Q926A), SpCas9-HF2 (with an additional D1135E substitution), SpCas9-HF3 (harbouring additional L169A) and SpCas9-HF4 (with additional Y450A substitution) (Kleinstiver et al., 2016). These SpCas9 mutants have mutations at positions which are believed to be involved in non-specific hydrophobic interactions with the target DNA on its PAM location (the Y450 residue is notable for participating in a base pair interaction with the sgRNA). All these high-fidelity variants retained 70% or more cleavage activity of wild-type (wt) SpCas9 when tested with the canonical 5'-NGG-3' PAM. Therefore, this study demonstrates that mutating non-specific DNA nucleotides could

be highly effective and could increase SpCas9 specificity. Furthermore, other variants were developed, capable of recognising NRN (where R is A or G) PAM compared to the canonical NGG motif (8). In particular, SpG (L1111R/A1322R) derivate and other mutants containing additional A61R, N1317R and R1333P substitutions, enable targeting of sites with NRN PAMs. Other SpCas9 mutants were engineered in order to induce loss of the base-specific interaction with the third G in the NGG motif; the new introduced non-base specific interactions enabled the recognition of more relaxed NG PAMs (9). In particular, the VRVRFRR variant, with R1335V/L1111R/D1135V/G1218R/E1219F/A1322R/T1337R substitutions, enhance the recognition of TGA, TGT and TGC PAMs. Among SpCas9 mutants which have been engineered to broaden PAM specificity there is also the xCas9 3.7 (Hu et al., 2018). This variant is characterised by several point mutations: A262T/R324L/S409I/E480K/E543D/M694I/E1219V and allow the cleavage also with NG, GAA, GAT PAMs.

Despite the fact that these SpCas9 variants do not remove the limitation of the PAM requirement, they demonstrate that the design of new orthologues could allow the DNA cleavage also with non-canonical PAM sequences, supporting the application of CRISPR-Cas tool to a wider range of targets.

1.2.4 Cas12a (formerly Cpf1)

CRISPR-Cas12a (formerly Cpf1) is a Class 2 CRISPR system, the same as the well-characterised Cas9. However, these two systems differ for some aspects (Tang et al., 2017; Zetsche et al., 2015).

Cas12a requires a thymidine-rich PAM sequence (5'-TTTN-3'), which expands the range of sequences that could be recognised beyond the guanosine-rich PAM required by the SpCas9; Cas12a induces 5' staggered ends, that facilitate gene replacement using NHEJ; it cleaves at sites distal to the PAM; it is characterised by repetitive cleavage which promotes also large fragment deletions and this multiplexed genome editing works using a single customised CRISPR array (Zetsche et al., 2016). Furthermore, Cas12a needs one single crRNA (versus the crRNAs and tracrRNAs required by the Cas9) (figure 1.9) that is much shorter than the one required by the Cas9 (Tang et al., 2017; Zetsche et al., 2015). This characteristic promotes its easier packaging (Hur et al., 2016) to be delivered where required.

Moreover, CRISPR-Cas12a system could easily be applied *in-vitro* or in living cells, which supports its development as attractive tool for genome editing, which currently finds application especially in plants.

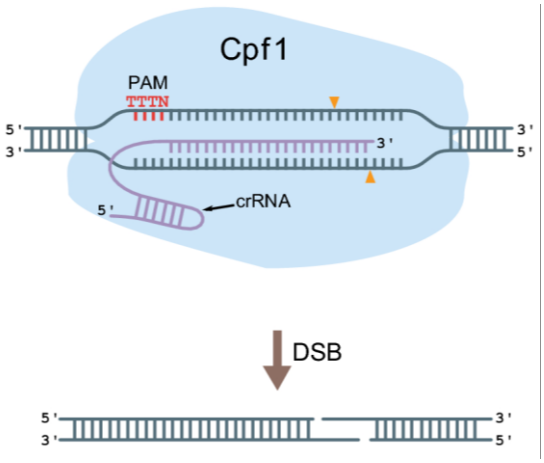


Figure 1.9: CRISPR-Cas12a (Cpf1) system

The CRISPR-Cas12a system is characterised by the Cas12a (formerly Cpf1) protein that induces staggered cut at the target site; furthermore, this protein requires a TTN PAM, which expands the targeting beyond the NGG sequence needed by the Cas9 PAM. Moreover, Cas12a works in combination with a short single crRNA, it does not require the tracrRNA component as the Cas9.

Source: Vanegas et al., Cpf1 enables fast and efficient genome editing in Aspergilli, *Fungal Biology and Biotechnology*, 2019

The main limitation regarding Cas12a application is the same as the Cas9: the requirement of a specific PAM sequence, that for Cas12a is the canonical TTTN. Therefore, in order to expand the targeting range of CRISPR-Cas12a system, this endonuclease has been engineered, introducing specific mutations in the protein.

Referring as wild type sequence the one derived from the *Acidaminococcus sp. BV3L6* (AsCpf1), new variants have been derived, carrying S542R/K607R and S542R/K548V/N552R mutations and recognising TYCV and TATV PAMs respectively both *in-vitro* and in human cells (Gao et al., 2017).

Therefore, the approach of engineering new Cas12a variants with alternative sequences seem to be a good approach to increase the targeting range of this endonuclease.

1.3 CRISPR-Cas applications

CRISPR-Cas9 systems represent a powerful tool for gene editing thanks to the possibility to easily modify targeted sequences in various organisms, from plants to animals. Consequently, it could easily be applied in various research fields, such as biomedical research, therapeutics, nanotechnologies, bioengineering, precision medicine, food and agriculture (Barrangou et al., 2016; Hsu et al., 2014; Shao et al., 2016; Hendriks et al., 2020).

1.3.1 Human diseases

Genetic diseases represent one of the most studied branches for CRISPR-Cas application, considering currently approaches for the treatment of these diseases are more focused on reduction of symptoms rather that working on the biological causes.

Among this field, Duchenne muscular dystrophy (DMD) represents one of the diseases more likely to be treated via CRISPR-Cas9 (64) or CRISPR-Cpf1 (Zhang et al., 2017) applications. DMD is caused by a mutation that significantly alters skeletal function; therefore, CRISPR tools could be used to repair those sites responsible for the alteration and restore the normal protein function (Nelson et al., 2016; Bengtsson et al., 2017; Mollanoori et al., 2020).

Cystic fibrosis (CF) is another genetic disease that could potentially be treated via CRISPR application: like DMD, also CF is characterised by specific mutations which repress the translation of a protein responsible for skeletal functions. Therefore, if these mutations are corrected, muscular activity could be restored (Sanz et al., 2017; Hodges et al., 2019; Marangi et al., 2018; Schwank et al., 2013).

Cancer research represents one of the most challenging field regarding CRISPR application: this tool could be used to selectively modify those mutations that induce the growing of abnormal tissues, knocking out genes whom activation stimulates cancer development, or to destroy tumoral cells via apoptosis (Rosenblum et al., 2020). Regarding the delivery of CRISPR components to the target tissue, efficient approaches include the use of virus-like nanoparticle that co-deliver CRISPR components combined with small molecule drug. This association results in the synergistic regulation of multiple cancer-associated pathways (Liu et al., 2020).

A recent therapy that has been studied for some tumour is based on the use of mutated T cells using CRISPR-Cas9; in these patients, T cells are edited in a way that they enhance anti-tumour effects (Lu et al., 2020). However, because of possible off-targets events and other risky consequences that are still unknown, this approach is actually studied only for very severe cancers and exclusively for those patients that does not respond to other therapies.

Cancer treatment via CRISPR application could represent also an example of personalised medicine (Jiang et al., 2019; Huang et al., 2018; Xing et al., 2019), that in the future could lead to the possibility to target specific cancerous genes related to each patient, that represents the main purpose of precision medicine.

1.3.2 Vector-borne diseases

Another interesting field where CRISPR-Cas9 finds application regards vector-borne diseases; in this case, animal species responsible for the transmission could be genetically modified, in order to affect the spread of that disease.

For example, malaria is caused by the transmission of parasites to humans through the bites of infected *Anopheles gambiae* mosquitoes. In particular, only females of this species are responsible for the transmission; therefore, CRISPR-Cas9 could be applied as a strategy to control the spreading of this disease via genetic modification of its vector.

Following this approach, gene drive systems (Burt, 2003) have been developed: when the DSB induced by the Cas9 is repaired using homologous chromosome as donor template, the heterozygote is converted into a homozygote via “homing” process. Ongoing studies based on self-propagating gene drives are used to insert specific mutations in the genome of *A. gambiae* mosquitoes and, in particular, targeting specific genes responsible for sex differentiation or fertility. These two approaches lead to diverse scenarios among the progeny: female sterility phenotype (Hammond et al., 2016; Kyrou et al., 2018) or male bias (Galizi et al., 2016) and both, in large scale, aim at the population collapse due to lack or female their sterility. The proof-of-principle of this strategy has already been showed in laboratory, supporting the idea that gene drives could represent an important cost-effective technology that, combined with other countermeasures, such as vaccine and insecticides, could significantly contribute to the fight against malaria (Hammond

and Galizi, 2017). However, some limitations of this mechanism still need to be controlled, such as the development of mutations resistant to gene drives (Hammond et al., 2017) that could alter the whole system.

1.3.3 Food and agriculture

Gene drive systems could be designed also against harmful pests that are vectors for plant pathogens and cause worldwide economic losses (Fasulo et al., 2010; McFarlane et al., 2018). Just as the strategy adopted to fight malaria, also this approach aims to eliminate specific insect populations (that are plant parasites) targeting their sexual chromosomes. Moreover, other non-gene drive methods based on CRISPR application could be used to control or suppress devastating pest insects, inducing mutations that lead to knockout of targeted genes (Ward et al., 2010; Heu et al., 2020; Barakate et al., 2016; Ahmed et al., 2019; Douris et al., 2020). This strategy is important to confer protection against viruses and other parasites that in normal conditions would destroy the plants. Therefore, gene editing represents an alternative to standard strategies, such as the release of chemical pesticides, introducing a new cost-effective approach against insect pests.

Another area of interest where CRISPR tool could be used in agriculture regards the improvement of crop production. In fact, because of many factors, such as increase of world population, spread of insect pests, climate change, the global request of food is significantly increasing. However, hostile geographic area where plantations are set and social-economic aspects in those places often affect quality and yield production. Given this background, genetic modifications could represent a strategy for the insertion of specific characteristics in the plants, that support their growth even in very drastic conditions. For example, CRISPR could be used for generating genetically modified cereals or vegetables, letting their growth in tropical regions, where very hot temperatures and harsh would affect the production in normal circumstances. Therefore, these engineered plants are stronger and more resistant. Moreover, another interesting application of CRISPR in agriculture regards the possibility to increase nutritional values of some crops (Su et al., 2019) generating genetically modified organisms (GMO), mainly cereals. This may represent a future strategy against food shortage in poor countries.

1.4 Gene regulation via CRISPR-Cas system

The secondary structure of the CRISPR-Cas complex, which includes the Cas9, the gRNA and the target, is crucial to allow the DNA cleavage: only if the three parts have a specific conformation the nuclease could induce the DSB. Therefore, changes in this structure or the addition of effector/silencer molecules could alter the Cas9 activation, affecting the whole system.
The Cas9 and the RNA components of CRISPR system could specifically be engineered to interfere with the nuclease activity, leading to a programmable inactivation or activation of the protein, with consequent effects on gene regulation. Therefore, CRISPR could be used as a flexible DNA binding platform for the recruitment of proteins or RNAs that have a key role in the targeting of genes for expression or repression, with a resulting regulation of related functions (Gilbert et al., 2013).

1.4.1 Catalytically inactive Cas9 (dCas9)

Among these approaches, one is based on altering the transcription process using a different endonuclease: the catalytically dead Cas9 (dCas9). This protein is a Cas9 with point mutations in its nuclease domain which remove the endonuclease activity. Essentially, the dCas9 still keeps the ability to interact and bind the target DNA, just cannot cleave it.
Therefore, taking advantage of its ability to bind other factors, when the dCas9 is combined with other molecules which could alter the transcription process (such as effectors or repressors), the CRISPR-dCas9 system is becoming a powerful tool about gene regulation (Dong et al., 2017).
CRISPR interference (CRISPRi) is a method which could repress the expression of targeted genes in host cells (bacterial or mammalian) and it is based on dCas9 application (Zhao et al., 2017; Depardieu et al., 2020). It is highly specific, simple and cost-effective compared with other methods.
When the dCas9 protein is expressed together with a sgRNA, as canonical CRISPR systems, recognises and interacts with a target sequence. However, the resulting dCas9:sgRNA:DNA target complex blocks that specific targeted site and the RNA polymerase cannot bind the promoter of the targeted gene or move forward on the DNA to continue the transcription; therefore, the cleavage could not happen (figure 1.10).

Consequently, this obstruction interferes with all the process required for gene expression, such as transcriptional elongation, RNA polymerase binding or transcription factor binding, and the downstream gene is silenced (Peters et al., 2016, Qi et al., 2013). At the same time, non-targeted genes are minimally impacted.

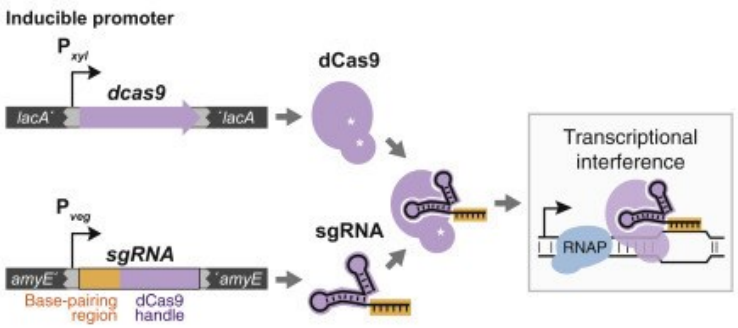


Figure 1.10: CRISPRi process

CRISPRi system is activated when a dCas9 is expressed together with a sgRNA and interacts with a targeted site; this binding leads to the dCas9:sgRNA:target complex formation, which constitutes an obstruction to the transcription process. Consequently, the downstream gene expression is blocked, and that specific gene silenced.

Source: Peters et al., A Comprehensive, CRISPR-based Functional Analysis of Essential Genes in Bacteria, *Cell*, 2016

However, the effective knockout of an essential gene is not permanent: the silencing effect of dCas9:sgRNA:DNA target can be induced and reversed. For example, the obstruction could be released in presence of an inducible promoter, such as the tetracycline (Tc)-inducible promoter (Rodriguez-Garcia et al., 2005). In presence of the inducer, this promoter is activated and, consequently, it could allow the expression of the gene which was previously silenced; in this way its transcription can continue and the CRISPR system is reverted.

Moreover, it was demonstrated that the CRISPRi system is also able to simultaneously suppress the transcription level of multiple genes in *S. aureus* (Dong et al., 2017; Zhao et al., 2017); this means that diverse knockout strains could be downregulated, significantly increasing the application of this tool.

1.4.2 RNA molecules

In addition to protein-based tools, other RNA-mediated regulatory mechanisms could be applied for gene regulation. Non-coding RNA (ncRNAs) molecules are considered key regulators for the modulation of transcription and translation of targeted genes, due to their ability to catalyse biochemical reactions, control protein activity. Their mechanism is based on Watson-Crick base pairing via intramolecular and intermolecular interactions with the target DNA, mRNA or other ncRNAs. Diverse classes of ncRNAs have been engineered so far: antisense RNA; small transcription activating RNA; toehold switch; CRISPRi and activation; riboswitch, aptamer and aptazyme; non-coding RNA (Lee et al., 2018). All of these systems could be directed to regulate the expression of a selected gene, altering its downstream function.

1.4.2.1 Riboswitches

In addition to protein-based tools, such as CRISPRi, other approaches are recently evolving. For example, in the field of synthetic biology, is emerging the development of designable genetic parts which, with their behaviour, interfere with the regulation of diverse cellular function.

RNA components, such as riboswitches, synthetic or natural small RNA molecules, could be programmed to promote or repress the expression of a targeted gene, based on their conformation. Therefore, these mRNAs could be designed to target any sequence in the genome, supporting a multitude of design and applications for gene editing (Galizi and Jaramillo 2019). Moreover, a significant benefit of their application is that RNAs are able to modulate chromosomal gene expression without modifying the chromosomal sequences.

Natural riboswitches contain both sensors and expression platform but could be re-engineered to allow the binding with synthetic triggers.

In the past few years, numerous synthetic riboswitches have been engineered, incorporating the aptamers in their structure; this allow the ligand bind which is mandatory to have the consequent conformational change to trigger the expression of a downstream gene. Moreover, RNA aptamer could be also modified to broaden the range of possible ligands and, consequently, increase the number of genes which could be regulated (Jang et al., 2018).

Example of synthetic riboswitches are the RNA-interacting guide RNAs (igRNAs), that are activated after binding a single-stranded trigger RNAs (trRNAs), another mRNA molecule that acts as a ligand (Galizi and Jaramillo 2019). In non-binding conditions, the igRNA is in its inactive status: on its sequence it carries the designed target site, but this is not accessible to the RNA polymerase because of the molecule conformation. Then, when the sensor igRNAs interacts with the trigger via base pair complementarity, there is a duplex formation, which induces a conformational change of the igRNA and turns it into its active status. Finally, the new released igRNA structure exposes the target site that becomes accessible to the RNA polymerase (figure 1.11) and the transcription could continue. Therefore, the sensor:trigger binding could significantly regulate the expression of a targeted gene, based on the association of this complex with a Cas9 or dCas9 (Galizi et al., 2020). Moreover, the lack of direct interaction between the sensor and the target DNA determines an allosteric regulation of the CRISPR system: in fact, the trigger interferes only with specific sequences of the igRNA, which non include the target site.

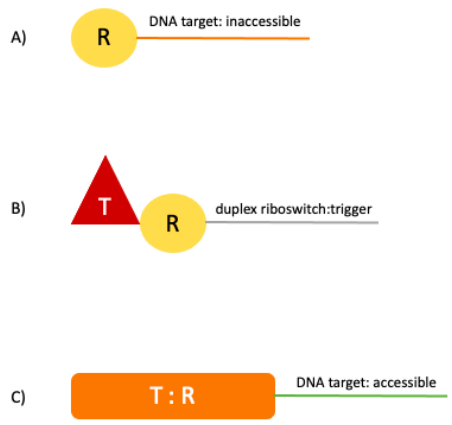


Figure 1.11: Representation of the riboswitch into its inactive and active status (after hybridization with the trigger)

The riboswitch (R) is activated after interaction with the trigger (T) via sequence hybridization. A) R in its inactive status: the target site, that is located on the R, is inaccessible to the RNA polymerase, because of the molecule conformation. B) Binding sequences of the R are complementary with the T. Once these two RNA components bind in a duplex, there is a conformational change of the R that is converted into its active status (released). C) When R is in its active status, the target site is accessible to the RNA polymerase, because the R conformation is released, allowing the transcription and inducing the gene expression.

Therefore, riboswitches represent powerful genetic editors which could selectively activate or inactivate the whole CRISPR system via induced conformational change of the complex.

1.4.2.2. Small RNAs (sRNAs)

Other gene regulators include bacterial small RNAs (sRNAs). These are antisense RNA molecules which function via RNA-RNA interaction with a sense RNA sequence included in the target mRNA. This interaction leads to a conformational change of the target site which affects the whole CRISPR system, based on the consequent process which could be performed. (Na et al., 2013; Chappell et al., 2015). For example, in case of translation, the target site rearrangement could expose ribosome binding sites (RBSs) while, in case of transcription, terminator hairpins.

Some sRNAs which repress the transcription, called attenuators, stop the transcriptional cascade generating gates with their own structure that obstruct the biochemical process, resulting in the suppression of downstream gene expression. On the contrary, sRNAs promoting the transcription are called sRNA transcriptional activators (STARs) and thy are able to activate the target site, inducing further transcription and translation.

Therefore, the possibility to use sgRNAs for different purposes, such as the induction or repression of gene expression, demonstrate the versatility of this class of RNA regulators.

1.5 CRISPR delivery into cells

The delivery method of the CRISPR components inside the cell influences their efficiency; Cas9s and gRNAs could be introduced in target cells and organs via viral or nonviral systems, depending on the desired application (Song, 2017; Wang et al., 2016; Liu et al., 2020).

Viral systems, such as adenovirus and lentivirus, are based on the viral infection of cells via endocytosis and the consequent release of the viral genome inside the cells. Viral delivery could work even in complicated *in-vivo* conditions thanks to the strength of viral-cell interactions.

Nonviral systems include several methods, such as electroporation, microinjections, lipidic nanoparticles, polymers, which deliver CRISPR components either as mRNA or as a complex of Cas9 protein and ribonucleoprotein (RNPs). However, proteins are rapidly degraded inside the cells, therefore RNPs are associated with reduced off-target effect to the DNA.

Methods based on the use of human or animal cells represent one of the most widespread techniques that have been used for CRISPR-Cas testing so far. In addition, there are also *in-vivo* procedures, which involve embryos or living animals, that allow the evaluation of the CRISPR complex activity directly in a specific living organism.

Using these diverse methods, the CRISPR components could be transported in the cytoplasm (via electroporation) or in the nucleus (via microinjection). In particular, to have the translocation into the nucleus, designed Cas9 constructs should include in their structure the nuclear localisation sequence (NLS) to ensure the component could enter the nucleus for genome editing.

Electroporation (Xu et al., 2018; Pawluk et al., 2016) is a technique widely used to transfer genetic material inside the cells, taking advantage of frequent division of immortalised cell lines (such as HEK293, HeLa, A549, Jurkat) that increase the possibility of CRISPR components to enter the nucleus. This method uses viral or non-viral reagents and is based on short pulses of high-voltage electric fields that allow the introduction inside the cell of desired molecules. The electric pulse induces the formation of transient pores in the phospholipidic bilayer of cells through with the components can be introduced (figure 1.12).

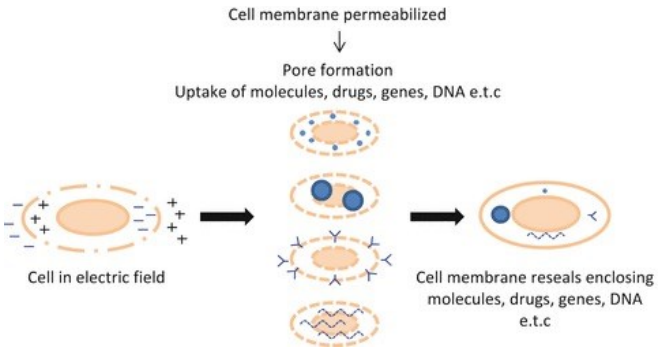


Figure 1.12: DNA integration by using cells: electroporation method

DNA encoding for the Cas9 and the gRNA could be integrated inside the genome of an electrocompetent cell using electroporation. This technique is based on the use of electric pulses that generate charge difference at cell's membrane and temporary increase the permeability of the bilayer; this leads to the entrance of negatively charged molecules, such as the DNA, inside the cell.

Source: Kalli et al., *Introduction of Genes via Sonoporation and Electroporation*, 2014

Specific fluorescent genes are also usually included in the plasmids that enter the cell, in order to detect monitor the nuclease activity. Then, based on the type of genome editing that is expected, in particular if the CRISPR system is induced or repressed (for example, when testing dCas9 or when factors that regulate the transcription-translation are added), different scenarios could be observed. Generally, a decrement of living cells should be observed in those cells characterised by nuclease activity; on the contrary, if the genome editing is inhibited, the number of living cells should not be altered.

Moreover, transfection is another method that requires cells but is not instrument-based as the electroporation: in fact, it uses viral and non-viral reagents to transfer DNA or mRNA inside the cells. Its advantages are that transfections are easy to be performed, could be applied to various types of cells and do not require additional chemical components.

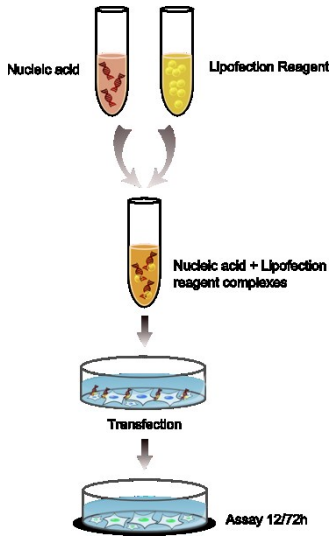


Figure 1.13: DNA integration by using cells: transfection method

Transfection method allows the entrance of DNA or RNA inside the cells using viral or non-viral (e.g., lipofection) reagents. It does not require specific instruments and could be applied to diverse cell lines, such as HEK293, Jurkat, HeLa).

At the same time, considering both cell electroporation and transfection, there are several disadvantages that limit the widespread of these techniques: transfection efficiency is very low in primary human cells; both reagents used for transfecting and electrochemical reactions induce significant cell toxicity, resulting in frequent very high mortality post transfection or post electric

pulses. Furthermore, another common limitation of these methods is that both require in parallel cell culture, that is expensive and time consuming.

Microinjection is another technique which allows the transfer of CRISPR components inside the nucleus or cytoplasm through a needle (Fuchs et al., 2012; Galizi et al., 2016) and it is mostly used in *in-vivo* experiments. In this way, the Cas9 and the gRNA could directly enter inside the embryo via the pore generated by the needle.

However, this technique is characterised by several limitation: after some time, embryos start becoming more resistant to injection due to strengthening of their surface; therefore, generation of germlines significantly decreases. It is a very low throughput method: in fact, it requires very skilled people to have a good efficiency, each cell must be injected individually, and the needle could damage the plasma or the nuclear membrane. Therefore, embryo injections are highly laborious and time consuming, limiting the scalability of the testing. Furthermore, this method requires in parallel maintenance of animals from which embryo derives, in addition to maintenance of transgenics.

Consequently, considering all these limitations of current methods used for CRISPR testing, it urges to find alternative ways that allow a scalable and rapid approach, in order to expand the application of this tool.

1.6 CRISPR-Cas testing: cell-free systems

Apart from above techniques, other *in-vitro* methods have been developed to evaluate the CRISPR activity without requiring cells or living organisms and, consequently, less time consuming.

Therefore, this approach is really useful at very early stages of an experiment when, for example, there is the evaluation of the best gRNA design for a specific Cas9 or the choice among different promoters. The possibility to rapidly test diverse candidates in a short time encourages the use of CRISPR tool in such more fields, expanding its application.

Common *in-vitro* techniques necessitate the previous synthesis and purification of all the CRISPR components (Rajagopalan et al., 2018), such as the endonuclease and the gRNA, resulting in extra time required to perform the whole CRISPR experiment.

Generally, this approach requires several steps: first of all, the lyophilised proteins have to be dissolved; then the digestion reaction has to be set, including the Cas9, the sgRNA and the targeting substrate and it could be performed via *in-vitro* reaction. Then, the digestion reaction could be analysed, generally via agarose gel electrophoresis (in this case, if the cleavage happens, two different bands are expected to appear) or fragment analyser system for a more advanced analysis of the generated fragments.

However, even if this *in-vitro* method does not require living cells or embryos, it still has some limitations, such as the previous and separate synthesis and purification of CRISPR components, before performing the actual reaction. Therefore, it is still expensive and time consuming, not ideal for experiments that have to be performed in large scale.

Starting from this analysis, in the last few years alternative approaches are being developed in order to increase the scalability of CRISPR reactions *in-vitro*.

Synthetic biology represents an interesting field that allows to engineer and reprogramme biological systems adding synthetic components, such as DNA or proteins, that could alter the standard behaviour of a specific complex and promote the desired one. Therefore, synthetic biology has provided a platform that could be applied in many contests at cellular and molecular levels, such as the regulation of gene networks or metabolomics (figure 1.14).

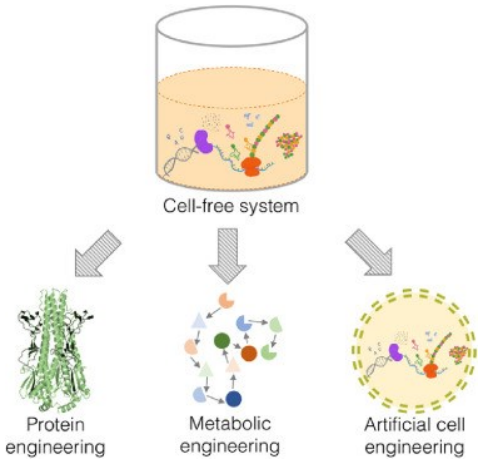


Figure 1.14: Applications of cell-free systems

Cell-free synthetic biology represents an interesting platform for synthesising biological parts that could have a key role at both cellular and molecular levels. Therefore, cell-free systems could be applied for cell, protein or metabolism engineering in order to modulate various biological pathways and functions without requiring living cells or organisms.

Source: Lu, Cell-free synthetic biology: Engineering in an open world, *Synthetic and Systems Biotechnology*, 2017

Cell-free systems allow to produce complex proteins, unnatural proteins, toxic proteins that cannot be synthetised via conventional cell-based approaches, expanding the number of libraries that could be analysed and supporting the discovery of novel enzymes or proteins that could have a key role in the control of biological pathways. Then, in addition to protein synthesis, cell-free systems could be used also for the regulation of various metabolisms, via production of desired substances that are required to activate a specific pathway rather than another one. Moreover, another challenge of these systems is artificial cell engineering. This method could be used to make synthetic cells via assemble of capsules that contain bioactive materials within a membrane. These artificial components could be designed and engineered to perform various functions; therefore, they may be useful in the future for studies related to environment, healthcare and biomaterials (Lu, 2017).

Cell-free systems are crude extracts derived from different organisms, such as bacteria, human cells, yeasts, and include all the components required for protein synthesis. Therefore, these systems include the RNA polymerase, that is the enzyme required to perform the transcription reaction, making RNA molecule from DNA template; buffer; salts; energy sources; ribosomes; amino acids (Dopp et al., 2019).

Possibly, if required, the final protein could undergo post-translational modifications that are necessary to perform its function, such as disulfide bonds to get a different conformation (figure 1.16).

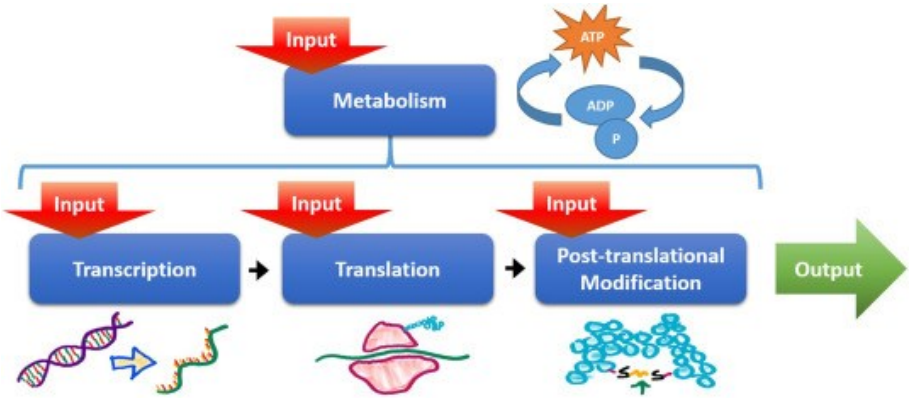


Figure 1.16: Reactions required for protein synthesis in a cell-free system

The addition of the DNA template represents the input to start the TXTL reactions in the cell extract. The first step is the transcription that transform the DNA into mRNA via RNA polymerase; then, the translation synthesises the protein correspondent to the mRNA, via ribosomes and tRNA that carries the required amino acids at each time. Finally, the produced protein could undergo post-translational modifications if required to perform its function.

Source: Dopp et al., Cell-free supplement mixtures: Elucidating the history and biochemical utility of additives used to support *in vitro* protein synthesis in *E. coli* extract, *Biotechnology Advances*, 2019

Recent applications of synthetic biology regard the development of cell-free systems where CRISPR components could directly be synthetised, and the related cleavage assay could be performed. This approach eliminates the need for the time-consuming process of *in-vitro* RNA synthesis and increases protein yield. Essentially, the Cas9 and the sgRNA proteins could be synthetised via TXTL reactions executed in the cell-free medium, starting from a DNA template and using RNA polymerase enzyme derived from bacteriophages (generally T7 and SP6) and with their respective promoters, leading to protein expression. Because the phage polymerase and promoters are very strong in inducing TXTL reactions *in-vitro*, in the last few years several cell-free systems have been developed. These are lysates derived from different organisms and, based on their origin, they could be used in diverse research contests (Kirill et al., 2014).

For example, the lysate derived from Human embryonic kidney cell line (HEK293) could be used to synthesise adenovirus, other viral vectors and human recombinant proteins, useful for virology studies and other research fields.

Moreover, also the use of cell extract derived from non-human organisms is significantly increasing to perform the *in-vitro* synthesis of proteins: for example, the lysates made of insect cells, Sf9 and Sf21 lines from *Spodoptera frugiperda*, Hi5 line from *Cabbage looper* and S2 line from *Drosophila*

melanogaster are currently used for expressing many proteins (Zitmann et al., 2017), especially membrane ones. Furthermore, because of the greater complexity of these insect cell extract, they are better compatible with eukaryotic proteins rather than bacterial lysate.

There are also other systems that are completely free from any cellular components, neither bacterial, human nor insect origin. This is the protein synthesis using recombinant elements (PURE) system, that is free of any cellular factors and chaperones for post-translational modification, therefore it is used to produce proteins free of modification or binding co-factors. Consequently, PURE system could be useful to understand the role of post-translational factors on nascent protein chains.

Crude extracts derived from bacteria are other systems widely used for cell-free gene expression (CFE) and cell-free protein synthesis (CFPS) applied to synthetic biology; in fact, they could be used for different purposes, such as the addition of non-native substrates, proteins, RNAs or recombinant DNA templates into a specific system. Bacterial polymerases generate sufficient mRNA *in-vitro* to activate protein synthesis, leading to the generation of genetic regulatory elements, such as riboregulators, riboswitches and other transcription-translation factors that could activate or repress the gene expression or other biosynthetic pathways, such as protein-protein interactions in organisms that are difficult to cultivate in the lab (Silverman et al., 2019).

CFE could also be used for the synthesis of biosensors that could detect infections, chemicals, pollutants or other components, based on the field of application of these systems: medical diagnosis, environmental monitoring, synthetic biology research. Generally, these biosensors are made of fluorescent reporter proteins that are expressed only if a specific target analyte is supplied. Moreover, in the therapeutic field, cell-free systems could be used to generate proteins that work as carriers for drug delivery in the organism. Because of the neutral structure of these synthetic proteins, they can easily drive medicines or other therapeutics directly in the organs without adverse effects.

Regarding the same field, cell-free lysates could also be used to make therapeutic proteins, such as monoclonal antibodies or other pharmaceutical compounds. In this case, cell-free systems represent a valid alternative for the synthesis of products of which source organisms are difficult to cultivate, therefore they only allow the production of the selected component (Lu et al., 2019).

Ideally, these *in-vitro* reactions should be performed by using a cell-free system that is close to the organism where the *in-vivo* step will then be performed. In this way, the whole testing will be

already optimised for that specific entity from the beginning, reducing the gap between the *in-vitro* and the *in-vivo* steps and expanding the use of cell-free tools to other synthetic biology applications. Moreover, the possibility to rapidly evaluate the activity of diverse cell extracts encourages the making of new lysates derived from various organisms (Moore et al., 2018), such as different bacteria, animals, plants.

