Functional characterization of CRISPR-Cas interactions with DNA

MARIALUCREZIA LOSITO

Functional characterization of CRISPR-Cas interactions with DNA

Marialucrezia Losito

Department of Chemistry

Durham University

Submitted in part fulfilment of the requirements for the degree of

Doctor of Philosophy

2022

STATEMENT OF COPYRIGHT

LIGHTDYNAMICS NETWORK

This project is part of the LightDyNamics consortium (https://www.lightdynamics.eu/), a multidisciplinary European Training Network (ETN) with funds from the European Commission, part of the Horizon Marie Sklodowska-Curie Action. The LightDyNamics network is composed by 21 partners from 9 different European countries, including both industries (such as AstraZenca, UK) and academic institutions (such as Dr Jan Verlet at Durham University and Dr David Rueda's team at Imperial College London). The research approach focuses on interdisciplinary training, development of transferable and business skills and international exposure.

LightDyNamics aims to study the interactions of light and soft matter, particularly multichromophore soft materials. DNA photochemistry and photophysics are used as a research tool to train Early Stage Researchers (ESRs) and to develop new technologies applicable not only to DNA, but soft materials in general. Specifically, DNA is chosen as the molecule of interest due to its fundamental biological role and the important applications of light-induced processes from a biomedical and optoelectronics perspective.

Part of LightDyNamics project also focuses on protein-DNA interaction and this work is in line with the aim of combining biophysical and cellular methods to characterize the molecular interactions of gene editing proteins with oligonucleotides. The work here described has been carried out in part at AstraZeneca, UK (Chapter 3 and 4) and in part at Imperial College London (Chapter 5 and 6).

DECLARATION OF ORIGINALITY

I hereby declare that all text, data, results and analysis contained in this thesis are my own except for the data presented in the first part of Chapter 6, Figures 6.1 - 6.2. This is also highlighted in the relevant figure captions. The data show preliminary results on AZ-Cas9 characterization and relative sgRNA optimization made by Marcello Maresca's team at AstraZeneca, Sweden. It is included as it provides the background into the single-molecule analysis I have performed on this nuclease.

I further declare that all sources, text and figures are appropriately cited.

ABSTRACT

The recent Nobel-prize-winning CRISPR technology has revolutionised the gene-editing field with its ability to precisely modify any target DNA sequence and its relevance in a wide range of applications, from functional characterisation of gene variants in biological sciences to mutation of plant genomes in agriculture. New prospectives for CRISPR-therapeutic applications are also rapidly advancing, particularly for the modification of disease-causing genes and *ex-vivo* editing of immune cells for cancer treatment. However, off-target editing exists as collateral damage and it represents a significant hurdle to realise CRISPR's full potential. To date, the scientific community lacks extensive knowledge regarding the impact of the eukaryotic cellular context on Cas nucleases activity, and how this affects Cas efficiency and specificity on human DNA. In this thesis, I use a combination of cellular and single-molecule assays to investigate this specific topic. First, I developed a strategy to demonstrate that Cas9 specificity is diminished by a local distortion of the DNA 3D structure in human cells. Next, by using the CRISPR system as a tool for gene expression regulation, I investigated the correlation between target transcription in cells and Cas9 editing efficiency. Together this first part of my work suggests that *in vivo* processes, which occur in eukaryotic cells and destabilise the DNA structure, have the potential to induce off-targets. However, this effect is highly dependent on the target itself and its genomic context. Finally, I extended my study to two other CRISPR-Cas systems: Cas12a and a new engineered Cas (AZ-Cas9). By applying single-molecule technologies, I contributed to the characterization of these nucleases, and obtained new information about the mechanisms underlining the DNA target search, the binding and cleavage kinetics and the off-target discrimination. These findings fill important knowledge gaps for future applications of these variants, and will be useful for the rational design of new high-fidelity nucleases.

ACKNOWLEDGEMENTS

I would first like to acknowledge my supervisor and mentor Dr Emanuela Cuomo, for her guidance over these years. It has been truly a pleasure knowing you, as a scientist and as a person, and working in your team. Also, thanks to Prof. Dr David Rueda, for his help and advice during my long secondment. The experience in your lab has been a great opportunity.

Many thanks to Dr Benjamin Taylor and all the Cell Biology team in AZ, UK, for their support. Thanks to Dr Ohad Yogev for being my first tutor and teaching me everything I needed when I first arrived in the lab. Also, thanks to Dr Marcello Maresca and his team in Göteborg for their collaboration.

I would like to thank the LightDyNamics network, PIs and students, for the excellent chance of being part of this project and for the amazing connections.

Thanks to the whole Rueda Lab, for being such a nice group to work with. I have been feeling welcomed from the very beginning and your help has been essential. Thank you, Dr Anita Meier, Dr Artur Kaczmarczyk, Dr Paul Girvan, George Moore, Paul Poudevigne-Durance, Quentin Smith, Dr Haruki Lino, Dr Erin Cutt, Dr Tianyi Zhang, and Dr Matthew Newton: I wish you all the best for your future, and thanks for the coffee breaks! In particular, thank you Quentin for the collaboration on our first paper. Also, thank you George for sharing with me all the best of the last part of this journey, it has been great having you by my side/desk.

Non smetterò mai di essere grata alla mia famiglia, per il loro amore e la loro fiducia. Grazie Mamma e Papà per essere esempio di forza e tenacia, per il sostegno e la dolcezza di cui mi circondate. Grazie Antonio perchè dove sono io, lì sei tu.

Grazie ai miei zii Paolo e Santina, e alle mie cugine Ezia e Chiara. Il vostro affetto e supporto mi fanno sentire fortunata e amata.

Grazie Isa perché il confronto e la condivisione con te sono sempre stati fondamentali e so che continueranno ad esserlo.

Many thanks to you, Maggie. You are one of the most beautiful and amazing gifts I got from this experience and our friendship is genuinely precious.

Grazie alle mie compagne di vita e al gruppo più giusto che ci sia perchè vi ritrovo sempre, o meglio, non vi perdo mai.

Marilú

"I am among those who think that science has great beauty."

Marie Salomea Skłodowska–Curie

Alla mia famiglia, forte e meravigliosa.

A Papà, Mamma e Antonio,

con amore.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 CRISPR-Cas introduction

In nature, the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)–CRISPR-associated protein (Cas) system is a bacterial defence mechanism against phage infection and plasmid transfer (Barrangou et al., 2007; Koonin and Makarova, 2019). This simple two-components CRISPR-Cas system, has been adapted for genome editing purposes and has become a versatile technology with innovative applications in basic biology research, medicine, and biotechnology. The CRISPR-revolution is set to impact our society in a wide range of fields including agriculture, human health, and infectious disease (Hille et al., 2018; Knott and Doudna, 2018). In general, the system relies on the activity of an RNA-guided nuclease which can be directed to any genomic sequence using a specific 20 nucleotides long guide RNA (Jinek et al., 2012a)

1.1.1 CRISPR in prokaryotic cell

Across species and all living organisms, from bacteria to human beings, there has always been a constant challenge in protecting ourselves from viruses. To defend against these threats, indeed, many organisms have developed innate mechanisms to make the infections harder to occur (e.g., by impeding viruses' entry through the cell membrane) or that respond more generally to the presence of infections (e.g., through inflammatory response) (Labrie et al., 2010). Even though those innate defences are crucial and essential, they likely cannot be as effective as mechanisms that target the specific infection. The challenge for developing targeted responses is to evolve a mechanism that can identify each incident virus and remember the specific targets (Bradde et al., 2020). In their adaptive immune system, vertebrates developed a learning-and-memory process, which produces novel antibodies through a mechanism of random genomic recombination (Flajnik and Kasahara, 2010). Those are then used to select effective immune elements when they bind to invaders, and then to maintain a memory pool to protect against future invasions. Recently, an adaptive immunity system has also been discovered in some bacteria and archaea. This system has been called with the acronym CRISPR, as the CRISPR loci contain short, partially palindromic DNA repeats that occur at regular intervals and form loci that alternate repeated elements (CRISPR repeats) and variable sequences (CRISPR spacers) (Figure 1.1). These specific loci were first observed in 1987 (Ishino et al., 1987), but they received little attention until similar loci were described in several microbial genomes (Jansen et al., 2002). Also, these loci have been found to be typically flanked by CRISPR-associated (Cas) genes (Haft et al., 2005). Their biological role remained ambiguous until 2005, when three groups reported that the spacers were homologous to foreign genetic elements, including viruses and plasmids (Bolotin et al., 2005; Mojica et al., 2005; Pourcel et al., 2005). These evidences led to the hypothesis that CRISPRs might function as an immune system (Makarova et al., 2006). Shortly after, data for CRISPR-mediated immune function were provided (Barrangou et al., 2007), and subsequent studies established that CRISPR-mediated defence involves sequence-specific, RNA-mediated targeting (Brouns et al., 2008) of mostly DNA (Marraffini and Sontheimer, 2008), and occasionally RNA (Hale and Duff, 2009). Since then, many efforts have been made to unveil the CRISPR-Cas system genetics, the molecular mechanism, and possible applications.

Overall, CRISPR-Cas immune systems function in three steps (Figure 1.1) (Behler and Hess, 2020). The first step is *adaptation*, in which new spacers are acquired from exogenous nucleic acid into the CRISPR locus (Figure 1.1) (Hille et al., 2018; McGinn and Marraffini, 2018). The adaptation step is followed by *crRNA biogenesis* (or *expression*), which comprises the expression of the CRISPR array from a transcription start site (TSS) that is usually located within the leader sequence upstream of the CRISPR array (Behler and Hess, 2020). These long transcripts, termed precursor-crRNAs (pre-crRNAs, Figure 1.1), are processed by specific endoribonucleases within the repeat sequences, which often form a hairpin structure due to the presence of palindromic sequence elements. Some CRISPR-Cas systems rely on an additional trimming or secondary cleavage step that is required to generate the active, mature crRNAs (Figure 1.1) (Behler and Hess, 2020). These mature crRNAs are composed of a repeat portion (tracrRNA) that is recognized by different Cas proteins and a spacer portion which mediates target specificity (Charpentier et al., 2015; Hille et al., 2018). Depending on the CRISPR class, several Cas proteins (in Class 1 systems), or a single Cas protein (in Class 2 systems), recruit the mature crRNA guide for DNA or RNA interference (Wang et al., 2022). The final step is *targeting* (or *interference*), in which, when a second infection occurs, the Cas nuclease enzyme is able to direct a sequence-specific destruction of the invading DNA thanks to the hybridization between the crRNA spacer and a complementary foreign target sequence (protospacer) (Figure 1.1) (Barrangou and Marraffini, 2014; Fineran and Charpentier, 2012; Garneau et al., 2010b; Marraffini, 2013; Sorek et al., 2013). Notably, the presence of a short-conserved sequence motif (2-5 bp) in close proximity to the crRNA-targeted sequence on the invading DNA, known as PAM, is essential. This motif prevents self-targeting of the complementary spacer within the CRISPR array, and it plays a crucial role in target DNA selection and degradation in most CRISPR-Cas systems (Bolotin et al., 2005; Mojica et al., 2009).

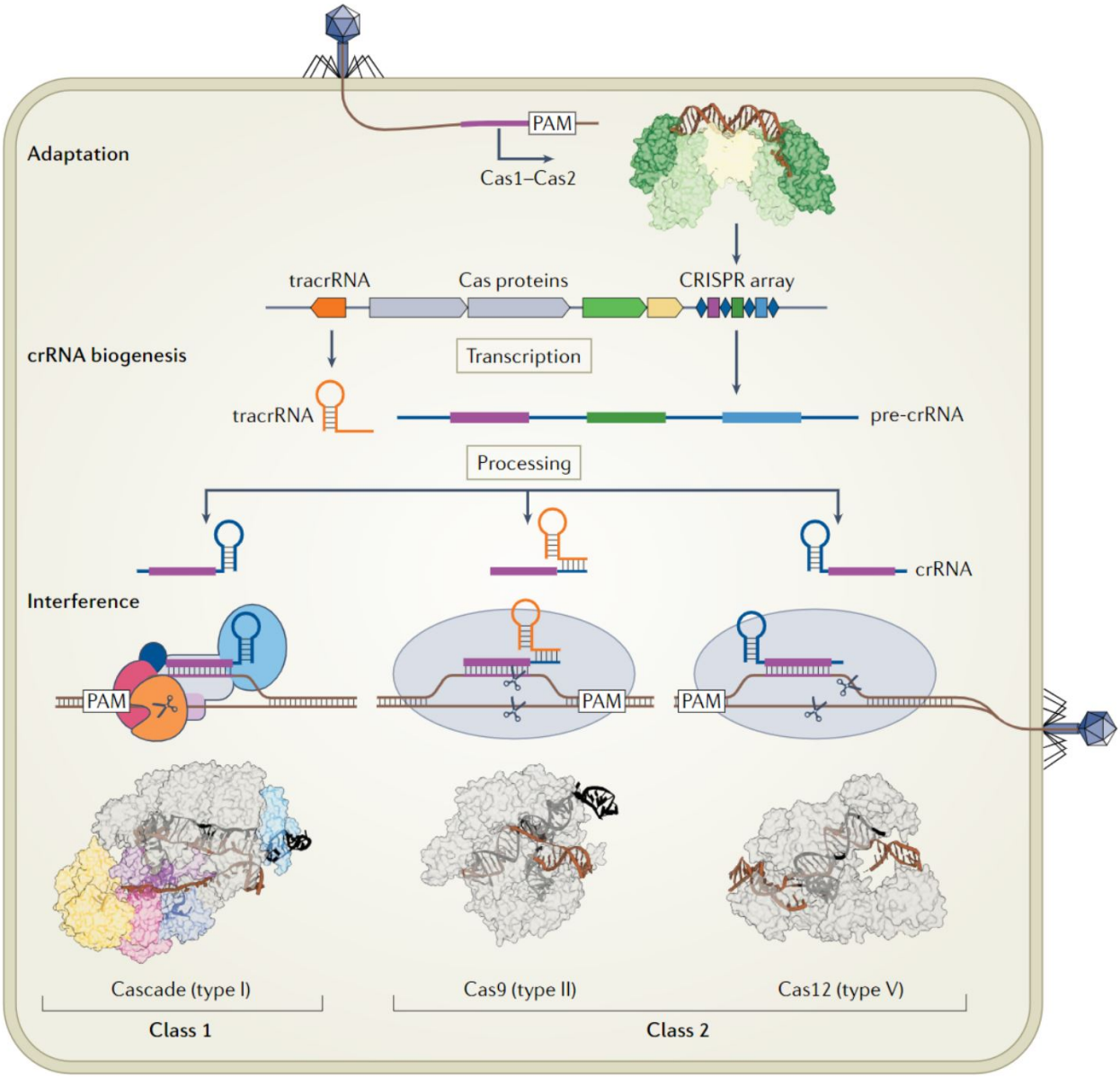


Figure 1.1: Schematic of CRISPR-Cas system in prokaryotic cells. A typical CRISPR locus consists of an array of repetitive sequences (diamonds) interspaced by short non-repetitive sequences (spacers, coloured squares). The array is associated with a set of CRISPR-associated (Cas) genes (coloured arrows) and trans-activating CRISPR RNA (tracrRNA) gene. The bacterial immune response is divided in three steps: adaptation, CRISPR RNA (crRNA) biogenesis and interference. During the *adaptation*, Cas1-Cas2 complex inserts foreign genetic elements into the CRISPR array as new spacers. During *crRNA biogenesis*, the CRISPR array is transcribed into pre-crRNA, which is further processed and then annealed with the tracrRNA to form mature crRNA. During the *interference* step, the mature gRNA engages Cas effector protein or complex to form a surveillance complex that recognizes and degrades foreign genetic elements containing a 20-nt crRNA complementary sequence preceding the PAM sequence. Class 1 systems have multisubunit effector complexes, whereas class 2 systems have single-subunit effector proteins. Cas1-Cas2 structure: PDB-5DS4; Cascade-Cas3 bound to crRNA and target DNA: PDB-6C66; Cas9 bound to guide RNA and target DNA: PDB-4UN3; Cas12a bound to crRNA and target DNA: PDB-5NFV. Image from Wang et al. 2022.

1.1.2 CRISPR classification

According to their protein sequence homology and CRISPR repeat identity, the CRISPR-Cas systems have been categorized into 2 major classes, 6 types and 33 subtypes (Figure 1.2) (Makarova et al., 2015; Shmakov et al., 2015; Makarova et al., 2019). The two classes radically differ with respect to the architectures of their effector modules involved in crRNA processing and interference. The class 1 systems have effector modules composed of multiple Cas proteins, some of which form crRNA-binding complexes (such as the Cascade complex in type I systems) that, with contributions from additional Cas proteins, mediate pre-crRNA maturation (Figure 1.2) (Makarova et al., 2019). On the other side, class 2 systems include a single, multidomain crRNA-binding protein (such as Cas9 in type II systems) that combines all activities required for interference and, in some variants, also those involved in pre-crRNA processing (Figure 1.2) (Koonin and Makarova, 2019). The class 1 CRISPR system is divided into types I, III, and IV, while the class 2 includes types II, V, and VI (Figure 1.2). Class 1 systems, primarily types I and III, are the most abundant in nature, being present in more than 90% of sequenced genomes of both bacteria and archaea (Makarova et al., 2015; Zheng et al., 2020), whereas class 2 only constitutes ∼10% of all CRISPR–Cas systems (Chylinski et al., 2014). Type III systems appear to be ancestral to all Class 1 systems, as proved by bioinformatic and structural analysis showing a high degree of similarity among these Cas effectors (Rouillon et al., 2013; Koonin and Makarova, 2013). The nucleic acid substrates recognized and cleaved also differ among the different CRISPR-Cas systems (Terns, 2018). Types I, II, V, and likely IV target double or single strand DNA, while Type III systems support cleavage of both DNA and RNA. Type VI systems appear to exclusively target RNA (Terns, 2018). Furthermore, several class 1 and class 2 CRISPR-Cas variants appeared to lack targeted cleavage activity and thus likely perform functions distinct from adaptive immunity (Koonin and Makarova, 2019). Such derived CRISPR-Cas systems include type IV, several variants of type I and at least one type V system variant.

In general, due to their simplicity, the class 2 CRISPR-Cas systems have been mostly employed for genome-engineering applications and in particular, SpCas9 from *Streptococcus pyogenes* (here referred as Cas9) is one of the biochemically and structurally best-characterized effector nucleases (Le Rhun et al., 2019).

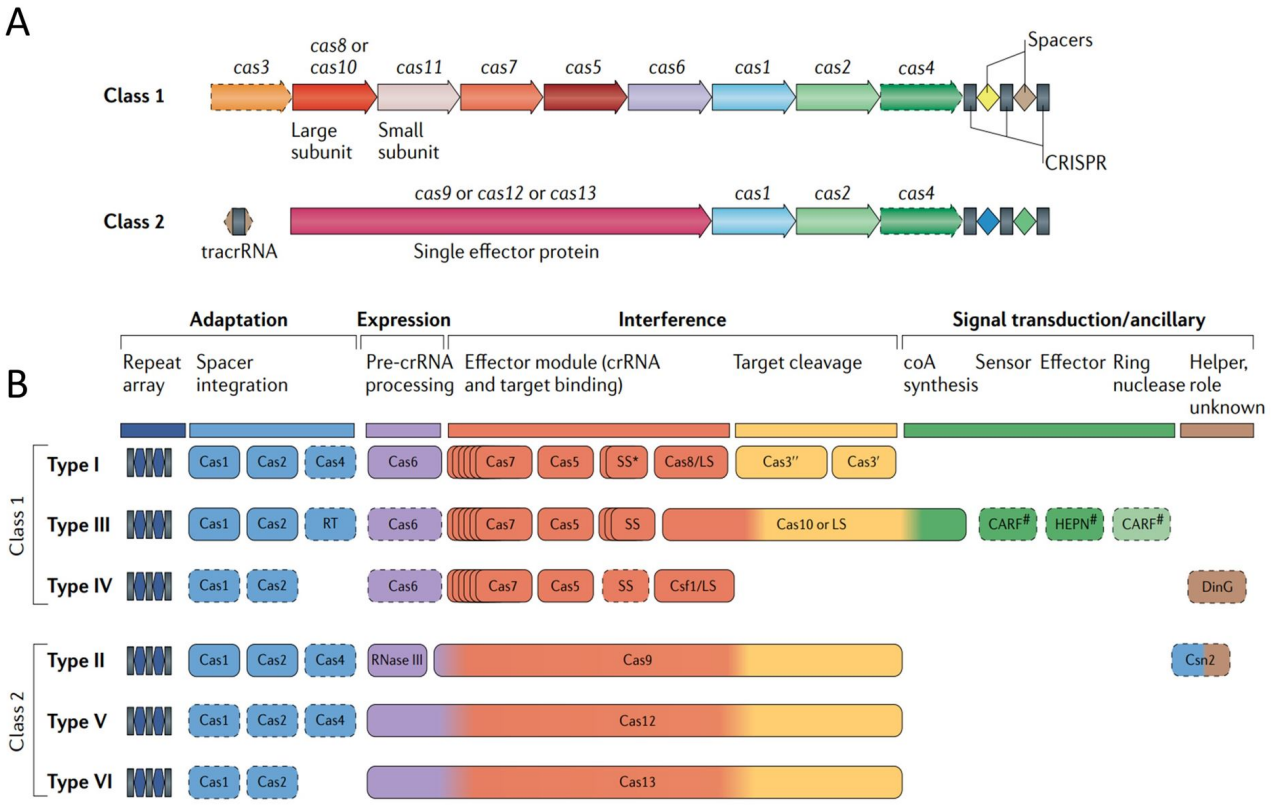


Figure 1.2: CRISPR systems classification. **(A)** Schematic organizations of generic CRISPR loci in Class 1 and Class 2 systems. Class 1 CRISPR–Cas systems have effector modules composed of multiple Cas proteins that form a crRNA-binding complex while Class 2 systems have a single, multidomain crRNA-binding protein. **(B)** Functional modules of CRISPR–Cas systems. The scheme shows the typical relationships between the genetic, structural and functional organizations of the six types of CRISPR–Cas systems. An asterisk indicates the putative small subunit that might be fused to the large subunit. The # symbol indicate that other unknown protein families could be involved in the same signalling pathway. Dispensable components are indicated by dashed outlines. Image from Makarova et al. 2019.

1.1.3 Class 2 systems

The DNA-targeting class 2 CRISPR–Cas system represents the first - and most studied - Cas effector used for precision genome editing. Class 2 enzymes, including Cas9 and Cas12, are now used widely for genome engineering across a broad range of species and have been adopted as programmable DNA-targeting modules for biotechnological applications (Knott and Doudna, 2018). Cas9 and Cas12 rely on similar principles to recognize and cut DNA, based on an RNA-guide complementary DNA sequence (Swarts and Jinek, 2018). These RiboNucleoProtein effectors (RNP, Cas:guide complex) become nuclease active only upon recognition of a cognate target DNA sequence. To identify the target, the RNP recognizes the PAM (which is G-rich for Cas9 but T-rich for Cas12) to initiate an ATP-independent DNA unwinding process that allows pairing of the DNA 'target' strand (TS) to the RNA guide (Gleditzsch et al., 2019). During the RNA-DNA hybridization process, the DNA 'non-target' strand (NTS) is unpaired from the TS, and the Mg^{2+}- dependent endonuclease cuts both DNA strands individually, using either two separate active sites (HNH and RuvC) producing a blunt ends cut as in SpCas9, or a single active site (RuvC) producing a staggered ends cut as in Cas12a (Swarts and Jinek, 2018; Stella et al., 2017).

1.1.3.1 Type II CRISPR-Cas9

Type II CRISPR-Cas9 proteins are multidomain enzymes that range in size from 700 amino acids (subtype II-D) to more than 1,700 amino acids (subtype II-C) (Wang et al., 2022). SpCas9 is a 1,368-amino-acid protein (158,441 Da) (Figure 1.3 A) (Jiang and Doudna, 2017a). Physiologically, the functional RNP includes Cas9 in complex with a 2-pieces-guide RNA (gRNA) formed by the crRNA hybridized to a tracrRNA scaffold (Jinek et al., 2012a). However, a synthetic single-guide RNA (sgRNA) has been engineered for experimental purposes and it consist of a direct fusion of the crRNA and the tracrRNA pieces (Jinek et al., 2012a). Structural analysis of Cas9 protein architectures shows that it exhibits a two distinct lobed organization, comprising a recognition (REC) lobe and a nuclease (NUC) lobe (Figure 1.3 A, B) (Jinek et al., 2014; Nishimasu et al., 2014). The REC lobe recognises the crRNA-TS duplex and it is composed of the bridge helix (BH), REC1, REC2, and REC3 domains, while the NUC lobe consists of the PAM-interacting domain (PI) and the nuclease domains RuvC and HNH

(Figure 1.3 A) (Jinek et al., 2014; Nishimasu et al., 2014). In the absence of a guide RNA, apo-Cas9 resides in an open conformation (Figure 1.3 B) that can eventually transition into a closed conformation upon guide recruitment, which stabilizes the flexible PAM-interacting domain for PAM identification (Figure 1.3 C) (Anders et al., 2014; Jiang et al., 2015; Jinek et al., 2014; Nishimasu et al., 2014). When the interaction between the target DNA and the gRNA occurs, it is stabilized by charge interactions, since the negatively-charged gRNA:target DNA heteroduplex is accommodated in a positively-charged groove at the interface between the REC and NUC lobes (Figure 1.3 C) (Swarts and Jinek, 2018). The gRNA binds the target DNA to form a T-shaped architecture comprising a guide:target heteroduplex, a repeat:anti-repeat duplex, and stem loops 1–3 (Figure 1.3 C) (Swarts and Jinek, 2018). Moreover, the RuvC domain interfaces with the PI domain to form a positively-charged surface that interacts with the 3 tail of the sgRNA. Accordingly, the hybridization of the gRNA with the TS DNA is accompanied by further conformational changes in REC2 and REC3 domains that accommodate the RNA–DNA hybrid in a central channel surrounded by the REC and NUC lobes (Figure 1.3 D). R-loop formation occurs directionally from the PAM-proximal side to the PAM-distal side. At this point, the SpCas9:gRNA complex further unwinds the remaining protospacer DNA, displaces the NTS DNA and examines the complementarity of the TS DNA with the gRNA (Figure 1.3 D) (Swarts and Jinek, 2018). The RuvC and HNH nuclease domains then shift their relative conformations in concert with the Rec lobe to enable nuclease activation (Jiang et al., 2016; Pacesa and Jinek, 2021; Sternberg et al., 2015; Zhu et al., 2019), in which the NTS DNA is bound in the catalytic pocket of RuvC domain, while the TS DNA is engaged by the HNH domain (Nishimasu et al., 2014). Cas9 RuvC domain has four catalytic residues (Asp10, Glu762, His983 and Asp986) with Asp10 being critical for the cleavage of the NTS DNA (Figure 1.4 A, B) (Swarts and Jinek, 2018). Cas9 HNH domain has three catalytic residues (Asp839, His840, and Asn863), with His840 being critical for the cleavage of the TS DNA (Figure 1.4 B) (Swarts and Jinek, 2018). Finally, Mg^{2+} divalent cations coordinated by the HNH and RuvC domain catalytic sites mediate a simultaneous cleavage of both strands, generating a DSB that is blunt-ended or has a single-nucleotide 5' overhangs (Jinek et al., 2012a; Nishimasu et al., 2014).

Regarding the target site search and recognition mechanism, it has been shown that it occurs

through three-dimensional collisions, in which Cas9 rapidly dissociates from DNA that does not contain the appropriate PAM sequence, and dwell time depends on the complementarity between guide RNA and adjacent DNA when a proper PAM is present (Sternberg et al., 2014; Knight et al., 2015; Ma et al., 2016; Newton et al., 2019; Globyte et al., 2019). Three-dimensional diffusion dominates Cas9 searching *in vivo*, however searching and binding at the target site seems to be dependent of the local chromatin environment, with less sampling and slower movement within heterochromatin (Knight et al., 2015). In a recent study disulfide crosslinking of an N^4-cystamine-modified DNA to a cysteine-functionalized SpCas9 (T1337C) allowed trapping of the transient interrogation state and revealed the mechanism by which Cas9 'reads' DNA (Cofsky et al., 2022). Cryo-EM data proved that the RNP associates with the PAM in an 'open conformation', then Cas9 twists and bends the DNA to flip the PAM-adjacent bases for DNA interrogation (Cofsky et al., 2022). The guide:TS pairing, results in a 'linear' crRNA-DNA hybrid intermediate that transits to a 'kinked' duplex conformation upon complete target binding (Bravo et al., 2022; Pacesa and Jinek, 2021). This kinked conformation would favours docking of the HNH domain to the guide:TS hybrid and also the binding of the NTS with the RuvC for cutting (Bravo et al., 2022; Pacesa and Jinek, 2021; Zhang et al., 2020). The presence of an excessive numbers of mismatches is thought to hinder the kinked conformation, thus resulting in inhibition of off-target DNA cutting (Bravo et al., 2022). However, recent studies are investigating further this 'conformational checkpoint', as other factors seem to be involved in Cas9 off-target discrimination mechanism. Indeed, it has been proved that the destabilization of the target DNA 3-dimensional structure can restore cleavage even in the presence of a high number of mismatches (Newton et al., 2019; Okafor et al., 2019; Singh et al., 2018b). Those findings would be in line with a 'kinetic model' describing Cas9 off-target discrimination through the energy landscape of target unwinding and RNA:DNA hybridisation during R-loop formation (Eslami-Mossallam et al., 2022; Klein et al., 2018; Ivanov et al., 2020).

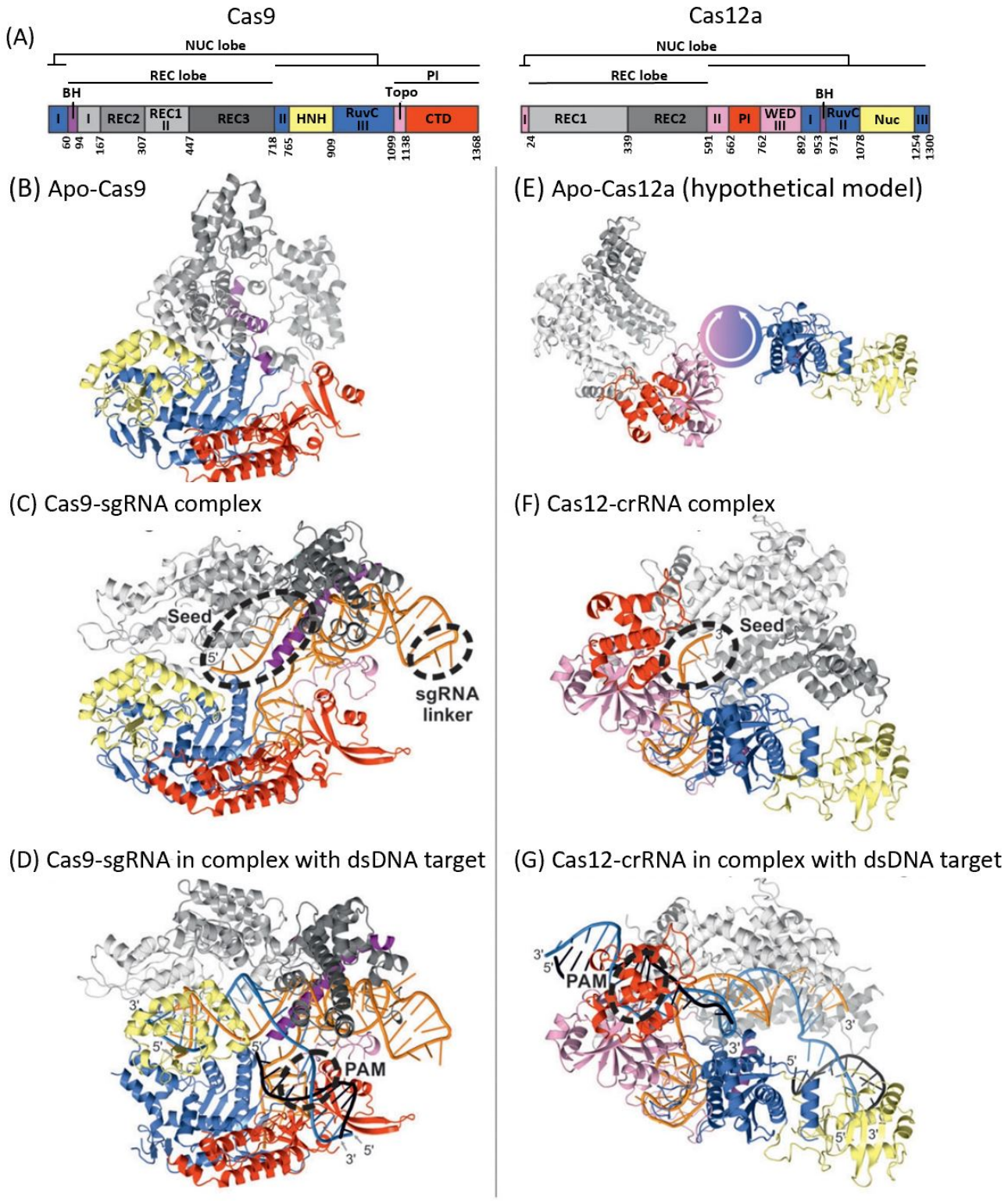

(A)
Cas9
NUC lobe
REC lobe
PI
BH
Topo
I
I
REC2
REC1
II
REC3
II
HNH
RuvC
III
CTD
60
94
167
307
447
718
765
909
1099
1138
1368
Cas12a
NUC lobe
REC lobe
BH
I
REC1
REC2
II
PI
WED
III
I
RuvC
II
Nuc
III
24
339
591
662
762
892
953
971
1078
1254
1300
(B) Apo-Cas9
(E) Apo-Cas12a (hypothetical model)
(C) Cas9-sgRNA complex
Seed
5'
sgRNA
linker
(F) Cas12-crRNA complex
Seed
3'
(D) Cas9-sgRNA in complex with dsDNA target
3'
5'
5'
PAM
3'
5'
(G) Cas12-crRNA in complex with dsDNA target
3'
5'
PAM
3'
3'
5'
5'
3'

Figure 1.3: Domain structure of SpCas9 (from *Streptococcus pyogenes*) and FnCas12a (from *Francisella novicida*). **(A)** Schematic representation of the domain organization of the two proteins. **(B)** Apo-Cas9 (PDB: 4CMP), without guide. **(C)** Cas9 bound to a sgRNA (PDB: 4ZT0). **(D)** Cas9–sgRNA complex bound to a DNA (PDB: 4UN3), showing a cleavage-competent conformation. **(E)** "Open" model of guide-free Cas12a (Dong et al., 2016). **(F)** Cas12 bound to a sgRNA, shows conformational changes, resulting in a "closed" conformation (PDB: 5NG6). **(G)** Cas12 bound to a DNA target (PDB: 5FNV) showing a cleavage-competent conformation. Image from Swarts and Jinek 2018.

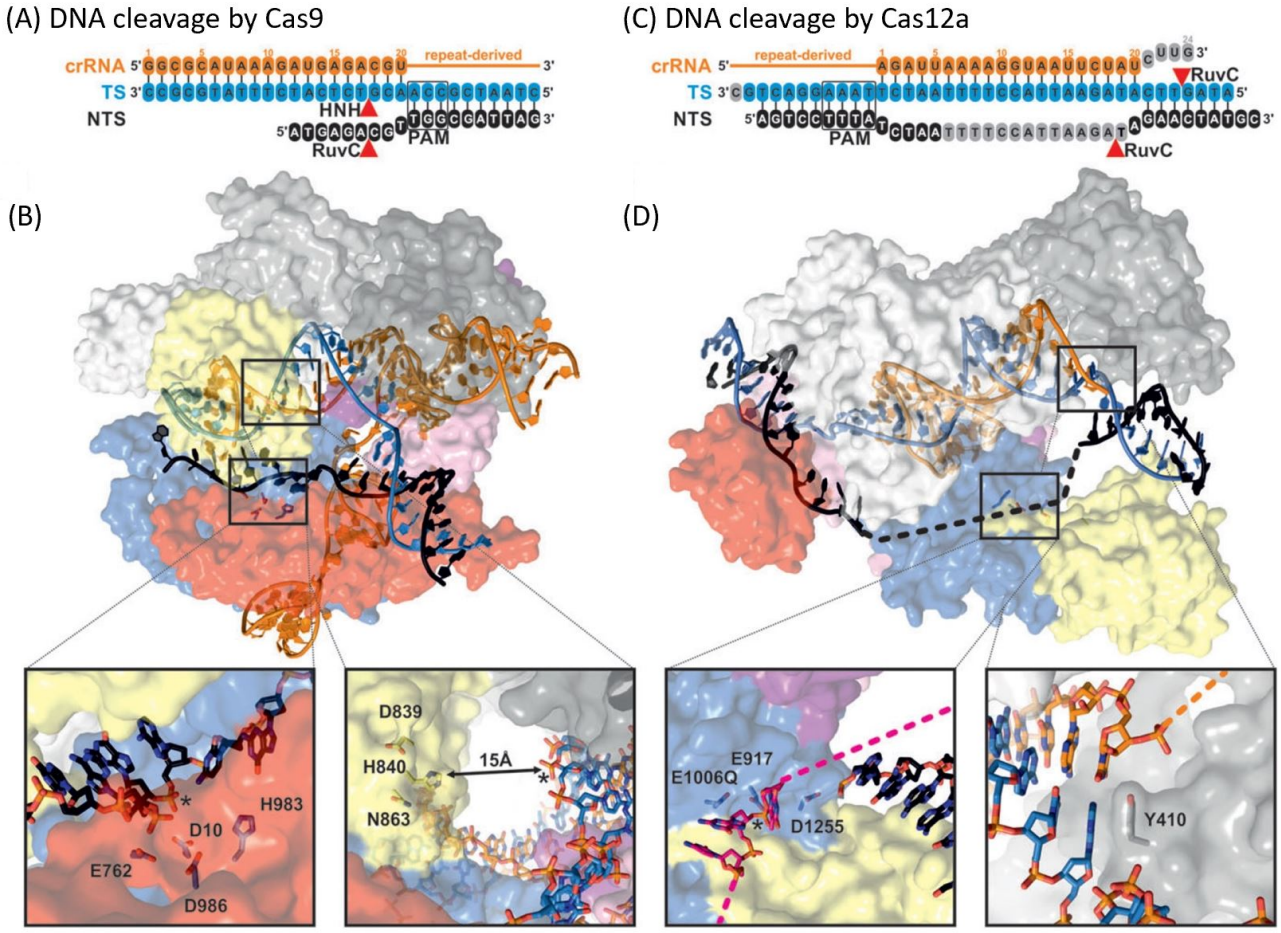


Figure 1.4: Domain structure of SpCas9 and FnCas12a: focus on target DNA binding and cleavage. **(A)** Schematic representation of DNA cleavage by SpCas9. **(B)** Catalytic residues located in the HNH and RuvC domains of SpCas9 (PDB: 5F9R). Close-up panels: Catalytic residues in the RuvC domain and in the HNH domain of SpCas9. **(C)** Schematic representation of DNA cleavage by FnCas12a. **(D)** Catalytic residues in the RuvC domain of FnCas12a (PDB: 5NFV). The black line indicates the hypothetical path of the NTS DNA. Left close-up panels: Catalytic residues in the RuvC domain of the FnCas12a (E1006Q,R1218A catalytic mutant). The pink line indicates the hypothetical path of the modeled TS DNA. Right close-up panel: Aromatic residue Y410 caps the RNA-TS heteroduplex in FnCas12a. The orange line indicates the hypothetical path of the 3' end of the crRNA. Image from Swarts and Jinek 2018.

1.1.3.2 Type V CRISPR-Cas12a

Among the Class 2 Cas-variants, the Type V CRISPR-Cas12a system (formerly Cpf1) (Zetsche et al., 2015) has been repurposed into an alternative and promising gene editing tool based on its substantial differences with Cas9. In general, Type V CRISPR Cas12 proteins are multidomain enzymes that range in size from 400-800 amino acids (subtype V-F; Cas14) up to 1,100–1,500 amino acids (subtype V-B; Cas12a) (Figure 1.3 A) (Wang et al., 2022). Cas12 proteins are classified into more than 10 subtypes (subtypes V-A to V-K, and subtype V-U) (Makarova et al., 2019), which differ in their RNP biogenesis pathways, RNP composition and in some instances in their nucleic acid target preference (Wang et al., 2022). Functional RNPs include a Cas12 protein bound either to one single crRNA guide, as the case for Cas12a (Zetsche et al., 2015), or to a crRNA and tracrRNA hybrid or a crRNA and short-complementarity untranslated RNA (scoutRNA) hybrid (Wang et al., 2022). The mature crRNA for Cas12a is 42-44 nt in length, with the first 19/20 nt corresponding to the repeat sequence and the remaining 23-25 nt to the spacer sequence (Zetsche et al., 2015). After the pre-crRNA has been transcribed during the expression stage, Cas12a cuts it 4 nt upstream of the hairpin structures formed by the CRISPR repeats, producing intermediate crRNA molecules which undergo further processing into mature crRNAs (Dong et al., 2016; Fonfara et al., 2016; Gao et al., 2016; Swarts et al., 2017). The ability of Type-V Cas12a to process its own pre-crRNA into mature crRNAs, without the requirement of a tracrRNA, makes it a unique effector protein with both endoribonuclease and endonuclease activities. From crystallographic and cryo-EM analysis, it has been possible to understand that Cas12a adopts a Cas9-analogous bilobed architecture formed by the REC and Nuc lobes (Figure 1.3 A, E) (Yamano et al., 2016). The REC lobe is comprised of REC1 and REC2 domains, and the Nuc lobe is comprised of the RuvC (RuvC I-III), the PAM-interacting domain (PID) and the WED (Wedge) domains, and additionally, the bridge helix (BH) (Figure 1.3 A, E) (Yamano et al., 2016). In the absence of a guide RNA, Cas12a assumes a flexible 'open' conformation (Figure 1.3 E) (Dong et al., 2016), which upon crRNA binding transitions to a 'closed' conformation (Figure 1.3 F, G) (Dong et al., 2016; Swarts et al., 2017; Yamano et al., 2016) that is poised for PAM recognition (Gao et al., 2016; Swarts et al., 2017). In the closed conformation, the Rec domains structurally occlude the RuvC active site for nuclease repression (Figure 1.3 G) (Dong et al., 2016; Swarts

et al., 2017). The interaction with the crRNA is stabilized through connections with the WED, RuvC and REC2 domains of the endonuclease, as well as two hydrated Mg^{2+} ions (Figure 1.3 F) (Dong et al., 2016; Stella et al., 2017; Swarts et al., 2017). Upon PAM-dependent unwinding of a dsDNA target and hybridization of the TS to the crRNA, the REC domains rearrange themselves to accommodate the heteroduplex (Figure 1.3 G) (Gao et al., 2016; Swarts et al., 2017). Notably, the Cas12a "seed" segment is shorter than Cas9, being 5–6 nucleotides long and it is located at the 5' end of the spacer-derived segment of the crRNA. Once the complementarity with the TS is checked, a cascade of conformational changes leads to the opening and activation of the RuvC active site, for the cleavage step to occur (Figure 1.3 G) (Cofsky et al., 2022; Stella et al., 2017; Swarts et al., 2017). Since the active site can only be occupied by only one DNA strand at a time, the two substrate DNA strands are presumably cut sequentially. Even though the precise mechanism of DNA cleavage in Cas12a nuclease is still under study, the hypothesis is that after NTS cleavage and dissociation from the catalytic site, the PAM distal part of the R-loop structure unwinds, allowing the TS DNA to be "pushed" into the RuvC catalytic centre (Saha et al., 2020). This sequential mechanism of DNA cleavage by a single active site results in a staggered-end DNA break with 5' overhangs, typical of Cas12 nucleases (Cofsky et al., 2022; Stella et al., 2017; Swarts et al., 2017). Importantly, this peculiar cleavage mechanism and the presence of overhangs at the cut site, is considered advantageous for gene-editing purpose since it may promote site-directed integration events.

The target search mechanism is currently under study and more information is needed to complete the full picture of this mechanism. However, it appears that Cas12a undergoes random and bidirectional diffusion while interacting with dsDNA (Jeon et al., 2018). Together with a 3D random collision mechanism, Cas12a exhibits this unique 1D diffusion mechanism of target search, which seems to be predominant. A solid understanding of the DNA interrogation mechanism will be fundamental for further development of Cas12-based genome editing tools. An additional main difference from Type-II Cas9 system is represented by Type-V Cas12a ability to carry out indiscriminate ssDNA degradation. Recent structural data revealed the ssDNA substrate bound to two magnesium cofactors within the RuvC active site, can trigger the catalysis mechanism of Cas12 enzymes (Huang et al., 2020; Pausch et al., 2021). After cis DNA cutting, the RuvC domain remains activated for nonspecific ssDNA 'shredding' in

trans (Chen et al., 2018; Swarts and Jinek, 2019). This "collateral activity" is common to all Cas12a orthologs and degrades any available ssDNA molecule into single/double nucleotides (Chen et al., 2018). Structural analysis revealed that the lid region, within the PI domain, is involved in this mechanism (Stella et al., 2018). Upon the formation of the crRNA-DNA hybrid, and after the TS cleavage, the lid undergoes a conformational change which allows for new interactions with the crRNA of the hybrid assembly, thus dissociating the polar interactions and making the catalytic pocket again accessible and able to cleave ssDNA indiscriminately (Stella et al., 2018). In addition, recent studies have reported non-specific nicking of target sequences bearing mismatches in distal regions of the target DNA (Murugan et al., 2020), suggesting that this could be a problem for potential applications.

1.2 Genome editing application

1.2.1 Method

The purpose of gene-editing technologies is to precisely change DNA sequences at specific target sites, and, in such a manner, study the function of the gene variant at the endogenous level, under endogenous regulation mechanisms. The extension of genome editing for gene functionality studies is to edit genomes to modify plants or organisms for research purposes or to edit human genes by correcting gene mutations causing specific genetic diseases (Li et al., 2020). The key step for editing a gene is to cause a break in both strands of DNA double helix, known as Double Strand Break (DSB) (Figure 1.5). When this happens, neither strand can serve as template to repair the other, so the cell will activate one of the alternative DNA damage responses (DDR) pathways to physically repair the damage (Damage et al., 2006). The fastest and most efficient mechanism that the cell can use to restore the break is via Non-Homologous End-Joining (NHEJ). In such a case, the DNA is repaired by simply joining the two broken ends together, without trying to find a matching sequence. However, this mechanism can be error-prone and inaccurate, resulting in the generation of an early stop-codon with consequent gene Knock Out (KO) due to the incorporation or loss of random nucleotides (indels) (Figure 1.5). A more precise process, termed Homolog-Direct Repair (HDR), can occur when the DNA finds a suitable replacement template nearby. In this case, the cell will usually copy and insert the available homologous sequence where the DSB occurred. This is at the basis of performing precise gene modification, such as gene Knock-In (KI), deletion, correction, or mutagenesis (Damage et al., 2006) (Figure 1.5). The design of a new gene editing technology requires two main considerations. First, it is necessary to identify the right enzyme that could cut the DNA, a so-called "nuclease", and create a DSB. Next, it is essential to identify a guide that would navigate the enzyme to the precise target in the cell's DNA, where the DSB is required. The first enzyme used for the gene editing purpose was FokI (Kim et al., 1996), which is an endonuclease working as a dimer. This nuclease was initially fused to the so-called zinc-fingers proteins (ZNFs), which are able to recognize a trinucleotide DNA sequence and thus serve as guide for FokI cleavage. Accordingly, other technologies have made use of transcription activator-like effector nucleases (TALENs) which can also be engineered to recognise and cleave

a specific DNA sequence (Christian et al., 2010). In this case a single TALEN motif can be engineered to recognize one nucleotide but an array of TALENs can be associated with a longer sequence. Just when TALENs were being perfected, CRISPR came along and gained popularity due to its simplicity. One of the main differences that makes the CRISPR-system particularly advantageous is that the guide is not a protein but a RNA molecule and the customization of just the crRNA part, which can be then easily annealed with the tracrRNA *in vitro* to form the full gRNA, enables specificity in every CRISPR experiment. Additionally, the development of the synthetic sgRNA also increased the ease of the experimental assays (Jinek et al., 2012a). While ZNFs and TALENs require the construction of a new protein for every target gene to be modified, which is difficult, expensive and time consuming, each sgRNA can be synthetically generated or made *in vitro* or *in vivo* from a DNA template. The discovery of the CRISPR-Cas system and all the progresses that have been made (and are still ongoing) to convert it in an appropriate, efficient and safe genome editing tool, are revolutionizing the gene editing and gene therapy fields.

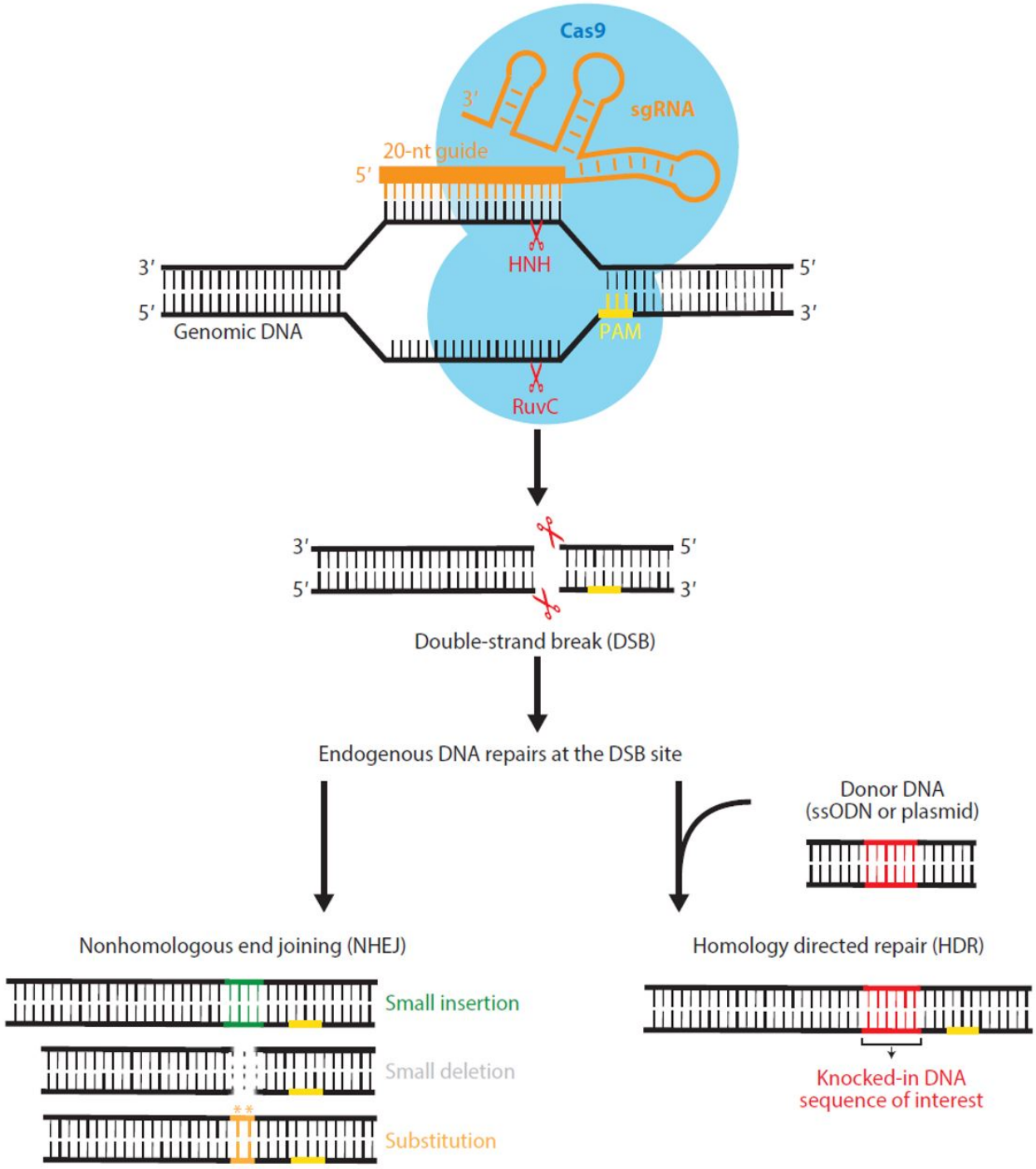


Figure 1.5: Mechanism of CRISPR-Cas9-mediated genome engineering. Cas9 cleavage is directed via sgRNA at a specific genomic site and a double-strand break (DSB) is introduced. In the absence of a DNA template, the cell activates the non-homologous end-joining (NHEJ) pathway, which is error-prone and causes random insertions and deletions (indels) or even substitutions at the DSB site. This frequently results in the disruption of gene function (KO). In the presence of a donor template containing a sequence of interest flanked by homology arms, the cell can initiate the homology directed repair (HDR) pathway. This is used to create desired mutations through homologous recombination. Image from Jiang and Doudna 2017b.

1.2.2 Cas9 therapeutics

As already mentioned, CRISPR-Cas9 has a great potential for therapeutic applications (You et al., 2019). In 2017 and 2018, CRISPR-based therapeutic applications were assessed in large animals and human embryos, and a number of CRISPR-related clinical trials have been set in place since then. The first CRISPR Phase 1 clinical trial was opened in 2017 in the US (Baylis and McLeod, 2017) with the intent to use CRISPR-Cas9 to edit autologous T cells for cancer immunotherapy against several cancers with no other curative treatment options (Uddin et al., 2020). After all the consideration of the case about the risk to benefit ratio, this trial was approved by the United States Food and Drug Administration (FDA). Moreover, the first clinical trial using CRISPR to catalyse gene disruption for therapeutic benefit was opened for patients with sickle-cell anaemia (SCD) and β-thalassemia (Frangoul et al., 2021), by Vertex Pharmaceuticals and CRISPR Therapeutics in US. On the other hand, *in vivo* CRISPR-based gene therapy approaches have been less extensively employed, but an interesting study has been recently launched with a clinical trial using *in vivo* delivery of CRISPR-Cas9 in patients affected of Leber Congenital Amaurosis (LCA) (Maeder et al., 2019), a debilitating monogenic disease that results in childhood blindness. However, numerous controversies still exist, mainly linked to the potential off-target effects (Section 1.3.3), the carcinogenic effect of CRISPR components, and the immunogenicity of Cas9 nucleases, which require further extensive studies. Indeed, there is an urgent need to focus on in-depth understanding of the intracellular mechanisms underlying CRISPR-Cas9-mediated gene editing for appropriate and efficient clinical applications of such technologies. Additionally, of high relevance is the ethical question related to the possibility of using this system to treat less severe diseases, as the risks of the technology are better understood (You et al., 2019). A concern remains whether this will lead to human genome editing for non-medical purposes, such as altering genes in human embryos to create offspring with certain aesthetic traits. The scientific community, together with public institutions, have to make important decisions about how the ethical boundaries of CRISPR applications should best be drawn (Brokowski and Adli, 2019; Caplan et al., 2015).

1.2.3 Cas12a-based gene editing

Even though Cas9-based genome editing is at the moment the most established, among Cas12 nucleases, the use of FnCas12a (from *Francisella Novicida*) , AsCas12a (from *Acidaminococcus sp.*), and LbCas12a (from *Lachnospiraceae bacterium*) as genome editing tools has been recently investigated. Firstly, it has been found that Cas12a functions as an efficient genome editing tool in various bacterial species and yeast, while the utilization of Cas9 for genome editing in certain bacteria is sometimes toxic (Swiat et al., 2017; Ungerer and Pakrasi, 2016). Cas12 plant-editing also has been successful and, moreover, it revealed and proved that Cas12a-mediated genome editing appears to be more efficient when Cas12a is provided with unprocessed pre-crRNAs compared to when crRNAs with fully processed 3' ends are provided (Xu et al., 2017). AsCas12a has successfully been used for genome editing in *Drosophila melanogaster* (Port et al., 2020), while LbCas12a in *Danio rerio* (Zebrafish) and *Xenopus tropicalis* (Moreno-Mateos et al., 2017). In general, the prospective of using Cas12 systems as a valid alternative to Cas9 is promising but more details need to be unveiled about its mechanism of action. In particular, for gene editing purposes, even if the lower off-target activity is certainly advantageous (Kleinstiver et al., 2016b; Kim et al., 2016), the non-specific nature of the ssDNA degradation and nicking activities have to be further investigated and properly addressed, along with the genomic state and DNA topology effect.

1.2.4 The CRISPR toolbox

1.2.4.1 dCas9 and Base Editor

Since its discovery, the CRISPR-Cas system has become famous as the "Molecular scissor" of the genome-editing field. However, thanks to its robustness and flexibility, this tool is having a great impact also in many other alternative application areas for genome and chromatin manipulation too, being more similar to a "Molecular Swiss army knife" (Mans et al., 2015). Cas9 has been modified to improve some of its functionalities, such as to recognize different PAMs or to be more specific (Sections 1.3.2 and 1.3.4), but also decreasing its activity has been used as a strategy to make it more suitable for different gene-editing purposes (Adli, 2018; Doudna, 2020) (Figure 1.6). Many applications, for example, use Cas9 DNA biding properties without its nuclease activity. To generate a so called "dead" (dCas9), two point mutations in the two catalytic domains HNH and RuvC (H840A and D10A) are introduced, giving rise to a nuclease defective protein (Jinek et al., 2012b). This dCas9 is still able to find and bind its target, but it will not cut the DNA strands. Notably, a single point mutation in either of these domains, results in a nickase enzyme. These Cas9 variants have been adapted for a wide range of genome-targeting purposes and one of the key progresses in the field has been the development of base-editing technologies. These are known as second-generation genome-editing tools and are able to precisely convert a single base into another without causing DNA DSBs (Figure 1.6 B) (Adli, 2018). The nickase Cas9 is the foundational platform for the base editor tools that enable C to T or A to G conversion at the target site (Figure 1.6 B) (Gaudelli et al., 2017; Komor et al., 2016). C to T conversion is obtained by fusing the nickase Cas9 to an APOBEC1 deaminase enzyme and Uracyl Glycosylase Inhibitor (UGI) protein, while A to G conversion is achieved with the fusion of a transfer RNA adenosine deaminase. These novel base-editing approaches significantly expand the scope of genome targeting and, in the last few years, new discoveries are contributing to this aim. As an example, Prime Editing is a recent genome editing technology that can introduce all 12 possible types of point mutations, small insertions and small deletions in a precise and targeted manner with favourable editing to indel ratios (Anzalone et al., 2019). This technology utilizes Prime Editors (PE), which are fusion proteins between a Cas9 nickase domain (inactivated HNH nuclease) and an engineered Reverse Transcriptase (RT) domain. PE is targeted to the editing site by an engineered prime editing

guide RNA (pegRNA), which not only specifies the target site in its spacer sequence, but also encodes the desired edit in an extension that is typically at the 3' end of the pegRNA. Upon target binding, Cas9 nicks the PAM-containing DNA strand and then uses the newly liberated 3' end at the target DNA site to prime reverse transcription using the extension in the pegRNA as a template. Moreover, a recent study showed that a Prime Editor nuclease (PEn), which combines RT and the wild-type SpCas9 nuclease, can be used to perform prime at DSBs sites by utilizing DNA end joining repair pathways (Peterka et al., 2022).

1.2.4.2 CRISPRi and CRISPRa

The dead Cas variants are also applied to the field of gene expression regulation (Figure 1.6 C) (Adli, 2018; Xu and Qi, 2019). Interestingly, since dCas9 strongly binds to the DNA target sequence, this tight binding event interferes with the activity of other DNA binding proteins such as endogenous transcription factors and RNA Polymerase II (Qi et al., 2013). This mechanism has been turned to account the development of the CRISPR interference (CRISPRi) approach, in which dCas9 binding activity blocks the transcriptional process and thus knocks down gene expression (Figure 1.6 C) (Qi et al., 2013). Notably, fusing a strong repressor complex such as Kruppel-associated Box (KRAB) to dCas9 results in a stronger and specific gene repressor than dCas9 alone (Gilbert et al., 2014). These transcriptional regulators are used to further recruit additional co-repressor proteins such as KRAB-box-associated protein-1 (KAP-1) and epigenetic modifiers such as heterochromatin protein 1 (HP1) proteins to repress genes (Friedman et al., 1996). The KRAB-mediated gene repression mechanism is partly mediated by local epigenetic reprograming of histone modifications at the promoter level (loss of histone H3-acetylation and an increase in H3 lysine 9 trimethylation) (Groner et al., 2010). In agreement with these findings, the dCas9-KRAB fusion complex results in reduced chromatin accessibility and increased H3K9me3 levels at both targeted gene promoters as well as distal enhancers (Thakore et al., 2015). Similarly to dCas9-KRAB-mediated gene repression, the dCas9-targeting platform can be used to recruit strong transcriptional activators and it results in robust induction of gene expression (Figure 1.6 C) (Adli, 2018). To this end, dCas9 has been fused to VP64 (Maeder et al., 2014; Mali et al., 2013; Perez-Pinera et al., 2013), which is composed of four tandem copies of a 16-amino-acid-long transactivation domain (VP16) of

the Herpes simplex virus (Wysocka and Herr, 2003). The dCas9-VP64-mediated gene activation strategy has been then improved by a number of second-generation CRISPR-based gene activation platforms. This include the fusion of dCas9 with a tripartite transactivation complex, which is composed of VP64, P65 (a transcription activation domain of the mammalian NF-κB transcription factor), and Rta (a transactivator from the Epstein-Bar virus). This VPR, as a complex of proteins, allows to achieve a higher and robust gene induction (Chavez et al., 2015). Notably, in addition to the dCas9 direct fusion, the effector domains can also be recruited through the sgRNA scaffold (Figure 1.7) (Adli, 2018). For these approaches, the sgRNA scaffold is engineered to contain RNA modules such as MS2 hairpin aptamers that can bind to specific RNA binding proteins such as bacteriophage MS2-coat protein (MCP) (Figure 1.7) (Peabody, 1993; Sugiyama and Nakada, 1967). The engineered sgRNA-MS2 scaffold is then used to recruit MCP-fused VP64 or other trans-activation complexes to activate expression from an endogenous locus. In another approach, called synergistic activation mediator (SAM) complex, dCas9-VP64 fusion complex has been used simultaneously to the recruitment of MCP-fused P65-HSF1 transactivation domains (HSF1: heat shock transcription activator) to the target site through the engineered sgRNA scaffold (Konermann et al., 2014). Lastly, in a strategy termed SunTag, dCas9 fused protein scaffold that contains repeating peptide array, is used to recruit multiple copies of an antibody fused effector protein (Figure 1.7) (Tanenbaum et al., 2014). Precise spatial and temporal control over the dynamics of gene expression from a target locus has great therapeutic potential and the flexibility of the CRISPR-system is showing a promising potential in this regard.

