

## Seek and Destroy

### Single-molecule perspective on the target search and recognition by the Cas9 endonuclease

Globyte, Viktorija

#### DOI

[10.4233/uuid:2c9a92d6-0c3c-4e1c-ab56-8038c0677156](https://doi.org/10.4233/uuid:2c9a92d6-0c3c-4e1c-ab56-8038c0677156)

#### Publication date

2019

#### Document Version

Final published version

#### Citation (APA)

Globyte, V. (2019). *Seek and Destroy: Single-molecule perspective on the target search and recognition by the Cas9 endonuclease*. [Dissertation (TU Delft), Delft University of Technology].  
<https://doi.org/10.4233/uuid:2c9a92d6-0c3c-4e1c-ab56-8038c0677156>

#### Important note

To cite this publication, please use the final published version (if applicable).  
Please check the document version above.

#### Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

#### Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.  
We will remove access to the work immediately and investigate your claim.

Seek and Destroy:  
Single-molecule perspective on the target search  
and recognition by the Cas9 endonuclease



Seek and Destroy:  
Single-molecule perspective on the target search  
and recognition by the Cas9 endonuclease

Dissertation

for the purpose of obtaining the degree of doctor  
at Delft University of Technology  
by the authority of the Rector Magnificus Prof.dr.ir. T.H.J.J. van der Hagen  
chair of the Board for Doctorates  
to be defended publicly on  
Thursday 26 September 2019, 10:00 AM

by

VIKTORIJA GLOBYTĖ  
Master of Science in Physics  
University of St Andrews, Scotland, United Kingdom  
Born in Šiauliai, Lithuania



This dissertation has been approved by the:

Promotor: dr. C. Joo

Copromotor: dr. S. M. Depken

Composition of the doctoral committee:

Rector Magnificus chairperson

Dr. C. Joo Delft University of Technology, promotor

Dr. C. S. M. Depken Delft University of Technology, copromotor

Independent members:

Prof. dr. N. H. Dekker Delft University of Technology

Prof. dr. J. van Noort Leiden University

Dr. C. Penedo-Esteiro University of St Andrews

Dr. S. J. J. Brouns Delft University of Technology

Dr. J. Lebbink Erasmus University Medical Center

Reserve:

Prof. dr. G. Koenderijk Delft University of Technology



Keywords: Single-molecule FRET, CRISPR, Cas9, target search, RNA, DNA.

Printed by: Gildeprint

Front & Back: Viktorija Globytė, Rusnė Jaugelaitė

Copyright © 2019 by V. Globytė

Casimir PhD Series 2019-12

ISBN 978-90-8593-417-2

An electronic version of this dissertation is available at <http://repository.tudelft.nl/>

*To my mother and my daughter*



# Contents

<b>Preface .....</b>	<b>xix</b>
<b>General Introduction .....</b>	<b>1</b>
1.1. The macromolecules of life .....	2
1.1.1. Deoxyribonucleic acid (DNA).....	2
1.1.2. Ribonucleic acid (RNA).....	3
1.1.3. Protein.....	4
1.1.4. Small non-coding RNA .....	4
1.2. Single-molecule Techniques.....	6
1.2.1. Magnetic Tweezers.....	6
1.2.1. Total internal reflection fluorescence microscopy (TIRFM)....	6
1.3. Thesis outline.....	8
References.....	10
<b>Single-molecule view on RNA-guided target search mechanisms .....</b>	<b>13</b>
2.1. Introduction .....	14
2.2. Argonaute .....	14
2.2.1. Seed recognition.....	14
2.2.2. Lateral diffusion .....	17
2.2.3. Conformational change .....	18
2.2.4. Cooperativity.....	18
2.3. CRISPR adaptive immunity .....	18
2.3.1. Cascade protein complex.....	19
2.3.2. Cas9 endonuclease.....	22
2.4. Integrated view on target search and recognition.....	27
2.4.1. Modes of target search.....	27
2.4.2. Mechanism of kinetic proofreading.....	29
References.....	32

<b>Cas9 searches for a prospacer adjacent motif using lateral diffusion....</b>	<b>37</b>
3.1. Introduction .....	38
3.2. Results.....	38
3.2.1. Single-molecule observation of Cas9 PAM search.....	38
3.2.2. 1-Dimensional Diffusion Used for PAM and Target Search.....	45
3.2.3. Mechanism of lateral diffusion .....	47
3.2.4. PAM multiplicity delays on-target binding.....	48
3.3. Discussion .....	50
3.4. Materials and methods.....	53
3.4.1. Recombinant SpCas9 purification.....	53
3.4.2. Biotinylation of the recombinant SpCas9 .....	53
3.4.3. Preparation of the single-guide RNA .....	54
3.4.4. In-vitro DNA cleavage assay with wild-type and biotinylated SpCas9 .....	54
3.4.5. Labeling of nucleic acids .....	54
3.4.6. Single-molecule two-color FRET .....	55
3.4.7. Data acquisition and analysis.....	55
3.5. Supplementary information .....	56
3.5.1. Supplementary figures .....	56
3.5.2. Supplementary tables.....	60
References.....	65
 <b>Single-molecule FRET methods to study the Cas9 endonuclease .....</b>	 <b>69</b>
4.1. Introduction .....	70
4.1.1. CRISPR bacterial adaptive immune system .....	70
4.1.2. Cas9 endonuclease.....	70
4.2. TIRF-based single-molecule FRET .....	71
4.2.1. single-molecule FRET .....	71
4.2.2. Total Internal Reflection Fluorescence Microscopy.....	72
4.2.3. Microscope slide assembly.....	72
4.3. Preparation for single-molecule experiments .....	73
4.3.1. Fluorescent labeling of nucleic acids.....	73
4.3.2. Cas9 biotinylation .....	74
4.3.3. Buffers and reagents .....	75
4.4. DNA-immobilization based assays .....	75
4.4.1. Experimental procedure.....	76
4.4.2. Assessing the effects of target truncations .....	79

4.4.3. Monitoring Cas9-induced DNA duplex dynamics .....	80
4.5. Cas9 immobilization based assays .....	81
4.5.1. Experimental procedure.....	81
4.5.2. Cas9-PAM interactions.....	82
4.5.3. Target search.....	85
4.6. Data analysis .....	86
4.6.1. Video processing.....	86
4.6.2. FRET efficiency and dwelltime analysis.....	86
4.7. Concluding remarks.....	88
References.....	89
<b>Small RNA molecules inhibit the activity of SpCas9 in vitro.....</b>	<b>93</b>
5.1. Introduction .....	94
5.2. Results.....	94
5.2.1. Small RNA inhibit DNA cleavage by Cas9 .....	94
5.2.2. Single-molecule fluorescence reveals the mechanism of inhibition .....	96
5.3. Discussion .....	99
5.4. Materials and methods.....	100
5.4.1. sgRNA production .....	100
5.4.2. Cleavage assays .....	101
5.4.3. Electrophoretic mobility shift assay (EMSA).....	101
5.4.4. Recombinant SpCas9 purification.....	101
5.4.5. Biotinylation of the recombinant SpCas9 .....	102
5.4.6. Single-molecule two-color fluorescence .....	102
5.4.7. Data acquisition and analysis.....	103
5.5. Supplementary Information.....	104
5.5.1. Supplementary Figures.....	104
5.5.2. Supplementary tables.....	106
References.....	107
<b>Overview of different Cas9 variants and concluding remarks.....</b>	<b>109</b>
6.1. Considerations of SpCas9 structure in relation to single-molecule observations .....	110
6.2. PAM specificity and recognition of Cas9 orthologs.....	113
6.2.1. Staphylococcus aureus Cas9 .....	113
6.2.3. Francisella novicida Cas9.....	114
6.2.2. Campilobacter jejuni Cas9 .....	116

6.3. Engineered SpCas9 variants .....	116
6.4. Concluding remarks .....	118
References.....	120
<b>Summary.....</b>	<b>123</b>
<b>Samenvatting.....</b>	<b>127</b>
<b>Acknowledgements .....</b>	<b>133</b>
<b>Curriculum Vitae .....</b>	<b>143</b>
<b>List of Publications .....</b>	<b>145</b>







# Preface

Through centuries the greatest minds on Earth have tried to answer the question “What is life?”. The lack of consensus on what constitutes life illustrates how complex it really is. A physicist might describe all living things as thermodynamic systems that are able to decrease their internal entropy. This definition possesses the simplistic beauty so common in physics, yet is much too incomplete. In biology, a science that most readily deals with life in all its forms, the definition is much broader and rather descriptive. According to biologists, all living things are able to organize their internal environments (homeostasis), convert chemicals and energy into cellular components at a rate higher than the decomposition of these components (metabolism), adapt to their environments (evolution), produce new individual organisms (reproduction), respond to stimuli and grow. It is quite fascinating that this definition describes the essence of all life forms, from simple single-celled bacteria to us humans. No matter how complex we might be, the processes that go on in our bodies on the smallest scale are the same as in every other organism – a fact that is both fascinating and humbling.

We do not understand life yet and I am not sure we ever will. In trying to do so, however, we as humans have shown incredible cooperation and creativity. The use of chemical and physical techniques has advanced our knowledge of living systems so rapidly that a century ago no one could have imagined that one day we could peer into what lies beyond the physical limitations of what we can observe. The advent of single-molecule techniques has done just that. Is it not amazing that we are able to look at and manipulate tiny molecules far beyond the diffraction limit, investigate their properties to build our understanding of the very mechanisms that drive the processes required for life to exist? We might not fully understand life, but we are getting closer with every protein discovered and analysed, every molecular mechanism described.

While the 20th century is often regarded as the century of physics, some regard 21st century to be the century of biology. Indeed, the rapid development of technology towards the end of the 20th century has enabled us to answer the complex questions in our quest to understand life and equipped us with tools to dig deeper. The invention of magnetic resonance imaging has helped us to better understand how our brain works, super resolution

microscopy allowed us to see what the cells look like in great detail, the advent of DNA sequencing has led to the rise of the field of genetics, helping us understand how genes govern life. The latter has been recently revolutionized by the discovery of Cas9, a single protein that can edit DNA and is so much simpler than any other gene editing tool previously developed. We can now manipulate the very instructions for all living things that are written in the DNA. Without fully understanding life and the said instructions, this brings us into a dangerous new world. The least we can do to make sure gene editing does not go horribly wrong is to understand the underlying mechanisms of this wonderful tool, Cas9. It is my hope that this thesis will add to our understanding of these mechanisms, even if by an infinitesimal amount.

Viktorija Globytė





# 1

## GENERAL INTRODUCTION

## 1.1. THE MACROMOLECULES OF LIFE

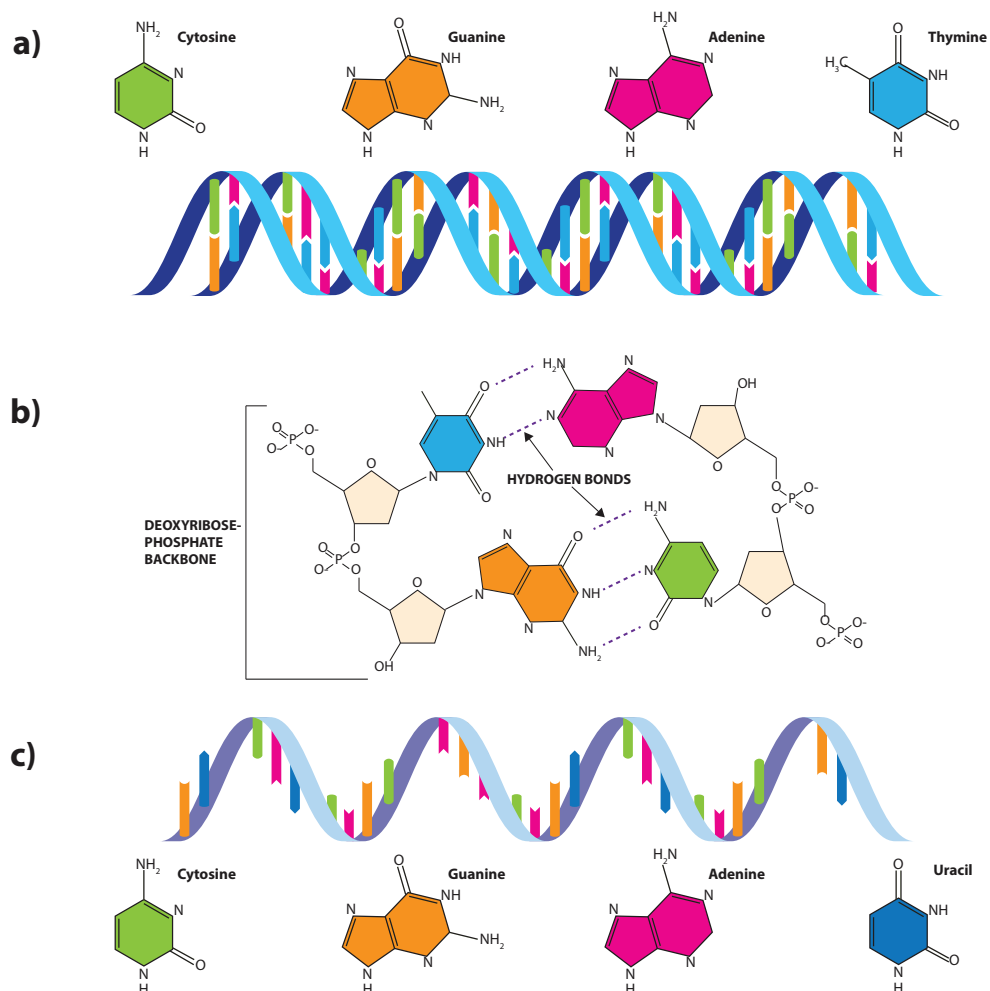
In order to grow a tree, we do not plant a microscopic tree in the ground, but rather a seed. Therefore, the seed must contain the instructions necessary to make the tree. This observation, although simple and ancient is nothing short of genius. The simple rules of heredity were established as early as the 19th century. However, up until the 1940s it was thought that those rules were written into proteins. In 1944 Oswald Avery and his colleagues published their work showing that the “transforming principle” that caused non-pathogenic bacteria to transform into a different, pathogenic strain was the deoxyribonucleic acid or DNA [1]. In 1952, the Hershey-Chase experiments furthered this notion and in 1953 the revolutionary structure of the double helix was published by Francis Crick and James Watson [2, 3]. In 1957, Francis Crick laid out the central dogma, essentially stating that information goes from DNA to RNA to protein, and the world of molecular biology was changed forever (Figure 1.1.).



Figure 1.1. The central dogma of molecular biology.

### 1.1.1. DEOXYRIBONUCLEIC ACID (DNA)

DNA is a biological molecule composed of two helical chains that coil around each other forming a double helix (Figure 1.2. a). It consists of four nucleotides that in turn consist of a sugar deoxyribose, a phosphate group and one of four nucleobases: adenine (A), thymine (T), guanine (G) and cytosine (C). The bases are grouped into those containing two-carbon nitrogen rings (purines) and one-carbon nitrogen ring (pyrimidines). Adenine and guanine are purines while thymine and cytosine are pyrimidines. The nucleotides join together forming covalent bonds between the sugar and phosphate groups forming what is known as the sugar-phosphate backbone (Figure 1.2. b) The bases themselves pair together (A pairs with T and G pairs with C) through hydrogen bonding and thus form a double helix. It is those four bases that encode the instructions for all life on Earth. DNA is a stable molecule and therefore the information stored in it is passed from generation to generation.



**Figure 1.2. Structure of the nucleic acids.** a) Schematic illustration of the DNA double helix and the four nucleobases. b) schematic illustration of the chemical structure of the DNA double helix. c) schematic illustration of the single-stranded RNA and the four nucleobases

### 1.1.2. RIBONUCLEIC ACID (RNA)

RNA, or ribonucleic acid, is another form of nucleic acid (Figure 1.2. c). Unlike DNA, its nucleotides have a ribose group instead of the deoxyribose and instead of a thymine base it possesses uracil. RNA is transcribed from DNA by an enzyme RNA polymerase and is a messenger (mRNA) molecule that carries the instructions, encoded in the DNA, to the machinery that makes the proteins. Non-coding RNA molecules take part in gene silencing



through a process known as RNA interference and also are part of a prokaryotic immune response [4, 5]. RNA is a short-lived molecule, however, it performs more tasks than the information storage performed by DNA.

### 1.1.3. PROTEIN

The final molecule in the central dogma is the protein. Proteins are large biomolecules which consist of amino acids. There are total 20 naturally occurring amino acids and they are encoded in the DNA. The amino acids are joined into a long chain and are bonded together by peptide bonds. Proteins take many different shapes and sizes and perform a variety of functions such as DNA replication[6], transporting molecules within cells [7], taking part in the cells' immune response [5, 8] and many others. Proteins are the workhorses of the cell and are by far the most versatile biomolecules. This thesis deals with the way a protein, assembled with RNA molecules, finds and interacts with DNA molecules, encompassing all three elements of the central dogma.

### 1.1.4. SMALL NON-CODING RNA

A revolutionary discovery of small noncoding RNAs opened up a new perspective of RNA regulation of gene expression [4]. For example, microRNA (miRNA) molecules play a role in translation inhibition and subsequent

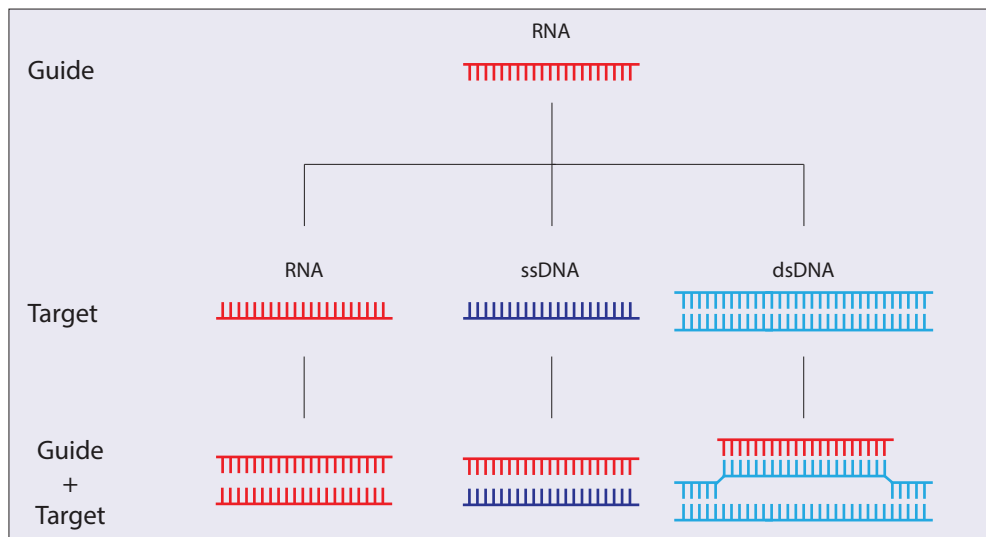


Figure 1.3. Schematic of different RNA guides and the nucleic acids they can target.

degradation of messenger RNA molecules in eukaryotes. Small interfering RNA (siRNA) molecules have a similar length to miRNA, but their full complementarity with their targets leads to direct cleavage of the messenger RNA (mRNA). Both miRNA and siRNA associate with proteins belonging to the Argonaute family. These small RNA molecules guide eukaryotic Argonaute to the target site, where recognition occurs via Watson-Crick base pairing between guide and target (Figure 1.3.) (Table 1). In prokaryotes, small regulatory RNA and DNA molecules not only can regulate gene expression but also can act as a defense mechanism against invading phage genomes and plasmids (Table 1). For example, some prokaryotic Argonaute proteins associate with DNA guides to find and destroy complementary DNA target sequences [8].

Another famous example is the CRISPR (clustered regularly interspaced short palindromic repeats) adaptive prokaryotic immune system where CRISPR-associated (Cas) proteins assemble with guide RNAs to find and destroy invaders by cleaving DNA or RNA target sites complementary to the guide. CRISPR immunity consists of multiple stages. Ensuing an infection by mobile genetic elements, short fragments of the invader's DNA are integrated into the CRISPR locus in the host genome as short spacers[5, 9]. This first stage of CRISPR immunity is known as the adaptation stage. During this stage, a genetic memory is created that is later used to destroy the invader upon reinfection [5]. In the second stage of immunity, transcription of the CRISPR locus and further maturation of the transcript produces short CRISPR RNAs

<div>TARGET PROTEIN</div>	RNA		DNA		DNA (CO-TRANSCRIPTIONAL)
<b>eAgo/ PIWI</b>	Gene silencing by miRNA [11, 12] and piRNA [13, 14]  RNA interference by siRNA [4]				Heterochromatin formation by small RNA[15]  Genome rearrangement by scnRNA[16]
<b>pAGO</b>			DNA interference by diRNA[17]		
<b>CRISPR</b>	Cas13a	Gene silencing by crRNA of Type VI [18, 19]	Cascade/ Cas9/Cpf1	DNA interference by crRNA of Type I [5, 10, 20], II [21, 22], and V [23]	
<b>Other systems</b>	Hfq	Gene regulation by sRNA[24, 25]			

Table 1.1 Small regulatory RNA systems

1 (crRNAs) [10]. These crRNAs associate with Cas proteins and destroy the returning invader upon recognition of the target sequence in the third stage, known as the interference stage.

## 1.2. SINGLE-MOLECULE TECHNIQUES

An invaluable tool to study DNA, RNA and protein is the single-molecule techniques. These techniques allow one to investigate the properties of individual molecules, as opposed to ensemble measurements where only an average of those properties can be measured. The first single-molecule measurements have been performed as early as the 1970s on ion channels, however, it was not until the 1990s until the field of single-molecule spectroscopy really took off [26]. There are two main types of single-molecule techniques: force-based techniques and fluorescence-based techniques. Although the work described in this thesis has been performed using the latter, the former has been just as important in establishing the current status quo of the knowledge in the field of CRISPR proteins and their molecular mechanisms.

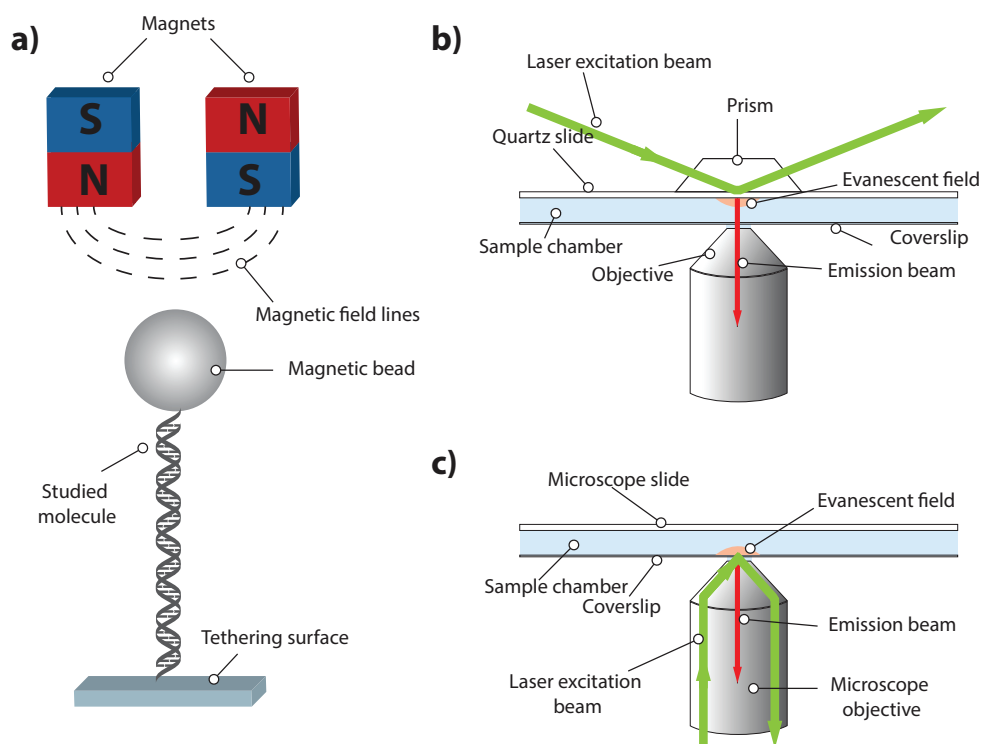
### 1.2.1. MAGNETIC TWEEZERS

Magnetic tweezers, a form of force spectroscopy, make use of magnets and a magnetic bead that is attached to a molecule of interest (Figure 1.4. a) [27]. By trapping the bead in the magnetic field, researchers can manipulate the molecule in question, and the force and torque exerted on the molecule can be obtained by measuring the height of the magnetic bead. Magnetic tweezers assays used in studies described in this thesis make use of supercoiling long DNA molecules. DNA molecules are bound to magnetic beads and immobilized on the surface of the flow cells. When the magnets are turned, torsional stress is applied to the DNA molecule. At low forces, DNA supercoils forming plectonemes, which decreases the extension of the DNA molecule in a symmetric fashion for both positive and negative supercoiling. In such a setting, the position of the bead is very sensitive to even the slightest changes of the length of the DNA molecule that occur by unzipping a portion of the DNA.

### 1.2.1. TOTAL INTERNAL REFLECTION FLUORESCENCE MICROSCOPY (TIR-FM)

In TIRFM, only a shallow depth ( $\sim 100$  nm) below the slide surface is illu-

minated [28]. This is achieved by directing the excitation beam at an angle where it undergoes total internal reflection at the interface between the glass slide and solvent in the flow cell (Figure 1.4 b, c). This is perfect for low-light imaging, including single-molecule detection [29]. In studies discussed in this thesis, TIRFM was used in fluorescence, Förster resonance energy transfer (FRET) (as in the experiments I performed myself), and DNA curtains as-says [30-32]. FRET techniques offer nanometer resolution through resonance energy transfer and allow one to observe processes otherwise impossible because of the physical diffraction limit. DNA curtains involve anchoring a long fluorescently labeled DNA molecule and then stretching it by a laminar flow. DNA curtains can be used to observe long-distance movement that cannot be tracked using techniques such as single-molecule FRET (smFRET). This technique is limited by the physical diffraction limit, and processes that occur on length scales smaller than  $\sim 250$  nm cannot be observed



**Figure 1.4. Single-molecule techniques.** a) Schematic illustration of a conventional magnetic tweezer setup. b) Schematic illustration of a prism-type TIRF setup b) Schematic illustration of an objective-type TIRF setup

## 1.3. THESIS OUTLINE

This thesis describes work done using single-molecule fluorescence and FRET techniques, which focused on the target search and recognition on *Streptococcus pyogenes* Cas9.

**Chapter 2** (Pages 13-35): "Single-molecule view on RNA-guided target search mechanisms"

In chapter 2 we present a review, which focuses on single-molecule advances in studying the target search and recognition mechanisms of Argonaute and CRISPR systems. Different stages of target search and recognition are described: initial weak interactions, such as PAM search and seed recognition, 1-dimensional and 3-dimensional diffusion, protein conformational changes and kinetic proofreading.

**Chapter 3** (Pages 37-67): "Cas9 searches for a protospacer adjacent motif using lateral diffusion"

In chapter 3 we present a single-molecule FRET study on Cas9 PAM and target search. We show that Cas9 exhibits two distinct binding modes when interacting with a PAM sequence, characterized by two distinct dwelltimes. The second, longer dwelltime was found to increase when the number of PAM sequences increased, suggesting a synergistic effect. Furthermore, we directly demonstrate that Cas9 can laterally diffuse between the PAM sequences without dissociation and that this diffusion can lead to on-target binding. We also demonstrate that Cas9 is able to laterally diffuse between adjacent target sites, showing that the protein uses a mixture of three-dimensional and short-range one-dimensional diffusion during its target search. Finally, we show that despite the fact that lateral diffusion between PAM sites can lead to finding the target, it delays on-target binding, acting as a decoy binding site.

**Chapter 4** (Pages 69-91): "Single-molecule FRET methods to study the Cas9 endonuclease."

In chapter 4 we describe single-molecule FRET techniques to investigate the Cas9 protein. The techniques described are based on the immobilization of either the target DNA or the protein. We discuss different types of measurements, namely long videos and snapshots, and demonstrate the types of data that can be acquired using each type of measurement. Finally, we

discuss aspects of image processing and data analysis.

**Chapter 5** (Pages 93-107) "Small RNA molecules inhibit the activity of SpCas9 in vitro"

In chapter 5 we present a biochemical and single-molecule study which demonstrates that small RNA molecules can efficiently inhibit the DNA cleavage by Cas9 in vitro. We show that Cas9 stably assembles with a sgRNA molecule that is hybridized with an inhibitor molecule. In addition, we show that this inhibitor molecule can be removed from the ribonucleoprotein complex if it is situated at the end region of the guide. Finally, our findings demonstrate that, despite the inhibitor dissociating from the protein, Cas9 remains inactive and cannot interact with target DNA molecules.

**Chapter 6** (Pages 109-121): "Overview of different Cas9 variants and concluding remarks"

In this final chapter an overview of different Cas9 orthologs and their PAM recognition mechanisms together with considerations of what individual mechanisms could mean for target search are discussed. In addition, an overview of SpCas9 mutants with altered PAM specificities is provided.

## REFERENCES

1. Avery, O.T., C.M. Macleod, and M. McCarty, *Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type Iii*. J Exp Med, 1944. **79**(2): p. 137-58.
2. Hershey, A.D. and M. Chase, *Independent functions of viral protein and nucleic acid in growth of bacteriophage*. J Gen Physiol, 1952. **36**(1): p. 39-56.
3. Watson, J.D. and F.H. Crick, *Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid*. Nature, 1953. **171**(4356): p. 737-8.
4. Fire, A., et al., *Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans*. Nature, 1998. **391**(6669): p. 806-11.
5. Barrangou, R., et al., *CRISPR provides acquired resistance against viruses in prokaryotes*. Science, 2007. **315**(5819): p. 1709-12.
6. Bell, S.P. and A. Dutta, *DNA replication in eukaryotic cells*. Annu Rev Biochem, 2002. **71**: p. 333-74.
7. Vale, R.D., *The molecular motor toolbox for intracellular transport*. Cell, 2003. **112**(4): p. 467-80.
8. Swarts, D.C., et al., *DNA-guided DNA interference by a prokaryotic Argonaute*. Nature, 2014. **507**(7491): p. 258-261.
9. Amitai, G. and R. Sorek, *CRISPR-Cas adaptation: insights into the mechanism of action*. Nat Rev Microbiol, 2016. **14**(2): p. 67-76.
10. Brouns, S.J., et al., *Small CRISPR RNAs guide antiviral defense in prokaryotes*. Science, 2008. **321**(5891): p. 960-4.
11. Reinhart, B.J., et al., *The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans*. Nature, 2000. **403**(6772): p. 901-6.
12. Lee, R.C., R.L. Feinbaum, and V. Ambros, *The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14*. Cell, 1993. **75**(5): p. 843-54.
13. Aravin, A., et al., *A novel class of small RNAs bind to MILI protein in mouse testes*. Nature, 2006. **442**(7099): p. 203-7.
14. Girard, A., et al., *A germline-specific class of small RNAs binds mammalian Piwi proteins*. Nature, 2006. **442**(7099): p. 199-202.
15. Verdel, A., et al., *RNAi-mediated targeting of heterochromatin by the RITS complex*. Science, 2004. **303**(5658): p. 672-6.
16. Mochizuki, K., et al., *Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in tetrahymena*. Cell, 2002. **110**(6): p. 689-99.
17. Olovnikov, I., et al., *Bacterial argonaute samples the transcriptome to identify foreign DNA*. Mol Cell, 2013. **51**(5): p. 594-605.
18. Abudayyeh, O.O., et al., *C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector*. Science, 2016. **353**(6299): p. aaf5573.
19. East-Seletsky, A., et al., *Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection*. Nature, 2016. **538**(7624): p. 270-273.
20. Marraffini, L.A. and E.J. Sontheimer, *CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA*. Science, 2008. **322**(5909): p. 1843-5.

21. Sapranaukas, R., et al., *The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli*. Nucleic Acids Res, 2011. **39**(21): p. 9275-82.
22. Jinek, M., et al., *A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity*. Science, 2012. **337**(6096): p. 816-21.
23. Zetsche, B., et al., *Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system*. Cell, 2015. **163**(3): p. 759-71.
24. Storz, G., J. Vogel, and K.M. Wassarman, *Regulation by small RNAs in bacteria: expanding frontiers*. Mol Cell, 2011. **43**(6): p. 880-91.
25. Updegrove, T.B., A. Zhang, and G. Storz, *Hfq: the flexible RNA matchmaker*. Curr Opin Microbiol, 2016. **30**: p. 133-8.
26. Neher, E., B. Sakmann, and J.H. Steinbach, *The extracellular patch clamp: a method for resolving currents through individual open channels in biological membranes*. Pflugers Arch, 1978. **375**(2): p. 219-28.
27. Gosse, C. and V. Croquette, *Magnetic tweezers: micromanipulation and force measurement at the molecular level*. Biophys J, 2002. **82**(6): p. 3314-29.
28. Ambrose, E.J., *A surface contact microscope for the study of cell movements*. Nature, 1956. **178**(4543): p. 1194.
29. Ambrose, W.P., P.M. Goodwin, and J.P. Nolan, *Single-molecule detection with total internal reflection excitation: comparing signal-to-background and total signals in different geometries*. Cytometry, 1999. **36**(3): p. 224-31.
30. Greene, E.C., et al., *DNA curtains for high-throughput single-molecule optical imaging*. Methods Enzymol, 2010. **472**: p. 293-315.
31. Ha, T., *Single-molecule fluorescence methods for the study of nucleic acids*. Curr Opin Struct Biol, 2001. **11**(3): p. 287-92.
32. Ha, T., *Single-molecule fluorescence resonance energy transfer*. Methods, 2001. **25**(1): p. 78-86.





# 2

## Single-molecule view on RNA-guided target search mechanisms

Most everyday processes in life involve a necessity for an entity to locate its target. On a cellular level, many proteins have to find their target to perform their function. From gene-expression regulation to DNA repair to host defence, numerous nucleic acid-interacting proteins use distinct target search mechanisms. Several proteins achieve that with the help of short RNA strands known as guides. This focus on single-molecule advances studying the target search and recognition mechanism of Argonaute and CRISPR (clustered regularly interspaced short palindromic repeats) systems. Different steps involved in search and recognition will be discussed, from the initial complex prearrangement into the target-search competent state to the final proofreading steps. This chapter focuses on target search mechanisms that range from weak interactions, to one- and three-dimensional diffusion, to conformational proofreading.

---

This article has been published as Globyte, V., S.H. Kim, and C. Joo, Single-Molecule View of Small RNA-Guided Target Search and Recognition. *Annu Rev Biophys*, 2018. **47**: p. 569-593.

## 2.1. INTRODUCTION

2 Target search is intrinsically a complex process that involves weak interactions and protein conformational changes. Proteins searching for their target have to diffuse through cytosol in three-dimensional (3D) fashion before encountering a DNA/RNA molecule that they weakly associate with, checking for a target site. Such weak interaction can lead to either quick dissociation or lateral diffusion by sliding or hopping, the latter of which in theory would speed up the target search process. However, there exists a limit beyond which lateral diffusion along the nucleic acid strand would slow down the search process and the protein dissociates. Dissociation again can lead to 3D diffusion or, in some systems, jumping, where the protein can move to another target site that is physically in close proximity but far in sequence, owing to the supercoiling or coiled conformation of the nucleic acid. In addition, fast target search, involving mainly weak interactions, and specific target recognition, being mostly stable interactions, require a conformational change in the protein [13-16]. This chapter aims to give a comprehensive overview of how these different search modes and mechanisms are combined in the target search and recognition of Argonaute and CRISPR/Cas proteins at the single-molecule level.

## 2.2. ARGONAUTE

Argonaute proteins (see Box 2.1 "Argonaute structure") are highly conserved in all forms of life. Eukaryotic Argonaute proteins play a central role in gene expression through processes referred to as RNA interference, whereas prokaryotic Argonautes participate in host defense via DNA interference [18, 19]. In animals, Argonaute proteins loaded with miRNA as guides bind to the 3' UTR of mRNA and prevent the production of proteins via several pathways that usually involve destabilization of mRNAs, and they require partial complementarity between guiding miRNA and target mRNA [21, 22]. Some Argonaute proteins [e.g., human Argonaute 2 (hAGO2)] are able to degrade mRNA without recruiting additional factors [25]. This endonucleolytic activity is mediated by siRNAs and requires full complementarity between the target and the guide.

### 2.2.1. SEED RECOGNITION

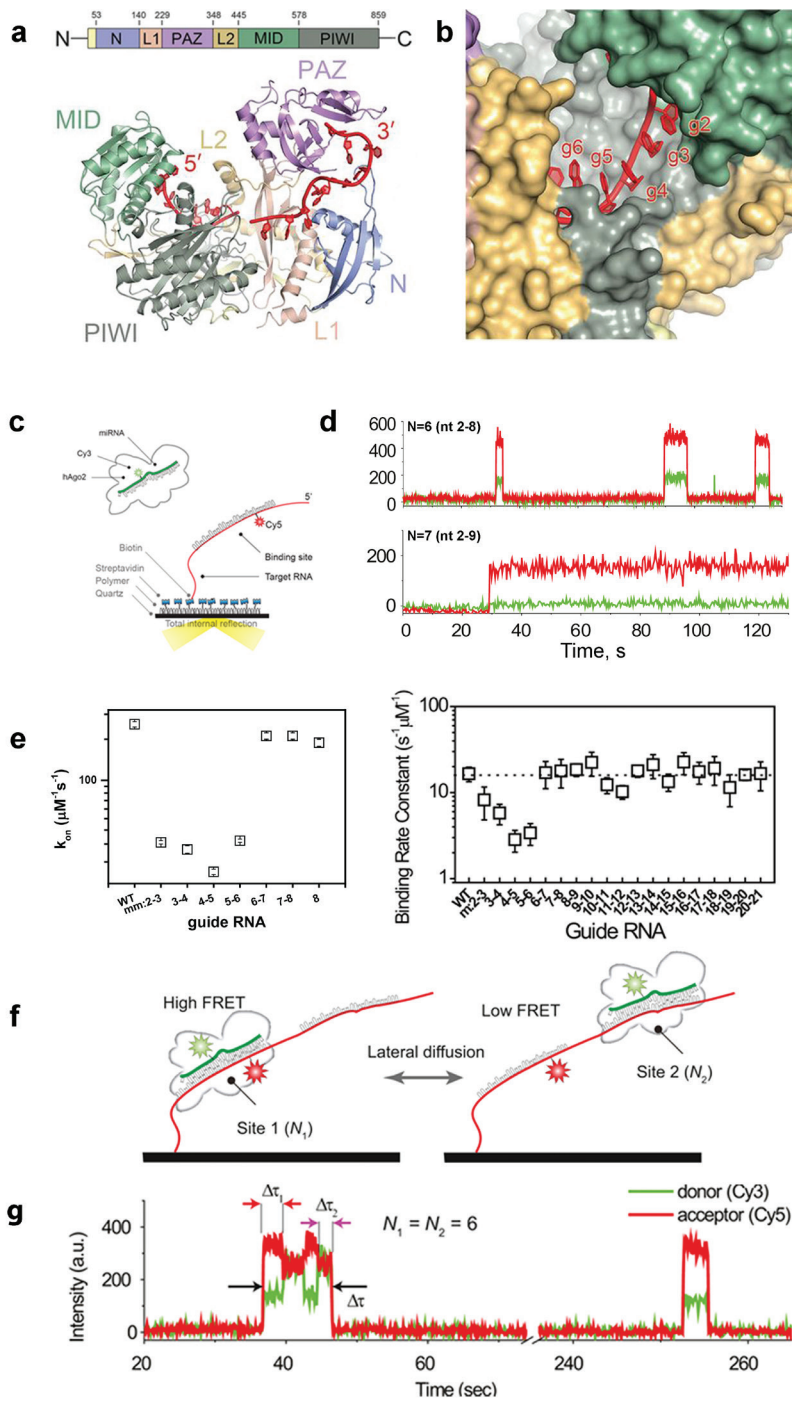
Biochemical and bioinformatics studies showed that human Argonaute proteins divide their guide molecules into five distinct domains: 5' anchor (first nucleotide), seed region (nucleotides 2-8), central region (nucleotides

**Box 2.1. ARGONAUTE STRUCTURE**

Argonaute proteins have a bilobed architecture with four domains: MID (middle), PIWI (P-element induced wimpy testis), PAZ (PIWI/Argonaute/Zwille), and N-terminal domains (Figure 2.1 a). These domains are highly conserved between eukaryotic and prokaryotic proteins. The MID domain interacts with the 5' phosphate of the guide [3, 4]. The PIWI domain contains an RNase H-like active site and catalyzes the slicing activity [6-8]. The PAZ domain binds the 3' end of the guide. This interaction protects the guide from being degraded, especially in eukaryotic Argonautes [9, 10]. The N-terminal domain is important for target cleavage and the dissociation of the cleavage products [11, 12]. This section describes target search mechanisms in animal Argonaute proteins, in particular human and mouse Argonaute.

9-12), the 3'supplementary region (nucleotides 13-16), and the 3' tail (nucleotides 17-22) [26-29]. The seed region plays the key role in target search. Structures of humanAGO2 have shown that the seed nucleotides 2-6 in the guide are preordered in an A-form helix and exposed to the solvent [30](Figure 2.1 b). Such preordering helps Argonaute overcome the entropic cost of target base pairing. A single-molecule fluorescence study on mouse AGO2 has shown that Argonaute increases the rate with which RNA binds to its target to levels limited by diffusion, compared to naked guides binding their targets alone, confirming that preordering facilitates seed recognition [31].