Therefore, for all these advantages, *in-vitro* TXTL tools are becoming one of the most convenient experimental platforms for the building of biochemical systems outside of living organisms (Noireaux et al., 2020).

1.6.1 Bacterial crude extracts applied to CRISPR-Cas reactions: MyTXTL®

Among commercially available cell extracts, MyTXTL® (Arbor Biosciences) represents a well characterised system derived from *E. coli* cytoplasm. As previously mentioned cell-free systems, this lysate uses the endogenous bacterial machinery to synthesise proteins *in-vitro* with high versatility, that is useful for biological or synthetic applications (Garamella et al., 2016) (figure 1.17).

Figure 1.17 Workflow of MyTXTL® cell extract

MyTXTL® master mix includes all the component required for the synthesis of a protein, such as RNA polymerase, coenzymes, transcription factors, energy buffer, amino acids. It only requires the addition of the DNA template (linear or plasmid) encoding for the Cas9, the sgRNA and the fluorescent reporter system. The solution is then incubated for a few hours, leading to the protein synthesis via TXTL reactions performed *in-vitro* in the cell extract. Then, possibly, the protein could be purified or could directly undergo the analytic step.

Source: Arbor Biosciences website, page visited on December 2020

This cell extract is an all-in-one reaction that includes the T7 RNA polymerase and the sigma factor 70 (σ^{70}) as transcription factor; it only requires the gene of interest that should be expressed under transcriptional control of a specific σ70 promoter (P_{70}).

Essentially, MyTXTL® platform could be used also to synthesise proteins constituting the CRISPR complex, such as Cas9 and sgRNA; then, the endonuclease activity could be analysed by using a specific P70a-eGFP vector plasmid. Basically, the target template should be located immediately upstream a fluorescent gene, usually the enhanced green fluorescent protein (eGFP), a truncated version of the green fluorescent protein (GFP), and downstream the lambda phage promoter Pr (P_{70a}). Because the P_{70a} controls the gene expression of both the target and the eGFP, this specific design allows to monitor the cleavage of the target by the Cas protein based on the eGFP expression. In fact, if the cleavage happens in correspondence with the target site, the P_{70a} promoter could not induce the eGFP expression anymore; consequently, in cleavage conditions, the fluorescence is expected to decrease (figure 1.18).

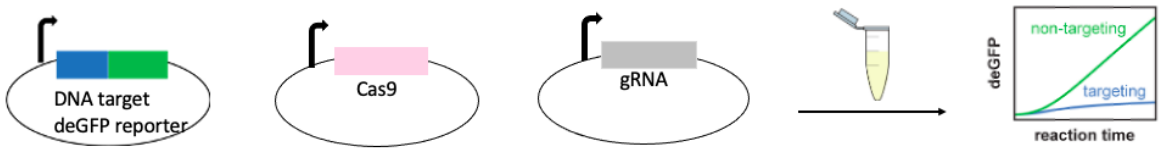


Figure 1.18: Reagents required to perform *in-vitro* CRISPR reactions using MyTXTL® cell extract

MyTXTL® is a bacterial cell extract that could be used to both synthesise and test various genetic editors. The *in-vitro* synthesis is based on TXTL reactions performed in the mix that already includes all the components required to make a protein. If a CRISPR reaction should be performed, three reagents are required: the genes encoding for the nuclease, the gRNA and the fluorescent protein. Once the corresponding proteins are generated in the mix, the nuclease activity could be analysed via deGFP fluorescence produced during the incubation in MyTXTL®.

The reporter expression in MyTXTL® is associated with a fluorescence production that rapidly reveals if the nuclease cleaves in correspondence with the target site. A strong eGFP expression in the cell extract induces a strong fluorescent signal, that is important to have practical real-time monitoring of the activity of the genetic editors. Another important characteristic of the reporter is

that its design should be flexible: the PAM and target have to be interchangeable. This customisable design allows to perform the TXTL testing with a great variety of endonucleases, such as Cas9 variants, Cas12a, ZFNs and other synthetic proteins, thanks to the possibility to easily adapt the target and/or the PAM based on the type of CRISPR rection that is set.

Because of its great versatility, MyTXTL® could be used to rapidly characterise diverse CRISPR components, such as Cas mutants, alternative gRNA design, PAM variants, inhibition of the cleavage by adding genetic regulators (Marshall et al., *Biotechnology and Bioengineering*, 2017; Marshall et al., *STAR Protocols*, 2020; Wandera et al., 2020), with all the advantages of an *in-vitro* testing that does not require protein purification or other time-consuming steps, such as culturing and transformation of living cells.

Another positive aspect of MyTXTL® is that this system could be used with both plasmids and linear DNA templates, that expands the possibility to test also those templates of which is difficult to obtain the related circular DNA. At the beginning, a limitation of linear templates was their recurrent degradation in the cell extract medium because of endogenous nucleases of *E. coli*, especially the RecBCD complex that is the major pathway for double-strand break repair in *E. coli*. In these conditions, a DNA fragment was less stable that a plasmid, limiting the use of both DNA formats. The activity of RecBCD complex is regulated by the χ sequence (5'-GCTGGTGG-3'), that is repeated throughout the *E. coli* genome numerous times; such sequence interferes with DNA sites via homologous recombination, stabilising the linear DNA. Therefore, the addition of short linear χ sequence to a TXTL reaction reduces the template degradation because the DNA template is better stabilised, without any interference with the protein synthesis (Marshall et al., *Biotechnology and Bioengineering*, 2017; Yim et al., 2020).

However, one limitation of MyTXTL® that remains regards the information about the quantification of the DNA in the medium: in fact, the concentration of the CRISPR reagents is influenced by several parameters, such as promoter strengths, UTR strengths and lengths of gene; moreover, because each cell-free system is different, model predictions should be measured for each platform (Marshall et al., 2019).

1.7 A platform for the *in-vitro* testing of CRISPR genetic editors in large scale

A fundamental advantage of cell-free systems is that they allow to perform complete synthesis reactions in a few hours without requiring extra components, apart from the DNA template that induces the protein synthesis. Therefore, these systems represent an interesting field to perform early-stage evaluations in large scale, prior to move to the *in-vivo* approaches.

Given this background, cell extracts could be used as testing platforms for the screening of different CRISPR components. In fact, the possibility to both synthesise and analyse a variety of endonucleases, gRNAs, reporter systems in only one reaction significantly increases the scalability of this system versus the traditional ones. Moreover, cell extract methods do not require cell or animal maintenance, meaning that they are more time saving, cheaper and do not need specific facilities: they could easily be performed in most research laboratories.

Potentially, *in-vitro* TXTL reactions could lead to the screening of a great amount of CRISPR components in a short time; however, in order to increase the scalability of the testing and have high-throughput methods, other techniques should be associated with the cell extract.

1.7.1 High-throughput cloning methods: BASIC assembly

One of the main requirements to expand the scalability of *in-vitro* CRISPR reactions is that a great quantity of reagents (Cas proteins, gRNAs, DNA target) should be provided in a short time and this is a limitation of standard cloning methods, such as Gibson or Golden Gate reactions.

If using MyTXTL® as cell extract source, the templates related to each CRISPR component could be added as linear amplicons or plasmids. The use of plasmids represents a significant constraint to large scale experiments when a specific DNA sequence is difficult to be cloned; this could happen because of common cloning problems, such as toxicity of the template in bacteria and struggling growth of that specific plasmid.

The Biopart Assembly Standard for Idempotent Cloning (BASIC) is a technique based on type II restriction enzymes that allow to assemble plasmids via cleavage and ligation of nucleotides with single stranded overhangs reporting type II (BsaI) restriction sites. These overhangs derived from specific prefix and suffix that are added to the original DNA template in a preliminary reaction that leads to the generation of so-called bioparts (BASIC-assembly website). Once all the templates are

in the same biopart format, they could be combined together in order to constitute the final plasmid (figure 1.19).

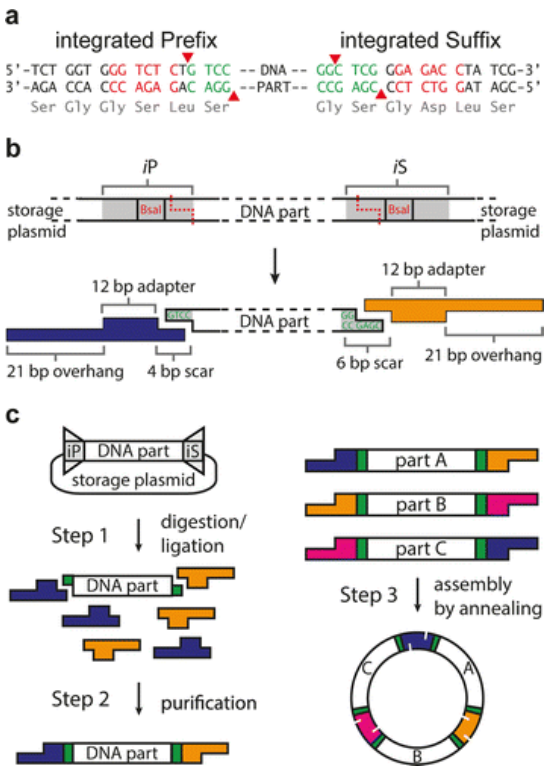


Figure 1.19: Mechanism of the BASIC assembly

(a) Specific prefix and suffix are added to the original DNA template, leading to the generation of overhangs that include the BsaI restriction sites. The template linked together with prefix and suffix constitutes a biopart (b). It is mandatory that all the regions that constitute the final plasmid, such as promoter, template, terminator have to be in the biopart format; in the end, the different bioparts could be assembled together because of BsaI restriction sites that are included in the prefix and suffix. Consequently, the final plasmid is generated (c).

Source: Storch et al., BASIC: A New Biopart Assembly Standard for Idempotent Cloning Provides Accurate, Single-Tier DNA Assembly for Synthetic Biology, *ACS Synthetic Biology*, 2015

Another important advantage of BASIC assembly is that within the synthetic linkers could be included parts that encode ribosome bind sequences (RBS) which regulate the further protein translation. Therefore, plasmids with RBS of different strengths could be generated, creating a

library of expression variants and providing an additional flexibility to the synthesis of a specific protein (Storch et al., 2015; Storch et al., 2016).

This makes the BASIC method highly versatile, that encourages the combination of various bioparts (promoter, RBS, gene variant, spacers, terminator) with each other. For example, different promoters could be associated with the same gene variant, generating various plasmids related to the same template; this allows to compare diverse assemblies and evaluate the plasmid variant that is characterised by minor toxicity in bacteria or best expression in the cell extract.

Another significant advantage of the BASIC assembly is that this cloning method could be used to explore dynamics related to cell-free protein expression, thanks to the possibility to rapidly set up and compare various plasmids characterised by the same template and different promoters. Therefore, the evaluation of a set of diverse promoters associated with the same gene could be useful to better understand how the transcription range is influenced (Reyes et al., 2017).

For example, the association of silenced promoters with the BASIC assembly represents a useful strategy to overcome toxicity-related problems when making a plasmid, especially during the transformation step. If the promoter is inactivated, the toxicity of the related template is neutralised at that stage because that gene is not expressed anymore and the plasmid growth is permitted.

The generation of plasmids using silenced promoters works with genetically modified bacteria that have a genome integration of a sensor (a repressor) for that specific promoter; then this sequence is able to bind the promoter, leading to its inactivation.

Afterwards, once the plasmid is generated via common bacterial transformation, it could undergo the further step, that is the cell-free reaction. In the bacterial lysate the repressor is not included: consequently, at this step, the promoter is not silenced, and it could drive the gene expression of its template. In the end, the bacterial lysate allows the *in-vitro* synthesis of the related protein, despite the toxicity of the original DNA.

Examples of inducible promoters that could be included in BASIC plasmids are the ones of the tetracycline (Tet) family; this gene regulation system allows to control the gene expression using anhydrotetracycline (ATc) as inducer to activate or deactivate the promoter repression. The whole complex includes a Tet promoter (TetP), an expression cassette for the repressor (TetR) and a chemical inducer, ATc; then, specific amino acids residues of TetR are important for tetracycline-dependent repressor. In the absence of ATc, TetR binds the TetP and suppresses the transcription; on the contrary, once ATc is present, it binds TetR and induces a conformational change that leads

to the dissociation of the repressor from the promoter, activating the gene expression (Ehrt et al., 2005) (figure 1.20).

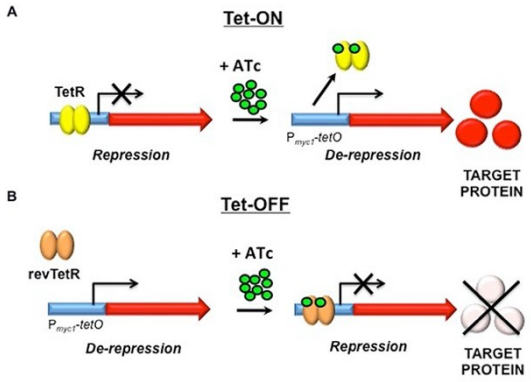


Figure 1.20: Tetracycline ON/OFF system

Tetrcycline inducible promoter could be used for gene regulation based on the configuration of the system. A) In the Tet-ON configuration, the repressor (TetR) binds the promoter and there is the repression of the transcription. In the presence of anhydrotetracycline (ATc), this inducer binds the TetR, inducing a conformational change that leads to the dissociation of the TetR from the promoter, allowing the transcription. B) In the Tet-OFF configuration, in the presence of ATc, the TetR binds again the promoter, silencing it and repressing the transcription.

Source: Evans et al., The application of tetracyclineregulated gene expression systems in the validation of novel drug targets in *Mycobacterium tuberculosis*, *Frontiers in Microbiology*, 2015

Therefore, TetP inducible promoter could be used in the assemble of BASIC plasmids in association with specific bacterial cells, such as DH10β Marionette (Meyer, 2018), that have an integration of a TetR inside their genome.

Consequently, the use of silenced promoters in the BASIC assembly increases the number of possible plasmids that could undergo further TXTL reactions, otherwise impossible to be cloned with standard promoters. In particular, with this system, could be generated even plasmids derived from toxic templates, that could be Cas9 mutants or particular gRNAs.

All these advantages make the BASIC a high-throughput method for generating plasmids that is conceptually simple at both design and experimental levels: in fact, it only requires a single-entry vector that then follows an easy hierarchical protocol that leads to parallel assembly of multiple parts in the same reaction. Moreover, the BASIC is highly accurate for up to seven parts per cloning

reaction, confirming that it is a powerful tool versus standard cloning methods, such as Gibson and Golden Gate, that could struggle when the assembly regards numerous fragments.

1.7.2 Automated platform for the CRISPR testing in large scale

Further improvements to perform CRISPR reactions in large scale could be the use of a completely automated DNA assembly tool that is able to assemble and make plasmids in large scale, at low cost and high efficiency.

Since the main problem is represented by cost-related issues because most laboratories could not afford expensive equipment, the development of open-source workflow by biofundries could be a valid compromise. In fact, these technology alliances provide an infrastructure that enables the design, the construction and the testing of large projects for biotechnology research, offering rapid progress in scalability, that is impossible to reach by a single entity (Hillson et al., 2019). This is the principle of the design-build-test-learn (DBTL) biological engineering cycle, that represents the framework of criteria followed for the development a synthetic biology application (figure 1.21) and is based on more sustainable and circular process, that is useful for the building of a future bioeconomy.

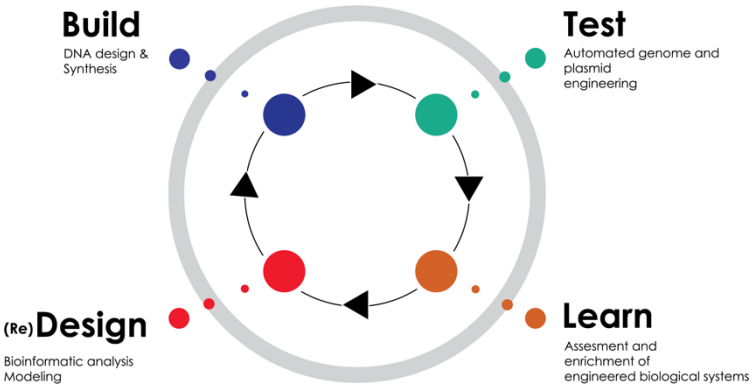


Figure 1.21: Design-Build-Test-Learn (DBTL) biological engineering cycle

The cycle starts with the D (design) which defines the desired function of a synthetic biology application; this step is mainly based on computational design of genetic parts or biological pathways. Then, there is the B (build) step: at this stage there is the physical generation of such genetic parts. The T (test) follows: this step involves the evaluation of the previously build parts in living cells or cell-free systems (such as cell extracts) at different scales. Finally, there is the L (learn): at this stage there is the development of modelling of D based on data derived from T. The whole cycle represents a sustainable and circular economy.

The *in-vivo* application should always be considered as the final aspiration of an experiment; however, prior to arrive at this final stage, there is a massive work behind, that is based on the screening among numerous components that, in case of CRISPR reactions are genetic editors. In particular, the activity of different Cas9 variants, PAM sequences, diverse gRNAs have to be analysed and only the most promising could be considered as valid candidates to undergo the final *in-vivo* application.

The association of MyTXTL® with an automated liquid handler, in particular the Echo 525 (Beckman Coulter), has showed to ensure a high throughput setting up of TXTL reactions (Bailey et al.; Marshall et al.), reducing the quantity of reagents required for each testing and, consequently, significantly cutting the costs.

Consequently, the possibility to have a completely automated platform that combines robotics, cell-free systems and high-throughput cloning methods represents an interesting approach for future uses of genetic editors in large scale. This method could apport numerous advantages, compared to standard techniques (mostly requiring living organisms) that are laborious and time consuming, limiting the scalability of the testing.

Therefore, setting up a computerised system that is able to both produce and test a massive amount of CRISPR components in a short time represents a valid solution for future applications in synthetic biology.

2. Aims and objectives

CRISPR-Cas9 is a technology widely used to edit the DNA at a precise location and finds applications in different fields, such as bioengineering, agriculture, medicine and it is of great interest for the prevention and treatment of human illnesses or to control vector-borne diseases. This system represents a powerful tool that, thanks to the possibility to easily modify targeted sequence in the genome, allows to insert desired mutations in a variety of organisms, such as plants and animals.
Based on the final application, designed CRISPR components, such as the nuclease and the gRNA, could be delivered inside cells and organs where they could perform their function.
Different methods based on living organisms have been developed for inducing gene editing by using cells or embryos, that are mostly used in research for the treatment of genetic disease.
In recent years many CRISPR tools have been developed and are now available, which encourages the use of this technology in a variety of *in-vivo* applications, such as gene therapy, vector-borne diseases, agriculture. However, most of these applications require living organisms, such as cells or animals, or expensive and time-consuming methods, that limit the scalability and the actual use of this tool.
In order to overcome this constraint, this research project focuses on the development of a novel approach for a standardised and scalable testing and of genetic editors. In particular, we aim to build a platform that would allow the *in-vitro* screening of CRISPR components in large scale, leading to a selection of those showing an activity that could, therefore, undergo further *in-vivo* applications.

This high throughput screening platform combines:

- Bacterial cell extract, which includes all the components responsible for synthesising and testing *in-vitro* CRISPR genetic editors;

- Standardised and scalable cloning method, based on a customisable design that leads to highly efficient and accurate assembly;

- Methods to overcome cell toxicity that may interfere with molecular cloning or testing in living cells or organisms;

- Robotics, to increase the scalability of the testing through the use of an automated system that could set up multiple CRISPR reactions at the same time.

First, I focused on the engineering of new Cas9 variants, in order to broaden the constraint of the PAM specificity and, consequently, expand the CRISPR targeting whatever in the genome, in an ideal PAMless scenario.

I designed new SpCas9-Protamine variants and I selected further nucleases to be analysed, that derived from diverse bacteria species (*Lactobacillus rhamnosus* and *Bacteroides fragilis*). After I obtained via computational analysis diverse PAM variants for each nuclease, I evaluated *in-vitro* the related cleavage activity of the Cas9s by using a cell-free system.

Then, I optimised this *in-vitro* testing by improving several parameters of the CRISPR reaction, such as the template format (linear or circular DNA, sgRNA or Cr-TrRNAs), the incubation temperature, the introduction of high-throughput cloning methods to generate the reagents and an automated system for setting up the CRISPR reactions.

All these phases, optimised and combined together, represent a whole platform that allows to design, test and finally screen in large scale various genetic editors.

This approach will encourage the design of novel genetic editors, which represents a high priority for a huge number of *in-vivo* applications. Moreover, it may facilitate the use of CRISPR tools in organisms where gene editing is more challenging (e.g., insects that have a very long-life cycle).

3. Materials and Methods

3.1 Bacterial cultures

Cloning methods were performed to make the reagents required for the *in-vitro* reactions. The final gRNA and reporter plasmids were generated via transformation with One Shot® TOP10 Escherichia coli strains (Invitrogen). Bacteria were grown in Laura-Bertani (LB) broth (10g/L peptone, 5 g/L yeast extract, 10 g/L NaCl) or LB agar (with 15 g/L agar for solid medium) with specific antibiotics to allow the selective growth of the required plasmid. The incubations were performed at 37C or 30C: the lower temperature was required for the L1, L2, L3, L4, B1, B2 and B3 reporter plasmids which did not grow at the standard temperature of 37C.

Moreover, to make the SpCas9 and LrCas9 plasmids, we used other chemically competent cells. In fact, following the BASIC assembly protocol (see section 3.9), we performed transformations with DH5α® cells with high transformation efficiency (10^9 CFU/μg pUC19, New England Biolabs) and DH10β Marionette cells homemade prepared by the Baldwin Lab at Imperial College London. The Marionette strain is genetically modified for silencing a specific promoter, reducing the toxicity of the plasmids during the transformation.

3.2 Polymerase Chain Reaction (PCR)

PCR reactions were performed as starting point to make the final plasmids or to generate the SpCas9, LrCas9 and BfCas9 amplicon templates and all the PCR were run in a Veriti Applied Biosystem thermocycler. The reactions were assembled by using Phusion® High-Fidelity PCR Master Mix (New England Biolabs), containing Phusion® DNA polymerase at a final concentration of 1 unit/50 μl, 1.5 mM $MgCl_2$, 200 μM of each dNTP and Phusion® HF buffer to perform high-fidelity amplification. The reactions were carried out in a final volume of 50 μl with ~ 10 ng of plasmid DNA template. Furthermore, specific oligonucleotide forward and reverse primers were added at a final concentration of 0.5 μM. Each PCR cycle started with an initial denaturation of 30 seconds at 98°C, followed by an annealing temperature set at +3°C of the lower melting temperature (Tm) primer or, when the primer length was less than 20 nucleotides, it was set at the same Tm of the lower primer.

Then, the extension temperature was set at 72°C for a time of ~ 15 seconds per kb, sometimes with a little increment in the time. In the termination step, the extension was carried out for other 10 min. Each PCR cycle was repeated 25 or 30 times, depending on the reaction. Afterward, PCR products were or directly purified by using the QIAquick PCR Purification Kit (Qiagen) or loaded on an agarose gel and then the band of interest was purified with the QIAquick Gel Extraction Kit (Qiagen).

3.3 Digestion with restriction enzymes

Digestions enzymes (New England Biolabs or Fast Digest by Thermo Fisher Scientific) were used to perform specific cuts via recognition sites, in order to get fragments required for the cloning process (e.g. to obtain the backbone which should then be assembled with the Cas9 or the gRNA sequences, to select specific promoters or terminators from other plasmids). Generally, 1μl enzyme was used to digest 1μg DNA and, in a final volume of 20μl reaction, we added also 10X Fast Digest Green Buffer. Each digestion was carried out for about 30 min at 37°C. After the incubation, the fragments were loaded on an agarose gel and the band of interest was purified by using the QIAquick Gel Extraction Kit (Qiagen).

3.4 Agarose gel electrophoresis

The DNA coming from both the PCR and the digestion with the restriction enzymes was loaded on a 1%-2% agarose gel (w/v) containing ethidium bromide (0.5 μg/mL). The run was performed in 0.5 x TAE buffer (20 mM Tris-acetate, 0.5 mM EDTA [pH 8.0], pH 8.3) at 70-120V, in order to separate the various bands based on their length. To estimate the size of each fragment, we included 6μl of HyperLadder™ 1kb (Bioline), 1kb plus DNA Ladder (Invitrogen) or HyperLadder™ 100 bp (Bioline), depending on the band of interest. Finally, the DNA was detected under UV light at 365 nm on a Gel Doc System (C280-Azure Biosystem).

3.5 Purification of DNA fragments

3.5.1 PCR Purification

DNA amplicons derived by the PCR reactions were cleaned by using the QIAquick PCR Purification Kit (Qiagen) following the manufacturer's instructions. To minimize the loss of DNA during the purification, we passed through the column the same sample twice. Eventually we eluted it in 30μl RNase-free H_2O.

3.5.2 Extraction and purification from agarose gel

After the electrophoresis, the selected bands of interest were extracted from the gel and purified by using the QIAquick Gel Extraction Kit (Qiagen) following the manufacturer's instructions. The final elution was performed in 30μl RNase-free H_2O.

3.6 Oligo Annealing

The oligo annealing was performed to make double-stranded DNA fragments having specific sequences required for the final plasmids. First, the two oligo strands were put together in equal molar amounts. Then, we added 1M NaCl: Na^+ ions were used to stabilize duplex formation by shielding the negatively charged DNA, allowing the two oligos to come into closer proximity. We heated the mixed oligonucleotides to 95°C for 5 minutes and we gradually cooled, reducing the temperature of ten degrees every 2 minutes, up to 20°C. Following the annealing step, the oligos were ready to be ligated with the other fragments.

3.7 Ligations

Ligations were performed to assemble the various fragments, leading to the generation of the final plasmid. We used two different enzymes, based on the kind of reaction: Takara ligation mix (Takara) and the T4 DNA Ligase (New England Biolabs), following the manufacturer's instructions.

3.7.1 Takara Ligation

The Takara mix was a single solution including all the components required for the ligation (e.g. enzyme and buffer); reactions were performed at R.T. for 10 min in a final volume of 10 10μl (5μl ligation mix and 5μl vector plus insert). The insert: vector molar ratios were 3:1 or 2:1, depending on the fragment lengths and concentrations.

3.7.2 T4 DNA Ligation

When using the T4 DNA Ligase (New England Biolabs) in addition to the enzyme, we added the T4 DNA Ligase Buffer (New England Biolabs), the insert and, generally, 75 ng vector in a final volume of 20μl. Each T4 ligation was performed at R.T. for 2 hours and followed by a further incubation at 65°C to inactivate the enzyme.

3.8 Golden Gate Assembly

The Golden Gate reaction was performed to allow the assembly of more fragments into a backbone vector, using the simultaneous activity of Type IIS restriction endonucleases (REases) and T4 DNA ligase. The inserts and the vectors were designed with the IIS recognition site distal to the cleavage site; consequently, when the REases cut the DNA, the recognition site was eliminated from the ligated product, allowing a precise, seamless assembly. Each Golden Gate reaction was performed with 50 or 75 ng backbone vector, based on the type of plasmid which had to be assembled, 100 ng

insert or 10mM annealed oligos, 2 mg/mL Bovine Serum Albumin (BSA), T4 DNA Ligase (New England Biolabs), T4 DNA Ligase Buffer (New England Biolabs) and the specific Type IIS restriction enzyme (purchased by Thermo Fisher Scientific or New England Biolabs). In particular, in this project, we included BsaI and AarI REases. For those Golden Gate reactions which required AarI, we used the GeneArt® Type IIS Assembly Kit (Thermo Fisher Scientific), which already included all the enzymes and buffer required for the assembly. We only needed to add 75ng recipient vector and 75ng insert or PCR fragments in the reaction. Each Golden Gate was carried out with 30 cycles of 1 minute at 37°C -1 minute at 16°C repeats, which were the optimal working temperatures of the Type IIS REase and the ligase. Then, according to the ligation - Golden Gate cloning protocol, we continued with the inactivation of both the enzymes: first, we incubated the reaction at 65°C (AarI and BsaI inactivation) for 5 minutes, then at 80°C (ligase inactivation) for 5 minutes.

3.9 BASIC Assembly

The Biopart Assembly Standard for Idempotent Cloning (BASIC) was performed to generate the Cas9s derived from *Lactobacillus rhamnosus* and *Streptococcus pyogenes* (positive control Cas9), in order to overcome the toxicity of LrCas9 which occurred during the transformation reaction. This assembly could be divided in two steps: first, the various parts composing the plasmid (e.g. promoter, backbone, DNA template) had to be converted in BASIC bioparts, after binding specific linkers (prefix and suffix). Then, there was the actual assembly, with the combination of all the bioparts together. I kindly thank the Baldwin Lab at Imperial College London that gifted us all the linkers (Biolegio) required for the assembly.

3.9.1 BASIC reaction

First of all, we designed specific primers having in the sequence the BsaI recognition sites and encoding the prefix and suffix required for the plasmid; then we subcloned them in the BASIC vector, the Cas9 and the promoter, making the three bioparts composing the plasmid. The BASIC linker ligation reaction was performed in 30µl total volume and included T4 buffer (Promega), T4 ligase 1-3 U/µl (Promega), prefix linker, suffix linker, BsaI-HF v2 (New England Biolabs). We performed 20

cycles of 2 minutes incubation at 37°C and 1 minute at 20°C, to allow the enzyme activity. Then, we continued with their inactivation: 5 minutes at 55°C (for the BsaI-HF v2) and 20 minutes at 80°C (for the T4 ligase). We repeated this protocol to make the bioparts for the Cas9, the promoter and the backbone. Once we got the bioparts, we magbead purified them (see 3.16); later we performed the final assembly.

3.9.2 Assembly reaction

The final BASIC assembly allowed to get the final plasmid, putting together the bioparts. Each reaction required 50ng biopart per 1kb final plasmid; it was performed in 10µl final volume which included CutSmart buffer 10x (New England Biolabs) and the three-linker ligated BASIC parts (promoter, Cas9 and backbone). We run the assembly reaction at 50°C for 45 minutes and we concluded the cloning with the transformation, by using DH5α® (to make every single biopart in form of plasmid prior to the final assembly) or DH10β Marionette (to make the final assembly) chemically competent cells (see sections 3.10.3 and 3.10.4).

3.10 Transformations

3.10.1 Transformation of TOP10 chemically competent *E. coli*

Most of the DNA constructs were transformed using One Shot® TOP10 (Invitrogen) chemically competent cells. 2.5µl of each ligation reaction were added in 25µl of competent cells. After 30 minutes incubation on ice, a heat shock was performed, placing the vial containing the cells in a water bath set at 42°C for 30 seconds and then it was placed again in the ice. After 2 minutes, 400µl S.O.C. medium (supplied by the kit) were added to the bacteria-construct mixture and then the vial was placed at 37°C in a shaking incubator (Stuart) for 1 hour at 225 rpm. Finally, ~125µl from each transformation vial were spread on LB agar plates supplemented with a specific antibiotic and

inverted plates were placed in an incubator (Thermo Fisher Scientific) at 37°C overnight or at R.T. for 3 days when transforming the L1, L2, L3, L4, B1, B2 and B3 reporter plasmids.

3.10.2 Transformation of Stbl3™ chemically competent *E. coli*

Some DNA construct were transformed using One Shot™ Stbl3™ (Invitrogen) chemically competent *E. coli*. ~2.5μl of each ligation reaction were added in 50μl of competent cells. After 30 minutes incubation on ice, a heat shock was performed, placing the vial containing the cells in a water bath set at 42°C for 45 seconds and then it was placed again in the ice. After 2 minutes, 400μl S.O.C. medium (supplied by the kit) were added to the bacteria-construct mixture and then the vial was placed at 37°C in a shaking incubator (Stuart) for 1 hour at 225 rpm. Finally, ~125μl from each transformation vial were spread on LB agar plates supplemented with a specific antibiotic and inverted plates were placed in an incubator (Thermo Fisher Scientific) at 37°C overnight.

3.10.3 Transformation of DH5α® chemically competent *E. coli* (high transformation efficiency)

DH5α® *E. coli* cells were used as transient hosts during the BASIC assembly, in particular for the transformation which allowed to get the separate bioparts in form of plasmids. 5μl of BASIC DNA assembly were added in 50μl of competent cells. After 20 minutes incubation on ice, a heat shock was performed, placing the vial containing the cells in a water bath set at 42°C for 45 seconds and then it was placed again in the ice for 2 minutes. 200μl S.O.C. medium (supplied by the kit) were added to the bacteria-construct mixture and then the vial was placed at 37°C in a shaking incubator (Stuart) for 1 hour at 225 rpm. Finally, ~250μl from each transformation vial were spread on LB agar plates and inverted plates were placed in an incubator (Thermo Fisher Scientific) at 37°C overnight.

3.10.4 Transformation of DH10β Marionette chemically competent *E. coli*

DH10β Marionette chemically competent cells were used in the last step of the BASIC cloning, following the final assembly reaction. This strain had been kindly provided and homemade prepared by the Baldwin Lab at Imperial College London. DH10β Marionette cells had a genome integration of a Tet repressor, in order to repress the ProTet promoter and reduce the cytotoxicity of some plasmids. 5µl of BASIC DNA assembly were added in 50µl of competent cells. After 20 minutes incubation on ice, a heat shock was performed, placing the vial containing the cells in a water bath set at 42°C for 45 seconds and then it was placed again in the ice for 2 minutes. 200µl S.O.C. medium (supplied by the kit) were added to the bacteria-construct mixture and then the vial was placed at 37°C in a shaking incubator (Stuart) for 1 hour at 225 rpm. Finally, ~250µl from each transformation vial were spread on LB agar plates and inverted plates were placed in an incubator (Thermo Fisher Scientific) at 37°C overnight.

3.11 Colony Screening

A first checking of bacterial colonies grown after the transformation was performed via colony screening. Primers were used to amplify specific sequences of the plasmid, generally the most critical regions such as close to the ligation or insertion points. Each colony from the LB agar plate was picked up with a pipette tip, added into a PCR tube and then streaked onto another LB agar plate which was used to regrow the same colony. The PCR reactions were performed by using the Fast Cycling PCR kit (Qiagen); each reaction mix included 2x Fast Cycling PCR Master Mix (containing HotStarTaq *Plus* DNA Polymerase, 1x Fast Cycling Buffer, 200µM of each dNTP, optimized Mg^{2+} concentration), 10x CoralLoad Dye and 0.5µM forward and reverse primers. First, the mix was incubated at 95°C for 5 minutes for an initial heat inactivation, then 30 cycles were performed with the following steps: 5 seconds at 96°C (denaturation), 5 seconds at a temperature approximately 5°C below T_m of primers, generally ~55°C (annealing), ~30 seconds (3 seconds for 100 bp) at 68°C (extension). Usually, 3 positive colonies related to the same plasmid were regrown in order to be sent for sequencing. Each colony was picked up with a pipette tip and placed in 5mL LB medium containing 5µl antibiotic; then the tube was left overnight at 37°C in a shaking incubator at 225 rpm

(Stuart) or at 30°C in a ThermoMixer (Eppendorf) at 300 rpm, in case of L1, L2, L3, L4, B1, B2 and B3 reporter plasmids which required a lower growing temperature.

