

**Single Cells in the Spotlight**  
**Probing the kinetics of CRISPR Adaptation and Interference**

McKenzie, R.

**DOI**

[10.4233/uuid:61ab1f7f-fccc-480b-b20f-9010c19d990c](https://doi.org/10.4233/uuid:61ab1f7f-fccc-480b-b20f-9010c19d990c)

**Publication date**

2021

**Document Version**

Final published version

**Citation (APA)**

McKenzie, R. (2021). *Single Cells in the Spotlight: Probing the kinetics of CRISPR Adaptation and Interference*. [Dissertation (TU Delft), Delft University of Technology]. <https://doi.org/10.4233/uuid:61ab1f7f-fccc-480b-b20f-9010c19d990c>

**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.

The background is dark gray with a bright, circular spotlight effect in the center. Scattered across the background are numerous small, interconnected nodes and lines, resembling a network or phylogenetic tree diagram. The nodes are small circles, some solid and some hollow, connected by thin lines, some solid and some dashed.

# **Single Cells in the Spotlight:**

**Probing the kinetics of CRISPR  
Adaptation and Interference**

**Rebecca E. McKenzie**



# **Single Cells in the Spotlight: Probing the kinetics of CRISPR Interference and Adaptation**

## **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

Friday 15 October 2021 at 10:00 o'clock

by

**Rebecca Elizabeth MCKENZIE**

Bachelor of Science (Honours) in Microbiology and Immunology,

University of Otago, New Zealand

born in Wellington, New Zealand.



This dissertation has been approved by the promotor.

Composition of the doctoral committee:

Rector Magnificus	chairperson
Dr.ir. S.J.J. Brouns	Delft University of Technology, promotor
Prof. Dr. ir. S.J. Tans	Delft University of Technology, copromotor

Independent members:

Dr. ir. C. van Houte	University of Exeter, United Kingdom
Prof. dr. C.L. Beisel	University of Wurzburg, Germany
Dr. D.C. Swarts	Wageningen University & Research
Prof. dr. J.W. Veening	University of Lausanne, Switzerland
Prof. dr. B. Rieger	Delft University of Technology

Reserve member:

Prof. dr. C. Dekker	Delft University of Technology
---------------------	--------------------------------



Cover by: Rebecca McKenzie

Printed by: Gildeprint

© 2021 Rebecca McKenzie

Casimir PhD series 2021-27

ISBN: 978-90-8593-493-6

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

For Mum & Dad

You made anything and everything feel possible



# Table of Contents

<b>Chapter 1</b>	General Introduction	7
<b>Chapter 2</b>	CRISPR-Cas: Adapting to change	25
<b>Chapter 3</b>	Using CAPTURE to detect adaptation in native arrays	53
<b>Chapter 4</b>	Direct visualization of native CRISPR target search in live bacteria reveals Cascade DNA surveillance mechanism	87
<b>Chapter 5</b>	Single cell variability of CRISPR-Cas adaptation and interference	139
	Summary	194
	Samenvatting	197
	About the author	200
	List of publications	201
	Acknowledgements	202



The background features a repeating pattern of abstract geometric shapes, including triangles and lines, in various shades of gray. A large, bold, black number '1' is positioned in the upper right quadrant.

# 1

## General Introduction

## 1.1 Immunity: Why do bacteria need it?

As has become very apparent in the world recently, immunity against viruses is important to survival. Just like humans many bacteria face the chance of infection by viruses known as bacteriophages (from here on phages). In this introductory chapter we will briefly review the ever-evolving bacterial immune system repertoire, zooming in on CRISPR and in particular the type I-E system central to the research covered here. Additionally, we will introduce the molecular genetics and microscopy techniques that have been used to study CRISPR defense in both populations and single cells thus far, and are further built upon in this thesis.

The basis of an immune system relies on the ability of the host cell to discriminate 'host' from 'non-host' or foreign invaders. The ability to do so then allows specific targeting of the threat and its clearance from the cell. This distinction may be particularly challenging for bacteria as, unlike in eukaryotic cells where DNA is located in the nucleus <sup>1</sup>, initially there is often no separation between the injected phage DNA and the bacterial chromosome. In addition, phages have been estimated to outnumber their bacterial hosts by 10-fold, contributing to at least 20 % of the daily bacterial mortality <sup>2</sup>. As a result of this constant peril, bacteria have evolved many sophisticated mechanisms to provide immunity <sup>3,4</sup>. Unsurprisingly, in response phages have developed a plethora of ways to ensure successful replication <sup>5,6</sup>, describing what we refer to as a never ending co-evolutionary arms-race. Such evolution is driven by the rapid adaptation of phages to selective pressures, such as host defense barriers, and has resulted in bacteria equipped to survive a wide range of threats and competitive environments <sup>4,7</sup>.

Much like the human immune system bacteria can initiate both innate and adaptive immune responses against invaders. Innate immunity is non-specific, resulting in an immediate but generic response to an incoming threat. A wide arsenal of innate mechanisms can be put in play to target all phases of the viral lifecycle including: preventing phage adsorption to the cells through loss or modification of surface receptors or production of an extracellular matrix <sup>8–10</sup>; prevention of DNA entry via superinfection exclusion (Sie) systems <sup>11,12</sup>; restraining replication through bacteriophage exclusion BREX <sup>13</sup>; destruction of the phage genome through restriction-modification (R-M) systems <sup>14</sup>; and premature or programmed cell death to limit phage propagation via toxin-antitoxin (TA) systems and abortive infection (Abi) systems respectively <sup>15–17</sup>. While innate immunity provides a strong first line of defense, it can lack the specificity and adaptability to combat certain invaders. Previously, it was thought

that adaptive immunity existed only in eukaryotes, however in recent years adaptive immunity has also been found in prokaryotes (bacteria and archaea) <sup>4,18</sup>.

## 1.2 A more specific example: CRISPR-Cas

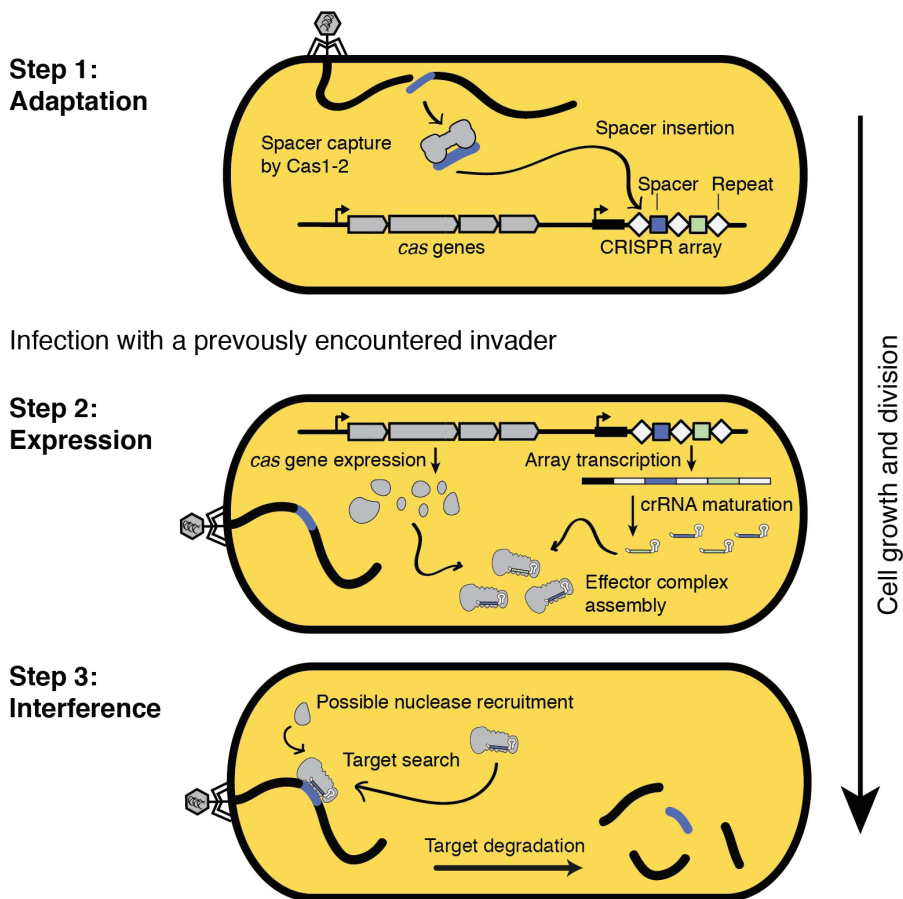
The first known example of an adaptive immune system in prokaryotes is CRISPR-Cas, a system which has become well known for its potential in genome editing <sup>19,20</sup>. The system was first discovered as a number of interspaced genomic repeat sequences during sequencing of the *iap* gene in *Escherichia coli* <sup>21</sup>. In 2002 these repeat regions interspaced with short sequences were aptly named CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) arrays and identified to reside nearby a *cas* (CRISPR associated) gene locus. However, it was not until in 2005 that the sequences interspacing the repeats were linked to viral genomes <sup>22–24</sup>, and a role in cell immunity was proposed. Following this, the first experimental evidence of adaptive immunity was published in 2007 <sup>25</sup>, resulting in a new, exciting and quickly advancing field of research to uncover the mechanistic nature bacterial adaptive immunity.

The CRISPR-Cas system is unique in its capacity to provide both specific and adaptive immunity to the bacterial cell. In short CRISPR defense can be defined in three phases (**Fig. 1.1**). The first stage, adaptation, involves the acquisition of a short sequence (termed spacer) from the invader and its storage in the CRISPR array for future defense <sup>23,25</sup>. The second stage involves expression of the *cas* genes, and processing of the array transcript into short guide sequences known as CRISPR RNAs (crRNAs) <sup>26</sup>. These crRNAs associate with the Cas proteins forming surveillance complexes that in the third stage, known as interference, are able to scan the cell, locate the invader and facilitate their degradation <sup>27–29</sup>.

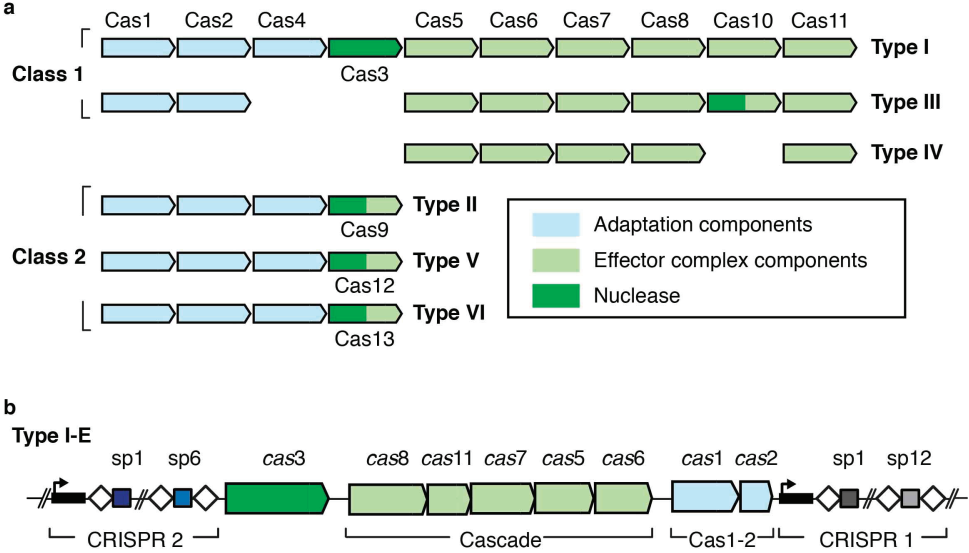
Since its initial discovery in depth computational work has revealed a huge diversity of CRISPR-Cas variants, that have now been divided into two distinct classes, six types and several subtypes <sup>30</sup> (**Fig. 1.2a**). While Class 1 systems are characterized by the formation of a large multi-subunit effector complexes, in contrast, Class 2 systems require only a single crRNA-surveillance subunit with multiple domains <sup>31</sup>. The types are then further divided based on the presence of certain signature Cas proteins, such as the Cas3 nuclease, which is conserved in all Type I systems <sup>32,33</sup> (**Fig. 1.2a**). Finally, sub-types are determined based on the composition of the effector complexes and operon organization <sup>31</sup>. In this thesis we focus on the DNA targeting type I-E CRISPR-Cas system (**Fig. 1.2b**), which is the most commonly found system in bacteria <sup>34</sup> and is often found in



*Enterobacteriaceae*<sup>35</sup> making it a convenient model organism for investigating CRISPR-Cas immunity<sup>36</sup>.



**Figure 1.1: CRISPR-Cas adaptive immunity consists of three main steps.** Step 1 Adaptation; Upon infection a small fragment of the invaders DNA is captured and inserted into the CRISPR array by the Cas1-2 complex. Step 2 Expression; Expression of the *cas* genes and transcription and processing of the CRISPR arrays into crRNAs enables assembly of effector complexes which in Step 3 Interference; find and degrade the invading DNA.



**Figure 1.2: Classification of Class 1 and Class 2 CRISPR-Cas systems.** **a**, CRISPR-Cas systems can be divided into two classes based on the presence of a multi-subunit effector complex (Class 1) or a single effector protein (Class 2). Further division into types is then based on the presence of signature genes. **b**, The type I-E CRISPR-Cas system consists of two CRISPR arrays and *cas* gene cluster encoding the Cascade effector complex (light green), the Cas3 nuclease-helicase (dark green) and the Cas1-2 adaptation module (light blue).

1.3 The type I-E system

In recent years, thorough investigations into the type I-E CRISPR-Cas system have made important progress in identifying the sequence of steps of defense<sup>26,27,36,37</sup>, the kinetics of molecular interactions<sup>38,39</sup>, and their structural basis<sup>40–44</sup>. In *E. coli* K12 the CRISPR system consists of two CRISPR arrays supplemented by a *cas* gene cluster which encodes proteins for the 11 subunit Cascade effector complex, the Cas1-2 adaptation complex and the Cas3 nuclease-helicase fusion protein<sup>45,46</sup> (**Fig. 1.2b**). Native expression levels of the *cas* genes in this system are too low to provide defense against invading mobile genetic elements<sup>47,48</sup>, consequently, *in vivo* studies have had to manipulate or engineer strains to allow further investigation and the wealth of knowledge that is now available<sup>26,37,45,48–52</sup>.

## 1.4 CRISPR interference

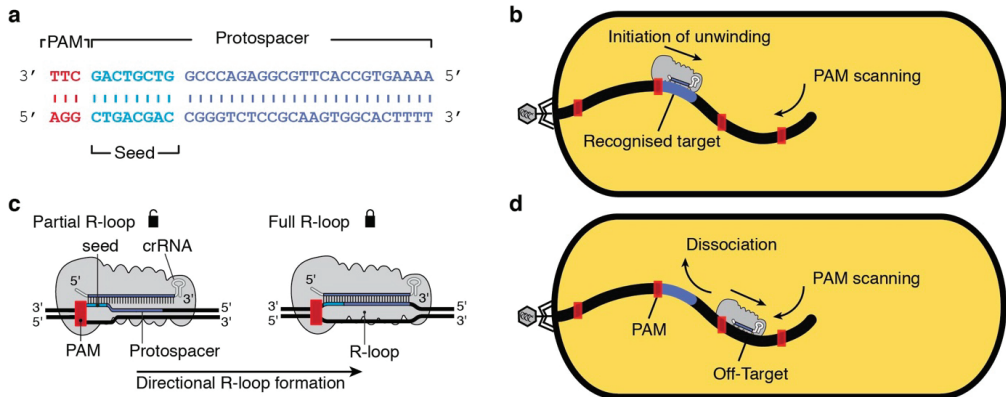
# 1

While acquisition of a specific spacer is first required before a specific CRISPR response against an invader can be carried out (details below), once the spacer has been stored in the CRISPR array the cell is able to mount an immediate response, termed direct interference, upon re-infection. During direct interference one of the main challenges faced by the I-E CRISPR system is the requirement to quickly locate the target amid all DNA in the cell within a limited time window. Interference is a tug-of-war between invader replication rates and CRISPR target search and degradation rates<sup>53,54</sup>. If the Cascade complexes locate the target too slowly, the invader will have the chance to replicate to a level that either exceeds the degradation rates (e.g. plasmid establishment at 100 copies)<sup>55</sup> or in the case of a phage induce cell death by lysis facilitating the release of progeny into the population<sup>4,56</sup>. Fast, low-fidelity target location could however result in accidental targeting of complementary sequences in the cells own genome. Fortunately, Cascade is able to differentiate between self and non-self through a small 3 nt sequence 5'-CTT termed the PAM (Protospacer Adjacent Motif). This sequence is present only on the invader, adjacent to the target, and not in the CRISPR array<sup>57–59</sup>.

In addition, a number of studies have shown that Cascades uses the PAM to scan the cellular DNA faster, checking crRNA complementarity with the DNA only when a PAM is recognized<sup>41,54,60–64</sup> (**Fig. 1.3a,b**). This unwinding subsequently allows for examination of the PAM adjacent DNA, known as the seed region<sup>65</sup>, for complementarity to the crRNA guide (**Fig. 1.3a,c**). The base pairing between a complementary crRNA and target sequence displaces the non-target strand of DNA and ultimately results in the formation of an R-loop<sup>42</sup>, expanding from the PAM proximal (or seed) region to the PAM distal region<sup>64</sup>(**Fig. 1.3c**). The directional formation of the R-loop allows Cascade to proofread the target sequence and often rapidly dissociate when a mismatch or off-target is present<sup>66,67</sup> (**Fig. 1.3d**). Full formation of the R-loop is considered to result in a conformational change in Cascade to a “locked” state and subsequently the recruitment of Cas3 helicase-nuclease to the target site (**Fig. 1.3c**)<sup>39–42,66</sup>. Once at the target site Cas3 proceeds to unwind the DNA and cleave the resulting ssDNA, ultimately degrading the invader and thus clearing it from the cell<sup>27,68–70</sup>.

Though Cascade complexes in *E. coli* have been observed to have a wide-ranging specificity for PAMs<sup>41,49,59</sup>, plasmids and phages with mutations in the PAM and seed region were able to escape CRISPR interference<sup>50,65</sup>. Though it has been shown that the invader can often still be degraded despite mutations

in the PAM, seed or target sequence<sup>49,52</sup>, the reduced binding affinity of the Cascade complex<sup>71</sup> and in some cases reduced Cas3 degradation rates<sup>38,52</sup> leads to a reduced target degradation rate, that will often not be able to overcome the invader replication rate to eliminate all invader copies from the cell in time<sup>51,53</sup>.



**Figure 1.3: Directional formation of the R-loop by Cascade.** **a**, Schematic of the sequence elements required for full R-loop formation. The PAM (red) is adjacent to the protospacer (highlighted in blue and purple) on the target. The first 8 nt of the protospacer is known as the seed sequence (blue). **b**, The Cascade complex uses the PAM to scan the target DNA efficiently, initiating unwinding upon PAM recognition. **c**, The R-loop is formed directionally from the PAM proximal to PAM distal region. The partial R-loop is in an unlocked state and allowing proofreading of the target sequence (left). Upon full R-loop formation Cascade undergoes a conformational change to a locked state (right). **d**, If a mismatch or off-target is detected during R-loop formation Cascade can rapidly dissociate.

## 1.5 CRISPR adaptation

The acquisition of a new spacer in the type I-E system can happen via two pathways, naïve adaptation or primed adaptation<sup>72–74</sup>. Naïve adaptation occurs from a previously unencountered invader and requires only the well conserved Cas1 and Cas2 proteins which form the adaptation complex<sup>37</sup>. Primed adaptation can occur for example when an escape target (i.e. PAM or seed mutation) is partially recognised by a pre-existing spacer promoting the acquisition of new spacers<sup>65</sup>. It has been shown that primed adaptation occurs 1000 times more frequently than naïve acquisition, suggesting this is the main pathway of acquiring new spacers in the type I-E system<sup>75</sup>. Regardless of which pathway is taken adaptation is comprised of a series of subsequent steps

including pre-spacer generation, selection for a consensus PAM and further processing by Cas1-2, followed by insertion of this fragment into the CRISPR array at the leader proximal end<sup>37,43,76,77</sup>.

Primed adaptation followed by interference with the new spacer (together called priming) is present in class I systems, and considered to be feedback loop in which mismatched or mutated targets are able to stimulate rapid spacer acquisition when the cell is under threat<sup>50,73,74</sup>. Two models of pre-spacer generation interference-dependent, or -independent, have been proposed and investigated experimentally<sup>36</sup>. The interference-dependent model describes pre-spacer production through low-level CRISPR interference. In this case despite the lowered binding affinity for a mutant target Cascade is still able to bind and initiate degradation at a slower rate<sup>49,51,78</sup>. Further, a number of studies have provided evidence linking these CRISPR interference products and adaptation through the characterization of pre-spacer substrates produced by Cas3<sup>51,52,79,80</sup>. Kunne *et al* 2016 demonstrated that Cas3 cuts fragments both between 30 – 100 nt in length and enriched for thymine stretches at the 3' end, thus enhancing the chance of selection and insertion of a spacer with a 5'-CTT consensus PAM into the CRISPR array<sup>52</sup>.

The interference-independent pathway on the other hand, suggests that upon binding a mutant target Cas3 is recruited by Cascade in a Cas1, Cas2 dependant manner. This results in the formation of a primed acquisition complex (PAC) coupling pre-spacer production and capture by Cas1-2 for integration<sup>38,39,61</sup>. Recently pre-spacers were shown to be associated simultaneously with both Cas1 (assumed to be in complex with Cas2) and Cas3 *in vivo*, further supporting the formation of a PAC complex<sup>81</sup>. While a large number of groups have investigated both the interference-dependant and -independent pathways of adaptation in type I-E CRISPR-Cas system, the majority of studies have only been able to consider a small selection of the 64 possible PAM sequences, often making a direct comparison of results difficult and leaving many details to be further elucidated.

## 1.6 Bacterial population studies

Due to the ease of access to the techniques, equipment required and the ability to rapidly screen a large number of cells, numerous population-based studies of CRISPR defense have been carried out. Direct interference has been investigated using techniques such as patch plating assays, conjugation assays and plaque assays allowing us to determine the conditions, for example the PAM

and CRISPR components, under which cells are able to successfully clear an invader<sup>25,26,29,82</sup>. Adaptation on the other hand, due to its low frequency in some systems is more difficult to study on a population level. Classical polymerase chain reaction (PCR) screening methods require acquisition to have occurred in at least 1% of the cells in the population in order to successfully detect an effect using gel electrophoresis<sup>37</sup>. Nevertheless, initial studies screening single colonies with this technique, were able to determine the protein and sequence requirements for spacer selection and insertion<sup>37,74,83</sup>. Further, sequencing of the PCR expanded arrays allowed mapping of spacers back to the invader, determining where they were selected from, leading to the first proposed mechanistic models for primed adaptation<sup>73,74,84</sup>. Since then, a number of methods have been developed to improve the sensitivity of detection, including but not limited to; PCR methods involving extra selection steps<sup>85,86</sup>, more specific primers binding only new spacers<sup>37,83,85,87</sup>, frameshift systems for identification of cells that have acquired a spacer<sup>88,89</sup> and deep sequencing of arrays from cells that successfully eliminate the invader<sup>50,86,90,91</sup>. The constant advance of methods to study CRISPR at the population level will allow further advancement in our understanding of both well characterized systems and new previously inaccessible systems. However, population studies in general are limited in their resolution and unable to provide insight on the individual stages of the CRISPR processes and their timing in single cells.

## 1.7 Single molecule studies

In addition to the population assays described above, advancements in fluorescence and forced based microscopy techniques<sup>92</sup> have enabled high resolution information to be obtained about individual molecules and their kinetics during CRISPR defense. Force based microscopy techniques involving magnetic tweezers, optical tweezers have shed light on details such as target recognition, R-loop formation and target cleavage<sup>66,93–95</sup>. Furthermore, a number of fluorescence based microscopy techniques such as single particle tracking<sup>54,96,97</sup>, single-molecule fluorescence resonance energy transfer (smFRET)<sup>39,70,76,98</sup>, and DNA curtains in combination with total internal fluorescence microscopy (TIRF)<sup>38,61,99</sup> have shed light on the details of target search, the molecular interactions that occur at the target site and the mechanisms of the adaptation machinery. Single molecule studies however, occur on short time scales (seconds to hours) and are limited to the study of only a few components in the cell due to the advanced set up required for multi fluorophore imaging.

## 1.8 Time-lapse microscopy

Time-lapse microscopy describes the ability to image cells frequently for long periods of time monitoring changes in cell features of interest perhaps through cell morphology, for example cell length, or fluorescent labelling. Originally, time-lapse experiments were carried out on gel pads soaked in a medium of interest<sup>100–102</sup> however, this set up allowed imaging of only a few generations due to possible desiccation of the cells or the formation of multiple cells layers, preventing further imaging.

Technology such as microfluidic flow cells have enabled researchers to cross this barrier, allowing long-term imaging due to a constant supply of media, dimensions designed to maintain a single layer of cells, and the removal of excess cells by flow<sup>103–105</sup>. The ability to image and follow individual cells within a dynamic population has enabled the characterization of many biological processes and the variation that exists within them<sup>101,106–110</sup>.

Heterogeneity of the CRISPR response in individual cells is likely to play a role in the overall protection of a population against an invader. Moreover, when measuring population survival in bulk it can be difficult to establish whether protection is determined by a resistant sub-population or the majority of cells undergoing successful defense. Recent work has challenged the view of CRISPR-Cas as a single cell defense mechanism<sup>56,111</sup> exhibiting the need for further elucidation of the single cell dynamics of CRISPR defense. In this thesis we introduce time-lapse microscopy in combination with microfluidics to investigate CRISPR interference and adaptation on the single cell level. Further investigation of single cells, and development of these imaging techniques in the field will be critical to obtaining an overall picture of population survival mechanisms in particular in native settings where populations are likely more complex.

## 1.9 Thesis Outline

The goal of this thesis was to develop new methods and further elucidate the dynamics of the CRISPR interference and adaptation processes in the type I-E CRISPR-Cas system of *Escherichia coli*.

### Chapter 2 CRISPR-Cas: Adapting to change |

This chapter focuses on CRISPR adaptation; the detailed process in which cells hosting a CRISPR system can insert small pieces of foreign DNA into their



CRISPR array to be stored for future defense. The process of adaptation requires several mechanistic steps and is crucial for ensuring protection of both the population and host. Here, we review the current mechanistic understanding of the adaptation process and the role of the well conserved Cas1-Cas2 protein complex.

### **Chapter 3 Using CAPTURE to detect spacer acquisition in native CRISPR arrays |**

Acquisition of spacers, to allow specific defense against certain invaders by the host CRISPR-Cas system, is a tightly controlled and low frequency process. In Chapter 3 building on previous techniques, we present a new method developed to detect rare spacer acquisition events in native settings. The method harnesses PCR amplification in combination with size extraction and specific primer design to allow detection of spacer acquisition in just 1 in  $10^5$  cells.

### **Chapter 4 Direct visualization of native CRISPR target search in live bacteria reveals Cascade DNA surveillance mechanism |**

In this chapter we focus on the target search process of Cascade, the crRNA guided surveillance complex of the type I-E CRISPR-Cas system in *Escherichia coli*. Using single-molecule tracking we visualize Cascade abundance and DNA probing kinetics in live cells. We observe that Cascade scans DNA through both PAM-dependent and -independent reactions, spending approximately 50% of its time bound to DNA. In addition, we investigate the relationship between Cascade copy number and interference levels finding more than 20 copies of Cascade are required for clearance in the majority of cells. Finally, we identify that the copy number dependence of Cascade follows a model where there is a tug-of-war between target replication and target search time upon invasion.

### **Chapter 5 Single cell variability in CRISPR-Cas adaptation and interference|**

In this chapter we develop a time-lapse microscopy approach that in combination with microfluidics and fluorescent reporters allows monitoring of the CRISPR response against a foreign plasmid. We show that while the priming process is highly variable, direct interference is comparatively deterministic, clearing the plasmid from all cells in a number of hours. Moreover, by comparing the kinetics of direct interference and primed interference we are able to define spacer acquisition as the origin of the variation in the priming process. Further, we experimentally identify a number of cellular factors such as growth rate, interdivision time and Cascade concentration that effect the efficiency of the cells response to foreign DNA. Finally, we develop a minimal agent-based model that can accurately reproduce our experimental data. This allows us to further probe



important additional factors of the primed adaptation process, giving insight into the influence of target copy number and Cascade variation within the population.

## 1

## References

1. Vellai, T. & Vida, G. The origin of eukaryotes: The difference between prokaryotic and eukaryotic cells. *Proceedings of the Royal Society B: Biological Sciences* (1999) doi:10.1098/rspb.1999.0817.
2. Suttle, C. A. Marine viruses - Major players in the global ecosystem. *Nat. Rev. Microbiol.* (2007) doi:10.1038/nrmicro1750.
3. van Houte, S., Buckling, A. & Westra, E. R. Evolutionary Ecology of Prokaryotic Immune Mechanisms. *Microbiol. Mol. Biol. Rev.* (2016) doi:10.1128/mmbr.00011-16.
4. Hampton, H. G., Watson, B. N. J. & Fineran, P. C. The arms race between bacteria and their phage foes. *Nature* **577**, 327–336 (2020).
5. De Smet, J., Hendrix, H., Blasdel, B. G., Danis-Wlodarczyk, K. & Lavigne, R. *Pseudomonas* predators: Understanding and exploiting phage-host interactions. *Nature Reviews Microbiology* (2017) doi:10.1038/nrmicro.2017.61.
6. Samson, J. E., Magadán, A. H., Sabri, M. & Moineau, S. Revenge of the phages: Defeating bacterial defences. *Nature Reviews Microbiology* (2013) doi:10.1038/nrmicro3096.
7. Koskella, B. & Brockhurst, M. A. Bacteria–phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiol. Rev.* **38**, 916–931 (2014).
8. Labrie, S. J., Samson, J. E. & Moineau, S. Bacteriophage resistance mechanisms. *Nature Reviews Microbiology* (2010) doi:10.1038/nrmicro2315.
9. Doron, S. *et al.* Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* (80-. ). (2018) doi:10.1126/science.aar4120.
10. Gao, L. *et al.* Diverse enzymatic activities mediate antiviral immunity in prokaryotes. *Science* (80-. ). (2020) doi:10.1126/science.aba0372.
11. Lu, M. J. & Henning, U. Superinfection exclusion by T-even-type coliphages. *Trends Microbiol.* (1994) doi:10.1016/0966-842X(94)90601-7.
12. Mahony, J., McGrath, S., Fitzgerald, G. F. & Van Sinderen, D. Identification and characterization of lactococcal-prophage-carried superinfection exclusion genes. *Appl. Environ. Microbiol.* (2008) doi:10.1128/AEM.01053-08.
13. Goldfarb, T. *et al.* BREX is a novel phage resistance system widespread in microbial genomes. *EMBO J.* (2015) doi:10.15252/embj.201489455.
14. Oliveira, P. H., Touchon, M. & Rocha, E. P. C. The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. *Nucleic Acids Res.* **42**, 10618–10631 (2014).
15. Depardieu, F. *et al.* A Eukaryotic-like Serine/Threonine Kinase Protects *Staphylococci* against Phages. *Cell Host Microbe* **20**, 471–481 (2016).
16. Chopin, M. C., Chopin, A. & Bidnenko, E. Phage abortive infection in lactococci:

- Variations on a theme. *Current Opinion in Microbiology* (2005) doi:10.1016/j.mib.2005.06.006.
17. Page, R. & Peti, W. Toxin-antitoxin systems in bacterial growth arrest and persistence. *Nat. Chem. Biol.* (2016) doi:10.1038/nchembio.2044.
  18. Barrangou, R. & Marraffini, L. A. CRISPR-Cas Systems: Prokaryotes Upgrade to Adaptive Immunity. *Mol. Cell* **54**, 234–244 (2014).
  19. Ledford, H. CRISPR treatment inserted directly into the body for first time. *Nature* **579**, 185–185 (2020).
  20. Wright, A. V., Nuñez, J. K. & Doudna, J. A. Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. *Cell* (2016) doi:10.1016/j.cell.2015.12.035.
  21. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakata, A. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J. Bacteriol.* **169**, 5429–33 (1987).
  22. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Soria, E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* **60**, 174–82 (2005).
  23. Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S. D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**, 2551–2561 (2005).
  24. Pourcel, C., Salvignol, G. & Vergnaud, G. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* **151**, 653–663 (2005).
  25. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–12 (2007).
  26. Brouns, S. J. J. *et al.* Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science* (80-. ). **321**, 960–964 (2008).
  27. 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) doi:10.1016/j.molcel.2012.03.018.
  28. Garneau, J. E. *et al.* The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67–71 (2010).
  29. Marraffini, L. A. & Sontheimer, E. J. CRISPR Interference Limits Horizontal Gene Transfer in *Staphylococci* by Targeting DNA. *Science* (80-. ). **322**, 1843–1845 (2008).
  30. Mohanraju, P. *et al.* Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science* (80-. ). **353**, aad5147 (2016).
  31. Makarova, K. S. *et al.* An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* **13**, 722–736 (2015).
  32. Makarova, K. S. *et al.* Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.* **18**, 67–83 (2020).
  33. Makarova, K. S. *et al.* Evolution and classification of the CRISPR–Cas systems. *Nat. Rev. Microbiol.* **9**, 467–477 (2011).
  34. Pourcel, C. *et al.* CRISPRCasdb a successor of CRISPRdb containing CRISPR arrays and cas genes from complete genome sequences, and tools to download

- and query lists of repeats and spacers. *Nucleic Acids Res.* (2019) doi:10.1093/nar/gkz915.
35. Medina-Aparicio, L., Dávila, S., Rebollar-Flores, J. E., Calva, E. & Hernández-Lucas, I. The CRISPR-Cas system in Enterobacteriaceae. *Pathog. Dis.* **76**, (2018).
  36. Xue, C. & Sashital, D. G. Mechanisms of Type I-E and I-F CRISPR-Cas Systems in Enterobacteriaceae. *EcoSal Plus* **8**, (2019).
  37. Yosef, I., Goren, M. G. & Qimron, U. Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. *Nucleic Acids Res.* **40**, 5569–5576 (2012).
  38. Dillard, K. E. *et al.* Assembly and Translocation of a CRISPR-Cas Primed Acquisition Complex. *Cell* **175**, 934–946.e15 (2018).
  39. Blosser, T. R. *et al.* Two Distinct DNA Binding Modes Guide Dual Roles of a CRISPR-Cas Protein Complex. *Mol. Cell* **58**, 60–70 (2015).
  40. Jackson, R. N. *et al.* Crystal structure of the CRISPR RNA-guided surveillance complex from Escherichia coli. *Science (80-. )*. **345**, 1473–1479 (2014).
  41. Hayes, R. P. *et al.* Structural basis for promiscuous PAM recognition in type I-E Cascade from E. coli. *Nature* **530**, 499–503 (2016).
  42. Jore, M. M. *et al.* Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat. Struct. Mol. Biol.* **18**, 529–536 (2011).
  43. Nuñez, J. K., Harrington, L. B., Kranzusch, P. J., Engelman, A. N. & Doudna, J. A. Foreign DNA capture during CRISPR-Cas adaptive immunity. *Nature* **527**, 535–8 (2015).
  44. Wang, J. *et al.* Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems. *Cell* (2015) doi:10.1016/j.cell.2015.10.008.
  45. Pougach, K. *et al.* Transcription, processing and function of CRISPR cassettes in Escherichia coli. *Mol. Microbiol.* **77**, 1367–1379 (2010).
  46. Díez-Villaseñor, C., Almendros, C., García-Martínez, J. & Mojica, F. J. M. Diversity of CRISPR loci in Escherichia coli. *Microbiology* **156**, 1351–1361 (2010).
  47. Pul, U. *et al.* Identification and characterization of E. coli CRISPR-cas promoters and their silencing by H-NS. *Mol. Microbiol.* **75**, 1495–512 (2010).
  48. Westra, E. R. *et al.* H-NS-mediated repression of CRISPR-based immunity in Escherichia coli K12 can be relieved by the transcription activator LeuO. *Mol. Microbiol.* **77**, 1380–93 (2010).
  49. Musharova, O. *et al.* Systematic analysis of Type I-E Escherichia coli CRISPR-Cas PAM sequences ability to promote interference and primed adaptation. *Mol. Microbiol.* **111**, 1558–1570 (2019).
  50. Fineran, P. C. *et al.* Degenerate target sites mediate rapid primed CRISPR adaptation. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E1629–38 (2014).
  51. Semenova, E. *et al.* Highly efficient primed spacer acquisition from targets destroyed by the Escherichia coli type I-E CRISPR-Cas interfering complex. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 7626–31 (2016).
  52. Künne, T. *et al.* Cas3-Derived Target DNA Degradation Fragments Fuel Primed CRISPR Adaptation. *Mol. Cell* **63**, 852–864 (2016).
  53. Severinov, K., Ispolatov, I. & Semenova, E. The Influence of Copy-Number of Targeted Extrachromosomal Genetic Elements on the Outcome of CRISPR-Cas

- Defense. *Front. Mol. Biosci.* **3**, (2016).
54. Vink, J. N. A. *et al.* Direct Visualization of Native CRISPR Target Search in Live Bacteria Reveals Cascade DNA Surveillance Mechanism. *Mol. Cell* **77**, 39-50.e10 (2020).
  55. Mamontov, V. *et al.* Long-Term Persistence of Plasmids Targeted by CRISPR Interference in Bacterial Populations. *bioRxiv* (2021).
  56. Strotskaya, A. *et al.* The action of Escherichia coli CRISPR-Cas system on lytic bacteriophages with different lifestyles and development strategies. *Nucleic Acids Res.* (2017) doi:10.1093/nar/gkx042.
  57. Deveau, H. *et al.* Phage Response to CRISPR-Encoded Resistance in *Streptococcus thermophilus*. *J. Bacteriol.* **190**, 1390–1400 (2008).
  58. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733–740 (2009).
  59. Leenay, R. T. *et al.* Identifying and Visualizing Functional PAM Diversity across CRISPR-Cas Systems. *Mol. Cell* **62**, 137–147 (2016).
  60. Sashital, D. G., Wiedenheft, B. & Doudna, J. A. Mechanism of Foreign DNA Selection in a Bacterial Adaptive Immune System. *Mol. Cell* **46**, 606–615 (2012).
  61. Redding, S. *et al.* Surveillance and Processing of Foreign DNA by the Escherichia coli CRISPR-Cas System Single-molecule analysis of the bacterial Cascade complex reveals two distinct pathways leading to differential regulation of Cas3 DNA translocase and nuclease activities. *Cell* **163**, 854–865 (2015).
  62. Xue, C., Zhu, Y., Zhang, X., Shin, Y.-K. & Sashital, D. G. Real-Time Observation of Target Search by the CRISPR Surveillance Complex Cascade. *Cell Rep.* **21**, 3717–3727 (2017).
  63. van Erp, P. B. G. *et al.* Mechanism of CRISPR-RNA guided recognition of DNA targets in Escherichia coli. *Nucleic Acids Res.* **43**, 8381–8391 (2015).
  64. Xiao, Y. *et al.* Structure Basis for Directional R-loop Formation and Substrate Handover Mechanisms in Type I CRISPR-Cas System. *Cell* **170**, 48-60.e11 (2017).
  65. Semenova, E. *et al.* Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 10098–103 (2011).
  66. Rutkauskas, M. *et al.* Directional R-Loop Formation by the CRISPR-Cas Surveillance Complex Cascade Provides Efficient Off-Target Site Rejection. *Cell Rep.* **10**, 1534–1543 (2015).
  67. Klein, M., Eslami-Mossallam, B., Arroyo, D. G. & Depken, M. Hybridization Kinetics Explains CRISPR-Cas Off-Targeting Rules. *Cell Rep.* **22**, 1413–1423 (2018).
  68. Hochstrasser, M. L. *et al.* CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. *Proc. Natl. Acad. Sci.* **111**, 6618–6623 (2014).
  69. Mulepati, S. & Bailey, S. In vitro reconstitution of an Escherichia coli RNA-guided immune system reveals unidirectional, ATP-dependent degradation of DNA Target. *J. Biol. Chem.* (2013) doi:10.1074/jbc.M113.472233.
  70. Loeff, L., Brouns, S. J. J. & Joo, C. Repetitive DNA Reeling by the Cascade-Cas3

- Complex in Nucleotide Unwinding Steps. *Mol. Cell* **70**, 385-394.e3 (2018).
71. Cooper, L. A., Stringer, A. M. & Wade, J. T. Determining the Specificity of Cascade Binding, Interference, and Primed Adaptation In Vivo in the Escherichia coli Type I-E CRISPR-Cas System. *MBio* **9**, e02100-17 (2018).
  72. Jackson, S. A. *et al.* CRISPR-Cas: Adapting to change. *Science* (80-. ). **356**, eaal5056 (2017).
  73. Swarts, D. C., Mosterd, C., van Passel, M. W. J. & Brouns, S. J. J. CRISPR Interference Directs Strand Specific Spacer Acquisition. *PLoS One* **7**, e35888 (2012).
  74. Datsenko, K. A. *et al.* Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat. Commun.* **3**, 945 (2012).
  75. Stringer, A. M., Cooper, L. A., Kadaba, S., Shrestha, S. & Wade, J. T. Characterization of primed adaptation in the escherichia coli type I-E CRISPR-cas system. *bioRxiv* (2020) doi:10.1101/2020.02.10.942821.
  76. Kim, S. *et al.* Selective loading and processing of prespacers for precise CRISPR adaptation. *Nature* **579**, 141–145 (2020).
  77. Yosef, I. *et al.* DNA motifs determining the efficiency of adaptation into the Escherichia coli CRISPR array. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 14396–401 (2013).
  78. Xue, C. *et al.* CRISPR interference and priming varies with individual spacer sequences. *Nucleic Acids Res.* **43**, 10831–10847 (2015).
  79. Musharova, O. *et al.* Spacer-length DNA intermediates are associated with Cas1 in cells undergoing primed CRISPR adaptation. *Nucleic Acids Res.* (2017) doi:10.1093/nar/gkx097.
  80. Shiriaeva, A. A. *et al.* Detection of spacer precursors formed in vivo during primed CRISPR adaptation. *Nat. Commun.* **10**, 4603 (2019).
  81. Musharova, O. *et al.* Prespacers formed during primed adaptation associate with the Cas1–Cas2 adaptation complex and the Cas3 interference nuclease–helicase. *PNAS* (2021) doi:https://doi.org/10.1073/pnas.2021291118.
  82. Edgar, R. & Qimron, U. The Escherichia coli CRISPR system protects from  $\lambda$  lysogenization, lysogens, and prophage induction. *J. Bacteriol.* (2010) doi:10.1128/JB.00644-10.
  83. Heler, R. *et al.* Cas9 specifies functional viral targets during CRISPR–Cas adaptation. *Nature* **519**, 199–202 (2015).
  84. Richter, C. *et al.* Priming in the Type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. *Nucleic Acids Res.* **42**, 8516–26 (2014).
  85. McKenzie, R. E., Almendros, C., Vink, J. N. A. & Brouns, S. J. J. Using CAPTURE to detect spacer acquisition in native CRISPR arrays. *Nat. Protoc.* **14**, (2019).
  86. Staals, R. H. J. *et al.* Interference-driven spacer acquisition is dominant over naive and primed adaptation in a native CRISPR-Cas system. *Nat. Commun.* **7**, 12853 (2016).
  87. Modell, J. W., Jiang, W. & Marraffini, L. A. CRISPR–Cas systems exploit viral DNA injection to establish and maintain adaptive immunity. *Nature* **544**, 101–104 (2017).
  88. Amlinger, L., Hoekzema, M., Wagner, E. G. H., Koskiniemi, S. & Lundgren, M.

- Fluorescent CRISPR Adaptation Reporter for rapid quantification of spacer acquisition. *Sci. Rep.* **7**, 10392 (2017).
89. Díez-Villaseñor, C., Guzmán, N. M., Almendros, C., García-Martínez, J. & Mojica, F. J. M. CRISPR-spacer integration reporter plasmids reveal distinct genuine acquisition specificities among CRISPR-Cas I-E variants of *Escherichia coli*. *RNA Biol.* **10**, 792–802 (2013).
  90. Schmidt, F., Cherepkova, M. Y. & Platt, R. J. Transcriptional recording by CRISPR spacer acquisition from RNA. *Nature* **562**, 380–385 (2018).
  91. Jackson, S. A., Birkholz, N., Malone, L. M. & Fineran, P. C. Imprecise Spacer Acquisition Generates CRISPR-Cas Immune Diversity through Primed Adaptation. *Cell Host Microbe* **25**, 250-260.e4 (2019).
  92. Ha, T. Single-molecule methods leap ahead. *Nat. Methods* **11**, 1015–1018 (2014).
  93. Newton, M. D. *et al.* DNA stretching induces Cas9 off-target activity. *Nat. Struct. Mol. Biol.* (2019) doi:10.1038/s41594-019-0188-z.
  94. Krivoy, A. *et al.* Primed CRISPR adaptation in *Escherichia coli* cells does not depend on conformational changes in the Cascade effector complex detected in Vitro. *Nucleic Acids Res.* **46**, 4087–4098 (2018).
  95. Szczelkun, M. D. *et al.* Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc. Natl. Acad. Sci.* **111**, 9798–9803 (2014).
  96. Martens, K. J. A. *et al.* Visualisation of dCas9 target search in vivo using an open-microscopy framework. *Nat. Commun.* (2019) doi:10.1038/s41467-019-11514-0.
  97. Jones, D. L. *et al.* Kinetics of dCas9 target search in *Escherichia coli*. *Science (80-. )*. (2017) doi:10.1126/science.aah7084.
  98. Globyte, V., Lee, S. H., Bae, T., Kim, J. & Joo, C. CRISPR/Cas9 searches for a protospacer adjacent motif by lateral diffusion. *EMBO J.* e99466 (2018) doi:10.15252/embj.201899466.
  99. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62–67 (2014).
  100. Elowitz, M. B. Stochastic Gene Expression in a Single Cell. *Science (80-. )*. **297**, 1183–1186 (2002).
  101. Kiviet, D. J. *et al.* Stochasticity of metabolism and growth at the single-cell level. *Nature* **514**, 376–379 (2014).
  102. Young, J. W. *et al.* Measuring single-cell gene expression dynamics in bacteria using fluorescence time-lapse microscopy. *Nat. Protoc.* **7**, 80–88 (2012).
  103. Wang, P. *et al.* Robust growth of *Escherichia coli*. *Curr. Biol.* (2010) doi:10.1016/j.cub.2010.04.045.
  104. Locke, J. C. W. & Elowitz, M. B. Using movies to analyse gene circuit dynamics in single cells. *Nature Reviews Microbiology* (2009) doi:10.1038/nrmicro2056.
  105. Bennett, M. R. & Hasty, J. Microfluidic devices for measuring gene network dynamics in single cells. *Nature Reviews Genetics* (2009) doi:10.1038/nrg2625.
  106. Uphoff, S. *et al.* Stochastic activation of a DNA damage response causes cell-to-cell mutation rate variation. *Science (80-. )*. **351**, 1094–1097 (2016).
  107. Balaban, N. Q., Merrin, J., Chait, R., Kowalik, L. & Leibler, S. Bacterial

- persistence as a phenotypic switch. *Science* **305**, 1622–5 (2004).
108. Cerulus, B., New, A. M., Pougach, K. & Verstrepen, K. J. Noise and Epigenetic Inheritance of Single-Cell Division Times Influence Population Fitness. *Curr. Biol.* **26**, 1138–1147 (2016).
  109. Hashimoto, M. *et al.* Noise-driven growth rate gain in clonal cellular populations. *Proc. Natl. Acad. Sci.* **113**, 3251–3256 (2016).
  110. Moormeier, D. E., Bose, J. L., Horswill, A. R. & Bayles, K. W. Temporal and Stochastic Control of *Staphylococcus aureus* Biofilm Development. *MBio* **5**, (2014).
  111. Watson, B. N. J. *et al.* Type I-F CRISPR-Cas resistance against virulent phages results in abortive infection and provides population-level immunity. *Nat. Commun.* (2019) doi:10.1038/s41467-019-13445-2.





# 2

**CRISPR-Cas:**

**Adapting to change**

Published as: S. A. Jackson\*, [R. E. McKenzie\\*](#), R. D. Fagerlund, S. N. Kieper, P. C. Fineran, S. J. J. Brouns, CRISPR-Cas: Adapting to change. *Science* (80-. ). **356**, eaal5056 (2017).

\* These authors contributed equally



## **Abstract**

Bacteria and archaea are engaged in a constant arms race to defend against the ever-present threats of viruses and invasion by mobile genetic elements. The most flexible weapons in the prokaryotic defense arsenal are the CRISPR-Cas adaptive immune systems. These systems are capable of selective identification and neutralization of foreign DNA and or RNA. CRISPR-Cas systems rely on stored genetic memories to facilitate target recognition. Thus, to keep pace with a changing pool of hostile invaders, the CRISPR memory banks must be regularly updated by the addition of new information, through a process termed CRISPR adaptation. In this review, we outline the recent advances in our understanding of the molecular mechanisms governing CRISPR adaptation. Specifically, the conserved protein machinery, Cas1-Cas2, provides the cornerstone of adaptive immunity in a range of diverse CRISPR-Cas systems.

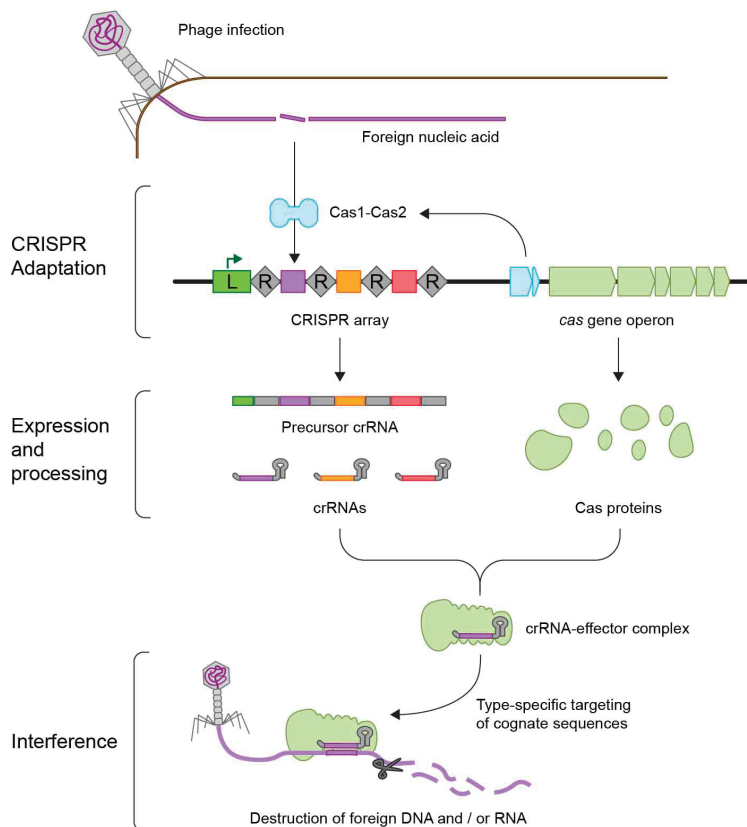
## 2.1 Introduction

Bacteria and archaea are constantly threatened by phage infection and invasion by mobile genetic elements (MGEs) through conjugation and transformation. In response, a defense arsenal has evolved, including various ‘innate’ mechanisms and the CRISPR-Cas adaptive immune systems<sup>1-3</sup>. CRISPR-Cas systems are widely distributed, occurring in 50% and 87% of complete bacterial and archaeal genomes, respectively. These systems function as RNA-guided nucleases that provide sequence-specific defense against invading MGEs<sup>4,5</sup>. The repurposing of these sequence-specific Cas nucleases, particularly Cas9, has stimulated a biotechnological revolution in genome editing that has resulted in breakthroughs across many biological fields<sup>6</sup>. In native hosts, the advantage conferred by CRISPR-Cas systems over innate defenses lies in the ability to update their resistance repertoire in response to infection (termed CRISPR adaptation). CRISPR adaptation is achieved by incorporating short DNA fragments from MGEs into CRISPR arrays to form memory units termed spacers. Early bioinformatic studies showed many spacers were of foreign origin, hinting that CRISPR loci may act as a form of memory for a prokaryotic immune system<sup>7-10</sup>. Subsequent confirmation of the link between spacers and resistance to phage and MGEs was gained experimentally<sup>4,5,11</sup>. An overview of CRISPR-Cas-mediated defense and CRISPR adaptation is provided in **Figure 2.1**.

### 2.1.1 Red Queen CRISPR adaptation

The ability to keep defenses up to date, by acquiring new spacers, is central to the success of CRISPR-Cas systems. Typically, new spacers are inserted at a specific end of the CRISPR array, adjacent to a ‘leader’ region that contains conserved sequence motifs<sup>4,12-14</sup>. The leader usually also contains the promoter driving CRISPR transcription and it has been demonstrated that integration of new spacers at the leader end enhances defense against phages and MGEs encountered recently<sup>15</sup>. This ‘polarized’ addition of spacers into CRISPR loci produces a chronological account of the encounters between phages and bacteria that can provide insights into phage-host co-occurrences, evolution and ecology<sup>16,17</sup>. However, phage and MGE variants with genetic mutations can avoid detection by existing CRISPR spacers – these evaders are termed ‘escape mutants’. Additionally,

spacers can be lost from CRISPR arrays by recombination between the repeats<sup>16,18</sup>.

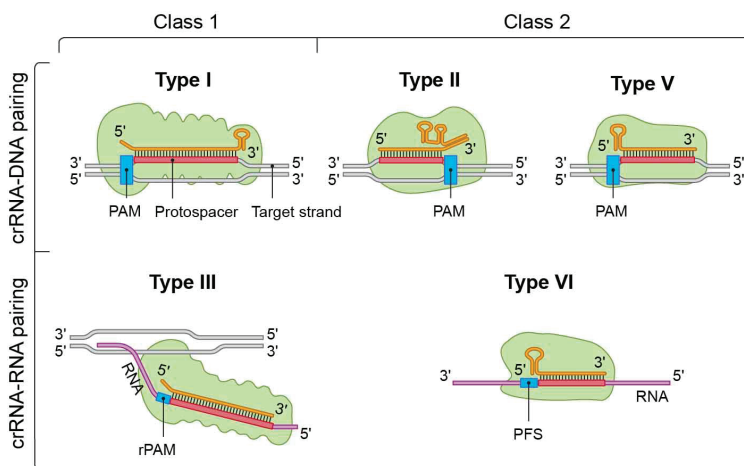


**Figure 2.1: A roadmap of CRISPR-Cas adaptation and defense.** In the example illustrated, a bacterial cell is infected by a bacteriophage. The first stage of CRISPR-Cas defense is CRISPR adaptation. This involves the incorporation of small fragments of DNA from the invader into the host CRISPR array. This forms a genetic “memory” of the infection. The memories are stored as spacers (colored squares) between repeat sequences (R) and new spacers are added at the leader-proximal (L) end of the array. The Cas1 and Cas2 proteins, encoded within the *cas* gene operon, form a Cas1-Cas2 complex (blue) – the ‘workhorse’ of CRISPR adaptation. In this example, the Cas1-Cas2 complex catalyzes the addition of a spacer from the phage genome (purple) into the CRISPR array. The second stage of CRISPR-Cas defense involves transcription of the CRISPR array and subsequent processing of the precursor transcript to generate CRISPR RNAs (crRNAs). Each crRNA contains a single spacer unit that is typically flanked by parts of the adjoining repeat sequences (grey). Individual crRNAs assemble with Cas effector proteins (green) to form crRNA-effector complexes. The crRNA-effector complexes catalyze the sequence-specific recognition and destruction of foreign DNA and / or RNA elements. This process is known as interference.

Thus, maintenance of CRISPR-Cas defense is reliant on the addition of new spacers into CRISPR arrays <sup>19,20</sup>. The continuous competition between host CRISPR adaptation and MGE escape, akin to Red Queen dynamics, have been exposed in several recent metagenome studies <sup>21,22</sup>. Individual cells within a prokaryotic community acquire different, and often multiple spacers during CRISPR adaptation <sup>23,24</sup>. The diversity of CRISPR loci within cell populations optimizes defense by limiting the reproductive success of mutants that escape the CRISPR-Cas defenses of individual cells <sup>25</sup>. Furthermore, the resulting polymorphisms in CRISPR loci enable fast and accurate differentiation of species subtypes, which may prove to have economic and clinical benefits - for example, enabling tracing of pathogens during outbreaks <sup>7,26</sup>.

### 2.1.2 Origins of CRISPR adaptation

According to their constituent Cas proteins, CRISPR-Cas systems are classified into two major classes consisting of six types and nineteen subtypes (**Fig 2.2**) <sup>27,28</sup>. Comparative genomics indicate that all known CRISPR-Cas systems evolved from a single ancestor <sup>27,28</sup>. The more compact class 2 CRISPR-Cas systems likely evolved from class 1 ancestors, through acquisition of genes encoding new single-subunit effector proteins and loss of additional *cas* genes <sup>28</sup>. However, despite the divergence of CRISPR-Cas systems into several types, the proteins primarily responsible for catalyzing spacer acquisition, namely Cas1 and Cas2, remain relatively conserved, and the genes encoding these proteins are associated with nearly all CRISPR-Cas systems <sup>27</sup>. Indeed, as long as spacers can be acquired from MGEs, unique effector machineries capable of utilizing the information stored in CRISPRs are likely to arise. Knowledge of the structure and function of CRISPR-Cas effector complexes has advanced rapidly in recent years <sup>28,29</sup>. In addition, significant progress has been made lately toward elucidating the molecular basis of how, when and why CRISPR adaptation occurs. Here, we review these recent findings and highlight the insights they shed on the function of different CRISPR adaptation mechanisms employed by diverse CRISPR-Cas systems.



**Figure 2.2: Target interactions and the PAMs of diverse CRISPR-Cas types.** Recognition of the invading DNA target by the crRNA-Cas effector complexes of types I, II and V, results in the formation of an RNA-DNA hybrid in which the non-target DNA strand is displaced. The target strand contains the protospacer (red), which is complementary to the spacer sequence (crRNA, orange). The protospacer adjacent motif (PAM, blue) is located at either the 3' end of the protospacer (type I and type V) or the 5' end (type II). Type III and VI recognize RNA targets, with type III exhibiting additional transcription-dependent DNA targeting. Some type III systems require an RNA-based PAM (rPAM). Type VI systems exhibit specificity for a protospacer flanking sequence (PFS) motif, which is analogous to a PAM.

## 2.2 The molecular basis of CRISPR adaptation

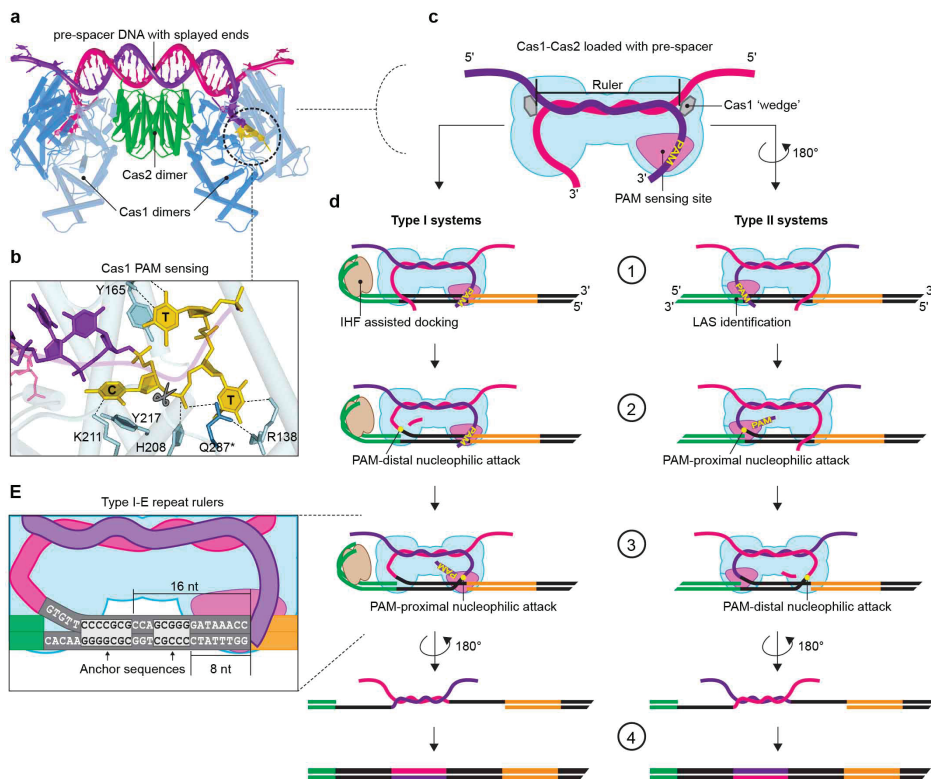
CRISPR adaptation requires the integration of new spacers into CRISPR loci and duplication of the associated repeat sequences. The Cas1 and Cas2 proteins, which form a Cas1<sub>4</sub>-Cas2<sub>2</sub> complex<sup>30,31</sup> (hereafter Cas1-Cas2) constitute the 'workhorse' of spacer integration. Spacers added to CRISPR arrays must be compatible with the diverse range of type-specific effector complex machinery (**Fig. 2.2**). Thus, despite being near-ubiquitous amongst CRISPR-Cas types, Cas1-Cas2 homologues meet the varied requirements for the acquisition of appropriate spacer sequences in different systems. For example, the effector complexes of several CRISPR-Cas types only recognize targets containing a specific sequence adjacent to where the crRNA basepairs with the target-strand of a MGE (**Fig. 2.2**)<sup>32</sup>. The crRNA-paired target sequence is termed the protospacer and the adjacent target-recognition motif is called a protospacer adjacent motif (PAM)<sup>33</sup>. PAM-based

target discrimination prevents the unintentional recognition and self-destruction of the CRISPR-locus by the crRNA-effector complex, yet canonical PAM sequences vary between, and sometimes even within systems.