The crystal structure of hAGO2 bound to a guide and target proposed a stepwise mechanism for target binding: Pairing of nucleotides 2-5 with the target promotes a conformational change exposing nucleotides 2-8 and 13-16 for target base pairing [32]. An smFRET study of hAGO2 has shown that the first three nucleotides in the seed are the most important in determining the binding rate (Figure 2.1 c). The number of consecutive complementary nucleotides does not affect binding rate at all as long as there are no mismatches in the seed. However, the stability of binding is determined by the degree of seed complementarity[33]. By varying the number of consecutive complementary nucleotides between guide and target, Chandradoss et al. [33] have shown that binding to the first six nucleotides has a small effect on the dwell time of hAGO2 on the target RNA. Increasing complementarity past seven consecutive nucleotides has a drastic effect with binding events lasting throughout the whole measurement (~300 s) (Figure 2.1 d). Therefore, hAGO2 uses the first seed nucleotides that are exposed to the solvent to probe potential target sites, and the rest of the seed further stabilizes binding, confirming the stepwise model proposed by previous biochemical and bioinformatics studies.



**Figure 2.1. Target search mechanism of eukaryotic Argonaute.** a) Crystal structure of human Argonaute with guide RNA [32]. b) Argonaute exposes the first three nucleotides of the seed to the solvent [33]. c) Schematic of single-molecule hAGO2 FRET assay [33]. d) Representative time traces showing hAGO2 binding to the target with different seed-target complementarity [33]. e) Binding rate of mouse and human Argonaute proteins with dinucleotide mismatched guide RNAs [31, 34]. f) Schematic of hAGO2 tandem target FRET assay [33]. N1 and N2 represent the number of complementary nucleotides at the first and second binding sites, respectively. g) Representative time trace showing hAGO2 transitioning between two FRET states [33].

Other single-molecule fluorescence studies have also shown that seed region is important in achieving fast target search. A study on mouse AGO2 explored the effects of dinucleotide mismatches along the seed region [31]. It was shown that the lack of guide and target complementarity within the first six seed nucleotides decreases the binding rate dramatically. A similar result was obtained by an smFRET study on hAGO2 that explored the effect of dinucleotide mismatches along the full guide [34]. In particular, both studies have found that mismatches in the middle and 3' end of the seed reduced binding rate more than mismatches on the 5' end of the seed (Figure 2.1 e). These findings are in contrast with Chandradoss et al. [33], who showed that the first three nucleotides in the seed are the most important. This disparity might arise from the fact that, in the two studies described above, the remainder of the guide was fully complementary to the target and could possibly compensate for the mismatches at the beginning of the seed. Despite different findings, all three studies show that the seed region of the guide RNA is crucial for binding rate and stability.

### 2.2.2. LATERAL DIFFUSION

Argonaute proteins have to find their target in a large pool of cellular RNAs. Furthermore, miRNA target sites are often found in the 3' UTR of mRNA, which can be several kilobases long. It is possible that Argonaute target search is facilitated by lateral diffusion as was first hinted by biochemical studies [26]. To test this hypothesis, Chandradoss et al. [33] designed a tandem target assay where two identical target sites were separated by 22 nucleotides on a single RNA strand (Figure 2.1 f). Binding to one side would yield a high FRET value, and binding to the second target site would show low FRET. Using a target of six-nucleotide complementarity with the guide, it was observed that over 70% of binding events showed rapid FRET changes, suggesting that Argonaute is shuttling between the target sites (Figure 2.1 g). It still remains to be determined whether the observed FRET changes are due to sliding (the protein contains constant contact with the RNA), hopping (multiple association and dissociation events are correlated along the con-

tour of the target RNA), or jumping (the protein can jump to another binding site that is physically close but far in sequence).

### 2.2.3. CONFORMATIONAL CHANGE

Full pairing of Argonaute guide and target requires a protein conformational change. Eukaryotic and archaeal Argonautes introduce a kink in their guide by an alpha helix, termed helix 7, which would have to shift to accommodate pairing between nucleotides 6 and 7[30]. Indeed, pairing of the first five nucleotides relieves helix 7 and allows further nucleotide pairing [32, 35]. However, for further basepairing, the central cleft, where guide and target molecules are accommodated, has to widen, as narrowing of the central cleft restricts pairing past guide nucleotide 8. The conformational changes after the turn of the helix 7 may not be important for silencing activity, but the movement of the PAZ domain and opening of the channel between PAZ and N-terminal domains are necessary to accommodate full guide-target base pairing for RNA slicing.

### 2.2.4. COOPERATIVITY

The possibility that neighboring target sites can act cooperatively to retain the AGO-RNA complex on the target RNA has been previously hinted at by biochemical studies, but the mechanism was not clear [36-38]. A possible explanation is provided by the observation of lateral diffusion. The smFRET study using tandem target assay shows a drastic difference between the dwell times on a single target and tandem target with residence times on the tandem target constructs being nearly 10 times higher than those on a single target. This result confirmed that the neighboring target sites act synergistically to retain AGO-miRNA on the target strand. Remaining bound to the target RNA for longer could decrease the energetic cost that comes with protein having to change conformation upon associating and dissociating with the target multiple times. In addition, increased dwell time gives more time to recruit other proteins necessary for degradation of mRNA.

## 2.3. CRISPR ADAPTIVE IMMUNITY

Bacteria and archaea use the RNA-mediated adaptive CRISPR/Cas immune system to defend against invading bacteriophages and plasmids [39, 40]. Different organisms have evolved distinct CRISPR systems. These systems are grouped into two main classes that are subdivided into six main

types [41, 42]. The signature of Class 1 CRISPR systems (types I and III) is the use of a multi-subunit protein complex for target recognition and degradation, whereas Class 2 systems (types II, V, and VI) use a single protein for this task. This section of the chapter focuses on the two best understood systems: type I [CRISPR/Cascade (CRISPR-associated complex for antiviral defense)] and type II [CRISPR/Cas9].

### 2.3.1. CASCADE PROTEIN COMPLEX

A majority of CRISPR systems found in nature belong to the type I CRISPR family. These types of CRISPR systems use a multi-subunit protein complex, Cascade, for the recognition of invading foreign DNA [43]. Cascade alone is unable to degrade target DNA and instead relies on the recruitment of a Cas3 endonuclease [44, 45]. Cascade complexes play a role not only in interference but also in a process called priming during which CRISPR memory is rapidly updated to fight escape mutants [46].

#### 2.3.1.1. Structural arrangements enable target search.

*Escherichia coli* Cascade consists of 11 Cas proteins (one copy of Cse1, two copies of Cse2, six copies of Cas7, one copy of Cas5, and one copy of Cas6) and adopts a sea horse-like structure [47, 48] (Figure 2.2 a,b). Its crRNA has a less intricate architecture than that of Cas9, consisting only of a 3' stem loop and a protospacer region. Similar to Cas9, crRNA in Cascade is pre-ordered in a pseudo A-form helical configuration. The spacer sequence is divided into segments by the flipping out of every sixth nucleotide. The nucleotides within the first two segments are crucial to target binding and are defined as the seed [49]. Mismatches in the other segments are much better tolerated and can still lead to successful interference.

#### 2.3.1.2. PAM search

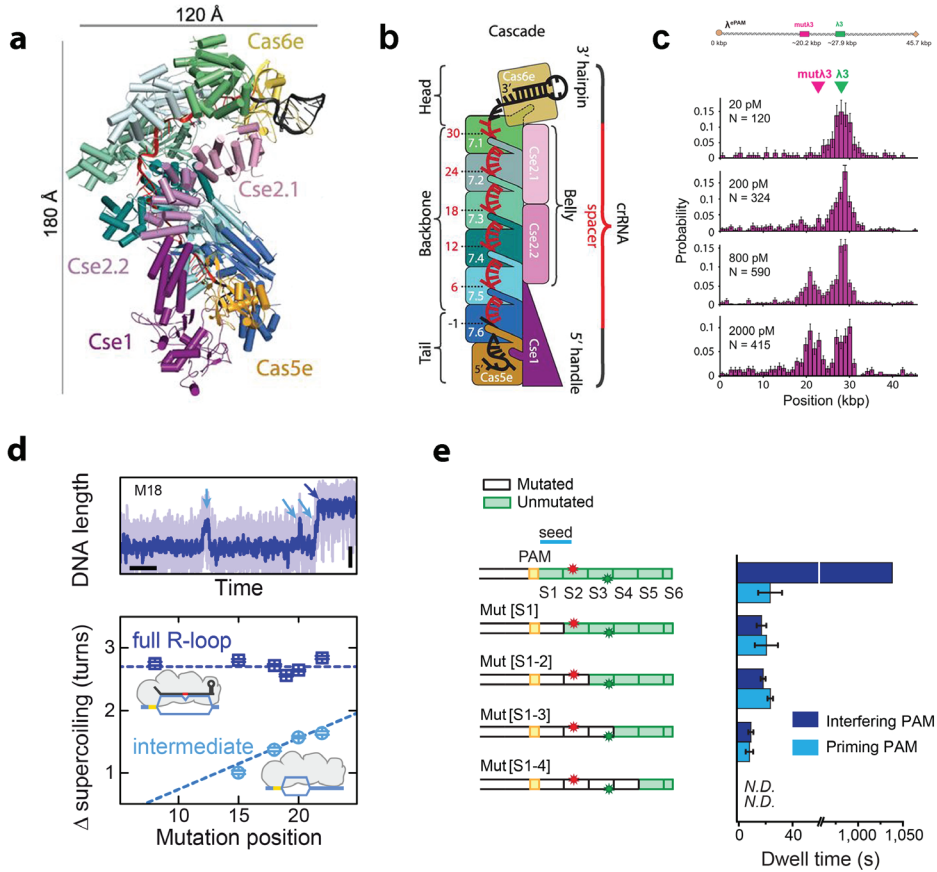
Cascade recognizes a PAM (Protospacer adjacent motif) sequence on the 5' end of the target, on the non-target strand. Cascade PAM recognition is promiscuous with at least five interfering PAM sequences identified for *E. coli* Cascade [44, 51, 52]. The PAM is recognized by Cse1 in double-stranded form, from the minor-groove site. Such mode of the minor-groove recognition indicates that mutated PAMs can be tolerated as long as target sequence is optimal [53]. This is supported by a DNA curtain study where Cascade was



able to bind a fully matching protospacer that was lacking a PAM but with a much reduced binding rate [54] (Figure 2.2 c).

### 2.3.1.3. Directional R-loop formation

Upon binding a PAM site, Cascade interrogates the target DNA and forms an RNA-DNA R-loop in a directional fashion as shown by magnetic tweezers experiments [55]. Point mutations in the seed region required higher negative supercoiling for R-loop formation than PAM-distal mutations. This



**Figure 2.2. Target search of Cascade protein complex.** a) Crystal structure of type I-E *Escherichia coli* Cascade with guide RNA [47]. b) Schematic of Cascade [47]. c) Histogram showing binding probability to cognate target and a protospacer lacking a PAM at different Cascade concentrations in DNA curtains experiments [54]. d) R-loop formation trajectory for protospacer with a point mutation at position 18 and mean supercoiling changes associated with full and intermediate R-loop formation [55]. e) Schematic of protospacer mutations for smFRET assay and dwell times for each construct [57].

agrees with the CHAMP (chip-hybridized association-mapping platform) assay, where it was found that the farther the point mutations are from the PAM, the more tolerated they are [56]. In addition, in magnetic tweezers experiments, Cascade stalling at mutations was observed, which needed appropriate supercoiling (approximately fourfold higher for mutations close to PAM compared to wild type) to overcome the intermediate state for full R-loop formation (Figure 2.2 d). Reversible binding events were also observed with DNA curtains, suggesting Cascade might make multiple attempts before it stably engages with the protospacer.

#### 2.3.1.4. Non-directional binding

Despite magnetic tweezers assays suggesting only directional formation of R-loop, an smFRET study has shown that Cascade is able to bind DNA in a sequence specific but non-directional manner [57]. Targets with a cognate PAM and fully complementary DNA sequence exhibited two types of binding events—long events characterized by an initial high FRET that soon transitioned to a low FRET state, and short-lived events exhibiting low- or mid- FRET state, corresponding to partially unwound DNA. Targets with a mutated seed region showed only the second type of events (Figure 2.2. e). This is in contrast to Cas9, where seed mutations completely abolish target binding. It was confirmed *in vivo* that the constructs that exhibited noncanonical binding modes triggered the priming response, which allows CRISPR memory to be rapidly updated.

#### 2.3.1.5. Cas3 recruitment

Cascade does not degrade the target itself but rather recruits the Cas3 nuclease [45]. In the magnetic tweezers experiments, it was shown that R-loop locking is required for the recruitment of Cas3, regardless of any mutations in the protospacer. However, mutations of the PAM significantly affected Cas3 cleavage, even if the R-loop was fully formed and was in its locked state, implying a dual signaling mechanism upon target recognition. Consistent with these findings, DNA curtains assay has shown that Cascade bound to a target flanked by a PAM could readily recruit Cas3 nuclease for DNA degradation. However, at PAM-lacking sites, Cascade could not directly recruit Cas3. Finally, the CHAMP assay also suggested that Cas3 is recruited in a DNA sequence-dependent manner.

### 2.3.2. Cas9 ENDONUCLEASE

Since the year 2012, the type II *Streptococcus pyogenes* Cas9 (SpyCas9) protein has been at the center of attention for genome engineering purposes, owing to its simplicity and programmability [58, 59]. This protein is a large endonuclease, consisting of 1,368 amino acids and multiple domains (Figure 2.3 a). SpyCas9 recognizes a 3-nucleotide PAM (protospacer adjacent motif) adjunct to the 3' end of the 20-nucleotide target sequence and cleaves the target three basepairs downstream from the PAM (see box 2.2 "PAM recognition"). Unlike most other CRISPR systems, SpyCas9 needs two RNA molecules, namely crRNA and trans-activating RNA (tracrRNA), to find and destroy the target [17, 60]. For genome editing purposes, the two RNA molecules can be fused into one single-guide RNA that maintains full functionality of the effector complex [61, 62]. Binding to RNA is crucial for Cas9 targeting, as it enables structural rearrangements necessary to accommodate a DNA target and contains the guide sequence, which is complementary to the target (see box 2.3 "Cas9 preorganization and structural rearrangement for target search")

#### 2.3.2.1. PAM search

PAM recognition is the first step in Cas9 target search and is an intrinsically complex protein-DNA interaction. Binding to the canonical PAM triggers local melting of the DNA at the PAM-adjacent nucleation site [20]. This is followed by the directional formation of RNA-DNA hybrid and the displacement of the nontarget strand (R-loop formation) [55, 63, 64]. Single nucleotide mutations in the PAM are able to slow down or abolish binding and R-loop formation, as shown by magnetic tweezers DNA-supercoiling assays [64]. In these experiments, performed with *Streptococcus thermophilus* Cas9 (StCas9), DNA was negatively supercoiled to assist R-loop formation (Figure 2.3 d,e). When PAM was mutated four nucleotides away from the seed, R-loop formation was still observed but at a much lower rate (Figure 2.3 f). PAM mutations in the positions closer to the seed slowed down R-loop formations even more. Therefore, mutations in the PAM alter R-loop formation by kinetic instability, which renders Cas9 unable to recognize the target and start R-loop formation.

DNA curtain experiments (Figure 2.3. j) showed that PAM recognition involves intrinsically weak interactions [5]. While Cas9 remains stably bound to a bona fide target site, only short-lived interactions are observed with off-targets. The off-target binding distribution correlates with the PAM distribution on the lambda DNA, consistent with other studies, showing that Cas9 samples PAM sites before it finds and stably associates with its target

**Box 2.2 PAM RECOGNITION**

CRISPR systems target specific sequences using Watson-Crick base pairing between guide RNA and target DNA to recognize and cleave the target [1]. In addition to the target sequence complementary to the guide RNA, specific Cas proteins involved in DNA interference recognize a PAM (protospacer adjacent motif) sequence as the first step of target search [17]. Although different in sequence and placement between different CRISPR systems, a PAM sequence is always present adjacent to the target site. The main role of PAM is to act as an indicator for self-nonself discrimination: The spacer sequences integrated in the host genome are identical to those in the invading DNA; hence the host could recognize and cleave its own DNA, which would be fatal to the cell [24]. In contrast, the protospacer sequences in the invader's genome are always flanked by a PAM sequence, which is not integrated in the CRISPR locus. Therefore, upon recognizing the PAM sequence as the first step prior to recognizing the target via Watson-Crick base pairing, the host ensures that the invader is destroyed and the integrity of its own genome remains protected.

site. The short-lived interactions are characterized by a double-exponential decay, indicating that Cas9 has at least two distinct modes when searching for PAM (Figure 2.3 k).

An insight into different binding modes is provided by smFRET experiments that can probe local interactions around PAM sites [65] (Figure 2.3 h). If there are neither PAM nor target sites present on the DNA, Cas9 binding is random and short lived ( $<0.5$  s). However, if there is at least a single PAM site present, Cas9 exhibits two distinct types of behavior: short transient binding to a PAM site ( $<0.5$  s) and more stable binding ( $\sim 2$  s). This implies that, upon binding a PAM site, Cas9 can either dissociate quickly upon failing to form an RNA-DNA duplex or diffuse locally around the PAM, looking for adjacent PAM sites and trying to form an R-loop there. Therefore, Cas9 might use a combination of short-lived 3D diffusion and long-lived 1D diffusion for PAM search. In vivo, Cas9 has been found to spend approximately a subsecond on PAM sites [66, 67].

However, some Cas9 molecules stay stably bound for longer than 5 s, despite the fact that there is no target present, indirectly suggesting that Cas9 may be searching for adjacent PAM sites flanked by a cognate target sequence, potentially using lateral diffusion. Biochemical data have revealed another layer of complexity, showing that Cas9 is able to bind DNA substrates with no target, but multiple PAM sites in electrophoretic mobility shift assays [5]. This peculiar observation may be explained by local diffu-

sion on the target strand that creates a synergistic effect between neighboring PAM sites.

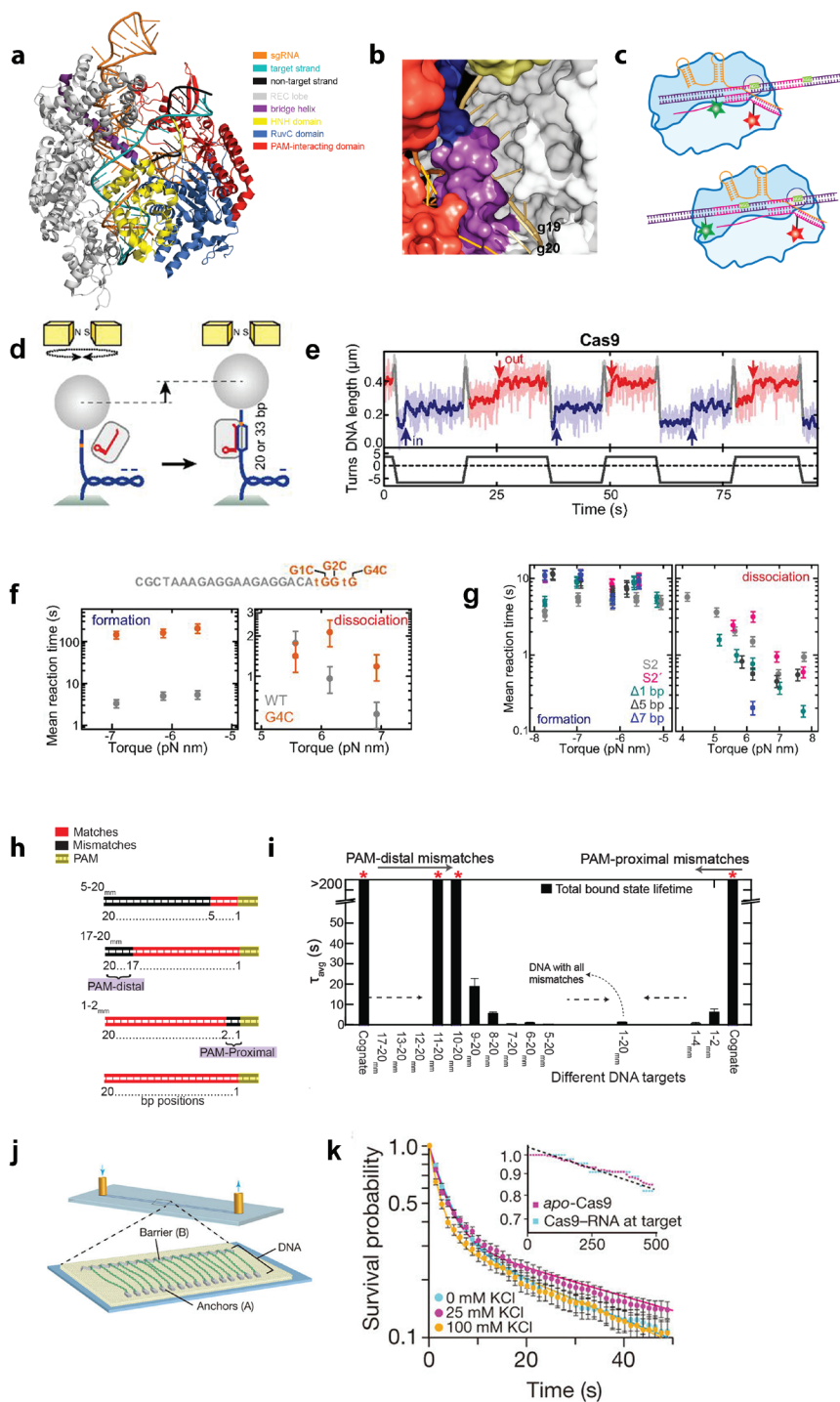
### 2.3.2.2. Seed recognition

The next step after recognizing the correct PAM is the recognition of a seed sequence on the target DNA. This sequence is the first 8–12 nucleotides downstream from the PAM. Recognition of the seed via Watson-Crick base-pairing between guide RNA and target DNA is crucial for stable binding. A magnetic tweezers study investigated the effect of protospacer truncations on the stability of the RNA-DNA R-loop using StCas9[64]. For 1- or 5-base-pair (bp) truncations from the PAM-distal end, R-loop stability was slightly reduced. R-loops were detected for truncations up to 7 bp with little change in the association rate but were not detected for 9-bp truncations (Figure 2.3 g). R-loops of 11 bp or shorter were not formed, revealing directional R-loop formation.

Additional evidence for the directionality of R-loop formation by Cas9 has been shown by an smFRET study [68]. Here, immobilized DNA is labeled with a donor dye at the PAM-distal end and RNA is labeled at the 5' end with an acceptor dye. Upon full complementarity, a high FRET state was observed. However, it was also possible to capture an intermediate FRET state that corresponds to only PAM-proximal base pairing before a high FRET state was reached. These findings show substeps in guide-target pairing and confirm the directionality of R-loop formation between RNA and DNA.

Another smFRET study has explored the effects that both PAM-proximal and PAM-distal mismatches have on Cas9 protospacer binding [69] (Figure 2.3 h). Even 2-bp PAM proximal mismatches are able to severely decrease binding, and 4-bp PAM-proximal mismatches decrease binding to the levels of fully mismatched targets. In contrast, PAM-distal mismatches are much better tolerated, with up to 12 mismatches showing the binding stability as the cognate target (Figure 2.3 i). Together, these results underscore the

**Figure 2.3. Cas9 target search and recognition.** a) Crystal structure of SpyCas9 with single-guide RNA and target DNA b) SpyCas9 preorders the first 10 guide nucleotides in a helical configuration and exposes nucleotides 19–20 to the solvent. c) Schematic of Cas9 multiple-PAM assay [65] d) Schematic of magnetic tweezers assay [64] e) Time trajectory of the DNA length. Cas9 binds at negative supercoiling, thus increasing DNA length and dissociates at positive supercoiling also increasing DNA length [64] f) Mean reaction times for R-loop formation and dissociation as a function of torque for WT protospacer and a PAM (G4C) mutant [64] g) Mean reaction times for R-loop formation and dissociation as a function of torque for different protospacer truncations [64] h) Protospacer mutation scheme for smFRET assay [69] i) Dwell times for DNA targets with different lengths and positions of mutations [69] j) Schematic of DNA curtains assay [5] k) Survival probabilities for off-target binding events are represented by a double-exponential decay [70]





importance of the seed region, which in this study has been shown to be eight nucleotides. This shows that early mismatched regions are able to stop R-loop formation and abolish binding regardless of the sequence downstream of the mismatched region.

### 2.3.2.3. Final stages of target recognition

Initial target search of Cas9 is a complex process involving multiple binding modes. Although finding the seed is enough for stable binding, Cas9 cleavage requires more stringent Watson-Crick base pairing. If a target is extensive enough for stable binding but not extensive enough for cleavage, Cas9 undergoes dynamic conformational changes. Single-molecule and bulk FRET experiments where the two nuclease domains, HNH and RuvC, were labeled have shown that the movement of the HNH domain from the PAM-distal end to the cleavage site is possible only when the complementa-

#### Box 2.3. Cas9 REORGANIZATION AND STRUCTURAL REARRANGEMENT FOR TARGET SEARCH

##### Apo-Cas9

Apo-Cas9 has a bilobed architecture with one lobe (nuclease lobe) containing the HNH, RuvC, and C-terminal domains and the other (recognition lobe) containing a large helical domain [2]. Apo-Cas9 is able to bind DNA; however, it displays no sequence specificity, as shown by DNA curtain assays [5]. The nonspecific DNA binding showed strikingly long lifetimes. However, in the presence of heparin or guide RNA as competitors, Apo-Cas9 quickly dissociates from the DNA strand.

##### Structural Arrangement for PAM and Seed Recognition

TracrRNA activates the Cas9-RNA complex. An important rearrangement upon binding crRNA and tracrRNA occurs in the C-terminal domain, also known as the PAM-interacting domain, which then forms a groove that can accommodate the PAM sequence in its DNA duplex form. Binding to guide RNA therefore enables Cas9 to look for PAM sites in a sequence-specific manner [20]. Similar to Argonaute (Ago) systems, the first 10 nucleotides in the seed region of the crRNA are preordered in an A-form helix, with the first nucleotides exposed to the solvent for initial DNA interrogation (Figure 3b). In addition, a kink is introduced into the guide RNA by an insertion of an amino acid (Tyr) between nucleotides 15 and 16, which is relieved upon target binding [2, 23].

rity between guide RNA and target DNA is no less than 18 nucleotides [70, 71]. Four PAM-distal mismatches are enough to stop Cas9 from reaching the final conformation state, leaving Cas9 transitioning between the initial and intermediate states.

## 2.4. INTEGRATED VIEW ON TARGET SEARCH AND RECOGNITION

Protein target search is a complicated process involving different search modes comprising weak interactions and protein conformational changes. Despite the difference in function, target search mechanisms of proteins from different families like Argonaute and CRISPR share a lot of similarities. To begin with, these proteins recognize a short nucleic acid sequence as an initial recognition step: Argonaute recognizes the first 3 nucleotides of the seed while CRISPR/Cas proteins recognize a PAM sequence [5, 33, 50, 54] (Figure 2.4; Table 2.1). Furthermore, all described proteins use a mixture of 3D and 1D diffusion to efficiently locate their targets [5, 33, 54, 65, 72]. Strikingly, all proteins can laterally diffuse approximately 10 bp in length. Another similarity is that Argonaute and Cas9 proteins do not require full target complementarity for stable base pairing, but instead, binding is stabilized by the pairing of the first 7–12 nucleotides, depending on the protein [33, 69] (Figure 2.4). In all described systems, the competent state requires more extensive base pairing, with Argonaute and CRISPR proteins requiring the full length and near full length of the guide respectively [34, 68, 70, 73]. In addition, Argonaute and CRISPR proteins undergo conformational changes during different stages of target recognition [35, 57, 70, 74–76] (Figure 2.4). All in all, despite subtle differences, these different protein families make use of the same core principles to find their targets in a fast and efficient manner.

### 2.4.1. MODES OF TARGET SEARCH

Target search is an essential part of the functioning of many different proteins. Despite differences in function, any target search and recognition should be both rapid and specific. The optimum way to achieve this is to use a combination of 3D diffusion and 1D diffusion while minimizing time spent on off-targets. See Table 2.1. for a summary. Human AGO2 of the Argonaute protein family achieves this by exposing only the first few nucleotides of the seed region and using them to probe potential target sequences. An smFRET study has shown that exposing the first three nucleotides facilitates target



search by lateral diffusion in which the hAGO2-miRNA does not dissociate from the RNA strand even if a fully matching target is not found but diffuses laterally to a neighboring target site [33]. Adjacent target sites also act synergistically to keep the protein from dissociating, thereby increasing the probability of finding a cognate target nearby.

	MOTIF FOR INITIAL RECOGNITION	REQUIREMENTS FOR STABLE BINDING	REQUIREMENTS FOR COMPETENT STATE	CONFORMATIONAL CHANGE UPON (...)	FACILITATED TARGET SEARCH
AGO	First 3 nt of seed[31, 33]	Seed pairing (7 nt)[33]	Full length guide pairing (~22nt) [34, 73]	Seed recognition[35] Full-length guide recognition[74, 76]	Lateral diffusion (>10 nt)[33]
Cas9	PAM (3 nt)[5, 78]	PAM[5, 64, 69]; Seed pairing (8 nt)[69] (13 nt) [64]	PAM[68]; Near-full length guide pairing (18-20 nt) [68]	Seed recognition[70] Full-length guide recognition[70]	Lateral diffusion (>10 nt)[65]
Cas9	PAM (3 nt)[54]	Near full-length guide pairing (28-32nt)[55, 64]	PAM[54, 55, 64] Near full-length guide (28-32 nt) [55, 64]	Near-full length guide recognition[57], [75]	Lateral diffusion (>100 bp) [72]

Table 2.1. Summary of Argonaute and CRISPR target search systems

Similar to Argonaute, Cas9 uses a mixture of 3D and lateral diffusion to find its target. First, the protein weakly associates with a PAM site, interrogating the adjacent sequence for complementarity, as shown in DNA curtain experiments [5]. An smFRET study has additionally shown that Cas9 has two binding modes, one being a specific PAM and guide-target-mediated interaction and another, termed sampling mode, being a search mode that does not involve RNA-DNA base pairing and is likely protein sampling the DNA for a PAM site—with one of the ways possibly being lateral diffusion [69]. Another smFRET study has shown that Cas9 is able to transition without dissociation between adjacent PAMs or PAMs with partial target sites [65]. The CRISPR type I Cascade protein complex has been shown to use 3D diffusion to find its target [54]. Recently, it has also been shown that the Cascade protein complex from a thermophilic organism (*Thermobifida fusca*) also uses 1D diffusion during its target search [72].

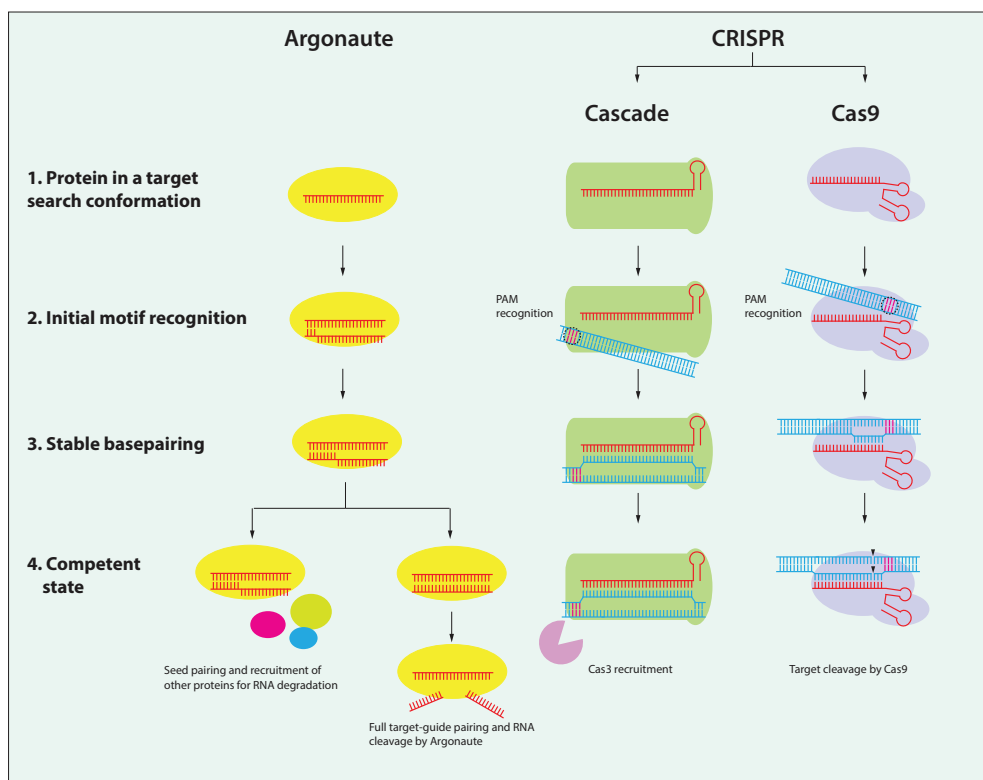
#### 2.4.2. MECHANISM OF KINETIC PROOFREADING

Initial weak interactions with the subseed region in the case of Argonaute and with PAM in the case of CRISPR proteins aid in fast dismissal of off-target sites. When the initial target recognition step is successful, further proofreading occurs via intricate protein conformational changes.

In the case of human Argonaute, the first conformational change the protein undergoes under binding to a cognate target is the release of the kink between guide nucleotides 6 and 7, which allows for further guide-target base pairing [32, 35]. This kink is released by rotating the alpha helix 7 by 44° and is possible only if guide nucleotides 2–5 are fully complementary to the target. Such conformational rearrangement allows for full seed basepairing and sharply increases affinity (human AGO2 residence times on its target). Seed basepairing, however, is not enough to trigger the catalytic activity of this protein. Binding to the seed widens the channel between the PAZ and N-terminus domains, which in turn allows for the disordered supplemental region of the guide RNA (nucleotides 13–16) to adopt an A-helical conformation [32]. Such preordering would decrease the entropic cost of target pairing even further. Base pairing in the mid-region has also been shown to be crucial for target cleavage, as shown by biochemical and single-molecule studies [34].

Despite full complementarity, Argonaute is unable to cleave some targets [34]. A possible explanation could be that certain sequences are unable to trigger the final conformational change in the middle region that would position the catalytic residues next to the cleavage site. It is also suggested that off-target site rejection is assisted by the interactions with the 3' end of the

guide and the PAZ domain of the protein. Single-molecule FRET studies have shown that modifications at this end of the guide slow down the protein dissociation from the target, which could lead to potential cleavage of off-target sites [74]. Further single-molecule and structural studies will reveal the full conformational proofreading mechanism of hAGO2 even further and answer lingering questions such as why some sequences cannot be cleaved despite full complementarity.



**Figure 2.4.** Graphic summary of the target search mechanisms of Argonaute, Cascade and Cas9 proteins

The Cas9 protein also undergoes extensive conformational changes throughout its target search. The major conformational change occurs upon binding guide RNA, which enables Cas9 to search for PAM in a sequence-specific manner [20]. Upon binding to the correct PAM, Cas9 bends the DNA so that the duplex could be unzipped and interrogated. The first 10 nucleotides of the seed are preordered in an A-helical configuration, thus prepaying the entropic cost for target basepairing [23]. As in the case with Argonaute, binding to the seed is enough to stabilize the Cas9-RNA-DNA complex [69]. However, biochemical data have shown that cleavage for such targets is very

inefficient. Further structural and single-molecule studies have revealed that Cas9 undergoes another conformational change as the complementarity between guide and target increases [63, 70]. Initially, the HNH domain that cleaves the target strand is positioned at the PAM-distal end, far from the cleavage site. However, when a full target has been found, the HNH domain moves to the cleavage site, thus achieving a catalytically active conformation. Single-molecule and bulk FRET experiments showed that at least 18 of 20 nucleotides between the target and guide have to be complementary for Cas9 to achieve target cleavage [70, 71]. As the number of mismatches is increased, the HNH domain is unable to pass through an intermediate conformation and cleave the target strand. The nontarget strand is cleaved by the RuvC domain, which is initially already positioned close to the cleavage site. However, without the movement of HNH domain, the nontarget strand also cannot be cleaved, indicating a signaling mechanism between the two domains as the final checkpoint before target cleavage.

The Cascade protein complex does not have a prominent conformational proofreading mechanism for binding. This could be a potential explanation why it can bind targets without a PAM or with significant mismatches in the seed. However, binding to a fully complementary target “locks” Cascade and stabilizes the R-loop [55, 64]. It has been shown that this locking and the presence of a correct PAM sequence is required to recruit the Cas3 nuclease for target degradation [54, 55]. Binding to partial targets instead triggers a priming response, where CRISPR memory is rapidly updated to fight escape mutants [57]. In this response, it is likely that the Cas1-Cas2 protein complex, which is responsible for the spacer integration in the CRISPR locus, is necessary to recruit Cas3 [54, 56].

## REFERENCES

1. Garneau, J.E., et al., *The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA*. *Nature*, 2010. **468**(7320): p. 67-71.
2. Jinek, M., et al., *Structures of Cas9 endonucleases reveal RNA-mediated conformational activation*. *Science*, 2014. **343**(6176): p. 1247997.
3. Boland, A., et al., *Crystal structure and ligand binding of the MID domain of a eukaryotic Argonaute protein*. *EMBO Rep*, 2010. **11**(7): p. 522-7.
4. Frank, F., N. Sonenberg, and B. Nagar, *Structural basis for 5'-nucleotide base-specific recognition of guide RNA by human AGO2*. *Nature*, 2010. **465**(7299): p. 818-22.
5. Sternberg, S.H., et al., *DNA interrogation by the CRISPR RNA-guided endonuclease Cas9*. *Nature*, 2014. **507**(7490): p. 62-7.
6. Parker, J.S., *How to slice: snapshots of Argonaute in action*. *Silence*, 2010. **1**(1): p. 3.
7. Parker, J.S., S.M. Roe, and D. Barford, *Crystal structure of a PIWI protein suggests mechanisms for siRNA recognition and slicer activity*. *EMBO J*, 2004. **23**(24): p. 4727-37.
8. Song, J.J., et al., *Crystal structure of Argonaute and its implications for RISC slicer activity*. *Science*, 2004. **305**(5689): p. 1434-7.
9. Lingel, A., et al., *Nucleic acid 3'-end recognition by the Argonaute2 PAZ domain*. *Nat Struct Mol Biol*, 2004. **11**(6): p. 576-7.
10. Ma, J.B., K. Ye, and D.J. Patel, *Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain*. *Nature*, 2004. **429**(6989): p. 318-322.
11. Hauptmann, J., et al., *Turning catalytically inactive human Argonaute proteins into active slicer enzymes*. *Nat Struct Mol Biol*, 2013. **20**(7): p. 814-7.
12. Kwak, P.B. and Y. Tomari, *The N domain of Argonaute drives duplex unwinding during RISC assembly*. *Nat Struct Mol Biol*, 2012. **19**(2): p. 145-51.
13. Bauer, M. and R. Metzler, *Generalized facilitated diffusion model for DNA-binding proteins with search and recognition states*. *Biophys J*, 2012. **102**(10): p. 2321-30.
14. Berg, O.G., R.B. Winter, and P.H. von Hippel, *Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory*. *Biochemistry*, 1981. **20**(24): p. 6929-48.
15. Klein, M., et al., *Why Argonaute is needed to make microRNA target search fast and reliable*. *Semin Cell Dev Biol*, 2017. **65**: p. 20-28.
16. Slutsky, M. and L.A. Mirny, *Kinetics of protein-DNA interaction: facilitated target location in sequence-dependent potential*. *Biophys J*, 2004. **87**(6): p. 4021-35.
17. Jinek, M., et al., *A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity*. *Science*, 2012. **337**(6096): p. 816-21.
18. Makarova, K.S., Y.I. Wolf, and E.V. Koonin, *Comparative genomics of defense systems in archaea and bacteria*. *Nucleic Acids Res*, 2013. **41**(8): p. 4360-77.
19. Shabalina, S.A. and E.V. Koonin, *Origins and evolution of eukaryotic RNA interference*. *Trends Ecol Evol*, 2008. **23**(10): p. 578-87.
20. Anders, C., et al., *Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease*. *Nature*, 2014. **513**(7519): p. 569-73.