3.12 Mini-preparation of plasmid DNA

A preparation of the plasmid in small volume was obtained from bacterial cultures following the QIAprep Spin Miniprep kit (Qiagen). The final elution was performed in 30µl RNase-free H_2O.

3.13 DNA Sequencing

Samples derived from the Mini-preparation step were sent to be sequenced to Eurofins Genomics. 50ng/µl plasmid were prepared in a final volume of 15µl (if 1-4 reactions) or 20µl (if 5-8 reactions). Furthermore, up to 8 different primers were sent at a concentration of 10pmol/µl. The sequencing was performed via Sanger method.

3.14 Maxi-preparation of plasmid DNA

DNA plasmids which obtained positive results from the sequencing were prepared in large scale, in order to be used for further experiments. Maxi-preparations were made by using the ZymoPURE™ II Plasmid maxiprep kit (Zymo Research). The previous day of the maxi-preparation a pre-inoculation was performed in the morning: each colony was picked up with a pipette tip and placed in 5mL LB medium containing 5µl antibiotic; then the tube was left ~6-7 hours at 37°C in a shaking incubator at 225 rpm (Stuart) or at 30°C in a ThermoMixer (Eppendorf) at 300 rpm, in case of L1, L2, L3, L4, B1, B2 and B3 reporter plasmids. In the evening, the final inoculation was carried out: 100mL LB medium containing 100µl antibiotic were placed in a 500 mL flask or 2L flask (in case of L1, L2, L3, L4, B1, B2 and B3 reporter plasmids only). Then 2.5µl from the pre-inoculation were added and the

final mix was left overnight in a hot room set at 37°C on a shaking incubator at 225 rpm (Stuart) or at 30°C in a shaking incubator at 225 rpm (for L1, L2, L3, L4, B1, B2 and B3 reporter plasmids). The next day, the Maxi-preparations were made following the manufacturer's instructions and the final elution was performed in 200μl RNase-free H_2O.

3.15 DNA Quantification

3.15.1 Nanodrop Spectrophotometer

The concentration and the purity of DNA plasmids were measured by using a Nanodrop spectrophotometer (Azure). 1μl sample was loaded, and the absorbance was measured at wavelengths of 260nm and 280nm simultaneously. At 260nm, the absorbance was used to calculate the concentration of the dsDNA, while the ratio between the reading at 260nm and 280nm (OD_{260}:OD_{280}) represented the purity; an OD_{260}:OD_{280} ratio of ~1.8 could generally be considered as high-quality DNA.

3.15.2 Qubit Fluorometer

Because many molecules absorb at 260nm, the spectrophotometer may be inaccurate, due to potential contaminations. Therefore, we decided to measure the final concentrations of the samples which were used as reagents in the TXTL (see section 3.17.1) reaction with a fluorometer. This technique used fluorescent dyes which bound specifically the nucleic acid of interest (in our case double-stranded DNA), providing more accurate quantification. The samples were quantified with the Qubit 4 Fluorometer (Thermo Fisher Scientific) by using the Qubit™ dsDNA Broad Range Assay Kit (Thermo Fisher Scientific). The assay is highly selective for double-stranded DNA (dsDNA) and is designed to be accurate for initial sample concentration from 100pg/μl–1,000ng/μl. 1μl sample was measured following the manufacturer's instructions and the concentration was detected.

3.16 Magbead Purification

Before being tested in TXTL, all PCR and plasmid samples (Cas9s, gRNAs, reporters) were previously purified with AMPure XP beads (Beckman Coulter). This technique was based on magnetic separation and utilized Solid Phaze Reversible Immobilization magnetic beads-based technology. The cleaning carried out following the manufacturer's instructions by using a magnetic rack and the final elution was performed in RNase-free H_2O, in a volume based on the starting and the final concentrations.

3.17 CRISPR *in vitro* testing

3.17.1 Cell-free protein expression: myTXTL® Sigma 70 Master Mix

To evaluate the activity of several CRISPR components, we used myTXTL® Sigma 70 Master Mix (Arbor Biosciences), an all-in-one system which allowed rapid cell-free protein expression. This Master Mix provided all the components required for *in vitro* transcription-translation (TX-TL), such as *E. coli* cell extract, RNA polymerase, sigma factor 70 (σ^{70}), basic transcription factors, amino acids and energy buffer. For the expression of a protein (Cas9, gRNA and reporter) it only required the addition of the specific DNA template under transcriptional control of a σ^{70} specific promoter. TXTL reactions were set up with 12µl final volume, by combining 9µl myTXTL® Sigma 70 with 3µl of a mix including the three reagents together(Cas9, gRNA (sgRNA or Cr-TrRNAs) and reporter), all of them at 5nM concentration. The sgRNAs, Cr-TrRNAs and the reporters were always added as plasmid templates; the SpCas9 and the LrCas9 were added as both plasmid and linear DNA, based on the experiment; the BfCas9 was only added as linear template.

Prior to perform each TXTL experiment, all the sample related to the Cas9s, the gRNAs and the reporters (both linear and plasmid formats) were purified by using AMPure XP beads (Beckman Coulter) (see section 3.16 above).

The final reactions were loaded in 384-well plates (Greiner) and the plate was sealed with MicroAmp™ Optical Adhesive Film (Thermo Fisher Scientific), in order to prevent sample evaporation. TXTL reactions were performed at 29°C or 37°C in a FLUOstar Omega microplate

reader (BMG Labtech) and carried out for ~5 hours at 500rpm orbital shaking. After each cycle, the related fluorescence was measured from the bottom of the plate at 485nm excitation filter and 520nm emission filter.

3.17.2 Stabilization of linear DNA templates: myTXTL® GamS Protein

When the in vitro gene expression was carried from linear template instead than its circular equivalent, we added in the reaction mix also myTXTL® GamS Protein (Arbor Biosciences). They were inhibitors of the endogenous exonucleases from *E. coli*, in particular RecBCD which better stabilized the linear DNA. Therefore, even with amplicon templates, we got a protein production efficiency comparable to reactions performed with plasmids. The ready-to-use stock solution (conc. 150µM) was directly added in the myTXTL® Sigma 70 Master Mix (Arbor Biosciences) to a final concentration of 10µM. After incubating for 5 min on ice, the DNA template was added to the mixture at a final concentration of 5nM.

3.18 Automation of the *in vitro* testing: Echo 525 Acoustic Liquid Handler

In order to perform high throughput experiments, we set up some of the TXTL reactions by using an acoustic liquid handler: the Echo 525 (Beckman Coulter), which was located at Imperial College London, White City Campus. This machine was able to transfer all the reagents (Cas9, gRNA, reporter, myTXTL® Sigma 70 Master Mix) having different viscosity with high speed, precision and accuracy, thanks to the sound energy. When using the Echo 525, TXTL reactions were set up in 384 Well Small Volume™ LoBase Microplates (Greiner) in 4µl final volume, reducing sample consumption. Once the plate was loaded with all the reagents, the TXTL reaction was carried out in the FLUOstar Omega microplate reader (BMG Labtech), as when using manual pipetting.

4.1. TXTL *in-vitro* reactions for the testing and screening of CRISPR genetic editors

After various attempts, I successfully made a vector plasmid to be used as precursor for the generation of the reporter plasmids for TXTL *in-vitro* reactions.
This vector, via simple Golden Gate reaction with AarI restriction enzyme, leads to the generation of all desired reporters, after removing the cloning spacer and inserting the PAM and the target site (figure 4.1).

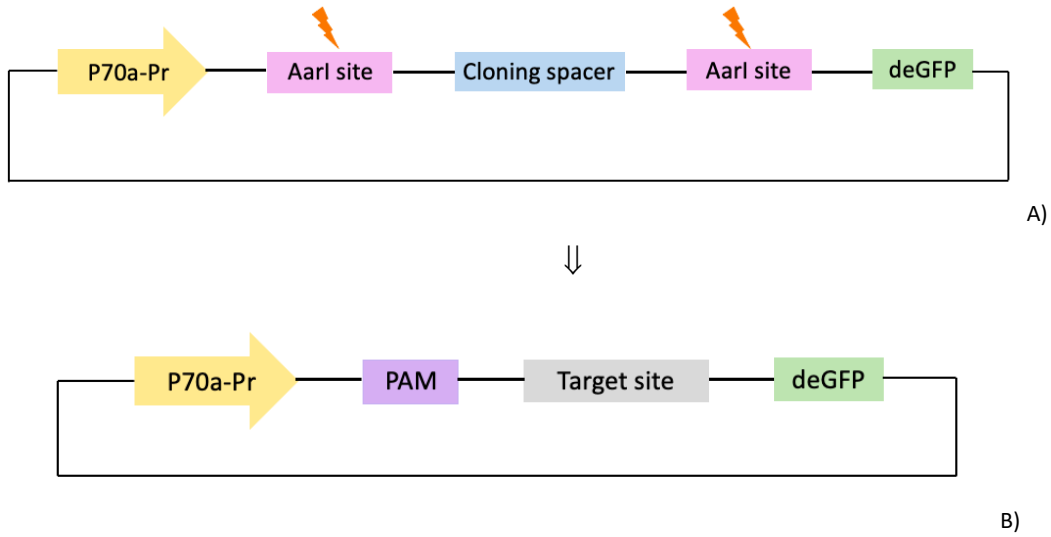


Figure 4.1: Engineering of customisable reporter systems

All the reporter plasmid generated in this study derive from the same ancestral vector (A). This plasmid includes the bacterial promoter P70a, which is responsible for the gene activation in the bacterial cell extract; two AarI restriction sites; the cloning spacer sequence, located between the AarI sites and, in the end of the concatenated sequence, the deGFP. This specific design allows to easily remove the cloning spacer via Golden Gate reaction with AarI digestion enzyme: consequently, that region could be replaced by the PAM-target site sequence (B), leading to the generation of reporter systems customised for all possible nucleases.

The specific reporter design explains also how each CRISPR reaction should be analysed: via monitoring the fluorescence that is produced during the TXTL incubation. In fact, when the endonuclease cuts in correspondence with the target site, the P70a promoter could not induce the

deGFP expression anymore: consequently, there is a decrement in the fluorescence rate. On the contrary, if the endonuclease does not cut the DNA, the P70a promoter could continue to activate the deGFP expression. Consequently, the fluorescence rate in non-cleavage conditions increases (figure 4.2).

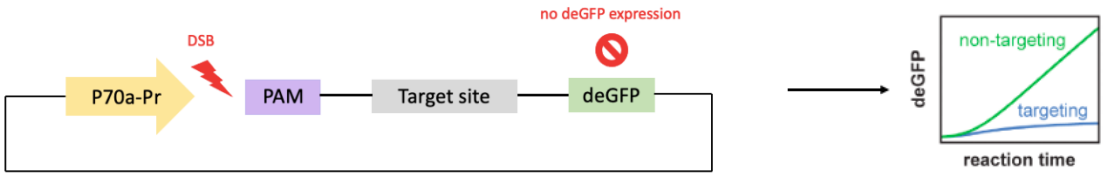


Figure 4.2: The role of the reporter system in monitoring the nuclease activity

The reporter system is used as index for analysing the cleavage activity of a specific nuclease in CRISPR reactions performed in MyTXTL® cell extract. In targeting conditions, the nuclease induces the DSB in correspondence with the target site and the cleavage stops the deGFP expression under the P70a promoter and the consequent fluorescence reaches its maximum value (blue line). In non-targeting conditions, there is no cleavage, the promoter still induces the deGFP expression and there is an increment of the fluorescence rate (green line). Consequently, the higher or lower fluorescence produced during the TXTL incubation is representative of the nuclease activity.

Thanks to this customisable design (figure 4.1), starting from the same ancestral vector, I generated and tested various reporter systems associated with PAMs and targets related to different Cas9 variants and orthologues. Moreover, for each nuclease, given a specific target site, I could explore new PAM sequences, trying to overcome this constraint, aiming to apply the CRSIPR tool wherever it could be required in the genome of a specific organism.

In order to test the PAM specificity of different Cas9 variants, I first set up an easy monitoring system that allowed to analyse in TXTL the cleavage activity; then I selected the nuclease proteins to be tested, such as SpCas9-Protamine variants and Cas9s derived from *Lactobacillus rhamnosus* and *Bacteroides fragilis* bacteria. I compared their ability to cleave the DNA with both their canonical PAMs and also with other derived sequences in different conditions, in order to broaden the PAM specificity and, ideally, expand their targeting.

First, I used the *in-vitro* TXTL for analysing the activity of the WT SpCas9, in order to validate the system prior to expand its application to other nucleases and genetic editors.

In standard conditions the testing should be performed at 29C, that is the temperature value where the components included in MyTXTL® cell extract have their best activity.

Each CRISPR reaction was performed in both targeting and non-targeting conditions. In targeting conditions, the gRNA is designed to guide the Cas9 in correspondence with the target site and, therefore, the DNA cleavage is expected. On the contrary, a non-targeting gRNA does not allow the cutting, consequently this was considered as negative control.

Prior to test *in-vitro* the various combinations of Cas9/gRNA/PAM, I validated if each designed reporter construct had enough strength to be used for quickly monitoring the CRISPR experiments A strong reporter is associated with a good deGFP expression in the bacterial cell extract, leading to a high fluorescent signal.

I made five reporter constructs for the SpCas9, two reporting the NGG PAM (they were designed with two different target sites), and other three characterised by NAG, NTG and NCC PAM sequences; furthermore, I generated other reporters for the orthologues LrCas9 and BfCas9. The designed PAM mutants were TTTTT, TTCCT, CCAAT and TTTCT for LrCas9 and NGGAACG, NAGTACA, GTTCAAT for BfCas9.

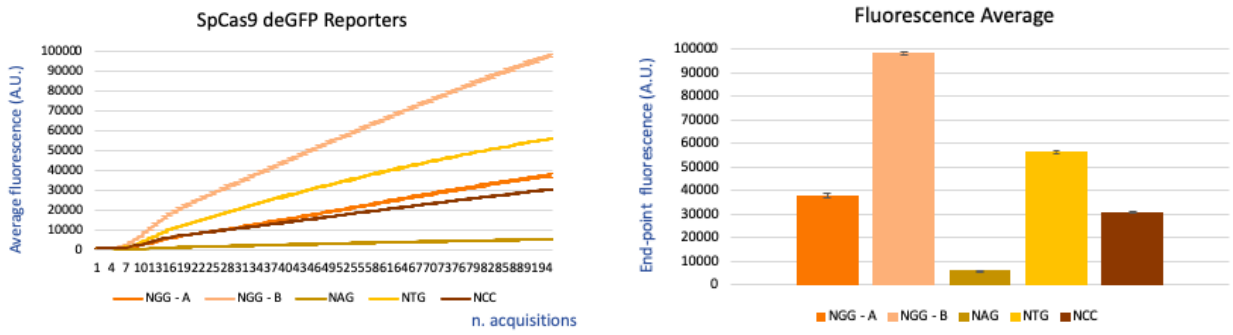


Figure 4.3: SpCas9 deGFP reporters

The fluorescence rate produced during the *in-vitro* synthesis of the reporter constructs associated with the SpCas9 allows to analyse the strength of each reporter. Each reporter plasmid was incubated with the TXTL cell extract at 29C and the deGFP expression was monitored. Line graphs represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 96 acquisitions (indicated in the X-axis) at an interval of 3 minutes. Bar graphs represent the average end-point value of fluorescence intensity (A.U.) ± SD (at ~5 hours from start of reaction).

Analysing the figure 4.3, I can compare the various SpCas9 reporter strengths. The deGFP genes associated with each construct induced the *in-vitro* synthesis of the corresponding deGFP protein, that could be confirmed by the generated fluorescence curves. This means that all these constructs could be used for monitoring the nuclease activity when using MyTXTL®, thanks to the fluorescent signal that they generate.

However, the fluorescence production rate of these constructs was different: the reporter carrying the NCC PAM gave the lowest fluorescence rate, due to the lowest deGFP expression. On the contrary, the reporter with the NGG PAM (B) appeared to be the strongest construct, giving the highest deGFP expression. The reason why these reporters are characterised by different expression may be changes in their RBS strengths, major toxicity of some structures more than others or any other factors that could affect the *in-vitro* synthesis of such reporters.

I generated other customisable reporter plasmids for the orthologues LrCas9 and BfCas9, carrying different PAM variants that were selected for those nucleases. The designed sequences were TTTTT, TTCCT, CCAAT and TTTCT for LrCas9 and NGGAACG, NAGTACA, GTTCAAT for BfCas9 and they were tested in the same reaction conditions as SpCas9 (incubation at 29C, 96 acquisitions at an interval of 3 minutes).

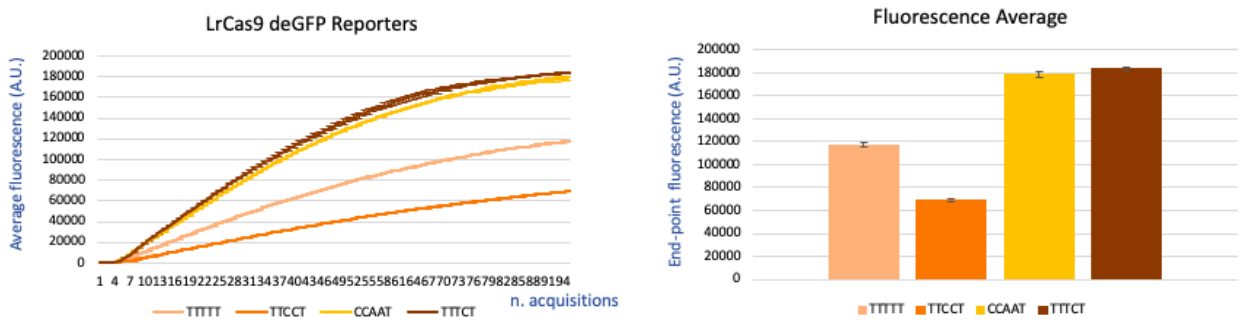


Figure 4.4: LrCas9 deGFP reporters

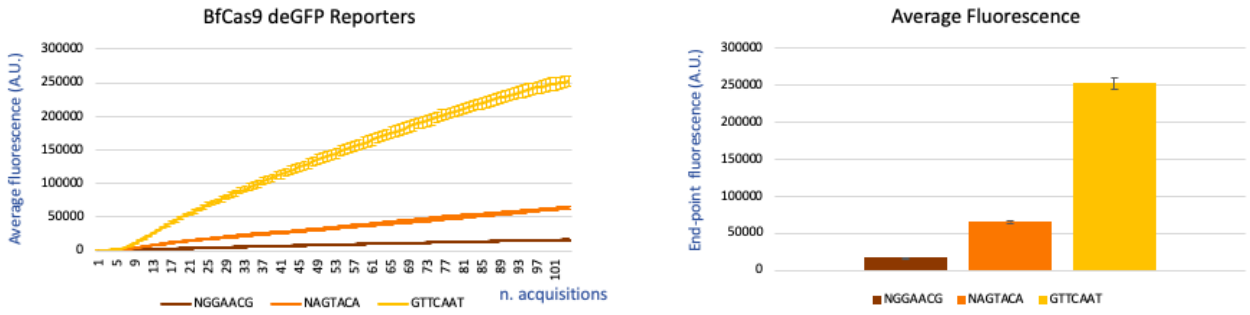


Figure 4.5: BfCas9 deGFP reporters

The deGFP expression related to the LrCas9 (figure 4.4) and BfCas9 (figure 4.5) reporter constructs was measured analysing the fluorescence generated by the TXTL reaction. Each reporter plasmid was incubated with the TXTL cell extract at 29C and the deGFP expression was monitored. Line graphs represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 96 acquisitions (indicated in the X-axis) at an interval of 3 minutes. Bar graphs represent the average end-point value of fluorescence intensity (A.U.) ± SD (at ~5 hours from start of reaction).

All the LrCas9 and BfCas9 reporters gave good fluorescence values in the same reaction conditions (29C and ~5 hours incubation). Analysing in detail the LrCas9 reporter plasmids (figure 4.4), the construct carrying the CCAAT and TTTCT PAMs were those which gave the highest deGFP expressions, with very similar fluorescence values, while the construct reporting the TTCCT PAM gave the lowest deGFP expression.

Then, focusing on BfCas9 reporters (figure 4.5), the construct carrying the GTTCAAT PAM appeared to be the strongest one, while the one with the NGGAACG PAM was the weakest.

In the end, all designed reporter for SpCas9, LrCas9 and BfCas9 were validated as good reference systems for monitoring the cleavage activity during CRISPR *in-vitro* reactions. Consequently, I could proceed with the testing of all the components together: the Cas9, the gRNA, and the reporter system carrying the PAM and the target site.

4.1.1 SpCas9 tested *in-vitro* at 29C

The sgRNA plasmid (figure 4.7) included bacterial promoter (J23119) and terminator (aspA) to induce the *in-vitro* expression of the gRNA in the *E. coli* cell extract, leading to the synthesis of the corresponding protein that constituted the CRISPR complex together with the SpCas9.

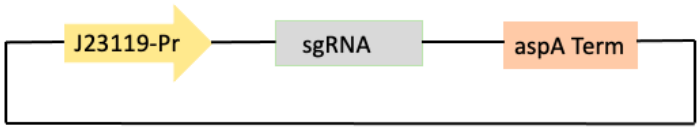


Figure 4.6: Design of the SpCas9 sgRNA plasmid

The gRNA plasmid includes the bacterial promoter J23119, which is responsible for the gene activation in the bacterial cell extract, the sgRNA sequence, designed to guide the spCas9 in correspondence with the target site, and the bacterial terminator aspA, which stops the transcriptional machinery.

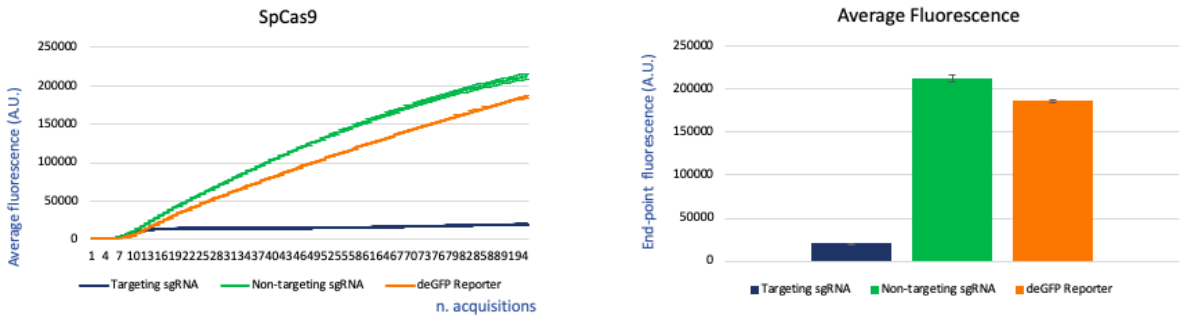
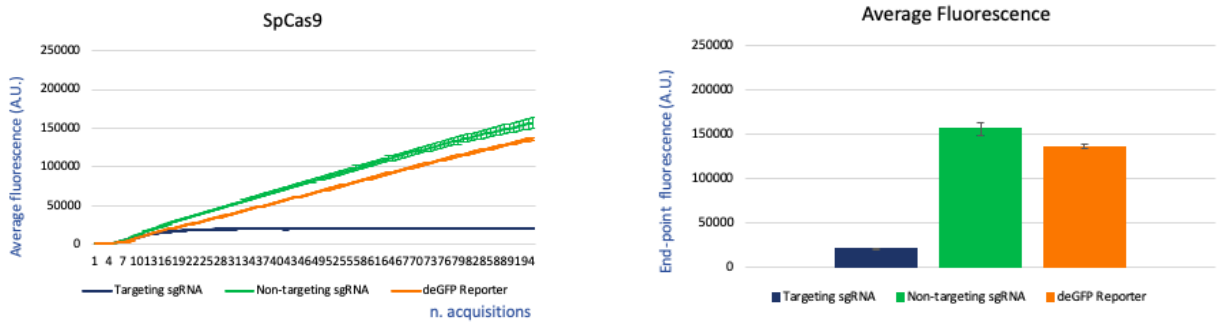


Figure 4.7: SpCas9 sgRNA at 29C

The activity of the SpCas9 was measured analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract In each testing, the SpCas9 plasmid was incubated with the sgRNA (targeting or non-targeting), the deGFP reporter and the TXTL cell extract at 29C and the deGFP expression was monitored. Line graphs represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 96 acquisitions (indicated in the X-axis) at an interval of 3 minutes. Bar graphs represent the average end-point value of fluorescence intensity (A.U.) ± SD (at ~5 hours from start of reaction).

Figure 4.7 clearly reports as the DNA was cleaved by the SpCas9. Analysing both targeting (blue line) and non-targeting (green line) sgRNA curves, the fluorescence follows two opposite trends, with a significant decrement of the deGFP expression in targeting conditions. After few cycles, the targeting curve reaches its maximum value, which means that the cut stops the deGFP expression. On the contrary, in non-targeting conditions, the SpCas9 could not cleave the DNA; consequently, the deGFP continued to be expressed under transcriptional control of the P70a promoter.

Therefore, I successfully validated the *in-vitro* expression and activity of this specific sgRNA. These results also confirmed that the J23119 bacterial promoter induced the sgRNA expression in TXTL, suggesting that it could be used as default for the design of new gRNA variants.



Figures 4.8: SpCas9 sgRNA at 29C, 2 steps reaction

The activity of the SpCas9 was measured analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. In the 2 steps reaction, the SpCas9 was previously pre-incubated at 29C for 30 min with the gRNA (targeting or non-targeting) and the TXTL cell extract only, without the deGFP reporter. After 30 min, the deGFP reporter was added to the pre-incubation reaction and the final reaction was incubated at 29C and the deGFP expression was monitored. Line graphs represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 96 acquisitions (indicated in the X-axis) at an interval of 3 minutes. Bar graphs represent the average end-point value of fluorescence intensity (A.U.) ± SD (at ~5 hours from start of reaction).

Using the same SpCas9/gRNA/deGFP reporter combination, I performed a second CRISPR reaction adding a pre-incubation step of all the components without the reporter.

I wanted to validate this 2 steps reaction to investigate if a strong gene expression in the cell extract could be associated with a major consumption of elements required to synthesise the protein, such as enzymes, coenzymes, transcription factors and other cell extract components. Because the reporter plasmid is characterised by a stronger promoter than the SpCas9 and the sgRNA ones, I hypothesised that if the reporter was temporarily removed from the mix, the SpCas9 and the sgRNA plasmids could use the cell extract components to guarantee the synthesis of their associated proteins, reducing the risk that the reporter could rapidly consume most of them to synthesise its own protein. However, the pre-incubation step was mainly considered to improve the TXTL at 37C (see further experiments, from section 4.4.1 below in this chapter).

Usually, in the TXTL cell extract, the reporter expression is stronger than that one of the other components, the Cas9 and the gRNA. Therefore, the pre-incubation allowed to start the *in-vitro* synthesis of these two components before adding the deGFP reporter. In this way, the master mix components responsible for the transcription-translation reactions (enzymes, coenzymes, transcription factors, amino acids) were distributed among all the reagents in a balanced way, avoiding their excessive consumption by the reporter for its own expression, due to its greater strength.

The 2 steps reaction showed positive results (figure 4.8): targeting and non-targeting sgRNAs clearly confirmed that the SpCas9 managed to cleave the DNA. Analysing both one (figure 4.7) and two steps (figure 4.8) reactions, the SpCas9 activity is equivalent for the targeting sgRNA in both reactions, there is no significant differencein the reactions performed with and without the pre-incubation.

4.1.1 Testing of new synthetic SpCas9 variants: Cas9-PRM1 fusion proteins

Ideally, CRISPR-Cas system could be engineered to target any location in the genome, provided that a specific nucleotide sequence (PAM) is located downstream the target site. The PAM constraint limits the actual CRISPR application to any genomic site.

SpCas9 variants and orthologues have already been engineered in the attempt to overcome this limitation and some of them have improved the precision of targeted DNA cleavage (Anders et al., 2016; Kleinstiver et al., 2016; Walton et al., 2020; Nishimasu et al., 2019; Hu et al., 2018; Kleinstiver et al., 2015); however, these variations often restrict the range of targetable sequences.

Nuclease variants could be generated binding other proteins to one of the Cas9 domains, in order to increase the nuclease precision, allowing specific genome editing, that is ideal for precision medicine applications (Bolukbasi et al., 2015). Apart from expand the targeting loci in the genome, Cas9 fusion proteins could improve the efficiency of correction/mutation, reducing error prone NHEJ events at the target site and improving HDR frequency (Jayavaradhan et al., 2019; Liu et al., 2018).

An ideal PAM-less CRISPR system would allow to operate directly wherever it is required, theoretically at any genomic locus, with high precision and without any restrictions. This would be

ideal for significantly expanding CRISPR applications in many fields because there would be no restrictions about the target selection.

Therefore, in order to broaden the PAM specificity and, consequently, improve the nuclease targeting, I engineered new SpCas9 variants that, ideally, could allow the CRISPR application even in those loci in the genome that do not have a NGG sequence located downstream their position.

The design of these alternative nucleases was based on the addition of sequences derived from DNA-binding proteins characterised by structural similarities with the Cas9 sequence. In this way, the affinity with the genome should not have been interfered but, at the same time, the new nuclease variants could have had a different affinity with non-canonical PAMs, leading to a potential targeting expansion.

Among the DNA-binding proteins I focused on the Human Sperm Protamine 1 (PRM1). This is one of the main proteins responsible for DNA compaction and it has high affinity for nucleic acids DNA and RNA thanks to its positive charge, as a result of an Arginine rich region involved in strong binding to negatively charged DNA (Brewer et al., 2003; Balhorn et al., 2007; Brewer et al., 1999).

Therefore, to interact with the genome, the PRM1 requires the same amino acid than the SpCas9, were R1333 and R1335 are responsible for PAM interaction.

Based on these similarities, I engineered SpCas9-PRM1 orthologues: SpCas9 proteins fused with the PRM1 and other variants with insertion/mutations derived from the PRM1 in the sequence of the WT SpCas9. In details, the WT SpCas9 was modified at two points: the PAM Interacting domain and the C-terminal (figure 4.10).

With the introduction of these mutations, I aimed to expand the canonical NGG PAM constraint, with consequent benefits for the DNA targeting. By modifying those specific domains in the WT SpCas9 sequence I wanted to investigate if a change in the 1333 and 1335 area could have improved the targeting flexibility or if those amino acid positions must be kept unmodified to allow the DNA cleavage. Regarding the insertions at the C-terminal of the WT SpCas9, I wanted to evaluate if the addition of a tag sequence at the end of the WT SpCas9 could have modified the PAM specificity.

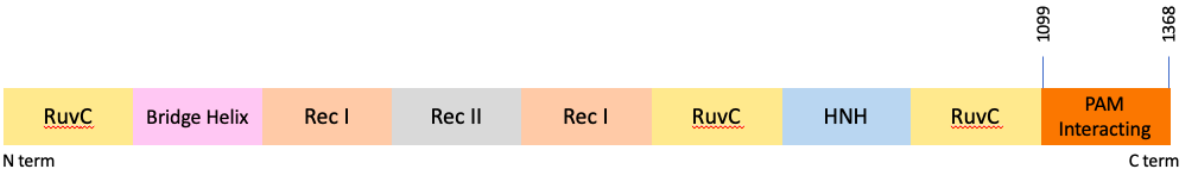


Figure 4.10: Wild type SpCas9 domains

The wild type SpCas9 protein is comprised of 1368 amino acids and is organised in multiple domains, each with distinct function. The PAM Interacting domain, located at the C-terminal, is responsible for the recognition of the PAM sequence by the Cas9. If there is this recognition, the protein could bind the DNA and, therefore, cleave in correspondence with the target site.

Figure 4.11: SpCas9-Protamine mutants engineered by modifying the PAM Interacting domain and the C-terminal of the wild-type SpCas9

The four SpCas9 orthologues have insertions at both the PI domain (SpCas9-Protamine 1 and 2) and the C-terminal (SpCas9-Protamine 3 and 4) of the wild-type SpCas9. The inserted mutations derived from the DNA binding domain of the PRM1 or from the whole PRM1 sequence itself.

In the SpCas9-Protamine 1 mutant (figure 4.11), I inserted the PRM1 DNA binding region immediately after the SpCas9 R1335; in this variant I did not change the position of the two fundamental arginine required in positions 1333 and 1335.

In the SpCas9-Protamine 2 mutant, I inserted the mutation in between R1333 and R1335, removing the lysine K1334. The inserted sequence was the PRM1 DNA binding region, but with a further mutation. To reduce the risk of disulphide bridges which could occurs with the cysteine (C) in the sequence, C was substituted with a glutamine (Q).

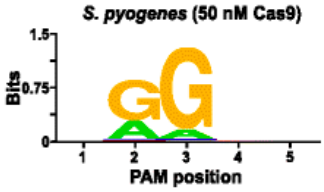
In the SpCas9-Protamine 3 variant, the PRM1 DNA binding region was fused at the C-terminal, after the final aspartic acid D1368, keeping unaltered the PI domain. Also, in this mutant, the C derived from the PRM1 was substituted with a Q (as for the SpCas9-Protamine 2).

Finally, I engineered the SpCas9-Protamine 4 variant, binding at the C-terminal the whole sequence of the PRM1, not only the DNA binding region. In this mutant I kept the C of the PRM1, and I did not modify the SpCas9 PI.

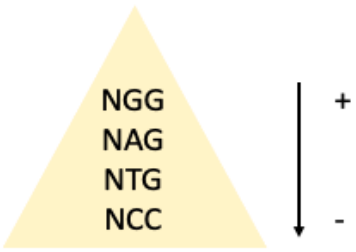
In order to evaluate if these SpCas9-PRM1 mutants had different PAM specificities compared to the WT nuclease, I searched for alternative PAM sequences that could have been tested with the variants I engineered.

Previous studies had derived the frequency of nucleotides at each PAM position for the SpCas9 (Karvelis et al., 2015), obtained from bioinformatic analysis (Mojica et al., 2009; Karvelis et al., 2017). The prediction of nucleotides at each position of the sequence was extrapolated using a position frequency matrix (PFM), that gave the probable composition of alternative PAMs for SpCas9 plotted in a WebLogo. This logo includes graphical representations of sequence conservation obtained by multiple alignments. The height of each letter indicates the conservation of the relative amino acid at that position (Crooks et al., 2004).

Based on the WebLogo for SpCas9 PAMs (figure 4.12-A) and starting from the canonical NGG sequence, I gradually substituted nucleotides at each position aiming to select new putative PAMs for the PRM1-SpCas9 orthologues. A and G were the nucleotides more likely to work with SpCas9, because of their frequent conservation in the alignments. Therefore, thanks to the collaboration with Nace Kranjc, PhD student in the Crisanti Lab at Imperial College London that performed the bioinformatic analysis, we derived new potential PAM sequences which could be tested in addition with the canonical NGG: NAG, NTG and NCC (figures 4.12-A and 4.12-B).