Much of what we know about the Cas1-Cas2 molecular structure and function has been gained from studies in the *Escherichia coli* type I-E system. Within the Cas1<sub>4</sub>-Cas2<sub>2</sub> complex the Cas1 subunits form two dimers that are bridged by a central Cas2 dimer (**Fig. 2.3a**)<sup>30,34,35</sup>. Cas1-Cas2-mediated spacer integration prefers dsDNA substrates and proceeds via a mechanism resembling retroviral integration<sup>36,37</sup>. In addition to Cas1-Cas2, at least one CRISPR repeat, part of the leader sequence<sup>12,13,15,38</sup>, and several host factors for repair of the insertion sites (e.g., DNA polymerase) are required<sup>39</sup>. Spacer acquisition involves three main processes: substrate capture, recognition of the CRISPR locus and integration within the array.

### 2.2.1 Cas1-Cas2 substrate capture

During substrate capture, Cas1-Cas2 is loaded with an integration-compatible pre-spacer, which is thought to be partially duplexed dsDNA<sup>36</sup>. For type I systems, the presence of a canonical PAM within the pre-spacer substrate increases the affinity for Cas1-Cas2 binding, yet is not requisite<sup>34</sup>. Details of how pre-spacer substrates are produced from foreign DNA is discussed later. For the *E. coli* type I-E Cas1-Cas2:pre-spacer complex, the ends of the dsDNA pre-spacer are splayed by tyrosine wedges in each Cas1 dimer, which lock open the DNA branch points while fixing in place a core 23 bp dsDNA region. The 3' single-stranded ends of the pre-spacer extend into active subunits of each corresponding Cas1 dimer (**Fig. 2.3a-c**)<sup>34,35</sup>. The length of new spacers is governed by the fixed distances between the two Cas1 wedges and from the branch points to the integrase sites. Many CRISPR-Cas systems display highly consistent, yet system-specific, spacer lengths and it is likely that analogous wedge-based Cas1-Cas2 'molecular rulers' exist in these systems to control pre-spacer length<sup>34,35</sup>. However, in some systems, such as type III, the length of spacers found within CRISPR arrays appears more variable and studies of Cas1-Cas2 structure and function for these systems are currently lacking.



**Figure 2.3: Cas1-Cas2-mediated spacer acquisition.** **a**, The Cas1-Cas2 protein complex loaded with a pre-spacer substrate (*E. coli* type I-E structure shown; PDB 5DQZ). **b**, The Cas1 protospacer adjacent motif (PAM) sensing site shows the canonical type I-E PAM (CTT, yellow), residue-specific interactions (a residue from the non-catalytic Cas1 monomer is annotated with \*) and the site of PAM processing (scissors). **c**, A schematic representation of the substrate loaded Cas1-Cas2 protein complex with the active PAM sensing site highlighted (light purple) and a partially duplexed DNA pre-spacer substrate (strands are purple and pink). The ruler mechanism determining spacer length for the *E. coli* type I-E system uses two conserved tyrosine residues (i.e. the “Cas1 wedge”, grey hexagons). **d**, Spacer integration proceeds as follows: 1) the Cas1-Cas2:pre-spacer complex binds to the leader (green) and first repeat (black). For type I and type II systems, respectively, Cas1-Cas2 docking to the leader is assisted by integration host factor (IHF) or recognition of the leader adjacent sequence (LAS). 2) The first nucleophilic attack most likely occurs at the leader-repeat junction and gives rise to a half-site intermediate. 3) The second nucleophilic attack occurs at the repeat-spacer (orange) boundary, resulting in full site integration. 4) Host DNA repair enzymes fill the integration site. **(E)** The type I-E repeat is magnified (lower left) to indicate the inverted repeats within its sequence and highlight the anchoring sites of the molecular rulers that determine the point of integration.

### 2.2.1 Cas1-Cas2 substrate capture

During substrate capture, Cas1-Cas2 is loaded with an integration-compatible pre-spacer, which is thought to be partially duplexed dsDNA <sup>36</sup>. For type I systems, the presence of a canonical PAM within the pre-spacer substrate increases the affinity for Cas1-Cas2 binding, yet is not requisite <sup>34</sup>. Details of how pre-spacer substrates are produced from foreign DNA is discussed later. For the *E. coli* type I-E Cas1-Cas2:pre-spacer complex, the ends of the dsDNA pre-spacer are splayed by tyrosine wedges in each Cas1 dimer, which lock open the DNA branch points while fixing in place a core 23 bp dsDNA region. The 3' single-stranded ends of the pre-spacer extend into active subunits of each corresponding Cas1 dimer (**Fig. 2.3a-c**) <sup>34,35</sup>. The length of new spacers is governed by the fixed distances between the two Cas1 wedges and from the branch points to the integrase sites. Many CRISPR-Cas systems display highly consistent, yet system-specific, spacer lengths and it is likely that analogous wedge-based Cas1-Cas2 'molecular rulers' exist in these systems to control pre-spacer length <sup>34,35</sup>. However, in some systems, such as type III, the length of spacers found within CRISPR arrays appears more variable and studies of Cas1-Cas2 structure and function for these systems are currently lacking.

### 2.2.2 Recognition of the CRISPR array

Prior to integration, the substrate-bound Cas1-Cas2 complex must locate the CRISPR leader-repeat sequence. Specific sequences upstream of CRISPR arrays direct leader-polarized spacer integration, both via direct Cas1-Cas2 recognition and assisted by host proteins. The Cas1-Cas2 complexes of several systems display intrinsic affinity for the leader-repeat region in vitro <sup>36,40</sup>, yet this is not always wholly sufficient to provide the specificity observed in vivo. It was recently discovered that for the type I-E system, leader-repeat recognition is assisted by the integration host factor (IHF) heterodimer <sup>41</sup>. IHF binds the CRISPR leader in a sequence-specific manner and induces ~120° DNA bending, providing a cue to accurately localize Cas1-Cas2 to the leader-repeat junction <sup>41,42</sup>. A conserved sequence motif upstream of the IHF pivot is proposed to stabilize the Cas1-Cas2-leader-repeat interaction and increase adaptation efficiency, supporting binding of the adaptation complex to DNA sites either side of bound IHF <sup>42</sup>.



IHF is absent in many prokaryotes, including archaea, suggesting other leader-proximal integration mechanisms exist. Indeed, type II-A Cas1-Cas2 from *Streptococcus pyogenes* catalyzed leader-proximal integration in vitro, at a level of precision comparable to the type I-E system with IHF<sup>40,41</sup>. In type II systems, a short leader-anchoring site (LAS) adjacent to the first repeat and  $\leq 6$  bp of this repeat are essential for adaptation<sup>15,38,40</sup> and are conserved in systems with similar repeats. Placement of an additional LAS in front of a non-leader repeat resulted in adaptation at both sites<sup>38</sup>, whereas LAS deletion caused ectopic integration at a downstream repeat adjacent to a spacer containing a LAS-like sequence<sup>15</sup>. Hence, in contrast to type I-E, type II-A systems appear to rely solely on intrinsic sequence specificity for the leader-repeat.

### 2.2.3 Integration into the CRISPR array

For CRISPR-Cas types that are reliant on PAM sequences for recognition of targets, the acquisition of interference-proficient spacers requires processing of the pre-spacer substrate at a specific position relative to the PAM. Each of the four Cas1 monomers in the Cas1-Cas2 complex contains a PAM sensing domain. The presence of a PAM in the active site of just one of the Cas1 monomers is sufficient to appropriately position the substrate and PAM relative to the cleavage site (**Fig. 2.3b**)<sup>34,35</sup>. Furthermore, the presence of a PAM within the pre-spacer substrate ensures integration into the CRISPR in the correct orientation<sup>23,43-45</sup>. This directional fidelity is critical because otherwise the PAM in the MGE target would lie at the wrong end of the crRNA target binding site, thus abrogating target recognition (**Fig. 2.2**). To avoid premature loss of the PAM directional cue, processing of the pre-spacer likely occurs after Cas1-Cas2 orients and docks at the leader-repeat (**Fig. 2.3d**). Cas1-mediated processing of the pre-spacer creates two 3'OH ends required for nucleophilic attack on each strand of the leader-proximal repeat<sup>36,37,46</sup>. The initial nucleophilic attack most likely occurs at the leader-repeat junction and forms a half-site intermediate, then a second attack at the existing repeat-spacer junction generates the full-site integration product (**Fig. 2.3d**). The precise order of the pre-spacer processing and integration steps remains to be fully determined, yet considerable progress toward elucidating the reaction mechanisms has been made.

Following the first nucleophilic attack, the intrinsic sequence-specificity of the Cas1-Cas2 complex defines the site of the second attack and ensures accurate repeat duplication. CRISPR repeats are often semi-palindromic,

containing two short inverted repeat (IR) elements, but the location of these can vary<sup>47</sup>. In type I-B and I-E systems, the IRs occur close to the center of the repeat (**Fig. 2.3e**) and are important for spacer acquisition<sup>48,49</sup>. In the type I-E system, both IRs act as anchors for the Cas1-Cas2 complex, which contains two ‘molecular rulers’ to position the Cas1 active site for the second nucleophilic attack at the repeat-spacer boundary<sup>48</sup>. However, in the type I-B system from *Haloarcula hispanica*, only the first IR is essential for integration and a single molecular ruler, directed by an anchor between the IRs, has been proposed<sup>49</sup>. In the type II-A systems of *Streptococcus thermophilus* and *S. pyogenes* the IRs are located distally within the repeats, suggesting these short sequences may directly position the nucleophilic attacks without a need for molecular rulers<sup>38,40</sup>. Although these recent findings suggest that leader-repeat regions at the beginning of CRISPR arrays contain sequences to ensure appropriate Cas1-Cas2 localization, further work is required to determine how the spacer integration events are specifically orchestrated in the diverse range of CRISPR-Cas types.

## 2.3 Production of pre-spacers from foreign DNA

Despite the elegance of memory-directed defense, CRISPR adaptation is not without complications. For example, the inadvertent acquisition of spacers from host DNA must be avoided because this will result in cytotoxic self-targeting – akin to autoimmunity in eukaryotic adaptive immune systems<sup>50,51</sup>. Therefore, production of pre-spacer substrates from MGEs should outweigh production from host DNA. In the following sections, we outline the routes to pre-spacer generation in different CRISPR-Cas systems.

### 2.3.1 Naïve CRISPR adaptation

Acquisition of spacers from MGEs that are not already catalogued in host CRISPRs is termed naïve CRISPR adaptation (**Fig. 2.4**)<sup>52</sup>. To facilitate naïve CRISPR adaptation, pre-spacer substrates are generated from foreign material and loaded onto Cas1-Cas2. Currently, the main known source of these precursors is the host RecBCD complex<sup>53</sup>. Stalled replication forks that occur during DNA replication can result in double strand breaks (DSBs), which are repaired via RecBCD-mediated unwinding and degradation of the dsDNA ends back to the nearest Chi sites<sup>54</sup>. During this repair

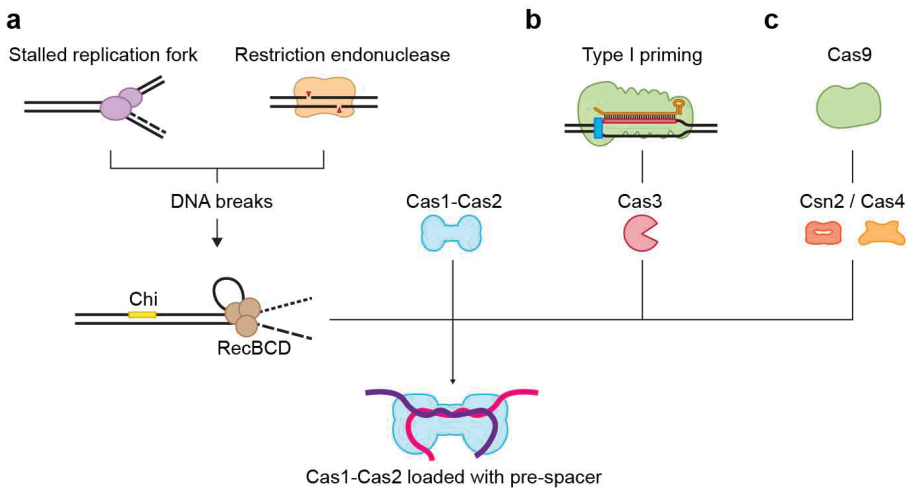
process, RecBCD produces ssDNA fragments, which have been proposed to subsequently anneal to form partially duplexed pre-spacer substrates for Cas1-Cas2<sup>53</sup>. The greater number of active origins of replication and the paucity of Chi sites on MGEs, compared with the host chromosome, biases naïve adaptation toward foreign DNA. Furthermore, RecBCD recognizes the unprotected dsDNA ends that are commonly present in phage genomes upon injection or prior to packaging, which theoretically provides an additional phage-specific source of naïve pre-spacer substrates<sup>53</sup>.

Despite the role of RecBCD in substrate generation, naïve CRISPR adaptation can occur in its absence, albeit with reduced bias towards foreign DNA<sup>53</sup>. Thus, events other than DSBs might also stimulate naïve CRISPR adaptation, such as R-loops that prime plasmid replication, lagging ends of incoming conjugative elements<sup>55</sup>, and even CRISPR-Cas-mediated spacer integration events themselves<sup>23,53</sup>. Furthermore, we do not know whether all CRISPR-Cas systems display an intrinsic bias towards production of pre-spacers from foreign DNA. In high throughput studies of native systems, the frequency of acquisition of spacers from host genomes is likely to be underestimated, because the autoimmunity resulting from self-targeting spacers means these genotypes are typically lethal<sup>23,50,51,56</sup>. For example, in the *S. thermophilus* type II-A system, spacer acquisition appears biased toward MGEs, yet nuclease-deficient Cas9 (dCas9) fails to discriminate between host and foreign DNA<sup>56</sup>. It is unknown whether CRISPR adaptation in type II systems is reliant on DNA break repair. Further studies in a range of host systems are required to clarify how diverse CRISPR-Cas systems balance the requirement for naïve production of pre-spacers from MGEs against the risk of acquiring spacers from host DNA.

### 2.3.2 crRNA-directed CRISPR adaptation (Priming)

Mutations in the target PAM or protospacer sequences can abrogate immunity, allowing MGEs to escape CRISPR-Cas defenses<sup>57-59</sup>. Furthermore, the protection conferred by individual spacers varies: often several MGE-specific spacers are required to mount an effective defense<sup>24,60</sup> and to prevent proliferation of escape mutants<sup>17,25</sup>. Thus, CRISPR-Cas systems need to undergo CRISPR adaptation faster than MGEs can evade targeting. Indeed, type I systems have evolved a mechanism known as primed CRISPR adaptation (priming) to facilitate rapid spacer acquisition<sup>45,61</sup>, even against highly divergent invaders<sup>59</sup> (**Fig. 2.4**). Priming utilizes

MGE target recognition that is facilitated by pre-existing spacers to trigger the acquisition of additional spacers from previously-encountered elements.



**Figure 2.4: Cas1-Cas2 substrate production pathways.** **a**, Naïve generation of substrates by RecBCD activity on DNA ends resulting from DSBs that occur as a result of stalled replication forks, innate defenses such as restriction endonuclease activity, or from the ends of phage genomes (not shown). **b**, Primed pre-spacer production in type I systems, which requires Cas3 helicase and nuclease activity. **c**, Cas9-dependent spacer selection in type II systems, which for some subtypes is dependent on the activity of accessory proteins, such as Csn2 or Cas4. The PAM specificity of the Cas9 protein determines the selection of PAMs in pre-spacer substrates.

Thus, priming is advantageous when MGE replication within the host cell exceeds current defense capabilities. This can occur when cells are infected by MGE escape mutants, or when the levels of CRISPR-Cas activity are insufficient to provide complete immunity using only the existing spacers – even in the absence of MGE escape mutations<sup>45,59,61-64</sup>.

Priming begins with target recognition by crRNA-effector complexes. Therefore, factors that influence target recognition (i.e. the formation and stability of the crRNA-DNA hybrid – see **Fig. 2.2**), including PAM sensing and crRNA:target complementarity, affect the efficiency of primed adaptation<sup>59,60,65-70</sup>. Furthermore, these same factors can induce conformational rearrangements in the target-bound crRNA-effector complex that result in favoring of either the interference or priming pathways<sup>66-68,71</sup>. In type I-E systems, the Cas8e (Cse1) subunit of Cascade can adopt one of two conformational modes<sup>68,71</sup>, which may promote either direct or Cas1-Cas2-stimulated recruitment of the effector Cas3 nuclease<sup>66,67,71</sup>.

Cas3, found in all type I systems, exhibits 3' to 5' helicase and endonuclease activity that nicks, unwinds and degrades target DNA <sup>72-74</sup>. In vitro activity of the type I-E Cas3 produces ssDNA fragments of ~30-100 nucleotides that are enriched for PAMs in their 3' ends and that anneal to provide partially duplexed pre-spacer substrates <sup>65</sup>. The spatial positioning of Cas1-Cas2 during primed substrate generation has not been clearly established, although Cas1-Cas2-facilitated recruitment of Cas3 would imply the CRISPR adaptation machinery is localized close to the site of pre-spacer production <sup>66,71</sup>. In type I-F systems Cas3 is fused to the C-terminus of Cas2 and forms a Cas1-Cas2-3 complex <sup>31</sup> that couples the CRISPR adaptation machinery directly to the source of pre-spacer generation during primed adaptation <sup>23,75</sup>.

Despite different target recognition modes favoring distinct Cas3 recruitment routes, primed CRISPR adaptation can be provoked by MGE escape mutants and non-escape (interference-proficient) targets <sup>23,45,61,76</sup>. However, when the intracellular copy-number influences of the MGE are excluded, interference-proficient targets promote greater spacer acquisition than escape mutants <sup>23,76</sup>. This forms a positive feedback loop, reinforcing immunity against recurrent threats even in the absence of escapees <sup>23,45</sup>. If the copy-number of the MGE within the host cell is factored in, then escape mutants actually trigger more spacer acquisition. This is because interference rapidly clears targeted MGEs from the cell, whereas escape mutants that evade immediate clearance by existing CRISPR-Cas immunity persist for longer. Over time, the prolonged presence of the escape MGE, combined with the priming-centric CRISPR-Cas target recognition mode, results in higher net production of pre-spacer substrates and spacer integration <sup>23,64,65,76</sup>.

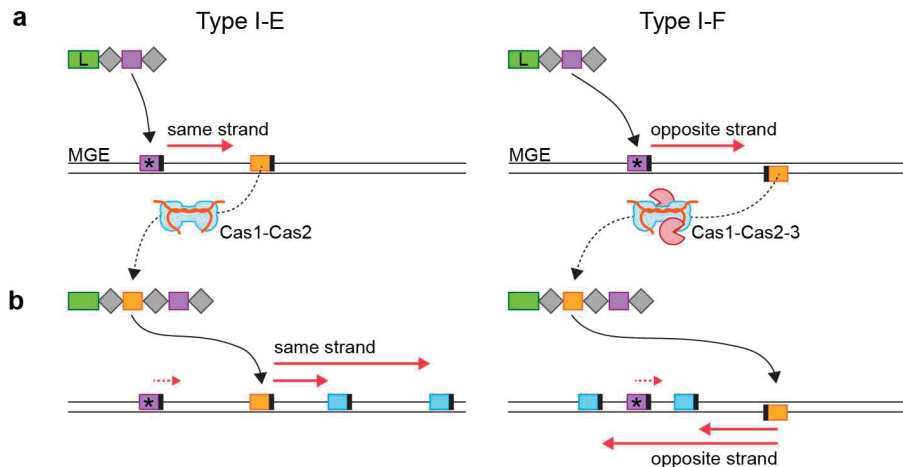
Because priming is initiated by site-specific target recognition (i.e. targeting a 'priming' protospacer), Cas1-Cas2 compatible pre-spacers are subsequently produced from MGEs with locational biases (**Fig. 2.5**). Mapping the MGE sequence positions, and strands targeted by newly acquired spacers (i.e. their corresponding protospacers), has revealed subtype-specific patterns and has provided much of our insight into the mechanisms of primed CRISPR adaptation <sup>23,44,45,61,75,77,78</sup>. In type I-E systems, new protospacers typically map to the same strand <sup>44,45</sup> as the priming protospacer (**Fig. 2.5**). For type I-B priming, Cas3 is predicted to load onto either strand at the priming protospacer, resulting in a bidirectional distribution of new protospacers <sup>77</sup>. For type I-F priming, the first new protospacer typically maps to the strand opposite the priming protospacer, in a direction consistent with Cas3 loading and 3' to 5' helicase activity on

the non-target strand. Furthermore, once the first spacer is acquired, two protospacer targets in the MGE will be recognized and pre-spacer production can be driven from both locations <sup>23,75</sup> (**Fig. 2.5**). However, priming is stimulated more strongly from the interference-proficient protospacer than from the original 'priming' protospacer. Thus, subsequent spacers (i.e. second and following) result from targeting by the first new spacer and are typically located back towards the original priming protospacer <sup>23</sup> (**Fig. 2.5**). The dominance of the first new spacer also holds true for type I-E <sup>45,76</sup> and likely all other systems that display priming. However, these are generalized models and many questions remain unresolved, such as the mechanisms resulting in strand selection and why some spacer sequences are more highly acquired from MGEs than others. Further analyses of priming in different systems, particularly the order of new spacers acquired, will greatly inform our understanding of primed Cas1-Cas2 substrate production.

### 2.3.3 Cas protein-assisted production of spacers

Given the apparent advantages conferred by priming in type I systems, analogous mechanisms to stimulate 'primed-like' CRISPR adaptation, are likely to exist in other CRISPR-Cas types. For example, DNA breaks induced by interference activity of class 2 CRISPR-Cas effector complexes could trigger host DNA repair mechanisms (e.g. RecBCD), thereby providing substrates for Cas1-Cas2. In agreement with a model for DNA break-stimulated enhancement of CRISPR adaptation, restriction enzyme activity can stimulate RecBCD-facilitated production of pre-spacer substrates <sup>53</sup>. RecBCD activity may also partially account for the enhanced CRISPR adaptation observed during phage infection of a host possessing an innate defense restriction-modification system <sup>79</sup>. Whether in this example, the enhanced CRISPR adaptation was RecBCD-dependent is unknown. In a CRISPR-Cas-induced DNA break model, the production of pre-spacer substrates is preceded by sequence-specific target recognition, hence this could be considered related to 'priming' <sup>80</sup>. Although direct evidence to support this concept is lacking, CRISPR adaptation in type II-A systems requires Cas1-Cas2, Cas9, as well as a trans-activating crRNA (tracrRNA; a cofactor for crRNA processing and interference in type II systems) and Csn2 <sup>56,80</sup>. The PAM-sensing domain of Cas9 enhances the acquisition of spacers with interference-proficient PAMs <sup>80</sup>. However, Cas9 nuclease activity is dispensable <sup>56</sup> and existing spacers are not strictly necessary <sup>80</sup>, suggesting that the PAM interactions of Cas9 could be sufficient to select appropriate new spacers. Some Cas9 variants can also function with non-CRISPR RNAs

and tracrRNA<sup>81</sup>. This raises the possibility that host or MGE-derived RNAs might direct promiscuous Cas9 activity, resulting in DNA breaks, or replication fork stalling that could potentially result in pre-spacer generation.



**Figure 2.5: Primed CRISPR adaptation from a multi-copy MGE by type I-E and I-F CRISPR-Cas systems.** **a**, An existing spacer (purple) with homology to an MGE sequence that has escaped interference (the ‘priming’ protospacer denoted with an asterisk) directs target recognition. The PAM adjacent to the protospacer is shown in black. The crRNA-effector complex recruits Cas3 (or Cas1-Cas2-3 for type I-F) and the 3’ to 5’ helicase activity (illustrated by the red arrow) results in the acquisition of a new spacer from a site distal to the initial priming location. The new spacer maps to an interference-proficient protospacer (orange). Spacer acquisition in the type I-E system requires the Cas1-Cas2 complex and spacer acquisition in the type I-F system uses a Cas1-Cas2-3 complex. **b**, The new spacer (orange) perfectly matches the MGE sequence at the orange protospacer location and facilitates targeting of the MGE and recruitment of Cas3. Hence, subsequent spacers (mapping to blue protospacers) typically originate from Cas3 activity (red arrows) beginning at this location.

## 2.4 Roles of accessory Cas proteins in CRISPR adaptation

Although Cas1 and Cas2 play a central role in adaptation, type-specific variations in *cas* gene clusters occur. In many systems, Cas1-Cas2 is assisted by accessory Cas proteins, which are often mutually exclusive and type-specific<sup>27</sup>. For example, in the *S. thermophilus* type II-A system, deletion of *csn2* impaired the acquisition of spacers from invading phages<sup>4</sup>.



Direct interaction between Cas1 and Csn2 also suggests a role for Csn2 in conjunction with the spacer acquisition machinery<sup>82</sup>. Csn2 multimers cooperatively bind to the free ends of linear dsDNA and can translocate by rotation-coupled movement<sup>83,84</sup>. Given that substrate-loaded type II-A Cas1-Cas2 is capable of full-site spacer integration in vitro<sup>40</sup>, Csn2 may be required for pre-spacer substrate production, selection or processing. Potentially, Csn2 binding to the free ends of dsDNA provides a cue for nucleases to assist in pre-spacer generation<sup>83</sup>.

Cas4, another ring-forming accessory protein, is found in type I, II-B and V systems<sup>27</sup>. Confirming its role in CRISPR adaptation, Cas4 is necessary for type I-B priming in *H. hispanica*<sup>77</sup> and interacts with a Cas1-2 fusion protein in the *Thermoproteus tenax* type I-A system<sup>85</sup>. Fusions between Cas4 and Cas1 are found in several systems, which indicates a functional association with the spacer acquisition machinery. Cas4 contains a RecB-like domain and four conserved cysteine residues, which are presumably involved in the coordination of an iron-sulfur cluster<sup>86</sup>. However, Cas4 proteins appear to be functionally diverse with some possessing uni- or bi-directional exonuclease activity, while others exhibit ssDNA endonuclease activity and unwinding activity on dsDNA<sup>86,87</sup>. Due to its nuclease activity, Cas4 is hypothesized to be involved in pre-spacer generation.

In type III systems, spacers complementary to RNA transcribed from MGEs are required to provide immunity (**Fig. 2.2**)<sup>88,89</sup>. Some bacterial type III systems contain fusions of Cas1 with reverse transcriptase domains (RTs) that provide a mechanism to integrate spacers from RNA substrates<sup>90</sup>. The RT-Cas1 fusion from *M. mediterranea* can integrate RNA precursors into an array, which are subsequently reverse transcribed to generate DNA spacers<sup>90</sup>. However, integration of DNA-derived spacers also occurs, indicating that the RNA derived-spacer route is not exclusive<sup>90</sup>. Hence, the combined integrase and reverse transcriptase activity of RT-Cas1-Cas2 enhances CRISPR adaptation against highly transcribed DNA MGEs and potentially from RNA-based invaders.

Other host proteins may also be necessary for pre-spacer substrate production. For example, RecG is required for efficient primed adaptation in both type I-E and I-F systems, but its precise role remains speculative<sup>39,91</sup>. Additionally, it remains enigmatic why some CRISPR-Cas systems require accessory proteins, while closely related types do not. For example, type II-C systems lack *cas4* or *csn2* that assist CRISPR adaptation in type II-A and II-B systems, respectively. These type-specific differences exemplify the diversity that has arisen during the evolution of CRISPR-Cas systems.



## 2.5 The genesis of adaptive immunity in prokaryotes

Expanding knowledge of the molecular mechanisms underlying CRISPR adaptation has led to a promising theory for the evolutionary origin of CRISPR-Cas systems<sup>92</sup>. Casposons are transposon-like elements typified by the presence of Cas1 homologs, casposases, which catalyze site-specific DNA integration and result in the duplication of repeat sites analogous to spacer acquisition<sup>93,94</sup>. It is possible that ancestral innate defenses gained DNA integration functionality from casposases, thus seeding the genesis of prokaryotic adaptive immunity<sup>95</sup>. The innate ancestor remains unidentified, but is likely to be a nuclease-based system. Co-occurrence of casposon-derived terminal inverted repeats and casposases in the absence of full casposons might represent an intermediate of the CRISPR signature repeat-spacer-repeat structures<sup>96</sup>. However, the evolutionary journey from the innate immunity-casposase hybrid to full adaptive immunity remains unclear. Evolution of diverse CRISPR-Cas types would have required stringent co-evolution of the Cas1-Cas2 spacer acquisition machinery, PAM and leader-repeat sequences, crRNA processing mechanisms and effector complexes.

In some systems, mechanisms to enhance the production of Cas1-Cas2 compatible pre-spacers from MGEs, such as priming, might have arisen because naïve CRISPR adaptation is an inefficient process with a high probability of acquiring spacers from host DNA. However, it was recently revealed that promiscuous binding of crRNA-effector complexes to the host genome results in a basal level of lethal ‘self-priming’ in a type I-F system<sup>23</sup>. Host *cas* gene regulation mechanisms might have arisen to balance the likelihood of self-acquisition events against the requirement to adapt to new threats – for example, when the risk of phage infection or horizontal gene transfer is high<sup>97,98</sup>. Alternatively, it has been proposed that selective acquisition of self-targeting spacers could provide benefits such as invoking altruistic cell death<sup>99</sup>, rapid genome evolution<sup>51</sup>, regulation of host processes<sup>100,101</sup>, or even preventing the uptake of other CRISPR-Cas systems<sup>102</sup>.

## 2.6 Outlook

The past four years has seen rapid progress in understanding mechanisms of CRISPR adaptation. Despite this progress, many facets of CRISPR

adaptation need more work. Synergy between innate defense systems and adaptation is relatively unexplored, but two aspects may be interrelated. First, DNA breaks<sup>53</sup> could stimulate generation of substrates for spacer acquisition (**Fig. 2.4**) and second, the stalling of infection could 'buy time' for CRISPR adaptation<sup>79,103,104</sup>. Analogously, it remains to be determined whether interference by CRISPR-Cas systems other than type I can also stimulate primed CRISPR adaptation. If not, the benefits of priming might provide an explanation for why type I systems are the most prevalent and diverse CRISPR-Cas type.

It is also unclear why many CRISPR-Cas systems have more than one CRISPR array that is used by a single set of Cas proteins. Given that Cas1-Cas2 is directed to leader-repeat junctions during integration, multiple arrays might provide additional integration sites, increasing CRISPR adaptation efficiency. In addition, parallel CRISPR arrays should increase crRNA production from spacers that were acquired recently (i.e. due to the polarized insertion of new spacers next to the promoter-containing CRISPR leaders)<sup>15</sup>. Whereas some strains have multiple CRISPR arrays belonging to the same type, other hosts have several different types of CRISPR-Cas systems simultaneously. The benefits of harboring multiple CRISPR-Cas systems are not entirely clear, but can result in CRISPRs being shared by different systems to extend targeting to both RNA and DNA<sup>105</sup>. From a CRISPR adaptation perspective, multiple systems might also enable a wider PAM repertoire to be sampled during spacer selection. Additional systems in a single host could also be a result of phage- and MGE-encoded anti-CRISPR proteins, which can inhibit both interference and primed CRISPR adaptation<sup>106-108</sup>. Alternatively, additional systems may allow some systems to function in defense, while others perform non-canonical roles<sup>101</sup>.

While Cas effector nucleases, such as Cas9, have been harnessed for many biotechnological applications, the use of repurposed CRISPR-Cas adaptation machinery has yet to be widely exploited. The sequence-specific integrase activity of Cas1-Cas2 holds promise in synthetic biology, such as for the insertion of specific sequences (or barcodes) to mark and track cells in a population. In *E. coli* the feasibility of such an approach is evident<sup>43</sup>, but transition to eukaryotic systems will provide the greatest utility where lineage tracking and cell fate could be followed. A similar approach has been demonstrated by exploiting Cas9 nuclease activity<sup>109</sup>. The elements required for leader-specific integration must be carefully considered for the introduction of CRISPR-Cas spacer acquisition machinery into eukaryotic cells, as unintended ectopic integrations could be problematic given the larger eukaryotic sequence space. Ultimately, our understanding of CRISPR

adaptation in prokaryotes may lead to applications where entire CRISPR systems are transplanted into eukaryotic cells to prevent viral invaders. As we begin to comprehend CRISPR adaptation in more detail the opportunities to repurpose other parts of these remarkable prokaryotic immune systems is increasingly becoming reality.

## 2

## 2.7 Acknowledgments

Work in the Fineran laboratory on CRISPR-Cas is supported by a Rutherford Discovery Fellowship (PCF) from the Royal Society of New Zealand (RSNZ), the Marsden Fund (RSNZ), the University of Otago and the Bio-protection Research Centre (Tertiary Education Commission). The Brouns laboratory is funded by an ERC starting grant (639707), an NWO VIDI grant (864.11.005), and FOM projectruimte (15PR3188). We thank members of the Fineran and Brouns groups for useful feedback on the manuscript.

## References

- 1 Dy, R. L., Richter, C., Salmond, G. P. & Fineran, P. C. Remarkable Mechanisms in Microbes to Resist Phage Infections. *Annu Rev Virol* **1**, 307-331, doi:10.1146/annurev-virology-031413-085500 (2014).
- 2 Samson, J. E., Magadán, A. H., Sabri, M. & Moineau, S. Revenge of the phages: defeating bacterial defences. *Nat Rev Microbiol* **11**, 675-687, doi:10.1038/nrmicro3096 (2013).
- 3 Marraffini, L. A. CRISPR-Cas immunity in prokaryotes. *Nature* **526**, 55-61, doi:10.1038/nature15386 (2015).
- 4 Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709-1712, doi:10.1126/science.1138140 (2007).
- 5 Brouns, S. J. *et al.* Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960-964, doi:10.1126/science.1159689 (2008).
- 6 Wright, A. V., Nuñez, J. K. & Doudna, J. A. Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. *Cell* **164**, 29-44, doi:10.1016/j.cell.2015.12.035 (2016).
- 7 Pourcel, C., Salvignol, G. & Vergnaud, G. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* **151**, 653-663, doi:10.1099/mic.0.27437-0 (2005).
- 8 Mojica, F. J., Díez-Villaseñor, C., García-Martínez, J. & Soria, E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign

- genetic elements. *J Mol Evol* **60**, 174-182, doi:10.1007/s00239-004-0046-3 (2005).
- 9 Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S. D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**, 2551-2561, doi:10.1099/mic.0.28048-0 (2005).
  - 10 Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I. & Koonin, E. V. 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* **1**, 7, doi:10.1186/1745-6150-1-7 (2006).
  - 11 Marraffini, L. A. & Sontheimer, E. J. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**, 1843-1845, doi:10.1126/science.1165771 (2008).
  - 12 Yosef, I., Goren, M. G. & Qimron, U. Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Res* **40**, 5569-5576, doi:10.1093/nar/gks216 (2012).
  - 13 Díez-Villaseñor, C., Guzmán, N. M., Almendros, C., García-Martínez, J. & Mojica, F. J. CRISPR-spacer integration reporter plasmids reveal distinct genuine acquisition specificities among CRISPR-Cas I-E variants of *Escherichia coli*. *RNA Biol* **10**, 792-802, doi:10.4161/rna.24023 (2013).
  - 14 Erdmann, S. & Garrett, R. A. Selective and hyperactive uptake of foreign DNA by adaptive immune systems of an archaeon via two distinct mechanisms. *Mol Microbiol* **85**, 1044-1056, doi:10.1111/j.1365-2958.2012.08171.x (2012).
  - 15 McGinn, J. & Marraffini, L. A. CRISPR-Cas Systems Optimize Their Immune Response by Specifying the Site of Spacer Integration. *Mol Cell* **64**, 616-623, doi:10.1016/j.molcel.2016.08.038 (2016).
  - 16 Lopez-Sanchez, M. J. *et al.* The highly dynamic CRISPR1 system of *Streptococcus agalactiae* controls the diversity of its mobilome. *Mol Microbiol* **85**, 1057-1071, doi:10.1111/j.1365-2958.2012.08172.x (2012).
  - 17 Andersson, A. F. & Banfield, J. F. Virus population dynamics and acquired virus resistance in natural microbial communities. *Science* **320**, 1047-1050, doi:10.1126/science.1157358 (2008).
  - 18 Horvath, P. *et al.* Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *J Bacteriol* **190**, 1401-1412, doi:10.1128/JB.01415-07 (2008).
  - 19 Amitai, G. & Sorek, R. CRISPR-Cas adaptation: insights into the mechanism of action. *Nat Rev Microbiol* **14**, 67-76, doi:10.1038/nrmicro.2015.14 (2016).
  - 20 Sternberg, S. H., Richter, H., Charpentier, E. & Qimron, U. Adaptation in CRISPR-Cas Systems. *Mol Cell* **61**, 797-808, doi:10.1016/j.molcel.2016.01.030 (2016).
  - 21 Sun, C. L., Thomas, B. C., Barrangou, R. & Banfield, J. F. Metagenomic reconstructions of bacterial CRISPR loci constrain population histories. *ISME J* **10**, 858-870, doi:10.1038/ismej.2015.162 (2016).

- 22 Paez-Espino, D. *et al.* Uncovering Earth's virome. *Nature* **536**, 425-430, doi:10.1038/nature19094 (2016).
- 23 Staals, R. H. *et al.* Interference-driven spacer acquisition is dominant over naive and primed adaptation in a native CRISPR-Cas system. *Nat Commun* **7**, 12853, doi:10.1038/ncomms12853 (2016).
- 24 Paez-Espino, D. *et al.* Strong bias in the bacterial CRISPR elements that confer immunity to phage. *Nat Commun* **4**, 1430, doi:10.1038/ncomms2440 (2013).
- 25 van Houte, S. *et al.* The diversity-generating benefits of a prokaryotic adaptive immune system. *Nature* **532**, 385-388, doi:10.1038/nature17436 (2016).
- 26 Liu, F. *et al.* Novel virulence gene and clustered regularly interspaced short palindromic repeat (CRISPR) multilocus sequence typing scheme for subtyping of the major serovars of *Salmonella enterica* subsp. *enterica*. *Appl Environ Microbiol* **77**, 1946-1956, doi:10.1128/AEM.02625-10 (2011).
- 27 Makarova, K. S. *et al.* An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol* **13**, 722-736, doi:10.1038/nrmicro3569 (2015).
- 28 Mohanraju, P. *et al.* Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science* **353**, aad5147, doi:10.1126/science.aad5147 (2016).
- 29 Wright, A. V., Nunez, J. K. & Doudna, J. A. Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. *Cell* **164**, 29-44, doi:10.1016/j.cell.2015.12.035 (2016).
- 30 Nuñez, J. K. *et al.* Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nat Struct Mol Biol* **21**, 528-534, doi:10.1038/nsmb.2820 (2014).
- 31 Richter, C., Gristwood, T., Clulow, J. S. & Fineran, P. C. In vivo protein interactions and complex formation in the *Pectobacterium atrosepticum* subtype I-F CRISPR/Cas System. *PLoS One* **7**, e49549, doi:10.1371/journal.pone.0049549 (2012).
- 32 Leenay, R. T. & Beisel, C. L. Deciphering, Communicating, and Engineering the CRISPR PAM. *J Mol Biol* **429**, 177-191, doi:10.1016/j.jmb.2016.11.024 (2017).
- 33 Mojica, F. J., Díez-Villaseñor, C., García-Martínez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733-740, doi:10.1099/mic.0.023960-0 (2009).
- 34 Wang, J. *et al.* Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems. *Cell* **163**, 840-853, doi:10.1016/j.cell.2015.10.008 (2015).
- 35 Nuñez, J. K., Harrington, L. B., Kranzusch, P. J., Engelman, A. N. & Doudna, J. A. Foreign DNA capture during CRISPR-Cas adaptive immunity. *Nature* **527**, 535-538, doi:10.1038/nature15760 (2015).
- 36 Nuñez, J. K., Lee, A. S., Engelman, A. & Doudna, J. A. Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature* **519**, 193-198, doi:10.1038/nature14237 (2015).

- 37 Arslan, Z., Hermanns, V., Wurm, R., Wagner, R. & Pul, U. Detection and characterization of spacer integration intermediates in type I-E CRISPR-Cas system. *Nucleic Acids Res* **42**, 7884-7893, doi:10.1093/nar/gku510 (2014).
- 38 Wei, Y., Chesne, M. T., Terns, R. M. & Terns, M. P. Sequences spanning the leader-repeat junction mediate CRISPR adaptation to phage in *Streptococcus thermophilus*. *Nucleic Acids Res* **43**, 1749-1758, doi:10.1093/nar/gku1407 (2015).
- 39 Ivančić-Baće, I., Cass, S. D., Wearne, S. J. & Bolt, E. L. Different genome stability proteins underpin primed and naive adaptation in *E. coli* CRISPR-Cas immunity. *Nucleic Acids Res* **43**, 10821-10830, doi:10.1093/nar/gkv1213 (2015).
- 40 Wright, A. V. & Doudna, J. A. Protecting genome integrity during CRISPR immune adaptation. *Nat Struct Mol Biol* **23**, 876-883, doi:10.1038/nsmb.3289 (2016).
- 41 Nuñez, J. K., Bai, L., Harrington, L. B., Hinder, T. L. & Doudna, J. A. CRISPR Immunological Memory Requires a Host Factor for Specificity. *Mol Cell* **62**, 824-833, doi:10.1016/j.molcel.2016.04.027 (2016).
- 42 Yoganand, K. N., Sivathanu, R., Nimkar, S. & Anand, B. Asymmetric positioning of Cas1-2 complex and Integration Host Factor induced DNA bending guide the unidirectional homing of protospacer in CRISPR-Cas type I-E system. *Nucleic Acids Res* **45**, 367-381, doi:10.1093/nar/gkw1151 (2017).
- 43 Shipman, S. L., Nivala, J., Macklis, J. D. & Church, G. M. Molecular recordings by directed CRISPR spacer acquisition. *Science* **353**, aaf1175, doi:10.1126/science.aaf1175 (2016).
- 44 Shmakov, S. *et al.* Pervasive generation of oppositely oriented spacers during CRISPR adaptation. *Nucleic Acids Res* **42**, 5907-5916, doi:10.1093/nar/gku226 (2014).
- 45 Swarts, D. C., Mosterd, C., van Passel, M. W. & Brouns, S. J. CRISPR interference directs strand specific spacer acquisition. *PLoS One* **7**, e35888, doi:10.1371/journal.pone.0035888 (2012).
- 46 Rollie, C., Schneider, S., Brinkmann, A. S., Bolt, E. L. & White, M. F. Intrinsic sequence specificity of the Cas1 integrase directs new spacer acquisition. *Elife* **4**, eLife.08716, doi:10.7554/eLife.08716 (2015).
- 47 Kunin, V., Sorek, R. & Hugenholtz, P. Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol* **8**, R61, doi:10.1186/gb-2007-8-4-r61 (2007).
- 48 Goren, M. G. *et al.* Repeat Size Determination by Two Molecular Rulers in the Type I-E CRISPR Array. *Cell Rep* **16**, 2811-2818, doi:10.1016/j.celrep.2016.08.043 (2016).
- 49 Wang, R., Li, M., Gong, L., Hu, S. & Xiang, H. DNA motifs determining the accuracy of repeat duplication during CRISPR adaptation in *Haloarcula hispanica*. *Nucleic Acids Res* **44**, 4266-4277, doi:10.1093/nar/gkw260 (2016).

- 50 Stern, A., Keren, L., Wurtzel, O., Amitai, G. & Sorek, R. Self-targeting by CRISPR: gene regulation or autoimmunity? *Trends Genet* **26**, 335-340, doi:10.1016/j.tig.2010.05.008 (2010).
- 51 Vercoe, R. B. *et al.* Cytotoxic chromosomal targeting by CRISPR/Cas systems can reshape bacterial genomes and expel or remodel pathogenicity islands. *PLoS Genet* **9**, e1003454, doi:10.1371/journal.pgen.1003454 (2013).
- 52 Fineran, P. C. & Charpentier, E. Memory of viral infections by CRISPR-Cas adaptive immune systems: acquisition of new information. *Virology* **434**, 202-209, doi:10.1016/j.virol.2012.10.003 (2012).
- 53 Levy, A. *et al.* CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature* **520**, 505-510, doi:10.1038/nature14302 (2015).
- 54 Wigley, D. B. Bacterial DNA repair: recent insights into the mechanism of RecBCD, AddAB and AdnAB. *Nat Rev Microbiol* **11**, 9-13, doi:10.1038/nrmicro2917 (2013).
- 55 Westra, E. R. *et al.* CRISPR-Cas systems preferentially target the leading regions of MOBF conjugative plasmids. *RNA Biol* **10**, 749-761, doi:10.4161/rna.24202 (2013).
- 56 Wei, Y., Terns, R. M. & Terns, M. P. Cas9 function and host genome sampling in Type II-A CRISPR-Cas adaptation. *Genes Dev* **29**, 356-361, doi:10.1101/gad.257550.114 (2015).
- 57 Semenova, E. *et al.* Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc Natl Acad Sci U S A* **108**, 10098-10103, doi:10.1073/pnas.1104144108 (2011).
- 58 Deveau, H. *et al.* Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J Bacteriol* **190**, 1390-1400, doi:10.1128/JB.01412-07 (2008).
- 59 Fineran, P. C. *et al.* Degenerate target sites mediate rapid primed CRISPR adaptation. *Proc Natl Acad Sci U S A* **111**, E1629-1638, doi:10.1073/pnas.1400071111 (2014).
- 60 Xue, C. *et al.* CRISPR interference and priming varies with individual spacer sequences. *Nucleic Acids Res* **43**, 10831-10847, doi:10.1093/nar/gkv1259 (2015).
- 61 Datsenko, K. A. *et al.* Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat Commun* **3**, 945, doi:10.1038/ncomms1937 (2012).
- 62 Savitskaya, E., Semenova, E., Dedkov, V., Metlitskaya, A. & Severinov, K. High-throughput analysis of type I-E CRISPR/Cas spacer acquisition in *E. coli*. *RNA Biol* **10**, 716-725, doi:10.4161/rna.24325 (2013).
- 63 Patterson, A. G., Chang, J. T., Taylor, C. & Fineran, P. C. Regulation of the Type I-F CRISPR-Cas system by CRP-cAMP and GalM controls spacer acquisition and interference. *Nucleic Acids Res* **43**, 6038-6048, doi:10.1093/nar/gkv517 (2015).
- 64 Severinov, K., Ispolatov, I. & Semenova, E. The Influence of Copy-Number of Targeted Extrachromosomal Genetic Elements on the Outcome of



- CRISPR-Cas Defense. *Front Mol Biosci* **3**, 45, doi:10.3389/fmolb.2016.00045 (2016).
- 65 Künne, T. *et al.* Cas3-Derived Target DNA Degradation Fragments Fuel Primed CRISPR Adaptation. *Mol Cell* **63**, 852-864, doi:10.1016/j.molcel.2016.07.011 (2016).
- 66 Redding, S. *et al.* Surveillance and Processing of Foreign DNA by the *Escherichia coli* CRISPR-Cas System. *Cell* **163**, 854-865, doi:10.1016/j.cell.2015.10.003 (2015).
- 67 Blosser, T. R. *et al.* Two distinct DNA binding modes guide dual roles of a CRISPR-Cas protein complex. *Mol Cell* **58**, 60-70, doi:10.1016/j.molcel.2015.01.028 (2015).
- 68 Hayes, R. P. *et al.* Structural basis for promiscuous PAM recognition in type I-E Cascade from *E. coli*. *Nature* **530**, 499-503, doi:10.1038/nature16995 (2016).
- 69 van Erp, P. B. *et al.* Mechanism of CRISPR-RNA guided recognition of DNA targets in *Escherichia coli*. *Nucleic Acids Res* **43**, 8381-8391, doi:10.1093/nar/gkv793 (2015).
- 70 Li, M., Wang, R. & Xiang, H. *Haloarcula hispanica* CRISPR authenticates PAM of a target sequence to prime discriminative adaptation. *Nucleic Acids Res* **42**, 7226-7235, doi:10.1093/nar/gku389 (2014).
- 71 Xue, C., Whitis, N. R. & Sashital, D. G. Conformational Control of Cascade Interference and Priming Activities in CRISPR Immunity. *Mol Cell* **64**, 826-834, doi:10.1016/j.molcel.2016.09.033 (2016).
- 72 Mulepati, S. & Bailey, S. In vitro reconstitution of an *Escherichia coli* RNA-guided immune system reveals unidirectional, ATP-dependent degradation of DNA target. *J Biol Chem* **288**, 22184-22192, doi:10.1074/jbc.M113.472233 (2013).
- 73 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* **46**, 595-605, doi:10.1016/j.molcel.2012.03.018 (2012).
- 74 Huo, Y. *et al.* Structures of CRISPR Cas3 offer mechanistic insights into Cascade-activated DNA unwinding and degradation. *Nat Struct Mol Biol* **21**, 771-777, doi:10.1038/nsmb.2875 (2014).
- 75 Richter, C. *et al.* Priming in the Type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. *Nucleic Acids Res* **42**, 8516-8526, doi:10.1093/nar/gku527 (2014).
- 76 Semenova, E. *et al.* Highly efficient primed spacer acquisition from targets destroyed by the *Escherichia coli* type I-E CRISPR-Cas interfering complex. *Proc Natl Acad Sci U S A* **113**, 7626-7631, doi:10.1073/pnas.1602639113 (2016).
- 77 Li, M., Wang, R., Zhao, D. & Xiang, H. Adaptation of the *Haloarcula hispanica* CRISPR-Cas system to a purified virus strictly requires a priming process. *Nucleic Acids Res* **42**, 2483-2492, doi:10.1093/nar/gkt1154 (2014).



- 78 Rao, C. *et al.* Active and adaptive *Legionella* CRISPR-Cas reveals a recurrent challenge to the pathogen. *Cell Microbiol* **18**, 1319-1338, doi:10.1111/cmi.12586 (2016).
- 79 Hynes, A. P., Villion, M. & Moineau, S. Adaptation in bacterial CRISPR-Cas immunity can be driven by defective phages. *Nat Commun* **5**, 4399, doi:10.1038/ncomms5399 (2014).
- 80 Heler, R. *et al.* Cas9 specifies functional viral targets during CRISPR-Cas adaptation. *Nature* **519**, 199-202, doi:10.1038/nature14245 (2015).
- 81 Sampson, T. R., Saroj, S. D., Llewellyn, A. C., Tzeng, Y. L. & Weiss, D. S. A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. *Nature* **497**, 254-257, doi:10.1038/nature12048 (2013).
- 82 Ka, D. *et al.* Crystal Structure of *Streptococcus pyogenes* Cas1 and Its Interaction with Csn2 in the Type II CRISPR-Cas System. *Structure* **24**, 70-79, doi:10.1016/j.str.2015.10.019 (2016).
- 83 Arslan, Z. *et al.* Double-strand DNA end-binding and sliding of the toroidal CRISPR-associated protein Csn2. *Nucleic Acids Res* **41**, 6347-6359, doi:10.1093/nar/gkt315 (2013).
- 84 Lee, K. H. *et al.* Identification, structural, and biochemical characterization of a group of large Csn2 proteins involved in CRISPR-mediated bacterial immunity. *Proteins* **80**, 2573-2582, doi:10.1002/prot.24138 (2012).
- 85 Plagens, A., Tjaden, B., Hagemann, A., Randau, L. & Hensel, R. Characterization of the CRISPR/Cas subtype I-A system of the hyperthermophilic crenarchaeon *Thermoproteus tenax*. *J Bacteriol* **194**, 2491-2500, doi:10.1128/JB.00206-12 (2012).
- 86 Zhang, J., Kasciukovic, T. & White, M. F. The CRISPR associated protein Cas4 Is a 5' to 3' DNA exonuclease with an iron-sulfur cluster. *PLoS One* **7**, e47232, doi:10.1371/journal.pone.0047232 (2012).
- 87 Lemak, S. *et al.* Toroidal structure and DNA cleavage by the CRISPR-associated [4Fe-4S] cluster containing Cas4 nuclease SSO0001 from *Sulfolobus solfataricus*. *J Am Chem Soc* **135**, 17476-17487, doi:10.1021/ja408729b (2013).
- 88 Hale, C. R. *et al.* RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* **139**, 945-956, doi:10.1016/j.cell.2009.07.040 (2009).
- 89 Goldberg, G. W., Jiang, W., Bikard, D. & Marraffini, L. A. Conditional tolerance of temperate phages via transcription-dependent CRISPR-Cas targeting. *Nature* **514**, 633-637, doi:10.1038/nature13637 (2014).
- 90 Silas, S. *et al.* Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase-Cas1 fusion protein. *Science* **351**, aad4234, doi:10.1126/science.aad4234 (2016).
- 91 Heussler, G. E., Miller, J. L., Price, C. E., Collins, A. J. & O'Toole, G. A. Requirements for *Pseudomonas aeruginosa* Type I-F CRISPR-Cas Adaptation Determined Using a Biofilm Enrichment Assay. *J Bacteriol* **198**, 3080-3090, doi:10.1128/JB.00458-16 (2016).
- 92 Krupovic, M., Makarova, K. S., Forterre, P., Prangishvili, D. & Koonin, E. V. Casposons: a new superfamily of self-synthesizing DNA transposons at the

- origin of prokaryotic CRISPR-Cas immunity. *BMC Biol* **12**, 36, doi:10.1186/1741-7007-12-36 (2014).
- 93 Hickman, A. B. & Dyda, F. The casposon-encoded Cas1 protein from *Aciduliprofundum boonei* is a DNA integrase that generates target site duplications. *Nucleic Acids Res* **43**, 10576-10587, doi:10.1093/nar/gkv1180 (2015).
- 94 Beguin, P., Charpin, N., Koonin, E. V., Forterre, P. & Krupovic, M. Casposon integration shows strong target site preference and recapitulates protospacer integration by CRISPR-Cas systems. *Nucleic Acids Res* **44**, 10367-10376, doi:10.1093/nar/gkw821 (2016).
- 95 Koonin, E. V. & Krupovic, M. Evolution of adaptive immunity from transposable elements combined with innate immune systems. *Nat Rev Genet* **16**, 184-192, doi:10.1038/nrg3859 (2015).
- 96 Krupovic, M., Shmakov, S., Makarova, K. S., Forterre, P. & Koonin, E. V. Recent Mobility of Casposons, Self-Synthesizing Transposons at the Origin of the CRISPR-Cas Immunity. *Genome Biol Evol* **8**, 375-386, doi:10.1093/gbe/evw006 (2016).
- 97 Patterson, A. G. *et al.* Quorum Sensing Controls Adaptive Immunity through the Regulation of Multiple CRISPR-Cas Systems. *Mol Cell* **64**, 1102-1108, doi:10.1016/j.molcel.2016.11.012 (2016).
- 98 Hoyland-Kroghsbo, N. M. *et al.* Quorum sensing controls the *Pseudomonas aeruginosa* CRISPR-Cas adaptive immune system. *Proc Natl Acad Sci U S A* **114**, 131-135, doi:10.1073/pnas.1617415113 (2017).
- 99 Koonin, E. V. & Zhang, F. Coupling immunity and programmed cell suicide in prokaryotes: Life-or-death choices. *Bioessays* **39**, 1-9, doi:10.1002/bies.201600186 (2017).
- 100 Li, R. *et al.* Type I CRISPR-Cas targets endogenous genes and regulates virulence to evade mammalian host immunity. *Cell Res* **26**, 1273-1287, doi:10.1038/cr.2016.135 (2016).
- 101 Westra, E. R., Buckling, A. & Fineran, P. C. CRISPR-Cas systems: beyond adaptive immunity. *Nat Rev Microbiol* **12**, 317-326, doi:10.1038/nrmicro3241 (2014).
- 102 Almendros, C., Guzman, N. M., Garcia-Martinez, J. & Mojica, F. J. Anti-cas spacers in orphan CRISPR4 arrays prevent uptake of active CRISPR-Cas I-F systems. *Nat Microbiol* **1**, 16081, doi:10.1038/nmicrobiol.2016.81 (2016).
- 103 Makarova, K. S., Anantharaman, V., Aravind, L. & Koonin, E. V. Live virus-free or die: coupling of antiviral immunity and programmed suicide or dormancy in prokaryotes. *Biol Direct* **7**, 40, doi:10.1186/1745-6150-7-40 (2012).
- 104 Dupuis, M. E., Villion, M., Magadán, A. H. & Moineau, S. CRISPR-Cas and restriction-modification systems are compatible and increase phage resistance. *Nat Commun* **4**, 2087, doi:10.1038/ncomms3087 (2013).
- 105 Elmore, J., Deighan, T., Westpheling, J., Terns, R. M. & Terns, M. P. DNA targeting by the type I-G and type I-A CRISPR-Cas systems of *Pyrococcus*

*furiosus*. *Nucleic Acids Res* **43**, 10353-10363, doi:10.1093/nar/gkv1140 (2015).

106 Bondy-Denomy, J., Pawluk, A., Maxwell, K. L. & Davidson, A. R. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature* **493**, 429-432, doi:10.1038/nature11723 (2013).

107 Pawluk, A. *et al.* Inactivation of CRISPR-Cas systems by anti-CRISPR proteins in diverse bacterial species. *Nat Microbiol* **1**, 16085, doi:10.1038/nmicrobiol.2016.85 (2016).

108 Vorontsova, D. *et al.* Foreign DNA acquisition by the I-F CRISPR-Cas system requires all components of the interference machinery. *Nucleic Acids Res* **43**, 10848-10860, doi:10.1093/nar/gkv1261 (2015).

109 Perli, S. D., Cui, C. H. & Lu, T. K. Continuous genetic recording with self-targeting CRISPR-Cas in human cells. *Science* **353**, aag0511, doi:10.1126/science.aag0511 (2016).



# 3

## Using CAPTURE to detect adaption in native arrays

Published as: R. E. McKenzie, C. Almendros, J. N. A. Vink, S. J. J. Brouns, Using CAPTURE to detect spacer acquisition in native CRISPR arrays. Nat. Protoc. 14 (2019), doi:10.1038/s41596-018-0123-5.

## Abstract

CRISPR-Cas systems are able to acquire immunological memories (spacers) from bacteriophages and plasmids in order to survive infection, however this often occurs at low frequency within a population making it difficult to detect. Here we have developed CAPTURE (CCRISPR Adaptation PCR Technique Using Re-amplification and Electrophoresis) a versatile and adaptable protocol to detect spacer acquisition events by electrophoresis imaging, with a sensitivity that can identify spacer acquisition in 1 in  $10^5$  cells. Our method harnesses two simple PCR steps, separated by automated electrophoresis and extraction of size-selected DNA amplicons, allowing the removal of unexpanded arrays from the sample pool, and enabling a 1000 times more sensitive detection of new spacers than existing PCR protocols. CAPTURE is a straightforward method requiring only one day to enable detection of spacer acquisition in all native CRISPR systems and facilitate studies aimed both at unravelling the mechanism of spacer integration and more sensitive tracing of integration events in natural ecosystems.

### 3.1 Introduction

Microbes have the unique capacity to acquire resistance against bacteriophage and plasmids by the incorporation of small DNA fragments (spacers), derived from these invaders, into the genome of the infected cell<sup>1,2</sup>. Cells that successfully incorporate these spacers in their CRISPR array become immunized against further infection by this specific invader<sup>1,2</sup>. The inserted spacers are expressed and processed into CRISPR RNAs (crRNAs) and act as guides for Cas protein effector complexes which help find, bind and cleave matching invader nucleic acid sequences, resulting in invader elimination and host survival<sup>2,3</sup>.

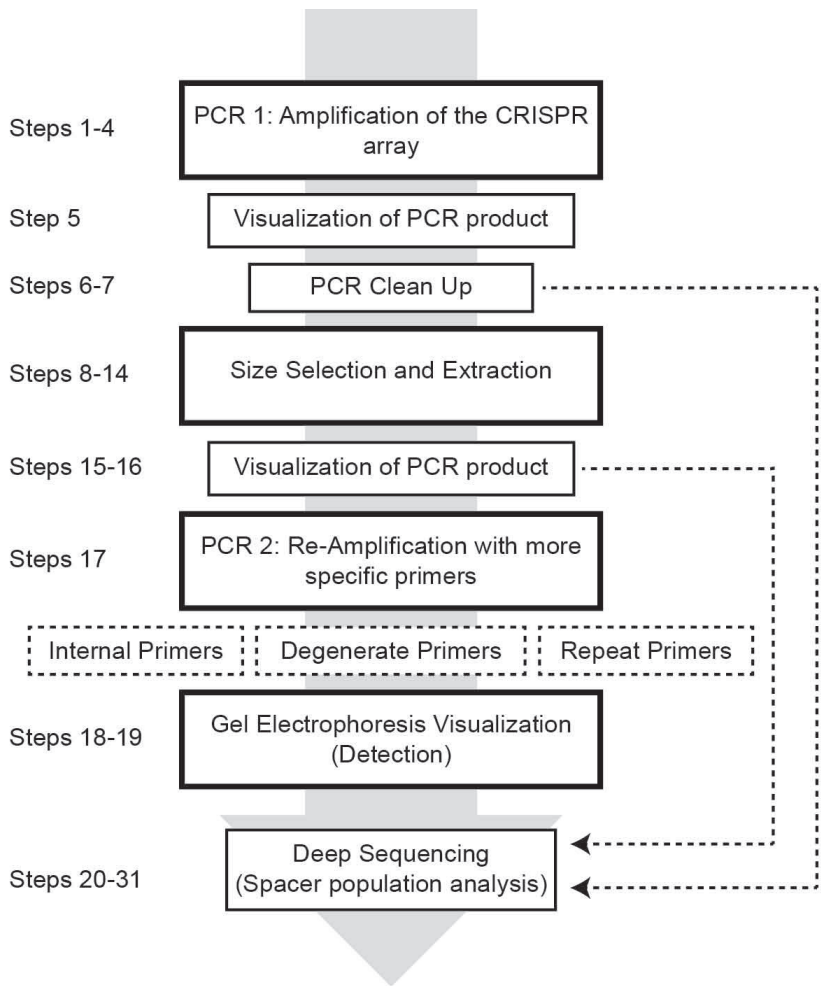
To date a number of evolutionarily diverse CRISPR-Cas systems have been identified and classified into 2 main classes (I and II) and six main types (I – VI) based on their *cas* gene repertoire<sup>4</sup>. Due to the rarity of spacer acquisition events, which are estimated to occur in 1 in  $10^7$  cells in some CRISPR-Cas systems<sup>5,6</sup>, adaptation remains hard to observe. Spacer acquisition was first shown in the type II-A system of *Streptococcus thermophilus*<sup>1</sup> and subsequently in a number of type I systems<sup>7,8</sup>, including the well characterized type I-E system<sup>9–11</sup>. Some type III systems were shown to acquire spacers by converting RNA into DNA through a reverse transcriptase mechanism<sup>12</sup> but spacer acquisition in type III systems seems to be very rare.