21. Ambros, V., *The functions of animal microRNAs*. Nature, 2004. **431**(7006): p. 350-5.
22. Bartel, D.P., *MicroRNAs: genomics, biogenesis, mechanism, and function*. Cell, 2004. **116**(2): p. 281-97.
23. Jiang, F., et al., *STRUCTURAL BIOLOGY. A Cas9-guide RNA complex preorganized for target DNA recognition*. Science, 2015. **348**(6242): p. 1477-81.
24. Marraffini, L.A. and E.J. Sontheimer, *Self versus non-self discrimination during CRISPR RNA-directed immunity*. Nature, 2010. **463**(7280): p. 568-71.
25. Meister, G., et al., *Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs*. Mol Cell, 2004. **15**(2): p. 185-97.
26. Ameres, S.L., J. Martinez, and R. Schroeder, *Molecular basis for target RNA recognition and cleavage by human RISC*. Cell, 2007. **130**(1): p. 101-12.
27. Bartel, D.P., *MicroRNAs: target recognition and regulatory functions*. Cell, 2009. **136**(2): p. 215-33.
28. Khorshid, M., et al., *A biophysical miRNA-mRNA interaction model infers canonical and noncanonical targets*. Nat Methods, 2013. **10**(3): p. 253-5.
29. Wee, L.M., et al., *Argonaute divides its RNA guide into domains with distinct functions and RNA-binding properties*. Cell, 2012. **151**(5): p. 1055-67.
30. Schirle, N.T. and I.J. MacRae, *The crystal structure of human Argonaute2*. Science, 2012. **336**(6084): p. 1037-40.
31. Salomon, W.E., et al., *Single-Molecule Imaging Reveals that Argonaute Reshapes the Binding Properties of Its Nucleic Acid Guides*. Cell, 2015. **162**(1): p. 84-95.
32. Schirle, N.T., J. Sheu-Gruttadauria, and I.J. MacRae, *Structural basis for microRNA targeting*. Science, 2014. **346**(6209): p. 608-13.
33. Chandradoss, S.D., et al., *A Dynamic Search Process Underlies MicroRNA Targeting*. Cell, 2015. **162**(1): p. 96-107.
34. Jo, M.H., et al., *Human Argonaute 2 Has Diverse Reaction Pathways on Target RNAs*. Mol Cell, 2015. **59**(1): p. 117-24.
35. Klum S. M., C.S.D., Schirle N. T., Joo C. and MacRae I. J., *Helix-7 in Argonaute2 shapes the microRNA seed region for rapid target recognition*. EMBO J, 2017.
36. Broderick, J.A., et al., *Argonaute protein identity and pairing geometry determine cooperativity in mammalian RNA silencing*. RNA, 2011. **17**(10): p. 1858-69.
37. Grimson, A., et al., *MicroRNA targeting specificity in mammals: determinants beyond seed pairing*. Mol Cell, 2007. **27**(1): p. 91-105.
38. Saetrom, P., et al., *Distance constraints between microRNA target sites dictate efficacy and cooperativity*. Nucleic Acids Res, 2007. **35**(7): p. 2333-42.
39. Barrangou, R., et al., *CRISPR provides acquired resistance against viruses in prokaryotes*. Science, 2007. **315**(5819): p. 1709-12.
40. Marraffini, L.A., *CRISPR-Cas immunity in prokaryotes*. Nature, 2015. **526**(7571): p. 55-61.
41. Makarova, K.S., et al., *An updated evolutionary classification of CRISPR-Cas systems*. Nat Rev Microbiol, 2015. **13**(11): p. 722-36.
42. Mohanraju, P., et al., *Diverse evolutionary roots and mechanistic variations of the*

*CRISPR-Cas systems*. Science, 2016. **353**(6299): p. aad5147.

43. Brouns, S.J., et al., *Small CRISPR RNAs guide antiviral defense in prokaryotes*. Science, 2008. **321**(5891): p. 960-4.
44. Mulepati, S. and S. Bailey, *In vitro reconstitution of an Escherichia coli RNA-guided immune system reveals unidirectional, ATP-dependent degradation of DNA target*. J Biol Chem, 2013. **288**(31): p. 22184-92.
45. Sinkunas, T., et al., *In vitro reconstitution of Cascade-mediated CRISPR immunity in Streptococcus thermophilus*. EMBO J, 2013. **32**(3): p. 385-94.
46. Fineran, P.C., et al., *Degenerate target sites mediate rapid primed CRISPR adaptation*. Proc Natl Acad Sci U S A, 2014. **111**(16): p. E1629-38.
47. Jackson, R.N., et al., *Structural biology. Crystal structure of the CRISPR RNA-guided surveillance complex from Escherichia coli*. Science, 2014. **345**(6203): p. 1473-9.
48. Zhao, H., et al., *Crystal structure of the RNA-guided immune surveillance Cascade complex in Escherichia coli*. Nature, 2014. **515**(7525): p. 147-50.
49. Semenova, E., et al., *Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence*. Proc Natl Acad Sci U S A, 2011. **108**(25): p. 10098-103.
50. Mojica, F.J., et al., *Short motif sequences determine the targets of the prokaryotic CRISPR defence system*. Microbiology, 2009. **155**(Pt 3): p. 733-40.
51. Hochstrasser, M.L., et al., *CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference*. Proc Natl Acad Sci U S A, 2014. **111**(18): p. 6618-23.
52. Westra, E.R., et al., *CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3*. Mol Cell, 2012. **46**(5): p. 595-605.
53. Hayes, R.P., et al., *Structural basis for promiscuous PAM recognition in type I-E Cascade from E. coli*. Nature, 2016. **530**(7591): p. 499-503.
54. Redding, S., et al., *Surveillance and Processing of Foreign DNA by the Escherichia coli CRISPR-Cas System*. Cell, 2015. **163**(4): p. 854-65.
55. Rutkauskas, M., et al., *Directional R-Loop Formation by the CRISPR-Cas Surveillance Complex Cascade Provides Efficient Off-Target Site Rejection*. Cell Rep, 2015.
56. Jung, C., et al., *Massively Parallel Biophysical Analysis of CRISPR-Cas Complexes on Next Generation Sequencing Chips*. Cell, 2017. **170**(1): p. 35-47 e13.
57. Blosser, T.R., et al., *Two distinct DNA binding modes guide dual roles of a CRISPR-Cas protein complex*. Mol Cell, 2015. **58**(1): p. 60-70.
58. Barrangou, R. and J.A. Doudna, *Applications of CRISPR technologies in research and beyond*. Nat Biotechnol, 2016. **34**(9): p. 933-941.
59. Jiang, W., et al., *RNA-guided editing of bacterial genomes using CRISPR-Cas systems*. Nat Biotechnol, 2013. **31**(3): p. 233-9.
60. Gasiunas, G., et al., *Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria*. Proc Natl Acad Sci U S A, 2012. **109**(39): p. E2579-86.
61. Doudna, J.A. and E. Charpentier, *Genome editing. The new frontier of genome engineering with CRISPR-Cas9*. Science, 2014. **346**(6213): p. 1258096.



62. Hsu, P.D., E.S. Lander, and F. Zhang, *Development and applications of CRISPR-Cas9 for genome engineering*. Cell, 2014. **157**(6): p. 1262-78.
63. Jiang, F., et al., *Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage*. Science, 2016. **351**(6275): p. 867-71.
64. Szczelkun, M.D., et al., *Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes*. Proc Natl Acad Sci U S A, 2014. **111**(27): p. 9798-803.
65. Globyte, V., et al., *CRISPR/Cas9 searches for a protospacer adjacent motif by lateral diffusion*. EMBO J, 2019. **38**(4).
66. Jones, D.L., et al., *Kinetics of dCas9 target search in Escherichia coli*. Science, 2017. **357**(6358): p. 1420-1424.
67. Knight, S.C., et al., *Dynamics of CRISPR-Cas9 genome interrogation in living cells*. Science, 2015. **350**(6262): p. 823-6.
68. Lim, Y., et al., *Structural roles of guide RNAs in the nuclease activity of Cas9 endonuclease*. Nat Commun, 2016. **7**: p. 13350.
69. Singh, D., et al., *Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9*. Nat Commun, 2016. **7**: p. 12778.
70. Dagdas, Y.S., et al., *A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9*. Sci Adv, 2017. **3**(8): p. eaao0027.
71. Sternberg, S.H., et al., *Conformational control of DNA target cleavage by CRISPR-Cas9*. Nature, 2015. **527**(7576): p. 110-3.
72. Dillard, K.E., et al., *Assembly and Translocation of a CRISPR-Cas Primed Acquisition Complex*. Cell, 2018. **175**(4): p. 934-946 e15.
73. Yao, C., et al., *Single-Molecule Analysis of the Target Cleavage Reaction by the Drosophila RNAi Enzyme Complex*. Mol Cell, 2015. **59**(1): p. 125-32.
74. Jung, S.R., et al., *Dynamic anchoring of the 3'-end of the guide strand controls the target dissociation of Argonaute-guide complex*. J Am Chem Soc, 2013. **135**(45): p. 16865-71.
75. Xue, C., N.R. Whitis, and D.G. Sashital, *Conformational Control of Cascade Interference and Priming Activities in CRISPR Immunity*. Mol Cell, 2016. **64**(4): p. 826-834.
76. Zander, A., et al., *Single-molecule FRET supports the two-state model of Argonaute action*. RNA Biol, 2014. **11**(1): p. 45-56.





# 3

## Cas9 searches for a protospacer adjacent motif using lateral diffusion

The *Streptococcus pyogenes* CRISPR/Cas9 nuclease (SpyCas9) has been widely applied in genetic engineering. Despite its importance in genome editing, aspects of the precise molecular mechanism of Cas9 activity remain ambiguous. In particular, because of the lack of a method with high spatio-temporal resolution, transient interactions between Cas9 and DNA could not be reliably investigated. It therefore remains controversial how Cas9 searches for protospacer adjacent motif (PAM) sequences. We have developed single-molecule Förster resonance energy transfer (smFRET) assays to monitor transient interactions of Cas9 and DNA in real time. Our study shows that Cas9 interacts with the PAM sequence weakly, yet probing neighboring sequences via facilitated diffusion. This dynamic mode of interactions leads to translocation of Cas9 to another PAM nearby and consequently an on-target sequence. We propose a model in which lateral diffusion competes with 3-dimensional diffusion and thus is involved in PAM finding and consequently on-target binding. Our results imply that the neighboring sequences can be very important when choosing a target in genetic engineering applications.

---

This chapter has been published as: Globyte, V., et al., *CRISPR/Cas9 searches for a protospacer adjacent motif by lateral diffusion*. EMBO J, 2019. **38**(4).

### 3.1. INTRODUCTION

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)-Cas systems are adaptive prokaryotic immune systems that provide bacteria and archaea with a defense mechanism against invading foreign genetic elements [1-6]. Upon infection, fragments of the invader's DNA are incorporated into the CRISPR locus in the host genome [7-10]. Those fragments are then transcribed into short CRISPR RNAs (crRNAs) which assemble with CRISPR-associated (Cas) proteins in order to recognize and destroy the invader when it returns [11]. The most famous of the discovered CRISPR systems is the type II system where the DNA of the invader is recognized and destroyed by the Cas9 protein which assembles with two RNA molecules, namely crRNA and trans-activating crRNA (tracrRNA) [4, 5, 10, 12-14]. The most widely researched Cas9 ortholog, *Streptococcus pyogenes* Cas9, recognizes a 20-nt target which is flanked by a PAM (protospacer adjacent motif) sequence on the 3' end of the target [12, 13]. The PAM sequence for SpCas9 is 5'-NGG-3'. CRISPR-Cas9 system has gained enormous attention due to its use in genome editing owing to its simplicity and programmability [15-18]. In order to edit a gene, Cas9 first has to find a small 23-nt sequence in a genome containing kilo-bases or mega-bases of DNA, in a crowded cellular environment. This process has been demonstrated to be slow, with a single Cas9 protein requiring 6 hours to locate a single target in a bacterial cell [19]. This example shows that efficient targeting of specific genes requires an advanced knowledge of Cas9 target search mechanism.

A recent single-molecule study using "DNA curtains" has shown that Cas9 uses only 3-dimensional diffusion to locate its target [20]. Due to diffraction limit (~100 nm) of the DNA curtain technique, it remains unknown whether the model of exclusive 3-dimensional target search is valid for the length scale of nucleotides. Other single-molecule studies have shown that RNA-guided proteins such as Argonaute or CRISPR type I Cascade protein complex use facilitated 1-dimensional diffusion during their target search [21, 22]. Using single-molecule FRET (Foerster Resonance Energy Transfer), we investigate the target search mechanism of SpCas9 and demonstrate that it uses 1-D diffusion along the DNA strand during its target search.

### 3.2. RESULTS

#### 3.2.1. SINGLE-MOLECULE OBSERVATION OF Cas9 PAM SEARCH

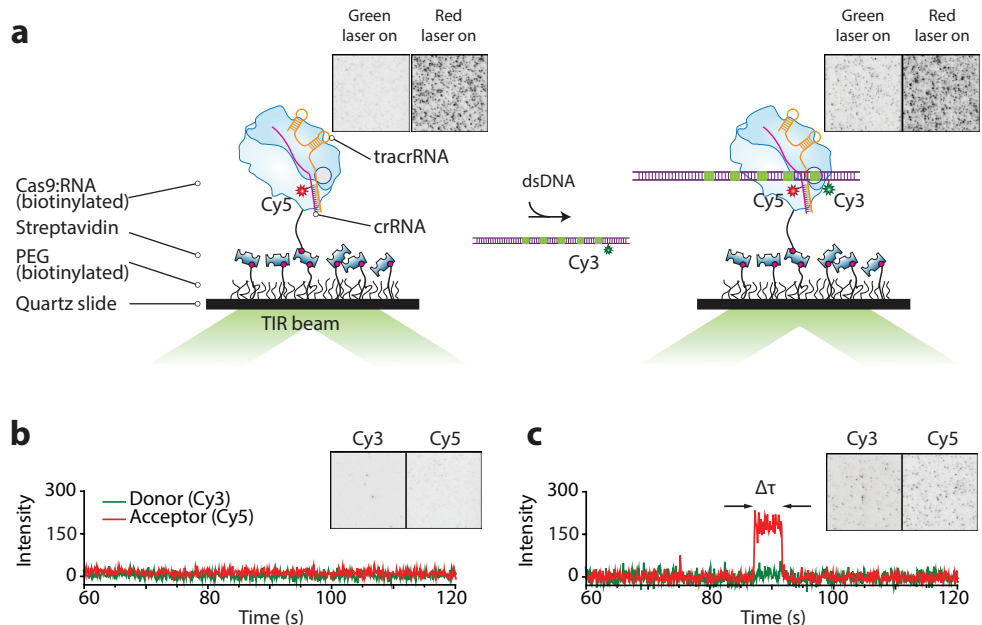
To visualize Cas9 target search process on a nanometer scale, we used single-molecule Förster resonance energy transfer (smFRET) technique. Biotinylated Cas9 (Figure S3.1a) was immobilized on a PEG-coated quartz surface

for long-term observation (Figure 3.1a) [21]. Biotinylation was shown not to affect Cas9 catalytic activity (Figure S3.1 b). The Cas9 protein was pre-incubated with dye-labeled crRNA and tracrRNA for 20 minutes before surface immobilization. Free-floating molecules, such as any unattached RNA or Cas9 molecules, were washed away, and remaining immobilized Cas9:RNA complexes could be directly imaged using total internal reflection microscopy (Figure 3.1a). Synthetic DNA and crRNA substrates were labeled with Cy3 and Cy5 dyes, respectively, such that the FRET efficiency between them would report on the position where Cas9 localizes on the DNA. Addition of Cy3-labeled DNA substrates did not affect the immobilization of Cas9:RNA complexes (Figure 3.1a). When excited with a green laser, binding events appear as spots on the CCD image (Cy3 signal on the left side and Cy5 signal on the right side) (Figure 3.1b, c). In fluorescence time traces binding events are characterized by increase in fluorescence intensity due to either direct excitation of donor or energy transfer from donor to acceptor (Figure 3.1 b, c).

The initial step in Cas9 target search is finding and recognizing a PAM sequence. It has been shown that, without encountering a cognate PAM, Cas9 cannot start R-loop formation, despite the full complementarity between the guide RNA and the target [23]. Therefore, in order to elucidate the mechanism by which Cas9 finds and recognizes PAM alone, we first investigated how the protein interacts with DNA strands containing only PAM sequences when no target sequence is present. In particular, multiple binding sites in close proximity have been shown to cause a synergistic effect in another RNA-guided target search system [24]. This effect emerges when the interaction between a searcher and a target is characterized by more than simple one-step binding and dissociation [21, 25]. Such a synergistic effect, if observed, would be an indication that Cas9 uses an additional mechanism to 3D diffusion when searching for PAM sequences.

To investigate how Cas9 interacts with multiple PAM sites and whether such synergistic effect exists in the CRISPR-Cas9 system, we designed DNA constructs containing 0, 1, 2, 3, 4 and 5 equidistant PAM sites (Figure 3.2a). DNA was labeled with a Cy3 dye at position -8 with respect to the first guanine on the first PAM site on the 5' end of the target DNA strand (Figure 3.2 a). crRNA was labeled with a Cy5 dye outside the guide region such that Cas9:RNA complex binding to the first PAM site would yield a high-FRET value (Figure 3.2 a, b). Binding to other PAM sequences would increase distance between dyes and therefore were expected to yield lower and distinct FRET values for each site, thus allowing to distinguish which PAM site was bound (Figure 3.2 b).

Negative control construct containing no PAM sites showed a very low number of binding events with random FRET values (Figure 3.2c, Figure



**Figure 3.1. Single-molecule assay for observation of Cas9-DNA interactions.** a) Illustrated schematic representation of our single-molecule assay. Cas9 assembled with tracrRNA and Cy5-labelled crRNA is immobilized on the surface via biotin-streptavidin links (left). After addition of Cy3-labelled DNA substrate interaction between Cas9 and DNA brings the dyes together resulting in FRET signal (right). EmCCD images show Cas9:RNA immobilized on the surface with green and red laser illumination (left). No fluorescence is seen when illuminated with green laser. Immobilized Cas9 can be seen when illuminated with red laser. After addition of Cy3-labelled DNA spots appear on the EmCCD in the green channel when illuminated with green laser (right). Size of EmCCD images is  $25\mu\text{m} \times 25\mu\text{m}$ . b) Example time trace when no binding event occurs (green laser excitation). EmCCD image shows signal from green and red channels before the addition of Cy3-labelled DNA when illuminated with green laser. Size of EmCCD image is  $25\mu\text{m} \times 50\mu\text{m}$ . c) Example time trace showing a binding event (green laser excitation). EmCCD image shows signal from green and red channels after the addition of Cy3-labelled DNA when illuminated with green laser. Size of EmCCD image is  $25\mu\text{m} \times 50\mu\text{m}$ .

S3.2a). Binding to DNA containing a single PAM yielded a narrow FRET distribution centered at 0.84 (Figure 3.2 c). Binding events showing different FRET states appeared when the number of PAM sites increased (Figure 3.2b). FRET histogram of all binding events for the construct containing 2 PAM sites shifted to a lower FRET value of 0.75 (Figure 3.2 c). Lower FRET states appeared for DNA constructs containing 3 and 4 PAMs with FRET histograms broadening significantly and the peaks of the histograms shifting to 0.73 and 0.69 respectively. (Figure 3.2 c). Finally, a histogram of all binding events for the construct containing 5 PAM sites was broad and centered at 0.53 – an average of all FRET values yielded by specific binding to any of the five PAM sites on the DNA strand (Figure 3.2 c). Furthermore, multiple bind-

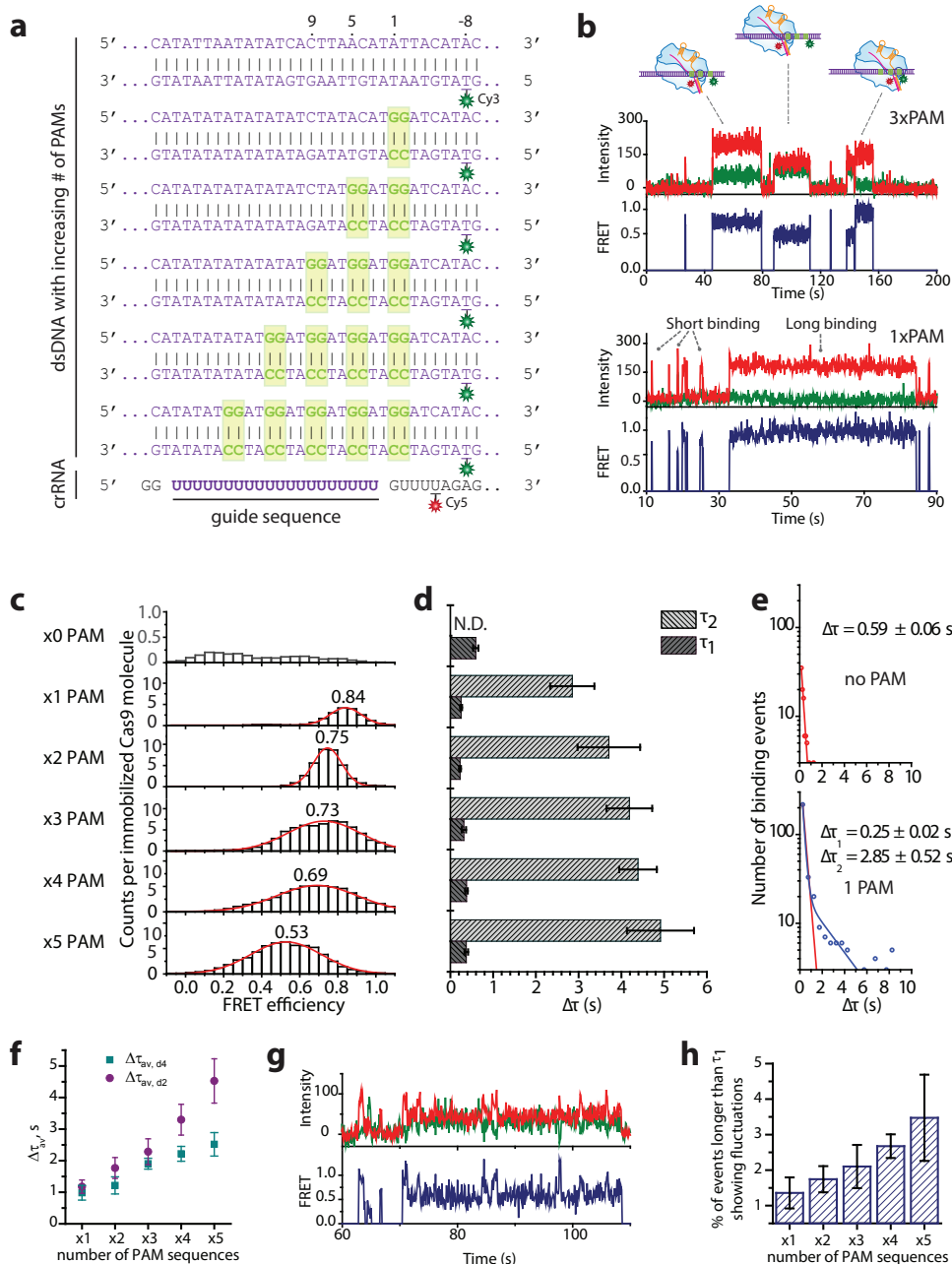
ing events showing FRET efficiencies corresponding to different PAMs could be observed in a single FRET trace showing that a single Cas9 can bind and dissociate from different PAM sites, as expected (Figure 3.2b).

In addition to broadening of the FRET histograms, the dwelltime was observed to increase with increasing number of PAM sequences (Figure 3.2 d). Experiments on DNA constructs without any PAM or target sequences only showed short-lived binding ( $\Delta\tau_1$ ), with a dwelltime of  $0.59 \pm 0.06$  s (Figure 3.2 d, e). Short binding events were also observed with PAM-containing DNA constructs ranging between  $0.25 \pm 0.02$  s to  $0.37 \pm 0.04$  s – similar for all constructs, regardless of the number of PAM sequences (Figure 3.2 d, Figure S3.2 b). Interestingly, a second type of longer binding events was observed for the constructs containing PAM sites as opposed to the negative control (Figure 3.2 b, d, e, Figure S3.2 b). It is further noted that the second dwelltime ( $\Delta\tau_2$ ) increased with increasing number of adjacent PAM sites from  $2.85 \pm 0.52$  s for the construct with a single PAM to  $4.91 \pm 0.78$  s for the construct with 5 PAM sites (Figure 3.2 d). In addition to  $\Delta\tau_2$ , the average dwelltime ( $\Delta\tau_{av}$ ) also was found to increase with increasing number of PAM sites from  $1.18 \pm 0.21$  s for a single PAM to  $4.53 \pm 0.70$  s for the construct containing 5 PAM sequences (Figure 3.2 f). Further analysis showed no correlation between FRET values and dwelltimes, suggesting that it is not the position of a PAM site, but rather the number of PAM sites in close proximity that caused this increase (Figure S3.2 a).

The observation that even for a single PAM the binding time is characterized by a double exponential distribution suggests that Cas9 uses another mechanism in addition to 3-D diffusion during its target search, as processes following exclusively 3-D diffusion follow one-step dissociation kinetics [26]. Furthermore, the increase of  $\tau_2$  implies that, due to the presence of multiple PAM sites in close proximity, Cas9 experiences a synergistic effect which causes it to stay bound to DNA for longer. We therefore hypothesize that upon encountering a PAM, Cas9 can follow two pathways. First, Cas9 can dissociate from DNA in a 3-dimensional fashion upon failing to form an RNA-DNA R-loop (corresponding to  $\tau_1$ ). Secondly, Cas9 can locally diffuse in a 1-dimensional fashion probing adjacent PAM sites (corresponding to  $\tau_2$ ).

To further investigate whether the observed increase in  $\tau_2$  could indeed be the result of 1-dimensional diffusion between the PAM, DNA sequences where the PAM sites were placed further apart (4 nucleotides) were designed (Figure S3.2 c). In this case, the separation between the two furthestmost PAM sites is 24 nucleotides. Dwelltime analysis revealed that the dwelltime distribution for each construct was again characterized by a double-exponential decay. Further analysis has shown that values of  $\tau_1$  were similar for all constructs, ranging between  $0.20 \pm 0.02$  s and  $0.55 \pm 0.10$  s (Figure S3.2 d).

These values also are similar to those observed in the case of 2 nucleotide separation between PAM sites (Figure 3.2 d), further supporting the hypothesis that  $\tau_1$  results from weak interactions with a single PAM site, followed by dissociation from the DNA strand. In addition, an almost identical increase



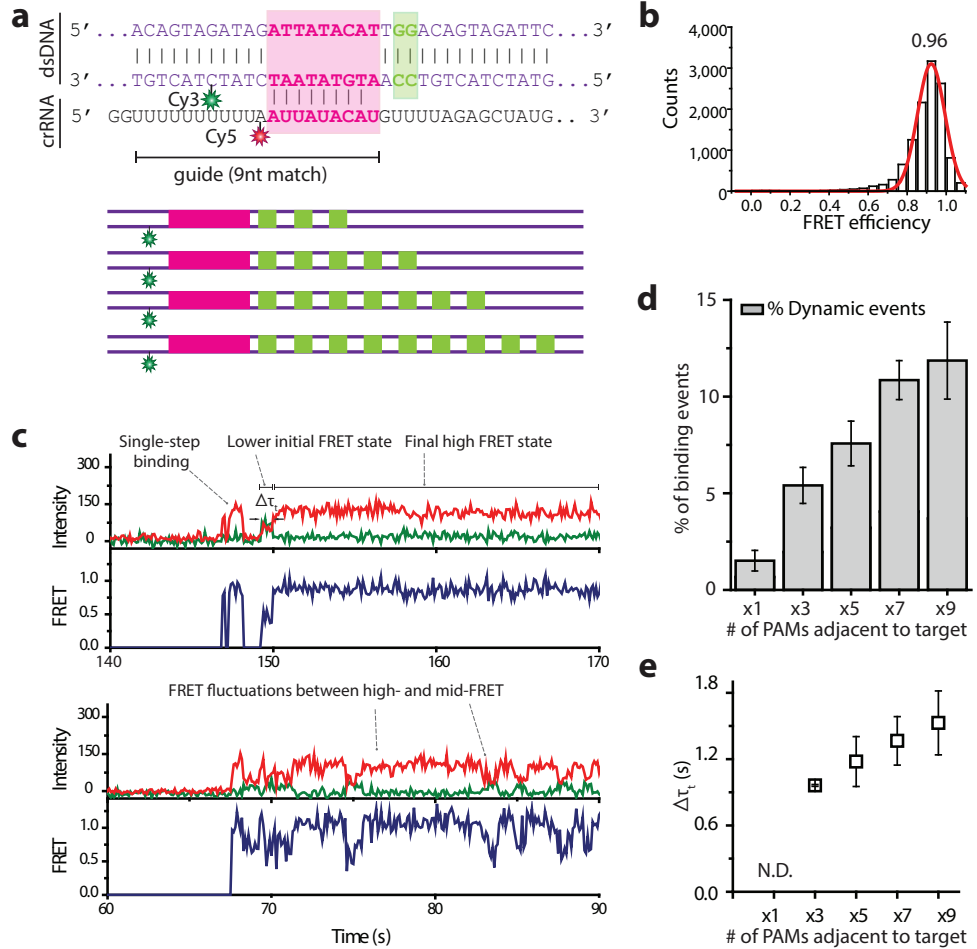
**Figure 3.2. PAM search mechanism.** a) DNA and crRNA sequences used for multiple PAM experiments with indicated labeling positions. Di-nucleotide GG PAMs are shown in boxes. b) Example time traces when DNA containing 3 PAM sequences is added (top) and when DNA containing a single PAM is added showing both short and long binding events (bottom). c) FRET histograms of all binding events for each construct. Y axis represents total number of counts, divided by the number of immobilized molecules in the field of view. d) Bar plot of all dwelltime values for all constructs showing short binding events for all DNA and long events only for the constructs containing PAM sequences. The times shown are mean values of dwelltimes obtained during four different experiments on four different days. Error bars represent standard error of the mean. e) Example dwelltime histograms from negative control and a construct containing a single PAM. Negative control dwelltime distribution is characterized by a single exponential decay (top) and dwelltime distribution for PAM-containing DNA is fitted by a double-exponential decay (blue line) (bottom). Equation for the double exponential decay fit:  $y = A_1 e^{-t/\tau_1} + A_2 e^{-t/\tau_2}$ ;  $\Delta A_1 = 796.4$   $\Delta A_2 = 28.3$ . Red line shows what a single-exponential decay fit for such distribution would be. Errors represent standard error of the mean. Dwelltime distributions for remaining DNA constructs are shown in Supplementary Fig 3.2c. f) Scatter plot showing the average dwelltime ( $\Delta\tau_{av}$ ) values for the cases of 2 (d2) and 4 (d4) nucleotide separation between the PAM sites. The values are averages of four measurements over four different days for each case. Error bars represent the standard error of the mean. The average dwelltime was obtained using the equation  $\tau_{av} = (A_1 \Delta\tau_1^2 + A_2 \tau_2^2) / (A_1 \Delta\tau_1 + A_2 \Delta\tau_2)$  g) Example trace showing direct FRET transitions for a 5-PAM containing DNA construct with separation between PAM sites being 4 nucleotides. h) Histogram showing the percentage of events that are longer than  $\tau_1$  showing FRET fluctuations for multiple PAM constructs with PAM separation of 4 nucleotides. The values are averages of four measurements over four different days. The error bars represent standard error of the mean.

in  $\tau_2$  was observed as in the case of 2 nucleotide separation (Figure S3.2 d). The values were found to increase from  $2.72 \pm 0.44$  s for a single PAM to  $4.89 \pm 0.70$  s for 5 PAMs (Figure S3.2 d). Average dwelltime analysis also revealed an increase in  $\Delta\tau_{av}$ : from  $1.22 \pm 0.27$  s for a single PAM to  $2.52 \pm 0.37$  s for the construct containing 5 PAMs (Figure 3.2 f). The increased separation between the PAMs results in a slower rise in  $\Delta\tau_{av}$  which further suggests that this increase is due to 1-dimensional diffusion, as increased distance between binding sites lowers the probability of finding a target by diffusing laterally along the DNA.

The increased separation between the PAM sites results in the greater separation of FRET efficiency values corresponding to binding to each individual PAM site. As a result, a type of binding event showing FRET fluctuations was observed in the case of 4-nucleotide separation between the PAMs (Figure 3.2 g). Furthermore, upon further analysis it was found that the percentage of binding events that show fluctuations and are longer than  $\tau_1$  increases with increasing number of neighboring PAM sites from  $1.36 \pm 0.44\%$  to  $3.47 \pm 1.21\%$  (Figure 3.2 h). This data directly shows that Cas9 laterally diffuses between PAM sites, thus providing an explanation for the observed increase in  $\tau_2$  and  $\Delta\tau_{av}$ . The overall low occurrence of such events can be explained by the nature of these interactions. Cas9-PAM interactions are intrinsically weak and 3-dimensional dissociation dominates, leading to a



large population of short-lived binding events characterized by a single FRET efficiency value.



**Figure 3.3 Multiple PAM influence on binding to a target site.** a) DNA and crRNA sequence used for experiments of a single target and multiple PAMs. DNA sequence is shown for the single-target single-PAM construct and schematic representation is shown for multiple-PAM containing constructs. b) FRET histogram of all binding events from the construct containing a single PAM next to the partial target. Y axis represents total number of counts. c) Example FRET traces from single-target multiple-PAM experiments. Top trace shows an example of a single-step binding event and a binding event that starts at a lower FRET state before transitioning to productive binding high-FRET state. Bottom trace shows an example of FRET fluctuations between the high-FRET state and lower FRET states. d) A histogram showing the percentage of dynamic. The percentage values were obtained from five

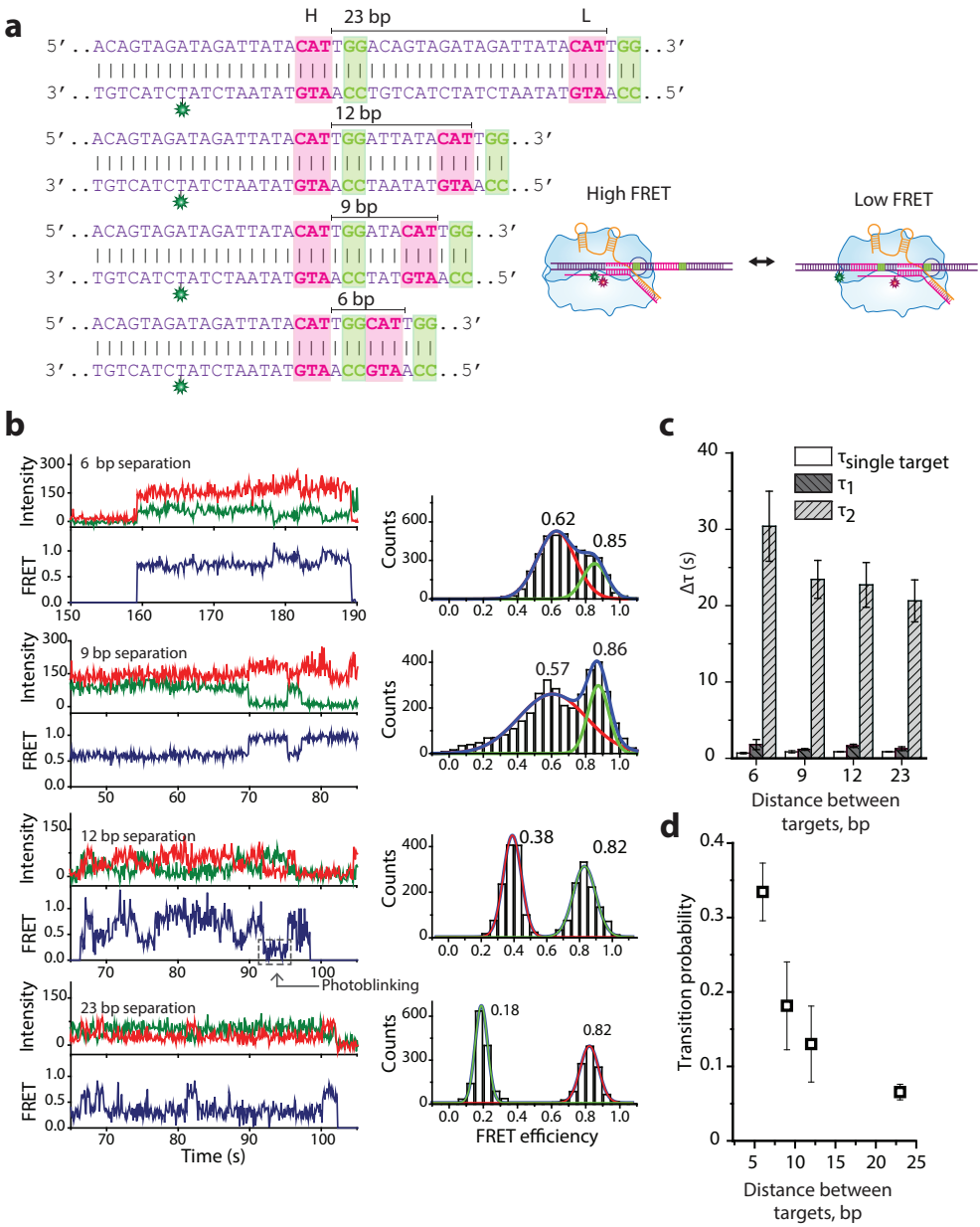
experiments on five different days. Error bars represent standard error of the mean. The percentage of events showing dynamic binding behavior was calculated by dividing the number of binding events that showed such behavior by the total number of binding events. e) A scatter plot showing the time Cas9 spends in a lower FRET state before transitioning to a high-FRET state. The dwelltimes increase with the number of PAMs adjacent to the target. Such dwelltime could not be determined for a construct containing a single PAM. The values were obtained from five experiments on five different days. Error bars represent standard error of the mean.

### 3.2.2. 1-DIMENSIONAL DIFFUSION USED FOR PAM AND TARGET SEARCH

The observation that Cas9 stays bound to a DNA strand for longer when multiple neighboring PAM sequences are present and that the incidence of events showing FRET fluctuations increases with increasing number of PAM sequences suggested that this effect could be caused by lateral diffusion between the PAM sites [26]. To investigate the possibility that Cas9 is able to scan PAM sequences in a 1-D fashion and to explore whether such probing could lead to binding to a neighboring target site, we designed DNA constructs containing a partial target of 9 nucleotides and an increasing number of PAM sites adjacent to it: 1xPAM, 3xPAM, 5xPAM, 7xPAM and 9xPAM (Figure 3.3 a). DNA was labeled at position +13 on the target strand and crRNA at position +10 relative to the first complementary nucleotide, such that high FRET efficiency would only be observed upon productive binding to the partial target site (Figure 3.3 a). Partial complementarity was chosen in order to allow for observation of multiple binding events [23, 27].

Binding to DNA containing a single PAM next to the partial target resulted in single-step events showing a stable expected high FRET efficiency of 0.96 (Figure 3.3 b, c). However, increasing number of adjacent PAM sites next to the target displayed an increasing percentage of binding events that either start at a lower FRET state before transitioning to the productive binding FRET state or show fluctuations between a clearly defined high FRET state and various lower FRET states (Figure 3.3 c, d). In particular, the percentage of events that show either fluctuations or 2-step binding (“dynamic events”) shows a 6-fold increase from  $2.03 \pm 1.61\%$  for the 1xPAM construct to  $13.9 \pm 2.51\%$  for the 9xPAM construct (Figure 3.3 d). Furthermore, the time Cas9 spends in an initial low FRET state before transitioning to high FRET state ( $\Delta\tau_i$ ), which indicates on-target binding, was found to increase with increasing number of PAM sequences adjacent to the target (Figure 3.3 e). Together with the observation of the increasing dwelltime when the number of neighboring PAM sites increases in Figure 3.2 d, these results further suggest that Cas9 not only uses 3-D diffusion alone during target search, but can also find a target site by laterally probing neighboring PAM sites. Therefore, upon failing to form a stable R-loop Cas9 does not necessarily dissociate from the

DNA strand but can go back to scanning the PAM sequences in a 1-dimensional fashion.



**Figure 3.4 Mechanism of lateral diffusion.** a) DNA sequences used in tandem target experiments. Complementary sequences to crRNA are marked in pink boxes and PAMs are marked in green boxes. Distances between the protospacers are indicated above the sequences. An illustration of Cas9 binding to each target site resulting in either high FRET (target H) or low FRET (target L) is shown on the right. b) Example fluorescence traces showing Cas9 transitioning between H and L targets with increasing target distances (right). FRET histograms constructed from events that show fluctuations showing two FRET peaks corresponding to binding to either target. High FRET peak stays the same in each histogram, but lower FRET peak moves to lower FRET values as the distance between targets increases. c) A histogram showing all dwelltimes from single-target controls and tandem-target experiments.  $\tau_1$  is similar for single-target control and tandem target experiments.  $\tau_2$  only appears in tandem target experiments and is over 20-fold higher than  $\tau_1$  or  $\tau_{st}$ . The values represent the mean dwelltimes that were obtained by fitting dwelltime distributions by either single or double exponential decay function from five measurements. Error bars represent standard error of the mean. d) A scatter plot showing the probability of Cas9 transitioning between targets before dissociation. The values are averages from five measurements and error bars represent standard error of the mean. The transition probability ( $p = T / (T+D)$ ) was calculated by summing up all the number of FRET transitions (T) and dividing T by the sum of T and the total number of dissociation events (D).

### 3.2.3. MECHANISM OF LATERAL DIFFUSION

The observations of increasing  $\tau_2$  with increasing number of neighboring PAM sequences together with the observed transitions from lower FRET state to high FRET state when a partial target was present suggest that lateral diffusion may indeed be involved in Cas9 target search. In order to investigate this mechanism more systematically, we designed tandem target DNA constructs, where identical partial targets were placed at different distances: 6, 9, 12, and 23bp (Figure 3.4 a). In this assay, binding to one target (H) would yield a high FRET value and binding to the second target (L) would correspond to a lower value (Figure 3.4 a). We used a partial match (3 nt) between crRNA and DNA target to investigate how the presence of a second target site influences Cas9 binding dynamics and whether it can laterally diffuse between the two target sites. Control experiments with a single target at each distance showed no fluctuations in FRET efficiency (Supplementary Figure 3.3 a).