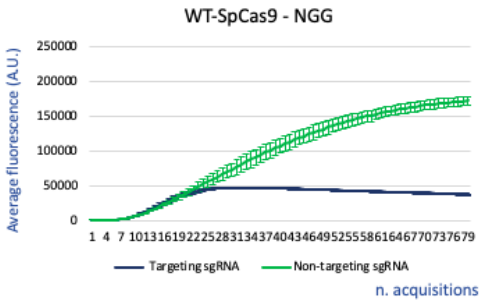
A

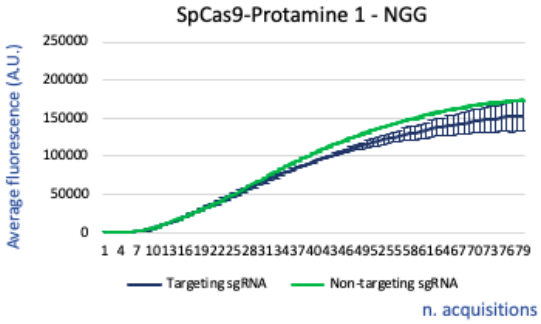


B

Figure 4.12-A and 4.12-B: PAM preferences for SpCas9 protein and derived sequences to be tested

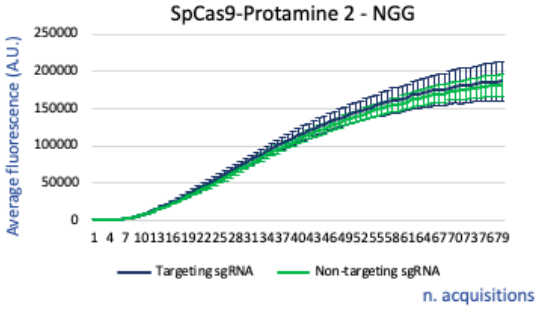
Frequency of nucleotides at each PAM position (figure 4.12-A) was calculated using a position frequency matrix (PFM) and plotted as a WebLogo (Source: Karvelis et al., Rapid characterization of CRISPR-Cas9 protospacer adjacent motif sequence elements, *Genome Biology*, 2015). From this logo, I derived the sequences NAG, NTG and NCC (figure 4.12-B), which were selected as putative PAM variants for the SpCas9. NCC was selected as negative control.

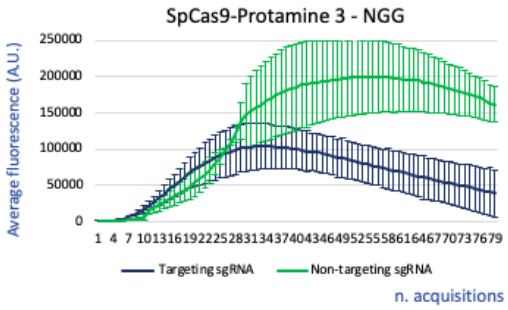




Figures 4.13 (left) and 4.14 (right). The wild-type SpCas9 (4.13) and the SpCas9-Protamine 1 orthologue (4.14) tested with the PAM NGG

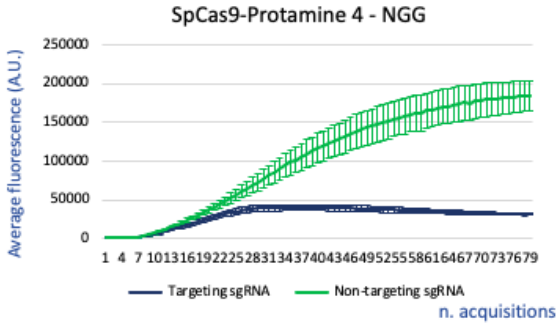
The activity of the SpCas9 Protamine orthologues was measured analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. In each testing, the Cas9 plasmid was incubated with the sgRNA (targeting or non-targeting), the deGFP reporter and the TXTL cell extract at 29C and the deGFP expression was monitored. Line graphs (figures 4.13 and 4.14) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 96 acquisitions (indicated in the X-axis) at an interval of 3 minutes.

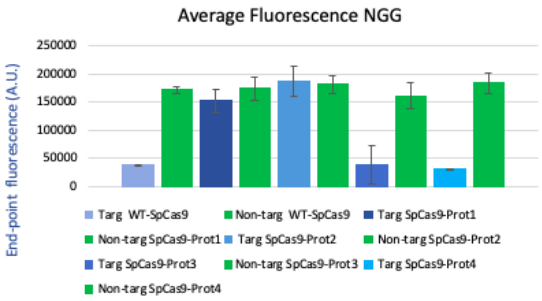




Figures 4.15 (left) and 4.16 (right): The SpCas9-Protamine 2 (4.15) and SpCas9-Protamine 3 orthologues (4.16) tested with the PAM NGG

The activity of the SpCas9 Protamine orthologues was measured analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. In each testing, the Cas9 plasmid was incubated with the sgRNA (targeting or non-targeting), the deGFP reporter and the TXTL cell extract at 29C and the deGFP expression was monitored. Line graphs (figures 4.15 and 4.16) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 96 acquisitions (indicated in the X-axis) at an interval of 3 minutes.

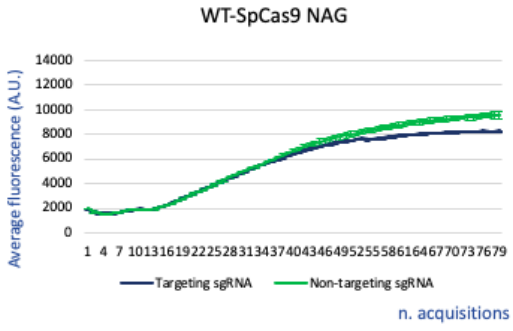


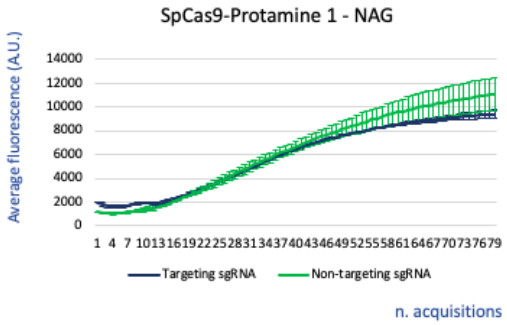


Figures 4.17 (left) and 4.18 (right): The SpCas9 orthologue P4 (4.17) and the average of the fluorescence produced by all the SpCas9-Protamine orthologues (4.18) tested with the PAM NGG

The activity of the SpCas9-Protamine orthologues was measured analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. In each testing, the Cas9 plasmid was incubated with the sgRNA (targeting or non-targeting), the deGFP reporter and the TXTL cell extract at 29C and the deGFP expression was monitored. Line graphs (figure 4.17) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 96 acquisitions (indicated in the X-axis) at an interval of 3 minutes. Bar graphs (figure 4.18) represent the average end-point value of fluorescence intensity (A.U.) ± SD (at ~5 hours from start of reaction).

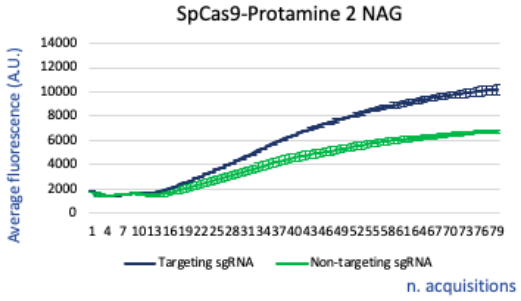
Analysing the TXTL reactions performed with all the SpCas9 orthologues, some mutants gave positive results in targeting conditions. In particular, set as positive control the wild-type SpCas9 (figure 4.13), I observed a similar trend in the orthologues P3 and P4 (figures 4.16 and 4.17). The other mutants, SpCas9-Protamine 1 and SpCas9-Protamine 2 (figures 4.14 and 4.15), gave similar results in both targeting and non-targeting conditions instead. This trend could be justified by the different locations were these orthologues had the mutations. SpCas9-Protamine 3 and SpCas9-Protamine 4 were characterized by insertions at the C-terminal of the SpCas9 protein; consequently, this did not interfere with the PI domain, which was free to recognise the PAM sequence and allow the DNA cleavage. On the contrary, SpCas9-Protamine 1 and SpCas9-Protamine 2, which had insertions in the PI domain, did not show any kind of activity. Therefore, mutations in that region reduce the Cas9 activity.

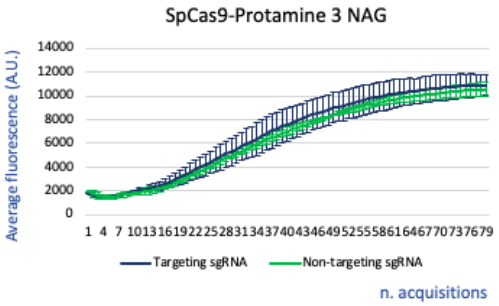




Figures 4.19 (left) and 4.20 (right): The wild-type SpCas9 (4.19) and the orthologue SpCas9-Protamine 1 (4.20) tested with the PAM NAG

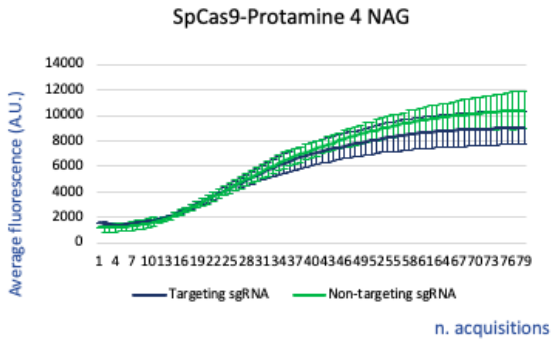
The activity of the SpCas9 Protamine orthologues was measured analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. In each testing, the Cas9 plasmid was incubated with the sgRNA (targeting or non-targeting), the deGFP reporter and the TXTL cell extract at 29C and the deGFP expression was monitored. Line graphs (figures 4.30 and 4.31) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 96 acquisitions (indicated in the X-axis) at an interval of 3 minutes.

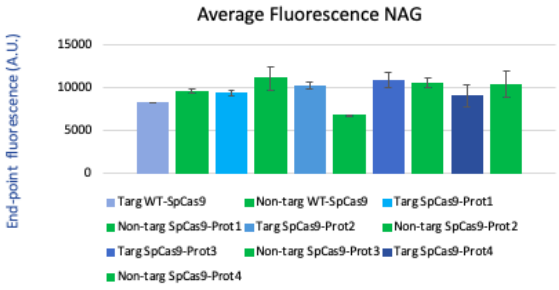




Figures 4.21 (left) and 4.22 (right): The SpCas9-Protamine 2 (4.21) and SpCas9-Protamine 3 orthologues (4.22) tested with the PAM NAG

The activity of the SpCas9 Protamine orthologues was measured analysing the fluorescence generated by the reaction. In each testing, the Cas9 plasmid was incubated with the sgRNA (targeting or non-targeting), the deGFP reporter and the TXTL cell extract at 29C and the deGFP expression was monitored. Line graphs (figures 4.21 and 4.22) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 96 acquisitions (indicated in the X-axis) at an interval of 3 minutes.



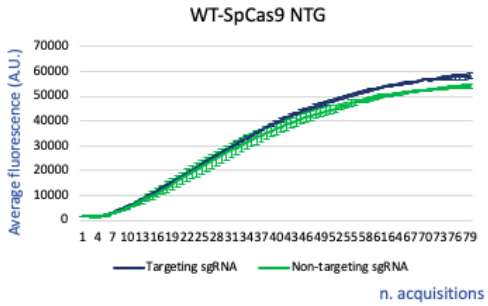


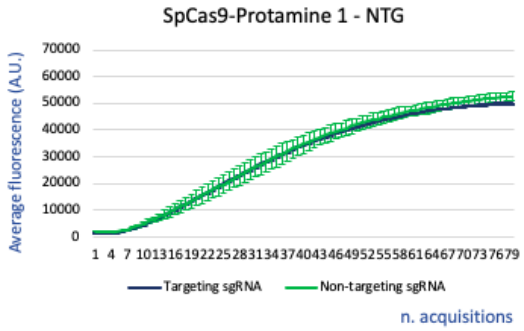
Figures 4.23 (left) and 4.24 (right): The SpCas9-Protamine 4 orthologue (4.23) and the average of the fluorescence produced by all the SpCas9-Protamine orthologues (4.24) tested with the PAM NAG

The activity of the SpCas9 Protamine orthologues was measured analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. In each testing, the Cas9 plasmid was incubated with the sgRNA (targeting or non-targeting), the deGFP reporter and the TXTL cell extract at 29C and the deGFP expression was monitored. Line graphs (figure 4.34) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 96 acquisitions (indicated in the X-axis) at an interval of 3 minutes. Bar graphs (figure 4.35) represent the average end-point value of fluorescence intensity (A.U.) ± SD (at ~5 hours from start of reaction).

Focusing of the testing with the PAM NAG, only the wild-type SpCas9 (figure 4.19) and the SpCas9-Protamine 4 (figure 4.23) variant showed a slight difference between their targeting and non-

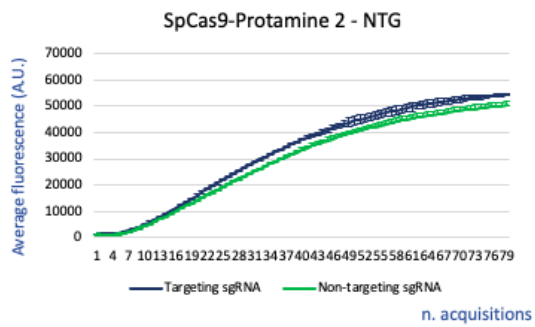
targeting trends; however, a clear activity could not be detected. Therefore, the PAM NAG was a negative variant when analysing the activity of the SpCas9 and of any derived nucleases (SpCas9-Protamine 1,2,3 and 4).

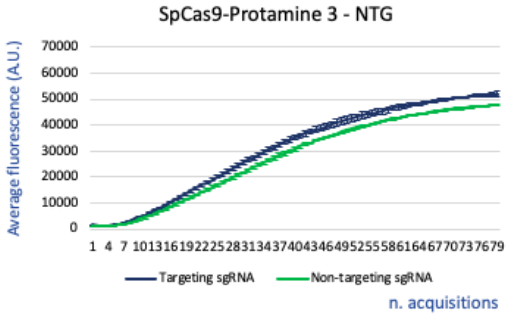




Figures 4.25 (left) and 4.26 (right): The wild-type SpCas9 (4.25) and the SpCas9-Protamine 1 orthologue (4.26) tested with the PAM NTG

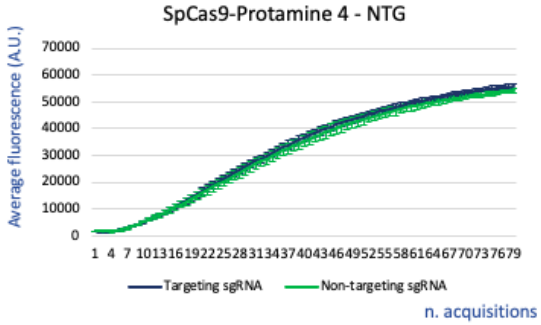
The activity of the SpCas9 Protamine orthologues was measured analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. In each testing, the Cas9 plasmid was incubated with the sgRNA (targeting or non-targeting), the deGFP reporter and the TXTL cell extract at 29C and the deGFP expression was monitored. Line graphs (figures 4.25 and 4.26) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 96 acquisitions (indicated in the X-axis) at an interval of 3 minutes.

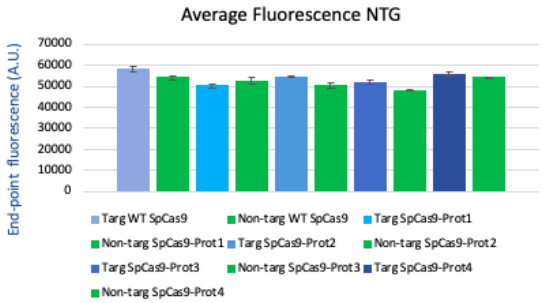




Figures 4.27 (left) and 4.28 (right): The SpCas9-Protamine 2 (4.27) and 3 orthologues (4.28) tested with the PAM NTG

The activity of the SpCas9 Protamine orthologues was measured analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. In each testing, the Cas9 plasmid was incubated with the sgRNA (targeting or non-targeting), the deGFP reporter and the TXTL cell extract at 29C and the deGFP expression was monitored. Line graphs (figures 4.27 and 4.28) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 96 acquisitions (indicated in the X-axis) at an interval of 3 minutes.

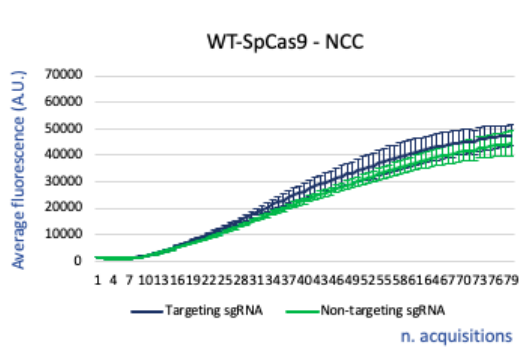


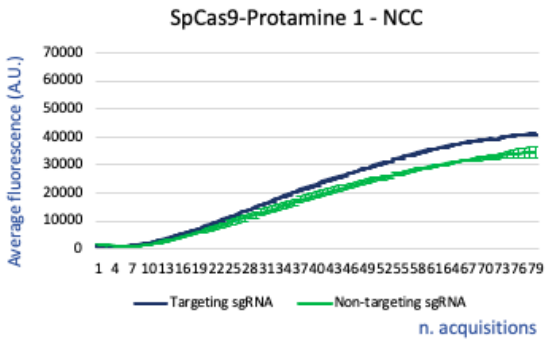


Figures 4.29 (left) and 4.30 (right): The SpCas9-Protamine 4 orthologue (4.29) and the average of the fluorescence produced by all the SpCas9-Protamine orthologues (4.30) tested with the PAM NTG

The activity of the SpCas9 Protamine orthologues was measured analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. In each testing, the Cas9 plasmid was incubated with the sgRNA (targeting or non-targeting), the deGFP reporter and the TXTL cell extract at 29C and the deGFP expression was monitored. Line graphs (figure 4.29) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 96 acquisitions (indicated in the X-axis) at an interval of 3 minutes. Bar graphs (figure 4.30) represent the average end-point value of fluorescence intensity (A.U.) ± SD (at ~5 hours from start of reaction).

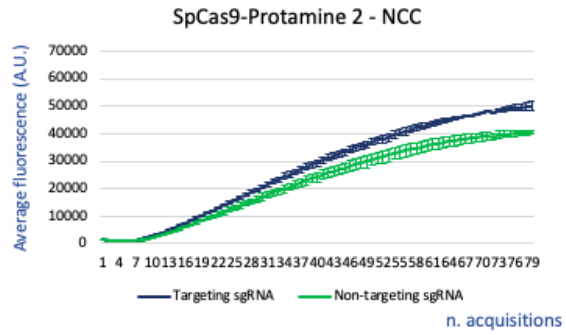
No endonuclease activity could be detected analysing data from the TXTL performed with the PAM NTG. The wild-type SpCas9 and all the derived SpCas9-Protamine variants gave the same trends in both targeting and non-targeting conditions, there was no significant difference in the fluorescence production. Therefore, as for the PAM NAG, also the NTG sequence did not promote the activity of the SpCas9 mutants.

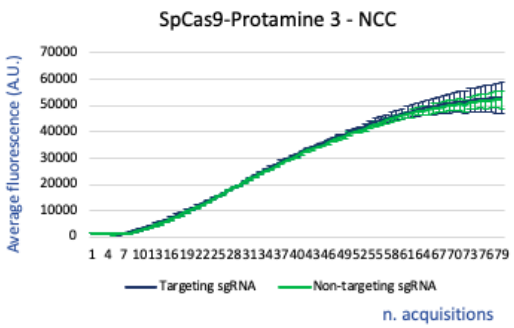




Figures 4.31 (left) and 4.32 (right): The wild-type SpCas9 (4.31) and the SpCas9-Protamine 1 orthologue (4.32) tested with the PAM NCC

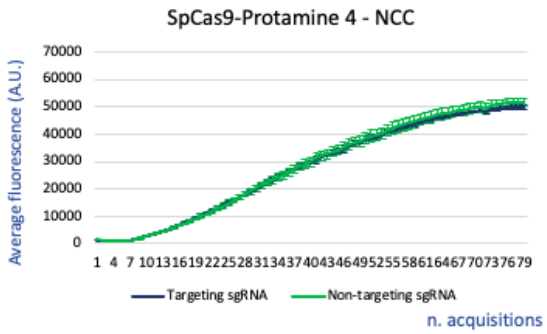
The activity of the SpCas9 Protamine orthologues was measured analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. In each testing, the Cas9 plasmid was incubated with the sgRNA (targeting or non-targeting), the deGFP reporter and the TXTL cell extract at 29C and the deGFP expression was monitored. Line graphs (figures 4.31 and 4.32) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 96 acquisitions (indicated in the X-axis) at an interval of 3 minutes.

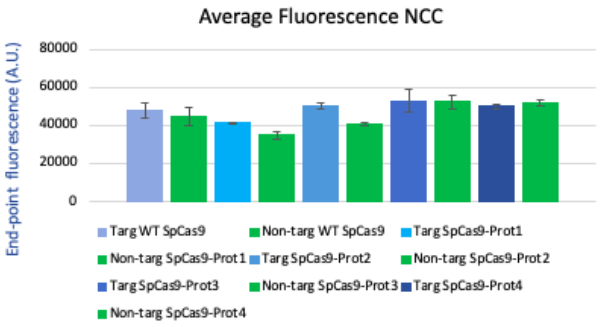




Figures 4.33 (left) and 4.34 (right): The SpCas9-Protamine 2 (4.33) and 3 orthologues (4.34) tested with the PAM NCC

The activity of the SpCas9 Protamine orthologues was measured analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. In each testing, the Cas9 plasmid was incubated with the sgRNA (targeting or non-targeting), the deGFP reporter and the TXTL cell extract at 29C and the deGFP expression was monitored. Line graphs (figures 4.33 and 4.34) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 96 acquisitions (indicated in the X-axis) at an interval of 3 minutes.





Figures 4.35 (left) and 4.36 (right): The SpCas9-Protamine 4 orthologue (4.35) and the average of the fluorescence produced by all the SpCas9-Protamine orthologues (4.36) tested with the PAM NCC

The activity of the SpCas9 Protamine orthologues was measured analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. In each testing, the Cas9 plasmid was incubated with the sgRNA (targeting or non-targeting), the deGFP reporter and the TXTL cell extract at 29C and the deGFP expression was monitored. Line graphs (figure 4.35) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 96 acquisitions (indicated in the X-axis) at an interval of 3 minutes. Bar graphs (figure 4.36) represent the average end-point value of fluorescence intensity (A.U.) ± SD (at ~5 hours from start of reaction).

The negative control PAM NCC gave expected results: no SpCas9 activity could be detected (figures from 4.31 to 4.36). Targeting and non-targeting conditions showed the same fluorescence trends, confirming that in these conditions the orthologues could not recognize the target site.

4.1.2 Testing of Cas9 orthologues: LrCas9

In order to expand the application of CRISPR/Cas system at any location in the genome, common limitations related to the well characterised SpCas9 should be overcome.

The requirement of a specific PAM restricts the possibility to edit those targeted loci that do not have the canonical NGG sequence downstream their position: this constraint represents a huge limit to the actual use of CRISPR in many research fields, from gene therapy to vector-borne disease and agriculture.

Therefore, I focused on other Cas9 orthologues, trying to overcome the PAM requirement to edit a specific location in the genome. I selected the Cas9s derived from *Lactobacillus rhamnosus* (LrCas9)

and *Bacterioides fragilis* (BfCas9) as possible alternative to the WT SpCas9. (The BfCas9 will be analysed in the section 4.2.3 later in this chapter).

I focused on LrCas9 because this orthologue already showed PAM flexibility, inducing DNA cleavage in bacteria with the PAM TTTCT (Crawley et al., 2018). This discovery supported the idea that this orthologue could be capable of targeting a broader range of sequences, enabling more precise targeting when associated with novel gRNA design.

In order to improve the PAM specificity, I derived new PAM sequences by using a computational pipeline. I got a sequence logo (figure 4.37-B), which represents the conservation of nucleotide positions in the 5'-end flanking sequences of the target sites. Finally, I could determine the PAM sequences for LrCas9. TTTTT was the most obvious one and I selected also TTCCT, which was a slightly different variant that would still be supported by PAM logo. In addition to these two variants, I tested *in-vitro* also another sequence, TTTCT (figure 4.37-A), which had previously been tested in bacteria, where it supported the nuclease activity (Crawley et al., 2018). CCAAT was selected as negative control instead. I thank Nace Kranjc, PhD student in the Crisanti Lab at Imperial College London, for the bioinformatic analysis.

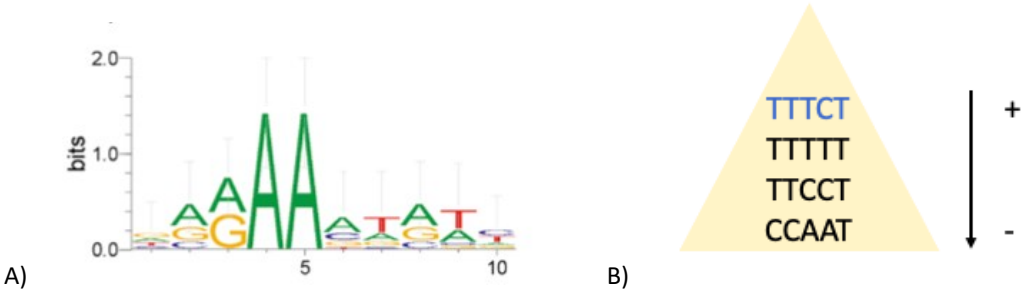
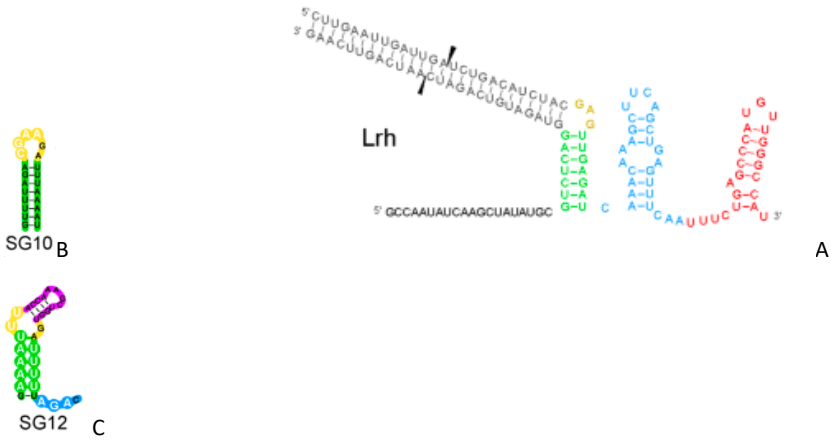


Figure 4.37-A (on the left) and 4.37-B (on the right). PAM preferences for LrCas9 protein and derived sequences to be tested

Frequency of nucleotides at each PAM position (figure 4.37-A) was calculated using an *in-silico* pipeline. Based on this logo, I derived the sequences TTTTT and TTCCT (figure 4.37-B), which were selected as putative PAM variants for the LrCas9. CCAAT was selected as negative control, while TTTCT was another possible positive sequence which had been already tested in bacteria with LrCas9.

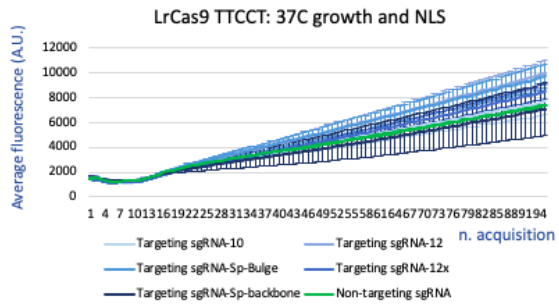
I tested each PAM sequence with five sgRNAs: sgRNA-10, sgRNA-12, sgRNA-sp-Bulge, sgRNA-12x, sgRNA with the SpCas9 backbone. Their design derived from the consensus structure of the crispr:tracrRNA duplex for the SpCas9 and published gRNA variants for the LrCas9 (figure 4.38-A)

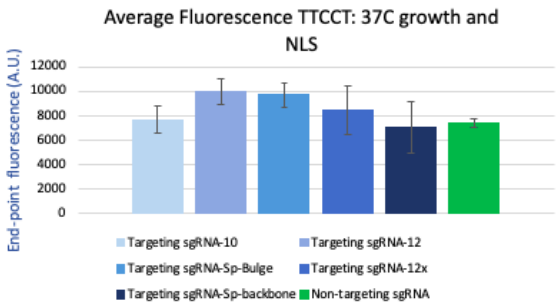
(Crawley et al., 2018); these variants derived from mutations made in the bulge (yellow) and the upper stream (purple) modules of the gRNA structure (figures 4.38-B and 4.38-C).
The sgRNA plasmids had the same structure as SpCas9 sgRNAs (figure 4.6 in this section).



Figures 4.38-A-B-C: Crispr:Tracr complex for LrCas9 (A) and mutations in the bulge (SG10, B) and upper stream (SG12, C) modules
The sgRNAs and the Cr-TrRNAs derived from the consensus structure of the LrCas9 Crispr:Tracr complex. Some gRNAs reported specific motifs in the bulge and in the upper stream modules compared to the wild-type structure.

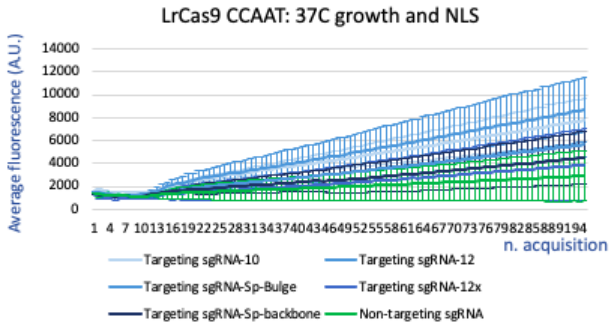
I perform the pilot experiment with the LrCas9 testing the sgRNAs with only two PAM sequences: TTCCT and CCAAT (negative control). The reason why I could not test also the other PAMs, TTTCT and TTTTT, was that I did not manage to generate the corresponding plasmid reporters. During the standard transformation with *E. coli*, required for making plasmids, I struggled with getting positive colonies: I always found plates with zero or white colonies only. White colonies meant that the deGFP was not expressed in *E. coli* and this could not be acceptable for a reporter plasmid that has to be expressed in a *E. coli* cell extract (MyTXTL®). I confirmed that the obtained plasmids were not strong promoters: the results were not good in terms of fluorescence production *in-vitro*.

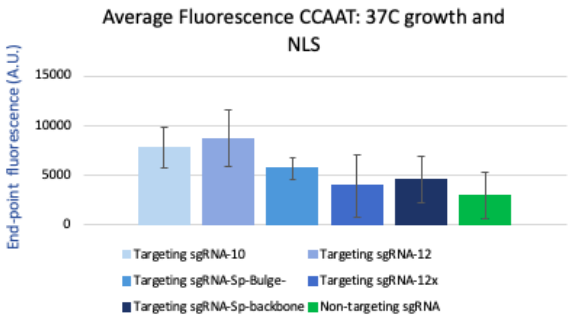




Figures 4.39 (on the left) and 4.40 (on the right). LrCas9 sgRNAs tested with the PAM TTCCT. The reporter had a different design compared to all the graphs from figure 4.46 (see below).

Same description as figures 4.41 and 4.42 below.





Figures 4.41 (on the left) and 4.42 (on the right). LrCas9 sgRNAs tested with the PAM TTCCT. The reporter had a different design compared to all the graphs below (from figure 4.46)

The activity of the LrCas9 was measured analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. In each testing, the Cas9 plasmid was incubated with the sgRNA (targeting or non-targeting), the deGFP reporter and the TXTL cell extract at 29C and the deGFP expression was monitored. Line graphs (figure 4.39 and 4.41) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 96 acquisitions (indicated in the X-axis) at an interval of 3 minutes. Bar graphs (figure 4.40 and 4.42) represent the average end-point value of fluorescence intensity (A.U.) ± SD (at ~5 hours from start of reaction).

Analysing figures from 4.39 to 4.42, TTCCT and CCAAT (negative control) reactions were characterized by very low fluorescence curves, values are comparable with usual basic noise. This could be justified because the two plasmid constructs were not strong enough to report the sgRNA and the LrCas9 activity. Therefore, I understood that having a good reporter system is mandatory

to analyse the whole CRISPR system in MyTXTL®: the reaction depends on the expression of all the components in the cell extract.

Consequently, I tried to improve the design of these reporters, in order to improve the deGFP expression in *E. coli* lysate that leads to a high fluorescence rate during the CRISPR reaction.

After some attempts, I successfully discovered that two changes in the plasmid design and in the transformation process allowed to generate a strong reporter and overcome the previous problems. Removing the nuclear localization sequence (NLS) in the plasmid and growing the plasmids in *E. coli* at 30C instead of at the standard 37C led to new reporter systems characterised by a higher deGFP expression in TXTL (figures 4.43 and 4.44).

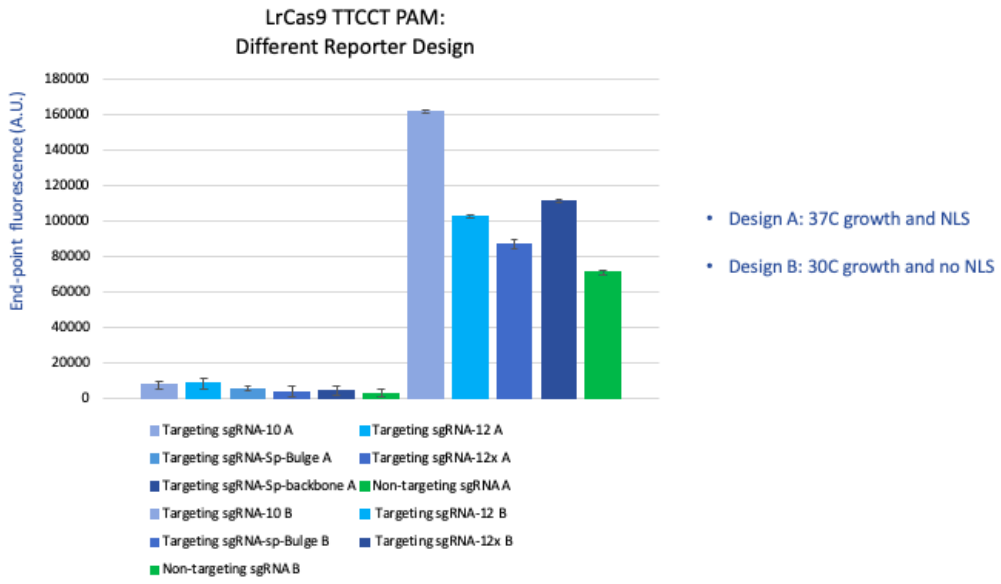


Figure 4.43: Fluorescence rate related to LrCas9 reporters characterised by two different design and carrying the PAM TTCCT

The activity of the LrCas9 was measured analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. The LrCas9 plasmid was incubated with the sgRNA (targeting or non-targeting), the deGFP reporter and the TXTL cell extract at 29C and the deGFP expression was monitored. The new reporter design did not include the NLS sequence and the plasmid growth in *E. coli* was set at 30C instead that at 37C. Bar graphs represent the average end-point value of fluorescence intensity (A.U.) ± SD (at ~5 hours from start of reaction).