### 3.2 Comparison to Existing Approaches

Current methods to detect CRISPR spacer acquisition utilize the conserved aspects of CRISPR array structure and the spacer acquisition mechanism. CRISPR loci are generally comprised of repeat-spacer arrays preceded by an AT-rich leader sequence, which acts as the transcriptional promoter and binding site for factors assisting with spacer acquisition<sup>13</sup>. Spacers are typically between 30 and 40 base pairs, and are inserted at the leader proximal end of the array duplicating the first repeat in the process<sup>2</sup>. The insertion of a spacer results in an increase in CRISPR array size which can be assessed by PCR and harnessed as an indicator for acquisition events<sup>5,7,9–11</sup>. In addition to these conventional PCR approaches, reporter systems have been developed that give a detectable signal upon spacer insertion due to a frameshift<sup>14,15</sup>. While these reporter systems provide a quantifiable and sensitive method of detection, their reliance upon a

frameshift requires strain engineering and limits the number of CRISPR types and systems this technique is applicable to<sup>14,15</sup>. Conventional array amplifying PCR methods are simple and user friendly but often lack the sensitivity that is required to study many aspects of CRISPR adaptation.

Advances in the study of adaptation were made in parallel with the increasing availability of deep sequencing allowing detection of spacer acquisition in minor fractions of the bacteria present in the population<sup>7</sup>. However, massive sequencing efforts suffer from high levels of non-informative PCR amplicons, representing unexpanded CRISPR arrays from cells that did not acquire new spacers. Recently published work, showed the development of a method, SENECA, to eliminate such unexpanded arrays from sequencing samples. This method requires the presence of an engineered restriction endonuclease site flanking the first repeat, which can be harnessed to prevent PCR amplification of non-expanded CRISPR arrays in a second PCR round<sup>16</sup>. Here we present a general strategy termed CAPTURE (CCRISPR Adaptation PCR Technique Using Re-Amplification and Electrophoresis), which can be adapted to any native CRISPR array. The method has been applied in recently published work from our lab<sup>17</sup> and builds on, combines and improves existing tricks to selectively amplify expanded CRISPR arrays<sup>5,7,9–11</sup>. Our method harnesses two PCR steps, separated by automated electrophoresis and extraction of size-selected DNA amplicons, allowing the removal of unexpanded arrays from the sample pool (**Fig. 3.1**). Following this, PCR re-amplification of the captured DNA pool with spacer acquisition specific primers allows visible detection of spacer acquisition within just 1 in 10<sup>5</sup> cells (0.01%) of the population, and potentially up to 1 in 10<sup>8</sup> cells after deep sequencing. With this method we enable the direct study of CRISPR-Cas systems with low acquisition rates, which appears to be a common feature of a large number of bacteria and archaea<sup>5,6</sup>.



**Figure 3.1: Overview of the CAPTURE protocol.** Outline of the major steps required in order to detect spacer acquisition in a population of cells with high sensitivity. Dashed lines indicate steps where choices have to be made before proceeding to the next step.



### 3.3 Potential Applications

The study of CRISPR adaptation reveals the physical interactions of hosts with their invaders, and can be utilized in computational methods to couple viral invaders to unknown hosts and to track spacer acquisition events over time, in sequenced strains<sup>18,19</sup>. Because each bacterial strain has a different history, differences in spacer content can be used to distinguish closely related pathogenic strains, such as *Mycobacterium* and *Yersinia* species<sup>20</sup>. Typically only a small fraction of spacers can be mapped back to known invaders suggesting a vast amount of unexplored invader diversity in nature<sup>19</sup>.

Aside from strain typing, the sequence of new spacers provides mechanistic information for example helping to identify the protospacer adjacent motif (PAM), a critical sequence motif in the DNA that authenticates DNA for cleavage by Cas proteins in type I, II and V CRISPR systems and prevents self-targeting of the CRISPR array<sup>7–11,17</sup>. Here, we aimed to develop a highly sensitive, simple and generally applicable method that could be used to detect the occurrence of such rare spacer acquisition events and unravel further mechanistic details of uncharacterized CRISPR systems. The use of CAPTURE can provide a wealth of information of both uncharacterized, newly discovered and well known CRISPR systems. The method helps to elucidate the role of Cas proteins, the identity of the PAM, the preferred spacer substrates of Cas1-Cas2, and the minimal requirements for both adaptation and defense as exemplified by Kieper et al. where the role of the Cas4 protein in adaptation was recently uncovered<sup>17</sup>.

### 3.4 Limitations

The use of CAPTURE includes the requirement of a gel extraction machine or more laborious manual extraction methods<sup>7</sup>. In order to increase both the utility of this protocol as a tool for the detection of spacer acquisition and to remove non-expanded CRISPR arrays, multiple PCR amplification steps are required. Re-amplification in PCR 2 with primer sets 2, 3 or 4 may introduce PCR bias, which prohibits using absolute abundance levels of new spacers with high confidence. Typically, only unique spacer sequences are used for downstream analysis (**Table 3.1**). It is therefore important that all users take

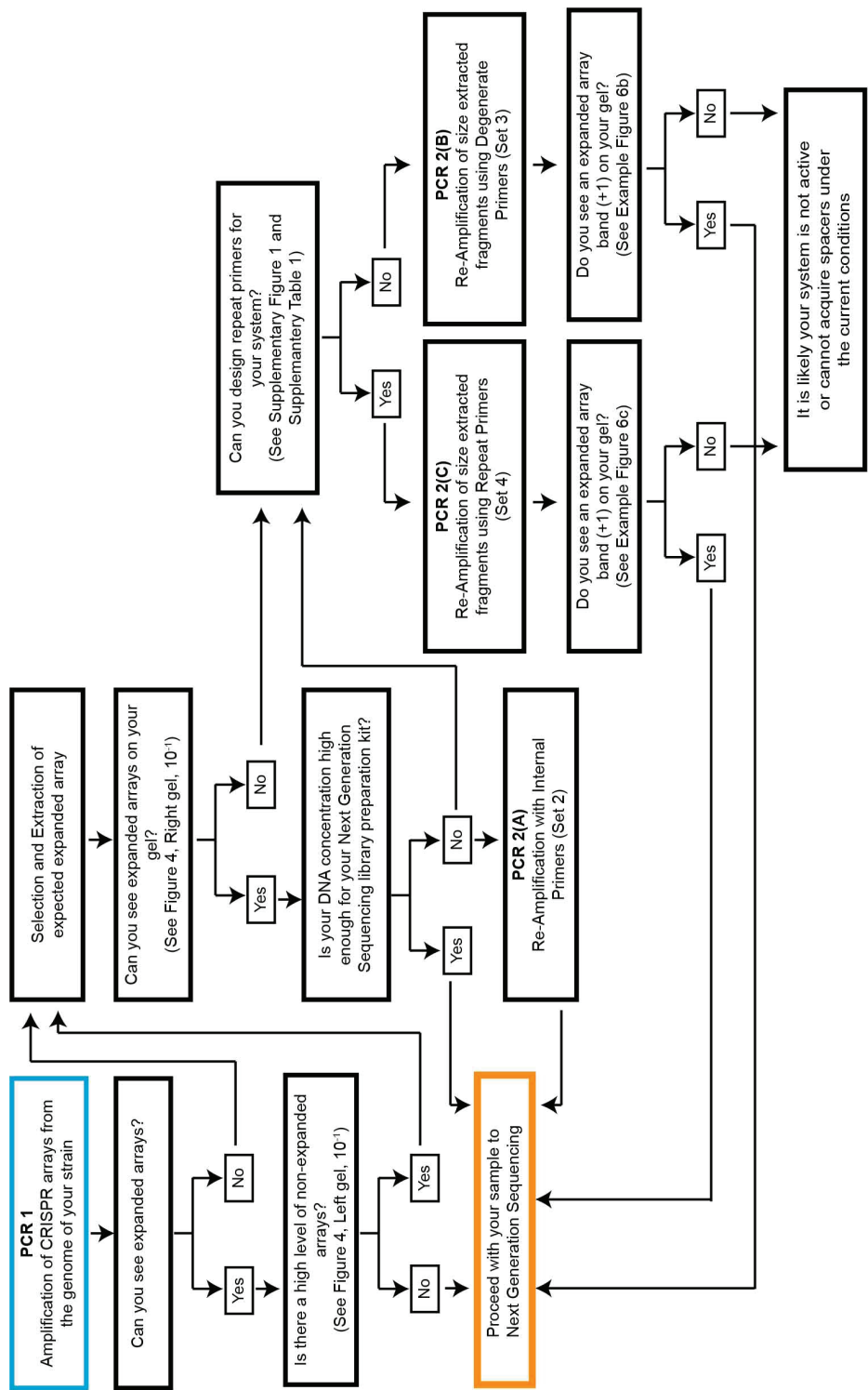
this into consideration when analysing sequencing results. The removal of duplicates during the processing of sequencing reads is recommended and ensures that re-amplification bias does not influence the prevalence of certain spacer sequences in the population. We advise users to consider using biological triplicates and to compare the results obtained from all unique and non-unique spacers. If the degenerate primers are used, it is possible to carry out normalization to correct for the bias introduced by the fixed 3' nucleotides of the primer<sup>21</sup>. This allows the user to draw accurate conclusions about the prevalence of certain spacers in the population.

### 3.5 Experimental Design

This protocol contains a series of steps involving size selection and extraction in combination with two PCR amplifications, PCR 1 and PCR 2. The options provided for these steps can be used in various combinations both to detect if the CRISPR adaptation module in your strain is active, as well as to aid the preparation of a deep sequencing sample containing a majority of arrays that have acquired spacers. Not all steps of the protocol need to be completed if the desired outcome can be reached at an earlier step (**Fig. 3.1 and 3.2**). Prior to use of this protocol, primer design and re-amplification options should be carefully considered, as these will differ depending on both the CRISPR array of the strain used and further actions to be taken with the final sample (**Fig. 3.2, Supplementary Fig. 3.1 and Supplementary Table 3.1**).

Table 3.1: Primer sets and their limitations

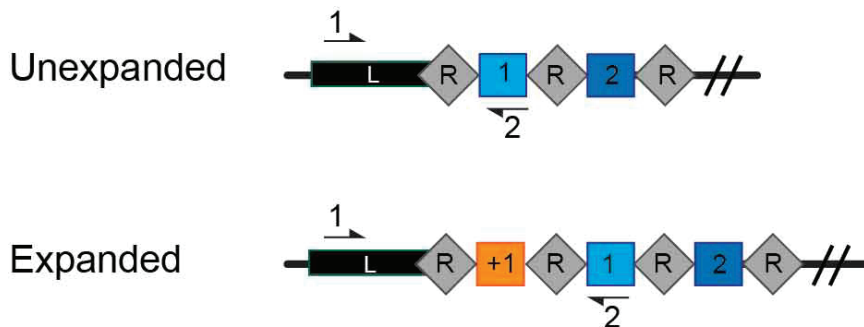
Step in protocol	Primer Set	Desired Outcome for Sample	Limitations	Solution
4	Set 1, Initial Primers	Detection	Does not select specifically for expanded arrays. Can only detect expanded arrays when adaptation occurs in at least 10% of the entire bacterial population	Try size selection and re-amplification
17(A)	Set 2, Internal Primers	Detection	Does not specifically bind expanded arrays and thus does not provide an extra level of detection	Try primer set 3 or 4 for more specific binding of expanded arrays
17(A)	Set 2, Internal Primers	Sequencing	Creates a PCR bias in deep sequencing results and could strongly influence the prevalence of certain spacer sequences	Remove duplicate sequences during analysis. Use biological replicates to confirm observed trends
17(B)	Set 3, Degenerate Primers	Sequencing	Creates a PCR bias due to both stronger G-C annealing at the 3' end and exclusion of the primer ending with the same 3' nucleotide as the existing leader proximal spacer	Complete a normalization during analysis that accounts for the bias introduced by omitting a primer containing the same 3' nucleotide as the first spacer <sup>21</sup> .
17(C)	Set 4, Repeat Primers	Sequencing	Cannot be used for all systems due to the differing repeats sequences. The primers designed need to bind in opposite directions specifically within the repeat sequence. Subsequently these primers may be very short and if the repeat region for a system is AT-rich it may be difficult to design primers with an appropriate annealing temperature.	Try Set 3, degenerate primers



**Figure 3.2: CAPTURE decision chart.** This chart indicates by following the arrows which steps should be taken and which primers should be designed for detection or preparing a sequencing sample.

### 3.5.1 PCR 1, Amplifying CRISPR arrays from a population (Step 4)

The described protocol is designed with the assumption some sequence knowledge or metagenomics data has previously been obtained for the bacterial strain or population of interest. Exact sequence information of the CRISPR array to be studied is required for the primer design step. Primers for PCR 1, the initial array amplification, must be specifically designed for each strain to bind within the leader sequence and where possible within the closest known spacer to the leader, as the sequence of both binding sites can differ greatly between species and subtypes of the CRISPR system. Careful design of primer Set 1 allows amplification of all expanded arrays present within the population and enables elimination of all repeats downstream from spacer 1 (Fig. 3.3). Elimination of excess repeats is important for later steps of the protocol because if a second re-amplification is required, the primer set may bind the repeats specifically, resulting in amplification only when a new spacer is inserted and two repeats are present. For an example of the primers that can be designed, here for the Type I-E CRISPR system of *Escherichia coli* BW25113, see (Supplementary Fig. 3.1 and Supplementary Table 3.1).



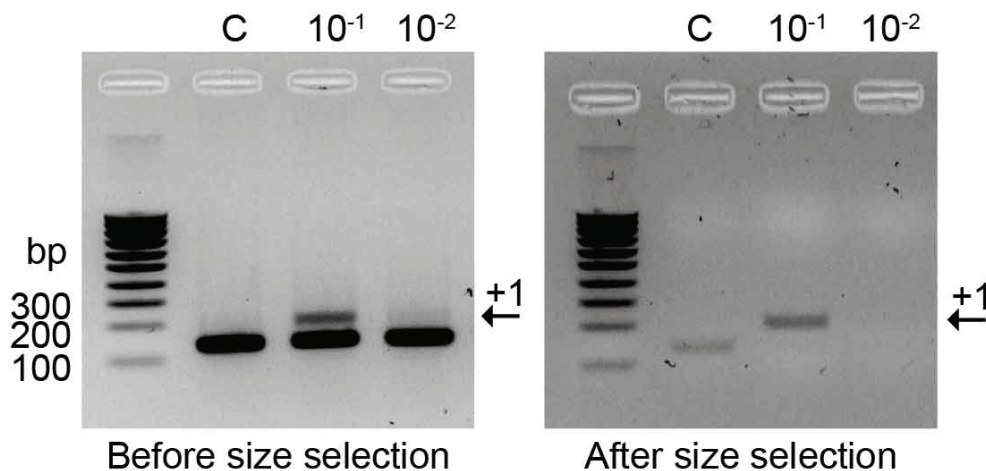
**Figure 3.3: Primer design for amplifying the CRISPR arrays within a population.** Schematic of the CRISPR array consisting of a leader sequence (L) and repeats (R) interspaced by short DNA sequences termed spacers (coloured squares). When acquisition occurs an additional spacer (+1) is added at the leader proximal end (orange) of the array. Primers indicated by black arrows (numbered 1 and 2) are designed to allow PCR amplification of both expanded and unexpanded arrays within the population.

### 3.5.2 Size Selection and Extraction Parameters (Steps 8-14)

In addition to considering which part of the CRISPR array is amplified it is also important to consider the amplified array product size in relation to the following size selection and extraction steps. Smaller bands will allow a higher degree of size separation during use of the Blue Pippin cassette. The highest percentage of agarose available in a Blue Pippin cassette is 3% and allows selection within the range of 100 – 250 bp. This equally applies to users completing the size selection and extraction manually from an agarose gel. For manual extraction we recommend use of a 3% agarose gel to increase expanded and non-expanded array separation. Following this the expanded array band must be excised from the gel, extracted and purified. For optimal selection of only the expanded arrays we suggest both the gel electrophoresis separation and subsequent DNA extraction to be completed twice<sup>7</sup>.

### 3.5.3 Controls (All Steps)

We recommend the use of an unexpanded (native) array control sample throughout the entire procedure to allow assessment of size extraction success and proportion of unexpanded arrays in the final DNA sample. The control should be the same as the CRISPR array of the strain of interest, in which the reverse primer (for example primer 2) will bind in the most leader proximal spacer (**Supplementary Fig. 3.1 and Supplementary Table 3.1**). Either the parental strain with no new spacers or a synthetic DNA construct with the sequence of the unexpanded array could be used as controls throughout this protocol. Use of the original strain of interest as a control can begin at Step 1 with the samples to be tested. A synthetic DNA control should be added as a separate sample during size extraction (Steps 8 -15). Such a control determines the carryover of the non-expanded arrays in the size selection and extraction step and allows optimization of the parameters for both use of the blue pippin and manual extraction (**Fig. 3.4**). The control also aids in the identification of PCR amplification artefacts that can arise due to unspecific annealing temperatures for primer binding or, internal binding of the repeats creating larger PCR amplicons. These artefacts can appear as expanded arrays in the control after size selection and can be removed with small optimizations (see **Table 3.2**)

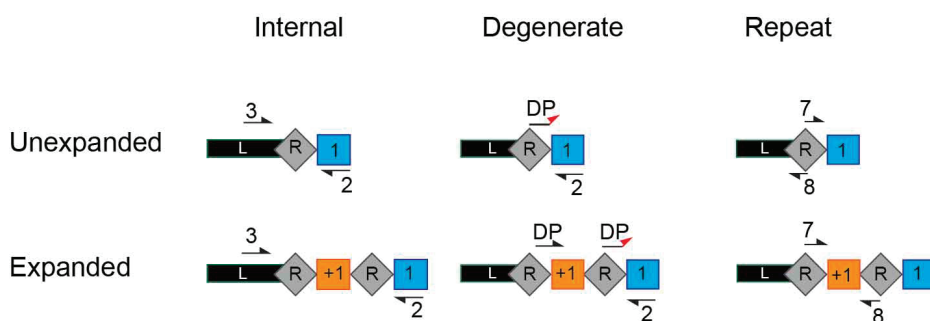


**Figure 3.4: Gel Electrophoresis images of CRISPR arrays before (Step 5) and after (Step 15) size selection.** The proportion of expanded CRISPR arrays (indicated by the black arrow labelled +1) in the sample (10<sup>-1</sup>) is greatly increased after size selection (right) compared to the control (C).

### 3.5.4 PCR 2, Re-Amplification after Size Selection (Step 17)

Re-amplification is an important additional step to this protocol as often the initial amplification (PCR 1) is not sensitive enough to detect acquisition occurring at a low frequency in the population (**Fig. 3.4**). In addition, when high frequency spacer acquisition can be detected in the PCR 1 there is often still a large amount of non-expanded arrays present in the sample, these non-expanded arrays can act as a mask to low frequency acquisition events which are revealed after size selection and re-amplification (**Fig. 3.5**). In this protocol through size selection and re-amplification we aim to maximise both the detection limit and the number of expanded array sequences obtained from deep sequencing. There are three options, Step 17 A, B or C, using primer sets 2, 3 or 4 respectively, for the second amplification step (PCR 2) of this method. The three primer sets we refer to as; (Set2) internal primers (Set 3) degenerate primers and (Set 4) repeat primers, should be carefully designed based on the desired output, subsequent plans for the sample pool and sequence analyses (**Fig. 3.2 and 3.5, Supplementary Fig. 3.1 and Supplementary Table 3.1**). The internal primers (Set 2) allow re-amplification of the expanded arrays enabling confirmation of successful spacer acquisition taking place in just 1 in 10<sup>3</sup> cells in the population of

interest (**Fig. 3.6a**). The degenerate primers (Set 3) allow extremely sensitive detection of acquisition at an occurrence of just 1 in  $10^5$  cells (**Fig. 3.6b**). This amplification is based on the use of 3 forward primers, these primers anneal their 3' nucleotide only with new spacers beginning with a nucleotide different from the original spacer<sup>5</sup> (**Fig. 3.5, Supplementary Fig. 3.1**). The repeat primer (Set 4) strategy ensures that the short 60 bp array products can only be amplified in the presence of 2 repeats (i.e. an expanded array), this process relies on the removal of a repeat during the first PCR amplification but allows high sensitivity detection of acquisition occurring in just 1 in  $10^6$  cells and a final sample population ready for further sequencing (**Fig 3.5 and 3.6c**).



**Figure 3.5: Primer design options for re-amplification of expanded CRISPR arrays.**

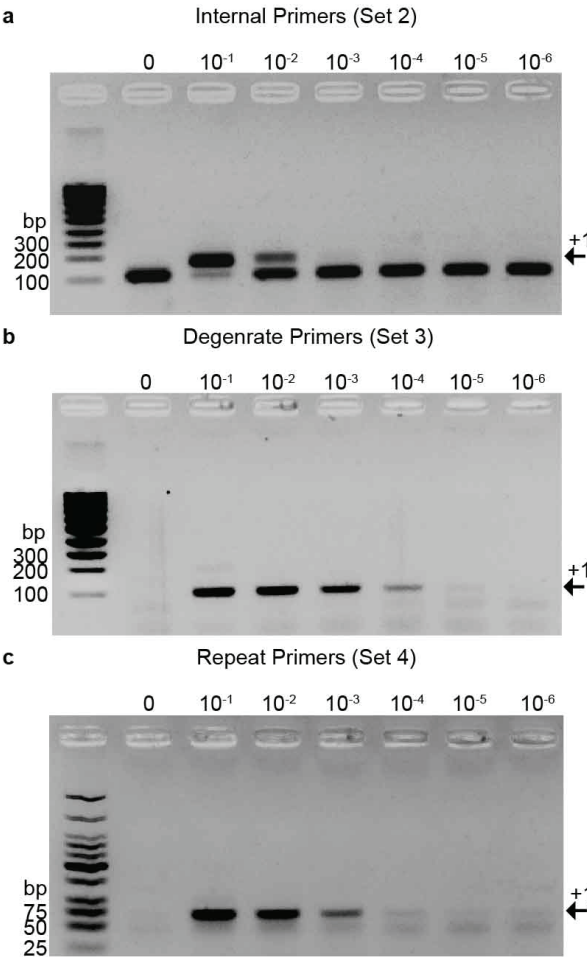
Schematic of the three possible primer options for the second PCR amplification step. The internal primers (left) bind in spacer 1 (primer #2) and the leader sequence (#3) internally of primer #1 used in the first PCR step before size selection. The degenerate primers (middle) consist of 3 forward primers (DP) that anneal their 3' nucleotide only with new spacers (orange) starting with a nucleotide different from the original spacer (blue), in combination with #2. The repeat primers (right) bind within the repeat (#7 and #8) orientated such that a product is only amplified when two repeats are present i.e. the array is expanded.

### 3.5.5 Sequencing (Steps 20-31)

A user may choose to sequence new spacers to determine spacer length distribution (can be consistent or variable), spacer source (bacteriophage genome, plasmid or host genome), or to retrieve the PAM of the newly acquired spacers. Deep sequencing may be used to increase the sensitivity at which new spacers are detected to enable discovery of rare spacer acquisition events. When deep sequencing is carried out it is advisable to re-



amplify expanded PCR amplicons using the CAPTURE method to minimize the presence of non-expanded CRISPR array amplicons. A user may choose not to sequence new spacers, by following the protocol up to and including step 15, when only general spacer acquisition activity is monitored over time, or for example when trying to answer questions about the general effect of *cas* gene mutants on the spacer acquisition activity.



## 3.6 Materials

### 3.6.1 Biological Materials

**CAUTION** Use gloves and a lab coat when working with biological samples.

- *Escherichia coli* K12 BW25113 (Parental strain KEIO Collection, National BioResource Project (NBRP) *E. coli* Strain)<sup>22</sup>.
- *E. coli* PIM5, a variant of the KEIO strain *E. coli* K12 JW12252 (ref. 10, Brouns lab collection). CRITICAL: The protocol described below permits the use of any bacterial strain in which sequence information of the CRISPR array is known or is suspected to be similar to a known array to allow the primer design.

### 3.6.2 Reagents

Lysogeny Broth (LB) Medium (Sigma Aldrich, cat. no. L3022)

GeneJET Genomic DNA Extraction Kit (Thermo Fisher Scientific, cat. no. K0721)

**CAUTION** Components of the kit are dangerous if swallowed and harmful to aquatic life if released into the environment.

OneTaq® Quick-Load® 2X Master Mix with Standard Buffer (Bioké, cat. no. M0486S)

Agarose (Promega, cat no. V3125)

40X TAE buffer stock (Promega, cat no. V4281)

SYBR Safe (Thermo Fisher Scientific, cat no. S33102)

SYBR Gold (Thermo Fisher Scientific, cat no. S11494)

SmartLadder SF (Eurogentec, cat no. MW-1800-04)

Quick-Load® Purple Low Molecular Weight DNA Ladder (NEB, cat no. N0557S)

Primer sets (Figure 3, Figure 5, Supplementary Figure 1 and Supplementary Table 1.) (Integrated DNA Technologies)

Gibco Distilled Water (Thermo Fisher Scientific, cat. No. 10977035)

Absolute Ethanol (Sigma Aldrich, cat. no. 46139) **CAUTION** Flammable, so keep away from heat sources.

GeneJET PCR Purification Kit (Thermo Fisher Scientific, cat. no. K0701)

**CAUTION** Components of the kit are dangerous if swallowed and harmful to aquatic life if released to the environment.

Isopropanol (Sigma Aldrich, cat no. 33539)

Qubit™ dsDNA HS Assay Kit (Thermo Fisher Scientific, cat no. Q32854)

Blue Pippin Cassette kit 3% agarose, including TAE Buffer and Internal Marker Q3 (Sopachem, cat. no. BDF3010)

**CRITICAL** Cassettes must be stored at room temperature (~20°C). The TAE buffer and marker should be stored separately at 4°C.

**CRITICAL** If the Blue Pippin machine is not available for this protocol the selection and extraction step can be carried out manually, additionally requiring:

GeneJET Gel Extraction Kit (Thermo Fisher Scientific, cat. no. K0691)

Next Generation Sequencing library preparation requires:

NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® (Bioke, cat no. E7645S)

NEBNext® Multiplex Oligos for Illumina (Index Primers Set 1) (Bioke, cat no. E7335S)

Agencourt AMPure XP Magnetic Beads 5 mL (Thermo Fisher Scientific, cat no. NC9959336)

2100 Bioanalyzer Instrument (Agilent Technologies, cat no. G2939BA)

### 3.6.3 Equipment

PCR tubes (Sarstedt, cat no. 72.991.002)

Implen NanoPhotometer NP80

Incubators (New Brunswick, cat. no. M1335-0002)

Microcentrifuge tubes (Eppendorf, cat. no. Z606901)

Pipette tips (Thermo Fisher Scientific, cat. no. 7325)

Pipettes (Thermo Fisher Scientific, cat. no. 4642080)

Biometra TOne Thermocycler

Centrifuge

BioRad Gel Doc XR+

VortexBlue Pippin (Sopachem, cat. no. BLU0001)

**CRITICAL** If you do not have the blue pippin machine on offer in your facility, gel extraction of the expanded arrays can be done manually by first concentrating the DNA samples for each condition. This is followed by manual gel extraction; the entire sample including 1 µl loading dye per 6 µl of sample should be loaded in a 3% agarose gel and run at 100 V for ~ 1 hr to maximise band separation, the expanded array (larger) band should be excised from the gel as accurately as possible and DNA extracted using the GeneJET Gel Extraction Kit. The manual gel extraction can be repeated a second time to further remove remaining non-expanded arrays.

Qubit™ 4 Fluorometer (Thermo Fisher Scientific, cat no. Q33226)

Qubit™ Assay Tubes (Thermo Fisher Scientific, cat no. Q32856)

DynaMag™-96 Bottom Magnet (Magnetic Stand) (Thermo Fisher Scientific, cat no. 12332D)

### 3.6.4 Reagent set up

#### Fresh Lysogeny Broth (LB) Medium

To produce 1L of media weigh 5 g/L yeast extract, 10 g/L NaCl, 10 g/L tryptone and mix well with 800 mL distilled H<sub>2</sub>O. After mixing thoroughly add the final 200mL of distilled H<sub>2</sub>O. The media must then be sterilized by autoclaving at 121°C for 20 mins and can be stored short term at room temperature.

#### Degenerate Primer Master Mix

To create the degenerate primer master mix used in Step 16, mix the three forward degenerate primers containing differing 3' nucleotides that do not match the existing spacer in your system (See Experimental Design and Supplementary Table 1) equally to give a final concentration of ~3.3 µM of each primer and total concentration of 10 µM. The primer master mix can be stored in the – 20°C freezer indefinitely.

#### 1X TAE Buffer

Dilute a 40X TAE Buffer stock using MilliQ H<sub>2</sub>O. 1X TAE buffer can be stored at room temperature for a recommended maximum of 3 years.

#### Agarose Gels

Make fresh when needed by combining agarose and 1X TAE buffer wt/vol. The most commonly used gels in this protocol are 2% i.e. 1 g agarose and 50 mL 1X TAE. Mix and heat these components for ~ 1 min. Then add 5 µL of Sybr safe (10,000X) to allow visualization of the DNA and pour into a gel mold with a comb. Leave to set for at least 10 mins.

## 3.7 Procedure

### 3.7.1 Preparation of samples for use as PCR template

#### TIMING ~12 hrs + 1 hr

1| Set up an overnight culture of your strain of interest at the optimal temperature for growth (here, 37°C, 180 rpm).

2| **(Optional)** Extract the genomic DNA of your sample using the GeneJET Genomic DNA Extraction Kit (ThermoFisher Scientific) following the supplied protocol. Elute the DNA in 200 µl of molecular biology grade H<sub>2</sub>O or less where a low concentration is expected.

**CRITICAL STEP** If your sample is acquired from the environment or a different source, there may be a more appropriate kit to complete this step.

**CRITICAL STEP** Genomic extraction is recommended when your sample has a low cell density or may contain contaminants that will affect subsequent PCR reactions.

3| **(Optional)** Measure the DNA concentration in ng/μl with a NanoPhotometer.

**PAUSE POINT** Can be stored at 4°C overnight or frozen at -20°C for long term storage.

3.7.2 Amplification of CRISPR arrays present in sample population

**TIMING** ~3 hrs

4| Using either the overnight culture from Step 1 directly, or the sample prepared in optional Steps 2 and 3 as a template, prepare a PCR reaction as follows:

Reagent	Volume Added	Final Concentration
OneTaq® Quick-Load® 2X Master Mix with Standard Buffer	25 μl	1X
FWD Primer Set 1 (10μM)	2 μl	0.4 μM
REV Primer Set 1 (10μM)	2 μl	0.4 μM
Template DNA or Overnight Culture of your strain of interest	x μl	~ 100 ng
Molecular biology grade H <sub>2</sub> O	Up to 50 μl	

Then perform a PCR amplification under the following conditions:

Cycles	Denature	Anneal	Elongate	Extend
35	94°C for 30 s	58°C* for 30 s	68°C for 30 s*	68°C 2 min

\* Conditions must be adapted to your specific primers and or product length

**CRITICAL STEP** If you are using an overnight culture as template for the initial PCR increasing your denaturation time to 5 mins at 98 °C can help aid cell lysis making more genomic DNA available as a template in PCR the reaction tube.

5| Run the PCR products (5-10 μl) on a 2% (wt/vol) agarose gel containing SYBR safe (1 μl / 10 ml) at 100 V for ~30min to check for the presence of a single band of your desired size (gel extraction or PCR optimization is needed if more bands are present).

**CRITICAL STEP** The gel percentage can be adapted for your expected fragment size. A higher percentage of agarose allows larger separation of smaller bands.

? See Troubleshooting (3.7.5)

6| Clean and concentrate the PCR(s) using the ThermoFisher Scientific GeneJET PCR Purification kit, following the manufacturer's instructions. Elute in 30 µl molecular biology grade H<sub>2</sub>O.

**CRITICAL STEP** Add isopropanol in equal proportion to the DNA binding buffer when purifying fragments smaller than 500 bp.

**CRITICAL STEP** The PCR reaction must be purified and H<sub>2</sub>O must be used during the elution of samples from kits, as salts in the PCR reaction or elution buffer can interfere in future steps using the Blue Pippin Machine.

7| Measure the concentration of DNA (ng/µl) on the NanoPhotometer. In the next steps large quantities of DNA can be lost so a concentration of at least 30 ng/µl in 30 µl is advised; run additional PCRs and pool if more DNA is needed.

**CRITICAL STEP** Between 20 and 50% of the DNA sample can be lost during the Blue Pippin process so it is better to load as much as possible without exceeding the maximum of 5 µg of DNA

**PAUSE POINT** DNA can be stored at 4°C overnight or frozen at -20°C for long term storage.

#### Size Selection and Extraction of Expanded Arrays

**TIMING** ~2 hrs 45 mins

**CRITICAL** In the absence of the Blue Pippin, manually extracting the expanded arrays from a gel will take ~ 4 hrs

8| Based on the size (or sizes if multiple spacers were acquired) of your arrays visualized in Step 5 choose an appropriate Blue Pippin Cassette. For arrays between 100 – 250 bp the 3% Gel Cassette with Internal Standards Marker Q3 (Cat. no. BDF3010) is advised.

9| Calibrate the Blue Pippin machine as described in the Sage Science Blue Pippin Operations Manual.

10| Load your cassette of choice into the machine, exchange the elution buffer and check buffer levels as described in the Sage Science Blue Pippin Operations Manual.

**CRITICAL STEP** The loading buffer must be at room temperature (~20°C) before use in the Blue Pippin gel cassette

11| Prepare your samples for loading into the Blue Pippin (final volume 40  $\mu$ l) by adding 10  $\mu$ l of assigned marker or loading solution. **CRITICAL STEP:** This step is cassette dependant.

**CRITICAL STEP** Samples should be vortexed well and spun down before loading. The loading solution contains a densifying agent that will allow the sample to sink below the electrophoresis buffer layer in the well, increasing the size selection accuracy.

**CRITICAL STEP** The fluorescein labels (on the DNA marker) will degrade at room temperature, minimize time spent on the bench or keep on ice.

12| Program the selection range as outlined in the Sage Science Blue Pippin Operations Manual to collect the size range in which all expanded arrays are included.

**CRITICAL STEP** Programming a wider selection range can improve DNA yield (For example ~20 bp either side of your desired fragment size).

13| Load samples slowly into appropriate lanes and run the program as described in the Sage Science Blue Pippin Operations manual.

14| Collect the DNA extracted for the defined size range from the elution chamber.

15| **(Optional)** Visualize DNA by gel electrophoresis (Figure 4). Run the extracted DNA products (5-10  $\mu$ l) on a 2% (wt/vol) agarose gel containing SYBR safe (1  $\mu$ l / 10 ml) at 100 V for ~30 min and image to check for the presence of expanded arrays.

**CRITICAL STEP** This step should be carried out if the user simply desires to check if their CRISPR system actively incorporates new spacers or not. Visualization of the extracted products allows confirmation of the ability to acquire spacers, and can be followed by optional deep sequencing.

**CRITICAL STEP** If the user desires to carry out deep sequencing without further re-amplification this step should be carried out. Visualization of the extracted products allows assessment of the success of the extraction and verification of the presence of expanded arrays in the sample before sequencing.

**? See Troubleshooting (3.7.5)**

16| Measure the concentration of DNA from Step 14 using the Nanophotometer to assess which step needs to be taken next (See Figure 2 and Experimental Design).

**? See Troubleshooting (3.7.5)**

**PAUSE POINT DNA** can be stored at 4°C overnight or frozen at -20°C for long term storage.

**CRITICAL STEP** If your DNA concentration is high enough immediately after this step to be used as input DNA for next generation sequencing library preparation with your kit of choice it is possible to proceed directly to sequencing (Step 20). However, it is important to note that higher starting concentrations are often recommended for library preparation kits as DNA can be lost when using magnetic beads.

**3.7.3 Re-Amplification of the size selected arrays to increase detection sensitivity**

**TIMING ~2 hrs**

17| Use the size selected DNA fragments as template DNA for a PCR amplification with one of the primer sets outlined in the Experimental Design section 'Re-Amplification after Size Selection' (Figure 5, Supplementary Figure 1 and Supplementary Table 1). Follow option A for re-amplification with internal primers, option B for re-amplification with degenerate primers, or option C for re-amplification with repeat primers. Option A should be used when a high acquisition rate is observed for your strain (Figure 6a). Options B and C (degenerate and repeat primers respectively) are capable of detecting rare events, enabling sensitive detection on gel, further detection through deep sequencing and removing the unexpanded background (Figure 6b and c). However, only the option C maintains full spacer diversity (Table 1).

**CRITICAL STEP** Primer set must be system specific and chosen carefully see details in Experimental Design and recommendations below.

**(A) Re-Amplification with Internal Primers**

(i) Prepare a PCR reaction as follows, using the fragments collected in Step 14 as a template:

Reagent		Volume Added		Final Concentration
OneTaq®	Quick-Load®	2X	25 µl	1X
Master Mix with Standard Buffer				
FWD Internal Primer Set	(10µM)	2	2 µl	0.4 µM
REV Internal Primer Set	(10µM)	2	2 µl	0.4 µM
Size Extracted Template DNA		x µl		
Molecular Biology Grade H <sub>2</sub> O		Up to 50 µl		

Perform PCR amplification under the following conditions:



Cycles	Denature	Anneal	Elongate	Extend
25	94°C for 30 s	62°C* for 30 s	68°C for 30 s*	68°C 2 min

\* Conditions must be adapted to your specific primers and or product length

**(B) Re-Amplification with Degenerate Primers**

(i) Prepare a PCR reaction as follows, using the fragments collected in Step 14 as a template:

Reagent	Volume Added	Final Concentration
OneTaq® Quick-Load® 2X Master Mix with Standard Buffer	25 µl	1X
FWD Degenerate Primer Master Mix, Set 3 (10µM)	2 µl	0.4 µM
REV Degenerate Primer Set 3 (10µM)	2 µl	0.4 µM
Size Extracted Template DNA	x µl	
Molecular biology grade H2O	Up to 50 µl	

Perform PCR amplification under the following conditions:

Cycles	Denature	Anneal	Elongate	Extend
25	94°C for 30 s	62°C* for 30 s	68°C for 30 s*	68°C 2 min

\* Conditions must be adapted to your specific primers and or product length

**(C) Re-Amplification with Repeat Primers**

(i) Prepare a PCR reaction as follows, using the fragments collected in Step 14 as a template:

Reagent	Volume Added	Final Concentration
OneTaq® Quick-Load® 2X Master Mix with Standard Buffer	25 µl	1X
FWD Repeat Primer Set 4 (10µM)	1 µl	0.2 µM
REV Repeat Primer Set 4(10µM)	1 µl	0.2 µM
DMSO	1.5 µl	3 %
Size Extracted Template DNA	0.5 µl	
Molecular biology grade H2O	Up to 50 µl	

Perform PCR amplification under the following conditions:

Cycles	Denature	Anneal	Elongate	Extend
20	94°C for 15 s	58°C* for 30 s	68°C for 20 s*	68°C 2 min

\* Conditions must be adapted to your specific primers and or product length

3.7.4 Visualizing spacer acquisition by Gel Electrophoresis

**TIMING ~45mins**

18| Run the PCR products (5-10 µl) on a 2% (wt/vol) agarose gel containing SYBR safe (1 µl / 10 ml) at 100 V for ~30 min to check for the presence of a single band of your desired size (gel extraction or PCR optimization is needed if more bands are present).

? See Troubleshooting (3.7.5)

19| Assess whether adaptation has occurred (presence of a band with the expected size) and when next generation sequencing is planned assess if your sample pool has a greater proportion of expanded arrays compared to the unexpanded control array (Figure 6).

**Next generation sequencing to investigate spacer content (Optional)**

**TIMING ~ 1 day sample library preparation, 1-2 days sequencing**

CRITICAL: Sequencing of samples with a next generation sequencing (NGS) platform allows even greater detection and more importantly sequence specific information for the spacers acquired within the population. There are many kits and sequencing platforms that could be used to sequence expanded array PCR amplicons, below we provide the preferences from our lab.

20| Measure the concentration of DNA (ng/µl) in your samples from Step 17 using the Fluorometric Qubit machine for higher accuracy. Prepare samples with the dsDNA HS kit (10 pg/µL to 100 ng/µL) in a thin walled tube to allow accurate measurement.

21| Choose the appropriate library preparation kit for your samples input concentrations and barcoding required. Here we use the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® in combination with NEBNext® Multiplex Oligos for Illumina.

22| Prepare your samples with the chosen kit to add flow cell binding adapters and where applicable individual barcodes to each sample. This involves end preparation (A-tailing), adaptor ligation, size selection using magnetic beads, PCR amplification to add barcodes and a final clean up with magnetic beads, following the manufacturer's instructions.

**CRITICAL STEP** Beads need to be at room temperature before use. Remove from 4 °C 30 mins before use.

23| Assess the quality of your sample using the Qubit (DNA concentration in ng/μl) and where available Bioanalyzer (fragment size and molar concentration).

24| If you will be sequencing your sample in house: Based on the guidelines for your sequencing instrument (here Miseq) pool all barcoded sample libraries at the desired proportions to the recommended final concentration here (2 nM). Denature your sample with an equal amount of freshly diluted 0.2 N NaOH and dilute to the appropriate loading concentration and volume for the flow cell following the Illumina guidelines for the specific platform. **CRITICAL:** The entire sample is not required for sequencing and it is therefore recommended to store a portion of your sample library before dilution and denaturation at -20 °C for long term storage.

**CRITICAL STEP** Use equimolar amounts of all barcoded libraries to maintain equal depth among all samples sequenced.

**CRITICAL STEP** The recommended loading concentration will differ for each sequencing machine. The recommended concentration for the MiSeq is 6 – 10 pM, here we use 4 pM, as we load very small fragments that cluster more efficiently than the recommended minimum 250 bp fragments.

**PAUSE POINT** Prepared sample library can be stored long term at 20 °C until sequencing will be carried out.

25| Following the specific instructions for the sequencing instrument load the appropriately diluted sample into the reagent cartridge. Next clean and load the flow cell into the machine, check for any additional solutions required and start the run. Alternatively, frozen libraries prepared above can be sent for custom sequencing.

26| In order to analyse sequencing data for each sample, the sequences must first be quality assessed, normalized, merged where necessary and de-multiplexed (separated by unique barcodes into samples). This can be carried out at a basic level by the sequencing instrument itself, or a software package or more detailed pipeline can be used<sup>7,17,21</sup>.

27| Once de-multiplexed, for each sample all spacer-containing sequences can be extracted by identifying conserved CRISPR array regions such as the 3' end of the degenerate primer and the 5' end of the repeat. The sequences flanked by these regions can be extracted as spacer sequences<sup>23</sup>.

28| In order to assess spacer content extracted sequences are aligned to potential spacer sources such as the host genome, other known factors in the cell such as plasmids or when the potential sources are unknown, sequences can be screened using BLAST (Basic Local Alignment Search Tool) or other sequence databases<sup>24</sup>.

29| After extraction of all spacer sequences duplicate spacers should be removed to determine the number of unique spacers acquired and to remove the duplication bias created by PCR amplification. In addition, normalization should be applied to the sequences if any form of degenerate primer was used for re-amplification in PCR 2 (See Table 1)<sup>21</sup>.

30| The orientation of the extracted spacers then needs to be determined to enable extraction of the PAM containing nucleotides adjacent to the target sequence<sup>7,17,25</sup>. Web tools such as WebLogo can then be used to further aid determination of the strongest PAM sequence for the system of interest<sup>9,26</sup>.

31| Finally, spacer content (nucleotide sequence) and length of the extracted sequences can be analysed and compared between conditions<sup>17</sup>.

### 3.7.5 Troubleshooting

**Table 3.2. Troubleshooting**

Step	Problem	Possible Reason	Solution
5	Multiple bands seen after PCR amplification	Primers are binding to multiple regions	Annealing temperature of primers must be adjusted (increased) to be as specific as possible
5	No PCR products	Primers were not optimized to specific conditions	Consider a gradient PCR or decreased annealing temperature
16	Low DNA yield after size selection	Up to 50% of DNA can be lost during the extraction procedure	<p>A Blue Pippin cassette in which your DNA fragment size is in the middle of the selection range should be chosen</p> <p>Programming a wider selection range should also increase your yield</p> <p>Increase the concentration of DNA loaded into the blue pippin cassette by pooling multiple PCRs</p>

15	A large proportion of unexpanded arrays are present in the extracted sample	Selection range was too close to the unexpanded array fragment size	<p>Increase the size of the collection starting point to as close to the expanded array size as possible</p> <p>To widen the overall range and increase the yield as mentioned previously you can instead increase the end collection size</p>
18	A large proportion of unexpanded arrays are present after internal primer re-amplification	The acquisition rate may be very low	Try using the degenerate or repeat primers to re-amplify the extracted fragment pool
18	No PCR product can be visualized after re-amplification of size extracted fragments	Low DNA yield from the extraction step	Run the extracted fragment pool directly on a gel to ensure product was extracted
18	No or low PCR yield	<p>Not enough template used for PCR</p> <p>Primers were not optimized to specific conditions</p>	<p>The template DNA concentration as indicated in the protocol can be increased</p> <p>Consider a gradient PCR or lower annealing temperature</p>
18	High levels of the expanded band can be seen in the control after re-amplification	<p>There may be carryover of longer products amplified in PCR 1 from the entire array.</p> <p>There could be a contamination or an amplification artefact, caused by repeats binding to each other, present in the sample</p>	<p>Increase the annealing temperature to increase the primer binding specificity in PCR 1</p> <p>Increase the annealing temperature in PCR 2</p> <p>Lower the number of cycles and template used for the re-amplification PCR and check control levels</p>
18	High levels of the unexpanded array band can be seen in the control after re-amplification	A large proportion of unexpanded or control arrays are still present after size extraction	<p>Lower the number of cycles and template used for the re-amplification PCR and check control levels</p> <p>Increase the annealing temperature to increase the primer binding specificity</p>

			<p>Consider changing primer set for the re-amplification</p> <p>If this does not work repeat the size extraction using the previously extracted sample pool</p> <p>Or alternatively use the size extracted DNA directly for deep sequencing</p>
18	Bands are not clearly distinguishable enough on the agarose gel to determine expanded: unexpanded array ratio	Gel was not run long enough or DNA yield is low and hard to visualize	<p>Increase the percentage of your gel and run the products for longer</p> <p>If this does not work consider staining with SYBR Gold for clearer bands</p>

### 3.7.6 Timing

Step 1, Overnight growth ~12 hrs

Steps 2 and 3 (optional), Genomic DNA extraction 1 hr

Step 4, PCR amplification 1 2 hrs

Steps 5-7, Gel electrophoresis and purification of sample 1 hr

Steps 8-14, Extraction of expanded arrays 2 hrs

Step 15 (optional), Visualization of extracted DNA 45 mins

Steps 16, Determination of DNA concentration 10 mins

Step 17, PCR amplification 2 ~2 hrs

Steps 18 and 19, Gel electrophoresis to visualize expanded arrays and determine if sequencing is necessary 45 mins

Steps 20-24 (optional), Sample preparation for next generation sequencing ~1 day

Step 25-31 (optional), Next generation sequencing sequencing and analysis of prepared sample ~1-2 days

### 3.8 Anticipated Results

The outlined protocol provides a method enabling detection of spacer acquired by active CRISPR systems even at extremely low levels. To test the sensitivity of CAPTURE we mimicked a population in which only a small fraction of the cells had acquired one new spacer. To achieve this we used both the wild type strain *Escherichia coli* K12 BW25113 containing a type I-E CRISPR system<sup>27</sup> and a variant of this strain, known to contain an extra 32 bp spacer<sup>10</sup>, referred to as unexpanded and expanded respectively (Figure. 3). The low prevalence of acquisition was simulated by the serial dilution of genomic DNA extracted from the expanded strain to  $10^{-6}$  in DNA extracted from the unexpanded strain. Such a dilution series allowed us to test the true sensitivity of our method by simulating acquisition rates as low as 1 in  $10^6$  expanded arrays (0.0001% of the population) with an upper limit of 1 in 10 expanded arrays (10% of the population) acquiring a spacer (**Fig. 3.6**). In our experience if all steps are followed this procedure can detect spacers acquired in just 1 in  $10^5$  cells of a population (**Fig. 3.6**), a sensitivity which can be increased even further with the subsequent use of deep sequencing on the prepared sample<sup>17</sup>.

### 3.9 Acknowledgements

SJJB is supported by European Research Council (ERC) StG grant 639707, NWO Vidi grant 864.11.005, and a TU Delft start-up grant. This work was supported by the Netherlands Organisation for Scientific Research (NWO/OCW), as part of the Frontiers of Nanoscience (NanoFront) program. The authors would like to thank members of the Brouns Lab for discussions on the manuscript and feedback.

### 3.10 Author Contributions

Conceived and designed the experiments: REM, CA, JV and SJJB. Performed the experiments: REM, CA and JV. Analyzed the data: REM, CA, JV and SJJB. Wrote the paper: REM and SJJB with input from all authors.

### 3.10 Supplementary Information

**Supplementary Table 3.1: Oligonucleotide sets used for this protocol**

Primer Set	Step Primers are used in	Primer name (in this protocol)	Sequence 5' – 3'	Description
Set 1 (Initial Primers)	4	1	GGATGTGTTGTTGTGTGATAC	FWD Anneals in the leader sequence of CRISPR2.3
		2	ACGCCTTTTTCGATTGC	REV Anneals in Spacer 1 of CRISPR 2.3
Set 2 (Internal Primers)	17 (A)	3	GTTGGTAGATTGTGACTGGC	FWD Anneals in the leader sequence of CRISPR2.3 (Internal)
		2	ACGCCTTTTTCGATTGC	REV Anneals in Spacer 1 of CRISPR 2.3
Set 3 (Degenerate Primers)	17 (B)	4	AGCGGGGATAAACCG <u>C</u>	FWD Anneals in the type I-E repeat with C at the 3' end
		5	AGCGGGGATAAACCG <u>A</u>	FWD Anneals in the type I-E repeat with A at the 3' end
		6	AGCGGGGATAAACCG <u>T</u>	FWD Anneals in the type I-E repeat with T at the 3' end
		2	ACGCCTTTTTCGATTGC	REV Anneals in Spacer 1 of CRISPR 2.3
Set 4	17 (C)	7	GCTGGCGCGGGGAACAC	FWD Anneals in



(Repeat Primers)				type I-E repeat
		8	GCCAGCGGGGATAAACC	REV Anneals in type I-E repeat

\*Underlined character indicates a unique spacer binding nucleotide



## References

1. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–12 (2007).
2. Jackson, S. A. *et al.* CRISPR-Cas: Adapting to change. *Science* **356**, eaal5056 (2017).
3. Brouns, S. J. J. *et al.* Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960–4 (2008).
4. Koonin, E. V., Makarova, K. S. & Zhang, F. Diversity, classification and evolution of CRISPR-Cas systems. *Current Opinion in Microbiology* **37**, 67–78 (2017).
5. Heler, R. *et al.* Cas9 specifies functional viral targets during CRISPR–Cas adaptation. *Nature* **519**, 199–202 (2015).
6. Hynes, A. P., Villion, M. & Moineau, S. Adaptation in bacterial CRISPR-Cas immunity can be driven by defective phages. *Nat. Commun.* **5**, 4399 (2014).
7. Staals, R. H. J. *et al.* Interference-driven spacer acquisition is dominant over naive and primed adaptation in a native CRISPR–Cas system. *Nat. Commun.* **7**, 12853 (2016).
8. Datsenko, K. A. *et al.* Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat. Commun.* **3**, 945 (2012).
9. Yosef, I., Goren, M. G. & Qimron, U. Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Res.* **40**, 5569–5576 (2012).
10. Swarts, D. C., Mosterd, C., van Passel, M. W. J. & Brouns, S. J. J. CRISPR Interference Directs Strand Specific Spacer Acquisition. *PLoS One* **7**, e35888 (2012).
11. Yosef, I. *et al.* DNA motifs determining the efficiency of adaptation into the *Escherichia coli* CRISPR array. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 14396–401 (2013).
12. Silas, S. *et al.* Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase-Cas1 fusion protein. *Science (80-. )*. **351**, aad4234–aad4234 (2016).
13. Nuñez, J. K., Bai, L., Harrington, L. B., Hinder, T. L. & Doudna, J. A. CRISPR Immunological Memory Requires a Host Factor for Specificity. *Mol. Cell* **62**, 824–833 (2016).
14. Amlinger, L., Hoekzema, M., Wagner, E. G. H., Koskiniemi, S. & Lundgren, M. Fluorescent CRISPR Adaptation Reporter for rapid quantification of spacer acquisition. *Sci. Rep.* **7**, 10392 (2017).
15. Díez-Villaseñor, C., Guzmán, N. M., Almendros, C., García-Martínez, J. & Mojica, F. J. M. CRISPR-spacer integration reporter plasmids reveal distinct genuine acquisition specificities among CRISPR-Cas I-E variants of *Escherichia coli*. *RNA Biol.* **10**, 792–802 (2013).
16. Schmidt, F., Cherepkova, M. Y. & Platt, R. J. Transcriptional recording by CRISPR spacer acquisition from RNA. *Nature* **562**, 380–385 (2018).

17. Kieper, S. N. *et al.* Cas4 Facilitates PAM-Compatible Spacer Selection during CRISPR Adaptation. *Cell Rep.* **22**, 3377–3384 (2018).
18. Edwards, R. A., McNair, K., Faust, K., Raes, J. & Dutilh, B. E. Computational approaches to predict bacteriophage–host relationships. *FEMS Microbiol. Rev.* **40**, 258–272 (2016).
19. Snyder, J. C., Bateson, M. M., Lavin, M. & Young, M. J. Use of Cellular CRISPR (Clusters of Regularly Interspaced Short Palindromic Repeats) Spacer-Based Microarrays for Detection of Viruses in Environmental Samples. *Appl. Environ. Microbiol.* **76**, 7251–7258 (2010).
20. Shariat, N. & Dudley, E. G. CRISPRs: Molecular Signatures Used for Pathogen Subtyping. *Appl. Environ. Microbiol.* **80**, 430–439 (2014).
21. Modell, J. W., Jiang, W. & Marraffini, L. A. CRISPR–Cas systems exploit viral DNA injection to establish and maintain adaptive immunity. *Nature* **544**, 101–104 (2017).
22. Baba, T. *et al.* Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2**, 2006.0008 (2006).
23. Biswas, A., Gagnon, J. N., Brouns, S. J. J., Fineran, P. C. & Brown, C. M. CRISPRTarget: bioinformatic prediction and analysis of crRNA targets. *RNA Biol.* **10**, 817–27 (2013).
24. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
25. Biswas, A., Staals, R. H. J., Morales, S. E., Fineran, P. C. & Brown, C. M. CRISPRDetect: A flexible algorithm to define CRISPR arrays. *BMC Genomics* **17**, 356 (2016).
26. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–90 (2004).
27. Díez-Villaseñor, C., Almendros, C., García-Martínez, J. & Mojica, F. J. M. Diversity of CRISPR loci in Escherichia coli. *Microbiology* **156**, 1351–61 (2010).





# 4

## Direct visualization of native CRISPR target search in live bacteria reveals Cascade DNA surveillance mechanism

Published as: J. N. A. Vink, K. J. A. Martens, M. Vlot, R. E. McKenzie, C. Almendros, B. Estrada Bonilla, D. J. W. Brocken, J. Hohlbein, S. J. J. Brouns, Direct Visualization of Native CRISPR Target Search in Live Bacteria Reveals Cascade DNA Surveillance Mechanism. *Mol. Cell.* **77**, 39-50.e10 (2020).

## **Abstract**

CRISPR-Cas systems encode RNA-guided surveillance complexes to find and cleave invading DNA elements. While it is thought that invaders are neutralized minutes after cell entry, the mechanism and kinetics of target search and its impact on CRISPR protection levels have remained unknown. Here we visualized individual Cascade complexes in a native type I CRISPR-Cas system. We uncovered an exponential relationship between Cascade copy number and CRISPR interference levels, pointing to a time-driven arms race between invader replication and target search, in which 20 Cascade complexes provide 50% protection. Driven by PAM-interacting subunit Cas8e, Cascade spends half its search time rapidly probing DNA (~30 ms) in the nucleoid. We further demonstrate that target DNA transcription and CRISPR arrays affect the integrity of Cascade and impact CRISPR interference. Our work establishes the mechanism of cellular DNA surveillance by Cascade that allows the timely detection of invading DNA in a crowded, DNA-packed environment.

## 4.1 Introduction

RNA-guided CRISPR-Cas surveillance complexes have evolved to specifically and rapidly recognize sequences of previously catalogued mobile genetic elements (MGEs) <sup>1</sup>. Target DNA recognition depends on CRISPR RNA (crRNA) – DNA complementarity and on the presence of a protospacer adjacent motif (PAM), a short nucleotide sequence flanking the target site <sup>2,3</sup>. To work effectively, the complexes need to find their targets fast enough to prevent an MGE from becoming established in the cell, which can occur within minutes upon cell entry <sup>4</sup>. Target search inside a cell faces a multitude of challenges: Firstly, cells are packed with DNA, and crRNA surveillance complexes need to find the needle in a haystack before an invading element takes control of the cell. PAM scanning and crRNA-seed interactions with the target have been suggested to speed up the search process by drastically reducing the number of potential target sites in the genome <sup>5,6</sup>. Several studies have shown that crRNA-effector complexes spend more time probing PAM rich regions, which is indicative of its function as the first recognition site <sup>7-9</sup>. The *Escherichia coli* K12 genome contains 127,081 preferred PAMs (CTT) that are recognized by the crRNA-effector complex Cascade in the Type I-E CRISPR-Cas system <sup>10</sup>. This large number of PAMs suggests that the interaction with the PAM needs to be sufficiently fast to cover enough sequence space to find an invading DNA sequence in time. A second challenge is posed by the action of other proteins present in the cell such as DNA binding proteins, DNA or RNA polymerases that may interfere with target search and formation of target bound crRNA complexes <sup>5,11</sup>. Some invading MGEs even use specialized anti-CRISPR proteins to inhibit crRNA-effector complexes and impair the target search process <sup>12,13</sup>. A third challenge that microbes face is to produce appropriate levels of Cascade complexes loaded with one particular crRNA to provide protection against a single invading element. While adding more and more spacers to CRISPR arrays will have the benefit of recognizing many invaders, the tradeoff is that long CRISPR arrays will dilute the number of Cascade complexes loaded with a particular crRNA, potentially decreasing the CRISPR response against that target. These cellular challenges raise the question how Cascade can navigate the crowded cell sufficiently fast to find DNA targets, and how many copies of Cascade are required to do so.

Here, we report the visualization of single-molecule Type I-E Cascade complexes in a native *E. coli* CRISPR-Cas system *in vivo*. We found that the probability of successful CRISPR protection depends exponentially on Cascade copy numbers, which leads to a time-driven arms race model



between Cascade target search and invader replication. The localization of Cascade shows the complex is enriched inside the nucleoid. We determined that 60% of the Cas8e subunit is incorporated into Cascade complexes and that Cascade DNA probing interactions are very rapid ( $\sim 30$  ms) and are driven by Cas8e. Furthermore, transcription of targets and CRISPR arrays reduce the number of functional complexes in the cell. Our work sheds light on target search and dynamical assembly of Cascade complexes in their native cellular environment, and describes how these processes impact CRISPR protection levels.

## 4.2 Results

### 4.2.1 Visualizing Cascade abundance and target search at the single-molecule level

To investigate how microbes deal with these challenges at the cellular level we used intracellular single-particle tracking Photo-Activated Localization Microscopy (sptPALM)<sup>14,15</sup>, a technique capable of following the movement and abundance of individual fluorescently-tagged proteins in cells with high precision. By genetically fusing a photoactivatable fluorescent protein (PAmCherry2,<sup>16</sup>) to the N-terminus of Cascade-subunit Cas8e (**Fig. 4.1a**), which was the only subunit for which labeling had no influence on the CRISPR interference ability of this strain (**Fig. 4.1b**), we were able to monitor the mobility and abundance of Cascade complexes in *E. coli* cells.