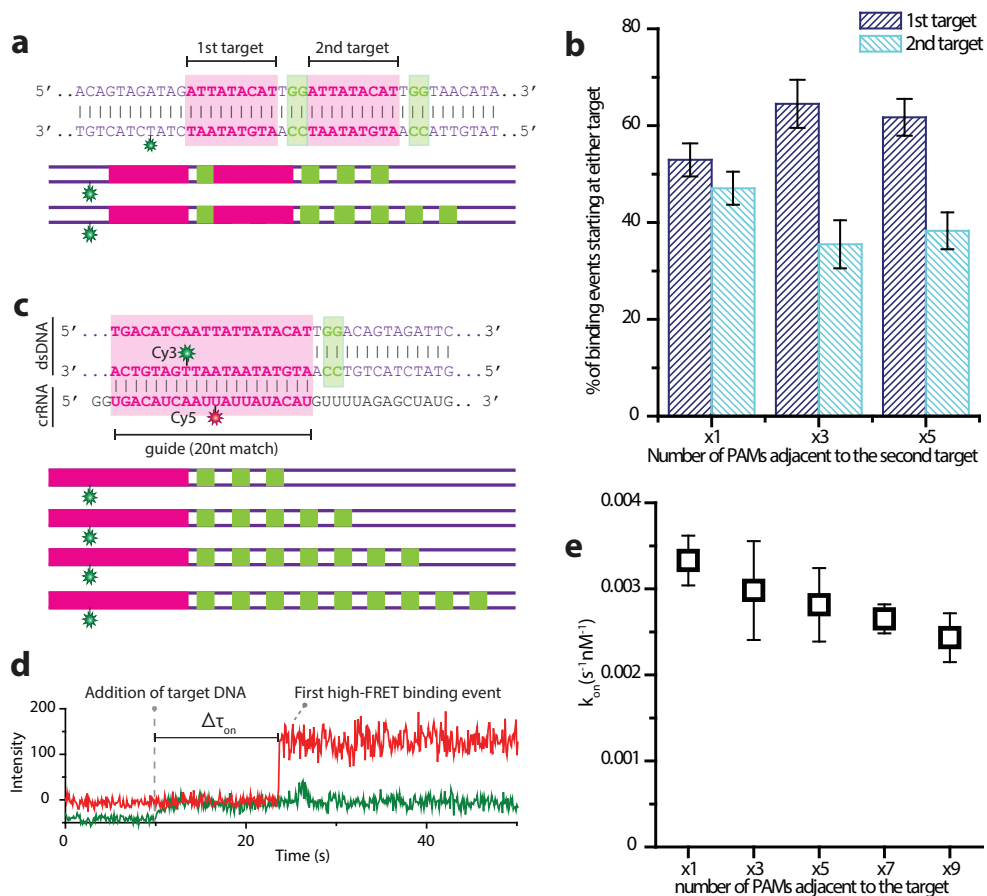
In the tandem target assay, Cas9 was directly observed to switch between two FRET states in a single binding event for each target separation (Figure 3.4 b). The observed two FRET peaks in histograms from transition events agree with the values from single target controls (Figure 3.4 b, Figure S3.3 a), confirming that the fluctuations are arising due to Cas9 shuttling between two target sites. The probability of translocating to a neighboring target before dissociation that arose due to 1-D diffusion was highest at ~0.35 when the distance between protospacers was 6 bp (Figure 3.4 d). At 9-bp and 12-bp separation the probability dropped to ~0.19 and ~0.13, respectively. At 23-

bp distance, the probability dropped to  $\sim 0.07$  – a 5-fold decrease compared to 6-bp separation.

The measured dwelltimes for a single target at each distance from the dye (Figure 3.4 c, Figure S3.3 b) followed a single-exponential decay with dwelltimes ( $\tau_{st}$ ) lower than 1 s which is in agreement with literature [27]. In contrast, the dwelltime distribution for tandem target constructs was characterized by a double exponential decay (Figure 3.4 c, Figure S3.4 a). The dwelltimes for target binding were obtained by measuring the binding times of events that have a FRET value, corresponding to on-target binding, which allowed any non-specific or PAM-only interactions to be excluded from analysis. Short binding events with a similar dwelltime as single target controls were observed with all constructs regardless of the distance between protospacers ( $\tau_1$ ). However, a second type of events observed had strikingly long dwelltimes ( $\tau_2$ ) of 20-30 s (Figure 3.4 c). The presence of such binding events with more than a 20-fold increase in dwelltime compared to a single-target suggests that Cas9 is experiencing a strong synergistic effect when two targets are neighboring each other.

### 3.2.4. PAM MULTIPLICITY DELAYS ON-TARGET BINDING

While the results indicate that lateral diffusion is involved in PAM and consequently target search, it remains unclear whether the presence of multiple PAM sites next to a target would promote on-target binding or act as a decoy binding site, thus delaying target recognition. To investigate the effects PAM multiplicity has on the on-target binding, we designed 3 tandem target constructs. A single PAM was always adjacent to the first target, while the second target had 1, 3 and 5 neighboring PAMs (Figure 3.5 a). As in the previous tandem target experiments (target separation 12 bp), binding to the first target resulted in a high FRET state ( $\sim 0.86$ ) and binding to the second target resulted in a lower FRET state ( $\sim 0.5$ ) (Figure 3.4 b). Complementarity between crRNA and DNA was chosen to be 9 nucleotides for greater binding stability. Analysis of individual binding events revealed that for the symmetric case where both target sites are flanked by a single PAM, the binding events that begin at either target are equally distributed:  $52.9\% \pm 3.4\%$  begin at the first target site and  $47.1\% \pm 3.4\%$  begin at the second target site (Figure 3.5 b). When the number of PAM sites adjacent to the second target site was increased to 3 and consequently 5 PAMs, the distribution of events changed as in both cases more than 60% of events now started at the first target site having a single PAM (Figure 3.5 b). These results suggest that having multiple PAM sites adjacent to the target can deter the protein from binding the target site.



**Figure 3.5 PAM multiplicity delays on-target binding.** a) DNA and crRNA sequence used for experiments with tandem target-multiple PAM constructs. DNA sequence is shown for the symmetric construct with each target having only one adjacent PAM and schematic representation is shown for multiple-PAM adjacent to the second target containing constructs. b) Histogram showing the percentage of binding events starting at either target. The values are averages of five measurements over three different days. Error bars represent standard error of the mean. c) DNA and crRNA sequence used for experiments with a full target and multiple PAMs. DNA sequence is shown for the single-target single-PAM construct and schematic representation is shown for multiple-PAM containing constructs. d) An example trace showing the addition of Cy3-labeled DNA target and the first binding event.  $\Delta\tau_{on}$  is indicated. e) Scatter plot showing the on-target binding rate ( $k_{on-target}$ ) for each construct. Binding rate decreases moderately with increasing number of adjacent PAMs. The values are averages of four measurements over three days. Error bars represent standard error of the mean.  $k_{on-target}$  was calculated using the equation  $k_{on-target} = 1 / (C \times \Delta\tau_{on})$  where C is the molarity of DNA target in the channel.

To further investigate this effect, we designed constructs with full complementarity (20 nt) between DNA and crRNA and an increasing number of neighboring PAM sites: x1, x3, x5, x7 and x9 (Figure 3.5 c). Flow experiments were performed and for each construct, the binding rate to the target (kon-target) was obtained by measuring the time between the addition of DNA to the flow chamber and the first high FRET binding event (Figure 3.5 d). The binding rate values were found to decrease moderately with increasing number of PAM sites (Figure 3.5 e). Once a high FRET state was achieved indicating on-target binding, no further FRET fluctuations were observed. This indicates that although multiple PAM sites cause the binding rate to the target to go down, they cannot compete with target binding once the target has been recognized. When all events, such as zero- or low-FRET peaks were included in the analysis, the binding (kon-total) rate was found to remain constant for all constructs (Figure S3.4 b). Together with the results showing that a target with a single neighboring PAM is preferred in tandem target experiments, this data hints that while PAM multiplicity does not affect overall binding behavior, it delays on-target binding with PAM clusters acting as decoy binding sites for Cas9.

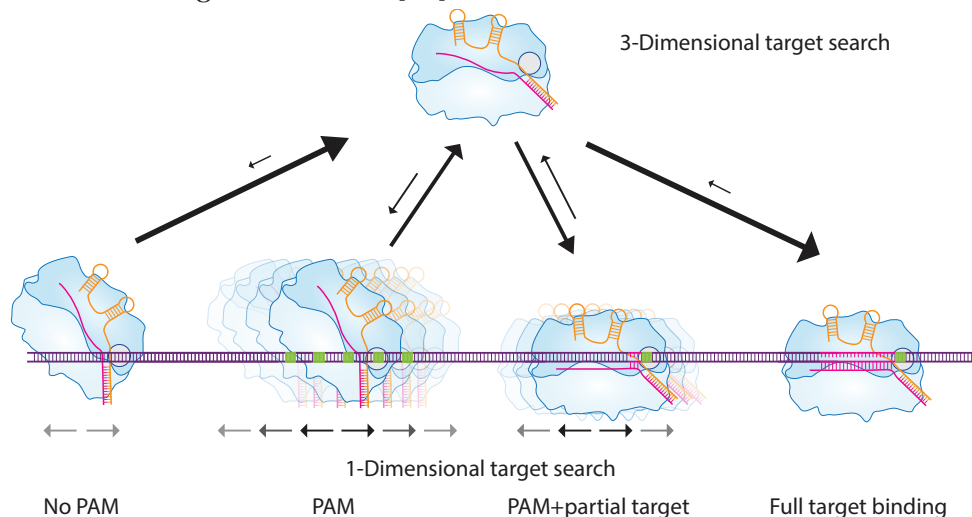
### 3.3. DISCUSSION

As a means of prokaryotic defense against invading foreign genetic elements, Cas9 has to be able to find its target in a crowded cellular environment, among kilo-bases of DNA. The target search becomes even more complicated when Cas9 is applied in eukaryotic cells as a genome engineering tool [15, 18]. In such situations where a protein needs to sample a myriad of sequences before finding a cognate target, facilitated diffusion has been shown to speed up target search as opposed to 3-dimensional diffusion alone [26, 28-33]. We propose that, once Cas9 finds a PAM sequence by 3-D collisions, it is able to diffuse laterally on a DNA strand. By competing with the dissociation process, this lateral diffusion mode intervenes in PAM finding and consequently target recognition. We determined that lateral diffusion of Cas9 primarily occurs in a local manner of ~20 basepairs, when searching for both, the PAM and partial complementarity between DNA and crRNA. This explains the disagreement with previous studies which suggested lateral diffusion does not occur in Cas9 target search, since such distances could not be investigated due to the diffraction limit of other microscopy techniques [20].

We speculate that a limiting factor for Cas9 diffusion may not only be distance, but also the need to open the DNA duplex which is energetically unfavorable if a protein without a helicase domain were to laterally diffuse



long distances. Structural data showed that Cas9 interacts with PAM sites directly, without opening up the double-stranded structure or the involvement of DNA-RNA interactions [34]. Thereby, the lateral diffusion for PAM search would be more effective than for PAM and partial complementarity. This speculation is supported by our observation that multiple neighboring PAM sites provide a binding site for Cas9 and allow Cas9 to interrogate an adjacent target site. This observation is in contrast to rapid dissociation from a PAM when an adjacent target is not present [27]. In addition, we show evidence of Cas9 laterally diffusing between individual PAM sites, further supporting the hypothesis that lateral diffusion is used for PAM search. Our data also provides explanation as to why PAM-rich DNA stretches can be efficiently bound *in vivo* even if no target is present nearby[35]. Our work is also in agreement with DNA curtains studies, which show that Cas9 localizes on PAM-rich regions on  $\lambda$ -DNA[20].



**Figure 3.6 Model for Cas9 target search.** Upon binding to DNA without PAM, Cas9 rapidly dissociates. Upon binding to a PAM Cas9 can dissociate or diffuse locally until another PAM site is found. If a matching DNA sequences flanks the PAM, Cas9 checks for complementarity and if it is not sufficient for stable binding, it can dissociate or diffuse laterally until another PAM is found. Such a process repeats until a target with a high enough degree of complementarity (9-12nt) is found and Cas9 cannot further dissociate.

Based on our findings we propose a model in which PAM sequences drive lateral diffusion as the protein directly interacts with them, as shown by structural studies (Figure 3.6) [34]. If upon binding to a PAM site a matching target is not found, Cas9 can dissociate or diffuse locally on the DNA strand until another PAM site is found. If a matching DNA sequences flanks the PAM, Cas9 checks for complementarity and if it is not sufficient for sta-



3 ble binding, it can dissociate or again, diffuse laterally until another PAM is found, as shown by our tandem target assays. Such a process repeats until a target with a sufficiently high degree of complementarity ( $>12\text{nt}$ ) is found and Cas9 cannot further dissociate. Therefore, we expand the knowledge of Cas9 target search mechanism by showing that it is a combination of 3-dimensional and 1-dimensional diffusion along the DNA strand and that the PAM sequences are not only important as the first step of target recognition but also drive target search.

In order to increase efficiency in genome editing, an in-depth knowledge of how Cas9 finds its target is crucial. Therefore, our results will be important when choosing target sites during genome editing. Neighboring PAM sites close to a desired target will cause the protein to stay bound to the PAM cluster for longer, thereby delaying on-target binding in a case where Cas9 expression levels are high and therefore the likelihood of finding a target by 3-D collisions is high, thus greatly decreasing total search time [19]. If Cas9 expression levels are low, it is likely that by keeping Cas9 bound to a neighboring region for longer, PAM-rich sites could increase the chance of a Cas9 molecule finding the target faster via 1-D diffusion. If no target is present next to a PAM-rich DNA site, such PAM clusters could be used as decoy binding sites by phages in order to prevent Cas9 binding to a cognate target during Cas9 DNA interference. In addition to importance in genetic engineering, our results suggest that the strong interaction and lateral diffusion between PAM sites could be important in bacterial defense against phages. Cas9 has been shown to be important in recognizing the PAM during the CRISPR adaptation step, together with Cas1-Cas2-Csn2 complex [36]. Therefore, PAM density in the invader's genome could potentially play a role in selecting which targets will be integrated in the CRISPR locus. Further in vivo studies will provide an answer to whether PAM clusters are beneficial for the invader, by acting as decoys and delaying target recognition, or for the host, by increasing the efficiency of functional spacer selection during CRISPR adaptation.

### 3.4. MATERIALS AND METHODS

#### 3.4.1. RECOMBINANT SpCas9 PURIFICATION

The pET plasmid encoding (6x)His-tagged Cas9 was transformed into BL21 (DE3), Rosetta. Transformed bacterial cells were moved to a 400ml of fresh LB medium containing 50ug/ml kanamycin. Incubate the culture with shaking (200rpm) at 18°C for 24 hours. Optical density was monitored and Cas9 protein expression was induced ( $A_{550}=0.6$ ) by using 0.5mM IPTG at 18°C for 24hours. After the cells were harvested by centrifugation (5000xg) for 10minutes (at 4°C), bacterial cells were resuspended with lysis buffer [20 mM Tris-HCl (pH 8.0), 400mM NaCl, 10mM b-mercaptoethanol, 1% Triton X-100, 50mg aprotinin, 50mg antipain, 50mg bestatin, 1mM PMSF (phenyl-methylsulfonyl fluoride)] (Sigma-Aldrich) and sonicated on ice. The lysate was centrifuged at 6000 rcf for 10min(4°C) and supernatant solution was mixed with 2ml of Ni-NTA slurry (Qiagen) at 4°C for 1 and half hour. The lysate/Ni-NTA mixture was loaded onto a column (Biorad) with capped bottom outlet. Loaded sample was washed multiple times with pre-made wash buffer [20 mM Tris-HCl (pH 8.0), 400mM NaCl, 10mM b-mercaptoethanol] and (6x)His-tagged SpCas9 was eluted with Elution Buffer [20 mM Tris-HCl (pH 8.0), 400mM NaCl, 10mM b-mercaptoethanol, 200mM Imidazole] (Figure EV1a). Finally, buffer containing eluted SpCas9 protein was changed to storage buffer [10mM HEPES-KOH (pH 7.5), 250mM KCl, 1mM MgCl<sub>2</sub>, 0.1mM EDTA, 7mM b-mercaptoethanol and 20% glycerol] by using centrifugal filter (Amicon Ultra 100K). The purified SpCas9 protein was frozen with liquid nitrogen and stored at -80°C.

#### 3.4.2. BIOTINYLATION OF THE RECOMBINANT SpCas9

The process of linking biotin to the recombinant protein was carried out in-vitro and proceeded during the process of protein purification. After loading the SpCas9 over-expressed bacterial lysate and Ni-NTA mixture onto a column (Biorad), mixed sample was washed multiple times with wash buffer [20 mM Tris-HCl (pH 8.0), 400mM NaCl]. Then we added 10-fold molar excess of maleimide-biotin (Sigma-Aldrich) to SpCas9 solution and incubate for overnight at 4°C (mix gently with rotator). To get rid of unbound maleimide-biotin chemicals, mixed sample was washed sufficiently with wash buffer [20 mM Tris-HCl (pH 8.0), 400mM NaCl]. Finally, biotinylated SpCas9 protein was eluted with elution Buffer [20 mM Tris-HCl (pH 8.0), 400mM NaCl, 200mM Imidazole], then the protein concentration was measured by spectrophotometer (Nanodrop 2000, Thermo Fisher Scientific). Eluted Sp-Cas9 protein was further purified with size exclusion chromatography. The

biotinylation degree of the wild-type SpCas9 protein was calculated with commercial kit (Pierce) and it reached about 100% for two Cysteine sites (Cys80/Cys574). Biotinylated SpCas9 protein was stored in storage buffer [10mM HEPES-KOH (pH 7.5), 250mM KCl, 1mM MgCl<sub>2</sub>, 0.1mM EDTA, 7mM b-mercaptoethanol and 20% glycerol] and purified protein was frozen in liquid nitrogen and store at -80°C.

### 3.4.3. PREPARATION OF THE SINGLE-GUIDE RNA

We used In-vitro RNA transcription (DNA template, T7 RNA polymerase (NEB) 5ul, 10x Buffer 10ul, rNTP mix (2.5mM each), MgCl<sub>2</sub> 10mM, DTT 1mM, H<sub>2</sub>O up to 100ul, RNase inhibitor (NEB) 0.5ul, Total 100ul reaction) to generate single-guide RNAs. DNA template contains X20 target protospacer sequence which is complementary to the RNA strand. After RNA transcription, DNA template was removed by DNase (NEB) treatment. Then pure single-guide RNA was purified and concentration was measured by spectrophotometer (Nanodrop 2000, Thermo Fisher Scientific).

### 3.4.4. IN-VITRO DNA CLEAVAGE ASSAY WITH WILD-TYPE AND BIOTINYLATED SP-CAS9

In-vitro cleavage experiments were performed with sgRNA and SpCas9 proteins purified at high purity (Figure EV2). DNA containing the target site was prepared by PCR, and higher molar concentration of the SpCas9 protein and biotinylated SpCas9 was treated at the same molarity. The sgRNA was added at a molar ratio three times greater than the protein (final molar ratio, DNA: protein: sgRNA = 1:3 :9) with a complementary sequence to the target site. Target DNA, SpCas9 protein and sgRNA were mixed and incubated at 37°C for 1hour. The cleaved DNA product was separated on the 1.5% agarose gel and cleavage ratio was calculated by ImageJ software.

### 3.4.5. LABELING OF NUCLEIC ACIDS

Nucleic acids were labeled using NHS-ester chemistry. DNA and RNA strands with a C6 amine modification on a Thymine or Uracil base were ordered synthetic from companies Ella biotech and IBA Lifesciences respectively. 1 mM DNA or RNA samples were mixed with ~20 mM dye (GE healthcare) and labeling buffer (Sodium bicarbonate, 8.4 mg/ml) in a volume ratio 1:1:5 and incubated for 6 hours at room temperature with gentle mixing in

the dark. Full labeling procedure can be found in [37].

### 3.4.6. SINGLE-MOLECULE TWO-COLOR FRET

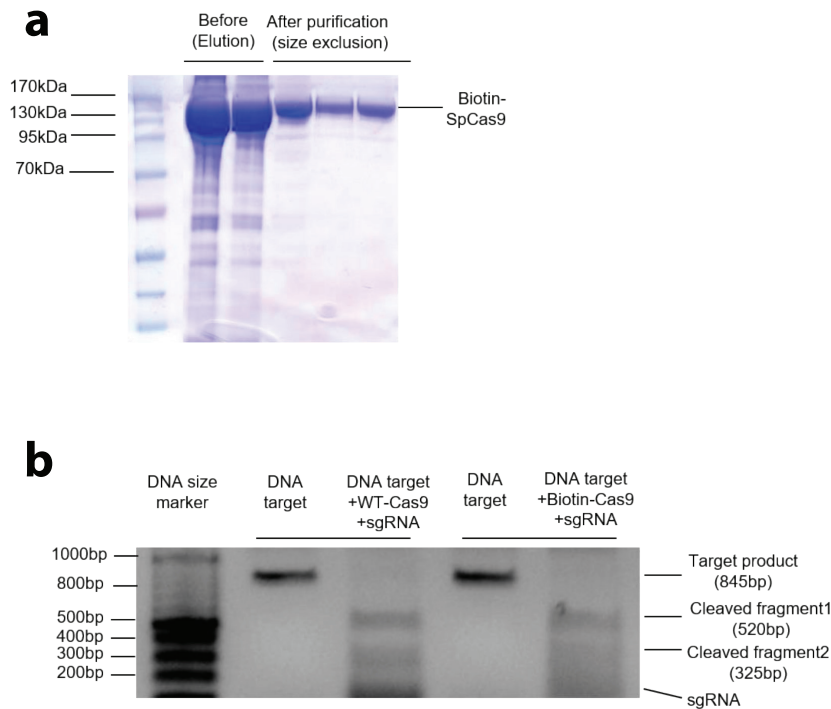
Single-molecule fluorescence measurements were performed with a prism-type total internal reflection fluorescence microscope. 0.1mg/ml Streptavidin was added to a polyethylene glycol-coated quartz surface and incubated for 2 minutes before being washed with T50 (10 mM Tris-HCl (pH 8.0), 50 mM NaCl). Biotinylated Cas9 was pre-incubated with Cy5-labeled crRNA and tracrRNA (ratio 1:2:4) at 37 degrees for 20 minutes in NEB buffer 3 (100 mM NaCl, 50 mM Tris-HCl, 10 mM MgCl<sub>2</sub>, 1 mM DTT) and then added to the chamber containing Streptavidin. After 2 minute incubation unbound Cas9 and RNA molecules were washed away with an imaging buffer (50mM HEPES-NaOH [pH7.5], 10mM NaCl, 2mM MgCl<sub>2</sub>, 1% glucose (Dextrose monohydrate), 1mM Trolox (2.5mg/10ml), 1mg/ml glucose oxydase [Sigma], 170ug/ml catalase [Merck]). 8 nM Cy3 labeled DNA substrate in imaging buffer was added to the channel. A reference video of immobilized Cy5-labeled Cas9:RNA complexes were made. Following the reference video, Cy3-labeled DNA molecules were excited using a 532 nm diode laser. Fluorescence signals of Cy3 and Cy5 were collected through a 60× water immersion objective (UplanSApo, Olympus) with an inverted microscope (IX73, Olympus). The 532 nm laser scattering was blocked out by a 532 nm long pass filter (LPD01-532RU-25, Semrock). The Cy3 and Cy5 signals were separated with a dichroic mirror (635 dcxr, Chroma) and imaged using an EM-CCD camera (iXon Ultra, DU-897U-CS0-#BV, Andor Technology). RNA and DNA sequences used can be found in Appendix Tables S1 and S2.

### 3.4.7. DATA ACQUISITION AND ANALYSIS

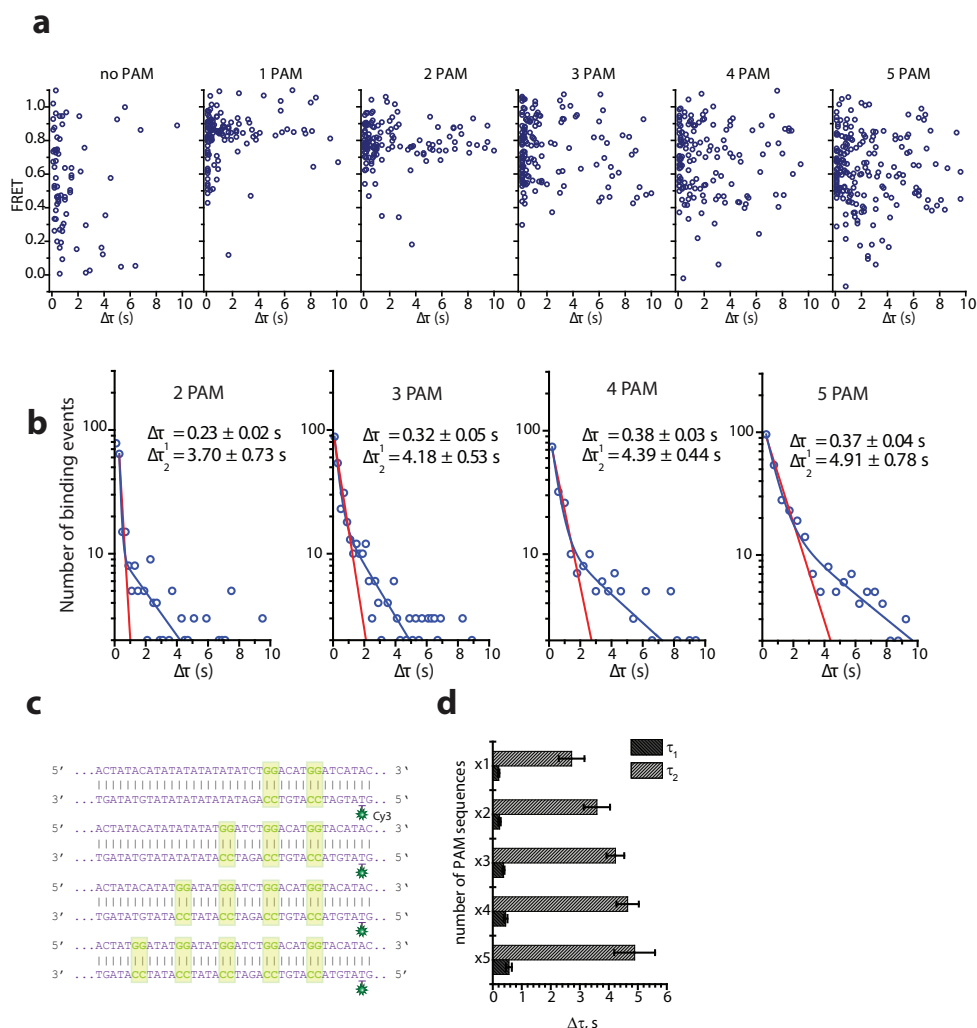
Using a custom-made program written in Visual C++ (Microsoft), a series of CCD images of time resolution 0.1s was recorded. The time traces were extracted from the CCD image series using IDL (ITT Visual Information Solution) employing an algorithm that looked for fluorescence spots with a defined Gaussian profile and with signals above the average of the background signals. Colocalization between Cy3 and Cy5 signals was carried out with a custom-made mapping algorithm written in IDL. The extracted time traces were processed using Matlab (MathWorks) and Origin (Origin Lab).

3.5. SUPPLEMENTARY INFORMATION

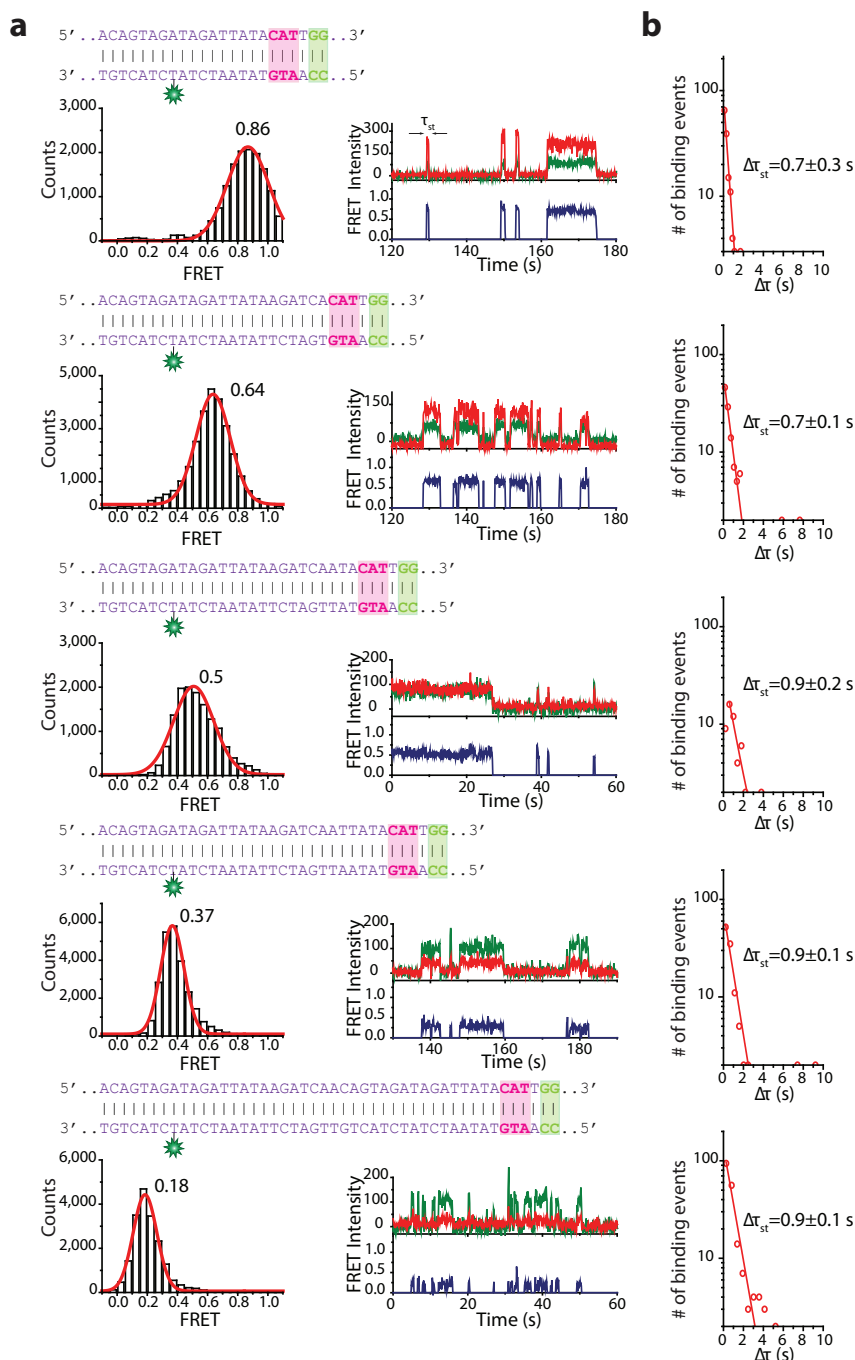
3.5.1. SUPPLEMENTARY FIGURES



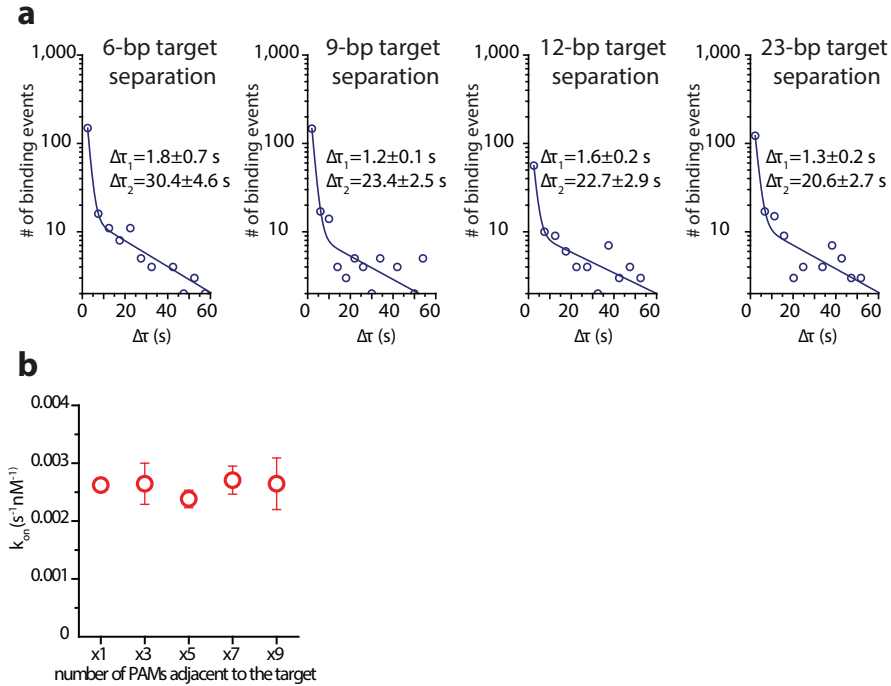
**Figure S3.1. Biotinylation and activity test of the SpCas9 protein.** a) SDS-PAGE data of the biotinylated SpCas9 protein. b) In-vitro cleavage assay for activity validation of the biotinylated Sp-Cas9. Target DNA strand was cleaved by equivalent molar ratio of the SpCas9 and biotinylated Sp-Cas9. Cleaved fragments sizes are 520bp and 325bp respectively. Primers used: CCR5 forward primer 5'-GAGGTGAGAGGATTGCTTGAGCCC-3', CCR5 reverse primer 5'-ATTGTCAGGAG-GATGATGAAGAAG-3'.



**Figure S3.2. Evidence of synergistic effect and lateral diffusion between PAMs** a) A scatter plot showing FRET vs dwelltime for each construct containing 0, 1, 2, 3, 4, 5 PAMs. No correlation between FRET values and dwelltimes is observed for either construct b) Dwelltime distributions for constructs containing 2, 3, 4 and 5 neighboring PAMs. Blue line shows a double-exponential decay fit and red line shows what a single-exponential decay fit would look like for these dwelltime distributions. Equation of the double-exponential decay fit:  $y = A_1 e^{-t/\tau_1} + A_2 e^{-t/\tau_2}$ . c) DNA sequences used in the multiple PAM experiments where the PAM separation is 4 nucleotides. d) Bar plot of all dwelltime values for all constructs (d4). The times shown are mean values of dwelltimes obtained during four different experiments on four different days. Error bars represent standard error of the mean.



**Figure S3.3. Single-target control experiments** a) DNA sequences used in single-target control experiments, complementary sequence is marked in pink box and PAMs are marked in green box. Corresponding FRET histograms and example FRET traces are shown below each DNA sequence. b) Dwelltime distributions for each DNA construct fitted with a single-exponential decay function.



**Figure S3.4. Tandem target dwelltime and single target binding rate** a) Dwelltime histograms for tandem-target construct for each target separation, fitted with a double-exponential decay function. Equation of the fit:  $y = A_1 e^{-t/\tau_1} + A_2 e^{-t/\tau_2}$ . A values for each tandem target construct: d6 -  $A_1 = 640.3$ ,  $A_2 = 8.0$ ; d9 -  $A_1 = 654.7$ ,  $A_2 = 8.8$ ; d12 -  $A_1 = 144.6$ ,  $A_2 = 6.4$ ; d23 -  $A_1 = 635.3$ ,  $A_2 = 5.6$ . Average dwelltime for each construct: d6 -  $\tau_{av} = 12.3$  s; d9 -  $\tau_{av} = 6.9$  s; d12 -  $\tau_{av} = 10.7$  s; d23 -  $\tau_{av} = 7.6$  s. Average dwelltimes were calculated using equation  $\tau_{av} = (A_1 \Delta\tau_1^2 + A_2 \Delta\tau_2^2) / (A_1 \Delta\tau_1 + A_2 \Delta\tau_2)$ . Scatter plot showing the total binding rate ( $k_{on}$ ) for each single-target, multiple PAM construct. Binding rate stays constant with increasing number of adjacent PAMs. The values are averages of four measurements over three days. Error bars represent standard error of the mean.  $k_{on}$  was calculated using the equation  $k_{on} = 1 / (C \Delta\tau_{on})$  where C is the molarity of DNA target in the channel.



## 3.5.2. SUPPLEMENTARY TABLES

sgRNA (bulk cleavage)	5'GGU GAC AUC AAU UAU UAU ACA UGU UUU AGA GCU AGA AAU AGC AAG UUA AAA UAA GGC UAG UCC GUU AUC AAC UUG AAA AAG UGG CAC CGA GUC GGU GCU UUU U 3'
tracrRNA (smFRET)	5'GGA ACC AUU CAA AAC AGC AUA GCA AGU UAA AAU AAG GCU AGU CCG UUA UCA ACU UGA AAA AGU GGC ACC GAG UCG GUG CUU UUU UU 3'
crRNA_+0 mPAM (smFRET)	5'GGU UUU UUU UUU UUU UUU UUU UGU <u>UUU</u> AGA GCU AUG CUG UUU UG 3'
crRNA_+9 mPAM+target sm(FRET)	5'GGU UUU UUU UUU <u>A</u> AU UAU ACA UGU UUU AGA GCU AUG CUG UUU UG 3'
crRNA+3 tandem target (smFRET)	5'GGU UUU UUU UUU <u>UUU</u> UUU UCA UGU UUU AGA GCU AUG CUG UUU UG 3'
crRNA_+20 mPAM+target flow (smFRET)	5'GGU GAC AUC AAU <u>UAU</u> UAU ACA UGU UUU AGA GCU AUG CUG UUU UG 3'

**Table S3.1. RNA sequences.** Red underlined nucleotide shows labeling position

mPAMx0 non-target strand	5' CAG AAT AGT AAT CTA ACA GTG TCA AGT ATA ATC TAT AAC ATC ATA ACT ATA CTA TAC ATA TAT ATA TAT ATC TAT ACA TAT ATC ATA CTA GCA TCG TT 3'
mPAMx0 target strand	5' AAC GAT GCT AGT ATG ATG ATA TAT GTA TAG ATA TAT ATA TAT ATG TAT AGT ATA GTT ATG ATG TTA TAG ATT ATA CTT GAC ACT GTT AGAT TAC TAT TCT G 3'
mPAMx1 non-target	5' CAG AAT AGT AAT CTA ACA GTG TCA AGT ATA ATC TAT AAC ATC ATA ACT ATA CTA TAC ATA TAT ATA TAT ATC TAT ACA TGG ATC ATA CTA GCA TCG TT 3'
mPAMx1 target	5' AAC GAT GCT AGT ATG ATC CAT GTA TAG ATA TAT ATA TAT ATG TAT AGT ATA GTT ATG ATG TTA TAG ATT ATA CTT GAC ACT GTT AGA TTA CTA TTC TG 3'
mPAMx2 non-target	5' CAG AAT AGT AAT CTA ACA GTG TCA AGT ATA ATC TAT AAC ATC ATA ACT ATA CTA TAC ATA TAT ATA TAT ATC TAT GGA TGG ATC ATA CTA GCA TCG TT 3'
mPAMx2 target	5' AAC GAT GCT AGT ATG ATC CAT CCA TAG ATA TAT ATA TAT ATG TAT AGT ATA GTT ATG ATG TTA TAG ATT ATA CTT GAC ACT GTT AGA TTA CTA TTC TG 3'
mPAMx3 non-target	5' CAG AAT AGT AAT CTA ACA GTG TCA AGT ATA ATC TAT AAC ATC ATA ACT ATA CTA TAC ATA TAT ATA TAT ATG GAT GGA TGG TAC ATA CTA GCA TCG TT 3'
mPAMx3 target	5' AAC GAT GCT AGT ATG TAC CAT CCA TCC ATA TAT ATA TAT AGT ATA GTA TAG TTA TGA TGT TAT AGA TTA TAC TTG ACA CTG TTA GAT TAC TAT TCT G 3'
mPAMx4 non-target	5' CAG AAT AGT AAT CTA ACA GTG TCA AGT ATA ATC TAT AAC ATC ATA ACT ATA CTA TAC ATA TAT ATA TGG ATG GAT GGA TGG TAC ATA CTA GCA TCG TT 3'
mPAMx4 target	5' AAC GAT GCT AGT ATG TAC CAT CCA TCC ATC CAT ATA TAT ATG TAT AGT ATA GTT ATG ATG TTA TAG ATT ATA CTT GAC ACT GTT AGA TTA CTA TTC TG 3'
mPAMx5 non-target	5' CAG AAT AGT AAT CTA ACA GTG TCA AGT ATA ATC TAT AAC ATC ATA ACT ATA CTA TAC ATA TAT GGA TGG ATG GAT GGA TGG TAC ATA CTA GCA TCG TT 3'
mPAMx5 target	5' AAC GAT GCT AGT ATG TAC CAT CCA TCC ATC CAT CCA TAT ATG TAT AGT ATA GTT ATG ATG TTA TAG ATT ATA CTT GAC ACT GTT AGA TTA CTA TTC TG 3'
1xPAM+9 Non-target	5' GTCAAGTATAATCTATAACATCATAACTAT ACAGTAGATAGATTATACAT TGG ACAGTAGATAGTAATATAGA TCATAACATATTACATACAGACATGTTA 3'
1xPAM+9 target	5' TAACATGTCTGTATGTAATATGTTATGAT CTATA TTAATCTACTGTCC A ATTGTATAATCTACTACTGT ATAGTTATGATGTTATAGATTATACTTGAC ' 3
3xPAM+9 Non-target	5' GTCAAGTATAATCTATAACATCATAACTAT ACAGTAGATAGATTATACAT TGG ACGGTAGGTAGTA ATATAGATCATAACATATTACATAC AGACATGTTA 3'
3xPAM+9 Target	5' TAACATGTCTGTATGTAATATGTT ATGATC TATATT ACTACCTACCGT CCA ATTGTATAATCTACT ACTG TATAGTTATGA TGTTATAGATTATACTTGAC ' 3
5xPAM+9 Non-target	5' GTCAAGTATAATCTATAACATCATAACTAT ACAGTAGATAGATTATACAT TGG ACGGTAGGTAGGAA GGATAGATCATAACATATTACATACAGACATGTTA 3'
5xPAM+9 Target	5' TAACATGTCTGTATGTAATATGTTATGATC TACCTTCCCTACCGT CCA ATTGTATAATCTACTACTGT ATAGTTATGATGTTATAGATTATACTTGAC ' 3