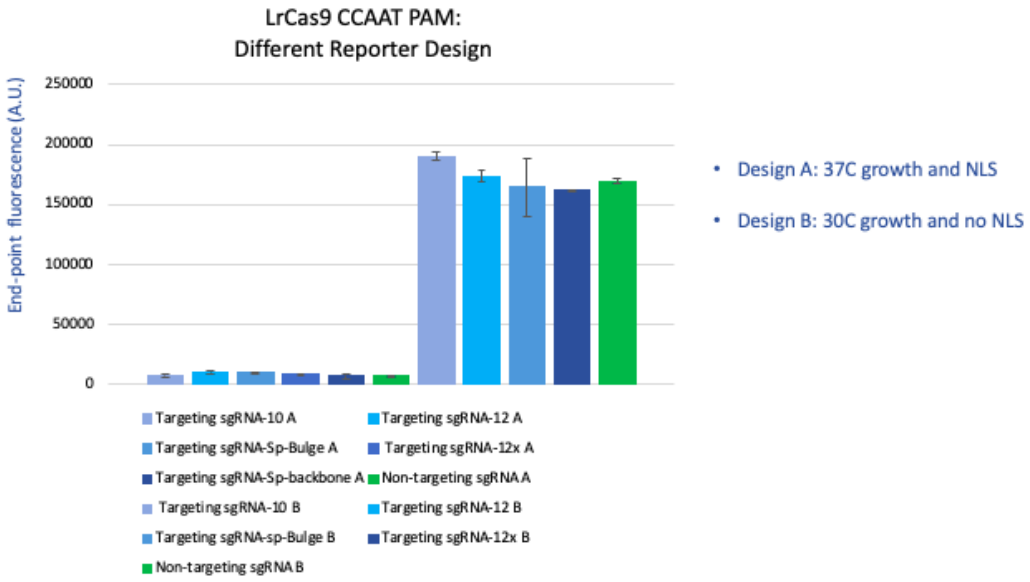
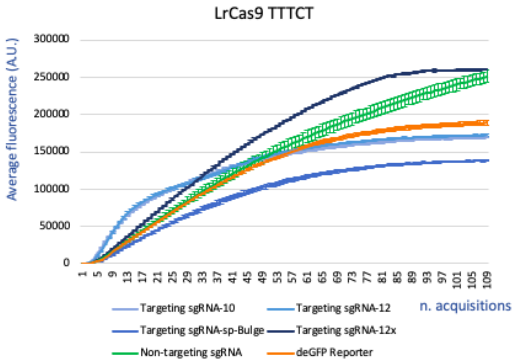
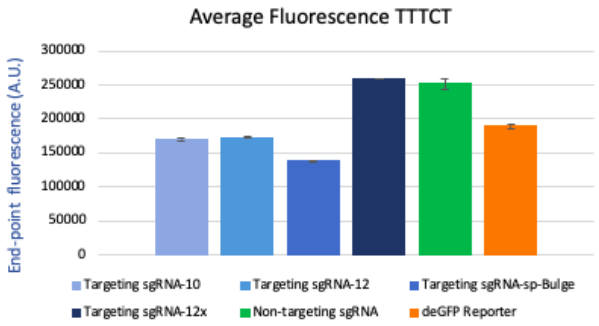


Figure 4.44: Fluorescence rate related to LrCas9 reporters characterised by two different design and carrying the PAM CCAAT

The activity of the LrCas9 was measured analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. The setting reaction was the same as figure 4.43. The new reporter design did not include the NLS sequence and the plasmid growth in *E. coli* was set at 30C instead that at 37C. Bar graphs represent the average end-point value of fluorescence intensity (A.U.) ± SD (at ~5 hours from start of reaction).

Once I improved the plasmid design, I could test the LrCas9 with all the selected PAM variants: TTTCT, TTTTT, TTCCT and CCAAT.

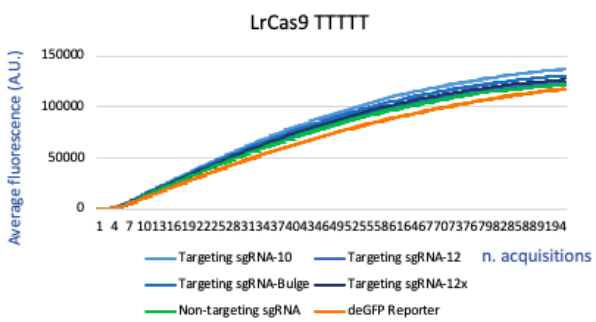


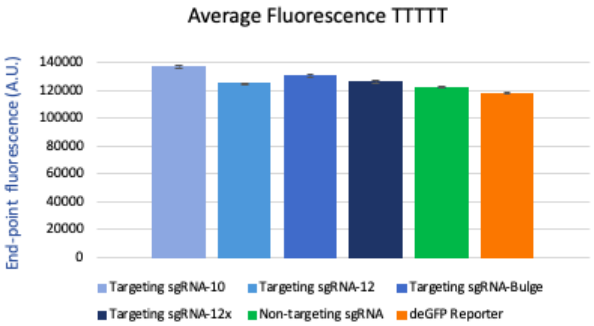


Figures 4.45 (on the left) and 4.46 (on the right). LrCas9 sgRNAs tested with the PAM TTTCT

The activity of the LrCas9 was measured analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. In each testing, the Cas9 plasmid was incubated with the sgRNA (targeting or non-targeting), the deGFP reporter and the TXTL cell extract at 29C and the deGFP expression was monitored. Line graphs (figure 4.45) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 96 acquisitions (indicated in the X-axis) at an interval of 3 minutes. Bar graphs (figure 4.46) represent the average end-point value of fluorescence intensity (A.U.) ± SD (at ~5 hours from start of reaction).

The LrCas9 tested with the PAM TTTCT gave positive results with the sgRNAs (figures 4.45 and 4.46). In particular, the best nuclease activity was detected with the sgRNA-sp-Bulge, but also sgRNA-10 and sgRNA-12 worked well in TXTL. At the same time, the sgRNA-12x gave negative results because it did not allow the DNA cleavage.

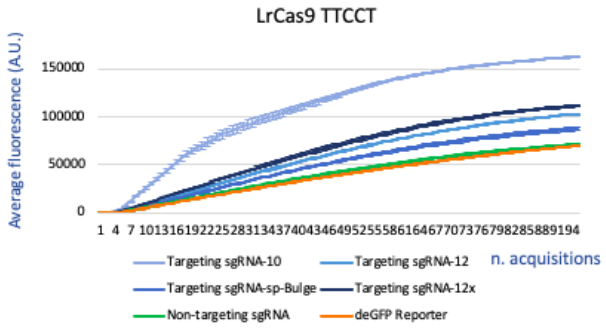


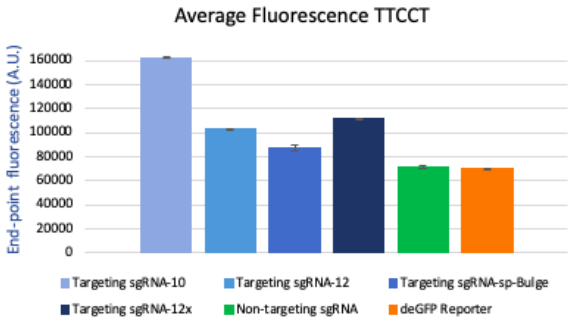


Figures 4.47 (on the left) and 4.48 (on the right). LrCas9 sgRNAs tested with the PAM TTTTT

The activity of the LrCas9 was measured analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. In each testing, the Cas9 plasmid was incubated with the sgRNA (targeting or non-targeting), the deGFP reporter and the TXTL cell extract at 29C and the deGFP expression was monitored. Line graphs (figure 4.47) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 96 acquisitions (indicated in the X-axis) at an interval of 3 minutes. Bar graphs (figure 4.48) represent the average end-point value of fluorescence intensity (A.U.) ± SD (at ~5 hours from start of reaction).

TTTTT PAM, which was the most promising sequence from the *in-silico* prediction (figures 4.37 A-B), did not gave the expected results. Surprisingly, when the LrCas9 was tested with this sequence (figures 4.47 and 4.48) no cleavage activity was detected *in-vitro*. Therefore, or the computational prediction regarding TTTTT was incorrect, or there was something in the sequence or in the secondary structure of the sgRNAs that could not allow a good expression of the gRNA in TXTL; consequently, also the synthesis of the related protein could have been affected.

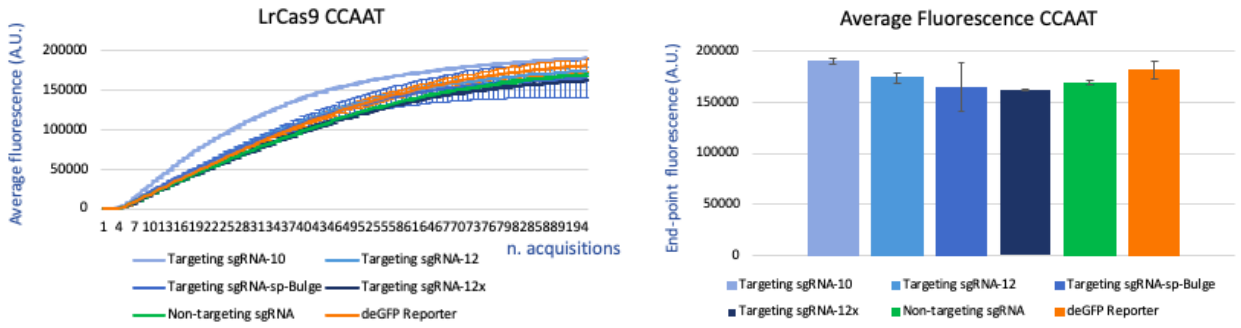




Figures 4.49 (on the left) and 4.50 (on the right). LrCas9 sgRNAs tested with the PAM TTCCT

The activity of the LrCas9 was measured analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. In each testing, the Cas9 plasmid was incubated with the sgRNA (targeting or non-targeting), the deGFP reporter and the TXTL cell extract at 29C and the deGFP expression was monitored. Line graphs (figure 4.49) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 96 acquisitions (indicated in the X-axis) at an interval of 3 minutes. Bar graphs (figure 4.50) represent the average end-point value of fluorescence intensity (A.U.) ± SD (at ~5 hours from start of reaction).

The other PAM sequence I derived from the *in-silico* analysis was TTCCT. This variant did nor give positive results when tested with all the sgRNAs (figures 4.49 and 4.50), the LrCas9 did not cleave the DNA. It may be that the design of the sgRNAs should be improved or, also in this case, the PAM prediction was incorrect. To answer this question, new gRNA variants could be engineered and their testing with the predicted PAMs TTTTT and TTCCT would confirm if those *in-silico* predictions were confirmed *in-vitro*.



Figures 4.51 (on the left) and 4.52 (on the right). LrCas9 sgRNAs tested with the PAM CCAAT

The activity of the LrCas9 was measured analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. In each testing, the Cas9 plasmid was incubated with the sgRNA (targeting or non-targeting), the deGFP reporter and the TXTL cell extract at 29C and the deGFP expression was monitored. Line graphs (figure 4.51) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 96 acquisitions (indicated in the X-axis) at an interval of 3 minutes. Bar graphs (figure 4.52) represent the average end-point value of fluorescence intensity (A.U.) ± SD (at ~5 hours from start of reaction).

The CRISPR reaction with the PAM CCAAT gave expected results. Focusing on figures 4.51 and 4.52, the curves related to targeting and non-targeting sgRNAs were comparable, no cleavage activity was detected *in-vitro*. Consequently, I could confirm that the CCAAT PAM was a good prediction as negative control for LrCas9.

4.2 Overcoming cell toxicity during Cas9 cloning: using amplicon templates for Cas9 expression

4.2.1 SpCas9 linear template at 29C

One of the main limitations which reduces the range of constructs which could potentially be tested in TXTL is the cytotoxicity in bacteria. To make a plasmid, there is a step that is performed in *E. coli*, the transformation, that leads to the generation of bacterial colonies that include the plasmid of interest. When plasmids are very toxic in *E. coli*, their transformation fails, colonies are nor generated and, consequently, the final construct required for the reaction could not be obtained. In

these conditions, all templates whose plasmids could not be generated, should not be tested in MyTXTL®, reducing the application of this tool in large scale.

In order to overcome this constraint, I introduced the use of linear templates for *in-vitro* CRISPR reactions: this approach does not require the cloning process, the gene of interest could be added in the TXTL mix directly in amplicon format. The only requirement should be the use of specific promoters and terminators, which represent the initial/end points for the template amplification. Because the gene of interest had to be expressed in the *E. coli* cell extract, I selected the J23119 and the rrnC bacterial promoter and terminator as default for all the amplifications (figure 4.53).

The choice of these specific promoter and terminator was set because this combination gave the best Cas9 amplification compared with other promoter-terminator associations (in particular, versus the same templates combined with SpCas9 promoter or P70 promoter and tonB terminator or T500 terminator).

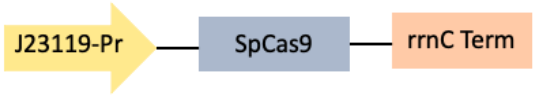


Figure 4.53: Structure of the SpCas9 linear template

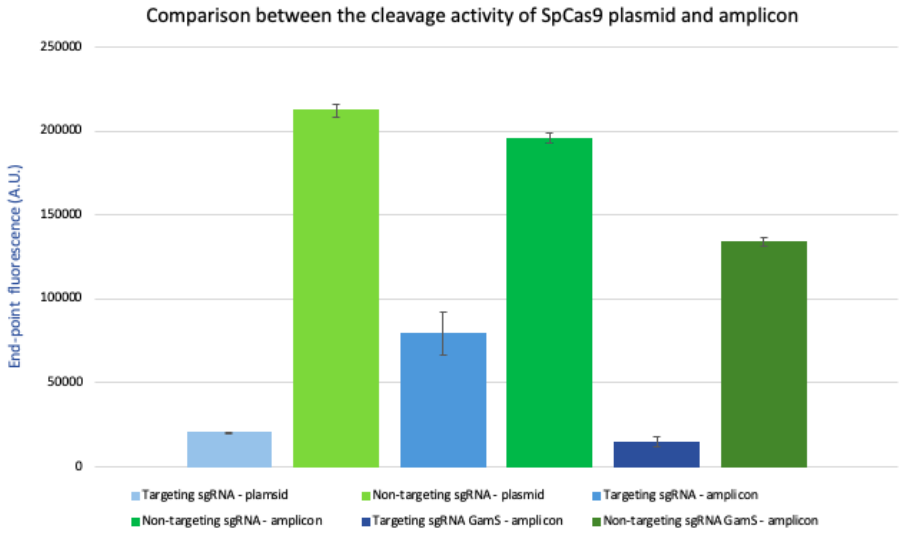
The SpCas9 in linear format is characterised by only three components: the J23119 promoter, which is responsible for directing the bacterial RNA polymerase in MyTXTL® cell extract; the Cas9 sequence; the bacterial terminator rrnC, which stops the transcriptional machinery in the bacterial lysate.

Compared to circular DNA, one of the risks when using the amplicons could be their greater degradation in the master mix medium because of the less stability. In order to prevent the DNA degradation, in combination with the TXTL *in-vitro* expression system, the linear DNA could be stabilized at the terminal ends with proteins which are inhibitors of the *E. coli* endogenous exonucleases RecBCD (myTXTL® GamS Protein, Arbor Biosciences), removing any possible interference from the background.

Therefore, an easy template amplification allows to make also those CRISPR components, such as Cas9, gRNAs, that were difficult to be obtained via cloning assembly, increasing the number of reactions which could potentially be performed with this platform.

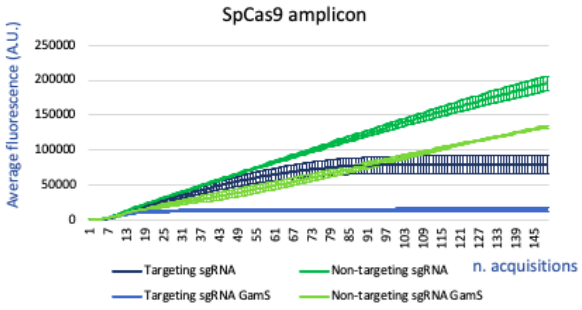
In order to validate the TXTL with amplicon templates, first I performed the reaction with the SpCas9, considering I already tested it also in the plasmid version (figure 4.6 above in this chapter); therefore, I had a positive control regarding this Cas9/gRNA/reporter combination.

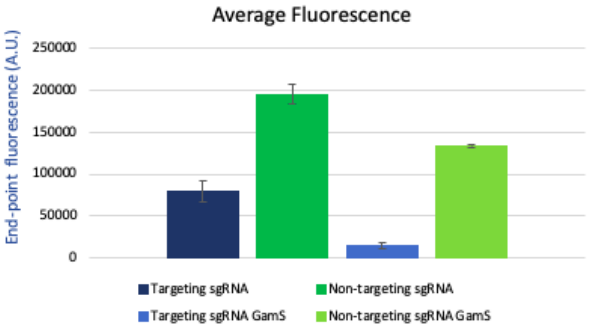
The TXTL performed with the SpCas9 amplicon confirmed the same results I previously obtained with the same endonuclease/sgRNA/reporter combination, but with the SpCas9 in plasmid format (figure 4.54): this gRNA design associated with this target site induced the DNA cleavage.



Figures 4.54: SpCas9 related cleavage activity in both plasmid and amplicon formats

The activity of the SpCas9 was measured analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. In each reaction, the SpCas9 template was tested as amplicon or plasmid and was incubated with the sgRNA plasmid (targeting or non-targeting), the deGFP reporter plasmid and the TXTL cell extract at 29C and the deGFP expression was monitored. Bar graphs represent the average end-point value of fluorescence intensity (A.U.) ± SD (at ~5 hours from start of reaction).





Figures 4.55 (left) and 4.56 (right). SpCas9 amplicon at 29C

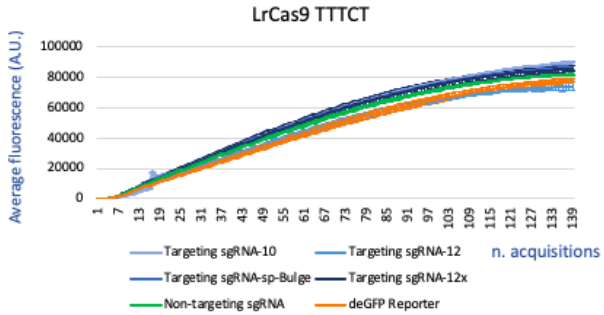
The activity of the SpCas9 was measured analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. In each testing, the SpCas9 amplicon was incubated with the sgRNA plasmid (targeting or non-targeting), the deGFP reporter plasmid and the TXTL cell extract at 29C and the deGFP expression was monitored. Line graphs (figure 4.55) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 150 acquisitions (indicated in the X-axis) at an interval of 3 minutes. Bar graphs (figure 4.56) represent the average end-point value of fluorescence intensity (A.U.) ± SD (at ~5 hours from start of reaction).

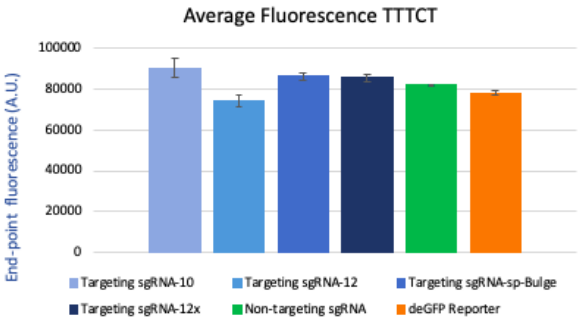
I repeated the same reaction with and without the inhibitors of the *E. coli* endogenous exonucleases RecBCD (myTXTL® GamS Protein, Arbor Biosciences). Analysing figures 4.55 and 4.56, I could observe how in targeting conditions the sgRNA allowed the DNA cleavage both with and without the GamS protein. In particular, when these proteins were added, the trends of both targeting and non-targeting conditions were lower than the curves where they were not used. This meant that the linear DNA was better stabilised, keeping the fluorescence lower than the reaction without the GamS.

Once I assessed that the CRISPR *in-vitro* testing could be performed with both plasmid and linear templates, I proceeded with the testing of other amplicons.

4.2.2 LrCas9 linear template at 29C

I proceeded with the *in-vitro* testing of LrCas9 amplicon with both related sgRNAs and Cr-TrRNAs, considering I already tested this endonuclease as plasmid, so the obtained results could be compared. I evaluated the cleavage activity directed to the same target site and associated with the same PAMs I derived *in-silico*: TTTCT, TTTTT, TTCCT and CCAAT (negative control).



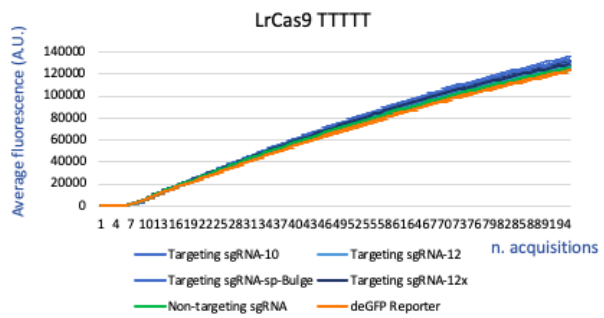


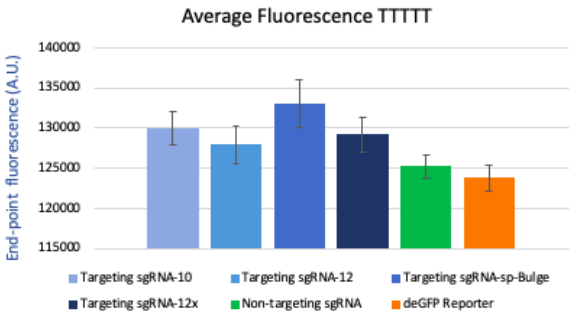
Figures 4.57 (on the left) and 4.58 (on the right). LrCas9 amplicon tested with the PAM TTTCT and the sgRNAs

The activity of the LrCas9 amplicon was measured analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. In each testing, the Cas9 amplicon was incubated with the gRNA amplicon (targeting or non-targeting), the deGFP reporter plasmid and the TXTL cell extract at 29C and the deGFP expression was monitored. Line graphs (figure 4.57) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 96 acquisitions (indicated in the X-axis) at an interval of 3 minutes. Bar graphs (figures 4.58) represent the average end-point value of fluorescence intensity (A.U.) ± SD (at ~5 hours from start of reaction).

Focusing on LrCas9 amplicon tested with the PAM TTTCT, I could not confirm the same results I previously got with the plasmid template and the same Cas9/gRNA/reporter combinations (figures 4.45 and 4.46). When testing in *E. coli* cell extract the LrCas9 in amplicon format, (figures 4.57 and 4.58), the fluorescence related to sgRNA-10, 12 and sp-Bulge did not decrease, in comparison with the non-targeting sgRNA. No cleavage activity was detected when testing these sgRNAs.

I should better investigate why I got these different trends after testing a plasmid or a linear template related to the same Cas9. Considering in this experiment with the amplicons I did not add the GamS proteins to stabilise the linear DNA, the Cas9 template may have been decomposed at a very early stage in the reaction because of its major instability in the cell extract; consequently, the related protein could not have been synthesised in sufficient quantity to evaluate the cleavage activity.

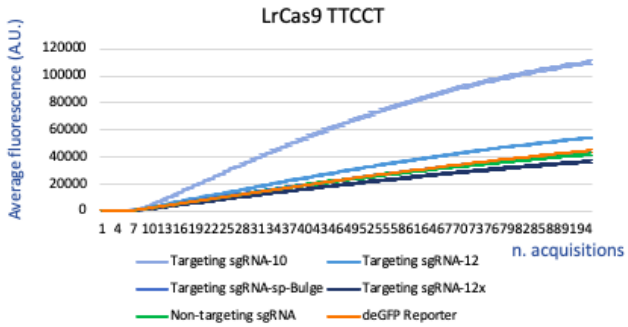


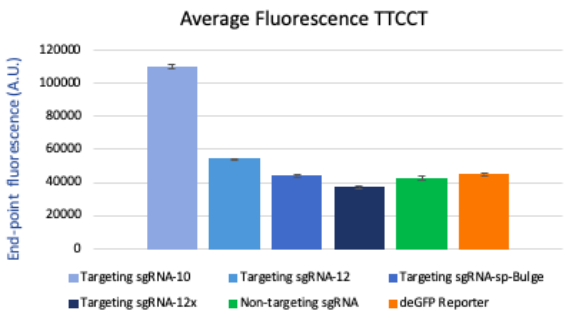


Figures 4.59 (on the left) and 4.60 (on the right). LrCas9 amplicon tested with the PAM TTTTT and the sgRNAs

The activity of the LrCas9 amplicon was measured analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. In each testing, the Cas9 amplicon was incubated with the gRNA plasmid (targeting or non-targeting), the deGFP reporter plasmid and the TXTL cell extract at 29C and the deGFP expression was monitored. Line graphs (figure 4.59) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 96 acquisitions (indicated in the X-axis) at an interval of 3 minutes. Bar graphs (figure 4.60) represent the average end-point value of fluorescence intensity (A.U.) ± SD (at ~5 hours from start of reaction).

Results related to the LrCas9 amplicon tested with the PAM TTTTT (figures 4.59 and 4.60) confirmed the results I obtained with the same Cas9/gRNA/reporter combination but with the plasmid LrCas9 as template (figures from 4.47 to 4.48). With this PAM sequence, all the designed sgRNAs were not able to guide the Cas9 and allow the DNA cleavage.

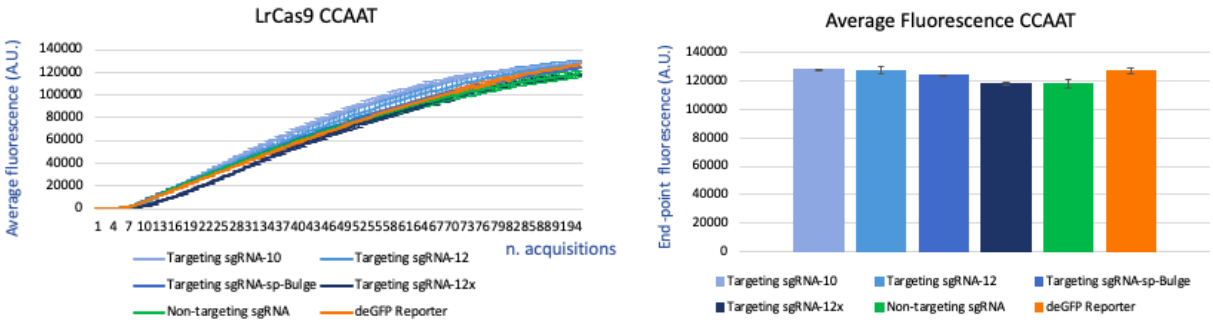




Figures 4.61 (on the left) and 4.62 (on the right). LrCas9 amplicon tested with the PAM TTCCT and the sgRNAs

The activity of the LrCas9 amplicon was measured analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. In each testing, the Cas9 amplicon was incubated with the gRNA plasmid (targeting or non-targeting), the deGFP reporter plasmid and the TXTL cell extract at 29C and the deGFP expression was monitored. Line graphs (figure 4.61) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 96 acquisitions (indicated in the X-axis) at an interval of 3 minutes. Bar graphs (figure 4.62) represent the average end-point value of fluorescence intensity (A.U.) ± SD (at ~5 hours from start of reaction).

With the PAM TTCCT, the CRISPR testing performed with the sgRNAs gave exactly the same fluorescence curves as with the LrCas9 plasmid (figures 4.49 and 4.50). In fact, in both reactions, I got almost the same trends for the sgRNAs 12, sp-Bulge and 12x compared to the non-targeting curve (figures 4.61 and 4.62); furthermore, the sgRNA-10 gave the highest fluorescence values, even more than the non-targeting sgRNA, and the same previously happened with the LrCas9 plasmid (figures 4.49 and 4.50). This phenomenon should be further investigated, because there could have been an issue with the *in-vitro* expression of the non-targeting sgRNA, compromising the synthesis of its related protein.



Figures 4.63 (on the left) and 4.64 (on the right). LrCas9 amplicon tested with the PAM CCAAT and the sgRNAs

The activity of the LrCas9 amplicon was measured analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. In each testing, the Cas9 amplicon was incubated with the gRNA plasmid (targeting or non-targeting), the deGFP reporter plasmid and the TXTL cell extract at 29C and the deGFP expression was monitored. Line graphs (figure 4.63) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 96 acquisitions (indicated in the X-axis) at an interval of 3 minutes. Bar graphs (figure 4.64) represent the average end-point value of fluorescence intensity (A.U.) ± SD (at ~5 hours from start of reaction).

Data related to the LrCas9 testing with the PAM CCAAT (figures 4.63 and 4.64) reported that this sequence did not allowed the DNA cleavage: the fluorescence curves were comparable for both targeting and non-targeting sgRNAs (same as for LrCas9 plasmids, figures 4.51 and 4.52). This confirmed that CCAAT could be considered as a negative control for further PAM investigations.

4.2.3 BfCas9 linear template at 29C

I introduced the use of amplicons in TXTL reactions in order to allow the *in-vitro* testing of all possible templates, that would be ideal in a large-scale screening platform. Amplicons might easily be used when the related plasmid could not be generated because of cytotoxicity occurring during the transformation step in *E. coli*.

Amplicon validation allowed us to perform the CRISPR testing of another Cas9 orthologue whose PAM sequence and sgRNA have not been characterized yet: Bacteroides fragilis Cas9 (BfCas9).

As for the LrCas9 orthologue, our aim was to broaden the PAM specificity for the endonuclease, trying to discover PAM-less CRISPR systems that would allow gene editing without any limitation. Therefore, I tried to analyse the BfCa9 cleavage activity associated with different PAM variants.

By using a computational pipeline based on sequence alignments, I got a sequence logo (figure 4.65-A) which reports the conservation of nucleotide positions in the 5'-end flanking sequences of the target sites. Starting from this *in-silico* derived logo, I determined the PAM variants for BfCas9: GGGAACG, GAGTACA and GTTCAAT (this sequence was set as negative control) (figure 4.65-B). The putative PAM specificities of BfCas9 showed an extended 5'-NGG-3' PAM, with a T and A preference on position 4. I thank Nace Kranjc, PhD student in the Crisanti Lab at Imperial College London, for the bioinformatic analysis.

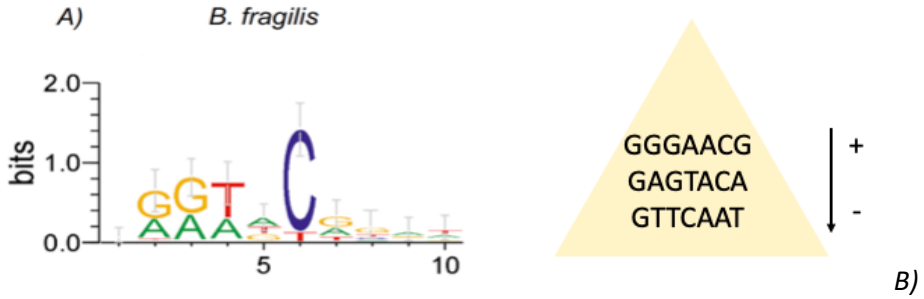


Figure 4.65-A-B: PAM preferences for BfCas9 protein and derived sequences to be tested

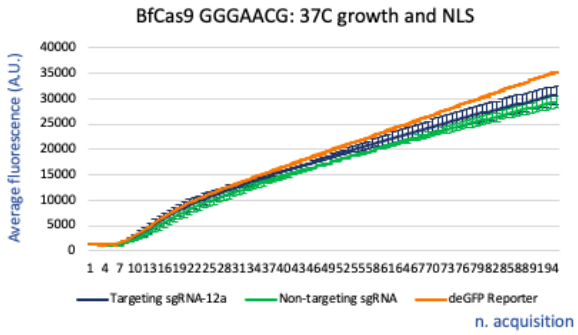
Putative PAM sequences identified for BfCas9 orthologue using *in-silico* PAM identification pipeline (B): GGGAACG, GAGTACA and GTTCAAT (negative control). Sequence logos represent conservation on nucleotides positions in the 5—end flanking regions (A).

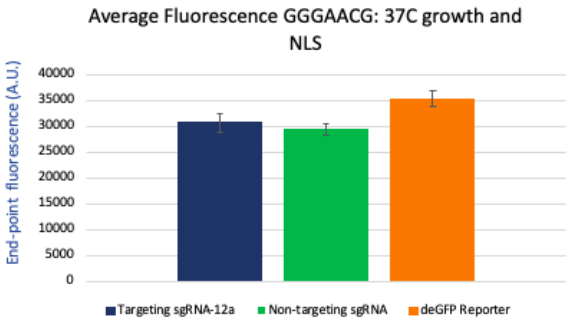
Once I derived the PAM variants (figure 4.65-B), I designed four sgRNAs for BfCas9 based on the sequence and secondary structure of the SpCas9 sgRNA and published gRNAs variants for LrCas9 (Crawley et al., 2018). The gRNAs I designed derived from mutations within the lower stem, bulge and nexus domains; therefore, I generated BfCas9 sgRNA-long, sgRNA-10, sgRNA-12, sgRNA-12a and sgRNA-sp-Bulge. The design of the sgRNAs was the same as for SpCas9 and LrCas9 (figure 4.6 in this section).

As it happened also at the beginning for LrCas9 (see section 4.1.2), I had significant problems about the expression of both the sgRNAs and the reporters I designed for BfCas9. In the pilot experiment to analyse the activity of BfCas9, I tested the sgRNA-12a with only the reporter carrying the PAM NGGAACG. I did not manage to generate the other reporter plasmids with the GAGTACA and GTTCAAT sequences; therefore, I could not compare this result with the negative control PAM.

The main problems I had while cloning the reporter plasmids were the frequent mutations I found in the final sequences and I got zero or few green colonies with the standard bacterial growth at 37C.

Consequently, I managed to perform only one reaction with BfCas9 orthologue.





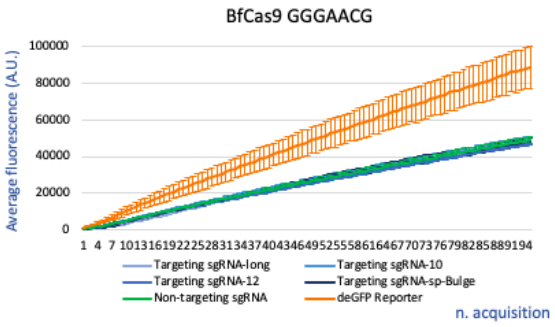
Figures 4.66 (on the left) and 4.67 (on the right). BfCas9 sgRNAs tested with the PAM NGGAACG.

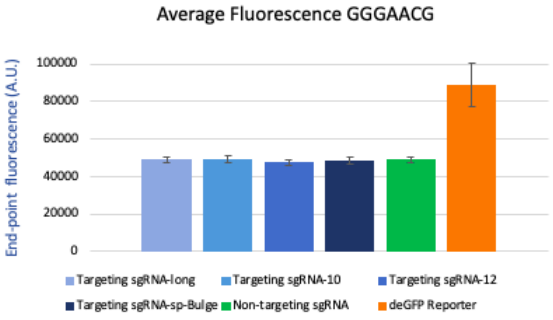
The activity of the BfCas9 was measured analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. In each testing, the Cas9 amplicon was incubated with the sgRNA (targeting or non-targeting), the deGFP reporter and the TXTL cell extract at 29C and the deGFP expression was monitored. Line graphs (figure 4.66) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 96 acquisitions (indicated in the X-axis) at an interval of 3 minutes. Bar graphs (figure 4.67) represent the average end-point value of fluorescence intensity (A.U.) ± SD (at ~5 hours from start of reaction).