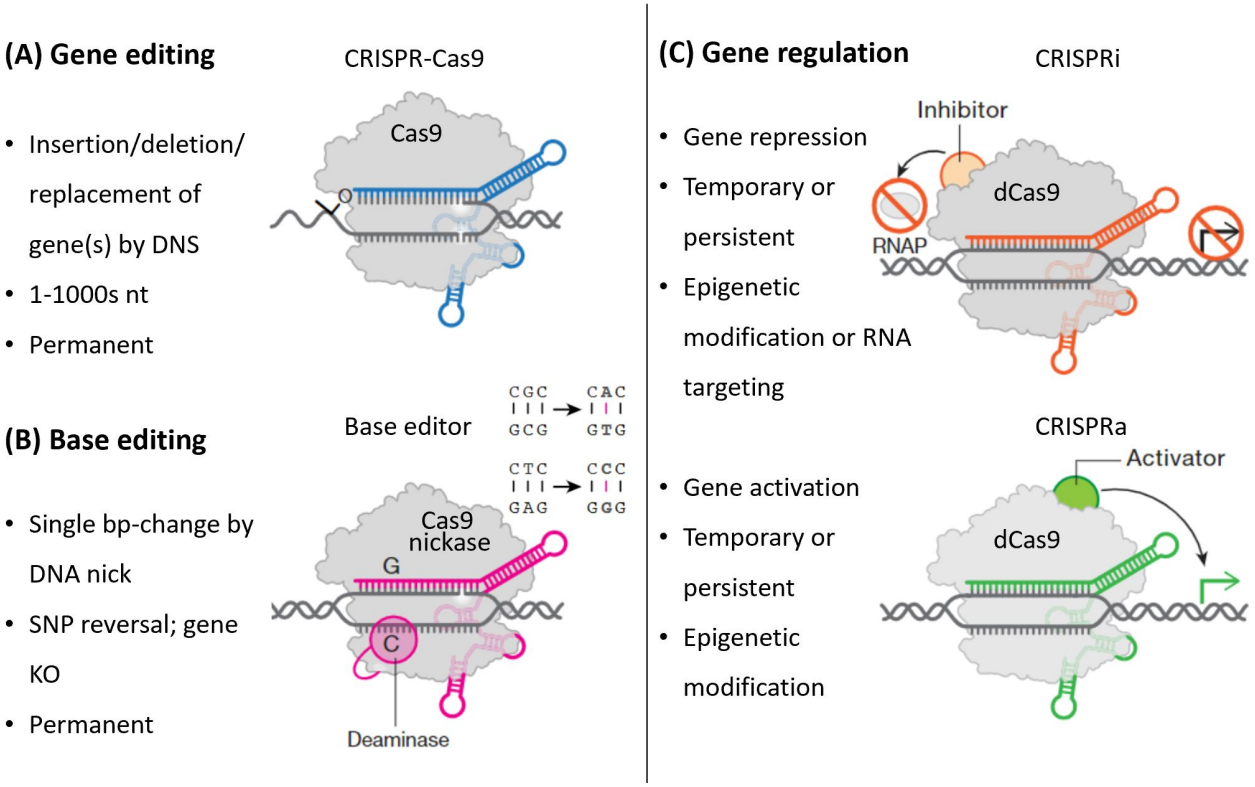


Figure 1.6: CRISPR-based tools. **(A)** Gene editing: CRISPR–Cas9 generates a double-stranded break (DSB) at the target site to simulate endogenous DNA repair. **(B)** Base editing: nickase-Cas9 with a fused domain replaces a single base through deamination and DNA replication or repair. **(C)** Gene regulation: dCas9 CRISPR-mediated gene repression or interference (CRISPRi) sterically blocks the RNA polymerase and induces heterochromatinization. dCas9 CRISPR-mediated gene activation (CRISPRa) is used to recruit the transcription machinery to increase expression of the target region and leads to direct epigenetic modifications. Image adapted from Doudna 2020.

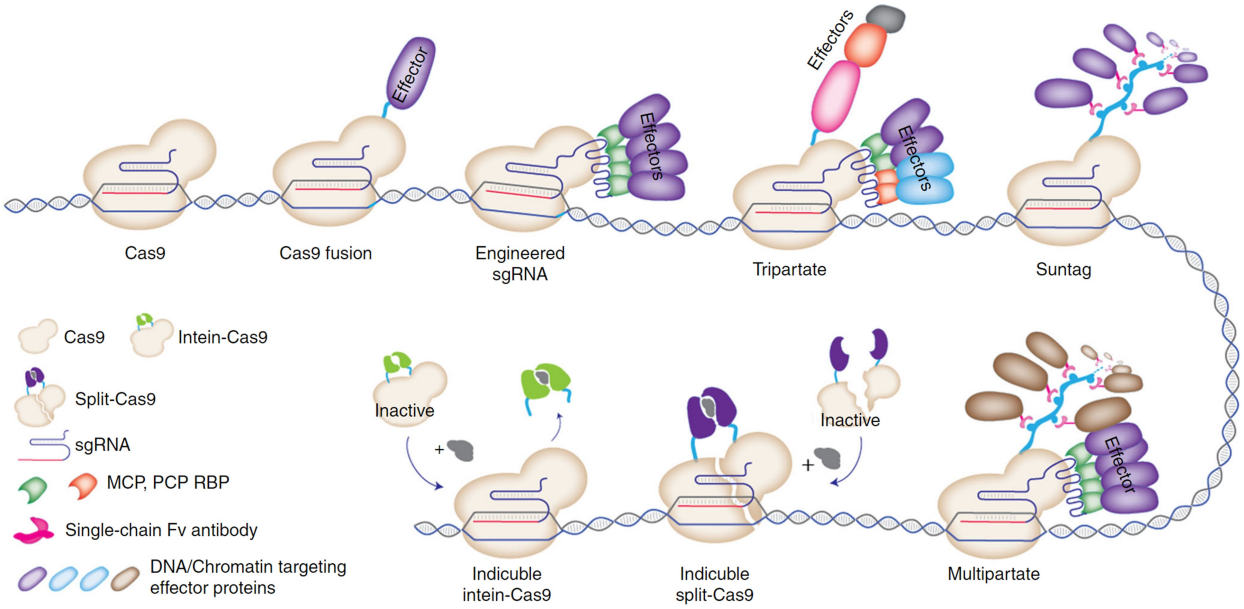


Figure 1.7: Effector proteins recruitment strategies. Effector proteins can be directly fused to active Cas9 or dCas9 through a linker peptide. The sgRNA scaffold can be engineered to contain RNA aptamers that specifically bind to a known RNA binding proteins (MCP or PCP). In the Tripartite strategy, multiple different effectors are recruited through dCas9 as well as engineered sgRNA scaffold. The SunTag approach utilizes a repeating peptide array of protein scaffold to recruit multiple copies of an antibody-fused effector protein. Additionally, chemically inducible strategies enable temporal control over the activity of Cas9 or Cas9-fused effector proteins. In the split Cas9 approach, each half of Cas9 protein can be induced to form functional complex. In the intein-Cas9 approach, the intein protein segment can be chemically induced to excise from Cas9 and result in its activation. Image from Adli 2018.

1.2.4.3 CRISPR-based epigenome editing

Finally, dCas9 has been efficiently used for the important purpose of epigenome editing, which involves all post-translational modifications and other chromatin features associated with regulatory elements in the genome (Adli, 2018). There are multiple layers of epigenetic regulatory mechanisms operating in the genome such as DNA methylation, histone post-translational modifications, and non-coding RNAs (short and long). One of the most studied epigenetic mechanisms of gene regulation is DNA methylation, which is generally associated with transcriptional repression (Moore et al., 2012). Moreover, aberrant DNA methylation is implicated in a number of pathological diseases, including cancer. Thus, with the aim of manipulate this epigenomic feature and deposit DNA methylation at a specifically targeted locus, dCas9 has been fused to the catalytic domain of eukaryotic DNA methyl transferase (DNMT3A) (Liu et al., 2016; Vojta et al., 2016) or prokaryotic DNA methyltransferase (MQ3) (Lei et al., 2017). Those strategies lead to substantial DNA methylation and are able to alter gene expression at specific target site, together with recruitment of additional components of repressive epigenetic machinery that further enhanced gene repression. In addition to targeted DNA methylation, dCas9-TET1 (Ten-Eleven Translocation proteins which mediate DNA demethylation) fusion complex can be used for efficient DNA demethylation and substantial increase in mRNA expression at the target sites (Liu et al., 2016; Morita et al., 2016; Xu et al., 2016). Finally, another mechanism to consider is based on post-translational modifications on the histone tails (Bowman and Poirier, 2015), which also constitute an important epigenomic features for regulatory activity of genomic elements. It has been shown that dCas9-mediated locus-specific recruitment of LSD1 (histone demethylase) results in substantial local reduction in the active enhancer markers H3K4me2 and H3K27ac and it was able to alter the expression of target genes (Kearns et al., 2015). Moreover, by using dCas9 fused to histone deacetylases (HDAC), it is possible to effectively reduce the H3K27ac and reduce the gene expression of the target loci (Kwon et al., 2017). On the other hand, recruitment of histone acetyl transferase P300 through dCas9 fusion results in a significant increase in local H3K27ac levels at enhancer elements, promoting gene expression (Hilton et al., 2015). Identifying the link between epigenetic marks and gene expression remains a central goal of chromatin biology and to this end, dCas9 guidable platform is now an indispensable tool to achieve this (Adli, 2018).

1.2.4.4 CRISPR-based diagnostic methods

In addition to the above mentioned CRISPR-tool applications, the properties of different Cas enzymes have been adopted for diagnostic assays development (Kaminski et al., 2021). Nucleic-acid-based diagnostics have become the gold standard for various acute and chronic conditions, particularly those caused by infectious diseases (Kaminski et al., 2021). In fact, nucleic-acid-based biomarkers associated with disease are essential for diagnostics because DNA and RNA can be amplified, enabling their highly specific detection via the pairing of complementary nucleotides (Yang and Rothman, 2004). One of the first developed CRISPR-based diagnostic methods is NASBACC (Nucleic Acid Sequence-Based Amplification (NASBA)-CRISPR Cleavage) (Pardee et al., 2016). This method combines nucleic acid sequence-based amplification (Compton, 1991) for the isothermal preamplification of targets, and Cas9 cleavage for PAM-dependent target detection with a toehold sensor for the readout. This has been efficiently used for the detection of the Zika Virus in infected monkey plasma at concentrations in the low femtomolar range (Pardee et al., 2016). A second Cas9-based method, referred as LEOPARD (for Leveraging Engineered tracrRNAs and On-target DNAs for PArallel RNA Detection), has been recently developed (Jiao et al., 2021) and this is based on the observation that tracrRNAs can hybridize to cellular RNAs, creating 'non-canonical' crRNAs. By reprogramming tracrRNAs to bind cellular transcripts of interest while allowing for the formation of Cas9–crRNA complexes, the resulting non-canonical crRNAs enable the guiding of Cas9 to a DNA target (Jiao et al., 2021; Kaminski et al., 2021). More interestingly, the "collateral activity" of Cas12 protein, which can trigger non-specific cleavage of ssDNA upon target recognition, has been used for diagnostic purposes. This is based on nucleic acids detection through signal amplification and allows for various readouts through the addition of functionalized reporter nucleic acids. One of the first Cas12-based detection methods reported was DETECTR (for DNA Endonuclease-TargEted CRISPR Trans Reporter) (Chen et al., 2018). In this method, collateral cleavage, upon target recognition, of short ssDNA reporters carrying a fluorophore and a quencher, leads to the separation of the two, resulting in a fluorescence signal. Other Cas12-based techniques include HOLMES (for one-HOur Low-cost Multipurpose highly Efficient System) (Li et al., 2018), which employs PCR as preamplification, and HOLMESv2 (Li et al., 2019), which uses Loop-mediated isothermal AMplification (LAMP) in a one-pot reaction. As soon as the

COVID-19 outbreak became pandemic, the immediate need for a simple, rapid, early and sensitive point-of-care testing pushed for the optimization of a number of CRISPR-detection methods (Ganbaatar and Liu, 2021). Particularly, Cas12- (Broughton et al., 2020) and Cas13-based (SHERLOCK, for RNA cleavage)(Kellner et al., 2019) have been widely used during the emergency as detection tools for SARS-CoV-2 (Ganbaatar and Liu, 2021).

1.3 CRISPR current challenges

CRISPR technologies are now considered easy to use, accurate, target-specific and multi-potential (Yang et al., 2021); however, several limitations and concerns still exist, which need to be solved. First of all, for therapeutic applications, the successful administration of CRISPR-Cas9 requires an efficient and safe delivery of both Cas9 and the gRNA to the target cell. So far, vast efforts have been made to improve this important aspect and move forward in this regard (Section 1.3.1).

Additionally, a key limitation to the use of CRISPR-Cas proteins for genome editing and other applications is the requirement of a PAM sequence to be present at the target site. The development of well-optimized and engineered Cas nucleases is significantly improving the system prospective in terms of PAM limitations (Section 1.3.2).

However, the major concern for the CRISPR technology to be used in complex eukaryotic organisms, is still represented by the off-target effects (Section 1.3.3) (Zhang et al., 2015). Several *in vivo* and *in vitro* off-target detection methods have been developed to further investigate Cas9 specificity (Sections 1.3.3.1 and 1.3.3.2) (Atkins et al., 2021) and those have also been used to evolve *in silico* off-target prediction models, such as CRISPRoff (Alkan et al., 2018) and uCRISPR (Zhang et al., 2019). In parallel, protein engineering is rapidly advancing also in the creation of new nucleases with less off-target activity (Section 1.3.4) (Naeem et al., 2020). Lastly, a great challenge is also represented by the effect of genomic and chromatine context on Cas proteins, which still requires to be elucidated (Section 1.3.5).

1.3.1 CRISPR delivery

For genetic modification of populations of cultured cells, CRISPR-Cas components can be delivered in DNA, mRNA or RNP (Cas9:gRNA complex) forms, with transfection typically achieved by nucleofection (Mali et al., 2013), electroporation (Ding et al., 2017) or lipofectamine treatment (Cong et al., 2013) (Figure 1.8). DNA-based delivery typically consists of one or multiple plasmids encoding the Cas protein and sgRNA that have to be transcribed by the hosting cells for the gene editing system to be active (Figure 1.8). On the contrary, direct delivery of the Cas mRNA allows immediate translation on entering the cytoplasm (Figure 1.8). Direct transfection of the active RNP bypasses even the translation step and offers the fastest editing kinetics (Figure 1.8) (Kim et al., 2014; Liang et al., 2015). However, another consideration is the stability of the construct and resulting intracellular concentration of Cas protein. DNA-based delivery results in a prolonged and relatively stable expression of Cas nuclease, increasing the on-target editing efficiency but with the potential increased risk of unwanted mutagenic effects (Lino et al., 2018). In contrast, RNP delivery, as a shorted lived event, allows the most direct control of the Cas concentration administered (Kim et al., 2014; Liang et al., 2015). Viral-based vectors is another option for Cas delivery both *in vitro* and *in vivo*. The editing machinery is packaged as DNA into the viral particles which are then used to infect cells, delivering the DNA cargo (Banskota et al., 2022). In this method, the limitation is the capacity of the viral capsid, which is enough to hold the RNP system, but it then leaves little room for additional markers, tags or a DNA template for repair. Alternative approaches include separate packaging of the sgRNA and Cas protein (Swiech et al., 2015), or using Adeno-Associated Vectors (AAV) based methods, which also have been successful with the additional advantage of tissue specificity by utilising specific AAV serotypes (Lisowski et al., 2015; Wu et al., 2006). Alternative viral vectors include adenovirus and lentivirus, both of which offer significantly larger capsid capacities than AAV, making them well suited to deliver all the required CRISPR components simultaneously (Maddalo et al., 2014; Platt et al., 2014). Other issues in viral delivery technologies reside in undesired immune responses reaction (Ahi et al., 2011), and unplanned mutagenic consequences due to the integration into the host genome (Lino et al., 2018). However, those can be controlled with the removal of viral genes (Brunetti-Pierri and Ng, 2017; Palmer and Ng, 2003) and the development of non-integrating lentivirus vectors

(Apolonia et al., 2007; Choi et al., 2016). Additional non-viral delivery methods include the use of lipofectamine, which is a positively charged synthetic lipid capable of forming lipid vesicles around negatively charged molecules. This is already a well-established technique in isolated cells and has also been demonstrated for *in vivo* Cas9 delivery (Zuris et al., 2014). Another common approach to deliver plasmids encoding for the CRISPR system components is the use synthetic, non-liposomal transfection reagents, such as FuGENE (Kennedy et al., 2014), which shows high efficiency and low toxicity. Gold nanoparticles, AuNPs, can also be functionalised to bind DNA molecules or electrostatically interact with proteins (Lee et al., 2017), and thus be used for RNPs and DNA delivery. These recent innovations with CRISPR-Cas gene editing and delivery systems have significantly advanced the efforts to develop human gene therapies (Lino et al., 2018; Zhang et al., 2021; Behr et al., 2021).

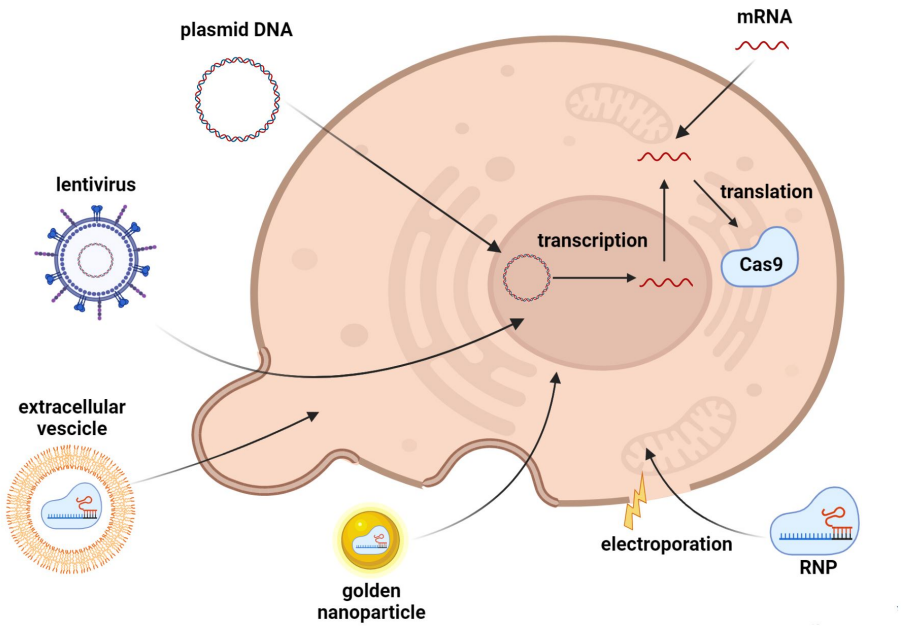


Figure 1.8: Delivery methods for CRISPR-Cas system. CRISPR-Cas components can be delivered in the forms of DNA, mRNA, or protein/RNP complex. DNA-based delivery is used to deliver one or more plasmids encoding the Cas protein and sgRNA that has to be transcribed by the hosting cells. mRNA delivery allows to skip the transcription step and directly translate for Cas protein in the cytoplasm. The RNP format enables the immediate action of the Cas nuclease when it is in the nucleus and electroporation is the most established method used for its delivery. Viral-based vector, lipofectamine or nanoparticles are also valid alternatives for both *in vivo* and *in vitro* application. Image made in Biorender.

1.3.2 Engineered Cas variants with alternative PAMs

As previously mentioned, the primary requirement for Cas protein binding is the presence of a PAM sequence. A number of Cas9 and Cas12 orthologs that recognize various PAM sequences have been discovered (Cebrian-Serrano and Davies, 2017), and this natural diversity has been harnessed to enable targeting of a greater fraction of genome sequences. PAM availability, however, remains a frequent limitation even with the use of natural Cas protein orthologs, since only a minor fraction of total PAM space can be accessed by all of the natural Cas proteins shown to function in mammalian cells (Anzalone et al., 2020). Because of these challenges, many researchers have engineered or evolved Cas9 or Cas12 variants with less restrictive PAM compatibilities. By installing specific mutation which disrupts a critical interaction between wild-type SpCas9 and the third nucleotide of its canonical PAM sequence (NGG), mutants of SpCas9 have been created to recognize NGAG, NGA and NGCG PAM sequences (Hirano et al., 2016; Kleinstiver et al., 2015, 2016a). On the other side, the T-rich PAM sequences recognized by Cas12 proteins can be used to complement the only G-rich PAMs preferred by many Cas9 orthologs. However, canonical Cas12 PAMs are often somehow restrictive, such as the TTTV PAM (where V indicates A, C or G) recognized by LbCas12a and AsCas12a (Zetsche et al., 2015). Therefore, Cas12a variants have been also engineered using structure-guided mutagenesis to recognize either TATV or TYCV PAM sequences (where Y indicates C or T) (Gao et al., 2017). Although substantial effort has been focused on SpCas9 variants, the discovery and engineering of Cas12 variants that recognize T-rich PAM sequences will likely continue to expand the targetability of Cas effectors. The set of natural and engineered Cas protein variants reported to date can collectively recognize over half of the PAM sequence space, which represents a vast progress in the field from its beginnings with SpCas9 (Anzalone et al., 2020). In general, the creation of robust Cas variants that maintain high activity while accessing certain challenging PAM sequences, remains an important opportunity for future advancement.

1.3.3 Off-target activity

Since the first application of CRISPR-system as a gene-editing tool, there was evidence that mismatches between the guide and target could be tolerated and still lead to target degradation (Garneau et al., 2010a; Deveau et al., 2008). Indeed, any guide designed to a target unique site in a whole organism genome would be predicted to have up to hundreds of potential off-target sites, and this could significantly hinder the potential for therapeutic applications. Off-target detection methodologies are fundamental since the functionality of Cas nucleases is still not fully understood and the extent of this off-target activity appears to be highly variable and dependent to many factors (e.g., the target site itself, the location and the type of mismatches, the cell type and the delivery method). Experimental observation and detection of off-target cleavage can be generally divided in two categories: biased and unbiased methods (Anton et al., 2018). Methods from the first category rely on *a priori* knowledge for the detection of site-specific mutation and sequence validation to check for mutations at expected off-target sites (e.g., sites with high sequence homology to the gRNA) (Hsu et al., 2013; Doench et al., 2016; Tsai and Joung, 2016; Aryal et al., 2018). One of the first studies used mismatched gRNA to investigate how it impacted Cas9 editing efficiencies in cells (Lin et al., 2014). Additionally, an alternative *in vitro* approach, used synthetic oligomers to generate sequence libraries that contained 1012 potential off-target sites derived from the sequence of 4 gRNAs (Pattanayak et al., 2013). These experiments provided the first evidences that mismatches in the 10-12 nt proximal to the PAM site (defined as "seed region") dramatically reduced target cleavage, but multiple mismatches at the PAM distal end of the target could be tolerated. On the other hand, unbiased methods can be used to survey the full genome for cleavage events, allowing detection of unpredicted off-target cleavage events (Koo et al., 2015; Tsai and Joung, 2016). Although biased techniques require minimal equipment, are often easier to implement and with a lower cost compared to others (Atkins et al., 2021), the need to efficiently survey the whole-genome for nuclease off-target affinity, is pushing for the development of a range of more efficient unbiased techniques.

1.3.3.1 *In vitro* off-target detection

Whole genome sequencing (WGS) is a straight forward approach for unbiased survey of the full genome for off-target nuclease activity (Figure 1.9 A) (Atkins et al., 2021; Yin et al., 2019). This relies on sequence-based evidence of nuclease activity (such as indels, inversions, translocations, and large deletions) left on genomic DNA upon endogenous repair mechanisms activation. Deep sequencing allows the identification of those repaired sites and WGS has been used to detect unpredicted Cas9 off-target cleavage in clonal cell populations and animal models (Smith et al., 2014; Veres et al., 2014). WGS is definitely advantageous since it detects the behaviour of the nucleases in a cellular environment but, at the same time, it is considered inefficient due to a low signal to noise ratio (Atkins et al., 2021). Moreover, the majority of sequence data collected during WGS represents unedited genomic DNA and this affects the depth-of-coverage for sequence locations of interest. Thus, WGS is limited by throughput, cost, and efficiency compared to whole-genome methods which incorporate target enrichment strategies (e.g., GUIDE-seq) (Atkins et al., 2021). One of the first unbiased, *in vitro* off-target cleavage detection technique was Digenome-seq (Figure 1.9 B) (Kim et al., 2015). In this assay, purified genomic DNA is incubated with Cas9, then fragmented and sequenced by NGS. After aligning reads to the known genomic sequence, sites that have been cut are identified by the presence of reads all terminating at the same location. Nuclease cleavage sites are distinctly characterized therefore by repeated detection of DSBs at the same sequence location (Kim et al., 2015). Digenome-seq achieves target enrichment by introducing a distinct signature to nuclease cleavage targets which improves the resolution of cleavage detection to single-nucleotide precision. The sensitivity of this method is limited by the error rate of NGS, and requires high read numbers (Atkins et al., 2021). However, this method can detect off-target sites with frequencies down to $\sim 0.1\%$ (Kim et al., 2015). An improved method, called SITE-Seq (Figure 1.9 C), has increased sensitivity and requires lower numbers of reads compared to Digenome-seq due to the addition of a step enriching for cut DNA by ligation of biotinylated DNA to the cut end and biotin pull-down (Cameron et al., 2017). Regardless, up to date, CIRCLE-seq is one of the most sensitive unbiased methods (Figure 1.10) (Tsai et al., 2017; Lazzarotto et al., 2018). The assay requires fragmentation of genomic DNA via sonication, end-repair, and self-ligation of fragments for intramolecular circularization. After circularization, the remaining

linear DNA is digested using a plasmid-safe ATP-dependent DNase. The resulting library of circularized fragments of genomic DNA is then digested using Cas9 and the gRNA to be profiled for off-target affinity. During Cas9 digestion, circles containing on-target and off-target sequence are linearized and prepared for NGS. CIRCLE-seq has 180,000-fold higher signal-to-noise ratio than Digenome-seq (Atkins et al., 2021), and this is due to the specific process of enrichment which ensures that only the cleavage-target sequences are prepared for deep sequencing. Recently, an updated version of the CIRCLE-seq methodology has been published (Lazzarotto et al., 2020). The modified technique is called CHANGE-seq and it utilizes a tagmentation reaction in the early steps of the protocol which reduces the labour and preparation time. Thus, compared to CIRCLE-seq, CHANGE-seq allows more rapid sample processing for higher-throughput experiments and shows a high potential for future applications (Atkins et al., 2021).

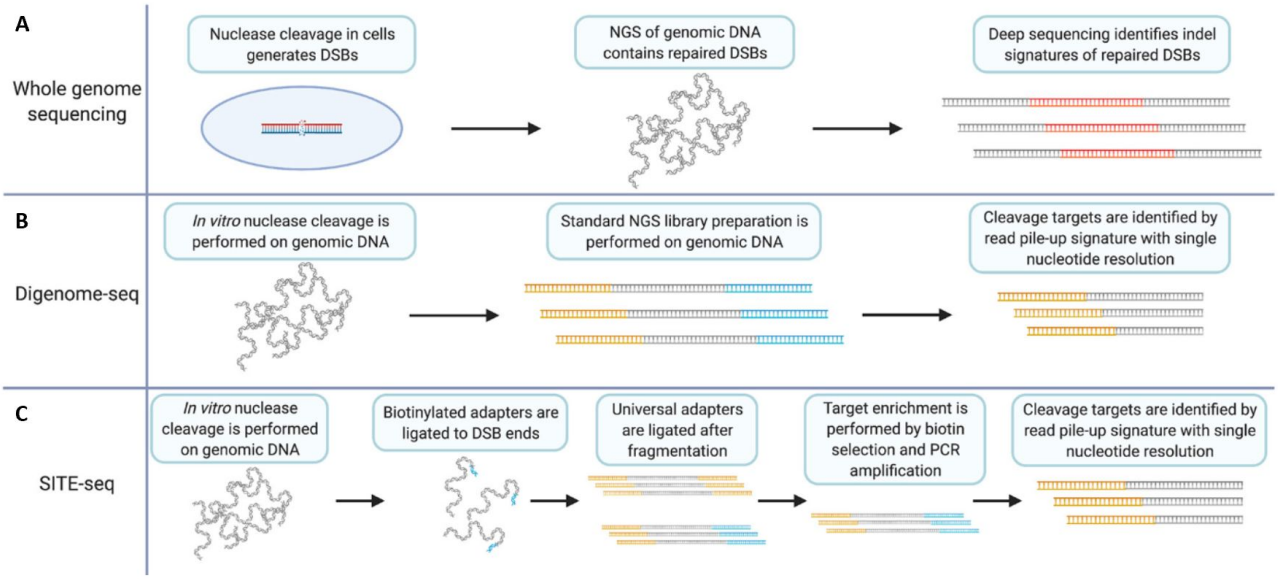


Figure 1.9: *In vitro* off-target detection methods. **(A)** CRISPR-Cas9 off-target detection via Whole Genome Sequencing (WGS) involves nuclease induction of DSBs in cells. Next, all sequence including indels and genomic rearrangements introduced during DSB repair are captured via deep sequencing of genomic DNA. **(B)** Digenome-seq starts with extraction of genomic DNA and *in vitro* CRISPR-Cas9 treatment, precluding DSB repair. Samples are then prepared by standard NGS methods where adapters (in yellow) are added. Cleavage targets are identified by a read pile-up signature with single-nucleotide resolution. **(C)** SITE-Seq also involves *in vitro* Cas9 digestion. Next, biotinylated adapters (in blue) are ligated to DSB ends followed by fragmentation and universal adapter ligation. Target enrichment is achieved by biotin selection and PCR amplification. Cleavage targets are identified by read pile-up during analysis. Image adapted from Atkins et al. 2021.

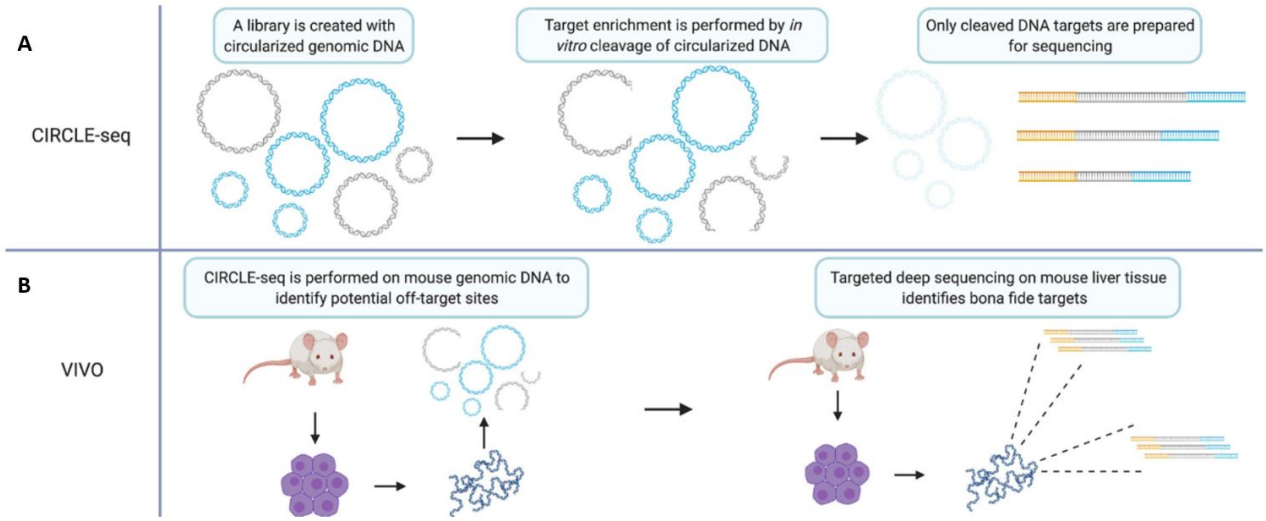


Figure 1.10: CIRCLE-Seq and VIVO detection methods. **(A)** CIRCLE-seq involves generation of a circularized *in vitro* library created from genomic DNA. Unrecognized targets remain circularized (in blue) while cleaved target sequences are linearized and can be prepared for sequencing (in gray). **(B)** VIVO starts with CIRCLE-seq on genomic DNA from mouse primary cells. The CIRCLE-seq results inform targeted deep sequencing with genomic DNA from mouse primary cells derived from CRISPR/Cas9 treated mice thereby validating bona fide off-target sites. Image adapted from Atkins et al. 2021.

1.3.3.2 *In vivo* off-target detection

Among the *in vivo* detection techniques, based on cellular assays, GUIDE-seq is a method for tagging and enriching the sequence surrounding DSBs for deep sequencing (Tsai et al., 2015). In this assay, cells are transfected with a plasmid coding for Cas9 and a gRNA and co-transfected with a blunt, double-stranded oligodeoxynucleotide (dsODN). This dsODN will be incorporated into DSBs during NHEJ repair process, thus tagging DSB sites with a short, known sequence (Tsai et al., 2015). The genomic DNA is then extracted and fragmented, and the resulting fragments undergo end-repair, dA-tailing, and ligation of a universal adapter sequence which is added to both ends of all fragments. Target enrichment is achieved by PCR amplification of fragments containing the dsODN only. By testing GUIDE-seq with 13 different guide sequences in two different cell types, HEK293 and U2OS, up to 150 off-target sites were identified with some guides, whereas others displayed no detectable off-targets (Tsai et al., 2015). This demonstrates the highly variable levels of promiscuity of different guides. Importantly, there were also significant differences in the frequencies of off-target editing between the two different cell types, clearly demonstrating that sequence alone is only one factor which affects off-target activity (Tsai et al., 2015). Additionally, HTGTS (High Throughput Genome-wide Translocation Sequencing) methodology can be used to detect and sequence translocations resulting from DSBs. Initially published to study the mechanism of translocation (Chiarle et al., 2011), this method has been adapted for detection of nuclease off-target activity and the protocol was modified to enhance the adapter-PCR target-enrichment methodology (Chiarle et al., 2011; Frock et al., 2014). The modified method is called linear amplification mediated (LAM)-HTGTS and it can detect translocations that occur due to DSB repair between the on-target site and any off-target sites. This assay was used to identify off-targets of 4 guide RNA designed for a specific target gene. One guide resulted in 33 detected translocation events, another guide in 2 events, while the other two guides displayed no translocation event. Once again, this demonstrates that off-targets are highly guide sequence dependant and even though Cas9 editing can induce translocations, the majority of DSBs are repaired locally and these would not be detected. Finally, early attempts to develop unbiased survey of Cas9 activity on a genome-wide scale are represented by studies which imply chromatin immunoprecipitation (ChIP) and subsequent deep sequencing of captured DNA fragments (ChIP-seq) (Atkins et al.,

2021). Multiple studies used ChIP-seq with catalytically inactive Cas9 (dCas9) to pull down Cas9 binding sites (Duan et al., 2014; Kuscu et al., 2014a; Wu et al., 2014a; Knight et al., 2015; O'Geen et al., 2015). However, ChIP-seq using dCas9 is limited with respect to off-target detection since it has been shown to yield abundant false positives (Kuscu et al., 2014a; Wu et al., 2014a; Knight et al., 2015; Tsai et al., 2015). DISCOVER-Seq is an additional off-target detection method which is based on ChIP-seq analysis and selectively amplifies Cas9 cleavage sites by detecting the signature of endogenous DNA repair processes (Wienert et al., 2019). DISCOVER-Seq advances the ChIP-seq method by utilizing meiotic recombination 11 homolog 1 (MRE11), a DNA repair protein which is localised to DSBs at the early stage of the repair pathway. Edited cultured cells or tissue samples are first fixed with formaldehyde to cross-link bound proteins to the DNA, the nuclei are extracted and the DNA bound to the protein is extracted and fragmented. The Mre11 bound fragments are then isolated by pull-down with specific antibodies and the DNA purified and sequenced to generate genomic profiles of Mre11 localisation (Wienert et al., 2019). This allows to determine the cut site with single base resolution. Although limited to the detection of unrepaired sites only, DISCOVER-seq has much higher sensitivity than previous approaches, such as BLESS (Breaks Labelling, Enrichment on Streptavidin and next-generation Sequencing), in which the detection of DSBs occurs upon in situ labelling with biotinylated linker DNA. Lastly, it is worth mentioning that a method has been published that is termed verification of *in vivo* off-targets (VIVO) (Figure 1.10 C) which combines CIRCLE-seq to identify off-target candidate sites with targeted deep sequencing to validate those sites *in vivo* (Akcakaya et al., 2018). This is a hybrid technique which has been applied for the detection of induced off-target in mice. However, the sensitivity is still limited by the detection limit of next-generation sequencing, and has an estimated sensitivity of 0.13% (Akcakaya et al., 2018), which will exclude potential off-targets sites with cutting frequencies lower than 0.1%.

1.3.4 Engineered high-fidelity Cas variants

While some Cas9 and Cas12 orthologs have been found to be naturally high-fidelity effectors, another major research focus has been laid on the development of engineered Cas variants that have greater discrimination for on-target sequences over off-target sequences (Hu et al., 2018; Lee et al., 2018; Vakulskas et al., 2018). To this extent, eSpCas9(1.1), SpCas9-HF1 and HypaCas9 (Chen et al., 2017b; Kleinstiver et al., 2016a; Slaymaker et al., 2016), have been generated by rational design approaches mainly focusing on mutations of positively charged residues within the REC domain that line the NTS binding groove, with the hypothesis that interrupting interactions between these residues and the negatively charged nucleic acid backbone would decrease binding affinity. It has indeed been proved that those mutated variants spend a larger fraction of time occupying the nuclease-inactive conformation when bound to off-target DNA sequences. An additional high-fidelity variant, evoCas9, consists of a quadruple mutant and it has been developed via a yeast-based assay to screen a library of SpCas9 effectors with mutations in the REC3 domain (Casini et al., 2018). However, although those Cas9 variants can discriminate better against mismatches, their enhanced specificity comes at the cost of severely reduced rates of on-target DNA cleavage. Most recently, a SuperFi-Cas9 (Bravo et al., 2022) variant has been engineered by mutating specific residues within RuvC that contact and stabilize a typical distorted configuration only present in the presence of PAM-distal mismatches and absent in on-target structures. Therefore, this SuperFi-Cas9 seems to have the ability to discriminate between on- and off-target DNA substrates without compromising DNA cleavage efficiency.

In general, Cas12a variants have been reported to naturally generate fewer off-target editing events by comparison to many of the Cas9 orthologs (Kim et al., 2016; Kleinstiver et al., 2016b). Nonetheless, high-fidelity Cas12a variants have also been engineered: an example is an high-fidelity version of an enhanced activity AsCas12a variant (enAsCas12a-HF1) which was found to be highly efficient and with a low off-target editing activity (Kleinstiver et al., 2019).

1.3.5 Genomic context and DNA topology effect

Up to date many studies have been focused on better understanding the mechanisms behind the different phases of CRISPR-Cas process, such as how Cas9 binds and cleaves sites along the genome (Kuscu et al., 2014b; Tsai et al., 2014; Wu et al., 2014b), what are the preferred repair pathways upon Cas9 cleavage (Schep et al., 2022), or trying to predict mutation outcomes of the genome editing process (Allen et al., 2018; Janssen et al., 2019; Shen et al., 2018). However, the intracellular Cas9 behaviour still remains poorly characterized and its efficiency cannot entirely be predicted. It has been shown that some genomic regions are associated with higher editing efficiencies (Knight et al., 2015; Verkuijl and Rots, 2019) and the genetic properties of the genomic target site seems to play a critical role in Cas9 activity (Daer et al., 2017; Verkuijl and Rots, 2019). Nevertheless, it is still not fully understood which factors contribute to this efficiency variability, which can be found even when targeting the same sequence in different cell types. There are numbers of studies that demonstrated that the way DNA is modified and packaged in the eukaryotic nucleus - the chromatin state - can have significant effects on Cas9 binding and may help to explain the variation in terms of efficiency (Hinz et al., 2015; Horlbeck et al., 2016; Isaac et al., 2016; Yarrington et al., 2018). It is worth pointing out that, Cas9, being prokaryotic in origin, did not evolve to deal with the complex chromatinised environment of the eukaryotic genome. In bacteria, indeed, genomic DNA generally consists of supercoiled circular molecules associated with proteins, while eukaryotic chromosomes are linear, packaged with histone octamers into nucleosomes and further organized into high-order structures. By using the artificial Widom 601 positioning sequence, which exhibits high affinity for the histone octamers, a few studies proved that stably positioned nucleosomes act as a barrier to Cas9 binding and function, with Cas9 nuclease activity being particularly inhibited within the nucleosome core but not at adjacent linker sequences (Hinz et al., 2015; Horlbeck et al., 2016). Moreover, an *in vitro* study used nucleosome assembled on the native 5S ribosomal sequence and found that Cas9 sites near those entry/exit sites are cleaved 700-fold better than the corresponding sites within the Widom 601 nucleosomes. This result may be explained from the observation that DNA breathing occurs at least 100-fold more in nucleosomes assembled on a natural sequence than on the Widom 601 (Isaac et al., 2016). Those studies explain that Cas9 gains access to nucleosomal DNA when the DNA is transiently unpeeled from the his-

tone octamer and this may clarify why sequences that are closer to the entry/exit sites are cut more readily by Cas9 than sites near the core dyad (Anderson and Widom, 2000). Similarly, Cas12a cleavage kinetics has been investigated on nucleosomal substrates (Strohkendl et al., 2021). It has been shown that the unwrapping dynamics of a 601-assebled mononucleosome regulate Cas12a cleavage efficiency. It appeared that Cas12a PAM recognition and R-loop propagation are inhibited by the nucleosome edge, with an overall binding inhibition of several orders of magnitude compared to nucleosome-free region (Strohkendl et al., 2021). With the aim to understand better the genomic context impact on the CRISPR-Cas system, a more recent study used multi-target gRNAs (mgRNAs) to direct Cas9 to a pre-determined number of well-mapped sites, with epigenetically diverse features, and investigate on the dynamics and cellular responses (Zou et al.). This system provided evidence that Cas9 binding is enhanced at chromatin-accessible regions, and Cas9 cleavage is more efficient near transcribed regions. Overall, the effect of DNA topology on Cas nucleases activity still needs to be properly interpreted. It is becoming increasingly clear that a more "open" DNA structure promotes Cas binding, meaning that every cellular processes which implies a mechanical modification of the double helix structure (including transcription, epigenetic modification, DNA repair, recombination) will somehow have an effect on Cas efficiency at a particular target site. As a consequence, this can also impact Cas off-target activity, since mismatched sequences previously less tolerated, become more easy to bind. Interestingly, it has been shown that the binding of a dCas9 at specific sites can be used to open previously inaccessible regions of chromatin, even though the underlying mechanism of this process is not fully characterized. A so called Proxy-system (Figure 1.11)(Chen et al., 2017a) has been developed to enhance the editing efficiency at adjacent region of a proxy-dCas9 site. This seems to be valid for Cas12 too, as the binding of a dCas9, also increases Cas12a access at a proximal target site (Strohkendl et al., 2021). Unravelling the variables that modify the influence of chromatin and DNA topology on CRISPR activity can be powerful to better understand and predict the CRISPR-Cas systems activity, which becomes particularly important for the off-target repercussion. In this perspective, developing new strategies which could make CRISPR-Cas systems more efficient and precise by acting on the chromatin state and DNA structure of its target, is an exciting opportunity.

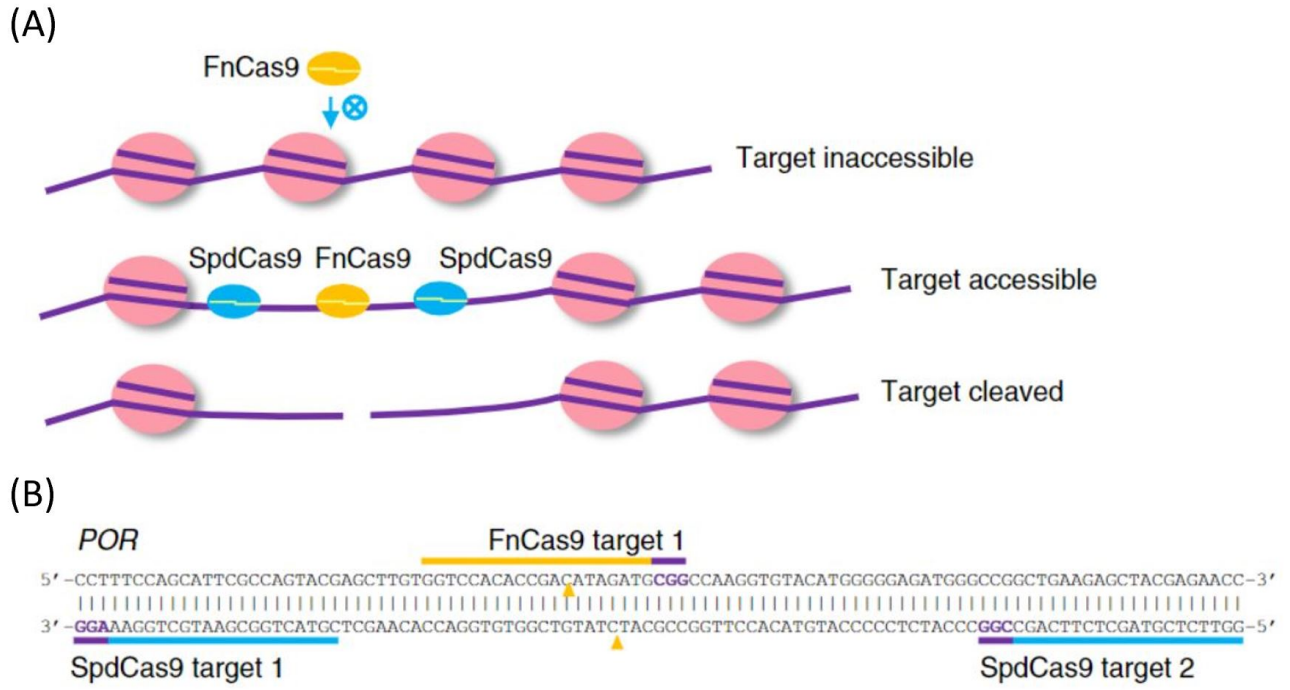


Figure 1.11: Schematic of Proxy-CRISPR system. **(A)** FnCas9 is unable to access endogenous targets in a certain chromatin configuration, but the binding of one or two dCas9 at proximal locations alters the local chromatin status and enables FnCas9 to access and cleave the otherwise inaccessible targets. **(B)** FnCas9 and SpdCas9 target sites in the exon 8 of POR gene, used to test the strategy. Image from Chen et al. 2017a.

1.4 Single Molecule studies

Single molecule studies have played an important role in the mechanistic understanding of Cas systems, particularly on SpCas9 (Cuculis and Schroeder, 2017) and, more recently, on Cas12a nucleases. Single molecule methods have been used extensively to observe many aspects of the functionality of Cas proteins (Cuculis and Schroeder, 2017), such as the on-target search and binding mechanisms, the cleavage activity, the possible conformational changes involved in RNP assembly and DNA interaction, and the off-target discrimination.

These methods are particularly attractive and powerful for the study of complex biological systems for several reasons (Deniz et al., 2008). First, the molecular properties are measured one molecule at a time, meaning that their distributions are more directly measured, in contrast with conventional ensemble experiments (Deniz et al., 2008). Thus, single molecule-experiments allow for the detection of rare species or states, which would be averaged in an ensemble (Deniz et al., 2008). Moreover, for single molecule analysis, the dynamics of systems are measured under equilibrium conditions and this is advantageous for those dynamics in systems that cannot be easily synchronized (Petrov et al., 2012). An example is the motion of DNA enzymes along their tracks, which can be complex and impracticable to synchronize for significant distances. Furthermore, by using single-molecule methods, it is possible to directly measure the kinetic rate constants and the relations between different states of a system with multiple discrete states (Schmid and Hugel, 2017). By operating at low concentrations or with immobilized molecules, those methods exhibit an exceptionally high sensitivity, allowing the properties of the monomeric species to be monitored at equilibrium (Deniz et al., 2008). Additionally, single-molecule manipulation permits the direct measurement of molecular forces, as well as the molecular structural and functional responses to mechanical manipulation (Neuman and Nagy, 2008).

1.4.1 Single Molecule methods

Broadly speaking, single-molecule techniques can be divided into two main categories: imaging-based approaches and force-based approaches (Gruszka, 2021). Within this classification, imaging-based methods are typically carried out via fluorescence microscopy, whereas force-based approaches generally rely on optical tweezers, magnetic tweezers, or atomic force microscopy (AFM) for precise mechanical control of single biomolecules. Single-molecule fluorescence microscopy (SMFM) is used to detect single fluorescent probes attached to biomolecules (Hao et al., 2020), and allows to localise them with spatial precision (∼300 nm) and to observe molecular events at relevant length scales for proteins, DNA, or RNA (25–100 nm) (Thompson et al., 2002). An additional method is single molecule Forster Resonance Energy Transfer (smFRET), which is considered a powerful "molecular ruler" since it allows to direct observe small-scale changes (1–10-nm) in intermolecular or intramolecular distances (Roy et al., 2008). On the other hand, single-molecule force-based measurements offer the possibility to apply and/or measure subpiconewton to nanonewton forces, which became relevant for the study and manipulation of individual biomolecules (Neuman and Nagy, 2008). Among this methods, AFM is a scanning probe method which provides force-based measurements or manipulation of single biomolecules as well as single-molecule imaging (Engel and Müller, 2000; Hansma et al., 2012). Furthermore, optical tweezers and magnetic tweezers represent alternative single molecule technologies which are based on the principle of particle trapping using an external force field. In the optical tweezers set-up, a high-powered laser beam is focused through a high–numerical aperture objective lens, generating a tight focus and large optical gradient (Figure 1.12) (Moffitt et al., 2008). In such a manner, dielectric particles positioned near the focus of the beam, experience a restoring force toward the trap center, which enables stable trapping, and this is used to manipulate beads (Figure 1.12) (Heller et al., 2014; Bustamante et al., 2021). Optical tweezers experiments can be performed a single- or dual-trap mode (Bustamante et al., 2021). In single-trap measurements, one end of the molecule is attached to the trapped bead and the other end is bound to the chamber surface or to a second bead held by a micropipette. In dual-trap systems, the molecule is instead tethered between two optically trapped beads (Bustamante et al., 2021). Magnetic tweezers operate by applying an external magnetic field to manipulate magnetic beads. This can also be used to generate torque on

trapped particles, by rotation of two permanent magnets, which is useful for inducing supercoiling in DNA. Rotor Bead Tracking (RBT) is a method in which a magnetic bead is used to apply forces (in the pN-scale) to stretch DNA, while a secondary bead bound to the side of the molecule (the "rotor bead") reads out the molecular twist and the extension (Bryant et al., 2012). RBT can be combined with single-molecule fluorescence and this allows to monitor conformational dynamics of nucleoprotein complexes in real time under controlled force and torque on the nucleic acid substrate.

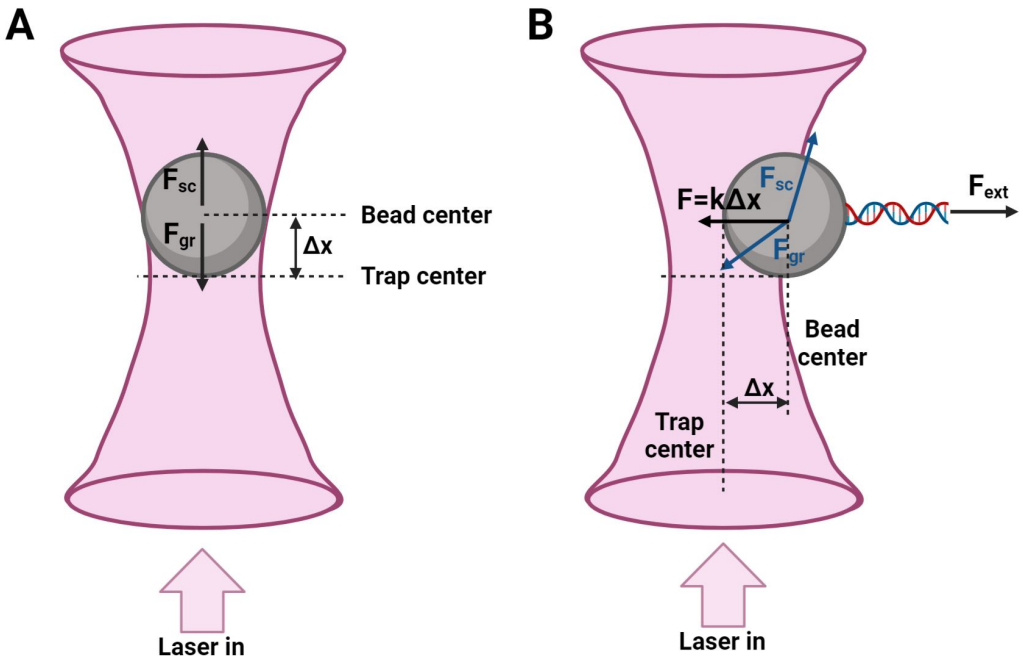


Figure 1.12: Principles of optical tweezers. The schematic shows the forces acting on optically trapped beads. **(A)** When a bead is trapped in a single-beam optical trap (pink cone) two main forces are acting on it: the scattering force, F_{sc}, and the gradient force, F_{gr}. F_{sc} comes from the photons hitting the particle and being scattered, and it acts in the direction of light propagation. F_{gr} results from photons passing through the bead and it acts along the electric field gradient toward the center of the laser focus. For stable trapping, the gradient force must be greater than the scattering force to prevent the latter pushing the particle out of the focus. In the schematic, the two forces cancel each other, which leads to a stable position that is displaced by Δz from the trap center in the axial direction. **(B)** When an external force, F_{ext}, acts on the bead (such as by using a tethered DNA molecule) this causes a lateral displacement, Δx. The external force is opposed by the combination of gradient and scattering forces, leading to a restoring force that can be described by Hooke's law, $F = k\ \Delta x$, where k represents the effective stiffness of the optical trap. Image adapted in BioRender from Heller et al. 2014.

1.4.2 Cas9 studies

The first Cas9 single molecule study used DNA curtains with Total Internal Reflection Fluorescence (TIRF) microscopy to explore Cas9 target search and binding mechanisms (Sternberg et al., 2014). In this approach, a microscope slide was prepared with chromium "barriers" and then coated with a lipid bilayer (Figure 1.13 A). Phage-λ DNA can be tethered to the lipid bilayer by biotin-streptavidin interactions and, by applying flow across the slide, the tethered DNA can diffuse through the bilayer and thus be trapped by the barriers, generating DNA curtains (Fazio et al., 2008). FLAG-tagged Cas9 was complexed with sgRNAs targeting specific sites on the DNA and then flowed over the DNA curtains, allowing for the visualization of DNA:RNP interactions by TIRF microscopy (Figure 1.13 B) (Sternberg et al., 2014). In this study, it was possible to observe an efficient on-target binding when an on-target guide was used to assemble the RNP. Stable interactions were also observed at off-target sites, even though at a much lower frequency, and appeared to correlate with local PAM site density (Figure 1.13 C) (Sternberg et al., 2014). Similar experiments were performed in the absence of flow in order to study the target search mechanism. Those experiments suggested a 3D collision mechanism of target search, since no evidence of 1D diffusion was observed (Sternberg et al., 2014).
In addition, smFRET TIRF approaches have been used to directly measure Cas9 binding and dissociation kinetics (Singh et al., 2016). In this experiment, the target DNA was labelled with Cy3 close to the PAM site and immobilised on microscope slide, while the RNP complex, labelled with Cy5 close to the 3' end of the crRNA, was flowed over the DNA (Figure 1.14 A). Cas9 binding was observed by FRET between the labelled crRNA and target DNA (Figure 1.14 B) and detailed kinetic information were calculated by analysing the dwell times in the presence and in the absence of mismatches (Singh et al., 2016). When PAM proximal mismatches were present, even if the binding was still observed, a much rapid dissociation was detected compared to a perfectly matched sequences. Similar experiments have been also performed by using an array of DNA templates with engineered mismatches (Singh et al., 2016). Results showed that PAM-distal mismatches are significantly better tolerated than PAM proximal mismatches and that an increase in Cas9 binding life-times is detected when >7 base pairs of heteroduplex (PAM-proximal target match) formation were permitted. SmFRET TIRF assay has also been used to demonstrate Cas9 directional R-loop formation, which is observed as an increase of the

FRET intensity as the RNA anneals with the DNA from the PAM proximal to PAM distal end of the target (Lim et al., 2016). Moreover, those studies showed that the presence of mismatches prevents full R-loop formation proving the role of R-loop formation in stable binding of the Cas9 complex.

Furthermore, FRET experiments were used to detect the movement of HNH domain, as an essential step to reach the cleavage competent conformation, and to analyse differences in the conformational change upon target binding of dCas9 versus Cas9, that may be attributed to additional conformational changes during cleavage (Dagdas et al., 2017; Osuka et al., 2018; Sternberg et al., 2015). It has been proved that this HNH conformation change can be observed upon binding to an on-target site but not in the presence of off-target. These studies suggested the presence of the so called "conformational checkpoint", involving the HNH domain, which provides additional discrimination of off-target sites, preventing cleavage even in the presence of mismatches that accommodate stable R-loop formation.

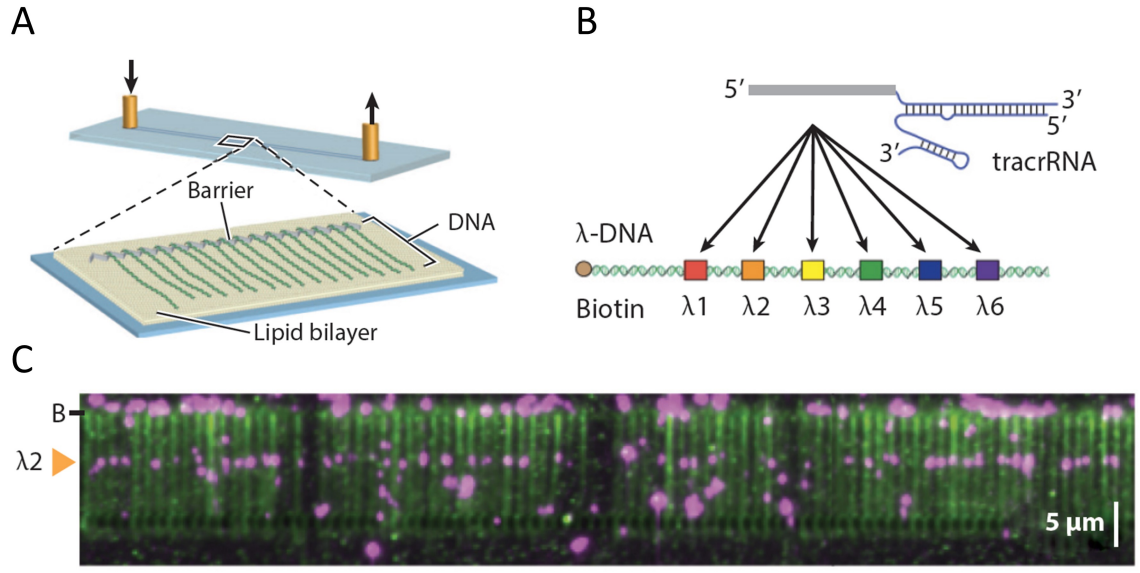


Figure 1.13: DNA-curtains assay for investigation of Cas9 binding and cleavage mechanisms. **(A)** Schematic of DNA curtains microscope slide. **(B)** Several sgRNAs were designed to target different specific site on λ-DNA. FLAG-tagged Cas9 and anti-FLAG-functionalized quantum dots were used to directly visualize Cas9-DNA interaction. **(C)** Single-molecule images of Cas9 binding to the DNA-curtains at the $\lambda 2$ site. Off-target binding sites are also detectable at a lower frequency. Image from Sternberg et al. 2014.