7xPAM+9 Non-target	5'GTCAAGTATAATCTATAACATCATAACTATACAGTAGATAG <b>ATTATACATTGG</b> AC <b>GGTAGG</b> TAG <b>GA</b> <b>AGGTAGGTCGGA</b> ACATATTACATACAGACATGTTA 3'
7xPAM+9 Target	5' TAACATGCTCTGTATGTAATATGTT <b>CCGACC</b> TAC <b>CTTCTACCTTCCGTCCA</b> <b>ATGTATAAT</b> CTA <b>TCT</b> ACTGTATAGTTATGATGTTA TAGATTATACCTTGAC 3'
9xPAM+9 Non-target	5'GTCAAGTATAATCTATAACA TCATAACTATACAGTAGATAG <b>ATTATACATTGG</b> AC <b>GGTAGG</b> TAG <b>GA</b> <b>AGGTAGGTCGGAAGGTAGG</b> ACATACAGACATGTTA 3'
9xPAM+9 target	5'TAACATGCTCTG TATGT <b>CCTACCTTCCGA</b> <b>CCTACCTTCTACCTTCCGTCCA</b> <b>ATGTATAAT</b> CTA <b>TCT</b> ACTGTATAG TTATGATGTTATAGATTATACCTTGAC 3'
Tandem target d_23 non-target	5' GTCAAGTATAATCTATAACATCATAACTAT ACAGTAGATAGATTATAC <b>CAT</b> <b>TGG</b> ACAGTAGATAGATTATAC <b>CAT</b> <b>TGG</b> TAACATATTACATACAGACATGTT 3'
Tandem target d_23 target	5' AACATGCTCTGTATGTAATATGTTA <b>CCA</b> <b>ATGTATAATCTTCTACTGT</b> <b>CCA</b> <b>ATGTATAATCTA</b> <b>TCT</b> ACTGT ATAGTTATGATGTTATAGATTATACTTGAC '3
Tandem target d_12 non-target	5' GTCAAGTATAATCTATAACATCATAACTAT ACAGTAGATAGATTATAC <b>CAT</b> <b>TGG</b> ATTATAC <b>CAT</b> <b>TGG</b> TAACATATTACATACAGACATGTTA 3'
Tandem target d_12 target	5' TAACATGCTCTGTATGTAATATGTTA <b>CCA</b> <b>ATGTATAAT</b> <b>CCA</b> <b>ATGTATAATCTA</b> <b>TCT</b> ACTGT ATAGTTATGATGTTATAGATTATACTTGAC 3'
Tandem target d_9 non-target	5'GTCAAGTATAATCTATAACATCATAACTAT ACAGTAGATAGATTATAC <b>CAT</b> <b>TGG</b> ATA <b>CAT</b> <b>TGG</b> TAACATATTACATACAGACATGTTA 3'
Tandem target d_9 target	5'TAACATGCTCTGTATGTAATATGTTA <b>CCA</b> <b>ATGTAT</b> <b>CCA</b> <b>ATGTATAATCTA</b> <b>TCT</b> ACTGT ATAGTTATGATGTTATAGATTATACTTGAC 3'
Tandem target d_3 non-target	5'GTCAAGTATAATCTATAACATCATAACTAT ACAGTAGATAGATTATAC <b>CAT</b> <b>TGG</b> <b>CAT</b> <b>TGG</b> TAACATATTACATACAGACATGTTA 3'
Tandem target d_3 target	5'TAACATGCTCTGTATGTAATATGTTA <b>CCA</b> <b>ATG</b> <b>CCA</b> <b>ATGTATAATCTA</b> <b>TCT</b> ACTGT ATAGTTATGATGTTATAGATTATACTTGAC 3'
Single target d_23 non-target	5' GTCAAGTATAATCTATAACATCATAACTAT ACAGTAGATAGTAATATAGA TCA ACAGTAGATAGATTATAC <b>CAT</b> <b>TGG</b> TAACATATTACATACAGACATGTTA 3'
Single target d_23 target	5' TAACATGCTCTGTATGTAATATGTTA <b>CCA</b> <b>ATGTATAATCTTCTACTGT</b> TGA TCTATATTACTA <b>TCT</b> ACTGT ATAGTTATGATGTTATAGATTATACTTGAC 3'
Single target d_12 non-target	5' GTCAAGTATAATCTATAACATCATAACTAT ACAGTAGATAGTAATATAGA TCA ATTATAC <b>CAT</b> <b>TGG</b> TAACATATTACATACAGACATGTTA 3'
Single target d_12 target	5' TAACATGCTCTGTATGTAATATGTTA <b>CCA</b> <b>ATGTATAAT</b> TGA TCTATATTACTA <b>TCT</b> ACTGT ATAGTTATGATGTTATAGATTATACTTGAC 3'
Single target d_9 non-target	5' GTCAAGTATAATCTATAACATCATAACTAT ACAGTAGATAGTAATATAGA TCA ATA <b>CAT</b> <b>TGG</b> TAACATATTACATACAGACATGTTA 3'
Single target d_9 target	5' TAACATGCTCTGTATGTAATATGTTA <b>CCA</b> <b>ATG</b> TATTGA TCTATATTACTA <b>TCT</b> ACTGTATAGTTATGATGTTATAGATTATACTTGAC 3'

Single target d_6 non-target	5' GTCAGTATAATCTATAACATCATAACTATACAGTAGATAGTAATATAGATCA <u>CAT</u> <u>TGG</u> TAACATATTACATACAGACATGTTA 3'
Single target d_6 target	5' TAACATGTCTGTATGTAATATGTTA <u>CCA</u> <u>ATG</u> TGATCTATATTACTA <u>T</u> CTACTGT ATAGTTATGATGTTATAGATTATACTTGAC 3'
Tandem target d_12 PAMx3 Non-target	5' GTCAGTATAATCTATAACATCATAACTAT ACAGTAGATAG <u>ATTATACAT</u> <u>TGG</u> <u>ATTATACAT</u> <u>TGG</u> TAGGAT <u>GG</u> TACATACAGACATGTTA 3'
Tandem target d_12 PAMx3 target	5' TAACATGTCTGTATGTA <u>CCATCCTA</u> <u>CCA</u> <u>ATGTATAAT</u> <u>CCA</u> <u>ATGTATAAT</u> CTA <u>T</u> CTACTGT ATAGTTATGATGTTATAGATTATACTTGAC 3'
Tandem target d_12 PAMx5 Non-target	5' GTCAGTATAATCTATAACATCATAACTAT ACAGTAGATAG <u>ATTATACAT</u> <u>TGG</u> <u>ATTATACAT</u> <u>TGG</u> TAGGAT <u>GGTAGGGTAGGG</u> ACATGTTA 3'
Tandem target d_12 PAMx5 target	5' TAACATGTCT <u>CCTACCTACC</u> ATCCTA <u>CCA</u> <u>ATGTATAAT</u> <u>CCA</u> <u>ATGTATAAT</u> CTA <u>T</u> CTACTGT ATAGTTATGATGTTATAGATTATACTTGAC 3'
1xPAM+ 20 Non-target	5' GTCAGTATAATCTATAACATCATAACTAT <u>TGACATCAATTATTATACAT</u> <u>TGG</u> ACAGTAGATAGTAATATAGATCAT AACATATTACATACAGACATGTTA 3'
1xPAM+ 20 target	5' TAACATGTCTGTATGTAATATGT TATGATCTATATTACTATCTACTGT <u>CCA</u> <u>ATGTATAATAATTGATGTCA</u> ATAGTTATGATGTTATAGATTATACTTGAC 3'
3xPAM+ 20 Non-target	5' GTCAGTATAATCTATAACATCATAACTAT <u>TGACATCAATTATTATACAT</u> <u>TGG</u> AC <u>GGTAGGGTAG</u> TAATATAGATC ATAACATATTACATACAGACATGTTA 3'
3xPAM+ 20 target	5' TAACATGTCTGTATGTAATATGTTATG ATCTATATTACTA <u>CCTACCGT</u> <u>CCA</u> <u>ATGTATAATAATTGATGTCA</u> ATAGTTATGATGTTATAGATTATACTTGAC 3'
5xPAM+ 20 Non-target	5' GTCAGTATAATCTATAACATCATAACTAT <u>TGACATCAATTATTATACAT</u> <u>TGG</u> AC <u>GGTAGGGTAGGAA</u> <u>GGTAG</u> AT CATAACATATTACATACAGACATGTTA 3'
5xPAM+ 20 target	5' TAACATGTCTGTATGTAATATGTTATGATCTA <u>CCCTC</u> <u>CTACCTACCGT</u> <u>CCA</u> <u>ATGTATAATAATTGATGTCA</u> ATAGTTATGATGTTATAGATTATACTTGAC 3'
7xPAM+ 20Non- target	5' GTCAGTATAATCTATAACATCATAACTAT <u>TGACATCAATTATTATACAT</u> <u>TGG</u> AC <u>GGTAGGGTAGGAA</u> <u>GGTAGGTCGG</u> AA CATATTACATACAGACATGTTA 3'
7xPAM+ 20 target	5' TAACATGTCTGTATGTAATATGTT <u>CCGACCTAC</u> <u>CTTCTACCTACCGT</u> <u>CCA</u> <u>ATGTATAATAATTGATGTCA</u> ATAGTTATGATGTTATAGATTATACTTGAC 3'
9xPAM+ 20 Non-target	5' GTCAGTATAATCTATAACATCATAACTAT <u>TGACATCAATTATTATACAT</u> <u>TGG</u> AC <u>GGTAGGGTAGGAA</u> <u>GGTAGGTCGGA</u> <u>AGGTAGG</u> ACATACAGACATGTTA 3'
9xPAM+ 20 target	5' TAACATGTCTGTATGT <u>CCTACCTTCCGAC</u> <u>CTACCTTCTACCTACCGT</u> <u>CCA</u> <u>ATGTATAATAATTGATGTCA</u> ATAGTTATGATGTTATAGATTATACTTGAC 3'
mPAMx2 d4 target	5' AACGATGCTA <u>G</u> TATGAT <u>CCATGTCC</u> AGATATATATATATATGTATAGT ATAGTTATGATGTTATAGATTATACTTGA CACTGTTAGATTACTATTCTG 3'
mPAMx2 d4 Non-target	5' CAGAATAGTAATCTAACAGTGTCAGTATAATCTATAACATCATAACTAT ACTATACATATATATATATATCT <u>GGACATGG</u> ATCATAC TAGCATCGTT 3'
mPAMx3 d4 target	5' AACGATGCTA <u>G</u> TATGT <u>ACCATGTCC</u> AGAT <u>CC</u> ATATATATATGTATAGT ATAGTTATGATGTTATAGATTATA CTTGACACTGTTAGATTACTATTCTG 3'

mPAMx3 d4 Non-target	5' CAGAATAGTAATCTAACAGTGTCAAGTAT AATCTATAACATCATAACTAT ACTATACATATATATATATGGATCTGGACATGGTACATAC TAGCATCGTT 3'
mPAMx4 d4 target	5' AACGATGCTA GTATGTACCATGTCCAGATCCATATCCATATGTATAGT ATAGTTATGATGTTATAGATT ATACTTGACACTGTTAGATTACTATTCTG 3'
mPAMx4 d4 Non-target	5' CAGAATAGTAATCTAACAGTG TCAAGTATAATCTATAACATCATAACTAT ACTATACATATGGATATGGATCTGGACATGGTACATAC TAGCATCGTT 3'
mPAMx5 d4 target	5' AACGATGCTA GTATGTACCATGTCCAGATCCATATCCATATCCATAGT ATAGTTATGATGTTAT AGATTATACTTGACACTGTTAGATTACTATTCTG 3'
mPAMx5 d4 Non-target	5' CAGAATAGTAATCTAACAGTGTCAAGTATAATCTATAACAT CATAACTAT ACTATGGATATGGATATGGATCTGGACATGGTACATAC TAGCATCGTT 3'

**Table S3.2. DNA sequences used in single-molecule experiments.** Sequences in green represent PAMs (GG and CC). Sequences in Pink show regions complementary to guide RNA. Red underlined nucleotide indicates the nucleotide with C6 amine modification for dyelabeling.

## REFERENCES

1. Barrangou, R., et al., *CRISPR provides acquired resistance against viruses in prokaryotes*. Science, 2007. **315**(5819): p. 1709-12.
2. Marraffini, L.A., *CRISPR-Cas immunity in prokaryotes*. Nature, 2015. **526**(7571): p. 55-61.
3. Makarova, K.S., et al., *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*. Biol Direct, 2006. **1**: p. 7.
4. Makarova, K.S., et al., *An updated evolutionary classification of CRISPR-Cas systems*. Nat Rev Microbiol, 2015. **13**(11): p. 722-36.
5. Mohanraju, P., et al., *Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems*. Science, 2016. **353**(6299): p. aad5147.
6. Wiedenheft, B., S.H. Sternberg, and J.A. Doudna, *RNA-guided genetic silencing systems in bacteria and archaea*. Nature, 2012. **482**(7385): p. 331-8.
7. Amitai, G. and R. Sorek, *CRISPR-Cas adaptation: insights into the mechanism of action*. Nat Rev Microbiol, 2016. **14**(2): p. 67-76.
8. Bolotin, A., et al., *Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin*. Microbiology, 2005. **151**(Pt 8): p. 2551-61.
9. Mojica, F.J., et al., *Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements*. J Mol Evol, 2005. **60**(2): p. 174-82.
10. van der Oost, J., et al., *Unravelling the structural and mechanistic basis of CRISPR-Cas systems*. Nat Rev Microbiol, 2014. **12**(7): p. 479-92.
11. Brouns, S.J., et al., *Small CRISPR RNAs guide antiviral defense in prokaryotes*. Science, 2008. **321**(5891): p. 960-4.
12. Jinek, M., et al., *A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity*. Science, 2012. **337**(6096): p. 816-21.
13. Gasiunas, G., et al., *Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria*. Proc Natl Acad Sci U S A, 2012. **109**(39): p. E2579-86.
14. Deltcheva, E., et al., *CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III*. Nature, 2011. **471**(7340): p. 602-7.
15. Hsu, P.D., E.S. Lander, and F. Zhang, *Development and applications of CRISPR-Cas9 for genome engineering*. Cell, 2014. **157**(6): p. 1262-78.
16. Jiang, W., et al., *RNA-guided editing of bacterial genomes using CRISPR-Cas systems*. Nat Biotechnol, 2013. **31**(3): p. 233-9.
17. Barrangou, R. and J.A. Doudna, *Applications of CRISPR technologies in research and beyond*. Nat Biotechnol, 2016. **34**(9): p. 933-941.
18. Cong, L., et al., *Multiplex genome engineering using CRISPR/Cas systems*. Science, 2013. **339**(6121): p. 819-23.
19. Jones, D.L., et al., *Kinetics of dCas9 target search in Escherichia coli*. Science, 2017. **357**(6358): p. 1420-1424.
20. Sternberg, S.H., et al., *DNA interrogation by the CRISPR RNA-guided endonuclease Cas9*. Nature, 2014. **507**(7490): p. 62-7.

21. Chandradoss, S.D., et al., *Surface passivation for single-molecule protein studies*. J Vis Exp, 2014(86).
22. Brown, M.W., et al., *Assembly and translocation of a CRISPR-Cas primed acquisition complex*. bioRxiv, 2017.
23. Szczelkun, M.D., et al., *Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes*. Proc Natl Acad Sci U S A, 2014. **111**(27): p. 9798-803.
24. Chandradoss, S.D., et al., *A Dynamic Search Process Underlies MicroRNA Targeting*. Cell, 2015. **162**(1): p. 96-107.
25. Grimson, A., et al., *MicroRNA targeting specificity in mammals: determinants beyond seed pairing*. Mol Cell, 2007. **27**(1): p. 91-105.
26. Berg, O.G., R.B. Winter, and P.H. von Hippel, *Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory*. Biochemistry, 1981. **20**(24): p. 6929-48.
27. Singh, D., et al., *Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9*. Nat Commun, 2016. **7**: p. 12778.
28. Halford, S.E. and J.F. Marko, *How do site-specific DNA-binding proteins find their targets?* Nucleic Acids Res, 2004. **32**(10): p. 3040-52.
29. Riggs, A.D., S. Bourgeois, and M. Cohn, *The lac repressor-operator interaction. 3. Kinetic studies*. J Mol Biol, 1970. **53**(3): p. 401-17.
30. Hammar, P., et al., *The lac repressor displays facilitated diffusion in living cells*. Science, 2012. **336**(6088): p. 1595-8.
31. Leith, J.S., et al., *Sequence-dependent sliding kinetics of p53*. Proc Natl Acad Sci U S A, 2012. **109**(41): p. 16552-7.
32. Gorman, J., et al., *Single-molecule imaging reveals target-search mechanisms during DNA mismatch repair*. Proc Natl Acad Sci U S A, 2012. **109**(45): p. E3074-83.
33. Ragunathan, K., C. Liu, and T. Ha, *RecA filament sliding on DNA facilitates homology search*. Elife, 2012. **1**: p. e00067.
34. Anders, C., et al., *Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease*. Nature, 2014. **513**(7519): p. 569-73.
35. O'Geen, H., et al., *A genome-wide analysis of Cas9 binding specificity using ChIP-seq and targeted sequence capture*. Nucleic Acids Res, 2015. **43**(6): p. 3389-404.
36. Heler, R., et al., *Cas9 specifies functional viral targets during CRISPR-Cas adaptation*. Nature, 2015. **519**(7542): p. 199-202.
37. Joo, C. and T. Ha, *Labeling DNA (or RNA) for single-molecule FRET*. Cold Spring Harb Protoc, 2012. **2012**(9): p. 1005-8.







# 4

## Single-molecule FRET methods to study the Cas9 endonuclease

Since its discovery, the CRISPR-Cas9 system has been in the center of attention for its promising applications in genome editing. However, in order to apply this system successfully in genetic engineering, all aspects of its molecular mechanism have to be well understood. One of the best ways to investigate the intricacies of the molecular mechanism is single-molecule studies as they allow real-time observation of the kinetic processes and provide high spatio-temporal resolution. This chapter describes single-molecule fluorescence resonance energy transfer experiments carried out to investigate the Cas9 protein from *Streptococcus Pyogenes*, providing information on the effects of target truncation, Cas9-induced double-stranded DNA dynamics and target search.

---

This chapter has been published as: Globyte, V. and C. Joo, *Single-molecule FRET studies of Cas9 endonuclease*. *Methods Enzymol*, 2019. **616**: p. 313-335.

## 4.1. INTRODUCTION

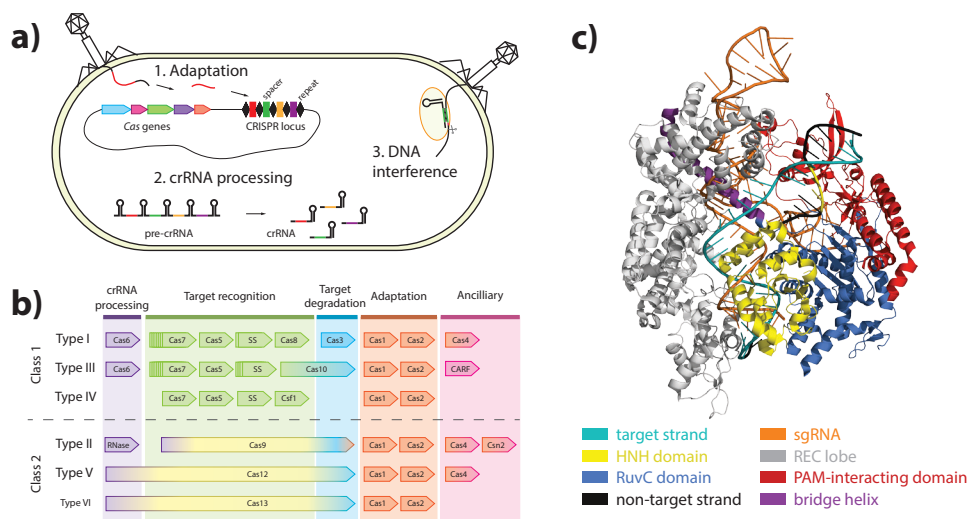
### 4.1.1. CRISPR BACTERIAL ADAPTIVE IMMUNE SYSTEM

Similar to eukaryotes, prokaryotes are under constant attack by mobile genetic elements, such as bacteriophages and plasmids. Through millions of years of evolution, prokaryotes have evolved distinct strategies to defend against invaders. One of these defense mechanisms is the CRISPR (Clustered Regularly Interspace Short Palindromic Repeats) adaptive immune system [1, 2]. Processes involved in CRISPR immunity can be grouped into three stages (Figure 4.1 a). In the first stage, known as the adaptation stage, short pieces of the invading DNA are integrated into the CRISPR locus in the bacterial genome, thus forming a genetic memory of the infection [1, 3]. In the second stage, the CRISPR locus is transcribed into precursor CRISPR RNA (crRNA) and the transcript is then processed to make short mature crRNA molecules [4]. In the final stage, known as the interference stage, crRNAs assemble with CRISPR-associated (Cas) proteins and guide them to destroy the invader when it returns.

Based on how DNA interference is performed in the final stage, CRISPR systems are grouped into two classes (Figure 4.1 b)[5, 6]. The main feature of Class 1 systems is a multi-subunit protein complex which assembles with crRNA and performs target recognition and/or degradation. This class is further subdivided into three types: type I, type III and type IV. In Class 2 systems, comprising type 2, type V and type VI systems, a single protein assembles with guide RNA to find and destroy invaders.

### 4.1.2. CAS9 ENDONUCLEASE

The most famous of the CRISPR systems is the type II CRISPR system and its signature protein Cas9 [7, 8]. This system gained widespread attention due to its simplicity and ease-of-use in gene-editing applications [9-11]. The Cas9 protein is a dual-RNA-guided endonuclease (Figure 4.1 c). Cas9 consists of two lobes (REC lobe and Nuclease lobe) and undergoes large conformational changes upon binding the duplex trans-activating RNA (tracrRNA) and crRNA [12]. The two RNA molecules can be fused into a single-guide RNA [13]. While Cas9 extensively interacts with tracrRNA which has an intricate architecture, the main role of crRNA is to guide the protein to its 20-nucleotide long target in the invader's genome. The first step of target recognition is recognizing a short 3 nucleotide motif called Protospacer Adjacent Motif (PAM) which is 5'-NGG-3' for *Streptococcus Pyogenes* Cas9 and is located on the non-target strand [14]. Recognizing the PAM melts the DNA duplex and assists the R-loop formation between crRNA and DNA [15, 16]. Finally,



**Figure 4.1. CRISPR systems.** a) Illustration of the three main steps of CRISPR immunity: adaptation, crRNA processing and DNA interference. b) CRISPR classes and types c) Crystal structure of the *Streptococcus pyogenes* Cas9 protein in complex with guide RNA and target DNA.

both the target and non-target strand are cleaved by the HNH and RuvC domains respectively [17]. This chapter describes single-molecule FRET studies of interactions between Cas9 and its DNA targets, focusing on the detailed descriptions on how to design and carry out experiments to answer different research questions concerning Cas9. The examples described in this chapter are assessing the effects of target truncations, Cas9-induced DNA duplex dynamics and Cas9 target search.

## 4.2. TIRF-BASED SINGLE-MOLECULE FRET

### 4.2.1. SINGLE-MOLECULE FRET

FRET (Forster Resonance Energy Transfer) is a non-radiative energy transfer between two fluorescent molecules (Figure 4.2b). If the emission spectrum of the donor fluorophore overlaps with the excitation spectrum of the acceptor fluorophore, once excited, the donor can then non-radiatively transfer its energy to the acceptor molecule through dipole-dipole coupling. The efficiency of the energy transfer is inversely proportional to the distance to the sixth power and thus can be used as a molecular ruler to determine the distance between two fluorescent dyes. Single-molecule FRET (smFRET) is an excellent technique to study biological processes in great detail as it offers great spatiotemporal resolution [18]. Biomolecules such as nucleic acids or proteins can be labelled by two or more fluorescent dyes

and the interactions between them can be monitored. Information that can be obtained through smFRET measurements includes, but is not limited to, affinity between two molecules, binding frequency (binding rate), binding location, protein inter-domain movements and many more.

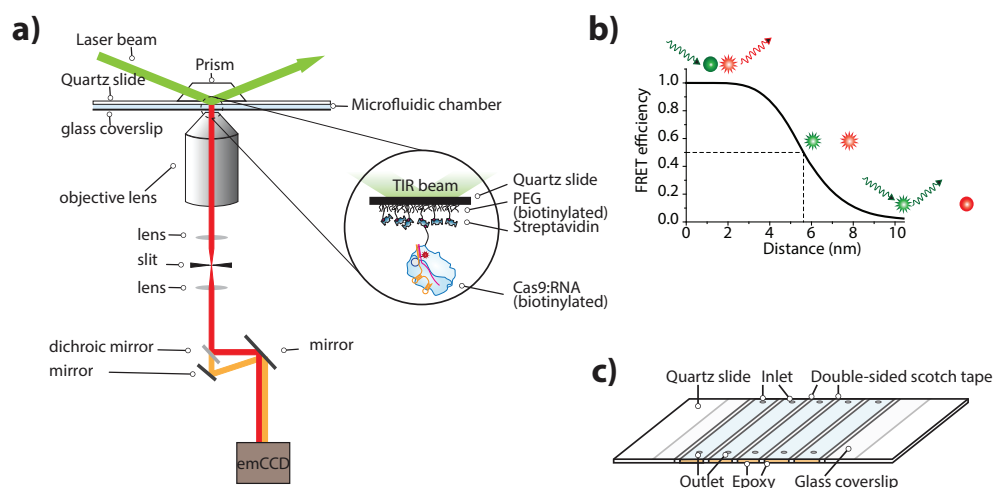
#### 4.2.2. TOTAL INTERNAL REFLECTION FLUORESCENCE MICROSCOPY

One of microscopy techniques used for smFRET is Total Internal Reflection Fluorescence (TIRF) Microscopy [19]. When the laser beam hits the sample chamber at a critical angle, it is totally internally reflected at the interface between the microscope slide and the microfluidic chamber. This gives rise to an evanescent field, which penetrates into the sample chamber. This technique is ideal because it has low background noise as only ~100 nm of the sample is illuminated by the evanescent field and it allows for parallel observation of several hundreds of molecules [20]. Two main types of TIRF microscopes exist – prism- and objective-type TIRF. The experiments described in this chapter rely on prism-type TIRF microscope, however, objective-type TIRF can also be used.

Schematic of a prism-type microscope is shown in Figure 4.2 a. The TIRF microscope is built around a commercial inverted microscope (Olympus). A quartz slide containing the sample is placed on the microscope stage. A drop of immersion oil is placed on the slide for index matching and a quartz prism (Eksma Optics) is placed on top. A laser beam hits the side of the prism and is directed to the quartz slide such that the beam hits the sample at the critical angle and is totally internally reflected. Evanescent electromagnetic field excites fluorescent molecules on the surface and the emitted fluorescence is collected by the objective lens. The imaging field of view is adjusted with a slit. The light beam is then collimated again and separated by dichroic mirrors (cut-off wavelength 630nm). Finally, the separated light beam is directed onto and recorded with an electron-multiplied charge coupled device (EMCCD) camera (Andor iXon Ultra 897).

#### 4.2.3. MICROSCOPE SLIDE ASSEMBLY

For single-molecule measurements, biotinylated DNA or protein are immobilized on the inner surface of a quartz microscope slide (Figure 4.2a). The slide is assembled by sandwiching double-sided scotch tape between a quartz slide and a coverslip (Figure 4.2c). Prior to assembly, holes are drilled in the quartz slide that act as inlet and outlet of the flow chamber. After drilling, slides are cleaned and coated with poly-ethylene glycol (Lysan) to



**Figure 4.2. Single-molecule FRET** a) Schematic of a prism type TIRF microscope b) Illustrative curve of FRET efficiency as a function of distance c) Schematic of a quartz slide used in single-molecule FRET experiments

avoid non-specific adsorption of proteins and nucleic acids to the slide. A detailed description of the PEGylation procedure can be found in [21]. For immobilization, a small amount (1-2%) of biotin-PEG is mixed in the PEG solution prior to PEGylation. Prior to experiments, a second round of PEGylation (at least 20 minutes) is performed to further improve the quality of the slide surface. If the protein used is very sticky, the flow chambers can be incubated with 5% Tween 20 (Sigma) for 10 minutes after slide assembly to decrease non-specific adsorption.

## 4.3. PREPARATION FOR SINGLE-MOLECULE EXPERIMENTS

### 4.3.1. FLUORESCENT LABELLING OF NUCLEIC ACIDS

Single-molecule FRET measurements require the addition of a FRET dye pair to the molecules studied. In the experiments described in this chapter, it is the nucleic acids that are fluorescently labelled [22]. Organic Cyanine dyes Cy3 and Cy5 (GE healthcare) are chosen for best signal-to-noise ratio. The separation between dyes must be such that energy transfer can occur (~2-10 nm or ~1-20 basepairs). For NHS-ester labelling, DNA and RNA strands with a C6 amine modification on a Thymine or Uracil base can be purchased commercially from companies such as IBA Life Sciences and IDT (Integrated DNA Technologies). The labelling procedure is performed as follows:

1. DNA or RNA samples are diluted to 1 mM using milliQ.
2. The sample is then mixed with ~20 mM dye and labelling buffer (Sodium bicarbonate, 8.4 mg/ml) in a volume ratio 1:1:5.
3. The labelling reaction is then incubated for 6 hours with gentle mixing in the dark.
4. Labelling is followed by ethanol precipitation.
5. The pellet is then carefully washed with 70% cold ethanol in order to remove free dye.
6. The pellet is dried and dissolved in a desired volume of a chosen solvent.
7. The UV-visible absorbance of nucleic acids (260 nm) and dyes (550 nm for Cy3 and 650 nm for Cy5) is then measured using a spectrometer (Nanodrop 2000). Concentration (molarity) is then calculated using Beer-Lambert law

$$c = \frac{A}{\epsilon l}$$

Here,  $c$  is the concentration,  $A$  is the absorbance,  $\epsilon$  is the extinction coefficient and  $l$  is the path length. Labelling efficiency is obtained by dividing the calculated dye concentration by the sample concentration and is typically above 90%. If there is too much free dye, ethanol precipitation can be repeated. If labelling efficiency is not high enough, a second round of labelling reaction can be done following ethanol precipitation.

#### 4.3.2. CAS9 BIOTINYLATION

In protein immobilization assays, biotinylated variant of recombinant Sp-Cas9 is used. The process of linking biotin to the recombinant protein is carried out in-vitro during protein purification [23]. The procedure is as follows:

1. After loading the SpCas9 over-expressed bacterial lysate and Ni-NTA mixture onto a column (Biorad), mixed sample is washed multiple times with wash buffer.
2. A 10 fold molar excess of maleimide-biotin (Sigma-Aldrich) is added to SpCas9 solution and incubated overnight at 4°C (mix gently with rotator).

3. To get rid of unbound maleimide-biotin chemicals, mixed sample is washed sufficiently with wash buffer.
4. Biotinylated SpCas9 protein is eluted with elution buffer.
5. The protein concentration is measured using a spectrophotometer (Nanodrop 2000, Thermo Fisher Scientific).
6. Eluted SpCas9 protein is further purified with size exclusion chromatography.
7. Biotinylated SpCas9 protein is stored in storage buffer and purified protein is snap-frozen in liquid nitrogen for long-term storage at -80°C.

### 4.3.3. BUFFERS AND REAGENTS

- Wash buffer
  - 20 mM Tris-HCL (pH 8.0)
  - 40 mM NaCl
- Elution buffer
  - 20 mM Tris-HCl (pH 8.0)
  - 400 mM NaCl
  - 200 mM Imidazole
- Storage buffer
  - 10 mM HEPES-KOH (pH 7.5)
  - 250 mM KCl
  - 1 mM MgCl<sub>2</sub>
  - mM EDTA
  - 7 mM b-mercaptoethanol
  - 20 % glycerol

## 4.4. DNA-IMMOBILIZATION BASED ASSAYS

In the experiments described in this section, assays based on the immobilization of DNA are used. The experiments explore the effects of target truncations on Cas9-DNA interactions and the Cas9-induced DNA duplex



dynamics. In the first case, dyes are placed on both DNA and crRNA to report binding. These experiments were carried out by making short snapshots of the molecules on the surface, which give ensemble FRET information over a long period of time. In the second case, dual-labelled DNA is used with labelling positions chosen at a distance in the most sensitive FRET range. The type of data obtained is real-time fluorescence traces, which can be further analyzed to obtain FRET histograms. This analysis in turn offers the insight into the processes after Cas9 binds DNA. Description of data analysis is provided in section 4.6. DNA immobilization procedure described in this section can be extended to other DNA- or RNA- binding proteins.

#### 4.4.1. EXPERIMENTAL PROCEDURE

DNA immobilization:

1. To immobilize DNA, 20  $\mu$ l 0.1 mg/ml streptavidin (ThermoFisher scientific) is added to the flow chambers and incubated for 2 min.
2. The chambers are then washed with 40  $\mu$ l T50.
3. Following streptavidin incubation, 25  $\mu$ l of 100 pM biotinylated fluorescently labelled DNA is injected to the flow slide and incubated for 2 min.
4. The chamber is then washed with 30  $\mu$ l imaging buffer, which contains oxygen scavenger (glucose oxidase, Sigma-Aldrich) and triplet state quencher (trolox, Sigma-Aldrich).

The experiments described in this section assume the flow chamber capacity to be 10-15  $\mu$ l. If the chamber is bigger, the volumes of reagents added need to be rescaled. The microscope slide is then placed on the TIRF microscope and the camera settings are adjusted for maximum signal-to-noise ratio. Typically, 300-600 fluorescent spots are observed per field of view. If the molecule density is too low, the DNA addition step can be repeated. If the molecule density is too high, it is best to choose a new channel and use a lower DNA concentration.

Preparation and addition of Cas9:RNA mix:

1. Following DNA immobilization, Cas9, crRNA and tracrRNA are mixed in imaging buffer such that final concentrations are 4 nM, 8 nM and 16 nM respectively if fluorescently labelled crRNA is used. If neither Cas9 nor any of the RNA molecules are labelled, the concentrations can be higher, as long as the ratio

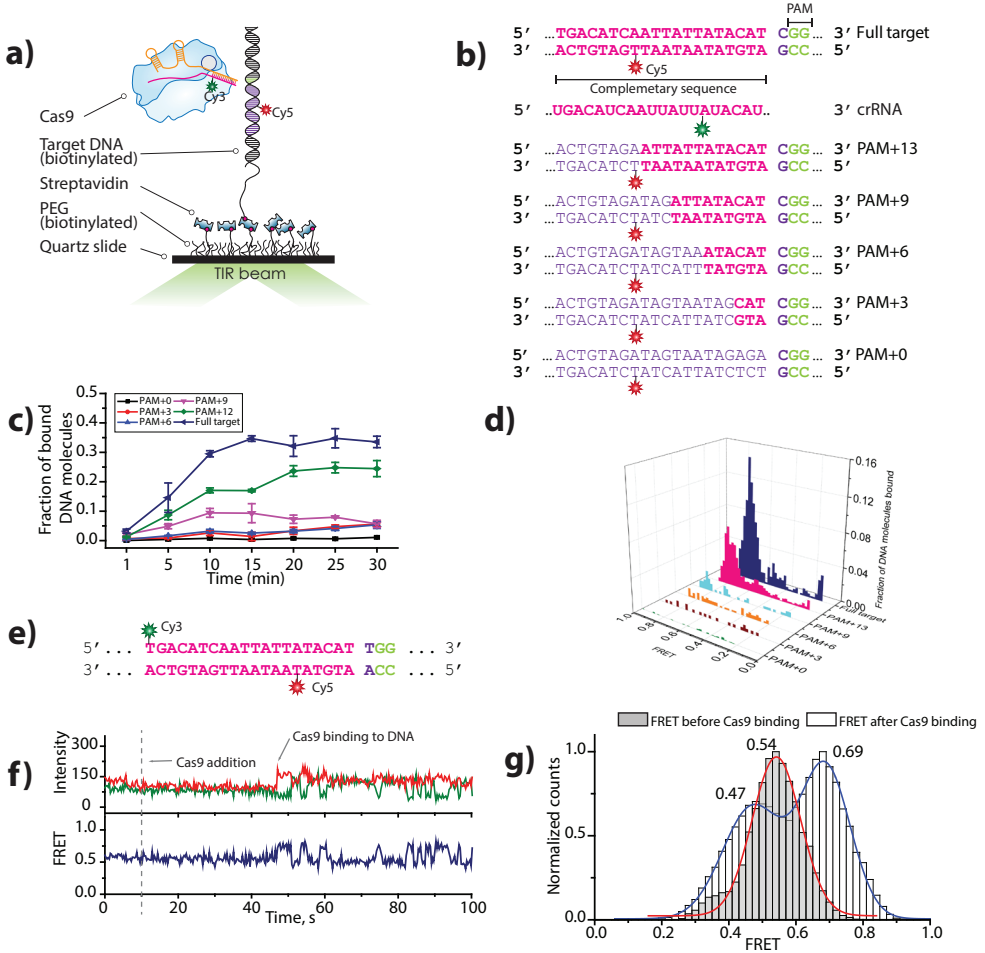
between Cas9:crRNA:tracrRNA is 1:2:4.

2. The mix is then incubated at 37C for 10 minutes and added to the flow slide.
3. The interactions between DNA and Cas9 are then monitored by either taking long videos with a typical exposure time of 100ms or by taking short ~10 frame snapshots.

The first type of measurement yields fluorescence time traces in which dynamic short-lived interactions can be further investigated (Figure 4.3 f). Such measurements, however, are subject to photobleaching, therefore it is important to adjust the laser power such that at least half of the molecules immobilized on the surface do not photobleach until the end of the measurement. The second type of measurement yields ensemble FRET histograms of single molecules which can be used to observe heterogeneous populations in the sample that cannot be observed in bulk measurements (Figure 4.3 d). This type of measurement can also be used to track the binding of the protein or other changes over long periods of time by taking snapshots at specific time intervals as this type of measurement is much less sensitive to photobleaching (Figure 6.3 c).

#### 4.4.1.1. Buffers and reagents

- T50
  - 10 mM Tris-HCl (pH 8.0)
  - 50 mM NaCl
- Cas9 imaging buffer
  - 50mM HEPES-NaOH (pH7.5)
  - 200mM NaCl
  - 10mM MgCl<sub>2</sub>
  - 0.8% glucose (Dextrose monohydrate)
  - 1mM Trolox (2.5mg/10ml)
  - 100x gloxy (glucose oxidase, (Sigma-Aldrich)) is added just before imaging (final 10x)



**Figure 4.3. DNA-immobilization-based experiments.** a) Illustration of the experimental scheme b) DNA and crRNA constructs used in experiments exploring target truncations c) Cas9 binding curves for each DNA construct over 30 minutes. Error bars represent standard error of the mean between 3 measurements. d) FRET efficiency histograms for each construct after 30 minutes of incubation in the microscope chamber. f) Representative fluorescence time trace of dual-labelled DNA experiments. Cas9:RNA complex was added at t=10s. g) FRET histogram showing the FRET states before and after Cas9 binding to dual-labelled DNA.

#### 4.4.2. ASSESSING THE EFFECTS OF TARGET TRUNCATIONS

The Cas9 protein is known to be very effective in rejecting off-targets [24, 25]. The directionality of R-loop formation implies that mutations at the beginning of the target sequence are less tolerated by the protein and that there exists a minimal guide-target basepairing length requirement for stable binding [15, 26]. In this section, sm-FRET experiments investigating the effects target truncations have on DNA-Cas9 binding are described.

##### 4.4.2.1. DNA and crRNA design

To observe what effects target truncations have on Cas9:DNA binding, we designed 6 DNA constructs with the following number of nucleotides matching the crRNA: PAM+20 (full target), PAM+13, PAM+9, PAM+6, PAM+3, PAM+0 (fully mismatched target) (Figure 4.3 b). The target region (23 nucleotides) is flanked by a 15-nucleotide random sequence on each side. The non-target strand is biotinylated and has a 24-nucleotide poly-Thymine overhang. The target strand is labelled with Cy5 at position +13 with respect to the PAM. crRNA is chosen to match the full target sequence and labelled with Cy3 at position +10 with respect to the first complementary nucleotide, such that a basepairing between crRNA and target DNA would yield a high FRET value. The high FRET value is an indication of Cas9 on-target binding as binding elsewhere on the DNA would give rise to lower FRET values.

##### 4.4.2.2. Effects of target truncations

Snapshot experiments conducted over 30 minutes reveal that Cas9 binding stability increases with increasing number of complementary nucleotides between crRNA and target DNA (Figure 4.3 c). The most drastic effect can be seen when the basepairing between guide and target exceeds 9 nucleotides. It can also be noted that for more extensive basepairing (13nt and full target), most binding events occur in the first 10 minutes after Cas9 has been introduced to the chamber. In addition, the FRET histogram 30 minutes after Cas9 introduction to the channel shows that while PAM+13 and full target constructs show expected high-FRET values for on-target binding, binding to DNA with less extensive guide-target pairing shows random FRET values, revealing that Cas9 does not efficiently bind truncated targets (Figure 4.3 d).

### 4.4.3. MONITORING CAS9-INDUCED DNA DUPLEX DYNAMICS

Upon PAM recognition, Cas9 is known to bend the DNA in order to assist the R-loop formation between guide RNA and target DNA [14]. In addition, Cas9 undergoes several conformational changes as the degree of basepairing between the guide and the target increases as it gears up for DNA cleavage [27, 28]. This section describes experiments carried out to monitor what changes Cas9 binding induces to the otherwise stable dsDNA.

#### 4.4.3.1. DNA design

In this assay, DNA and crRNA sequences are chosen the same as the full target sequence in section 4.2. However, this time dual-labelled DNA is used – non-target strand is labelled with Cy3 and target strand is labelled with Cy5 (Figure 4.3 e). The labelling positions are chosen on nucleotide 20 and nucleotide 6 (14 basepair separation) – in the most sensitive FRET range in order to allow for observation of any small changes in the inter-dye distance. The crRNA used is unlabeled.