Analysing figures 4.66 and 4.67, I have no evidence of BfCas9 activity with the NGGAACG PAM. However, prior to understand if the issue was that the BfCas9 was not properly expressed *in-vitro* or if I had to improve the design of the sgRNA, first I focused on the main problem: I did not manage to clone the reporters carrying the other PAMs. Different approaches could be introduced to overcome cloning problems: the use of amplicons also for the reporter component in MyTXTL® reactions, that would completely avoid the whole cloning process, or modifications during the plasmid assembly.

Therefore, I tried to introduce the same changes that allowed us to generate strong reporter systems for LrCas9, such as remotion of the NLS sequence in the plasmid and bacterial growth of the construct at 30C instead that at 37C. The NLS was initially added in the plasmid because I wanted a reporter that could be used both in TXTL as well cell/*in-vivo* testing. With this new design and new growing conditions, I managed to obtain several green colonies on *E. coli* plates for all the reporter constructs; green colonies confirmed that the deGFP was expressed in these bacteria, that is promising for further testing in MyTXTL®.

These changes in the cloning protocol allowed us to get all the missing reporters without any mutations: I could proceed with their use in BfCas9 *in-vitro* testing.

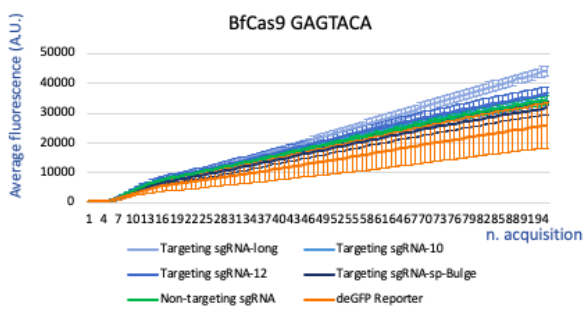


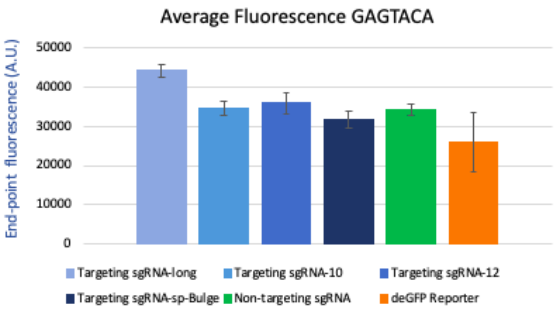


Figures 4.68 (on the left) and 4.69 (on the right). BfCas9 sgRNAs tested with the PAM NGGAACG

The activity of the BfCas9 was measured analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. In each testing, the Cas9 amplicon was incubated with the sgRNA (targeting or non-targeting), the deGFP reporter and the TXTL cell extract at 29C and the deGFP expression was monitored. Line graphs (figure 4.68) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 96 acquisitions (indicated in the X-axis) at an interval of 3 minutes. Bar graphs (figure 4.69) represent the average end-point value of fluorescence intensity (A.U.) ± SD (at ~5 hours from start of reaction).

Results derived from the *in-vitro* testing of BfCas9 with the PAM GGGAACG (figures 4.68 and 4.69) did not report DNA cleavage. In fact, there was no difference between targeting and non-targeting sgRNAs: their curves were overlapping.



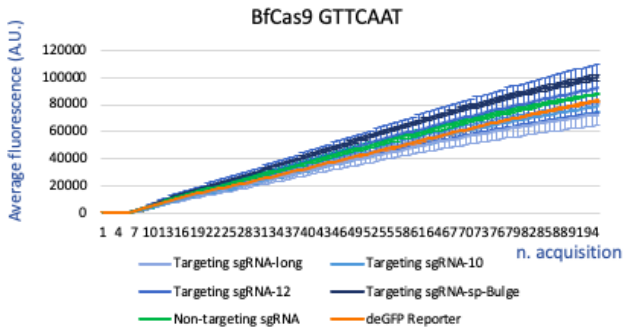


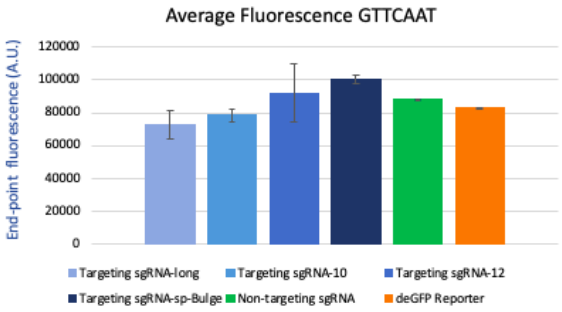
Figures 4.70 (on the left) and 4.71 (on the right). BfCas9 sgRNAs tested with the PAM NAGTACA

The activity of the BfCas9 was measured analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. In each testing, the Cas9 plasmid was incubated with the sgRNA (targeting or non-targeting), the deGFP reporter and the TXTL cell extract at 29C and the deGFP expression was monitored. Line graphs (figure 4.70) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 96 acquisitions (indicated in the X-axis) at an interval of 3 minutes. Bar graphs (figure 4.71) represent the average end-point value of fluorescence intensity (A.U.) ± SD (at ~5 hours from start of reaction).

Not even the second PAM mutant GAGTACA allowed the DNA cleavage. All the targeting sgRNAs gave the same trends, comparable with that of the non-targeting sgRNAs (figures 4.70 and 4.71). Considering both derived PAMs gave the same results, I should investigate if the main problem is the PAM variant which is not the good sequence to allow the cutting or if the gRNA design should be improved.

An idea could be the making of other reporter variants: given that the derived PAM motives for BfCas9 were NGGAACG and NAGTACA, the assumption N=G I used for the first position in the sequence could be modified with other nucleotides. This approach will allow to analyse different PAM variants derived from the same ancestral motif, trying to understand possible changes in the BfCas9 activity with a single substitution in the PAM.





Figures 4.72 (on the left) and 4.73 (on the right). BfCas9 sgRNAs tested with the PAM NTTCAAT

The activity of the BfCas9 was measured analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. In each testing, the Cas9 plasmid was incubated with the sgRNA (targeting or non-targeting), the deGFP reporter and the TXTL cell extract at 29C and the deGFP expression was monitored. Line graphs (figure 4.72) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 96 acquisitions (indicated in the X-axis) at an interval of 3 minutes. Bar graphs (figure 4.73) represent the average end-point value of fluorescence intensity (A.U.) ± SD (at ~5 hours from start of reaction).

The PAM I selected as negative control, GTTCAAT, gave no cutting evidence (figures 4.72 and 4.73). Based on all these experiments with the BfCas9, I cannot exclude that for some reasons the Cas9 was not properly produced in the TXTL master mix; therefore, these results could both mean that the main problem was the inexpression of the nuclease or if the design of some components (the gRNAs, the Cas9 or the reporter) should be improved.

However, even if from these data I could not find a good sgRNA-reporter combination which induced the cleavage by the Cas9, I successfully improved the making of most of these plasmids, which was a significant progress.

4.3 Evaluating the activity of Crispr and Tracr-RNAs

The gRNA has a key role in driving the nuclease in correspondence with a specific target site and it comprises a duplex including two RNA components: crRNAs and tracrRNAs. The first fragment interacts with the nuclease, while the tracr base pair with the crispr to form a unique secondary structure that allows the DNA cleavage. The development of a synthetic sgRNA that combines both

crispr and tracr components into a single RNA molecule by fusing the 3' end of the crRNA to the 5' of the tracrRNA has represented a significant improvement for expanding the CRISPR targeting. This customisable structure allows to easily modify only the gRNA component that depends on the target (the crispr-RNA), allowing to potentially direct the CRISPR system to cleave any DNA sequence (Jiang et al., 2017; Jinek et al., 2012). Moreover, the single guide is characterised by a shorter complex sequence compared to the Cr-TrRNA; this means that its use leads to a reduction of steric effects that could interfere with the Cas9/gRNA interaction.

However, there are several Cas9 orthologues which still do not have a characterised sgRNA structure and the Cr-Tr design is the only available for their application.

Different aspects could induce a diverse behaviour of sgRNA and Cr-TrRNA derived from the same guide RNA template: some Cr-TrRNA design could have a long sequence that causes steric effects; the *in-silico* sgRNA prediction, derived from the crispr:tracr duplex of already known gRNA variants, could be incorrect; other design could have particular secondary structures that influence their function in the CRISPR system.

It is important to compare the activity of both the sgRNA and the Cr-TrRNAs, in order to understand if a different secondary structure could affect the gRNA activity in MyTXTL®, leading to the selection of the design which gave the best activity *in-vitro*. Then, this construct could be used as default for the development of new gRNA variants.

A platform for the screening in large scale of genetic editors should be validated with all possible design of gRNA constructs, including both single guide and Cr-TrRNAs. This would be ideal especially when testing new Cas9 orthologues that could not have a sgRNA structure already well characterised.

4.3.1 Cr-TrRNAs designed for SpCas9

Since I always validated various aspects of this testing platform with a well-characterised nuclease as proof of principle, the SpCas9, first I derived two Cr-TrRNAs design associated with this protein. Then, these plasmids were used for the development of Cr-TrRNAs constructs associated with other Cas9 orthologues.

I derived the sequence of the SpCas9 Cr-TrRNAs (G1) from the design of the SpCas9 sgRNA plasmid (figure 4.6) that worked well in MyTXTL®. I assembled the new G1 construct with the same bacterial promoter, J23119, which previously allowed the sgRNA *in-vitro* expression in the *E. coli* cell extract. In addition to the targeting G1 Cr-TrRNA, I generated also a non-targeting version, by modifying the sequence of the Crispr RNA. This plasmid was referred as negative control.

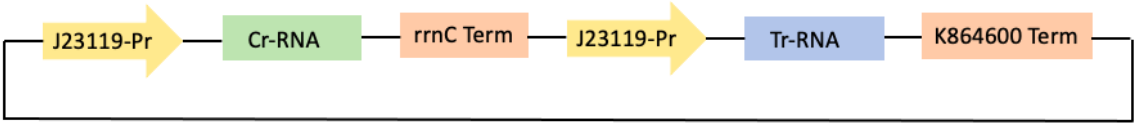
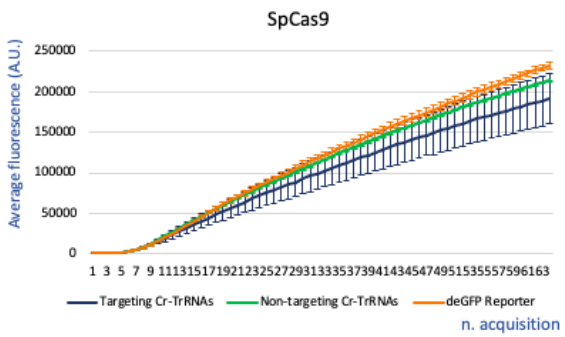
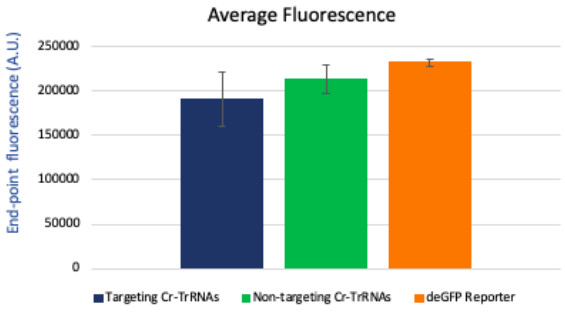


Figure 4.74. Design of the G1 plasmid, the SpCas9 Cr-TrRNAs

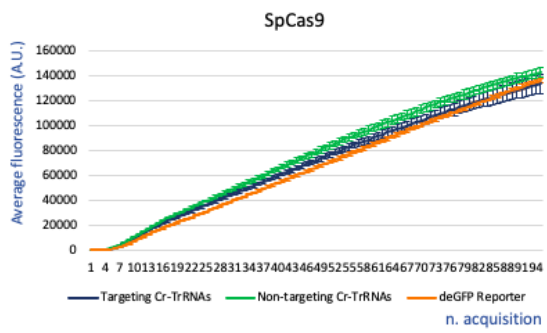
The sequence contains the main regions of the Cr-TrRNA plasmid G1: upstream the Crispr and the Tracr RNAs is located the bacterial promoter J23119. It is responsible for the gene activation in the bacterial cell extract; then, there are the two guide sequences, each of them followed by a bacterial terminator: the rrnC for the Crispr RNA and the K864600 for the Tracr RNA.

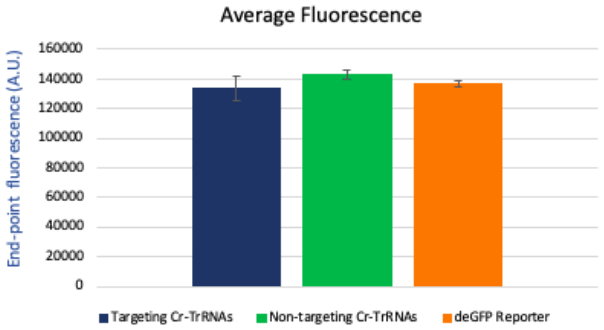




Figures 4.75 (left) and 4.76 (right). SpCas9 Cr-TrRNAs G1 at 29C

The activity of the SpCas9 was detected analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. In each testing, the SpCas9 plasmid was incubated with the Cr-TrRNAs (targeting or non-targeting), the deGFP reporter and the TXTL cell extract at 29C and the deGFP expression was monitored. Line graphs (figure 4.75) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 65 acquisitions (indicated in the X-axis) at an interval of 3 minutes. Bar graphs (figure 4.76) represent the average end-point value of fluorescence intensity (A.U.) ± SD (at ~5 hours from start of reaction).





Figures 4.77 (left) and 4.78 (right). SpCas9 Cr-TrRNAs G1 at 29C, 2 steps reaction

The activity of the SpCas9 was detected analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. In the 2 steps reaction, the SpCas9 was previously pre-incubated at 29C for 30 min with the gRNA (targeting or non-targeting) and the TXTL cell extract only, without the deGFP reporter. After 30 min, the deGFP reporter was added to the pre-incubation reaction and the final reaction was incubated at 29C and the deGFP expression was monitored. Line graphs (figure 4.77) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 96 acquisitions (indicated in the X-axis) at an interval of 3 minutes. Bar graphs (figure 4.78) represent the average end-point value of fluorescence intensity (A.U.) ± SD (at ~5 hours from start of reaction).

As for the sgRNA reactions, also with the Cr-TrRNAs I performed both one and two step reactions, with and without the pre-incubation step when the reporter was temporary removed from the mix (see section 4.1.1 above in this chapter). However, in both cases (figures from 4.75 to 4.78) I could not observe any activity from the SpCas9. This might be due to the design of the Cr-TrRNAs which did not allow a good gRNA expression (e.g., wrong promoter or terminators) or there could be something in the sequence itself that affects the secondary structure of the gRNA, influencing its activity in the cell extract.

Therefore, using as default the sgRNA which worked well in TXTL (figure 4.6), I decided to optimise the design of the Cr-TrRNA plasmid trying to make it more similar to the single guide structure. The main difference, apart from the gRNA sequences, was the terminator. In fact, in the sgRNA (figure 4.7) there was the bacterial terminator aspA, while in the plasmid G1 (figure 4.74) there were two different terminators, rrnC for the Crispr RNA and K864600 for the Tracr RNA. Consequently, I decided to design another plasmid, the G2 (figure 4.81), where I inserted the same terminator than the single guide (aspA), located downstream both the Crispr and Tracr RNA sequences.

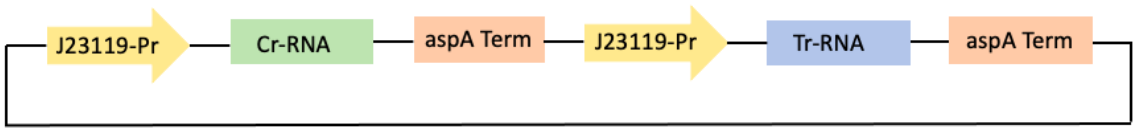
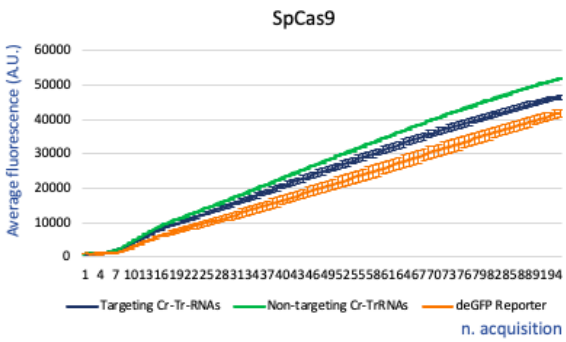
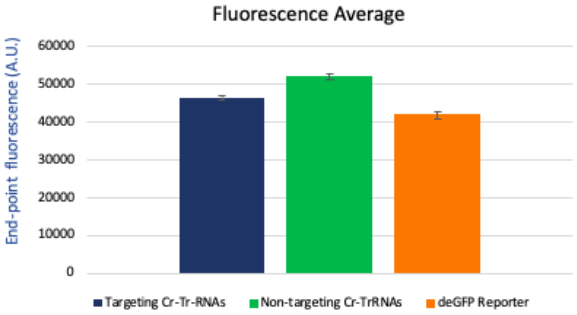


Figure 4.79. Design of the G2 plasmid, the SpCas9 Cr-TrRNAs

The sequence contains the main regions of the Cr-TrRNA plasmid G2: upstream the Crispr and the Tracr RNAs is located the bacterial promoter J23119. It is responsible for the gene activation in the bacterial cell extract; then, there are the two guide sequences, each of them followed by the bacterial terminator aspA.





Figures 4.80 (left) and 4.81 (right). SpCas9 Cr-TrRNAs G2 at 29C

The activity of the SpCas9 was detected analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. In each testing, the SpCas9 plasmid was incubated with the Cr-TrRNAs (targeting or non-targeting), the deGFP reporter and the TXTL cell extract at 29C and the deGFP expression was monitored. Line graphs (figure 4.80) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 96 acquisitions (indicated in the X-axis) at an interval of 3 minutes. Bar graphs (figure 4.81) represent the average end-point value of fluorescence intensity (A.U.) ± SD (at ~5 hours from start of reaction).

Focusing on figures 4.80 and 4.81, I could see as no cleavage activity was observed.

Even if I modified the design of the plasmid by using the same bacterial terminator as the sgRNA, I did not improve the Cr-TrRNA activity. Targeting and non-targeting conditions were comparable, which means that also the G2 structure was not able to guide the SpCas9 in correspondence with the target site, where the cleavage was expected.

Consequently, there should be something else in the sequence of both G1 and G2 Cr-TrRNAs which obstructed the whole CRISPR machinery *in-vitro*; probably, their secondary structures, more

complex than the sgRNA one, have a key role in the final shape of the protein, with a consequent effect on the gRNA activity.

4.3.2 Cr-TrRNAs designed for LrCas9

Making Cr-TrRNA variants is important especially for those Cas9 that still do not have a characterised sgRNA. New structures, new design have to be tested, in order to find a valid plasmid that allows the DNA cleavage. Therefore, apart from the SpCas9, I focused on another orthologue that I previously tested with different sgRNA: LrCas9.

In parallel with the G1 construct (figure 4.74), I generated also a Cr-TrRNA variant carrying the crispr and tracr sequences for LrCas9 (figure 4.82), that I derived from published gRNA variants for lactobacilli (Crawley et al., 2018).

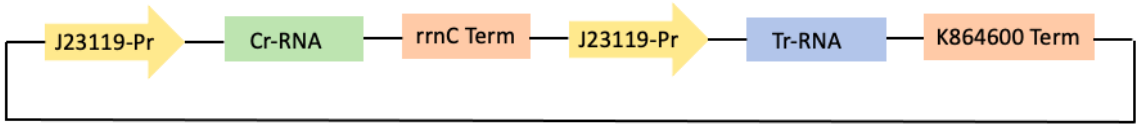
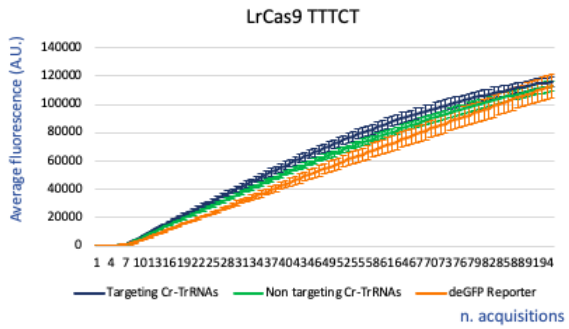
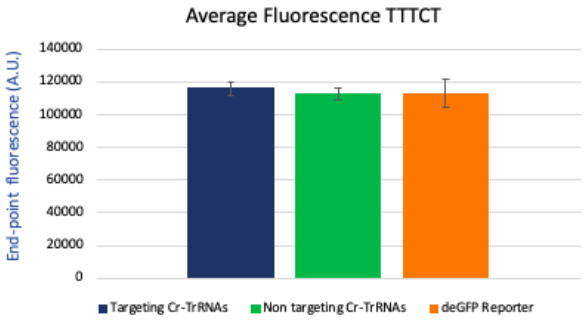


Figure 4.82. Design of the LrCas9 Cr-TrRNA plasmid

The sequence contains the main regions of the LrCas9 Cr-TrRNA plasmid: upstream the Crispr and the Tracr RNAs is located the bacterial promoter J23119. It is responsible for the gene activation in the bacterial cell extract; then, there are the two guide sequences, each of them followed by a bacterial terminator: the rrnC for the Crispr RNA and the K864600 for the Tracr RNA.

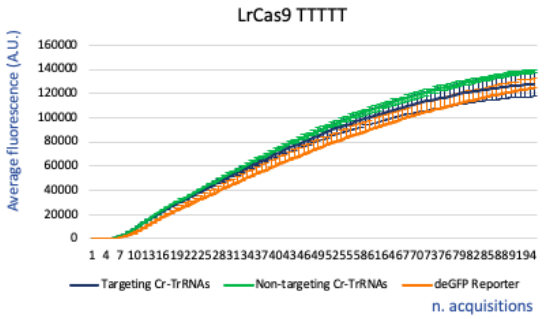
Once I generated the LrCas9 Cr-TrRNAs, I proceeded with the testing of previously *in-silico* derived PAM sequences: TTTCT, TTTTT, TTCCT and CCAAT (negative control).

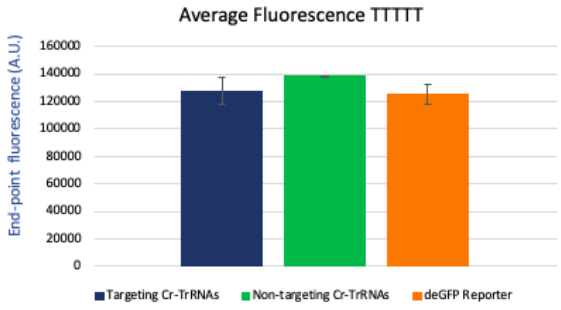




Figures 4.83 (on the left) and 4.84 (on the right). LrCas9 Cr-TrRNAs tested with the PAM TTTCT

The activity of the LrCas9 was detected analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. In each testing, the Cas9 plasmid was incubated with the Cr-TrRNAs (targeting or non-targeting), the deGFP reporter and the TXTL cell extract at 29C and the deGFP expression was monitored. Line graphs (figure 4.83) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 96 acquisitions (indicated in the X-axis) at an interval of 3 minutes. Bar graphs (figure 4.84) represent the average end-point value of fluorescence intensity (A.U.) ± SD (at ~5 hours from start of reaction).



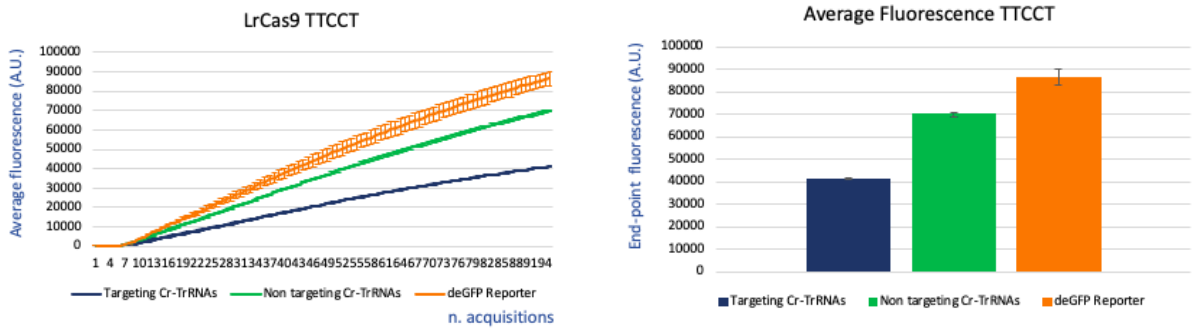


Figures 4.85 (on the left) and 4.86 (on the right). LrCas9 Cr-TrRNAs tested with the PAM TTTTT

The activity of the LrCas9 was detected analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. In each testing, the Cas9 plasmid was incubated with the Cr-TrRNAs (targeting or non-targeting), the deGFP reporter and the TXTL cell extract at 29C and the deGFP expression was monitored. Line graphs (figure 4.85) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 96 acquisitions (indicated in the X-axis) at an interval of 3 minutes. Bar graphs (figure 4.86) represent the average end-point value of fluorescence intensity (A.U.) ± SD (at ~5 hours from start of reaction).

Analysing the LrCas9 Cr-TrRNAs tested with the PAM TTTCT (figures 4.83 and 4.84) and TTTTT (figures 4.85 and 4.86), I could see no Cas9 activity, both targeting and non-targeting curves gave similar fluorescence values.

However, considering the Cr-TrRNA design did not work even with the SpCas9 that should have been the proof of principle for the validation of this gRNA design, maybe I should better investigate if the sequences of these Cr-TrRNAs should be modified because their secondary structure did not allow the DNA cleavage. Another reason could be that the *in-silico* predictions for LrCas9 PAMs (figures 4.37-A-B) are not confirmed *in-vitro*.



Figures 4.87 (on the left) and 4.88 (on the right). LrCas9 Cr-TrRNAs tested with the PAM TTCCT

The activity of the LrCas9 was detected analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. In each testing, the Cas9 plasmid was incubated with the Cr-TrRNAs (targeting or non-targeting), the deGFP reporter and the TXTL cell extract at 29C and the deGFP expression was monitored. Line graphs (figure 4.87) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 96 acquisitions (indicated in the X-axis) at an interval of 3 minutes. Bar graphs (figure 4.88) represent the average end-point value of fluorescence intensity (A.U.) ± SD (at ~5 hours from start of reaction).

The second PAM sequence I derived from the *in-silico* analysis was TTCCT (figures 4.37-A-B). A reduction of the deGFP expression could be detected with the Cr-TrRNAs in targeting conditions (figures 4.87 and 4.88); targeting and non-targeting gRNAs showed different trends, with lower fluorescence values for the targeting Cr-TrRNAs. However, when the same PAM sequence was tested with all the sgRNAs (figures 4.61 and 4.62), the LrCas9 did not cleave the DNA. Therefore, I could perform again the testing with the Cr-TrRNAs and see if these results were a false positive or not. I should monitor when the plateau related to the targeting curve (blue line) is reached, because that was not observed in figures 4.89 and 4.90 but it appears if the nuclease cleaves the DNA (probably in this case it could not be observed because the CRISPR reaction should have continued for more cycles).

4.4 Performing TXTL *in-vitro* testing at non-standard temperatures

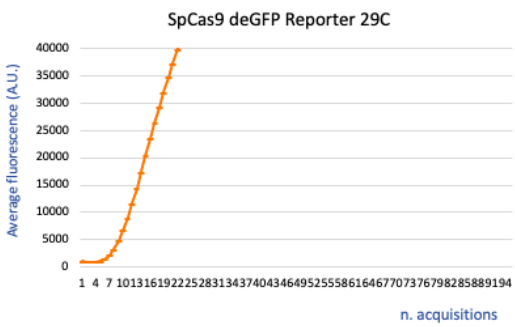
29C is the suggested temperature to get a good yield of the gene expression with MyTXTL®, so I set this value as default for CRISPR reactions performed in standard conditions (experiments from section 4.1 to section 4.3.2).

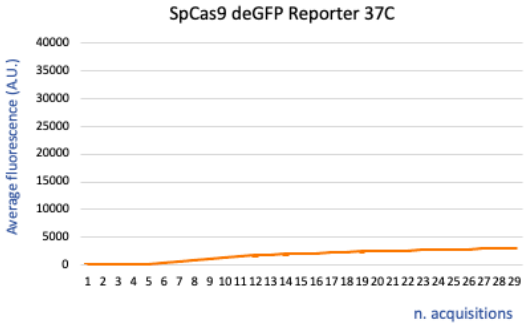
However, some nucleases or other genetic editors could be temperature-sensitive, meaning that their functional activity requires specific and diverse temperature ranges. For example, this is the case of LrCas9 and BfCas9 orthologues, that have an optimal temperature value at 37C, if I consider as their final applications the bacteria whose they derive, *Lactobacillus rhamnosus* and *Bacteroides fragilis*.

The standard value of 29C could be good if the further *in-vivo* application would be performed in insects, animals or other organisms that have 29C as their usual body temperature; the incubation value should be chosen based on that one which is required for the ultimate *in-vivo* step, when the constructs and whole CRISPR system would be tested in living organisms. Therefore, in order to expand the range of temperature at whom the *in-vitro* reactions could be conducted and validate the TXTL testing at a different setting, I performed the CRISPR reactions also at non-standard temperatures.

I decided to set the *in-vitro* reactions also at 37C because most organisms, such as humans, bacteria, animals, have their bodies set at 37C; therefore, this value represents a wider field of application for genetic editors, that increases the number of CRISPR components that could be tested at their suggested temperature setting.

Moreover, another justification to the choice of 37C is due to the orthologues that I are testing in this project: LrCas9 and BfCas9: the bacteria whose these proteins derive have their temperature setting at 37C. Considering I did not obtain great results in term of DNA cleavage of LrCas9 and BfCas9 at 29C, I believed that setting at 37C the CRISPR reactions would have improved their nuclease activity too.





Figures 4.89 (left) and 4.90 (right). SpCas9 deGFP reporters tested at 29C and 37C

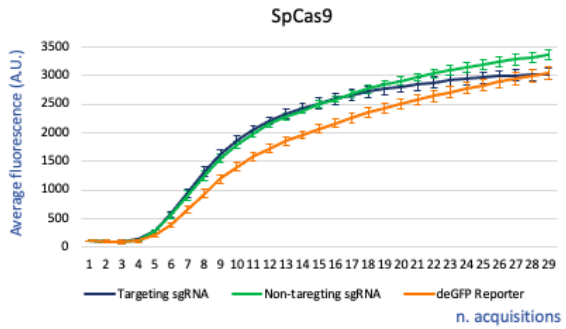
The deGFP expression related to the SpCas9 plasmids was detected analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. The reporter plasmid was incubated with the TXTL cell extract at 37C and the deGFP expression was monitored. Line graphs represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 96 acquisitions for the testing at 29C (figure 4.89) and 30 acquisitions at 37C (figure 4.90) (indicated in the X-axis) at an interval of 3 minutes.

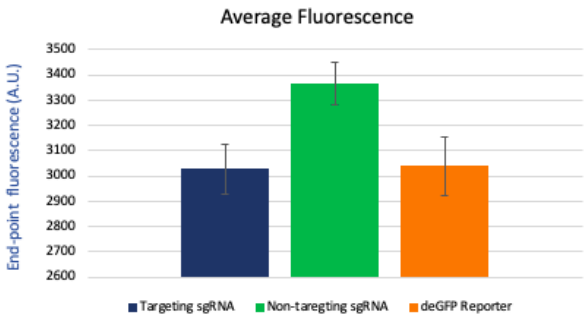
The first component I tested in MyTXTL® was the deGFP reporter. In fact, a good deGFP expression *in-vitro* is fundamental to monitor the whole CRISPR reaction and, consequently, the Cas9 activity. We tested the same SpCas9 reporter, the S1, at both 29C and 37C and I compared their trends. Analysing figures 4.89 and 4.90, at 29C the fluorescence value was very high, meaning that the deGFP expression was very good. However, I got completely different results for the same reporter tested at 37C instead: after few cycles, the reaction already reached the plateau, which was very low compared to the values at 29C. The plateau is an index of when the system reaches the saturation because of the consumption of reagents required for the protein synthesis.

In general, what I observe in figure 4.90 is not ideal to perform a TXTL reaction, considering the nuclease activity is monitored analysing the fluorescence production and a strong deGFP reporter characterised by high fluorescence rate is fundamental to have a good monitoring of the CRISPR reaction.

Despite of low fluorescence production at 37C, I proceeded with the testing of SpCas9 and LrCas9 with related gRNAs and Cr-TrRNAs at this non-standard setting.

4.4.1 SpCas9 tested at 37C





Figures 4.91 (left) and 4.92 (right). SpCas9 sgRNA at 37C

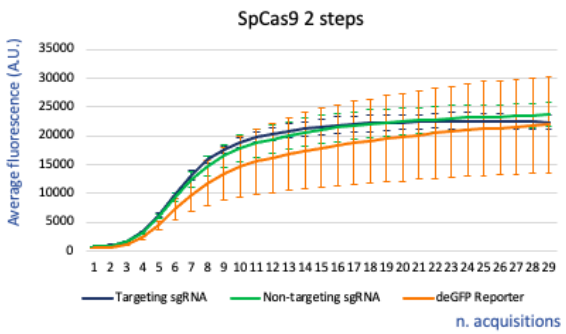
The activity of the SpCas9 was detected analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. In each testing, the SpCas9 plasmid was incubated with the sgRNA (targeting or non-targeting), the deGFP reporter and the TXTL cell extract at 37C and the deGFP expression was monitored. Line graphs (figure 4.91) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 30 acquisitions (indicated in the X-axis) at an interval of 3 minutes. Bar graphs (figure 4.92) represent the average end-point value of fluorescence intensity (A.U.) ± SD (at ~1.5 hours from start of reaction).

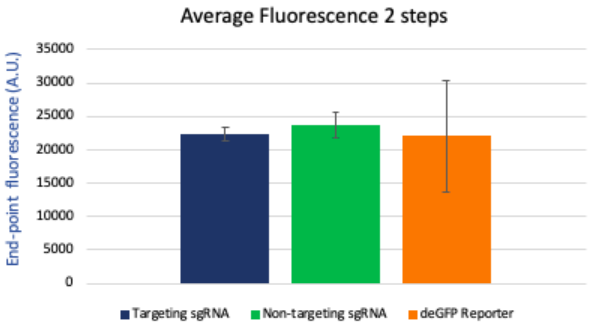
Increasing the temperature from 29C to 37C induced a significant decrement of the master mix activity. In fact, comparing figure 4.7 (29C) with figures 4.91 and 4.92 (37C), the same Cas9/sgRNA/reporter combination worked well at 29C: in those conditions, the DNA cleavage was clearly detected. On the contrary, when the same TXTL reaction was performed at 37C (figure 4.91 and 4.92) no Cas9 activity was shown. Moreover, even if the reagent plasmids were the same (SpCas9, sgRNA and deGFP reporter plasmids), the general *in-vitro* expression was very low in both targeting and non-targeting conditions. This means that the whole TXTL machinery at 37C did not work as good as at 29C.