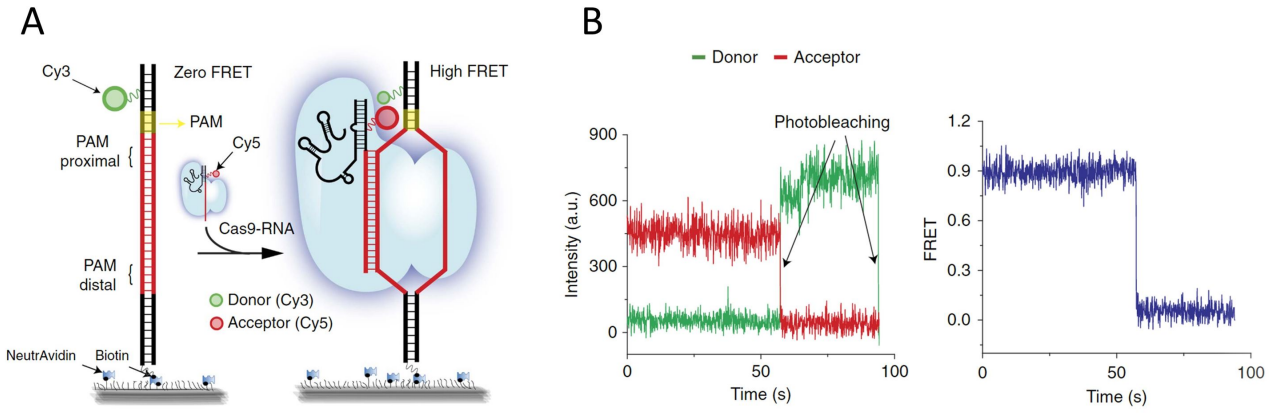


Figure 1.14: smFRET assay to study Cas9 binding mechanism. **(A)** Schematic of single-molecule FRET assay. Custom-designed DNA templates were labeled with a FRET donor (Cy3) and immobilized on the microscope slide via neutravidin-biotin interaction (cognate DNA segment is shown in red, PAM sequence is shown in yellow). The sgRNA was labelled with a FRET acceptor (Cy5) and complexed with Cas9 to form the RNP. **(B)** A representative smFRET time trajectory showing the moment of Cas9 binding to the DNA which results in high-FRET signal. Image from Singh et al. 2016.

Interestingly, a single-molecule-based approach has been used to demonstrate how Cas9 specificity can be tuned and improved by modulating the hybridisation energy of guide and target DNA, through incorporation of DNA bases into the RNA guide molecules (Figure 1.15) (Rueda et al., 2017). In this study, an in bulk-cleavage assay first demonstrated that Cas9 can form a complex with a fully DNA guide; however, this complex is not active for cleavage. By rational design, 7 bases were identified to make important RNA specific contacts with the Cas9 protein, and by restoring these sites to RNA in the crDNA guide it was possible to recover the on-target cleavage activity (Rueda et al., 2017). Interestingly, the hybrid guides with DNA bases showed reduced off-target activity while retaining on-target activity *in vitro*, consistent with the excess energy hypothesis for rational design of high-fidelity Cas9 complexes. Single molecule FRET was then used to study the R-loop formation of those hybrid guides (Figure 1.15) and showed that, although the binding of the hybrid guided complex was dramatically reduced, molecules were observed reaching the high FRET state, consistent with their ability to cleave target DNA *in vitro*. However, the emergency of a mid-FRET state population (Figure 1.15) suggested that the presence of DNA bases in the guide destabilises the R-loop formation, and this would explain the reduced binding affinity, but also the improved specificity.

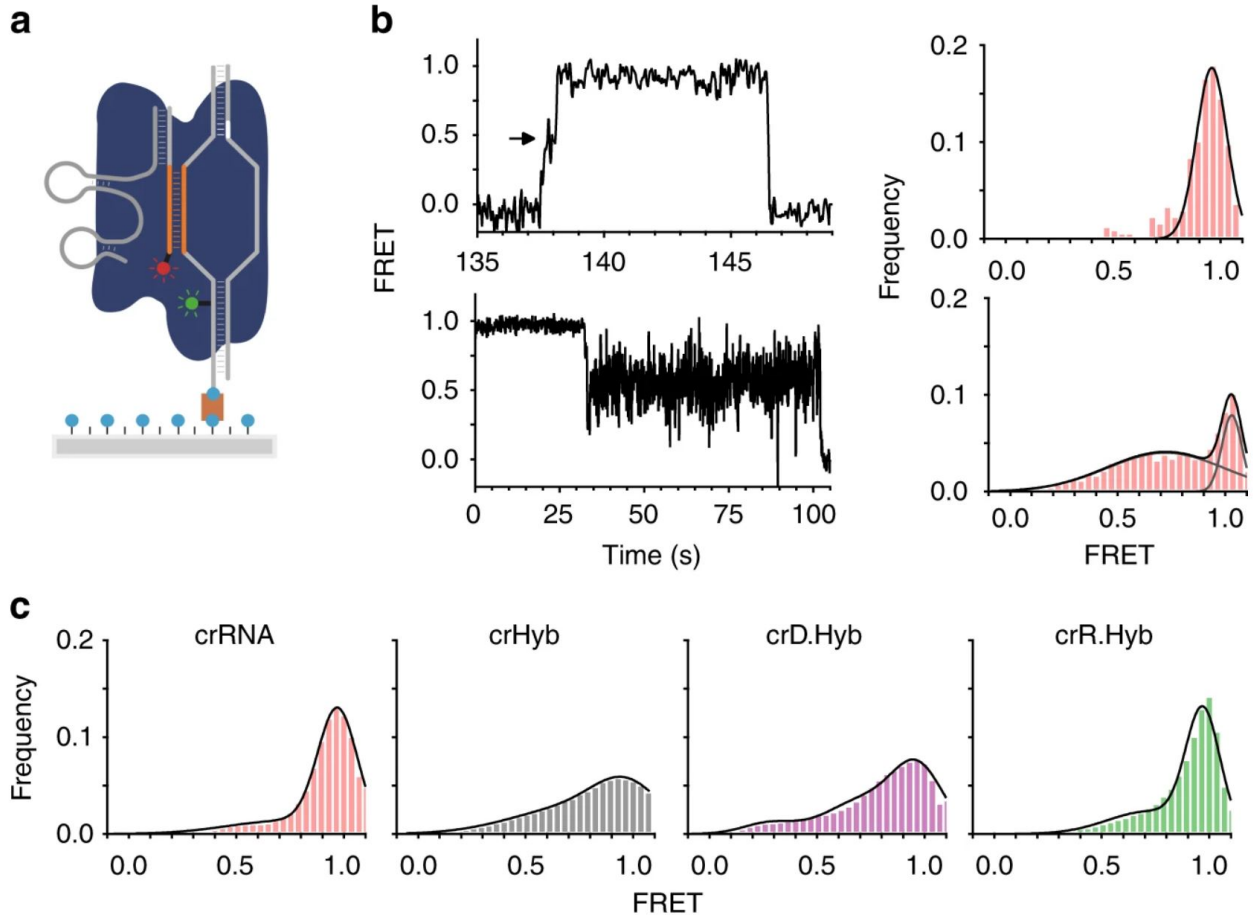


Figure 1.15: Conformational dynamics of target-bound dCas9 complexes with hybrid guides. **(A)** Single-molecule FRET experimental design. Surface-immobilised Cy3-labelled (green) DNA containing the target site (orange) and PAM site (white) binds Cy5-labelled (red) dCas9 complex (blue) to yield a FRET signal. **(B)** Representative single-molecule FRET trajectories show dCas9–crRNA complex binding in heteroduplex conformation as an increase to high FRET (top) through a transient intermediate (arrow). Bottom: Representative single-molecule FRET trajectory shows an irreversible conformational change into a dynamic intermediate FRET state. Right: Corresponding time-binned FRET histograms, and Gaussian fits (black curves). **(C)** Time-binned FRET histograms for each guide: crRNA (red); different hybrid crRNAs (grey, magenta, and green). Image from Rueda et al. 2017.

Finally, optical and magnetic tweezers, with their ability to finely apply forces and also precisely twist macromolecules, have been used to investigate how force and supercoiling on DNA affect Cas9 R-loop formation, off-target binding and cleavage activity. Optical tweezers measurements showed for the first time that Cas9 binding and cleavage specificity is affected by high stretching forces (Newton et al., 2019). The mechanical distortion of the DNA-duplex under high stretching forces, induces stable off-target binding at sequence-specific sites across the DNA (Figure 1.16) (Newton et al., 2019). This result becomes particularly important if related to the physiological cellular environment. Indeed, as previously mentioned, during many fundamental processes (DNA transcription, replication, repair or chromatin remodelling), cellular DNA is constantly unwound and single-stranded "bubbles" can be transiently exposed, meaning that Cas9 complexes binding at off-target sites can be favoured (Newton et al., 2019). RBT has been used to study dCas9 R-loop formation and dissociation dynamics under torque, both at the on-target and in the presence of sequence mismatches (Figure 1.17) (Ivanov et al., 2020). Those studies are again notably relevant since local changes in superhelical density arise dynamically from fundamental cellular processes as described above. Studies in eukaryotes have identified gene- and locus-specific variations in supercoiling, with regions of high negative super helicity strongly correlated with promoter sites (Corless and Gilbert, 2017; Naughton et al., 2013). Therefore, the response of Cas9 to torsional strain is fundamental to understanding its function in cellular contexts. Taken together, many recent single-molecule studies have provided new and fundamental insights into the DNA-Cas9 interaction mechanism, strongly contributing to a better understanding of the kinetics of this prominent gene-editing system.

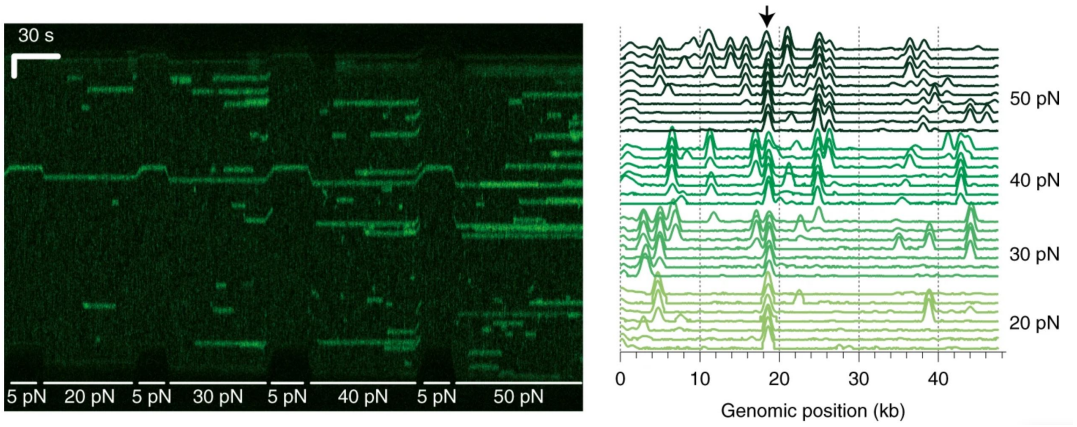


Figure 1.16: Optical tweezers experiment demonstrated force-induced Cas9 off-target activity. Left: Kymograph of force-stretched λ-DNA in the presence of 10 nM λ2-crRNA-dCas9 (green) between 5 and 50 pN, with off-target binding occuring at >20 pN. Right: Time-binned intensity histogram shows force dependence and genomic off-target binding locations (arrow marks on-target site). Image from Newton et al. 2019.

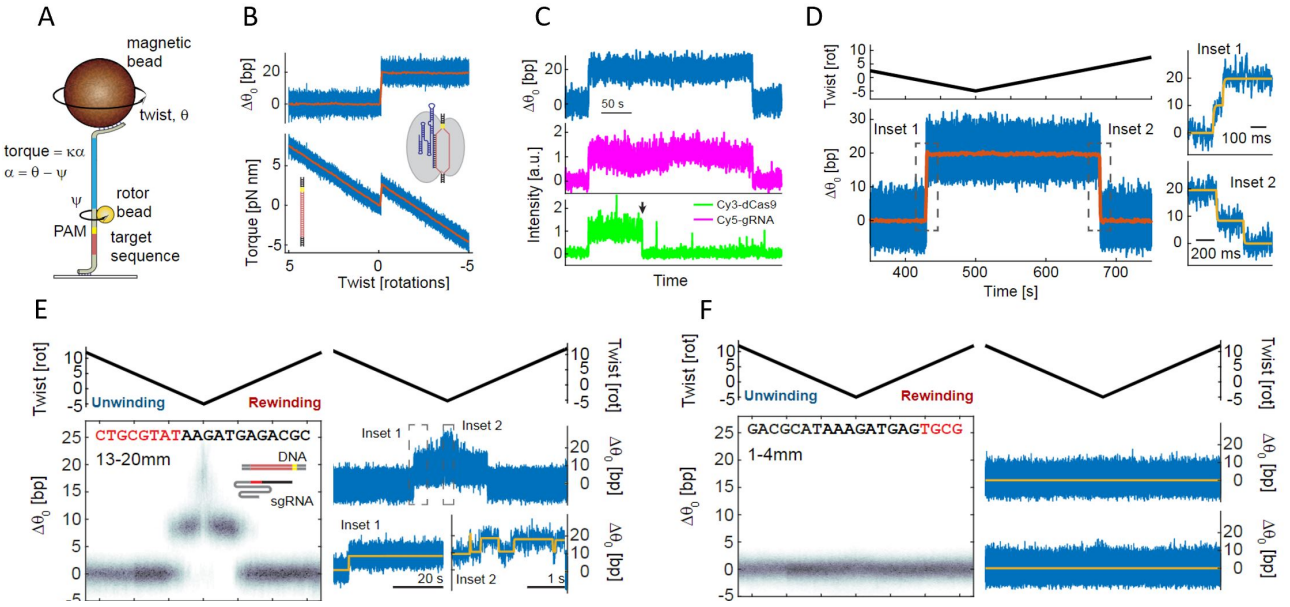


Figure 1.17: RBT assay for dCas9 R-loop detection. **(A)** A magnetic bead is used to apply tension and modulate the DNA total twist, θ. The applied torque is measured via the angle ψ of a small rotor bead that reports on the deflection, α, of the blue DNA segment. **(B)** A jump in the torque-twist curve indicates the R-loop formation, as a consequence of the change in the equilibrium twist, $\Delta\theta_0$. **(C)** Fluorescent labelled dCas9 and gRNA allow to visualize the arrival of the dCas9-gRNA complex and R-loop formation (black arrow indicates Cy3-dye bleaching). **(D)** R-loop formation with a fully matching guide shows the presence of a transient intermediate on the way to complete hybridization (Inset 1) or dissociation (Inset 2). **(E)** 8 PAM-distal mismatched nucleotides (13-20mm) prevents full R-loop formation on relaxed DNA and leads to stable dwells at the intermediate. **(F)** Mutations of the first four nucleotides of the guide (1-4mm) leads to no R-loop formation, meaning that the closed state remains favoured over mechanical distortion. Image from Ivanov et al. 2020.

1.4.3 Cas12a studies

SmFRET experiments (Singh et al., 2018a) proved that Cas9 and Cas12a share a two-step mechanism of sampling for PAM sequence followed by directional RNA-DNA heteroduplex extension. This suggests that there is a general target identification mechanism among these CRISPR systems. Single-molecule experiments showed that ultra-stable binding of Cas12a requires the same extent of sequence match as target cleavage (17-bp PAM-proximal matches) (Jeon et al., 2018). This contrasts with Cas9 system, which requires only 9 bp proximal matches for ultra-stable binding and 16 bp for cleavage. In this study, Cas12a cleavage rate has been found to be reduced with increasing PAM-distal mismatches, even when the mismatches do not affect stable binding, suggesting that shorter RNA-DNA heteroduplexes result in slower conformational changes required for cleavage activation (Singh et al., 2018a). Moreover, perturbing DNA rigidity by introducing a nick near the PAM, slowed down the cleavage rate, possibly impeding the Cas12a-induced DNA bending, necessary for DNA cutting (Singh et al., 2018a). TIRF single-molecule assay additionally showed that target searching on dsDNA molecules is probably carried out through 1D hopping dynamic (Jeon et al., 2018), and FRET analysis have been used for direct, real time, observation of Cas12a conformational dynamics at different stages of the cleavage reaction (Figure 1.18) (Stella et al., 2018) The same study also elucidated Cas12a "collateral activity" mechanism, proving the crucial role of the lid region (Stella et al., 2018). Magnetic tweezers experiments on Cas12a also observed a slower rate constant for TS cleavage that could be due to a slow conformational transition that is required to engage the TS in the RuvC active site (van Aelst et al., 2019). Recently, RBT approach has been also applied to Cas12a system to characterize differences in dynamics, mechanics, intermediates, and mismatch sensitivities between Cas9 and Cas12a (Aris et al., 2022).

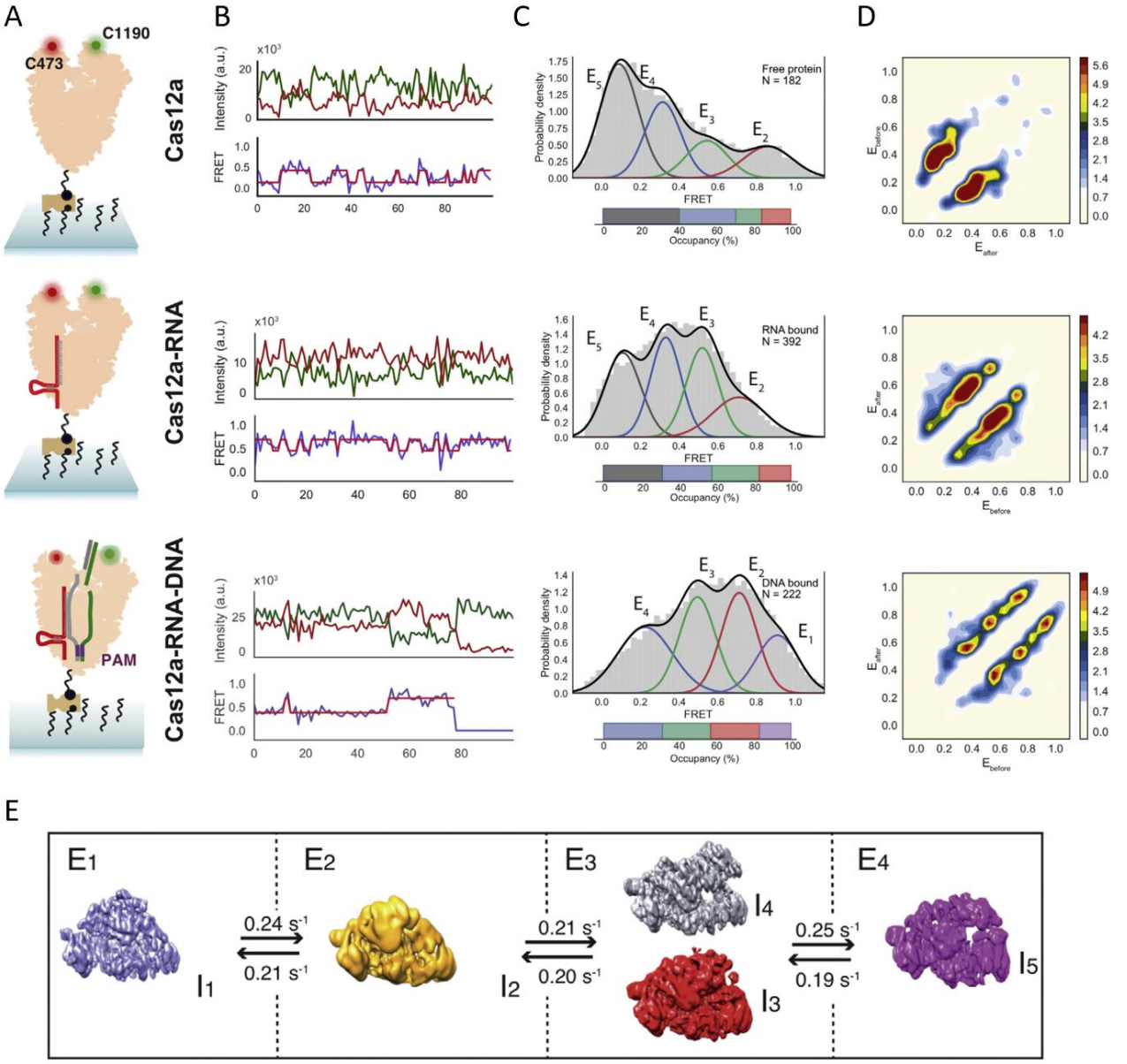


Figure 1.18: SmFRET assay to investigate Cas12a conformational dynamics at different stages of the cleavage reaction. **(A)** Schematic of the smFRET set-up showing fluorophore labeling sites. **(B)** Representative smFRET time trajectories showing reversible transitions between conformational states (Top: donor and acceptor intensity. Bottom: FRET.). **(C)** Distribution of FRET efficiency for Cas12a only, Cas12a-crRNA, and Cas12a-crRNA-DNA-bound form. Each histogram displays 4 distinct FRET states (E2-E5) and an additional state is detected for the ternary complex bound to the DNA (E1). The bar chart below each histogram represents the population of the conformations. **(D)** Transition Density Plots (TDPs) showing the frequency of transitions between states. Initial and final FRET values for each transition are accumulated into 2D histograms for apo, crRNA, and DNA-bound states. **(E)** Schematic of sequential transitions between E1, E2, E3, and E4 FRET states with forward and reverse rates. The high-FRET state, E1, is sampled exclusively by the ternary complex and it can be assigned to the more compact structure (I1 conformation). The medium-FRET states, E2 and E3, correspond to an incomplete crRNA-DNA hybrid. Finally, the low-FRET state, E4, can be attributed to the most open conformation (I5) observed. Cas12a only and the binary complex also sample a very extended conformation (E5 FRET state) only found in the absence of the crRNA and target DNA. Image from Stella et al. 2018.

1.5 Project Aims

There are many important opportunities and also crucial challenges for the CRISPR-Cas technologies to be optimized as a successful gene-editing tool. As mentioned in the Introduction, the advantages of this system over the previous utilized tools, together with the incredibly fast improvement driven by worldwide interest, have pushed the development of CRISPR technologies with methods easy to use, more precise, and multi-potential. However, several limitations and concerns still exist, and need to be addressed. In this thesis, I present my work on the CRISPR-Cas systems, done with the aim of deepening the existing knowledge on Cas9 activity on DNA and unravelling the mechanistic aspects of new Cas variants.

Within chapter 3 and 4, I describe two strategies developed to study the effect of DNA topology in the cellular context on Cas9 activity. First, I focus on showing that distorted DNA promotes Cas9 sequence-specific off-target binding in cells. To this extent, by using a cellular assay which utilizes an adapted Proxy-CRISPR, I investigate the effect of local conformational changes of the DNA structure (Newton et al., 2022. Manuscript submitted). Next, I have designed a method, based on CRISPRa, to create a promoter-localised distortion and chromatin "opening" of a number of target genes, to study if induced gene transcription causes enhancement of Cas9 on-target editing efficiency on previously inaccessible genes.

In Chapter 5 and 6, to extend the mechanistic knowledge of Cas9 kinetics to other CRISPR-systems, I study Cas12a (Losito et al., 2021) and AZ-Cas9 dynamics on dsDNA, at single-molecule level. In this part of the work, I gain new information on the on-target search mechanism, the off-target discrimination, the cleavage kinetics and high-fidelity properties of these nucleases' variants.

Bibliography

M. Adli. The CRISPR tool kit for genome editing and beyond. *Nature Communications 2018 9:1*, 9(1):1–13, 5 2018. ISSN 2041-1723. doi: 10.1038/s41467-018-04252-2. URL https://www.nature.com/articles/s41467-018-04252-2.

Y. S. Ahi, D. S. Bangari, and S. K. Mittal. Adenoviral Vector Immunity: Its Implications and circumvention strategies. *Current gene therapy*, 11(4): 307, 8 2011. ISSN 1875-5631. doi: 10.2174/156652311796150372. URL /pmc/articles/PMC4009923//pmc/articles/PMC4009923/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4009923/.

P. Akcakaya, M. L. Bobbin, J. A. Guo, J. Malagon-Lopez, K. Clement, S. P. Garcia, M. D. Fellows, M. J. Porritt, M. A. Firth, A. Carreras, T. Baccega, F. Seeliger, M. Bjursell, S. Q. Tsai, N. T. Nguyen, R. Nitsch, L. M. Mayr, L. Pinello, M. Bohlooly-Y, M. J. Aryee, M. Maresca, and J. K. Joung. In vivo CRISPR editing with no detectable genome-wide off-target mutations. *Nature*, 561(7723):416, 9 2018. ISSN 14764687. doi: 10.1038/S41586-018-0500-9. URL /pmc/articles/PMC6194229/https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6194229/.

F. Alkan, A. Wenzel, C. Anthon, J. H. Havgaard, and J. Gorodkin. CRISPR-Cas9 off-targeting assessment with nucleic acid duplex energy parameters. *Genome Biology*, 19 (1):1–13, 10 2018. ISSN 1474760X. doi: 10.1186/S13059-018-1534-X/FIGURES/6. URL https://genomebiology.biomedcentral.com/articles/10.1186/s13059-018-1534-x.

F. Allen, L. Crepaldi, C. Alsinet, A. J. Strong, V. Kleshchevnikov, P. De Angeli, P. Páleníková, A. Khodak, V. Kiselev, M. Kosicki, A. R. Bassett, H. Harding, Y. Galanty, F. Muñoz-Martínez, E. Metzakopian, S. P. Jackson, and L. Parts. Predicting the mutations generated by repair of Cas9-induced double-strand breaks. *Nature Biotechnology 2018 37:1*, 37(1): 64–72, 11 2018. ISSN 1546-1696. doi: 10.1038/nbt.4317. URL https://www.nature.com/articles/nbt.4317.

C. Anders, O. Niewoehner, A. Duerst, and M. Jinek. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature*, 513

(7519):569, 9 2014. ISSN 14764687. doi: 10.1038/NATURE13579. URL /pmc/articles/PMC4176945//pmc/articles/PMC4176945/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4176945/.

J. D. Anderson and J. Widom. Sequence and position-dependence of the equilibrium accessibility of nucleosomal DNA target sites. *Journal of Molecular Biology*, 296(4):979–987, 3 2000. ISSN 0022-2836. doi: 10.1006/JMBI.2000.3531.

T. Anton, E. Karg, and S. Bultmann. Applications of the CRISPR/Cas system beyond gene editing. *Biology Methods and Protocols*, 3(1):1–10, 2018. ISSN 23968923. doi: 10.1093/biomethods/bpy002.

A. V. Anzalone, P. B. Randolph, J. R. Davis, A. A. Sousa, L. W. Koblan, J. M. Levy, P. J. Chen, C. Wilson, G. A. Newby, A. Raguram, and D. R. Liu. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature 2019 576:7785*, 576(7785):149–157, 10 2019. ISSN 1476-4687. doi: 10.1038/s41586-019-1711-4. URL https://www.nature.com/articles/s41586-019-1711-4.

A. V. Anzalone, L. W. Koblan, and D. R. Liu. Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors. *Nature biotechnology*, 38(7):824–844, 7 2020. ISSN 1546-1696. doi: 10.1038/S41587-020-0561-9. URL https://pubmed.ncbi.nlm.nih.gov/32572269/.

L. Apolonia, S. N. Waddington, C. Fernandes, N. J. Ward, G. Bouma, M. P. Blundell, A. J. Thrasher, M. K. Collins, and N. J. Philpott. Stable Gene Transfer to Muscle Using Non-integrating Lentiviral Vectors. *Molecular Therapy*, 15(11):1947–1954, 11 2007. ISSN 1525-0016. doi: 10.1038/SJ.MT.6300281.

K. D. Aris, J. Cofsky, A. V. Wright, I. E. Ivanov, A. Balaji, J. A. Doudna, and Z. Bryant. Single-molecule analysis of DNA interrogation by Cas9 and Cas12a on supercoiled DNA. *Biophysical Journal*, 121(3):288a, 2 2022. ISSN 0006-3495. doi: 10.1016/J.BPJ.2021.11.1308.

N. K. Aryal, A. R. Wasylishen, and G. Lozano. CRISPR/Cas9 can mediate high-efficiency off-target mutations in mice in vivo. *Cell Death & Disease 2018 9:11*, 9(11):1–3, 10 2018. ISSN

2041-4889. doi: 10.1038/s41419-018-1146-0. URL https://www.nature.com/articles/s41419-018-1146-0.

A. Atkins, C.-H. Chung, A. G. Allen, W. Dampier, T. E. Gurrola, I. K. Sariyer, M. R. Nonnemacher, and B. Wigdahl. Off-Target Analysis in Gene Editing and Applications for Clinical Translation of CRISPR/Cas9 in HIV-1 Therapy. *Frontiers in Genome Editing*, 0:16, 8 2021. ISSN 2673-3439. doi: 10.3389/FGEED.2021.673022.

S. Banskota, A. Raguram, S. Suh, S. W. Du, J. R. Davis, E. H. Choi, X. Wang, S. C. Nielsen, G. A. Newby, P. B. Randolph, M. J. Osborn, K. Musunuru, K. Palczewski, and D. R. Liu. Engineered virus-like particles for efficient in vivo delivery of therapeutic proteins. *Cell*, 185 (2):250–265, 1 2022. ISSN 0092-8674. doi: 10.1016/J.CELL.2021.12.021.

R. Barrangou and L. A. Marraffini. CRISPR-Cas systems: prokaryotes upgrade to adaptive immunity. *Molecular cell*, 54(2):234, 4 2014. ISSN 10974164. doi: 10.1016/J.MOLCEL.2014.03.011. URL /pmc/articles/PMC4025954//pmc/articles/PMC4025954/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4025954/.

R. Barrangou, C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D. A. Romero, and P. Horvath. CRISPR provides acquired resistance against viruses in prokaryotes. *Science (New York, N.Y.)*, 315(5819):1709–1712, 3 2007. ISSN 1095-9203. doi: 10.1126/SCIENCE.1138140. URL https://pubmed.ncbi.nlm.nih.gov/17379808/.

F. Baylis and M. McLeod. First-in-human Phase 1 CRISPR Gene Editing Cancer Trials:Are We Ready? *Current Gene Therapy*, 17(4):309, 11 2017. ISSN 15665232. doi: 10.2174/1566523217666171121165935. URL /pmc/articles/PMC5769084//pmc/articles/PMC5769084/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC5769084/.

J. Behler and W. R. Hess. Approaches to study CRISPR RNA biogenesis and the key players involved. *Methods*, 172:12–26, 2 2020. ISSN 1046-2023. doi: 10.1016/J.YMETH.2019.07.015.

M. Behr, J. Zhou, B. Xu, and H. Zhang. In vivo delivery of CRISPR-Cas9 therapeutics: Progress and challenges. *Acta pharmaceutica Sinica. B*, 11(8):2150–2171, 8 2021. ISSN

2211-3835. doi: 10.1016/J.APSB.2021.05.020. URL https://pubmed.ncbi.nlm.nih.gov/34522582/.

A. Bolotin, B. Quinquis, A. Sorokin, and S. Dusko Ehrlich. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*, 151(8):2551–2561, 2005. ISSN 13500872. doi: 10.1099/mic.0.28048-0.

G. D. Bowman and M. G. Poirier. Post-translational modifications of histones that influence nucleosome dynamics. *Chemical Reviews*, 115(6):2274–2295, 3 2015. ISSN 15206890. doi: 10.1021/CR500350X/ASSET/IMAGES/LARGE/CR-2014-00350X{_}0015.JPEG. URL https://pubs.acs.org/doi/full/10.1021/cr500350x.

S. Bradde, A. Nourmohammad, S. Goyal, and V. Balasubramanian. The size of the immune repertoire of bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 117(10):5144–5151, 3 2020. ISSN 10916490. doi: 10.1073/PNAS.1903666117/SUPPL{_}FILE/PNAS.1903666117.SAPP.PDF. URL www.pnas.org/cgi/doi/10.1073/pnas.1903666117.

J. P. Bravo, M. S. Liu, G. N. Hibshman, T. L. Dangerfield, K. Jung, R. S. McCool, K. A. Johnson, and D. W. Taylor. Structural basis for mismatch surveillance by CRISPR–Cas9. *Nature 2022 603:7900*, 603(7900):343–347, 3 2022. ISSN 1476-4687. doi: 10.1038/s41586-022-04470-1. URL https://www.nature.com/articles/s41586-022-04470-1.

C. Brokowski and M. Adli. CRISPR ethics: moral considerations for applications of a powerful tool. *Journal of molecular biology*, 431(1):88, 1 2019. ISSN 10898638. doi: 10.1016/J.JMB.2018.05.044. URL /pmc/articles/PMC6286228//pmc/articles/PMC6286228/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC6286228/.

J. P. Broughton, X. Deng, G. Yu, C. L. Fasching, V. Servellita, J. Singh, X. Miao, J. A. Streithorst, A. Granados, A. Sotomayor-Gonzalez, K. Zorn, A. Gopez, E. Hsu, W. Gu, S. Miller, C. Y. Pan, H. Guevara, D. A. Wadford, J. S. Chen, and C. Y. Chiu. CRISPR–Cas12-based detection of SARS-CoV-2. *Nature Biotechnology 2020 38:7*, 38(7):870–874, 4 2020. ISSN 1546-1696. doi: 10.1038/s41587-020-0513-4. URL https://www.nature.com/articles/s41587-020-0513-4.

S. J. Brouns, M. M. Jore, M. Lundgren, E. R. Westra, R. J. Slijkhuis, A. P. Snijders, M. J. Dickman, K. S. Makarova, E. V. Koonin, and J. Van Der Oost. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, 321(5891):960–964, 8 2008. ISSN 00368075. doi: 10.1126/SCIENCE.1159689/SUPPL{_}FILE/BROUNS.SOM.PDF. URL https://www.science.org/doi/full/10.1126/science.1159689www.genomatix.de/www.sciencemag.org/cgi/content/full/1160342/DC1.

N. Brunetti-Pierri and P. Ng. Gene therapy with helper-dependent adenoviral vectors: lessons from studies in large animal models. *Virus Genes 2017 53:5*, 53(5):684–691, 6 2017. ISSN 1572-994X. doi: 10.1007/S11262-017-1471-X. URL https://link.springer.com/article/10.1007/s11262-017-1471-x.

Z. Bryant, F. C. Oberstrass, and A. Basu. Recent developments in single-molecule DNA mechanics. *Current opinion in structural biology*, 22(3):304, 6 2012. ISSN 0959440X. doi: 10.1016/J.SBI.2012.04.007. URL /pmc/articles/PMC3726534//pmc/articles/PMC3726534/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3726534/.

C. J. Bustamante, Y. R. Chemla, S. Liu, and M. D. Wang. Optical tweezers in single-molecule biophysics. *Nature reviews. Methods primers*, 1(1), 12 2021. doi: 10.1038/S43586-021-00021-6. URL /pmc/articles/PMC8629167//pmc/articles/PMC8629167/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC8629167/.

P. Cameron, C. K. Fuller, P. D. Donohoue, B. N. Jones, M. S. Thompson, M. M. Carter, S. Gradia, B. Vidal, E. Garner, E. M. Slorach, E. Lau, L. M. Banh, A. M. Lied, L. S. Edwards, A. H. Settle, D. Capurso, V. Llaca, S. Deschamps, M. Cigan, J. K. Young, and A. P. May. Mapping the genomic landscape of CRISPR–Cas9 cleavage. *Nature Methods 2017 14:6*, 14(6):600–606, 5 2017. ISSN 1548-7105. doi: 10.1038/nmeth.4284. URL https://www.nature.com/articles/nmeth.4284.

A. L. Caplan, B. Parent, M. Shen, and C. Plunkett. No time to waste—the ethical challenges created by CRISPR. *EMBO reports*, 16(11):1421–1426, 11 2015. ISSN 1469-3178. doi: 10.15252/EMBR.201541337. URL https://onlinelibrary.wiley.com/doi/full/10.15252/embr.201541337https://onlinelibrary.wiley.com/doi/abs/10.15252/embr.201541337https://www.embopress.org/doi/abs/10.15252/embr.201541337.

A. Casini, M. Olivieri, G. Petris, C. Montagna, G. Reginato, G. Maule, F. Lorenzin, D. Prandi, A. Romanel, F. Demichelis, A. Inga, and A. Cereseto. A highly specific SpCas9 variant is identified by in vivo screening in yeast. *Nature biotechnology*, 36(3):265, 3 2018. ISSN 15461696. doi: 10.1038/NBT.4066. URL /pmc/articles/PMC6066108//pmc/articles/PMC6066108/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC6066108/.

A. Cebrian-Serrano and B. Davies. CRISPR-Cas orthologues and variants: optimizing the repertoire, specificity and delivery of genome engineering tools. *Mammalian Genome*, 28(7):247, 6 2017. ISSN 14321777. doi: 10.1007/S00335-017-9697-4. URL /pmc/articles/PMC5569134//pmc/articles/PMC5569134/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC5569134/.

E. Charpentier, H. Richter, J. van der Oost, and M. F. White. Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity. *FEMS Microbiology Reviews*, 39(3):428–441, 5 2015. ISSN 0168-6445. doi: 10.1093/FEMSRE/FUV023. URL https://academic.oup.com/femsre/article/39/3/428/2467856.

A. Chavez, J. Scheiman, S. Vora, B. W. Pruitt, M. Tuttle, E. P R Iyer, S. Lin, S. Kiani, C. D. Guzman, D. J. Wiegand, D. Ter-Ovanesyan, J. L. Braff, N. Davidsohn, B. E. Housden, N. Perrimon, R. Weiss, J. Aach, J. J. Collins, and G. M. Church. Highly-efficient Cas9-mediated transcriptional programming. *Nature methods*, 12(4):326, 3 2015. ISSN 15487105. doi: 10.1038/NMETH.3312. URL /pmc/articles/PMC4393883/https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4393883/.

B. Chen, W. Zou, H. Xu, Y. Liang, and B. Huang. Efficient labeling and imaging of protein-coding genes in living cells using CRISPR-Tag. *Nature Communications*, 9(1), 2018. ISSN 20411723. doi: 10.1038/s41467-018-07498-y. URL http://dx.doi.org/10.1038/s41467-018-07498-y.

F. Chen, X. Ding, Y. Feng, T. Seebeck, Y. Jiang, and G. D. Davis. Targeted activation of diverse CRISPR-Cas systems for mammalian genome editing via proximal CRISPR targeting. *Nature Communications*, 8, 4 2017a. ISSN 20411723. doi: 10.1038/ncomms14958.

J. S. Chen, Y. S. Dagdas, B. P. Kleinstiver, M. M. Welch, A. A. Sousa, L. B. Harrington,

S. H. Sternberg, J. K. Joung, A. Yildiz, and J. A. Doudna. Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature*, 550(7676):407, 10 2017b. ISSN 14764687. doi: 10.1038/NATURE24268. URL /pmc/articles/PMC5918688//pmc/articles/PMC5918688/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC5918688/.

R. Chiarle, Y. Zhang, R. L. Frock, S. M. Lewis, B. Molinie, Y. J. Ho, D. R. Myers, V. W. Choi, M. Compagno, D. J. Malkin, D. Neuberg, S. Monti, C. C. Giallourakis, M. Gostissa, and F. W. Alt. Genome-wide Translocation Sequencing Reveals Mechanisms of Chromosome Breaks and Rearrangements in B Cells. *Cell*, 147(1):107–119, 9 2011. ISSN 0092-8674. doi: 10.1016/J.CELL.2011.07.049.

J. G. Choi, Y. Dang, S. Abraham, H. Ma, J. Zhang, H. Guo, Y. Cai, J. G. Mikkelsen, H. Wu, P. Shankar, and N. Manjunath. Lentivirus pre-packed with Cas9 protein for safer gene editing. *Gene Therapy 2016 23:7*, 23(7):627–633, 4 2016. ISSN 1476-5462. doi: 10.1038/gt.2016.27. URL https://www.nature.com/articles/gt201627.

M. Christian, T. Cermak, E. L. Doyle, C. Schmidt, F. Zhang, A. Hummel, A. J. Bogdanove, and D. F. Voytas. Targeting DNA Double-Strand Breaks with TAL Effector Nucleases. *Genetics*, 186(2):757, 10 2010. ISSN 00166731. doi: 10.1534/GENETICS.110.120717. URL /pmc/articles/PMC2942870//pmc/articles/PMC2942870/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC2942870/.

K. Chylinski, K. S. Makarova, E. Charpentier, and E. V. Koonin. Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Research*, 42(10):6091, 6 2014. ISSN 13624962. doi: 10.1093/NAR/GKU241. URL /pmc/articles/PMC4041416//pmc/articles/PMC4041416/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4041416/.

J. C. Cofsky, K. M. Soczek, G. J. Knott, E. Nogales, and J. A. Doudna. CRISPR–Cas9 bends and twists DNA to read its sequence. *Nature Structural & Molecular Biology 2022 29:4*, 29(4):395–402, 4 2022. ISSN 1545-9985. doi: 10.1038/s41594-022-00756-0. URL https://www.nature.com/articles/s41594-022-00756-0.

J. Compton. Nucleic acid sequence-based amplification. *Nature 1991 350:6313*, 350(6313):

91–92, 1991. ISSN 1476-4687. doi: 10.1038/350091a0. URL https://www.nature.com/articles/350091a0.

L. Cong, F. A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. D. Hsu, X. Wu, W. Jiang, L. A. Marraffini, and F. Zhang. Multiplex genome engineering using CRISPR/Cas systems. *Science*, 339(6121):819–823, 2 2013. ISSN 10959203. doi: 10.1126/science.1231143.

S. Corless and N. Gilbert. Investigating DNA supercoiling in eukaryotic genomes. *Briefings in Functional Genomics*, 16(6):379–389, 11 2017. ISSN 2041-2649. doi: 10.1093/BFGP/ELX007. URL https://academic.oup.com/bfg/article/16/6/379/3749256.

L. Cuculis and C. M. Schroeder. A Single-Molecule View of Genome Editing Proteins: Biophysical Mechanisms for TALEs and CRISPR/Cas9. *https://doi.org/10.1146/annurev-chembioeng-060816-101603*, 8:577–597, 6 2017. ISSN 19475438. doi: 10.1146/ANNUREV-CHEMBIOENG-060816-101603. URL https://www.annualreviews.org/doi/abs/10.1146/annurev-chembioeng-060816-101603.

R. M. Daer, J. P. Cutts, D. A. Brafman, and K. A. Haynes. The Impact of Chromatin Dynamics on Cas9-Mediated Genome Editing in Human Cells. *ACS Synthetic Biology*, 6(3):428–438, 3 2017. ISSN 21615063. doi: 10.1021/acssynbio.5b00299.

Y. S. Dagdas, J. S. Chen, S. H. Sternberg, J. A. Doudna, and A. Yildiz. A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. *Science Advances*, 3(8), 8 2017. ISSN 23752548. doi: 10.1126/SCIADV.AAO0027. URL /pmc/articles/PMC5547770//pmc/articles/PMC5547770/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC5547770/.

G. Damage, S. Centre, M. O. . Driscoll, and P. A. Jeggo. The role of double-strand break repair — insights from human genetics. *Nature Reviews Genetics 2006 7:1*, 7(1):45–54, 1 2006. ISSN 1471-0064. doi: 10.1038/nrg1746. URL https://www.nature.com/articles/nrg1746.

A. A. Deniz, S. Mukhopadhyay, and E. A. Lemke. Single-molecule biophysics: at the interface of biology, physics and chemistry. *Journal of the Royal Society Interface*, 5(18):15, 1 2008. ISSN 17425662. doi: 10.1098/RSIF.2007.1021.

URL /pmc/articles/PMC2094721//pmc/articles/PMC2094721/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC2094721/.

H. Deveau, R. Barrangou, J. E. Garneau, J. Labonté, C. Fremaux, P. Boyaval, D. A. Romero, P. Horvath, and S. Moineau. Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. *Journal of Bacteriology*, 190(4):1390–1400, 2 2008. ISSN 00219193. doi: 10.1128/JB.01412-07/ASSET/6B6A6E24-97BB-4EE0-B02E-76FCAC6D7BD8/ASSETS/GRAPHIC/ZJB0040875380003.JPEG. URL https://journals.asm.org/doi/full/10.1128/JB.01412-07.

X. Ding, M. P. Stewart, A. Sharei, J. C. Weaver, R. S. Langer, and K. F. Jensen. High-throughput Nuclear Delivery and Rapid Expression of DNA via Mechanical and Electrical Cell-Membrane Disruption. *Nature biomedical engineering*, 1(3), 3 2017. ISSN 2157846X. doi: 10.1038/S41551-017-0039. URL /pmc/articles/PMC5602535//pmc/articles/PMC5602535/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC5602535/.

N. Doench, JohnFusi, M. Sullender, M. Hegde, E. W. Vaimberg, K. F. Donovan, I. Smith, Z. Tothova, C. Wilen, R. Orchard, H. W. Virgin, J. Listgarten, and D. E. Root. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnology*, 34(2):184–191, 2016. ISSN 1650-1519. doi: 10.1038/nbt.3437.Optimized. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4744125/pdf/nihms-740733.pdf.

D. Dong, K. Ren, X. Qiu, J. Zheng, M. Guo, X. Guan, H. Liu, N. Li, B. Zhang, D. Yang, C. Ma, S. Wang, D. Wu, Y. Ma, S. Fan, J. Wang, N. Gao, and Z. Huang. The crystal structure of Cpf1 in complex with CRISPR RNA. *Nature 2016 532:7600*, 532(7600):522–526, 4 2016. ISSN 1476-4687. doi: 10.1038/nature17944. URL https://www.nature.com/articles/nature17944.

J. A. Doudna. The promise and challenge of therapeutic genome editing. *Nature 2020 578:7794*, 578(7794):229–236, 2 2020. ISSN 1476-4687. doi: 10.1038/s41586-020-1978-5. URL https://www.nature.com/articles/s41586-020-1978-5.

J. Duan, G. Lu, Z. Xie, M. Lou, J. Luo, L. Guo, and Y. Zhang. Genome-wide identification

of CRISPR/Cas9 off-targets in human genome. *Cell Research 2014 24:8*, 24(8):1009–1012, 7 2014. ISSN 1748-7838. doi: 10.1038/cr.2014.87. URL https://www.nature.com/articles/cr201487.

A. Engel and D. J. Müller. Observing single biomolecules at work with the atomic force microscope. *Nature structural biology*, 7(9):715–718, 9 2000. ISSN 1072-8368. doi: 10.1038/78929. URL https://pubmed.ncbi.nlm.nih.gov/10966636/.

B. Eslami-Mossallam, M. Klein, C. V. Smagt, K. V. Sanden, S. K. Jones, J. A. Hawkins, I. J. Finkelstein, and M. Depken. A kinetic model predicts SpCas9 activity, improves off-target classification, and reveals the physical basis of targeting fidelity. *Nature Communications*, 13(1):1367–1367, 3 2022. ISSN 2041-1723. doi: 10.1038/S41467-022-28994-2. URL https://europepmc.org/articles/PMC8924176https://europepmc.org/article/pmc/pmc8924176.

T. Fazio, M. L. Visnapuu, S. Wind, and E. C. Greene. DNA Curtains and Nanoscale Curtain Rods: High-Throughput Tools for Single Molecule Imaging. *Langmuir : the ACS journal of surfaces and colloids*, 24(18):10524, 9 2008. ISSN 07437463. doi: 10.1021/LA801762H. URL /pmc/articles/PMC3033740//pmc/articles/PMC3033740/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3033740/.

P. C. Fineran and E. Charpentier. Memory of viral infections by CRISPR-Cas adaptive immune systems: Acquisition of new information. *Virology*, 434(2):202–209, 12 2012. ISSN 00426822. doi: 10.1016/J.VIROL.2012.10.003. URL http://dx.doi.org/10.1016/j.virol.2012.10.003.

M. F. Flajnik and M. Kasahara. Origin and evolution of the adaptive immune system: genetic events and selective pressures. *Nature reviews. Genetics*, 11(1):47, 1 2010. ISSN 14710056. doi: 10.1038/NRG2703. URL /pmc/articles/PMC3805090//pmc/articles/PMC3805090/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3805090/.

I. Fonfara, H. Richter, M. BratoviÄ, A. Le Rhun, and E. Charpentier. The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature 2016 532:7600*,

532(7600):517–521, 4 2016. ISSN 1476-4687. doi: 10.1038/nature17945. URL https://www.nature.com/articles/nature17945.

H. Frangoul, D. Altshuler, M. D. Cappellini, Y.-S. Chen, J. Domm, B. K. Eustace, J. Foell, J. de la Fuente, S. Grupp, R. Handgretinger, T. W. Ho, A. Kattamis, A. Kernytsky, J. Lekstrom-Himes, A. M. Li, F. Locatelli, M. Y. Mapara, M. de Montalembert, D. Rondelli, A. Sharma, S. Sheth, S. Soni, M. H. Steinberg, D. Wall, A. Yen, and S. Corbacioglu. CRISPR-Cas9 Gene Editing for Sickle Cell Disease and β-Thalassemia. *New England Journal of Medicine*, 384(3):252–260, 1 2021. ISSN 0028-4793. doi: 10.1056/NEJMOA2031054/SUPPL{_}FILE/NEJMOA2031054{_}DATA-SHARING.PDF. URL https://www.nejm.org/doi/full/10.1056/NEJMoa2031054.

J. R. Friedman, W. J. Fredericks, D. E. Jensen, D. W. Speicher, X. P. Huang, E. G. Neilson, and F. J. Rauscher. KAP-1, a novel corepressor for the highly conserved KRAB repression domain. *Genes & development*, 10(16):2067–2078, 1996. ISSN 0890-9369. doi: 10.1101/GAD.10.16.2067. URL https://pubmed.ncbi.nlm.nih.gov/8769649/.

R. L. Frock, J. Hu, R. M. Meyers, Y. J. Ho, E. Kii, and F. W. Alt. Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nature Biotechnology 2014 33:2*, 33(2):179–186, 12 2014. ISSN 1546-1696. doi: 10.1038/nbt.3101. URL https://www.nature.com/articles/nbt.3101.

U. Ganbaatar and C. Liu. CRISPR-Based COVID-19 Testing: Toward Next-Generation Point-of-Care Diagnostics. *Frontiers in Cellular and Infection Microbiology*, 11:373, 4 2021. ISSN 22352988. doi: 10.3389/FCIMB.2021.663949/BIBTEX.

L. Gao, D. B. Cox, W. X. Yan, J. C. Manteiga, M. W. Schneider, T. Yamano, H. Nishimasu, O. Nureki, N. Crosetto, and F. Zhang. Engineered Cpf1 variants with altered PAM specificities increase genome targeting range. *Nature biotechnology*, 35(8):789, 8 2017. ISSN 15461696. doi: 10.1038/NBT.3900. URL /pmc/articles/PMC5548640//pmc/articles/PMC5548640/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC5548640/.

P. Gao, H. Yang, K. R. Rajashankar, Z. Huang, and D. J. Patel. Type V CRISPR-Cas Cpf1 endonuclease employs a unique mechanism for crRNA-mediated target DNA recognition.

Cell Research 2016 26:8, 26(8):901–913, 7 2016. ISSN 1748-7838. doi: 10.1038/cr.2016.88. URL https://www.nature.com/articles/cr201688.

J. E. Garneau, M. Dupuis, M. Villion, D. A. Romero, R. Barrangou, P. Boyaval, C. Fremaux, P. Horvath, A. H. Magadán, and S. Moineau. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature 2010 468:7320*, 468(7320):67–71, 11 2010a. ISSN 1476-4687. doi: 10.1038/nature09523. URL https://www.nature.com/articles/nature09523.

J. E. Garneau, M. Dupuis, M. Villion, D. A. Romero, R. Barrangou, P. Boyaval, C. Fremaux, P. Horvath, A. H. Magadán, and S. Moineau. The CRISPR/cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, 468(7320):67–71, 2010b. ISSN 00280836. doi: 10.1038/nature09523.

N. M. Gaudelli, A. C. Komor, H. A. Rees, M. S. Packer, A. H. Badran, D. I. Bryson, and D. R. Liu. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature 2017 551:7681*, 551(7681):464–471, 11 2017. ISSN 1476-4687. doi: 10.1038/nature24644. URL https://www.nature.com/articles/nature24644.

L. A. Gilbert, M. H. Larson, L. Morsut, Z. Liu, A. Gloria, S. E. Torres, N. Stern-ginossar, O. Brandman, H. Whitehead, J. A. Doudna, W. A. Lim, and S. Jonathan. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. 154(2):442–451, 2014. doi: 10.1016/j.cell.2013.06.044.CRISPR-Mediated.

D. Gleditzsch, P. Pausch, H. Müller-Esparza, A. Özcan, X. Guo, G. Bange, and L. Randau. PAM identification by CRISPR-Cas effector complexes: diversified mechanisms and structures. *RNA Biology*, 16(4):504, 4 2019. ISSN 15558584. doi: 10.1080/15476286.2018.1504546. URL /pmc/articles/PMC6546366//pmc/articles/PMC6546366/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC6546366/.

V. Globyte, S. H. Lee, T. Bae, J.-S. Kim, and C. Joo. CRISPR/Cas9 searches for a protospacer adjacent motif by lateral diffusion. *The EMBO Journal*, 38 (4):e99466, 2 2019. ISSN 1460-2075. doi: 10.15252/EMBJ.201899466. URL https://onlinelibrary.wiley.com/doi/full/10.15252/embj.201899466https:

//onlinelibrary.wiley.com/doi/abs/10.15252/embj.201899466https://www.embopress.org/doi/abs/10.15252/embj.201899466.

A. C. Groner, S. Meylan, A. Ciuffi, N. Zangger, G. Ambrosini, N. Dénervaud, P. Bucher, and D. Trono. KRAB–Zinc Finger Proteins and KAP1 Can Mediate Long-Range Transcriptional Repression through Heterochromatin Spreading. *PLoS Genetics*, 6(3):1000869, 3 2010. ISSN 15537390. doi: 10.1371/JOURNAL.PGEN.1000869. URL /pmc/articles/PMC2832679//pmc/articles/PMC2832679/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC2832679/.

D. T. Gruszka. Biochemistry: one molecule at a time. *Essays in Biochemistry*, 65(1):1, 4 2021. ISSN 00711365. doi: 10.1042/EBC20210015. URL /pmc/articles/PMC8056033//pmc/articles/PMC8056033/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC8056033/.

D. H. Haft, J. Selengut, E. F. Mongodin, and K. E. Nelson. A Guild of 45 CRISPR-Associated (Cas) Protein Families and Multiple CRISPR/Cas Subtypes Exist in Prokaryotic Genomes. *PLoS Computational Biology*, 1(6):e60, 11 2005. ISSN 1553-7358. doi: 10.1371/JOURNAL.PCBI.0010060. URL /pmc/articles/PMC1282333//pmc/articles/PMC1282333/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC1282333/.

C. R. Hale and M. O. Duff. RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. 139(5):945–956, 2009. doi: 10.1016/j.cell.2009.07.040.RNA-Guided.

H. G. Hansma, L. I. Pietrasanta, I. D. Auerbach, C. Sorenson, R. Golan, and P. A. Holden. Probing biopolymers with the atomic force microscope: A review. *http://dx.doi.org/10.1163/156856200743940*, 11(7):675–683, 1 2012. ISSN 15685624. doi: 10.1163/156856200743940. URL https://www.tandfonline.com/doi/abs/10.1163/156856200743940.

R. Hao, Z. Peng, and B. Zhang. Single-Molecule Fluorescence Microscopy for Probing the Electrochemical Interface. *ACS Omega*, 5(1):89–97, 1 2020. ISSN 24701343. doi: 10.1021/ACSOMEGA.9B03763/ASSET/IMAGES/LARGE/AO9B03763{_}0007.JPEG. URL https://pubs.acs.org/doi/full/10.1021/acsomega.9b03763.

I. Heller, T. P. Hoekstra, G. A. King, E. J. Peterman, and G. J. Wuite. Optical tweezers analysis of DNA-protein complexes. *Chemical Reviews*, 114(6):3087–3119, 3 2014. ISSN 15206890. doi: 10.1021/CR4003006/ASSET/IMAGES/CR4003006.SOCIAL.JPEG{_}V03. URL https://pubs.acs.org/doi/abs/10.1021/cr4003006.

F. Hille, H. Richter, S. P. Wong, M. Bratovič, S. Ressel, and E. Charpentier. The Biology of CRISPR-Cas: Backward and Forward. *Cell*, 172(6):1239–1259, 3 2018. ISSN 1097-4172. doi: 10.1016/J.CELL.2017.11.032. URL https://pubmed.ncbi.nlm.nih.gov/29522745/.

I. B. Hilton, A. M. D'Ippolito, C. M. Vockley, P. I. Thakore, G. E. Crawford, T. E. Reddy, and C. A. Gersbach. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nature Biotechnology*, 33(5):510–517, 5 2015. ISSN 15461696. doi: 10.1038/nbt.3199.

J. M. Hinz, M. F. Laughery, and J. J. Wyrick. Nucleosomes Inhibit Cas9 Endonuclease Activity in Vitro. *Biochemistry*, 54(48):7063–7066, 12 2015. ISSN 15204995. doi: 10.1021/acs.biochem.5b01108.

S. Hirano, H. Nishimasu, R. Ishitani, and O. Nureki. Structural Basis for the Altered PAM Specificities of Engineered CRISPR-Cas9. *Molecular Cell*, 61(6):886–894, 3 2016. ISSN 1097-2765. doi: 10.1016/J.MOLCEL.2016.02.018.

M. A. Horlbeck, L. B. Witkowsky, B. Guglielmi, J. M. Replogle, L. A. Gilbert, J. E. Villalta, S. E. Torigoe, R. Tjian, and J. S. Weissman. Nucleosomes impede Cas9 access to DNA in vivo and in vitro. 2016. doi: 10.7554/eLife.12677.001.

P. D. Hsu, D. A. Scott, J. A. Weinstein, F. A. Ran, S. Konermann, V. Agarwala, Y. Li, E. J. Fine, X. Wu, O. Shalem, T. J. Cradick, L. A. Marraffini, G. Bao, and F. Zhang. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology 2013 31:9*, 31(9): 827–832, 7 2013. ISSN 1546-1696. doi: 10.1038/nbt.2647. URL https://www.nature.com/articles/nbt.2647.

J. H. Hu, S. M. Miller, M. H. Geurts, W. Tang, L. Chen, N. Sun, C. M. Zeina, X. Gao, H. A. Rees, Z. Lin, and D. R. Liu. Evolved Cas9 variants with broad PAM compatibility and high

DNA specificity. *Nature 2018 556:7699*, 556(7699):57–63, 2 2018. ISSN 1476-4687. doi: 10.1038/nature26155. URL https://www.nature.com/articles/nature26155.

X. Huang, W. Sun, Z. Cheng, M. Chen, X. Li, J. Wang, G. Sheng, W. Gong, and Y. Wang. Structural basis for two metal-ion catalysis of DNA cleavage by Cas12i2. *Nature Communications 2020 11:1*, 11(1):1–14, 10 2020. ISSN 2041-1723. doi: 10.1038/s41467-020-19072-6. URL https://www.nature.com/articles/s41467-020-19072-6.

R. S. Isaac, F. Jiang, J. A. Doudna, W. A. Lim, G. J. Narlikar, and R. Almeida. Nucleosome breathing and remodeling constrain CRISPR-Cas9 function. *eLife*, 5(APRIL2016), 4 2016. ISSN 2050084X. doi: 10.7554/ELIFE.13450.

Y. Ishino, H. Shinagawa, K. Makino, M. Amemura, and A. Nakatura. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *Journal of Bacteriology*, 169(12):5429, 1987. ISSN 00219193. doi: 10.1128/JB.169.12.5429-5433.1987. URL /pmc/articles/PMC213968/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC213968/.

I. E. Ivanov, A. V. Wright, J. C. Cofsky, K. D. Palacio Aris, J. A. Doudna, and Z. Bryant. Cas9 interrogates DNA in discrete steps modulated by mismatches and supercoiling. *Proceedings of the National Academy of Sciences of the United States of America*, 117(11):5853–5860, 3 2020. ISSN 10916490. doi: 10.1073/PNAS.1913445117/-/DCSUPPLEMENTAL. URL /pmc/articles/PMC7084090//pmc/articles/PMC7084090/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC7084090/.

R. Jansen, J. D. Van Embden, W. Gaastra, and L. M. Schouls. Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular Microbiology*, 43(6):1565–1575, 3 2002. ISSN 1365-2958. doi: 10.1046/J.1365-2958.2002.02839.X. URL https://onlinelibrary.wiley.com/doi/full/10.1046/j.1365-2958.2002.02839.xhttps://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2958.2002.02839.xhttps://onlinelibrary.wiley.com/doi/10.1046/j.1365-2958.2002.02839.x.