### 4.2.2 Twenty Cascade complexes provide 50% CRISPR protection

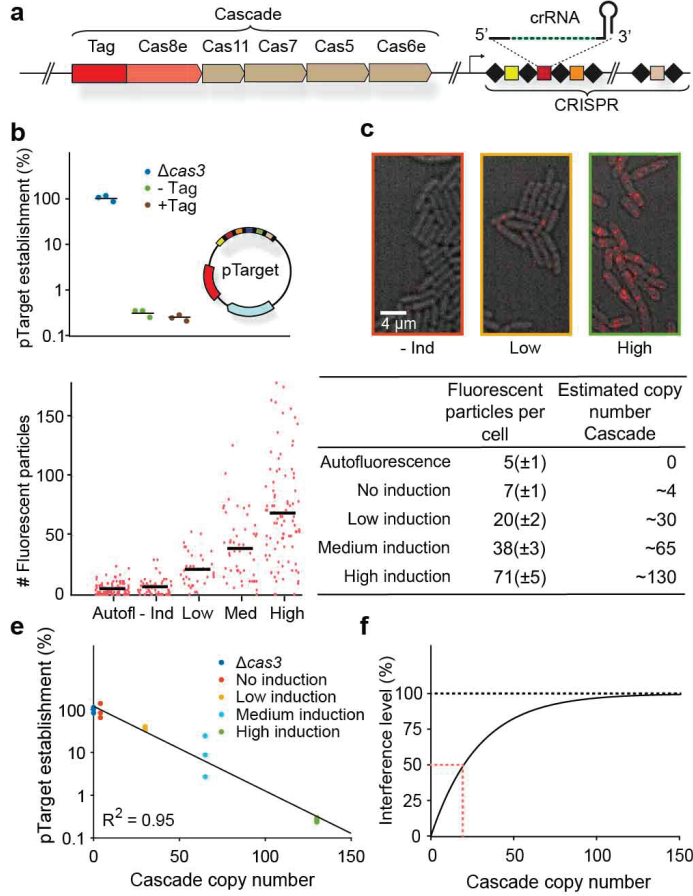
We first wanted to link the copy number of Cascade to successful target search, and established an assay that measures the level of CRISPR protection in cells at the time of cell entry by a mobile genetic element (MGE). In this assay all Cascade complexes present in the cell must be able to target the incoming MGE and Cascade target search has to be rate limiting. To meet the first requirement, we constructed a high copy plasmid (pTarget; **Fig. 4.1b**) containing target sites for all 18 spacers found in the genomic arrays of *E. coli* K12, such that all Cascade complexes would be targeting the incoming plasmid. Secondly, we ensured that Cascade copy numbers

were rate limiting <sup>17</sup> by equipping cells with a low copy plasmid expressing the nuclease Cas3 (pCas3, adapted from <sup>18</sup>).

We achieved different expression levels of Cascade in the cell by tuning the expression of the native regulator LeuO <sup>18</sup> (**Fig. 4.1c**). The copy numbers of Cascade under these varying levels of LeuO induction were estimated from the number of fluorescent particles present in the cell, taking complex assembly (see following section), growth rate (Table S1) and maturation time of PAmCherry into account (**Fig. 4.1d; Methods**). We found that the average number of Cascade complexes per cell in the absence of LeuO induction was low (~4 copies) and that copy numbers increased more than 30-fold for the highest induction level (~130 copies). We measured the interference ability under these conditions by determining the probability that pTarget becomes established in a cell. We observed that establishment of pTarget decreases sharply with increasing copy numbers of Cascade (**Fig. 4.1e**). However, even with 130 Cascade complexes present, we still observed a level of pTarget survival (~0.5%).

To explain these observations, we modelled the probability that an invading MGE becomes established in the cell depending on the number of Cascade complexes that target this specific MGE. The model is based on multi-copy plasmids and phage systems, where the DNA clearance is most likely to occur when an invader enters as a single copy, as the concentration of invading DNA increases over time. Therefore, depending on the invader and the level of CRISPR interference, there will be a critical time point ( $t_c$ ) beyond which the invader is permanently established inside the cell and can no longer be cleared <sup>19</sup>. Our model describes the probability that it takes a certain copy number of proteins ( $n$ ) each with an average search time ( $\bar{t}_s$ ) to find the target before  $t_c$  is reached.

Our model accurately predicted that pTarget establishment decreases exponentially with increasing copy numbers of Cascade (**Fig. 4.1e, Methods**). When we translated these establishment probabilities into interference levels, we could deduce that around 20 Cascade complexes are required to reach a CRISPR interference level of 50% (**Fig. 4.1f**). The exponential relationship further entails every subsequent 20 complexes halve the number of cells not able to achieve interference, which means that 40 Cascade complexes can provide 75% interference; 60 Cascade complexes 87.5%.



**Figure 4.1: Cascade copy number vs CRISPR protection.** **a**, Chromosomal locus of the Cascade subunits and integration site of the photoactivatable fluorescent protein upstream of *cas8e*. **b**, pTarget establishment, calculated from the ratio of transformation of pTarget/pGFPuv, is a measure for the interference level of the CRISPR system. To test whether tagged Cascade complexes were able to function normally, we compared the tagged strain to the untagged and the  $\Delta cas3$  strain. pTarget (bottom right) contains protospacers for all spacers in the K12 genome (colored, not all depicted) and are flanked by a 5'-CTT-3' PAM (black bars). **c**, Overlay of brightfield image of cells (grey) and single molecule signal (red) from a single representative frame for different induction levels. **d**, Number of fluorescent particles measured in each cell plotted for different levels of Cascade expression (left). The mean number of fluorescent particles ( $\pm$  standard deviation; table left column) was converted to a Cascade copy number (table right column, Methods). **e**, pTarget establishment plotted for different copy numbers of Cascade. The data points were fitted with an exponential decay function. pTarget establishment =  $e^{-an}$ , where  $n$  equals Cascade copy number and  $a$  the fitted coefficient. In our model  $a = \bar{\tau}_s / \tau_c$ . **f**, The fitted exponential decay on the left converted into an interference level (Interference level =  $1 - \text{pTarget establishment}$ ). Indicated in red (dashed) is the amount of Cascade copies required for 50% interference.

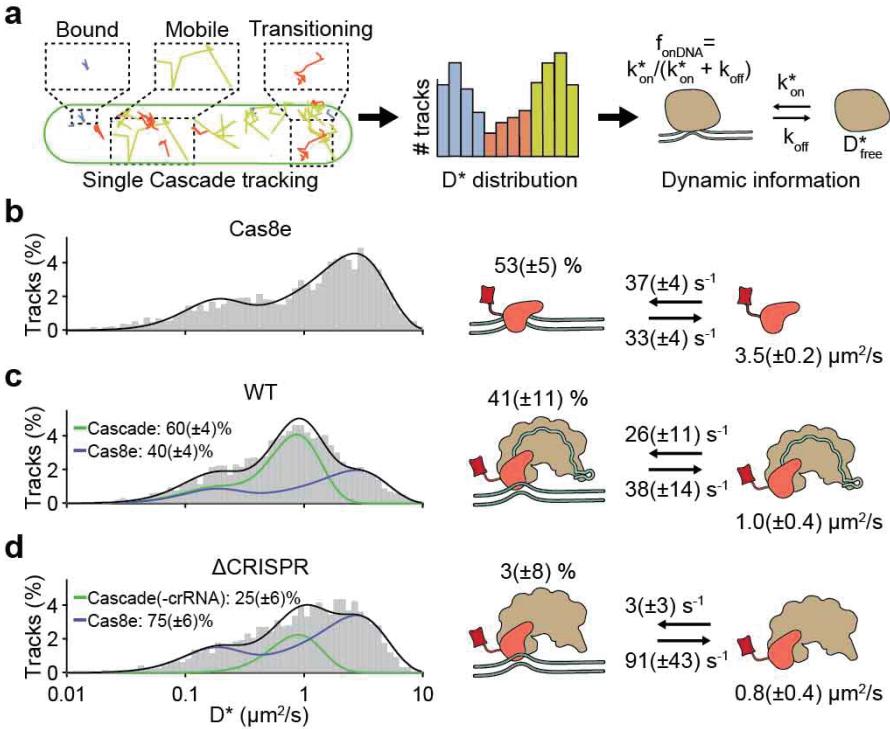
It becomes very unlikely for the CRISPR system to destroy multiple genetic copies of the MGE if it has failed to destroy the single copy that was present at the start before replication. Therefore, we can approximate  $t_c$ , with the replication time of the plasmid in the absence of copy number control (~3 min, <sup>20</sup>), which allows us to retrieve an estimated search time of ~90 minutes for one Cascade complex to find a single target in the cell (**Methods**). In contrast to pTarget establishment, which decreases exponentially, the average search time decreases linearly with increasing copy numbers of Cascade. Therefore 10 Cascade complexes require approximately 9 minutes to find a single target, while 90 Cascade complexes could achieve this within a minute.

To summarize, we found a direct relation between the number of Cascade complexes and the establishment probability of an MGE. The native *E. coli* system requires 20 Cascade complexes loaded with a cognate crRNA to obtain 50% CRISPR interference levels. This relation depends on the replication rate of the invading MGE and the average search time of a single complex and demonstrates the importance of rapid target search on CRISPR interference ability.

#### 4.2.3 The majority of Cas8e assembles into the Cascade complex

To quantify the dynamics of target search, we traced the diffusion paths of thousands of individual complexes in the bacterial cell (**Fig. 4.2a**). The apparent diffusion coefficient  $D^*$ , a measure for mobility, of Cascade was calculated by extracting the displacement of each fluorescent particle for four consecutive 10 ms steps, allowing us to investigate the abundance, mobility and behavior of individual complexes and subunits in the cell. To minimize the influence of spurious autofluorescent particles in *E. coli* <sup>21</sup>, we used expression levels with the highest estimated Cascade copy numbers (~130 copies, high induction; **Fig. 4.1d**).

To distinguish diffusion of Cascade complexes from monomeric Cas8e subunits, we first measured the diffusion of the tagged Cas8e fusion protein in a strain lacking genes of the other four Cascade subunits in the genome (Cas11, Cas7, Cas5, and Cas6e). Based on the role of Cas8e in non-specific DNA binding <sup>22–24</sup>, we expected to find mobile and DNA-bound populations of Cas8e.



**Figure 4.2: Diffusion behavior of Cas8e and Cascade.** **a**, Tracks with small (blue), intermediate (orange) and large (yellow) displacements from a single cell of the WT strain (left). The most likely state for three tracks is indicated, although, due to limited track length and fast transitions, states cannot be assigned confidently to every individual track. The  $D^*$  distribution (middle), from a large population of tracks, enables reliable extraction of DNA interaction kinetic parameters (pseudo-first order on-rate ( $k_{\text{on}}^*$ ), off-rate ( $k_{\text{off}}^*$ ) and the apparent free diffusion coefficient ( $D_{\text{free}}^*$ )) by using analytical diffusion distribution analysis (DDA; right). These parameters further allow the calculation of the fraction DNA bound ( $f_{\text{onDNA}}$ ). **b-d**,  $D^*$  distributions for **b**, Cas8e, **c**, Cascade and **d**, ΔCRISPR strain. Total (black), Cas8e (blue) and Cascade (green) fractions fits are indicated by lines. Parameters (right) of Cas8e (b) were used to fit the Cas8e fraction in Cascade (c-d). Error estimation is based on bootstrapping (± standard deviation). See also Figure S1, S2 and S3.

However, we were unable to describe the data accurately by static two-state models of non-interconverting fractions (**Supplementary Fig. 4.1**). We therefore hypothesized that rapid DNA binding and unbinding events of Cascade on a timescale similar to the framerate (~10-40 ms) would lead to time-averaging of a mobile state (high  $D^*$  values) and a DNA-bound state (low  $D^*$  values), giving rise to intermediate  $D^*$  values (**Fig. 4.2a**). We accounted for these events by developing a generally applicable analysis method called analytical Diffusion Distribution Analysis (analytical DDA), which is useful for proteins with fast transitioning kinetics between states with

different diffusion coefficients, such as DNA-interacting proteins. The distribution of  $D^*$  values is not only affected by the fraction of the time spent bound and freely diffusing, but furthermore changes depending on the absolute transition rates (**Supplementary Fig. 4.2**). Therefore this method allows us to extract quantitative information on DNA binding kinetics and enables the study of fast transition rates previously inaccessible to sptPALM (**Methods**).

When we applied the analytical DDA on the Cas8e diffusional data, we retrieved an average residence time of ~30 ms on DNA and a similar average time spent (~30 ms) rapidly diffusing ( $D^* \sim 3.5 \mu\text{m}^2/\text{s}$ , as expected for a protein of 82 kDa; **Methods**), indicating that Cas8e is bound to DNA for ~50% of the time (**Fig. 4.2b**). The  $D^*$  distribution of Cas8e then allowed us to extract the diffusion behavior of the Cascade complex as a whole. We estimated the fraction of free Cas8e and Cascade-containing Cas8e at 40% and 60%, respectively (**Fig. 4.2c**). This finding suggests that Cas8e is produced in excess<sup>18</sup> or somehow involved in a dynamic interaction with the core Cascade subunits (crRNA, Cas11, Cas7, Cas5, Cas6e)<sup>22,23</sup>.

Surprisingly, we found that the DNA binding kinetics of Cascade were similar to Cas8e alone, indicating that Cas8e is an important driver of DNA probing characteristics of the Cascade complex. Furthermore, the DNA residence times are on average ~30 ms and are thereby considerably shorter than the 0.1-10 s that have been reported for *in vitro* studies previously<sup>7,24,25</sup>. As expected, we found a smaller diffusion coefficient for unbound Cascade complexes (~1.0  $\mu\text{m}^2/\text{s}$ ) (**Methods**) due to their larger size. Together, our analysis shows that more than half of the Cas8e proteins are part of intact Cascade complexes, and that the DNA interacting behavior of Cascade is largely determined by the properties of Cas8e.

The probing kinetics that we measured determine the number of sites Cascade can scan every minute. The total time Cascade needs to probe a single site includes the average time the complex is bound to a DNA site and the average time it requires to find the next DNA site. The Cascade probing time *in vivo* sums up to roughly 60 ms ( $1/k_{\text{off}} + 1/k_{\text{on}}^*$ ), which implies that the complex is able to scan approximately 1000 DNA sites per minute. The probing kinetics of single sites are furthermore linked to the distributions of target search times, and with simulations we could verify that our model of Cascade DNA scanning indeed leads to the expected distribution of interference levels (**Supplementary Fig. 4.3**). Using our previous estimate of the overall target search time for a single Cascade of ~ 90 min, we

calculate that the complex scans 90.000 DNA sites in the cell before finding a target (**Methods**).

To investigate the role of crRNAs in Cascade complex assembly, we deleted all CRISPR arrays in the K12 genome ( $\Delta$ CRISPR). The resulting diffusion behavior can be described by fractions of free Cas8e and with Cascade-like diffusion behavior (**Fig. 4.2d**) that almost entirely lacks interaction with DNA ( $f_{\text{onDNA}} = 3\%$ ). This indicates that although Cascade (sub)complex formation does not strictly require the presence of crRNA <sup>26,27</sup>, Cascade assembly is greatly enhanced by crRNA. Taken together, the majority of Cas8e proteins are incorporated in Cascade complexes in the presence of crRNA, and this gives Cascade DNA interacting properties.

#### 4.2.4 Cascade is enriched but not exclusively present in the nucleoid

Not all potential DNA interaction sites in the host chromosome might be accessible to Cascade. The host DNA is concentrated in the middle of the cell in the nucleoid and is very compact which excludes large complexes such as ribosomes <sup>28</sup>. Nucleoid exclusion would reduce the amount of DNA available for scanning and increase the amount of freely diffusing Cascade complexes. To investigate whether the DNA-bound fraction is governed by affinity properties of Cascade for DNA rather than a restricted search space outside the DNA-containing nucleoid region, we studied the spatial distribution of Cascade localizations. Nucleoid-excluded ribosomes are enriched away from the central long axis of the cell <sup>29</sup>. For Cascade, we found a homogeneous spatial distribution throughout the cell (**Fig. 4.3a**), indicating that Cascade is small enough to freely scan the nucleoid for target sites.

We furthermore used the spatial distribution of Cascade to extract quantitative information on the DNA-bound fraction. To that purpose, we created a DNA-free environment in the cell by adding cephalixin <sup>30</sup>. This antibiotic affects cell wall synthesis and causes cells to elongate, forming DNA-free cytoplasmic space between nucleoids without condensing the nucleoid (**Fig. 4.3b**). The time Cascade is bound to DNA is inherently linked to the relative amount it spends in DNA-free and DNA containing regions. Therefore, by calculating the relative amount of localizations in both regions (Enrichment Factor; *EF*) we can extract the fraction of time spent on DNA independently from the DDA analysis. Cascade was only moderately enriched (*EF* of  $1.8 \pm 0.2$  fold) in the nucleoid regions (**Fig. 4.3c**), indicating that Cascade spends a considerable amount of time diffusing in the

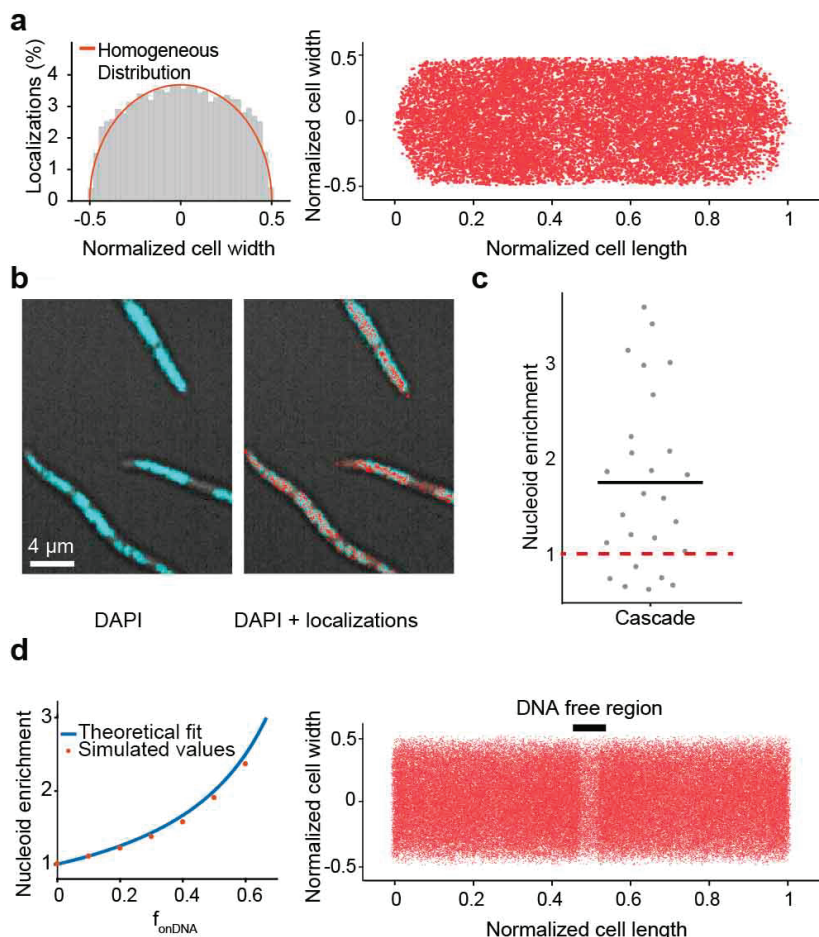


cytoplasm while not associated with DNA. From the enrichment factor, the fraction of Cascade complexes bound to DNA can be approximated to 45% (**Fig. 4.3d**; for derivation see **Methods**). This value is consistent with the ~50% value we extracted from the DDA distribution of Cascade (**Fig. 4.2c**). However, it strongly contrasts other DNA binding proteins such as Fis and RNA polymerase, which show a much higher nucleoid enrichment<sup>30,31</sup>. The above findings indicate that Cascade inherently spends more time freely diffusing the cell and that this is caused by the nature of DNA-Cascade interactions and not by size-based nucleoid exclusion, as is the case for ribosomes<sup>29</sup>. Therefore, we decided to study the nature of the DNA interactions in more depth.

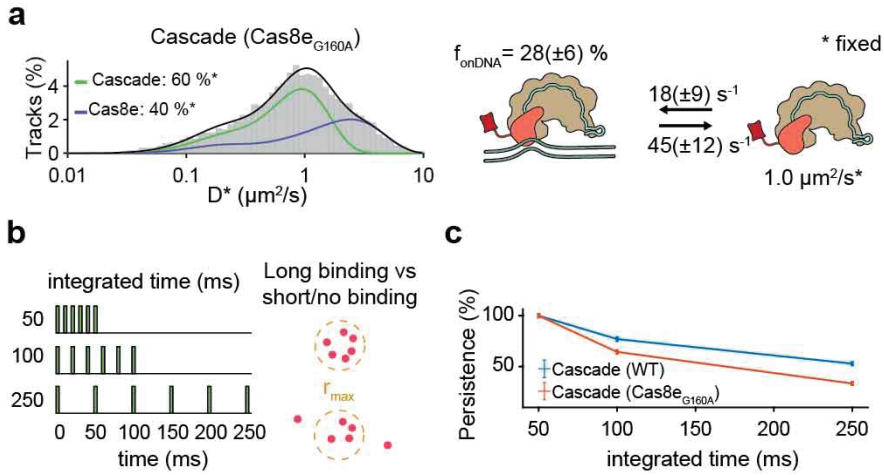
#### 4.2.5 Cascade-DNA interactions are not only PAM-dependent

Next, we assessed how PAM interactions contributed to DNA binding by introducing mutation G160A in the Cas8e subunit which abolishes the interaction with the PAM<sup>32</sup>. This G160A mutation decreased the fraction of DNA-bound Cascade from  $41 \pm 11$  to  $28 \pm 6\%$  (**Fig. 4.4a**) without fully inhibiting DNA binding, suggesting that PAM-independent interactions<sup>32–34</sup> play a role in DNA probing as well. To assess the contribution of these different types of interactions to the average DNA residence time found previously, we measured the persistence of Cascade-DNA interactions by increasing the dark time between exposures (**Fig. 4.4b**). Our data showed that sustained binding events at longer time scales (100 – 250 ms) were more frequently observed for WT Cascade than for the PAM binding mutant complex Cascade-Cas8e<sub>G160A</sub> (**Fig. 4.4c**). Together with the increased off-rate of the mutated complex (**Fig. 4.4a**), this finding demonstrates that PAM-dependent interactions of Cascade with DNA last longer than PAM-independent interactions.





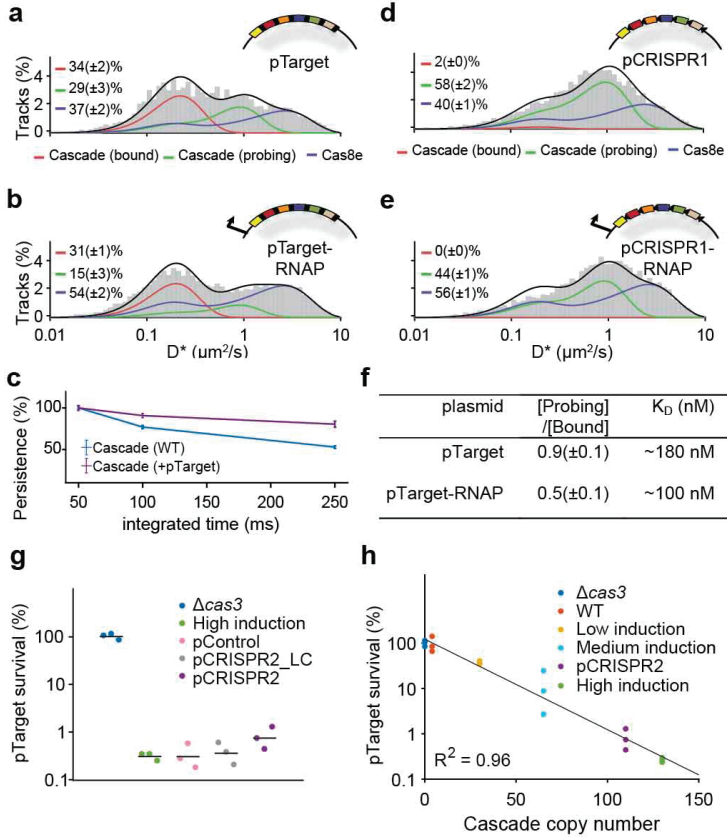
**Figure 4.3: Cascade localization inside the cell.** **a**, Localization of Cascade in the cell. Left: Distribution of Cascade over the cell width ( $n = 33$  cells; 15428 localizations): in orange is indicated the expected distribution in case of a homogeneous localization within the cell. Right: same localizations plotted within dimensions of single cell in which the cell length and cell width of each cell was normalized. **b**, Overlay of DAPI fluorescence and brightfield image (left) with Cascade localizations (right) in cephalixin treated cells. **c**, The nucleoid enrichment in the WT strain (27 subregions in 18 cells). The average ratio is indicated with a black bar. The expected ratio if Cascade has no interaction with DNA is indicated in red (dashed). **d**, Relation between DNA bound fraction and nucleoid enrichment. Left: A theoretical relation between nucleoid enrichment and DNA bound fraction was derived (Methods) and compared to simulated values for different amounts of  $f_{\text{onDNA}}$ . Right: Localizations of simulated Cascade proteins ( $n = 50.000$ ) diffusing through part of an elongated cell are plotted on top of long cell axis. A DNA-free region (black bar) is visible due to enrichment of Cascade binding to DNA in nucleoid regions. Simulations of particles were performed with off-rate of  $38 \text{ s}^{-1}$  and an on-rate of  $26 \text{ s}^{-1}$  to reach a nucleoid enrichment of 1.8, similar to the average that was found for Cascade.



**Figure 4.4: PAM-dependent and PAM-independent DNA probing.** **a**,  $D^*$  distributions for Cascade and Cas8e with a mutation (G160A) deficient in PAM binding. To compare kinetic rates, we assumed that the relative Cas8e-Cascade fractions and the diffusion of free Cascade and Cas8e were not altered by the mutation and those values were fixed. **b**, Depiction of persistence analysis. Increasing the integration time while keeping exposure time constant and counting the number of localizations within a certain radius allow the calculation of the persistence of binding events. **c**, The relative amount of long binding events (6 consecutive localizations within  $r_{\text{max}}$ : 1 pixel (0.128  $\mu\text{m}$ ) of the mean position) for WT and PAM binding mutant Cascade normalized to 50 ms integration time. Error estimation in (a) and (c) is based on bootstrapping ( $\pm$  standard deviation).

#### 4.2.6 Target DNA binding is influenced by the cellular environment

After establishing intrinsic DNA probing characteristics of Cascade, we next investigated its diffusion behavior in the presence of targets (**Fig. 4.5**). To prevent target DNA degradation by Cas3 nucleases, we deleted the *cas3* gene and verified that the deletion did not alter Cascade diffusion behavior (**Supplementary Fig. 4.4**). To verify that all Cascade complexes could bind a target, we measured the copy number of pTarget to be  $\sim 400$  copies/cell (**Supplementary Fig. 4.5**). As the native *E. coli* CRISPR arrays contain 18 spacers, this resulted in  $\sim 7000$  target sites per cell which far outnumbers Cascade copy numbers under our growth conditions ( $\sim 130$ , **Fig. 4.1d**).



**Figure 4.5: Cascade - DNA interactions in the presence of targets.** (a-b),  $D^*$  distribution for the  $\Delta cas3$  strain carrying pTarget (a) and pTarget-RNAP (b). pTarget contains protospacers for all spacers in the K12 genome (colored, not all depicted) and are flanked by a 5'-CTT-3' PAM (black bars). Cascade (probing) (green) and Cas8e (blue) fractions were fitted with parameters from Figure 1C and 1D, and a new target-bound fraction (Cascade (bound)) was introduced as a single diffusion state ( $D^* = 0.06 \mu m^2/s$  ( $+\sigma^2/t$ ); red). **c**, The abundance of sustained binding events as in Figure 3C, but for WT and pTarget-carrying cells. (**d-e**),  $D^*$  distribution for the  $\Delta cas3$  strain carrying pCRISPR1 (d) and pCRISPR1-RNAP (e). pCRISPR1 contains the same protospacers as pTarget that are now flanked by repeat PAMs. **f**, *In vivo*  $K_D$  estimates based on the ratio between Probing/Bound Cascade and the plasmid copy number (Figure S5; Methods). **g**, pTarget establishment for  $\Delta cas3$  (blue), WT (high induction; green), an empty high copy plasmid (pControl; pink), and low or high copy plasmids carrying CRISPR arrays (pCRISPR2\_LC/pCRISPR2; grey/purple). Each dot represents an independent biological replicate. **h**, pTarget establishment plotted for different copy numbers of Cascade. Same as Figure 1E but with addition of pCRISPR2. The Cascade copy number of the pCRISPR2 strain was estimated from the relative abundance of the Cascade (probing) fraction in the WT (high induction; Figure 2c) and pCRISPR2 (Figure S4) strain. Each dot represents an independent biological replicate. Error estimation in (a-f) is based on bootstrapping ( $\pm$  standard deviation). See also Figure S4, S5 and S6.

Compared to a non-targeted control plasmid (**Supplementary Fig. 4.4**), the introduction of pTarget in cells decreased the fraction of free Cascade complexes (from  $60 \pm 4$  to  $29 \pm 3\%$ ), and gave rise to a  $34 \pm 2\%$  immobile, target-bound Cascade fraction ( $D_{\text{Cascade}(\text{bound})}^* = 0.06 \mu\text{m}^2/\text{s}$ ) (Figure 5A). As expected, addition of pTarget increased the persistence of sustained binding events, indicating specific DNA target binding (**Fig. 4.5c**). The combined information of plasmid copy number and the ratio of probing to target bound Cascade enabled us to determine a cellular  $K_D$  value for the affinity of Cascade for targets of  $\sim 180$  nM (**Fig. 4.5f; Methods**), indicating that the affinity *in vivo* is around 10 times lower than what has been observed *in vitro*<sup>32</sup>.

We hypothesized that transcription of DNA along target sites would be one of the main factors influencing Cascade target DNA binding. To investigate the effects of transcription by host RNA polymerase (RNAP), we introduced a (lac) promoter in front of the pTarget sequence. To our surprise, we observed that the affinity of Cascade for target sites that undergo transcription ( $\sim 100$  nM) was higher than for non-transcribed target sites ( $\sim 180$  nM; probing/target bound Cascade from  $0.5 \pm 0.1$  to  $0.9 \pm 0.1$ ). In addition, we observed an increased fraction of free Cas8e subunits (from  $37 \pm 2\%$  to  $54 \pm 2\%$ ) in the strain containing transcribed pTarget (**Fig. 4.5b**). Collectively, these findings suggest that transcription of a target DNA sequence somehow facilitates target search and increases the affinity of a target. In addition, it appears that collisions of RNAP with target-bound Cascade result in changes in the Cascade assembly, likely by dissociation of the Cas8e subunit from the complex upon collision with RNA polymerase, which potentially dissociates Cascade from the target. The relatively dynamic association of Cas8e within the Cascade complex has been observed previously *in vitro*<sup>23</sup> and was more recently also observed upon binding to the CRISPR array<sup>35</sup>. We hypothesized that this dynamic behavior might be a functional characteristic and will also occur upon encountering CRISPR arrays inside the cell. To test this hypothesis, we made a variant of pTarget where all 18 interference PAMs were replaced by the trinucleotide sequence matching the repeats of the CRISPR array (pCRISPR1). Cascade did not show any interaction with the non-transcribed pCRISPR1 plasmid (**Fig. 4.5d**). However, when we added a promoter sequence in front of the pCRISPR1 array of targets, we observed moderately enhanced levels of free Cas8e (from  $40 \pm 1$  to  $56 \pm 1\%$ ) (**Fig. 4.5e**), reminiscent of Cas8e expulsion from the complex upon collision with RNA polymerase, or from targets with repeat like PAMs<sup>35</sup>. Effectively this shows that transcribed CRISPR arrays may function as target decoys in the cell and can therefore potentially influence the levels of functional Cascade complexes in the cell.

To test whether CRISPR array really form decoys in the cell and could impact interference levels, we constructed a compatible high copy number plasmid pCRISPR2 containing a normal CRISPR array (**Supplementary Fig. 4.6**). While the introduction of pCRISPR2 into cells containing pTarget only led to a small decrease in the number of Cascade complexes (15% less) (**Supplementary Fig. 4.4**), the CRISPR interference levels were reduced by as much as 50% (**Fig. 4.5g**). This effect was not observed with low copy variant of pCRISPR2 (pCRISPR2\_LC) or with a high copy plasmid lacking CRISPR arrays (pControl), indicating that this effect comes from the presence of a large number of CRISPR arrays in the cell (**Fig. 4.5g**). We further found that the observed impact of CRISPR arrays on Cascade copy number and interference level fits well with our previously predicted relation between Cascade copy numbers and probability of successful MGE establishment (**Fig. 4.5h**). It furthermore demonstrates how relatively small changes in Cascade copy numbers (15%) can have a big impact on CRISPR interference levels (50%). Taken together, our data indicate that Cascade target search and binding is strongly influenced by the action of RNA polymerase and that CRISPR arrays form target decoys in the cell, which can affect CRISPR interference levels.

### 4.3 Discussion

How crRNA-effector complexes can achieve timely detection of incoming mobile genetic elements in the crowded environment of the cell is an intriguing aspect of CRISPR biology that remains poorly understood. We provide first insights into the fundamental kinetics of the surveillance behavior of type I crRNA-effector complexes in their native cellular environment. We determined how many copies of Cascade are required to establish effective immunity and uncovered how Cascade complexes navigate the crowded bacterial cell packed with DNA. Our results indicate that Cascade does not restrict its search space to parts of the cell, for example the nucleoid-free periphery, but instead is occupied scanning the entire host nucleoid for a match. Depending on genome size of a microbe and the number of copies of the genome in the cell, the nucleoid size may vary widely. To cover this vast sequence space sufficiently fast, the Cascade complex interrogates DNA sequences by using a combination of PAM-dependent and PAM-independent interactions which on average last only 30 ms. This probing interaction is much faster than previously reported interaction times determined of type I Cascade complexes by *in vitro* methods, which range between 0.1 and 10 s<sup>7,24,25</sup>. The ability to rapidly

probe DNA sequences for potential matches with the crRNA, and to move from one place in the nucleoid to the next, may explain how a relatively low number of Cascade complexes in *E. coli* may still confer CRISPR immunity. Interestingly, the average probing time of 30 ms for Cascade matches values found for *Streptococcus pyogenes* dCas9 in *E. coli*<sup>5,36</sup>, suggesting that DNA probing interactions of crRNA-effector complexes from both Class I and II systems may have evolved independently to take place at this time scale.

The probing kinetics we measured for Cascade will allow the complex to scan 1000 DNA sites per minute. Given the abundance of PAMs in the host DNA, this interaction time would lead to a search time in the order of hours. This value matches our independently calculated estimate of 1.5 hours for a single Cascade to find a single DNA target in the cell, which is four times faster than dCas9 search time estimates of 6 hours<sup>5</sup>. However, our data also indicates that Cascade not only probes PAMs, the complex also spends a considerable amount of time engaged in PAM-independent DNA interactions. These might be constituted by direct crRNA – DNA interactions<sup>37,38</sup>, or electrostatic interactions of Cascade with the DNA<sup>33,39</sup>. This suggests an even larger DNA sequence space needs to be covered, creating the need for even more efficient and functionally flexible surveillance solutions. This more flexible probing behavior would be required to recognize targets with mutations in the PAM or protospacer in order to trigger a CRISPR memory update pathway called priming<sup>40,41</sup>, which appears to be unique for type I CRISPR-Cas systems.

One possibility to reconcile Cascade DNA probing characteristics to the overall search time could be that Cascade undergoes facilitated 1D DNA sliding, where Cascade probes multiple sites per DNA binding event. We have shown that Cascade spends 50% of its search time on DNA, and the other 50% diffusing to a new site in the cytoplasm. This value may seem low compared to other DNA interacting proteins such as transcription factor LacI, which is DNA bound for 90% of the time (Elf et al., 2007). However, 50% has been theoretically derived as the optimum for a target search process involving one-dimensional DNA sliding and 3D translocation/hopping (Slutsky and Mirny, 2004). Indeed, recently it has been shown *in vitro* that Cascade and Cas9 can slide along the DNA in search of targets<sup>9,24</sup>. If this also occurs *in vivo*, this would be a striking example of a DNA binding protein having an optimized time division between DNA-bound and freely mobile states to survey the DNA content of the cell.

The relatively high abundance (50%) of freely diffusing Cascade complexes may have benefits as well, as this will lead to more Cascade complexes in



the periphery of the cell outside of the nucleoid. By surveying these peripheral regions more frequently, Cascade may be able to detect incoming bacteriophage or plasmid DNA more rapidly when these genetic elements enter the cell.

Besides the chromosomal host DNA, other cellular constituents also affect target DNA binding properties. We found a much higher  $K_D$  value *in vivo* (180 nM) than was reported earlier using *in vitro* methods (20 nM)<sup>32</sup>. The discrepancy in binding affinity between *in vivo* and *in vitro* measurements may be caused by an increase in target search time (i.e. a lower on-rate) or an increase in target dissociation rate (i.e. a higher off-rate) *in vivo*. In any scenario, this discrepancy highlights the strong role of the crowded cellular environment on target binding.

Counterintuitively, we have found that Cascade binds transcribed target sites with higher affinity (100 nM) than non-transcribed target sites (180 nM). Previous studies have shown that negative-supercoiling is required for Cascade binding<sup>42</sup>, and that increased negative super-coiling accelerates the rate of R-loop formation<sup>43</sup>. As transcribed regions cause more negative supercoiled regions in the DNA<sup>44</sup>, this could explain the increase in the affinity for transcriptionally active sites. Rates of spacer acquisition were also found to be higher for transcriptionally active regions<sup>45</sup>, so together these effects may influence the abundance and effectiveness of spacers in nature.

Next to the positive effect of transcription on target search, we have also found that collisions between RNAP and target-bound Cascade lead to Cascade disassembly, where the Cas8e subunit is expelled from the Cascade core. Furthermore, CRISPR arrays themselves can trigger Cascade disassembly, indicating they form target decoys in the cell. When present at high copy number, CRISPR arrays can even impact CRISPR interference levels (**Fig. 4.5g**). The loose association of Cas8e with the core Cascade complex as observed *in vitro*<sup>23</sup>, might serve a biological role in cells to recycle Cascade from off-targets including the CRISPR array, and may prevent Cas3 recruitment and subsequent self-targeting<sup>46</sup>.

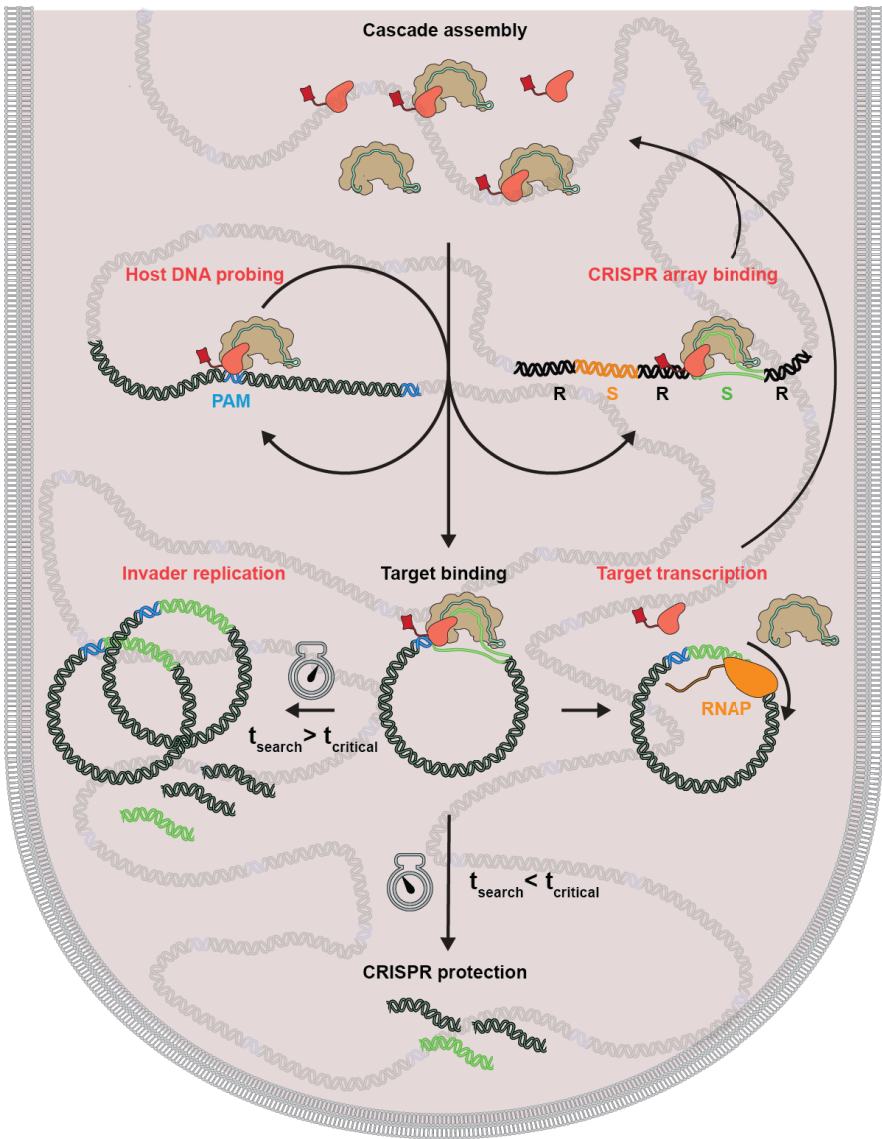
By measuring cellular copy numbers, and accurately measuring CRISPR interference levels, we could uncover an exponential relationship between the number of Cascade complexes in the cell and CRISPR interference. This relationship describes that every 20 Cascade complexes loaded with one crRNA can provide 50% more protection from an invading DNA element (i.e. 20 copies provide 50%, 40 copies 75% protection). Therefore at constant Cas protein production and degradation levels, the effective concentrations

of Cascade complexes loaded with one type of crRNA will become diluted when CRISPR arrays become longer. The size of the CRISPR array is therefore a tradeoff between the higher protection levels of a few spacers, and lower protection levels of many spacers. With our findings we can test optimality of this tradeoff under different conditions and help explain the observed sizes of CRISPR arrays found in nature <sup>47</sup>.

The initial entry is the most vulnerable time for the invader, but invading MGEs have the possibility to outrun CRISPR-Cas immunity by replicating faster than being found. In the native cellular environment, we have found that scanning of host DNA, binding to CRISPR arrays and encountering transcribing RNA polymerases can prevent Cascade from finding the target before the critical time ( $t_c$ ) is reached and the invader is permanently established (**Fig. 4.6**). We therefore hypothesize the presence of a kinetic arms race, in which invaders have evolved to replicate increasingly fast upon cell entry, while CRISPR-systems have evolved to increase the rate at which they are able to find the target. A recent study has indeed shown that the replication rate of foreign elements affects CRISPR interference levels <sup>48</sup>. Many bacteriophages use a two-stage injection <sup>49,50</sup>, which may have evolved to limit the amount of time their DNA is exposed to intracellular defense mechanisms, while already allowing the production of proteins to replicate phage DNA, control host takeover, or to inhibit host defense (e.g. anti-CRISPR proteins) <sup>51</sup>. It has been previously shown that the host can counter this strategy by selectively targeting early injected DNA regions, maximizing the time available to look for targets <sup>52</sup>.

The specificity and kinetics of the CRISPR-system inside the crowded cellular environment is remarkable. Our study has observed very rapid scanning of DNA sites by Cascade complexes and our model predicts the impact of probing kinetics and copy numbers of Cascade on protection levels of CRISPR-Cas systems. We believe that not only specificity and evasion strategies such as anti-CRISPRs but also target search and infection kinetics have played an important role in the evolution of this immune system. The target search equations established here could be expanded to the population level, allowing to model how individual variability in Cascade expression levels and replication rates can impact the survival of entire populations. Therefore, our data provides an important framework for further quantitative cellular studies that will address how CRISPR systems optimally deal with the challenges of cost-effective and rapid target search.





**Figure 6.6: Model of how Cascade protects the cell.** Successful protection against an invader requires Cascade target search to circumvent several potential diversions (red). After Cascade is assembled, the complex probes the host DNA by rapidly binding and dissociating. It uses PAM-dependent and PAM-independent DNA interactions and scans the entire nucleoid region. If it binds to a CRISPR array (S: spacer; R: Repeat), the complex disintegrates. When it has found its target, it depends on the search time ( $t_{\text{search}}$ ) and the critical time ( $t_{\text{critical}}$ ) whether the invader is cleared and the cell protected, or the invader can replicate and establish itself in the cell. Moreover, transcription by RNA polymerase (RNAP) can still remove bound complexes, compromising CRISPR protection.

## 4.4 Acknowledgements

The authors thank Dr. A. Košmrlj (Princeton University) for deriving equation 29 presented in Methods. We acknowledge S. Creutzberg for supplying plasmid pSC020 and M. Siliakus for supplying plasmid pMS011 and all members of the Hohlbein and the Brouns groups for input during group discussions. We thank Jaap Keijsers, Fiona Murphy and Stan van de Wall for providing preliminary measurements and scripts for data analysis. S.B. is supported by the European Research Council (ERC) Stg grant 639707, and by a Vici grant of the Netherlands Organisation for Scientific Research (NWO). R.M. is supported by the Frontiers of Nanoscience (NanoFront) program from NWO/OCW.

## 4.5 Author contributions

S.B. and J.H. conceived and supervised the project; J.V., M.V. R.M., C.A., D.B., B.B. did the experimental work; J.V. and J.H. derived the theory; J.V. and K.M. wrote analysis scripts; J.V., K.M. and J.H. established microscopy workflow, J.V., J.H. and S.B. wrote the manuscript with input from all authors.

## 4.6 Methods

### Cloning

The inserts to create pTarget and pCRISPR1 plasmids were purchased as synthetic constructs from Gen9 (pTarget insert and pCRISPR1 insert; Table S3). To increase the copy number of targets in the cell, the constructs were cloned into a pUC19 backbone with XbaI and KpnI restriction sites, yielding pTarget-RNAP and pCRISPR1-RNAP. The lac promoter was removed for both plasmids by digestion with Sall and PciI, creating blunt ends with Klenow Fragment and subsequently religated to yield pTarget and pCRISPR1. CRISPR arrays were amplified from the K12 BW25113 strain (primers BN383 and BN384; BN370 and BN385 for CRISPR array 2.1 and 2.3 respectively) and cloned into pJPC-12 plasmid containing the pSC101 ori with KpnI and Sall sites (for CRISPR array 2.1) and Sall and EcoRV sites (for CRISPR array 2.3). The copy number of the plasmid could be varied by introducing mutations in the *repA* gene with site-directed mutagenesis PCR (BN373-375). The E96R mutation of RepA yields a reported copy number of

~240/cell (pCRISPR2) compared to the WT RepA (pCRISPR2\_LC) copy numbers of ~7/cell<sup>53</sup>. A plasmid was made from the high copy-variant that did not contain any CRISPR arrays (pControl). All constructs were verified by sequencing.

## Recombination

The strains used in this study were created by using Lambda red recombineering<sup>54</sup>. Strains harbouring the pSC020 plasmid that contains both the Lambda red recombinase and Cre-recombinase were grown at 30 °C. Before transformation of an insert containing an antibiotic resistance marker, the expression of Red recombinase was induced with 0.2% L-Arabinose. Colonies on the specific antibiotic plate were verified with PCR and sequencing and subsequently Cre recombinase expression was induced with 1 mM IPTG at 37 °C to promote plasmid and antibiotic resistance gene loss. The strain was subsequently patch plated to screen for resistance sensitivity due to plasmid loss.

If the scar that is left after lox-site recombination is directly upstream or downstream of a gene it might influence gene transcription/termination. In the design of constructs for *pamcherry2*<sup>16</sup> the lox-cat-lox sequence was placed upstream of the IGR (Intergenic region) that is present between *cas3* and *cas8e*. To allow for correct termination of *cas3*, a part of the IGR was also added at the 5' end of the antibiotic resistance marker. The 3' flank of the constructs overlapped with the *cas8e* gene. The 5' flank of the constructs matched a sequence upstream and downstream of *cas3* (PAMCherry ins; Table S3). Amplification of the constructs with a forward primer matching the downstream region kept *cas3* intact upon insertion (BG7128), whereas a primer matching the upstream region deleted the *cas3* gene allowing measurements in the presence of targets (BG7129). The insert also contained a part of the *cas8e* sequence containing a G160A mutation. This mutation could be introduced into the gene simultaneously with the fluorescent protein, depending on the reverse primer that was used for insert amplification (BG7130 for WT, BG7131 for G160A).

Knockouts of the CRISPR arrays and Cas gene subunits of the K12 strain were made by amplifying a lox-*kan*-lox or lox-cat-lox sequence with flanks matching the specific sequences and introducing them into the strain as described above (BG7366+BG7367 for CRISPR array 2.1; BG7368+BG7369 for CRISPR array 2.2+2.3; BG8366+BG8367 for  $\Delta(cas11-cas6e)$ ). A full overview of the sequences of these inserts is given in Table S3.

## Growth conditions

To prevent the high-copy target plasmids from influencing the growth rate of the strains and therefore changing the fraction of matured PAMCherry complexes we used a rich defined medium with minimal autofluorescence. Strains were grown in M9 minimal medium containing the following supplements: 0.4% glucose, 1x EZ amino acids supplements (M2104 Teknova), 20 µg/ml uracil (Sigma-Aldrich), 1mM MgSO<sub>4</sub> (Sigma-Aldrich) and 0.1 mM CaCl<sub>2</sub> (Sigma-Aldrich) (further referred to as M9 medium). Strains were inoculated o/n from glycerol stocks and 200x diluted in fresh medium the next day. Cells were always grown with the required antibiotics. The expression level of Cascade for strains carrying the pKEDR13 plasmid could be tuned by different expression levels of LeuO. The expression level referred to in the text as low induction was achieved by leaky expression of LeuO (no addition of IPTG), medium induction was achieved by addition of low levels of IPTG (0.01 mM), whereas high induction was achieved by addition of 1 mM IPTG upon dilution of the o/n culture. For all sptPALM measurements the high induction condition was used. The cells were grown for ~2.5 hours to an OD of 0.1 before use. For enforced elongation of cells, cephalixin (40 µg/ml) was added 0.5 hour after fresh inoculation and grown for two more hours. When required, DAPI for staining of DNA was added right before imaging (0.5 mg/ml).

## Transformation assay

Each culture was grown under conditions described above and 30 ml were used to create competent cells. Cells were washed 3 times in ice-cold 10% glycerol solution and the final culture was reduced to 250 µl. The cells were aliquoted and stored at -80 °C. A mixture of pTarget (10 pg/µl) and pGFPuv (10pg/µl) was transformed into 40 µl of culture. In case of strong interference levels, the ratio was adjusted to a 100:1 (pTarget (100 pg/µl):pGFPuv (1 pg/µl)). The transformability of strains was linear in these concentration regimes, allowing these different relative concentrations to be used.

Electroporated cells were immediately plated in two dilutions on plates containing ampicillin (100 µg/ml) and glucose (0.4%). Glucose was added to prevent premature expression of GFPuv which would cause a decrease in fitness of cells containing this plasmid. The next day, 96 colonies from each replicate were reinoculated in 96-wells plate with LB containing ampicillin (100 µg/ml) and IPTG (1 mM). After overnight incubation, the 96 well colonies were analysed in a plate-reader (Synergy H1, Biotek). pTarget establishment was defined as

$$P_{\text{establishment}} = \frac{\# \text{ pTarget colonies [pGFPuv Transformed]}}{\# \text{ GFPuv colonies [pTarget Transformed]}} \quad (1)$$

pTarget establishment was further normalized to the interference level of a  $\Delta\text{cas3}$  strain.

### qPCR

Each culture grew under conditions described above and 2 ml were used to extract the DNA. DNA was isolated with the Genejet Genomic DNA kit (Thermo Scientific) and concentrations were measured with the Qubit dsDNA HS Assay kit (Thermo Scientific). qPCR was performed with primers that have been used before in plasmid copy determination (BG8677-BG8680)<sup>30</sup>. The Ct value of the PCR amplifying the *dxs* gene and the *bla* gene was a measure for the ratio between chromosomal and plasmid DNA. 1 ng of genomic DNA and 0.5  $\mu\text{M}$  of each primer were added to the iTaq<sup>TM</sup> SYBR Green SYBR Green PCR reaction mixture. A standard curve for the amplification efficiency was made by a dilution series of pMS011, a plasmid containing one copy of the *dxs* and the *bla* gene.

### Slide preparation

In order to work with very clean slides, an extensive cleaning procedure was used (modified from<sup>55</sup>). Slides were burned in the oven at 500 °C for two hours, and stored in aluminium foil until the day of usage. Slides were subsequently sonicated in MilliQ, Acetone and KOH, incubated in Piranha Solution (75%  $\text{H}_2\text{SO}_4$ , 7.5%  $\text{H}_2\text{O}_2$ ) and afterwards rinsed with MilliQ. 1% Agarose slabs containing the growth medium were hardened between two cleaned glass slides, spaced slightly apart using parafilm. After hardening, a concentrated culture of cells was added in between the slab and one of the slides. The agarose slab was always prepared within 20 minutes of the measurement to prevent desiccation.

### Microscope set-up

For the acquisition of microscopy data, a home-build TIRF microscope was used, which is described in more detail elsewhere<sup>36</sup>. Briefly, four lasers with different wavelengths (405, 473, 561 and 642 nm) are situated in a Lighthub laser box (Omicron, Germany), and are transformed in a collimated beam via a reflective collimator and an optical fibre. Stroboscopic illumination was used to allow for 2 ms excitation in the temporal middle of the captured 10 ms long frame<sup>56</sup>. The excitation laser is focused on the backfocal plane of a 100x oil immersion SR/HP objective (NA = 1.49, Nikon, Japan), and the emission is captured on a Zyla 4.2 plus sCMOS camera (Andor, UK). 2x2

pixel binning was used, resulting in 128x128 nm pixels. Data acquisition was performed using MicroManager<sup>57</sup>. Measurements were performed at room temperature (21 °C)

### Single-molecule Measurements

The cells were imaged with a brightfield light and 405 and 561 nm lasers. First brightfield images were taken to find contours of the cells. The 405 nm laser was used to stochastically activate PAmCherry and the laser intensity was slowly increased during the measurement up to 10  $\mu$ W. The laser intensities were measured directly after the reflective collimator. With increasing the laser intensity of the 405 nm laser during the measurements, we aimed at keeping the number of activated molecules relatively constant (~1-10 per FOV). The 561 nm laser was used to excite the fluorescent protein tags (40 mW pulses with 2ms pulse width, leading to average exposure intensity of 8 mW).

To measure Cascade localization in cephalixin-treated cells that were stained with DAPI, we took an alternative approach. To prevent DAPI fluorescence from influencing the fluorescence measurements of the single molecules, we briefly activated a subset of particles with the 405 nm laser and subsequently tracked Cascade for a couple of frames with 561 nm excitation, repeatedly doing this, until most fluorescent proteins were photobleached.

### Analysis

#### Detection, localization and tracking

Analysis was done with home-built software, adapted from<sup>58,59</sup>. The sCMOS camera we used has pixel dependent offset, gain and variance, which we took into account to minimize the detection of false positive localisations. We estimated these parameters by measuring 60.000 dark frames and 20.000 homogeneously illuminated frames with increasing levels of intensity<sup>60</sup>. To further optimize our detection, we implemented a temporal median filter (time window 400 frames) for background estimation<sup>61</sup>. The background estimate was not directly subtracted from the image, but photon statistics were incorporated in a likelihood-ratio test that calculated the probability of a scenario with and without an emitter for each pixel in every frame. Briefly, a raw image was first converted into photon counts by using the camera offset and gain maps. Subsequently for every pixel the intensity ( $I_{\text{tot}}$ ) of a potential emitter was estimated by Gaussian-weighted ( $\sigma = 1$  pixel) summation of a 7x7 window to a background subtracted image. Subsequently, potential emitters of more than 50 photons were preselected and were further

subjected to a ratio test. The ratio test uses the probability defined for pixel  $i$  to have a transformed value  $v$  in the 7x7 region around the preselected pixels as previously described<sup>62</sup>:

$$p_{\text{sCMOS}}(v) = \frac{[(d_i - o_i)/g_i + \text{var}_i/g_i^2]|\mu_i, \text{var}_i, g_i, o_i|}{\Gamma(v+1)} = \frac{e^{-(\mu_i + \text{var}_i/g_i^2)}(\mu_i + \text{var}_i/g_i^2)^v}{\Gamma(v+1)} \quad (2)$$

Where  $d_i$  is the raw image value,  $g_i$  is the gain,  $\text{var}_i$  the variance and  $o_i$  the offset for pixel  $i$ . The ratio test calculates the product of the probability of all pixels in the subregion in case of an emitter  $\mu_i = b_i + I_i$ , where  $b_i$  is the estimated background and  $I_i$  is the estimated intensity of the emitter at pixel  $i$  (which was estimated by a Gaussian from the centre of the 7x7 subregion with emitter intensity  $I_{\text{tot}}$ ) divided by the product of the probability of all pixels in the subregion in case of absence of an emitter  $\mu_i = b_i$ .

We set the likelihood to a level that achieved approximately one false positive per frame of 512 x 512 pixels. This method allowed the detection efficiency to be more robust across and between FOVs and independent of manual thresholding for each measurement. Detected particles were subsequently localized with MLE-sCMOS software as previously described<sup>62</sup>.

The localized particles were subsequently linked. Localizations in subsequent frames which were closer to each other than 6 pixels in length (0.78  $\mu\text{m}$ ) were assigned as a track. Particles were allowed to disappear for one frame (due to blinking/moving out of focus), but these steps were not used in the calculation of the apparent diffusion coefficient,  $D^*$ .

### Determination of diffusion coefficients

Several methods were employed to extract diffusion states and their abundances from the analysed tracks. The distribution of the apparent diffusion coefficients can be fitted to an analytical equation as reported earlier<sup>31,63</sup>. These equations depend on the number of steps that is used to generate the average diffusion coefficients of each particle. We used tracks containing a minimum of four steps and only four steps were used in longer tracks.

For a single diffusion coefficient fitting becomes:

$$f_D(x; D, n) = \frac{\left(\frac{n}{D + \sigma^2/dt}\right)^n x^{n-1} e^{-\frac{nx}{D + \sigma^2/dt}}}{(n-1)!} \quad (3)$$

With multiple states this equation becomes:

$$f_D(x; A_i, D_i, n) = \sum_{i=1}^N A_i \frac{\left(\frac{n}{D_i + \sigma^2/dt}\right)^n x^{n-1} e^{-\frac{nx}{D_i + \sigma^2/dt}}}{(n-1)!} \quad (4)$$



Where  $A_i$  are the fractions ( $\sum A_i = 1$ ),  $D_i^*$  are the apparent diffusion coefficients of the different states and  $n$  are the number of steps. The localization error ( $\sigma$ ) was found to be 40 nm, based on the apparent diffusion of the slowest moving fraction in our global data set and similar to other studies using the same fluorescent protein<sup>31,59</sup> or set-up<sup>36</sup>. This equation was fitted to our track distributions with a Maximum Likelihood Estimation algorithm. The uncertainty in the fit was estimated with Bootstrap resampling. The list of  $D^*$  values was resampled 20.000 times with replacement to the size of the original data set. Each resample was then fitted with the same Maximum Likelihood Estimation algorithm.

### Analytical Diffusion Distribution Analysis (DDA)

$D^*$  Distributions have been fitted in numerous studies of DNA binding proteins (see above)<sup>31,63</sup>, making use of distributions developed by Qian *et al.*<sup>64</sup>. The goal is to find the distribution of measured  $D^*$  values ( $x$ ), for a certain number of underlying states that each have a probability  $A_i$  and a diffusion coefficient  $D_i$ . It is derived from repeated convolution of the exponential distribution of displacement, resulting in a gamma function for each state. These distributions assume, however, that there is no transitioning occurring between states.

In order to incorporate dynamics of state transitions into our fitting, we incorporated statistics coming from photon distribution analysis (PDA) that is used for single molecule FRET diffusion coefficient distributions<sup>65–67</sup>. This method, that we term Diffusion Distribution Analysis (DDA), describes the distribution of time spent in each state given a certain  $k_{on}^*$ ,  $k_{off}$  and the integrated time  $t_{int}$ . Here we discuss the analytical way to find this distribution.

Firstly, the probability distribution function for time can be calculated by three equations corresponding to 0, an odd and an even number of transitions<sup>65</sup>:

$$W_{contS1}(t_{S1} = t_{int} | k_{off}, t_{int}) = e^{-k_{off}t_{int}} \quad (5)$$

$$W_{oddS1}(t_{S1} | k_{off}, k_{on}^*, t_{int}) = k_{off} e^{-k_{off}t_{S1} - k_{on}^*t_{S2}} I_0(2\sqrt{k_{off}k_{on}^*t_{S1}t_{S2}}) \quad (6)$$

$$W_{evenS1}(t_{S1} | k_{off}, k_{on}^*, t_{int}) = \sqrt{k_{off}k_{on}^*t_{S1}/t_{S2}} e^{-k_{off}t_{S1} - k_{on}^*t_{S2}} I_1(2\sqrt{k_{off}k_{on}^*t_{S1}t_{S2}}) \quad (7)$$

Where  $t_{S1}$  and  $t_{S2}$  are times spent in state S1 and state S2 and  $I_0$  and  $I_1$  are Bessel functions of order zero and one respectively. Note that  $t_{S1} + t_{S2} = t_{int}$ . Equations for starting in state 2 ( $W_{contS2}$ ,  $W_{oddS2}$  and  $W_{evenS2}$ ), can be found by exchanging  $k_{off}$  for  $k_{on}^*$  and  $t_{S1}$  for  $t_{S2}$  and vice versa in equations 5-7.