I thought that one of the reasons why the *in-vitro* expression was very low at higher temperature could have been the rapid consumption of the master mix components which did not allow the synthesis of the Cas9 and the gRNA in time to detect the activity. At 37C the transcription and the translation could work in a faster way than at 29C, meaning that the gene expression of all the components is rapidly performed. Then, considering the reporter expression is generally very strong because of the promoter P70a, it could determine the consumption of fundamental enzymes, cofactors, nucleotides, transcription factors, in favour of its own expression despite of the one of the other reagents (Cas9 and gRNA). Consequently, the system reaches the saturation (which

corresponds to the plateau, the highest value of the fluorescence curves) and the *in-vitro* synthesis of all the reagents stops.

I tried to overcome this problem introducing two steps reactions. In this way, I previously performed a pre-incubation of 30 min at 37C with the Cas9 and the gRNA together with MyTXTL®, without the reporter. During this step, the *in-vitro* synthesis of the two reagents already started using the master mix components. Then, after 30 min, the reporter was added, and all the components were incubated together, again at 37C. At this point, even if the reporter expression was stronger than the one of the Cas9 and the gRNA, it did not prevent their expression, because the other two reagents had already been partly synthesised during the pre-incubation in MyTXTL®.



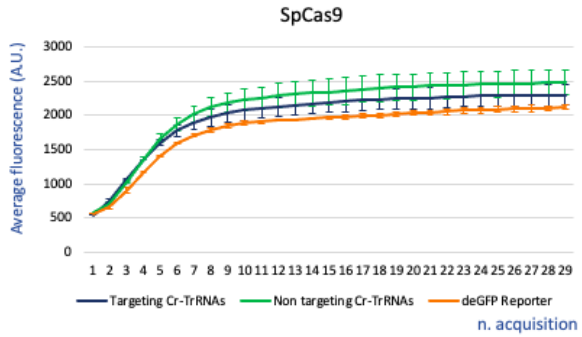


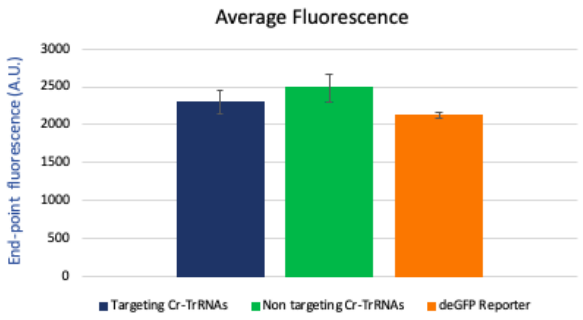
Figures 4.93 (left) and 4.94 (right). SpCas9 sgRNA at 29C, 2 steps reaction

The activity of the SpCas9 was detected analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. In the 2 steps reaction, the SpCas9 was previously pre-incubated at 37C for 30 min with the gRNA (targeting or non-targeting) and the TXTL cell extract only, without the deGFP reporter. After 30 min, the deGFP reporter was added to the pre-incubation reaction and the final reaction was incubated at 37C and the deGFP expression was monitored. Line graphs (figure 4.93) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 30 acquisitions (indicated in the X-axis) at an interval of 3 minutes. Bar graphs (figure 4.94) represent the average end-point value of fluorescence intensity (A.U.) ± SD (at ~1.5 hours from start of reaction).

However, even with the pre-incubation step, I did not improve the results: targeting and non-targeting sgRNAs still gave the same values and the SpCas9 activity was not detected. The only improvement was observed in terms of expression of all the reagents: the saturation was reached at higher values than without the pre-incubation (figures 4.91 and 4.92 vs figures 4.93 and 4.94).

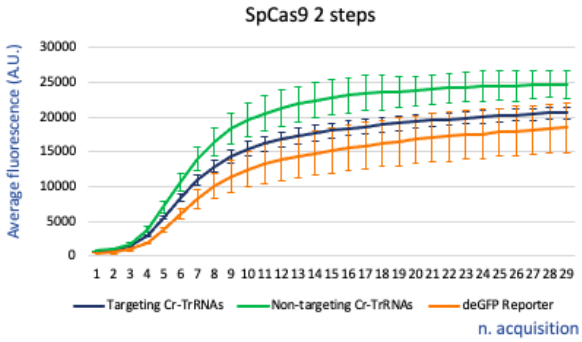
In addition to the SpCas9 sgRNA, I tested at 37C also the same Cr-TrRNAs I already tested at 29C: G1 and G2 (see figures 4.74 and 4.79 for plasmid structures).

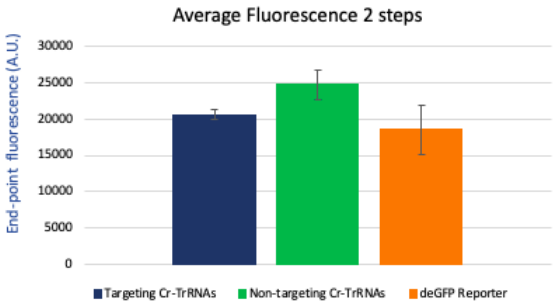




Figures 4.95 (left) and 4.96 (right). SpCas9 Cr-TrRNA G1 at 37C

The activity of the SpCas9 was detected analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. In each testing, the SpCas9 plasmid was incubated with the Cr-TrRNAs (targeting or non-targeting), the deGFP reporter and the TXTL cell extract at 37C and the deGFP expression was monitored. Line graphs (figure 4.95) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 30 acquisitions (indicated in the X-axis) at an interval of 3 minutes. Bar graphs (figure 4.96) represent the average end-point value of fluorescence intensity (A.U.) ± SD (at ~1.5 hours from start of reaction).





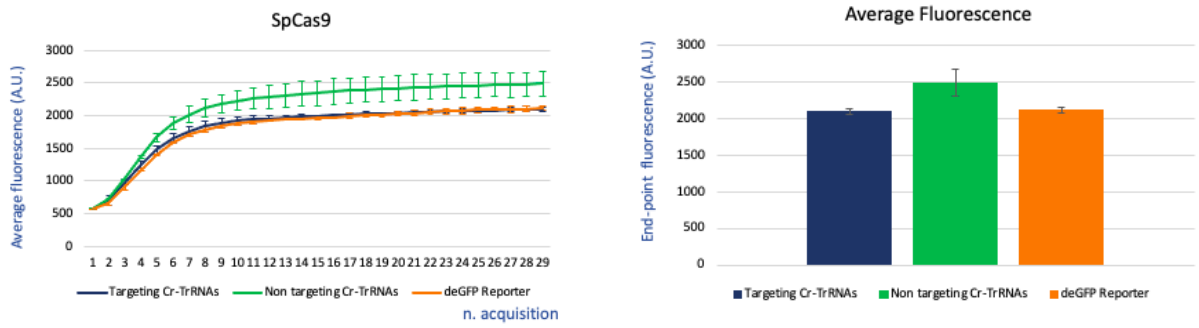
Figures 4.97 (left) and 4.98 (right). SpCas9 Cr-TrRNA G1 at 37C, 2 steps reaction

The activity of the SpCas9 was detected analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. In the 2 steps reaction, the SpCas9 was previously pre-incubated at 37C for 30 min with the Cr-TrRNAs (targeting or non-targeting) and the TXTL cell extract only, without the deGFP reporter. After 30 min, the deGFP reporter was added to the pre-incubation reaction and the final reaction was incubated at 37C and the deGFP expression was monitored. Line graphs (figure 4.97) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 30 acquisitions (indicated in the X-axis) at an interval of 3 minutes. Bar graphs (figure 4.98) represent the average end-point value of fluorescence intensity (A.U.) ± SD (at ~1.5 hours from start of reaction).

Regarding the Cr-TrRNA G1 (figures 4.95 and 4.96), I could not detect any DNA cleavage and, actually, this is the same result I got at 29C with the same gRNA (figures from 4.75 to 4.78). Therefore, the reason why no Cas9 activity was detected could be that the G1 gRNA secondary structure cannot allow the cleavage because of steric effects, therefore the gRNA sequence should be improved.

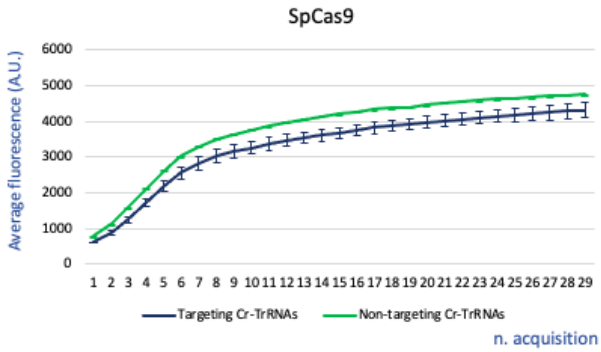
However, as for the testing with the sgRNA, I got a significant improvement regarding the *in-vitro* expression at 37C with the pre-incubation step (figures 4.97 and 4.98): fluorescence values were higher (~ 10 times higher) in both targeting and non-targeting conditions.

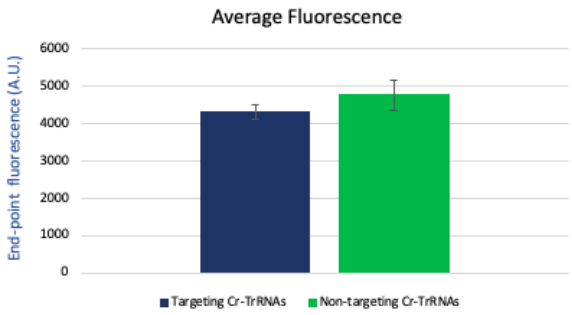
I tried to improve the design of the Cr-TrRNAs modifying the terminators located downstream both the gRNA sequences. I substituted the two different ones which were located in the plasmid G1 (rrnC and K864600) with the aspA for both crispr and tracr RNAs. Basically, I used the same terminator which constituted the SpCas9 sgRNA (see figure 4.79 for the plasmid structure).



Figures 4.99 (left) and 4.100 (right). SpCas9 Cr-TrRNA G2 at 37C

The activity of the SpCas9 was detected analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. In each testing, the SpCas9 plasmid was incubated with the Cr-TrRNAs (targeting or non-targeting), the deGFP reporter and the TXTL cell extract at 37C and the deGFP expression was monitored. Line graphs (figure 4.99) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 30 acquisitions (indicated in the X-axis) at an interval of 3 minutes. Bar graphs (figure 4.100) represent the average end-point value of fluorescence intensity (A.U.) ± SD (at ~1.5 hours from start of reaction).





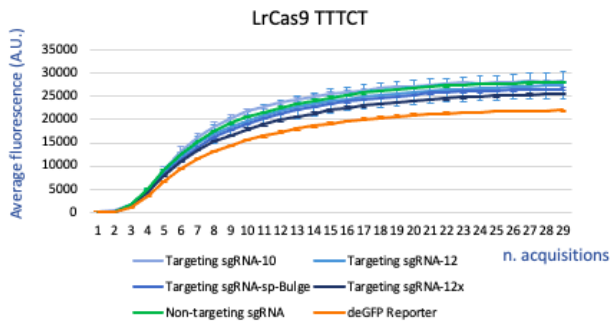
Figures 4.101 (left) and 4.102 (right). SpCas9 Cr-TrRNA G2 at 37C, 2 steps reaction

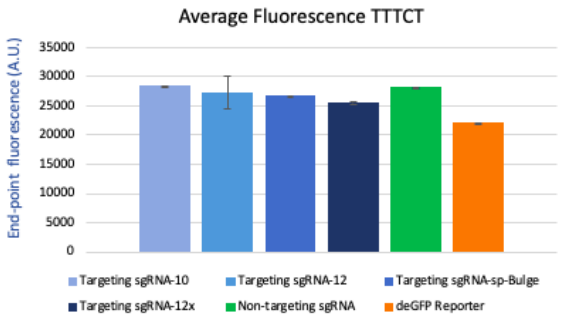
The activity of the SpCas9 was detected analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. In the 2 steps reaction, the SpCas9 was previously pre-incubated at 37C for 30 min with the Cr-TrRNAs (targeting or non-targeting) and the TXTL cell extract only, without the deGFP reporter. After 30 min, the deGFP reporter was added to the pre-incubation reaction and the final reaction was incubated at 37C and the deGFP expression was monitored. Line graphs (figures 4.99 and 4.101) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 30 acquisitions (indicated in the X-axis) at an interval of 3 minutes. Bar graphs (figure 4.100 and 4.102) represent the average end-point value of fluorescence intensity (A.U.) ± SD (at ~1.5 hours from start of reaction).

Regarding the DNA cleavage, even the plasmid G2 did not gave positive results at 37C (figures 4.99 and 4.100).

The deGFP expression was very low in both targeting and non-targeting conditions, as for the SpCas9 and the Cr-TrRNA G1 previously analysed (figures 4.95 and 4.96). Furthermore, in the 2 steps reaction (figures 4.101 and 4.103) I did not see the same significant increment in term of *in-vitro* expression, as it happened for the sgRNA and the Cr-TrRNA G1 (figures 4.97 and 4.98). In fact, when the pre-incubation was performed, targeting and non-targeting curves showed a trend just a bit higher than the reaction performed with one step only. Therefore, also the G2 design could not detect the SpCas9 activity.

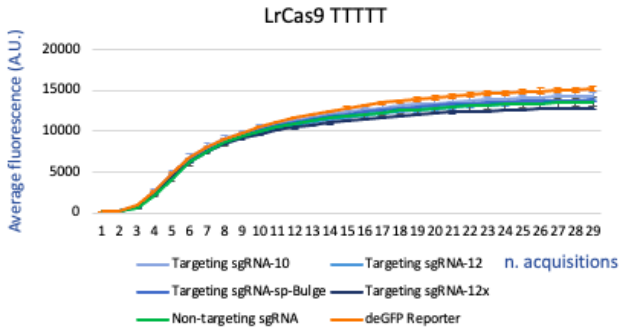
4.4.2 LrCas9 tested at 37C

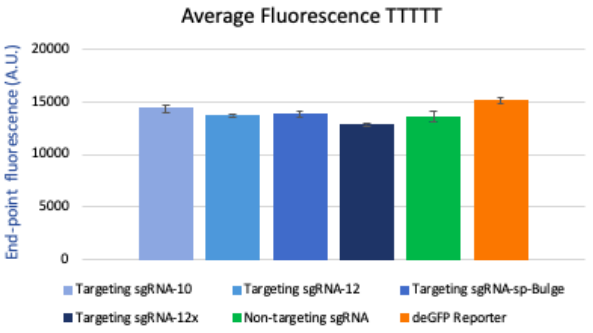




Figures 4.103 (on the left) and 4.104 (on the right). LrCas9 sgRNAs tested with the PAM TTTCT at 37C

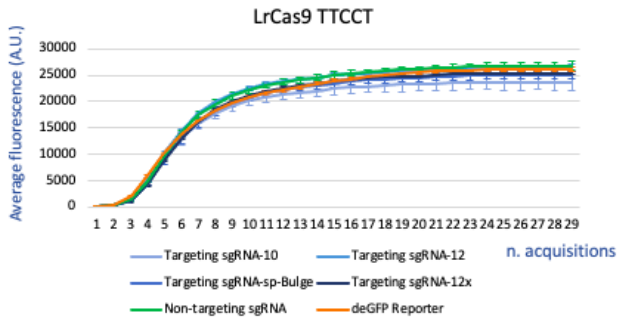
The activity of the LrCas9 was detected analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. In each testing, the Cas9 plasmid was incubated with the sgRNA (targeting or non-targeting), the deGFP reporter and the TXTL cell extract at 37C and the deGFP expression was monitored. Line graphs (figure 4.103) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 30 acquisitions (indicated in the X-axis) at an interval of 3 minutes. Bar graphs (figure 4.104) represent the average end-point value of fluorescence intensity (A.U.) ± SD (at ~1.5 hours from start of reaction).

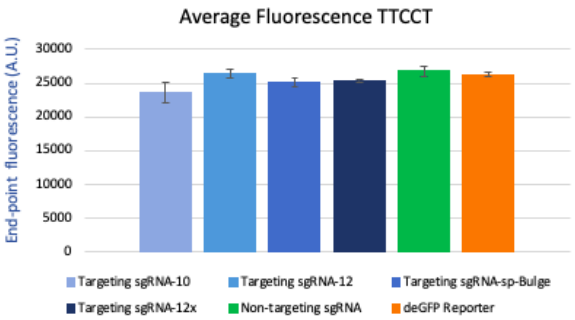




Figures 4.105 (on the left) and 4.106 (on the right). LrCas9 sgRNAs tested with the PAM TTTTT at 37C

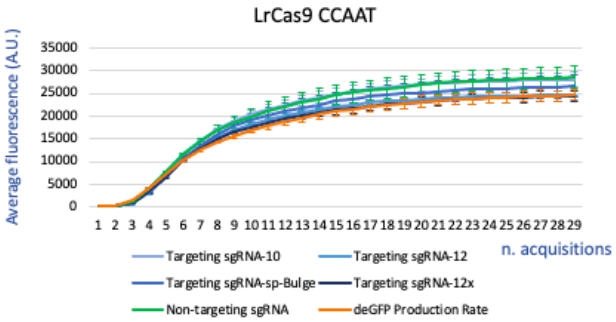
The activity of the LrCas9 was detected analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. Line graphs (figure 4.105) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 30 acquisitions (indicated in the X-axis) at an interval of 3 minutes. Bar graphs (figure 4.106) represent the average end-point value of fluorescence intensity (A.U.) ± SD (at ~1.5 hours from start of reaction).

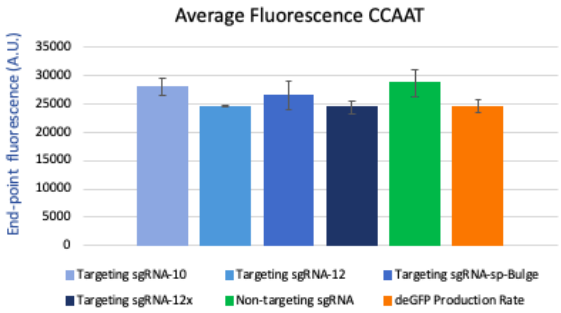




Figures 4.107 (on the left) and 4.108 (on the right). LrCas9 sgRNAs tested with the PAM TTCCT at 37C

The activity of the LrCas9 was detected analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. Line graphs (figure 4.107) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 30 acquisitions (indicated in the X-axis) at an interval of 3 minutes. Bar graphs (figure 4.108) represent the average end-point value of fluorescence intensity (A.U.) ± SD (at ~1.5 hours from start of reaction).



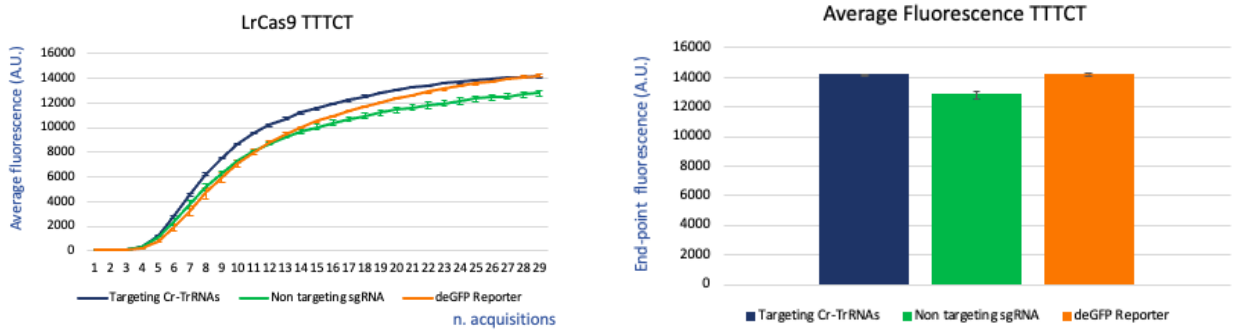


Figures 4.109 (on the left) and 4.110 (on the right). LrCas9 sgRNAs tested with the PAM CCAAT at 37C

The activity of the LrCas9 was detected analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. Line graphs (figure 4.109) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 30 acquisitions (indicated in the X-axis) at an interval of 3 minutes. Bar graphs (figure 4.110) represent the average end-point value of fluorescence intensity (A.U.) ± SD (at ~1.5 hours from start of reaction).

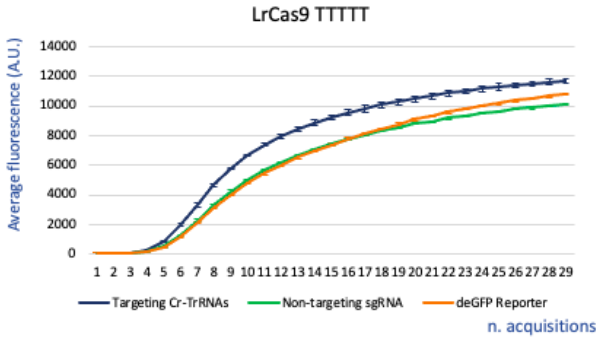
I decided to perform the LrCas9 testing at 37C in one step only; in fact, the only improvement I obtained for the SpCas9 with the pre-incubation step was a major *in-vitro* expression of all the components, but nothing changed in terms of Cas9 activity.

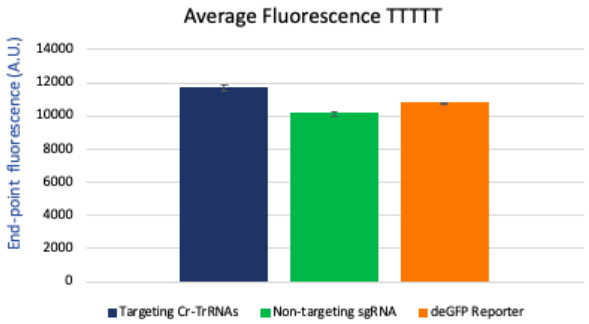
Analysing data related to the testing of the LrCas9 at 37C (figures from 4.103 to 4.110), none sgRNA allowed the DNA cutting, not even the sgRNA-sp-Bulge, sgRNA-10 and sgRNA-12 which allowed the cleavage activity when LrCas9 was tested with the PAM TTTCT at 29C (figures 4.44 and 4.45).
Comparing the LrCas9/sgRNA testing at 37C with the one obtained at 37C with the SpCas9, I got higher fluorescence values in both targeting and non-targeting conditions with all the tested PAMs. This could be explained by the greater strength of the LrCas9 deGFP reporters compared to the SpCas9 ones, leading to a better *in-vitro* expression.
Then, I tested at 37C also the LrCas9 Cr-TrRNAs (see figure 4.82 for the plasmid structure).



Figures 4.111 (on the left) and 4.112 (on the right). LrCas9 Cr-TrRNAs tested with the PAM TTTCT at 37C

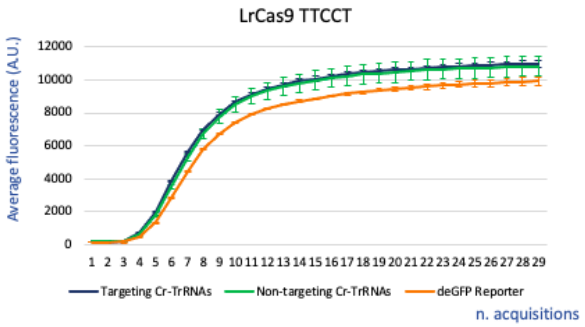
The activity of the LrCas9 was detected analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. Line graphs (figure 4.111) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 30 acquisitions (indicated in the X-axis) at an interval of 3 minutes. Bar graphs (figure 4.112) represent the average end-point value of fluorescence intensity (A.U.) ± SD (at ~1.5 hours from start of reaction).

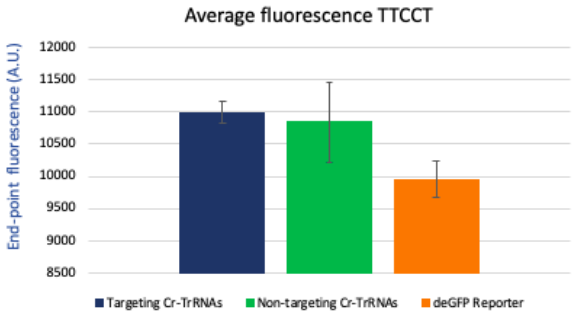




Figures 4.113 (on the left) and 4.114 (on the right). LrCas9 Cr-TrRNAs tested with the PAM TTTTT at 37C

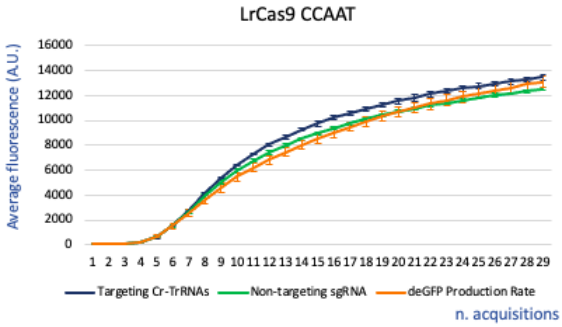
The activity of the LrCas9 was detected analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. Line graphs (figure 4.113) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 30 acquisitions (indicated in the X-axis) at an interval of 3 minutes. Bar graphs (figure 4.114) represent the average end-point value of fluorescence intensity (A.U.) ± SD (at ~1.5 hours from start of reaction).

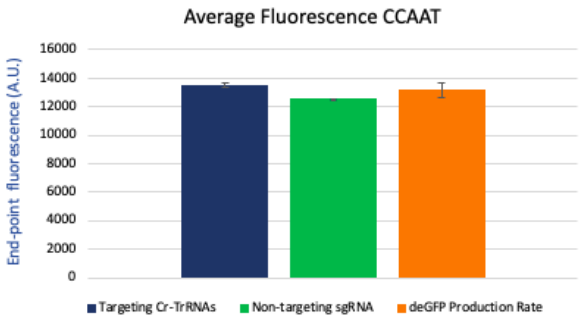




Figures 4.115 (on the left) and 4.116 (on the right). LrCas9 Cr-TrRNAs tested with the PAM TTCCT at 37C

The activity of the LrCas9 was detected analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. Line graphs (figure 4.115) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 30 acquisitions (indicated in the X-axis) at an interval of 3 minutes. Bar graphs (figure 4.116) represent the average end-point value of fluorescence intensity (A.U.) ± SD (at ~1.5 hours from start of reaction).





Figures 4.117 (on the left) and 4.118 (on the right). LrCas9 Cr-TrRNAs tested with the PAM CCAAT at 37C

The activity of the LrCas9 was detected analysing the fluorescence generated by the CRISPR reaction performed in MyTXTL® cell extract. Line graphs (figure 4.117) represent the average intensity ± SD of deGFP fluorescence, as arbitrary unit (A.U.), over time for a total of 30 acquisitions (indicated in the X-axis) at an interval of 3 minutes. Bar graphs (figure 4.118) represent the average end-point value of fluorescence intensity (A.U.) ± SD (at ~1.5 hours from start of reaction).

In the end, CRISPR reactions performed at 37C with both the SpCas9 and the LrCas9 with relative PAMs did not gave positive results in terms of nuclease activity.

Apart from the design of the Cr-TrRNAs and its secondary structure that may interfere with the nuclease activity, the main issue of the CRISPR testing at 37C could be just the increment of the temperature. Considering the recommended value for MyTXTL® cell extract was 29C, it is highly probable that at 37C the RNA polymerase, the coenzymes, the transcription factors do not work properly, interfering with the activity of all the genetic editors added in the reaction. Therefore, if this is confirmed, the CRISPR *in-vitro* testing at 37C could not be performed if as cell extract is used MyTXTL®.

5. Discussion

In this work I evaluated and validated new approaches for the *in-vitro* testing of CRISPR genetic editors.

CRISPR-Cas is a technology widely used to edit the DNA at a precise location and involves an endonuclease (e.g., Cas9, Cas12a) which, guided by a RNA sequence, recognises specific nucleotides (PAM) flanking the target site and induces a DSB in the genome. Furthermore, additional genetic editors have been discovered or developed in the laboratory, such as other nuclease variants (Hu et al., 2018; Kleinstiver et al., 2015), transcription/translation activators and inhibitors. Some of the proposed designs also allow to modulate CRISPR activity, for example to induce or repress DNA cleavage in the presence of specific molecular triggers or inhibitors (Galizi et al., 2020; Galizi and Jaramillo 2019).

In the last few years, this technology has been applied in many fields: theoretically, CRISPR-Cas9 offers efficacy in all cell types, that allows its usage in food, agriculture, bioengineering, nanotechnology, therapeutics.

Gene therapy based on the use of CRISPR-Cas system represents a valid approach for the treatment of genetic and degenerative diseases caused by mutations that alter normal functions in the organism. Moreover, the possibility to target and edit specific sequences in the genome allows precision medicine and personalised therapy applications.

The expansion of genetic manipulation in such fields is supported by the possibility to easily design genetic editors that could be used in many different contests, that makes CRISPR/Cas a very flexible tool.

These characteristics encourage the discovery of new components to implement genetic manipulation technologies and, at the same time, to develop new approaches which allow the rapid testing of this tool.

I designed and made numerous genetic editors, aiming to expand the PAM specificity for some Cas9 variants (WT SpCas9, SpCas9-PRM1 fusion proteins) and orthologues (LrCas9 and BfCas9).

In order to monitor the CRISPR-Cas reactions in MyTXTL® cell extract, I engineered a flexible design for the reporter constructs: the structure of the original cloning spacer vector could easily be switched for inserting specific PAM sequences and target sites at any time.

The customisable reporter vector was fundamental to test in MyTXTL® Cas9 orthologues with different PAM specificity, such as SpCas9-PRM1 variants, LrCas9, BfCas9. The reporter vector I

generated would also allow to test other endonucleases or genetic editors such as catalytically inactivated Cas proteins, ZFNs, TALENs etc., expanding the field of diagnostic and therapeutic research.

After the successful development of suitable fluorescent reporters using the above-mentioned customisable vector, I proceeded with the *in-vitro* testing of various Cas9/gRNA/deGFP reporter combinations.

First, I tested WT SpCas9, SpCas9-PRM1 fusion proteins and LrCas9 at 29C, in standard condition in MyTXTL® cell extract.

Then, I wanted to explore the possibility to use our reporter system to test new synthetic Cas9 variants. One aspect that is of high interest in CRIPR research is directed to the broadening of the PAM recognition constrain. I therefore decided to explore the option to link the Human Sperm Protamine I to the WT SpCas9. I developed different variants and tested the relative cleavage activity using our reporter assay (see results chapter 4, section 4.1.1). Once I generated these SpCas9 variants, I searched for alternative PAMs too, in order to expand the canonical NGG constraint and see if the new Cas9s were able to recognise alternative motifs. I derived the putative PAM variants from a WebLogo calculated using a position frequency matrix (PFM) (Karvelis et al. 2015) that showed conservation on nucleotide positions at flanking sequences extracted from the 3'-ends of the target site. The selected putative PAM motifs were NAG, NTG and NCC that were tested *in-vitro* with the same sgRNA.

The results from TXTL testing showed that the WT SpCas9, as well as the SpCas9-Protamine orthologues 3 and 4 showed cleavage activity when tested against the canonical NGG PAM sequence. However, no activity was shown when the same gRNA target was combined with non-canonical PAMs for all the variants, including the original SpCas9 as expected (Karvelis et al. 2015). These results indicate that the addition of PRM1-derived sequences as fusion proteins of the WT SpCas9 does not adversely affect PAM recognition. Our results also indicate that the addition of the PRM1 and PRM1-derived sequences to the C-terminus of the nuclease has positive effect on SpCas9 activity, whilst the addition of mutations in the PI domain completely inhibits its activity. Beside demonstrating the utility of our reporter system to test various PAMs in fully standardised settings (i.e., same target) our results will be also useful to guide further engineering of new fusion Cas9 proteins for various applications (e.g., to generate inducible proteins through the addition of stabilising/destabilising peptides or tags for immunodetection).

In addition to engineering new SpCas9 variants, I focused also on the testing of other Cas9 orthologues, expected to have different PAM specificity and therefore targetability. In fact, the requirement of a specific PAM sequence downstream the target site limits the possibility to edit the genome at any desirable location. Consequently, the choice of CRISPR targets may be limited, for example, for the treatment of specific point mutations causing genetic diseases.

Moreover, new Cas9 orthologues could be characterised by different levels of activity when tested in different organisms and/or temperature settings, therefore requiring testing at various temperature conditions. In this study, I focused on the testing of *Lactobacillus rhamnosus* and *Bacteroides fragilis* Cas9s. I derived from an *in-silico* pipeline (Kranjc et al., unpublished) the putative PAMs and the resulting sequences were GGGAACG, GAGTACA PAMs for BfCas9 and AAAAA, AGGAA for LrCas9. I also included the TTTCT PAM for LrCas9 according to the recent characterisation in the LrCas9 bacteria (Karvelis et al., 2015) and CCAAT as negative control, because this sequence differed from the derived sequences at each position.

The cloning of the LrCas9 and the BfCas9 in standard vector (carrying the same regulatory sequences used to express the SpCas9 protein) was unsuccessful after several attempts, indicating possible toxicity following the transformation of these constructs in *E. coli*. This was probably due to the toxicity of the templates or of that specific design of the constructs.

Introducing the BASIC assembly (Storch et al., 2015; Kleinstiver et al., 2015), I could make the LrCas9 plasmid in a new conditionally expressed format, that revealed to be more tolerated in bacteria.

The improvement in cloning efficiency by using the BASIC assembly suggests that this method could be used for making plasmids in large scale, overcoming common toxicity issues occurring with standard assembly processes. Moreover, this high throughput cloning technique could be used to generate plasmids carrying the same template but different promoters. This could be interesting for studies focused on the effect of promoter strength on plasmid toxicity, aiming to find the best assembly format for each template, that would be ideal to know when performing large-scale cloning.

Moreover, the use of a high throughput assembly could significantly increase the number of endonucleases, gRNAs and other genetic editors which could be tested *in-vitro* and, consequently, the number of CRISPR reactions which could be performed.

In parallel with the use of BASIC assembly, I validated the *in-vitro* CRISPR testing also with linear DNA templates. This could be another successful strategy to overcome toxicity issues related to plasmid constructs: this alternative approach does not require the cloning process and the template

of interest could directly be added in the cell extract in amplicon format. Therefore, the use of linear DNA significantly reduces the time required to set up CRISPR reactions in MyTXTL® and allows to potentially test any possible template.

In this study I applied the use of amplicons for the testing of BfCas9 orthologue; the template only carried the Cas9 sequence and bacterial promoter and terminator responsible for the RNA polymerase activation/transcription termination.

By combining the BASIC (applied to LrCas9) and the use of amplicons (applied to BfCas9) I successfully managed to test in MyTXTL®a variety of PAM sequence that were previously derived *in-silico* for these orthologues.

Regarding the BfCas9, no cleavage activity was detected when the sgRNAs were tested with all the PAM variants. Therefore, I should better investigate if the main problem was the gRNA design that has to be improved because of structural problems occurring during the CRISPR reaction (such as steric effects) or if the PAM variants I derived *in-silico* cannot be confirmed as positive sequences.