J. M. Janssen, X. Chen, J. Liu, and M. A. Gonçalves. The Chromatin Structure of CRISPR-Cas9 Target DNA Controls the Balance between Mutagenic and Homology-Directed Gene-

Editing Events. *Molecular Therapy - Nucleic Acids*, 16:141–154, 6 2019. ISSN 21622531. doi: 10.1016/j.omtn.2019.02.009.

Y. Jeon, Y. H. Choi, Y. Jang, J. Yu, J. Goo, G. Lee, Y. K. Jeong, S. H. Lee, I. S. Kim, J. S. Kim, C. Jeong, S. Lee, and S. Bae. Direct observation of DNA target searching and cleavage by CRISPR-Cas12a. *Nature Communications*, 9(1), 12 2018. ISSN 20411723. doi: 10.1038/s41467-018-05245-x.

F. Jiang and J. A. Doudna. CRISPR-Cas9 Structures and Mechanisms. 2017a. doi: 10.1146/annurev-biophys. URL `https://doi.org/10.1146/annurev-biophys-`.

F. Jiang and J. A. Doudna. CRISPR–Cas9 Structures and Mechanisms. *https://doi.org/10.1146/annurev-biophys-062215-010822*, 46:505–529, 5 2017b. ISSN 19361238. doi: 10.1146/ANNUREV-BIOPHYS-062215-010822. URL `https://www.annualreviews.org/doi/abs/10.1146/annurev-biophys-062215-010822`.

F. Jiang, K. Zhou, L. Ma, S. Gressel, and J. A. Doudna. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science*, 348(6242):1477–1481, 6 2015. ISSN 10959203. doi: 10.1126/SCIENCE.AAB1452/SUPPL{_}FILE/JIANG.SM.PDF. URL `https://www.science.org/doi/full/10.1126/science.aab1452`.

F. Jiang, D. W. Taylor, J. S. Chen, J. E. Kornfeld, K. Zhou, A. J. Thompson, E. Nogales, and J. A. Doudna. Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science*, 351(6275):867–871, 2 2016. ISSN 10959203. doi: 10.1126/science.aad8282.

C. Jiao, S. Sharma, G. Dugar, N. L. Peeck, T. Bischler, F. Wimmer, Y. Yu, L. Barquist, C. Schoen, O. Kurzai, C. M. Sharma, and C. L. Beisel. Noncanonical crRNAs derived from host transcripts enable multiplexable RNA detection by Cas9. *Science*, 372 (6545):941–948, 5 2021. ISSN 10959203. doi: 10.1126/SCIENCE.ABE7106/SUPPL{_}FILE/ABE7106{_}TABLE{_}S1.XLSX. URL `https://www.science.org/doi/full/10.1126/science.abe7106`.

M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, and E. Charpentier. A programmable dual RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (New York, N.Y.)*, 337(6096):816, 8 2012a. ISSN 10959203. doi: 10.1126/SCIENCE.1225829.

URL /pmc/articles/PMC6286148//pmc/articles/PMC6286148/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC6286148/.

M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, and E. Charpentier. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337(6096):816–821, 2012b. ISSN 10959203. doi: 10.1126/science.1225829.

M. Jinek, F. Jiang, D. W. Taylor, S. H. Sternberg, E. Kaya, E. Ma, C. Anders, M. Hauer, K. Zhou, S. Lin, M. Kaplan, A. T. Iavarone, E. Charpentier, E. Nogales, and J. A. Doudna. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science*, 343(6176):1–28, 2014. ISSN 10959203. doi: 10.1126/science.1247997.

M. M. Kaminski, O. O. Abudayyeh, J. S. Gootenberg, F. Zhang, and J. J. Collins. CRISPR-based diagnostics. *Nature Biomedical Engineering 2021 5:7*, 5(7):643–656, 7 2021. ISSN 2157-846X. doi: 10.1038/s41551-021-00760-7. URL https://www.nature.com/articles/s41551-021-00760-7.

N. A. Kearns, H. Pham, B. Tabak, R. M. Genga, N. J. Silverstein, M. Garber, and R. Maehr. Functional annotation of native enhancers with a Cas9 -histone demethylase fusion. *Nature methods*, 12(5):401, 4 2015. ISSN 15487105. doi: 10.1038/NMETH.3325. URL /pmc/articles/PMC4414811//pmc/articles/PMC4414811/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4414811/.

M. J. Kellner, J. G. Koob, J. S. Gootenberg, O. O. Abudayyeh, and F. Zhang. SHERLOCK: nucleic acid detection with CRISPR nucleases. *Nature Protocols 2019 14:10*, 14(10):2986–3012, 9 2019. ISSN 1750-2799. doi: 10.1038/s41596-019-0210-2. URL https://www.nature.com/articles/s41596-019-0210-2.

E. M. Kennedy, A. V. R. Kornepati, M. Goldstein, H. P. Bogerd, B. C. Poling, A. W. Whisnant, M. B. Kastan, and B. R. Cullen. Inactivation of the Human Papillomavirus E6 or E7 Gene in Cervical Carcinoma Cells by Using a Bacterial CRISPR/Cas RNA-Guided Endonuclease. *Journal of Virology*, 88(20):11965, 10 2014. ISSN 0022-538X. doi: 10.1128/JVI.01879-14. URL /pmc/articles/PMC4178730//pmc/articles/PMC4178730/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4178730/.

D. Kim, S. Bae, J. Park, E. Kim, S. Kim, H. R. Yu, J. Hwang, J. I. Kim, and J. S. Kim. Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nature Methods 2015 12:3*, 12(3):237–243, 2 2015. ISSN 1548-7105. doi: 10.1038/nmeth.3284. URL https://www.nature.com/articles/nmeth.3284.

D. Kim, J. Kim, J. K. Hur, K. W. Been, S. H. Yoon, and J. S. Kim. Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nature Biotechnology 2016 34:8*, 34(8):863–868, 6 2016. ISSN 1546-1696. doi: 10.1038/nbt.3609. URL https://www.nature.com/articles/nbt.3609.

S. Kim, D. Kim, S. W. Cho, J. Kim, and J. S. Kim. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Research*, 24(6):1012, 2014. ISSN 15495469. doi: 10.1101/GR.171322.113. URL /pmc/articles/PMC4032847//pmc/articles/PMC4032847/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4032847/.

Y. G. Kim, J. Cha, and S. Chandrasegaran. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences of the United States of America*, 93(3):1156, 2 1996. ISSN 00278424. doi: 10.1073/PNAS.93.3.1156. URL /pmc/articles/PMC40048/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC40048/.

M. Klein, B. Eslami-Mossallam, D. G. Arroyo, and M. Depken. Hybridization Kinetics Explains CRISPR-Cas Off-Targeting Rules. *Cell Reports*, 22(6):1413–1423, 2 2018. ISSN 2211-1247. doi: 10.1016/J.CELREP.2018.01.045.

B. P. Kleinstiver, M. S. Prew, S. Q. Tsai, V. V. Topkar, N. T. Nguyen, Z. Zheng, A. P. Gonzales, Z. Li, R. T. Peterson, J. R. J. Yeh, M. J. Aryee, and J. K. Joung. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature 2015 523:7561*, 523(7561):481–485, 6 2015. ISSN 1476-4687. doi: 10.1038/nature14592. URL https://www.nature.com/articles/nature14592.

B. P. Kleinstiver, V. Pattanayak, M. S. Prew, S. Q. Tsai, N. T. Nguyen, Z. Zheng, and J. K. Joung. High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target

effects. *Nature 2015 529:7587*, 529(7587):490–495, 1 2016a. ISSN 1476-4687. doi: 10.1038/nature16526. URL https://www.nature.com/articles/nature16526.

B. P. Kleinstiver, S. Q. Tsai, M. S. Prew, N. T. Nguyen, M. M. Welch, J. M. Lopez, Z. R. McCaw, M. J. Aryee, and J. K. Joung. Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. *Nature Biotechnology*, 34(8):869–874, 1 2016b. ISSN 15461696. doi: 10.1038/nbt.3620.

B. P. Kleinstiver, A. A. Sousa, R. T. Walton, Y. E. Tak, J. Y. Hsu, K. Clement, M. M. Welch, J. E. Horng, J. Malagon-Lopez, I. Scarfò, M. V. Maus, L. Pinello, M. J. Aryee, and J. K. Joung. Engineered CRISPR–Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing, 3 2019. ISSN 15461696.

S. C. Knight, L. Xie, W. Deng, B. Guglielmi, L. B. Witkowsky, L. Bosanac, E. T. Zhang, M. E. Beheiry, J. B. Masson, M. Dahan, Z. Liu, J. A. Doudna, and R. Tjian. Dynamics of CRISPR-Cas9 genome interrogation in living cells. *Science*, 350(6262):823–826, 2015. ISSN 10959203. doi: 10.1126/science.aac6572.

G. J. Knott and J. A. Doudna. CRISPR-Cas guides the future of genetic engineering. *Science (New York, N.Y.)*, 361(6405):866, 8 2018. ISSN 10959203. doi: 10.1126/SCIENCE.AAT5011. URL /pmc/articles/PMC6455913//pmc/articles/PMC6455913/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC6455913/.

A. C. Komor, Y. B. Kim, M. S. Packer, J. A. Zuris, and D. R. Liu. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Proteomics*, 87(1): 124–138, 2016. doi: 10.1016/j.neuron.2015.05.044.HTR7.

S. Konermann, M. D. Brigham, A. E. Trevino, J. Joung, O. O. Abudayyeh, C. Barcena, P. D. Hsu, N. Habib, J. S. Gootenberg, H. Nishimasu, O. Nureki, and F. Zhang. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature 2014 517:7536*, 517(7536):583–588, 12 2014. ISSN 1476-4687. doi: 10.1038/nature14136. URL https://www.nature.com/articles/nature14136.

T. Koo, J. Lee, and J. S. Kim. Measuring and Reducing Off-Target Activities of Programmable Nucleases Including CRISPR-Cas9. *Molecules and Cells*, 38(6):475–481, 6 2015.

ISSN 02191032. doi: 10.14348/MOLCELLS.2015.0103. URL https://www.molcells.org/journal/view.html?doi=10.14348/molcells.2015.0103.

E. V. Koonin and K. S. Makarova. CRISPR-Cas: Evolution of an RNA-based adaptive immunity system in prokaryotes. *RNA Biology*, 10(5):679, 5 2013. ISSN 15558584. doi: 10.4161/RNA.24022. URL /pmc/articles/PMC3737325//pmc/articles/PMC3737325/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3737325/.

E. V. Koonin and K. S. Makarova. Origins and evolution of CRISPR-Cas systems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 374(1772), 5 2019. ISSN 14712970. doi: 10.1098/RSTB.2018.0087. URL /pmc/articles/PMC6452270//pmc/articles/PMC6452270/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC6452270/.

C. Kuscu, S. Arslan, R. Singh, J. Thorpe, and M. Adli. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nature Biotechnology 2014 32:7*, 32(7):677–683, 5 2014a. ISSN 1546-1696. doi: 10.1038/nbt.2916. URL https://www.nature.com/articles/nbt.2916.

C. Kuscu, S. Arslan, R. Singh, J. Thorpe, and M. Adli. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nature biotechnology*, 32(7):677–683, 2014b. ISSN 1546-1696. doi: 10.1038/NBT.2916. URL https://pubmed.ncbi.nlm.nih.gov/24837660/.

D. Y. Kwon, Y. T. Zhao, J. M. Lamonica, and Z. Zhou. Locus-specific histone deacetylation using a synthetic CRISPR-Cas9-based HDAC. *Nature Communications*, 8, 5 2017. ISSN 20411723. doi: 10.1038/NCOMMS15315. URL /pmc/articles/PMC5437308//pmc/articles/PMC5437308/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC5437308/.

S. J. Labrie, J. E. Samson, and S. Moineau. Bacteriophage resistance mechanisms. *Nature Reviews Microbiology 2010 8:5*, 8(5):317–327, 3 2010. ISSN 1740-1534. doi: 10.1038/nrmicro2315. URL https://www.nature.com/articles/nrmicro2315.

C. R. Lazzarotto, N. T. Nguyen, X. Tang, J. Malagon-Lopez, J. A. Guo, M. J. Aryee, J. K. Joung, and S. Q. Tsai. Defining CRISPR–Cas9 genome-wide nuclease activities with CIRCLE-seq. *Nature Protocols 2018 13:11*, 13(11):2615–2642, 10 2018. ISSN 1750-2799. doi: 10.1038/s41596-018-0055-0. URL https://www.nature.com/articles/s41596-018-0055-0.

C. R. Lazzarotto, N. L. Malinin, Y. Li, R. Zhang, Y. Yang, G. H. Lee, E. Cowley, Y. He, X. Lan, K. Jividen, V. Katta, N. G. Kolmakova, C. T. Petersen, Q. Qi, E. Strelcov, S. Maragh, G. Krenciute, J. Ma, Y. Cheng, and S. Q. Tsai. CHANGE-seq reveals genetic and epigenetic effects on CRISPR–Cas9 genome-wide activity. *Nature Biotechnology 2020 38:11*, 38(11): 1317–1327, 6 2020. ISSN 1546-1696. doi: 10.1038/s41587-020-0555-7. URL https://www.nature.com/articles/s41587-020-0555-7.

A. Le Rhun, A. Escalera-Maurer, M. Bratovič, and E. Charpentier. CRISPR-Cas in Streptococcus pyogenes. *RNA Biology*, 16(4):380, 4 2019. ISSN 15558584. doi: 10.1080/15476286.2019.1582974. URL /pmc/articles/PMC6546361//pmc/articles/PMC6546361/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC6546361/.

J. K. Lee, E. Jeong, J. Lee, M. Jung, E. Shin, Y. h. Kim, K. Lee, I. Jung, D. Kim, S. Kim, and J. S. Kim. Directed evolution of CRISPR-Cas9 to increase its specificity. *Nature Communications 2018 9:1*, 9(1):1–10, 8 2018. ISSN 2041-1723. doi: 10.1038/s41467-018-05477-x. URL https://www.nature.com/articles/s41467-018-05477-x.

K. Lee, M. Conboy, H. M. Park, F. Jiang, H. J. Kim, M. A. Dewitt, V. A. Mackley, K. Chang, A. Rao, C. Skinner, T. Shobha, M. Mehdipour, H. Liu, W. C. Huang, F. Lan, N. L. Bray, S. Li, J. E. Corn, K. Kataoka, J. A. Doudna, I. Conboy, and N. Murthy. Nanoparticle delivery of Cas9 ribonucleoprotein and donor DNA in vivo induces homology-directed DNA repair. *Nature biomedical engineering*, 1(11):889, 11 2017. ISSN 2157846X. doi: 10.1038/S41551-017-0137-2. URL /pmc/articles/PMC5968829//pmc/articles/PMC5968829/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC5968829/.

Y. Lei, X. Zhang, J. Su, M. Jeong, M. C. Gundry, Y. H. Huang, Y. Zhou, W. Li, and M. A. Goodell. Targeted DNA methylation in vivo using an engineered dCas9-MQ1 fu-

sion protein. *Nature Communications 2017 8:1*, 8(1):1–10, 7 2017. ISSN 2041-1723. doi: 10.1038/ncomms16026. URL https://www.nature.com/articles/ncomms16026.

H. Li, Y. Yang, W. Hong, M. Huang, M. Wu, and X. Zhao. Applications of genome editing technology in the targeted therapy of human diseases: mechanisms, advances and prospects. *Signal Transduction and Targeted Therapy 2020 5:1*, 5(1):1–23, 1 2020. ISSN 2059-3635. doi: 10.1038/s41392-019-0089-y. URL https://www.nature.com/articles/s41392-019-0089-y.

L. Li, S. Li, N. Wu, J. Wu, G. Wang, G. Zhao, and J. Wang. HOLMESv2: A CRISPR-Cas12b-Assisted Platform for Nucleic Acid Detection and DNA Methylation Quantitation. *ACS Synthetic Biology*, 8(10):2228–2237, 10 2019. ISSN 21615063. doi: 10.1021/ACSSYNBIO.9B00209/SUPPL{_}FILE/SB9B00209{_}SI{_}001.PDF. URL https://pubs.acs.org/doi/abs/10.1021/acssynbio.9b00209.

S. Y. Li, Q. X. Cheng, X. Y. Li, Z. L. Zhang, S. Gao, R. B. Cao, G. P. Zhao, J. Wang, and J. M. Wang. CRISPR-Cas12a-assisted nucleic acid detection. *Cell Discovery 2018 4:1*, 4(1):1–4, 4 2018. ISSN 2056-5968. doi: 10.1038/s41421-018-0028-z. URL https://www.nature.com/articles/s41421-018-0028-z.

P. Liang, Y. Xu, X. Zhang, C. Ding, R. Huang, Z. Zhang, J. Lv, X. Xie, Y. Chen, Y. Li, Y. Sun, Y. Bai, Z. Songyang, W. Ma, C. Zhou, and J. Huang. CRISPR/Cas9-mediated gene editing in human tripronuclear zygotes. *Protein & Cell*, 6(5):363, 5 2015. ISSN 16748018. doi: 10.1007/S13238-015-0153-5. URL /pmc/articles/PMC4417674//pmc/articles/PMC4417674/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4417674/.

Y. Lim, S. Y. Bak, K. Sung, E. Jeong, S. H. Lee, J. S. Kim, S. Bae, and S. K. Kim. Structural roles of guide RNAs in the nuclease activity of Cas9 endonuclease. *Nature Communications 2016 7:1*, 7(1):1–8, 11 2016. ISSN 2041-1723. doi: 10.1038/ncomms13350. URL https://www.nature.com/articles/ncomms13350.

Y. Lin, T. J. Cradick, M. T. Brown, H. Deshmukh, P. Ranjan, N. Sarode, B. M. Wile, P. M. Vertino, F. J. Stewart, and G. Bao. CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. *Nucleic Acids Re-*

search, 42(11):7473–7485, 6 2014. ISSN 0305-1048. doi: 10.1093/NAR/GKU402. URL https://academic.oup.com/nar/article/42/11/7473/1454761.

C. A. Lino, J. C. Harper, J. P. Carney, and J. A. Timlin. Delivering CRISPR: a review of the challenges and approaches. *Drug Delivery*, 25(1):1234, 2018. ISSN 15210464. doi: 10.1080/10717544.2018.1474964. URL /pmc/articles/PMC6058482//pmc/articles/PMC6058482/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC6058482/.

L. Lisowski, S. S. Tay, and I. E. Alexander. Adeno-associated virus serotypes for gene therapeutics. *Current Opinion in Pharmacology*, 24:59–67, 10 2015. ISSN 1471-4892. doi: 10.1016/J.COPH.2015.07.006.

X. S. Liu, H. Wu, X. Ji, Y. Stelzer, X. Wu, S. Czauderna, J. Shu, D. Dadon, R. A. Young, and R. Jaenisch. Editing DNA methylation in the mammalian genome. *Cell*, 167(1):233, 9 2016. ISSN 10974172. doi: 10.1016/J.CELL.2016.08.056. URL /pmc/articles/PMC5062609//pmc/articles/PMC5062609/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC5062609/.

M. Losito, Q. M. Smith, M. D. Newton, M. E. Cuomo, and D. S. Rueda. Cas12a target search and cleavage on force-stretched DNA. *Physical chemistry chemical physics : PCCP*, 23(47):26640–26644, 12 2021. ISSN 1463-9084. doi: 10.1039/D1CP03408A. URL https://pubmed.ncbi.nlm.nih.gov/34494640/.

H. Ma, L. C. Tu, A. Naseri, M. Huisman, S. Zhang, D. Grunwald, and T. Pederson. CRISPR-Cas9 nuclear dynamics and target recognition in living cells. *Journal of Cell Biology*, 214(5):529–537, 2016. ISSN 15408140. doi: 10.1083/jcb.201604115.

D. Maddalo, E. Manchado, C. P. Concepcion, C. Bonetti, J. A. Vidigal, Y. C. Han, P. Ogrodowski, A. Crippa, N. Rekhtman, E. D. Stanchina, S. W. Lowe, and A. Ventura. In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system. *Nature 2014 516:7531*, 516(7531):423–427, 10 2014. ISSN 1476-4687. doi: 10.1038/nature13902. URL https://www.nature.com/articles/nature13902.

M. L. Maeder, S. J. Linder, V. M. Cascio, Y. Fu, Q. H. Ho, and J. K. Joung. CRISPR RNA-

guided activation of endogenous human genes. 10(10):977–979, 2014. doi: 10.1038/nmeth.2598.CRISPR.

M. L. Maeder, M. Stefanidakis, C. J. Wilson, R. Baral, L. A. Barrera, G. S. Bounoutas, D. Bumcrot, H. Chao, D. M. Ciulla, J. A. DaSilva, A. Dass, V. Dhanapal, T. J. Fennell, A. E. Friedland, G. Giannoukos, S. W. Gloskowski, A. Glucksmann, G. M. Gotta, H. Jayaram, S. J. Haskett, B. Hopkins, J. E. Horng, S. Joshi, E. Marco, R. Mepani, D. Reyon, T. Ta, D. G. Tabbaa, S. J. Samuelsson, S. Shen, M. N. Skor, P. Stetkiewicz, T. Wang, C. Yudkoff, V. E. Myer, C. F. Albright, and H. Jiang. Development of a gene-editing approach to restore vision loss in Leber congenital amaurosis type 10. *Nature medicine*, 25(2):229–233, 2 2019. ISSN 1546-170X. doi: 10.1038/S41591-018-0327-9. URL `https://pubmed.ncbi.nlm.nih.gov/30664785/`.

K. S. Makarova, N. V. Grishin, S. A. Shabalina, Y. I. Wolf, and E. V. Koonin. A putative RNA-interference-based immune system in prokaryotes: Computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biology Direct*, 1(1):1–26, 3 2006. ISSN 17456150. doi: 10.1186/1745-6150-1-7/FIGURES/8. URL `https://biologydirect.biomedcentral.com/articles/10.1186/1745-6150-1-7`.

K. S. Makarova, Y. I. Wolf, O. S. Alkhnbashi, F. Costa, S. A. Shah, S. J. Saunders, R. Barrangou, S. J. Brouns, E. Charpentier, D. H. Haft, P. Horvath, S. Moineau, F. J. Mojica, R. M. Terns, M. P. Terns, M. F. White, A. F. Yakunin, R. A. Garrett, J. Van Der Oost, R. Backofen, and E. V. Koonin. An updated evolutionary classification of CRISPR-Cas systems. *Nature Reviews Microbiology*, 13(11):722–736, 2015. ISSN 17401534. doi: 10.1038/nrmicro3569.

K. S. Makarova, Y. I. Wolf, J. Iranzo, S. A. Shmakov, O. S. Alkhnbashi, S. J. Brouns, E. Charpentier, D. Cheng, D. H. Haft, P. Horvath, S. Moineau, F. J. Mojica, D. Scott, S. A. Shah, V. Siksnys, M. P. Terns, Venclovas, M. F. White, A. F. Yakunin, W. Yan, F. Zhang, R. A. Garrett, R. Backofen, J. van der Oost, R. Barrangou, and E. V. Koonin. Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nature Reviews Microbiology 2019 18:2*, 18(2):67–83, 12 2019. ISSN 1740-1534. doi: 10.1038/s41579-019-0299-x. URL `https://www.nature.com/articles/s41579-019-0299-x`.

P. Mali, J. Aach, P. B. Stranges, K. M. Esvelt, M. Moosburner, S. Kosuri, L. Yang, and G. M. Church. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nature biotechnology*, 31(9):833, 9 2013. ISSN 10870156. doi: 10.1038/NBT.2675. URL /pmc/articles/PMC3818127//pmc/articles/PMC3818127/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3818127/.

R. Mans, H. M. van Rossum, M. Wijsman, A. Backx, N. G. Kuijpers, M. van den Broek, P. Daran-Lapujade, J. T. Pronk, A. J. van Maris, and J. M. G. Daran. CRISPR/Cas9: A molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in Saccharomyces cerevisiae. *FEMS Yeast Research*, 15(2):1–15, 2015. ISSN 15671364. doi: 10.1093/femsyr/fov004.

L. A. Marraffini. CRISPR-Cas Immunity against Phages: Its Effects on the Evolution and Survival of Bacterial Pathogens. *PLoS Pathogens*, 9(12):1–4, 2013. ISSN 15537374. doi: 10.1371/JOURNAL.PPAT.1003765. URL /pmc/articles/PMC3861508/https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3861508/.

L. A. Marraffini and E. J. Sontheimer. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science*, 322(5909):1843–1845, 12 2008. ISSN 00368075. doi: 10.1126/SCIENCE.1165771/SUPPL{_}FILE/MARRAFFINI-SOM.PDF. URL https://www.science.org/doi/full/10.1126/science.1165771.

J. McGinn and L. A. Marraffini. Molecular mechanisms of CRISPR–Cas spacer acquisition. *Nature Reviews Microbiology 2018 17:1*, 17(1):7–12, 8 2018. ISSN 1740-1534. doi: 10.1038/s41579-018-0071-7. URL https://www.nature.com/articles/s41579-018-0071-7.

J. R. Moffitt, Y. R. Chemla, S. B. Smith, and C. Bustamante. Recent Advances in Optical Tweezers. *Annual Review of Biochemistry*, 77:205–228, 6 2008. ISSN 00664154. doi: 10.1146/ANNUREV.BIOCHEM.77.043007.090225. URL https://www.annualreviews.org/doi/abs/10.1146/annurev.biochem.77.043007.090225.

F. J. Mojica, C. Díez-Villaseñor, J. García-Martínez, and E. Soria. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of Molec-*

ular Evolution, 60(2):174–182, 2 2005. ISSN 00222844. doi: 10.1007/S00239-004-0046-3/TABLES/6. URL https://link.springer.com/article/10.1007/s00239-004-0046-3.

F. J. Mojica, C. Díez-Villaseñor, J. García-Martínez, and C. Almendros. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology*, 155(3): 733–740, 2009. ISSN 13500872. doi: 10.1099/mic.0.023960-0.

L. D. Moore, T. Le, and G. Fan. DNA Methylation and Its Basic Function. *Neuropsychopharmacology 2013 38:1*, 38(1):23–38, 7 2012. ISSN 1740-634X. doi: 10.1038/npp.2012.112. URL https://www.nature.com/articles/npp2012112.

M. A. Moreno-Mateos, J. P. Fernandez, R. Rouet, C. E. Vejnar, M. A. Lane, E. Mis, M. K. Khokha, J. A. Doudna, and A. J. Giraldez. CRISPR-Cpf1 mediates efficient homology-directed repair and temperature-controlled genome editing. *Nature Communications 2017 8:1*, 8(1):1–9, 12 2017. ISSN 2041-1723. doi: 10.1038/s41467-017-01836-2. URL https://www.nature.com/articles/s41467-017-01836-2.

S. Morita, H. Noguchi, T. Horii, K. Nakabayashi, M. Kimura, K. Okamura, A. Sakai, H. Nakashima, K. Hata, K. Nakashima, and I. Hatada. Targeted DNA demethylation in vivo using dCas9–peptide repeat and scFv–TET1 catalytic domain fusions. *Nature Biotechnology 2016 34:10*, 34(10):1060–1065, 8 2016. ISSN 1546-1696. doi: 10.1038/nbt.3658. URL https://www.nature.com/articles/nbt.3658.

K. Murugan, A. S. Seetharam, A. J. Severin, and D. G. Sashital. CRISPR-Cas12a has widespread off-target and dsDNA-nicking effects. *The Journal of Biological Chemistry*, 295(17):5538, 4 2020. ISSN 1083351X. doi: 10.1074/JBC.RA120.012933. URL /pmc/articles/PMC7186167//pmc/articles/PMC7186167/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC7186167/.

M. Naeem, S. Majeed, M. Z. Hoque, and I. Ahmad. Latest Developed Strategies to Minimize the Off-Target Effects in CRISPR-Cas-Mediated Genome Editing. *Cells*, 9(7), 7 2020. ISSN 20734409. doi: 10.3390/CELLS9071608. URL /pmc/articles/PMC7407193//pmc/articles/PMC7407193/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC7407193/.

C. Naughton, N. Avlonitis, S. Corless, J. G. Prendergast, I. K. Mati, P. P. Eijk, S. L. Cockroft, M. Bradley, B. Ylstra, and N. Gilbert. Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures. *Nature structural & molecular biology*, 20(3):387, 3 2013. ISSN 15459993. doi: 10.1038/NSMB.2509. URL /pmc/articles/PMC3689368//pmc/articles/PMC3689368/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3689368/.

K. C. Neuman and A. Nagy. Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nature Methods 2008 5:6*, 5(6):491–505, 5 2008. ISSN 1548-7105. doi: 10.1038/nmeth.1218. URL https://www.nature.com/articles/nmeth.1218.

M. D. Newton, B. J. Taylor, R. P. Driessen, L. Roos, N. Cvetesic, S. Allyjaun, B. Lenhard, M. E. Cuomo, and D. S. Rueda. DNA stretching induces Cas9 off-target activity. *Nature Structural and Molecular Biology*, 26(3):185–192, 3 2019. ISSN 15459985. doi: 10.1038/s41594-019-0188-z.

H. Nishimasu, F. A. Ran, P. D. Hsu, S. Konermann, S. I. Shehata, N. Dohmae, R. Ishitani, F. Zhang, and O. Nureki. Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell*, 156(5):935, 2 2014. ISSN 10974172. doi: 10.1016/J.CELL.2014.02.001. URL /pmc/articles/PMC4139937//pmc/articles/PMC4139937/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4139937/.

H. O'Geen, I. M. Henry, M. S. Bhakta, J. F. Meckler, and D. J. Segal. A genome-wide analysis of Cas9 binding specificity using ChIP-seq and targeted sequence capture. *Nucleic Acids Research*, 43(6):3389–3404, 3 2015. ISSN 0305-1048. doi: 10.1093/NAR/GKV137. URL https://academic.oup.com/nar/article/43/6/3389/2453430.

I. C. Okafor, D. Singh, Y. Wang, M. Jung, H. Wang, J. Mallon, S. Bailey, J. K. Lee, and T. Ha. Single molecule analysis of effects of non-canonical guide RNAs and specificity-enhancing mutations on Cas9-induced DNA unwinding. *Nucleic acids research*, 47(22):11880–11888, 12 2019. ISSN 1362-4962. doi: 10.1093/NAR/GKZ1058. URL https://pubmed.ncbi.nlm.nih.gov/31713616/.

S. Osuka, K. Isomura, S. Kajimoto, T. Komori, H. Nishimasu, T. Shima, O. Nureki, and S. Uemura. Real-time observation of flexible domain movements in CRISPR–Cas9. *The EMBO Journal*, 37(10), 5 2018. ISSN 0261-4189. doi: 10.15252/EMBJ.201796941. URL /pmc/articles/PMC5978321//pmc/articles/PMC5978321/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC5978321/.

M. Pacesa and M. Jinek. Mechanism of R-loop formation and conformational activation of Cas9. *bioRxiv*, page 2021.09.16.460614, 9 2021. doi: 10.1101/2021.09.16.460614. URL https://www.biorxiv.org/content/10.1101/2021.09.16.460614v1https://www.biorxiv.org/content/10.1101/2021.09.16.460614v1.abstract.

D. Palmer and P. Ng. Improved system for helper-dependent adenoviral vector production. *Molecular Therapy*, 8(5):846–852, 11 2003. ISSN 1525-0016. doi: 10.1016/J.YMTHE.2003.08.014.

K. Pardee, A. A. Green, M. K. Takahashi, D. Braff, G. Lambert, J. W. Lee, T. Ferrante, D. Ma, N. Donghia, M. Fan, N. M. Daringer, I. Bosch, D. M. Dudley, D. H. O'Connor, L. Gehrke, and J. J. Collins. Rapid, Low-Cost Detection of Zika Virus Using Programmable Biomolecular Components. *Cell*, 165(5):1255–1266, 5 2016. ISSN 0092-8674. doi: 10.1016/J.CELL.2016.04.059.

V. Pattanayak, S. Lin, J. P. Guilinger, E. Ma, J. A. Doudna, and D. R. Liu. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nature biotechnology*, 31(9):839, 9 2013. ISSN 10870156. doi: 10.1038/NBT.2673. URL /pmc/articles/PMC3782611//pmc/articles/PMC3782611/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3782611/.

P. Pausch, K. M. Soczek, D. A. Herbst, C. A. Tsuchida, B. Al-Shayeb, J. F. Banfield, E. Nogales, and J. A. Doudna. DNA interference states of the hypercompact CRISPR-CasΦ effector. *Nature structural & molecular biology*, 28(8):652, 8 2021. ISSN 15459985. doi: 10.1038/S41594-021-00632-3. URL /pmc/articles/PMC8496406//pmc/articles/PMC8496406/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC8496406/.

D. S. Peabody. The RNA binding site of bacteriophage MS2 coat protein. *The EMBO*

Journal, 12(2):595, 1993. ISSN 02614189. doi: 10.1002/j.1460-2075.1993.tb05691.x. URL /pmc/articles/PMC413242/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC413242/.

P. Perez-Pinera, D. D. Kocak, C. M. Vockley, A. F. Adler, A. M. Kabadi, L. R. Polstein, P. I. Thakore, K. A. Glass, D. G. Ousterout, K. W. Leong, F. Guilak, G. E. Crawford, T. E. Reddy, and C. A. Gersbach. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nature methods*, 10(10):973, 10 2013. ISSN 15487091. doi: 10.1038/NMETH.2600. URL /pmc/articles/PMC3911785//pmc/articles/PMC3911785/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3911785/.

M. Peterka, N. Akrap, S. Li, S. Wimberger, P.-P. Hsieh, D. Degtev, B. Bestas, J. Barr, S. van de Plassche, P. Mendoza-Garcia, S. Šviković, G. Sienski, M. Firth, and M. Maresca. Harnessing DSB repair to promote efficient homology-dependent and -independent prime editing. *Nature Communications 2022 13:1*, 13(1):1–9, 3 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-28771-1. URL https://www.nature.com/articles/s41467-022-28771-1.

A. Petrov, J. Chen, S. O'Leary, A. Tsai, and J. D. Puglisi. Single-Molecule Analysis of Translational Dynamics. *Cold Spring Harbor Perspectives in Biology*, 4(9), 2012. ISSN 19430264. doi: 10.1101/CSHPERSPECT.A011551. URL /pmc/articles/PMC3428772//pmc/articles/PMC3428772/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3428772/.

R. J. Platt, S. Chen, Y. Zhou, M. J. Yim, L. Swiech, H. R. Kempton, J. E. Dahlman, O. Parnas, T. M. Eisenhaure, M. Jovanovic, D. B. Graham, S. Jhunjhunwala, M. Heidenreich, R. J. Xavier, R. Langer, D. G. Anderson, N. Hacohen, A. Regev, G. Feng, P. A. Sharp, and F. Zhang. CRISPR-Cas9 knockin mice for genome editing and cancer modeling. *Cell*, 159(2):440–455, 10 2014. ISSN 10974172. doi: 10.1016/J.CELL.2014.09.014/ATTACHMENT/F9704CC5-59D4-4F8D-9A20-752901BAA1E5/MMC4.PDF. URL http://www.cell.com/article/S0092867414011635/fulltexthttp://www.cell.com/article/S0092867414011635/abstracthttps://www.cell.com/cell/abstract/S0092-8674(14)01163-5.

F. Port, M. Starostecka, and M. Boutros. Multiplexed conditional genome editing with Cas12a

in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America*, 117(37):22890–22899, 9 2020. ISSN 10916490. doi: 10.1073/PNAS.2004655117. URL www.pnas.org/cgi/doi/10.1073/pnas.2004655117.

C. Pourcel, G. Salvignol, and G. Vergnaud. CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology*, 151(3):653–663, 3 2005. ISSN 13500872. doi: 10.1099/MIC.0.27437-0/CITE/REFWORKS. URL https://www.microbiologyresearch.org/content/journal/micro/10.1099/mic.0.27437-0.

L. S. Qi, M. H. Larson, L. A. Gilbert, J. A. Doudna, J. S. Weissman, A. P. Arkin, and W. A. Lim. Repurposing CRISPR as an RNA-γuided platform for sequence-specific control of gene expression. *Cell*, 152(5):1173–1183, 2013. ISSN 00928674. doi: 10.1016/j.cell.2013.02.022.

C. Rouillon, M. Zhou, J. Zhang, A. Politis, V. Beilsten-Edmands, G. Cannone, S. Graham, C. V. Robinson, L. Spagnolo, and M. F. White. Structure of the CRISPR interference complex CSM reveals key similarities with cascade. *Molecular Cell*, 52(1):124–134, 10 2013. ISSN 10972765. doi: 10.1016/j.molcel.2013.08.020.

R. Roy, S. Hohng, and T. Ha. A Practical Guide to Single Molecule FRET. *Nature methods*, 5(6):507, 6 2008. ISSN 15487091. doi: 10.1038/NMETH.1208. URL /pmc/articles/PMC3769523//pmc/articles/PMC3769523/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3769523/.

F. O. Rueda, M. Bista, M. D. Newton, A. U. Goeppert, M. E. Cuomo, E. Gordon, F. Kröner, J. A. Read, J. D. Wrigley, D. Rueda, and B. J. Taylor. Mapping the sugar dependency for rational generation of a DNA-RNA hybrid-guided Cas9 endonuclease. *Nature Communications*, 8(1), 12 2017. ISSN 20411723. doi: 10.1038/s41467-017-01732-9.

A. Saha, P. R. Arantes, R. V. Hsu, Y. B. Narkhede, M. Jinek, and G. Palermo. Molecular Dynamics Reveals a DNA-Induced Dynamic Switch Triggering Activation of CRISPR-Cas12a. *Journal of Chemical Information and Modeling*, 60(12):6427–6437, 12 2020. ISSN 15205142. doi: 10.1021/ACS.JCIM.0C00929/SUPPL{_}FILE/CI0C00929{_}SI{_

}001.PDF. URL /pmc/articles/PMC7605327//pmc/articles/PMC7605327/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC7605327/.

R. Schep, C. Leemans, E. K. Brinkman, T. van Schaik, and B. van Steensel. Protocol: A Multiplexed Reporter Assay to Study Effects of Chromatin Context on DNA Double-Strand Break Repair. *Frontiers in Genetics*, 12:2488, 1 2022. ISSN 16648021. doi: 10.3389/FGENE.2021.785947/BIBTEX.

S. Schmid and T. Hugel. Efficient use of single molecule time traces to resolve kinetic rates, models and uncertainties. *The Journal of Chemical Physics*, 148(12):123312, 12 2017. ISSN 0021-9606. doi: 10.1063/1.5006604. URL https://aip.scitation.org/doi/abs/10.1063/1.5006604.

M. W. Shen, M. Arbab, J. Y. Hsu, D. Worstell, S. J. Culbertson, O. Krabbe, C. A. Cassa, D. R. Liu, D. K. Gifford, and R. I. Sherwood. Predictable and precise template-free CRISPR editing of pathogenic variants. *Nature 2018 563:7733*, 563(7733):646–651, 11 2018. ISSN 1476-4687. doi: 10.1038/s41586-018-0686-x. URL https://www.nature.com/articles/s41586-018-0686-x.

S. Shmakov, O. O. Abudayyeh, K. S. Makarova, Y. I. Wolf, J. S. Gootenberg, E. Semenova, L. Minakhin, J. Joung, S. Konermann, K. Severinov, F. Zhang, and E. V. Koonin. Discovery and functional characterization of diverse Class 2 CRISPR-Cas systems HHS Public Access Graphical abstract. *Mol Cell*, 60(3):385–397, 2015. doi: 10.1016/j.molcel.2015.10.008. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4660269/pdf/nihms-729607.pdf.

D. Singh, S. H. Sternberg, J. Fei, J. A. Doudna, and T. Ha. Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9. *Nature Communications 2016 7:1*, 7(1):1–8, 9 2016. ISSN 2041-1723. doi: 10.1038/ncomms12778. URL https://www.nature.com/articles/ncomms12778.

D. Singh, J. Mallon, A. Poddar, Y. Wang, R. Tippana, O. Yang, S. Bailey, and T. Ha. Real-time observation of DNA target interrogation and product release by the RNA-guided endonuclease CRISPR Cpf1 (Cas12a). *Proceedings of the National Academy of Sciences of the United States of America*, 115(21):5444–5449, 5 2018a. ISSN 10916490.

doi: 10.1073/PNAS.1718686115/SUPPL{_}FILE/PNAS.1718686115.SAPP.PDF. URL www.pnas.org/cgi/doi/10.1073/pnas.1718686115.

D. Singh, Y. Wang, J. Mallon, O. Yang, J. Fei, A. Poddar, D. Ceylan, S. Bailey, and T. Ha. Mechanisms of improved specificity of engineered Cas9s revealed by }single molecule FRET analysis. *Nature structural & molecular biology*, 25 (4):347, 4 2018b. ISSN 15459985. doi: 10.1038/S41594-018-0051-7. URL /pmc/articles/PMC6195204//pmc/articles/PMC6195204/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC6195204/.

I. M. Slaymaker, L. Gao, B. Zetsche, D. A. Scott, W. X. Yan, and F. Zhang. Rationally engineered Cas9 nucleases with improved specificity. *Science (New York, N.Y.)*, 351(6268):84, 1 2016. ISSN 10959203. doi: 10.1126/SCIENCE.AAD5227. URL /pmc/articles/PMC4714946//pmc/articles/PMC4714946/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4714946/.

C. Smith, A. Gore, W. Yan, L. Abalde-Atristain, Z. Li, C. He, Y. Wang, R. A. Brodsky, K. Zhang, L. Cheng, and Z. Ye. Whole-genome sequencing analysis reveals high specificity of CRISPR/Cas9 and TALEN-based genome editing in human iPSCs. *Cell Stem Cell*, 15(1):12–13, 7 2014. ISSN 18759777. doi: 10.1016/J.STEM.2014.06.011/ATTACHMENT/D72FD21E-B222-4C2B-834C-363D3F0A3D53/MMC2.XLSX. URL http://www.cell.com/article/S1934590914002598/fulltexthttp://www.cell.com/article/S1934590914002598/abstracthttps://www.cell.com/cell-stem-cell/abstract/S1934-5909(14)00259-8.

R. Sorek, C. M. Lawrence, and B. Wiedenheft. CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annual review of biochemistry*, 82:237–266, 6 2013. ISSN 1545-4509. doi: 10.1146/ANNUREV-BIOCHEM-072911-172315. URL https://pubmed.ncbi.nlm.nih.gov/23495939/.

S. Stella, P. Alcón, and G. Montoya. Structure of the Cpf1 endonuclease R-loop complex after target DNA cleavage. *Nature 2017 546:7659*, 546(7659):559–563, 5 2017. ISSN 1476-4687. doi: 10.1038/nature22398. URL https://www.nature.com/articles/nature22398.

S. Stella, P. Mesa, J. Thomsen, B. Paul, P. Alcón, S. B. Jensen, B. Saligram, M. E. Moses, N. S. Hatzakis, and G. Montoya. Conformational Activation Promotes CRISPR-Cas12a Catalysis and Resetting of the Endonuclease Activity. *Cell*, 175(7):1856–1871, 12 2018. ISSN 0092-8674. doi: 10.1016/J.CELL.2018.10.045.

S. H. Sternberg, S. Redding, M. Jinek, E. C. Greene, A. Jennifer, M. Biophysics, C. Biology, B. Division, and L. Berkeley. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*, 507(7490):62–67, 2014. doi: 10.1038/nature13011.DNA.

S. H. Sternberg, B. Lafrance, M. Kaplan, and J. A. Doudna. Conformational control of DNA target cleavage by CRISPR–Cas9. *Nature*, 527(7576):110, 11 2015. ISSN 14764687. doi: 10.1038/NATURE15544. URL /pmc/articles/PMC4859810//pmc/articles/PMC4859810/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4859810/.

I. Strohkendl, F. A. Saifuddin, B. A. Gibson, M. K. Rosen, R. Russell, and I. J. Finkelstein. Inhibition of CRISPR-Cas12a DNA targeting by nucleosomes and chromatin. *Science Advances*, 7(11):6030–6040, 3 2021. ISSN 23752548. doi: 10.1126/SCIADV.ABD6030. URL https://www.science.org.

T. Sugiyama and D. Nakada. Control of translation of MS2 RNA cistrons by MS2 coat protein. *Proceedings of the National Academy of Sciences of the United States of America*, 57(6):1744–1750, 1967. ISSN 00278424. doi: 10.1073/PNAS.57.6.1744/ASSET/675DCAA9-493C-4B5C-9346-FF567DF75826/ASSETS/PNAS.57.6.1744.FP.PNG. URL https://www.pnas.org.

D. C. Swarts and M. Jinek. Cas9 versus Cas12a/Cpf1: Structure–function comparisons and implications for genome editing. *Wiley Interdisciplinary Reviews: RNA*, 9(5):e1481, 9 2018. ISSN 1757-7012. doi: 10.1002/WRNA.1481. URL https://onlinelibrary.wiley.com/doi/full/10.1002/wrna.1481https://onlinelibrary.wiley.com/doi/abs/10.1002/wrna.1481https://wires.onlinelibrary.wiley.com/doi/10.1002/wrna.1481.

D. C. Swarts and M. Jinek. Mechanistic Insights into the Cis- and Trans-acting Deoxyribonuclease Activities of Cas12a. *Molecular cell*, 73(3):589, 2 2019. ISSN 10974164. doi: 10.1016/

J.MOLCEL.2018.11.021. URL /pmc/articles/PMC6858279//pmc/articles/PMC6858279/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC6858279/.

D. C. Swarts, J. van der Oost, and M. Jinek. Structural basis for guide RNA processing and seed-dependent DNA targeting and cleavage by CRISPR-Cas12a. *Molecular cell*, 66(2):221, 4 2017. ISSN 10974164. doi: 10.1016/J.MOLCEL.2017.03.016. URL /pmc/articles/PMC6879319//pmc/articles/PMC6879319/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC6879319/.

M. A. Swiat, S. Dashko, M. Den Ridder, M. Wijsman, J. Van Der Oost, J. M. Daran, and P. Daran-Lapujade. FnCpf1: a novel and efficient genome editing tool for Saccharomyces cerevisiae. *Nucleic Acids Research*, 45(21):12585, 12 2017. ISSN 13624962. doi: 10.1093/NAR/GKX1007. URL /pmc/articles/PMC5716609//pmc/articles/PMC5716609/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC5716609/.

L. Swiech, M. Heidenreich, A. Banerjee, N. Habib, Y. Li, J. Trombetta, M. Sur, and F. Zhang. In vivo interrogation of gene function in the mammalian brain using CRISPR-Cas9. *Nature biotechnology*, 33(1):102, 1 2015. ISSN 15461696. doi: 10.1038/NBT.3055. URL /pmc/articles/PMC4492112//pmc/articles/PMC4492112/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4492112/.

M. E. Tanenbaum, L. A. Gilbert, L. S. Qi, J. S. Weissman, and R. D. Vale. A protein tagging system for signal amplification in gene expression and fluorescence imaging. *Cell*, 159(3):635, 10 2014. ISSN 10974172. doi: 10.1016/J.CELL.2014.09.039. URL /pmc/articles/PMC4252608//pmc/articles/PMC4252608/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4252608/.

M. P. Terns. CRISPR-based Technologies: Impact of RNA-targeting Systems. *Molecular cell*, 72(3):404, 11 2018. ISSN 10974164. doi: 10.1016/J.MOLCEL.2018.09.018. URL /pmc/articles/PMC6239212//pmc/articles/PMC6239212/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC6239212/.

P. I. Thakore, A. M. D'Ippolito, L. Song, A. Safi, N. K. Shivakumar, A. M. Kabadi, T. E. Reddy, G. E. Crawford, and C. A. Gersbach. Highly specific epigenome editing by CRISPR-

Cas9 repressors for silencing of distal regulatory elements. *Nature Methods 2015 12:12*, 12 (12):1143–1149, 10 2015. ISSN 1548-7105. doi: 10.1038/nmeth.3630. URL https://www.nature.com/articles/nmeth.3630.

R. E. Thompson, D. R. Larson, and W. W. Webb. Precise Nanometer Localization Analysis for Individual Fluorescent Probes. *Biophysical Journal*, 82(5):2775–2783, 5 2002. ISSN 0006-3495. doi: 10.1016/S0006-3495(02)75618-X.

S. Q. Tsai and J. K. Joung. Defining and improving the genome-wide specificities of CRISPR–Cas9 nucleases. *Nature Reviews Genetics 2016 17:5*, 17(5):300–312, 4 2016. ISSN 1471-0064. doi: 10.1038/nrg.2016.28. URL https://www.nature.com/articles/nrg.2016.28.

S. Q. Tsai, Z. Zheng, N. T. Nguyen, M. Liebers, V. V. Topkar, V. Thapar, N. Wyvekens, C. Khayter, A. J. Iafrate, L. P. Le, M. J. Aryee, and J. K. Joung. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nature Biotechnology 2014 33:2*, 33(2):187–197, 12 2014. ISSN 1546-1696. doi: 10.1038/nbt.3117. URL https://www.nature.com/articles/nbt.3117.

S. Q. Tsai, Z. Zheng, N. T. Nguyen, M. Liebers, V. V. Topkar, V. Thapar, N. Wyvekens, C. Khayter, A. J. Iafrate, L. P. Le, M. J. Aryee, and J. K. Joung. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nature Biotechnology*, 33(2):187–198, 2015. ISSN 15461696. doi: 10.1038/nbt.3117.

S. Q. Tsai, N. T. Nguyen, J. Malagon-Lopez, V. V. Topkar, M. J. Aryee, and J. K. Joung. CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets. *Nature methods*, 14(6):607–614, 5 2017. ISSN 1548-7105. doi: 10.1038/NMETH.4278. URL https://pubmed.ncbi.nlm.nih.gov/28459458/.

F. Uddin, C. M. Rudin, and T. Sen. CRISPR Gene Therapy: Applications, Limitations, and Implications for the Future. *Frontiers in Oncology*, 10:1387, 8 2020. ISSN 2234943X. doi: 10.3389/FONC.2020.01387/BIBTEX.

J. Ungerer and H. B. Pakrasi. Cpf1 Is A Versatile Tool for CRISPR Genome Editing Across

Diverse Species of Cyanobacteria. *Scientific Reports 2016 6:1*, 6(1):1–9, 12 2016. ISSN 2045-2322. doi: 10.1038/srep39681. URL https://www.nature.com/articles/srep39681.

C. A. Vakulskas, D. P. Dever, G. R. Rettig, R. Turk, A. M. Jacobi, M. A. Collingwood, N. M. Bode, M. S. McNeill, S. Yan, J. Camarena, C. M. Lee, S. H. Park, V. Wiebking, R. O. Bak, N. Gomez-Ospina, M. Pavel-Dinu, W. Sun, G. Bao, M. H. Porteus, and M. A. Behlke. A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human haematopoietic stem and progenitor cells. *Nature medicine*, 24(8):1216, 8 2018. ISSN 1546170X. doi: 10.1038/S41591-018-0137-0. URL /pmc/articles/PMC6107069//pmc/articles/PMC6107069/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC6107069/.

K. van Aelst, C. J. Martínez-Santiago, S. J. Cross, and M. D. Szczelkun. The Effect of DNA Topology on Observed Rates of R-Loop Formation and DNA Strand Cleavage by CRISPR Cas12a. *Genes*, 10(2), 2 2019. ISSN 2073-4425. doi: 10.3390/GENES10020169. URL https://pubmed.ncbi.nlm.nih.gov/30813348/.

A. Veres, B. S. Gosis, Q. Ding, R. Collins, A. Ragavendran, H. Brand, S. Erdin, M. E. Talkowski, and K. Musunuru. Low incidence of off-target mutations in individual CRISPR-Cas9 and TALEN targeted human stem cell clones detected by whole-genome sequencing. *Cell stem cell*, 15(1):27–30, 7 2014. ISSN 1875-9777. doi: 10.1016/J.STEM.2014.04.020. URL https://pubmed.ncbi.nlm.nih.gov/24996167/.

S. A. Verkuijl and M. G. Rots. The influence of eukaryotic chromatin state on CRISPR–Cas9 editing efficiencies, 2 2019. ISSN 18790429.

A. Vojta, P. Dobrinic, V. Tadic, L. Bockor, P. Korac, B. Julg, M. Klasic, and V. Zoldos. Repurposing the CRISPR-Cas9 system for targeted DNA methylation. *Nucleic Acids Research*, 44(12):5615, 7 2016. ISSN 13624962. doi: 10.1093/NAR/GKW159. URL /pmc/articles/PMC4937303//pmc/articles/PMC4937303/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4937303/.

J. Y. Wang, P. Pausch, and J. A. Doudna. Structural biology of CRISPR–Cas immunity and genome editing enzymes. *Nature Reviews Microbiology 2022*, pages 1–16, 5 2022. ISSN

1740-1534. doi: 10.1038/s41579-022-00739-4. URL https://www.nature.com/articles/s41579-022-00739-4.

B. Wienert, S. K. Wyman, C. D. Richardson, C. D. Yeh, P. Akcakaya, M. J. Porritt, M. Morlock, J. T. Vu, K. R. Kazane, H. L. Watry, L. M. Judge, B. R. Conklin, M. Maresca, and J. E. Corn. Unbiased detection of CRISPR off-targets in vivo using DISCOVER-Seq. *Science (New York, N.Y.)*, 364(6437):286, 4 2019. ISSN 10959203. doi: 10.1126/SCIENCE.AAV9023. URL /pmc/articles/PMC6589096//pmc/articles/PMC6589096/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC6589096/.

X. Wu, D. A. Scott, A. J. Kriz, A. C. Chiu, P. D. Hsu, D. B. Dadon, A. W. Cheng, A. E. Trevino, S. Konermann, S. Chen, R. Jaenisch, F. Zhang, and P. A. Sharp. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nature Biotechnology 2014 32:7*, 32(7):670–676, 4 2014a. ISSN 1546-1696. doi: 10.1038/nbt.2889. URL https://www.nature.com/articles/nbt.2889.

X. Wu, D. A. Scott, A. J. Kriz, A. C. Chiu, P. D. Hsu, D. B. Dadon, A. W. Cheng, A. E. Trevino, S. Konermann, S. Chen, R. Jaenisch, F. Zhang, and P. A. Sharp. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nature Biotechnology*, 32(7):670–676, 2014b. ISSN 15461696. doi: 10.1038/nbt.2889.

Z. Wu, A. Asokan, and R. J. Samulski. Adeno-associated Virus Serotypes: Vector Toolkit for Human Gene Therapy. *Molecular Therapy*, 14(3):316–327, 9 2006. ISSN 1525-0016. doi: 10.1016/J.YMTHE.2006.05.009.

J. Wysocka and W. Herr. The herpes simplex virus VP16-induced complex: the makings of a regulatory switch. *Trends in Biochemical Sciences*, 28(6):294–304, 6 2003. ISSN 0968-0004. doi: 10.1016/S0968-0004(03)00088-4.

R. Xu, R. Qin, H. Li, D. Li, L. Li, P. Wei, and J. Yang. Generation of targeted mutant rice using a CRISPR-Cpf1 system. *Plant Biotechnology Journal*, 15(6):713, 6 2017. ISSN 14677652. doi: 10.1111/PBI.12669. URL /pmc/articles/PMC5425385//pmc/articles/PMC5425385/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC5425385/.

X. Xu and L. S. Qi. A CRISPR–dCas Toolbox for Genetic Engineering and Synthetic Biology. *Journal of Molecular Biology*, 431(1):34–47, 1 2019. ISSN 10898638. doi: 10.1016/J.JMB.2018.06.037.

X. Xu, Y. Tao, X. Gao, L. Zhang, X. Li, W. Zou, K. Ruan, F. Wang, G. L. Xu, and R. Hu. A CRISPR-based approach for targeted DNA demethylation. *Cell Discovery 2016 2:1*, 2(1): 1–12, 5 2016. ISSN 2056-5968. doi: 10.1038/celldisc.2016.9. URL `https://www.nature.com/articles/celldisc20169`.

T. Yamano, H. Nishimasu, B. Zetsche, H. Hirano, I. M. Slaymaker, Y. Li, I. Fedorova, T. Nakane, K. S. Makarova, E. V. Koonin, R. Ishitani, F. Zhang, and O. Nureki. Crystal structure of Cpf1 in complex with guide RNA and target DNA. *Cell*, 165(4):949, 5 2016. ISSN 10974172. doi: 10.1016/J.CELL.2016.04.003. URL `/pmc/articles/PMC4899970//pmc/articles/PMC4899970/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4899970/`.

S. Yang and R. E. Rothman. PCR-based diagnostics for infectious diseases: uses, limitations, and future applications in acute-care settings. *The Lancet Infectious Diseases*, 4(6):337–348, 6 2004. ISSN 1473-3099. doi: 10.1016/S1473-3099(04)01044-8.

Y. Yang, J. Xu, S. Ge, and L. Lai. CRISPR/Cas: Advances, Limitations, and Applications for Precision Cancer Research. *Frontiers in Medicine*, 8:115, 3 2021. ISSN 2296858X. doi: 10.3389/FMED.2021.649896/BIBTEX.

R. M. Yarrington, S. Verma, S. Schwartz, J. K. Trautman, and D. Carroll. Nucleosomes inhibit target cleavage by CRISPR-Cas9 in vivo. *Proceedings of the National Academy of Sciences of the United States of America*, 115(38):9351–9358, 9 2018. ISSN 10916490. doi: 10.1073/pnas.1810062115.

R. Yin, C. K. Kwoh, and J. Zheng. Whole Genome Sequencing Analysis. *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, 1-3:176–183, 1 2019. doi: 10.1016/B978-0-12-809633-8.20095-2.

L. You, R. Tong, M. Li, Y. Liu, J. Xue, and Y. Lu. Advancements and Obstacles of CRISPR-Cas9 Technology in Translational Research. *Molecular Therapy: Methods & Clinical Devel-*

opment, 13:359–370, 2019. doi: 10.1016/j.omtm.2019.02.008. URL https://doi.org/10.1016/j.omtm.2019.02.008.

B. Zetsche, J. S. Gootenberg, O. O. Abudayyeh, I. M. Slaymaker, K. S. Makarova, P. Essletzbichler, S. E. Volz, J. Joung, J. Van Der Oost, A. Regev, E. V. Koonin, and F. Zhang. Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell*, 163(3):759–771, 10 2015. ISSN 10974172. doi: 10.1016/J.CELL.2015.09.038/ATTACHMENT/D7072A0B-0438-4D32-89BF-8FCD238CBAEA/MMC3.XLSX. URL http://www.cell.com/article/S0092867415012003/fulltexthttp://www.cell.com/article/S0092867415012003/abstracthttps://www.cell.com/cell/abstract/S0092-8674(15)01200-3.

D. Zhang, T. Hurst, D. Duan, and S. J. Chen. Unified energetics analysis unravels SpCas9 cleavage activity for optimal gRNA design. *Proceedings of the National Academy of Sciences of the United States of America*, 116(18):8693–8698, 4 2019. ISSN 10916490. doi: 10.1073/PNAS.1820523116/SUPPL{_}FILE/PNAS.1820523116.SAPP.PDF. URL www.pnas.org/cgi/doi/10.1073/pnas.1820523116.

S. Zhang, J. Shen, D. Li, and Y. Cheng. Strategies in the delivery of Cas9 ribonucleoprotein for CRISPR/Cas9 genome editing. *Theranostics*, 11(2):614–648, 2021. ISSN 1838-7640. doi: 10.7150/THNO.47007. URL https://pubmed.ncbi.nlm.nih.gov/33391496/.

X. H. Zhang, L. Y. Tee, X. G. Wang, Q. S. Huang, and S. H. Yang. Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. *Molecular Therapy - Nucleic Acids*, 4(11):e264, 1 2015. ISSN 2162-2531. doi: 10.1038/MTNA.2015.37.

Y. Zhang, H. Zhang, X. Xu, Y. Wang, W. Chen, Y. Wang, Z. Wu, N. Tang, Y. Wang, S. Zhao, J. Gan, and Q. Ji. Catalytic-state structure and engineering of Streptococcus thermophilus Cas9. *Nature Catalysis 2020 3:10*, 3(10):813–823, 9 2020. ISSN 2520-1158. doi: 10.1038/s41929-020-00506-9. URL https://www.nature.com/articles/s41929-020-00506-9.

Y. Zheng, J. Li, B. Wang, J. Han, Y. Hao, S. Wang, X. Ma, S. Yang, L. Ma, L. Yi, and W. Peng. Endogenous Type I CRISPR-Cas: From Foreign DNA Defense to Prokaryotic

Engineering. *Frontiers in Bioengineering and Biotechnology*, 8:62, 3 2020. ISSN 22964185. doi: 10.3389/FBIOE.2020.00062/BIBTEX.

X. Zhu, R. Clarke, A. K. Puppala, S. Chittori, A. Merk, B. J. Merrill, M. Simonović, and S. Subramaniam. Cryo-EM structures reveal coordinated domain motions that govern DNA cleavage by Cas9. *Nature Structural & Molecular Biology 2019 26:8*, 26(8):679–685, 7 2019. ISSN 1545-9985. doi: 10.1038/s41594-019-0258-2. URL https://www.nature.com/articles/s41594-019-0258-2.