We can convert the time spent in the mobile state ( $t_{S2}$ ) to the diffusion coefficient by the following equation:

$$D = \frac{D_{\text{free}} t_{S2}}{t_{\text{int}}} \quad (8)$$

It follows that the probability distribution functions can be converted by:

$$W(D) = W\left(t_{S2} = \frac{Dt_{\text{int}}}{D_{\text{free}}}\right) \quad (9)$$

Furthermore, the chance that the particle at the start is in state 1 or state 2 is provided by:

$$p_{S1} = \frac{k_{\text{on}}^*}{k_{\text{on}}^* + k_{\text{off}}} \quad (10)$$

$$p_{S2} = \frac{k_{\text{off}}}{k_{\text{on}}^* + k_{\text{off}}} \quad (11)$$

To correctly describe the distribution over a certain number of frames, we first calculated the distribution over a single time frame  $t_f$ . Within a single frame, a particle started in that state can either end in the same state or in a different state. Therefore, in a two-state system the probability function for four scenarios have to be calculated:

$$W(D|k_{\text{off}}, k_{\text{on}}^*, t_f)_{S1 \rightarrow S1} = W_{\text{even}S1}(D) + W_{\text{cont}S1} \quad (12)$$

$$W(D|k_{\text{off}}, k_{\text{on}}^*, t_f)_{S1 \rightarrow S2} = W_{\text{odd}S1}(D) \quad (13)$$

$$W(D|k_{\text{off}}, k_{\text{on}}^*, t_f)_{S2 \rightarrow S1} = W_{\text{odd}S2}(D) \quad (14)$$

$$W(D|k_{\text{off}}, k_{\text{on}}^*, t_f)_{S2 \rightarrow S2} = W_{\text{even}S2}(D) + W_{\text{cont}S2} \quad (15)$$

Subsequently the probability to find a certain diffusion coefficient ( $x$ ) for a single time step given the underlying average diffusion coefficient ( $D$ ) is given by  $f_D(x|D, 1)$  (Eq. 3). Then we find the distribution of measured diffusion coefficients for a single frame by:

$$W(x|k_{\text{off}}, k_{\text{on}}^*, t_f)_{Si \rightarrow Sj} = f_D(x|D, 1) W(D|k_{\text{off}}, k_{\text{on}}^*, t_f)_{Si \rightarrow Sj} \quad (16)$$

$i = j = 1, 2$

Now that we have the distribution for a single time step, we need to find the distribution for the average of multiple frames. For this we use the same method as Qian *et al.* <sup>64</sup>, namely repeated convolution of the distribution for a single frame, while keeping track of the start and end state. The probability distributions are therefore:

$$W(x|2t_f)_{S1 \rightarrow S1} = \sum_{i=1,2} (W(x|t_f)_{S1 \rightarrow Si} * W(x|t_f)_{Si \rightarrow S1}) \quad (17)$$

$$W(x|2t_f)_{S1 \rightarrow S2} = \sum_{i=1,2} (W(x|t_f)_{S1 \rightarrow Si} * W(x|t_f)_{Si \rightarrow S2}) \quad (18)$$

$$W(x|2t_f)_{S2 \rightarrow S1} = \sum_{i=1,2} (W(x|t_f)_{S2 \rightarrow Si} * W(x|t_f)_{Si \rightarrow S1}) \quad (19)$$

$$W(x|2t_f)_{S2 \rightarrow S2} = \sum_{i=1,2} (W(x|t_f)_{S2 \rightarrow Si} * W(x|t_f)_{Si \rightarrow S2}) \quad (20)$$

For 4 frames, the distributions found for 2 frames can be convoluted again. The full distribution is then found by summing up each of the partial distributions multiplied by the chance they start in  $S1$  or  $S2$ :

$$W_{tot} = p_{S1}(W(x|4t_f)_{S1 \rightarrow S2} + W(x|4t_f)_{S1 \rightarrow S1}) + p_{S2}(W(x|4t_f)_{S2 \rightarrow S1} + W(x|4t_f)_{S2 \rightarrow S2}) \quad (21)$$

We then have to further correct for the broadening of the distribution of immobile particles where the apparent step size comes from localization error (Figure S2). As localization error, in contrast to diffusion, is correlated<sup>68</sup>, the distribution is not described by a gamma distribution, or any other known exact solution. We find very close agreement with simulations when we subtract the fraction of immobile particles after four time steps ( $W_{contS1}(t_{S1} = 4t_f)$ , Eq.5 ) multiplied with the distribution of expected  $D^*$  for four time steps  $f_D(x|0,4)$  (Eq. 3) and replace it with the same fraction of immobilized particles multiplied with the distribution of expected  $D^*$  for 2.9 time steps  $f_D(x|0,2.9)$ . This value stems from the variance found for correlated MSD values due to localization error<sup>68</sup>.

The assumptions that underlie this model are as follows:

- Each diffusing species can be in two states, namely an immobile and a mobile state.
- The immobile state in our case includes all species bound to chromosomal DNA, including potential 1D sliding events, for which the diffusion is at such a low relative speed that we can consider them as immobile. Our model therefore cannot distinguish between bound and 1D sliding species.
- The immobile state is still perceived as diffusing due to a localization error,  $\sigma$ , which in our case is 40 nm. As the distribution of sequential localization errors differs from sequential diffusion steps we correct for this (Figure S2).
- The mobile state is defined by the parameter  $D_{free}$ , which is the diffusion coefficient of a species in the absence of interactions with DNA. All slowing down in the motion because of transient DNA interaction are captured in our model by the introduction of transitions and do not affect the value found for this parameter.
- The transition between the two states for each species is Markovian, meaning that transition rates are independent of past or future states.

For each species that you fit there are four degrees of freedom, namely the abundance of the species in the total population and the three kinetic parameters  $k_{on}^*$ ,  $k_{off}$ , and  $D_{free}$ . However because the sum of all fractions of species is one ( $\sum c = 1$ ) and the sum of the average time spent freely

diffusing multiplied by the free diffusion coefficient for each species is equal to the average measured Diffusion Coefficient ( $\langle D \rangle = \sum c_i \frac{k_{\text{off},i}}{k_{\text{off},i} + k_{\text{on},i}} D_{\text{free},i}$ ), the amount of free fitting parameters is reduced by two. This means that for a single diffusing species (in our case monomeric Cas8e) we only need to fit two parameters and for a two-species distribution (in our case Cascade) for which one is already known (Cas8e) we need to fit three parameters (8 degrees of freedom – 3 already known Cas8e kinetic parameters – 2 from the above described equations). We found that the uncertainty of our fit, determined by bootstrapping and simulations, is reasonable up to three fitting parameters, therefore we designed our experiments in a way, that in the presence of multiple species (such as pTarget (Fig. 5)) we already predetermined the kinetic parameters for most species to limit the required fitting parameters to three.

### Copy number determination

The copy number of the Cascade complex was determined by generating cell outlines from brightfield images (only well separated cells were chosen). The cell outlines were made with the Oufi software <sup>69</sup>. The total number of tracks that were found in the outlined cells generated a copy number (Figure 1D). Because single localization events can partly stem from false positives, the total amount of tracks was estimated based on the distribution of tracks longer than 1 step and subsequently this distribution was fitted with an exponential to calculate the amount of particles that only had a single localization before bleaching. Similarly, as we know the false positive rate was approximately one per frame, we could also subtract the number of frames from the single step tracks and in this way estimate the total number of tracks. This approach yielded comparable results.

The copy number of proteins in cells are hard to quantify <sup>70</sup>. Currently, protein copy numbers can be estimated either by western blot or by single-molecule fluorescence based methods both of which have specific drawbacks. Although single molecule studies are regarded as the most accurate method, especially at low copy numbers <sup>71</sup>, there are a lot of variables that can lead to over- or underestimation. Underestimation can originate from maturation time of the protein, misfolded/inactivated protein, false negative detections, overlap of PSFs and linking of two separate molecules in a single track. Overestimation can come from failed linking of tracks, false positive detections and blinking fluorescent proteins.

As has been done in previous studies, we take the underestimations stemming from maturation time (23 min for PAmCherry <sup>16</sup>), close to growth rate of 31 min) and estimated *in vivo* folding efficiency (50% <sup>72</sup>) into account

<sup>59</sup>. We also consider that an estimated 40% of the particles we observed come from Cas8e subunits not active complexes. Taken together, the number of particles we observe are subtracted by the amount of estimated autofluorescent particles and subsequently multiplied by a compensation factor of two to reach our estimated copy number values.

We believe that the assumptions made in this study could maximally lead to over- or underestimating our estimated copy numbers by two to three-fold. We note that the relative amounts we observed between the different expression levels will be independent of these assumptions.

### Cascade in DNA-containing/DNA-free regions

To get an independent measure of the total time fraction spent probing DNA, Cascade was visualized in cells that were elongated by addition of cephalixin. The drug cephalixin disabled the ability of the cells to divide, creating elongated cells where nucleoids were separated by DNA-free spaces <sup>30</sup>. Subregions of cell outlines were manually selected and further refined with the Oufi software <sup>69</sup>. The relative amount of localizations of DNA-free and DNA-containing regions was not calculated for entire cells, as differences in illumination intensity between parts of the FOV could also change the amount of localizations detected for different parts of the cell. Each subregion contained one nucleoid free region, flanked by two nucleoid containing regions with a total length of around 4  $\mu\text{m}$ . Segments of 0.1  $\mu\text{m}$  divided along the long axis of the cell are separated into nucleoid or DNA-free segments based on the sum of the DAPI fluorescence within each segment. The average number of localizations of Cascade molecules in nucleoid segments divided by the average number of localizations Cascade molecules in DNA-free segments could be used to infer the DNA bound time fraction (see below,  $f_{\text{onDNA}}$  from nucleoid enrichment).

### Persistence sustained binding events for different integrated times

To estimate how long binding events last, one could plot the number of particles remaining within a certain radius from the first frame position for different number of steps. However, particles can diffuse away when they are released from DNA or be lost due to photobleaching. To account for bleaching rates, previous studies increased dark time between exposures, while keeping exposure times the same <sup>73,74</sup>. This approach uses the data of all time steps, including only single time steps.

As we are investigating lifetime of binding events on a subsecond timescale this approach fails, as single steps of slow-moving particles, which can be clearly separated from bound particles on larger timescales ( $t_{\text{int}} > 1 \text{ s}$ ), will be counted as bound particles leading to overestimated off-rates. At these

timescales, it is more reliable to use tracks of at least 5 steps to distinguish bound from moving particles. As we are interested in how many of these events we observe, depending on the framerate, normalization is required. For this we cannot use the sum of all tracks observed at each framerate, as a larger amount of fast moving molecules diffuse further than the maximum tracking distance of  $0.78 \mu\text{m}$  between two exposures, and are also more affected by confinement with increasing integrated time. Therefore, the number of moving particles of certain track length is not an accurate normalization when comparing different frame times. However, as we used similar exposure for all frame times, the number of detected localizations per protein is unaffected. Furthermore, bound molecules are not affected by confinement or linking errors with increasing frame rates.

The most robust normalization procedure was therefore to normalize the number of localizations within sustained bound tracks (all localizations within 1 pixel of the mean location of the track) to the total number of localizations, as those do not depend on the length of introduced dark time between exposures. A further increase of the dark time was not possible as on longer time scales the movement of the plasmid ( $D_{\text{free}}^* = 0.06 \mu\text{m}^2/\text{s}$ ) made plasmid bound particles diffuse further than 1 pixel.

### Confinement and localization error simulation

To verify whether our new transitional  $D^*$  analysis yielded accurate parameter predictions and investigate the influence of localization error and confinement on the parameters of the fit, we simulated particles moving and transitioning between bound and free moving states within the dimensions of an *E. coli* cell, adapted from methodology used in (34). At every time step particles were simulated to be either in a bound state  $S1$  ( $D = 0 \mu\text{m}^2/\text{s}$ ), or a mobile state  $S2$  ( $D = D_{\text{free}}$ ). At the starting time point, states were assigned to each particle according to the equilibrium probability  $p_{S1}$  and  $p_{S2}$  (Eq. 10 + 11). Subsequently, at following time steps of 0.1 ms, particles in state  $S1$  were assigned to  $S2$  with a probability of  $p_{S1 \rightarrow S2} = k_{\text{off}} t_{\text{step}}$  (where  $t_{\text{step}} = 0.0001 \text{ s}$ ) and particles in state  $S2$  were assigned to  $S1$  with a probability of  $p_{S2 \rightarrow S1} = k_{\text{on}}^* t_{\text{step}}$ . Displacements in three dimensions at each time step were taken from a standard normal distribution multiplied with  $\sqrt{2Dt_{\text{step}}}$  (where  $D$  is either 0 for particles in state  $S1$  or  $D_{\text{free}}$  for particles in state  $S2$ ). Steps beyond the boundaries of a cell were rejected and new displacements were randomly drawn.

The 2D projection of five localizations at 10 ms time intervals for each molecule was generated as output and was analysed in our tracking software. Localization error was included in the simulation by addition of a

random displacement for each position taken from a Gaussian distribution ( $\sigma = 40$  nm). It was found that changes in outcome of the simulation were not sensitive to cell length in the range of our bacteria (3–6  $\mu\text{m}$ ), decreasing less than 5% for the smallest size. Most of the confinement effect is caused by the cell width, which was relatively constant between all the cells measured.

### Cascade nucleoid enrichment simulation

The simulation above was adapted to simulate the movement in DNA-free and DNA-containing regions. Particles were simulated to move inside of a cell of 10  $\mu\text{m}$  in length and 1  $\mu\text{m}$  in width consisting of 100 segments without endcaps (0.1  $\mu\text{m}$  per segment). Five segments were modelled as DNA-free segments and the rest of the segments as DNA-containing segments.

Cascade molecules were randomly placed throughout the cell and subsequently were simulating with similar time steps as described above, except that moving particles were only allowed to transition to S1 (bound state) inside of the nucleoid containing regions. Before recording the position of the simulated particles, the simulation ran for 100,000 time steps (10 s) so that equilibrium was reached. Localization error was added in the same way as described above.

### Expected free diffusion coefficients

The diffusion coefficient of molecules in classic (Newtonian) fluids can generally be estimated by the Stokes-Einstein equation. A study measuring the diffusion of GFP multimers inside the *E. coli* cytoplasm has shown good agreement with the predictions of this equation<sup>75</sup>, whereas a second study found a different relation attributed to the complex nature of the cytoplasmic fluid<sup>76</sup>. To compare our findings of the apparent free diffusion coefficient of Cas8e ( $\sim 3.5 \mu\text{m}^2/\text{s}$ ) and Cascade ( $\sim 1.0 \mu\text{m}^2/\text{s}$ ), we therefore looked for reported free cytoplasmic diffusion coefficient values of proteins of similar size inside *E. coli* cells. For Cas8e, two proteins have been studied with a similar size to PamCherry-Cas8e (82 kDa), namely CFP-CheR-YFP (86 kDa)<sup>77</sup> and TorA-GFP3 (84 kDa)<sup>75</sup>, which have reported values of  $1.7 \mu\text{m}^2/\text{s}$  and  $6 \mu\text{m}^2/\text{s}$ . Our estimate for Cas8e lies within the range of these values. For Cascade (430 kDa), the closest reported protein in size is RNA polymerase, for which the  $D_{\text{free}}^*$  was found to be  $1.1 \mu\text{m}^2/\text{s}$  (400 kDa core enzyme, 470 kDa holoenzyme)<sup>31</sup>. Furthermore larger proteins such  $\beta$ -Gal-GFP<sub>4</sub> (582 kDa;  $0.6 \mu\text{m}^2/\text{s}$ )<sup>78</sup>, and 30S ribosome subunits (900 kDa  $0.4 \mu\text{m}^2/\text{s}$ )<sup>29</sup> were reported with lower diffusion coefficients as expected. These findings support the free apparent diffusion value we found for Cascade ( $\sim 1.0 \mu\text{m}^2/\text{s}$ ).

**$f_{\text{onDNA}}$  from nucleoid enrichment**

The distribution of Cascade in nucleoid-free and nucleoid containing regions depends on the time Cascade spends on DNA. We divided the cell up along the long axis into segments of 100 nm wide. During the time Cascade is bound to DNA it can only be inside of the nucleoid regions whereas, when it is not bound to DNA Cascade can be anywhere within the cell. Therefore, the average number of particles in a DNA-containing segment is given by:

$$\overline{N_{\text{DNA}}} = \left( \frac{f_{\text{onDNA}}}{sm_{\text{DNA}}} + \frac{1-f_{\text{onDNA}}}{sm_{\text{tot}}} \right) N_{\text{tot}} \quad (22)$$

and the average number of particles in a DNA-free segment is given by

$$\overline{N_{\text{DNA-free}}} = \frac{1-f_{\text{onDNA}}}{sm_{\text{tot}}} N_{\text{tot}} \quad (23)$$

Where  $f_{\text{onDNA}}$  is the fraction of time bound to DNA,  $sm_{\text{DNA}}$  and  $sm_{\text{tot}}$  are the number of DNA segments and the total number of segments respectively and  $N_{\text{tot}}$  is the total number of particles in a cell. The ratio, which is equal to the enrichment factor  $EF$ , can then be expressed as:

$$EF = \frac{\overline{N_{\text{DNA}}}}{\overline{N_{\text{DNA-free}}}} = \left( \frac{f_{\text{onDNA}}}{sm_{\text{DNA}}} + \frac{(1-f_{\text{onDNA}})}{sm_{\text{tot}}} \right) / \frac{1-f_{\text{onDNA}}}{sm_{\text{tot}}} \quad (24)$$

If the number of DNA-free segments is much less than the number of DNA segments  $sm_{\text{DNA}} \approx sm_{\text{tot}}$  the expression above can be simplified to:

$$EF = \frac{1}{1-f_{\text{onDNA}}} \quad (25)$$

This equation allows extraction of  $f_{\text{onDNA}}$  from  $EF$  directly and implies that this value does not depend on the diffusion coefficients of the mobile population.

***In vivo*  $K_D$  values**

The  $K_D$  value is a commonly calculated affinity constant used for binding kinetics of proteins and assembly of multicomponent systems <sup>79</sup>, but the  $K_D$  has also been used as an estimate for *in vivo* binding affinity <sup>80</sup>. In the reaction scheme  $A + B \rightleftharpoons AB$ , the  $K_D$  is calculated as

$$K_D = [A][B]/[AB] \quad (26)$$

For Cascade the reaction scheme is as follows: [Cascade (probing)] + [free target sites]  $\rightleftharpoons$  [Cascade (bound)]. The concentration of a single entity inside of a cell of length 4  $\mu\text{m}$  and width 1  $\mu\text{m}$  with hemispherical endcaps is approximately 0.5 nM. The copy number for pTarget was estimated by qPCR to be approximately 100 plasmids per chromosome. As the number of chromosomes in actively dividing cells is generally higher than one, we used literature values for the number of chromosomes/cell found in <sup>81</sup>, providing 4/cell which also used a glucose and amino acid enriched M9 medium as growth medium. This brings the copy number of pTarget to 400/cell, which is equal to 200 nM. For a Cascade complex carrying one of several crRNAs

in the cell, the amount of free target sites is equal to the copy number of the plasmid pTarget minus the amount of already occupied target sites of that crRNA, but as the copy number of each target (400) is much higher than the number of Cascade complexes potentially carrying that crRNA (on average  $130/18 \approx 7$ ),  $[\text{free targets}] \approx [\text{pTarget}]$ . The  $K_D$  value was then calculated as:

$$\begin{aligned} K_D &= [\text{pTarget}][\text{Cascade}(\text{probing})]/[\text{Cascade}(\text{bound})] \\ &= 200 \text{ nM } [\text{Cascade}(\text{probing})]/[\text{Cascade}(\text{bound})] \end{aligned} \quad (27)$$

### Theoretical model interference level vs copy number

In the case where the interference level is limited by the target search of the proteins, we can model the relation based on the distribution of search times of single proteins. The search time for a single protein, because it is the arrival time of a recurring independent random event, is exponentially distributed and characterized by the average search time,  $\langle t_s \rangle$ :

$$p_1(t_s) = 1/\langle t_s \rangle e^{-t_s/\langle t_s \rangle} \quad (28)$$

We have verified given our kinetic model of Cascade with simulations that this is the case (Figure S3). The chance that one of  $n$  proteins finds the target at search time  $t_s$  while the other proteins have not yet found the target is:

$$p_n(t_s) = np_1(t_s) \left( \int_{t_s}^{\infty} p_1(t) dt \right)^{n-1} = n/\langle t_s \rangle e^{-nt_s/\langle t_s \rangle} \quad (29)$$

We have verified this derivation with simulations (Figure S3). The establishment probability of the plasmid is equal to the likelihood for all search times larger than  $t_{\text{critical}} (t_c)$ , the time point at which the cell can no longer clear the invader. Therefore:

$$P_{\text{establishment}}(t_c) = \int_{t_c}^{\infty} p_n(t) dt = e^{-nt_c/\langle t_s \rangle} \quad (30)$$

As the chance of targeting after replication is low, we assume in our model that Cascade is only able to clear the foreign DNA before replication. Therefore  $t_c$  is equal to the replication time of the plasmid  $t_R$ .

As we found that 20 copies of Cascade reduce interference level by half, this leads to

$$\ln(0.5) = -20t_R/\langle t_s \rangle \quad (31)$$

or

$$t_R/\langle t_s \rangle = 0.035 \quad (32)$$

Right after transformation, the negative regulators of copy numbers are absent, so replication in that instant is faster than the growth rate of the cell. Replication time of pTarget has not been measured so far, but by using a temperature-dependent ori, Olsson *et al.* measured a replication time of 3 min for a slightly larger plasmid in the absence of copy number control<sup>20</sup>. If we assume pTarget replication occurs on a similar time scale, we get an estimated search time for one Cascade to find a single target of ~90 minutes.



We can further describe the relationship between the average search time  $\langle t_s \rangle$  and the  $k_{\text{off}}$  and  $k_{\text{on}}^*$  that were measured for Cascade. This relationship is found by multiplying the amount of time spent for each binding event times the average amount of binding events required to find the target. The amount of time spent for each binding event is equal to the sum of the time spent on binding ( $1/k_{\text{off}}$ ) and the time spent on diffusing to the next site ( $1/k_{\text{on}}^*$ ).

Therefore the average search time is:

$$\langle t_s \rangle = \left( \frac{1}{k_{\text{off}}} + \frac{1}{k_{\text{on}}^*} \right) \frac{\text{\#DNA binding events}}{\text{\#DNA target sites}} \quad (33)$$

We have again verified this description by using our simulations of our kinetic model of Cascade target search (Figure S3).

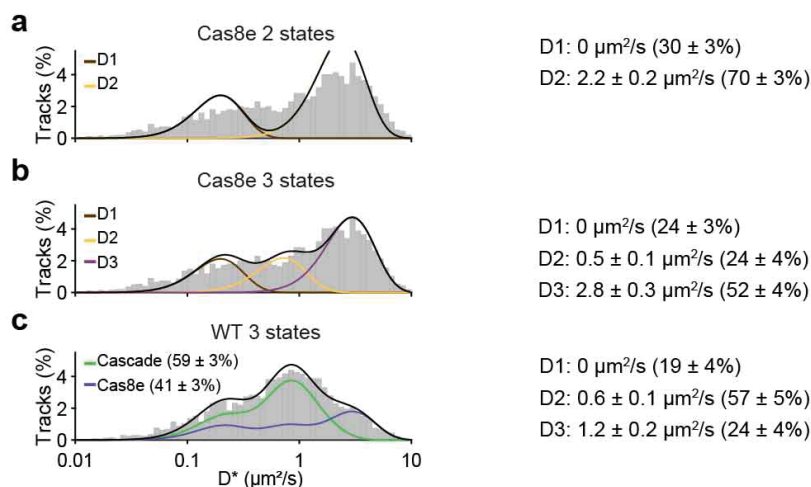
It must be emphasized that the number of binding events is different from the number of binding sites in the fact that if a single binding event scans multiple sites (during 1D sliding), the number of binding sites probed per event are more than one. Using Eq. 30 and 33, the chance of establishment of a single invader in the cell with multiple Cascade copies is therefore as follows:

$$p_{\text{establishment}} \sim e^{-n/\text{\#DNA binding events} \left( \frac{1}{k_{\text{off}}} + \frac{1}{k_{\text{on}}^*} \right)} \quad (34)$$

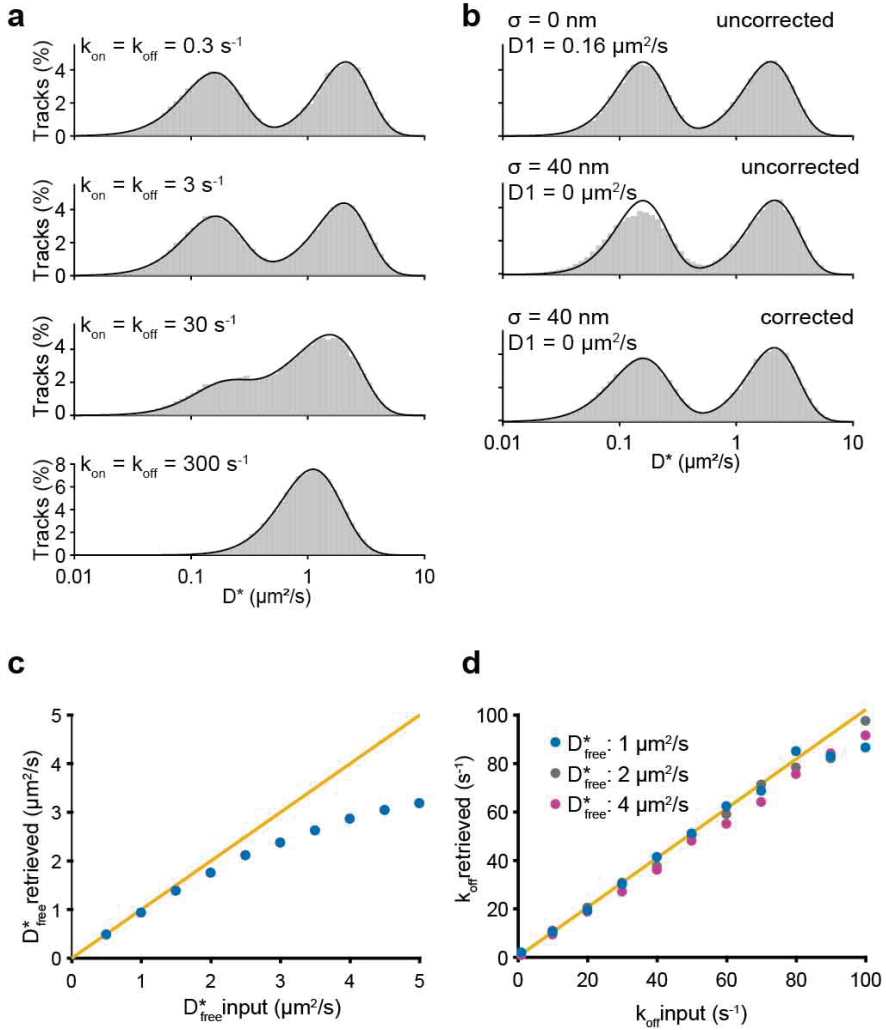
### Simulation Cascade search times

To see whether the above described theoretical model was compatible with our kinetic model of Cascade search, we simulated the search times of Cascade. To do so, we simulated Cascade probing DNA sites as was described above (See Simulation Localization and Confinement). Subsequently every time Cascade changed from a mobile state to a bound state we added with a certain probability that the newly probed site is the target (1/90.000). When Cascade located the target the simulation for that particle was stopped and the search time was recorded for each individual Cascade complex. To simulate the search time for 5 Cascades, we grouped the single search times in multiples of five and took the fastest search time of 5 Cascades.

## 4.7 Supplementary Information



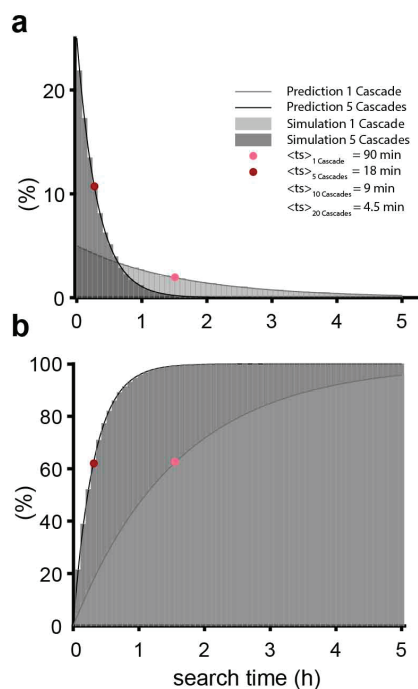
**Supplementary Figure 4.1: Static  $D^*$  fitting, Related to Figure 4.2.** **a**,  $D^*$  distribution (left) of the Cas8e strain (Fig. 4.2b) fitted with two static states with extracted  $D^*$  value of each fraction on the right (relative abundance). The slowest state (D1; brown) was fixed to  $0 \mu\text{m}^2/\text{s}$ . **b**, Same as (a) but then for three static states. **c**,  $D^*$  distribution (left) of the WT strain (Fig. 4.2c). Cas8e distribution from Supplementary Fig. 4.1b was taken and used to fit the distribution with additional three states for Cascade diffusion. The relative abundance of Cas8e and Cascade estimated from static  $D^*$  fitting is similar to that found for dynamic fitting (60 and 40%), even though the distributions of Cascade and Cas8e are different. Error estimation in (a-c) is based on bootstrapping ( $\pm$  standard deviation).



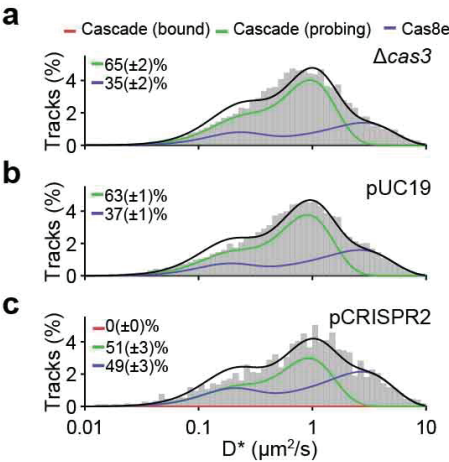
**Supplementary Figure 4.2: Performance of analytical DDA, Related to Figure 4.2. a,**

Comparison of simulation to the theoretical distribution (black line) found with the newly developed analysis method. 50.000 particles were simulated to move without boundaries and position was recorded for 4 consecutive steps. Particles were simulated with  $D^*_{\text{free}} = 2 \mu\text{m}^2/\text{s}$  and increasing on- and off-rates (from 0.3 to 300  $\text{s}^{-1}$ ). The theoretical model (black line) is directly plotted on top of the histogram of simulated  $D^*$  values. A localization error drawn from a Gaussian distribution with  $\sigma = 40 \text{ nm}$  was added to both the model and the simulation. **b**, Influence of localization error. Distribution of an average of consecutive displacements that are offset by a localization error are correlated, which is why in the absence of localization error in the simulation (top) there is no requirement for correction. However immobile particles offset by localization error with the same mean apparent diffusion coefficient are slightly differently distributed (middle). Correction (described in Methods) for the immobile particles is sufficient to restore the fit (bottom). **c**, Influence of

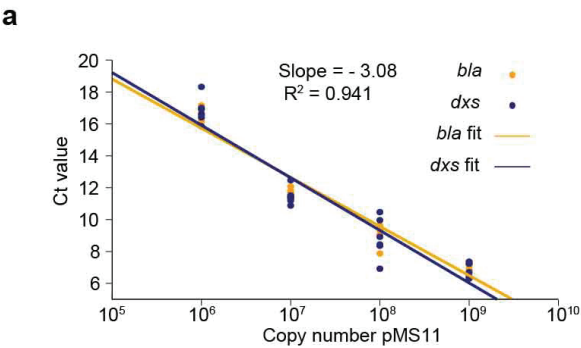
confinement. Particles were simulated inside of a cell 4  $\mu\text{m}$  long and of 1  $\mu\text{m}$  diameter. Simulations were run through analysis software to retrieve parameters.  $D_{\text{free}}^*$  estimates are influenced by confinement where fast moving particles appear to be slower. **d**, The off-rate is not as influenced by effects of confinement and stays the same even for the fastest moving particles (purple). Estimates become more unreliable for much faster or slower transitions than are measured in the integrated time of typical tracks.



**Supplementary Figure 4.3: Comparison between theoretical prediction and independent simulations of Cascade search times, Related to Figure 4.2. a**, Probability density function (PDF) and **b**, cumulative distribution function (CDF) of the distribution of search times. The equations of search time distributions for single and multiple (5) copies (Eq. 28 and Eq. 29) were tested against a simulation of Cascade search times using parameters of the kinetic model that were found experimentally ( $k_{\text{off}} = 38 \text{ s}^{-1}$ ;  $k_{\text{on}}^* = 26 \text{ s}^{-1}$ ) and an estimated 90.000 DNA sites to be scanned before reaching the target (See Methods, theoretical model interference level vs copy number). For the theoretical prediction, the average search time was calculated by using Equation 33. Both the theoretical prediction and simulations indicate that the average search time (indicated by a dot), decreases sharply from 90 minutes for a single Cascade copy to 4.5 minutes for 20 Cascade copies.



**Supplementary Figure 4.4:  $D^*$  Histograms other conditions, Related to Figure 4.5.**  $D^*$  distributions for **a**,  $\Delta cas3$  strain, **b**,  $\Delta cas3$  strain + pUC19, the empty variant of pTarget-RNAP and pCRISPR1-RNAP and **c**,  $\Delta cas3$  strain + pCRISPR2. The amount of available Cascade complexes in the interference assay for strain pCRISPR2 targeting (Fig. 4.5h) were extracted from the relative amount of Cascade complexes in this strain (51%) divided by the number of complexes in the WT strain (60%).

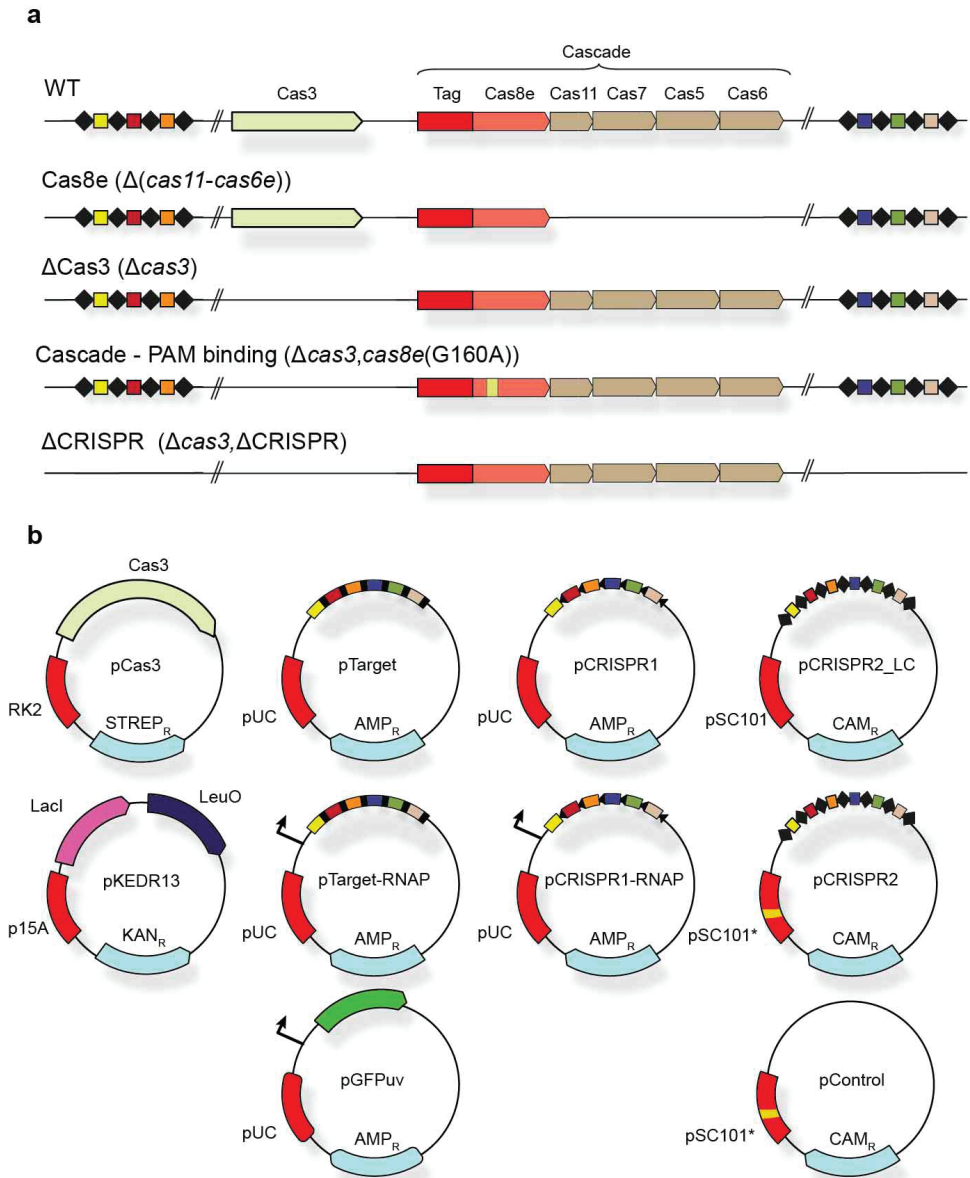


**Supplementary Figure 4.5: Plasmid copy number determination, Related to Figure 4.5.**

**a**, Calibration curve of *dxs* and *bla* primer amplification with dilution series of pMS11 (plasmid containing both *dxs* and *bla* gene). The regression of six technical replicates was used to make the calibration curve for both primer sets (regression parameters of *bla* and *dxs* gene in orange and purple respectively). **b**, The Ct values of *bla* and *dxs* gene amplifications were calculated from biological triplicates. These Ct values were converted to absolute copy numbers

	Ct (CN) <i>bla</i>	Ct (CN) <i>dxs</i>	PCN
$\Delta cas3$ +pTarget	$8.82 \pm 0.09$ ( $10^{-5.75}$ )	$15.15 \pm 0.11$ ( $10^{-7.76}$ )	$102 \pm 10$
$\Delta cas3$ +pCRISPR1	$8.85 \pm 0.07$ ( $10^{-5.76}$ )	$15.20 \pm 0.22$ ( $10^{-7.78}$ )	$104 \pm 15$

(CN) by using the regression values from the calibration curve. The plasmid copy number per chromosome (PCN/chromosome) was calculated by dividing the copy number of the *bla* gene by the copy number of the *dxs* gene. The plasmid copy number per cell was estimated by multiplying PCN/chromosome by the expected number of chromosomes per cell (4) based on a literature value <sup>81</sup>



**Supplementary Figure 4.6: Strains and plasmids used, Related to Figure 4.5.** **a**, Strains used in this study, strains were constructed with lambda recombination and verified by sequencing. Only part of each CRISPR array indicated (total 18 spacers). **b**, Plasmids used in this study. Indicated are the ori (red), antibiotic resistance marker (light blue) and other components on the plasmid. Only part of the total 18 spacers are indicated for pTarget, pCRISPR1 and pCRISPR2. For sequences and descriptions see Supplementary Tables 4.3 and 4.4

Supplementary Tables

	Doubling time
K12 BW25113	24.9 ± 0.1 min
WT + pKEDR13	24.5 ± 0.4 min
WT + pKEDR13 + 1 mM IPTG	31.7 ± 0.6 min
WT + pCas3 + pKEDR13 + 1 mM IPTG	33.3 ± 0.2 min
ΔCas3 + pKEDR13 + pTarget + 1 mM IPTG	31.8 ± 0.4 min

**Supplementary Table 4.1: Growth rate of *E. coli* strains used in this study.** Growth rates were determined in a plate reader where cells were inoculated in similar conditions as described in Methods. The instantaneous growth rate was determined at  $t = 2.5$  hours, which represented the growth rate at the time of the microscope studies. Three independent cultures were measured to get the mean and standard error values.

Supplementary Table 4.2: Primers used in this study

Name	Description	Sequence (5'-3')
BG7128	PAmCherry ( <i>lox-cam-lox</i> ) insert fw (WT)	GGAGGCTATTAAAGGTGCACAAT
BG7129	PAmCherry ( <i>lox-cam-lox</i> ) insert fw ( $\Delta cas3$ )	GTCTCTTCTTTGCAGGGAGG
BG7130	PAmCherry ( <i>lox-cam-lox</i> ) insert rv (WT)	TATCGTCACGGGGCAAAC
BG7131	PAmCherry ( <i>lox-cam-lox</i> ) insert rv (G160A)	AGCAGGTATAGACTCATTGGACT
BG7366	$\Delta CRISPR1$ insert ( <i>lox-kan-lox</i> ) fw	GCAGAGGCGGGGGAACCTCCAAGTGA TATCCATCATCGCATCCAGTGCGCCG GTGTCCTTTTTACCTGTTTGACC
BG7367	$\Delta CRISPR1$ insert ( <i>lox-kan-lox</i> ) rv	GGTTGTTTTATGGGAAAAATGCTT TAAGAACAAATGTATACTTTTAGATT CTACCTCTGGTGAAGGAGTTG
BG7368	$\Delta CRISPR2+3$ insert ( <i>lox-kan-lox</i> ) fw	TAAGTGAGAAGGCCGGGCGGGAAC TGCCCGGCCTGAACATACCTGAATTA GAGTCGGACTTCGCGTTTCGC
BG7369	$\Delta CRISPR2+3$ insert ( <i>lox-cam-lox</i> ) rv	GATTGTGACTGGCTTAAAAATCATT AATTAATAATAGGTTATGTTTAGAGCT AGTTATTGCTCAGCGGTGG
BG8366	$\Delta(cas11-cas6e)$ insert ( <i>lox-kan-lox</i> ) fw	TTGAGTGGAATGGGATTAAGGGGAA GCCAGGTCATTTTATTACACCTCAAG GTGTCCTTTTTACCTGTTTGAC
BG8367	$\Delta(cas11-cas6e)$ insert ( <i>lox-kan-lox</i> ) rv	ACAAACATTTACGGGAGTTAAACCG CAAGGAGGGCCATCAAATGGCTGAT TCCTACCTCTGGTGAAGGAGTTG
BG8677	qPCR <i>bla</i> fw	CTACGATACGGGAGGGCTTA
BG8678	qPCR <i>bla</i> rv	ATAAATCTGGAGCCGGTGAG
BG8679	qPCR <i>dxs</i> fw	CGAGAACTGGCGATCCTTA
BG8680	qPCR <i>dxs</i> rv	CTTCATCAAGCGGTTTCACA
BN370	pCRISPR2 (array2.3) rv	GTGAGCTGATACCGCTCGCCTGAAC CTCTC TGGCATGGA
BN383	pCRISPR2 (array2.1) fw	TGCTTTAAGAACAAATGTATACTTTTA G
BN384	pCRISPR2 (array2.1) rv	TCTAAACATAACCTATTATTACCAAGT GATA TCCATCATCGC
BN385	pCRISPR2 (array2.3) fw	GCGATGATGGATATCACTTGGTAATA ATAG GTTATGTTTAGA
BN373	Site-directed mutagenesis RepA HC fw	TGGTTAAAGGCTTTTCGGATCTTCCAG
BN374	Site-directed mutagenesis RepA LC fw	TGGTTAAAGGCTTTGAGATCTTCCAG
BN375	Site-directed mutagenesis RepA HC+LC rv	AAGGATTCCTGATTTCCACAGTTC



Supplementary Table 4.3: Synthetic DNA inserts used in this study

Description	Sequence (5'-3')
PAmyCherry ins	TGGCGTTAAGCATTTCGCGAGGTTCCAGATGGACAAAAGCCCCAGGCGAT ATTTCTATCAACCTGAGGCCAGCGTTCGAACCCAAACAATTGGAATGTTAG TCTCTTCTTTGCAGGGAGGCAAGACATGTGTATATCACTGTAATTCGATAT TTATGAGCAGCATCGAAAAATAGCCCGCTGATATCATCGATAATACTAAAA AAACAGGGGAGGCTATTAAAGGTGCACAATGTACATCTTCTTTAATTTCCC GGTATGAGATTTTATTTACAGTATGAATATTTTATGTAATAAAATTCATG GTAATTATTATAACTAAAAGTTTCTTTAATAATAAGGCGCCCTAGGTACC GTTCTGTATAATGTATGCTATACGAAGTTATGAGCTGTTGACAATTAATCAT CGGCTCGTATAATGTGTGGGCAATGAGCTTGCACTGCAGAACTTTGATAT ACCATGGAGAAAAAATCACTGGATATACCACCGTTGATATATCCCAATGG CATCGTAAAGAACATTTTGAGGCATTTCACTCAGTTGCTCAATGTACCTAT AACCAGACCGTTGAGCTGGATATTACGGCCTTTTAAAGACCGTAAAGAAA AATAAGCACAAAGTTTTATCCGGCCTTTATTACATTCTTGCCCGCCTGATG AATGCTCATCCGAATTCCTGTATGGCAATGAAAGACGGTGAGCTGGTGAT ATGGGATAGTGTTACCCCTTGTACACCGTTTTCCATGAGCAAACCTGAAAC GTTTTCATCGCTCTGGAGTGAATACCACGACGATTTCCGGCAGTTTCTACA CATATATTCGCAAGATGTGGCGTGTTACGGTGAAAACCTGGCCTATTTCC CTAAAGGGTTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGA GTTTCACCAAGTTTTGATTTAAACGTGGCCAATATGGACAACCTTCTTCGCC CCGTTTTCACTATGGGCAAATATTATACGCAAGGCGACAAGGTGCTGATG CCGCTGGCGATTCAAGTTTCATCATGCCGTTTGTGATGGCTTCCATGTCGG CAGAAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAGGGCGGG GCGTAAATAACTTCGTATAATGTATGCTATACGAACGGTATCTAGACTTCG GGAATGATTGTTATCAATGACGATAATAAGACCAATAACGGTTTATCCCTA CTTAAGTAGGGAAGGTGCACAATGTACATCTTCTTTAATTTCCCGGTATG AGATTTTATATTCACAGTATGAATATTTTATGTAATAAAATTCATGGTAATTA TTATACTAAAAGTTTCTTTAATAATAAAACGAATAACTTGCAGATTTGAAA TGCATGCATTATTGTCTTTAAACAATTCACACATCTTAATATATGTATAGG TTAATTGTATTAACCAATGAATATATTTTGCAGTGAATGTGATTATTGAA TTAATTACGCCGTATTTTTCTTTGTTTTACCGATAACGGAAGTGCGCGA CGTATAGAAATGAGGAGAAATGTCGGAGCATGAAAGGAGAAACAATGG TGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCG CTTCAAGGTGCACCTGGAGGGGTCCGTGAACGGCCACGAGTTCGAGATC GAGGGCGAGGGCGAGGGCCGCCCTACGAGGGCACCCAGACCGCCAA GCTGAAGGTGACCAAGGGTGGCCCCCTTGCCCTTCGCCTGGGACATCCTG TCCCCTCAGTTCATGTACGGCTCCAATGCCTACGTGAAGCACCCCGCCGA CATCCCCGACTACTTTAAGCTGTCTTCCCCGAGGGCTTCAAGTGGGAGC GCGTGATGAACCTCGAAGACGGCGGCGTGGTGACCGTGACCCAGGACTC CTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACC AACTTCCCCTCCGACGGCCCCGTAAATGCAGAAAGACCATGGGCTGGG AGACCTCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGAG AGCTCAAGGCGAGGACGAAGCTGAAGGACGGCGGCCACTATGACACTGA GGTCAAGACCACTACAAGGCCAAGAAGCCCGTGCAGTTGACCTGGCGCC TACAACGTCAACCGCAAGTTGGATATCACTCCCACAACGAGGACTACAC CATCGTGGAACAGTACGAACGTGCCGAGGGCCTCACTCCACCGGCGGC ATGGACGAGCTGTACAAGCCCGGGGCGCTCATGGCTAATTTGCTTATTGA TAACTGGATCCCTGTACGCCCGCGAAACGGGGGGAAAGTCCAAATCATAA ATCTGCAATCGCTATACTGCAGTAGAGATCAGTGGCGATTAAAGTTTGCC CGTGACGATATGGAATGGCCGCTTTAGCACTGGTGGTTGCAATGGGCA AATTATCGCCCCGGCAAAAGATGACGTTGAATTTGACATCGCATAATGAA TCCGCTCACTGAAGATGAGTTTCAACAACCTCATCGCGCCGTGGATAGATA

	<p>TGTTCTACCTTAATCACGCAGAACATCCCTTTATGCAGACCAAAGGTGTCA  AAGCAAATGATGTGACTCCAATGGAAAACTGTTGGCTGGGGTAAGCGGC  GCGACGAATTGTGCATTTGTCAATCAACCGGGGCAGGGTGAAGCATTATG  TGGTGGATGCACTGCGATTGCGTTATTCAACCAGGCGAATCAGGCACCAG  GTTTTGGTGCCGGTTTTAAAGCGGTTTACGTGGAGGAACACCTGTAACA  ACGTTCTGTACGTGGGATCGATCTTCGTTCAACGGTGTTACTCAATGTCT  CACATTACCTCGTCTTCAAAAACAATTTCTAATGAATCACATACGGAAAA  CCAACTACCTGGATTAACCTATCAAGTCCAATGAGTCTATACCTGCTTC  GTCAATTGGGTTTGTCCGTGGTCTATTCTGGCAACCAGCGCATATTGAATT  ATGCGATCCCATTGGGATTGGTAAATGTTCTTGCTGTGGACAGGAAAGCA  ATTTGCGTTATACCGG</p>
pTarget insert	<p>TCTAGAGAATTGACAGAACGGCCTCAGTAGTCTCGTCAGGCTCCTTCTG  TTTTCGCAAATCTATGGACTATTGCTATTCTTGGGCGCACGGAATACAAAG  CCGTGTATCTGCTCTTTGGCTCTGCAACAGCAGCACCCATGACCACGTCT  TGAAATGCTGGTGAGCGTTAATGCCGCAAACACCTTATTACGCCTTTTTCG  GATTGCCCGGTTTTTGCCTTCCATGGCAGCGTCAGGCGTGAAATCTCACC  GTCGTTGCCTTTCGGTTACAGGCGTTGCAAACCTGGCTACCGGGCTTGTAG  TCCATCATTCCACCTATGTCTGAACCTCCCTCCGGGGGATAATGTTACGG  TCATGCGCCCCCTTTGGGCGGCTTGCTTGACGCCAGCTCCAGCAGCTG  TAAGCTGGCTGGCAATCTCTTTCGGGGTGAGTCCTTTAGTTTCCGTATCTC  CGGATTATATAAGCTGACTTGCAGGCGGCGACGCGCAGGGTATGCGCGA  TTCGCTTGCGACCGCTCAGAAATTCAGACCCGATCCAACTTTCAACATT  ATCAATTACAACCGACAGGGAGCCCTTAGCGTGTTCCGGCATCACCTTTGG  TTCGCGCTGCTTTGCGTGAGCGTATCGCCGCGCGTCTGCGAAAGCTTGG  TACC</p>
pCRISPR1 insert	<p>TCTAGAGAATTGACAGAACGGCCTCAGTAGTCTCGTCAGGCTCCGGCTG  TTTTCGCAAATCTATGGACTATTGCTATTCCGGGGGCGCACGGAATACAAA  GCCGTGTATCTGCTCGGTGGCTCTGCAACAGCAGCACCCATGACCACGT  CGGGAAATGCTGGTGAGCGTTAATGCCGCAAACACCGGATTACGCCTTTT  TGCGATTGCCCGGTTTTTGCCGGCCATGGCAGCGTCAGGCGTGAAATCT  CACCGTCGTTGCCGGTGGTTTCAGGCGTTGCAAACCTGGCTACCGGGCG  GGTAGTCCATCATTCCACCTATGTCTGAACCTCCCGGCCGGGGGATAATGT  TTACGGTCATGCGCCCCCGGTGGGCGGCTTGCTTGCAGCCAGCTCCA  GCAGCGGAAGCTGGCTGGCAATCTCTTTCGGGGTGAGTCCGGTAGTTTC  CGTATCTCCGATTTATAAAGCTGACGGGCAGGCGGCGACGCGCAGGGT  ATGCGCGATTGCGGGGCGACCGCTCAGAAATTCAGACCCGATCCAAAC  GGTCAACATTATCAATTACAACCGACAGGGAGCCCGAGCGTGTTTCGGC  ATCACCTTTGGCTTCGGCTGCGGTGCGTGAGCGTATCGCCGCGCGTCTG  CGAAAGCGGGGTACC</p>

Supplementary Table 4.4: Plasmids used in this study

Name in study	Name in storage	Description	Source
pKEDR13	pKEDR13	Expression plasmid LeuO	<sup>18</sup>
pGFPuv	pGFPuv	Expression plasmid GFPuv	Clontech
pMS011	pMS011	Plasmid containing <i>bla</i> and <i>dxs</i> gene (qPCR)	<sup>82</sup>
pSC020	pSC020	Plasmid containing Cre and lambda recombinase	S. Creutzberg (unpublished)
pTarget	pTU256	Target plasmid containing all 18 potential protospacers for flanked by 5'-CTT-3'	This study
pTarget-RNAP	pTU150	Target plasmid containing all 18 potential protospacers for flanked by 5'-CTT-3' and <i>p/ac</i> upstream	This study
pCRISPR1	pTU258	Target plasmid containing all 18 potential protospacers for flanked by 5'-CGG-3'	This study
pCRISPR1-RNAP	pTU152	Target plasmid containing all 18 potential protospacers for flanked by 5'-CGG-3' and <i>p/ac</i> upstream	This study
pCRISPR2_LC	pTU158	Plasmid containing all 18 potential protospacers for flanked by repeat sequences; low copy backbone variant of pSC101	This study
pCRISPR2	pTU160	Plasmid containing all 18 potential protospacers for flanked by repeat sequences; high copy backbone variant of pSC101	This study
pControl	pTU254	High copy backbone variant of pSC101	This study
pCas3	pTU255	Expression plasmid Cas3	This study

## Glossary

Full name	Symbol	Description
Apparent diffusion coefficient	$D^*$	Apparent due to confinement
Bound state	$S1$	
Dissociation constant	$K_D$	Constant which is a measure for the binding affinity of two objects with each other
DNA segments	$sm_{DNA}$	Amount of segments defined as containing DNA by DAPI staining
Enrichment Factor	$EF$	The number of localizations in DNA-containing segments divided by the number of localizations in DNA-free segments
Fraction DNA bound	$f_{onDNA}$	Fraction of the time DNA binding proteins spend on DNA is calculated from the off- and on-rate (Figure 1).
Frametime	$t_f$	Positions of simulated/measured particles are recorded for each frametime
Free diffusion coefficient	$D_{free}^*$	Diffusion coefficient in the absence of DNA binding. Apparent due to confinement.
Integrated time	$t_{int}$	Overall timescale: can be one or multiple frametimes
Localization error	$\sigma$	Average error in determination of particle position
Mobile state	$S2$	
off-rate	$k_{off}$	Rate DNA bound protein is released from DNA. Inverse of residence time
pseudo-first order on-rate	$k_{on}^*$	Rate mobile protein is binding to DNA. As the amount of potential DNA probing sites is very large, on-rate is independent of DNA concentration (pseudo-first order)
pTarget establishment	$P_{establishment}$	Measure for interference level calculated from the transformation ratio of pTarget and pGFPuv (Eq. 1)

Time step	$t_{\text{step}}$	Displacements of simulated particles are calculated for each time step
Total number of segments	$sm_{\text{tot}}$	Total number of segments in the cell

## References

1. Marraffini, L. A. CRISPR-Cas immunity in prokaryotes. *Nature* **526**, 55–61 (2015).
2. Deveau, H. *et al.* Phage Response to CRISPR-Encoded Resistance in *Streptococcus thermophilus*. *J. Bacteriol.* **190**, 1390–1400 (2008).
3. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733–740 (2009).
4. Shao, Q., Hawkins, A. & Zeng, L. Phage DNA Dynamics in Cells with Different Fates. *Biophys. J.* **108**, 2048–2060 (2015).
5. Jones, D. L. *et al.* Kinetics of dCas9 target search in *Escherichia coli*. *Science* (80-. ). (2017) doi:10.1126/science.aah7084.
6. Gleditzsch, D. *et al.* PAM identification by CRISPR-Cas effector complexes: diversified mechanisms and structures. *RNA Biol.* 15476286.2018.1504546 (2018) doi:10.1080/15476286.2018.1504546.
7. Redding, S. *et al.* Surveillance and Processing of Foreign DNA by the *Escherichia coli* CRISPR-Cas System. *Cell* **163**, 1–12 (2015).
8. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62–67 (2014).
9. Globyte, V., Lee, S. H., Bae, T., Kim, J. & Joo, C. CRISPR/Cas9 searches for a protospacer adjacent motif by lateral diffusion. *EMBO J.* e99466 (2018) doi:10.15252/embj.201899466.
10. Leenay, R. T. *et al.* Identifying and Visualizing Functional PAM Diversity across CRISPR-Cas Systems. *Mol. Cell* **62**, 137–147 (2016).
11. Vigouroux, A., Oldewurtel, E., Cui, L., Bikard, D. & van Teeffelen, S. Tuning dCas9's ability to block transcription enables robust, noiseless knockdown of bacterial genes. *Mol. Syst. Biol.* **14**, e7899 (2018).
12. Bondy-Denomy, J. *et al.* Multiple mechanisms for CRISPR-Cas inhibition by anti-CRISPR proteins. *Nature* **526**, 136–139 (2015).
13. Pawluk, A., Bondy-Denomy, J., Cheung, V. H. W., Maxwell, K. L. & Davidson, A. R. A new group of phage anti-CRISPR genes inhibits the type I-E CRISPR-Cas system of *Pseudomonas aeruginosa*. *MBio* **5**, (2014).
14. Manley, S. *et al.* High-density mapping of single-molecule trajectories with photoactivated localization microscopy. *Nat. Methods* **5**, 155–157 (2008).
15. English, B. P. *et al.* Single-molecule investigations of the stringent response

- machinery in living bacterial cells. *Proc. Natl. Acad. Sci.* **108**, 365–373 (2011).
16. Subach, F. V. *et al.* Photoactivatable mCherry for high-resolution two-color fluorescence microscopy. *Nat. Methods* **6**, 153–159 (2009).
17. Majsec, K., Bolt, E. L. & Ivančić-Baće, I. Cas3 is a limiting factor for CRISPR-Cas immunity in *Escherichia coli* cells lacking H-NS. *BMC Microbiol.* **16**, 28 (2016).
18. Westra, E. R. *et al.* H-NS-mediated repression of CRISPR-based immunity in *Escherichia coli* K12 can be relieved by the transcription activator LeuO. *Mol. Microbiol.* **77**, 1380–93 (2010).
19. Severinov, K., Ispolatov, I. & Semenova, E. The Influence of Copy-Number of Targeted Extrachromosomal Genetic Elements on the Outcome of CRISPR-Cas Defense. *Front. Mol. Biosci.* **3**, (2016).
20. Olsson, J. A., Berg, O. G., Dasgupta, S. & Nordström, K. Eclipse period during replication of plasmid R1: contributions from structural events and from the copy-number control system. *Mol. Microbiol.* **50**, 291–301 (2003).
21. Floc'h, K. *et al.* Bacterial cell wall nanoimaging by autoblinking microscopy. *Sci. Rep.* **8**, 14038 (2018).
22. Sashital, D. G., Wiedenheft, B. & Doudna, J. A. Mechanism of Foreign DNA Selection in a Bacterial Adaptive Immune System. *Mol. Cell* **46**, 606–615 (2012).
23. Jore, M. M. *et al.* Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat. Struct. Mol. Biol.* **18**, 529–536 (2011).
24. Brown, M. W. *et al.* Assembly and translocation of a CRISPR-Cas primed acquisition complex. *Cell* **175**, 1–13 (2018).
25. Xue, C., Zhu, Y., Zhang, X., Shin, Y.-K. & Sashital, D. G. Real-Time Observation of Target Search by the CRISPR Surveillance Complex Cascade. *Cell Rep.* **21**, 3717–3727 (2017).
26. Beloglazova, N. *et al.* CRISPR RNA binding and DNA target recognition by purified Cascade complexes from *Escherichia coli*. *Nucleic Acids Res.* **43**, 530–543 (2015).
27. Brouns, S. J. J. *et al.* Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science (80-. )*. **321**, 960–964 (2008).
28. Mondal, J., Bratton, B. P., Li, Y., Yethiraj, A. & Weisshaar, J. C. Entropy-based mechanism of ribosome-nucleoid segregation in *E. coli* Cells. *Biophys. J.* **100**, 2605–2613 (2011).
29. Sanamrad, A. *et al.* Single-particle tracking reveals that free ribosomal subunits are not excluded from the *Escherichia coli* nucleoid. *Proc. Natl. Acad. Sci.* **111**, 11413–11418 (2014).
30. Reyes-Lamothe, R. *et al.* High-copy bacterial plasmids diffuse in the nucleoid-free space, replicate stochastically and are randomly partitioned at cell division. *Nucleic Acids Res.* **42**, 1042–51 (2014).
31. Stracy, M. *et al.* Live-cell superresolution microscopy reveals the organization of RNA polymerase in the bacterial nucleoid. *Proc. Natl. Acad. Sci.* **112**, E4390–E4399 (2015).
32. Hayes, R. P. *et al.* Structural basis for promiscuous PAM recognition in type

- I-E Cascade from *E. coli*. *Nature* **530**, 499–503 (2016).
33. van Erp, P. B. G. *et al.* Mechanism of CRISPR-RNA guided recognition of DNA targets in *Escherichia coli*. *Nucleic Acids Res.* **43**, 8381–8391 (2015).
34. Xiao, Y. *et al.* Structure Basis for Directional R-loop Formation and Substrate Handover Mechanisms in Type I CRISPR-Cas System. *Cell* **170**, 48–60.e11 (2017).
35. Jung, C. *et al.* Massively Parallel Biophysical Analysis of CRISPR-Cas Complexes on Next Generation Sequencing Chips. *Cell* **170**, 35–47.e13 (2017).
36. Martens, K. J. A. *et al.* An open microscopy framework suited for tracking dCas9 in live bacteria. *bioRxiv* 437137 (2018) doi:10.1101/437137.
37. Xue, C., Whitis, N. R. & Sashital, D. G. Conformational Control of Cascade Interference and Priming Activities in CRISPR Immunity. *Mol. Cell* **64**, 826–834 (2016).
38. Blosser, T. R. *et al.* Two distinct DNA binding modes guide dual roles of a CRISPR-Cas protein complex. *Mol. Cell* **58**, 60–70 (2015).
39. Hochstrasser, M. L. *et al.* CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. *Proc. Natl. Acad. Sci.* **111**, 6618–6623 (2014).
40. Datsenko, K. A. *et al.* Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat. Commun.* **3**, 945 (2012).
41. Jackson, S. A. *et al.* CRISPR-Cas: Adapting to change. *Science* (80-. ). **356**, eaal5056 (2017).
42. 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) doi:10.1016/j.molcel.2012.03.018.
43. Szczelkun, M. D. *et al.* Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc. Natl. Acad. Sci.* **111**, 9798–9803 (2014).
44. Ma, J. & Wang, M. D. DNA supercoiling during transcription. *Biophysical Reviews* vol. 8 75–87 (2016).
45. Staals, R. H. J. *et al.* Interference-driven spacer acquisition is dominant over naive and primed adaptation in a native CRISPR-Cas system. *Nat. Commun.* **7**, 12853 (2016).
46. Xiao, Y., Luo, M., Dolan, A. E., Liao, M. & Ke, A. Structure basis for RNA-guided DNA degradation by Cascade and Cas3. *Science* **361**, eaat0839 (2018).
47. Martynov, A., Severinov, K. & Ispolatov, I. Optimal number of spacers in CRISPR arrays. *PLoS Comput. Biol.* **13**, (2017).
48. Høyland-Kroghsbo, N. M., Muñoz, K. A. & Bassler, B. L. Temperature, by Controlling Growth Rate, Regulates CRISPR-Cas Activity in *Pseudomonas aeruginosa*. *MBio* **9**, (2018).
49. Chen, Y. J. *et al.* Two-Stage Dynamics of in Vivo Bacteriophage Genome Ejection. *Phys. Rev. X* **8**, (2018).
50. Davison, J. Pre-early functions of bacteriophage T5 and its relatives.