#### 4.4.3.2. Real-time observation of Cas9-induced DNA duplex dynamics

An example FRET time-trace is shown in Figure 4.3 f. Initially, DNA is in its neutral state. At time  $t=10$ s, Cas9 is introduced in the channel. At time  $t=48$ s Cas9 binding to the immobilized DNA induces an increase in FRET efficiency. This could be due to either bending or twisting of the DNA which brings dyes closer together with a possible contribution from a phenomenon known as PIFE. The observed drop in FRET efficiency to a value of 0.47 corresponds to the unzipping of dsDNA and R-loop formation. The FRET histograms from multiple traces show that initial FRET value from DNA alone is 0.54 and after Cas9 binding the two observed states have a higher (0.71) and lower (0.47) FRET values corresponding to initial bending or twisting of DNA and unzipping the DNA respectively. The observed dynamics indicate that Cas9 binding significantly alters the state of the DNA target and that upon binding Cas9 is not stationary, but rather undergoes multiple dynamic conformational changes which induce the alterations in the FRET efficiency between dyes on the dsDNA.

## 4.5. CAS9 IMMOBILIZATION BASED ASSAYS

Another way to investigate protein-nucleic acid interactions is based on immobilizing the protein on the microscope slide surface. This can be achieved by the use of tags and antibodies (MBP-, SUMO-, His-tag, etc.) or by biotinylating the protein. In this section, experiments based on biotin-Cas9 immobilization are described (Figure 4.4 a). Immobilizing the protein is advantageous when weak interactions are investigated, since any free-floating labelled molecules, such as unbound RNA, are washed away during the immobilization step and therefore non-specific interactions are minimized. While the procedure described in this section is specific to SpCas9, protein immobilization is widely applicable to many different systems. Immobilizing the protein is also advantageous when two or more proteins can bind one substrate and give rise to unwanted effects. This section describes in detail the design of DNA and RNA used in experiments investigating Cas9-PAM interactions and Cas9 target search. Data obtained from both types of experiments are presented and discussed. Data analysis is described in more detail in section 4.6.

### 4.5.1. EXPERIMENTAL PROCEDURE

For Cas9 immobilization, biotinylated SpCas9 variant is used. The procedure is as follows:

1. Flow chambers on a microscope slide are incubated with streptavidin as in section 4.4.1.
2. Meanwhile, a Cas9-RNA mix is prepared and pre-incubated: Cy5-labelled crRNA and tracrRNA (final concentration 4 nM and 8 nM respectively) are mixed in NEB buffer 3 and incubated at room temperature for 10 minutes.
3. SpCas9 is added to the mix (final concentration 2 nM).
4. The mix is pre-incubated at 37C for 20 minutes.
5. After pre-incubation, 25  $\mu$ l Cas9-RNA is injected into a flow chamber and incubated for 2 min.
6. Following incubation, the channel is washed with 30  $\mu$ l imaging buffer as in section 4.4.1.
7. The slide is mounted on the microscope and camera settings are adjusted for maximum signal-to-noise ratio.

Typically, this yields 500-700 molecules on the surface. The density is

higher than in DNA immobilization scheme since not all crRNA-loaded Cas9 will be active. If the density of molecules is too low, more Cas9-RNA mix can be added to the channel. If the density of molecules on the surface is too high, it is best to dilute the remaining pre-incubated Cas9-RNA mix with 1x NEB buffer and add it to a new channel.

After Cas9 has been successfully immobilized, 30  $\mu$ l of 8nM DNA in imaging buffer are injected to the flow chamber and videos are recorded. In Cas9 immobilization assays, a low-salt imaging buffer is used.

#### 4.5.1.1. Buffers

- NEB buffer 3 1x
  - 100 mM NaCl
  - 50 mM Tris-HCl
  - 10 mM MgCl<sub>2</sub>
  - 1mM DTT
- Low salt imaging buffer
  - 50mM HEPES-NaOH [pH7.5]
  - 10mM NaCl
  - 2mM MgCl<sub>2</sub>
  - 1% glucose (Dextrose monohydrate)
  - 1mM Trolox (2.5mg/10ml)
  - Gloxy is added before imaging (final 10x)

#### 4.5.2. CAS9-PAM INTERACTIONS

PAM recognition is the first step in Cas9 target search [14, 29]. Binding to the guide RNA induces a conformational change in Cas9 after which a PAM-binding pocket is formed in the C-terminal domain of the protein [12]. PAM-recognition is independent of the guide sequence and is therefore a direct interaction between Cas9 and DNA, in which RNA only acts as an enabler by inducing Cas9 conformational change, priming the protein for PAM recognition. Magnetic tweezers and other smFRET studies have shown that without a cognate PAM Cas9 is unable to bind the target [15, 24]. This was explained by structural studies, which showed that PAM binding induces

a 30 degree bend of DNA and locally melts it to ease DNA unwinding [14]. DNA curtains studies have shown that aside from a target site, Cas9 transiently localizes on PAM-rich sites [25]. Furthermore, it has been shown in bulk that multiple-PAM containing DNA constructs act as good competitors for bona-fide targets [25]. This section describes experiments performed to investigate the Cas9 PAM-search process in detail.

#### 4.5.2.1. DNA and crRNA design

As mentioned previously, the guide or target sequence is not important for PAM binding. Therefore, in order to minimize non-specific interactions, a poly-U guide sequence was chosen (Figure 4.4 b). Six DNA sequences with the number of PAM sites increasing from 0 to 5 were designed. The sequence containing no PAM sites is a random 100 baspair long dsDNA construct. The sequences of the non-target strand, which contains the PAM, are as follows: (1) 5'-...NN GG NN..-3' (2) 5'-...NN GG NN GG NN..-3' (3) 5'-...NN GG NN GG NN GG NN..-3' (4) 5'-...NN GG NN GG NN GG NN GG NN..-3' (5) 5'-...NN GG NN GG NN GG NN GG NN GG NN..-3' (Figure 4.4 b). The constructs are 100bp long in total and NN can be any sequence except GG or CC. crRNA is labelled with Cy5 on the position -5 relative to the first nucleotide of the guide sequence and DNA is labelled with Cy3 at position -8 relative to the first PAM on the 3' end of the PAM sequence.

#### 4.5.2.2. PAM search

The chosen positions of the dyes in these experiments allow one to observe binding to an individual PAM separately, as binding to different PAM sites yields distinct FRET values. While binding to a random DNA sequence does not show any distinct FRET values and its dwelltime is characterized by a single-exponential decay distribution, binding to a PAM site shows a distinct peak which broadens and shifts to lower values as the number of accessible PAMs increases (Figure 4.4 c). In addition, the binding time distribution follows a double-exponential decay, meaning there are at least two distinct types of interactions occurring. The increase in the longer dwelltime with increasing number of PAM sites suggest the presence of multiple PAM sites in close proximity causes a synergistic effect (Figure 4.4 c). Further analysis of FRET and dwelltime values show no correlation between the two, meaning that the observed effects are not due to the position of the PAM site, but rather an intrinsic property of the system.





### 4.5.3. TARGET SEARCH

Unlike PAM search, which is the direct interaction between Cas9 and DNA, target recognition is guided by an RNA molecule (crRNA or the complementary segment of the sgRNA)[30]. The guide-target basepairing requirement leads to more stable interactions compared to those between the protein and the PAM. Some other RNA guided proteins, including a Cascade complex from type I CRISPR system, have been reported to use lateral diffusion to speed up target search [31, 32]. This section describes a tandem target assay to investigate the mechanics of target search and lateral diffusion in the SpyCas9 system.

#### 4.5.3.1. DNA and crRNA design

Tandem target DNA constructs have two identical partial targets separated by a certain number of basepairs. In the experiments described in this section, the two targets were separated by 23, 12, 9 and 6 basepairs - 4 constructs in total (Figure 4.4 d). The Cy3 dye on the DNA was placed on the target strand, on the 13th nucleotide with respect to the first PAM so as to avoid possible effects a dye between the two targets could have on Cas9-target interactions (Figure 4.4 d). Having a dye on the first target also allows changing the distance between the two binding sites while keeping the dye position constant. The crRNA was designed such that only the first 3 nucleotides are complementary to the target and the rest of the sequence was chosen to be poly-U to minimize the possibility of non-specific interactions. Binding between crRNA and full target DNA is very stable, therefore complementarity of only 3nt was chosen to allow for observation of multiple events and binding dynamics. The dye was placed on the 10th nucleotide with respect to the first complementary nucleotide. Such dye placement means that binding to the first target will yield a high-FRET value and binding to a second target will yield a low-FRET value. The low-FRET value decreases as the distance between the two target sites is increased.

#### 4.5.3.2. Direct observation of lateral diffusion

Recorded time traces revealed a type of dynamic events where the FRET value shifts continuously between two discrete values (Figure 4.4 e). Further examination revealed that the two FRET values are the expected values from binding to each of the two identical target sites (Figure 4.4 f). This observation is indicative of Cas9 diffusing laterally between the two target

sites. Analysis of translocation frequency revealed that transition probability decreases with increasing distance between target sites, from ~0.4 for 6 basepair distance to ~0.07 for 23 basepair distance (Figure 4.4 g). Furthermore, dwelltime distribution is characterized by a double-exponential decay (Figure 4.4 h), with the longer dwelltime being an order of magnitude longer than what is expected in the case of 3nt complementarity between crRNA and DNA target, indicating that two target sites in close proximity cause a strong synergistic effect [23].

## 4.6. DATA ANALYSIS

Single-molecule FRET experiments can give insight into various intricate processes involving molecules of interest. This information is extracted through careful data analysis. Most data analysis involves the processing of raw data to obtain files that can then be analyzed further to answer various research questions. This section describes the general data analysis principles used in the experiments described in this chapter, from video processing using IDL, to FRET and dwelltime analysis using MatLAB algorithms and OriginPro 9.0 graphing software.

### 4.6.1. VIDEO PROCESSING

The videos recorded by the EMCCD are processed in IDL using a custom written script to extract fluorescence intensities of molecules on the surface. Fluorescence from a single molecule is imaged as a point spread function on a CCD. The IDL algorithm finds intensity peaks in the recorded videos and extracts the fluorescence intensity from each molecule over the duration of the video. Extracted time traces are analyzed using a custom written MatLAB script.

$$E = \frac{I_A}{I_A + I_D \gamma}$$

### 4.6.2. FRET EFFICIENCY AND DWELLTIME ANALYSIS

FRET efficiency is calculated by the ratio of the acceptor intensity to the total intensity:

Here the  $\gamma$  factor accounts for the difference in intrinsic brightness of the

donor and acceptor molecules. The value for Cy3-Cy5 pair is approximately 1.

To extract FRET and dwelltime values, each individual binding event is selected manually, selecting the start and end of the event. Such manual analysis allows for greater flexibility, as any desired part of the event, such as the first few seconds or a FRET state of interest, can be selected. The values obtained by MatLAB analysis are written into files that can be analyzed using OriginPro or other graphing software. The files produced by MatLAB analysis from the experiments described in this chapter were plotted and analyzed further using OriginPro 9.0. Regardless of the software used, general guidelines for fitting histograms apply. For example, when dwelltime histograms are plotted and fitted using in-built algorithms, it is important to select a binning size which captures the decay, yet does not override the details. For example, for distributions where the majority of binding events are shorter than 1s, choosing the bin size larger than 1s runs the risk of losing important details, such as possible dual-exponential decay. If most binding events are several seconds long, choosing a smaller bin size will not give any extra information, but rather make the fitting more difficult and less reliable. One should also be critical about the values the fitting algorithm gives for dwelltime. For example, plotting and fitting should be repeated if the obtained value for dwelltime is shorter than the camera exposure time. When fitting FRET histograms, a large data set allows for a smaller bin size. However, if the data set is small, caution is needed as choosing a larger bin size can often hide smaller FRET peaks if multiple peaks are present in the histogram. FRET histograms are fit with standard Gaussian functions which are in-built in most graphing softwares. The FRET histograms shown in this chapter have been fit with a standard Gaussian function and plotted between FRET efficiency values between -0.1 and 1.1 with a bin size of 0.05.

A MatLAB script is also used to obtain FRET values from short snapshot videos in batch. Fluorescence intensity values of each molecule in the field of view are extracted from all videos per time point using IDL and numerical FRET values are then obtained using a MatLAB script. The FRET distributions are then plotted and fitted using OriginPro 9.0. However, unlike during individual-event-based analysis, the y-axis now is not counts, but the number of molecules by default.

Finally, translocation frequency analysis in tandem target experiments is performed by manually counting translocation events, considering each transition as a separate binding event and then dividing the obtained value by the total number of binding events (Figure 4.4 e).

The data analysis performed manually is time-consuming and labor-intensive, however, it allows for greater flexibility and accuracy, especially when

weak interactions are investigated. The data analysis can also be automated (when all events in time traces are selected automatically) or semi-automated (when one decides which parts of the time traces should be automatically analyzed). Which type of analysis will yield the best results depends on many factors, such as the type of data being analyzed, the quality of the data, previously obtained results, and therefore each case should be considered individually.

## 6.7. CONCLUDING REMARKS

Single-molecule FRET is a powerful tool to study the interactions between biomolecules, such as RNA-guided proteins and their targets. This chapter provides detailed descriptions of smFRET studies of the CRISPR-Cas9 system, with research questions ranging from effects of target truncations to target search. To address each of the research questions, DNA and RNA are carefully designed. Each process is thus controlled separately and measured using either DNA immobilization assay or protein immobilization assay. The wealth of information obtains during the experiments described in this chapter provides an insight into the molecular mechanism of the function of SpCas9. The assays and techniques described in this chapter can be used to address different aspects of the SpyCas9 molecular mechanism. The principle of these experiments can easily be applied to study other RNA- or DNA-guided proteins.

## REFERENCES

1. Barrangou, R., et al., *CRISPR provides acquired resistance against viruses in prokaryotes*. Science, 2007. **315**(5819): p. 1709-12.
2. Marraffini, L.A., *CRISPR-Cas immunity in prokaryotes*. Nature, 2015. **526**(7571): p. 55-61.
3. Amitai, G. and R. Sorek, *CRISPR-Cas adaptation: insights into the mechanism of action*. Nat Rev Microbiol, 2016. **14**(2): p. 67-76.
4. Brouns, S.J., et al., *Small CRISPR RNAs guide antiviral defense in prokaryotes*. Science, 2008. **321**(5891): p. 960-4.
5. Makarova, K.S., et al., *An updated evolutionary classification of CRISPR-Cas systems*. Nat Rev Microbiol, 2015. **13**(11): p. 722-36.
6. Mohanraju, P., et al., *Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems*. Science, 2016. **353**(6299): p. aad5147.
7. Jinek, M., et al., *A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity*. Science, 2012. **337**(6096): p. 816-21.
8. Gasiunas, G., et al., *Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria*. Proc Natl Acad Sci U S A, 2012. **109**(39): p. E2579-86.
9. Doudna, J.A. and E. Charpentier, *Genome editing. The new frontier of genome engineering with CRISPR-Cas9*. Science, 2014. **346**(6213): p. 1258096.
10. Barrangou, R. and J.A. Doudna, *Applications of CRISPR technologies in research and beyond*. Nat Biotechnol, 2016. **34**(9): p. 933-941.
11. Jiang, W., et al., *RNA-guided editing of bacterial genomes using CRISPR-Cas systems*. Nat Biotechnol, 2013. **31**(3): p. 233-9.
12. Jinek, M., et al., *Structures of Cas9 endonucleases reveal RNA-mediated conformational activation*. Science, 2014. **343**(6176): p. 1247997.
13. Hsu, P.D., E.S. Lander, and F. Zhang, *Development and applications of CRISPR-Cas9 for genome engineering*. Cell, 2014. **157**(6): p. 1262-78.
14. Anders, C., et al., *Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease*. Nature, 2014. **513**(7519): p. 569-73.
15. Szczelkun, M.D., et al., *Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes*. Proc Natl Acad Sci U S A, 2014. **111**(27): p. 9798-803.
16. Jiang, F., et al., *Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage*. Science, 2016. **351**(6275): p. 867-71.
17. Huai, C., et al., *Structural insights into DNA cleavage activation of CRISPR-Cas9 system*. Nat Commun, 2017. **8**(1): p. 1375.
18. Roy, R., S. Hohng, and T. Ha, *A practical guide to single-molecule FRET*. Nat Methods, 2008. **5**(6): p. 507-16.
19. Ambrose, E.J., *A surface contact microscope for the study of cell movements*. Nature, 1956. **178**(4543): p. 1194.
20. Ambrose, W.P., P.M. Goodwin, and J.P. Nolan, *Single-molecule detection with total internal reflection excitation: comparing signal-to-background and total signals in different geometries*. Cytometry, 1999. **36**(3): p. 224-31.

21. Chandradoss, S.D., et al., *Surface passivation for single-molecule protein studies*. J Vis Exp, 2014(86).
22. Joo, C. and T. Ha, *Labeling DNA (or RNA) for single-molecule FRET*. Cold Spring Harb Protoc, 2012. **2012**(9): p. 1005-8.
23. Globyte, V., et al., *CRISPR/Cas9 searches for a protospacer adjacent motif by lateral diffusion*. EMBO J, 2019. **38**(4).
24. Singh, D., et al., *Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9*. Nat Commun, 2016. **7**: p. 12778.
25. Sternberg, S.H., et al., *DNA interrogation by the CRISPR RNA-guided endonuclease Cas9*. Nature, 2014. **507**(7490): p. 62-7.
26. Lim, Y., et al., *Structural roles of guide RNAs in the nuclease activity of Cas9 endonuclease*. Nat Commun, 2016. **7**: p. 13350.
27. Dagdas, Y.S., *A Conformational Checkpoint Between DNA Binding And Cleavage By CRISPR-Cas9*. bioRxiv.
28. Sternberg, S.H., et al., *Conformational control of DNA target cleavage by CRISPR-Cas9*. Nature, 2015. **527**(7576): p. 110-3.
29. Karvelis, T., et al., *Rapid characterization of CRISPR-Cas9 protospacer adjacent motif sequence elements*. Genome Biol, 2015. **16**: p. 253.
30. Nishimasu, H., et al., *Crystal structure of Cas9 in complex with guide RNA and target DNA*. Cell, 2014. **156**(5): p. 935-49.
31. Dillard, K.E., et al., *Assembly and Translocation of a CRISPR-Cas Primed Acquisition Complex*. Cell, 2018. **175**(4): p. 934-946 e15.
32. Chandradoss, S.D., et al., *A Dynamic Search Process Underlies MicroRNA Targeting*. Cell, 2015. **162**(1): p. 96-107.







# 5

## Small RNA molecules inhibit the activity of SpCas9 *in vitro*

The Cas9 endonuclease from *Streptococcus pyogenes* (SpCas9) has been widely applied in genome editing. Despite the ease of application, the delivery of Cas9 to cells for therapeutics remains problematic. Most delivery methods include a plasmid-based approach where Cas9 and sgRNA are encoded in the same vector, delivery of mRNA for Cas9 translation together with a separate sgRNA and the introduction of an assembled ribonucleoprotein complex. Of the three strategies, the first two are easier to deliver, however they suffer from the lack of control over Cas9 activity and expression and are more susceptible to cellular factors that can decrease the activity of Cas9. Here we report that small RNA and DNA molecules efficiently inhibit the catalytic activity of Cas9 *in vitro* when hybridized with the guide RNA. We use single-molecule fluorescence to determine that Cas9 efficiently loads the duplex but is then stuck in an inactive conformation which prevents it from interacting with the target even after the inhibitor molecule is removed.

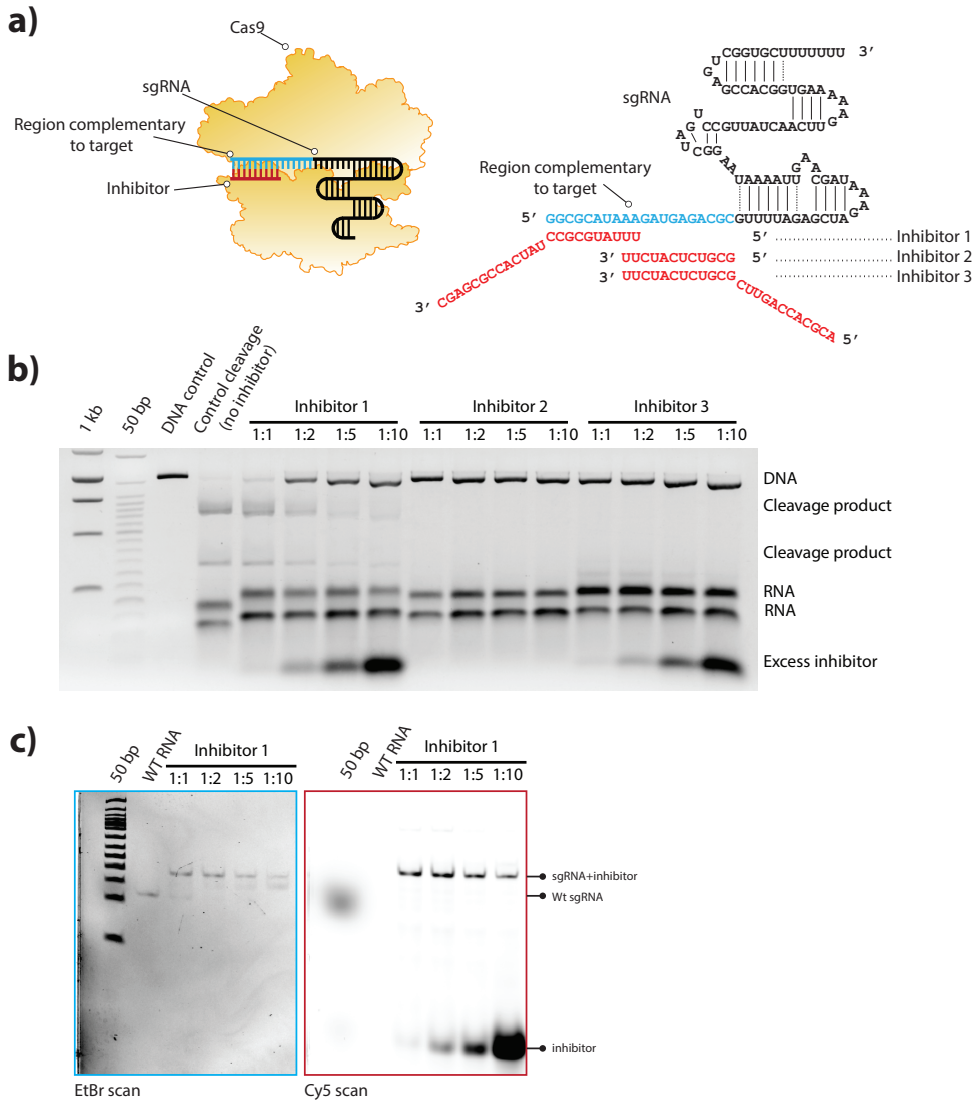
## 5.1. INTRODUCTION

Clustered regularly interspaced short palindromic repeat (CRISPR) systems and their associated (Cas) proteins are adaptive prokaryotic immune systems that provide bacteria and archaea with adaptive immunity against invading viruses and plasmids [1-5]. The type II-A CRISPR/Cas9 system from *Streptococcus pyogenes* is a well characterized RNA-guided DNA endonuclease that has been extensively repurposed for genome engineering due to its simplicity and programmability[6-9]. However, efficient delivery of Cas9 to cells and tissue remains problematic. Most delivery methods rely on one of the three following approaches: delivery of the DNA encoding Cas9 and sgRNA using a viral vector, delivery of sgRNA and mRNA for translating Cas9 and assembling the ribonucleoprotein (RNP) complex *in vivo*, and electroporating the assembled RNPs into cells [10-15]. While the third strategy offers the greatest control of the assembly of functional Cas9:RNA complexes and their concentration, it is limited to *ex vivo* applications. The other two main strategies can be applied *in situ*, however, they do not offer precise control over the expression of Cas9 and the assembly of functional RNPs. Furthermore, cellular components might interfere with either the RNA transcription, the translation of the protein or RNA loading into Cas9. In particular, Cas9 has been demonstrated to be an efficient strategy for the treatment of HIV when applied in combination with RNA interference[16]. In such cases, if Cas9 is to be applied with a different RNA-guided protein targeting a similar gene, we postulate that the RNA sequences may undergo transient interactions via Watson-Crick basepairing and may in turn affect the activity of Cas9 if the RNP is expressed and assembled *in vivo*. Here we demonstrate that small RNA molecules can completely abolish the target cleavage of Cas9 *in vitro* and use single-molecule fluorescence microscopy in order to determine the mechanism of inhibition.

## 5.2. RESULTS

### 5.2.1. SMALL RNA INHIBIT DNA CLEAVAGE BY CAS9

In order to determine if small RNAs have an effect on Cas9 function, we designed inhibitor RNA molecules complementary to either the PAM proximal or PAM distal regions of the guide RNA (Figure 5.1 a). The guide RNA was pre-incubated with increasing concentrations of the inhibitor RNAs. In order to determine if Cas9 is functional, we performed cleavage assays in the absence and presence of small RNA inhibitors. A control experiment without an inhibitor showed that Cas9 efficiently cleaves the target (Figure 5.1 b). However, the addition of the inhibitor at increasing ratios of 1:1; 1:2;



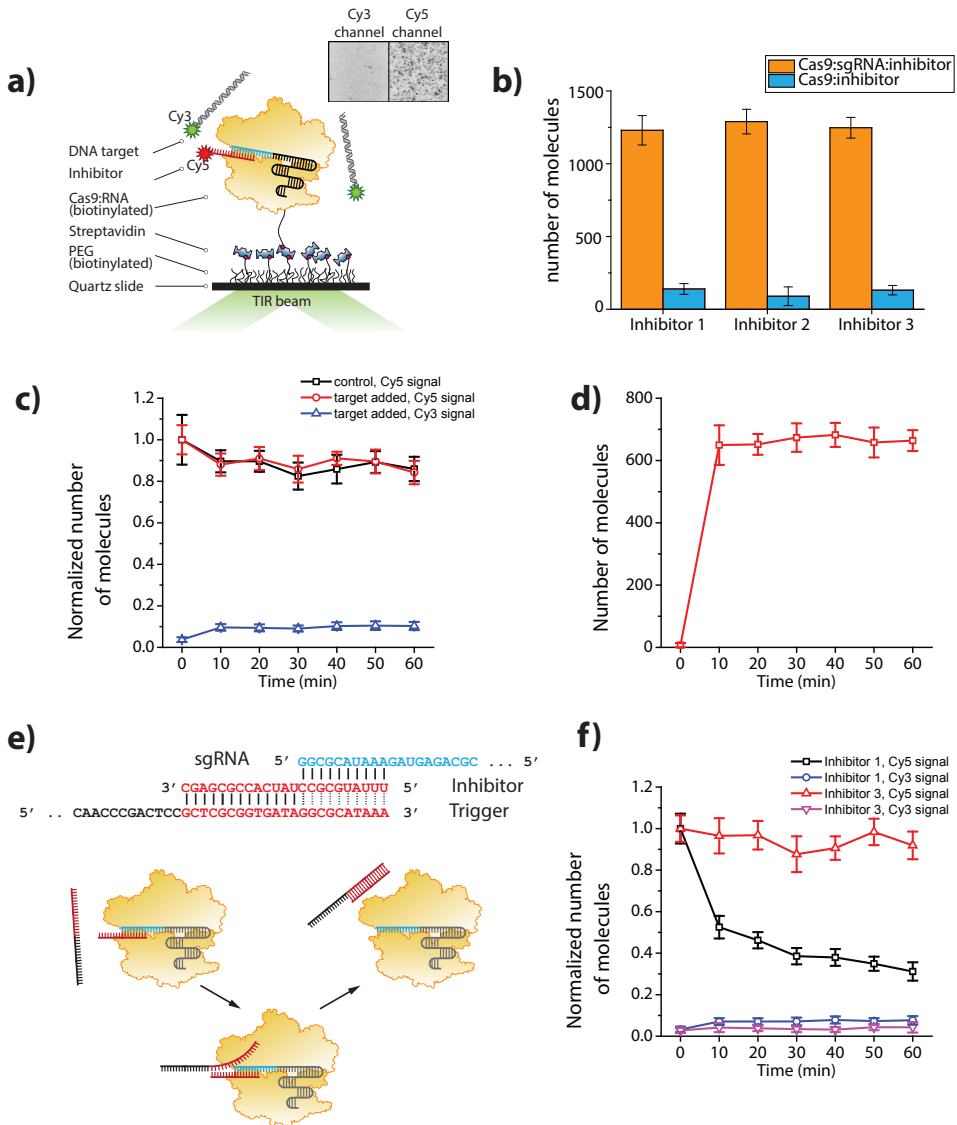
**Figure 5.1. Small RNAs inhibit Cas9 cleavage in vitro.** a) A schematic representation of the inhibitor molecules and Cas9. b) A 3% agarose gel showing an In vitro cleavage assay demonstrating the inhibition of cleavage. c) Electrophoretic mobility shift assay resolved using 10% PAGE showing that inhibitor 1 efficiently hybridizes with sgRNA.

1:5; 1:10 (to guide RNA) showed decreasing cleavage of target DNA (Figure 5.1 b). Inhibitor 1 was found to progressively inhibit cleavage as the ratio of inhibitor to guide RNA was increased. Inhibitors 2 and 3 completely abolished cleavage even at the lowest ratios (Figure 5.1 b). In order to test

whether the inhibitors can stably bind to the guide RNA, we performed an electrophoretic mobility shift assay (EMSA). The shift assay demonstrated that the inhibitors were stably hybridized to the guide RNA and shifted up in the gel, which suggested that the hybrid might be loaded by Cas9 (Figure 5.1 c, Figure S5.1 a, b).

### 5.2.2. SINGLE-MOLECULE FLUORESCENCE REVEALS THE MECHANISM OF INHIBITION

In order to further elucidate the mechanism by which the small RNA molecules inhibit the catalytic activity of Cas9, we developed a single-molecule fluorescence assay in which the Cas9 protein, pre-incubated with the sgRNA and the Cy5-labeled inhibitor, is immobilized on the surface of the microscope slide (Figure 5.2 a). After directly exciting the Cy5 fluorophores, we observe  $1229 \pm 100$  molecules per field of view on average for inhibitor 1,  $1289 \pm 84$  for inhibitor 2 and  $1247 \pm 71$  for inhibitor 3, indicating that Cas9 efficiently loads the guide-inhibitor hybrid (Figure 5.2 b). A control experiment where Cas9 was pre-incubated only with the inhibitor, but without sgRNA showed only around 100 molecules per field of view on average for each inhibitor (Figure 5.2b), indicating that the observed high density of the molecules in the previous cases are indeed Cas9:sgRNA:inhibitor complexes. In order to monitor the interactions between the complexes immobilized on the surface and a DNA target, a Cy3-labelled target DNA was added and the interactions were monitored for 60 minutes by taking short snapshots over multiple fields of view every 10 minutes. Snapshot experiments were chosen in order to minimize the photobleaching of the fluorophores. The molecules were simultaneously excited by a 532 nm and a 633 nm laser in order to monitor the behaviour Cy3-labeled target and Cy-5 labeled Cas9:sgRNA:inhibitor complexes independently. A control, where no target has been added, showed that the inhibitor remained stably bound over the imaging time (Figure 5.2c, Figure S5.2 a, b) with the number of molecules on the surface decreasing on average by  $15 \pm 5$  percent due to photobleaching. Interestingly, we observed similar levels of loss of Cy5 fluorescence in the microfluidic chamber where the target DNA was added, indicating that the presence of the target did not affect the Cas9:RNA:inhibitor complex (Figure 5.2c, Figure S5.2a, b). Incidentally, we observed very few molecules appearing in the Cy3 channel, which indicated that the Cas9:sgRNA:inhibitor complexes did not bind the target. No transient interactions were observed, which suggested that Cas9 did not interact with neither target nor PAM sequences when loaded with the guide:inhibitor duplex. A control, where no inhibitor was hybridized to the guide RNA, showed fast and efficient binding (Figure 5.2d), demonstrating that the Cas9 protein is active. This indicates that the



**Figure 5.2. Single-molecule fluorescence reveals the mechanism of inhibition** a) schematic representation of the single-molecule assay. b) a histogram showing the average number of molecules per field of view. The molecules were counted as fluorescent spots in the Cy5 channel. The average was taken over 15 snapshots over different fields of view. The error bars represent standard deviation. The data is the average of three independent experiments taken over three different days. c) a scatter plot showing the number of molecules on the surface over 60 minutes for inhibitor 1. The molecules were counted as fluorescent spots in the Cy5 channel for Inhibitor and in the Cy3 channel for target. The number of molecules was normalized by the number of molecules on the surface at time= 0. The average was

taken over 15 snapshots over different fields of view per time point. The error bars represent standard deviation. The data is the average of three independent experiments taken over three different days. d) a scatter plot showing the binding of DNA target to Cas9:sgRNA complexes on the microscope slide. Cy3-labeled DNA was imaged. The average was taken over 15 snapshots over different fields of view. The error bars represent standard deviation. The data is the average of three independent experiments taken over three different days. e) a schematic representation of the assay with the addition of the trigger molecule f) a scatter plot showing the number of molecules on the surface after the addition of the trigger molecule in combination with the target. the number of molecules was normalized by the number of molecules on the surface at time= 0. The average was taken over 15 snapshots over different fields of view. Inhibitor molecules were counted as fluorescent spots in the Cy5 channel and target molecules as fluorescent spots in the Cy3 channel. The error bars represent standard deviation. The data is the average of three independent experiments taken over three different days.

small RNA molecules inhibit the catalytic activity of Cas9 by preventing it from interacting with the target DNA.

5 To test whether the inhibitor molecule could be removed from the protein, we designed a DNA strand which is complementary to inhibitors 1 and 3 and may trigger the release of the inhibitor from the complex by competing with the basepairing between the inhibitor and the sgRNA (Figure 5.2 e). Time-lapse images over 60 minutes in the microfluidic chamber, where the trigger and the target were added together, showed a very rapid loss of Cy5 fluorescence, where the number of fluorescent molecules decreased by 50% over the first 10 minutes with the number of molecules on the surface decreasing more slowly over the next 50 minutes with 30% of the Cy5-labeled molecules remaining after 60 minutes. This was observed in the case of inhibitor 1, but not inhibitor 3 (Figure 5.2 f). The behavior of the complexes with inhibitor 3 when the trigger was present in the channel was observed to be identical to the control when neither trigger nor target was added to the channel (Figure 5.2 f, Figure S5.2 b). As the extent of basepairing between Inhibitor 1 and guide is 4nt shorter than in the case of Inhibitor 3, we set out to determine, whether it is the extent of basepairing between sgRNA and the inhibitor RNA or the position of the latter with respect to the guide, which gives rise to the loss of fluorescence. We designed two inhibitor molecules, Inhibitor 4 which has a 2nt longer basepaired region than Inhibitor 1, and inhibitor 5 which has a 2nt shorter basepaired region than Inhibitor 3 (Figure S5.2c). After performing the same experiments with the trigger molecules, we observed loss of fluorescence in the case of Inhibitor 4, identical to that of Inhibitor 1 (Figure S.2d). Inhibitor 5 did not show any loss of fluorescence, similar to Inhibitor 3. This shows that the trigger is only able to induce the release of an inhibitor only when it is situated at the end of the guide RNA, suggesting that the protein holds onto the inhibitor more strongly when it is hybridized to the seed region of the guide. This may explain why the cleavage inhibition is less efficient in the case of Inhibitor 1 (Figure 5.1 b). Interestingly, despite the loss of Cy5 fluorescence indicating the release of

inhibitor 1, a correlating appearance of molecules in the Cy3 channel was not observed. This suggests, that despite the inhibitor leaving the protein, the protein remains locked in an inactive conformation, unable to further interact with the target.

In order to further confirm that Cas9 strongly binds the sgRNA:inhibitor duplex we have attempted to perform an electrophoretic mobility shift assay. However, with no Cas9-bound DNA present, Cas9:RNA ribonucleoprotein complexes did not enter the gel. In addition, we have performed atomic force microscopy (AFM) imaging of the assembled RNPs in order to observe whether the inhibitor induces a substantial conformational change (data not shown). Unfortunately, we could not see any clear differences between the RNPs with and without inhibitor or any clear structural features as Cas9 is a small and globular protein. Therefore, in order to observe any structural intermediates, a technique such as cryo electron microscopy or X-ray crystallography should be used. Currently, we are performing *in vivo* plasmid loss assays to see whether the inhibition also takes place in living cells. Furthermore, an *in vitro* experiment with purified miRNA from HEK297T cells which could inhibit cleavage is under consideration, as a mimic of an *in vivo* condition.

### 5.3. DISCUSSION

Despite the enormous interest in Cas9 and its potential in therapeutics, regulating its activity remains challenging in areas where precise control of the process is required. Moreover, if Cas9 is to be applied in combination with another RNA-guided protein, programmed to target a similar sequence, the interactions between the RNA may potentially give rise to unexpected and unwanted effects. In addition, it has been demonstrated that guide RNA molecules that form hairpin structures at the PAM-distal region can increase the specificity of Cas9 by making it energetically unfavourable for it to pair with off-target sites, which decreases the overall activity of Cas9 [17]. We show that Cas9 cleavage of the target DNA can be efficiently inhibited by small RNA molecules which are hybridized to the sgRNA and loaded into Cas9. Using single-molecule fluorescence, we have demonstrated that the sgRNA:inhibitor hybrids are efficiently loaded into Cas9 and remain stable over at least 60 minutes. This suggests that the negatively charged groove between the REC and the Nuclease lobes of Cas9 can accommodate an RNA duplex without the necessity for structural rearrangements upon PAM recognition, which are necessary to accommodate the target DNA [18]. Furthermore, we did not observe interactions with the target even though a control experiment with bare RNA showed efficient target binding, which demonstrates that small RNA inhibits the function of Cas9 by preventing it from in-



5 interacting with the target. A trigger molecule complementary to the inhibitor was observed to be able to trigger the release of the inhibitor from the RNP. Surprisingly, no target binding was observed even after the inhibitor molecule has been removed, which further implies that the loading of inhibitor:sgRNA hybrid causes Cas9 to adopt an inactive conformation which it cannot change even when the inhibitor is no longer present. Interestingly, we did not observe the loss of fluorescence indicative of the inhibitor dissociating from the complex with inhibitor 3, which suggests that the protein interacts with it more strongly when it is placed in the seed region as the electrophoretic mobility shift assays showed that all sgRNA molecules hybridize to the inhibitors equally efficiently. In addition, a recent study demonstrated that riboregulated toehold-gated gRNA can be used to control Cas9 function by the addition of a trigger molecule which unfolds the toehold-gated gRNA [19]. Our work suggests that this is only possible before Cas9 is loaded with a hybridized or folded RNA molecule. Therefore, our findings demonstrate that delivery of DNA or sgRNA and mRNA for translating Cas9 and assembling the RNP complexes *in vivo* may be susceptible to unwanted effects such as possible inhibition by endogenous RNA molecules. However, further research will be necessary to elucidate the extent of this inhibition *in vivo*. Nevertheless, electroporating assembled RNP complexes into cells may be the strategy that will suffer the least from unexpected effects.

## 5.4. MATERIALS AND METHODS

### 5.4.1. sgRNA PRODUCTION

To make the sgRNA, we first PCR amplified a dsDNA template, which contains the target sequence from a DNA plasmid (pgRNA-bacteria plasmid from Addgen), using a primer that contains a T7 promoter. The following thermal cycling conditions were used to generate the PCR template: 98°C for 3 minutes; 98°C for 10 seconds; 65°C for 20 seconds; 72°C for 15 seconds; go to step 2 for 29 cycles and 72°C for 8 minutes. The PCR template was verified using gel electrophoresis (1,5% agarose, 1X TBE buffer, 120V for 90 minutes) and subsequently purified using the WizardSV Gel and PCR Clean-Up System (Promega) according to the manufacturer's instructions. A single gRNA was then transcribed from the PCR template using the RiboMax™ Large Scale RNA Production Systems kit (Promega) according to the manufacturer's instructions. Following transcription, RNA products were purified using the RNeasy MinElute Cleanup Kit (Qiagen) according to the manufacturer's instructions. RNA quality was verified using gel electrophoresis (Mini-Protean TBE-Urea Precast Gels (Bio-Rad), 200V for 30 minutes). Gels were visualized

under UV light in a Biorad ChemiDOCT MP imaging system.

#### 5.4.2. CLEAVAGE ASSAYS

For cleavage assays, sgRNA was prepared by heating up to 95°C for 10 minutes and slowly cooling down (1°C every 4 minutes until a final temperature of 4°C). Cas9 was pre-incubated with sgRNA in a 1X Cas9 Nuclease Reaction Buffer (New England Biolabs) in a molar ratio of 1:10 for 30 minutes at 25°C. Thereafter, target DNA was added in a molar ratio of 1:10 to Cas9, effecting a total molar ratio of 100:10:1 of sgRNA:dCas9:DNA. An excess ratio was used to ensure complete cleavage. The complex was incubated for 30 minutes at 37°C. Proteinase K was then used to digest Cas9 for 15 minutes at 37°C. Cleavage products were verified using gel electrophoresis (1,5% agarose, 1X TBE buffer, 120V for 90 minutes). Gels were visualized under UV light in a Biorad ChemiDOCT MP imaging system.

#### 5.4.3. ELECTROPHORETIC MOBILITY SHIFT ASSAY (EMSA)

For EMSA, sgRNA was incubated with different concentrations of inhibitor RNA (molar ratios of (1:1; 1:2; 1:5; 1:10) by heating up to 95°C for 10 minutes and slowly cooling down (1°C every 4 minutes until a final temperature of 4°C). the sgRNA-inhibitor hybrids were analyzed using gel electrophoresis (10% 1X TBE-Precast Gels (Invitrogen), 90V for 90 minutes).