R. S. Zou, A. Marin-Gonzalez, Y. Liu, H. B. Liu, L. Shen, R. Dveirin, J. X. J. Luo, R. Kalhor, and T. Ha. Massively parallel genomic perturbations with multi-target CRISPR reveal new insights on Cas9 activity and DNA damage responses at endogenous sites. doi: 10.1101/2022.01.18.476836. URL https://doi.org/10.1101/2022.01.18.476836.

J. A. Zuris, D. B. Thompson, Y. Shu, J. P. Guilinger, J. L. Bessen, J. H. Hu, M. L. Maeder, J. K. Joung, Z. Y. Chen, and D. R. Liu. Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nature Biotechnology 2014 33:1*, 33(1):73–80, 10 2014. ISSN 1546-1696. doi: 10.1038/nbt.3081. URL https://www.nature.com/articles/nbt.3081.

Chapter 2

Materials and Methods

2.1 Cell culture

Standard Operation Procedures (SOPs) have been followed for freezing, revival and maintenance of cells. All the cell lines used in this study were acquired from AZ Global Cell Bank (Alderley Park, UK) (see List of Reagents, Table 8.1) and are Mycoplasma free and authenticated by Short Tandem Repeat (STR) profiling. All cell culture work was performed in a laminar flow hood under appropriate aseptic conditions.

2.1.1 Cell counting

Cell counting was performed to determine an accurate number of viable cells in a given cell suspension, which was necessary in order to set up routinely maintenance of cell culture, electroporation, lentiviral transfections and FACS analysis. In general, an aliquot of cell suspension was diluted in proper media and the automated Vi-CELL™ XR Cell Viability Analyzer was used for counting. This system is based on the Trypan Blue Dye Exclusion method, which is a long-standing and widely used method to identify dead cells. Only cells with intact membranes can effectively exclude the dye, so dead cells with compromised membranes become stained.

2.1.2 Cell freeze and thaw

Frozen cell culture stocks were prepared for long term storage. Adherent cells were detached with Accutase® solution and collected (300 g x 5 min). Cells were washed with PBS and re-

collected (300 g x 5 min), then diluted in freshly prepared freezing serum (FBS + 10% DMSO). Usually, a typical concentration of 4x10^6 cells was frozen in 1 ml of freezing serum, in a 2 ml Corning Cryovial. Vials were frozen slowly using a polystyrene box at - 80 ℃, then moved after 24 hours at - 145 ℃ for long term storage. The thawing procedure, on the contrary, has to happen as fast as possible: vials were therefore set at 37 ℃ water bath for approximately 5 min, or until cells were thawed. Finally cells were resuspended in proper media and plated at the desired density in plate or flask.

2.2 Molecular Biology

2.2.1 Vector assembly and cloning

For sgRNA expression plasmid and lentivector preparation, Gibson assembly method was adopted, which allows for the joining of multiple DNA fragments in a single, isothermal reaction (Gibson et al., 2009). A Scarless reaction, using a single restriction enzyme (SfiI), allows to produce "Scarless" assemblies, whereby the joins between fragments have no unwanted artefacts left over from the DNA assembly process (See List of Reagents for guide sequences). The Gibson assembly mix (10 μl) and SfiI restriction enzyme (0.25 μl) were incubated with the receiving vector (60 ng) and the sgRNA (2.5 μM) (ordered as oligo with overlapping ends) for 15-60 min at 50 ℃. Heat shock transformation was performed in a competent E. Coli cells with 5 μl of the assembly according to the manufacturer's instructions. E. coli cells were recovered in 1 ml of SOC broth for 1 hour shaking at 1000 rpm. After transformation, 0.1 ml of the recovery culture was plated. After 24 hours, 1 colony was inoculated into 1.5 ml LB broth and incubate for 14-16 hours at 1000 rpm. Plasmid DNA was extracted using QIAprep Spin Miniprep Kit (QUIAGEN) according the manufacturer's instructions and Sanger sequencing performed on it. See List of Reagents for the list of vector used in this study.

2.2.2 Lentiviral transduction

Transfection of foreign DNA using lentiviral vector allows the generation of stable cell lines expressing a gene of interest. Unlike the short-term protein expression observed using transient transfection approaches, this procedure enables long-term protein expression studies. The specific lentivector for the expression of sgRNA (for CRISPRa with dCas9-VPR system) was assembled via a Gibbson assembly reaction and transfected in the host cells (HEK23-T) with additional two plasmids for virus packaging (psPAX2) and envelope (pMD2.G). The transfection mix (500 μl) was prepared mixing the expression vector (6 μg) with psPAX2 (3.6 μg), pMD2.G (1.2 μg) and Fugene6HD (30 μl) as transfection reagent. The mixture was incubated for 15 minutes at room temperature and added to 2x10^6 HEK293T cells in 10 cm plates. After 72 hours from the transfection, HEK293-T cell release the virus particles to the media which was collected, filtered through a 0.45-micron filter and stored at -80 ℃. Viral particles were

then used to infect the target cell lines. For cell infection, cells were diluted to 1x10^5/0.5 ml and placed in 6-well plates. 0.5 ml of diluted cells were mixed with 0.5 ml of viruses and 1 μl of polybrene. Six hours post transduction, 3 ml of media were added to each well. 72 hours post transduction, media was replaced with fresh media and cells kept in culture.

2.2.3 Doxycycline induction

Doxycycline is a synthetic tetracycline derivative that is the effector molecule for all Tet-On and Tet-Off Systems. These systems are inducible gene expression systems for mammalian cells that contain all the necessary components in a single plasmid, lentiviral, or retroviral vector. In this work, MCF7 cell line containing a Tet-On inducible system for dCas9 expression have been used for CRISPRa experiments. The cells express the Tet-On-3G transactivator protein and contain the dCas9 gene of under the control of a promoter with *tet* operator sequences. When bound by Doxycycline, the Tet-On 3G protein undergoes a conformational change that allows it to bind to *tet* operator sequences and to activate the transcription of the downstream dCas9 gene. Adherent cells were usually plated in 6-well plate, with a density of 3x10^5 cells per well. 100 ng/ml of Doxyccyline was dissolved in the proper media and added on top of the plated cells. Cells were collected after 48 hours of treatment.

2.2.4 Plasmid transfection

Vectors for sgRNA expression for CRISPRa (PCP/PH system) were previously described (Martella et al., 2019). Plasmid transfection results in transient transfection, which differs from that obtained by lentivirus transduction because the introduced nucleic acid exists in the cell only for a limited period of time and is not integrated into the genome. However, the high copy number of the transfected genetic material leads to high levels of expressed protein within the period that it exists in the cell. Depending on the construct used, transiently expressed transgene can generally be detected for 1 to 7 days, but transiently transfected cells are typically harvested 24 to 96 hours post-transfection. FuGENE® HD Transfection Reagent has been used for plasmids transfections in this work (CRISPRa with PCP/PH system) and FuGENE: DNA ratio has to be optimized for HEK293 (10C10 clone) at 3:1 ratio (1.5 μl reagent: 500 ng DNA). Cells were plated in a 6-well plate, with a density of 3x10^5 cells per well. After 24

hours, cells were washed with PBS and 1ml of OptiMem was added on top. 1.5 μl FuGene® HD was mixed with 500 ng DNA and OptiMem was added to the mixture up to 100 μl. The mixture was incubated for 10 minutes at room temperature and it was then slowly dripped on top of cells. Six hours post-transfection, media was replaced with fresh media. A mixture which contains a plasmid for GFP expression was prepared and transfected as control during each transfection.

2.2.5 Agarose Gel Electrophoresis

Conventional agarose gel electrophoresis was used to analyse DNA fragments. The electrophoretic mobility of DNA fragments mainly depends on the fragment size and to a lesser extent on the conformation of the DNA, type and concentration of agarose, as well as applied voltage and electrophoresis buffer. Agarose gels have greater range of separation and can resolve DNAs from 50 bp to 20 kbp in length. All agarose gels were prepared with SYBR™ Safe DNA for visualization of the DNA bands and were run in 1X TAE (Tris-Acetate-EDTA) buffer.

2.2.6 Genomic DNA extraction and PCR

Genomic DNA was extracted from cell cultures using Viagen DirectPCR® DNA Extraction System and amplified by direct PCR. For the reaction, 0.5 mg/ml of Proteinase K solution was added to DirectPCR Lysis Reagent (cell) solution, mixed and used for the lysis of cultured cells. Lysate was then incubated at 55 ℃ for 16 hours, warmed up to 85 ℃ for 45 minutes and then cooled down at 4 ℃. Genomic DNA was stored at -20 ℃. The typical PCR cycle has been run with the following program: 98 ℃, 10 minutes; (98 ℃, 1 minute; 55-60 ℃, 5 minutes; 72 ℃, 15 minutes) x 34 cycles; 72 ℃, 1 minute 4 ℃, ∞, varying the annealing temperature according to the primers. Analysis of PCR amplicon have been made in 1% agarose gel electrophoresis containing SYBR Safe DNA gel stain. All the Sanger sequencing reactions in this study were carried out by Eurofins.eu. See List of Reagents for list of primers for KO gene region amplification used in this study.

2.2.7 Total RNA extraction and qPCR

RNA extraction was performed using "RNAeasy Plus Mini Kit", following manufacturer's instructions. In this study, Quantitative Reverse Transcription PCR (qRT-PCR) has been adopted as a methodology to quantify the relative expression of the target genes compared to the expression of a housekeeping gene control (40S ribosomal protein S18). "GoScript Reverse Transcription System Kit" has been used for all qPCR reactions, following the manufacturer's instructions. The extracted RNA is reverse-transcribed into complementary DNA (cDNA) by reverse transcriptase. The cDNA is then used as a template for qPCR reaction which is done in the presence of TaqMan probes, labelled with a fluorescent dye on the 5' end. The use of fluorogenic probes enables the detection of a specific PCR product as it accumulates during PCR. By using the cDNA derived from genomic mRNA as a template for the assay and, by having one probe for the detection of a specific gene target and one probe for an internal control gene, it is possible to quantify the relative expression of mRNA of the target gene which was present in the starting cell pool. For the CRISPRa related studies, by repeating the quantification before and after the expression and activation of the CRISPRa system, it is possible to calculate the gene overexpression in term of fold change of the detected mRNA. Analysis of all the qRT-PCR reaction has been done using QuantStudio (TM) 6 and 7FlexRealTime PCR 1.3 software. See List of Reagents for TaqMan gene probes used in this study.

2.3 CRISPR-Cas9 editing in cells

2.3.1 Guide-RNA design

CRISPR 3 is an AZ in house tool developed by Quantitative Biology (QBio) team in AstraZeneca, and it allows for the design and investigation of guides for CRISPR experiments. The main viewing part of CRISPR3 is based around a graphical genome viewer called Genoverse. This was adapted from the Sanger Centre CRISPR design tool (Hodgkins et al., 2015) and further enhanced by QBio team with the calculation of the off-target probability (Qi et al., 2013). Given the target gene as imput, the software will display the results of the best guide designs and guide mappings. The best-predicted gRNAs for each gene KO to be tested in this study were chosen considering efficiency, confidence and off-target parameters prediction. Guide for CRISPRa are targeting the genomic region 200 bp upstream or downstream of the Transcription Starting Site (TSS). The on-target guide the Traffic Light Reporter was previously designed and tested in house, mismatched guides have been designed by changing one or more nucleotide at different positions, starting from the on-target guide sequence. Guides for the Proxy-system were designed to be 14 or 15 nt long and, targeting in proximity of TLR on-target guide, according to the PAM availability in the genomic sequence. See List of Reagents for the list of all guides used in this study.

2.3.2 RNP assembly

Cas9-gRNA complex was assembled *in vitro* prior to transfection into cells by electroporation. The two portions of the gRNA (crRNA and tracrRNA) were resuspended in a 1:1 ratio in Duplex DNA Buffer (IDT) and heated at 95℃ for 5 min, then cooled down gradually. This will result in the annealing of the two RNA portions. The full gRNA was then resuspended with purified Cas9 protein in a 1:1 ratio and left at room temperature for 15 minutes, which allows the formation of the complete RiboNucleoProtein (RNP). The RNP was added to the cells suspension and transfected into the cells following the electroporation protocol.

2.3.3 RNP electroporation

RNP electroporation for gene KO has been mainly conducted with the Neon Transfection System (ThermoFisher), which allows the use of a specific pipette with an electrode in the tip. The design of the electrode in pipette has been shown to produce a more uniform electric field which results in less toxicity to the cells and higher transfection efficiencies. The suspension of cells and DNA is loaded into the pipette and this is then connected to a power supply. Cells are subjected to a high-voltage electrical pulse of defined magnitude and length (depending on the cell line) and then directly placed in normal (non-selecting) cell growth medium to recover. Neon protocol has to be tested and adapted to each cell line used in this study. Generally, protocol n.24 has been used in this study for A549 and HEK cells, which applies 3 pulses with a pulse voltage of 1600 V and a pulse width of 10 ms. Protocol n.20 has been used for MCF7 cells, which applies 2 pulses with a pulse voltage of 1150 V and a pulse width of 30 ms. For Proxy-TRL experiments, RNP electroporation was performed with SE Cell Line 96-well NucleofectorTM Kit using Amaxa™ 96-well Shuttle™ (program CM120 for U2OS cells). Cells were resuspended in proper media and kept in culture. In general, $2.5x10^5$ cells were collected for each electroporation and mixed with final 1μM RNP.

2.3.4 TIDE analysis

TIDE analysis (Tracking of Indels by Decomposition) (http://shinyapps.datacurators.nl/tide/) is based on Sanger sequencing of the target which is supposed to be targeted by the nuclease (Figure 2.1) (Brinkman et al., 2014). TIDE requires as input a control sequence data file (in most cases here is genomic DNA from cells electroporated without RNP), a sample sequence data file (which is genomic DNA from a pool of cell electroporated with RNP) and a character string representing the gRNA sequence (20 nt). TIDE first aligns the gRNA sequence to the control sequence to determine the position of the expected Cas9 break site (Figure 2.1). Next, the control sequence region upstream of the break site is aligned to the experimental sample sequence in order to determine any offset between the two sequence reads (Figure 2.1). The software uses the peak heights for each base, as determined by the sequence analysis software to determine the relative abundance of aberrant nucleotides over the length of the whole sequence trace. Subsequently, the TIDE software decomposes the composite sequence trace into its

individual components by means of multivariate non-negative linear modelling, with the control sequence trace serving as a template to model the individual indel components (Figure 2.1). This decomposition results in an estimate of the relative abundance of every possible indel within a chosen size range. It is worth noting that the results quality depends on the quality of the sequence reads. Software creators recommend aiming for an average aberrant sequence signal strength before the break site <10% (both control and test sample), and R2 > 0.8 for the decomposition result.

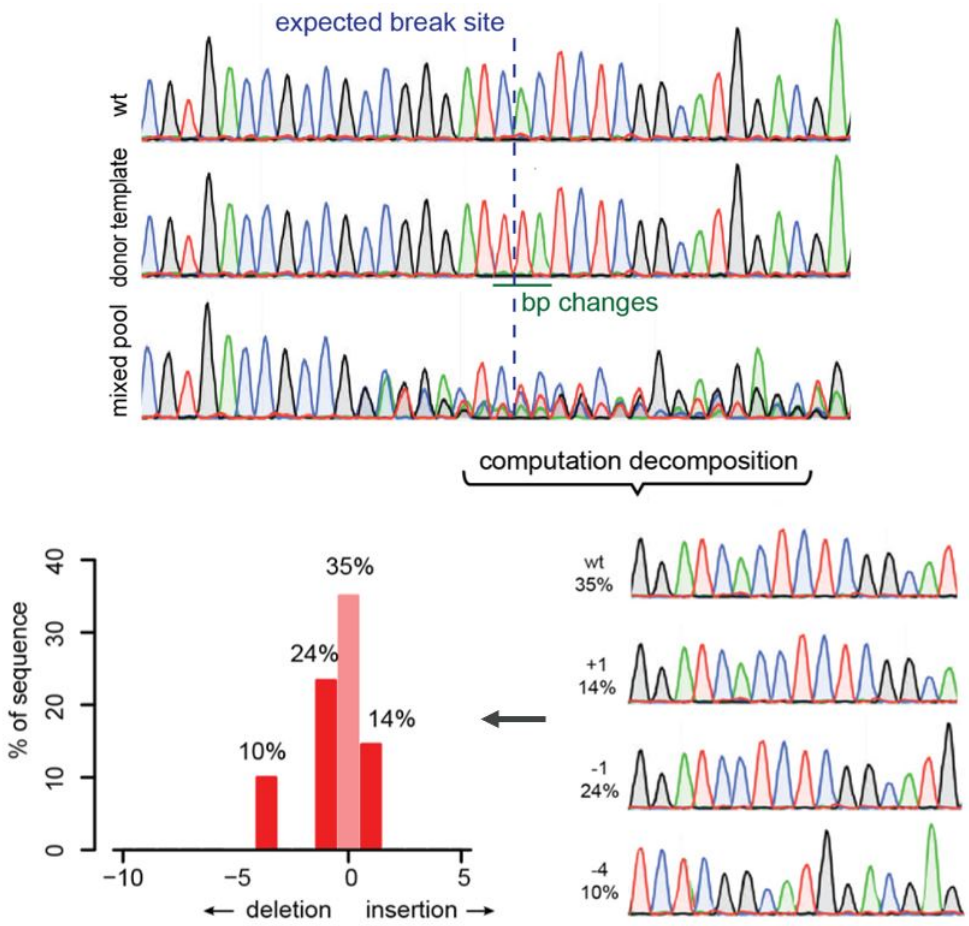


Figure 2.1: TIDE assay worklow. TIDE can precisely determine the spectrum and frequency of targeted mutations generated in a pool of cells by genome editing tools such as CRISPR-Cas9, but also TALENs and ZFNs. TIDE requires only standard molecular biology reagents and reaction such as standard PCR and standard capillary sequencing. By providing a wt-control sequence file and a donor template sequence file (crRNA) to detect the expected break site, the algorithm accurately reconstructs the spectrum of indels on the sample mixed pool traces. The web tool reports the identity of the detected indels and their frequencies. Image from http://shinyapps.datacurators.nl/tide/.

2.3.5 FACS analysis

Fluorescence-Activated Cell Sorting (FACS) is a technique based on flow cytometry which utilizes laser-based technology to count, sort, and profile cells in a heterogeneous fluid mixture (McKinnon, 2018). Using a flow cytometer machine, cells suspended in a liquid stream are passed through a laser light beam in single file fashion, and the interaction with the light is measured by an electronic detection apparatus as light scatter and fluorescence intensity. If a fluorescent label, or fluorochrome, is specifically and stoichiometrically bound to a cellular component, the fluorescence intensity will ideally represent the amount of that particular cell component. FACS analysis allows to simultaneously collect data on a biological sample and sort a specific population. For FACS analysis, adherent cells were detached Accutase® solution, harvested (300 *g* x 5 min), resuspended in FACS Buffer (RPMI medium, 5-10% FBS, GlutaMax supplement, 1% Pen/Strep) and filtered through a 70 μm or 100 μm nylon mesh (BD Falcon™ cell strainers). BD FACSMelody™ Cell Sorter has been used for all the experiments in this study and FlowJo-v10.6.1 software has been used for data analysis.

2.3.6 Traffic Light Reporter

U2OS cells with an adapted Traffic Light Reporter (Certo et al., 2011) cassette integrated in their genome were available in house. Within this reporter a double-strand break is produced at a specific nuclease cleavage site, and repair of the break generates distinct fluorescent signals upon resolution of the DNA damage either through HDR with an exogenous donor template or through NHEJ (Figure 2.2) (Certo et al., 2011). In the former case, to signal gene targeting, the open reading frame of a functional enhanced GFP (eGFP) is restored by providing an exogenous donor template (Porteus and Baltimore, 2003); in the latter case, to signal gene disruption, two out-of-frame monomeric (m)Cherry coding sequences (+/-1 or +/-2 bp frameshift) are placed back in-frame via an indels-caused frameshift (Certo et al., 2011). By design, the eGFP coding sequence contains a non-functional alternative +3 reading frame, and the T2A or P2A 'dis-linker' which enables the downstream-encoded mCherry to escape degradation of the misfolded protein encoded in this +3 reading frame of eGFP.

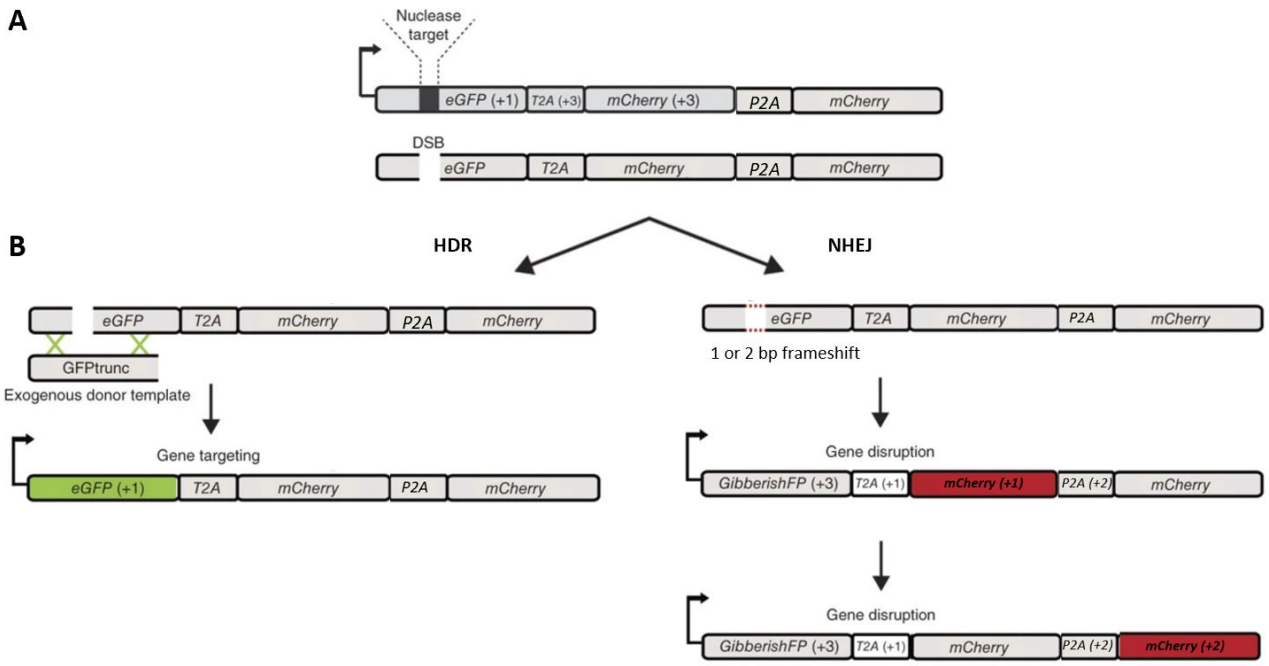


Figure 2.2: Traffic Light Reporter schematic. **(A)** Diagram of the TLR. Arrow represents promoter and initial eGFP start codon. Reading frames relative to the initial eGFP start codon are indicated in parentheses. **(B)** Schematic showing the two different engineering outcomes after the induction of a site specific double-strand break (DSB). If the break is resolved through the HDR pathway, the full eGFP sequence will be reconstituted, and cells will fluoresce green; if the break undergoes NHEJ, eGFP will be translated out of frame (gibberishFP, +3 reading frame) and the T2A/P2A and mCherry sequences are rendered in frame to produce red fluorescent cells. Image adapted from Certo et al. 2011.

2.4 AZ-Cas9 *in vitro* cleavage assay

For AZ-Cas9 study an *in vitro* cleavage assay was performed to check that the cleavage activity was not affected by the labelling of the gRNA. RNP was assembled *in vitro* as previously described (section). 10 nM of labelled DNA oligo containing the target sequence ($\lambda 2$) were incubated with 100 nM RNP at 37°C x 30 minutes total, in Digest Buffer (100 mM TRIS-HCl pH7.5, KCl 500 mM, 25% glycerol, 5 mM DTT, 50 mM $MgCl_2$). Every 5 minutes, the reaction was stopped with Quenching Solution (0.5 M EDTA, pH 8). Cleavage products were then analysed via gel electrophoresis in 1% agarose gel.

2.5 DNA and RNA synthesis, purification and labelling

2.5.1 PAGE purification

The integrity of all DNA and RNA for single molecule experiments were first checked by denaturing polyacrylamide gel electrophoresis (8 M urea, 15% 29:1 polyacrylamide: bis-acrylamide in 1 x Tris-Borate-EDTA buffer) with gels run at 20 W for 2 hours. Where additional bands below the full length product were observed, samples were purified by denaturing PAGE with extraction by crush and soak in elution buffer (40 mM Tris-HCl pH 8, 480 mM NaOAc pH 5, 0.5 mM EDTA) followed by ethanol precipitation.

2.5.2 Labelling of DNA and RNA

For in-house labelling of DNA and RNA, oligos were ordered from IDT with modified amino-linker bases (iAmMC6T, IDT) at the required sites for labelling and resuspended in ddH_2O. Mono-reactive Cy3 or Cy5 dye (GE Healthcare) was resuspended in DMSO (14 μl, 99%, SIGMA). Resuspended oligos (50 μl, 100 μM) were incubated in $NaHCO_3$ buffer (43 μl, 0.5 M, pH 9) with resuspended dye (7 μl) for at least two hours at 4°C. Unreacted dye was removed by ethanol precipitation, and labeled oligos were resuspended in ddH_2O (50 μl).

2.5.3 HPLC purification

After labelling, labelled DNA and RNA oligos were purified by HPLC (Discovery BIO Wide Pore C8 5μm column, 568323-U) in TEAA buffer (50 mM acetonitrile, 2.8 mM triethylamine, adjusted to pH 7 with glacial acetic acid) to remove the unlabelled population. Collected fractions containing the labelled population were vacuum dried and finally purified by ethanol precipitation. Where required, labelled oligos were further purified by denaturing PAGE.

2.6 Optical tweezer experiments

2.6.1 Biotinylation of phage λ-DNA

For the majority of experiments biotinylated λ-phage DNA, λ-DNA, was prepared by filling in the 5' 12-bp overhangs with biotinylated nucleotides (Figure 2.3) (Gross et al., 2010). Per 50 μl reaction: in a PCR tube, λ-DNA (21 μl, 500 μg) was combined with biotin-CTP (B-dCTP, 10 µl, 40 μM), biotin-ATP (B-dATP, 10 μl, 40 μM), dTTP (0.5 μl, 10 mM), dGTP (0.5 μl, 10 mM), NEB Buffer 2 (5 μl, 10X, NEB), ddH_2O (3 μl), and finally Klenow Polymerase (0.05 μl, 50U/μl). Pipetting of any solutions containing λ-DNA was done slowly and carefully with wide bore pipettes to reduce mechanical nicking of the DNA. This was transferred to a PCR machine and run with the following program: 37℃, 30 minutes; 70℃, 15 minutes; 4℃, ∞. This was then purified using QIAquick PCR Purification Kit (QIAGEN), eluting in 60 μl elution buffer and quantified using a BioDrop μLite+. Final concentration was typically 30 μg/ml. Biotinylated DNA was stored at 4℃ for > 1 month or aliquots flash frozen in liquid nitrogen and stored at -20℃, freeze thawing cycles were minimised to prevent nicking.

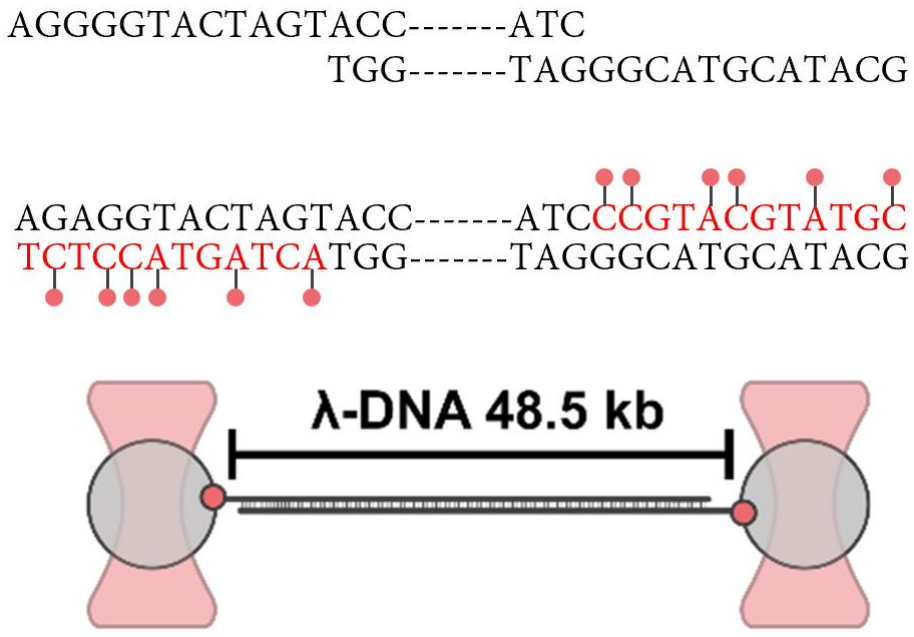


Figure 2.3: Biotinylation of phage-λ DNA. Klenow Polymerase reaction was prepared with biotinylated dCTP and dATP (biotin represented with red circle in the schematic) and used to fil the 5' overhangs of λ-DNA molecules. Beads coated with streptavidin can be then used to tether dsDNA molecule which will remain unconstrained.

2.6.2 C-TRAP microscope

All of the optical tweezers experiments were performed on C-trap® STED microscope (Figure 2.4) (LUMICKS A.G., Amsterdam). For confocal imaging, 532 nm laser was used for Cy3 excitation, with emission detected in three channels with blue (512(25) nm), green (585(75)) nm and red (640LP) filters. Experiments were performed at an equilibrium temperature of 30℃ measured at the objectives.

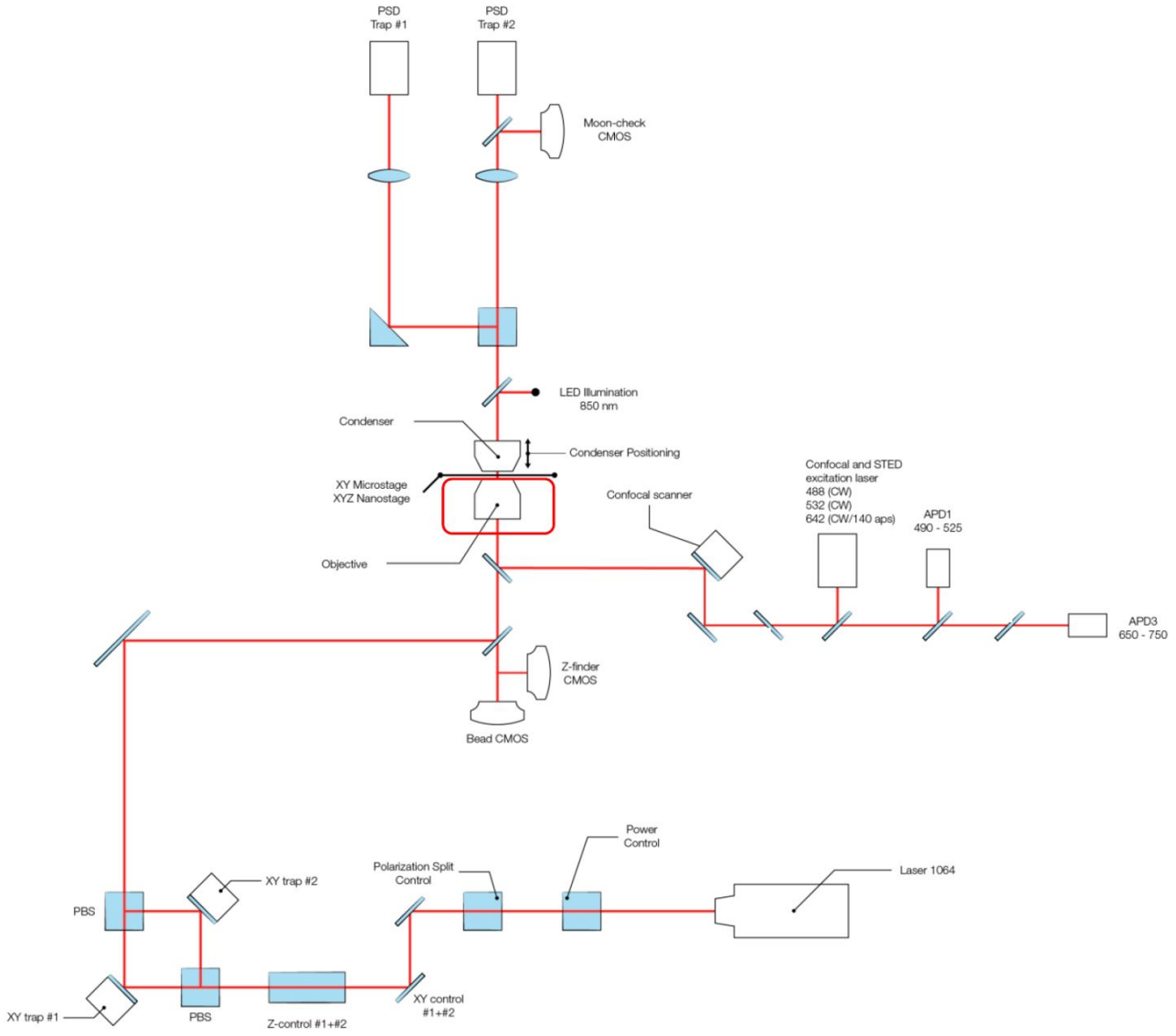


Figure 2.4: Optical trapping, Microscope and Confocal systems. The optical trapping system consist of a laser source (Laser 1064 nm wavelength), the optical power management systems (power control and polarization split), XY steering mirrors and a Force detector system (a LED illumination for camera imaging, a polarized beam splitter and two Position Sensitive Detectors - PSD - trap #1 and #2). The Microscopy system includes the objective, condenser and sample positioning system (XY microstage and XY-Z nanostage). The Confocal system consists of a laser source (3-channel light generator with 3 wavelenghts), fluorescence detectors (ADPs) and a confocal scanner.

2.6.3 Cleaning and passivation of the fluidics and flow cell

The syringes, tubing and flow cell (Figure 2.5) was cleaned with bleach (1 ml, 5%, NaClO), followed by ddH_2O (1 ml), sodium thiosulphate (1 ml, 100 mM in PBS) and finally ddH_2O, each flowed over approximately 10-15 minutes, with a pressure of 2 bar. All channels were passivated with Pluronic F128 (1 ml, 0.5% w/v in PBS) flowed over 30 minutes at 1 bar and protein channels further passivated with BSA (1 ml, 0.1% w/v in PBS) flowed over 30 minutes at 0.5 bar.

Figure 2.5: Microfluidics chip layout used in this study. A glass micro-fabricated chip (70 x 14 x 1.3 mm, 100 μm channel height) with the shown schematic was used in all the C-trap experiments in this study. Channel 1 was used to flow streptavidin coated beads. Channel 2 was used to flow and catch single biotinylated λ-DNA molecules. Channel 3 was used to flow buffer only and check the force-distance curve of the caught dsDNA molecule. Generally, Channel 4 was used to flow fluorescent labelled RNP in Ca^{2+} buffer and used to study the DNA-protein binding interaction. Channel 5 was used to flow fluorescent labelled RNP in Mg^{2+} buffer and used to study Cas nuclease activity on DNA.

2.6.4 DNA force-extension curve

Before each experiment, the force-distance plot for single DNA molecule tethered and stretched between the beads has been checked (Figure 2.6). Indeed, when a single dsDNA molecule is streatched, a characteristic and specific-force extension curve can be recorded (Figure 2.6). DNA elasticity can be explained using the worm-like chain (WLC) model (Kratky and Porod, 1949), which describes the elastic properties of a semi-flexible polymer through its persistence length Lp. This parameter describes the characteristic length scale of a polymer chain beyond which the molecule remains linear. The force-extension relation of dsDNA at low forces (<5 pN), is described by an interpolating formula of the WLC model. The approximation is given by:

$$L_{DNA} = L_c \left(1 - \frac{1}{2}\sqrt{\frac{K_B T}{F L_p}}\right) \tag{2.1}$$

Where L_c is the DNA molecule contour length (maximum end to end distance), F is the applied force, K_B is the Boltzman constant (4.11 pN.nM) and T is the temperature in Kelvin. To extend this model to the high-force regime, the extensible WLC model (eWLC) takes into account contributions from the stretch modulus (S) of the DNA, resulting in the following force–extension relation:

$$L_{DNA} = L_c \left(1 - \frac{1}{2}\sqrt{\frac{K_B T}{F L_p}} + \frac{F}{S}\right) \tag{2.2}$$

When the contour length, L_c, of the DNA is higher than the distance between the beads, the molecule is disordered and flexible. When the distance between the beads becomes similar to L_c, the DNA is straightened out, with energy required to overcome this loss of entropy (Figure 2.6). By increasing the distance over the contour length, the DNA molecule is stretched, with small changes in distance resulting in a large increase in force (Figure 2.6) (Bustamante et al., 2021). At very high forces (>50 pN) the DNA is overstretched and the base pairs

start to breakdown, with consequentially melting of the dsDNA, which will finally pass from double to single stranded DNA (Figure 2.6). Since the elastic properties of DNA have been well characterized, eWLC model optical tweezers can then be used in various experimental configurations to investigate DNA processing enzymes and other protein–DNA interactions (Bustamante et al., 2021).

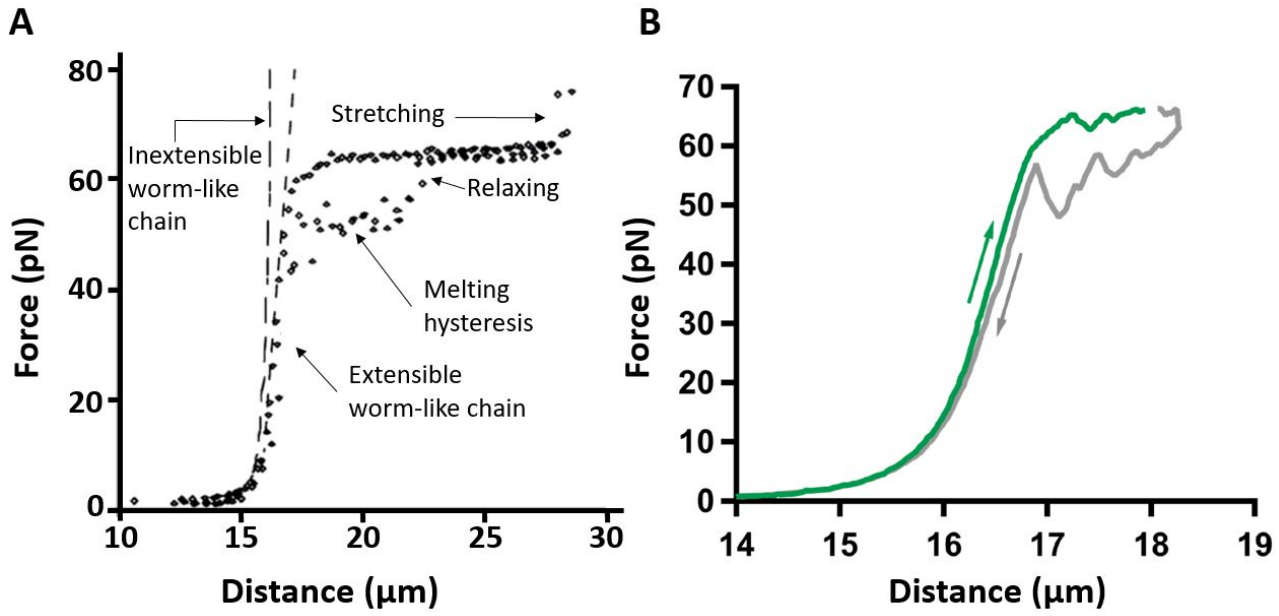


Figure 2.6: DNA force–extension curves. **(A)** Schematic of typical force–extension curves generated from force-ramp experiments. These curves can be used to analyse the elastic properties of the tethered DNA by fitting them to theoretical models such as the extensible worm-like-chain (eWLC) model of polymer elasticity. Image adapted from Bustamante et al. 2021. **(B)** Typical experimental force-extension curve, generated in this study, of a single λ-DNA molecule stretched (green) from 0-65 pN and relaxed (grey) down to 0 pN.

2.6.5 Cas12a and AZ-Cas9 optical tweezer experiments

Cas12a (IDT) and AZ-Cas9 were stored at -80℃ and 5 μM in storage buffer (20mM Tris-HCl, pH 7.5, 150 mM KCl, 10% glycerol and 1 mM tris(2-carboxyethyl)phosphine. Working fractions were aliquoted, snap frozen in liquid nitrogen. Experiments were performed in a microfluidics flow cell (Figure 2.5). SPHERO 4.89 μm Streptavidin Coated polystyrene beads were diluted to 0.005% w/v and flowed in channel 1 (Figure 2.5). Biotinylated λ-DNA was diluted to 30 ng/ml and flowed in channel 2 (Figure 2.5). The two part guide was annealed by mixing crRNA and tracrRNA in annealing buffer (100 mM NaCl, 50 mM Tris-HCl pH 8), heating to 90℃ for 2 minutes, and cooled at room temperature for 15 minutes before being put on ice. Cy3- crRNA (250 nM) was complexed with Cas12a or AZCas (either 250nM or 1μM) at room temperature for 30 min in complexing buffer (50 mM Tris-HCl, pH 7.5, 100 mM or 25 mM NaCl, 10 mM $CaCl_2$ or $MgCl_2$, 0.2 mg.ml1 BSA, 0.5 mM DTT), and subsequently diluted to 250 pM with either Ca^{2+}-buffer for binding experiments (50 mM Tris–HCl, pH 7.5, 100 mM or 25 mM NaCl, 10 mM CaCl2, 0.2 mg.ml1 BSA, 0.5 mM DTT, 5 mM protocatechuic acid (PCA), 100 nM protocatechuic dioxygenase (PCD) and saturated Trolox) or Mg^{2+}-buffer for cutting experiments (50 mM Tris-HCl, pH 7.5, 100 mM or 25 mM NaCl, 10 mM $MgCl_2$, 0.2 mg.ml1 BSA, 0.5 mM DTT, 5 mM PCA, 100 nM PCD and saturated Trolox). Before each experiment, the protein channel was further passivated with the RNP complex by flowing 0.5 ml over 30 minutes at 0.2 bar.

2.6.6 Data analysis

All data was exported as '.h5' files and analysed in python using the 'lumicks. pylake' python package and custom scripts. Images were extracted as .tiff files and either analysed with python or using FIJI to extract intensity profiles and quantify fluorescence. For diffusion analysis, a combination of custom python (https://github. com/singlemoleculegroup) and Igor Pro 8 (Wavemetrics) scripts has been used. Cas12a diffusion events were tracked and quantified using the python single particle tracking algorithm, and resulting trajectories were further processed in Igor Pro 8 to determine velocity trajectories and mean squared displacement plots.

2.7 TIRF smFRET

2.7.1 TIRF microscope

SmFRET experiments were performed on a home built prism-based total internal reflection microscope, TIRM, built around and Olympus IX71 microscope with 532 nm excitation (Vortran Stradus, US) and 542 nm long pass emission filter (Chroma 542LP ET). The acceptor and donor emission is split with a beam splitter containing a 635 nm long pass dichroic mirror (Chroma T635LPXR) and detected on an Andor iXon Ultra EMCCD camera.

2.7.2 Slides and fluorophore calibration

All the slides for FRET analysis were prepared in house from members of Rueda Lab. For calibration of the mapping of the Cy3 and Cy5 channels for smFRET TIRF experiments, a slide was prepared containing a low density of fluorescent beads with emission in both the Cy3 and Cy5 channels (FluoSpheres Carboxylate-Modified Microspheres, 0.2 μm, red fluorescence (580/605), 2% w/v). An ultra-low fluorescence quartz slide was cleaned with ddH_2O, methanol, again with ddH_2O and dried with N_2. A single layer of double sided tape was used to create a 2 cm wide channel, with a clean coverslip attached on top. Diluted beads were incubated in the channel for 15 minutes before sealing with epoxy glue. Before each smFRET TIRF experiment, a short video was recorded with $>$ 100 beads in the field of view and used to map the Cy3 and Cy5 signals when extracting smFRET traces (Zhao and Rueda, 2009).

2.7.3 SmFRET experiments

Constructed slide chambers were passivated and functionalised by incubation with 0.2 mg.ml-1 BSA in T50 buffer (50 mM Tris-HCl, pH 7.0, 50 mM NaCl) for 10 minutes, followed by a wash with T50 and incubation with Neutravidin (0.2 mg.ml-1 in T50 for 10 minutes. This was washed out with T50 then replaced with imaging buffer (100 mM NaCl, 50 mM Tris-HCl pH 8, 1mM $MgCl_2$, 0.2 mg.ml-1 BSA, 125 mM PCA, 10 μM PCD and saturating Trolox) ready for sample injection. Cy3 labelled DNA for immobilisation was diluted to 10 pM before injection into the slide chamber and incubated for 10 minutes before washing with imaging buffer. DNA was imaged on TIRF microscope to check density before injection of RNP complex (200 μl, 10

nM RNP in imaging buffer) and data collection.

2.7.4 Data acquisition and analysis

Videos were recorded with a 60 ms exposure time and 150 gain of 600x. Custom IDL scripts were used for alignment mapping of the Cy3 and Cy5 channels, background subtraction, molecule detection and extraction of single-molecule trajectories. Single-molecule trajectories were analysed using custom MATLAB or R scripts.

Bibliography

E. K. Brinkman, T. Chen, M. Amendola, and B. Van Steensel. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Research*, 42(22):e168–e168, 12 2014. ISSN 0305-1048. doi: 10.1093/NAR/GKU936. URL https://academic.oup.com/nar/article/42/22/e168/2411890.

C. J. Bustamante, Y. R. Chemla, S. Liu, and M. D. Wang. Optical tweezers in single-molecule biophysics. *Nature reviews. Methods primers*, 1(1), 12 2021. doi: 10.1038/S43586-021-00021-6. URL /pmc/articles/PMC8629167//pmc/articles/PMC8629167/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC8629167/.

M. T. Certo, B. Y. Ryu, J. E. Annis, M. Garibov, J. Jarjour, D. J. Rawlings, and A. M. Scharenberg. Tracking genome engineering outcome at individual DNA breakpoints. *Nature methods*, 8(8):671, 8 2011. ISSN 15487091. doi: 10.1038/NMETH.1648. URL /pmc/articles/PMC3415300//pmc/articles/PMC3415300/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3415300/.

D. G. Gibson, L. Young, R. Y. Chuang, J. C. Venter, C. A. Hutchison, and H. O. Smith. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods 2009 6:5*, 6(5):343–345, 4 2009. ISSN 1548-7105. doi: 10.1038/nmeth.1318. URL https://www.nature.com/articles/nmeth.1318.

A. Hodgkins, A. Farne, S. Perera, T. Grego, D. J. Parry-Smith, W. C. Skarnes, and V. Iyer. WGE: a CRISPR database for genome engineering. *Bioinformatics*, 31(18):3078–3080, 9 2015. ISSN 1367-4803. doi: 10.1093/BIOINFORMATICS/BTV308. URL https://academic.oup.com/bioinformatics/article/31/18/3078/240857.

O. Kratky and G. Porod. Diffuse small-angle scattering of x-rays in colloid systems. *Journal of Colloid Science*, 4(1):35–70, 2 1949. ISSN 0095-8522. doi: 10.1016/0095-8522(49)90032-X.

A. Martella, M. Firth, B. J. Taylor, A. Göppert, E. M. Cuomo, R. G. Roth, A. J. Dickson, and D. I. Fisher. Systematic Evaluation of CRISPRa and CRISPRi Modalities Enables Development of a Multiplexed, Orthogonal Gene Activation and Repression System. *ACS Synthetic Biology*, 8(9):1998–2006, 2019. ISSN 21615063. doi: 10.1021/acssynbio.8b00527.

K. M. McKinnon. Flow Cytometry: An Overview. *Current protocols in immunology*, 120:5.1.1, 2 2018. ISSN 1934368X. doi: 10.1002/CPIM.40. URL /pmc/articles/PMC5939936//pmc/articles/PMC5939936/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC5939936/.

M. H. Porteus and D. Baltimore. Chimeric nucleases stimulate gene targeting in human cells. *Science*, 300(5620):763, 5 2003. ISSN 00368075. doi: 10.1126/SCIENCE.1078395/SUPPL{_}FILE/PORTEUS.SOM.PDF. URL https://www.science.org/doi/full/10.1126/science.1078395.

L. S. Qi, M. H. Larson, L. A. Gilbert, J. A. Doudna, J. S. Weissman, A. P. Arkin, and W. A. Lim. Repurposing CRISPR as an RNA-γuided platform for sequence-specific control of gene expression. *Cell*, 152(5):1173–1183, 2013. ISSN 00928674. doi: 10.1016/j.cell.2013.02.022.

R. Zhao and D. Rueda. RNA folding dynamics by single-molecule fluorescence resonance energy transfer. *Methods*, 49(2):112–117, 10 2009. ISSN 1046-2023. doi: 10.1016/J.YMETH.2009.04.017.

Chapter 3

Effect of local DNA distortion on Cas9 off-target activity in cells

3.1 Introduction

Since the discovery of CRISPR-Cas9, and its implication for genome editing purposes, concerns about Cas9 off-target activity have presented a barrier to therapeutic applications. Moreover, despite the routine use of CRISPR-Cas technology, its efficiency and specificity cannot entirely be predicted. This is a central challenge for the use of this technology both in basic research and in clinical applications. Two major areas of research are focusing on the most relevant information that would impact our ability to precisely engineer any region of the DNA, namely how nucleotide mismatches impair Cas9 off-target binding and cleavage (Feng et al., 2021; Listgarten et al., 2018), and why some genomic regions are associated with different on-target editing efficiency (Knight et al., 2015; Verkuijl and Rots, 2019). This first part of my work aims to achieve a better understanding on the extent of the effect of nucleotide mismatches on Cas9 off-target binding and cleavage in cells, and how this is related to DNA tridimensional conformation within the cellular context. My work follows an extensive study (Newton et al., 2022, manuscript submitted to Molecular Cell), which includes single-molecule Cas9 activity characterization on negatively supercoiled λ-DNA and next-generation sequencing to detect negative supercoiling induced off-targets on cell-free DNA.

3.2 Results

3.2.1 Proxy-CRISPR and Traffic Light Reporter System

To understand if Cas9 efficiency and specificity is directly dependent on perturbation of the local DNA structure, I designed a strategy which combines a specific and local opening of the dsDNA with an easy readout of Cas9 cleavage at a proximal target site. I started from previous data, which showed that the binding of non-cleavage dCas9 complexes to sites proximal to the intended cleavage site increases the cleavage efficiency of wtCas9 at the on-target site (Proxy-CRISPR, Chen et al. 2017), although the mechanism underlying this increased efficiency is still not well understood. In my study, I modified and adapted the Proxy-CRISPR system by using a wtCas9 (instead of dCas9) in complex with a short gRNA (14-15 nt), which can still guide Cas9 to the specific proximal sites but prevents any nuclease cleavage activity (Kiani et al., 2015; Dagdas et al., 2017), therefore mimicking the effect achieved with the dCas9 in the original Proxy-CRISPR system (Figure 3.1). The main advantage of using a short guide is that it should prevent undesirable cutting event at the proximal site, which can occur in case of "guide swapping" between dCas9 and wtCas9. In addition, the use of a wtCas9 for the Proxy-CRISPR system allows to proceed with a single electroporation of two RNPs complexes, one functioning as a Proxy-CRISPR and the other targeting the actual site, reducing the toxic effects on cells caused by a sequential, double electroporation. To test Cas9 cleavage efficacy and calculate the KO efficiency, I utilised the Traffic-Light DSB repair Reporter (TLR) system (Materials and Methods, Certo et al. 2011) which enables single-cell quantitative detection of Cas9 cutting activity through a fluorescent readout (Figure 3.1). Following cleavage by Cas9 at the target site in the TLR, the repair pathway by non-homologous end-joining (NHEJ) results in a frame shift of an out-of-frame Red Fluorescent Protein (RFP) cassette, leading to RFP re-expression (Figure 3.1). The percentage of RFP expressing cells, can be quantified by fluorescence aided cell sorting (FACS) as a measure of Cas9 efficiency.

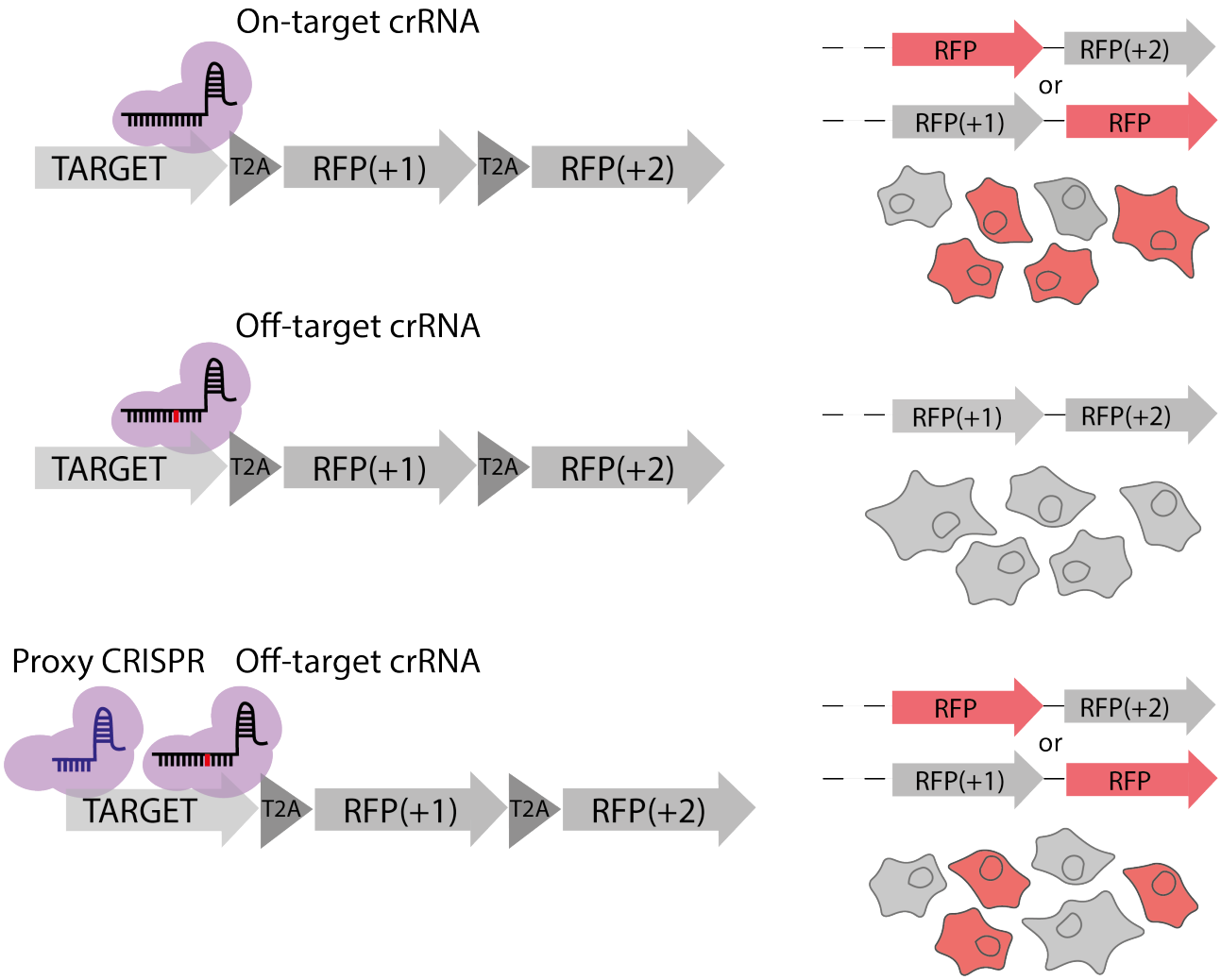


Figure 3.1: Schematic of Proxy-CRISPR on the Traffic-Light Reporter (TLR). Cas9 is targeted to the TLR site with a fully-matching gRNA generating a DSB. Repair of the break by NHEJ generates a frameshift which places a monomeric RFP coding sequence in-frame, and a red fluorescent signal allows for detection of Cas9 activity and gene disruption. When Cas9 is complexed with a mismatched gRNA, mimicking the off-target activity, gene editing is not detected. The local opening of the DNA double helix via Proxy-CRISPR system, enhances Cas9 off-target activity, as shown from the fluorescent signal.

3.2.2 Proximal DNA structure distortion enhances Cas9 off-target in cells

3.2.2.1 System and Method design for TLR

I first validated the TLR system in U2OS-TLR cells (cell line engineered by AZ Cell engineering team and available in house). Cas9 RNP was assembled *in vitro* with a fully matching gRNA, previously tested in house (List of Reagents) targeting the TLR cassette, then electroporated into the cells with 96-well Nucleofector™ Kit using the Amaxa™ 96-well Shuttle™ (Materials and Methods). Cells were analysed by FACS, 72 hours after electroporation, confirming an observed editing efficiency of 35% (Figure 3.2). Next, for the off-target study, I designed a number of mismatched guides, containing from 1 up to 4 mismatches at different positions (List of Reagents). Moreover, for the guides containing a single mismatch, I designed guides variants with all the three possible type of mismatches at the same position (List of Reagents). The efficiency of those guides was overall diminished compared to the fully-matched guide, and, this was correlated to the position, the type and the number of mismatches, being the guides with mismatches in the seed region or with an overall high number of mismatches, less or totally not efficient (Figures 3.3,3.4). Contextually, I designed 7 guides to be tested as Proxy-CRISPR (guides P1 to P7). In this assay, the design of the short gRNAs is mainly limited by the presence of PAMs sequences and the distance from the actual cleavage target site. To extend my analysis and be able to test more proxy-sites, I designed five guides, P1 to P5, which can recognise a sequence flanked by the canonical NGG PAM and two more guides, P6 and P7, which can recognise sequences flanked by a non-canonical PAM (ACG and AAG), shown to have a relatively high efficiency among the non-canonical PAM (Hsu et al., 2013; Kleinstiver et al., 2015). The distance of the short PROXY-gRNAs from the cleavage site ranges from 18 to 59 bp PAM to PAM, being positioned either upstream or downstream the cleavage site. In fact, it is important to consider that if the two Cas9:gRNA complexes are too close to each other, steric hindrance can compromise their activity and, at the same time, if the Proxy-CRISPR binds too far away from the active Cas9, its effect on proximal DNA site can be also limited. I then tested the short Proxy-gRNAs by electroporating U2OS-TLR cells – as described above - and confirmed that those guides cannot cleave their targeting site (Figure 3.2 A).

3.2.2.2 System and Method validation on TLR

I sequentially tested if the short guides can effectively function as Proxy-CRISPR and cause an enhancement in the editing efficiency of the TLR-KO site. Briefly, RNPs were assembled *in vitro* with the KO guide and the Proxy-guide separately, then electroporated simultaneously into the cell with a single electroporation. After 72 hours, cells were collected and analysed by FACS analysis for RFP expression quantification. To start with, I co-transfected the fully matched guide or one of the guide with 1 mismatch (5A>C, Figure 3.2 B) together with the short guides, one at a time. Results show that Proxy P1 outperforms among the 7 designed short guides, being able to enhance the efficiency of the mismatched guide, which now almost reaches the editing efficiency of the fully matching guide, going from 19% to 30% (Figure 3.2 B). P1 is 14 nucleotides long, and it recognises a sequence 37 bp upstream from the cleavage site, which is flanked by a canonical NGG PAM. Altogether, this appears to be the best combination of factors which can lead successful Proxy-CRISPR effect at the site of interest. Other short guides, P4, P6 and P7, also enhanced the editing efficiency of the same mismatched guide (Figure 3.2 B), therefore, as an additional test, I also tried to co-transfect those guides in combination with P1, hoping that they could act synergically and increase the overall Proxy-CRISPR effect (Figure 3.2 C). However, results show no enhancement of the Proxy-CRISPR effect with the combination of two short guides since I obtain the same results given from P1 alone (Figure 3.2 C). Hence, I performed the next experiments by using P1 to study the Proxy-CRISPR effect on Cas9 activity. Co-transfection of P1 with guides with a single mismatch at different positions leads to a significant increase in editing efficiency of up to 2-fold for 12 out of 15 guides tested (Figure 3.4). However, I cannot observe any meaningful increase in editing efficiency for guides with more than one mismatches (Figure 3.3), with the exception of one guide with 2 mismatches (5A>C, 8C>G, Figure 3.3). Interestingly, the efficiency of the fully-matching guide is also not increased by the Proxy-CRISPR system (Figures 3.3, 3.4), probably due to the fact that this guide is already performing at the highest possible efficiency for that target site. Together these data demonstrate that in cells, the local DNA deformation caused by the Proxy-CRISPR system can reduce Cas9 specificity by enabling an increase in mismatch tolerance and this effect is dependent on the position and the type of mismatch.