- Bacteriophage* **5**, e1086500 (2015).
51. De Smet, J., Hendrix, H., Blasdel, B. G., Danis-Wlodarczyk, K. & Lavigne, R. Pseudomonas predators: Understanding and exploiting phage-host interactions. *Nature Reviews Microbiology* (2017) doi:10.1038/nrmicro.2017.61.
52. Modell, J. W., Jiang, W. & Marraffini, L. A. CRISPR–Cas systems exploit viral DNA injection to establish and maintain adaptive immunity. *Nature* **544**, 101–104 (2017).
53. Peterson, J. & Phillips, G. J. New pSC101-derivative cloning vectors with elevated copy numbers. *Plasmid* **59**, 193–201 (2008).
54. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 6640–5 (2000).
55. Chandradoss, S. D. *et al.* Surface Passivation for Single-molecule Protein Studies. *J. Vis. Exp.* **86**, 4–11 (2014).
56. Farooq, S. & Hohlbein, J. Camera-based single-molecule FRET detection with improved time resolution. *Phys. Chem. Chem. Phys.* **17**, 27862–27872 (2015).
57. Edelstein, A., Amodaj, N., Hoover, K., Vale, R. & Stuurman, N. Computer Control of Microscopes Using µManager. in *Current Protocols in Molecular Biology* vol. Chapter 14 Unit14.20 (John Wiley & Sons, Inc., 2010).
58. Holden, S. J. *et al.* Defining the Limits of Single-Molecule FRET Resolution in TIRF Microscopy. *Biophys. J.* **99**, 3102–3111 (2010).
59. Uphoff, S., Reyes-Lamothé, R., De Leon, F. G., Sherratt, D. J. & Kapanidis, A. N. Single-molecule DNA repair in live bacteria. *Proc. Natl. Acad. Sci. U. S. A.* (2013) doi:10.1073/pnas.1301804110.
60. Vliet, L. Van, Sudar, D. & Young, I. Digital fluorescence imaging using cooled charge-coupled device array cameras. *Cell Biol.* **III**, 109–120 (1998).
61. Hoogendoorn, E. *et al.* The fidelity of stochastic single-molecule super-resolution reconstructions critically depends upon robust background estimation. *Sci. Rep.* **4**, 3854 (2015).
62. Huang, F. *et al.* Video-rate nanoscopy using sCMOS camera-specific single-molecule localization algorithms. *Nat. Methods* **10**, 653–658 (2013).
63. Vrljic, M., Nishimura, S. Y., Brasselet, S., Moerner, W. E. & McConnell, H. M. Translational diffusion of individual class II MHC membrane proteins in cells. *Biophys. J.* **83**, 2681–2692 (2002).
64. Qian, H., Sheetz, M. P. & Elson, E. L. Single particle tracking. Analysis of diffusion and flow in two-dimensional systems. *Biophys. J.* **60**, 910–921 (1991).
65. Palo, K., Mets, Ü., Loorits, V. & Kask, P. Calculation of photon-count number distributions via master equations. *Biophys. J.* **90**, 2179–2191 (2006).
66. Kalinin, S., Felekyan, S., Valeri, A. & Seidel, C. A. M. Characterizing Multiple Molecular States in Single-Molecule Multiparameter Fluorescence Detection by Probability Distribution Analysis. *J. Phys. Chem. B* **112**, 8361–8374 (2008).
67. Antonik, M., Felekyan, S., Gaiduk, A. & Seidel, C. A. M. Separating structural



- heterogeneities from stochastic variations in fluorescence resonance energy transfer distributions via photon distribution analysis. *J. Phys. Chem. B* **110**, 6970–6978 (2006).
68. Michalet, X. Mean square displacement analysis of single-particle trajectories with localization error: Brownian motion in an isotropic medium. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* **82**, 041914 (2010).
  69. Paintdakhi, A. *et al.* Oufi: An integrated software package for high-accuracy, high-throughput quantitative microscopy analysis. *Mol. Microbiol.* **99**, 767–777 (2016).
  70. Lee, S.-H., Shin, J. Y., Lee, A. & Bustamante, C. Counting single photoactivatable fluorescent molecules by photoactivated localization microscopy (PALM). *Proc. Natl. Acad. Sci.* **109**, 17436–17441 (2012).
  71. Huang, B. *et al.* Counting low-copy number proteins in a single cell. *Science* **315**, 81–84 (2007).
  72. Durisic, N., Laparra-Cuervo, L., Sandoval-Álvarez, Á., Borbely, J. S. & Lakadamyali, M. Single-molecule evaluation of fluorescent protein photoactivation efficiency using an in vivo nanotemplate. *Nat. Methods* **11**, 156–162 (2014).
  73. Ho, H. N., Van Oijen, A. M. & Ghodke, H. The transcription-repair coupling factor Mfd associates with RNA polymerase in the absence of exogenous damage. *Nat. Commun.* **9**, 1570 (2018).
  74. Knight, S. C. *et al.* Dynamics of CRISPR-Cas9 genome interrogation in living cells. *Science* **350**, 823–826 (2015).
  75. Nenninger, A., Mastroianni, G. & Mullineaux, C. W. Size dependence of protein diffusion in the cytoplasm of *Escherichia coli*. *J. Bacteriol.* **192**, 4535–4540 (2010).
  76. Mika, J. T. & Poolman, B. Macromolecule diffusion and confinement in prokaryotic cells. *Curr. Opin. Biotechnol.* **22**, 117–126 (2011).
  77. Kumar, M., Mommer, M. S. & Sourjik, V. Mobility of cytoplasmic, membrane, and DNA-binding proteins in *Escherichia coli*. *Biophys. J.* **98**, 552–559 (2010).
  78. Mika, J. T., Van Den Bogaart, G., Veenhoff, L., Krasnikov, V. & Poolman, B. Molecular sieving properties of the cytoplasm of *Escherichia coli* and consequences of osmotic stress. *Mol. Microbiol.* **77**, 200–207 (2010).
  79. McGuigan, J. A. S., Kay, J. W. & Elder, H. Y. Critical review of the methods used to measure the apparent dissociation constant and ligand purity in Ca<sup>2+</sup> and Mg<sup>2+</sup> buffer solutions. *Prog. Biophys. Mol. Biol.* **92**, 333–370 (2006).
  80. Zawadzki, P. *et al.* The Localization and Action of Topoisomerase IV in *Escherichia coli* Chromosome Segregation Is Coordinated by the SMC Complex, MukBEF. *Cell Rep.* **13**, 2587–2596 (2015).
  81. Wallden, M., Fange, D., Lundius, E. G., Baltekin, Ö. & Elf, J. The Synchronization of Replication and Division Cycles in Individual *E. coli* Cells. *Cell* **166**, 729–739 (2016).
  82. Caforio, A. *et al.* Converting *Escherichia coli* into an archaeobacterium with a hybrid heterochiral membrane. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 3704–3709 (2018).



# 5

## Single Cell Variability of CRISPR-Cas Interference and Adaptation

R. E. McKenzie\*, E. M. Keizer\*, J. N. A. Vink, J. van Lopik, F. Büke, V. Kalkman, C. Fleck,  
S. J. Tans, S. J. J. Brouns, bioRxiv, submitted, doi:10.1101/2021.07.21.453200.

\* These authors contributed equally

## Abstract

CRISPR-Cas defence is a combination of adaptation to new invaders by spacer acquisition, and interference by targeted nuclease activity. While these processes have been studied on a population level, their temporal dynamics and variability at the level of individual cells has remained unknown. Here, using a microfluidic device combined with time-lapse microscopy and mathematical modelling, we study invader clearance in a population of *Escherichia coli* across multiple generations. We observed that CRISPR interference is fast with a narrow distribution of clearance times. In contrast, for invaders with escaping PAM mutations we show large cell-to-cell variability of clearance times, which originates from primed CRISPR adaptation. Faster growth and cell division, as well as higher levels of Cascade, increase the chance of clearance by interference. In contrast, faster growth is associated with decreased chances of clearance by priming. We identify Cascade binding to the mutated invader DNA, rather than spacer integration, as the main source of priming heterogeneity. The highly stochastic nature of primed CRISPR adaptation implies that only subpopulations of bacteria are able to respond to invading threats in a timely manner. We conjecture that CRISPR-Cas dynamics and heterogeneity at the cellular level are crucial to understanding the strategy of bacteria in their competition with other species and phages.

## 5.1 Introduction

During the last decade, important progress has been made in identifying the sequence of steps and molecular interactions required for successful adaptive immunity by the model type I-E CRISPR-Cas system<sup>1–10</sup>. CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) immunity involves three main stages beginning with the acquisition of a spacer, a small piece of DNA derived from a foreign invader and stored in the CRISPR array for future defence<sup>11,12</sup>. This array is then transcribed and processed into small CRISPR RNAs (crRNAs) which guide a surveillance complex, formed from a number of Cas (CRISPR-associated) proteins, towards the invaders DNA<sup>13,14</sup>. For type I-E systems a 5'-CTT consensus PAM (Protospacer Adjacent Motif) sequence flanking the targeted site of the invader<sup>15,16</sup> is required for recognition and ultimately degradation of the invader, through a process called direct interference<sup>7,17–19</sup>. However, invaders can escape direct interference via mutation within the seed region of the target site or PAM<sup>15,20,21</sup>. In response, the I-E system can initiate priming, which promotes accelerated acquisition of new spacers due to a pre-existing partial match to the invader<sup>1,4</sup>. Primed adaptation is much faster than naïve adaptation<sup>22</sup>, and is required for the insertion of a new matching spacer with a consensus PAM allowing subsequent invader degradation, which we here refer to as primed interference.

At the level of individual cells however, much more is unknown. Interference is a tug-of-war between invader replication and degradation, which could result in complex and stochastic dynamics within single cells. Replication and degradation themselves may also display variability between cells in the population. For instance, invader degradation rates can be affected by stochastic processes such as the expression of CRISPR-Cas components, target localization, and nuclease recruitment<sup>3,23</sup>. Priming also depends on many processes in which the dynamical interplay is unclear, including the production of suitable fragments of DNA for spacer acquisition (pre-spacers), the assembly of adaptation complexes required for further spacer selection, and the processing and insertion of these pre-spacers into the CRISPR array<sup>6,10,24,25</sup>. Elucidating the cellular dynamics and heterogeneity of the CRISPR-Cas response is critical to understanding interference and adaptation mechanistically, and of direct importance to its natural function. For instance, upon invasion, cells are thought to have a limited time window to respond in order to escape invader replication, protein production, and cell death<sup>26–29</sup>.

A number of studies have investigated the interference process by collecting either population averages, or single-cell data on short time scales (<1 s)<sup>3,5,21,30–32</sup>. However, averaging within a population can conceal the variation between cells, as well as the dynamics within single-cells over time<sup>33,34</sup>, thus masking the underlying dynamics of CRISPR-Cas interference. In addition, investigations into the adaptation process have provided great insight into the diversity of spacers acquired<sup>30,35</sup>, possible mechanisms of target destruction<sup>1,36</sup>, and conditions under which adaptation most frequently occurs within a population<sup>37–39</sup>, however these studies could not observe any variation existing in each step of the adaptation process within individual cells.

Here we set out to investigate and quantify the dynamics and variability of the CRISPR-Cas response at the single-cell level. Using time-lapse microscopy and microfluidic devices, we followed individual cells over multiple rounds of division while simultaneously monitoring CRISPR-Cas protein expression and DNA degradation. Hence, we obtained individual lineages, the genealogical relations between them, as well as real-time data on the DNA clearance process, instantaneous growth rates, cell sizes, and division frequencies of individual cells. We determined that while direct interference occurs quickly and consistently, clearing the target from all cells within hours, priming is highly variable and much slower, taking over several tens of hours for some cells. We were able to define the adaptation and clearance stages of priming and identified primed adaptation as the source of the variation observed – more specifically the binding of Cascade to the mutated target DNA, rather than other complex processes including the integration of new spacer DNA fragments into the host genome.

## 5.2 Results

### 5.2.1 Time-lapse microscopy of the CRISPR response

Using two strains, KD615 (WT) and KD635 ( $\Delta cas1,2$ ) (**Supplementary Table 5.1**), we investigated priming and direct interference respectively. The strains contain an array with a leader, two repeats and a single previously characterized spacer, spacer8 (SP8)<sup>4,5</sup> (**Fig. 5.1a-c**). In addition, these strains are engineered to control *cas* gene expression using arabinose and IPTG induction, and hence initiation of the CRISPR-Cas response. Target plasmids were engineered to encode a constitutively expressed YFP or CFP

fluorescent protein<sup>40</sup> and contain a target sequence that is complementary to SP8 in the CRISPR array, allowing direct monitoring of target DNA presence in individual cells over time (**Fig. 5.1a-c**) (**Supplementary Table 5.1**). In order to investigate the direct interference process, we flanked the target sequence with a 5'-CTT consensus PAM<sup>16</sup> (**Fig 5.1a,b**). Further, to investigate the priming response we mutated the PAM to 5'-CGT (**Fig 5.1b,c**), a mutation known to allow mobile genetic elements (MGE) to escape interference, and invoke a primed adaptation response<sup>1,5,20</sup>.

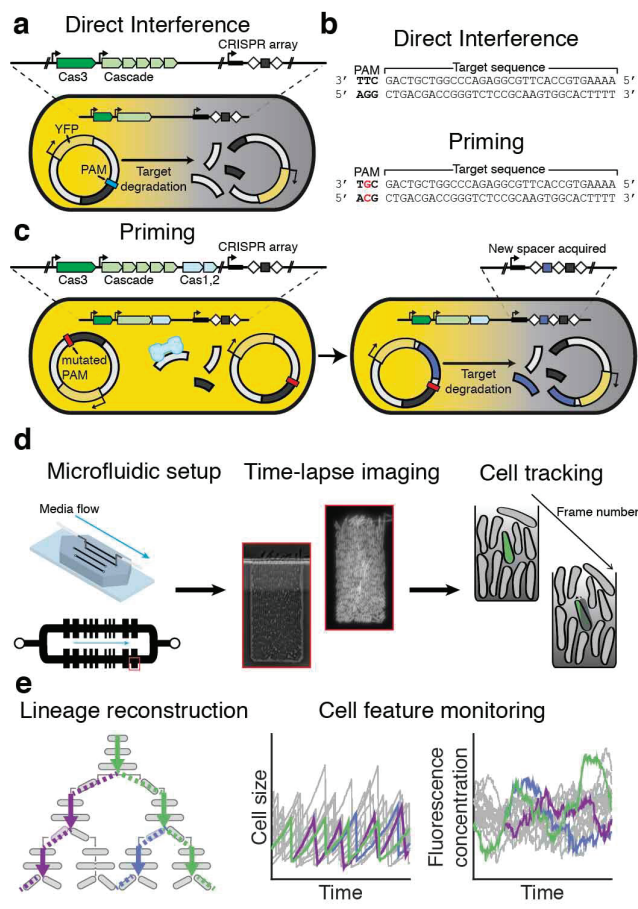
Use of a microfluidic device<sup>41</sup> enabled fluorescence time-lapse imaging for over 36 hours with the option for media exchange (**Fig. 5.1d**). The device contained chambers allowing observation of a single layer of cells, constant medium supply, removal of cells that no longer fit the chamber due to growth, and control of intracellular processes via induction. Image analysis software was used to segment and track all cells and their fluorescence signals, thus allowing the re-construction of lineage trees in a defined region at the bottom of the chamber (**Fig. 5.1d,e**)<sup>41-43</sup>.

### 5.2.2 Direct interference is fast and synchronous

We first investigated the direct interference response (**Fig. 5.1a**). Prior to cas gene induction, the images showed high YFP fluorescence in all cells, confirming the presence of the target plasmid (**Fig. 5.2a**) which decreased upon induction, indicating CRISPR-Cas mediated degradation of the target DNA (**Fig. 5.2a**). When the plasmid did not contain a target sequence (pControl) YFP levels did not decrease for over 35 hours (**Supplementary Fig. 5.1**), indicating that the plasmid loss was CRISPR-Cas dependent.

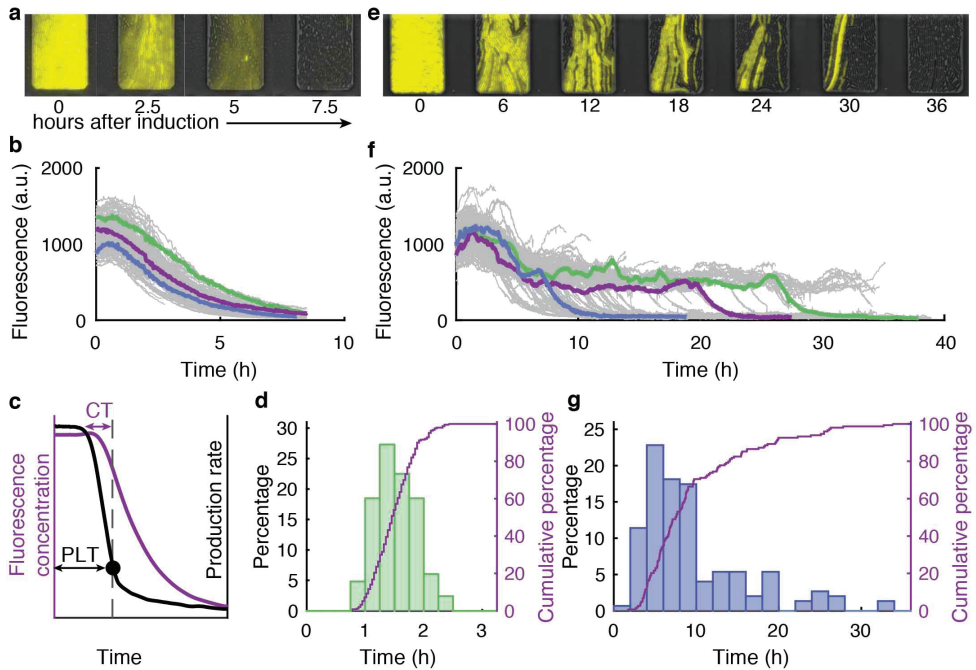
The mean YFP fluorescence per cell unit area (which estimates the YFP concentration) showed the decrease started after about 1 hour of induction, and then exhibited a smooth monotonic decline without significant fluctuations (**Fig. 5.2b**). Note that traces end upon the cells exiting the observation chamber. CRISPR mediated degradation of the target was thus efficient and synchronous, and in the case of a 5-copy plasmid could overcome the plasmid replication and copy number control. Hence, we surmised that the YFP fluorescence may decrease exponentially, as the YFP proteins are diluted exponentially due to volume growth upon clearance of the last plasmid. Indeed, we found the fluorescence decrease to be exponential (**Supplementary Fig. 5.2**).

Direct interference variability between cells also appeared limited (**Fig. 5.2b**). To address it more directly, we quantified the moment all plasmids are cleared by determining the YFP production rate as the change in total cellular fluorescence per unit of time <sup>44</sup>. The production rate scales with the number of target DNA copies, and shows the expression timing more precisely by suppressing slow dilution effects. Indeed, the YFP production rate decreased rapidly, and reached zero (the background level of cells not expressing YFP) when the mean fluorescence was still close to its maximum (**Fig. 5.2c**). This moment was identified as the plasmid loss time (PLT) (**Fig. 2c**). PLT was narrowly distributed between about 1 and 2.5 hours (**Fig. 5.2d**,  $CV^2 = 0.055$ ). Hence, in all cells the target was cleared. The clearance was rapid taking between 1 and 3 generations, and sometimes occurred in the same generation in which the CRISPR-Cas response was initiated by induction (**Supplementary Fig. 5.3**).



**Figure 5.1: Investigating single-cell behaviour during CRISPR defence using time-lapse microscopy.** **a**, Schematic of the direct interference process. The cell contains a I-E CRISPR system, as well as the CRISPR array with a single spacer targeting the plasmid (grey box). The plasmid encodes YFP and contains a sequence matching the spacer (grey), flanked by a consensus PAM (blue). Immediate targeting by the CRISPR system resulting in degradation of the plasmid and loss of the YFP in the cell. **b**, To invoke priming the 5'-CTT consensus PAM, flanking the target sequence located on the plasmid, is mutated by one nucleotide to a non-consensus PAM 5'-CGT. **c**, Schematic of the priming process. (Left cell) A mutation of the PAM (red) flanking the target sequence means the spacer in the CRISPR array can no longer initiate direct interference. Fragments in the cell can be captured and processed by Cas1,2 (light blue). (Right cell) The Cas1,2 complex integrates the fragment into the CRISPR array as a new spacer (purple), which can target the plasmid resulting in degradation and loss of YFP in the cell. **d**, To allow long term imaging cells are grown in a microfluidic chip that allows constant media supply. Cells within a single well are imaged frequently in phase contrast and fluorescence allowing segmentation and tracking of lineage history across frames. **e**, Variation in features of reconstructed single-cell lineages (left) such as size (middle) and YFP concentration (right) are continuously monitored enabling further investigation.





**Figure 5.2: Variation in target plasmid clearance times is much larger when CRISPR adaptation is required.** **a**, Overlay of fluorescent and phase contrast time-lapse images of cells targeting a matching plasmid with a consensus PAM. Presence of the target plasmid is tracked by its YFP production. Images are shown at 2.5 hour intervals starting from induction of *cas* gene expression. **b**, Reconstructed lineage traces of the imaged population (**a**) from induction of the CRISPR system over time (grey) lineages show some variation in plasmid clearance times (coloured). **c**, Production rate (black line) of the YFP is used to determine the plasmid loss time, PLT, (black dot, black arrow) allowing earlier detection than using the mean fluorescence (purple line). The time from first targeting of a single plasmid to production rate threshold is defined as the clearance time (CT, purple arrow). **d**, Distribution of plasmid loss times determined by the production rate during direct interference. **e**, Overlay of time-lapse images of cells targeting a matching plasmid with a mutated non consensus PAM. Presence of the target is tracked by YFP production. Images are shown at 6 hour intervals. **f**, Reconstructed lineage traces of the imaged population (**e**) from induction of the CRISPR system over time (grey) lineages show large variations in the time taken to clear the plasmid (coloured). **g**, Distribution of plasmid loss times calculated with the production rate during priming.

### 5.2.3 Primed adaptation is highly variable

Next, we studied plasmid clearance after adaptation from a target with a mutated PAM (**Fig. 5.1c**). Most notable in these priming experiments was the heterogeneity between lineages, with the clearance process ranging from 2-30 cellular generations (**Supplementary Fig. 5.3**). Upon induction, some lineages showed a decreasing trend in fluorescence as early as 4 hours (**Fig. 5.2e-f**), while others remained fluorescent after 35 hours (**Fig. 5.2f**). The PLT's were indeed broadly distributed and displayed a long tail towards large values (**Fig. 5.2g**,  $CV^2 = 0.458$ ). Of note, we did not observe plasmid clearance in the same generation in which the CRISPR-Cas system was induced (**Supplementary Fig. 5.3**).

The shapes of the YFP declines were exponential, similar to the direct interference data (**Fig. 5.2b and f, Supplementary Fig. 5.2**). Alignment of all production rate traces at the PLT yielded a similar averaged profile for direct interference and priming (**Supplementary Fig. 5.4**). In these data, the onset of the decrease is about 60 min before PLT in both cases, thus estimating the clearance time (CT), the duration of the target clearance process. In priming, clearance thus contributes much less to PLT variability than the preceding processes (**Fig. 5.2g**). These observations suggest that new spacers preceded by a consensus PAM are indeed acquired, and that the CRISPR adaptation phase is responsible for the observed temporal variability (**Fig. 5.2g**).

Spacer acquisition in the population was indeed confirmed by PCR of the CRISPR array in cells collected from the microfluidic device (**Supplementary Fig. 5.5**). Spacer acquisition was not observed with the  $\Delta cas1,2$  strain, consistent with Cas1 and Cas2 being required for acquisition<sup>45</sup>. In the absence of Cas1 and Cas2 however, low frequency plasmid loss was observed in 1.4% of the lineages over a 35-hour period (**Supplementary Fig. 5.6**). Hence, complete clearance is possible with a mutated PAM, even if highly inefficient.

### 5.2.4 Genealogical relations impact the CRISPR response

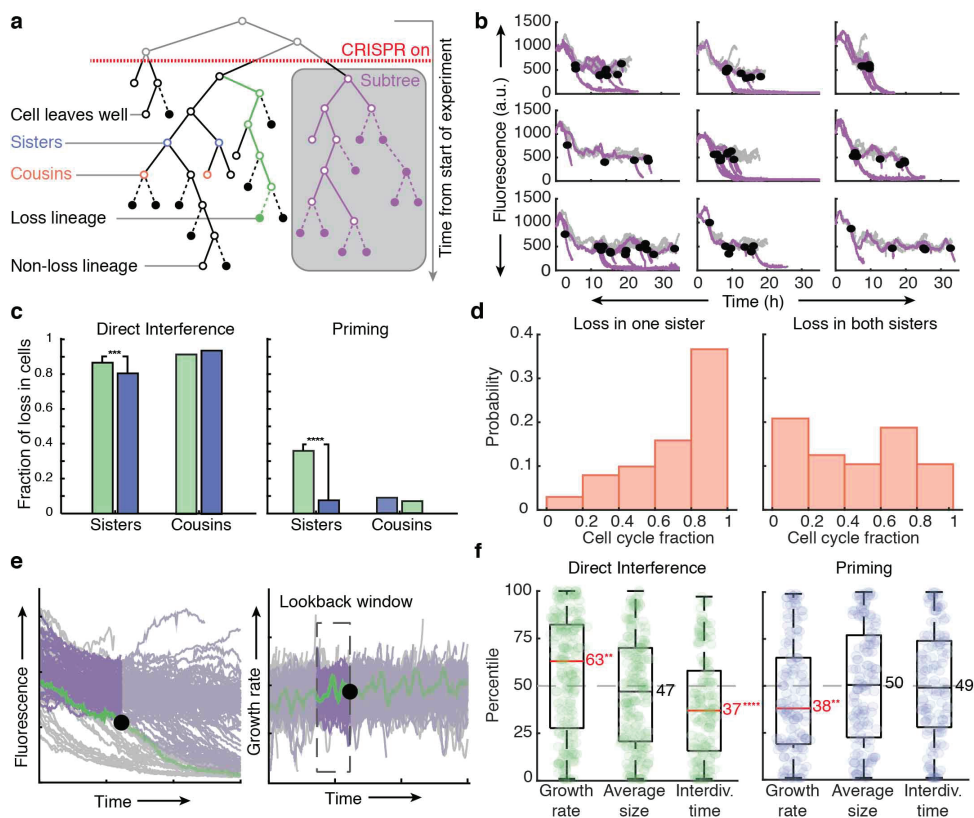
To study the role of genealogy in the CRISPR-Cas response, we took a more in depth look at the lineage history before plasmid loss (**Fig. 5.3a**). For primed adaptation, some subtrees showed all plasmid loss events occurring close together, however most subtrees showed a wide PLT variability (**Fig.**

**5.3b**, black dots), in line with lineages responding independently. However, statistical analysis showed that sisters cleared their plasmids within the same cell cycle more frequently than expected at random, and more strongly so for priming than for direct interference (**Fig. 5.3c**). Hence, inheritance plays a role in the CRISPR-Cas response (**Fig. 5.3c**).

These data led us to hypothesize that in priming, sisters correlate because spacer acquisition occurs in the mother, after which plasmid degradation (primed interference) continues into the daughters. If true, the moment of plasmid loss may be distributed throughout the daughter's cell cycle, depending on the acquisition moment within the mother's cell cycle. Conversely, when acquisition and clearance both manage to occur in the mother, hence yielding no correlated loss events, clearance should occur at the end of the mother's cell cycle, because of the time (~ 60 min) needed for plasmid clearance. To test this hypothesis, we divided the cell cycles into five fractions, and tabulated the observed loss event for each fraction. Indeed, loss events without correlated plasmid loss in the sister cell occurred more frequently towards the end of the cell cycle (**Fig 5.3d**), while the moment of loss was more randomly distributed when both sisters lost the plasmid (**Fig 5.3d**). Altogether this indicated that loss likely takes place more frequently in sisters than cousins (**Fig 5.3c**) because adaptation occurred in the mother.

### 5.2.5 The growth rate has opposing effects on adaptation and interference

To study if stochastic variations in cell cycle parameters affect the CRISPR-Cas response, we used a ranking analysis, as this suppresses long-term trends in the population (**Fig. 5.3e**). For instance, we ranked each lineage showing plasmid loss relative to lineages that had not lost their plasmids at that moment, with the ranking based on growth rate averaged over a 'lookback window' (**Fig. 5.3e**), determined using autocorrelation times (**Supplementary Fig. 5.7**). In direct interference, the 'loss lineages' exhibited a higher median growth rate than 'non-loss lineages', with their growth rate ranking in the 63<sup>rd</sup> percentile ( $p=0.01$ ) (**Fig. 5.3f**). These lineages also showed shorter interdivision times ( $p=0.0001$ ), but not a difference in cell size (**Fig. 5.3f**). These results were robust over a range of lookback window sizes (see **Supplementary Fig. 5.8**). We stress that growth is likely only one of the many factors affecting the CRISPR-Cas response, which is also reflected by the broad ranking distributions (**Fig. 5.3f**). Overall, the analysis



**Figure 5.3: Growth rate and interdivision times have an influence on direct interference and priming.** **a**, Schematic of key analysis structure and terminology. **b**, A comparison of 9 subtrees constructed from induction. **c**, The observed fraction of loss in cells (navy) related as either sisters or cousins during direct interference or priming is plotted against the fraction of expected loss events (cyan) in related cells when the events are randomized in the same time window. **d**, The cell cycle was divided into 5 equal fractions and plasmid loss times are plotted in the corresponding fraction where one sister alone lost the plasmid (left) or both sisters lost the plasmid (right). **e**, schematic explaining the rank-based analysis approach. For each detected loss event (left, black circle) the cell feature i.e. growth rate for that lineage (right, green) is averaged over a lookback window (right, dashed rectangle), and then ranked amongst all averages of non-loss lineages in the same window (violet, right). **f**, Boxplots of percentile rankings of all loss lineages that cleared a consensus target via direct interference (green, left), or a mutated target via priming (navy, right), for growth rate, birth size and interdivision respectively over a lookback window of 30 minutes. The median percentile ranking of loss lineages is indicated by a line and value, categories in which this value was significantly different from a ranking in the 50<sup>th</sup> percentile are indicated in red followed by asterisks. (\*\*\*\* $P < 0.0001$ , \*\*\* $P < 0.001$ , \*\* $P < 0.01$ )

indicated that faster growth in coordination with more frequent cell division has a positive effect on the rate of clearance of a consensus target.

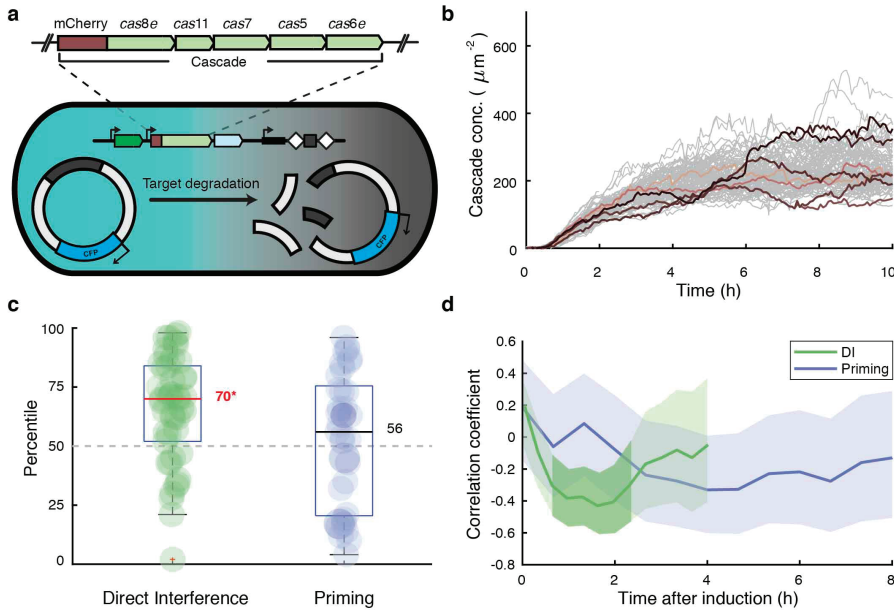
Primed adaptation showed a different picture. To probe the effects on spacer acquisition, which occur about 60 min before plasmid loss, we used a lookback window between 90 and 60 min before the PLT. While cell size and interdivision time did not show an effect (no significant deviation from the 50<sup>th</sup> percentile) the growth rate did, with loss lineages growing more slowly compared to non-loss lineages (38<sup>th</sup> percentile,  $p=0.01$ ) (**Fig. 5.3f**). This was robust to changes in the lookback window (**Supplementary Fig. 5.9**). Altogether, these findings indicated that, on average, slower growing cells achieved faster plasmid clearance through priming.

### 5.2.6 Cascade concentrations impact the CRISPR response

Apart from physiological determinants like growth<sup>46</sup>, Cascade expression levels may influence the speed of CRISPR-Cas defence. We fused mCherry (RFP) to the N-terminus of the Cas8e subunit of Cascade<sup>3</sup> (**Fig. 5.4a**). Using single particle fluorescence calibration, we estimated that the cells contain on average about 200 Cascade molecules/ $\mu\text{m}^2$  (**Fig. 5.4b** and **Supplementary Fig. 5.10**). Hence, we quantified the (stochastic) variations in Cascade abundance within single-cell-lineages upon induction (**Fig. 5.4b**).

Cascade levels fluctuated on a longer timescale than the cell cycle (200 min, **Supplementary Fig. 5.11**) and were strongly correlated between sisters and cousins ( $R=0.89$  and  $0.62$  respectively, **Supplementary Fig. 5.12**) indicating that Cascade levels are stable over several generations. We reasoned that lineages with high Cascade concentrations may target and clear the plasmids faster. Hence, we performed time-lapse experiments and used the ranking approach, now ranking lineages based on the average Cascade in a window 60 min prior to plasmid loss. For direct interference, loss lineages exhibited significantly higher Cascade levels than non-loss lineages, and ranked in the 70<sup>th</sup> percentile ( $p=0.03$ , **Fig. 5.4c**). Conversely, no differences in Cascade levels were observed between loss and non-loss lineages for priming, with the former ranking in the 56<sup>th</sup> percentile (**Fig. 5.4c**).

In priming however, the target search by Cascade occurs over a longer period of time prior to achieving plasmid loss, likely rendering the ranking approach less suitable. Hence, to test this notion, we quantified the total area under the Cascade concentration curves (**Fig. 5.4b**) in a certain time



**Figure 5.4: Growth rate and interdivision times influence direct interference and priming.** **a**, Schematic of the experimental set up adapted to allow visualization of target presence (CFP) and Cascade levels (mCherry) simultaneously. The expansion indicates the mCherry fluorescent tag was integrated upstream of the *cas8e* subunit. **b**, Cascade concentration of single-cell lineages over time from induction. **c**, Cascade concentrations were averaged over a 30 minute lookback window from the plasmid loss event for all loss lineages during direct interference (green) or priming (navy). The Cascade concentration of the loss lineages were ranked as percentile amongst the non-loss lineages and plotted here. The median percentile ranking of loss lineages is indicated by a line and value, categories in which this value was significantly different from a ranking in the 50<sup>th</sup> percentile ( $p < 0.05$ ) are indicated in red followed by an asterisk. **d**, The Pearson correlation coefficient of plasmid loss time versus total cumulative cascade concentration at that moment is plotted every 5 minutes (DI) or 10 minutes (Priming) starting from induction of the CRISPR system. The plotted line for both a target with a consensus PAM (green) and target with a mutant PAM (navy) are enveloped by a 95% confidence interval. Darker shading indicates where the correlation coefficient is significantly different from zero ( $p < 0.05$ ).

window. At 0-2 hours post induction, PLT and Cascade search hours indeed correlated negatively for direct interference but not for priming (**Fig. 5.4d**), in line with our earlier analysis (**Fig. 5.4c**). Interestingly, beyond 2 hours post induction, the correlation increases in magnitude for priming, even as the variability is high, and the correlation is not significant (**Fig. 5.4d**). In summary, stochastic variations in Cascade expression levels affect direct interference. For the priming process, the impact of Cascade levels appeared weaker. We hypothesize this could be due to the underlying

processes being less synchronized in time in comparison to direct interference, and hence masked by other stochastic variations.

### 5.2.7 Low Cascade-target binding affinity generates CRISPR-Cas response variability

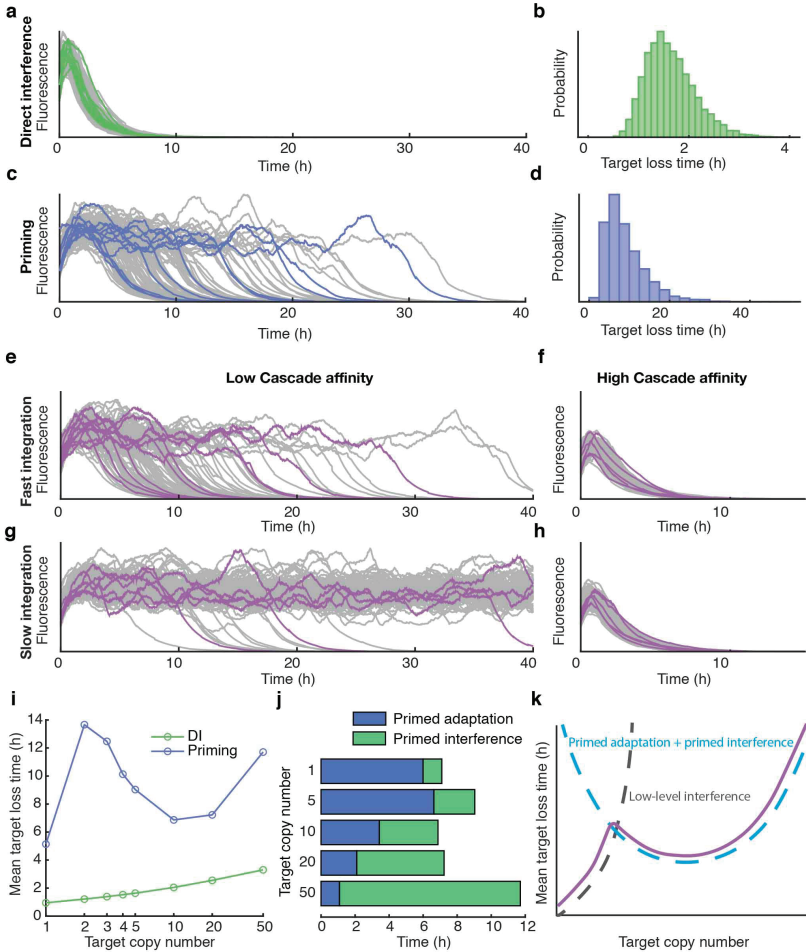
To gain insight into the variability and dynamics of the CRISPR-Cas defence we developed an agent-based simulation framework. Adaptive immunity in bacterial populations has been modelled previously<sup>47–49</sup> but to our knowledge none describe variability or single-cell behaviour. Briefly, we simulated 100 cells, their growth and division, plasmid maintenance, stochastic protein production and partitioning at division, spacer acquisition, and target DNA degradation (see Supplementary Methods for details). We found that with these minimal model ingredients, and by only changing the Cascade-target binding affinity due to the PAM mutation, the model reproduced both the low variability of direct interference (**Fig. 5.5a,b and Fig. 5.2b,d**), and the high variability of priming (**Fig. 5.5c,d and Fig. 5.2f,g**) from the experimental conditions. Specifically, we found a Cascade-target binding affinity reduction of two orders of magnitude for the PAM mutation, which is consistent with previous work<sup>50,51</sup> (**Supplementary Methods Table 5.2**).

The priming process can be conceptually understood as a two-step process, adaptation followed by interference, where a highly reduced rate of the first step is able to recreate the broadness of the PLT distribution (see Supplementary Methods for details). We hypothesized that variation of the primed adaptation response could originate from the low-affinity target search of Cascade, or the spacer integration. In the agent-based model, we were able to vary the rates of these two processes by a factor 100, while keeping the Cas3-mediated target destruction constant, and find that slow spacer integration alone is not enough to explain the observed variability (**Fig. 5.5h**). Conversely, reduced Cascade-target binding affinity is both necessary and sufficient to reproduce the observations (**Fig. 5.5e-h**) and is required to generate pre-spacers.

### 5.2.8 Competition between adaptation and low-level interference

In priming, low Cascade-target affinity and resulting sporadic target degradation can yield a low-level interference prior to adaptation, which in





**Figure 5.5: Results from the stochastic agent-based model of CRISPR adaptation and interference.** **a-d**, Example trajectories showing fluorescence concentration produced by target plasmids simulated with the agent-based model for the **(a)** direct interference and **(c)** priming condition, and corresponding target loss distribution **(b,d)** respectively. **e-h**, Example trajectories from 4 different parameter combinations. High Cascade affinity **(f,h)** corresponds an increase in target binding by a factor 100 as compared to low Cascade affinity **(e,g)**, slow integration **(g,h)** represents a 100-fold reduction in the spacer integration rate as compared to fast integration **(e,f)**. **i**, Mean target loss time of the population as a function of the average target copy number per cell for direct interference (green) and priming (navy). **j**, Breakdown of average time spent on primed adaptation (blue) and primed interference (green) for cells that clear targets through priming, for target copy numbers in the range 1-50. **k**, Schematic of alternative target loss pathways. At low copy numbers, targets can be cleared through low-level interference, which becomes increasingly rare as copy numbers increase. The priming process shows a u-shaped relationship with the target copy number, as a result of adaptation becoming faster as target copy numbers increase, and time required for interference increasing with target copy number.

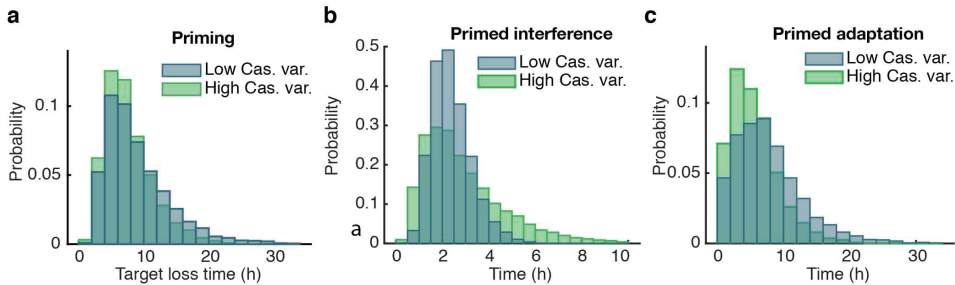


turn provides a continuous source of target DNA fragments that can act as pre-spacers<sup>2</sup>. Hence, we wondered whether target abundance affects this process. For direct interference, as expected, we found that the PLT increased monotonically in simulated trajectories as the average number of targets varies from 1 to 50 (**Fig. 5.5i**, see **Supplementary Fig. 5.13** for full range of distributions). Simulations of priming did not show such a monotonic trend: the PLT first went up, then down, and finally up again (**Fig. 5.5i**, **Supplementary Fig. 5.14**). This behaviour could be explained by splitting priming into adaptation and interference (**Fig. 5.5j**): while primed interference logically only speeds up with fewer targets, primed adaptation initially slows down with fewer targets because of the resulting fewer pre-spacers, but then speeds up for the lowest number of targets, because low-level interference is now sufficiently efficient, in combination with unequal partitioning upon division (**Supplementary Fig. 5.15**). Indeed, our experiments also showed such clearance of a 5-copy target by low-level interference without spacer acquisition (**Supplementary Fig. 5.3**). This alternative pathway competes with priming when there are few targets (**Fig. 5.5k**), and might explain the trend in **Fig. 5j** showing faster loss at 1 target as compared to 5 targets. Target abundance thus affects the balance between primed adaptation and primed interference, resulting in a non-monotonous trend for the target clearance probability.

### 5.2.9 Cascade expression stochasticity can accelerate CRISPR adaptation

Our experiments showed that CRISPR-Cas defence is affected by Cascade expression (**Fig. 5.4c-d**) which is stochastic in nature (**Fig. 5.4b**). For direct interference simulations, we changed the level of gene expression variability for Cascade to have 100-fold stronger expression bursts while maintaining average Cascade concentrations (see **Supplementary Methods** for details), which resulted in a higher mean PLT: while some cells could clear all targets earlier, many cells required more time to clear all targets as compared to lower-variability Cascade expression (**Supplementary Fig. 5.16**). Surprisingly, for priming the mean PLT became lower when the Cascade variability increased (**Fig. 5.6a**). The primed interference phase showed a trend similar to direct interference: a broadening of the PLT distribution yielding a slow-down on average (**Fig. 5.6b**). However, the entire distribution shifted to lower values for primed adaptation (**Fig. 5.6c**), yielding an overall speed-up. For mutated PAMs, pre-spacer production critically depends on high Cascade levels, even if transient, as the cumulative probability of a pre-

spacer integration event depends on the Cas concentration in a highly non-linear fashion (see Supplementary Methods for a more detailed illustration).



**Figure 5.6: Distribution of primed adaptation and primed interference time for high and low variability in Cascade concentration.** **a**, Target loss time distribution for two different levels of Cascade concentration variability for priming. At low variability (blue) Cascade proteins are produced in frequent, small bursts, whereas at high variability (green) proteins are synthesized more sporadically in large bursts (100-fold increase), keeping average Cascade concentration constant. **b-c**, The variability of primed interference times (**b**) for high Cascade variability (green) increases as compared to low Cascade variability (blue), whereas the variability of primed adaptation times (**c**) decreases with higher Cascade variability.

### 5.3 Discussion

In this study, we have investigated a previously unexplored question: what are the dynamics and variability of the CRISPR adaptation and interference responses in individual cells? Our time-lapse microscopy approach allowed real-time monitoring of invader presence, cell traits and inheritance in single-cell lineages. We found that direct interference, despite its dependence on various stochastic processes and poorly understood tug-of-war between replication of invading nucleic acids and degradation by CRISPR-Cas systems, is notably deterministic and efficient, with invader DNA clearance achieved in all cells within 1 to 3 generations. Conversely, the priming CRISPR-Cas response was highly variable, ranging from 2 to 30 generations before clearance. Our data show that direct interference and primed interference can in fact occur on comparable time scales, and identify the adaptation phase of priming as the origin of the variation.

Our direct observation of the CRISPR-Cas action and modelling approach revealed several factors that impact CRISPR-Cas response variability. The interaction between Cascade and the target DNA, which is characterized by

a low affinity owing to the PAM mutation, represents a key source of heterogeneity in the adaptation process of the priming response, rather than the complex spacer integration process. For direct interference we found that cells that cleared the target DNA earlier, grew and divided faster than the population mean. This may be explained by the fact that faster growth is known to reduce plasmid copy numbers<sup>52,53</sup>, while in slow growing cells plasmid maintenance mechanisms increase plasmid abundance<sup>54</sup>. For priming the reverse was found. Cells that primed and cleared the target DNA earlier, grew more slowly, perhaps due to higher target DNA availability. The abundance of Cascade also plays a key role in direct interference and priming. We note that slower growing cells had higher Cascade abundance (Supplementary Fig. 17). Although single cell variability was not monitored, spacer acquisition was found to occur more frequently during slower growth at late stationary phase<sup>37,38</sup>, in line with our findings.

Our finding that target copy number influences the efficiency of spacer acquisition has implications for phage invasion. It suggests that one genome copy of a single virulent phage with an escape PAM may not lead to efficient CRISPR adaptation. However upon replication of the phage genome, it may become abundant enough, though at this point in time it is likely that primed interference with a new spacer cannot successfully eliminate a virulent phage before cell lysis<sup>27,55,56</sup>. Despite this, it has been shown bioinformatically that priming by type I systems is widespread in nature<sup>57</sup>, especially against temperate phages<sup>58</sup>. Such events could occur due to low-level interference, in which a cell is able to simultaneously clear the invader while present as a single copy and acquire a spacer from the fragments produced. This would result in immunization of a single cell in the population, ultimately leading to a subpopulation of resistant cells that could limit further propagation of the same phage. Such a phenomenon may be more likely to occur when a defective phage infects the cell<sup>55</sup>.

The variation existing between single cells in a population is remarkable. Stochasticity or noise in gene expression and cellular components has been demonstrated to play crucial roles in many cellular processes<sup>59–61</sup>. We anticipate that the dynamics and heterogeneity of the CRISPR-Cas system, as studied here, play an important function in strategies that bacteria exploit and evolve in their continuous competition with phages, as well as with other species. For instance, CRISPR-Cas could contribute to bet-hedging strategies<sup>62</sup>, in which subpopulations develop to combat changes in the environment, such as phage exposure. A distinct subpopulation in which Cascade is highly expressed could allow faster elimination of an invading phage, and subsequent re-population. This may in turn increase the fitness

of the population, by reducing the overall burden of CRISPR-Cas expression and risk of autoimmunity<sup>30,63</sup>, and hence outcompete other bacterial strains. While such a strategy may not guarantee single cell survival, it is at large beneficial for the population as whole. Indeed, previous studies have shown CRISPR-Cas immunity in single cells acts to limit phage propagation throughout the population in an abortive infection like manner<sup>64–66</sup>. On the other hand, the survival of only a subpopulation of cells may result in population bottlenecks and an overall loss of diversity<sup>67</sup>. This may be disadvantageous in terms of spacer diversity, where it has been shown that populations containing a range of spacers are better able to combat and even facilitate the extinction of new invaders<sup>35,47</sup>.

While a number of studies have thoroughly investigated CRISPR-Cas systems through population and single molecule based experiments<sup>5,7,69,11,30,36–38,64,65,68</sup>, these findings do not provide insight into the cell-to-cell variability. Our work has begun to bridge this gap demonstrating how important the dynamics of CRISPR-Cas systems are to their functioning and the outcome of populations facing a threat. Further investigation into different CRISPR-Cas types and classes, fluctuating environments<sup>70</sup>, and conditions supporting the formation of subpopulations<sup>71</sup> will enhance the understanding of CRISPR-Cas dynamics on both the molecular and population scale.

## 5.4 Methods

### Cloning

Plasmid pTU166 targeted by KD615 and KD635, was created by amplifying the streptomycin resistance cassette from pCDFDuet-1 with primers BN831 and BN832 to add a 5'CTT-PS8 tail. The backbone of pVenus was amplified using primers BN833 and BN834 and both products were restricted with KpnI and HindIII enzymes. Overnight ligation at 16 °C and transformation into DH5a resulted in colonies selected to contain the plasmid. Plasmids pTU190 and pTU193 were created by PCR amplification of pTU166 using primer BN911 in combination with BN912 or BN891 respectively. Products were restricted with Sall, ligated and transformed into DH5a. Target plasmids pTU389 and pTU390 were PCR amplified from plasmid pTU265 a derivative of pVenus containing CFP using primers BN2278 in combination with BN2275 or BN2276 respectively. Products were restricted with NcoI, ligated and transformed into DH5a. All plasmids were confirmed by Sanger

sequencing (Macrogen). All plasmids used are listed in Supplementary Table 1. Primers used are listed in Supplementary Table 2.

### Creation of strains KD615mCherry-Cas8e and KD635mCherry-Cas8e

Strains were created using lambda red homologous recombination <sup>72</sup>. Plasmid pSC020, containing both Lambda red and the Cre-recombinase, was transformed by electroporation into strains KD615 and KD635. Strains were recovered at 30 °C for 1.5 h and plated on media containing 100 µg/ml ampicillin. Transformants were then grown overnight in liquid medium at 30 °C, with selection, and made competent the following day by inoculating 50 ml with 500 µl of overnight culture. Once the cells reached an OD<sub>600</sub> of 0.2 a final concentration of 0.2% L-Arabinose (Sigma-Aldrich) was added and cells were grown for another 1.5 h and subsequently washed with pre-cooled 10% glycerol.

The mCherry-cas8e G-block (IDT) (Supplementary Table 3) based on the design used in <sup>3</sup> was resuspended with ddH<sub>2</sub>O to a concentration of 50 ng/µl and transformed into the competent cells by mixing 2 µl DNA with 50 µl of cells and recovering at 30 °C for 1.5 h. After recovery cells were plated undiluted with selection for kanamycin and ampicillin. PCR verified colonies were then grown in liquid culture with 1 mM IPTG at 37 °C to promote the loss of the kanamycin resistance cassette and pSC020. Individual colonies were screened for plasmid loss by patching each colony onto three plates containing no antibiotics, only kanamycin and only ampicillin. Colonies exhibiting no resistance were then PCR screened with primers (Supplementary Table 2) BN2204 and BN2205 for the presence of the mCherry-Cascade fusion. Strains were confirmed by Sanger sequencing (Macrogen).

### Growth conditions

All strain and plasmid combinations (Supplementary Table 1) used were grown at 37 °C, shaking at 180 rpm, prior to microscopy. To avoid autofluorescence under the microscope a minimal M9 media was used containing the following supplements; 2% glycerol (Sigma-Aldrich), 1X EZ Supplements (M2104 Teknova), 20 µg/ml uracil (Sigma-Aldrich), 1 mM MgSO<sub>4</sub> (Sigma-Aldrich) and 0.1 mM CaCl<sub>2</sub> (Sigma-Aldrich), from here on called M9 media.

### Microfluidic device

The device used was developed by D.J. Kiviet in the Ackermann lab and has been previously used in the Tans lab <sup>41</sup>. The device contains a main flow channel 23.5  $\mu\text{m}$  high and 200  $\mu\text{m}$  wide that splits into two 100  $\mu\text{m}$  wide flow channels of the same height. Perpendicular to these flow channels are wells with a height of 0.75  $\mu\text{m}$ , widths of 1x 80  $\mu\text{m}$ , 1x 60  $\mu\text{m}$ , 2x 40  $\mu\text{m}$ , 3x 20  $\mu\text{m}$ , 3x 10  $\mu\text{m}$ , 3x 5  $\mu\text{m}$  and depths of 60  $\mu\text{m}$ , 30  $\mu\text{m}$ , 50  $\mu\text{m}$  and 40  $\mu\text{m}$ . These well sizes are repeated 5 times and are the location where the growth of microcolonies occurs during an experiment. The PDMS devices were made by casting them into an epoxy mould, a gift from D.J. Kiviet and the Ackermann lab.

The PDMS device was produced by mixing polymer and curing agent (Sylgard 184 elastomer, Dow Corning) in ratio of 1 mL of curing agent to 7.7 g of polymer. This mixture was poured into the epoxy mould and air bubbles were subsequently removed by use of a desiccator for 30 mins followed by baking at 80 °C for 1 h. After baking the device can be carefully removed from the mould with aid of a scalpel and holes were punched for liquid in-and outlets. For use under the microscope, the PDMS chip was covalently bound to a clean glass coverslip. This was done by treating both the PDMS and glass surface with 5-10 sweeps of a portable laboratory corona device (model BD-20ACV, Electro-Technic Products). After treatment the chip was placed carefully onto the glass slide and gently tapped to facilitate full contact between the PDMS and glass surface. Finally, the device was baked for another 1-2 h at 80°C and stored until the experiment was started.

### Loading and filling of microfluidic wells

Cells were initially grown overnight (for ~ 12 h) at 37 °C, 180 rpm in 10 mL M9 media with antibiotic selection (streptomycin 50  $\mu\text{g}/\text{ml}$ ) for the target plasmid. The following day 500  $\mu\text{l}$  of culture was passaged into fresh M9 medium (with selection for the target plasmid), approximately 3 h before microscope set up, and grown at 37 °C, 180 rpm. After 3 h of growth the cells were pelleted and resuspended in ~ 30  $\mu\text{l}$ .

To begin the experiment 2  $\mu\text{l}$  of 0.01% Tween20 ( $\text{dH}_2\text{O}$ ) solution is slowly pipetted into the selected media lane to allow the removal of air and flow of liquid into the wells perpendicular to the media lane. Following this, 2  $\mu\text{l}$  of concentrated bacterial culture was pipetted slowly into the same lane. Once liquid could be seen exiting at the opposite end of the media lane the syringes containing media (loaded on syringe pumps), the valve controller

and the waste collection flasks were attached to the chip by metal connectors and polyethylene tubing. Media was pumped into the chip at a flow rate of 0.5 mL/h allowing constant supply of nutrients to the cells. The rate of media flow was also important for removal of cells from the top of the well, to allow constant division and long-term tracking of cells located lower within the well.

### Media switches

All experiments were carried out with precise 37 °C temperature control and required the use of 2 different medias. For the first 12 h of the experiment (including loading of the chip) cells were grown in Media 1; M9 supplemented with both anhydrotetracycline (40 ng/ml) and Streptomycin (25 µg/ml) to induce the YFP and select for cells containing the target plasmid respectively. After 12 h of growth in the chip the media was switched via the valve controller (Hamilton, MPV valve positioner) to Media 2; M9 supplemented with anhydrotetracycline (40 ng/ml), 0.1% L-arabinose and 0.1 mM IPTG. This media change allowed removal of the selection for the target plasmid, continued induction of the YFP and induction of the CRISPR-Cas system after filling of the wells.

### Spacer acquisition detection from microfluidic chip output

Over the course of the experiment, the cells that flow out of the wells and subsequently the chip were collected in a sterile Erlenmeyer flask. The cells were then concentrated by centrifuging for 5 min at 2000 g. The supernatant was removed and cell were resuspended in 2 mL of M9 media. Colony PCR was performed 1 µl of culture using primers BN1530 and BN1531 (Supplementary Table 2) and the products were run on a 2% agarose gel at 100 V for 30 mins alongside the 100-1000 bp DNA Ladder (SmartLadder-SF, Eurogentec).

### Imaging and image analysis

For all time-lapse experiments, phase contrast images were acquired at 1 min intervals at a maximum of 2 positions. In experiments with a YFP target plasmid, fluorescent images were taken every 2 mins, with an exposure time of 500 ms. For experiments with a CFP target plasmid and the mCherry-Cascade fusion images were acquired every 4 mins with exposure times of 500 ms and 200 ms respectively. Images were acquired for the entire experiment including the first 12 hrs of growth. Cells were imaged with an inverted microscope (Nikon, TE2000), equipped with 100X oil immersion objective (Nikon, Plan Fluor NA 1.3), automated stage (Märzhäuser, SCAN



IM 120 3 100), high power LED light source with liquid light guide (Sutter, Lambda HPX-L5), GFP, mCherry, CFP and YFP filter set (Chroma, 41017, 49008, 49001 and 49003), computer controlled shutters (Sutter, Lambda 10-3 with SmartShutter), cooled CMOS camera (Hamamatsu, Orca Flash4.0) and an incubation chamber (Solent) allowing temperature control. In order to obtain images with a pixel size of  $0.041 \mu\text{m}$  an additional 1.5X lens was used. The microscope was controlled by MetaMorph software. A series of acquired phase contrast images were analyzed with a custom MATLAB (MathWorks) program, originally based on Schnitzcells software <sup>43</sup>, adapted to allow for automated segmentation of cells growing in a well <sup>41</sup>. Segmentation was inspected and corrected manually where necessary. All segmented cells were then tracked between frames using the pixel overlap between cells allowing the formation of lineage structures <sup>41</sup>.

### **Plasmid loss and clearance time detection using the fluorescent protein production rate**

Before screening for plasmid loss, we detect cell death in lineages by applying a moving average filter to the cellular growth rate. If the cellular growth rate reached zero and did not recover again, the remainder of the fluorescence time series after this point was excluded from the analysis. For each lineage, we computed the fluorescence production rate of the plasmid-encoded fluorophore from a cell's total fluorescence, cell area, cellular growth rate, and the rate of photobleaching of the fluorophore <sup>44</sup>. As there is always some amount of residual fluorescence produced by the cells, we selected an appropriate threshold for plasmid loss detection from the upper values of the distribution of production rates of plasmid-free cells. To detect plasmid loss in individual lineages we applied a moving average filter to the fluorescence production rate and detected the first instance of the production rate reaching a value below the threshold. This plasmid loss time (PLT) can be seen as an upper bound estimate, as some processes (transcription, translation, fluorophore maturation) still carry on for some time after the last plasmid has been cleared but could not be measured in our set up. The onset of the clearance time (CT), which signifies the start of the destruction of all plasmids through interference and ends at the plasmid loss time (PLT), is difficult to detect in individual lineages due to the naturally occurring fluctuations in the fluorescence production rate. To determine this quantity, we align all plasmid loss lineages at the PLT and compute the average trend. The CT per experimental condition is approximated as the duration from the point where the average production rate starts to decrease until the PLT.