#### 5.4.4. RECOMBINANT SpCas9 PURIFICATION

The pET plasmid encoding (6x)His-tagged Cas9 was transformed into BL21 (DE3), Rosetta. Transformed bacterial cells were moved to a 400ml of fresh LB medium containing 50ug/ml kanamycin. Incubate the culture with shaking (200rpm) at 18°C for 24 hours. Optical density was monitored and Cas9 protein expression was induced ( $A_{550}=0.6$ ) by using 0.5mM IPTG at 18C for 24hours. After the cells were harvested by centrifugation (5000xg) for 10minutes (at 4°C), bacterial cells were resuspended with lysis buffer [20 mM Tris-HCl (pH 8.0), 400mM NaCl, 10mM b-mercaptoethanol, 1% Triton X-100, 50mg aprotinin, 50mg antipain, 50mg bestatin, 1mM PMSF (phenyl-methylsulfonyl fluoride)] (Sigma-Aldrich) and sonicated on ice. The lysate was centrifuged at 6000 rcf for 10min(4°C) and supernatant solution was mixed with 2ml of Ni-NTA slurry (Qiagen) at 4°C for 1 and half hour. The lysate/Ni-NTA mixture was loaded onto a column (Biorad) with capped bottom outlet. Loaded sample was washed multiple times with pre-made wash

buffer [20 mM Tris-HCl (pH 8.0), 400mM NaCl, 10mM b-mercaptoethanol] and (6x)His-tagged SpCas9 was eluted with Elution Buffer [20 mM Tris-HCl (pH 8.0), 400mM NaCl, 10mM b-mercaptoethanol, 200mM Imidazole] . Finally, buffer containing eluted SpCas9 protein was changed to storage buffer [10mM HEPES-KOH (pH 7.5), 250mM KCl, 1mM MgCl<sub>2</sub>, 0.1mM EDTA, 7mM b-mercaptoethanol and 20% glycerol] by using centrifugal filter (Amicon Ultra 100K). The purified SpCas9 protein was frozen with liquid nitrogen and stored at -80°C.

#### 5.4.5. BIOTINYLATION OF THE RECOMBINANT SpCas9

The process of linking biotin to the recombinant protein was carried out in-vitro and proceeded during the process of protein purification. After loading the SpCas9 over-expressed bacterial lysate and Ni-NTA mixture onto a column (Biorad), mixed sample was washed multiple times with wash buffer [20 mM Tris-HCl (pH 8.0), 400mM NaCl]. Then we added 10-fold molar excess of maleimide-biotin (Sigma-Aldrich) to SpCas9 solution and incubate for overnight at 4°C (mix gently with rotator). To get rid of unbound maleimide-biotin chemicals, mixed sample was washed sufficiently with wash buffer [20 mM Tris-HCl (pH 8.0), 400mM NaCl]. Finally, biotinylated SpCas9 protein was eluted with elution Buffer [20 mM Tris-HCl (pH 8.0), 400mM NaCl, 200mM Imidazole], then the protein concentration was measured by spectrophotometer (Nanodrop 2000, Thermo Fisher Scientific). Eluted Sp-Cas9 protein was further purified with size exclusion chromatography. The biotinylation degree of the wild-type SpCas9 protein was calculated with commercial kit (Pierce) and it reached about 100% for two Cysteine sites (Cys80/Cys574). Biotinylated SpCas9 protein was stored in storage buffer [10mM HEPES-KOH (pH 7.5), 250mM KCl, 1mM MgCl<sub>2</sub>, 0.1mM EDTA, 7mM b-mercaptoethanol and 20% glycerol] and purified protein was frozen in liquid nitrogen and store at -80°C.

#### 5.4.6. SINGLE-MOLECULE TWO-COLOR FLUORESCENCE

Single-molecule fluorescence measurements were performed with a prism-type total internal reflection fluorescence microscope. 0.1mg/ml Streptavidin was added to a polyethylene glycol-coated quartz surface and incubated for 2 minutes before being washed with T50 (10 mM Tris-HCl (pH 8.0), 50 mM NaCl). Biotinylated Cas9 was pre-incubated with sgRNA and Cy5-labeled inhibitor RNA (ratio 1:2:10) at 37 degrees for 20 minutes in NEB buffer 3 (100 mM NaCl, 50 mM Tris-HCl, 10 mM MgCl<sub>2</sub>, 1 mM DTT) and then

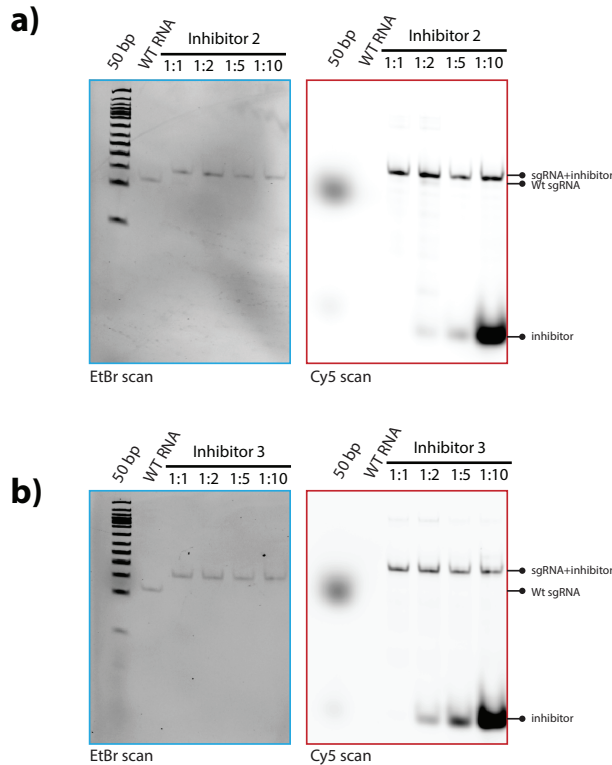
added to the chamber containing Streptavidin. After 2 minutes of incubation unbound Cas9 and RNA molecules were washed away with an imaging buffer (50mM HEPES-NaOH [pH7.5], 150mM NaCl, 2mM MgCl<sub>2</sub>, 1% glucose (Dextrose monohydrate), 1mM Trolox (2.5mg/10ml), 1mg/ml glucose oxidase [Sigma], 170ug/ml catalase [Merck]). 8 nM Cy3 labeled DNA substrate in imaging buffer was added to the channel. For the experiments with the trigger, 400nM of unlabeled trigger was added in addition to 8nM of Cy-3 labeled DNA. Simultaneous laser excitation (532 nm and 633nm) was used. Following the addition of DNA, 15 snapshots of different fields of view consisting of about 10 frames were collected every 10 minutes for a total of one hour. Fluorescence signals of Cy3 and Cy5 were collected through a 60× water immersion objective (UplanSApo, Olympus) with an inverted microscope (IX73, Olympus). The 532 nm laser scattering was blocked out by a 532 nm long pass filter (LPD01-532RU-25, Semrock). The Cy3 and Cy5 signals were separated with a dichroic mirror (635 dcxr, Chroma) and imaged using an EM-CCD camera (iXon Ultra, DU-897U-CS0-#BV, Andor Technology).

#### 5.4.7. DATA ACQUISITION AND ANALYSIS

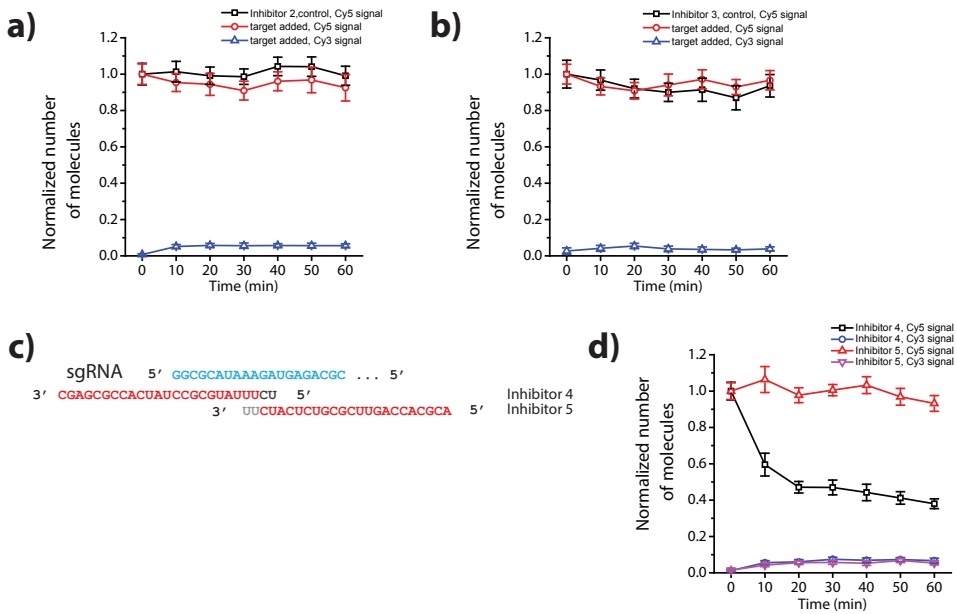
Using a custom-made program written in Visual C++ (Microsoft), a series of CCD images of time resolution 0.1s was recorded. The time traces were extracted from the CCD image series using IDL (ITT Visual Information Solution) employing an algorithm that looked for fluorescence spots with a defined Gaussian profile and with signals above the average of the background signals. Colocalization between Cy3 and Cy5 signals was carried out with a custom-made mapping algorithm written in IDL. The extracted time traces were processed using Matlab (MathWorks) and Origin (Origin Lab).

# 5.5. SUPPLEMENTARY INFORMATION

## 5.5.1. SUPPLEMENTARY FIGURES



**Figure S5.1. Inhibitors efficiently hybridize with sgRNA** a) Electrophoretic mobility shift assay resolved using 10% PAGE showing that inhibitor 2 efficiently hybridizes with sgRNA. b) Electrophoretic mobility shift assay resolved using 10% PAGE showing that inhibitor 3 efficiently hybridizes with sgRNA



**Figure S5.2. Behavior of Cas9:sgRNA:inhibitor complexes over 60 minutes.** a) a scatter plot showing the number of molecules on the surface over time for inhibitor 2. b) a scatter plot showing the number of molecules on the surface for inhibitor 3. In both cases The number of molecules was normalized by the number of molecules on the surface at time= 0. The average was taken over 15 snapshots over different fields of view. The error bars represent standard deviation. The data is the average of three independent experiments taken over three different days. c) Schematic showing Inhibitors 4 and 5 d) a scatter plot showing the number of molecules on the surface after the addition of the trigger molecule in combination with the target for Inhibitors 4 and 5. the number of molecules was normalized by the number of molecules on the surface at time= 0. The average was taken over 15 snapshots over different fields of view. Inhibitor molecules were counted as fluorescent spots in the Cy5 channel and target molecules as fluorescent spots in the Cy3 channel. The error bars represent standard deviation. The data is the average of three independent experiments taken over three different days.

5.5.2. SUPPLEMENTARY TABLES

Table S5.1. RNA sequences Used

sgRNA	5 'GGCGCAUAAAGAUGAGACGC GUUUUAGAGCUAGAAAUAGCAAGUAAAAUAAGGCUAGUCCG UUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU 3 '
INHIBITOR 1	5 ' UUUUUGCGCCUAUGCGCCUGUUGCC 3 '- <b>Cy5</b>
INHIBITOR 2	5 ' GCGUCUCAUCUU 3 '- <b>Cy5</b>
INHIBITOR 3	5 ' ACGCACCAGUUCGCGUCUCAUCUU 3 '- <b>Cy5</b>
INHIBITOR 4	5 ' UUUUUGCGCCUGUUGCC 3 '- <b>Cy5</b>
INHIBITOR 5	5 ' ACGCACCAGUUCGCGUCUCAUC 3 '- <b>Cy5</b>

Table S5.2 DNA Sequences used

PCR TEMPLATE FWD	5 ' TAATACGACTCACTATAGGTACGGTTATCCACAGAATCAGTT TTAGAGCTAGAAATAGCAAGTTAAAATAAGG 3 '
PCR TEM- PLATE REV	5 ' AAAAAAAGCACCGACTCGGTGCCAC 3 '
NON-TAR- GET STRAND	5' CCAGCTGTCTGCACAGGAGAAATCCCTGCT <b>GGCGCATAAAGATGAGACGC TGG</b> AGTACAAACGCCAGCTGGC TGCACTTGGCGACAAGGTTACGTATCAG 3 '
TARGET STRAND	5' CTGATACGTAACCTTGTGCGCAAGTGCAGCCAGCTGGCGTT TGTACT <b>CCAG CGTCTCATCTTTATGCGCC</b> AGCAGGGATTTC TCCTGTGCAGACAGCTGG 3 '- <b>Cy3</b>
TRIGGER 1	5' GCTCGCGGTGATAGGCGCATAAACGACTCCCATCCCGCCC AACCCGACT <b>CCGCTCGCGGTGATAGGCGCATAAA</b> 3 '
TRIGGER 3	5' AAGATGAGACGCGAACTGGTGCCTCGACTCCCATCCCGCC CAACCCGACTCC <b>AAGATGAGACGCGAACTGGTGCCT</b> 3 '

Bold letters shor the target site, red letters indicate the PAM. Blue regions in the triggers indicate the complementary region to inhibitor

## REFERENCES

1. Barrangou, R., et al., *CRISPR provides acquired resistance against viruses in prokaryotes*. Science, 2007. **315**(5819): p. 1709-12.
2. Marraffini, L.A., *CRISPR-Cas immunity in prokaryotes*. Nature, 2015. **526**(7571): p. 55-61.
3. Makarova, K.S., et al., *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*. Biol Direct, 2006. **1**: p. 7.
4. Mohanraju, P., et al., *Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems*. Science, 2016. **353**(6299): p. aad5147.
5. Wiedenheft, B., S.H. Sternberg, and J.A. Doudna, *RNA-guided genetic silencing systems in bacteria and archaea*. Nature, 2012. **482**(7385): p. 331-8.
6. Jinek, M., et al., *A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity*. Science, 2012. **337**(6096): p. 816-21.
7. Gasiunas, G., et al., *Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria*. Proc Natl Acad Sci U S A, 2012. **109**(39): p. E2579-86.
8. Sapranasauskas, R., et al., *The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli*. Nucleic Acids Res, 2011. **39**(21): p. 9275-82.
9. Doudna, J.A. and E. Charpentier, *Genome editing. The new frontier of genome engineering with CRISPR-Cas9*. Science, 2014. **346**(6213): p. 1258096.
10. Yang, H., et al., *One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering*. Cell, 2013. **154**(6): p. 1370-9.
11. Chew, W.L., et al., *A multifunctional AAV-CRISPR-Cas9 and its host response*. Nat Methods, 2016. **13**(10): p. 868-74.
12. Shalem, O., et al., *Genome-scale CRISPR-Cas9 knockout screening in human cells*. Science, 2014. **343**(6166): p. 84-87.
13. Heckl, D., et al., *Generation of mouse models of myeloid malignancy with combinatorial genetic lesions using CRISPR-Cas9 genome editing*. Nat Biotechnol, 2014. **32**(9): p. 941-6.
14. Paquet, D., et al., *Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9*. Nature, 2016. **533**(7601): p. 125-9.
15. Wang, T., et al., *Genetic screens in human cells using the CRISPR-Cas9 system*. Science, 2014. **343**(6166): p. 80-4.
16. Zhao, N., et al., *Combinatorial CRISPR-Cas9 and RNA Interference Attack on HIV-1 DNA and RNA Can Lead to Cross-Resistance*. Antimicrob Agents Chemother, 2017. **61**(12).
17. Kocak, D.D., et al., *Increasing the specificity of CRISPR systems with engineered RNA secondary structures*. Nat Biotechnol, 2019. **37**(6): p. 657-666.
18. Nishimasu, H., et al., *Crystal structure of Cas9 in complex with guide RNA and target DNA*. Cell, 2014. **156**(5): p. 935-49.
19. Siu, K.H. and W. Chen, *Riboregulated toehold-gated gRNA for programmable CRISPR-Cas9 function*. Nat Chem Biol, 2019. **15**(3): p. 217-220.





# 6

## Overview of different Cas9 variants and concluding remarks

The work described in this thesis focuses on the target search and recognition (or lack thereof) of the Cas9 protein from *Streptococcus pyogenes*. Due the enormous commercial interest in this protein, the Cas9 research is becoming its own field. Its applications in genome editing have be a separate topic from the start, however, the search for new Cas9 orthologs or engineering of Cas9 mutant proteins and characterizing them is diverging from the field of CRISPR biology. It is not surprising – as opposed to other CRISPR systems, the interest in Cas9 comes from its applicability in genome engineering. Therefore, the focus here lies on coming up with new Cas9 variants that can complement or outperform SpCas9. The process typically involves the identification of such proteins (or engineering of mutants), characterization of their PAM and guide requirements and verification of activity in eukaryotic cells. However, little to no research is done on the molecular mechanism of these proteins – PAM and target search, target recognition, and any structural rearrangements or proofreading mechanisms remain a mystery. As these proteins will likely be applied in therapeutics in the future, we face the dangers of using something we do not fully understand in situations where any mistake can be detrimental. Therefore, this final chapter is a discussion of some of the work that has been done on developing or identifying new Cas9 proteins and the considerations concerning their PAM and/or target search based on their structural data.

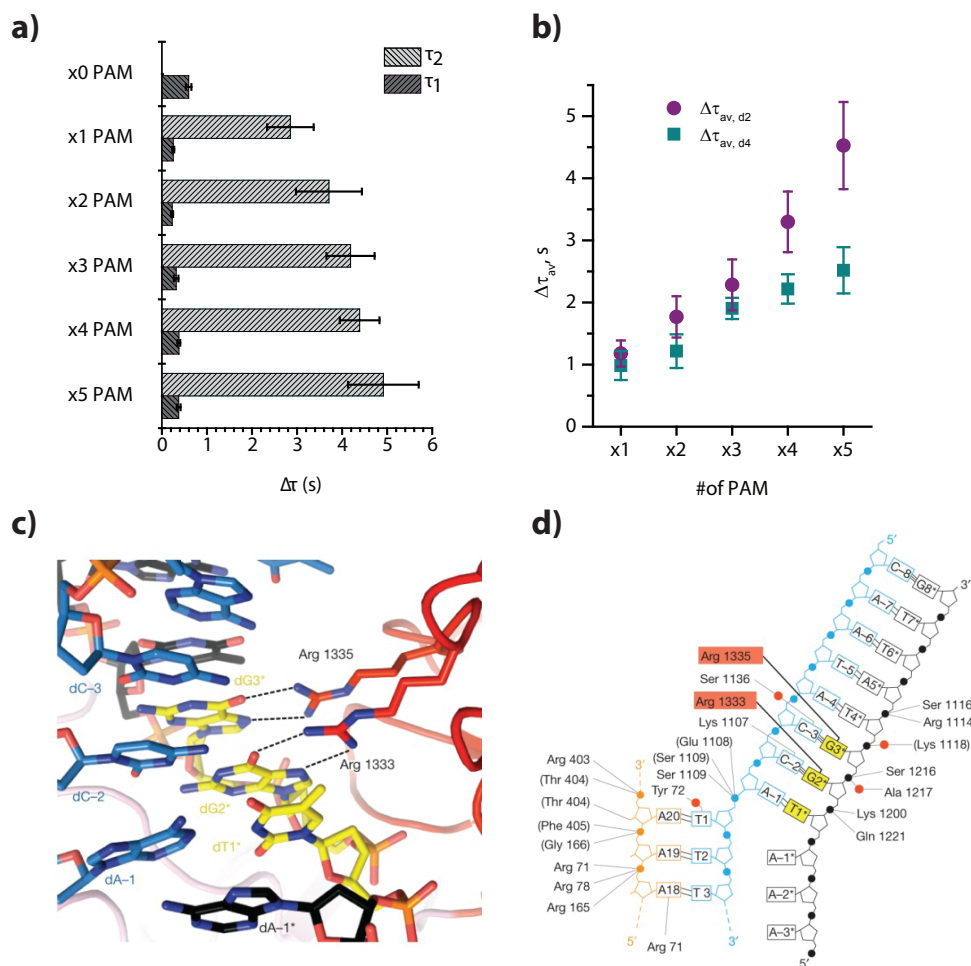
The work described in this thesis mostly focused on Cas9-target interactions, with the target search being the main focus. For a biophysicist, the importance of the knowledge of the target search processes is obvious – it can explain different levels of activity, help choose the best target sites or create models, which would predict the behavior of Cas9. Unfortunately, with the rapid growth of the Cas9 field the mechanistic studies have often stopped at solving the crystal structures of nucleic acid-bound Cas9 proteins. In their 2015 publication describing the structure of *Staphylococcus aureus* Cas9, Nishimasu et al argue that „Additional crystal structures of Cas9 orthologs bound to nucleic acids are critical to understand the potential mechanistic and structural conservations underlying RNA-guided DNA targeting by cas9“[1]. While structural data do reveal a wealth of information about the overall function of the protein, it fails to predict the accurate target search mechanism, as well as any mechanism of kinetic proofreading. While the target search mechanisms are remarkably similar in many nucleic acid guided proteins, the target search mechanism of Cas9 and other CRISPR proteins is complicated by the PAM requirement. Therefore, it is important to consider both - structural and single-molecule data to fully understand the underlying core principles of Cas9 PAM search. Thus, in the upcoming sections I will provide considerations between the structure of SpCas9 in relation to single-molecule observations described in this thesis and an overview of the PAM recognition mechanisms of different Cas9 variants as described in structural studies together with some of my speculations on implications for PAM search, based on our observations in single-molecule experiments.

## 6.1. CONSIDERATIONS OF SpCas9 STRUCTURE IN RELATION TO SINGLE-MOLECULE OBSERVATIONS

In chapter 3 of this thesis we explored the target and PAM search mechanisms of SpCas9. We have observed, that Cas9 has two distinct PAM binding modes, which are characterized by distinct dwelltimes of  $\sim 0.2\text{s}$  ( $\tau_1$ ) and  $\sim 2.5\text{s}$  ( $\tau_2$ ) [2]. In addition, it was found that while the first dwelltime stays the same regardless of how many PAM sequences are in succession, the second dwelltime increases with increasing number of neighbouring PAM sites (Figure 6.1 a).

Cas9 structural studies show that the di-nucleotide 5'-GG-3' PAM of Sp-Cas9 is read out from the major groove side by base-specific hydrogen bonding interactions with the Arg1333 and Arg1335 residues (Figure 6.1 c, d) [3]. The PAM-interacting domain interacts with the minor groove of the DNA duplex and orients the target strand for pairing with the guide RNA. The

Lys1107 residue interacts with the dC2 nucleotide (complementary to the dG2\* in the PAM) (Figure 6.1 c). The residues downstream of Lys1107 interact with the +1 phosphate of the target DNA on the target strand (the phosphate of the first nucleobase complementary to guide RNA). Interactions with the Lys1107-Ser1109 loop rotate the +1 phosphate which coincides with the distortion that allows the first nucleobase of the guide RNA to pair with the first nucleobase of the target DNA strand. As noted by Anders and colleagues, the interactions between the +1 phosphate and the loop are PAM dependent [3]. Therefore, it is most likely that the first measured dwelltime,



**Figure 6.1. PAM search and recognition by SpCas9.** a) bar plot showing the dwelltimes of Cas9 binding to constructs with different number of adjacent PAMs [2] b) Scatter plot showing the average dwelltime ( $\Delta\tau_{av}$ ) values for the cases of 2 (d2) and 4 (d4) nucleotide separation between the PAM sites. [2] c) detailed view of the major groove. Sequence-specific hydrogen-bonding interactions with the PAM are indicated with dashed lines. [3] d) schematic of Cas9 interactions with the PAM duplex. Water molecules are represented by red circles [3]

$\Delta\tau_1$ , is governed by the interactions between the Arg1335 and Arg1333 with the GG nucleobases and the interaction between the +1 phosphate and the phosphate lock loop. As the first nucleotide after the PAM in our experiments was not complementary to the guide RNA, no basepairing can occur after the distortion created by the phosphate-loop interactions, and, being energetically unfavourable, the DNA is quickly rejected by Cas9 as an off-target, corresponding to the short-lived interactions that have been observed.

In addition to the base-specific interactions between the arginine residues and the G nucleobases, the sugar-phosphate backbone of the non-target DNA strand makes numerous ionic and hydrogen-bond contacts with the protein (Figure 6.1 d). Furthermore, Cas9 does not interact with the target strand complement of the 5'-GG-3' PAM sequence. Considering this information, the longer dwelltime we have observed in our single-molecule experiments could potentially be explained as the combination of the base-specific and base-non-specific interactions and PAM-recognition intermediates, which were suggested in the structural studies [3]. The base-non-specific interactions together with the lack of interactions between the Cas9 protein and the target strand suggest that after abortive binding where no target adjacent to the PAM is found, the non-specific interactions can be strong enough to keep the DNA weakly associated with the protein while it may locally diffuse around the PAM in a 1-dimensional fashion. If another PAM is found, the base-specific interactions together with the phosphate lock loop and +1 phosphate interactions take place again before an off-target is rejected and the DNA can be expelled, or another round of local diffusion governed by weak base-non-specific interactions takes place. As Cas9 is already interacting with the DNA, it is possible that such interactions would be shorter than initial binding after the PAM was found by 3D diffusion. Such an iterative mechanism could also explain why we observe the dwelltime to go up with the increasing number of PAMs close to each other as the protein encounters more PAM „pit stops“ during the local 1-dimensional diffusion before it dissociates. It also provides a possible explanation why an increased distance between the PAMs does not affect  $\Delta\tau_2$ , but decreases  $\Delta\tau_{av}$  (Figure 6.1 b). If indeed  $\Delta\tau_1$  corresponds to base-specific interactions, when the PAM has been found by 3-dimensional collisions, then regardless of the number of PAMs these interactions stay the same. The local diffusion governed by weak interactions with the DNA backbone is likely very fast, thus if Cas9 is to undergo local diffusion, the total dwelltime is likely governed by the number of PAMs rather than the distance. As the interactions are weak, however, increasing distance increases the chance that Cas9 dissociates before it finds another GG sequence for base-specific interactions and thus the short-lived, base-specific interactions governed by 3-dimensional diffusion start dominating with increasing distance between the PAM sites, contributing to the

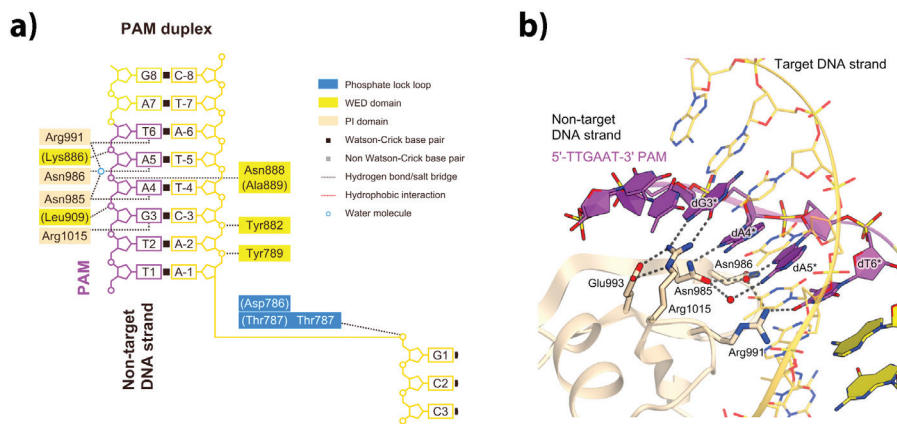
decrease in the average dwelltime. Of course, the above are speculations and extensive experimental work with multiple Cas9 mutants would be needed to confirm which protein-DNA interactions contribute to which observable behaviour. Nevertheless, it suggests that different binding modes could be governed by different types of interactions and help anticipate what could be expected from other Cas9 proteins, for which the crystal structure has been solved, which will be discussed in the following sections.

## 6.2. PAM SPECIFICITY AND RECOGNITION OF CAS9 ORTHOLOGS

Despite its popularity, SpCas9 has several drawbacks which limit its use in eukaryotic cells, the main ones being the stringent PAM requirement and its size. Although the -GG- sequence is estimated to occur every 8 basepairs in the human genome [4], it limits the precision with which a cut site can be chosen. As for the size, SpCas9 consists of 1368 amino acids, which makes it difficult to deliver it to cells using strategies such as lentiviral or adenoviral delivery. As a result, a lot of effort has been put into discovering new Cas9 orthologs which target different PAM sequences, are smaller, or both. In this section I will provide a brief overview of the Cas9 proteins from *Staphylococcus aureus* (SaCas9), *Francisella novicida* (FnCas9), and *Campilobacter jejuni* (CjCas9), and their PAM recognition mechanisms and implications for PAM search.

### 6.2.1. STAPHYLOCOCCUS AUREUS CAS9

The Cas9 protein from *Staphylococcus aureus* has been shown to function in eukaryotic cells and is smaller than SpCas9 consisting of only 1053 amino acids. Like SpCas9, SaCas9 belongs to the type II-A CRISPR family. It recognizes a 5'-NNGRRT-3' PAM (R represents a purine - A or G)[1, 5]. The PAM-interacting (PI) domain of SaCas9 can be divided into a topoisomerase-homology (TOPO) and C-terminal domains as in SpCas9 and has a similar core fold to that of SpCas9 [1]. As in SpCas9, SaCas9 recognizes the PAM from the major groove side. Furthermore, the third G in the 5'-NNGRRT-3'PAM is recognized by Arg1015 whereas in SpCas9 the third G is recognized by Arg1335 [3]. The distinct PAM specificities of the two proteins arise from different amino acid residues interacting with the PAM nucleotides: the 4th and 5th purine in the PAM form direct and water-mediated hydrogen bonds with Asn985 and Asn985/Asn986/Arg991 in SaCas9 (Figure 6.2 a, b). As in SpCas9, the phosphate backbone is recognized from the minor groove side



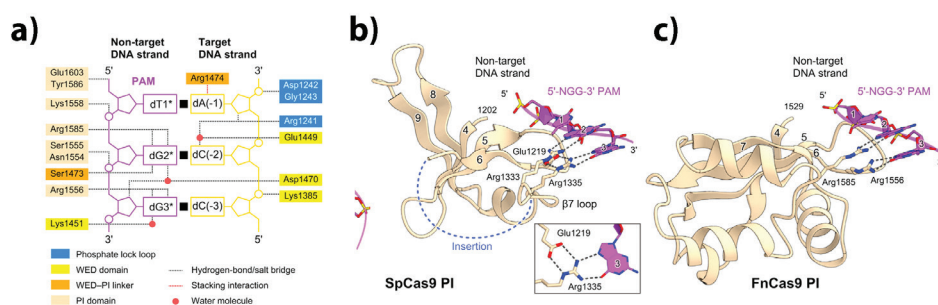
**Figure 6.2. Interactions between SaCas9 and the PAM duplex.** a) Schematic of Cas9 and PAM interactions. water-mediated hydrogen bonding interactions are omitted. [1] b) recognition of the 5'-TTGAAT-3' PAM. PAM sequence is highlighted in purple. Water molecules are shown as red spheres and hydrogen bonds are shown as dashed lines [1]

by base-non-specific interactions. Remarkably, the +1 phosphate in the target strand hydrogen bonds with Asp786 and Thr787 of SaCas9 which also results in the rotation of the +1 phosphate, which aids DNA unwinding (Figure 6.2 a). It is evident that distinct PAM specificities of SpCas9 and SaCas9 come from different amino acid sequences in the PI domain, but the mode of PAM recognition is remarkably similar between the two proteins. Therefore, I would expect that PAM-search mechanism of SaCas9 is very similar to that of SpCas9, with a possibility that a PAM-specific interaction could be slightly stronger in SaCas9 due to the larger number of interactions with the amino acid residues and the PAM. However, since SaCas9 also weakly interacts with the phosphate backbone of the target strand complement of the PAM, it is likely that the lateral diffusion mode would be different to that of SpCas9 as the non-specific interactions might be enough to keep the protein bound to DNA for longer, which would in turn allow for more local 1-dimensional diffusion before dissociation.

### 6.2.3. FRANCISELLA NOVICIDA Cas9

The Cas9 protein from *Francisella novicida* is one of the largest Cas9 orthologs consisting of 1629 amino acid residues [6, 7], from the type II-B CRISPR family. It consists of seven domains with the WED domain being most dis-

tinct compared to other Cas9 orthologs [8]. The PAM-interacting (PI) domain of FnCas9 has little sequence homology with SaCas9 and SpCas9, however, it adopts a similar core fold and recognizes a 5'-NGG-3' PAM. A 5'-NGA-3' is also permitted, however, protospacers with such a PAM are cleaved with a lower efficiency. Both SpCas9 and FnCas9 recognize the 5'-NGG-3' PAM with a pair of Arginine residues – Arg1333/Arg1335 and Arg1585/Arg1556 respectively (Figure 6.3 a, b, c). In SpCas9 the side chain of Arg1335 is anchored by a salt bridge with Glu1219, contributing to a specific recognition of the third G (Figure 6.3 b). However, in the case of FnCas9, the Arg1556 side chain does not form such contacts with proximal residues, hence the recognition of the third A in the PAM. In addition, the residues from the WED domain interact with the dC(-2) nucleotide of the target strand and the phosphate backbone of the non-target strand (Figure 6.3 a). FnCas9 phosphate lock loop also interacts with the +1 phosphate of the target strand, indicating that DNA unwinding mechanism is conserved between different Cas9 orthologs (Figure 6.3 a). Since FnCas9 forms only a single hydrogen bond with the third G in the PAM and interacts with the phosphate backbone of both strands, it is likely that local 1-D diffusion during PAM search also takes place in this system. Overall, the base-specific PAM-interacting mode might be slightly weaker than that of SpCas9 due to the lack of interactions between the Arg1556 side chain and other proximal amino acid residues, which confer the specificity towards a second G in the PAM of SpCas9. Even though the PAM is recognized in a different manner, I expect that the overall PAM search mechanism will be quite similar to that SpCas9.



**Figure 6.3. PAM recognition by FnCas9.** a) Schematic of the interactions between FnCas9 and the PAM. Water-mediated hydrogen bonds are omitted [8] b) Comparison of the PAM-interacting domains between b) SpCas9 and c) FnCas9

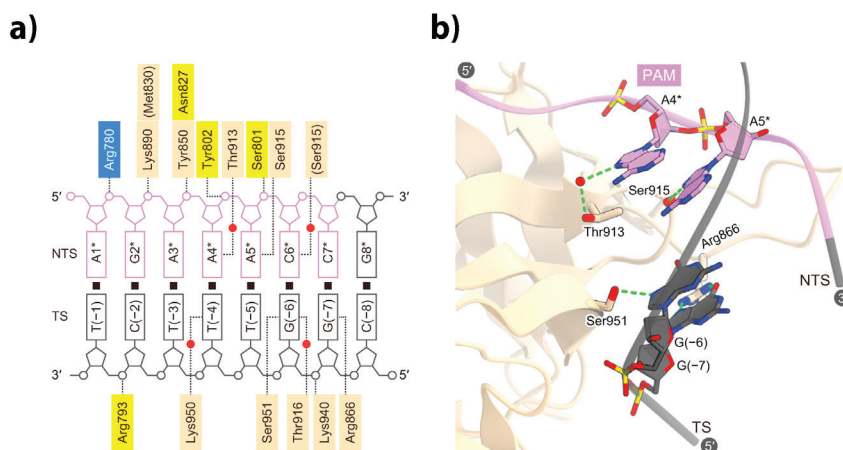


### 6.2.2. CAMPILOBACTER JEJUNI Cas9

CjCas9 is one of the smallest Cas9 orthologues, consisting of only 984 amino acids, from the type II-C CRISPR family[9]. It recognizes a 7bp PAM sequence 5'-NNNVRYM-3' (V is A/G/C; R is A/G; Y is T/C; M is A/C) and has a preference for T and C at positions 6 and 7 respectively. Furthermore, the crystal structure of CjCas9 has revealed that, unlike SpCas9, SaCas9 and FnCas9, it forms base-specific interactions with both the target and the non-target strand nucleotides in the PAM (Figure 6.4 a, b) [10]. Thr913 interacts with the V4\*, Ser801 interacts with R5\* and Ser915 interacts with Y6\* on the non-target strand and Lys950 interacts with the complement of V4\*, Ser951 and Thr916 interact with the complement of Y6\* and Arg866 interacts with the complement of M7\* (Figure 6.4 a, b). The phosphate backbone of both strands in the PAM also interacts with the protein in a base-non-specific manner (Figure 6.4 a). It is evident that CjCas9 has a very distinct PAM recognition mode compared to the three previously described orthologs. Another type II-C Cas9, the NmeCas9, is also known to interact with both strands of the PAM as CjCas9, indicating that type II-C PAM recognition is mechanistically distinct from that of type II-A and type II-B [11]. The promiscuity of the PAM together with the recognition of both DNA strands implies that PAM and subsequently target search of CjCas9 may be slower than that of SaCas9, SpCas9 and FnCas9. The extensive base-specific and base-independent interactions with the PAM suggest stronger binding to a PAM than in other orthologs. However, the non-stringent base requirement means that many more PAMs are permissible. The combination of the two effects might lead to CjCas9 taking a longer time to find its target as it would need to sample more PAM sites to which it associates more strongly than other Cas9 proteins. CjCas9 has been shown to work in vivo in eukaryotic cells, however is less efficient than SpCas9[10, 12]. While this could be due to many reasons, a distinct PAM recognition mode and a probably different PAM search mechanism are likely to play a role. Therefore, the CjCas9 target search mechanism should be more extensively studied using biophysical techniques before the protein is used in therapeutic applications.

## 6.3. ENGINEERED SpCas9 VARIANTS

Different Cas9 orthologs require different guide RNAs. This, together with the difference in size makes it difficult to use Cas9 to target several genes at once. As a result, a lot of effort has gone into producing new SpCas9 variants with altered PAM specificities, either by directed evolution or structure-guided mutagenesis.



**Figure 6.4. PAM recognition by CjCas9.** a) schematic of the PAM recognition by CjCas9. Hydrogen bonds are depicted as dashed lines and water molecules as red spheres. [10]. b) Recognition of the 5'-AGAAACG-3' PAM. Water molecules are shown as red spheres and hydrogen bonds as dashed lines [10].

One of the first attempts was to modify SpCas9 to recognize different PAMs using bacterial selection-based directed evolution [13]. The experiments yielded three mutants with altered PAM specificities – the VQR (D1135V/R1335Q/T1337R) mutant recognizing 5'-NGAN-3' and 5'-NGCG-3' PAMs, EQR (D1335E/ R1335Q/T1337R) mutant recognizing 5'-NGAG-3' PAM and a quadruple mutant VRER (D1335V/G1218R/R1335E/T1337R) recognizing 5'-NGCG-3' PAM. All mutations are spatially oriented near the PAM. Structural studies revealed that the non-canonical PAMs are recognized by an induced fit mechanism [14]. Binding of the non-canonical PAMs is achieved by structural remodelling of the PAM. The amino acid substitutions induce and accommodate structural changes in the DNA and provide base-specific contacts with the PAM. However, the DNA distortion is energetically unfavourable, contributing to lower cleavage efficiency[13]. Using this information, an SpCas9 variant recognizing an NAAG PAM has been engineered, which also showed decreased cleavage activity [14]. As no other parts of the protein have been changed, it shows that sufficiently strong interactions with the PAM are needed for efficient DNA cleavage. Amino acid substitutions may have affected not only PAM recognition but also PAM search, as a less efficient search mechanism would also contribute to the decrease in cleavage efficiency.

Another SpCas9 variant with an expanded PAM specificity has been evolved using phage-assisted continuous evolution [15]. A resulting protein named xCas9 bearing 7 mutations (A262T/R324L/S409I/E408K/E543D/

M694I/E1219V) recognizes 5'-NG3-3' PAMs. Since the whole protein sequence has been subjected to evolution, however, it is possible that there are more mutations, which may change the protein behaviour in unexpected ways. Therefore, this protein, its target search, cleavage and proofreading mechanism should be extensively studied using biophysical techniques, before it can be safely applied in therapeutics.

A single-nucleotide PAM is most desired as it allows for the greatest flexibility in choosing a target site. Therefore, another SpCas9 variant recognizing 5'-NG-3' PAMs has been engineered using structure-guided rational engineering [16]. The base-specific interactions between the Arg1335 and the third G in the PAM were removed and compensated by base-non-specific interactions. This yielded a protein with 7 mutations which recognizes a relaxed 5'-NG-3' PAM. Its cleavage activity is lower than that of WT SpCas9, indicating again that a sufficiently strong interaction with the PAM is needed for efficient DNA cleavage. As most of these interactions are base-non-specific, it is possible that this protein will undergo more extensive lateral diffusion than WT SpCas9. As we have shown that this diffusion between PAM sites contributes to delayed on-target binding, it is possible that this could be a factor in the decreased cleavage efficiency of Cas9-NG.

## 6.4. CONCLUDING REMARKS

The huge commercial and societal interest in the Cas9 protein has pushed for the rapid development of this field in order to overcome the inherent limitations of this protein. The stringent PAM requirement is one of the limitations of SpCas9 as it limits the number of target sites and the precise positioning of the cut site. Therefore, a lot of work has been done in trying to overcome this by exploring different Cas9 orthologs or engineering SpCas9 variants with different PAM specificities. Structural studies of Cas9 orthologs show that even though the PAM-interacting domain in all of them has a similar core fold, the recognition mode is slightly different in SaCas9 and FnCas9 compared to SpCas9 and very distinct in CjCas9, which recognizes both DNA strands in the PAM. All these differences may contribute to a different PAM search mechanism which in turn can lead to different activity levels.