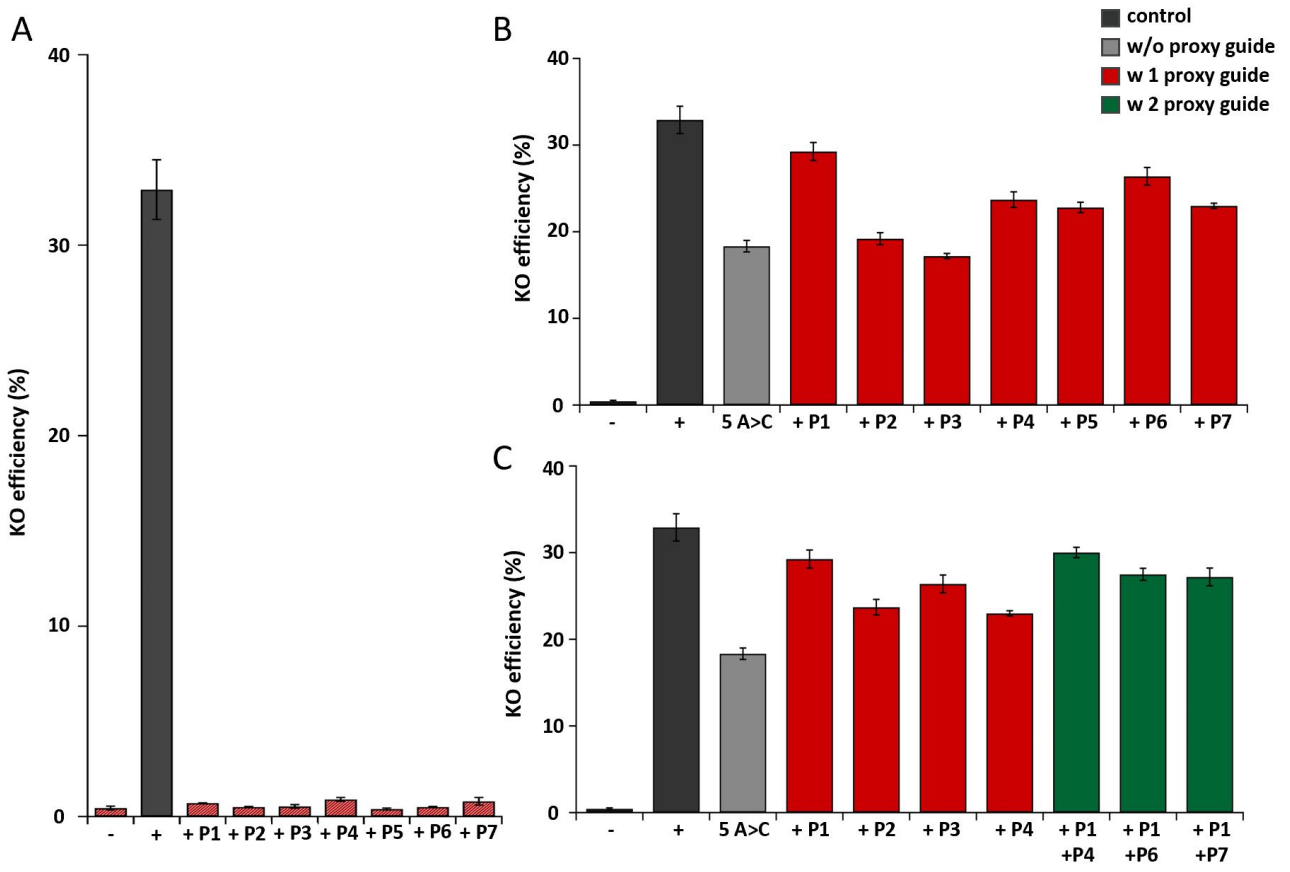


Figure 3.2: Proxy-CRISPR effect on TLR KO efficiency with a guide containing 1 mismatch. (−) Negative control: wtCas9 only, without any guide. (+) Positive control: wtCas9 with the fully matching guide. **(A)** Control: KO efficiency of Proxy-guides. wtCas9 was assembled *in vitro* with each Proxy-guide separately, then electroporated into the cell and editing efficiency quantified by FACS after 72 hours. **(B)** Proxy-CRISPR effect on the editing efficiency of a guide with 1 mismatch (5A>C). wtCas9 was assembled with the mismatched KO guide and with each Proxy-guide separately, then electroporated simultaneously with a single electroporation, one Proxy-guide at the time. The editing efficiency of the mismatched guide was quantified by FACS after 72 hours. Co-transfection with Proxy P1 leads to an enhanced efficiency up to 30%. **(C)** Synergistic Proxy-CRISPR effect on the editing efficiency of a guide with 1 mismacth (5A>C). wtCas9 was assembled with the mismatched KO guide and with each Proxy-guide separately, then electroporated simultaneously with a single electroporation, with a combination of 2 Proxy-guides at the time (n=3).

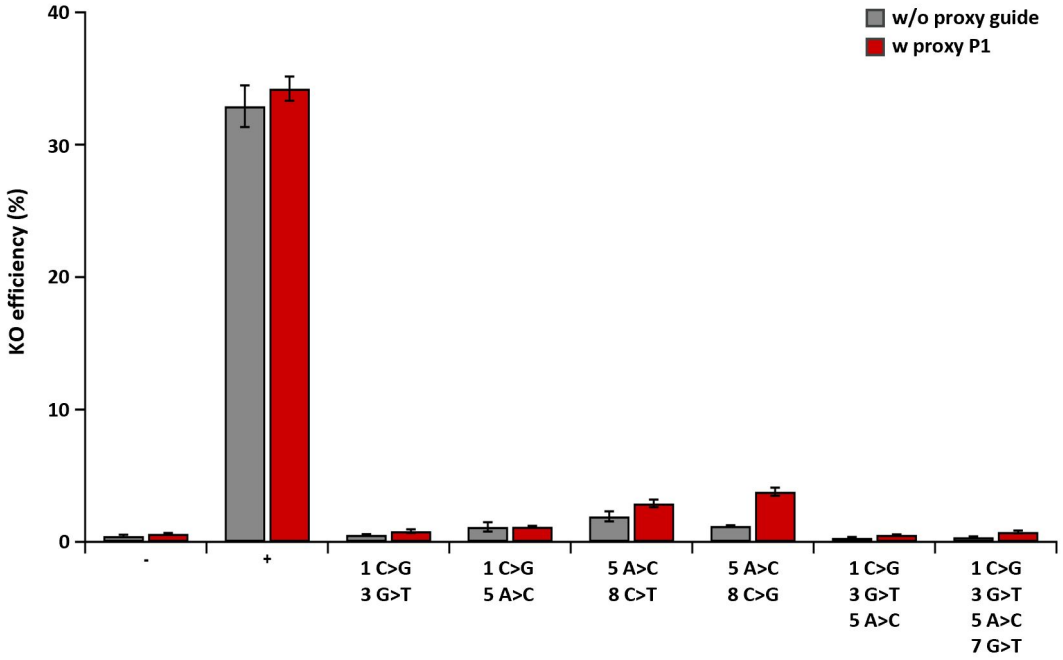


Figure 3.3: Proxy-CRISPR effect on TLR KO efficiency with guides containing more than 1 mismatch. (−) Negative control: wtCas9 only, without any guide. (+) Positive control: wtCas9 with the fully matching guide. wtCas9 was assembled *in vitro* with each mismatched KO guide (2, 3 or 4 mismatches, at the indicated position) and with P1-Proxy-guide separately, then electroporated simultaneously with a single electroporation, one mismatched guide at the time. The editing efficiency of each mismatched guide was quantified by FACS after 72 hours (n=3).

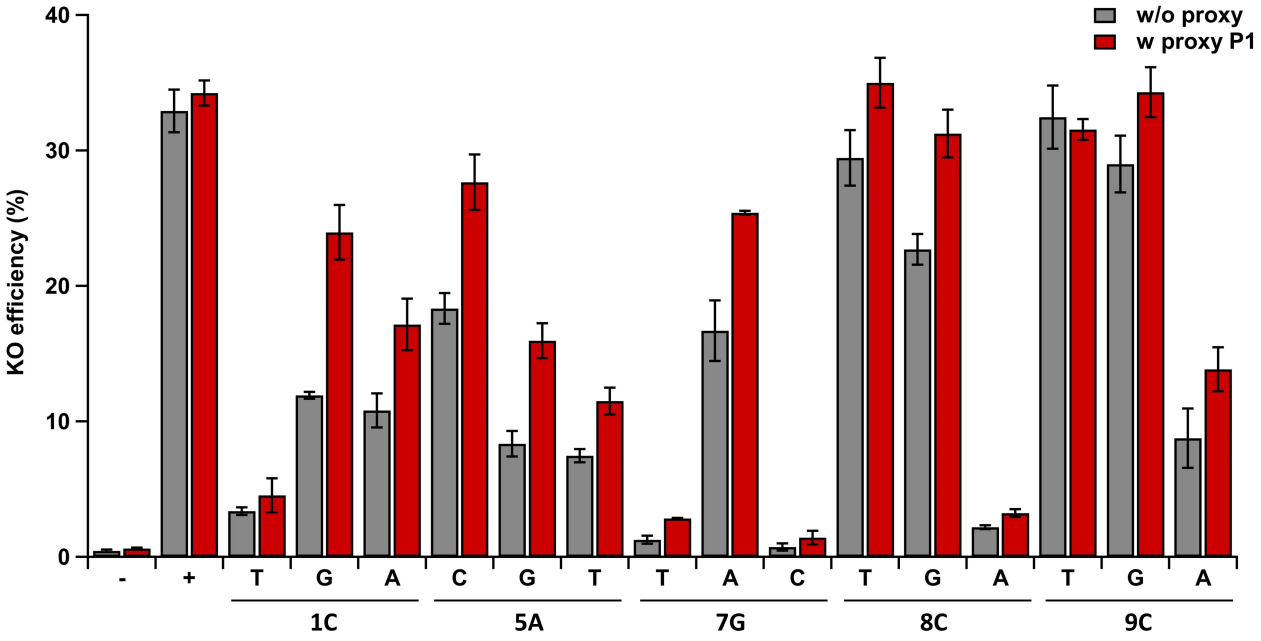


Figure 3.4: Proxy-CRISPR effect on TLR KO efficiency with guides containing 1 mismatch at different positions. (−) Negative control: wtCas9 only, without any guide. (+) Positive control: wtCas9 with the fully matching guide. wtCas9 was assembled *in vitro* with each mismatched KO guide (1 mismatch at the indicated position from the PAM) and with P1-Proxy-guide separately, then electroporated simultaneously with a single electroporation, one mismatched guide at the time. The editing efficiency of each mismatched guide was quantified by FACS after 72 hours (n=3).

3.2.2.3 System and Method validation on the endogenous target EMX1

In order to further confirm these results, I then tested this system on the endogenous target EMX1. I designed 3 guides for gene KO and I first checked the guide efficiency (Figure 3.5) by transfecting them via electroporation in U2OS cells, through the Neon Electroporator System (Material and Methods). This time, in the absence of the TLR, the KO efficiency was calculated via DNA genomic extraction (72 hours from electroporation), Sanger sequencing and TIDE analysis, as previously described (Materials and Methods). Among the 3 designed guides (List of Reagents), guide G3 outperformed over guides G1 and G2, editing the EMX1 target with an efficiency of ∼80% (Figure 3.5). I then designed a total of 15 mismatched guides starting from the G3 crRNA sequence, all containing one mismatch in the seed region (position 2, 5, 7, 8 and 11 from the PAM), and with all the possible type of mismatch at the same position. All the mismatched guides showed a diminished editing efficiency compared to the fully matching guide (Figure 3.6). I also designed 2 Proxy-guides (P1 and P2), both targeting a proximal site upstream the KO target site and respectively 70 bp and 27 bp far from it (PAM to PAM). I then tested P1 and P2, co-transfecting them separately with all the 15 mismatched guides one at the time (Figure 3.6). Results show an overall enhancement in KO editing efficiency of 11 out of the 15 tested mismatched guides with both the Proxy-guides.

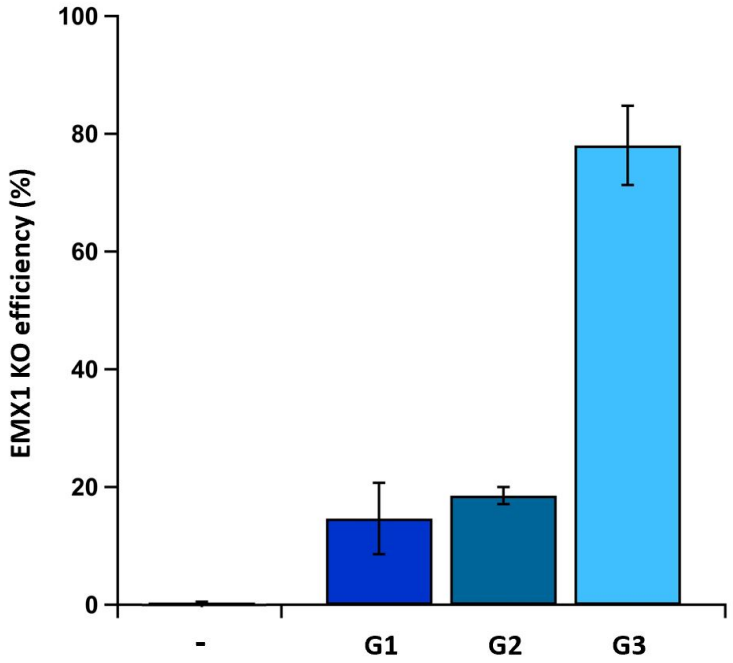


Figure 3.5: EMX1 KO efficiency guide test. (−) Negative control: wtCas9 only, without any guide. wtCas9 was assembled *in vitro* with each KO guide, then electroporated in U2OS cells separately. Genomic DNA was extracted after 72 hours, sequenced and editing efficiency was quantified by TIDE analysis (n=4).

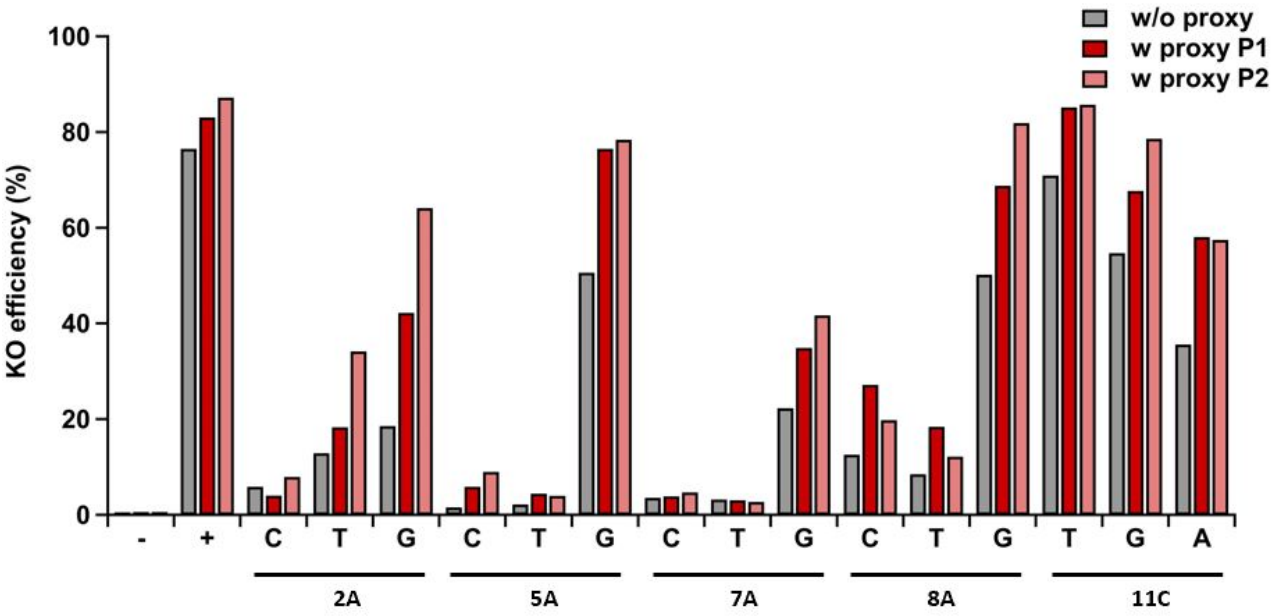


Figure 3.6: Proxy-CRISPR effect on EMX1 KO efficiency with guide containing 1 mismatch at different positions. (−) Negative control: wtCas9 only, without any guide. (+) Positive control: wtCas9 with the fully matching guide. wtCas9 was assembled *in vitro* with each mismatched KO guide (1 mismatch at the indicated position from the PAM) and with 2 Proxy-guides separately, then electroporated simultaneously with a single electroporation, one mismatched guide at the time with each Proxy-guide. Genomic DNA was extracted after 72 hours, sequenced and editing efficiency was quantified by TIDE analysis (Representative experiment).

3.3 Discussion

In order to confirm in the cellular context the molecular mechanism of recent findings on the effect of DNA distortion on Cas9 activity (Newton et al., 2022. Manuscript submitted to Molecular Cell), here I designed and tested an approach that allows to measure off-targets effect in cells in the presence of local distortion of DNA proximal to the on-target site. In Newton et al., novel results, obtained *in vitro* via optical tweezers experiments, demonstrate that the increment of negative supercoiling levels at the target DNA reduces the specificity of Cas9. In the same study, a next-generation sequencing approach based on the previously described CIRCLE-seq method (Tsai et al., 2017), has been used to test whether negative supercoiling in DNA could induce Cas9 off-target cleavage on cell-free DNA. For those experiments, Cas9 was assembled with EMX1 targeting guide and incubated with negatively supercoiled genomic DNA, obtained via gyrase treatment. Results reported over 13,000 off-target sites compared to samples in the absence of gyrase.

With this work, I aimed to confirm this finding in cells. Here, I describe the use of an adapted version of the Proxy-CRISPR system to recreate a local modification of the dsDNA structure via Cas9 binding to test the effect on off-target activity. To simplify the method and at the same time prevent undesirable cutting events at the proximal site, I describe for the first time the use of wtCas9 in complex with short gRNAs (14-15 nt) to target the proximal site in a Proxy-CRISPR approach. I tested this strategy both on an artificial reporter system (TLR) and on an endogenous gene (EMX1). The artificial reporter is integrated in the genome of the cell and therefore the results can be contextualised in an endogenous phenotype context. In both cases, the results corroborate the work *in vitro* showing that the local distortion of DNA affects Cas9 off-target activity specifically.

In my experiments, I find that the distortion caused by the Proxy-system, does not affect the on-target activity of Cas9, which was already in part discussed in Chen et al. 2017, where the Proxy system was first described. In that study, a dCas9 proximal binding was producing an effect on the on-target editing only with some gRNAs, while with other gRNAs targeting the same gene but at different loci, the Proxy-effect was diminished and insufficient. Together with my findings, this observation suggest that the distortion of DNA has a significant effect on the off-target recognition process specifically, while the on-target is less dependent or that other

factors are involved.

Interestingly, these results are in line with the hypothesis that Cas9 specificity and off-target discrimination can be better described by using a kinetic model wherein target cleavage is a function of the energy cost of DNA melting versus the energy recovered through RNA-DNA base pairing and cost of incorporation of mismatches. In such a model, DNA destabilisation and distortion can be thought to tilt the energy landscape, favouring target cleavage even in the presence of a significant energy cost from the accommodation of mismatches. Indeed, in my experiments, depending on the number, the position and the type of mismatch, a local opening of the double helix is proved to reduce Cas9 specificity by enabling an increased mismatches tolerance. However, the correlation between editing efficiency and both the mismatch identity and the mismatch position, further highlights the complexity of determining the underlying rules governing Cas9 specificity.

Additionally, by analysing in detail the type of mismatches and base-pairs used in the TLR experiments, and comparing them to the CIRCLE-seq data from Newton et al. (2022), we made important observations. For example, substitutions at positions 8 and 9 resulting in a dG-U base pairing (Figure 3.4, 8 C > U and 9 C > U) are very well tolerated, and this is in agreement with recent structural studies of Cas9 complexes bound to mismatched targets, revealing that wobble-base pairing and protonated wobble-base pairing are both well accommodated (Pacesa et al., 2021). This is also confirmed in the CIRCLE-seq data, which shows, in absence of supercoiling, that most mismatches involve the formation of wobble pairs (G-U/T or A-C). However, this requirement seems to be relaxed when the DNA is negatively supercoiled or distorted, and in my results, even the accommodation of purine-purine mismatches is observed. Indeed, mismatches at positions 1, 8 and 9 in TLR, and position 11 in EMX1, resulting in rG-dG base pairing, are some of the most highly tolerated (Figure 3.4).

Taken together, these findings elucidate the mechanism at the base of Cas9 off-target discrimination, adding further proofs on the recent hypothesis that it is highly affected by DNA topolology. This has to be considered and taken into account for future development of safer and more efficient CRISPR-tools.

Bibliography

M. T. Certo, B. Y. Ryu, J. E. Annis, M. Garibov, J. Jarjour, D. J. Rawlings, and A. M. Scharenberg. Tracking genome engineering outcome at individual DNA breakpoints. *Nature methods*, 8(8):671, 8 2011. ISSN 15487091. doi: 10.1038/NMETH.1648. URL /pmc/articles/PMC3415300//pmc/articles/PMC3415300/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3415300/.

F. Chen, X. Ding, Y. Feng, T. Seebeck, Y. Jiang, and G. D. Davis. Targeted activation of diverse CRISPR-Cas systems for mammalian genome editing via proximal CRISPR targeting. *Nature Communications*, 8, 4 2017. ISSN 20411723. doi: 10.1038/ncomms14958.

Y. S. Dagdas, J. S. Chen, S. H. Sternberg, J. A. Doudna, and A. Yildiz. A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. *Science Advances*, 3(8), 8 2017. ISSN 23752548. doi: 10.1126/SCIADV.AAO0027. URL /pmc/articles/PMC5547770//pmc/articles/PMC5547770/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC5547770/.

H. Feng, J. Guo, T. Wang, C. Zhang, and X. H. Xing. Guide-target mismatch effects on dCas9–sgRNA binding activity in living bacterial cells. *Nucleic Acids Research*, 49(3):1263–1277, 2 2021. ISSN 0305-1048. doi: 10.1093/NAR/GKAA1295. URL https://academic.oup.com/nar/article/49/3/1263/6121456.

P. D. Hsu, D. A. Scott, J. A. Weinstein, F. A. Ran, S. Konermann, V. Agarwala, Y. Li, E. J. Fine, X. Wu, O. Shalem, T. J. Cradick, L. A. Marraffini, G. Bao, and F. Zhang. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology 2013 31:9*, 31(9):827–832, 7 2013. ISSN 1546-1696. doi: 10.1038/nbt.2647. URL https://www.nature.com/articles/nbt.2647.

S. Kiani, A. Chavez, M. Tuttle, R. N. Hall, R. Chari, D. Ter-Ovanesyan, J. Qian, B. W. Pruitt, J. Beal, S. Vora, J. Buchthal, E. J. Kowal, M. R. Ebrahimkhani, J. J. Collins, R. Weiss, and G. Church. Cas9 gRNA engineering for genome editing, activation and repression. *Nature methods*, 12(11):1051, 11 2015. ISSN 15487105. doi: 10.1038/NMETH.3580.

URL /pmc/articles/PMC4666719//pmc/articles/PMC4666719/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4666719/.

B. P. Kleinstiver, M. S. Prew, S. Q. Tsai, V. V. Topkar, N. T. Nguyen, Z. Zheng, A. P. Gonzales, Z. Li, R. T. Peterson, J. R. J. Yeh, M. J. Aryee, and J. K. Joung. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature 2015 523:7561*, 523(7561):481–485, 6 2015. ISSN 1476-4687. doi: 10.1038/nature14592. URL https://www.nature.com/articles/nature14592.

S. C. Knight, L. Xie, W. Deng, B. Guglielmi, L. B. Witkowsky, L. Bosanac, E. T. Zhang, M. E. Beheiry, J. B. Masson, M. Dahan, Z. Liu, J. A. Doudna, and R. Tjian. Dynamics of CRISPR-Cas9 genome interrogation in living cells. *Science*, 350(6262):823–826, 2015. ISSN 10959203. doi: 10.1126/science.aac6572.

J. Listgarten, M. Weinstein, B. P. Kleinstiver, A. A. Sousa, J. K. Joung, J. Crawford, K. Gao, L. Hoang, M. Elibol, J. G. Doench, and N. Fusi. Prediction of off-target activities for the end-to-end design of CRISPR guide RNAs. *Nature biomedical engineering*, 2(1):38, 1 2018. ISSN 2157846X. doi: 10.1038/S41551-017-0178-6. URL /pmc/articles/PMC6037314//pmc/articles/PMC6037314/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC6037314/.

M. Pacesa, C.-H. Lin, A. Cléry, K. Bargsten, M. J. Irby, F. H. Allain, P. Cameron, P. D. Donohoue, and M. Jinek. Structural basis for Cas9 off-target activity. *bioRxiv*, page 2021.11.18.469088, 11 2021. doi: 10.1101/2021.11.18.469088. URL https://www.biorxiv.org/content/10.1101/2021.11.18.469088v1https://www.biorxiv.org/content/10.1101/2021.11.18.469088v1.abstract.

S. Q. Tsai, N. T. Nguyen, J. Malagon-Lopez, V. V. Topkar, M. J. Aryee, and J. K. Joung. CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets. *Nature methods*, 14(6):607–614, 5 2017. ISSN 1548-7105. doi: 10.1038/NMETH.4278. URL https://pubmed.ncbi.nlm.nih.gov/28459458/.

S. A. Verkuijl and M. G. Rots. The influence of eukaryotic chromatin state on CRISPR–Cas9 editing efficiencies, 2 2019. ISSN 18790429.

Chapter 4

Effect of gene-wide DNA distortion on Cas9 on-target activity in cells

4.1 Introduction

To investigate further the intracellular Cas9 behaviour, the following chapter describes a strategy designed to study the impact of open chromatin in actively transcribed genes on Cas9 functionality, directly in the cellular context. Indeed, with regard to our knowledge on the CRISPR system, the question remains open about how the genomic and chromatin contest affects and dictates Cas9 binding and cleavage events (Daer et al., 2017; Verkuijl and Rots, 2019). In the previous chapter, I show that a local DNA distortion can highly affect and increase Cas9 off-target activity. However, I also observe that Cas9 on-target activity does not seem to be affected when using the Proxy-CRISPR system (Figure 3.4, 3.6). Hence, I designed a method to analyse Cas9 nuclease activity on epigenetically "open" or "closed" genes, and study if and how active transcription influences the on-target activity specifically. Indeed, based on the evidence that non- or low- expressed genes are in many cases characterized by closed chromatin (heterochromatin) features along with both their promoter and coding regions (Grewal and Moazed, 2003), and Cas9 has less accessibility at those sites (Daer et al., 2017), I hypothesize the KO efficiency can be enhanced by forcing gene expression. I, therefore, decided to combine both CRISPRa (for gene upregulation) and CRISPRn (nuclease active) tools to test if the overexpression of a selected target gene leads to enhanced Cas9 KO efficiency.

4.2 Results

4.2.1 Correlation between gene expression level and gene KO efficiency

I designed a gene-wide approach and developed a method to open the chromatin structure at the promoter site, using CRISPRa tools, and to test the modulation of Cas9 on-target activity on that target-gene. By forcing the endogenous expression of epigenetically silenced targets that are poorly accessible to Cas9 nuclease, the whole gene chromatin should be remodelled in a process necessary for the gene transcription to happen. This strategy would allow study of the effect of chromatin remodelling on Cas9 activity directly in cells. To this end, I initially selected low expressed or silenced target genes that show low Cas9 editing efficiency. For this analysis, I consulted data from two available databases. I first examined the single-cell RNA sequencing dataset accessible through the Human Protein Atlas (https://www.proteinatlas.org/), which is based on a meta-analysis of the literature on single-cell RNA sequencing and single-cell databases that include healthy human tissue. This includes single-cell transcriptomics data for 25 tissues and peripheral blood mononuclear cells (PBMCs) – for a total of 69 different cell lines – and datasets retrieved from the Single Cell Expression Atlas, the Human Cell Atlas, the Gene Expression Omnibus, the Allen Brain Map, and the European Genome-phenome Archive. I then also analysed the gene expression level through Cancer Cell Line Encyclopedia (CCLE) data using the DepMap portal (https://depmap.org/portal/), which contains data from CRISPR knockout screens from project Achilles, as well as genomic characterization data from the CCLE project. This dataset is based on CCLE RNAseq gene expression data for 1019 cell lines. To identify genes with different expression levels in at least two cell lines, I looked over the RNA level per cell line of each potential target, starting from some targets which were already in study in house and cell lines available in house (Table 4.1). A selected target gene will show i) different expression patterns in different cell lines, ii) correlation between expression level and Cas9 nuclease activity, and iii) will be suitable to overexpression using CRISPRa. As a control, I also chose a gene that has similar expression levels across the cell lines (SRP14, Table 4.1).

Table 4.1: Selected target genes for the study.

Gene	Cell line	Expression level - Protein Atlas (nTPM)	Expression level - 22Q1 Public (nTPM)
CXCR4	A549	0.3	1.2
	JURKAT	323.6	337
	MCF7	1.5	42.2
	HEK293	6.9	0.0
	RH-30	62.6	51.9
	SH-SY5Y	26.1	207
SRP14	A549	163.7	216.7
	JURKAT	297.7	256
	MCF7	204.5	290
	HEK293	230.3	359.5
	RH-30	176.8	221.3
	SH-SY5Y	255.0	439.5
MET	A549	115.3	89.8
	JURKAT	0.2	1.2
	MCF7	5.0	7.8
	HEK293	23.4	59.7
	RH-30	95.3	30
	SH-SY5Y	0.0	2
ASCL1	A549	0.0	1.1
	JURKAT	0.0	0.0
	MCF7	0.0	5.6
	HEK293	0.0	0.0
	RH-30	0.0	0.0
	SH-SY5Y	29.0	10.5

Where nTPM: normalized Transcipt Per Milion

To test the correlation between gene expression level and gene KO efficiency in the cell lines available in house (A549, MCF7, HEK293 and SH-SY5Y) I designed and tested 1-4 gRNAs for each gene (Table 4.1, List of Reagents for sequences). To first analyze the guides efficiency and the gene-specific KO efficiency in the cell lines, the RNP complexes (Cas9:gRNA) were assembled *in vitro* with each guide and transfected into the cells via electroporation through the Neon Transfection System (Materials and Methods). Cells were then cultured for 48 hours, genomic DNA extracted, and PCR amplification performed for the specific expected cutting site (Materials and Methods). After sequencing, Cas9 KO efficiency was calculated by TIDE analysis (Materials and Methods). From the results, SRP14, which is highly expressed at a similar level in A549, MCF7, and HEK293 cell lines (Table 4.1) is cut by Cas9, with the 2 designed guides, with similar and relatively high editing efficiency (e.g., G2: 36% to 53%) in all the 3 cell lines tested (Table 4.2, Figure 4.1). On the other side, the lineage-specific gene CXCR4 is lowly expressed in A549 and MCF7, while it is highly expressed in Jurkat cells (Table 4.1). The editing results on this target gene show indeed a correlation between the expression level and the gene-editing efficiency; by using the same gRNA (G2, Table 4.2), I obtain only 12.6% editing efficiency in A549 and 4% editing efficiency in MCF7, while I obtain up to 66.2% editing efficiency in Jurkat cells (Figure 4.1). An additional example is MET, which shows medium or high expression levels in HEK293 and A549 cells where I detect 62.3% and 56.5% KO efficiency (Table 4.2, Figure 4.1). According to the consulted dataset, MET is low expressed in SH-SY5Y cells, and here I can detect a drop in the KO efficiency to 2% by using the same gRNA (G2, Table 1.2). Nonetheless, during the KO tests, I found a gRNA performing with high editing efficiency (65.8%) for ASCL1 target site in HEK293 cells (G1, Table 1.2), although this gene is not highly expressed in this specific cell line. I then questioned whether the gene overexpression would have an additional effect on its KO efficiency, and I decided to further test the CRISPRa system on this target in this cell line. It is also worth noticing that several guides for each gene were designed (Table 4.2). However, some guides (e.g. G1, G3, and G4 for CXCR4) have poor KO efficiency even in the presence of high gene expression, suggesting that guide design is a key factor in gene-editing strategy.

Table 4.2: Correlation between gene expression level and KO efficiency

Gene target	tested gRNA	Cell line	Gene expression level	KO efficiency (%)	TIDE R^2 value
CXCR4	G1	A549	0	2.6	0.98
		JURKAT	++++	5.9	0.8
	G2	A549	0	12.7	0.97
		JURKAT	++++	66.2	0.86
		MCF7	+	4.1	0.98
	G3	A549	0	1.3	0.99
		JURKAT	++++	1.3	0.98
	G4	A549	0	1.1	0.99
		JURKAT	++++	1.3	0.99
MET	G1	A549	++	11.4	0.79
		HEK293	+	9.5	0.85
	G2	A549	++	45.9	0.8
		HEK293	+	54.8	0.79
		SH-SY5Y	0	2	0.79
	G3	A549	++	49.8	0.97
		HEK293	+	66.7	0.95
SRP14	G1	A549	++	24	0.99
		HEK293	++	52	0.96
		MCF7	++	55	0.8
	G2	A549	++	36.4	0.9
		HEK293	++	52.3	0.95
		MCF7	++	57	0.87
ASCL1	G1	SH-SY5Y	++	-	-
		HEK293	0	65.8	0.88

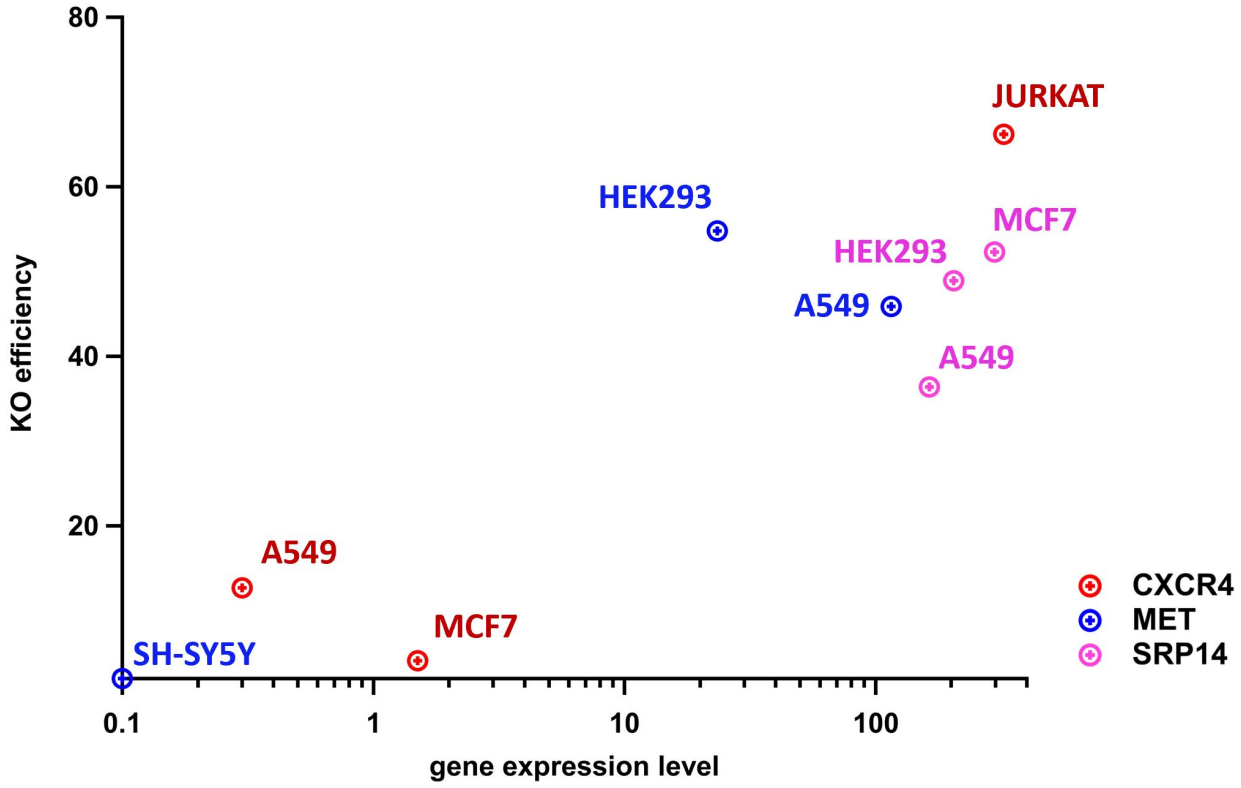


Figure 4.1: Correlation between gene target expression level and KO efficiency in different cell lines. RNPs were assembled *in vitro* with selected guides (Table 4.2: G2 for MET, G2 for CXCR4 and G2 for SRP14) and electroporated separately in selected cell lines via the Neon Electroporator system (Materials and Methods). Cells were collected after 48 hours, and the genomic DNA was extracted and sequenced. KO editing efficiency for each gene, in each cell line, was calculated via TIDE analysis. Each dot represents the KO efficiency (%) of a selected gene target, in each cell line, correlated to the level of expression in the log scale of that target in that cell line (according to Human Protein Atlas dataset).

4.2.2 CRISPRa cell line generation and gene expression upregulation

4.2.2.1 CRISPRa system and cell line generation

For those genes described above, which showed a correlation between expression level and cutting efficiency, I applied CRISPRa to modulate expression and test if increased expression would result in increased gene KO efficiency. To this extent, two types of CRISPRa systems were used:

- dCas9-VPR: consists of dCas9 fused to the tripartite activation factors VP64-p65-Rta (VPR) (Wiegand et al., 2015). The VP64-p65-Rta-dCas9 activator is created by modifying an existing dCas9 activator, in which a VP64 transcriptional activator is joined to the C terminus of dCas9. In the dCas9-VPR protein, the transcription factors p65 and Rta are added to the C terminus of dCas9-Vp64 resulting in increased expression of targeted genes.

- dCas9-pp7/PCP-PH: consists of a CRISPRa RNA scaffold recruitment system which uses the dCas9 and gRNA modified with the addition of pp7 hairpin aptamers. These aptamers recruit the viral RNA binding protein (RBP) PCP, which is fused to the transcriptional activator domains p65 and HSF1 (PH), to generate the fusion protein PCP-PH and increase target gene expression (Martella et al., 2019; Zalatan et al., 2015).

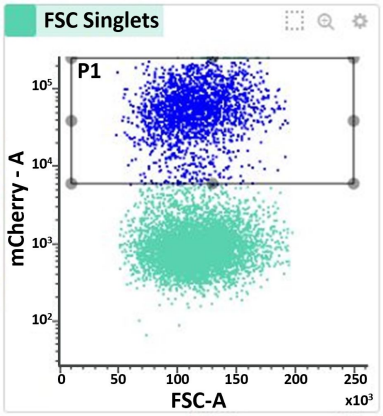
Based on their availability in house, I could select 3 cell lines to test the CRISPRa systems (A549, MCF7 and HEK293).

For CXCR4 overexpression, I first generated an A549 cell line inserting a dCas9-VPR system in the cell genome via lentivirus transduction (Materials and Methods). The lentiviral vector containing dCas9-VPR used in this work (List of Reagents) is provided with a mCherry cassette for sorting and enrichment of positively transduced cells, and Figure 4.2 shows the FACS setting and sorting of mCherry-positive A549 cells. I obtained ∼27% of positively transduced cells which were sorted, cultured and expanded. Next, I performed a second transduction with a lentivirus containing the sequence of a single CRISPRa guide, targeting CXCR4 promoter (See List of Reagents for guide sequences). This time, the used lentivector expresses a Blue Fluorescence Protein (BFP) for sorting and Figure 4.3 shows FACS setting and sorting results, where ∼86% of cells were sorted as BFP positive.

For CXCR4 target, I also tested a MCF7 cell line, available in house, containing the same dCas9-VPR system but doxycycline-inducible. In this case, dCas9-VPR expression is controlled and can only occur upon treatment with Doxycycline (Materials and Methods). The expression of the same single CRISPRa gRNA used in A549 cells is obtained via lentiviral transduction and positively transduced MCF7 cells were previously sorted in house.

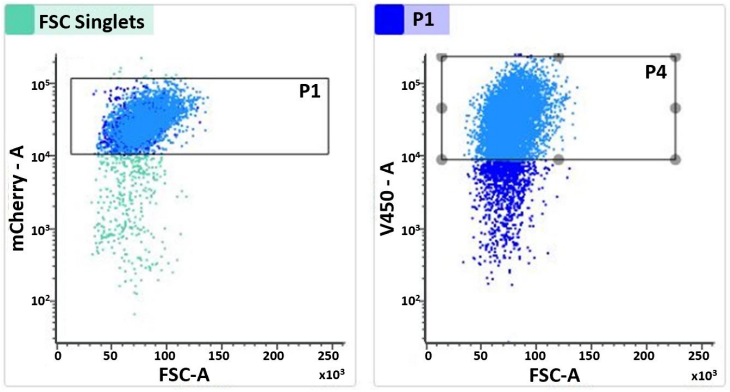
Differently, for MET upregulation I decided to test the CRISPRa activation in HEK293 cells available in house (clone 10C10), where this target gene is expressed at medium level (Table 4.1) and resulted to be already partially accessible to Cas9, according to the previous KO experiment (Table 4.2, Figure 4.1). In fact, it is possible that in this context, the overexpression effect on the editing efficiency would be enhanced compared to a cell line in which the gene is totally silenced, such as SH-SY5Y. In this case, the HEK293 (clone 10C10) cell line constitutively expresses the dCas9-pp7/PCP-PH CRISPRa system, and I used a pool of 3 CRISPRa gRNAs for MET promoter, delivered via plasmid transfection mediated by FuGENE Transfection reagent (Materials and Methods).

In addition, I tested CRISPRa activation of ASCL1 gene in the same HEK293 cell line, by delivering a single CRISPRa guide via FuGENE-mediated plasmid transfection (Materials and Methods).



Population	Events	% Parent	% Total
All Events	10,000		100.00%
Scatter	8,668	86.68%	86.68%
SSC Singlets	7,956	91.79%	79.56%
FSC Singlets	7,798	98.01%	77.98%
P1	2,084	26.72%	20.84%

Figure 4.2: A549-dCas9-VPR cell line sorting. A549 cells were transduced with a lentivector containing the dCas9-VPR gene and mCherry fluorescent cassette. Cells were cultured for 72 hours and then sorted via FACS. Left: mCherry positive gating plot vs Forward Scatter (FSC) Area, used to sort the positively transduced cells. Right: Table showing population percentage of the final sorting: P1 indicates the population of mCherry positive cells sorted from the initial pool, and it is 26.72% of the total % of Parental cells.



Population	Events	% Parent	% Total
All Events	10,000		100.00%
Scatter	8,704	87.04%	87.04%
SSC Singlets	8,223	94.47%	82.23%
FSC Singlets	8,219	99.95%	82.19%
P1	7,853	95.55%	78.53%
P4	6,750	85.95%	67.50%

Figure 4.3: A549-dCas9-VPR-CRISPRa cell line sorting. A549 cells sorted in Figure 1.4 were cultured and re-infected with a lentivector containing CXCR4-CRISPRa guide sequence and a Blue Fluorescent Protein (BFP) cassette. Cells were cultured for 72 hours and then sorted via FACS. Left: mCherry positive gating plot vs Forward Scatter (FSC) Area, used to sort the positively transduced cells with dCas9-VPR lentivector: 95.5% cells were sorted as mCherry positive (P1). BFP (V450-A) positive gating plot, used to sort the positively transduced cells with CRISPRa guide lentivector: 85.9% of cells were sorted as BFP positive (P4). Right: Table showing population percentage of the final sorting.

4.2.2.2 Gene expression upregulation

After CXCR4-CRISPRa-gRNA transduction and cell sorting, A549 and MCF7 were collected and qPCR analysis (Materials and Methods) was performed to test if the dCas9-VPR system with the designed CRISPRa-gRNA would induce CXCR4 over-expression. Total RNA was extracted from A549 cells while MCF7 cells were prior treated with Doxycycline to actually induce dCas9-VPR expression (Materials and Method). Total RNA was extracted from MCF7 after 48 hours of Doxycycline induction. For A549 cells, the fold change was calculated by comparing CXCR4 expression in cells transduced with CRISPRa-gRNA lentivector to the non-transduced cells while for MCF7 cells, the fold change was calculated by comparing CXCR4 expression in cells treated with Doxycycline versus untreated cells. The results in A549 cells show an upregulation of CXCR4 expression of 2.5-fold change compared to the parental line, while I detect a 5-fold higher expression in MCF7 cells (Figure 4.4 A).

For HEK293-EXPI dCas9-pp7/PCP-PH (10C10 clone), cells were cultured for 72 hours after plasmid transfection, and then total RNA was extracted for qPCR analysis (Materials and Methods). The expression level fold change was calculated by comparing MET and ASCL1 expression in cells transfected with CRISPRa-gRNAs to cells transfected with a Non-Targeting CRISPRa-gRNA (which has no complementary sequence in the human genome). Results show a 3-fold change up-regulation of MET compared to the control and an up-regulation of about 500-fold change in the case of ASCL1 (Figure 4.4 C). In general, those results confirm that I can enhance the expression of the selected target genes by using the CRISPRa tool, although I find the extent of this effect to be extremely cell and target-dependent.

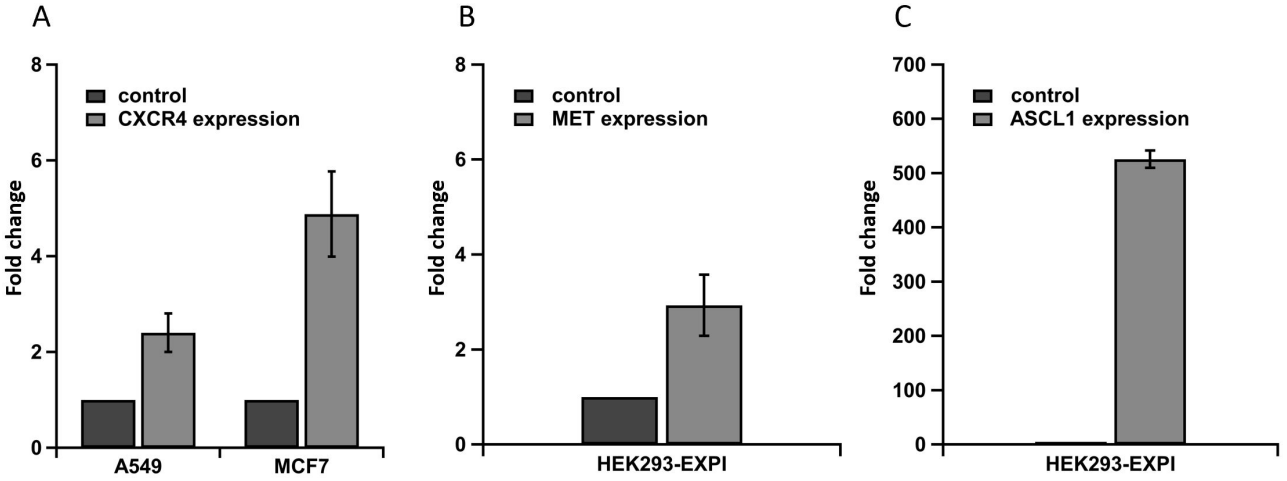


Figure 4.4: Target gene expression after CRISPRa upregulation. **(A)** CXCR4 expression fold change upon CRISRPa in A549 and MCF7. Total RNA was extracted from FACS-sorted A549 cells, 48 hours after lentivirus infection, and CXCR4 gene expression was calculated by qPCR analysis. Total RNA was extracted from MCF7 cells, 48 hours after Doxycycline induction and CXCR4 gene expression was calculated by qPCR analysis. CXCR4 overexpression upon CRISPRa was calculated by comparing the gene expression level to non-transduced cells (A549) or non-induced cells (MCF7). **(B)** MET expression fold change upon CRISPRa in HEK293-EXPI. HEK293-EXPI with a constitutive expression of dCas9-pp7/PCP-PH (10C10 clone) were transfected with a single CRISPRa guide for MET promoter, delivered via plasmid transfection mediated by FuGENE Transfection reagent. Gene overexpression was calculated by qPCR analysis, comparing the gene expression level to cells transfected with a non-targeting guide. **(C)** ASCL1 expression fold change upon CRISPRa in HEK293-EXPI (10C10 clone). A pool of 3 CRISPRa gRNAs for ASCL1 promoter, was delivered via plasmid transfection mediated by FuGENE Transfection reagent. Gene overexpression was calculated by qPCR analysis, comparing the gene expression level to cells transfected with a non-targeting guide.

4.2.3 CRISPRa upregulation shows a minimal effect on the on-target editing efficiency

RNP electroporation was performed to induce the KO 48 hours after CRISPRa activation. Briefly, the Cas9-gRNA complex was assembled *in vitro* using the same KO guide tested in Figure 4.1, then added to the cell suspension and transfected into the cells via electroporation through the Neon Transfection System (Materials and Methods). Cells were harvested after 72 hours, the genomic DNA was extracted and PCR amplification was performed for the specific expected cutting site (Materials and Methods). After sequencing, cutting efficiency was calculated using TIDE analysis (Materials and Methods).

CXCR4 KO efficiency does not result in a significant enhancement at the on-target site after CRISPRa activation. In fact, the 2.5-fold change upregulation detected in A549 (Figure 4.4 A) does not allow for a greater editing efficiency, which remains around ∼5% before and after CRISPRa (Figure 4.5 A). Similarly, the 5-fold change CXCR4 upregulation detected in MCF7 cells (Figure 4.4 A) does not induce a significant effect on the on-target KO efficiency, which is enhanced of ∼2% upon CRISPRa (Figure 4.5 A). Likewise, the enhancement of MET editing efficiency detected in HEK293 cells upon CRISPRa activation is only about 2.5% (Figure 4.5 B). For ASCL1 in HEK293 cells, as a result of the 500-fold change CRISPRa upregularion (Figure 4.4 C), I can detect a 10% increase in KO efficiency of the gene (Figure 4.5 C).

It is possible to speculate that a combination of factors is involved in the variability of these results, which definitely requires a deeper and broader analysis in terms of upregulation and KO efficiency of a statistically relevant number of genes across several cell lines. However, it is reasonable to think that, in a silenced or low expressed gene, even by forcing a promoter "opening", more factors are involved locally at different sites along the gene body (chromatin remodelling factors, histones post-translational modifications, DNA accessibility and three-dimensional DNA conformational change) which can impact Cas9 nuclease activity at different levels and which can add an additional degree of variability to the system. Despite the fact that this limited gene selection cannot be conclusive in terms of a genome-wide application, I envisage that multi-omics data analysis, in particular high-quality data for chromatin accessibility landscape for specific gene targets, will be pivotal in the identification of the best site for the CRISPRn system. In this regard, the guide design approach may need to be adapted and

precisely adjusted for each target and cell line.

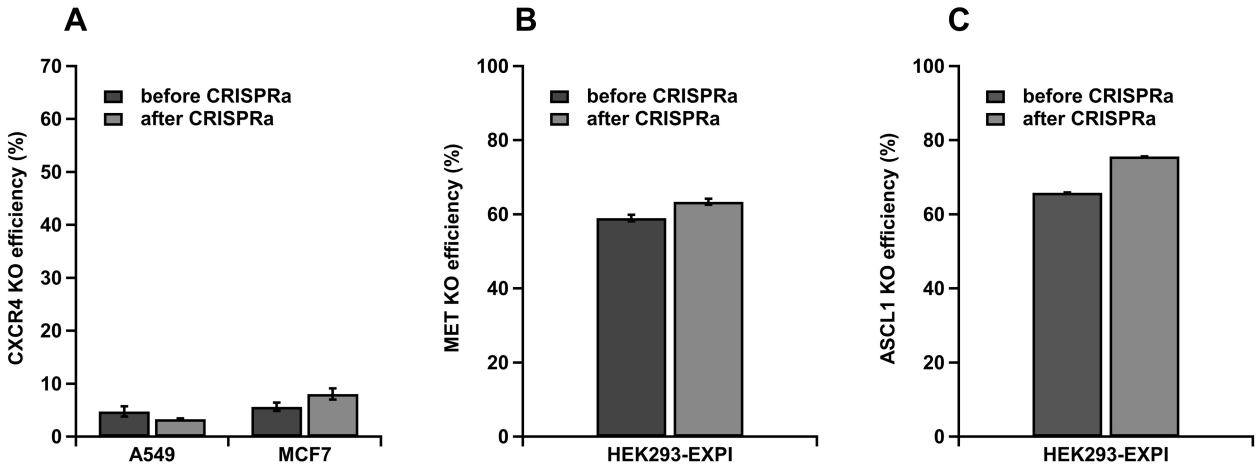


Figure 4.5: KO editing efficiency upon CRISPRa gene overexpression. RNP complexes were assembled *in vitro* and electroporated into the cells 48 hours after CRISPRa activation, via the Neon Electroporator system. Cells were then cultured for 48 hours. The genomic DNA was extracted and sequenced, and KO editing efficiency was calculated via TIDE analysis (Materials and Methods). **(A)** CXCR4 KO efficiency in A549 and MCF7 cells before and after CRISPRa-induced upregulation. **(B)** MET KO efficiency and **(C)** ASCL1 KO efficiency in HEK293 cells before and after CRISPRa-induced upregulation.

4.3 Discussion

Here, I describe a strategy based on the combination of CRISPRa (activation) and CRISPRn (nuclease active) tools to recreate a chromatin opening of silenced or low expressed genes and test Cas9 editing efficiency at the selected targets. In this work, I used a combination of computational and experimental approaches to select specific target genes, considering different aspects such as gene expression level in different cell lines, the correlation between the expression level and Cas9 cutting activity on that target, and the possibility to up-regulate the gene by CRISPRa system. Taking into consideration that Cas9 has been proved to be less efficient at silenced genes (Daer et al., 2017; Verkuijl and Rots, 2019), probably due to the diminished accessibility of the closed-chromatin, with this strategy I aimed to investigate this aspect and possibly improve editing efficiency by increasing Cas9 accessibility. I show results on CXCR4, MET, and ASCL1 target genes. Cell lines that constitutively express dCas9-VPR (or which have an inducible expression of dCas9-VPR) or with a PCP-PH recruitment system through a modified gRNA, have been used to induce the upregulation of the selected genes using the CRISPRa tool. CXCR4 and MET gene expression result in a modest upregulation (2.5-5 fold change) compared to the control (Figure 4.4 A, B), while ASCL1 expression is enhanced of about 500 fold change (Figure 4.4 C). Cas9 KO efficiency on CXCR4 and MET do not show a significant enhancement at the on-target site upon gene upregulation (Figure 4.5 A,B); this could be related to the low upregulation upon CRISPRa. On the other hand, CRISPRa on ASCL1 gives a much higher gene upregulation and, in this case, Cas9 KO efficiency is improved compared to the control. However, the enhancement in KO efficiency (10%) is still quite modest (Figure 4.5 C).

Here it is noteworthy that, by using dCas9 fused with transcriptional activators or recruited via modified gRNA, CRISPRa allows for gene expression upregulation which triggers an overall chromatin remodelling process involving the whole gene body, for the gene expression to take place. Within this process, CRISPRa is supposed to act both on the chromatin status of the target gene, leading to nucleosomes sliding or eviction and to changes in the histones post-translational modifications, and on the DNA confromation, creating DNA loops, and distorting ("opening") the helix structure at different sites, due to interaction of multiple DNA-binding proteins involved in the transcription process (Li et al., 2007). In my results I cannot detect any

significant effect on Cas9 on-target activity; however, this was also true for the Proxy-system discussed in Chapter 3. It is reasonable to think that CRISPRa could potentially have a higher effect on Cas9 off-target activity, rather than on the on-target, which can be investigated by using mismatched KO guides.

It should be also mentioned that there could be competition between Cas9 and dCas9 when simultaneously expressed in this system for the two different gRNAs (one targeting the promoter region for CRISPRa activation and one targeting the coding region for KO). This competition can result in reduced Cas9 activity at the cutting site or DSBs in the promoter region of the gene. The strategy to separate CRISPRa activation and CRISPRn in two sequential steps should limit the competition effect. However, this possibility still needs to be considered. Eventually, this limitation can be overcome by using a short guide for CRISPRa and thus avoiding DSBs occurring at the promoter site.

Additionally, since the CRISPRn efficiency at the KO site is highly dependent on the chromatin envirorment, and the chromatin compactness can vary considerably from genomic site to genomic site and from cell type to cell type, the chromatin status and accessibility at specific local site along the whole gene body can be investigated with a variety of technologies, such as Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE-seq, Giresi et al. 2007), DNase hypersensitive region sequencing (DNase-seq, Buenrostro et al. 2013), Assay for Transposase-Accessible Chromatin sequencing (ATAC-seq, Boyle et al. 2008) and Nicking Enzyme assisted sequencing (NicE-seq, Ponnaluri et al. 2017). This information can be used to implement the design of the gRNA.

Altogether, these are useful considerations for the refinement of the described method to obtain additional information about CRISPR activity and genomic context dependency.

Bibliography

A. P. Boyle, S. Davis, H. P. Shulha, P. Meltzer, E. H. Margulies, Z. Weng, T. S. Furey, and G. E. Crawford. High-Resolution Mapping and Characterization of Open Chromatin across the Genome. *Cell*, 132(2):311–322, 1 2008. ISSN 0092-8674. doi: 10.1016/J.CELL.2007.12.014.

J. D. Buenrostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, and W. J. Greenleaf. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods 2013 10:12*, 10(12):1213–1218, 10 2013. ISSN 1548-7105. doi: 10.1038/nmeth.2688. URL https://www.nature.com/articles/nmeth.2688.

R. M. Daer, J. P. Cutts, D. A. Brafman, and K. A. Haynes. The Impact of Chromatin Dynamics on Cas9-Mediated Genome Editing in Human Cells. *ACS Synthetic Biology*, 6(3):428–438, 3 2017. ISSN 21615063. doi: 10.1021/acssynbio.5b00299.

P. G. Giresi, J. Kim, R. M. McDaniell, V. R. Iyer, and J. D. Lieb. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Research*, 17(6):877, 6 2007. ISSN 10889051. doi: 10.1101/GR.5533506. URL /pmc/articles/PMC1891346//pmc/articles/PMC1891346/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC1891346/.

S. I. Grewal and D. Moazed. Heterochromatin and epigenetic control of gene expression. *Science*, 301(5634):798–802, 2003. ISSN 00368075. doi: 10.1126/science.1086887.

B. Li, M. Carey, and J. L. Workman. The Role of Chromatin during Transcription. *Cell*, 128(4):707–719, 2 2007. ISSN 0092-8674. doi: 10.1016/J.CELL.2007.01.015. URL http://www.cell.com/article/S0092867407001092/fulltexthttp://www.cell.com/article/S0092867407001092/abstracthttps://www.cell.com/cell/abstract/S0092-8674(07)00109-2.

A. Martella, M. Firth, B. J. Taylor, A. Göppert, E. M. Cuomo, R. G. Roth, A. J. Dickson, and D. I. Fisher. Systematic Evaluation of CRISPRa and CRISPRi Modalities Enables Development of a Multiplexed, Orthogonal Gene Activation and Repression System. *ACS Synthetic Biology*, 8(9):1998–2006, 2019. ISSN 21615063. doi: 10.1021/acssynbio.8b00527.

V. K. Ponnaluri, G. Zhang, P. O. Estève, G. Spracklin, S. Sian, S. y. Xu, T. Benoukraf, and S. Pradhan. NicE-seq: High resolution open chromatin profiling. *Genome Biology*, 18(1):1–15, 6 2017. ISSN 1474760X. doi: 10.1186/S13059-017-1247-6/FIGURES/7. URL https://genomebiology.biomedcentral.com/articles/10.1186/s13059-017-1247-6.

S. A. Verkuijl and M. G. Rots. The influence of eukaryotic chromatin state on CRISPR–Cas9 editing efficiencies, 2 2019. ISSN 18790429.

D. J. Wiegand, D. Ter-ovanesyan, J. L. Braff, and N. Davidsohn. Highly-efficient Cas9-mediated transcriptional programming. 12(4):326–328, 2015. doi: 10.1038/nmeth.3312.Highly-efficient.

J. G. Zalatan, M. E. Lee, Almeida, Ricardo, Gilbert, Luke A., H. Whitehead, M. La Russa, J. C. Tsai, J. S. Weissman, J. E. Dueber, L. S. Qi, and W. A. Lim. Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *cell*, 23(1):1–7, 2015. ISSN 15378276. doi: 10.1038/jid.2014.371. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2688697/pdf/nihms-110443.pdf.

Chapter 5

Cas12a characterization at single-molecule level

5.1 Introduction

Facilitated by the advancement across multiple disciplines such as bioinformatics, structural biology and high-throughput sequencing, the discoveries and engineering of various innovative CRISPR-Cas systems are rapidly expanding the CRISPR toolbox. As anticipated, Cas12a has surfaced as a promising alternative over the well-known SpCas9 for genomic applications and diagnostics, showing several advantages such as fewer off-target effects in cells (Kleinstiver et al., 2016; Kim et al., 2016) and staggered cleavage (Zetsche et al., 2015; Strohkendl et al., 2018). Up to date, the key mechanistic details of Cas12a specificity and dynamics must be understood in regard to DNA interaction at specific sites, off-target site discrimination and target recognition and cleavage. In the following chapter, I describe my work on Cas12a nuclease-DNA interaction and how the mechanical distortion of dsDNA affects the nuclease activity, therefore extending my investigation about this specific aspect to a second and relevant CRISPR-Cas system. To this end, I used a single-molecule optical tweezers assay (Newton et al., 2019) combined with confocal microscopy, enabling real-time observation of single-molecule Cas12a mechanism of action.