### Sister and cousin statistics

For each lineage that lost the plasmid, we wanted to compare the probability of loss in an unrelated cell and in a related cell. For related cells we counted the frequency of loss and non-loss in sister and cousin cells of the loss cell, but only if the sister or cousin divided (contained a complete cell cycle). For unrelated cells we counted the total number of loss events ( $i$ ) that occurred throughout the cell cycle of the related cell. For each loss event we counted how many cells ( $c_i$ ) still contained the plasmid up to that point. The probability of plasmid loss happening in an unrelated cell during the lifecycle of the related cell was subsequently calculated recursively using the following equations:

$$p_0 = 0$$

$$p_i = \frac{1-p_{i-1}}{c_i} + p_{i-1}$$

Where  $p_i$  is the probability of loss occurring within an unrelated cell given  $i$  plasmid loss events occurred within the cell cycle of the related cell and  $c_i$  stands for the number of cells still containing the plasmid at the same time as the  $i$ -th plasmid loss event.

### Cascade copy number determination

The control strain KD614 mCherry-Cas8e containing plasmid pTU265 (Supplementary Table 1) was prepared and loaded into the microfluidic chip as above. After 12 h a sterile tube was connected to the waste tubing and output from the chip was collected for 30 mins. The media was then switched to induce Cascade. Approximately 5 h after induction when Cascade levels are considered to be stabilized the output from the chip was again collected for 30 mins. To improve counting, cells were subsequently fixed with 2.5% paraformaldehyde solution at 22 °C for 45 mins<sup>73</sup>. Slides were cleaned by sonication in subsequent steps with MilliQ, acetone and KOH (1M). Next, 1 % agarose pads containing the M9 medium were prepared and hardened between two slides within 20 mins of measuring to prevent desiccation. The fixed cells were then spun down and resuspended in 5 µl of which 1 µl was pipetted onto a pre-prepared agarose pad.

The cells were imaged using a TIRF microscope (Olympus IX81; Andor Ixon X3 DU897 EM-CCD camera) using a high power 561 nm laser, which quickly bleached most mCherry molecules within a couple of frames. Intensity of

single molecules were measured with Thunderstorm starting from the thirtieth frame <sup>74</sup>. The total cell fluorescence was measured by segmenting the cells from the phase contrast image and sum fluorescence counts of all cell pixels (with background subtracted). The copy number was calculated by dividing the total cell fluorescence in the first frame by the average fluorescence intensity of the single molecules. We could then calculate the Cascade concentration  $\sim 200$  Cascade molecules/ $\mu\text{m}^2$  by dividing the population average of the mean summed RFP per cell by this copy number, which was applied to the cells in our time-lapse data.

### Model implementation

Stochastic simulations were performed using the adapted *Extrande* algorithm <sup>75</sup> implemented in C++. Each data point in Fig. 5i-j and Fig. 6a-c was obtained from 100 simulated experiments of up to  $10^4$  min. The population size of each simulation was fixed at 100 cells. See Supplementary Methods for model details and parameters.

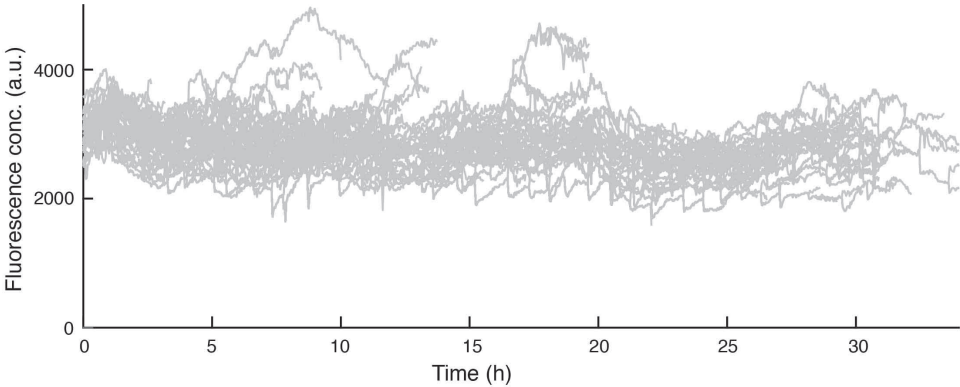
## 5.5 Acknowledgements

The authors would like to thank Martijn Wehrens for his help and advice throughout the project and all of the members of the Tans and the Brouns groups for input during group discussions. We acknowledge the group of Konstantin Severinov for the gift of strains KD615 and KD635. R.E.M. is supported by the Frontiers of Nanoscience (NanoFront) program from NWO/Ministry of Education (OCW). C.F. received funding from FET-Open research and innovation actions grant under the European Union's Horizon 2020 research and innovation programme (CyGenTiG; grant agreement 801041). Work in the group of S.J.T is supported by The Netherlands Organization for Scientific Research (NOW). S.J.J.B. has received funding from the European Research Council (ERC) CoG under the European Union's Horizon 2020 research and innovation programme (grant agreement No. [101003229]) and from the Netherlands Organisation for Scientific Research (NWO VICI; VI.C.182.027).

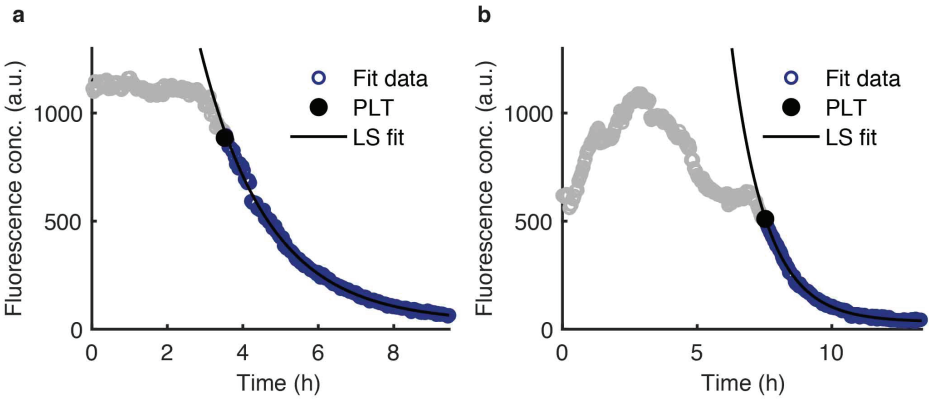
### Author contributions

R.E.M., S.J.J.B. and S.J.T. and conceived the project, R.E.M., J.V., J.V.L. and V.K. performed the experiments. E.M.K., J.V., R.E.M. and F.B. analysed the data. E.M.K. and C.F. performed the modelling, R.E.M., E.M.K., S.J.J.B., S.J.T., and C.F., wrote the manuscript with input from all authors.

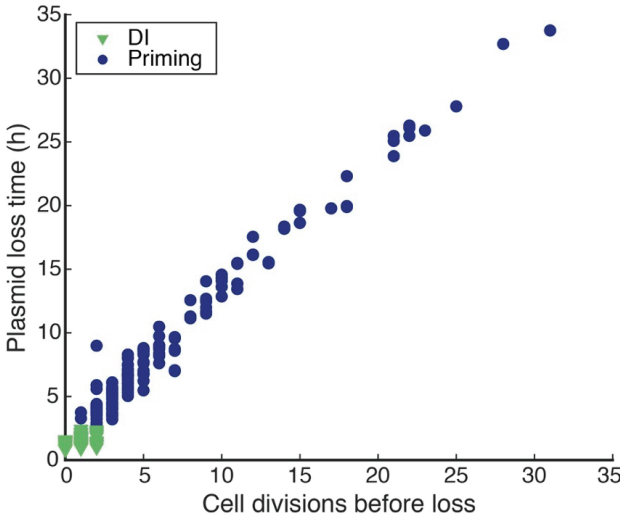
5.6 Supplementary Information



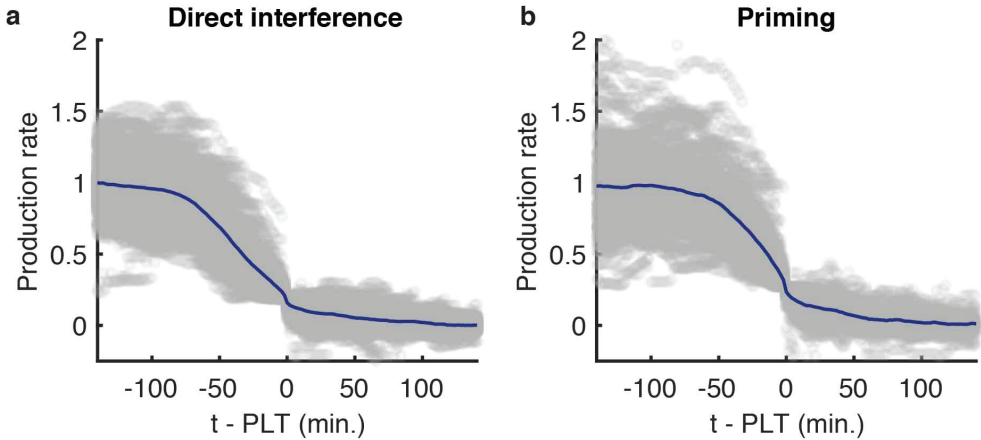
**Supplementary Figure 5.1: Plasmid loss is CRISPR-dependent.** The YFP fluorescence traces in arbitrary units (a.u.) of the WT strain harboring pControl a plasmid with no target for the CRISPR-Cas system. Time-lapse imaging was carried out for 35 hours post induction of the *cas* genes, and showed no plasmid loss.



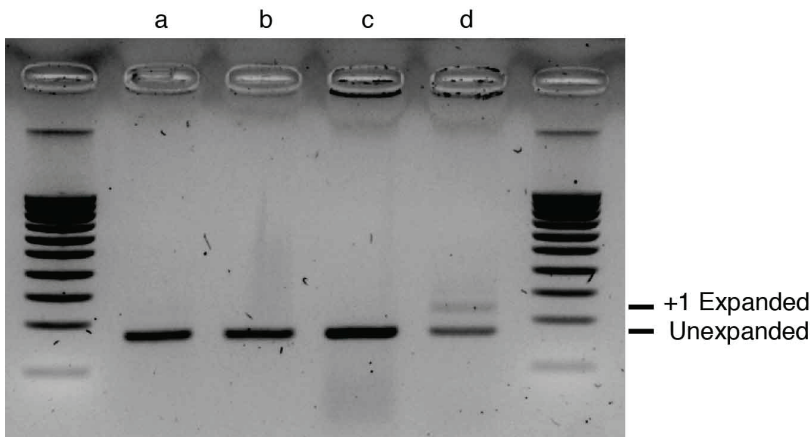
**Supplementary Figure 5.2: Decay of YFP fluorescence in both direct interference and priming follows exponential decay.** The fluorescence concentration (open circles) of (a) direct interference and (b) priming lineages can be described by exponential decay. This was evaluated by performing a least-squares (LS) fit of the fluorescence concentration data (purple open circles) after the moment of plasmid loss (PLT, black circle) to an exponential curve (LS fit, black line).



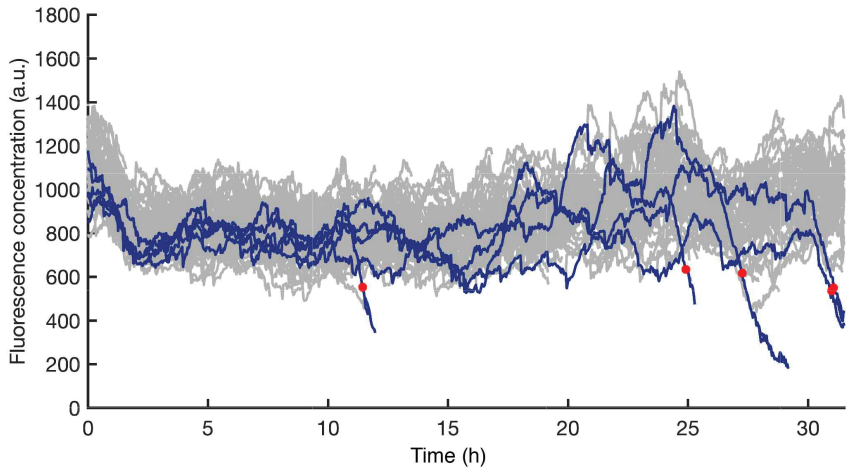
**Supplementary Figure 5.3: Cell divisions before plasmid loss highly correlates with plasmid loss time.** The number of cell divisions from the moment of induction until plasmid loss are plotted against the PLT in hours. Both consensus target clearance by direct interference (green) and mutant target clearance by priming (blue) are shown.



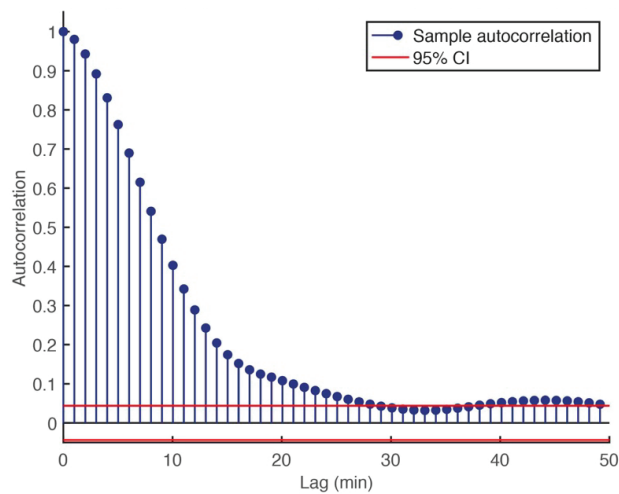
**Supplementary Figure 5.4: Plasmid loss during Direct interference and primed interference processes occur on a comparable timescale.** All production rate traces (grey) starting from 140 minutes prior to the detected plasmid loss time PLT from (a) direct interference and (b) priming were aligned at the PLT ( $t\text{-PLT}=0$ ) and the average trend (navy) normalized for comparison. From the average trend, we estimate the clearance time (CT), time taken from the initiation of plasmid clearance until elimination of all copies, to be in the order of 60 minutes for both direct interference and priming from the onset of the production rate decrease.



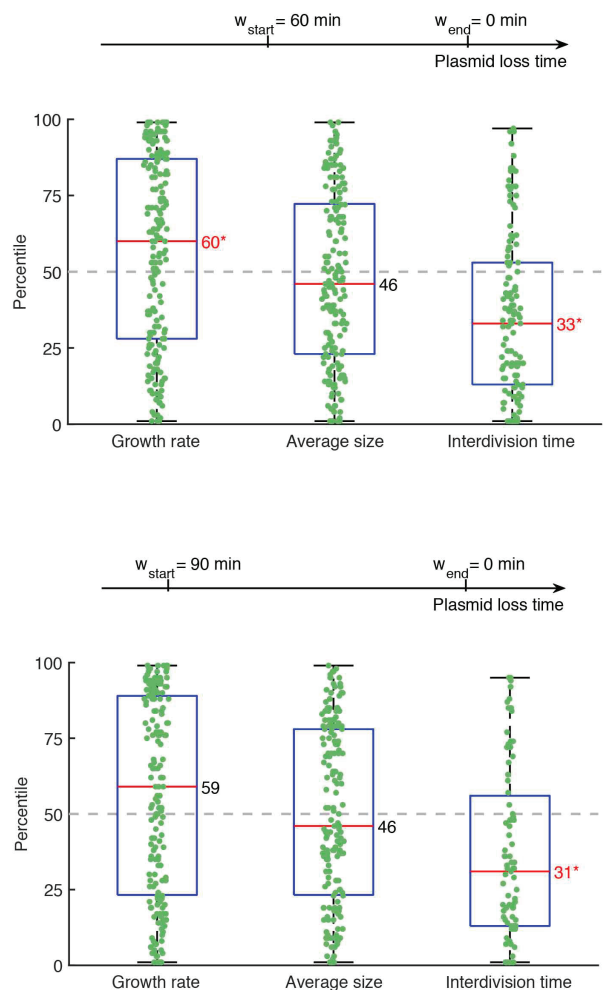
**Supplementary Figure 5.5: Spacer acquisition was only seen in the WT strain in the presence of a mutated PAM triggering priming.** Cells from the chip output were collected in a flask for each experiment and the CRISPR arrays were screened for expansion due to spacer acquisition by PCR amplification using primers BN1530 + BN1531 (Supplementary Table 2). The gel shows PCR amplified CRISPR arrays from each experiment **a**, WT + pControl (Control) **b**,  $\Delta cas1,2$  + pTarget (direct interference) **c**,  $\Delta cas1,2$  + pMutant **d**, WT + pMutant (priming). The presence of a larger band indicates array expansion and therefore successful spacer acquisition.



**Supplementary Figure 5.6: In the absence of Cas1 and Cas2 clearance of a target with a non-consensus PAM mutant occurs rarely.** The YFP fluorescence of the  $\Delta cas1,2$  strain containing pMutant was imaged for 34 hours after induction. Lineages that were able to clear the plasmid are highlighted in blue with the red dot indicating the moment of detection. 1.4% of lineages (5 unique events, red dot) cleared the plasmid.

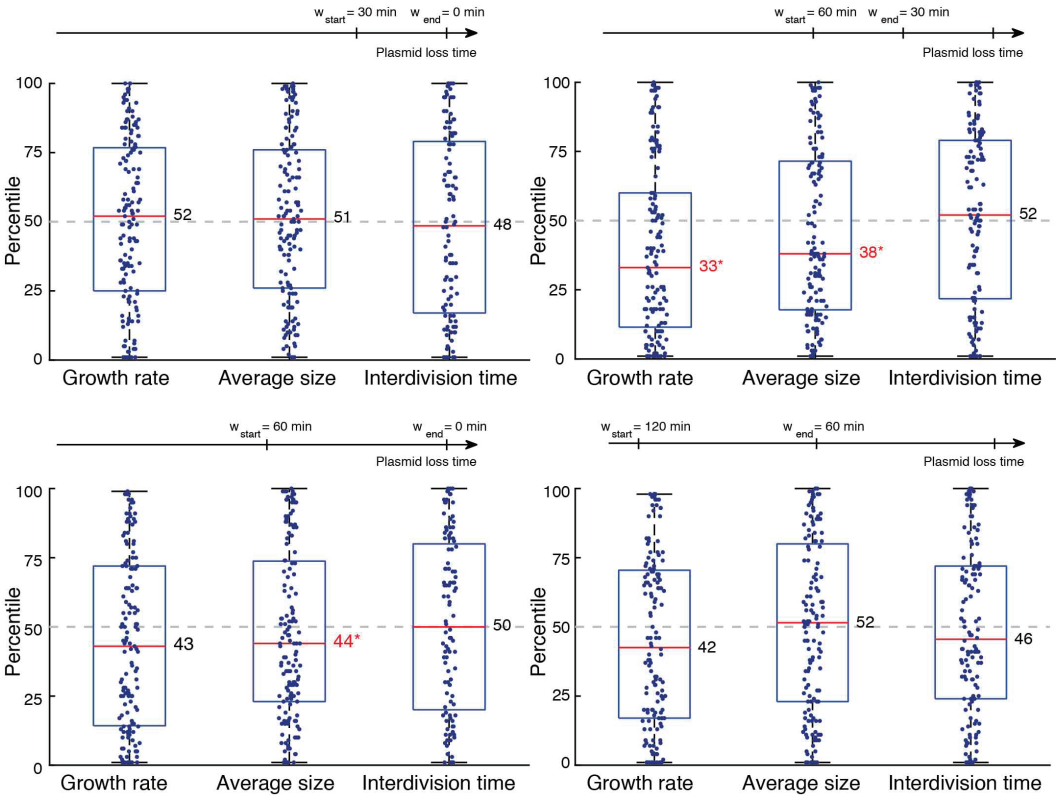


**Supplementary Figure 5. 7: Autocorrelation time of cellular growth rate.** The autocorrelation time was calculated for the cellular growth rate of the WT strain containing pControl by averaging the autocorrelation of cell growth as a function of time in individual lineages. After 10 mins the autocorrelation of cellular growth has decreased to 0.4. After approximately 30 minutes the autocorrelation has decreased to zero, as indicated by 95% confidence intervals (red lines).

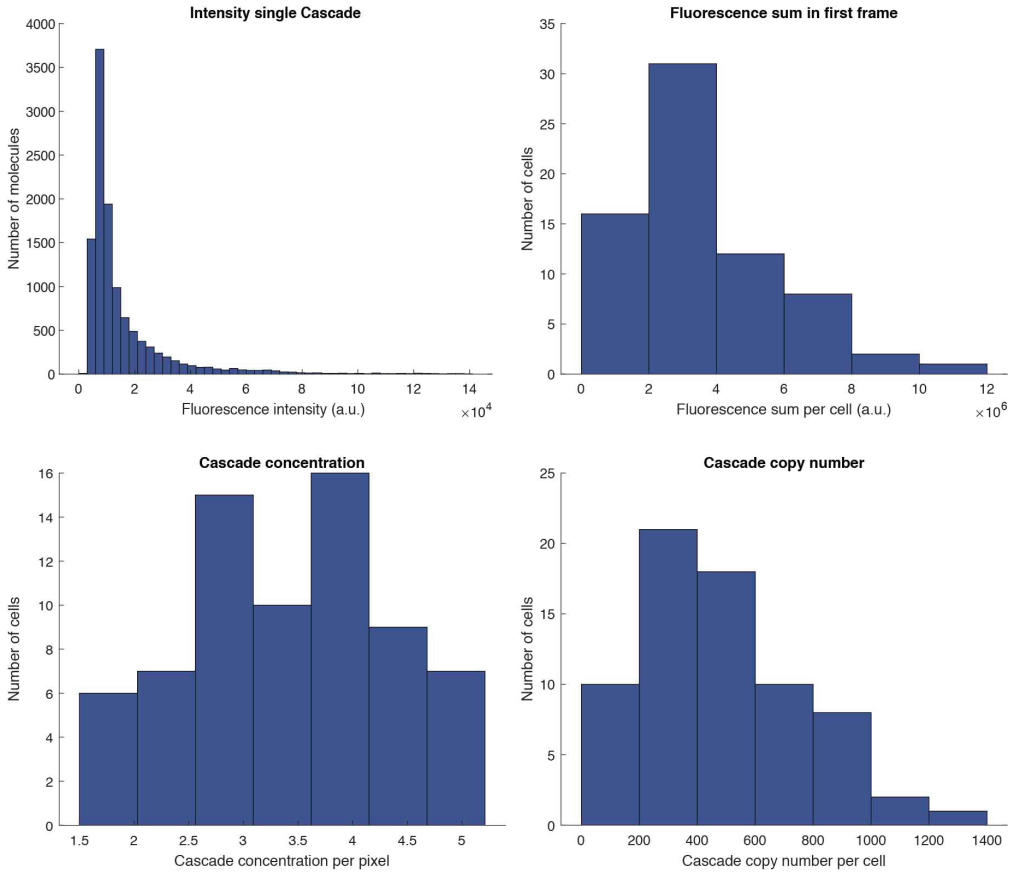


**Supplementary Figure 5.8: Growth rate, cell size and interdivision time of direct interference with different lookback windows.** Boxplots of growth rate, average cell size and interdivision time presented as the percentile rankings of all plasmid loss lineages (green) that cleared a known target via direct interference. The cell feature of interest (e.g. growth rate) was averaged over a lookback window chosen in relation to the time from plasmid loss of the lineage of interest. The same cell feature was then averaged for all non-loss lineages in the population at that same moment. The cell feature of interest was then ranked amongst the non-loss population as a percentile. We considered lookback windows of 60 minutes prior to plasmid loss (top) and 90 minutes prior to plasmid loss (bottom). The median percentile ranking of loss lineages is indicated by a red line and black text, categories in which this value was significantly different from a ranking in the 50<sup>th</sup> percentile (p-value<0.05) are indicated in red text followed by an asterisk.

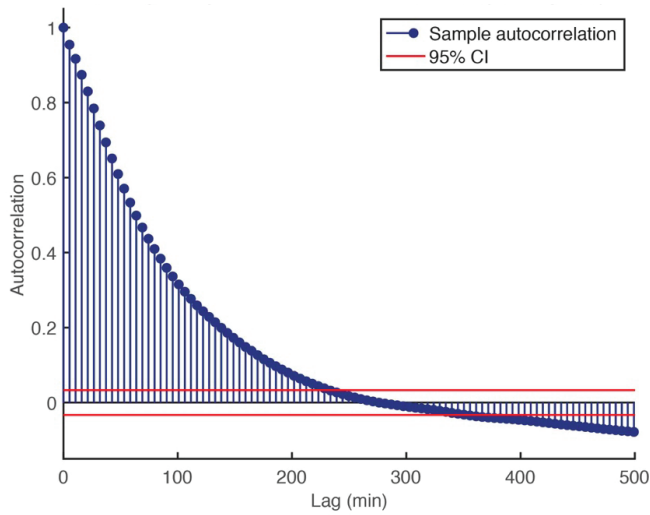




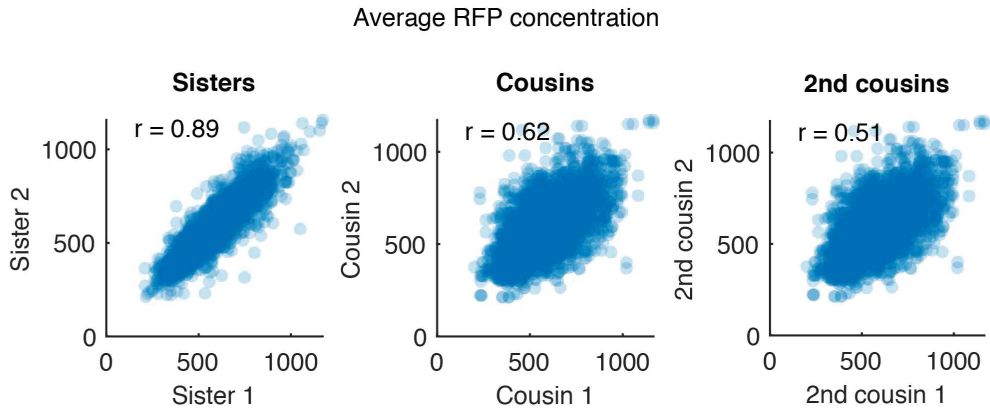
**Supplementary Figure 5.9: Growth rate, cell size and interdivision time of priming with different lookback windows.** Boxplots of growth rate, average cell size and interdivision time presented as the percentile rankings of all plasmid loss lineages (navy) that cleared a known target via priming. The cell feature of interest (e.g. growth rate) was averaged over a lookback window chosen in relation to the time from plasmid loss of the lineage of interest. The same cell feature was then averaged for all non-loss lineages in the population at that same moment. The cell feature of interest was then ranked amongst the non-loss population as a percentile. We considered a range of lookback windows. The median percentile ranking of loss lineages is indicated by a red line and black text, categories in which this value was significantly different from a ranking in the 50<sup>th</sup> percentile (p-value < 0.05) are indicated in red text followed by an asterisk.



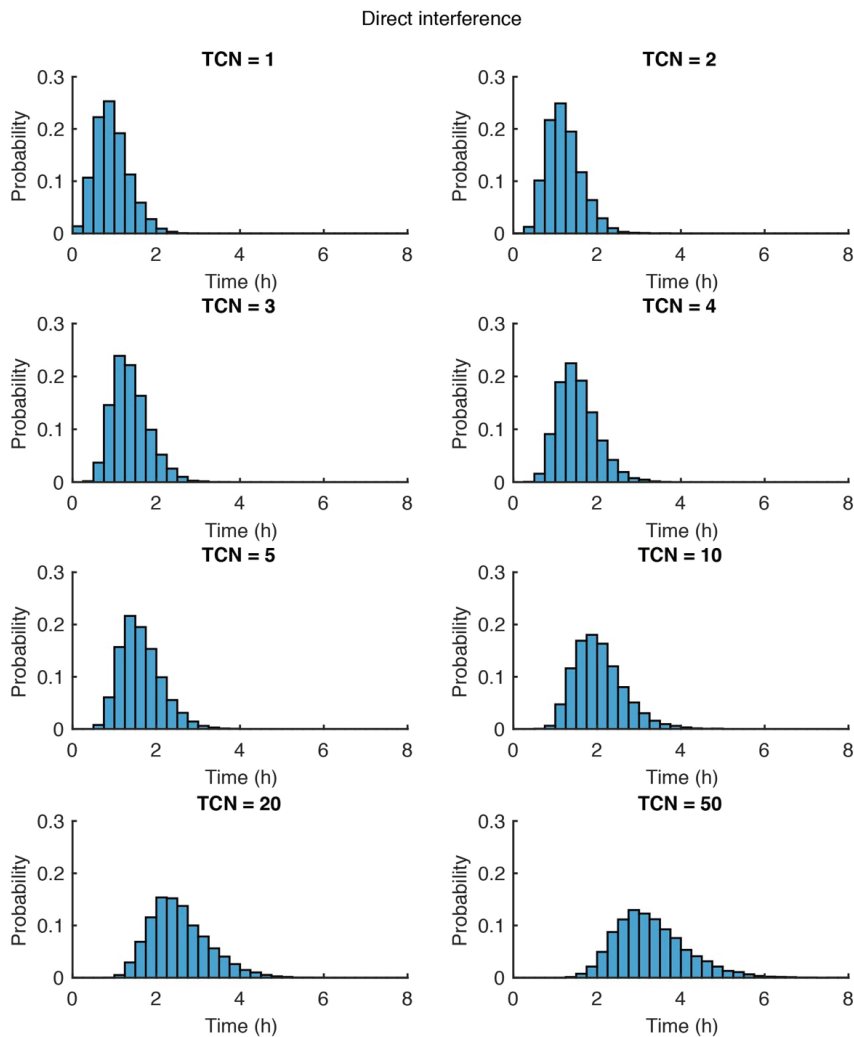
**Supplementary Figure 5.10: Cascade copy number determination.** **a**, The fluorescence sum (RFP) of each cell in the first frame was determined. **b**, The RFP molecules were then bleached until it was possible to determine the fluorescence intensity of a single molecule (representing a single Cascade). **c**, The Cascade copy number per cell was then determined by dividing the average fluorescence sum by the average intensity of a single Cascade molecule. **d**, The Cascade concentration per pixel was determined by dividing the fluorescence sum by the area of the cell in pixels.



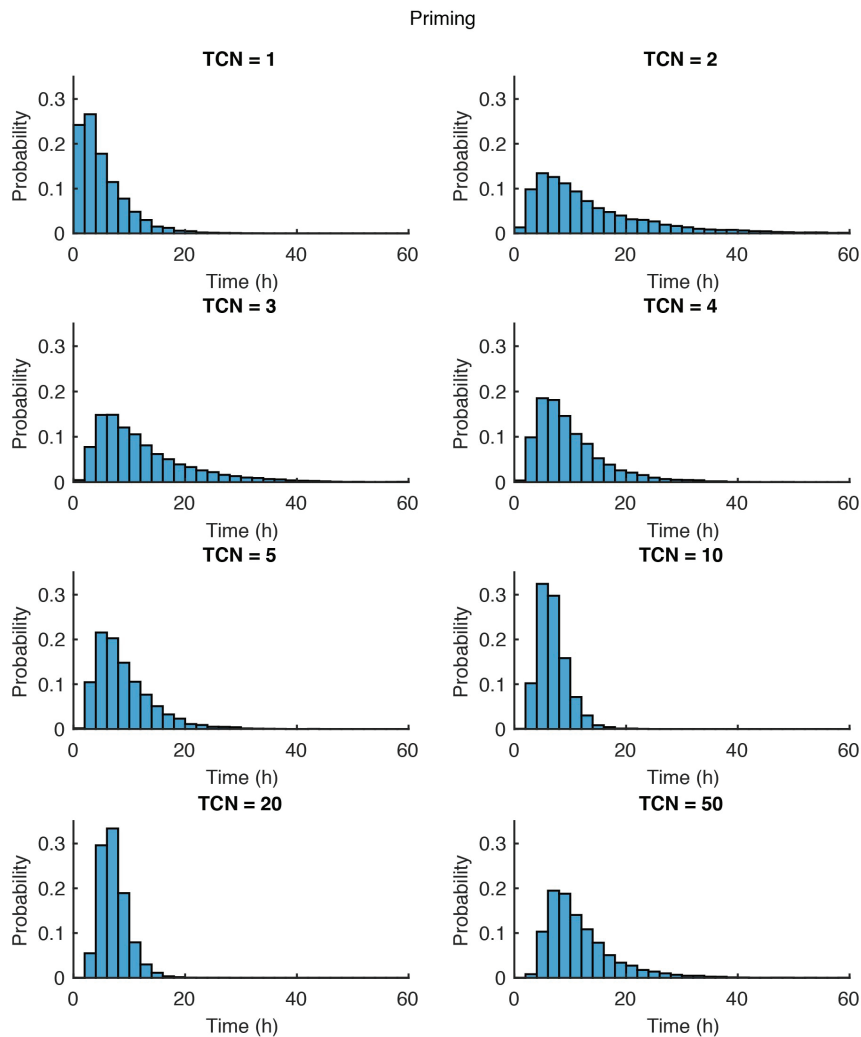
**Supplementary Figure 5.11: Autocorrelation of RFP (Cascade) concentration.** The autocorrelation was calculated by averaging over the autocorrelation of RFP concentration of the WT-mCherry strain in individual lineages. After approximately 200 minutes the autocorrelation has decreased to zero, as indicated by 95% confidence intervals (red lines). The long decay time of the autocorrelation function indicates that Cascade protein levels fluctuate on a time scale longer than the cell cycle.



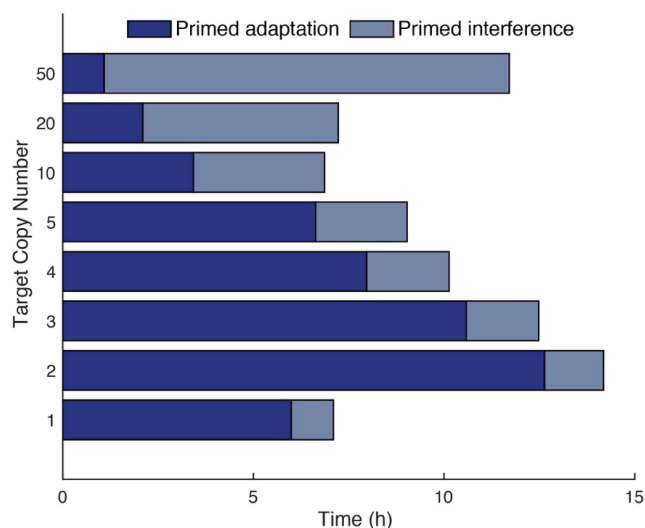
**Supplementary Figure 5.12: Correlation of RFP levels between cells related as sisters, cousins, and second cousins.** The levels of RFP (Cascade) are strongly correlated between sister, cousins, and second cousins. The correlation coefficient  $r$  decreases as the cells become less closely related.



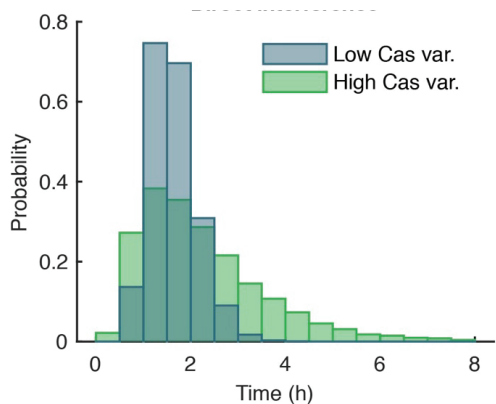
**Supplementary Figure 5.13: Distribution of target loss times resulting from simulations of the direct interference condition for average target copy numbers (TCN) per cell ranging from 1-50**



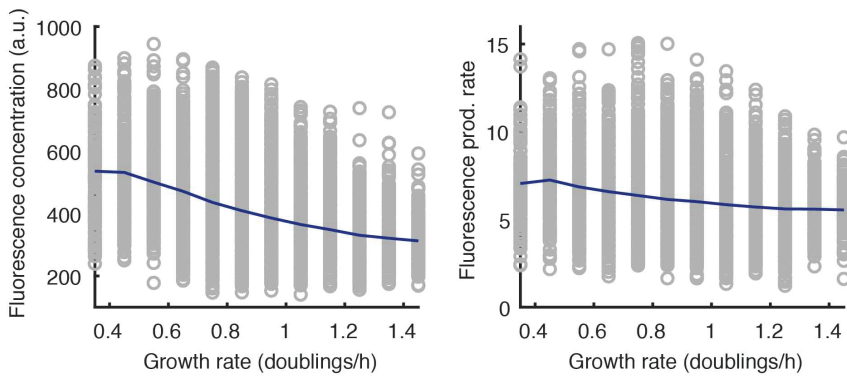
**Supplementary Figure 5.14: Distribution of target loss times resulting from simulations of the priming condition for average target copy numbers (TCN) per cell ranging from 1-50**



**Supplementary Figure 5.15: Target loss time as a function of the target copy number (TCN) as computed from simulated trajectories by the agent-based model for the priming condition.** Bar charts representing the time spent on primed adaptation (navy) and primed interference (grey) for cells clearing targets through priming with an average plasmid copy number ranging from 1-50.



**Supplementary Figure 5.16: Distribution of plasmid loss times in direct interference for high and low variability in Cascade concentration.** Target loss time distribution for two different levels of Cascade concentration variability resulting from simulated trajectories of the direct interference condition. At low variability (blue) Cascade proteins are produced in frequent, small bursts, whereas at high variability (green) proteins are synthesized more sporadically in large bursts (100-fold increase), keeping average Cascade concentration constant. The variability of PLT interference times for high Cascade variability increases as compared to low Cascade variability.



**Supplementary Figure 5.17: Slower growing cells have higher RFP (Cascade) concentrations.** Cascade concentration (left) and Cascade production rate (right) show an inverse relationship with cellular growth rate (grey circles), revealing slower growing cell on average (navy line) have a higher concentration of Cascade.

Supplementary Table 5.1: Strains and Plasmids used in this study

Strain	Description	Source
KD615	<i>E. coli</i> K12, F+, araBp8- <i>cse1</i> , lacUV- <i>cas3</i> , CRISPR I R-SP8-R, ΔCRISPR II+III	(Musharova et al., 2019)
KD635	<i>E. coli</i> K12, F+, araBp8- <i>cse1</i> , lacUV- <i>cas3</i> , CRISPR I R-SP8-R, Δ <i>cas1</i> ,2, ΔCRISPR II+III	(Musharova et al., 2019)
KD615mCherry-Cas8e	<i>E. coli</i> K12, F+, araBp8- <i>cse1</i> , lacUV- <i>cas3</i> , CRISPR I R-SP8-R, ΔCRISPR II+III, mCherry- <i>cas8e</i>	This study
KD634mCherry-Cas8e	<i>E. coli</i> K12, F+, araBp8- <i>cse1</i> , lacUV- <i>cas3</i> , CRISPR I R-SP8-R, Δ <i>cas1</i> ,2, ΔCRSIPR II+III, mCherry- <i>cas8e</i>	This study
Plasmid		
pTarget (pTU166)	pSC101, StrepR, TetR mVenus PS8 flanked by 'CTT' PAM	This work
pMutant (pTU190)	pSC101, StrepR, TetR mVenus PS8 flanked by 'CGT' PAM	This work
pControl (pTU193)	pSC101 ori, StrepR, TetR-mVenus, no target	This work
pVenus	pSC101 ori, KanR, mVenus-YFP	Bokinsky lab
pCDFDuet-1	pCloDF13 ori, StrepR	Lab Collection
pTU265	pSC101, StrepR, TetR-Cerulean, No target	This work
pTU389	pSC101, StrepR, TetR-Cerulean, PS8 flanked by 'CGT' PAM	This work
pTU390	pSC101, StrepR, TetR-Cerulean, PS8 flanked by 'CTT' PAM	This work
pSC020	Derivative of pKD46 containing Lamda red and the Cre-recombinase	This work



Supplementary Table 5.2: Oligonucleotides used in this study

Name	Description	Sequence
BN831	Streptomycin resistance and PS8 insertion into pVenus, Fw	TTTT <u>GGTACC</u> TTATTTGCCGACTACCTTGGTGATCTC
BN832	Streptomycin resistance and PS8 insertion into pVenus, Rv	TTTTAAGCTTAAAAGTGCCACTTGCGGAGACCCGGTCG TCAGCTTACATTCAAATATGTATCCGCTC
BN833	Backbone amplification pVenus, Rv	TTTTGGTACCGGACTCTGGGGTTCGAG
BN834	Backbone amplification pVenus, Fw	TTTTAAGCTTCGAAACGATCCTCATCCTG
BN891	Streptomycin resistance insertion (no target), Rv	TTTTAAGCTTACATTCAAATATGTATCCGCTC
BN911	Modify pTU166 PAM universal, Rv	TTTTGTCGACACATTCAAATATGTATCCGCTCATGAGAC
BN912	Modify pTU166 CTT PAM to CGT	TTTTGTCGACACGCTGACGACCGGGTC
BN1494	To amplify pTU193 Backbone minus yfp, Rv	TTTCTCGAGTAAGGATCTCCAGGCATC
BN1495	To amplify pTU193 Backbone minus yfp, Fw	TTTCTCGAGTAAGGATCTCCAGGCATC
BN1507	To amplify Cerulean from p15A, Fw	TTTGAATTCCAGAATTCAAAAGATCTAGGAGG
BN1508	To amplify Cerulean from p15A, Rv	TTTCTCGAGAGGATCCTTATTTATACAGCTCATCC
BN1513	To check Cerulean insertion and confirm pTU265 by sequence, Fw	CCTCATTAAGCAGCTCTAATGCGCTG
BN1530	To screen for CRISPR array amplification, Fw	GGTTTGAAAATGGGAGCTCG
BN1531	To screen for CRISPR array amplification, Rv	GTTACATTAAGGTTGGTGGGTTG
BN2202	To amplify mCherry-Cas8e gblock, Fw	ACAGAATCTGGATGGATGG
BN2203	To amplify mCherry-Cas8e gblock, Rv	CTGATCTCTACTGCAGTATAGC
BN2204	Screen for mCherry-cas8e knock in, Fw	GCGCTTGCACTTAATCGC
BN2205	Screen for mCherry-cas8e knock in, Rv	ACCAGCAGTGCTAAAGCG
BN2206	Screen for mCherry-cas8e knock in, Fw	CTTTCCGTCCGGTGTGTCAGG
BN2275	Insertion PS8 CGT PAM into pTU265, Fw	TTTCCATGGGAAAAGTGCCACTTGCGGAGACCCGGTCG TCAGCGTACATTCAAATATGTATCCGCTCAT

BN2276	Insertion PS8 CTT PAM into pTU265, Fw	TTT <b>CCATGG</b> AAAAAGTGCCACTTGCGGAGACCCGGTCGTCAG <b>CTT</b> ACATTCAAATATGTATCCGCTCAT
BN2278	Insertion of PS8 and PAM universal, Rv	TTT <b>CCATGG</b> CCTCATCCTGTCTCTTGATC

\*PAM sequences are indicated in bold sequences, underlined sequences indicate restriction sites

Supplementary Table 5.3: Synthetic DNA G-block used in this study

Name	Sequence
<i>cas8e</i> -mCherry insert	ACAGAATCTGGATGGATGGGTCTGGCAGGGTAACAGTATTGTTATTACCTATACAGGGGATGAAGGGATGACCAGAGTCATCCCTGCAAATCCCAAATAACCTGGAGCTGCAGATACCGTTTCGTATAATGTATGCTATACGAAGTTATAGATCTCTATTTGTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGCCATATTCAACGGGAAACGTCTTGCTCTAGGCCGCGATTAAATTCCAACATGGATGCTGATTATATGGGTAATAATGGGCTCGCGATAATGTCGGGCAATCAGGTGCGACAATCTATCGATTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAACTGGCTGACGGAATTTATGCTCTTCCGACCATCAAGCATTTTATCCGTACTCCTGATGACGCATGGTTACTCACACTGCGATCCCCGGGAAAAACAGCATTCCAGGTATTAGAAGAATATCCTGATTCAGGTGAAAAATATTGTTGATGCGCTGGCAGTGTTCTCGCCGGTTGCATTCCGATTCCTGTTTGTAATTGTCCTTTAACAGCGACCGCGTATTCGTCTCGCTCAGGCGCAATCACGAATGAATAACGGTTTGTTGATGCGAGTGATTTTGATGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGAAATGCACAACTTTTGCCATTCTACCGGATTCAGTCGTCACCTCATGGTGATTTCTCACTTGATAACCTATTTTGACGAGGGGAAATTAATAGTTGTATTGATGTTGGACGAGTCGGAATCGCAGACCGATACCAGGATCTTGCCATCCTATGGAAGTGCCTCGGTGAGTTTTCTCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAATTGCAGTTTCATTTGATGCTCGATGAGTTTTTCTAAGTCGACATAACTTCGTAATATGTATGCTATACGAACGGTAGAAAATTGCAATGCATCTGCCGAATGCCGTGTGGACGTAAGCGTGAACGTCAGGATCACGTTTCCCCGACCCGCTGGCATGTCAACAATACGGGAGAACACCTGTACCGCCTCGTTCGCCGCGCCACCATAAATCACCGCACCGTTCATCAGTACTTTCAGATAACACATCG

## 5.7 Supplementary Methods

### 5.7.1 Master Equation description of the probability of plasmid loss

In order to test whether the distribution of the target clearance times by direct interference can be reproduced by a simple one-step process, we consider a model using a compound probability for binding of Cascade to the target and subsequent target removal from the system. In bacteria the number of targets is subject to maintenance which delays the removal of  $M_0$  targets. For sake of simplicity we ignore this additional step, which has the advantage that the number of unknown parameters is kept to an absolute minimum. Because direct interference is a fast process, one can assume that target maintenance does not have a strong effect on the clearance time distribution. The Cascade number is not constant, but rather Cascade production is induced at the beginning of the experiment. This simplified model only depends on five parameters: the delay after induction for production of Cascade  $\tau_c$ , the Cascade production rate  $\sigma$ , the turn-over rate of Cascade  $\lambda$ , the number of targets per cell  $M$ , and the probability of a target removal event  $P_d$ . The number of targets in individual cells will be in general stochastic, however due to target maintenance one can assume that this distribution will be quite narrow. For this reason, we set  $M_0 = 5$ <sup>76</sup>.

The time dependent Cascade copy number is modelled as a production-degradation process with a delay  $\tau_c$ , and zero initial amount of Cascade: The bulk mean  $\mu(t)$  is given by:

$$\mu(t) = \frac{\sigma}{\lambda} \theta(t - \tau_c) \left(1 - e^{-\lambda(t - \tau_c)}\right).$$

By fitting this equation to Cascade concentration data for the bulk mean (**Fig.5.4b** of the main text), we estimate:  $\tau_c = 34$  min,  $\sigma = 3$  min<sup>-1</sup>, and  $\lambda = 0.0061$  min<sup>-1</sup> to obtain an average copy number of almost 500 Cascades per cell at steady state.

The removal of  $M_0$  targets from the system is a First-Passage-Time problem. We formulate the simple Master Equation (ME) for the conditional probability  $P_M(t)$  to find  $M$  targets in a cell at a given time  $t$ :

$$\frac{dP_M(t)}{dt} = \mu(t)p_d(M+1)P_{M+1} - \mu(t)p_d M P_M,$$

where  $P_d$  is the compound probability that within the time interval  $\Delta t$  a Cascade molecule binds to a target and the target is subsequently removed from the system.

To obtain the First-Passage-Time distribution we need to determine the survival probability  $S$  to find at least one target, which is simply given by  $S = 1 - P_0$ .  $P_0$  is obtained by solving the above ME with the initial condition  $P_M(t=0) = \delta_{M M_0}$ :

$$P_0(t|M_0) = \left[ 1 - e^{-p_d \int_0^t \mu(t') dt'} \right]^{M_0}.$$

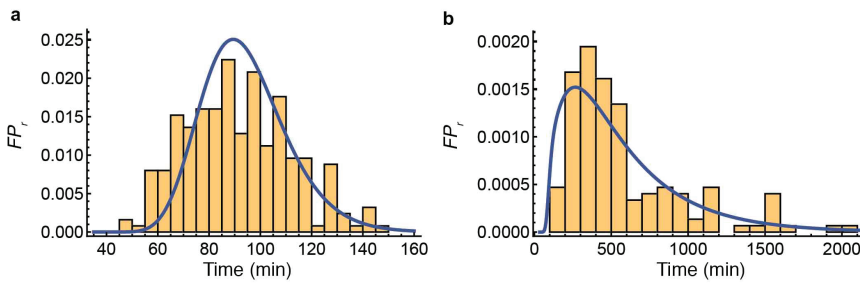
$P_0(t|M_0) = 0$  for  $t < \tau_c$  and because the state  $M = 0$  is naturally an adsorbing boundary we readily find  $\lim_{t \rightarrow \infty} P_0(t|M_0) = 1$ . The First-Passage-Time distribution  $FP_r(t|M_0)$  for target removal is given by the  $FP_r = -dS/dt = dP_0/dt$

$$FP_r(t|M_0) = M_0 p_d \mu(t) \left[ 1 - e^{-p_d \int_0^t \mu(t') dt'} \right]^{M_0-1}.$$

Fitting this distribution to the empirical data (**Fig. 2d** of the main text) gives rise to  $P_d = 4.4 \times 10^{-4} \text{ min}^{-1}$ . The average target removal time  $\tau$  is given by:

$$\tau = \int_0^{\infty} t' FP_r(t'|M_0) dt'.$$

Using the estimates for  $P_d$ ,  $\sigma$ ,  $\lambda$ ,  $\tau_c$  and  $M_0 = 5$  we obtain  $\tau \approx 94 \text{ min}$ . The fit of  $FP_r$  to the data can be seen in **Supplementary Methods Fig. 5.1a**.



**Supplementary Methods Figure 5.1: The fits of the simple one or two-step model to the data. a**, fit of  $FP_r$  (solid line) to the target removal time in the case of direct interference. **b**, fit of  $FP_r$  (solid line) to the target removal time in the case of priming.

The simplified model yields a decent fit to the direct interference data. What about the target clearance during priming? To investigate whether this can be conceptually understood by a two-step process, first spacer acquisition subsequently followed by primed interference, we condition  $FP_r$  on the time  $\tau_p$  needed for spacer acquisition.

$$FP_r(t|M_0, \tau_p) = M_0 p_d \theta(t - \tau_p) \mu(t) \left[ 1 - e^{-p_d \int_{\tau_p}^t \mu(t') dt'} \right]^{M_0 - 1}.$$

The rationale behind this is that a Cascade molecule needs to bind to a target to produce the prespacers necessary for spacer acquisition before primed interference can happen. It follows  $FP_r(t|M_0, \tau_p) = 0$  for  $t < \tau_p$ . Note that  $\tau_p \geq \tau_d$ , since in the absence of Cascade the probability of spacer acquisition is negligibly small. The distribution for  $\tau_p$  is given by the First-Passage-Time distribution for the passage  $M_0 \rightarrow M_0 - 1$ :  $FP_p = -d P_{M_0} / dt$ :

$$FP_p(\tau_p|M_0) = M_0 p_p \mu(\tau_p) e^{-M_0 p_p \int_0^{\tau_p} \mu(t') dt'},$$

Where  $p_p$  is the compound probability that within the time interval  $\Delta t$  one Cascade binds to a target, pre-spacers are produced and a spacer is integrated.

The distribution for the target removal times is given by:

$$FP_r(t|M_0) = \int_0^\infty FP_r(t|M_0 - 1, \tau_p) FP_p(\tau_p|M_0) d\tau_p.$$

The integral cannot be done analytically. Fitting  $FP_r(t|M_0)$  to the experimentally obtained data for the distribution of target loss times during priming (**Fig. 5.2e** of the main text) yields  $p_p = 10^{-6} \text{ min}^{-1}$ . The fit of  $FP_r$  to the data can be seen in **Supplementary Methods Fig. 5.1b**.

### 5.7.2 An agent-based model for stochastic biochemical kinetics of cell populations in microfluidic wells

Although a highly simplified description of our system, the results from the ME description show that the Cascade copy number is an important determinant in creating the variability in the PLT distribution in the case of direct interference. For priming, the distribution could be replicated by considering the process as the result of two subsequent steps, of which the spacer acquisition process creates the wide PLT distribution. However, this model of primed adaptation is highly simplified and does not give any mechanistic insight into the process of adaptation and interference in a

growing cell population. To better understand how cell-to-cell variability and population dynamics affect CRISPR-Cas defense, we have developed a stochastic, agent-based simulation framework to analyse the kinetics of spacer acquisition and target loss. An agent-based approach allows us to keep track of the biochemical composition of individual cells in a growing population, as well as the inheritance of molecules and other cellular features in lineages. In this type of model, each cell is an agent, and there is no interaction between cells. For computational efficiency and to emulate the experimental set-up, the size of the cell population is kept constant. Results for this type of set-up, where the population size is constant, are identical as for a population experiencing exponential population growth, as long as the population size is sufficiently large (100-1000 cells) <sup>77</sup>. The intracellular reactions are governed by stochastic reaction kinetics which can be described by the Chemical Master Equation (CME). As an exact solution to the CME exists only for a handful of simple reaction networks, we use the stochastic simulation algorithm (SSA) <sup>78</sup>, which provides trajectories which are consistent with the CME provided the rate constants are time-independent. When a reaction involves more than one molecular species, the propensity for this reaction to take place in some small time interval depends on the cellular volume. In our application, we are dealing with cells that are continuously in the exponential growth phase which violates the assumption of the SSA of constant propensities between reaction events. For this reason, we use the *Extrande* extension by Voliotis *et al.*, which allows us to efficiently simulate the reaction network containing time-dependent propensities <sup>75</sup>.

### 5.7.3 Model assumptions

Since the detailed mechanism of primed spacer acquisition in type I-E CRISPR-Cas systems is not yet completely known, we start out with a simplified model to see if this is sufficient to explain our data. Because primed adaptation is much more efficient than naive adaptation <sup>22</sup>, we assume that the rate of naive adaptation is negligibly small over the time course of the experiment. The spacer composition of the CRISPR array is not modeled in detail. Rather, we assume that we start out with a crRNA sequence that matches the target, but is flanked by a non-consensus PAM. The effector complexes containing this spacer can still bind to the target DNA <sup>5,20</sup>, but with a binding affinity that is decreased up to a factor 100 - 150 as compared to binding with a consensus PAM <sup>50,51</sup>. Once the effector complex is bound to the target, Cas3-catalysed destruction of the target takes place <sup>79</sup>. Thus, the level of interference is associated with the level of effector complex binding

<sup>50</sup>.

Cas3-mediated destruction of targets is a source of substrates for spacer acquisition machinery, the Cas1-Cas2 complex, during primed adaptation<sup>2,23</sup>. Intermediates of target DNA degradation are transient and quickly degrade after an initial burst. Abundant levels of Cas1 and Cas2 lead to robust spacer acquisition, by allowing Cas1-Cas2 to capture the transient intermediates of Cas3 action<sup>23</sup>. Since in our system Cas3, Cas1, and Cas2 are highly expressed, we assume the levels of these proteins are not rate-limiting within the scope of our model and thus do not explicitly model their abundances. Furthermore, in agreement with previously published work, we assume cells have a target maintenance system that is controlled by logistic dynamics in order to keep the target concentration at its target level<sup>56</sup>. In addition, targets and target-containing configurations are actively partitioned between daughter cells<sup>80,81</sup> according to a multi-hypergeometric distribution, with each daughter receiving on average half of the mother cell's targets. All other proteins are partitioned according to a Binomial distribution, where the ratio of daughter cell sizes determines the probability of each molecule ending up in one of two daughter cells. We model synthesis of CRISPR proteins as a Poisson process, in which proteins are produced in geometrically distributed bursts to capture the effect of transcriptional bursting<sup>82</sup>. We assume all molecular species are stable on the timescale of the experiment (i.e. no degradation), with the exception of the free crRNAs (not loaded in Cascade) and the DNA fragments that are the result of interference, which have a short lifetime.

### 5.7.4 Algorithm outline

For the agent-based model, we have adapted the First-Division Algorithm by Thomas<sup>83</sup> to include the Extrade extension to the SSA. Furthermore, we keep the population size constant by randomly selecting a cell to be removed from the population in the event of a cell division. The steps to replicate our experimental set-up are described below.

**1. Population initialization:** At time  $t = 0$ , initialize  $N$  cells by assigning to each cell an age  $t_i \sim U(-\log(2)/M_p, \log(2)/\mu_p)$ , a growth rate  $\mu_i \sim \text{Lognormal}(\mu_p, \sigma_p^2)$  and molecule count  $x_i$ . Select division size  $V_{d,i} \sim \text{Lognormal}(\mu_{V_D}, \sigma_{V_D}^2)$  and compute generation time  $t_{\text{gen},i}$  as  $\log(V_{d,i}/V_{b,i})/\mu_i$ , where  $V_{b,i}$  is the birth size. This determines the division time of the cell which is defined as  $t_{d,i} = t_i + t_{\text{gen},i}$

**2. Biochemical reactions:** Determine the next dividing cell:  $j = \operatorname{argmin}_i (t_{d,i} - t_i)$ . Determine  $\Delta t$  from  $\min(t_{d,j} - t_j, L)$ , where  $L$  is *Extrande*'s look-ahead horizon. Advance the molecule numbers of each cell independently from age  $t_i$  to  $t_i + \Delta t$  using the *Extrande* algorithm and advance time from  $t$  to  $t + \Delta t$ .

**3. Cell division:** When  $t = t_{d,j}$ , replace the dividing cell by two newborn daughter cells of zero age. The birth size of both daughters is determined as  $V_{b,D_1} = \operatorname{Normal}(\mu_{V_B}, \sigma_{V_B}) V_{d,j}$  and  $V_{b,D_2} = V_{d,j} - V_{b,D_1}$ . Assign to one of these a molecule number distributed according to the Binomial distribution (proteins) and the Multi-hypergeometric distribution (targets and target configurations), depending on the mother's molecule count  $x_j$  and the daughter's size ratio to the mother cell  $\frac{V_{b,D_1}}{V_{d,j}}$ , and assign the remaining molecules to the other daughter. Assign to each daughter independently a growth rate  $\mu_i$ , division volume  $V_{d,i}$ , and compute corresponding division time. To ensure a constant population size, randomly select a cell to be deleted from the population.

**4. Repeat:** Repeat from 2. until  $t = t_{\text{final}}$ .

### 5.7.5 Molecular mechanism and model parameters

Each cell in the population contains a pool of biochemical species that can interact with each other through biochemical reactions, as described in step 2. We distinguish between the targets  $P$ , the CRISPR array  $A$ , which codes for a spacer crRNA matching a sequence on the target, and the surveillance protein Cascade. Together with the crRNA, the Cascade protein makes up the effector complex  $E$ . When the effector complex encounters a target it can bind, albeit with a low affinity in the case of a non-consensus PAM on the target, forming a complex  $EP$ . Destruction of the target can then take place, producing DNA fragments  $F$ . One of these fragments can be integrated into the CRISPR array  $A$  as a new spacer, transforming the array to  $A^*$  which can now also express the newly acquired crRNA,  $\text{crRNA}^*$ , in addition to the spacer that was already present. The effector complex containing the new spacer,  $E^*$  has a higher binding affinity for the target. These biochemical reactions are governed by the equations described in **Supplementary Methods Table 5.1**.



The size of individual cells increases exponentially with a constant elongation rate throughout the cell cycle. Cellular length is used as a measure for cell size, as *E. coli* cell width remains approximately constant throughout the cell cycle and thus the cellular volume is linearly proportional to the cell length<sup>84</sup>. Growth parameters were chosen to be representative for our experimental data. As no kinetic data are available on individual reactions of the adaptation and interference processes, these parameters were calibrated to qualitatively agree with the experimentally determined target loss time distributions from the direct interference and priming conditions and previously published abundances of Cas proteins<sup>85</sup>. Unless stated otherwise, the growth parameters used were  $\mu_p = \log(2)/70$ ,  $\sigma_p = 0.2$ ,  $\mu_{V_B} = 0.5$ ,  $\sigma_{V_B} = 0.07$ .  $\mu_{V_D}, \mu_{V_D} = 3.9$ ,  $\sigma_{V_D} = 0.11$ .  $\mu_{V_B}, p^* = 5$ . To simulate the direct interference condition with the same model, we simply modify the initial state of the system such that the spacer array consists of crRNA\*, which is flanked by the consensus PAM sequence.

#### 5.7.6 Cascade variability impacts the probability of spacer acquisition

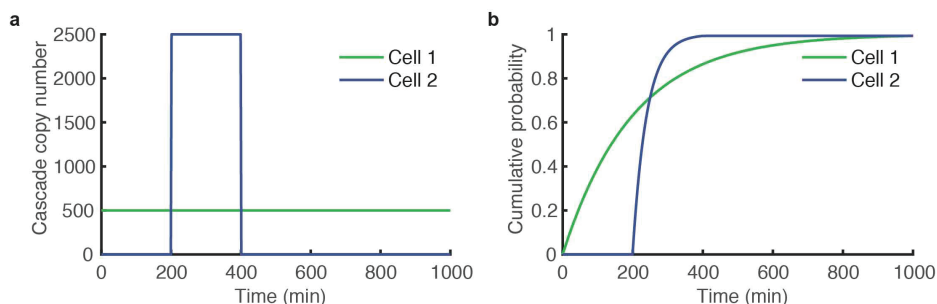
In the main text of the manuscript we have shown that in priming, increased variability in the expression of Cascade can lead to faster spacer acquisition on average (**Fig. 5.5c** of manuscript). In simulations of the agent-based model, variability of the Cascade protein concentration is controlled through the protein production rate  $k_1$  in coordination with the average protein burst size  $b_c$ : to modify Cascade variability while maintaining a constant concentration,  $b_c$  is multiplied by a factor  $a$  while  $k_1$  is multiplied by its inverse,  $\frac{1}{a}$ . In **Fig. 5.5**,  $a = 100$  which leads to an increase of the coefficient of variation of the Cascade concentration at steady state from  $CV = 0.02$  (low Cascade variability) to  $CV = 0.42$  (high Cascade variability).