As for LrCas9, the PAM TTTCT showed activity against sgRNAs derived from the crispr:tracrRNA duplex for the LrCas9 with mutations in the bulge and in the upper strem of the structure (Crawley et al., 2018); moreover, the sgRNA that mainly repressed the deGFP expression was the one carrying the bulge of the SpCas9 sgRNA. Therefore, these results related to the TTTCT PAM confirmed some *in-silico* prediction. However, the reduced activity with the other derived PAM sequences suggests that further optimization may be required in order to broaden PAM specificity.

However, these results confirmed that the combination of a high-throughput cloning method (BASIC assembly) or linear DNA template with a cell-free system (MyTXTL®) represents a valid approach that can be applied in the future for rapidly testing and screening multiple CRISPR components in parallel.

When using the SpCas9 in combination with a sgRNA targeting the reporter carrying the NGG PAM, I obtained a significant repression of the fluorescent signal, indicating high activity of the Cas9/gRNA complex. I then went to explore the possibility to express the targeting RNA as dimer (as naturally found in bacteria). This option would be particularly useful when testing new Cas9 variants for which sgRNA (as well as relative PAM specificity) have not yet been characterised. For example, the LrCas9 that I have tested here.

Whilst the Cas9/sgRNA combination showed high activity for SpCas9 against the NGG PAM and for LrCas9 against the TTTCT PAM, no cleavage activity was shown when testing Cr- TrRNAs with the

same orthologue/PAM combination; the LrCas9 Cr-TrRNA showed activity only against the PAM TTCCT.

I used the same promoter sequence for both single guide and Cr-TrRNA design, therefore also the duplex variants should be active in presence of the RNA polymerase in MyTXTL® cell extract. However, the level of transcription for the Cr-TrRNAs may be suboptimal for other reasons, probably related to the different structure of the construct compared to the single guide version.

Another possible explanations for these different behaviours could be the steric effects caused by the longer crispr:tracr complex sequence compared to the single guide one, with subsequent problems for the Cas9/gRNA interaction. It has already been demonstrated how the SpCas9 could be easily directed to any genomic target by using a chimeric sgRNA of 20 nucleotides (Jinek et al., 2012). This discovery significantly improved the design or further gRNAs because it showed how the whole length of both crispr and tracr RNA components was not required for directing the Cas9. There were crucial nucleotides that had a key role in the process, leading to the Cas9:gRNA:target complex. Therefore, the use of sgRNA should be highly recommended, compared to the Cr-TrRNAs, because the chimeric format already includes only those parts of the sequence that are fundamental for the gRNA function; on the contrary, the Cr-TrRNA variant still has unnecessary nucleotides that could lead to formations responsible for steric effects in the Cas9-gRNA interaction area.

Ideally, *in-vitro* reactions should be performed at temperature setting that is representative of the living organism where the final *in-vivo* application happens. This is important because many genetic editors could be temperature-sensitive, meaning that they should work only at specific temperature values and their activity may significantly be influenced by changes in setting conditions.

Therefore, I aimed to expand the TXTL *in-vitro* testing to other temperatures, different from the standard 29C value, that is recommended for MyTXTL® cell extract.

Most organisms, such as humans, bacteria, animals, have their body set at 37C; consequently, in order to broaden the number of genetic editors that could be tested with this *in-vitro* testing platform, I decided to set the TXTL reactions for SpCas9 and LrCas9 also at this value, using the same Cas9/gRNA/reporter combinations.

However, at 37C I could not confirm the same results I obtained at 29C, not even with those gRNAs that showed activity in standard conditions.

The plausible explanation for these different behaviours could be that all the components included in the cell extract, such as RNA polymerase, coenzymes, transcription factors, amino acids, are

extremely temperature-sensitive and they do not work properly when the TXTL reaction is set at different values than the standard 29C.

Moreover, at 37C, the plateau of the reaction, that represents the saturation of the cell extract components, is reached after few cycles of incubation: probably, at higher temperature, there is a more rapid consumption of the reagents required for the TXTL reaction.

In order to overcome this limitation, further steps could include the *in-vitro* testing at other temperatures around 29C and 37C. In this way, I could evaluate if there are other values that allow a good performance of CRISPR reactions in the cell extract.

Moreover, the use of different promoters could affect the *in-vitro* expression of the associated template and, consequently, the related consumption of TXTL components. I could therefore investigate if the use of alternative promoters could reduce the depletion of cell extract constituents.

Finally, another improvement for this screening platform could be the generation of a homemade cell extract derived from the organism where the final testing should be applied. In this way, the gap between the *in-vitro*/*in-vivo* steps could be reduced because the initial screening will be performed in conditions already optimised for the final designed system.

6. Conclusion – Implications and future directions

The rapid progress of genetic editing through the continuous development of new CRISPR tools urge the development of alternative methods for the rapid testing of the multitude of molecular components involved, such as endonucleases, target sites, gRNAs.

In this project I aimed to set the basis to develop a new approach for the customisable/standardisable, flexible, rapid and scalable screening of genetic editors.

This platform includes different techniques responsible for the synthesis and the consequent testing of a great variety of CRISPR components *in-vitro*, leading to a selection of those that show the expected activity and, consequently, could undergo further *in-vivo* steps. Considering CRISPR-Cas applications are rapidly increasing is many fields, such as gene therapy, agriculture, vector-borne diseases, it could be important to have a high-throughput system that allows to analyse in large scale and in a short time many genetic editors, leading to the selection of only most promising ones. This rapid pre-screening of active CRISPR components *in-vitro* may reduce time and costs of the whole experiment, compared to standard processes that are based on the validation of new genetic editors directly *in-vivo*.

In this study the screening of genetic editors was based on the use of a commercially available cell-free system (MyTXTL®) where CRISPR components were directly synthesised and tested, without requiring any living organisms. This approach was really useful because, avoiding all the time-consuming processes associated with the use of cells or animals, allowed to test and screen a great variety of genetic editors in a short time.

We generated different combinations of Cas9/gRNA complex, associated with new PAM variants, trying to broaden this constraint and, consequently, allow to target more sequences in the genome with the CRISPR system. In order to detect the nuclease activity, we successfully engineered a customisable vector plasmid that allowed to easily generate different fluorescent reporter constructs carrying the required PAM and target site. The flexible design of this vector was essential for the rapid evaluation of a variety of PAM sequences derived for the WT SpCas9, PRM1-SpCas9 variants and LrCas9 and BfCas9 orthologues.

We combined a scalable cloning method, the BASIC assembly, with amplicon templates to generate the reagents required for the CRISPR reaction, such as the Cas9, the gRNA and the reporter, taking advantages from the use of both formats. The BASIC allowed us to overcome common problems associated with standard cloning techniques. The possibility to assemble the final plasmid by using

different combinations of promoters, terminators and RBS was useful to generate final constructs characterised by stronger or weaker expression in the cell extract, with consequent effect on the *in-vitro* transcription and translation of the related template that these plasmids carried. Moreover, the BASIC format revealed to be more tolerated in bacteria during the transformation step: this technique allowed us to successfully generate the LrCas9 orthologue after several failing attempts of cloning the same orthologue with standard methods. This progress suggests the use of the BASIC assembly as default to generate plasmids in large scale and with high efficiency. A further improvement regarding the generation of CRISPR components was obtained with the introduction of linear templates in the TXTL reactions. With this approach, the cloning process was not required, with all the consequent advantages; in fact, a simple DNA amplicon related to any possible template could easily be generated, leading to a further increment in the number of genetic editors that may potentially be tested in MyTXTL®. Therefore, the combined use of both BASIC assembly and linear DNA could be a successful strategy for generating CRISPR components in large scale, that would be beneficial for the whole screening platform.

Then, we tried to optimise the design of the gRNAs required for the *in-vitro* testing. This component is essential to induce the DNA cleavage; consequently, the possibility to test both single guide and Cr-TrRNA formats would allow to expand the range of CRISPR reactions that could be performed. In fact, there are Cas9 orthologues that do not have an already characterised sgRNA and, in these cases, natural crispr:tracr duplex could be tested instead. However, improvements are still required, considering no DNA cleavage was detected in the TXTL reactions performed with Cr-TrRNA variants we engineered. In particular, more detailed studies are needed to investigate if the opposite behaviour between the single guide and the duplex format was due to the longer sequence of the Cr-TrRNAs and, consequently, possible effects of its secondary structure in correspondence with the target site.

Ideally, the *in-vitro* CRISPR reactions should be performed at specific setting that may be representative of the living organism designed for the final *in-vivo* application. Therefore, we focused on the expansion of temperature range of TXTL incubations, in order to allow the testing of also other genetic editors that are temperature sensitive. If this platform aims to test all possible CRISPR templates, it could be interesting to investigate on the range of values that allows to perform the *in-vitro* testing with MyTXTL®. For example, we already assessed that this cell extract has not a good performance at 37C, that is the set temperature for most living organisms, such as humans, animals, bacteria. However, it could be interesting to validate the testing at the whole range of

temperatures between 29C, that is the standard suggested value for MyTXTL®, and 37C. This analysis could be useful to discover which is the temperature when the cell extract activity starts to decrease, trying to find a compromise between a customisable setting and efficiency of the reaction. One of the reasons that could justify why the cell extract activity is influenced by the temperature may be that the RNA polymerase, the coenzymes and the energy buffer are very sensitive at changes in the reaction setting and, consequently, they do not work properly if the temperature value changes.

Finally, the combination of efficient cloning methods with linear templates and cell-free systems was significantly upgraded when TXTL reactions were set by using an automated liquid handler (Echo 525). This approach, led to improvements in scalability and precision, supports the development of a high-throughput screening platform.

In conclusion, in this study we generated several components required for CRISPR reactions and we overcame the toxicity of some of them, introducing changes in their structures. We successfully validated the association of a high-throughput cloning method (BASIC assembly) with a cell-free system (MyTXTL®) to quickly analyse *in-vitro* the activity of numerous genetic editors, such as new Cas9 variants and orthologues, derived PAM variants, targets and non-characterised gRNA design. Finally, among all these tested components, only those which induce an activity *in-vitro* should be selected and proceed towards the ultimate *in-vivo* applications.

However, some improvement should be introduced in this screening platform: regarding the high-throughput cloning, more BASIC plasmid variants should be assembled for the same template, in order to compare them and prepare full libraries of promoters, backbones, UTRs associated to each CRISPR component. This may be important to analyse the different expression of these components in the cell extract. Finally, to reduce the gap between *in-vitro*/*in-vivo* testing, instead of the bacterial cell extract, should be made another lysate that is specific for the organism where the *in-vivo* step should then be performed. This will allow to have a screening platform already optimised from the *in-vitro* stage for the final designed organism, that would be ideal.

7. Bibliography

Ahmed et al., Improvement and use of CRISPR/Cas9 to engineer a sperm-marking strain for the invasive fruit pest *Drosophila suzukii*, *BMC Biotechnology*, 2019

Anders et al., Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease, *Nature*, 2014

Anders et al., Structural plasticity of PAM recognition by engineered variants of the RNA-guided endonuclease Cas9, *Molecular Cell*, 2016

Bailey et al., Miniaturization and Rapid Processing of TXTL Reactions Using Acoustic Liquid Handling, Beckman Coulter

Balhorn et al., The protamine family of sperm nuclear proteins, *Genome Biology*, 2007

Barakate et al., An Overview of CRISPR-Based Tools and Their Improvements: New Opportunities in Understanding Plant–Pathogen Interactions for Better Crop Protection, *Frontiers in Plant Sciences*, 2016

Barrangou, CRISPR-Cas systems and RNA-guided interference, *WIREs RNA*, 2013

Barrangou et al., Applications of CRISPR technologies in research and beyond, *Nature Biotechnology*, 2016

Barrangou et al., CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes, *Science*, 2007

Barrangou et al., CRISPR-Cas Systems: Prokaryotes Upgrade to Adaptive Immunity, *Molecular Cell*, 2014

Bengtsson et al., Muscle-specific CRISPR/Cas9 dystrophin gene editing ameliorates pathophysiology in a mouse model for Duchenne muscular dystrophy, *Nature Communications*, 2017

Bolukbasi et al., DNA-binding domain fusions enhance the targeting range and precision of Cas9, *Nature Methods*, 2015

Brewer et al., Dynamics of Protamine 1 Binding to Single DNA Molecules, *Journal of Biological Chemistry*, 2003

Brewer et al., Protamine-Induced Condensation and Decondensation of the Same DNA Molecule, *Science*, 1999

Burt, Site-specific selfish genes as tools for the control and genetic engineering of natural populations, *The Royal Society Biological Sciences*, 2003

Cermak et al., Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting, *Nucleic Acids Research*, 2011

Chappell et al., Creating small transcription activating RNAs, *Nature Chemical Biology*, 2015

Chatterjee et al., Divergent PAM Specificity of a Highly-Similar SpCas9 Ortholog, *bioRxiv*, 2018

Chen et al., Application of the CRISPR/Cas9 System to Drug Resistance in Breast Cancer, *Advanced Science*, 2018

Christian et al., Targeting DNA Double-Strand Breaks with TAL Effector Nucleases, *Genetics*, 2010

Crawley et al., Characterizing the activity of abundant, diverse and active CRISPR-Cas systems in lactobacilli, *Scientific Reports*, 2018

Crooks et al., WebLogo: A Sequence Logo Generator, *Genome Research*, 2004

Depardieu et al., Gene silencing with CRISPRi in bacteria and optimization of dCas9 expression level, Methods, 2020

Dong et al., CRISPR/dCas9-mediated inhibition of gene expression in *Staphylococcus aureus*, *Journal of Microbiological Methods*, 2017

Dopp et al., Cell-free supplement mixtures: Elucidating the history and biochemical utility of additives used to support *in vitro* protein synthesis in *E. coli* extract, *Biotechnology Advances*, 2019

Douris et al., Using CRISPR/Cas9 genome modification to understand the genetic basis of insecticide resistance: *Drosophila* and beyond, *Pesticide Biochemistry and Physiology*, 2020

Ehrt et al., Controlling gene expression in mycobacteria with anhydrotetracycline and Tet repressor, *Nucleic Acids Research*, 2005

Evans et al., The application of tetracycline regulated gene expression systems in the validation of novel drug targets in *Mycobacterium tuberculosis*, *Frontiers in Microbiology*, 2015

Fasulo et al., A fly model establishes distinct mechanisms for synthetic CRISPR/Cas9 sex distorters, *PLOS Genetics*, 2020

Fineran et al., Memory of viral infections by CRISPR-Cas adaptive immune systems: Acquisition of new information, *Virology*, 2012

Fuchs et al., Mosquito Transgenic Technologies to Reduce *Plasmodium* Transmission, *Malaria*, 2012

Gaj et al., ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering, *Trends in Biotechnology*, 2013

Galizi and Jaramillo, Engineering CRISPR guide RNA riboswitches for *in vivo* applications, *Current Opinion in Biotechnology*, 2019

Galizi et al., A CRISPR-Cas9 sex-ratio distortion system for genetic control, *Scientific Reports*, 2016

Galizi et al., Engineered RNA-Interacting CRISPR Guide RNAs for Genetic Sensing and Diagnostics, *The CRISPR Journal*, 2020

Gao et al., Engineered Cpf1 variants with altered PAM specificities, *Nature Biotechnology*, 2017

Gao, The future of CRISPR technologies in agriculture, *Nature Reviews Molecular Cell Biology*, 2018

Garamella et al., The All *E. coli* TX-TL Toolbox 2.0: A Platform for Cell-Free Synthetic Biology, *ACS Synthetic Biology*, 2016

Gilbert et al., CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes, *Cell*, 2013

Gupta et al., An optimized two-finger archive for ZFN-mediated gene targeting, *Nature Methods*, 2012

Gupta et al., Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9, *The Journal of Clinical Investigation*, 2014

Hammond and Galizi, Gene drives to fight malaria: current state and future directions, *Pathogens and Global Health*, 2017

Hammond et al., A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector *Anopheles gambiae*, *Nature Biotechnology*, 2016

Hammond et al., The creation and selection of mutations resistant to a gene drive over multiple generations in the malaria mosquito, PLOS Genetics, 2017

Haque et al., Application of CRISPR/Cas9 Genome Editing Technology for the Improvement of Crops Cultivated in Tropical Climates: Recent Progress, Prospects, and Challenges, *Frontiers in Plant Sciences*, 2018

Hendriks et al., CRISPR-Cas Tools and Their Application in Genetic Engineering of Human Stem Cells and Organoids, *Cell Stem Cell*, 2020

Heu et al., CRISPR-Cas9-Based Genome Editing in the Silverleaf Whitefly (*Bemisia tabaci*), *The CRISPR Journal*, 2020

Hidalgo-Cantabrana et al., Characterization and Repurposing of Type I and Type II CRISPR–Cas Systems in Bacteria, *Journal of Molecular Biology*, 2019

Hille et al., The Biology of CRISPR-Cas: Backward and Forward, *Cell*, 2018

Hillson et al., Building a global alliance of biofoundries, *Nature Communications*, 2019

Hockemeyer et al., Genetic engineering of human pluripotent cells using TALE nucleases, *Nature Biotechnology*, 2011

Hodges et al., Delivering on the promise of gene editing for cystic fibrosis, *Genes & Diseases*, 2019

Hsu et al., Development and Applications of CRISPR-Cas9 for Genome Engineering, *Cell*, 2014

Hu et al., Evolved Cas9 variants with broad PAM compatibility and high DNA specificity, *Nature*, 2018

Huang et al., Applications of CRISPR-Cas Enzymes in Cancer Therapeutics and Detection, *Trends in Cancer*, 2018

Hur et al., Targeted mutagenesis in mice by electroporation of Cpf1 ribonucleoproteins, *Nature Biotechnology*, 2016

Jang et al., RNA-based dynamic genetic controllers: development strategies and applications, Current Opinion in Biotechnology, 2018

Jayavaradhan et al., CRISPR-Cas9 fusion to dominant-negative 53BP1 enhances HDR and inhibits NHEJ specifically at Cas9 target sites, *Nature Communications*, 2019

Jiang et al., Application of *CRISPR/Cas9* gene editing technique in the study of cancer treatment, *Clinical Genetics*, 2019

Jiang et al., Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage, *Science*, 2016

Jinek et al., A programmable dual RNA-guided DNA endonuclease in adaptive bacterial immunity, *Science*, 2012

Jinek et al., Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation, *Science*, 2014

Jiang et al., CRISPR-Cas9 Structures and Mechanisms, *Annual Review of Biophysics*, 2017

Karvelis et al., Methods for decoding Cas9 protospacer adjacent motif (PAM) sequences: A brief overview, Methods, 2017

Karvelis et al., Rapid characterization of CRISPR-Cas9 protospacer adjacent motif sequence elements, *Genome Biology*, 2015

Kirill et al., Cell-free Protein Synthesis, Methods and Protocols, *Methods in Molecular Biology,* 2014

Kleinstiver et al., Engineered CRISPR-Cas9 nucleases with altered PAM specificities, *Nature*, 2015

Kleinstiver et al., High-fidelity CRISPR-Cas9 variants with undetectable genome-wide off-targets, *Nature*, 2016

Kyrou et al., A CRISPR–Cas9 gene drive targeting *doublesex* causes complete population suppression in caged *Anopheles gambiae* mosquitoes, *Nature Biotechnology*, 2018

Lee et al., Design rules of synthetic non-coding RNAs in bacteria, *Methods*, 2018

Lim et al., Structural roles of guide RNAs in the nuclease activity of Cas9 endonuclease, Nature Communications, 2016

Liu et al., Application of Various Delivery Methods for CRISPR/dCas9, *Molecular Biotechnology*, 2020

Liu et al., Expanded base editing in rice and wheat using a Cas9-adenosine deaminase fusion, *Genome Biology*, 2018

Liu et al., Virus-like nanoparticle as a co-delivery system to enhance efficacy of CRISPR/Cas9-based cancer immunotherapy, *Biomaterials*, 2020

Loureiro et al., CRISPR-Cas: Converting A Bacterial Defence Mechanism into A State-of-the-Art Genetic Manipulation Tool, *Antibiotics*, 2019

Lu, Cell-free synthetic biology: Engineering in an open world, *Synthetic and Systems Biotechnology*, 2017

Lu et al., Cell-free protein synthesis: Recent advances in bacterial extract sources and expanded applications, *Biochemical Engineering Journal*, 2019

Lu et al., Safety and feasibility of CRISPR-edited T cells in patients with refractory non-small-cell lung cancer, *Nature Medicine*, 2020

Makarova et al., An updated evolutionary classification of CRISPR–Cas systems, *Nature Reviews Microbiology*, 2015

Marangi et al., Innovative Therapeutic Strategies for Cystic Fibrosis: Moving Forward to CRISPR Technique, *Frontiers in Pharmacology*, 2018

Marraffini, The CRISPR-Cas system of Streptococcus pyogenes: function and applications, in: Ferretti JJ, Stevens DL, Fischetti VA, editors, 2016

Marraffini et al., CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea, *Nature Reviews Genetics*, 2010

Marshall et al., High-throughput Microliter-Sized Cell-Free Transcription- Translation Reactions for Synthetic Biology Applications Using the Echo® 550 Liquid Handler, Labcyte website (visited on Dec 2020)

Marshall et al., Quantitative modeling of transcription and translation of an all-*E. coli* cell-free system, *Scientific Reports*, 2019

Marshall et al., Rapid and Scalable Characterization of CRISPR Technologies Using an *E. coli* Cell-Free Transcription-Translation System, *Molecular Cell*, 2017

Marshall et al., Short DNA containing χ sites enhances DNA stability and gene expression in *E. coli* cell-free transcription-translation systems, *Biotechnology and Bioengineering*, 2017

McFarlane et al., CRISPR-Based Gene Drives for Pest Control, *Trends in Biotechnology*, 2018

McGinn et al., Molecular mechanisms of CRISPR–Cas spacer acquisition, *Nature Reviews Microbiology*, 2019

Meyer, *Escherichia* coli "Marionette" strains with 12 highly optimized small-molecule sensors, *Nature Chemical Biology*, 2018

Millet et al., A TALE nuclease architecture for efficient genome editing, *Nature Biotechnology*, 2010

Min et al., ISPR-Cas9 corrects Duchenne muscular dystrophy exon 44 deletion mutations in mice and human cells, *Science Advances*, 2019

Mojica et al., Short motif sequences determine the targets of the prokaryotic CRISPR defence system, *Microbiology*, 2009

Mollanoori et al., Promising therapeutic approaches using CRISPR/Cas9 genome editing technology in the treatment of Duchenne muscular dystrophy, *Genes & Diseases*, 2020

Moore et al., Rapid acquisition and model-based analysis of cell-free transcription–translation reactions from non-model bacteria, *PNAS*, 2018

Na et al., Metabolic engineering of *Escherichia coli* using synthetic small regulatory RNAs, *Nature Biotechnology*, 2013

Nelson et al., Genome engineering: a new approach to gene therapy for neuromuscular disorders, *Nature Reviews Neorology*, 2017

Nelson et al., *In vivo* genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy, *Science*, 2016

Nishimasu et al., Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA, *Cell*, 2014

Nishimasu et al., Engineered CRISPR-Cas9 nuclease with expanded targeting space, *Science*, 2018

Noireaux et al., The New Age of Cell-Free Biology, *Annual Review of Biomedical Engineering*, 2020

Pawluk et al., Naturally Occurring Off-Switches for CRISPR-Cas9, *Cell*, 2016

Peters et al., A Comprehensive, CRISPR-based Functional Analysis of Essential Genes in Bacteria, *Cell*, 2016

Pickar-Oliver et al., The next generation of CRISPR–Cas technologies and applications, *Nature Reviews Molecular Cell Biology*, 2019

Qi et al., Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression, *Cell*, 2013

Qiao et al., Co-expression of Cas9 and single-guided RNAs in *Escherichia coli* streamlines production of Cas9 ribonucleoproteins, *Communications Biology*, 2019

Rajagopalan et al., A Two-Step Method for Obtaining Highly Pure Cas9 Nuclease for Genome Editing, Biophysical, and Structural Studies, *Methods and Protocols*, 2018

Reyes et al., Uncovering cell-free protein expression dynamics by a promoter library with diverse strengths, *BioRxiv*, 2017

Rodriguez-Garcia et al., Natural and synthetic tetracycline-inducible promoters for use in the antibiotic-producing bacteria *Streptomyces*, *Nucleic Acids Research*, 2005

Rosenblum et al., CRISPR-Cas9 genome editing using targeted lipid nanoparticles for cancer therapy, *Science Advances*, 2020

Sanz et al., Cas9/gRNA targeted excision of cystic fibrosis-causing deep-intronic splicing mutations restores normal splicing of *CFTR* mRNA, *PLOS One*, 2017

Schwank et al., Functional Repair of CFTR by CRISPR/Cas9 in Intestinal Stem Cell Organoids of Cystic Fibrosis Patients, *Cell Stem Cell*, 2013

Shabbir et al., CRISPR-cas system: biological function in microbes and its use to treat antimicrobial resistant pathogens, *Annals of Clinical Microbiology and Antimicrobials*, 2019

Shao et al., The big bang of genome editing technology: development and application of the CRISPR/Cas9 system in disease animal models, *Zoological Research*, 2016

Siksnys et al., Rewiring Cas9 to Target New PAM Sequences, *Molecular Cell*, 2016

Silverman et al., Cell-free gene expression: an expanded repertoire of applications, *Nature Reviews Genetics*, 2019

Song, The CRISPR/Cas9 system: Their delivery, *in vivo* and *ex vivo* applications and clinical development by startups, *Biotechnology Progress*, 2017

Sorek et al., CRISPR-Mediated Adaptive Immune System in Bacteria and Archaea, *Annual Review of Biochemistry*, 2013

Storch et al., BASIC: A New Biopart Assembly Standard for Idempotent Cloning Provides Accurate, Single-Tier DNA Assembly for Synthetic Biology, *ACS Synthetic Biology*, 2015

Storch et al., BASIC: A Simple and Accurate Modular DNA Assembly Method, *Synthetic DNA*, 2016

Su et al., The application of the CRISPR-Cas9 genome editing machinery in food and agricultural science: Current status, future perspectives, and associated challenges, *Biotechnology Advances*, 2019

Tang et al., A CRISPR–Cpf1 system for efficient genome editing and transcriptional repression in plants, *Nature Plants*, 2017

Urnov et al., Genome editing with engineered zinc finger nucleases, *Nature Reviews*, 2010

Van der Oost et al., Unravelling the structural and mechanistic basis of CRISPR–Cas systems, *Nature Reviews Microbiology*, 2014

Vanegas et al., Cpf1 enables fast and efficient genome editing in Aspergilli, *Fungal Biology and Biotechnology*, 2019

Vangah et al., CRISPR-Based Diagnosis of Infectious and Non-infectious Diseases, *Biological Procedures Online*, 2020

Walton et al., Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants, *Science*, 2020

Wandera et al., An enhanced assay to characterize anti-CRISPR proteins using a cell-free transcription-translation system, *Methods*, 2020

Wang et al., CRISPR/Cas9 in Genome Editing and Beyond, *Annual Review of Biochemistry*, 2016

Wang et al., The rapidly advancing Class 2 CRISPR-Cas technologies: A customizable toolbox for molecular manipulations, *Journal of Cellular and Molecular Medicine*, 2020

Ward et al., A generic white pupae sex selection phenotype for insect pest control, *bioRxiv*, 2020

Xing et al., CRISPR-cas9: a powerful tool towards precision medicine in cancer treatment, *Acta Pharmacologica Sinica*, 2019

Xu et al., Efficient homology-directed gene editing by CRISPR/Cas9 in human stem and primary cells using tube electroporation, *Scientific Reports*, 2018

Yim et al., Protecting Linear DNA Templates in Cell-Free Expression Systems from Diverse Bacteria, *ACS Synthetic Biology*, 2020

Zetsche et al., Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System, *Cell*, 2015

Zetsche et al., Multiplex gene editing by CRISPR–Cpf1 using a single crRNA array, *Nature Biotechnology*, 2016

Zhang et al., CRISPR-Cpf1 correction of muscular dystrophy mutations in human cardiomyocytes and mice, *Science Advances*, 2017

Zhang et al., *In silico* Method in CRISPR/Cas System: An Expedite and Powerful Booster, *Frontiers in Oncology*, 2020

Zhao et al., Construction of a Gene Knockdown System Based on Catalytically Inactive ("Dead") Cas9 (dCas9) in *Staphylococcus aureus*, *American Society for Microbiology Applied and Environmental Microbiology*, 2017

Zhu et al., Applications of CRISPR–Cas in agriculture and plant biotechnology, *Nature Reviews Molecular Cell Biology*, 2020

Zitmann et al., Process Optimization for Recombinant Protein Expression in Insect Cells, *New Insights into Cell Culture Technology*, 2017

8. Acronyms

- A: Alanine
- AsCpf1: *Acidaminococcus sp. BV3L6* Cpf1
- ATc: Anhydrotetracycline
- ATP: Adenosine Triphosphate
- BASIC: Biopart Assembly Standard for Idempotent Cloning
- BfCas9: *Bacteroides fragilis* Cas9
- Bp: Base Pairs
- BSA: Bovine Serum Albumin
- C: Cysteine
- Cas: CRISPR-Associated Protein
- CF: Cystic Fibrosis
- CFU: Colony-Forming Unit
- CFE: Cell-Free gene Expression
- CFPS: Cell-Free Protein Synthesis
- CoA: Coenzyme A
- CP: Creatine Phosphate
- CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats
- CRISPRi: CRISPR Interference
- crRNA: CRISPR RNA
- Cys: Cysteine
- D: Aspartic Acid
- DBTL: Design-Build-Test-Learn
- dCas9: Dead Cas9
- DNA: Deoxyribonucleic Acid
- DMD: Duchenne Muscular Dystrophy
- DSB: Double-Strand Break
- E: Glutamic Acid
- eGFP: Enhanced Green Fluorescent Protein
- G: Guanine

- GFP: Green Fluorescent Protein
- GMO: Genetically Modified Organism
- gRNA: Guide RNA
- HDR: Homology-Directed Repair
- HEK: Human Embryonic Kidney
- HF: High-Fidelity
- Hys: Histidine
- H_2O: Dihydrogen Monoxide
- I: Isoleucine
- igRNAs: RNA-interacting guide RNAs
- K: Lysine
- Kb: Kilobases
- L: Leucine
- LB: Laura-Bertani
- LrCas9: *Lactobacillus rhamnosus* Cas9
- M: Methionine
- $MgCl_2$: Magnesium Chloride
- mL: Millilitre
- mRNA: Messenger RNA
- N: Asparagine
- ncRNAs: Non-Coding RNAs
- NHEJ: Non-Homologous End Joining
- NLS: Nuclear Localisation Sequence
- nm: nanometer
- NUC: Nuclease Lobe
- OD: Optical Density
- ORF: Open Reading Frame
- P: Proline
- PAM: Protospacer Adjacent Motif
- PCR: Polymerase Chain Reaction
- pH: Potential of Hydrogen

- PI: PAM Interacting
- pg: picogram
- Pr: Promoter
- pre-crRNA: Precursor crRNA
- PRM1: Human Sperm Protamine 1
- PURE: Protein synthesis Using Recombinant Elements
- Q: Glutamine
- R: Arginine
- RBS: Ribosome Binding Site
- RBSs: Ribosome Binding Sites
- REC: Recognition Lobe
- RNA: Ribonucleic Acid
- RNPs: Ribonucleoproteins
- RPM: Revolutions Per Minute
- RT: Room Temperature
- RVD: Repeat-Variable Di-residue
- S: Serine
- sgRNA: Single Guide RNA
- S.O.C.: Super Optimal Broth
- SpCas9: *Streptococcus pyogenes* Cas9
- STARs: sRNA transcriptional activators
- sRNAs: Small RNAs
- T: Threonine
- TALEN: Transcription Activator-Like Effector Nuclease
- Tc: Tetracycline
- Tet: Tetracycline
- TetP: Tetracycline Promoter
- TetR: Tetracycline Repressor
- TL: Translation
- Tm: Melting Temperature
- tracrRNA: trans-activating CRISPR RNA

- tRNA: Transfer RNA
- trRNAs: trigger RNAs
- TX: Transcription
- TXTL: Transcription-Translation
- UV: Ultraviolet
- V: Volts
- V: Valine
- WT: Wild-Type
- w/v: weight/volume
- Y: Tyrosine
- ZFN: Zinc Finger Nuclease
- σ^{70}: sigma factor 70
- µg: Microgram
- µl: Microlitre
- µM: Micromolar

Acknowledgements

Grazie al Prof. Crisanti per avermi permesso di svolgere gran parte del lavoro di ricerca presso il suo laboratorio all'Imperial College London. Un'esperienza indelebile, ricca di scienza, confronti, sfide e miglioramenti. Grazie per questa bellissima opportunita' di crescita soprattutto personale.

Grazie alla Dott.ssa Immobile Molaro e alla Prof.ssa Santucci per avermi permesso di svolgere l'intero percorso di dottorato in un modo alternativo e ricco di nuove esperienze.

Grazie a Roberto per tutto. Per il supporto, non solo scientifico, per l'umanita', per i consigli, per la presenza anche a distanza. Grazie per la tanta pazienza, soprattutto durante i miei primi tempi a Londra. Grazie infinite per la fiducia e per aver sempre anteposto il lato umano a quello professionale.

Grazie a Giusy. La mia forza, la mia parte complementare. Resti sempre quella certezza, quell'approdo sicuro sempre pronto a sorreggermi quando penso di non farcela da sola. Siamo una squadra sempre piu' forte, ma non per gli obiettivi raggiunti, per il bello del nostro legame.

Grazie a Luca, Giulia, Silvia, Nayomi e Matteo. Voi sapete che questo e' molto piu' di un semplice grazie. Mi avete rialzato quando credevo di non farcela da sola grazie alla vostra presenza, alla vostra spontaneita', alla vostra energia. Grazie ai vostri sorrisi che sono stati tanto preziosi per far tornare a sorridere anche me. Grazie per tutti i momenti che abbiamo condiviso in questi anni, porto tutto con me.

Grazie a Martina e Antonella. Siete una presenza costante. Nonostante la lontananza, siete sempre cosi' vicine. E questa e' la conferma piu' bella della nostra amicizia.

Grazie a Gloria per tutto quello che abbiamo condiviso insieme, fin dal quel primo giorno a Londra. Grazie per saper ascoltare e per aver continuato ad esserci.

Thanks to the whole Crisanti Lab. Thanks for the endless love and support you gave me during these years, thanks for the suggestions, the inspiring talks, the happy moments we shared together. You have been the best teammates I could ask for this amazing journey, I always felt as part of the family. Love you guys.
A special thanks to Charlotte and Nace for the confidence, the patience and the useful advice from the beginning. Charl, you will always be the TXTL queen and Nace, thanks to have chosen to be my first (and only!) customer.

Thanks to Dr. Marko Storch for the practical support when using the Echo 525 automated liquid handler at London Biofoundry and for the useful suggestions regarding the use of the BASIC assembly in my PhD project.

Grazie Londra. Sei stata molto di piu' di quello che mi aspettassi e molto di piu' di una semplice citta'. Grazie per tutto cio' che mi hai fatto vivere. Resti un cerchio che non si chiude.