This work has been published in Physical Chemistry Chemical Physics (PCCP) Journal (Losito et al., 2021).

5.2 Results

5.2.1 Cas12a stably binds the on target site with high affinity

To monitor the interaction of Cas12a with DNA in real time and at single-molecule resolution, I used a combination of optical tweezers with confocal fluorescence microscopy and microfluidics (Materials and Methods). Cas12a complexes were assembled with a 5'-Cy3-labeled crRNA, unlabelled tracrRNA, and catalytically active wtCas12 from *Acidaminococcus sp.* (IDT) (Figure 5.1 A). The crRNA sequence was selected to targets Cas12a to a unique site at 33.5 kb on λ-DNA, in proximity of a TTTG-PAM (List of Reagents). This site is asymmetrically positioned, enabling the orientation of the DNA to be determined if on-target binding is observed; this is important as the DNA can be tethered between the beads in either orientation. Experiments were performed using a microfluidic flow cell shown in Figure 5.1A. First, two streptavidin coated polystyrene beads (4.84 μm) are trapped and a single piece of biotinylated λ-DNA is caught in between them. Next, the DNA is moved to the buffer only channel where a force-extension curve is recorded from 0-65 pN to confirm the presence of only a single piece of intact DNA (Materials and Methods). For these experiments the DNA is torsionally unconstrained, it is tethered by biotin-streptavidin interactions on only one strand at each end, resulting in the characteristic overstretching plateau at 65 pN (Materials and Methods). The DNA is then moved into the channel containing the labelled Cas12a complex and held at 5 pN (Figure 5.1 B), which is a sufficient force to extend the DNA to its contour length (16 µm), and allow confocal imaging, but does not alter the base pairing of the two strands. For the microfluidic setting, I used two channels to flow the labelled RNP complex in two different buffer condition, one containing Ca^{2+} and the other containing Mg^{2+} (Figure 5.1 A, Materials and Methods). This allows to always use an active protein but to prevent the DNA cleavage in the Ca^{2+} buffer channel, and thus to study the protein binding, or to allow the cleavage activity when moving in the Mg^{2+} containing buffer. In the presence of labelled RNP complex, in Ca^{2+} buffer, a single binding event is observed at the expected on-target location which remains stable for several minutes (Figure 5.1 B), consistent with previous observations that Cas12a complex binds target DNA with very high affinity. No binding is observed with Cy3-crRNA alone (Figure 5.2), confirming that such events represent stable binding of the holoenzyme complex.

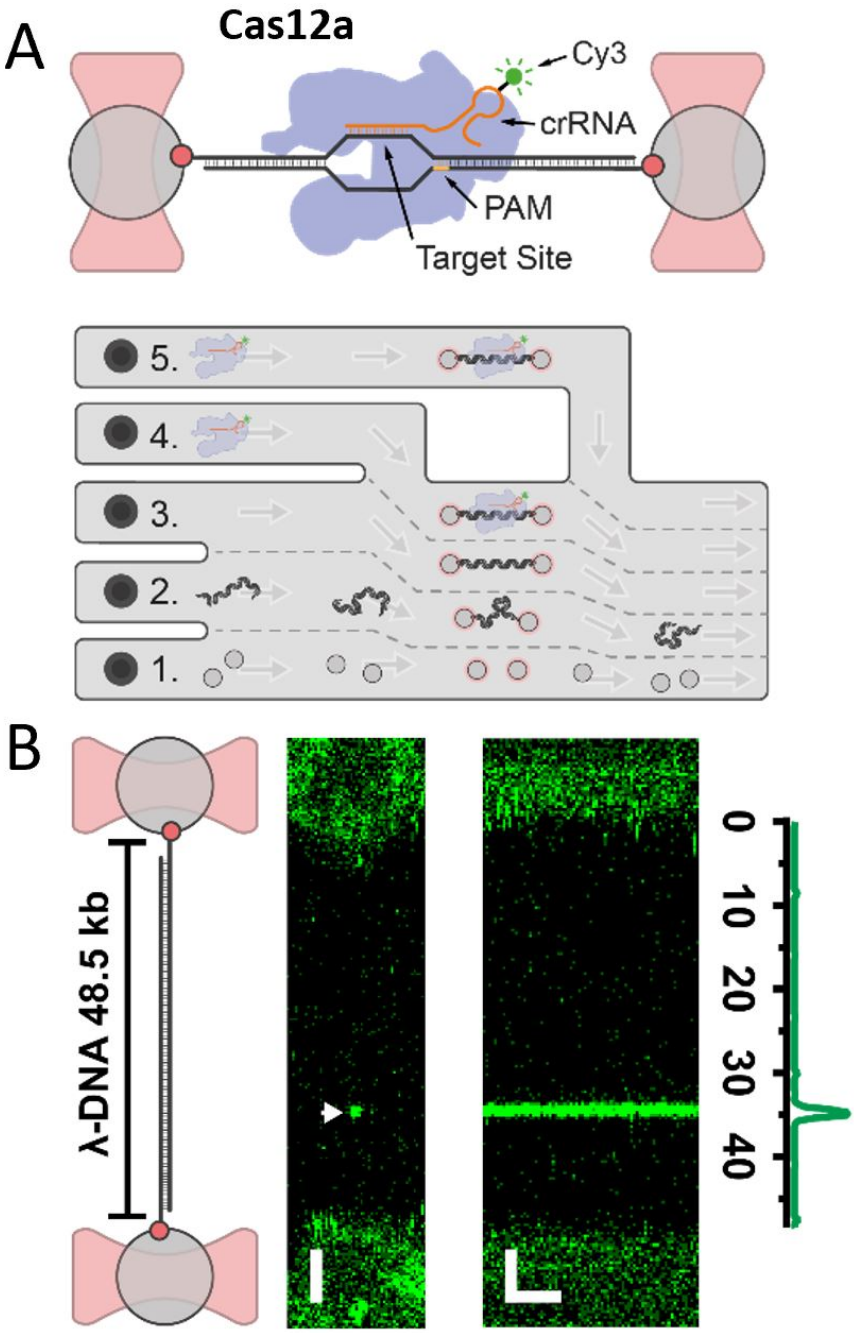


Figure 5.1: C-trap set up and Cas12 on-target binding analysis. **(A)** Biotinylated λ-DNA tethered between optically trapped beads, with bound Cy3-labelled Cas12a RNP complex (top). Microfluidics (bottom): (1) beads channel; (2) DNA channel; (3) buffer-only channel; (4) Cas12a in Ca^{2+}-buffer; (5) Cas12a in Mg^{2+}-buffer. **(B)** 2D-confocal image of λ-DNA at 5 pN with a single Cy3-Cas12a complex bound at expected target site (33.5 kb, arrow) in Ca^{2+} -buffer. Kymograph shows Cas12a stably bound on target for several minutes. Genomic location analysis confirms on-target binding. Scalebars = 2 µm and 2 min.

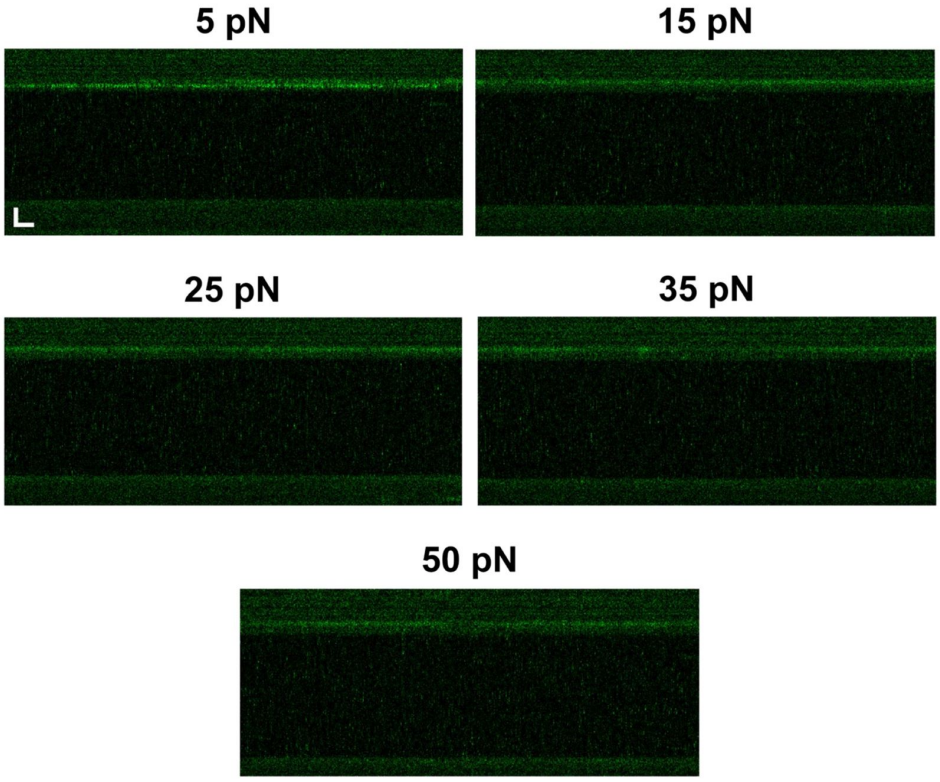


Figure 5.2: Control: Kymographs recorded in the presence of Cy3-labelled crRNA only, without Cas12a. Kymographs of force-stretched λ-DNA (5-50 pN) in the presence of 1 nM labelled-gRNA only (no Cas12a), showing that guide alone does not bind the DNA (unlabelled). Scale Bars = 2 μm, 2 s.

5.2.2 Cas12a specifically cleaves the on-target but force hinders cleavage

Next, I tested whether Cas12a is active on force-clamped-DNA by studying the protein-DNA interaction in the Mg^{2+}-buffer channel (Figure 5.3 A). From previous study it is known that fluorescent labelling does not affect activity in control bulk assays (Singh et al., 2018). A single Cas12a complex bounds its specific target, as expected, and cleavage is observed at 20 s as a drop in the force between the beads and a displacement of the force-clamped bead (Figure 5.3 A). I approximate the cleavage rate constant ($k_{cleave} = 0.021 \pm 0.001\ s^{-1}$) by measuring the dwell time at the target site from initial binding to cleavage and I repeated the measurements at forces ranging from 2 to 7 pN (Figures 5.3 B, 5.4). Results show that the dwell time between binding and cutting increases linearly with force, and the corresponding rate constants decreases exponentially, indicating that force hinders cleavage (Figure 5.3 B). Indeed, at 10 pN or higher, I could not observe cleavage within the 800 s experimental window. In order to obtain mechanistic information about the force dependency of the cleavage activity, I fitted the dwell times to the Bell–Evans equation (Evans, 2003; Mora et al., 2020) (Figure 5.3 B).

$$k_{cleave(F)} = k_{cleave(0)} e^{-\frac{F x^{\ddagger}}{k_B T}} \tag{5.1}$$

Where F is the force used to tether the DNA, k_B is the Boltzmann constant, T is the temperature, and $x^{\ddagger}$ is the distance to the transition state. From the intercept of the curve to the logarithmic scale k_{cleave} values, I could calculate the zero-force cleavage rate constant $k_{cleave(0)} = 0.08 \pm 0.01\ s^{-1}$, which is consistent with prior bulk measurements (Murugan et al., 2020; Strohkendl et al., 2018). Additionally, from the slope of the curve, I calculated the distance to the transition state $x^{\ddagger} = 1.8 \pm 0.3$ nm. This indicates a large, rate limiting conformational change between binding and complete cleavage and it is possible to hypothesize that this is consistent with the expected movement of the REC domain bringing the TS towards the cleavage site in RuvC. Interestingly, these domains are 2 nm apart in the various published structures (Dong et al., 2016; Gao et al., 2016; Stella et al., 2017; Swarts and Jinek, 2019; Ya-

mano et al., 2016). This conformational change is also in agreement with prior single-molecule FRET experiments and molecular dynamics calculations (Stella et al., 2018; Saha et al., 2020).

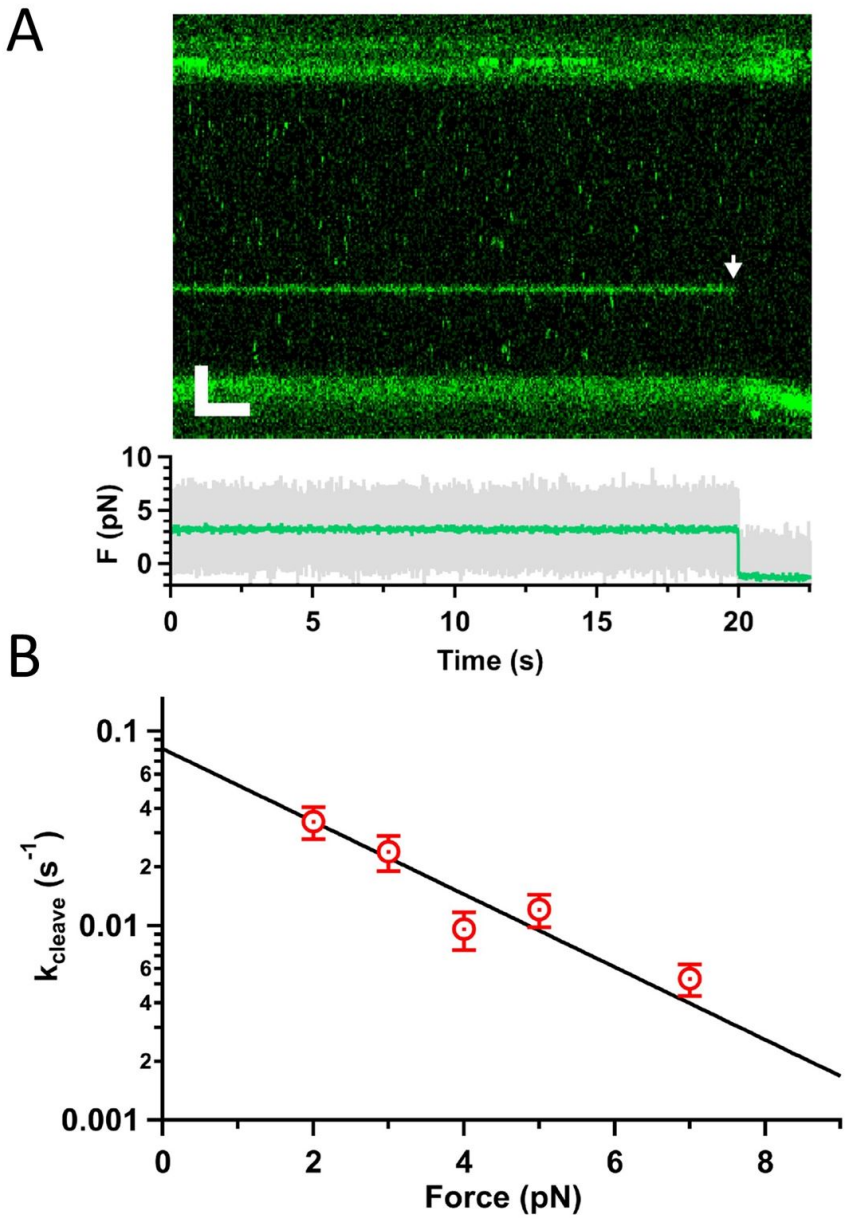


Figure 5.3: Cas12a on-target cleavge analysis and force correlation.**(A)** Kymograph of force clamped λ-DNA (3 pN) in 10 mM Mg^{2+} and 100 mM Na^{+} with Cas12a bound on target. Cleavage is observed at 20 s as fluorescence loss (arrow), displacement of the trapped bead, and drop in the force (bottom). Scale bars = 2 μm, 2 s. **(B)** Cleavage rate constant as a function of force (average and standard error of the mean (s.e.m.), n $\geq$ 17 each), and fit to the Bell–Evans equation (line).

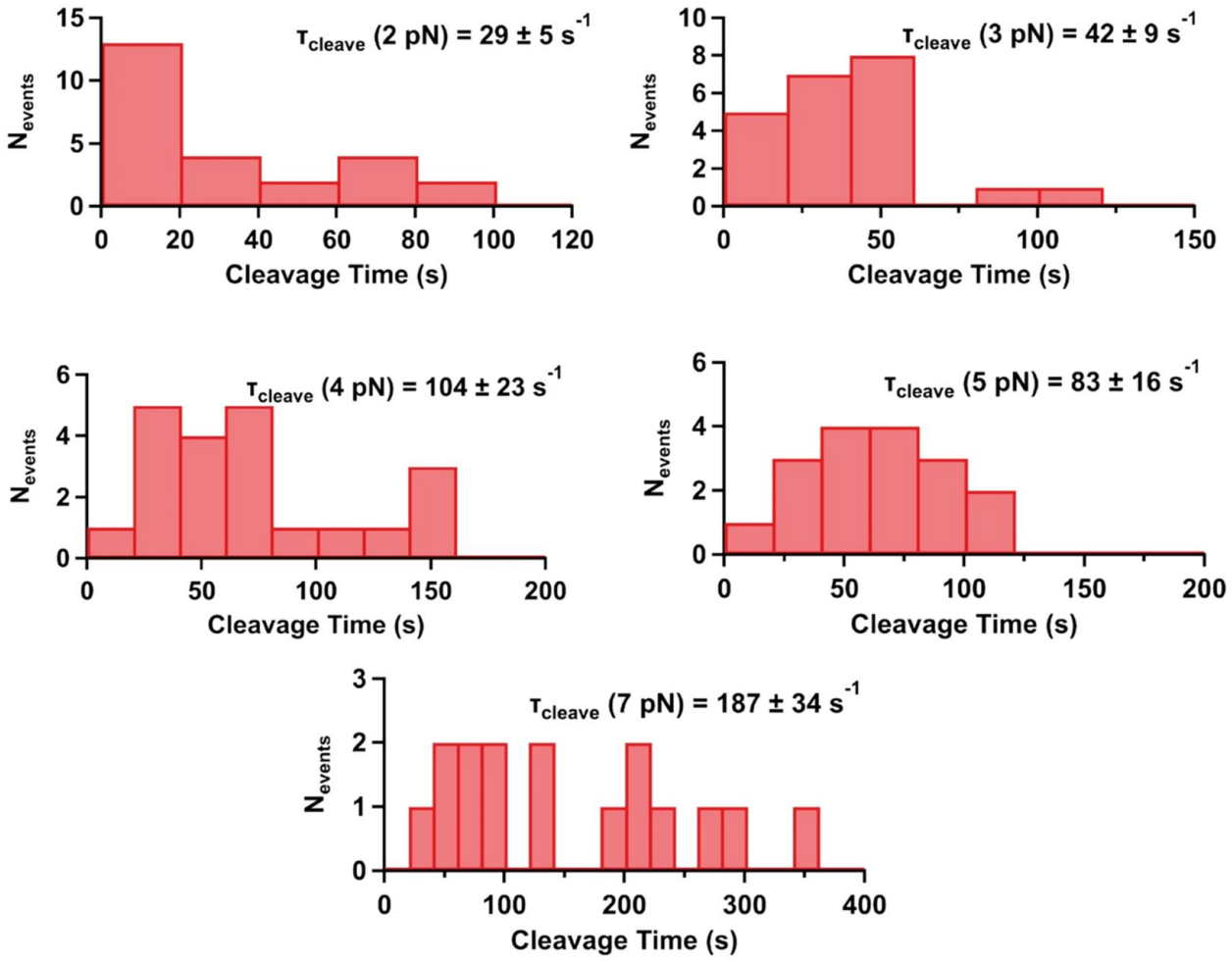


Figure 5.4: Distributions of dwell times between target binding and cutting as a function of force, and average dwell time and standard error of the mean, as indicated. The rate constant k_{cleave} (Figure 5.3 B) was approximated as the inverse of the average dwell time at each force.

5.2.3 Cas12a diffuses on dsDNA to locate its target and search mechanism involves hopping

Next, I further investigated on the target search mechanism of Cas12a. Indeed, in agreement with prior single-molecule observation (Jeon et al., 2018), my results show Cas12a undergoing random and bidirectional diffusion while interacting with the dsDNA. This mechanism is completely different from Cas9, which instead quickly run off from the DNA if the complementarity with the sgRNA is not encountered. In my hands, in approximately half of the traces (n = 26/50), I can observe the Cas12a complex binding away from the target site, then diffusing until the correct on-target site is located (Figure 5.5). When diffusing, the complex can pass the target site multiple times before binding it tightly (Figure 5.5). This can be likely due to Cas12a being unable of binding the target solidly until the PAM Interacting Domain (PID) recognizes the PAM sequence. Therefore, 1D diffusion is probably the main unique Cas12a mechanism for target location, together with a second mechanism, similar to Cas9 behaviour, which involves 3D random collision to the target site (Figure 5.6 B).

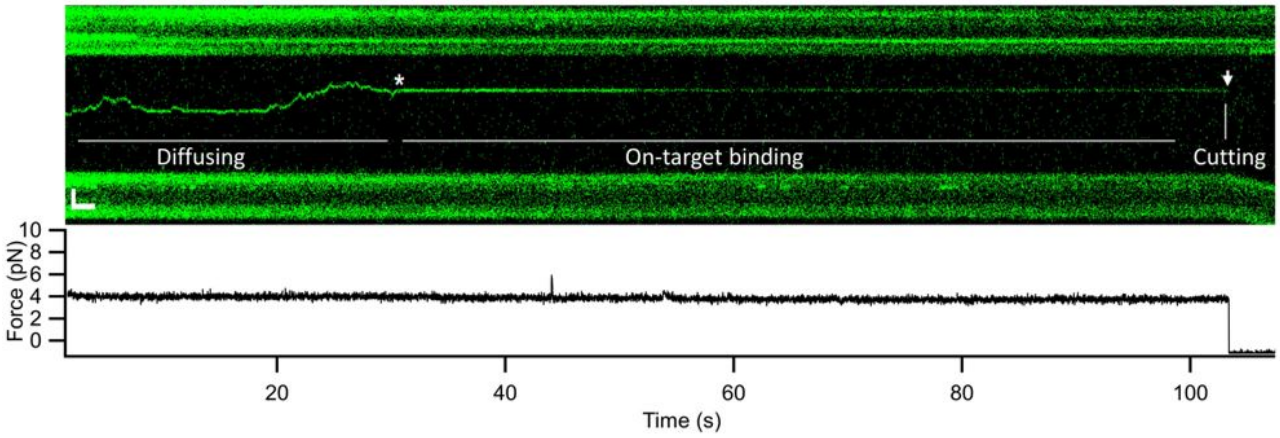


Figure 5.5: The full event from Cas12a target search to cleavage. Kymograph of force-clamped λ-DNA (4 pN) in 10 mM Mg^2 and 100 mM Na^+, showing the full event of Cas12a target search via 1D diffusion mechanism, on-target binding (star) and cleavage (arrow). Vertical scale bars = 2 µm.

Next, to characterise the diffusional behaviour of the Cas12a complex, I collaborated with Quentin Smith, from the Rueda Lab, for the extraction of diffusion trajectories from the kymographs, using a combination of custom python (https://github.com/ singlemoleculegroup) and IgorPro 8 (Wavemetrics) scripts. By tracking and quantifying Cas12a diffusion events using the python single particle tracking algorithm and further processing the trajectories in Igor Pro 8, we could calculate their Mean Square Displacement (MSD). First, I recorded a number of trajectories, in buffer containing 100 mM Na^+, of both static molecules, bound to the target, and diffusing complexes (Figure 5.7 A). The trajectories show that diffusing complexes can travel thousands of base pairs in just a few seconds, with average speed of ∼1 kbp per second compared with the static, target-bound molecules (Figure 5.7 C, D). The slope of the initial linear rise of the MSD curves yields the diffusion coefficient, D_{100} = (1.6 ± 0.9) x 10^6 bp^2s^{-1} (Figure 5.7 C). Next, I recorded several trajectories at lower ionic strength (25 mM Na^+) (Figure 5.7 B), and measured the relative diffusion coefficient (Figure 5.7 C, D). Indeed, the comparison of those values at different ionic strength, can give information about the mode of the diffusion mechanism, understanding if the enzyme is sliding or hopping. Since protein affinity to non-specific DNA is primarily determined by electrostatic interactions, varying the salt concentration in the experiments can modulate these interactions and allows to discriminate between the hopping and sliding models. If hopping is the underlying mechanism of translocation of the protein on DNA, then, at higher salt concentrations, the protein spends more time in solution, resulting in an increase in the measured diffusion coefficient. However, in the sliding mechanism, due to the continuous contact of the protein with DNA, the diffusion coefficient is independent of salt concentration (Tafvizi et al., 2011). My data show that the diffusion coefficient decreases by an order of magnitude with lower ionic strength (D_{25} = (1.7 ± 0.5) x 10^5 bp^2s^{-1}) (Figure 5.7 E), consistent with Cas12a hopping on the DNA, also in agreement with the previous single-molecule observations (Jeon et al., 2018).

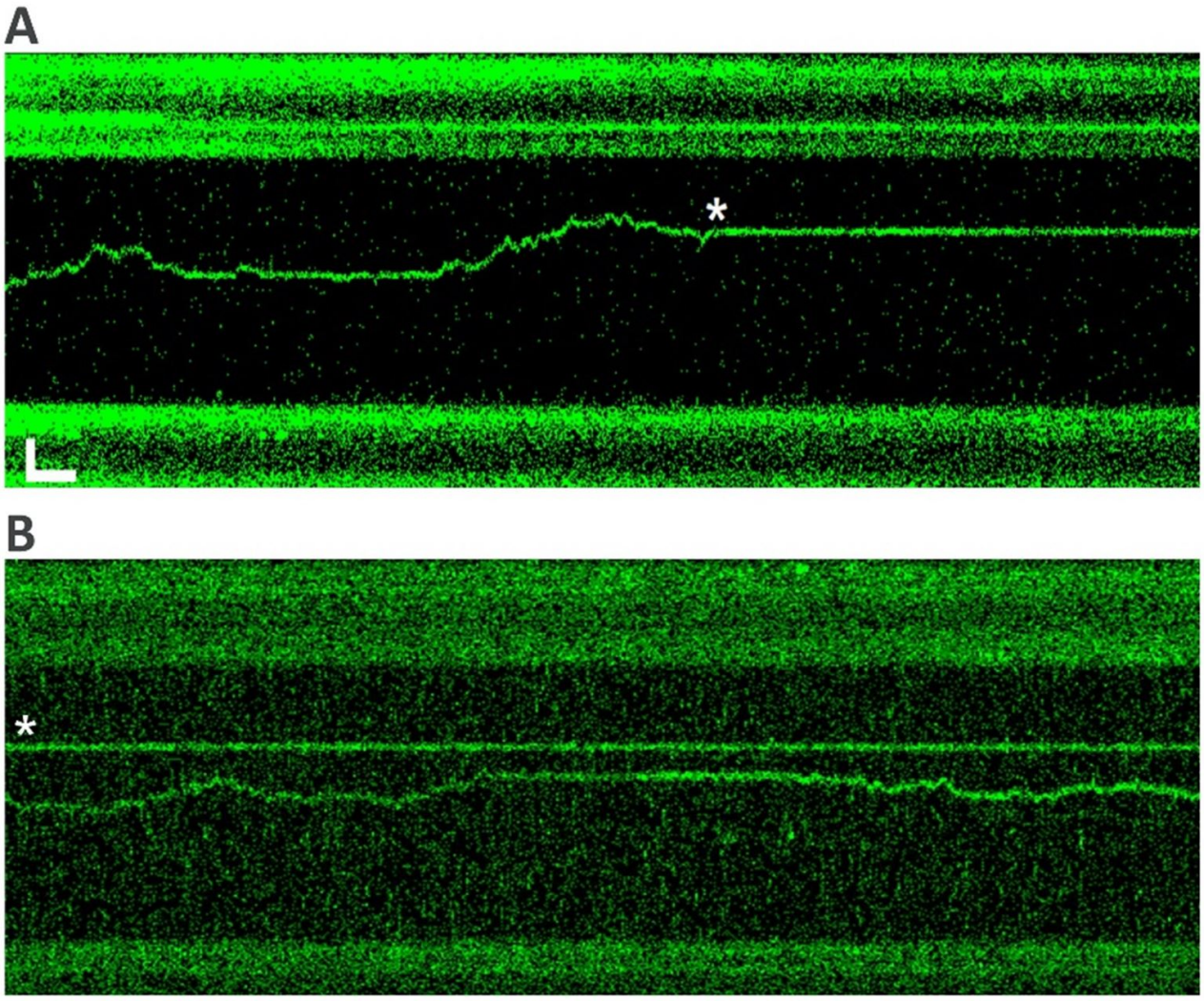


Figure 5.6: Cas12a diffuses on DNA to locate its target. **(A)** Kymograph of force-clamped λ-DNA (4 pN) in 10 mM Mg^{2+} and 100 mM Na^{+} with a diffusing Cas12a complex undergoing 1D diffusion to locate the on-target site. The complex diffuses twice past the target site (star) before binding tightly. **(B)** Kymograph of force-clamped λ-DNA (10 pN) in 10 mM Ca^{2+} showing a diffusing complex and a stable bound one at the on-target site, occupied after a 3D collision mechanism. Scale bars = 2 μm, 2 s.

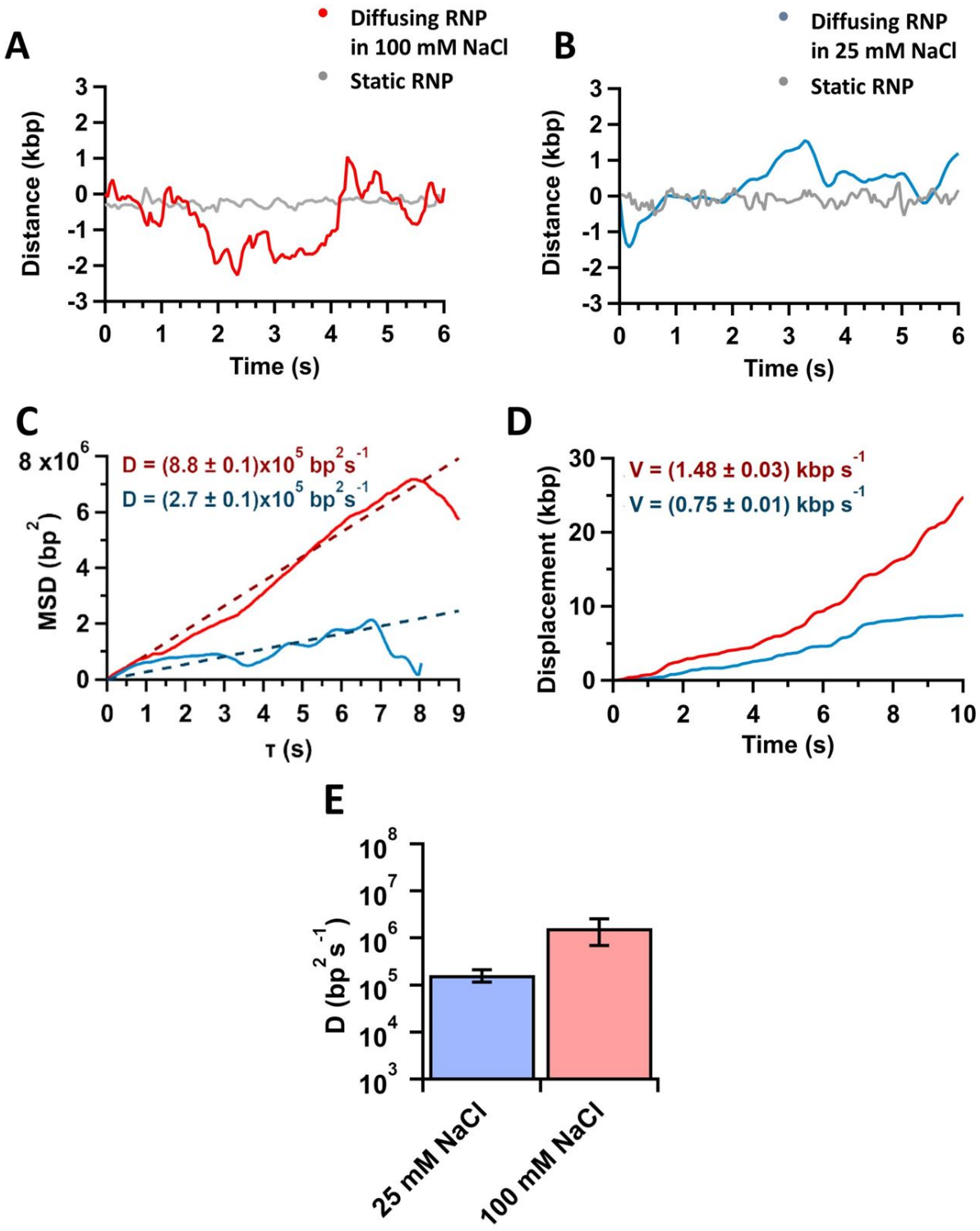


Figure 5.7: Cas12a diffusion analysis. **(A)** Trajectories of target bound (grey) and diffusing of Cas12a complex in 10 mM Ca^{2+} and 100 mM Na^{+}. **(B)** Trajectories of target bound (grey) and diffusing of Cas12a complex in 10 mM Ca^{2+} and 25 mM (blue) Na^{+}. **(C)** Mean-square displacement analysis and diffusion coefficients of trajectories in at 25 (blue) and 100 mM (red) Na^{+}. **(D)** Cumulative displacement analysis of trajectories in (B) at 25 (blue) and 100 mM (red) Na^{+}, and average diffusion velocities. **(E)** Diffusion coefficients in 25 and 100 mM Na^{+} (average and s.e.m., n = 20 each)

5.2.4 DNA stretching induces off-target binding and decreases the diffusion coefficient

Finally, I investigated how increasing the tension on the DNA affects Cas12a binding activity. Similarly to what has been already shown with Cas9 (Newton et al., 2019), using the labelled RNP complex, I obtain a single on-target binding event at low forces, 2-5 pN. However, when I start stretching the dsDNA with higher forces (< 25 pN), I can detect additional binding events at off-target locations, with molecules diffusing randomly on the stretched DNA. Increasing the force at ≥ 25 pN, results in additional stable binding events at off-target locations (Figure 5.7 A). To quantify this, I determined the number of off-target binding events per 10 000 bp as a function of force (Figure 5.8 B). The number of off-target bound complexes increases sigmoidally, which can be fit to the Hill-equation:

$$\theta = \theta_{max} \frac{[RNP]^n}{{K_d}^n + [RNP]^n} \tag{5.2}$$

Were θ is the measured binding events, θ_{max} is the maximum binding event on λ-DNA, n is the Hill coefficient, K_d is the dissociation constant and [RNP] indicates the concentration of the Cas12:guide complex. The fitting with experimental measurements, yield a saturation of 2.9 ± 1.4 complexes bound per 10 000 bp, and a mid-point force of 37 ± 15 pN. Interestingly, I also observed that, as the force increased to 30 pN, the diffusion coefficient decreases by 30-fold ($D = (0.5 \pm 0.2) \times 10^5$ bp^{2}s^{-1}, Figure 5.8 C) . It is reasonable to think that the bound Cas12a complexes eventually cease to diffuse and remain bound to their initial (off-target) binding site on the DNA, comparable to the behaviour of Cas9 (Newton et al., 2019).

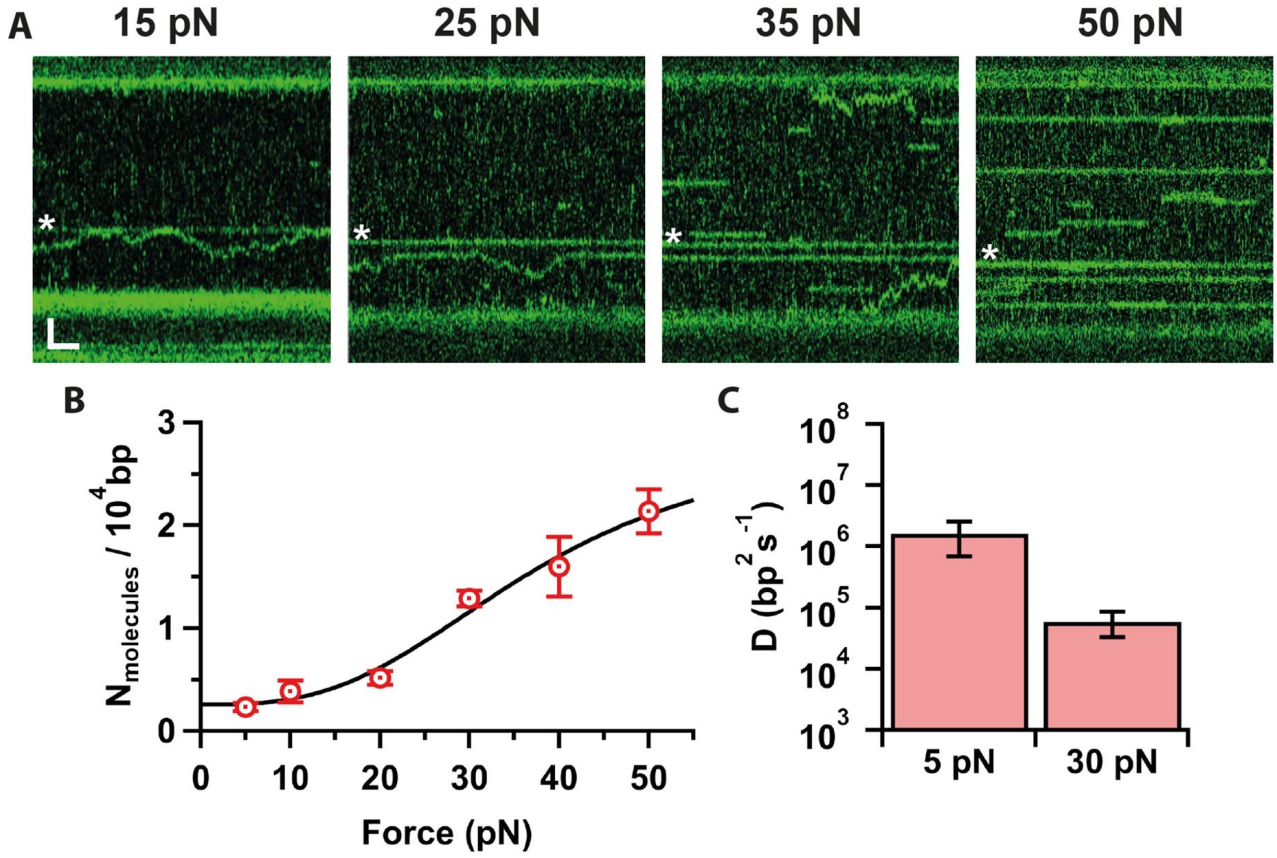


Figure 5.8: DNA streatching induces Cas12a off-target binding. **(A)** Kymographs of force stretched λ-DNA (15-50 pN) in 10 mM Ca^{2+} and 100 mM Na+, showing Cas12a bound on target (star) and off targets. Scale bars = 2 μm, 2 s. **(B)** Number of molecules bound per 10 000 bp as a function of force (average and s.e.m., n = 7 each), and fit to the Hill equation (line). **(C)** Cas12a complex diffusion coefficients at 5 and 30 pN (average and s.e.m., n = 20 each).

5.3 Discussion

In recent years, Cas12a has emerged as a salient alternative for genome editing applications to Cas9, particularly since it exhibits fewer off-target effects in cells and staggered cleavage which can be advantageous for Homologous Direct Repair mechanism to occur at the DSB site over Non Homologous End Joining. While Cas12a has already been investigated by some single-molecule studies (Jeon et al., 2018), the mechanistic and dynamic aspects of DNA interaction, target search and cleavage, and off-target binding remain poorly characterized. With the aim of elucidating those mechanisms and extending my investigation about the effect of DNA distortion on Cas nucleases activity, I performed a single molecule study using correlative single-molecule optical tweezers with fluorescent detection. Results show that the complex binds primarily away from the target on the DNA and diffuses randomly and bidirectionally until it engages with its target (Figures 5.5, 5.9). Target recognition likely requires the PID to recognize the PAM sequence as it diffuses over it, which can entail several attempts (Figures 5.5, 5.9). Once bound at the target site, the duplex is melted, the R-loop formed, and cleavage is observed within seconds (Figures 5.5, 5.9). Stretching the DNA with force hinders cleavage, which reveals a large amplitude ($\sim$1.8 nm) rate-limiting conformational change. This conformational change could involve a "closing" of the enzyme to bring the TS in REC domain into the active site in RuvC, and is consistent with known structures (Dong et al., 2016; Gao et al., 2016; Yamano et al., 2016; Swarts and Jinek, 2018; Stella et al., 2017) and prior observations (Stella et al., 2017; Saha et al., 2020). Diffusion analysis supports a search mechanism that involves hopping on the DNA, as previously proposed Jeon et al. (2018). The observed slow diffusion at higher forces, is consistent with the idea that Cas12a probes the DNA sequence at every hop. Since DNA stretching facilitates melting and R-loop formation, it is possible to speculate that mechanically distorting the DNA increases the dwell time at every "hop", thereby decreasing the observed diffusion coefficient, which in turn results in frequent off-target Cas12a binding in a force-dependent fashion. Interestingly, I also observe that off-target binding increases markedly at higher forces (>20 pN), similarly to Cas9 behaviour (Newton et al., 2019). This observation is consistent with the idea that force stretching lowers the DNA melting barrier and facilitates R-loop formation, even in the presence of multiple mismatches. However, Cas12a exhibits fewer off-target effects *in vivo*. The force-induced cleavage inhibition shown

here implies that those off-target bound complexes do not cleave, which will possibly have a low impact for off-target activity *in vivo*. In physiological environments, negatively supercoiled DNA, which is also underwound, may experience similar off-target binding, but it is also unlikely to be cleaved. Previous single-molecule studies (Newton et al., 2019) have indicated that Cas9 can be affected by mechanical perturbations, and here I additionally provide new insights of this particular aspect for Cas12a mechanism. Indeed, in this context, I can detect an opposing effect of DNA unwinding on off-target binding and cleavage, which can elucidate the mechanistic basis of its specificity.

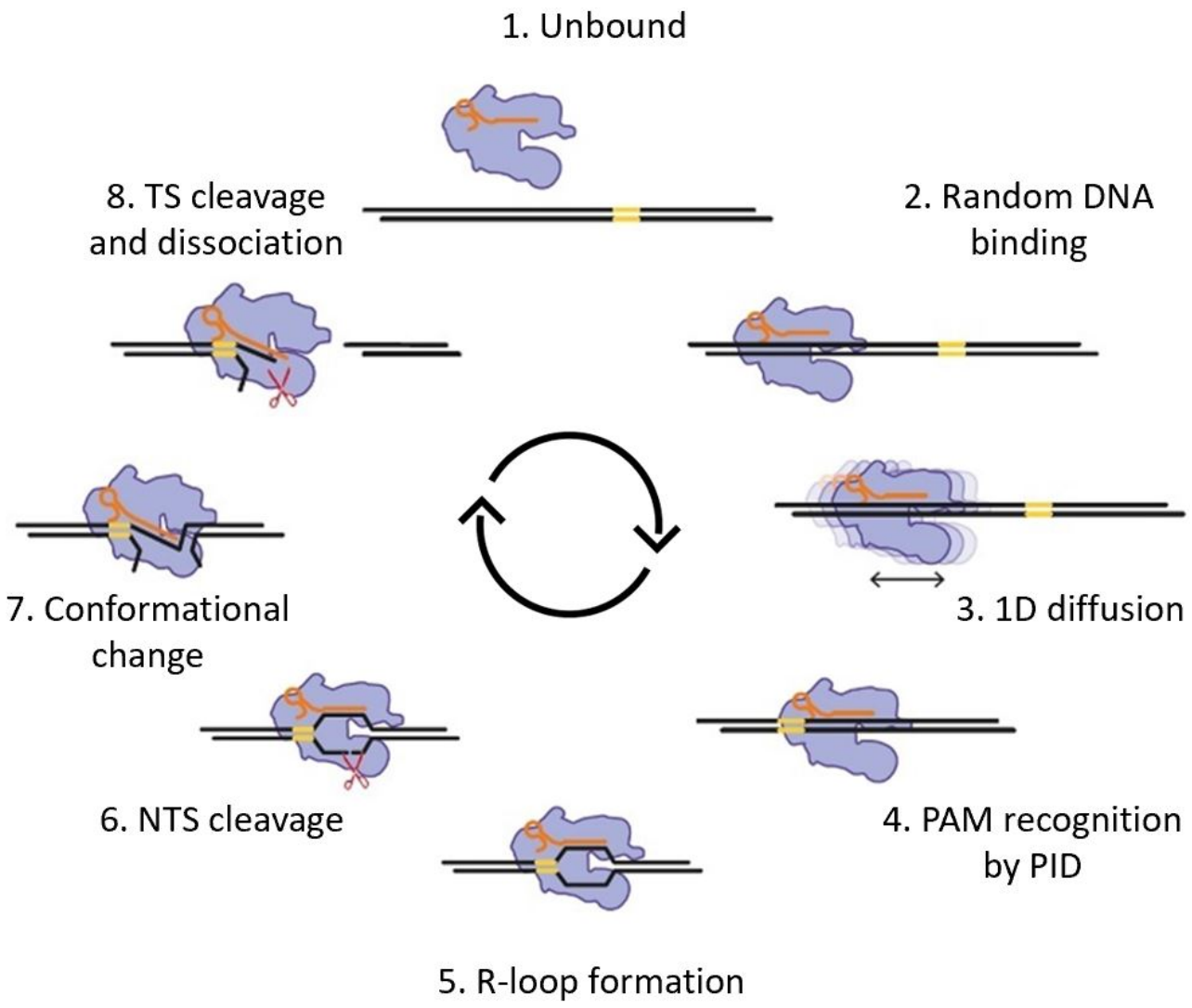


Figure 5.9: Proposed model for Cas12a target search and cleavage mechanisms. Cas12a binds DNA at a random location and undergoes 1D diffusion to locate its target. The PID recognizes the target's PAM sequence, which triggers a cascade of conformational changes to unwind the DNA duplex and form the R-loop. The NTS interacts with the RuvC domain triggering the first cleavage event. A rate limiting conformational change brings the TS into the RuvC active site enabling staggered TS cleavage. The DSB is complete and Cas12a is recycled for another round of catalysis.

Bibliography

D. Dong, K. Ren, X. Qiu, J. Zheng, M. Guo, X. Guan, H. Liu, N. Li, B. Zhang, D. Yang, C. Ma, S. Wang, D. Wu, Y. Ma, S. Fan, J. Wang, N. Gao, and Z. Huang. The crystal structure of Cpf1 in complex with CRISPR RNA. *Nature 2016 532:7600*, 532(7600):522–526, 4 2016. ISSN 1476-4687. doi: 10.1038/nature17944. URL `https://www.nature.com/articles/nature17944`.

E. Evans. Probing the Relation Between Force—Lifetime—and Chemistry in Single Molecular Bonds. *http://dx.doi.org/10.1146/annurev.biophys.30.1.105*, 30:105–128, 11 2003. ISSN 10568700. doi: 10.1146/ANNUREV.BIOPHYS.30.1.105. URL `https://www.annualreviews.org/doi/abs/10.1146/annurev.biophys.30.1.105`.

P. Gao, H. Yang, K. R. Rajashankar, Z. Huang, and D. J. Patel. Type V CRISPR-Cas Cpf1 endonuclease employs a unique mechanism for crRNA-mediated target DNA recognition. *Cell Research 2016 26:8*, 26(8):901–913, 7 2016. ISSN 1748-7838. doi: 10.1038/cr.2016.88. URL `https://www.nature.com/articles/cr201688`.

Y. Jeon, Y. H. Choi, Y. Jang, J. Yu, J. Goo, G. Lee, Y. K. Jeong, S. H. Lee, I. S. Kim, J. S. Kim, C. Jeong, S. Lee, and S. Bae. Direct observation of DNA target searching and cleavage by CRISPR-Cas12a. *Nature Communications*, 9(1), 12 2018. ISSN 20411723. doi: 10.1038/s41467-018-05245-x.

D. Kim, J. Kim, J. K. Hur, K. W. Been, S. H. Yoon, and J. S. Kim. Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nature Biotechnology 2016 34:8*, 34(8):863–868, 6 2016. ISSN 1546-1696. doi: 10.1038/nbt.3609. URL `https://www.nature.com/articles/nbt.3609`.

B. P. Kleinstiver, S. Q. Tsai, M. S. Prew, N. T. Nguyen, M. M. Welch, J. M. Lopez, Z. R. McCaw, M. J. Aryee, and J. K. Joung. Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. *Nature Biotechnology*, 34(8):869–874, 1 2016. ISSN 15461696. doi: 10.1038/nbt.3620.

M. Losito, Q. M. Smith, M. D. Newton, M. E. Cuomo, and D. S. Rueda. Cas12a target search and cleavage on force-stretched DNA. *Physical chemistry chemical physics : PCCP*,

23(47):26640–26644, 12 2021. ISSN 1463-9084. doi: 10.1039/D1CP03408A. URL https://pubmed.ncbi.nlm.nih.gov/34494640/.

M. Mora, A. Stannard, and S. Garcia-Manyes. The nanomechanics of individual proteins. *Chemical Society Reviews*, 49(19):6816–6832, 10 2020. ISSN 1460-4744. doi: 10.1039/D0CS00426J. URL https://pubs.rsc.org/en/content/articlehtml/2020/cs/d0cs00426jhttps://pubs.rsc.org/en/content/articlelanding/2020/cs/d0cs00426j.

K. Murugan, A. S. Seetharam, A. J. Severin, and D. G. Sashital. CRISPR-Cas12a has widespread off-target and dsDNA-nicking effects. *The Journal of Biological Chemistry*, 295(17):5538, 4 2020. ISSN 1083351X. doi: 10.1074/JBC.RA120.012933. URL /pmc/articles/PMC7186167//pmc/articles/PMC7186167/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC7186167/.

M. D. Newton, B. J. Taylor, R. P. Driessen, L. Roos, N. Cvetesic, S. Allyjaun, B. Lenhard, M. E. Cuomo, and D. S. Rueda. DNA stretching induces Cas9 off-target activity. *Nature Structural and Molecular Biology*, 26(3):185–192, 3 2019. ISSN 15459985. doi: 10.1038/s41594-019-0188-z.

A. Saha, P. R. Arantes, R. V. Hsu, Y. B. Narkhede, M. Jinek, and G. Palermo. Molecular Dynamics Reveals a DNA-Induced Dynamic Switch Triggering Activation of CRISPR-Cas12a. *Journal of Chemical Information and Modeling*, 60(12):6427–6437, 12 2020. ISSN 15205142. doi: 10.1021/ACS.JCIM.0C00929/SUPPL{_}FILE/CI0C00929{_}SI{_}001.PDF. URL /pmc/articles/PMC7605327//pmc/articles/PMC7605327/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC7605327/.

D. Singh, J. Mallon, A. Poddar, Y. Wang, R. Tippana, O. Yang, S. Bailey, and T. Ha. Real-time observation of DNA target interrogation and product release by the RNA-guided endonuclease CRISPR Cpf1 (Cas12a). *Proceedings of the National Academy of Sciences of the United States of America*, 115(21):5444–5449, 5 2018. ISSN 10916490. doi: 10.1073/PNAS.1718686115/SUPPL{_}FILE/PNAS.1718686115.SAPP.PDF. URL www.pnas.org/cgi/doi/10.1073/pnas.1718686115.

S. Stella, P. Alcón, and G. Montoya. Structure of the Cpf1 endonuclease R-loop complex after

target DNA cleavage. *Nature 2017 546:7659*, 546(7659):559–563, 5 2017. ISSN 1476-4687. doi: 10.1038/nature22398. URL https://www.nature.com/articles/nature22398.

S. Stella, P. Mesa, J. Thomsen, B. Paul, P. Alcón, S. B. Jensen, B. Saligram, M. E. Moses, N. S. Hatzakis, and G. Montoya. Conformational Activation Promotes CRISPR-Cas12a Catalysis and Resetting of the Endonuclease Activity. *Cell*, 175(7):1856–1871, 12 2018. ISSN 0092-8674. doi: 10.1016/J.CELL.2018.10.045.

I. Strohkendl, F. A. Saifuddin, J. R. Rybarski, I. J. Finkelstein, and R. Russell. Kinetic Basis for DNA Target Specificity of CRISPR-Cas12a. *Molecular cell*, 71(5):816, 9 2018. ISSN 10974164. doi: 10.1016/J.MOLCEL.2018.06.043. URL /pmc/articles/PMC6679935//pmc/articles/PMC6679935/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC6679935/.

D. C. Swarts and M. Jinek. Cas9 versus Cas12a/Cpf1: Structure–function comparisons and implications for genome editing, 9 2018. ISSN 17577012.

D. C. Swarts and M. Jinek. Mechanistic Insights into the Cis- and Trans-acting Deoxyribonuclease Activities of Cas12a. *Molecular cell*, 73(3):589, 2 2019. ISSN 10974164. doi: 10.1016/J.MOLCEL.2018.11.021. URL /pmc/articles/PMC6858279//pmc/articles/PMC6858279/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC6858279/.

A. Tafvizi, L. A. Mirny, and A. M. Van Oijen. Dancing on DNA: Kinetic aspects of search processes on DNA. *Chemphyschem : a European journal of chemical physics and physical chemistry*, 12(8):1481, 6 2011. ISSN 14397641. doi: 10.1002/CPHC.201100112. URL /pmc/articles/PMC4590286//pmc/articles/PMC4590286/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4590286/.

T. Yamano, H. Nishimasu, B. Zetsche, H. Hirano, I. M. Slaymaker, Y. Li, I. Fedorova, T. Nakane, K. S. Makarova, E. V. Koonin, R. Ishitani, F. Zhang, and O. Nureki. Crystal structure of Cpf1 in complex with guide RNA and target DNA. *Cell*, 165(4):949, 5 2016. ISSN 10974172. doi: 10.1016/J.CELL.2016.04.003. URL /pmc/articles/PMC4899970//pmc/articles/PMC4899970/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4899970/.

B. Zetsche, J. S. Gootenberg, O. O. Abudayyeh, I. M. Slaymaker, K. S. Makarova, P. Essletzbichler, S. E. Volz, J. Joung, J. Van Der Oost, A. Regev, E. V. Koonin, and F. Zhang. Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell*, 163(3):759–771, 10 2015. ISSN 10974172. doi: 10.1016/J.CELL.2015.09.038/ATTACHMENT/D7072A0B-0438-4D32-89BF-8FCD238CBAEA/MMC3.XLSX. URL http://www.cell.com/article/S0092867415012003/fulltexthttp://www.cell.com/article/S0092867415012003/abstracthttps://www.cell.com/cell/abstract/S0092-8674(15)01200-3.

Chapter 6

AZ-Cas9 characterization at single molecule level

6.1 Introduction

A recent addition to the CRISPR toolbox is a new Cas9 ortholog (MH0245Cas9), here referred to as AZ-Cas9. Currently structural and functional studies to determine AZ-Cas9 activity and specificity are undergoing. Here I discuss my work and contribution to the mechanistic characterization at the single-molecule level of AZ-Cas9. This part of my project is under AstraZeneca disclosure regulation. The information available about the new Cas nuclease discovery can be found within the description of the patent n. WO/2019/099943: "Compositions and methods for improving the efficacy of Cas9-based knock-in strategies".

6.2 AZ-Cas9 introduction

"AZ-Cas9" is a novel Cas9 ortholog (MH0245Cas9) identified in the sequenced gut metagenome and it derives from a bacterial species having a Type II-B CRISPR system (patent available online n. WO2019099943). MH0245Cas9 protein has been engineered by applying nuclear localization signals (NLS) to prevent its degradation and to force its nuclear localization in mammalian cells. As shown in Figure 6.1 A (from the patent), AZ-Cas9 is capable of cleaving a DNA fragment containing the target site *in vitro*. Sanger sequencing analysis of the cleaved products shows that AZ-Cas9 generates 5' overhangs *in vitro* (Figure 6.1 B, from the patent). Furthermore, it has been proved that AZ-Cas9 is also functional in the HEK293 human cell line (Figure 6.1 C, from the patent), generating staggered ends cut. Altogether, those experimental results suggests that AZ-Cas9 probably shares the same mechanism of cleavage of FnCas9 (from *Francisella Novicida*) since they both cause staggered ends cut; however, the analysis of Cas nucleases phylogenetic tree indicates that FnCas9 and AZ-Cas9 are fairly divergent.

The sequence of the gRNA scaffold from AZ-Cas9 has been characterized and a predicted conformation is shown in Figure 6.2. Interestingly, differences in the 3D structure of the AZ-Cas9 gRNA scaffold have been detected in comparison to other Cas nucleases gRNAs (patent n. WO2019099943). AZ-Cas9 gRNA shows a complex secondary structure, which will need to be confirmed by structural analysis, and positions 9-26 are predicted to form an extra stem loop, absent in SpCas9, FnCas9 or Cas12a (Figure 6.2). First attempts of guide engineering have been made in AZ to decrease the guide size and increase its activity. The last and most efficient version so far, is composed of an 83 nt long tracrRNA (whereby it is 93 nt long for SpCas9 gRNA) and a 22 nt long crRNA, 2 nt longer than SpCas9 (see List of Reagents for sequence details).

The preferred PAM sequence for AZ-Cas9 has also been investigated by using a pooled library of 64 plasmids, which covers various PAM sequence combinations and a target cleavage site. Results show that, although being able to recognize the "NGG" PAM, AZ-Cas9 has a less stringent preference for this canonical PAM sequence compared to SpCas9 (patent n. WO2019099943).

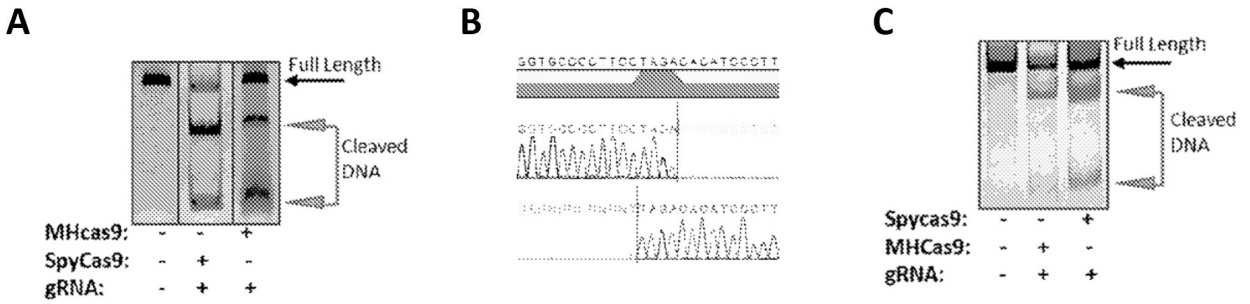


Figure 6.1: AZ-Cas9 cleavage assay. **(A)** Gel electrophoresis after *in vitro* cleavage assay, performed with SpCas9 and AZ-Cas9 assembled with their relative gRNAs. **(B)** Sanger sequencing analysis of cleaved fragment from (A) showing that AZ-Cas9 generates 5' overhangs *in vitro*. **(C)** Gel electrophoresis of genomic DNA at the cut site, after cleavage performed in HEK293 cells with SpCas9 and AZ-Cas9. Figures from patent n. WO2019099943.

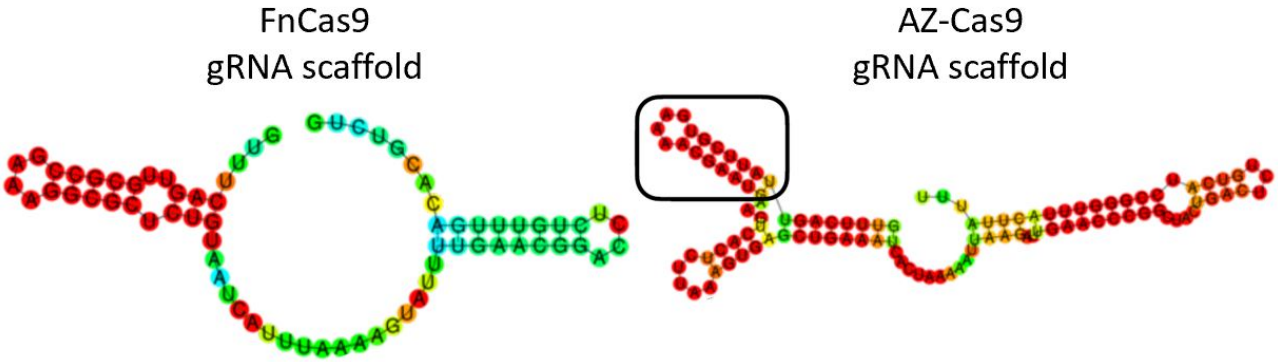


Figure 6.2: Structure of AZ-Cas9 gRNA scaffold. Comparison of 3D structures of FnCas9 and AZ-Cas9, which shows an extra stem loop (position 9-26) totally absent in both SpCas9 and Cas12a too. Figure from patent n. WO2019099943.