We will now illustrate how higher Cascade variability can lead to faster spacer acquisition by considering two scenarios, and comparing the cumulative probability of the time until spacer acquisition for the simplified two-step model, which is given by

$$FP_{SA}(t|M_0) = 1 - e^{-M_0 p_p \int_0^t \mu(t') dt'}.$$

First, we consider a cell which has a constant Cascade level of 500 copies at any point in time between  $t = 0$ -1000 min, and plot the corresponding cumulative spacer acquisition probability (**Supplementary Methods Fig. 5.2a**). Second, we consider a second cell in which Cascade is not constant

but rather appears as a shorter 'burst' of 2500 copies from  $t = 200$  min until  $t = 400$  min, and 0 copies at any other time (**Supplementary Methods Fig. 5.2a**). The cumulative spacer acquisition probability for the second cell reaches 1 faster than for the first cell (**Supplementary Methods Fig. 5.2b**), despite the two cells having the same average Cascade concentration over the course of 1000 minutes. This suggests that the effects of upwards fluctuations can outweigh the downward fluctuations.



**Supplementary Methods Figure 5.2: a, Cascade copy number of two cells with the same average over time. a,** Cell 1 has a constant copy number of 500 Cascades, while in cell 2 Cascade is present transiently at 2500 copies between 200 and 400 minutes. **b,** Cumulative probability of time until spacer acquisition for the two cells with Cascade copy numbers as described in panel a ( $\mu_0 = 1, p_p = 0.00001$ ).

**Supplementary Methods Table 5.1: Overview of the reactions in the model for primed adaptation**

Phase	Reactions
target replication	$P \xrightarrow{k_0/(1+((P/V_t)/p_0)^2)} 2P,$ $p_0 = \frac{p^*}{V_t} / \sqrt{\left(\frac{k_0}{\mu} - 1\right)}$
Expression	$G \xrightarrow{k_1(t)} G + b_P \cdot Cascade$ $k_1(t) = \frac{k_1}{1 + \exp(-k_c t)}$ <p>Before spacer integration</p> $A \xrightarrow{k_2} A + b_c \cdot crRNA$ <p>After spacer integration</p> $A^* \xrightarrow{k_2} A^* + b_c \cdot crRNA + b_c \cdot crRNA^*$ $crRNA + Cascade \xrightarrow{k_3} E$ $crRNA^* + Cascade \xrightarrow{k_3} E^*$
Interference	$E + P \xrightleftharpoons[k_5]{k_4} EP$ $E^* + P \xrightleftharpoons[k_7]{k_6} EP^*$ $EP \xrightarrow{k_8} E + b_F \cdot F$ $EP^* \xrightarrow{k_8} E^* + b_F \cdot F$ $F \xrightarrow{k_9} \emptyset$
Primed adaptation	$F + A \xrightarrow{k_{10}} A^*$

**Supplementary Methods Table 5.2: Reaction rates used in simulations**

Reaction	Parameter	Value [min <sup>-1</sup> ]
Target replication	$k_0$	0.125
Cascade production	$k_1$	2.4
crRNA/crRNA <sup>*</sup> transcription	$k_2$	10
crRNA/crRNA <sup>*</sup> degradation	$k_3$	0.014
crRNA-Cas/crRNA <sup>*</sup> -Cas effector complex formation	$k_4$	0.01
E-P binding affinity	$k_5$	$1 \text{ e}^{-5}$
EP dissociation	$k_6$	$1 \text{ e}^{-4}$
E <sup>*</sup> -P binding affinity	$k_7$	$1 \text{ e}^{-3}$
EP <sup>*</sup> dissociation	$k_8$	$1 \text{ e}^{-4}$
Target degradation	$k_9$	1
Fragment degradation	$k_{10}$	1
Spacer integration	$k_{11}$	0.25
Cascade burst size	$b_P$	3
crRNA/crRNA <sup>*</sup> burst size	$b_c$	3
DNA fragment burst size	$b_F$	5
Post-induction delay of protein production	$k_d$	0.025

## References

1. Datsenko, K. A. *et al.* Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat. Commun.* **3**, 945 (2012).
2. Künne, T. *et al.* Cas3-Derived Target DNA Degradation Fragments Fuel Primed CRISPR Adaptation. *Mol. Cell* **63**, 852–864 (2016).
3. Vink, J. N. A. *et al.* Direct Visualization of Native CRISPR Target Search in Live Bacteria Reveals Cascade DNA Surveillance Mechanism. *Mol. Cell* **77**, 39–50.e10 (2020).
4. Swarts, D. C., Mosterd, C., van Passel, M. W. J. & Brouns, S. J. J. CRISPR Interference Directs Strand Specific Spacer Acquisition. *PLoS One* **7**, e35888 (2012).
5. Musharova, O. *et al.* Systematic analysis of Type I-E Escherichia coli CRISPR-Cas PAM sequences ability to promote interference and primed adaptation. *Mol. Microbiol.* **111**, 1558–1570 (2019).
6. Kim, S. *et al.* Selective loading and processing of pre-spacers for precise CRISPR adaptation. *Nature* **579**, 141–145 (2020).
7. Xue, C. & Sashital, D. G. Mechanisms of Type I-E and I-F CRISPR-Cas Systems in Enterobacteriaceae. *EcoSal Plus* **8**, (2019).
8. Loeff, L., Brouns, S. J. J. & Joo, C. Repetitive DNA Reeling by the Cascade-Cas3 Complex in Nucleotide Unwinding Steps. *Mol. Cell* **70**, 385–394.e3 (2018).
9. Dillard, K. E. *et al.* Assembly and Translocation of a CRISPR-Cas Primed Acquisition Complex. *Cell* **175**, 934–946.e15 (2018).
10. Nuñez, J. K., Harrington, L. B., Kranzusch, P. J., Engelman, A. N. & Doudna, J. A. Foreign DNA capture during CRISPR-Cas adaptive immunity. *Nature* **527**, 535–8 (2015).
11. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–12 (2007).
12. Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S. D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**, 2551–2561 (2005).
13. Brouns, S. J. J. *et al.* Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science (80-. )*. **321**, 960–964 (2008).
14. Jackson, R. N. *et al.* Crystal structure of the CRISPR RNA-guided surveillance complex from Escherichia coli. *Science (80-. )*. **345**, 1473–1479 (2014).
15. Deveau, H. *et al.* Phage Response to CRISPR-Encoded Resistance in *Streptococcus thermophilus*. *J. Bacteriol.* **190**, 1390–1400 (2008).
16. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733–740 (2009).
17. 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) doi:10.1016/j.molcel.2012.03.018.
18. Garneau, J. E. *et al.* The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67–71 (2010).
  19. Leenay, R. T. *et al.* Identifying and Visualizing Functional PAM Diversity across CRISPR-Cas Systems. *Mol. Cell* **62**, 137–147 (2016).
  20. Semenova, E. *et al.* Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 10098–103 (2011).
  21. Fineran, P. C. *et al.* Degenerate target sites mediate rapid primed CRISPR adaptation. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E1629–38 (2014).
  22. Stringer, A. M., Cooper, L. A., Kadaba, S., Shrestha, S. & Wade, J. T. Characterization of primed adaptation in the *Escherichia coli* type I-E CRISPR-cas system. *bioRxiv* (2020) doi:10.1101/2020.02.10.942821.
  23. Semenova, E. *et al.* Highly efficient primed spacer acquisition from targets destroyed by the *Escherichia coli* type I-E CRISPR-Cas interfering complex. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 7626–31 (2016).
  24. Jackson, S. A. *et al.* CRISPR-Cas: Adapting to change. *Science* (80-. ). **356**, eaal5056 (2017).
  25. Wright, A. V. *et al.* Structures of the CRISPR genome integration complex. *Science* (80-. ). (2017) doi:10.1126/science.aao0679.
  26. Kutter, E. *et al.* From Host to Phage Metabolism: Hot Tales of Phage T4's Takeover of *E. coli*. *Viruses* **10**, 387 (2018).
  27. Davison, J. Pre-early functions of bacteriophage T5 and its relatives. *Bacteriophage* **5**, e1086500 (2015).
  28. Hampton, H. G., Watson, B. N. J. & Fineran, P. C. The arms race between bacteria and their phage foes. *Nature* **577**, 327–336 (2020).
  29. Shao, Q., Hawkins, A. & Zeng, L. Phage DNA Dynamics in Cells with Different Fates. *Biophys. J.* **108**, 2048–2060 (2015).
  30. Staals, R. H. J. *et al.* Interference-driven spacer acquisition is dominant over naive and primed adaptation in a native CRISPR-Cas system. *Nat. Commun.* **7**, 12853 (2016).
  31. Xue, C. *et al.* CRISPR interference and priming varies with individual spacer sequences. *Nucleic Acids Res.* **43**, 10831–10847 (2015).
  32. Jackson, S. A., Birkholz, N., Malone, L. M. & Fineran, P. C. Imprecise Spacer Acquisition Generates CRISPR-Cas Immune Diversity through Primed Adaptation. *Cell Host Microbe* **25**, 250–260.e4 (2019).
  33. Elowitz, M. B. Stochastic Gene Expression in a Single Cell. *Science* (80-. ). **297**, 1183–1186 (2002).
  34. Spudich, J. L. & Koshland, D. E. Non-genetic individuality: chance in the single cell. *Nature* **262**, 467–471 (1976).
  35. van Houte, S. *et al.* The diversity-generating benefits of a prokaryotic adaptive immune system. *Nature* **532**, 385–8 (2016).
  36. Richter, C. *et al.* Priming in the Type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. *Nucleic Acids Res.* **42**, 8516–26 (2014).
  37. Amlinger, L., Hoekzema, M., Wagner, E. G. H., Koskiniemi, S. & Lundgren,

- M. Fluorescent CRISPR Adaptation Reporter for rapid quantification of spacer acquisition. *Sci. Rep.* **7**, 10392 (2017).
38. Høyland-Kroghsbo, N. M., Muñoz, K. A. & Bassler, B. L. Temperature, by Controlling Growth Rate, Regulates CRISPR-Cas Activity in *Pseudomonas aeruginosa*. *MBio* **9**, (2018).
  39. Díez-Villaseñor, C., Guzmán, N. M., Almendros, C., García-Martínez, J. & Mojica, F. J. M. CRISPR-spacer integration reporter plasmids reveal distinct genuine acquisition specificities among CRISPR-Cas I-E variants of *Escherichia coli*. *RNA Biol.* **10**, 792–802 (2013).
  40. Kremers, G.-J., Goedhart, J., van Munster, E. B. & Gadella, T. W. J. Cyan and Yellow Super Fluorescent Proteins with Improved Brightness, Protein Folding, and FRET Förster Radius †, ‡. *Biochemistry* **45**, 6570–6580 (2006).
  41. Wehrens, M. *et al.* Size Laws and Division Ring Dynamics in Filamentous *Escherichia coli* cells. *Curr. Biol.* **28**, 972-979.e5 (2018).
  42. Kiviet, D. J. *et al.* Stochasticity of metabolism and growth at the single-cell level. *Nature* **514**, 376–379 (2014).
  43. Young, J. W. *et al.* Measuring single-cell gene expression dynamics in bacteria using fluorescence time-lapse microscopy. *Nat. Protoc.* **7**, 80–88 (2012).
  44. Levine, J. H., Fontes, M. E., Dworkin, J. & Elowitz, M. B. Pulsed Feedback Defers Cellular Differentiation. *PLoS Biol.* **10**, e1001252 (2012).
  45. Yosef, I., Goren, M. G. & Qimron, U. Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Res.* **40**, 5569–5576 (2012).
  46. Schwabe, A. & Bruggeman, F. J. Contributions of cell growth and biochemical reactions to nongenetic variability of cells. *Biophys. J.* (2014) doi:10.1016/j.bpj.2014.05.004.
  47. Martynov, A., Severinov, K. & Ispolatov, I. Optimal number of spacers in CRISPR arrays. *PLOS Comput. Biol.* **13**, e1005891 (2017).
  48. Iranzo, J., Lobkovsky, A. E., Wolf, Y. I. & Koonin, E. V. Evolutionary Dynamics of the Prokaryotic Adaptive Immunity System CRISPR-Cas in an Explicit Ecological Context. *J. Bacteriol.* **195**, 3834–3844 (2013).
  49. Bradde, S., Vucelja, M., Teșileanu, T. & Balasubramanian, V. Dynamics of adaptive immunity against phage in bacterial populations. *PLOS Comput. Biol.* **13**, e1005486 (2017).
  50. Cooper, L. A., Stringer, A. M. & Wade, J. T. Determining the Specificity of Cascade Binding, Interference, and Primed Adaptation In Vivo in the *Escherichia coli* Type I-E CRISPR-Cas System. *MBio* **9**, e02100-17 (2018).
  51. Jung, C. *et al.* Massively Parallel Biophysical Analysis of CRISPR-Cas Complexes on Next Generation Sequencing Chips. *Cell* **170**, 35-47.e13 (2017).
  52. Klumpp, S. Growth-Rate Dependence Reveals Design Principles of Plasmid Copy Number Control. *PLoS One* **6**, e20403 (2011).
  53. Lin-Chao, S. & Bremer, H. Effect of the bacterial growth rate on replication control of plasmid pBR322 in *Escherichia coli*. *Mol. Gen. Genet.* **203**, 143–9 (1986).

54. Ingmer, H., Miller, C. & Cohen, S. N. The RepA protein of plasmid pSC101 controls *Escherichia coli* cell division through the SOS response. *Mol. Microbiol.* **42**, 519–526 (2001).
55. Hynes, A. P., Villion, M. & Moineau, S. Adaptation in bacterial CRISPR-Cas immunity can be driven by defective phages. *Nat. Commun.* **5**, 4399 (2014).
56. Severinov, K., Ispolatov, I. & Semenova, E. The Influence of Copy-Number of Targeted Extrachromosomal Genetic Elements on the Outcome of CRISPR-Cas Defense. *Front. Mol. Biosci.* **3**, (2016).
57. Nicholson, T. J. *et al.* Bioinformatic evidence of widespread priming in type I and II CRISPR-Cas systems. *RNA Biol.* **16**, 566–576 (2019).
58. Nobrega, F. L., Walinga, H., Dutilh, B. E. & Brouns, S. J. J. Prophages are associated with extensive CRISPR-Cas auto-immunity. *Nucleic Acids Res.* (2020) doi:10.1093/nar/gkaa1071.
59. Balaban, N. Q., Merrin, J., Chait, R., Kowalik, L. & Leibler, S. Bacterial persistence as a phenotypic switch. *Science* **305**, 1622–5 (2004).
60. Uphoff, S. *et al.* Stochastic activation of a DNA damage response causes cell-to-cell mutation rate variation. *Science (80-. )*. **351**, 1094–1097 (2016).
61. Moormeier, D. E., Bose, J. L., Horswill, A. R. & Bayles, K. W. Temporal and Stochastic Control of *Staphylococcus aureus* Biofilm Development. *MBio* **5**, (2014).
62. Acar, M., Mettetal, J. T. & van Oudenaarden, A. Stochastic switching as a survival strategy in fluctuating environments. *Nat. Genet.* **40**, 471–475 (2008).
63. Westra, E. R. *et al.* Parasite Exposure Drives Selective Evolution of Constitutive versus Inducible Defense. *Curr. Biol.* **25**, 1043–1049 (2015).
64. Strotskaya, A. *et al.* The action of *Escherichia coli* CRISPR-Cas system on lytic bacteriophages with different lifestyles and development strategies. *Nucleic Acids Res.* (2017) doi:10.1093/nar/gkx042.
65. Watson, B. N. J. *et al.* Type I-F CRISPR-Cas resistance against virulent phages results in abortive infection and provides population-level immunity. *Nat. Commun.* (2019) doi:10.1038/s41467-019-13445-2.
66. Lopatina, A., Tal, N. & Sorek, R. Abortive Infection: Bacterial Suicide as an Antiviral Immune Strategy. *Annual Review of Virology* (2020) doi:10.1146/annurev-virology-011620-040628.
67. Moxon, R. & Kussell, E. The impact of bottlenecks on microbial survival, adaptation, and phenotypic switching in host–pathogen interactions. *Evolution (N. Y.)*. (2017) doi:10.1111/evo.13370.
68. Marraffini, L. A. & Sontheimer, E. J. CRISPR Interference Limits Horizontal Gene Transfer in *Staphylococci* by Targeting DNA. *Science (80-. )*. **322**, 1843–1845 (2008).
69. Patterson, A. G. *et al.* Quorum Sensing Controls Adaptive Immunity through the Regulation of Multiple CRISPR-Cas Systems. *Mol. Cell* (2016) doi:10.1016/j.molcel.2016.11.012.
70. Nguyen, J., Lara-Gutiérrez, J. & Stocker, R. Environmental fluctuations and their effects on microbial communities, populations and individuals. *FEMS Microbiol. Rev.* (2020) doi:10.1093/femsre/fuaa068.

71. Spormann, A. M. Physiology of Microbes in Biofilms. in *Current Topics in Microbiology and Immunology* 17–36 (2008). doi:10.1007/978-3-540-75418-3\_2.
72. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci.* **97**, 6640–6645 (2000).
73. Uphoff, S., Reyes-Lamothé, R., De Leon, F. G., Sherratt, D. J. & Kapanidis, A. N. Single-molecule DNA repair in live bacteria. *Proc. Natl. Acad. Sci. U. S. A.* (2013) doi:10.1073/pnas.1301804110.
74. Ovesný, M., Křížek, P., Borkovec, J., Švindrych, Z. & Hagen, G. M. ThunderSTORM: a comprehensive ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging. *Bioinformatics* **30**, 2389–2390 (2014).
75. Voliotis, M., Thomas, P., Grima, R. & Bowsher, C. G. Stochastic Simulation of Biomolecular Networks in Dynamic Environments. *PLOS Comput. Biol.* **12**, e1004923 (2016).
76. Thompson, M. G. *et al.* Isolation and characterization of novel mutations in the pSC101 origin that increase copy number. *Sci. Rep.* **8**, 1590 (2018).
77. Thomas, P. Intrinsic and extrinsic noise of gene expression in lineage trees. *Sci. Rep.* **9**, 474 (2019).
78. Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81**, 2340–2361 (1977).
79. Krivoy, A. *et al.* Primed CRISPR adaptation in *Escherichia coli* cells does not depend on conformational changes in the Cascade effector complex detected in Vitro. *Nucleic Acids Res.* **46**, 4087–4098 (2018).
80. Shao, B. *et al.* Single-cell measurement of plasmid copy number and promoter activity. *Nat. Commun.* **12**, 1475 (2021).
81. Meacock, P. A. & Cohen, S. N. Partitioning of bacterial plasmids during cell division: a cis-acting locus that accomplishes stable plasmid inheritance. *Cell* (1980) doi:10.1016/0092-8674(80)90639-X.
82. Golding, I., Paulsson, J., Zawilski, S. M. & Cox, E. C. Real-Time Kinetics of Gene Activity in Individual Bacteria. *Cell* **123**, 1025–1036 (2005).
83. Thomas, P. Making sense of snapshot data: Ergodic principle for clonal cell populations. *J. R. Soc. Interface* (2017) doi:10.1098/rsif.2017.0467.
84. Taheri-Araghi, S. *et al.* Cell-Size Control and Homeostasis in Bacteria. *Curr. Biol.* **25**, 385–391 (2015).
85. Djordjevic, M., Djordjevic, M. & Severinov, K. CRISPR transcript processing: a mechanism for generating a large number of small interfering RNAs. *Biol. Direct* **7**, 24 (2012).



## Summary |

The microbial warfare between prokaryotes (bacteria and archaea) known as the host, and their invaders (mobile genetic elements, MGEs) is never-ending and results in constant evolution. This evolution is circular and driven by the rapid adaptation of MGEs such as phage to avoid the cells immune response. In order to survive infection, prokaryotes are forced to develop and harbor a wide array of innate and adaptive immune systems that provide defense. The first described example of prokaryotic adaptive immunity is CRISPR-Cas an intricate system that consists of a memory bank (CRISPR array) to enable recognition of previously encountered threats, and an operon of CRISPR associated proteins (Cas proteins) that find and eliminate the invader. In this thesis we develop and apply methods to further probe the kinetics of the type I-E CRISPR-Cas system of *Escherichia coli* at the single cell level. **Chapter 1** gives a general introduction into bacterial immune systems, covering the discovery of CRISPR, the diversity of CRISPR systems and its basic mechanisms of defense. We then zoom in on the type I-E CRISPR system of *E. coli* specifically and the unique mechanistic details of interference and adaptation. Finally, we give an overview of a number of techniques used to study CRISPR-Cas so far, explain the limitations of these techniques, and highlight the methods we have contributed to the field.

In order to carry out a fast and specific response against an invader the CRISPR-Cas system is required to possess a small fragment of DNA from the invader, termed a spacer, in its CRISPR array. This spacer is selected and incorporated in well conserved and specific process coined adaptation. In **Chapter 2** we review the advances in the understanding of the CRISPR adaptation process in great detail focusing on the action of the well conserved Cas1 and Cas2 proteins. We cover the molecular steps of the process, mechanisms of the production of DNA fragments to be incorporated as spacers (pre-spacers), capture of these pre-spacers by the Cas1-Cas2 proteins, recognition of and insertion of spacers into the CRISPR array and the role of accessory proteins.

Due to the possible risk of the cell acquiring a spacer from its own genome the adaptation process must be highly regulated and can therefore be infrequent, particularly in the absence of an invader. In **Chapter 3** we describe CAPTURE (CCRISPR adaptation PCR technique using reamplification and electrophoresis) a method we developed to enhance the sensitivity of detection of spacer acquisition occurring in a population. We

combine previously described polymerase chain reaction (PCR) steps in combination with specific primer design and size extraction techniques to reach a detection limit of spacer acquisition that has occurred in 0.01% of cells in the population. A detailed outline of the experimental design, required components and reagents and a step by step protocol are given. Further, we compare our approach to existing techniques, discuss potential limitations and provide potential future applications for our method.

**Chapter 4** shifts the focus from the adaptation process to the interference process, elucidating key details about the target search dynamics of Cascade belonging to the type I-E system of *E. coli*. Here, by fusing a photoactivatable fluorescent tag to the *cas8e* subunit of Cascade, we use single-molecule particle tracking to investigate the relationship between *in vivo* interference levels and Cascade copy number. We observed that Cascade copy number dependence follows a linear trend suggesting a time-driven arms-race between invader replication and target search by Cascade. Further, we study the probing kinetics of Cascade and describe evidence for a subunit dissociation occurring upon binding to repeat-like targets, suggesting a possible mechanism for the avoidance of self-targeting. Taken together, this chapter provides a new depth of information on the target search process carried out by the Cascade surveillance complex.

In **Chapter 5** we use time-lapse microscopy in combination with microfluidics to study single-cell lineages over time, throughout the entire duration of CRISPR defense. Here, for the first time we were able to quantify the variation in time taken to clear an invader occurring between single cells within a population. Monitoring of the invader through a fluorescent reporter and use of the production rate to more accurately determine the moment of loss showed us that while priming is a highly variable process, direct interference is comparatively fast and deterministic. Further analysis of the experimental data and the use of an agent-based model allowed us to identify the spacer acquisition process as the source of the increased variation in priming. In addition, through comparing single-cell lineages that successfully eliminated the invader to those that did not, we were able to identify cellular growth rate, interdivision time and Cascade concentration as factors increasing the efficiency of defense. Finally, the model was used as a tool enabling us to investigate to what extent changes in a factor of interest, e.g. invader copy number, affect the variation seen in invader elimination times.

In conclusion, in the thesis we have developed a number of methods which allowed us to further investigate the highly dynamic type I-E CRISPR-Cas

immune response. Through this, we have provided insights into the variation seen in response times between individual cells within a population and contributed to elucidating the details of CRISPR adaptation and interference. Recently, the view that CRISPR defense acts only at the level of individual cells is beginning to be challenged. It will therefore be of further interest to apply these techniques to a broader range of CRISPR systems and environments to deepen our knowledge of CRISPR immunity in particular in natural settings.

# Samenvatting |

De microbiologische oorlogsvoering tussen prokaryoten (bacterien en archaea), de gastheer en hun indringers (mobile genetic elements, MGEs) is oneindig en leidt tot constante evolutie. Deze evolutie is circulair en wordt gedreven door de snelle adaptatie van MGEs zoals fagen om de immuunrespons van cellen te omzeilen. Om deze infectie te overleven, worden prokaryoten gedwongen een breed scala aan aangeboren en adaptieve immuunsystemen te ontwikkelen en onder te brengen. Het eerst beschreven voorbeeld van een prokaryoot immuunsysteem is CRISPR-Cas, een complex systeem dat bestaat uit een geheugenbank (CRISPR array), die herkenning van eerder tegengekomen indringers mogelijk maakt en een operon van CRISPR associated eiwitten (Cas eiwitten) die de indringer vindt en elimineert. In dit proefschrift ontwikkelen en passen we methodes toe die de kinetiek van het Type I-E CRISPR Cas systeem van *Escherichia coli* op het single-cell niveau bestuderen. **Hoofdstuk 1** geeft een algemene inleiding voor bacteriële immuunsystemen, de ontdekking van CRISPR, de diversiteit van CRISPR systemen en zijn algemene verdedigingsmechanismen. We bespreken vervolgens het Type I-E CRISPR systeem van *E. coli* en de unieke mechanistische aspecten van interferentie en adaptatie in meer detail. Tot slot geven we een overzicht van een aantal technieken die tot dusverre gebruikt zijn om CRISPR te bestuderen, leggen we de beperkingen van deze systemen uit en lichten we de methodes toe waarmee we een bijdrage hebben geleverd aan het onderzoeksveld.

Om een snelle en specifieke respons tegen de indringer uit te voeren, moet het CRISPR-Cas systeem in een CRISPR array een klein DNA fragment, genaamd de spacer, bevatten van de indringer. Deze spacer is geselecteerd en geïncorporeerd in een geconserveerd en specifiek proces genaamd adaptatie. In **hoofdstuk 2** geven we een overzicht van de vorderingen in het begrijpen van het CRISPR adaptatieproces, waar we ons richten op de geconserveerde Cas1 en Cas2 eiwitten. We beschrijven de moleculaire stappen van het proces, de mechanismen voor de productie van DNA fragmenten, die als spacers geïncorporeerd worden (pre-spacers), de binding van deze pre-spacers aan Cas1-Cas2, de herkenning en insertie van spacers in het CRISPR array en de rol van bijbehorende eiwitten.

Door het mogelijk risico van de cel om een spacer te verkrijgen uit zijn eigen genoom moet het adaptatieproces sterk gereguleerd zijn, waardoor het zeldzaam kan zijn, vooral in de afwezigheid van een indringer. In **hoofdstuk 3** beschrijven we CAPTURE (CRISPR adaptation PCR technique using

reamplification and electrophoresis), een methode die we ontwikkeld hebben om de gevoeligheid van spaceracquisitie in een populatie te verbeteren. We combineren eerder beschreven PCR stappen in combinatie met het ontwerp van specifieke primers en extractie technieken op basis van grootte om een detectielimiet te bereiken voor spaceracquisitie in 0.01% van de cellen in een populatie. Een gedetailleerd stappenplan voor de experimentele opzet, de benodigde ingrediënten en een protocol wordt hier beschreven. Verder vergelijken we onze methode met eerdere methodes, bespreken we de beperkingen en suggereren we mogelijke toekomstige toepassingen van onze techniek.

**Hoofdstuk 4** verlegt de focus van het adaptatieproces naar het interferentieproces, waar we belangrijke details over de dynamica van het zoekproces van type I-E Cascade in *E. coli* ophelderen. Door de fusie van een fotoactiveerbaar fluorescent label aan de cas8e subunit van Cascade, gebruiken we single-molecule particle tracking om de relatie tussen *in vivo* interferentie niveaus en het aantal Cascade complexen te bestuderen. We zagen een lineaire trend in deze relatie wat suggereert dat er een tijdsgedreven wapenwedloop tussen indringerreplicatie en aantal Cascades bestaat. We bestudeerden ook de kinetiek van het scannen van Cascade en vinden bewijs voor subunit dissociatie bij binding aan repeat-achtige sequenties, wat suggereert dat er een mogelijk mechanisme is om self-targeting te ontwijken. Alles tezamen verdiept dit hoofdstuk ons begrip over het zoekproces wat het Cascade complex uitvoert.

In **hoofdstuk 5** gebruiken we time-lapse microscopie in combinatie met microfluidica om de afkomst en nakomelingen van single cells te volgen over de gehele duur van CRISPR verdediging. We zijn hier voor het eerst in staat geweest de variatie te kwantificeren die bestaat in de tijd die het kost om een indringer te elimineren in afzonderlijke cellen van een populatie. Het monitoren van de indringer met behulp van een fluorescente reporter en het gebruik van de productiesnelheid om accurater te bepalen wanneer eliminatie plaatsvond liet ons zien dat waar priming een zeer variabel proces is, directe interferentie relatief snel en deterministisch is. Verdere analyse van de experimentele data en het gebruik van een agent-based model stelde ons in staat om het spaceracquisitieproces te identificeren als de bron van de toegenomen variatie in het priming proces. Bovendien konden we door een vergelijking te maken tussen cellijnen die de indringer elimineerden en cellijnen die dat niet deden, aantonen dat de cellulaire groeisnelheid, de delingssnelheid en de Cascadeconcentratie factoren zijn die een rol spelen in de efficiëntie van immuniteit. Ten slotte is het model gebruikt als tool om

te bestuderen in hoeverre een bepaalde factor, bijv. aantal indringers, effect heeft op de variatie in de tijd die het kost een indringer te elimineren.

Concluderend hebben we in dit proefschrift een aantal methoden ontwikkeld die ons in staat hebben gesteld de zeer dynamische immuunrespons van het type I-E CRISPR-Cas systeem te bestuderen. Hierdoor hebben we inzicht verkregen in de variatie in responstijden van het immuun systeem tussen individuele cellen in een populatie en bijgedragen aan de verdere opheldering van CRISPR adaptatie en interferentie. Recentlijk is de zienswijze dat het CRISPR immuunsysteem alleen maar van toepassing is op individuele cellen aan het veranderen. Het zal daardoor in de toekomst van belang zijn deze technieken toe te passen op een breder scala aan CRISPR systemen en omgevingen om zo onze kennis van CRISPR immuniteit vooral in natuurlijke omstandigheden te verdiepen.

### About the author

Rebecca Elizabeth McKenzie was born on the 29<sup>th</sup> of September 1991 in Wellington, New Zealand. In 2010, she began her studies in Microbiology and Immunology at the University of Otago, Dunedin, New Zealand. In 2013, she joined the research group of Peter Fineran, the beginning of her interest in CRISPR-Cas systems. One large learning curve and approximately 8 months of molecular lab work later, she submitted and defended her honours



thesis entitled "Acquisition of Resistance in the type I-F CRISPR-Cas System" to graduate with a Bachelor of Science with Honours. Following this Rebecca worked in the microbiology lab of Fonterra Clandeboyne, Temuka, New Zealand.

However, the call to return to scientific research was strong and in 2015, through a connection made by Peter Fineran, Rebecca was offered a position as a research assistant in the Brouns lab, Wageningen, The Netherlands. After relocation of the entire lab to Delft in 2017, she signed a PhD contract and the journey to assemble this thesis on type I-E CRISPR-Cas systems began. This project was a collaboration between two institutions, AMOLF and TU Delft, and co-supervised by Stan Brouns and Sander Tans. The main work involved the development and application of time-lapse microscopy to study the Type I-E CRISPR-Cas system in single cells as described here in Chapter 5 of this thesis. This change from molecular to microscopy techniques, while challenging, allowed her to develop a number of new skills, knowledge and collaborations that made her feel part of a team.

Rebecca loves to be outdoors, to travel, hike, camp, bake cakes, and most of all spend time with family and friends. After 6.5 years in The Netherlands she plans to return to New Zealand where she hopes to have a little extra time to enjoy the outdoors now her PhD is complete ;).

## List of publications

**R. E. McKenzie\***, E. M. Keizer\*, J. N. A. Vink, J. van Lopik, F. Büke, V. Kalkman, C. Fleck, S. J. Tans, S. J. J. Brouns, *bioRxiv* (2021), doi:10.1101/2021.07.21.453200. (Manuscript submitted)

J. N. A. Vink, K. J. A. Martens, M. Vlot, **R. E. McKenzie**, C. Almendros, B. Estrada Bonilla, D. J. W. Brocken, J. Hohlbein, S. J. J. Brouns, Direct Visualization of Native CRISPR Target Search in Live Bacteria Reveals Cascade DNA Surveillance Mechanism. *Mol. Cell.* 77, 39-50.e10 (2020).

A. N. Salazar, F. L. Nobrega, C. Anyansi, C. Aparicio-Maldonado, A. R. Costa, A. C. Haagsma, A. Hiralal, A. Mahfouz, **R. E. McKenzie**, T. van Rossum, S. J. J. Brouns, T. Abeel, An educational guide for nanopore sequencing in the classroom. *PLoS Comput. Biol.* 16 (2020), doi:10.1371/journal.pcbi.1007314.

C. Almendros, F. L. Nobrega, **R. E. McKenzie**, S. J. J. Brouns, Cas4-Cas1 fusions drive efficient PAM selection and control CRISPR adaptation. *Nucleic Acids Res.* 47 (2019), doi:10.1093/nar/gkz217.

**R. E. McKenzie**, C. Almendros, J. N. A. Vink, S. J. J. Brouns, Using CAPTURE to detect spacer acquisition in native CRISPR arrays. *Nat. Protoc.* 14 (2019), doi:10.1038/s41596-018-0123-5.

T. Künne, Y. Zhu, F. da Silva, N. Konstantinides, **R. E. McKenzie**, R. N. Jackson, S. J. J. Brouns, Role of nucleotide identity in effective CRISPR target escape mutations. *Nucleic Acids Res.* 46 (2018), doi:10.1093/nar/gky687.

S. N. Kieper, C. Almendros, J. Behler, **R. E. McKenzie**, F. L. Nobrega, A. C. Haagsma, J. N. A. Vink, W. R. Hess, S. J. J. Brouns, Cas4 Facilitates PAM-Compatible Spacer Selection during CRISPR Adaptation. *Cell Rep.* 22, 3377–3384 (2018).

S. A. Jackson\*, **R. E. McKenzie\***, R. D. Fagerlund, S. N. Kieper, P. C. Fineran, S. J. J. Brouns, CRISPR-Cas: Adapting to change. *Science* (80-. ). 356, eaal5056 (2017).

M. E. Wilkinson, Y. Nakatani, R. H. J. Staals, S. N. Kieper, H. K. Opel-Reading, **R. E. McKenzie**, P. C. Fineran, K. L. Krause, Structural plasticity and in vivo activity of Cas1 from the type I-F CRISPR-Cas system. *Biochem. J.* (2016), doi:10.1042/BCJ20160078.

C. Richter\*, R. L. Dy\*, **R. E. McKenzie\***, B. N. J. Watson, C. Taylor, J. T. Chang, M. B. McNeil, R. H. J. Staals, P. C. Fineran, Priming in the Type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. *Nucleic Acids Res.* 42, 8516–26 (2014).

\* These authors contributed equally



### Acknowledgements

**Ehara taku toa i te toa takitahi, engari he toa takitini** - Success is not the work of an individual, but the work of many. Finally, here the chance to thank all those who made this journey both possible and fun!

Firstly, I would like to thank the committee for your time and interest in my thesis and presence at my defense.

**Stan**, I want to thank you for the opportunity to move to the other side of the world and join your group first as a research assistant and later to complete a PhD. Over the past 6.5 years in your group I've been lucky enough to have the opportunity to learn, develop new set ups, organize many a lab event, participate in committees, optimize methods and completely change skill set. I'm grateful for the freedom you gave me to do these things and the independence I have developed as researcher under your supervision. While it was possibly quite exhausting always hearing my feedback about things that could change in the lab; be it group meetings, office arrangements, more social activities I hope some things have made a difference. It is strange to now watch history repeat as an entire new batch of Brouns lab members arrive – I wish you all the best with the next round!

**Sander**, I want to thank you for your critical thoughts, feedback and support throughout this project. It has been fun to introduce you and the Tans lab into the world of CRISPR! Although we did not meet often it was sometimes your few words and understanding such as “it's ok data analysis can be a slow process” that kept me motivated at the end.

**Peter**, although it's a long time now since you were my supervisor, without you I wouldn't have ended up here in the Brouns lab and life could be something else entirely! I know you were sad I did not stay to complete a PhD in NZ, but it really was your own fault - as I strongly remember in a 3rd year lecture you said “one of the things I like most about science is that you can make strong collaborations and have the opportunity to travel the world through your work”. I thought that sounded quite good ;) so I signed on to begin this journey in CRISPR, thank you for the good times in the Fineran lab and the advice I know I can always ask for. To the **Fineran lab members of 2013** and the **4<sup>th</sup> year micro gang** if not for you and the fun environment I may not have come back to research in 2015!

**Emma**, We did it! This paper took quite a lot more time, persistence and problem solving than we both initially expected but in the end I think we can be proud! I want to thank you for joining this project and making me feel part of a team! It was so nice to have a ‘partner in crime’ to discuss the results, the ups-and-downs of the project, what to do next and to juggle 3 PI's ;). It's been a pleasure to both work with and get to know you and I wish you all the best for your defense (I'm sorry to miss it) and the next steps!

To my students who I couldn't have finished this book without; **Alice**, thanks for being my first student and bearing with me on that huge learning curve! While you didn't always have the best luck with the experiments your perseverance paid off and I'm very happy to hear you're now a PhD yourself. **Jasper**, thanks for the fun times and team work while figuring out this new system together! I enjoyed the train rides to AMOLF and the opportunity to get to know you. All the best with your PhD! **Vera**, the ultra-organized! I have great appreciation for your colour coding, file naming system and refreshingly honest personality. As with Jasper I'm grateful for the train rides to AMOLF and the chance to get to know you, I particularly enjoyed the random facts you always knew such as 'the dutch are statistically the best drivers in the world'. Happy to hear you've found a job you enjoy! and **Chris** thank you for your hard work and perseverance when science and life wasn't easy! We learnt a lot together and I'm glad things are working out in your new role.

To the **Brouns group** past and present thanks for the fun times, the ever-dynamic atmosphere, the support and somehow making the weirdest, most isolated corner of building 58 fun!

Let's start at the beginning in a land far, far away (Wageningen): When I first joined the Brouns group it consisted of only **Tim**, **Marnix** and **Franklin**. **Tim** thank you for your kind words and calm presence in the office and for without hesitation translating the scary letters in a blue envelope I would receive. **Marnix** you taught me that google can always answer all my questions, thanks for always indulging me in the 'how was your weekend' office chat. **Franklin**, I'll never forget the day when you invited me to eat the classic dutch food of the Dreijen with you and Melvin, and the transition from colleagues to friends began. Luckily, we later realized we shared a love for more refined food ;). Thanks for the many fine dining escapes from Wageningen and later Delft, the endless hours of talking about life, about work, about the future. Your support got me through some of the harder moments and without you we (the lab) would never would have started the Swedish meatball tradition (that abruptly ended at some point?! You are the most ambitious person I know and I'm so happy to see your persistence and hard work pay off with your very own research group beginning! I hope you will remember that life is not all about working, don't forget to stop and smell the roses every now and then.

Although not in the Brouns group there were many people in **BacGen** who helped me on my way. **Melvin**, you answered every 'where is this' question I had with a smile on your face, an impressive feat! I would have otherwise been lost in the lab. **Nico**, I couldn't have asked for a better person to sit opposite in the lab. We talked for hours! I had not and have not had such lively company at the lab bench since. It seems in fact that one of your frequent topics, climate change and the need to reduce flying and meat intake, had a profound effect on me as I now find myself incorporating this into everyday life ;)! It's been great to watch you grow as a scientist and now

start up your own group back where it all began. **Emmy**, although we crossed paths for a very short time in Wageningen I am happy to have had the chance to experience your positive do-good attitude! I wish you and Nico all the best in your future adventures! **Alex**, from day one you were always up for a good laugh! I truly appreciate your kind but refreshingly no-bull\*\*\*\* attitude and I'm happy we've stayed in touch. Awesome that we managed to squeeze in a few trips together (thanks for the amazing tour of Slovenia) and I hope to climb a few more mountains with you in the future! **Prarthana**, it was so refreshing to meet someone so like-minded in such an unexpected place ;). I love and look forward to our catch ups and feel so fulfilled after we spend hours discussing everything that comes to mind, time flying by. Despite not seeing you so often, it always feels like yesterday when we get together - and I hope that will continue into the future even when we are a bit further apart! **Wen**, you are like a ray of sunshine and I'm happy we got the chance to spend some time together both socially and mulling over the complex details of adaptation. **Rob**, it was a relief to meet someone who understood my accent, weird expressions and somehow even some kiwi slang. Thanks for always being up for a beer, the Monday night sneak peek and a rant about life, academia and where to go next. I hope you'll make it across the ocean in the coming years to finally do the NZ part of your trip! **Cata**, you were a great comfort to me in Wageningen a fellow 'let's do things' gal. We had some great times shopping, exploring Utrecht, eating, drinking and travelling to Budapest (once you found your passport ;)). Your energy is infectious and your positivity in spite of the difficulties you faced was admirable! I hope we can keep in touch and meet again in the future for some reminiscing over a bottle of wine!

Of course, life in the Dreijen would not have been the same if there were no international people to eat lunch with each day so to **Basti, Maarten, Alex, Lara, Monika, Benoit, Kal, Javi, Irene, Cristina, Sven**, and **Juanan** and huge thanks for your all accepting attitudes and the many fun times and conversation topics. Finally, thanks also to all those who I thoroughly enjoyed a Friday night drink with or bumping into in the corridor for a chat, you were missed when we relocated to Delft **Hanne, Tom, Wim, Yifan, Brenda, Jeroen, Joyshree, Mark, Jorrit, Jasper, Tijn** and **Bas**.

Because life's not all about work I want to take a quick pause and acknowledge **Anne Vogel** for making the effort to welcome me to The Netherlands. Despite the fact we had never met you reached out to me at a time when loneliness was at its peak. Because of you I have a new awareness of what a difference these small acts of kindness can make to any new arrivals! In addition, to my first house mates: **Ben**, what were the chances that two kiwis end up in the same house in NL!? What good times we had building a "stadium" to watch and educate the locals about the rugby world cup! I hope to see you in NZ in the future. **Matthijs**, you're amazing cooking and genuinely caring personality were such a comfort when I was still adapting to my new world. **Michiel**, you were always the joker, thanks for keeping things ordered but also fun. I'll never again forget to switch off the lights ;p. It means a lot to me that you three, together with **Daniella** and

**Anne** (and **Laura** for giving me the room in the first place) created a real home away from home!

Back to work, in 2017 the Brouns lab was relocated to Delft. And while in Wageningen some new members had joined the group they had been mostly hiding in the basement, surfacing into my life in the Applied Sciences building of TU Delft ;): **Patrick**, those first weeks in Delft I was very grateful to have your company. You're always up for a chat about whatever and after all the relocations and office changes I was always very happy when on the odd occasion I would finally catch you alone in your office for a chat like the 'good old days'. I will never forget that amazing potato restaurant in Trier and I hope one day we can go back (if it's still there). I admire your strength and perseverance in the tough times and envy your seemingly overnight acquisition of bioinformatics skills. Your post PhD happiness is so nice to see and has been extremely motivating for me. **Lucy** thank you for the plant seeds, financial advice, bargain hunting websites and most of all an always welcoming presence that is fun to be around! Looking forward to a reunion with you both in Indonesia (or NZ). **Seb**, it was always nice to have a fellow Fineran lab compatriot to reminisce with. Thanks for making the office a homely place when no one else bothered to decorate. Over the years I have observed and appreciated your never-ceasing curiosity and desire to learn, both qualities that will get you far and that I'm sure you will have passed on to Mati. It's great to see you in a new role in industry taking the next steps and I wish you all the best.

**Anna**, thank you not only for keeping the lab in tip top shape - filled with all the things we could ever want and need - but also for your support. Whenever I had the PhD blues I could always count on you to listen and respond in a way that re-motivated me. The shared stories of your own PhD and the reminders of what really matters really got me through to the end! I will miss the many hours of chatting in the corridor / office or wherever I bumped into you. **Teunke**, thank you for your calm, kind directness I have learnt a lot from the way you express things. It was always a great day when you brought cake to work and I've missed your presence in the past year ☺. **Cristian**, you are a crazy bundle of positivity! It was great to have your upbeat, no-worries personality and music in the lab for the duration of my PhD. The lab would have been a whole lot quieter without you. The end can get tough and I hope you'll manage to keep these qualities at the surface through the final battle (I'm sorry to miss your defense)! Our shared desire for travel made hearing your stories over a beer a highlight. Looking forward to seeing where you end up in the future! **Rita**, your kind-heartedness is overwhelming and you are an ever-supportive presence to all those around you. Even in the toughest of times you came to work and smiled at others, I find your inner strength incredible! Thank you for all the chats in the lab and the office when I needed to get something off my chest and the many laughs we've shared about strange human behaviour. **Tobal**, your cooking skills blow me away! Every time we come to your place I know it's going to be a whole day feast where at the end I feel I can never eat again! I really

appreciated your presence in the office and having a fellow “how was your weekend” human around ;). It’s so great to see you and Patri building a home here in Delft, make sure Tuli doesn’t forget me :p. **Patri**, I feel you had to endure many chats about science during the winter drinks at the Bouw pub however, I’m glad you continued to join and we became friends! I’m in envy of your talent for painting and all things artistic. Thank you for all the small gifts you made us and ensuring that I never lost my lab coat due to my cute *E.coli*. I’m crossing all my fingers that delft blue becomes in demand once again! **Boris**, your margaritas are something else :O!! Thanks for keeping the 9-5 attitude alive and normal, and the fun times both in and outside of work. All the best to you and **Eva** for your future adventures and keeping your 100 plants alive ;p. **Benjamin**, you changed our lives with Terraforming Mars, finally giving me a game I could beat Jochem in ;). Although, I did not join I was always impressed by the number of people you managed to convince to do fitness during the work day. Thanks for all the cooking, games and good company when I was recovering from homesickness and jetlag. **Anouk** all the best with your studies, I look forward to the games you will design in the future!

It’s nice to see the Brouns lab finally take on some new members; **Marina** it’s been a joy to have this small crossover with you in The Netherlands before we meet again in NZ (you will get there don’t worry). Through our long days in the microscope I got to know your kind and caring side but also your inner fire. Don’t be afraid to say what you think; any group is lucky to have you! **Lucia** although we don’t really know each other I feel I know you very well (haha) thanks for your enthusiasm, openness and the magic touch which sold the house. All the best in NL I hope you enjoy it here as much as I have! **Sam** the deep conversationalist, although we did not have many opportunities due to the different phases of a PhD, I enjoyed our chats and philosophising about salaries, work conditions and human rights, to name a few. All the best with your future decisions to pursue science or philosophy. **Jelger**, you are such a generous person, always keeping the lab energised with treats. We didn’t see each other many times but whenever we did it was always with a smile and a joyful “how are you doing?” so thank you for that. And to the latest addition **Daan** all the best with your PhD journey.

To the **Tans lab** although I only intermittently joined your group meetings, I truly appreciated your warm and welcoming atmosphere despite my constant absence. **Vanda**, thank you for the great chats about travelling, climbing mountains and life. It was always nice to return to AMOLF after some time knowing the I could catch up with you in the office. To **Martijn Wehrens** I cannot thank you enough for your email responses and help at the beginning of my project – this saved me many a hurdle that you had already jumped. It is amazing that you took the time while writing up your thesis and expecting/preparing for a baby. I wish you and your family all the best! **Ferhat**, it was so comforting to have someone in a similarly odd situation to me experimenting in Delft and AMOLF. You were the best microscope companion I could have asked for thanks for all the cooperative

clean up / set up missions, data transfer and most of all a friendly, supportive colleague to email when there was panic that my images were suddenly illuminated incorrectly! **Marko**, your general knowledge of microscopes helped me in many times of need, thanks for always being willing to drop what you were doing to help me out!

To all the BN support staff I crossed paths with in the past 6 years thank you all for keeping both the department and science running smoothly! In particular **Anke**, thank you for all your hard work but also always stopping for a quick chat! **Sacha**, for always greeting me with a smile and making time for a catch up in the corridor. **Jeremie**, thank you for all your hard work setting up the microscope in Delft but most of all the fun (and frustration) we had together making PDMS chips and getting the fluid to stay in the microfluidics set up. It's a shame covid stopped our momentum but it was a refreshing moment at the end of my PhD to start working with you. **Roland** thanks for always being willing to help with tasks that required tools and equipment the Brouns lab didn't have and the laughs we had while solving these problems. And to **Chantal, Tahnee, Esther, Amanda, Nadine, Romy** and **Tracy** for your support in the organisation of social events in BN, without such events the atmosphere in the department would not have been the same and I would not have met all the people I acknowledge below.

Speaking of social events in the Bionanoscience department, it was through the monthly BN borrel that I met many people who started out as colleagues and ended up as great friends. A shout out to **Alberto** and that we actually enacted our crazy idea to form a social committee some years ago. Covid put a halt on this but we had great success, stress and fun organising a number of borrels enjoyed by many! On a more personal note **Alberto** and **Mireia** it's been great to meet two like-minded people, I only regret that we didn't get the chance to do more camping trips! I hope soon you'll be able to travel as you dream of and don't forget your always welcome in NZ with a free van and tour included ;)! Then a big thanks to the rest of the social committee **Nils, Cristian, Alessio**, the man of wicked birthday parties and great music and **Alejandro** the new enthusiast, I hope you can plan some events soon. Thanks also to past members **Misha**, you're a guy of many talents! It was always a pleasant moment when I'd see you around the halls of BN and **Eve** for joining the adventure and creating some fun that I always looked forward to on the first Friday of the month. Although not in the social committee **Sandro** I was always happy to see you for a chat in the corridor or at the Borrel. Thanks for our small scientific adventure to create some new strains, it's a shame they never survived in the microfluidic chip =|. **Wayne** you always had a good story to keep us entertained! **Mehran**, thanks for embracing the Brouns group at that first Kavli day and together with **Johannes** providing a solid stream of jokes throughout every evening. **Carsten**, thanks for all the in-depth chats at the occasional BN borrel they were always thought provoking (in a good way).



**Nils**, it's been fun to watch you turn into a father. Your passion for wind turbines and all things energy saving is amazing. I feel there is so much I can learn from you and I'm sad that we will be so far away. Thanks for joining us on some crazy bike rides, all the nice home cooked dinners, introductions to new board games, putting your organizational skills to work to arrange things with us and most of all just always being up to hang out! **Lola**, we seem to get on better these days, now you can hold up your own head. It's been crazy to watch you grow over the last year and I hope mum and dad will tell you all about me so you'll still recognise me in the future when we visit.

To the midweek drinkers we started with many more (so a shout out to the earlier crew as well) but ended with mostly **George, Anthony, Paola, Fede, Costi, Alberto, Mireia, Jochem, Sergii** and **Xin** thank you for the (mostly) Wednesday touch of sanity that sometimes got a little out of hand. Playing a large role in getting me through the lockdowns were the **Digital delicacies** team, this weekly online cooking event was a highlight – let's not forget to make that recipe book!!

**Fede**, you were my first friend in Delft during a period where I greatly needed some support! Thanks for the many evenings of cooking (although tiramisu is perhaps not our specialty), venting, ranting, laughing and repeating it all, all over again many times. **Costi**, it was so great to meet someone with a shared passion for pottery (although you are far, far better than I). Thanks for the amazing hospitality and tour of San Ginesio it was so great to experience the magical place where you grew up. **Anthony**, the man of staycations – I'm still very keen to try your optimized martini. Thank you for never giving up the organisation of mid week drinks despite times of low turnout! You are so unique and I greatly appreciate your humour which you give precisely at times when people need a laugh. Never stop laughing at your own jokes and all the best in your postdoc and future endeavours. **Paola**, thank you for your calm and joyful presence, for being keen to try out new spots for a beer/coffee/dinner and always having cool jackets to admire! I wish you and Anthony all the best.

Next up the girls of BN **Laura** your energy was always contagious, **Viktorija** the life of the party, **Mathia** kind, caring and with more dance moves than I've ever seen, **Lisa** the crazy sportswomen, phage microscopy companion and always full of surprises! **Helena**, you were a pivotal person in the first part of my PhD. I enjoyed every 'escape' to the big city, brunch, sushi night and G&T we managed to consume. Our friendship was instant and your openness and 'get out there and do things' attitude was inspiring. Watching you wrap up and get things done made me realise it would happen for me too and although it can be hard to keep in touch in such different phases of life I really enjoy our catch ups and hope to see you sometime soon! **Eve**, with your beautiful hair, contagious laugh and our shared love of camping, thanks for all the great convo. **Elena**, one of the most enthusiastic and realistic people I encountered on this journey. Keep keeping it real! **Nikki**, I

know we didn't cross over in BN for long but it's always interesting to catch up and hear where you've got to now. **Alicia**, you are such a joyous, colourful, happy, free and honest person. I always appreciate your take on a situation. And most of all it's impossible to not have fun when you're around. I joined mostly at the end of an era but thank you all for the much-needed cocktails, dinners and nights of dancing (and dancing on escalators) while you were all still around and later with those of you that remained!

Finally, it's your turn, you're only so close to the end because you're so important ;). **Nicole** and **Tanja**, my amazing paranymphs!! Thank you for saying yes to this role and most of all for helping me today and all the days that led to this day.

**Nicole**, it astounds even me how many hours we can talk for ;). I can really talk to you about anything and everything without judgement, with support and it always ends with a laugh and a positive outlook, even with the weirdest of stories. You have been a (very colourful) rock for me through the second half of my PhD and those moments where we had coffee or lunch at the Greek place were fundamental in me getting here. Your care and sense for how those around you are feeling is truly unique, such perception can be tough at times, but it is a great quality you should never lose! Most of all you are one of the funniest people I know with your entire repertoire of sounds, faces, mismatched socks, sparkly, shiny, patterned clothing! I'm sad to be heading across the world so far away from you but I know that the moment we get on the phone it'll feel as if we are in the same room.

**Tanja**, how funny that we met spontaneously as part of your interview! I'm so happy Jochem and I convinced you that BN would be a good choice ;p. I admire your ability to grow tomatoes, juggle Lola and a PhD, live such an organized life with weekly shopping (a great contrast to my own) and most of all your ability to think rationally when most others would fail. You make those around you feel good and have the ability to keep the atmosphere positive. For me you have been a strong pillar of reason and support and it was so nice to have you in my corner (literally) for a coffee or chat when needed. It has been so enjoyable to have you and Nils nearby in Delft to join us for activities! The weihnachts market in your garden (although my German wasn't so good), Christmas dinner, bike rides where we got soaked are all memories I hold close. I'm going to miss you, Nils and Lola! I wish you the best of luck with the end of your PhDs and hope I can still provide some digital support from far away. Big hug!

To the members of **Duivenrust** (I hope you didn't read all of the above searching for your names). First of all, thank you for choosing me and not placing a "sorry Dutch speaking only" advert on fb. I feel so lucky to have joined the house, and from that to have formed lasting friendships that have already and will continue to cross oceans! You guys were a real support to me during the challenging times when I moved to Delft and made the transition far easier. In addition, thank you for letting Jochem and his 'big feet' move in. **Luka**, thank you for all the beer-filled evenings (did we have evenings without beer?), the constant flow of general knowledge (I tried to



remember some) and for ensuring we occasionally had peach in the stir fry ;). **Anouk**, I was so happy when you arrived back from Spain after leaving me alone with Guus and Luka for so long ;P haha just kidding. But it did make a difference as you were the heart of the house and kept us all together - thanks for always taking the time to plan house events and dinners! I look forward to watching your future steps together and feel happy your move to Sydney has shown we can stay in touch while far away! **Guus** and **Enza**, first and foremost I want to thank you for being my go to “BBQ right now” people when the sun spontaneously shines! I appreciated all the games of pandemic that got us through the pandemic, it was nice to have something regular in such an irregular time. Our shared interest in travelling and weekendjes always keep me buzzing with new ideas for trips after seeing you. I truly hope you'll make it into New Zealand soon, I'm really looking forward to our future adventures, house + van swaps and catch ups over there! Fingers reeeally crossed!!

To **Marijn** and **Lilian** what a shame we only got to live close by for such a short time, regardless it was nice to have you in the neighbourhood! Thank you for the getaway to Rijswijk and the good food, wine and conversations we have round the dinner table. All the best on your new venture with Elias and I hope we keep in touch.

To **Kelly** and **Joel** how lucky were we to have such great neighbours! It was such a surprise to find we had neighbours with so many shared interests. Thank you for the drinks, dinners, BBQs and privacy despite the fact we stare directly into each other's kitchens! All the best in your new home and fingers crossed for great neighbours!

**Lotna** and **Steeff**, Thank you for the yearly adventures more often than not filled with surprises organised by the two of you! **Ceci**, I hope to do some mountain trekking with you soon! **Anna**, it was great to get to know you a little better when you ended up in our house. You're an extremely loyal person, curious about the world and why people behave how they do. Along these lines I enjoyed all our open conversations and evenings together. It's great to see you setting up your life in Belgium with **Bart**, a man of good champagne and travel stories I always enjoy. I feel at peace to know that we can play terraforming mars digitally and still catch up even when there's distance between us.

To all the **Kiwis** who brought me a slice of home over the last 6.5 years thank you!! Amazingly, I never ran out of watties or whittakers! Your visits meant a lot and kept me connected while being so far away <3

To **Will**, **Bart**, **Corrie**, **Marnix** and **Nicole**, I want to thank you for without hesitation despite all possible language or cultural barriers instantly making me feel part of the family!! My experience in The Netherlands would not have been the same without having the feeling of home and family here as well. **Will**, I will always remember your encouragement to sign a PhD contract and

stay in The Netherlands and most of all the enthusiastic response when I announced it was happening. Although our time together was short and sweet, I am so happy I had the opportunity to meet you, and hold those moments dear to my heart. **Bart**, there is so much to learn from you! I enjoy your history excerpts, hearing the stories from different times in your life and your talent for seeing things from many different perspectives. On another note your trip planning skills are something to admire and I feel very lucky that Jochem has inherited this from you ;). **Corrie**, it's wonderful to get to know you, to hear your music in the mornings and the stories of your childhood on the farm in the evenings (and the many other topics of conversation we cover) and of course to see Utrecht from the perspective of a true local. Thank you to you both for the warm hospitality and instant feeling of home you have provided Jochem and I for these last months in The Netherlands. I feel a little guilty to transport Jochem away almost as far as possible, but it's a great comfort to have your support and we look forward to not only when you visit us in New Zealand, but also when we can visit you here and feel at home in Utrecht again. **Marnix** and **Nicole**, I couldn't have been luckier in the family lottery! We have so much to talk about and share, your trips and plans always inspire new ideas for us and I always enjoy visiting you; your fridge full of a wide range of beers and your cupboards full of board games. **Marnix**, thank you for always taking care of Jochem (eg unplanned rides to the train) and therefore by default me as well, and also sharing many laughs about that ;). **Nicole**, thanks for always making the effort to check in with me at family events and the chatting and laughter fills the days/evenings as a result.

Finally, almost last but definitely not least. My amazing family **Callum**, **Catherine**, **Mum** and **Dad** thank you for your endless support no matter what I do, I can only imagine how hard it is to encourage and raise someone to do whatever they want only to end up having them halfway across the world. Even when it was tough you encouraged me, smiled and listened to my stories – it's from you that I have learnt to give the same unconditional support to others. **Callum**, I'm not sure we could be more different yet still somehow have many things in common. That is what I love about you, your financial and organised mind, your future planning and strength in knowing what you want. I'm so glad you have found **Catherine** you two are so in tune and I'm happy to see you travelling the world and building your life together. It's such a shame we didn't have more opportunities to be together on this side of the world. The trip to Belgium we just squeezed in will always be strong in my memories as a great time and the last trip before covid began. Looking forward to more adventures and time together in NZ =). **Mum**, I have dearly missed just hanging out with you shopping, drinking coffee going to the lighthouse cinema I can't wait for all that to begin again! You are my number one fan happy when I'm happy and sad when I'm sad, and through this I always knew whatever decision I made you would whole-heartedly support it. You always put others ahead of yourself a trait I admire and try to take on. Most importantly I have to thank you for teaching me to be the social person that I am today, the ability to connect, network and collaborate are

the skills I believe really enabled me to complete my PhD. Without your ~ 30 years of commitment to guiding me to be the best I can I may still have been the shy girl that cried in the mall when she saw her friends. **Dad** all those drives to Dunedin each February in which I so much enjoyed just being with you – and the ‘subtle’ brainwashing that Otago was the place to study – who would’ve thought those things would start a passion for microbiology that would end in me producing this book. You have been a stable pillar in my life who always gives the right advice, and helps me to prioritise what to spend my time and energy caring about. You have also been a role model who’s career I have followed with interest and has sometimes led me to question if I should be working in conservation instead. Thanks for taking us all to as many waterfalls as was humanly possible! I may have complained about it as a kid, but after so long in a country with no waterfalls I’ve realised just how special the nature in New Zealand is and how lucky I am to have spent my childhood seeing it. I’m looking forward to family holidays, days in Martinborough, weekend walks and coffees and really being in Wellington to spend time with you all again.

**Jochem**, here we are a product of CRISPR hahaha but I’m so happy it brought us together! It’s hard to put everything I feel into so few words but I’m quite certain without your daily support both scientifically and especially personally this PhD journey would have been much bumpier. I am constantly learning both from you and about you, your ability to transport yourself outside of a situation and look at it from an objective perspective is something I truly admire. This perspective helped me through the many times that my inclination was to react emotionally instead of rationally. You push me to be better but comfort me when it’s hard, you surprise me yet also always provide stability and a sense of being home. When I moved from NZ to The Netherlands for my big O.E. I never expected to end up building a life and home like we had in Delft. The evenings on the roof top terrace, the constant buzz of entertaining visitors, BBQ-ing to quell my homesickness and the mutual unspoken agreement that our home was always open to anyone. I know we are both excited for what’s next and I couldn’t have hoped for a better person to do it all with! <3

Woah that was long, if you actually read it all congrats! As you can see this book was not produced alone and to all those who crossed my path perhaps only momentarily and are not mentioned here I thank you as well! More than anything else that I’ve gained from my PhD, it is the friendships I’ve made and sense of community and support that I feel most grateful for.





ISBN 978-90-8593-493-6