SpCas9 variants with different PAM specificities also have different modes of PAM recognition compared to WT SpCas9 as in most of them, the lack of base-specific interactions has been compensated by introducing base-non-specific interactions between the protein and the DNA backbone. This is also likely to affect PAM search and contribute to the demonstrated lower cleavage efficiency. While PAM search may seem to be a minor fac-

tor in determining the cleavage efficiency, a study on Cas9-Cas9 chimeras shows that the cleavage activity of a PAM-attenuated Cas9 is restored when it is fused with another Cas9 protein (cleavage-inactivated) which guides it to the target site thus effectively performing the PAM search and increasing the effective concentration of the PAM-attenuated Cas9 at the cut site [17]. This implies that inefficient PAM search indeed contributes to low activity levels. Taken together, these studies show that the newly found or engineered Cas9 proteins should be studied in-depth in a systematic manner using biophysical tools to fully elucidate the target search, cleavage and proofreading mechanisms before such proteins can even be considered for therapeutic applications.

## REFERENCES

1. Nishimasu, H., et al., *Crystal Structure of Staphylococcus aureus Cas9*. Cell, 2015. **162**(5): p. 1113-26.
2. Globy, V., et al., *CRISPR/Cas9 searches for a protospacer adjacent motif by lateral diffusion*. EMBO J, 2019. **38**(4).
3. Anders, C., et al., *Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease*. Nature, 2014. **513**(7519): p. 569-73.
4. Karvelis, T., et al., *Rapid characterization of CRISPR-Cas9 protospacer adjacent motif sequence elements*. Genome Biol, 2015. **16**: p. 253.
5. Ran, F.A., et al., *In vivo genome editing using Staphylococcus aureus Cas9*. Nature, 2015. **520**(7546): p. 186-91.
6. Hsu, P.D., E.S. Lander, and F. Zhang, *Development and applications of CRISPR-Cas9 for genome engineering*. Cell, 2014. **157**(6): p. 1262-78.
7. Chylinski, K., A. Le Rhun, and E. Charpentier, *The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems*. RNA Biol, 2013. **10**(5): p. 726-37.
8. Hirano, H., et al., *Structure and Engineering of Francisella novicida Cas9*. Cell, 2016. **164**(5): p. 950-61.
9. Fonfara, I., et al., *Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems*. Nucleic Acids Res, 2014. **42**(4): p. 2577-90.
10. Yamada, M., et al., *Crystal Structure of the Minimal Cas9 from Campylobacter jejuni Reveals the Molecular Diversity in the CRISPR-Cas9 Systems*. Mol Cell, 2017. **65**(6): p. 1109-1121 e3.
11. Zhang, Y., et al., *DNase H Activity of Neisseria meningitidis Cas9*. Mol Cell, 2015. **60**(2): p. 242-55.
12. Kim, E., et al., *In vivo genome editing with a small Cas9 orthologue derived from Campylobacter jejuni*. Nat Commun, 2017. **8**: p. 14500.
13. Kleinstiver, B.P., et al., *Engineered CRISPR-Cas9 nucleases with altered PAM specificities*. Nature, 2015. **523**(7561): p. 481-5.
14. Anders, C., K. Bargsten, and M. Jinek, *Structural Plasticity of PAM Recognition by Engineered Variants of the RNA-Guided Endonuclease Cas9*. Mol Cell, 2016. **61**(6): p. 895-902.
15. Hu, J.H., et al., *Evolved Cas9 variants with broad PAM compatibility and high DNA specificity*. Nature, 2018. **556**(7699): p. 57-63.
16. Nishimasu, H., et al., *Engineered CRISPR-Cas9 nuclease with expanded targeting space*. Science, 2018. **361**(6408): p. 1259-1262.
17. Ma, D., et al., *Engineer chimeric Cas9 to expand PAM recognition based on evolutionary information*. Nat Commun, 2019. **10**(1): p. 560.





# Summary

The most commonly known relationship between the macromolecules of life – DNA, RNA and protein – is the central dogma which states that information flows from DNA to RNA through transcription and from RNA to protein through translation. The discovery of non-coding RNAs has altered and expanded this picture by revealing many more functions of RNA as well as different ways nucleic acids and proteins interact. For example, tRNAs act as a link between mRNA and ribosomes in translation, ribosomal RNA together with proteins make up the main translation machinery, riboswitches enable mRNA to regulate its own activity, miRNA and siRNA regulate gene expression through RNA interference. A notable recent addition to this list is the adaptive prokaryotic CRSIPR immune system where RNA molecules guide proteins to destroy the invading viral DNA. The CRISPR systems were made famous by the discovery of the Cas9 endonuclease, a single RNA-guided protein, which can perform target search, recognition and cleavage without other proteins involved. This has made Cas9 the golden tool for gene editing, making the latter cheaper, faster and easier than it has ever been before.

The first steps in Cas9 activity are target search and recognition and thus are important for efficient cleavage activity. These processes can involve a myriad of different types of interactions between the protein, its guide RNA and target DNA and are best investigated using single-molecule techniques as they help overcome the ensemble averaging that occurs in bulk measurements. **Chapter 1** introduces the macromolecules of life, together with non-coding RNAs and single-molecule techniques, which are most commonly used in the fields of CRISPR and target search.

Despite the difference in composition and function, the same core principles govern the target search and recognition mechanisms of RNA-guided proteins. For example, weak interactions help reject off-target sites quickly, a combination of 1-dimensional and 3-dimensional diffusion can speed up target search and structural rearrangements provide a good means for kinetic proofreading. **Chapter 2** therefore provides a detailed overview and comparison of the most studied RNA-guided target search systems – Argonaute and CRISPR families, with the focus being on eukaryotic Argonaute and Cas9 proteins and the Cascade protein complex.



An overview of the current state of the art sets us up for our own study of Cas9 target search. It is known that Cas9 cleaves a 20-nucleotide target sequence, which is flanked by a 3 nucleotide PAM sequence, but how does the protein find it? DNA-curtains studies have shown that Cas9 uses only 3-dimensional collisions to find its target, but are any different search modes hidden beyond the diffraction limit? In **Chapter 3** we use single-molecule FRET to answer these questions. We show that Cas9 exhibits two modes of interaction with PAM sequences which are characterized by different dwell-times, which differ by an order of magnitude. Furthermore, the longer dwell-time increases with an increasing number of neighboring PAM sites, which is suggestive of lateral diffusion. Indeed, by placing the PAM sites further apart on the DNA strand we were able to directly capture evidence of Cas9 laterally diffusing between PAM sequences and show that the occurrence of such events goes up with an increasing number of PAM sites. In addition, we directly show that Cas9 is able to slide between neighboring target sites. However, the probability to find the target by lateral diffusion decreases exponentially with increasing separation between the targets. Finally, we investigated if the lateral diffusion between PAM sites could act as a means to speed up target binding. By placing an increasing number of these sites close to a cognate target we have observed the opposite - a moderate decrease in binding rate with an increasing number of PAMs. This result was corroborated by an asymmetrical tandem target experiment which showed that Cas9 prefers to bind a target flanked by a single PAM site.

Single-molecule FRET is a versatile technique and can be used to explore different properties of various systems. **Chapter 4** describes single-molecule FRET methods to study the Cas9 nuclease. We describe methods based on two main immobilization schemes, namely target DNA immobilization and the immobilization of the Cas9 ribonucleoprotein (RNP) complex. In addition to the experiments, sample preparation steps such as slide cleaning, labeling of the nucleic acids and Cas9 biotinylation are described. Example data are provided for each immobilization scheme in addition to recommendations for when each scheme is more useful. For example, DNA immobilization is more useful to investigate how Cas9 binding perturbs the DNA duplex. Protein immobilization, on the other hand, is more useful to investigate weak interactions as all unbound RNA molecules are washed away before adding the target and DNA is less likely to stick to the slide surface non-specifically. In addition, we provide explanations of crRNA and DNA design for each type of experiment. Finally, a summary of image processing and data analysis is provided at the end of the chapter.

Aside from the fundamental research conducted on Cas9 which helps answer many questions about the protein activity and the type II CRISPR immune system as a whole, the main focus in the field of Cas9 is the applica-

tions of Cas9 in genome engineering. So far, various factors have been found to affect Cas9 cleavage activity. To name a few, secondary structures of target DNA such as G-quadruplexes can greatly decrease Cas9 activity as well as mismatches between guide RNA and target DNA, and certain sequences have been found to be preferable by Cas9 than others. In **Chapter 5** we demonstrate that Cas9 activity can be inhibited by small RNA molecules in vitro. Using biochemical assays, we show that sequences which are complementary to the seed sequence of the guide inhibit cleavage completely even at low excess ratios whereas sequences complementary to the end region of the guide inhibit cleavage progressively with increasing ratios of inhibitor to guide. Using single-molecule fluorescence we show that these hybrids are efficiently loaded by Cas9 and prevent interactions with the DNA target. Furthermore, we use a trigger molecule complementary to the inhibitor and demonstrate that the later can be released by the Cas9 RNP complex if it is hybridized to the end of the guide, but not the seed. Despite the inhibitor being released by Cas9, it remains unable to interact with target DNA which suggests that the protein is stuck in an inactive conformation. Such inhibition can be particularly important in therapeutic applications where miRNA can potentially basepair with the guide and inhibit cleavage, or transcriptionally active genes can be targeted less efficiently. To further examine the extent of this observation, in vivo experiments are being conducted.

Due to its applications in gene editing, Cas9 research has diverged from CRISPR biology and has become its own field. A lot of effort has been directed towards identifying new Cas9 variants or creating SpCas9 mutants which are able to target different PAM sequences. In this final **Chapter 6** an overview and comparison of the crystal structures and PAM-recognition modes of different Cas9 proteins are provided. Several Cas9 orthologs are described: *Staphylococcus aureus* Cas9 (SaCas9), *Francisella novicida* Cas9 (FnCas9) and *Campilobacter jejuni* Cas9 (CjCas9). Aside from the PAM recognition, the implications of the mechanism to target search are discussed using the observations described in chapter 3 as an example. In addition, engineered Cas9 mutants such as xCas9 or Cas9-NG are discussed. Different mutation schemes and amino acid substitutions can slow down PAM search and recognition and thus decrease the overall activity of Cas9. Therefore, this chapter concludes that Cas9 orthologs and especially mutants should be studied in-depth using biophysical tools before being applied in therapeutics to avoid any unwanted effects that can arise from perturbed PAM search and recognition mechanisms.



# Samenvatting

De meest bekende relatie tussen de macromoleculen van het leven – DNA, RNA en eiwit – is het centrale dogma welke zegt dat informatie wordt overgedragen van DNA naar RNA via transcriptie en van RNA naar eiwit door middel van translatie. De ontdekking van niet-coderend RNA heeft dit beeld veranderd en uitgebreid door te laten zien dat er veel meer functies zijn van RNA en veel meer manieren voor nucleïnezuren om met eiwitten interactie aan te gaan. Voorbeelden hiervan zijn tRNAs die een schakel vormen tussen mRNA en ribosomen tijdens translatie; ribosomaal RNA dat samen met eiwitten het hoofdbestanddeel uitmaakt van het translatie mechanisme; riboswitches die mRNA in staat stellen om eigen activiteit te reguleren; en miRNA en siRNA die genexpressie reguleren door middel van RNA interferentie. Een opmerkelijke recente toevoeging aan deze lijst is het adaptieve prokaryotische CRISPR immuunsysteem waar RNA moleculen eiwitten leiden naar binnendringend viraal DNA om dit te vernietigen. De CRISPR systemen zijn bekend geraakt door de ontdekking van de Cas9 endonuclease, een enkel RNA-geleid eiwit, welke in staat is om doelwitten te lokaliseren, te herkennen en te knippen zonder hulp van andere eiwitten. Dit heeft ertoe geleid dat Cas9 als uitzonderlijk gereedschap wordt gezien voor genoombewerking, wat dit laatste goedkoper, sneller en makkelijker maakt dan ooit tevoren.

De eerste stappen in Cas9 activiteit zijn de lokalisatie en herkenning van het doelwit en zijn daarom belangrijk voor efficiënte knip-activiteit. Bij deze processen kan een groot aantal verschillende interacties tussen het eiwit, het gids-RNA en het doelwit-DNA plaatsvinden en deze kunnen het beste onderzocht worden door het gebruik van technieken die individuele moleculen bekijken, aangezien hierbij geen middeling plaatsvindt van de individuele signalen zoals in -experimenten met massa's moleculen. **Hoofdstuk 1** introduceert de macromoleculen van het leven, samen met niet-coderend RNA en enkel-molecuul technieken, welke het meest gebruikt zijn in de onderzoeksgebieden van CRISPR en doelwit-lokalisatie.

Ondanks het verschil in compositie en functie worden de doelwit-lokalisatie- en herkenningsmechanismes van RNA-geleide eiwitten door dezelfde basisprincipes gestuurd. Zwakke interacties helpen bijvoorbeeld om foutieve doelwitten snel te kunnen verwerpen/afkeuren. Ook kan een combinatie van eendimensionale en driedimensionale diffusie doelwit-lokalisatie versnellen

en structurele herschikkingen kunnen een goede manier zijn voor kinetische proeflezing. **Hoofdstuk 2** geeft een gedetailleerd overzicht en vergelijking van de meest bestudeerde RNA-geleide doelwit-lokalisatie systemen – Argonaute en CRISPR families, met de focus op eukaryotische Argonauten en Cas9 en Cascade eiwitcomplexen.

Een overzicht van recent onderzoek om onze eigen studie van Cas9 doelwit-lokalisatie te bespreken. Het is bekend dat Cas9 een 20 nucleotide doelwit-sequentie knipt, die geflankeerd is door een 3-nucleotide lange PAM sequentie, maar hoe vindt het eiwit deze sequenties? DNA-gordijn studies hebben aangetoond dat Cas9 alleen driedimensionale botsingen gebruiken om zijn doelwit te vinden, maar zijn er nog andere zoekmethodes verborgen voorbij het diffractielimiet? In **hoofdstuk 3** gebruiken wij enkel-molecuul FRET om deze vragen te beantwoorden. We laten zien dat Cas9 twee verschillende manieren van interactie met PAM sequenties heeft, deze zijn gekarakteriseerd door verschillende bindingstijden welke een ordegrootte verschillen. Daarnaast neemt de langere bindingstijd toe bij een toenemend aantal naburige PAMs, wat laterale diffusie suggereert. Door de PAM sequenties verder van elkaar te plaatsen op het DNA zijn wij in staat om direct bewijs te leveren van laterale diffusie van Cas9 tussen PAM sequenties en laten we ook zien dat deze gebeurtenissen vaker voorkomen naarmate het aantal PAM sequenties toeneemt. Daarnaast laten wij ook direct zien dat Cas9 in staat is om tussen nabijgelegen doelwit-sequenties te schuiven. De kans om de doelwit-sequentie te vinden met laterale diffusie neemt echter exponentieel af met een toenemende afstand tussen de doelwit-sequenties. Tenslotte hebben wij onderzocht of laterale diffusie tussen PAM sequenties ook gebruikt kan worden om de binding aan doelwit-sequenties te versnellen. Door het plaatsen van een toenemend aantal van deze PAM sequenties naast een doelwit hebben wij het tegenovergestelde waargenomen – een gematigde afname in bindingssnelheid met een toenemend aantal PAM sequenties. Dit resultaat is gevalideerd met een asymmetrische tandem doelwit experiment welke liet zien dat Cas9 bij voorkeur bindt aan een doelwit met een enkele nabijgelegen PAM sequentie.

Enkel-molecuul FRET is een veelzijdige methode en kan gebruikt worden om verschillende eigenschappen van diverse systemen te bestuderen. **Hoofdstuk 4** beschrijft enkel-molecuul FRET methodes om het Cas9 nuclelease te bestuderen. We beschrijven methodes gebaseerd op twee manieren van immobilisatie, één waarbij het doelwit DNA wordt geïmmobiliseerd en één waarbij het Cas9-ribonucleïnezuurcomplex wordt geïmmobiliseerd. Naast de experimenten, zijn ook de voorbereidende stappen voor de monsters beschreven, zoals het schoonmaken van de slides, het labelen van de nucleïnezuuren en het biotinyleren van Cas9. Experimentele voorbeelddata wordt gegeven voor elke immobilisatie methode en daarbij worden aanbeve-

elingen gedaan voor situaties waarin deze methodes het meest geschikt zijn. DNA immobilisatie is bijvoorbeeld geschikter om te onderzoeken hoe Cas9 binding de DNA duplex verstoort. Eiwit immobilisatie is daarentegen meer geschikt om zwakke interacties te bestuderen aangezien alle ongebonden RNA moleculen weggespoeld zijn voordat doelwitsequenties worden toegevoegd en aangezien DNA minder snel aspecifiek op het glazen oppervlak blijft plakken. Daarnaast geven wij uitleg over het ontwerpen van crRNA en DNA voor elk experiment. Tenslotte wordt aan het einde van het hoofdstuk een samenvatting gegeven van beeldverwerking en de data analyse

Naast het fundamentele onderzoek naar Cas9 wat helpt om de vele vragen over eiwit activiteit en de type II CRISPR immuunsystemen als geheel, te beantwoorden, is de algemene focus in het onderzoeksgebied van Cas9 de toepassing in genoombewerking. Tot nu toe zijn verschillende factoren gevonden die de knipactiviteit van Cas9 beïnvloeden. Om er een paar te noemen: secundaire structuren van doelwitsequenties van DNA zoals de G-quadruplex kunnen de activiteit van Cas9 significant verlagen, net zoals een foutieve basecombinaties tussen het begeleidend-RNA en doelwit-DNA, ook verkiest Cas9 bepaalde sequenties boven die van anderen. In **hoofdstuk 5** laten wij zien dat Cas9 activiteit verhinderd kan worden door kleine RNA moleculen in vitro. Door het gebruik van biochemische assays laten we zien dat sequenties die complementair zijn aan de kiem-sequentie van het gids-RNA de knipactiviteit compleet stoppen, zelfs bij hele lage overmaat verhoudingen terwijl sequenties die complementair zijn aan het einde van de gids-RNA de knipactiviteit juist progressief verhinderen met toenemende ratio's van concurrerend RNA tot gids-RNA. Met behulp van enkel-molecuul fluorescentie laten we zien dat deze hybrides efficiënt geladen zijn door Cas9 en dat interacties worden voorkomen met het DNA-doelwit. Daarnaast gebruiken wij een molecuul dat complementair is aan het concurrerende RNA “ en laten zien dat dit laatste wordt losgelaten door het Cas9 ribonucleozuur-eiwit complex als het gehybridiseerd is aan het einde van het gids-RNA “, maar niet in het geval van hybridizatie aan de kiem-sequentie. Ondanks dat de inhibitor is losgelaten door Cas9, is Cas9 niet in staat om interactie aan te gaan met het doelwit-DNA, wat suggereert dat het eiwit blijft steken in een inactieve conformatie. Een zodanige verhoging van activiteit kan vooral belangrijk zijn in therapeutische toepassingen waar miRNA potentieel baseparen kan vormen met het gids-RNA en zo de knipactiviteit kan voorkomen, of genen die transcriptioneel gezien actief zijn kunnen op deze manier minder efficiënt worden verwerkt. Om te kijken in hoeverre deze observatie geldt, worden in vivo experimenten uitgevoerd.

Vanwege de toepassingen in genoombewerking is Cas9 onderzoek geëvolueerd van CRISPR biologie en is het nu een apart onderzoeksgebied geworden. Veel onderzoek wordt gedaan om nieuwe Cas9 varianten te vin-

den of om SpCas9 mutanten te creëren die in staat zijn om verschillende PAM sequenties te als doelwit kunnen gebruiken. In dit laatste **hoofdstuk 6** wordt een overzicht en vergelijking gegeven van de kristalstructuren en de PAM-herkenningsmodi. Verschillende Cas9 orthologen worden hier beschreven: *Staphylococcus aureus* (SaCas9), *Francisella novicida* Cas9 (FnCas9) en *Campilobacter jejuni* Cas9 (CjCas9). Naast herkenning van de PAM-sequentie worden ook de implicaties van het zoekmechanisme besproken met behulp van de observaties van hoofdstuk 3 als voorbeeld. Daarnaast worden genetische bewerkte Cas9 mutanten zoals xCas9 of Cas9-NG besproken. Verschillende mutaties en aminozuursubstituties kunnen de zoektocht naar en herkenning van PAM-sequenties vertragen en daarmee de algehele activiteit van Cas9 verlagen. Daarom concludeert dit hoofdstuk dat Cas9 orthologen en vooral mutanten grondig bestudeerd dienen te worden met biofysische methodes voordat ze worden toegepast als therapeutische technieken om ongewenste effecten te voorkomen welke op kunnen treden door verstoorde PAM zoek- en herkenningsmechanismen.







# Acknowledgements

The time has come to write my most dreaded chapter of this thesis. I am almost sure that I am going to tear up more than once in the process. And although it is a joyous occasion (PhD almost over, woohoo!), I cannot help but feel myself overtaken by nostalgia as I start to reminisce about the years that passed. I remember myself in my first year, reading the acknowledgements in the theses of those graduating and tried to imagine what I would write. Fast forward five years and here I am, giving my thanks and remembering all the amazing people I met and countless fantastic memories I gathered. Before I go on to more personal notes, hoping to not forget anyone, I just want to say that joining BN for my PhD has been the best possible decision I could have made – all of you made it such a stimulating environment, not only scientifically, but also personally, with so many different characters and points of view. It has been an honor to spend my last five years with you.

First and foremost, I would like to thank my supervisor and promotor, Chirlmin. I came here a complete newbie in the fields of biophysics and biology, and you believed in me and had more confidence in my ability to crack this biology stuff than I did. I admire your knowledge and humility, a combination rarely found in academia. As a true leader you always took care of us. And not only that – you took time and energy to notice our differences and did your best to supervise us, myself included, according to the type of person we are. I am grateful for you taking your time to listen and consider any criticism – again, something rarely encountered in the competitive environment of academia. We always managed to resolve our differences in opinion in a civilized and most importantly productive manner. You have created an atmosphere in the lab where people feel comfortable and are not afraid to express their opinion and made it a fantastic place for us to grow professionally and personally. I have learned a lot from you.

I would also like to thank my copromotor Martin. You were always helpful, ready to answer any questions and always up for a discussion about all the weird results we were getting. You are smart and knowledgeable and yet extremely approachable and down to earth. In addition, you are a great mentor to your students, someone who takes mentoring seriously, and supervises people with empathy and kindness. Please, never change.

I would like to thank the members of my doctoral committee (Prof. Nynke Dekker, Prof. John van Noort, Prof. Gijse Koenderijk, Dr. Stan Brouns, Dr. Carlos Penedo-Esteiro and Dr. Joyce Lebbink). Thank you for taking the time to read my thesis and for taking part in the defense ceremony. A special thanks to the members of BN – Nynke and Stan – who in their own way contributed to my PhD. Also, thank you, Carlos, for planting the seed in my interest in single-molecule spectroscopy and FRET.

An integral part of my journey have been the collaborators in the Seoul National University, Prof. Jin-Soo Kim and Dr. Seung Hwan Lee. Thank you for making this possible and providing the most important component of my research – Cas9. Seung Hwan, thank you for coming all the way to Delft to get the project moving. I still use the buffers you made.

A special thank you to the people who will stand with me during the defense – my Paranymps, Pawel and Aistė. Pawel, I had a hard time thinking whether I should mention you as my paranymp, or former colleague, or former lab technician – you are all those things, but most importantly you have become one of my closest friends. Thank you for all the conversations, dinners, crazy nights out and all your support. Aiste, pupa, you are one of my oldest and closest friends. Ačiū tau labai už tavo kantrybę klausantis mano dejonų ir nusiskundimų, už tavo draugystę ir pagalbą. Visada žinau, kad galiu tavim pasitikėti ir tu man visada padėdi, kuo tik gali. Tu esi nuostabi draugė, aš labai laiminga, kad tave turiu.

Moving to the Netherlands and starting my journey in BN was made easy by the fantastic administrative staff in our department. Jolijn, you were my first point of contact here and you were so incredibly helpful! I would also like to thank all the members, present and past, of the secretariat, who always helped with everything, from filling forms to organizing activities and parties: Emmylou, Dijana, Amanda, Tracey, Nadine, Tahnee. In addition, thank you to the management staff for keeping this department running smoothly: Angela, Liset, Marije, Chantal and Esther.

BN is a dynamic place with so many amazing researchers, combining such different techniques and research topics. This would not be possible without the amazing PIs in this department. Marileen, thank you for keeping this department running and always searching for a dialog with PhD students and postdocs. Cees, thank you for always having crazy, “blue sky” ideas. Stan, thank you for discussing CIRSPR projects with us and providing a biological perspective. Greg, you were a great teacher in biology for physicists. Bertus, thank you for Quo Vadis+. Thank you to other PIs, which I interacted less extensively with, for making BN such a vibrant environment: Timon, Nynke, Hyun, Marie-Eve, Liedewij, Arjen, Dimphna and Cristophe.

No PI, PhD student or postdoc would be able to do what they are doing without the amazing technical support in our department. Anna, you did so much “behind the scenes” work in our lab and spoiled us so much by taking care of everything! It was a real struggle when you left!/ Ilja, you are always so much fun, and so knowledgeable, I hope to have more projects with you! /Cecilia, I like your no-nonsense attitude and that you are always willing to help, not shying away from projects that are not familiar to you. /Jan, thank you for taking care of our ordering and always being so efficient and diligent! /Margreet, thank you for supporting us with the setups and software and helping out with ordering in times of need. /Jelle, you always brought a different perspective to the table when we had troubles with hardware and provided us with all kinds of parts which were so skillfully and carefully crafted. Although not part of the Joo lab, I had the pleasure of working with you, Jaco, and was impressed with your knowledge and friendliness. /Sacha, you were so so helpful during the move, when we had to rebuild our setups. In addition, you are very kind and friendly, and I always enjoyed our conversations. /A special thanks goes to the department technicians and safety officers for keeping everything in order and being patient with those, like myself, who sometimes break the safety rules: Susane, Anke, Erwin and Robbie. /Roland, you are the guardian of our data, the tech guy, we can always rely on you./ Essengul, I have not worked with you, but if I had, I’m sure it would have been a lot of fun! I also want to thank other technicians in the department, for being so open and approachable and ready to help: Theo, Eli, Niels, Jeremie, Roland, Jacob, Dimitri.

And now the time comes to thank those who had to endure me the most during the last five years – the JooC lab (oh oh here come the waterworks). First, our dinosaurs: Malwina, I came to you with questions about everything: wet lab, RNA, protein, microscope, Matlab, Origin – you name it – and you were always able to answer them or if not, willing to find the answer together. So kind and friendly and even mischievous – I remember you ordering Zubrovka during Kavli day in Brussels!/ Mohamed, oh Mo! My office mate, my foosball team mate, the person who coined the name “cucumbers”. You helped me a lot and were always ready to drop your own stuff and shown me how to do this or that. Kind and friendly, composed on the outside, but with such a wild sense of humor! I enjoyed sharing an office with you so much and I miss you every day. I haven’t really played foosball after you left, it’s just not the same./ Stanley, Stan-lee, my RNA-buddy, always there to remind us about the dangers of RNases. You always had crazy scientific ideas and a wild approach to research, and although I am told you were not the first one to do this, but I learned working from home from you. I enjoyed our conversations and trips to Wageningen. I wish you all the best in UK./ Jetty, you were my first office mate from the lab and made the transition for me very

easy. I felt very welcome and comfortable. You helped me with the setup which made it very easy for me to maintain it. Thank you for the fun chats in the office./ Luuk! My CRISPR buddy, my travel companion, my office mate and my foosball opponent. I really value your advice on all things scientific and am so happy that we have a similar traveling style – the trips with you were so much fun! I learned a lot from you. You were the one who taught me how to run a gel! You took me by the hand (well, not literally) and introduced me to the equipment and the way the lab is run, when I was just a newbie. The office and the lab are not the same without you.

Laura, Lau, you deserve a medal for answering all my stupid questions about the thesis and defense preparations. This date would have been much later if it hadn't been for your help with all the administrative and practical advice. And of course, you helped me because you are kind, friendly and so agreeable. You are so easy to talk to and very funny. I greatly enjoyed our conversations about all kinds of telenovelas – could not have ever imagined that I would meet someone who knows those here! I admire your bravery in starting a company – wish you the best of luck with it!/ Mike, if you are not here to make friends, I think you might be failing. Your joining the lab made it so much more dynamic, you organized the BBQs, bowling, dinners, even the lab trip. Always helpful, always reliable, funny and always up for a bak-kie. You're the only other person I know who enjoys Cola Zero as much as I do. I always enjoy the lunches, and the coffee conversations, and that I can talk to you about football. I admire your crazy ideas, your curiosity and your drive. And of course, the wall of quotes. I can say *ik ben gelukkig* to have met you. Good luck with the rest of your PhD, and when frustrated, remember – it's not even a road, it's just a roundabout./ Thijs, T-ice, Ice-tea, TJ, as my longest running office mate you had to listen to my never-ending rants, for which you deserve a medal. Is there a better proof of your patience? You were always up for a coffee or a lunch, or dinner, and always up for a laugh in the office. My fellow target search explorer, I thank you for all the scientific discussions and your advice, which often helped me make sense of my own data. Your easy-going attitude made it all the more enjoyable to share an office with you. You grew a lot in those four years (also physically with all the calisthenics!) and I hope you will defend in November (although now seems unlikely), if not, you know what will happen. Best of luck in the upcoming few months!/ Sungchul, Dr. Kim, Skimpy, Beercules, what a joker you are! So incredibly knowledgeable, smart and skilled and yet so crazy – no one else could pull off such a working schedule as you do. Thank you for all your advice in the lab – you are a never-ending treasure trove of knowledge. Thank you for all the things you recommend and the things you don't believe and for driving us for lunch or dinner, even if that means you can't enjoy being Beercules./ Ivo, you tried to escape us, but could not resist our

magnetic pull. I am glad you chose to join this lab. You are kind and relaxed, always up for a coffee and endure my rants without complaining. And that is not easy! Also, not easy is your project, I admire your determination! I hope you won't move to Leiden and abandon us, well at least no while I am still here! / Iasonas, the Greek guy, you seem so calm and composed, but you have such a crazy sense of humor! I always like the South park jokes and your bromance with Mike is just something else. You also have great tips on data analysis, and I admire your theoretical knowledge. / Adam, you're very friendly and easy going, as well as knowledgeable, I hope to learn a lot from you in the coming months. Good luck with all those single molecules!

Also, shout out to the students, who were supervised by other members of the lab, but who I had great fun with: Chun, so smart and funny, always cheerful, always helpful! / Isabel, someone who actually brought cake! I enjoyed sharing an office with you and hope to see you soon! / Mees, by far the craziest student we have ever had. / David, I admire your positivity, even I can't seem to bring you down – good luck in Sweden, you'll need that positivity during the dark winter days. / Sabina, so hardworking, but also easy going, it was a lot of fun having you in the lab! / Also, I'd like to thank the students I supervised – Soufian, Tim, Erwin and Justine, each of you have taught me something.

A special mention and thanks to the honorary members of the Joo lab – Michela, Mathia, Misha and Carlos. / Michy, you were in my first office and I always had a lot of fun chatting with you. I'm glad you're finally enjoying the sun in Israel. / Mathia, you're so friendly and trustworthy. I admire your positivity and your strength. You're such a pleasure to be around! / Misha, thank you for all the CRISPR discussions. You are smart and most importantly very kind. Good luck finishing your PhD. / Carlos, you're way too cool for us! I admire that you come all the way to Delft every week.

And now a time comes to remember all the great people of BN I had the honor to interact with in the last years. I will start with my former office mates – Pauline, Johannes, Ilyong, Peter and Zoreh. You were so friendly and helpful and made it very enjoyable to share the office with you. / Sam, I appreciate that you were always up for a chat or for a drink. / Richard, thank you for all the advice you have given me over the years, all the chats, smokes and drinks. I always enjoyed talking to you about science and about life in general. I wish you all the best in your pursuit of academia. / Orkide, you are so positive and so cheerful! I admired your down-to-earth attitude and how easy it is to talk to you. / Maarten, I really like your sense of humor! I enjoyed the dark jokes we shared and how you were a party animal in disguise. / Bojk, you are very knowledgeable, not just in science but in almost every other subject. Thank you for the discussions and the laughs we had. Sumit,

the most eloquent member of BN. I never cease to be amazed by your choice of words./ Kayley, it is great to see a fellow St Andrean in BN!/ Umberto, Mariana, Belen, Filip, Louis – thank you for always being friendly and kind./ Anthony, I always enjoyed our chats and beers. I like your relaxed demeanor and thank you for taking part in organizing all the surveys, career days and other events in BN./ Fede, so philosophical and poetic. You care very deeply about everything and everyone around you. Thank you for organising the BN concerts and for keeping me company multiple times during nights in the lab./ Ganji, so crazy and resourceful! I like how you never sugar-coat anything./ Fabai, so talented, yet down to Earth, I always had fun talking to you!/ Daniel V., your good mood and positivity is so admirable, as is your sense of humor. I miss running into you on my way to the microscope./ Sergii, my fellow Eastern European, the best dancer BN has seen. Thank you for our chats about politics and life in general./ Sandro, thank you for always being so nice and friendly and for enquiring how I am doing – means a lot./ Alessio, you're a lot of fun, thank you for the BN borrels and concerts. And of course, 549./ Wayne, you are kind and always interesting to talk to./ Jorine, a fierce lady who does not sugar coat anything! I'm sure you can pull any lipstick off! I miss chatting with you, good luck in your future career./ Kuba, I remember annoying you with reminding that "Cas9 is not CRISPR" and you were always tolerating my nagging./ Adi, you also don't like to sugar coat things. You were always good fun during borrels and late nights at work. I am very glad to have you in my office and I greatly enjoy our conversations and your sense of humor. And of course, enjoy the upcoming big adventure!/ Alberto, thank you for reviving the BN social life together with Becca./ Oskar, I have only met you recently, but can already tell that you are a very nice and friendly guy. I'm glad you joined the project and I wish I was as efficient and hardworking as you!/ Michel, mama! I am so glad we started this project, because no matter how scary I seemed, we became friends. You were my biggest cheerleader when Matilda was born, thank you so much for your support and the advice you have shared (and hope will continue sharing). You have such a great sense of humor, I love it when our meetings start with 5 minutes of laughter. You are also an incredibly strong person and such a great mama to Madi. Again, I'm so so happy we became friends./ Je-Kyung, JK, thank you for willing to help and giving a go at imaging Cas9 with AFM./ Sonja, thank you for always enquiring how the JooCs are doing./ Siddhart, I have enjoyed our conversations a lot./ Yoones, you have got some moves! I liked your provocative questions and always enjoyed the chats with you. Good luck in the corporate world./ Stephanie, I admire your grit and determination as well as how friendly you are./ Louis, such a cheerful guy! Your smile always brightened my day! I also thank other PhD students and post-docs of the CD and Marileen's lab, past and present, – Greg, Yaron, Maurits,



Kim, Ana, Calin, Felix for the conversations we had and everything I learned from you, whether directly or indirectly.

Moving corners of the department... Helena, thank you for always being up for a good party and the interesting conversations we had./ Nicole, I wish I was as chill and cool as you are. Always great fun to talk to you./ Mehran, I enjoyed our conversations about life and politics, and research. You always have something to say./ Hiran, thank you for taking part in organizing all kinds of BN things, career days, concerts, etc./ Fayeze, thank you for being so warm and friendly./ Marek, I was always amazed by your knowledge of the history of the Polish-Lithuanian commonwealth and enjoyed our discussions./ Eve, you are a lot of fun to be around and such a party animal! I always loved chatting with you and hope we can meet more often!/ Carsten, you were a big revelation, I could not have imagined what a fun guy you are when I just met you./ Lisa, you are always so warm and friendly./ Benjamin, keep up the good work, I'm sure you're going to conquer Mars./ Alicia, I remember how you kindly took me in for an evening when I locked myself out in my first year. A great illustration of your helpful and kind character./ Jonas, one of the old dinosaurs as myself, we started around the same time and I had many great conversations with you./ David, a fellow fresh parent of BN, you will be a very fun dad!/ Becca, thank you for taking the lead with Alberto to revive the BN dying social life. You're doing a great job. You're always relaxed and fun to be around./ Jochem, you're warm and welcoming and you always have something to say. I admire your determination./ Christian, you're such a genuine and friendly guy! I always enjoy having a beer with you, you're full of positive energy./ Cristobal, you are very knowledgeable and know how to enjoy life, doing the things you like. You and Patri are always welcome in our home./ Patrick, you may seem quiet, but you are funny, with a sense of humor I really appreciate. I admire your grit - your project has not been easy./ Da, what a voice! I enjoyed the borrels with you and admired your positive attitude./ Vanessa, Victor, Dominic, Sophie, even though you've left quite some time ago, your presence is still missed here. Finally, I'd like to thank the people that I did not have a chance to interact with extensively, but you make BN the amazing department it is: Ferhat, Diederik, Tim, Leila, Enzo, Els, Sophie, Werner, Christine, Yiteng, Diego, Nils, Mitasha, Nicola, Biswajit, Kui, Ewa, Magda, Eveline, Elisa, Henry, Sabina, Kevin, Eugene, Paola, Ramon, Behrouz, Afshin, Aafke, Duco, Zhanar, Franklin, Rita and anyone who I may have forgotten to mention (if so - my apologies).

I also made some friends outside the department. Christian, Ramiro, James, Deniz, Sandy - you made these years way more enjoyable than they would have been otherwise. Chris, I love our discussions about politics, and thank you so much for the jam sessions, even though I'm musically handicapped, I always had great fun./ Ramiro, you're so much fun and throw such



good parties! / James, I always enjoy singing my heart out to “Loch Lomond” with you. We share a similar sense of humor, I always enjoyed your jokes. / Deniz, you’re a pleasure to be around, so warm and friendly. / Sandy, thank you for your hospitality when I visited Canada and for being my confidant. I wish you the best of luck in your new adventure in the Netherlands.

I would also like to extend my gratitude to others who have also been there during this journey. Chris (Christopher), I know you would like to be here and see me defend, but I know you’ll be here in spirit. You have lifted my spirits numerous times, always there to listen and to help. I am very excited for you with your new adventure this autumn! / Rita!!!! Pupa, thank you for introducing me to the way of life in the Netherlands, among other things. Tu man tiek daug padėjai, davei daug patarimų, kaip išgyventi šitoj valstybėj. Ačiū tau labai, buvo nuostabu turėt artimą draugę taip arti, nors ir svetimoj šaly. / Dagmar, Hans-Christian and Rosina thank you for your hospitality and your support for both Sebastian and myself. Vielen dank!

Mama, ačiū tau už viską: už rupestį, už paramą, už tikėjimą manim. Tu rūpinaisi manim ir močiute, nepalikdama laiko sau. Aš labai tikiuosi, kad dabar nors kiek atsigriebsi už visą tą laiką. Aš labai džiaugiuosi, kad su tavim mes galim diskutuoti apie viską – apie meną, mokslą politiką ir visa kita! Esu laiminga, kad tave turiu. / Rimvydai, ačiū, kad rūpinatės mumis, kartu keliaujate. Ačiū už pagalbą man nuvežant į oro uostus, už nuvežimus prie jūros ir kitas vietas. Džiaugiuosi, kad esat mano gyvenime. / Iveta, nors mes neseniai susipažinom, labai džiaugiuosi tave pažinodama. Tiek daug pozityvo ir gero humoro jausmo!

Matilda, mano zuiki. Tu dar visai visai mažutė, bet gal kada nors perskaitysi šias eilutes. Tavo gimimas man parodė, kas iš tikrųjų svarbu ir padėjo teisingai susidėlioti prioritetus. Ir tu toks mielas ir juokingas kūdikis. Aš tave labai labai myliu ir džiaugiuosi, kad tave turiu.

Finally, Sebastian, meine liebe. I do not even know where to start. We have been through so much together and you have always been my rock. Thank you for always being there for me and for being such an amazing partner and father. I have never imagined that I could love someone so much and feel so loved, until I met you. You are so kind, caring and funny and you know me like no one else does. Life with you is so wonderful. I love our little family, I love you and I cannot wait to see what the future brings our way.





# Curriculum Vitae

## VIKTORIJA GLOBYTĖ

- 29-11-1991      Born in Šiauliai, Lithuania
- 1998-2010      Secondary Education,  
Šiaulių Lieporių Giminazija  
Šiauliai, Lithuania
- 2010-2014      M.Sc. Physics  
University of St Andrews  
St Andrews, Scotland
- 2014-2019      PhD Department of Bionanoscience  
Kavli institute of Nanoscience  
Delft University of Technology, The Netherlands  
Thesis:              Seek And Destroy: Single-molecule  
                         perspective on the target search and  
                         recognition by the Cas9 endonuclease  
*Promotor:*        Dr. C. Joo  
*Copromotor:*    Dr. S. M. Depken



# List of Publications

4. **Globyte, V.**, M. Bengtson, J. van der Torre, C. Dekker, C. Joo, "*Small RNA molecules inhibit the activity of Cas9 in vitro*" (*Manuscript in preparation*)
3. **Globyte, V.** and C. Joo, *Single-molecule FRET studies of Cas9 endonuclease*. *Methods Enzymol*, 2019. **616**: p. 313-335.
2. **Globyte, V.**, et al., *CRISPR/Cas9 searches for a protospacer adjacent motif by lateral diffusion*. *EMBO J*, 2019. **38**(4).
1. **Globyte, V.**, S.H. Kim, and C. Joo, *Single-Molecule View of Small RNA-Guided Target Search and Recognition*. *Annu Rev Biophys*, 2018. **47**: p. 569-593.