6.3 Single Molecule Results

6.3.1 AZ-Cas9 *in vitro* cleavage assay with fluorescent gRNA

I decided to investigate the interaction of AZ-Cas9 on λ-DNA, and, since the recognized PAM is the canonical NGG PAM, also recognized from SpCas9, the crRNA sequence was selected to target the unique site "λ2" already used for previous SpCas9 single molecule analysis (Newton et al., 2019; Sternberg et al., 2014). The crRNA complementary to λ2 site was ordered, purified and labelled in house with Cy3-fluorescent dye (Materials and Methods). The tracrRNA, which also differs in terms of sequence from tracrRNA for SpCas9 and Cas12a, was provided by AZ. In preparation for the single molecule studies, I firstly performed an *in vitro* cleavage experiment (Material and Methods) to confirm that the labelling of the guide does not affect the cleavage activity (Figure 6.3). I used a Cy3-labelled 90 nt long dsDNA oligo containing λ2 target sequence (List of Reagents), and I incubated the dsDNA with the active RNP, previously assembled *in vitro* with the Cy3-labelled crRNA (Materials and Methods), at 37 °C. I then stopped the reaction at distinct time points and analysed the cleavage products via gel electrophoresis (Material and Methods). After 1 minute the cleavage reaction is at 50% and after 5 minutes it is almost completed (Figure 6.3). I then concluded that the labelling does not affect the nuclease activity *in vitro*.

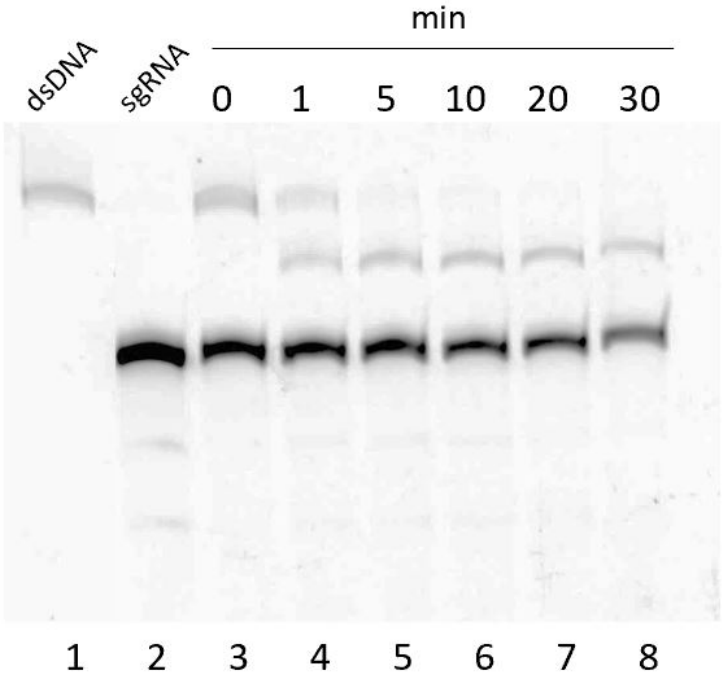


Figure 6.3: AZ-Cas9 *in vitro* cleavage assay. Cy3-labelled λ2-crRNA and tracrRNA were resuspended in a 1:1 ratio in annealing buffer (100 mM NaCl, 50 mM Tris-HCl pH 8), heating to 90 °C for 2 minutes, and cooled at room temperature for 15 minutes. The resulting sgRNA was then complexed with purified AZ-Cas9, at a final concentration of 1 μM. 10 nM of labelled Cy3-labelled-dsDNA oligo containing the λ2 target sequence, were incubated with 100 nM RNP at 37 °C x 30 minutes total, in Digest Buffer (100 mM TRIS-HCl pH 7.5, KCl 500 mM, 25% glycerol, 5mM DTT, 50 mM $MgCl_2$). Every 5 minutes, the reaction was stopped with Quenching Solution (0.5 M EDTA pH 8). Cleavage products were then analysed via gel electrophoresis in 1% Agarose gel, and imaged with 532 nm laser for Cy3 excitation. Lane 1: Cy3-labelled-dsDNA only. Lane 2: Cy3-labelled λ2-sgRNA only. Lanes 3-8: cleaved products at the indicated time points.

6.3.2 AZ-Cas9 stably binds the on target site with high affinity

Similarly to Cas12a analysis (Chapter 5), I used the C-trap microscope to study the interaction of AZ-Cas9 with DNA in real time and at single-molecule resolution (Materials and method). To assemble the RNP complex, unlabelled tracrRNA was first annealed with a 5'-Cy3-labeled crRNA targeting λ2 site, and the gRNA was then complexed with catalytically active AZ-Cas9 (Figure 6.4 A). λ2 site is located at ∼18 kb along the length of λ-DNA (∼48.5 kb). I performed the C-trap experiments using the microfluidic flow cell with the same set-up previously shown for Cas12a analysis (Figure 6.4 B), with channels 4 and 5 specifically containing buffer with Ca^{2+} or Mg^{2+}only (Figure 6.4 B) to be able to study separately the binding and the cleavage activity of the nuclease. Streptavidin-coated polystyrene beads (4.84 μm) and torsionally unconstrained biotinylated λ-DNA were used for the analysis (Material and Methods). Initially, I stretched the DNA to its contour length by using a force of 5 pN, in the presence of 250 pM labelled RNP complex in Ca^{2+} buffer, which was the RNP concentration used for Cas12a experiments. Under these conditions and RNP concentration I could barely detect any binding event. I then increased the RNP concentration and started to see single binding events at the expected on-target location with 20 nM RNP, a 100x times higher concentration than Cas12a and 2x higher than SpCas9 used in previous single molecule studies (Newton et al., 2019). This can be related to the PAM-recognition process and limitations, observed with the previously described *in vitro* plasmid cleavage assay, which showed less AZ-Cas9 affinity for the NGG-PAM. However, in my experiments, once the protein binds the on-target site, the binding is stable for minutes (Figure 6.4 C). I can then confirm that AZ-Cas9 is able to locate the on-target site and stably bind to it with high affinity, but, an excess of protein is required for the binding to occur, particularly when compared to Cas12a. Additionally, I can also observe a single mode of AZ-Cas9 binding, occurring via 3 dimensional collision at the target site and never diffusing on the dsDNA. This is the first observation which gives us information about the target search mechanism of this new Cas9 variant.

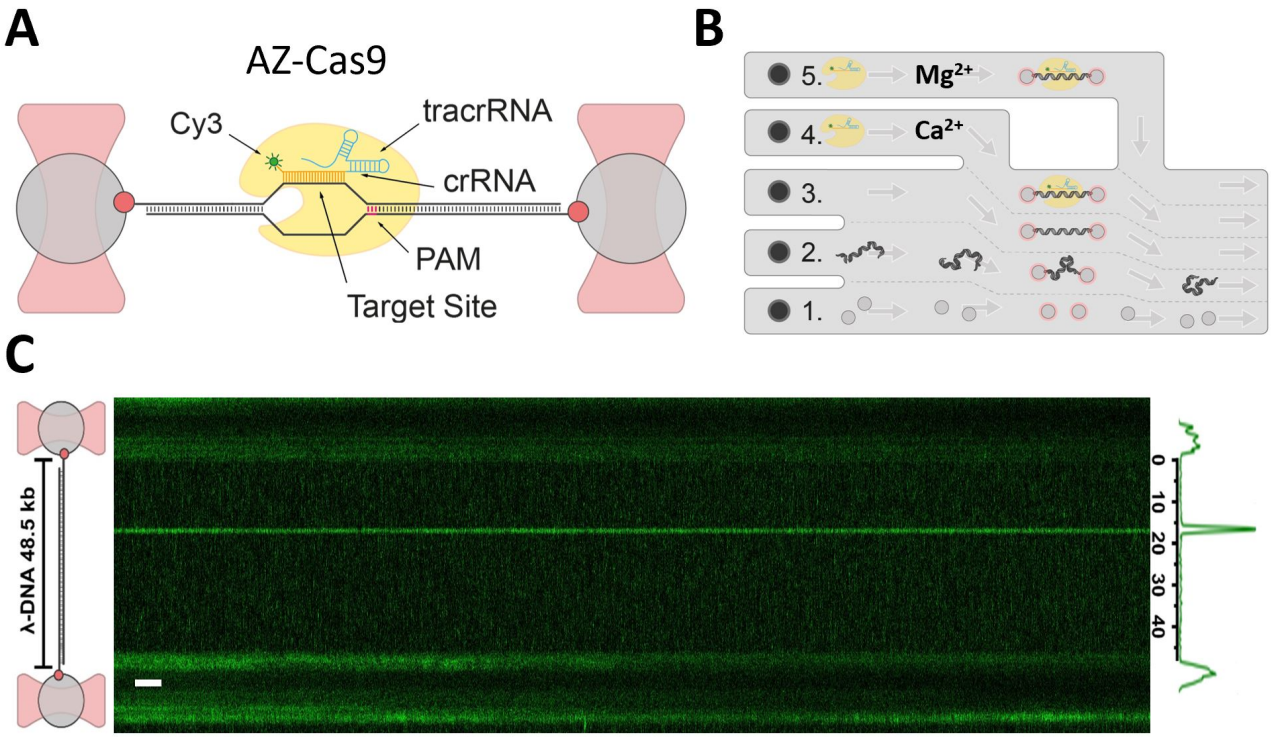


Figure 6.4: C-trap set up AZ-Cas9 on-target binding analysis. **(A)** Schematic of biotinylated λ-DNA tethered between two optically trapped beads, with bound Cy3-labelled AZ-Cas9 complex **(B)** C-trap microscope microfluidics : (1) beads channel; (2) DNA channel; (3) buffer-only channel; (4) AZ-Cas9 in Ca^{2+}-buffer; (5) Az-Cas9 in Mg^{2+} - buffer. **(C)** Kymograph of λ-DNA held at 5 pN in the presence of 20 nM Cy3-labelled RNP, in Ca^{2+} - buffer. A single AZ-Cas9 complex is shown stably bound at the on target for minutes. Genomic location analysis confirms the expected on-target at 18 kb. Scalebar = 2 s.

6.3.3 DNA stretching induces AZ-Cas9 off-target binding but with a fast dissociation rate

Next, I investigated how increasing the tension on the DNA affects AZ-Cas9 binding activity. When I increase the force of DNA stretching, AZ-Cas9 behaves as previously shown for Sp-Cas9 and Cas12a. As expected, the distortion of DNA due to mechanical stretching, favours additional binding events at off-target locations (Figure 6.5 A). To quantify this, I determined the number of off-target binding events per 10.000 bp as a function of force and I can detect a linear dependency (Figure6.5 B). Indeed, I found that the number of off-target binding events increase linearly with the force applied to stretch the dsDNA (Figure 6.5 B). According to my measurements, I can detect a maximum of ∼2 complexes bound per 10 000 bp, which is comparable to the previous study on Cas12a (Chapter 5, Figure 5.8 B). To investigate further on the specificity properties of this new variant and to understand the force dependence of the off-target binding kinetics, I measured the dwell times for >200 off-target binding events of the AZ-Cas9 RNP complex, in Ca^{2+} buffer, across λ-DNA held at 50 pN (Figure 6.5 A), and constructed a dwell-time histogram (Figure 6.5 C). The resulting cumulative probability distributions (Figure 6.5 D) revealed a fast and a slow population of off-target dissociation events. This was also previously shown for SpCas9 (Newton et al., 2019), and it is possible to speculate that both SpCas9 and AZ-Cas9 share an off-target binding kinetics which is happening with a two binding mode: a transient (short-lived) and a meta-stable (longer-lived) mode. However, for SpCas9, the average dwell time of binding events were respectively $T_{fast} = 9.0 \pm 0.6$ s and $T_{slow} = 61 \pm 8$ s (mean ± s.e.m.), and it was shown that the slow mode becomes dominant at higher forces (Newton et al., 2019). Interestingly, by comparing those results with my analysis on AZ-Cas9, I overall see much short lived off-target binding kinetic. Indeed, at high forces (50 pN), the off-target dwell time of both the transient and the stable events, are 10 time shorter than SpCas9, with a $T_{fast} = 1.5 \pm 0.1$ s, and the $T_{slow} = 6.4 \pm 0.7$ s (mean ± s.e.m.). This fast dissociation kinetic at the off-target sites can be at the base of AZ-Cas9 high-specificity properties.

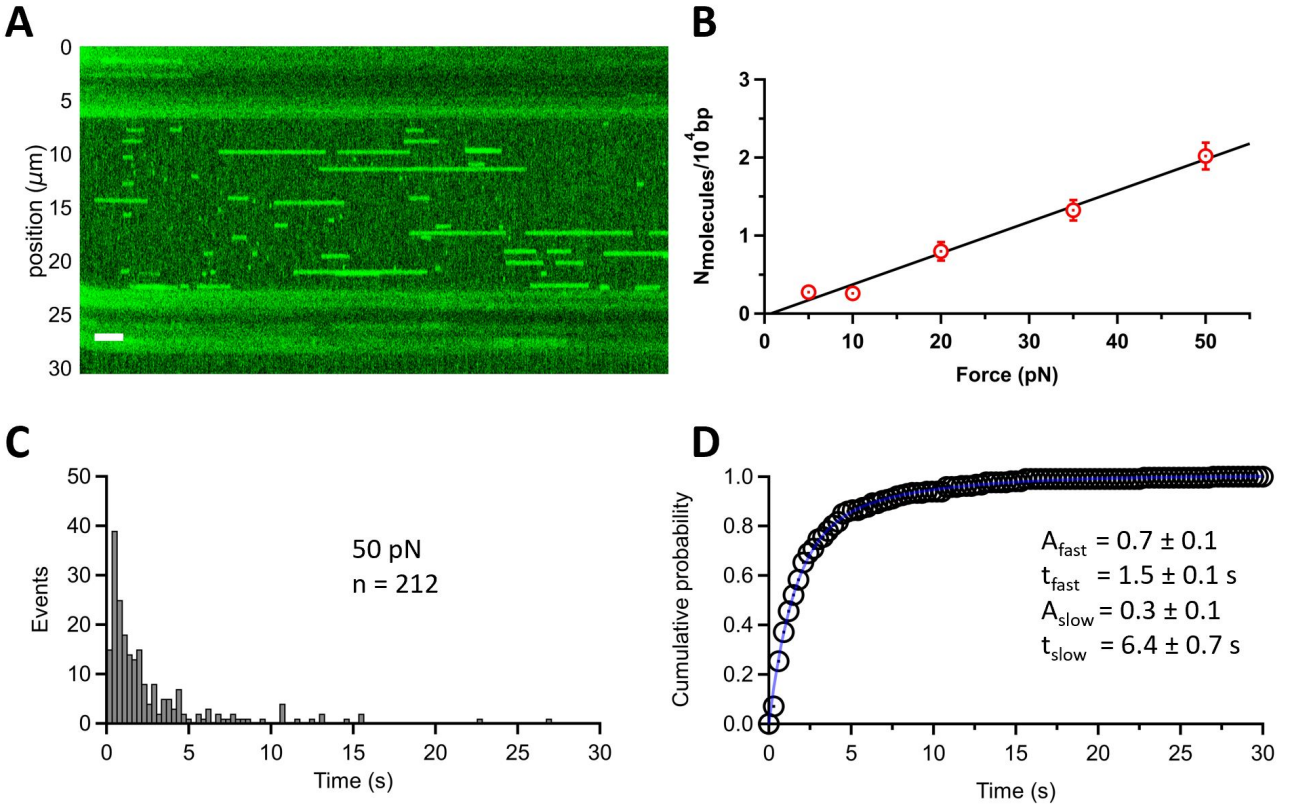


Figure 6.5: Force induces off-target binding of AZ-Cas9 but with a fast dissociation rate. **(A)** Kymograph of λ-DNA in the presence of 20 nM Cy3- λ2-RNP:AZ-Cas9 complex (green signal) held at 50 pN, in 10 mM Ca^{2+} and 100 mM Na^{+}. Many off-target binding events at different sites are observed. **(B)** Number of molecules bound per 10 000 bp as a function of force (average and s.e.m., n $\geq$ 7 each), and fit to the a linear (line). **(C)** Dwell time histogram and **(D)** cumulative probability of distribution for off-target binding events of λ2-crRNA induced at 50 pN.

Interestingly, I can detect same off-target locations appearing to be bound multiple times, suggesting an underlying sequence-dependence of the induced sites. To further test this observation, a single DNA molecule was pulled and held at 40 pN for 300 s and I time-binned the kymographs using 20 s windows. I obtained 13 intensity-plot traces (Figure 6.6) which I used to map the inter-beads distance to the λ-DNA sequence. The resulting time-binned intensity plot (Figure 6.6) shows a consistent peak around 18 kbp which I identify as the on-target site (Figure 6.6, red star) and a few off-target sites which appear consistently along the kymograph (Figure 6.6, grey dashed lines). Intriguingly, some of these off-target sites on λ-DNA were previously detected to be bound by λ2-crRNA:SpCas9 complex (Newton et al., 2019), such as peaks n1, n3 and n7 (Figure 6.6). The crRNA sequence used for AZ-Cas9 is identical to the one used for SpCas9, with the exception of 2 extra nucleotides, therefore, this correspondence in some of the preferred off-target sites between the two nucleases is probably dictated by the guide sequence itself. To confirm the sequence dependency of these off-target sites, I next compared the experimental results with the position of predicted mismatched sites. I used the Cas-OFFinder software, available online (http://www.rgenome.net/cas-offinder/), to check for off-target prediction. By using the λ2-22 nt crRNA sequence as input and inquiring for 5'-NRG-3' PAM availability (with R = A or G), the software found 11 possible off-target sites, with a total of 8 or 9 nucleotides mismatches (Table 6.1). These sites were compared to the experimental data for the λ2-crRNA:AZ-Cas9 complex binding sites on DNA stretched at 40 pN (Figure 6.6). Almost all the predicted off-target sites, had a corresponding binding peak in the experimental data, even if, due to the resolution of the density-binned plot, it is difficult to discern from possible off-target sequences which are close to each other (e.g. positions 1129 - 1107). Those data demonstrate that these force-induced off-target events are non-random but crRNA sequence-dependent, suggesting they are in part mediated by base pairing between the DNA and crRNA. These sites and the sequences showed in Table 6.1 can possibly be selected and used for further off-target kinetic analysis.

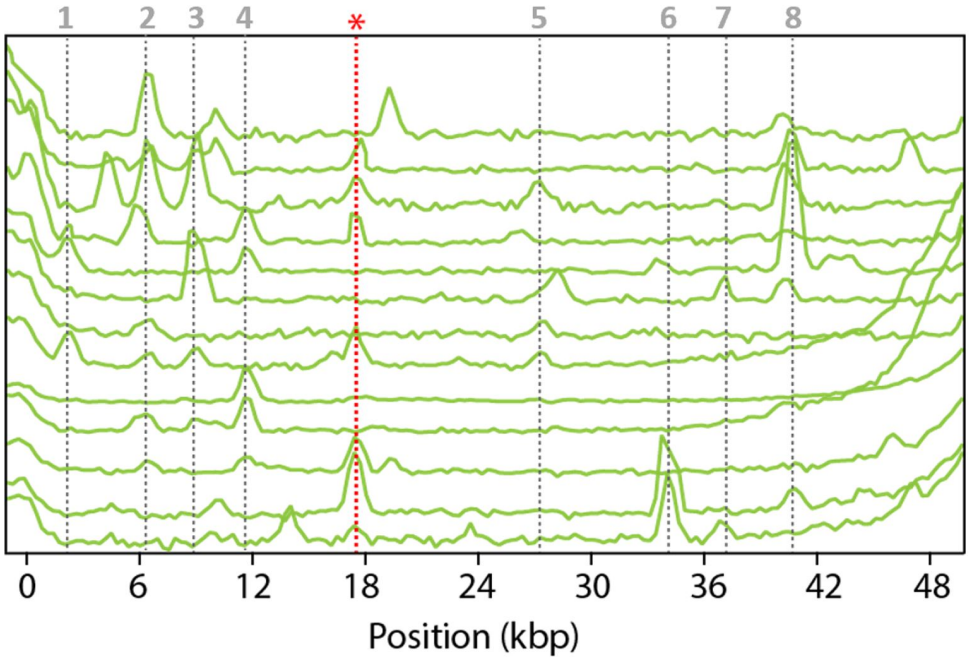


Figure 6.6: AZ-Ca9 off-target position analysis. Time-binned intensity plot AZ-Cas9 binding at 40 pN, mapped to λ-DNA, showing genomic positions of force-induced off-target binding. On-target site is indicated by a red star and red dashed line. Off-target sites which appear to be bound multiple times are indicated by grey dashed lines. Each trace represents 20 s of a 300 s total kymograph. A reference number has been assigned to each off-target site to be related to a predicted off-target sequence (Table 6.1).

Table 6.1: Predicted AZ-Cas9 off-target sites for $\lambda 2$ crRNA sequence.

cRNA	DNA	Position (bp)	mis-matches	direction	off-target ref. n.
CAGTGATAAGTGG AATGCCATGNRG	CAGTGATAAGTGG AATGCCATGTGG	18068	0	+	*
CAGTGATAAGTGG AATGCCATGNRG	CAGTaATtAcgGt gcTGCgcTGGAG	5001	9	+	2
CAGTGATAAGTGG AATGCCATGNRG	AaTGgccgGTGG cATGtCATGCAG	8362	8	+	3
CAGTGATAAGTGG AATGCCATGNRG	CAGTGAggAGaa tAAcatCAcGCAG	6601	9	+	3
CAGTGATAAGTGG AATGCCATGNRG	CcGgGATAAcacG ctcaCCATGAAG	1129	9	+	1
CAGTGATAAGTGG AATGCCATGNRG	CgGgGATAAcTtc AtTcCCgTcCGG	10435	9	-	4
CAGTGATAAGTGG AATGCCATGNRG	CgGTGcTtttTGc cATaCCAcGGGG	1107	9	-	1
CAGTGATAAGTGG AATGCCATGNRG	tAcTGATgtGatG gcTGCtATGGGG	43977	9	+	8
CAGTGATAAGTGG AATGCCATGNRG	CgtTGATAAGTc GcAgatCAgcAGG	38460	9	+	7
CAGTGATAAGTGG AATGCCATGNRG	CccTtATtgGgGGt AaGaCATGAAG	45166	9	+	8
CAGTGATAAGTGG AATGCCATGNRG	tAtcGAagAGTGc AAgGCgATcAAG	42961	9	+	8
CAGTGATAAGTGG AATGCCATGNRG	agGTttTAcGcaGA ATGgCAaGCAG	38841	9	-	7

6.3.4 AZ-Cas9 cleavage rate suggests high-fidelity properties

Next, I tested whether the AZ-Cas9 complex could cleave force-stretched λ-DNA in Mg^{2+} containing buffer and how force affects the cleavage activity. At low forces (6 pN, Figure 6.7) the AZ-Cas9 was observed bound to its specific on-target site, as expected, and cleavage was detected after ~300 s as a drop in the force between the beads and a displacement of the force-clamped bead (Figure 6.7 A). This result confirms that the AZ-Cas9 complex is active under these experimental conditions. To investigate further on the cleavage kinetics, I measured the dwell time at the target site from initial binding to cleavage, at different forces (5 to 30 pN) (Figure 6.7 B). Interestingly, I observed an inverse correlation compared to Cas12a behaviour since (Chapter 5, Section 5.2.2), as in this case, force seems to favour the cleavage. In order to obtain further mechanistic information about the force dependency of the cleavage activity, I calculated the rate constant k_{cleave} at each force as the inverse of the average dwell time from binding to cut. I then fit the k_{cleave} values to the Bell-Evans equation (Evans 2003; Mora et al. 2020, Equation 5.1, Chapter 5, Section 5.2.2). From the intercept of the curve to the logarithmic scale k_{cleave} values (figure 6.7 C), I could calculate the zero-force cleavage rate constant $k_{cleave(0)} = 0.003 \pm 0.01\ s^{-1}$, which is one order of magnitude slower than Cas12a $k_{cleave(0)}$ ($0.08 \pm 0.01\ s^{-1}$), previously calculated in similar experimental condition (Chapter 5, Section 5.2.2). Interestingly, previous kinetic analysis performed on high-fidelity Cas9 variants, HypaCas9 and Cas9-HF1, showed a dramatic decrease in the observed decay rate for cleavage of on-target DNA substrates of about 21- to 35-fold compared to SpCas9 ($0.028\ s^{-1}$ for HypaCas9 and $0.047\ s^{-1}$ for Cas9-HF1 vs $1\ s^{-1}$ for SpCas9) (Liu et al., 2020). In that study, it was proposed that HypaCas9 and Cas9-HF1 gain discrimination mainly through further slowing of the rate of DNA cleavage with off-target DNA, which shifts kinetic partitioning to favour release rather than cleavage of the bound DNA. For SpCas9, the conformational change is rate-limiting and determines specificity because R-loop formation is largely irreversible and is followed by fast DNA cleavage (Gong et al., 2018). According to these first measurements on AZ-Cas9 cleavage at the on-target site, I can speculate that this variant shares a similar mechanism with the other higher-fidelity variants, showing slower rates of DNA cleavage, which then will allow for release of the off-target substrate from the enzyme before the irreversible cleavage reaction. I hypothesise that this new variant can perform with high-fidelity properties, which

are presumably achieved by dramatically increasing the rate of dissociation of mismatched nucleotides, further enhanced by the mechanical distortion and opening of the dsDNA, in addition to decreasing the rate of catalysis.

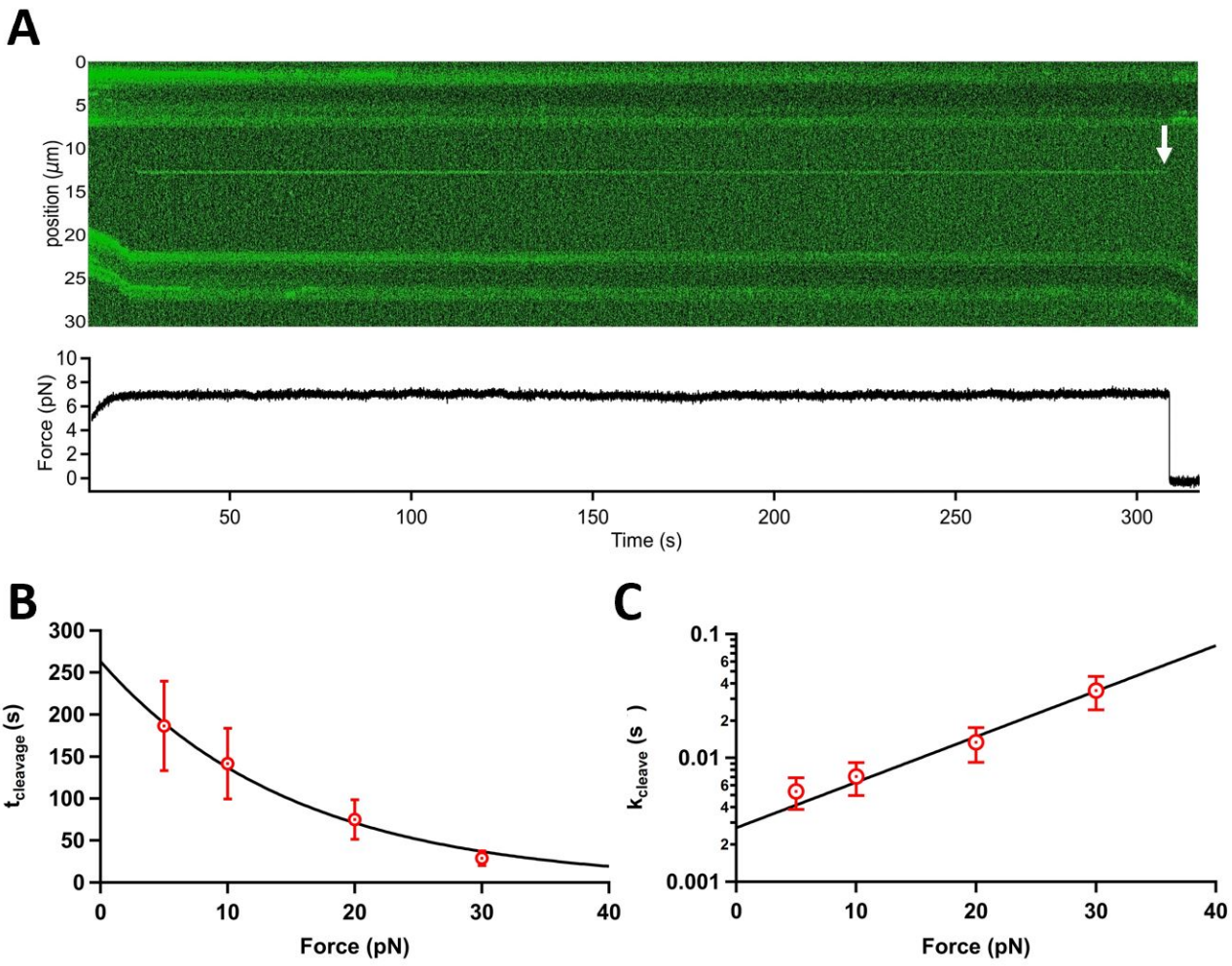


Figure 6.7: AZ-Cas9 rate cleavage analysis. **(A)** Kymograph of force clamped λ-DNA held at 6 pN in the presence of 20 nM Cy3-λ2-crRNA:AZ-Cas9, in 10 mM Mg^{2+} and 100 mM Na^{+}. A single RNP complex molecule is show bound at the expected target site. Cleavage is observed at 300 s as fluorescence loss (white arrow), displacement of the trapped bead, and drop in the force (bottom). **(B)** Average dwell time and standard error of the mean, at forces ranging from 5 to 50 pN. The rate constant k_{cleave} was approximated as the inverse of the average dwell time at each force ($n \geq 6$ each) **(C)** Cleavage rate constant as a function of force (average and standard error of the mean), and fit to the Bell-Evans equation (line).

6.3.5 FRET analysis for R-loop formation at the on-target

To further investigate the on-target binding mechanism and monitor the R-loop formation step, I used a previously described single-molecule FRET assay (Rueda et al., 2017) (Figure 6.8). A dsDNA oligo containing the on-target $\lambda 2$ sequence was labelled in house with Cy3 (Material and Methods) and immobilised on the slide surface, via biotin-streptavidin interaction (Materials and Methods). AZ-Cas9 was assembled *in vitro* with a Cy5-labelled sgRNA targeting the $\lambda 2$ site (Figure 6.8 A). Upon addition of the Cy5-RNP complex, the Cas binding at the target site, results in a FRET signal between the Cy3 (donor) and Cy5 (acceptor). In this assay, full R-loop formation upon binding to the on-target site, results in stable high FRET state ($\sim$1 FRET, Figure 6.8 B), which was detected in the majority of the single molecule trajectories (79%, n=61/77). In 27 of the 77 traces analysed, it was possible to detect a transient intermediate state ($\sim$0.6 FRET, Figure 6.8 B, black arrow), with an average dwell time in the range of ms. Based on previous studies on SpCas9, I assigned these trajectories to a possible initial binding in an open conformation ($\sim$0.6 FRET, arrow) followed by zipping into a full heteroduplex conformation ($\sim$1 FRET). A time-binned FRET histogram of 77 trajectories with the fully matching crRNA shows that the complex predominantly exists in the fully heteroduplex conformation (Figure 6.9 B), consistent with the tight binding previously observed. However, the histogram reveals the presence of another intermediate state ($\sim$0.3 FRET) which can be probably due to the RNP transiently exploring different states while interrogating for the guide sequence complementary. This is also confirmed by a cumulative density plot, which mainly shows the transient intermediate ($\sim$0.6 FRET) (Figure 6.9 A). As previously suggested, by performing this type of analysis with oligo target DNA containing the identified potential off-targets sequences (Table 6.1), differences in R-loop formation mechanism and dynamic can be elucidated and used to acquire new information about AZ-Cas specificity. Indeed, in the presence of mismatches, a decrease in the high-FRET population could be expected, with a consequent broadening of the intermediate-FRET distribution. This would possibly indicate that, at the off-target sites, the formation of the high-FRET conformation is hindered, as well as the cleavage.

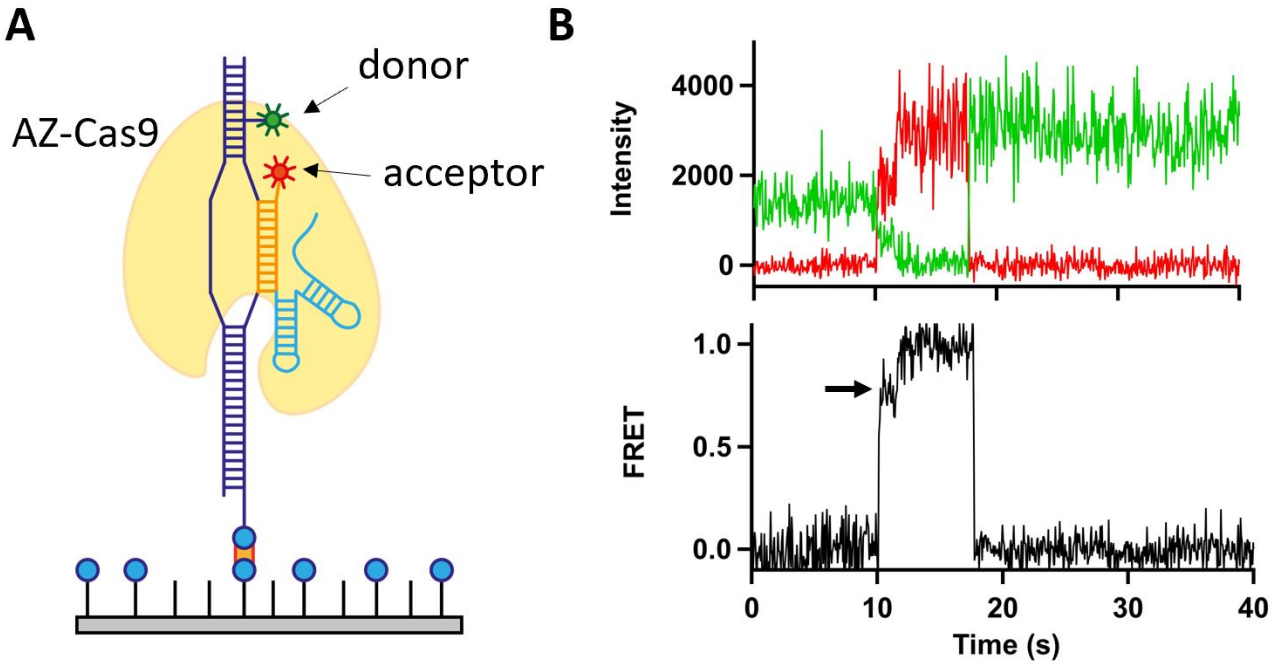


Figure 6.8: smFRET assay for AZ-Cas9 on-target binding analysis. **(A)** Schematic of smFRET AZ-Cas9 binding and R-loop formation experiments. Target DNA, labelled with Cy3 (green) on the target strand, is immobilised on the microscope slide. Binding and R-loop formation of AZ-Cas9 complex labelled with Cy5 (red) results in FRET. Target site is coloured in orange **(B)** Example single-molecule trajectory for Cy5-RNP complex binding to the Cy3 labelled λ2 on-target site (top). Corresponding calculated FRET trajectory (bottom). On-target binding results in high-FRET state. A fast intermediate step, corresponding to the dsDNA opening upon R-loop formation, is detectable (black arrow).

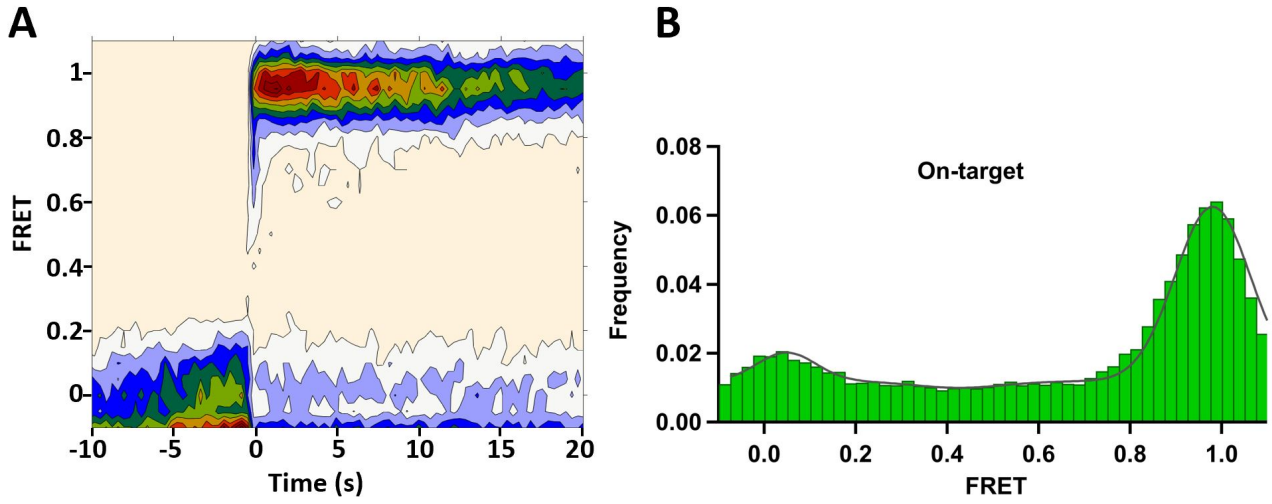


Figure 6.9: Cumulative Density Plot and FRET histogram **(A)** Cumulative Density Plot and **(B)** Time-binned FRET histogram of all binding events (n = 77) detected for AZ-Cas9 in complex with the fully matching crRNA, fit with a multiple-Gaussian (line).

6.4 Discussion

A new Cas9 ortholog, here referred to as AZ-Cas9, has been identified in the sequenced gut metagenome and engineered by the AstraZeneca Discovery Science team to be used as a gene-editing tool in mammalian cells. Preliminary *in vitro* and *in cellulo* analysis showed that AZ-Cas9 can interact with an NGG PAM, recognize 22-nt long target sites, and generate staggered ends upon cleavage. By using the previously described single-molecule assay, I investigated the dynamics of AZ-Cas9 interaction with dsDNA, its on-target search mechanism and R-loop formation, on-target cleavage activity, and off-target discrimination. The experiments here described demonstrate that the complex recognizes the on-target sequence through a 3D collision mechanism and, similarly to SpCas9, it never diffuses on the dsDNA. In this regard, it is worth mentioning that the analysis of the trajectories recorded within the C-trap microscope are subjected and limited to the resolution of the tracking algorithm. Indeed, the fluorescence intensity peaks tracking along the kymographs occurs at the level of 100 or 50 nm sized pixels, which correspond to $\sim$ 300-150 bp. The algorithm used in these analysis fits the pixel intensity values into a Gaussian curve, which further increase the resolution to values smaller than the pixel size. Moreover, an apparent diffusion of the protein can be detected as a consequence of the system itself not being fully static, since the DNA is tethered between the beads and it has its own elasticity, and tracking inaccuracy can exist due to fluorescence background. Hence, by looking at static traces, it is not possible to completely understand if the protein is stably and tightly bound or if it is moving of a few nucleotides. However, the trajectories recorded with AZ-Cas9 definitely exclude a 1D search mechanism similar to Cas12a (which travels $\sim$1 kbp per second) and it is the first observation giving us information about the mode of AZ-Cas9 target search dynamic. Once the on-target binding is established upon PAM recognition, the R-loop is formed and, in the absence of Mg^{2+}, the complex can remain bound for minutes. Nonetheless, the binding measurements show a lower binding efficiency at the on-target site compared to SpCas9, since a higher protein concentration is needed. This dynamic could be correlated to the PAM-recognition process, but further analysis of the PAM-binding process are required. Stretching the DNA with force leads to the detection of increased off-target binding events, similarly to SpCas9 and Cas12a behaviour (Newton et al. 2019, Chapter 5). Once again, this corroborates the hypothesis that force-induced DNA stretching facilitates Cas

nucleases binding in the presence of mismatches, by lowering the DNA melting energy barrier. Intriguingly, by measuring the dwell time of >200 off-target biding events on dsDNA stretched at 50 pN, I detected a dissociation constant which is 10 times faster than previously reported for SpCas9 (Newton et al., 2019). In addition, I was able to calculate the $k_{cleave(0)}$ of AZ-Cas9 at the on-target site and I found that this rate is in line with previous *in vitro* calculations done on High-Fidelity Cas nucleases (Liu et al., 2020). These results suggest that AZ-Cas9 can perform with high-fidelity accuracy, although further *in vitro* and *in vivo* analyses, together with structural studies, are required to confirm this hypothesis.

Bibliography

E. Evans. Probing the Relation Between Force—Lifetime—and Chemistry in Single Molecular Bonds. *http://dx.doi.org/10.1146/annurev.biophys.30.1.105*, 30:105–128, 11 2003. ISSN 10568700. doi: 10.1146/ANNUREV.BIOPHYS.30.1.105. URL `https://www.annualreviews.org/doi/abs/10.1146/annurev.biophys.30.1.105`.

S. Gong, H. H. Yu, K. A. Johnson, and D. W. Taylor. DNA Unwinding Is the Primary Determinant of CRISPR-Cas9 Activity. *Cell Reports*, 22(2):359–371, 1 2018. ISSN 2211-1247. doi: 10.1016/J.CELREP.2017.12.041.

M. S. Liu, S. Gong, H. H. Yu, K. Jung, K. A. Johnson, and D. W. Taylor. Engineered CRISPR/Cas9 enzymes improve discrimination by slowing DNA cleavage to allow release of off-target DNA. *Nature Communications*, 11(1), 12 2020. ISSN 20411723. doi: 10.1038/S41467-020-17411-1. URL `/pmc/articles/PMC7367838//pmc/articles/PMC7367838/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC7367838/`.

M. Mora, A. Stannard, and S. Garcia-Manyes. The nanomechanics of individual proteins. *Chemical Society Reviews*, 49(19):6816–6832, 10 2020. ISSN 1460-4744. doi: 10.1039/D0CS00426J. URL `https://pubs.rsc.org/en/content/articlehtml/2020/cs/d0cs00426jhttps://pubs.rsc.org/en/content/articlelanding/2020/cs/d0cs00426j`.

M. D. Newton, B. J. Taylor, R. P. Driessen, L. Roos, N. Cvetesic, S. Allyjaun, B. Lenhard, M. E. Cuomo, and D. S. Rueda. DNA stretching induces Cas9 off-target activity. *Nature Structural and Molecular Biology*, 26(3):185–192, 3 2019. ISSN 15459985. doi: 10.1038/s41594-019-0188-z.

F. O. Rueda, M. Bista, M. D. Newton, A. U. Goeppert, M. E. Cuomo, E. Gordon, F. Kröner, J. A. Read, J. D. Wrigley, D. Rueda, and B. J. Taylor. Mapping the sugar dependency for rational generation of a DNA-RNA hybrid-guided Cas9 endonuclease. *Nature Communications*, 8(1), 12 2017. ISSN 20411723. doi: 10.1038/s41467-017-01732-9.

S. H. Sternberg, S. Redding, M. Jinek, E. C. Greene, A. Jennifer, M. Biophysics, C. Biology, B. Division, and L. Berkeley. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*, 507(7490):62–67, 2014. doi: 10.1038/nature13011.DNA.

Chapter 7

Concluding Remarks

7.1 Summary

In this thesis, I combine cell-based assays and single-molecule techniques to study the effect of DNA distortion on Cas nucleases. In Chapter 3, I describe a cellular strategy which proves that a local DNA distortion, proximal to the cleavage site, causes an enhancement of Cas9 off-target activity. I use an adapted version of the previously described Proxy-CRISPR system, to target a wtCas9, complexed with a short sgRNA, to a site proximal to the intended KO site. The short guide is able to direct Cas9, allowing for binding but preventing cleavage. I demonstrate that the presence of Proxy-CRISPR affects Cas9 specificity at a nearby target site; in fact, multiple KO guides, with different types of mismatches in different positions, perform with higher editing efficiency when co-transfected with the Proxy-CRISPR. In my work, I demonstrate this effect both on an artificial reporter system (TLR), integrated into the genome of the cells, and on an endogenous target (EMX1). An effect similar to the one caused by the Proxy-CRISPR on the DNA conformation may occur naturally on the eukaryotic DNA at different stages of the cell cycle, caused by endogenous processes, such as transcription, epigenetic modification, DNA repair, recombination and supercoiling. In Chapter 4, I discuss a second method designed to look at DNA during transcription, when a gene-wide "DNA opening" occurs. Specifically, I act on the promoter of silenced or low expressed genes to investigate the effect on Cas9 on-target activity. In my experiments, I use CRISPRa tools for upregulation of gene transcription, either by using cell lines which expresses the fusion dCas9-VPR activation system or by using modified gRNA to recruit activators. Although limited to a selected number of gene targets

and cell lines, my data shows high variability in the CRISPRa effect on gene expression, with a weak effect on Cas9 on-target editing efficiency. Whereas Cas9 has been proved to have diminished accessibility at genes embedded in a closed-chromatin environment, here I find that this effect is highly dependent on the target gene itself, the cell line used, and the design of sgRNAs. The last two chapters describe single-molecule approaches for mechanistic and functional characterization of two different CRISPR systems. In Chapter 5 I focus on Cas12a activity on λ-DNA, and I report important information about the target search mechanism, the cleavage kinetics and how these are affected by force-induced DNA stretching. By using correlative single-molecule optical tweezers with fluorescent detection I show that Cas12a can bind away from the target on the DNA and, differently from Cas9, diffuses randomly and bidirectionally until it recognises its target. Furthermore, I show that Cas12a search mechanism involves hopping on the DNA. Stretching the DNA with force hinders cleavage, revealing a large amplitude rate-limiting conformational change, which can be correlated to a "closing" process of the enzyme to bring the target strand into the active nuclease RuvC site. Finally, in Chapter 6, I describe my contribution to the functional characterization of a new Cas9 variant, AZ-Cas9, and I uncover new details about its activity on DNA. I prove that, similarly to SpCas9, the main search mechanism of AZ-Cas9 occurs via 3D collision on the DNA and that once the on-target is recognized, the R-loop is quickly formed, and the binding is stable. AZ-Cas9 off-target binding activity is also affected by force-induced DNA stretching, but my findings also suggest possible high-fidelity properties.

7.2 Future directions

The first part of this work highlights how it is extremely important to understand the highly variable specificity of Cas9 depending on the DNA structure and the genomic context, particularly when considering the therapeutic application. Up to date, the prediction of potential off-target activity in cellular assays mainly relies on guides-sequence screening; however, this may not properly predict the potential for off-target activity across all the cell types and genomic states present in an entire organism. Moreover, the findings discussed in Chapter 3 which show how non-canonical wobble base pairing and purine-purine mismatches can be highly tolerated in the Proxy-TLR experiments, suggest that future prediction algorithms should incorporate these factors to better predict potential off-target sites. In both the Proxy results on the TLR and on EMX1, Cas9 editing efficiency is found to be distinctly dependent on the mismatch identity and also the position of the mismatch within the guide. Interestingly, in the Proxy and CRISPRa assays, Cas9 on-target activity seems to be less affected, but genome-wide and high throughput analysis could be performed to better understand the effect of genomic context and gene expression on Cas9 on-target binding and cleavage mechanisms. Altogether, this highlights the complexity of the rules governing Cas9 specificity and calls attention to the need for additional work to precisely define how the sequence requirements are altered in the context of DNA distortion. Additionally, as discussed, these results can be correlated to a mechanism of a tilt in the energy landscape at the base of Cas9 specificity, which is altered by the target DNA structure. This is thought to affect the threshold of the conformational checkpoint required for Cas9 cleavage, which is controlled by a sequence-mediated inspection but here appears to be also dependent on the target DNA conformation. In this perspective, using kinetic models which could take into account the tilting caused by supercoiling or mechanical stretching, would probably lead to a more accurate off-target prediction. Finally, it is interesting to contextualize those findings in the light of the endogenous Cas nucleases activity, within the natural bacterial immune system, considering that the CRISPR system evolved in the presence of a highly negatively supercoiled genome. It is possible to speculate that this system adapted an increased tolerance for mismatches to allow for the degradation of invasive virus genomes, even with imperfect sequence matching, which is helpful to prevent their single-base mutations. These considerations raise the intriguing possibility that in the endogenous context the CRISPR sys-

tems may be able to accommodate a higher number of mismatches than previously considered. On the other hand, the single-molecule characterization performed in this study on Cas12a and AZ-Cas9 enlighten their mechanism and dynamics, opening new perspectives on their application as gene-editing tools. Particularly, although we observe a similar mechanism between Cas9 and Cas12a, in regards to the increased off-target binding activity induced by DNA stretching, the force-induced cleavage inhibition shown for Cas12a implies that those off-target bound complexes do not possibly cleave, which will have minimal implications for off-target activity *in vivo*. Together with other advantages, such as the ability to recognise a different T-rich PAM, and to produce staggered-ends cut, this crucial quality raises the potential for Cas12a-based methods application. At the same time, the observation on AZ-Cas9 rate cleavage, found here to be comparable to other high-fidelity Cas proteins, emphasises the requirement to expand the characterization of this new variant. Further experiments are needed to confirm AZ-Cas9 high specificity in cells and this may promote the addition of an important component to the CRISPR systems comprehensiveness. Although the development and the ease of SpCas9-based systems led to their rapid adoption in a great range of applications, the extensive effort to discover and engineer new Cas proteins is being extremely successful and it is actively contributing to the exceptional CRISPR evolution.

8 List of Reagents

Table 8.1: Cell line used in this study

Cell line name	**Basal medium**	**Supplement**	**Cell type**
A549	F12-NutMix	10% FCS 1% Glutamax	Lung cancer
HEK293-EXPI (10C10clone)	DMEM	10% FCS 1% Glutamax	Embryonic Kidney
HEK293	DMEM	10% FCS 1% Glutamax	Embryonic Kidney
HEK293T	DMEM	10% FCS 1% Glutamax	Embryonic Kidney Highly transfectable
JURKAT	RPMI	10% FCS 1% Glutamax 10mM Hepes	Immortalized T lymphocyte
MCF7	RPMI w/o Phenol	10% FCS 1% Glutamax	Breast cancer
RH-30	RPMI	10% FCS 1% Glutamax 10mM Hepes	Pediatric rhabdomyosarcoma
SH-SY5Y	RPMI	10% FCS 1% Glutamax 10mM Hepes	Neuroblastoma
U2OS Traffic Light-Reporter	McCoy's Modified 5A	10% FCS 1% Glutamax	Sarcoma

Table 8.2: Cell line used in this study for CRISPRa system activation

Cell line name	**Features**	**Source**
A549-lenti-dCas9-VPR	Constitutive expression of dCas9-VPR	Generated in this study by lentiviral infection of A549 WT and sorted by FACS analysis.
A549-lenti-dCas9-VPR+ CXCR4 CRISPRa	Constitutive expression of dCas9-VPR and sgRNA for CRISPRa activation of CXCR4 gene	Generated in this study by lentiviral infection of A549-lenti-dCas9-VPR and sorted by FACS analysis.
MCF7-inducible-dCas9	Inducible expression of dCas9-VPR by Doxycycline treatment	AZ cell bank
MCF7-inducible-dCas9+ CXCR4 CRISPRa guide	Inducible expression of dCas9-VPR by Doxycycline treatment and constitutive expression of sgRNA for CRISPRa activation of CXCR4 gene	Generated in this study by lentiviral infection of MCF7-inducible-dCas9 and sorted by FACS analysis.
HEK293-EXPI(10C10 clone)	Constitutive expression of dCas9-PCP-PH system	AZ cell bank

Table 8.3: Vectors used in this study

Vector	**Dimension**	**Note**
pKLV2-U6gRNA5(BbsI)-PGKpuro2ABFP-ScarLess	10110 bp	CRISPR gRNA expression vector with an improved scaffold and puro/BFP markers
pKLV2-EF1a-dCas9-VPR-mCherry-(pKLV2-EF1aCas9FZt2aBsdW)	15176 bp	dCas9-VPR expression vector with BsdR/mCherry markers.

Table 8.4: Oligonucleotide for cloning into pKLV2 vector series

GTGGAAAGGACGAAACACCG -gRNA sequence- GTTTAAGAGCTATGCTGGAA

Table 8.5: Target sequence for crRNA used in this study

Gene target	CRISPRn guide (5'> 3')	CRISPRa guide (5'> 3')
ASCL1	AGAGCCCTCGTCCGACGAGT CGTTGGGGACGTGCTCCCGA GACTTGTGACCGCCCCCTGA	CGGGAGAAAGGAACGGGAGG
SRP14	TTTCCAGAAGTGCCGGACGT CGTCGCCGCGATGGTGTTGT	TTTCTCGCATGTGACGACCC
CXCR4	CAACCACCCACAAGTCATTG TCTTCTGGTAACCCATGACC GCTTCTACCCCAATGACTTG GTTCCAGTTTCAGCACATCA	TCTGGGAGGTCCTGTCCGGCTC
MET	GCCCCCGCTGTGCTTGCACC CCTCGTGCTCCTGTTTACCT GACAAATGTGTGCGATCGG	TGGCAGGGCAGCGCGCGTGT TGGTCGCCTGGCGGTGCCTC CAGCCGGGCATCGGCGCGCG

Table 8.6: TaqMan gene probes used in this study

Gene target	Gene probe	Supplier
CXCR4	Assay ID: Hs00607978-s1 (Cat. n. 4331182)	ThermoFisher
MET	Assay ID: Hs01565584-m1 (Cat. n. 4453320)	ThermoFisher
SRP14	Assay ID: Hs03043834-g1 (Cat. n. 4448892)	ThermoFisher
ASCL1	Assay ID: Hs00269932-m1 (Cat. n. 4331182)	ThermoFisher
S18	Assay ID: Hs99999901-s1 (Cat. n. 4448489)	ThermoFisher

Table 8.7: Target sequence for Proxy-crRNA used on TLR or EMX1

Guide name	Sequence (5'>3')	PAM	Distance from the on target PAM to PAM (bp)	Guide orientation with respect to the on-target guide
Proxy TLR-P1	CACGGGCAGCTTGC	NGG	37	Upstream
Proxy TLR-P2	GACGAGGGTGGGCC	NGG	19	Upstream
Proxy TLR-P3	TGACGAGGGTGGGCC	NGG	18	Upstream
Proxy TLR-P3	TCCGCCATGCCCGA	NGG	63	Downstream
Proxy TLR-P4	CTGGACGTAGCCTT	NGG	59	Downstream
Proxy TLR-P5	TCTGATAAAAGCAGC	ACG	46	Downstream
Proxy TLR-P6	AGCACGACTTCTTC	AAG	34	Downstream
Proxy EX-P1	GCCCAGGTGAAGGTG	NGG	71	Upstream
Proxy EX-P2	GGCAGAAGCTGGAGG	NGG	28	Upstream

Table 8.8: Target sequence for TLR-crRNA used for KO

Guide name	**Sequence (5'>3')**	**n. mismacthes**	**Position of the mismatches from the PAM**
TL-G0	AGCACTGCACGCCGTAGGTC	0	-
TL-G1	AGCACTGCACGCCGTAGGTG	1	1
TL-G1j	AGCACTGCACGCCGTAGGTA	1	1
TL-G1k	AGCACTGCACGCCGTAGGTT	1	1
TL-G1b	AGCACTGCACGCCGTCGGTC	1	5
TL-G1n	AGCACTGCACGCCGTGGGTC	1	5
TL-G1o	AGCACTGCACGCCGTTGGTC	1	5
TL-G1c	AGCACTGCACGCCTTAGGTC	1	7
TL-G1d	AGCACTGCACGCCATAGGTC	1	7
TL-G1p	AGCACTGCACGCCCTAGGTC	1	7
TL-G1e	AGCACTGCACGCTGTAGGTC	1	8
TL-G1f	AGCACTGCACGCGGTAGGTC	1	8
TL-G1q	AGCACTGCACGCAGTAGGTC	1	8
TL-G1g	AGCACTGCACGTCGTAGGTC	1	9
TL-G1h	AGCACTGCACGGCGTAGGTC	1	9
TL-G1r	AGCACTGCACGACGTAGGTC	1	9
TL-G2	AGCACTGCACGCCGTAGTTG	2	1,2
TL-G2b	AGCACTGCACGCCGTCGGTG	2	1,5
TL-G2c	AGCACTGCACGCTGTCGGTC	2	5,8
TL-G2d	AGCACTGCACGCGGTCGGTC	2	5,8
TL-G2e	AGCACTGCACGTGGTAGGTC	2	8,9
TL-G2f	AGCACTGCACGGAGTAGGTC	2	8,9
TL-G2g	AGCACTGCACGTCTTAGGTC	2	7,9
TL-G2h	AGCACTGCACGGCATAGGTC	2	7,9
TL-G2i	AGCACTGCACGCTATAGGTC	2	7,8
TL-G2l	AGCACTGCACGCGTTAGGTC	2	7,8
TL-G3	AGCACTGCACGCCGTCGTTG	3	1,3,5
TL-G4	AGCACTGCACGCCTTCGTTG	4	1,3,5,7

Table 8.9: Target sequence for EMX1-crRNA used for KO

Guide name	**Sequence (5'>3')**	**n. mismacthes**	**Position of the mismatches from the PAM**
EX-KO1	CAAACGGCAGAAGCTGGAGG	0	-
EX-KO2	AGGGCTCCCATCACATCAAC	0	-
EX-KO3	GAGTCCGAGCAGAAGAAGAA	0	-
EX-3-1	GAGTCCGAGCAGAAGAAGCA	1	2
EX-3-1b	GAGTCCGAGCAGAAGTAGAA	1	5
EX-3-1c	GAGTCCGAGCAGATGAAGAA	1	7
EX-3-1d	GAGTCCGAGCAGTAGAAGAA	1	8
EX-3-1e	GAGTCCGAGCAGAAGAAGTA	1	2
EX-3-1f	GAGTCCGAGCAGAAGAAGGA	1	2
EX-3-1g	GAGTCCGAGCAGAAGGAGAA	1	5
EX-3-1h	GAGTCCGAGCAGAAGCAGAA	1	5
EX-3-1i	GAGTCCGAGCAGACGAAGAA	1	7
EX-3-1j	GAGTCCGAGCAGAGGAAGAA	1	7
EX-3-1k	GAGTCCGAGCAGCAGAAGAA	1	8
EX-3-1l	GAGTCCGAGCAGGAGAAGAA	1	8
EX-3-1m	GAGTCCGAGTAGAAGAAGAA	1	9
EX-3-1n	GAGTCCGAGGAGAAGAAGAA	1	9
EX-3-1o	GAGTCCGAGAAGAAGAAGAA	1	9

Table 8.10: Oligo for crRNA used in the single-molecule experiments

Oligo name	**Sequence (5'>3')**
Cas12a on-target	AAUUUCUACUCU/AmC6T/GUAGAUATTCAGATGAT ATGACTATCAAGG
AZ-Cas9 on-target (λ2)	/5AmC12/CAGTGATAAGTGGAATGCCATGGUUUCAGU UAUUCGUGA

Table 8.11: DNA oligo used for AZ-Cas9 *in vitro* cleavage assay

Oligo name	**Sequence (5'>3')**	**note**
DNA λ2 target	CATGCGTACACGCGTTTGTACACCACATGGCATTC CACTTATCACTTGGACACCACACCTCCTGGTCGCG /iAmMC6T/CCTATAGTTTGTTTCCTGAGT	used for AZ-Cas9 *in vitro* cleavage
DNA biot-λ2 target	/5Biosg/TTTTTTCCCTCGTGACCACCCACATGGC ATTCCACTTATCACTGAG/iAmMC6T/CCGCTACC	used for AZ-Cas9 FRET assay

Table 8.12: List of primers for KO region amplification for each sgRNA

Guide name	**Primers Sequence (5'>3')**
ASCL1-G1	*FWD*:AACTTCAGCGGCTTTGGCTA
	REV:TCCTTTTCTCCCCCTCCCAA
ASCL1-G2	*FWD*:TTTTGCTCCCACTCTAAGAAGTC
ASCL1-G3	*REV*:GAGTTCAAGTCGTTGGAGTAGTT
SPR14-G1	*FWD*:ATTGTTACCAGAAGCAACCTAGG
	REV:AAGACAATGGAAGTCTAGTAGCG
SPR14-G2	*FWD*:TGCTTACACTTCTTCAAGGTGAT
	REV:AAATCGTACGCTGCTATCAAAAC
CXCR4-G1	*FWD*:AGGTTTTAAATTTGGCTCCAAGG
CXCR4-G2	*REV*:TAGCCCACTACTTCAGAATTTCC
CXCR4-G3	*FWD*:AGGTTTTAAATTTGGCTCCAAGG
CXCR4-G4	*REV*:TAGCCCACTACTTCAGAATTTCC
MET-G1	*FWD*:AGTCATGTCCAACCGCACAA
	REV:AAAGGACTTTGGCTCCCAGG
MET-G2	*FWD*:AGTCATGTCCAACCGCACAA
	REV:AAAGGACTTTGGCTCCCAGG
MET-G3	*FWD*:TCAGACAGGGCACAATGTCC
	REV:CATGGCAGGGATCACCAGTT